STUDY OF G-QUADRUPLEX FORMATION IN THE PROMOTER REGION OF ZINC TRANSPORTER hZIP1

Daniela Franco Bodek

Imperial College London

Department of Chemistry

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor in Philosophy

To Alexandro

DECLARATIONS

DECLARATION OF ORIGINALITY

I, Daniela Franco Bodek, declare that the work submitted in this Thesis is my own. All contributions and sources are duly acknowledged and appropriately referenced.

DECLARATION OF COPYRIGHT

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

ACKNOWLEDGEMENTS

First and foremost, I would like to deeply thank my supervisors Ramón Vilar Compte and Rüdiger Woscholski for their invaluable guidance and wisdom, their trust and support, patience and motivation, and the great opportunity they gave me of carrying out this project.

I would like to thank CONACyT for the sponsorship that made this work possible.

My sincere gratitude goes to the postdocs from the Vilar and Woscholski groups who also mentored me during these years: Oscar Mendoza Pomar, Beata Klejevskaja, Arun Shivalingam, Alexandra Alves Duarte, and Lok Mak.

I would like to acknowledge the generosity of Peter Makhov and Marco di Antonio for sharing with me key materials and knowledge.

I would also like to thank all group members, both past and present, who made the countless hours spent in the laboratory and office a joyful experience. Additionally, I want to thank the Ramirez family who have made me feel part of their family.

I would like to thank my parents, Pepe and Claudia, and my brother Tomás for their love, trust and support.

Finally, I would like to thank Alexandro for everything.

ABSTRACT

The formation of G-quadruplexes, non-canonical DNA structures, in the promoter region of genes has attracted a lot of attention in recent years as regulators of gene expression; these structures have become attractive targets for therapeutic design of molecules. In this thesis, the effect of a putative G-quadruplex sequence was investigated in the promoter region of *hZIP1*, an important zinc transporter that is down-regulated during the development of prostate cancer. Understanding the mechanisms of regulation of this gene is fundamental to unravel the processes that govern the malignancy progress. It was found that the G-rich sequence (named hZIP1-1) located near the transcription start site of *hZIP1* is capable of forming stable parallel intramolecular G-quadruplexes at physiological pH and potassium concentration *in vitro*. Mutations that disrupt the G-quadruplex potential were designed with emphasis on preventing the disruption of known and predicted transcription binding sites. The well-studied G-quadruplex ligands Pyridostatin, BRACO-19 and PhenDC3 were found to stabilize the G-quadruplex sequence while having little effect on the non-G-quadruplex mutant. *In cellulo* studies revealed that disrupting the G-quadruplex potential of the *hZIP1* promoter decreases transcription. Furthermore, treatment with Pyridostatin increases transcription. These results suggest that the G-quadruplex may function as an activator of gene expression that can be targeted using G-quadruplex selective small molecules, providing a new therapeutic route to be explored in the treatment of prostate cancer.

TABLE OF CONTENT

Declarations *3*

Declaration of Originality 3

Declaration of Copyright 3

Acknowledgements *4*

Abstract *5*

Table of Content *6*

Abbreviations *9*

Chapter 1: Introduction 12

1.1 G-quadruplexes *12*

1.1.1 DNA structure: G-quadruplex 12

1.1.2 Thermodynamic and kinetic properties 18

1.1.3 G-quadruplexes *in vivo* 20

1.1.4 Regulatory roles of G-quadruplexes 24

1.2 ZIP1 regulation *31*

1.2.1 Zinc, ZIP1 and prostate cancer 31

1.2.2 Regulation of *hZIP1* 34

1.3 Objectives *39*

1.3.1 General aims 39

1.3.2 Specific objectives 39

Chapter 2: Characterization of hZIP1 G-quadruplex 41

2.1 Bioinformatics analysis *41*

2.1.1 G4-DNA sequence in the promoter region of *hZIP1* 41

2.1.2 Mutant design for the study of hZIP1-1 44

2.1.3 Concluding remarks 48

2.2 Biophysical and biochemical characterization *49*

2.2.1 Annealing conditions studied by electrophoretic mobility shift assay and circular dichroism spectroscopy 49

2.2.2 Thermal difference spectra and polymerase stop assay 56

2.2.3 Concluding remarks 60

2.3 Thermodynamic and kinetic characteristics *61*

2.3.1 Melting and annealing curves 62
2.3.2 Double strand competition 73
2.3.3 Concluding remarks 78

2.4 Summary, discussion and conclusions 80

Chapter 3: G-Quadruplex stabilizing ligands and their effect on hZIP1 84

3.1 Ligand stabilisation 84
3.1.1 Small molecules and G-quadruplexes 84
3.1.2 FRET melting assay 89
3.1.3 Concluding remarks 93

3.2: Discussion and conclusions 94

Chapter 4: The effect of G4 in the promoter region of hZIP1 96

4.1 In vitro transcription 96
4.1.1 Plasmid construction 96
4.1.2 *In vitro* transcription of *hZIP1* promoter 97

4.2 Luciferase reporter assay 100
4.2.1 *In vivo* transcription of *hZIP1* promoter 100
4.2.2 *In vivo* transcription and the effect of G4 ligands 103

4.3: Discussion and conclusions 105

Chapter 5: Discussion and conclusions 107

5.1 General discussion and conclusions 107

5.2 Future work 111

Experimental procedures 112

Promoter analysis and mutant design 112
Buffer and reagent preparation 113
Gel Preparation 114
DNA preparation 115
Electrophoretic Mobility Shift Assay (EMSA) 116
Circular Dichroism Spectroscopy (CD) 116
Thermal Difference Spectra (TDS) 117
Polymerase Stop Assay (PSA) 117
Melting and Annealing 118
Kinetic Stability Assay 118

Ligands for G-quadruplex stabilization 119
Ligand stabilization using FRET Melting 119
Plasmid Construction 119
In vitro transcription, RNA purification and Quantification 121
Cell culture conditions 122
Luciferase Reporter Assay 122

References 124

Appendix 142

A.1 Transcription Factor Binding prediction 142
A.1.1 Wild Type Promoter 142
A.1.2 MutB Promoter 147
A.1.3 MutE Promoter 151
A.1.4 MutF promoter 155
A.1.5 MutG Promoter 159
A.1.6 MutH Promoter 163
A.2 Electrophoretic Gels 168
A.3 Circular Dichroism 172
A.4 Thermal Difference Spectra 175
A.5 Melting and annealing curves 178
A.6 FRET melting 182
A.7 In vitro Transcription 188
A.8 Luciferase Reporter Assay 190

ABBREVIATIONS

A	Adenine
AP-2α	Activating enhancer binding Protein 2 alpha
AR	Androgen Receptor
Bcl-2	B-cell CLL/lymphoma2
BLAST	Basic Local Alignment Search Tool
BLM	Bloom Syndrome RecQ-like helicase
bp	Base pairs
BRACO-19	N,N'-(9-(4-(Dimethylamino)phenylamino)acridine-3,6-diyl)bis(3-(pyrrolidin-1-yl)propanamide)
C	Cytosine
CD	Circular dichroism
CREB	cAMP responsive element binding protein
C_T	Cycle threshold
DIGP	tetrakis-(diisopropyl-guanidine)phthalocyanide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAseI	Deoxyribonuclease I
dNTP	Deoxynucleoside triphosphate
dsDNA	Double stranded DNA
E. coli	*Escherichia coli*
EBI	European Bioinformatics Institute
EDTA	Ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
EMSA	Electrophoretic mobility shift assay
EPDNew	Eukaryotic Promoter Database New
FRET	Fluorescent Resonance Energy Transfer
G	Guanine
G4-DNA	G-quadruplex DNA
GFP	Green fluorescent protein
H4TF1	Histone four transcription factor 1
HIF-1α	Hypoxia inducible factor 1α

HIV-1	Human Immunodeficiency virus type 1
hTERT	Human telomerase reverse transcriptase
Irt	Iron related transporter
K_a	Equilibrium constant
LRA	Luciferase reporter assay
MAZ	Myc-associated zinc finger protein
min	Minutes
miRNA	MicroRNAs
mM	Millimolar
MRE	Metal response element
mRNA	Messenger RNA
MTF1	MRE-binding transcription factor 1
N	Nucleotide
NF-κB	Nuclear factor kappa B
NMM	N-methyl mesoporphyrin IX
PARP-1	Poly(ADP-ribose) polymerase 1
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDS	Pyridostatin
Pif1	Pif1 5'-to-3' DNA helicase
Poly(G)	Polyguanylic acid
Poly(I)	Polyinosinic acid
POT	Protection of telomeres
PQS	Putative quadruplex sequence
PSA	Polymerase stop assay
qPCR	Real time polymerase chain reaction
RAP1	Repressor-activator protein 1
RNA	Ribonucleic acid
RNAPII	RNA Polymerase II
rNTP	Ribonucleoside triphosphate
RREB-1	Ras responsive element binding protein 1
RT-qPCR	Reverse transcription real time polymerase chain reaction
sec	Seconds

SLC39A1	Solute carrier family 39 member 1
SP1	Specificity protein 1
ssDNA	Single stranded DNA
T	Thymine
$T_{1/2}$	Temperature value at θ=0.5
TBE	Tris base, boric acid and EDTA buffer
TDS	Thermal difference spectra
TE	Tris-EDTA buffer solution
TF	Transcription factor
TGE	Tris-Glicine- EDTA buffer
T_m	Temperature of midtransition
TMPyP2	5,10,15,20-tetra-(N-methyl-2-pyridyl)porphine
TMPyP4	5,10,15,20-tetra-(N-methyl-4-pyridyl)porphine
Tris	Tris(hydroxymethyl)aminomethane
TSS	Transcription start site
VEGF	Vascular endothelial growth factor
ZIP	Zrt-/ Irt-like protein
Zrt	Zinc related transporter
Θ	Folded fraction

CHAPTER 1: Introduction

G-quadruplex DNA is a non-canonical DNA structure that can be formed in guanine rich sequences in the presence of potassium and other cations. In this chapter, the structural and thermodynamic properties of DNA G-quadruplexes are described and their possible formation *in vivo* is discussed with emphasis in their potential role in gene regulation. Later, the link between zinc transporter ZIP1 and prostate cancer is discussed, the current knowledge of the regulation of *hZIP1* is summarized and the possibility of a G-quadruplex in the promoter region is presented for the first time. This leads us to the key question addressed in this thesis: is the formation of a G-quadruplex involved in the regulation of *hZIP1*?

1.1 G-QUADRUPLEXES

1.1.1 DNA STRUCTURE: G-QUADRUPLEX

DNA is commonly found as a right-handed double helix described first in 1953[1,2] in which both strands are complementary to each other given a strict pairing of adenine with thymine and guanine with cytosine. This pairing is usually referred to as Watson-Crick[3] base pairing (see Figure 1.1.1).

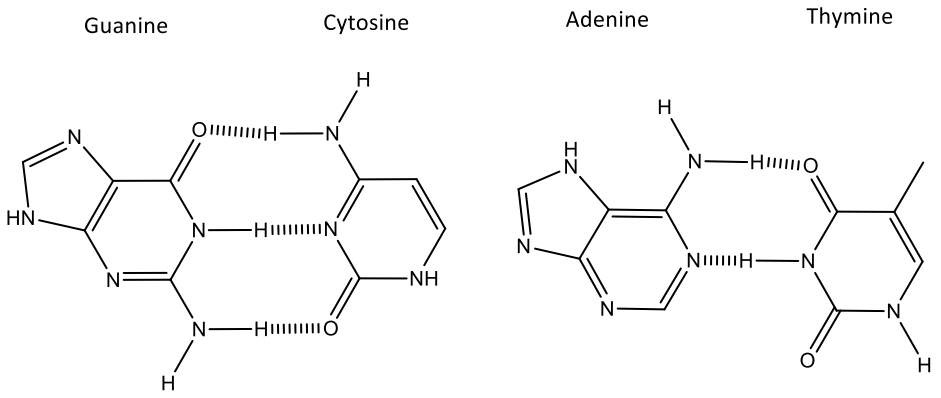


Figure 1.1.1 Nucleotide bases forming the classic Watson-Crick base paring.

Although the right-handed double helix, known as B-DNA, is the most commonly represented structure, DNA can adopt other structures (see Figure 1.1.2). A-DNA is also a right-handed double helix with a more compact structure, and Z-DNA is a left-handed double helix. Besides two stranded structures, three and four stranded structures can also be adopted by DNA like the i-motif, which relies in quadruplex formation of cytosine rich strands. [4]

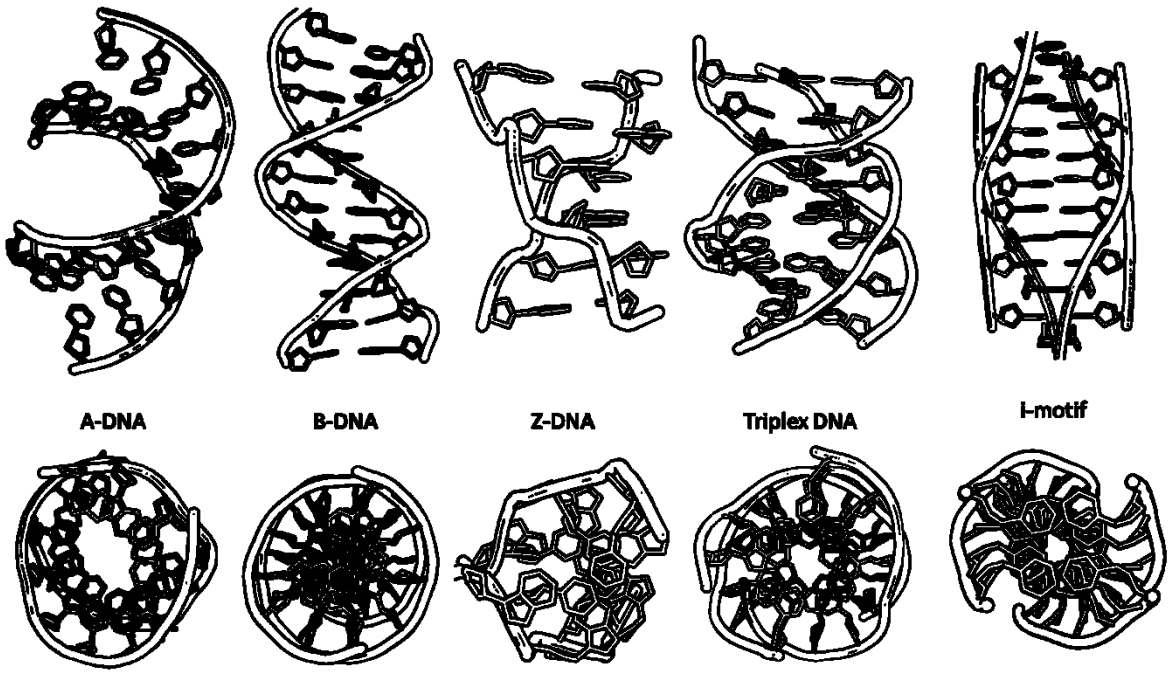


Figure 1.1.2 Some examples of different DNA structures viewed from side and from top. From left to right: A-form DNA (PDB 440D),[5] B-form DNA (PDB 1BNA),[6] Z-form DNA (PDB 3P4J),[7] triplex DNA (PDB 1D3X),[8] i-motif DNA (PDB 225D).[9]

In 1962, Gellert *et al*[10] built up on the observation made by Bang in 1910 that guanylic acid formed a gel when in concentrated solution, and using X-ray diffraction they proposed that the bases of the guanidylic acid were arranged as planar tetramers that could aggregate by stacking upon one another. Rich[11] had studied the structure of polyinosinic acid (Poly(I)) in 1958 using X-ray diffraction and concluded that the polymer probably formed a helical structure made of three strands. Thiele and Guschlbauer[12] analysed again the data in 1973, and paired it with their own studies of Poly(I) using circular dichroism spectroscopy and titration experiments, concluded that a four stranded structure was more likely. Figure 1.1.3 shows the molecular structure of guanidylic and inosinic acids.

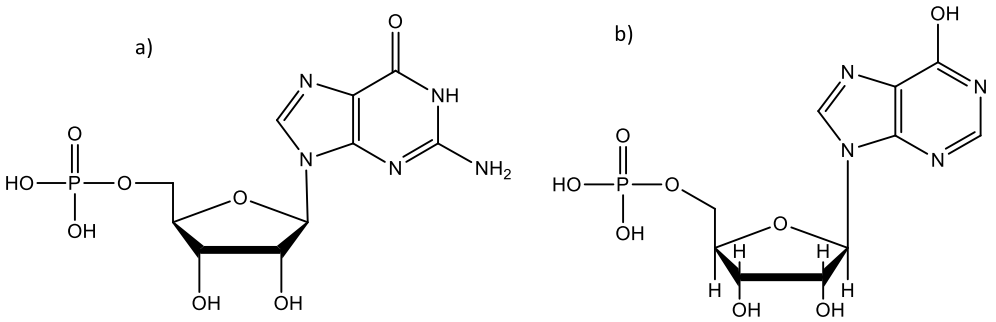


Figure 1.1.3 a) Guanidylic acid or guanosine monophosphate; b) Inosinic acid or inosine monophosphate.

In 1974, Arnott *et al*[13] published a study on the structure of Poly(I) and polyguanylic acid (Poly(G)) using X-ray diffraction analysis and concluded that, indeed, the polynucleotide chains formed four stranded structures, rejecting the three stranded structure proposed by Rich[11] in 1958. They noted that for Poly(G) this molecular model was made of the tetrads Gellert *et al*[10] had proposed before.

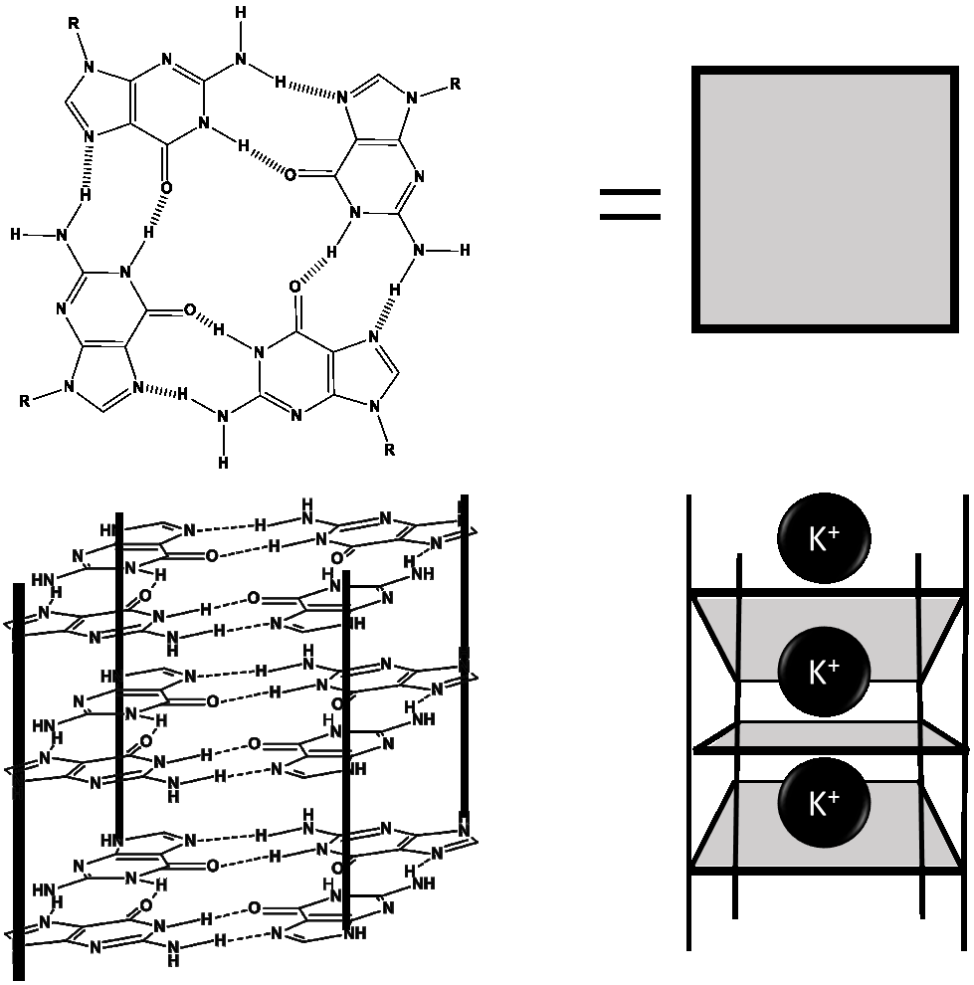


Figure 1.1.4 Top, a G-tetrad showing Hoogsteen hydrogen bonding. Bottom left, three G-tetrads stacked to form a G-quadruplex. Bottom right, location of stabilizing potassium cations between the tetrads.

Henderson *et al*[14] in 1987 recognised that G-rich oligonucleotides from the section of the end of the chromosomes, known as the telomere, formed an intramolecular structure stabilized by non-Watson-Crick base pairs but rather by G:G pairs and identified them as novel hairpins. In 1988, Sen and Gilbert[15] reported that single stranded DNA with G-rich sections would associate into four strands at physiological salt concentrations, coined the term "G4-DNA" and began to speculate on its biological implications including in the telomere region. They noted that the guanine hydrogen bonding would be best described in terms of Hoogsteen[16] hydrogen bonding given the involvement of N7 in the guanine, which is not involved in Watson-Crick pairing (see Figure 1.1.4). Sundquist and Klug[17]

reported in 1989 that two G-rich oligonucleotides forming hairpins could in turn associate and form a dimer by salt dependent tetrad formation. Shortly afterwards Panyutin *et al*[18] argued the formation of intramolecular tetrads in G-rich sequences and proposed their existence in telomeres. The structures stabilized by stacking tetrads are known as G-quadruplexes. Throughout this thesis I will refer to these structures both as G-quadruplexes and G4-DNA.

It was also noted by the authors of these many studies that the structural properties of the G-quadruplexes were heavily dependent on the ionic strength with a strong emphasis on the cation used. It was noted that the channel created by the tetrad formation seemed appropriate to accommodate water molecules, and the oxygen atom could coordinate cations;[13] this has been confirmed by solving crystal structures which showed the potassium sits inside of the channel and in between the tetrads as shown in Figure 1.1.4, while sodium being smaller can sit in the same plane as the tetrads or in between them. The alkali metals stabilize G4-DNA in the following order $K^+ > Na^+ > Rb^+ > NH_4^+ > Cs^+ > Li^+$.[19]

The folding of the structure depends on the stabilizing metal cation, strand concentration and the strand sequence itself. For G-quadruplexes to be formed, it is necessary for four guanine repeats to associate into tetrads, with a minimal unit of $N_xG_nN_x$ where N is any nucleotide repeated X-times, and G is a guanine run (G-run) of n number. Whether the G-runs are found in the same strand or different strands will affect molecularity.[19] G-quadruplexes can form as: a) intermolecular structures with four independent strands as a tetramer, also known as tetramolecular G-quadruplexes; b) with two strands as a dimer, also known as bimolecular G-quadruplexes; or c) as an intramolecular with a single strand also referred to as unimolecular G-quadruplex. Figure 1.1.5 exemplifies these different forms.

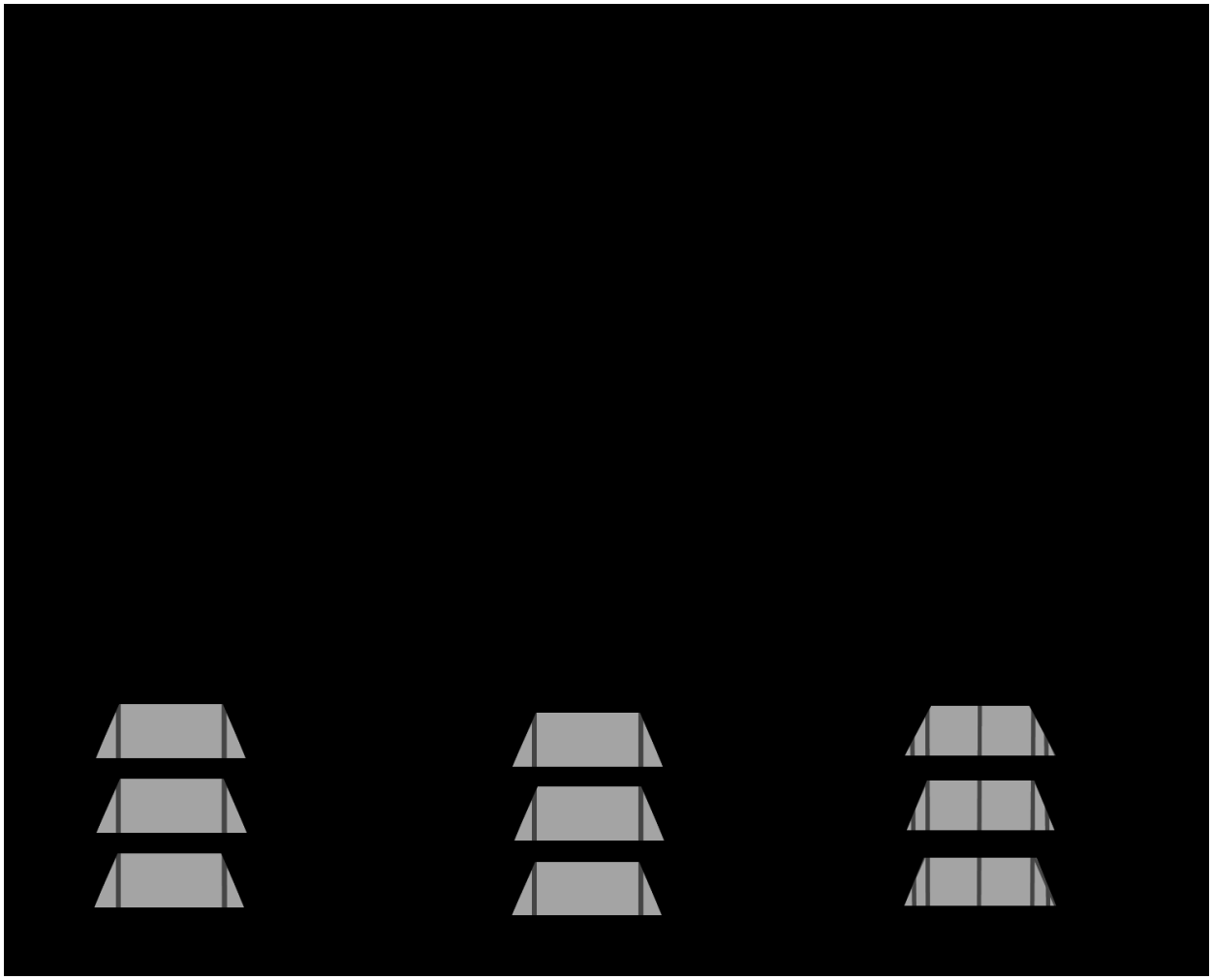


Figure 1.1.5, Representative structures of tetramolecular, bimolecular and unimolecular G-quadruplexes. From top to bottom: side view, top view, and schematic representation of the structure. From left to right: tetramolecular G-quadruplex (PDB 352D),[20] bimolecular G-quadruplex (PDB 1D59),[21] and unimolecular G-quadruplex (PDB 1KF1).[22] Cations have not been included for clarity.

The combination of molecularity, orientation of the strands, loop size and location, and number of stacked tetrads results in a wide variety of possible conformations. When the four strands of a G-quadruplex structure are facing the same direction it is referred to as parallel, if at least one of the strands runs in the opposite direction it is called antiparallel. The loops connecting the different tetrads and strands can be of three types: propeller, when they connect the bottom and the top G-tetrads; lateral, if they join together adjacent strands; or diagonal if they connect opposite strands. Parallel unimolecular structures feature only propeller loops, and therefore sometimes in the literature they are referred to as propeller structures, while unimolecular antiparallel structures can feature a mix of propeller, diagonal and lateral loops. Unimolecular antiparallel structures are often named according to their loops and connectivity (see Figure 1.1.6): if they have two strands running on each direction they might be a basket or chair conformation, while if they have three strands

running on the same direction they are known as hybrids or 3+1. Propeller loops may also be found in the literature as “strand-reversal” loops and lateral loops may be found as “edgewise” loops. [23]

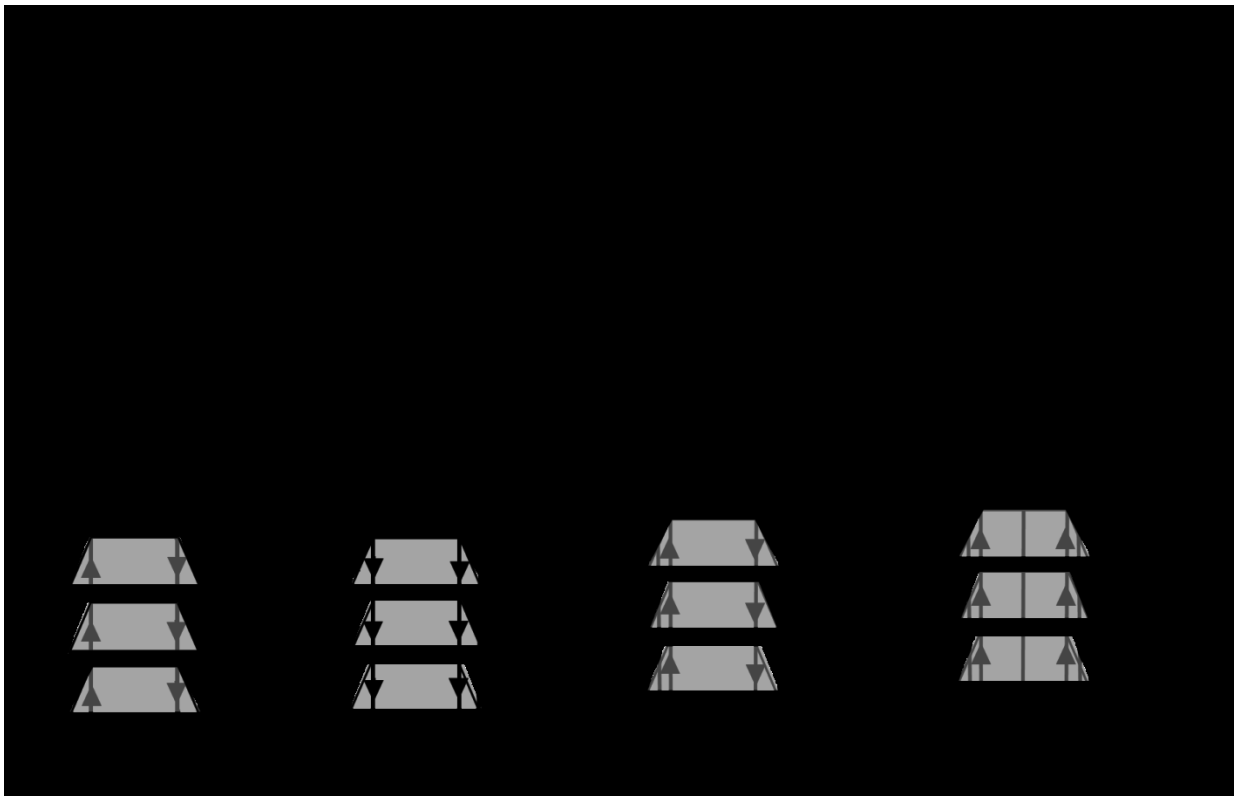


Figure 1.1.6 Representative examples of different intramolecular G-quadruplex topologies. From top to bottom: side view, top view and schematic representation of the different conformations. From left to right, antiparallel chair conformation featuring only lateral loops (PDB 1C35);[24] antiparallel basket conformation featuring lateral and diagonal loops (PDB 2KOW);[25] antiparallel 3+1 or hybrid conformation has three strands running in the same direction (PDB 2KZD);[26] parallel structure featuring only propeller loops (PDB 2KZE).[26] In the schematic representation propeller loops do not contribute to the G-quadruplex formation. Cations have not been included for clarity.

The glycosidic angles in parallel structures are all in *anti* conformations. On the other hand, antiparallel structures have a mixture of *anti* and *syn* conformations. Whether the structures have only *anti* or mixed *anti* and *syn* conformations will give rise to different spectroscopic characteristics. Figure 1.1.7 shows *syn* and *anti* conformations.[27]

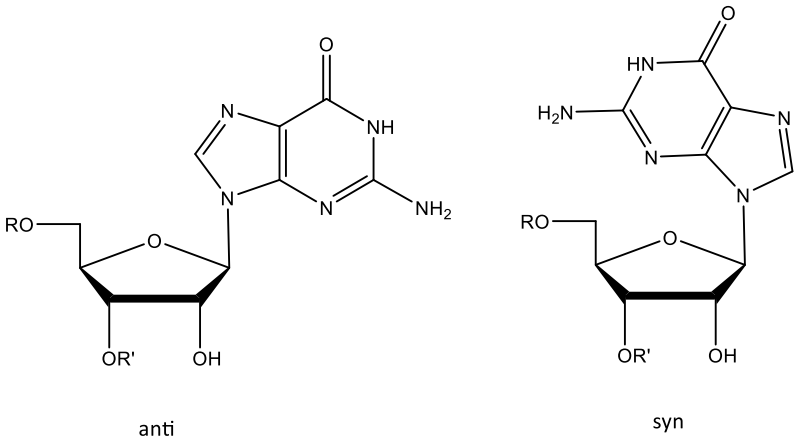


Figure 1.1.7 Glycosidic bond conformations *anti* (left) and *syn* (right) in guanosine.

The particular folding of a G-quadruplex will depend on the sequence itself and the cation used for stabilization. It is not unusual for a sequence to be able to adopt more than one structure. Indeed, some sequences can change structure from parallel to antiparallel depending on the metal used for stabilization.[28] What is more, it has been shown that divalent cations destabilize antiparallel structures and promote parallel structures even after the first has already been formed.[29] The polymorphism displayed by some sequences may not only be influenced by the cation; crowding conditions also seem to favour one structure over the other.[30] In a given set of conditions, more than one folding may be present simultaneously for the same sequence. Between different sequences, it is apparent that loop length, number of G-runs in a strand, and nucleotide overhang play a role in the adopted topology as well as the stability of the structure.[31]

While this work will only focus on DNA G-quadruplex structures, it is worth noting that RNA can also form quadruplexes. RNA has gained a lot of attention due to its single stranded nature and potential to form G-quadruplexes, with increasing evidence of their importance *in vivo.*[32] DNA G-quadruplex structures are also known to be aptamers with peroxidase-like activity when complexed with hemin[33], and this has been exploited for the design of a wide range of probes.[34]

1.1.2 THERMODYNAMIC AND KINETIC PROPERTIES

G-quadruplexes thermodynamic stability has been extensively studied. While G-quadruplexes have been found to be very stable, they are not necessarily more stable than their duplex counterparts. It has been argued that under identical conditions, G-quadruplexes are not more stable than DNA duplexes of a comparable size and that their unfolding rates are also not slower.[35]

Traditionally, G4-DNA sequences have been studied as short oligonucleotide chains of high purity. To study the stability of a structure typically a melting experiment is used: a characteristic spectroscopic

peak is monitored while increasing the temperature until the structure is melted. The temperature of midtransition from folded to unfolded (melting temperature, T_m) can be used to compare the stability of a given structure against others, and the melting curve can be used to extract important thermodynamic parameters like entropy (ΔS) , enthalpy (ΔH) and free energy (ΔG).[36] However, studies have found that the design of the oligonucleotide itself will affect the stability of the structure, making comparisons less obvious. Overhanging sequences in either end will typically decrease the stability and influence the folding by reducing the degrees of freedom.[37] Therefore studying the same G4-DNA core sequence but changing the length of the flanking nucleotides can yield different results from folding populations to the melting point. An additional complication for studying and comparing G-quadruplexes is that the addition of the labelling molecules may affect its stability.[38] The length and composition of the loops also change the stability of the studied G-quadruplex, meaning that related structures may have different melting temperatures by switching bases that are not involved in the tetrad formation.[39] The concentration and nature of the stabilizing ion will also have an important effect on the melting temperature.[31]

Intermolecular G-quadruplexes require the association of different strands which usually requires a long association time compared to the quick association of intramolecular G-quadruplexes.[31] Given that in biological processes the folding is more likely to be kinetically controlled due to the short time scale of the processes, the fast association of these structures is relevant.[40] Given the polymorphic nature of intramolecular strands, the folding and unfolding pathway is likely to include intermediaries which complicate the study of the kinetic parameters[41] (see Figure 1.1.8). To study the kinetic behaviour of intramolecular G-quadruplex structures a common technique is to monitor the G-quadruplex unfolding in the presence of its reverse complement strand such that the duplex formation traps the unfolded G-quadruplex structure. The modification of a sequence to facilitate the study may itself have an important impact in the folding and unfolding process[42], an overhang can also dramatically change the kinetic stability of the quadruplex, accelerating the opening of the structure by a reverse complement competitor.[43] Some parameters known to affect the G-quadruplex kinetic properties include: temperature, ionic strength, cation, G-run length and base overhang.[31]

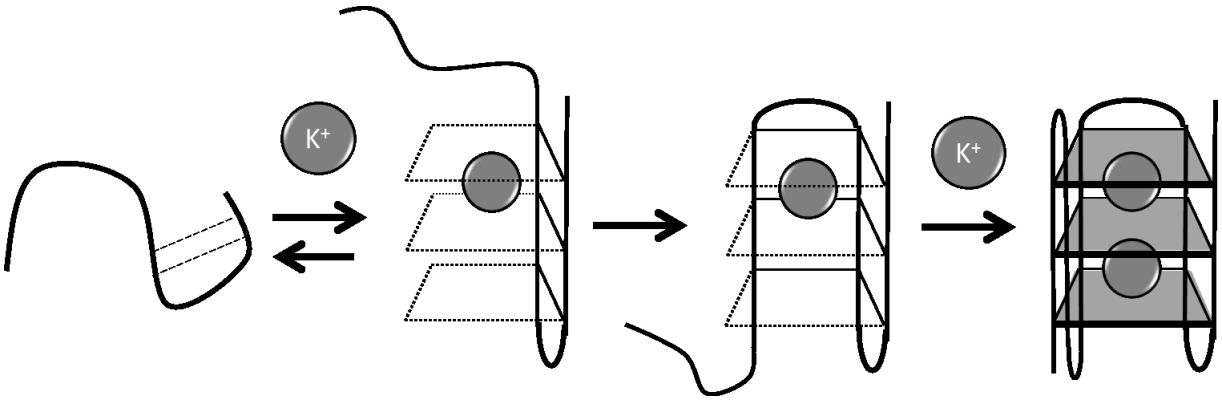


Figure 1.1.8 Folding pathway of an intramolecular G-quadruplex proposed by Zhang and Balasubramanian[41] featuring a hairpin and triplex intermediaries for G-quadruplex formation. The dashed line in the first species signifies that there is a secondary structure present. The dotted lines in the second and third species represent the position in which the G-tetrad will be formed, as shown in the fourth and final species. Figure adapted from the literature. [41]

1.1.3 G-QUADRUPLEXES *IN VIVO*

In order to determine the existence and role of G-quadruplexes *in vivo,* researchers have tried to address a number of different questions. The following discussion summarizes important issues that have been argued in favour of the existence and importance of these four stranded structures.

Are there sequences in the genome that can form G-quadruplexes and if so, where are these sequences located and how are they distributed in the genome?

To address this question, researchers[44,45] examined the human genome using bioinformatics techniques looking for sequences with the motif "$G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$" (where N can be any nucleotide including guanine) predicted to form G-quadruplexes *in vitro*. They found a very large amount of sequences that satisfy this criterion, and concluded that 376,000 could potentially exist simultaneously (but did not imply they do). Another study[46] found that promoter regions were enriched with these potential G-quadruplex sequences. While there has been an interesting association of possible quadruplex sequences in cancer related genes[46], association with other topics like cardiac related genes have also been reported.[47] Not all reports have been anthropocentric; a genome search in maize revealed a very large amount of putative G-quadruplex sequences[48], while the *E. coli* genome has shown a striking correlation between putative G-quadruplex sequences and gene regulation regions.[49] New bioinformatics studies have focused in searching for more complicated structures, like duplex stem-loop in quadruplexes, and found prevalence of sequences with the potential to form such structures in regulatory regions.[50]

Recently, a genome-wide experimental search for G-quadruplex structures was published by Chambers *et al*[51] which showed a surprisingly large number of these structures, more than 700,000, many of which were not predicted. The method, G4-Seq, combines high-throughput sequencing and polymerase stop assay while promoting G-quadruplex formation either with potassium or with Pyridostatin (a G4-DNA stabilizing ligand, see Chapter 3 section 3.1.1). They found that 60% of the computationally predicted G-quadruplexes were indeed found in the potassium conditions, and 70% of the observed G-quadruplexes were not predicted using a classical G4-DNA description because they include bulged structures and longer loops.

Several bioinformatics online tools have been developed to analyse the location of putative G-quadruplex regions like QGRS-H Predictor,[52] QGRS Mapper[53] and QuadBase,[54] while Quadparser[44] and G4 calculator[55] are available to download.

Are G-quadruplexes thermodynamically and kinetically plausible under physiological conditions, including cation concentration, pH, temperature, and molecular crowding?

It is well established that G-quadruplexes form *in vitro* at physiological concentrations of sodium and potassium cations, they are stable at neutral pH and room temperature, which was deemed very early on as an indicator of the possibility of these structures forming *in vivo.*[15] If a G-quadruplex is indeed formed, for example, in one of the strands of a promoter region, it would be a competitor to the formation of the canonical double helix. Studies of G-quadruplex/duplex equilibrium have shown that at physiological temperature, these two structures can co-exist and in some cases the G-quadruplex form will predominate.[56] Nevertheless, Lane[35] argues that when taking common conditions into account, G-quadruplexes reported in the literature are of comparable or inferior thermodynamic stability than duplex DNA, and they are also less stable kinetically than DNA duplexes of the same size. Indeed, G-quadruplexes *in vitro* can be opened by their reverse complementary strands and be trapped as duplexes for many sequences.[56,43] However, the double helix may open during transcription or on its own due to negative supercoiling, freeing the G-rich strand.[57] Supercoiling DNA increases torsional stress, and it has been proposed that non-canonical structures, such as G-quadruplexes, can alleviate this stress and have been shown to form under its influence.[58] Finally, it should be noted that the cellular environment is crowded by a wide variety of different molecules and this affects steric conformations and water activity which in turn can affect molecular folding of biomolecules. Molecular dynamic simulations have shown that in the absence of cations, G-quadruplexes can be stabilized by ethanol.[59] Research has shown that under crowding conditions G-quadruplex structure can be stabilized while DNA duplex can be destabilized.[60,61] It has also been shown that the structure obtained under kinetically favouring conditions may differ from the structure found in equilibrium,

and it was argued that the G-quadruplex structure formed in physiological conditions will not necessarily match the most thermodynamically stable structure.[61] Taken together, these studies show that G-quadruplex structures are viable under physiological conditions but it is not clear if they are stable enough to compete with the double helix.

What proteins, if any, can promote the formation, bind to and unwind these structures?

Genome instability refers to a process inclined to increase genomic alterations, so when the DNA is replicated changes may occur including mutations, deletions or amplifications, or even rearrangement of the chromosome.[62] Due to the high stability of G-quadruplexes and their spontaneous formation in the right conditions, and notably that polymerases cannot synthetize DNA once a secondary structure is formed, questions arise about the complications they could bring to replication,[63] especially in terms of genome stability. Helicases are proteins that can unwind DNA structures and separate the strands. Helicases have been reported to unwind G-quadruplexes and the loss of their activity can be very detrimental. Helicase Pif1, for example, has been shown to be able to strongly bind and unwind G-quadruplexes and prevent genome instability associated with G4-DNA.[64] Other helicases have also been reported to unwind G4-DNA: BLM is a helicase from the RecQ helicase family that is defective in Bloom's syndrome (which is characterized by genomic instability), and it has been shown to be able to unwind G4-DNA;[65] prokaryotic RecQ can also unwind G-quadruplexes as well as duplexes, and interestingly it is possible to selectively inhibit G4-DNA unwinding by using G-quadruplex stabilizing ligands;[66] Dna2 has a helicase and nuclease activity, it binds to G4-DNA of the telomeric repeat and is able to unwind it but its nuclease activity is attenuated when the structure is formed.[67]

Repressor-activator protein 1 (RAP1) is a key player in telomere regulation in yeast, it is able to bind to double stranded DNA and it can recognize and bind to both parallel and antiparallel G-quadruplexes formed by yeast telomeric sequences.[68] What is more, it has been argued that RAP1 promotes G-quadruplex formation.[69] Yeast protein G4P2 was also found to selectively bind G4-DNA.[70] Telomere end-binding proteins (TEBPs) have been shown to cooperate in the formation or antiparallel G-quadruplex in the telomere in protozoa.[71]

These are only a few examples of the several proteins that have been identified as being able to bind to G-quadruplexes.[72] Of particular interest is the finding that transcription factor SP1, which recognizes a G-rich binding site, can recognize and bind G-quadruplex structures.[73] Furthermore, the authors of this study also found that there is a significant overlap between SP1 binding sites and potential G-quadruplex structures. Interestingly, SP1 is not the only transcription factor that has a good overlap with G-quadruplex putative sequences. An analysis[74] of human, chimpanzee, mouse and rat

promoters found that nine transcription factors significantly occur within 100 bases of putative G4-DNA, including SP1, VDR, KROX, WT1, MAZ, Kid3, ZF5, AP-2 and ETF. Interestingly, all of these transcription factors are zinc-finger binding proteins except AP-2 and ETF. In a different study, transcription factor MAZ was shown to reduce the stability of a G-quadruplex that contained an experimentally validated MAZ binding site[75] and was able to unfold it.[76] This was a surprising result because the same authors had previously shown that MAZ stabilized a different G-quadruplex[77] that also contained a MAZ binding site.

If G-quadruplexes form *in vivo*, can they be observed directly?

Different probes and antibodies have been designed to bind selectively to G4-DNA in order to detect it. The first cellular visualisation of G4-DNA was done in 2001 using an *in vitro* generated antibody for a ciliate telomeric sequence, *Stylonychia lemnae*, which showed that telomere in this organism forms only antiparallel conformations and crucially is resolved during replication.[78] An antibody was engineered later in 2008 to have high affinity for human G-quadruplexes over duplex DNA,[79] and it was subsequently used to detect G-quadruplex in a pull-down assay and sequence the recover and sequence the strands.[80] In a different study, human fixed cells were treated with a G-quadruplex selective antibody and visualized by means of a labelled secondary antibody.[81] This work clearly shows formation of these structures in both telomeric and non-telomeric regions. With this technique the authors were also able to show differences of G4-DNA abundance during the cell cycle.

A fluorescently labelled derivative of Pyridostatin was used to visualise the interaction and location of Pyridostatin inside human fixed cell which showed localization in non-telomeric regions. There was a good overlap with sites that showed helicase GFP-hPif1 staining. This suggests the drug targeted naturally formed G-quadruplexes (as opposed to inducing formation).[82]

Recently, a different strategy was developed by using small ligands as fluorescent probes which display different fluorescence life-times upon binding to G4-DNA structures as compared to other topologies such as duplex and singly-stranded DNA.[83] This work was able to show for the first time the interaction of a small molecule with G4-DNA in live cells.

What roles could G-quadruplexes play in biology?

Several potential roles of these structures have been investigated. The relevance of telomeric G4-DNA has been made evident by the research mentioned before. Notably, the telomere is rich in G-repeats and it is known to form readily G-quadruplexes that inhibit telomerase activity[84]. This observation is of great interest given that inhibiting this protein, responsible for elongating the telomere in *ca.* 85% of cancer cells,[85] has been identified as a potential anticancer therapeutic approach. Increasing

evidence of the role of RNA G-quadruplexes show they may be involved in translational regulation.[86] Finally, G-quadruplex structures found in the promoter region have sparked a lot of interest for their potential role in gene regulation, and they are discussed in more detailed in the next section.

1.1.4 Regulatory roles of G-quadruplexes

As discussed in Section 1.1.3, bioinformatic studies have shown that putative G-quadruplex forming sequences are over-represented in promoter regions. Some examples of studied G-quadruplexes structures in promoters are shown in Figure 1.1.9. Interestingly, there is also an association with genes that are closely related with cancer. This has sparked a lot of interest into trying to understand the possible roles that this structure can have in gene transcription.



Figure 1.1.9 Examples of intramolecular G-quadruplexes structures that have been studied in promoters. From left to right, c-myc (PDB 1XAV),[87] c-kit (2O3M),[88] bcl-2 (PDB 2F8U).[89]

One of the most popular methods for studying the effects of G4-DNA is the luciferase reporter assay (LRA). A firefly luciferase gene[90] is cloned into a vector containing a mammalian promoter and transfected into cells, afterwards the cells are lysed and the luciferase protein is quantified by its catalytic luminescent activity. The amount of protein will depend on the activity of the promoter. This technique will be discussed further on Chapter 4 section 4.2.

In 2002, Siddiqui-Jain *et al*[91] published an article discussing the effect of a G-quadruplex sequence, "Pu27", in the transcriptional control of an important oncogene *c-Myc*. Using LRA they showed that

when the formation of the G-quadruplex was altered by mutations the activity of the promoter increased three-fold. This led to the idea that the formation of the G-quadruplex in the promoter region could function as a suppressor of transcription, and therefore it could become an attractive target for ligands which could stabilise it and consequently reduce the expression of said oncogene. This was further supported by the fact that G-quadruplex stabilizing ligand TMPyP4 (see Chapter 3, section 3.1.1) indeed reduced promoter activity while TMPyP2, a related compound that does not bind well to G-quadruplexes, did not have such strong effect. Human PARP-1 has been shown to be able to bind to the guanine rich sequence of human *c-Myc* promoter, influencing transcription by promoting B-DNA form and reducing formation of G4-DNA[92] and increasing transcription. The binding of the protein to DNA was inhibited by TMPyP4.

It has been suggested that G-quadruplexes in the promoter region can act as transcription suppressors by occluding the recognition site of transcription factors like SP1.[57] In support of this mechanism, MAZ has been shown to be able to unfold G-quadruplexes, SP1 and MAZ often interact synergistically.[76] Nevertheless, it has also been shown that SP1 is able to bind to G-quadruplexes[73], albeit with lower affinity than to duplexes, which suggests that SP1 is not necessarily inhibited by G-quadruplex formation.

New evidence of G-quadruplex sequences effect in promoter regions continue to appear in the literature. Other examples of G-quadruplexes in promoters that have been shown to act as repressors include the genes: *ADAM-15*[93] which is aberrantly expressed in various cancers; *PALB2*[94] which is associated with breast cancer, *HRAS*[76] overexpression correlates with tumour invasiveness; and the viral promoter HIV-1[95] which can be targeted for antiviral activity. Some of the putative sequences found can form several G-quadruplexes like *hTERT*, the catalytic domain of telomerase. This promoter contains 12 G-runs and can form a pair of consecutive G-quadruplexes[96], and interestingly one of them can adopt very different conformations.[26]

The promoter of *bcl-2* has also attracted a lot of attention because it is an apoptosis inhibitor that is over expressed in several cancers, has two promoters, and one of them (P1) is particularly G-rich so it can form several G-quadruplexes.[97,98] From the studied sequences, "Pu39"[97] exhibits repressor behaviour similar to *c-myc*, while "p32"[98] shows an almost opposite behaviour. When the promoter is mutated to disrupt G-quadruplex formation there is a sharp decrease in transcription, but upon treatment with a G-quadruplex stabilizing ligand only the wild type promoter exhibits a decrease in transcription.[98] This exemplifies that even if the G4-DNA is not itself found as a repressor element, it can still nevertheless be targeted using small molecules to decrease the transcription of a given gene.

Another good example of druggability of G-quadruplexes is the promoter region of the Androgen Receptor (AR). When the entire G4 sequence is deleted from the core promoter region, no transcriptional change is observed. However, upon G4-ligand treatment only the promoter that contains the putative G4-DNA sequence was targeted and the transcription was reduced, while the transcription of the promoter lacking the G4-DNA sequence was not affected by the G4 stabilizing ligand.[99]

G-quadruplex formation in the promoter of *c-kit*, a gene that encodes a receptor tyrosine kinase, has a G-rich promoter and more than one G-quadruplex forming sequences have been identified and studied. [73,100,101] This high propensity for G-quadruplex formation has made it an attractive target for ligand binding studies, and indeed, small molecules have been developed that can target and decrease *c-kit* expression.[102] Furthermore, a molecule that can destabilize the formation of G-quadruplexes has been shown to increase the expression of the gene.[103]

G-quadruplex forming sequences have also been described downstream of the transcription start site, for example in the *c-myb* promoter.[104] This sequence contains three regions with four GGA repeats each, and biological studies have revealed that the deletion of the first and last repeat increases the promoter's activity, while the deletion of the second does not affect transcription. However, if the three sections are deleted, the promoter is abrogated.[104] To understand these results, the authors examined the binding of transcription binding factor MAZ and found it can bind to both the double stranded and the G-quadruplex formed sequence.[104]

Figure 1.1.9 and Table 1.1.1 show examples of putative G-quadruplex sequences that have been reported in the literature as proved to form intramolecular G-quadruplex structures *in vitro*. Conclusions by the authors regarding the effect on gene regulation are also included on Table 1.1.1.

Table 1.1.1 Table of example characterized G-quadruplex sequences found in promoter regions in the literature.

Gene		Sequence	T_m(°C)	Observation in LRA	Ligand effect		Regulation	Ref
ADAM-15		TTGGGGCCGGGTGGGAGGGGGCGGGCCGGGGCGGGGCC	88.2	G→T mutant, ↑3-fold	TMPyP4	↓	Repressor	[93]
AR		AGGGGAGGGGAGAAAAGGAAAGGGGAGGGGAGGGA	72*	Delete G4, no change	MM45	↓	Target	[99]
Bcl-2	Pu39	AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG		G→A mutant, ↑2-fold	SYUIQ	↓	Repressor	[97]
	Mid39	F-CGGGCGCGGGAGGAAGGGGGCGGGAGC-T	48.5					
	p32	TGGGGTCCGCGACGGGGTGGGGGCTCCCGGGG	82	G→A mutant, ↓3-fold	PD[a]	↓	Activator	[98]
c-kit	c-kit1	AGGGAGGGCGCTGGGAGGAGGG	51	Stabilizing ligands		↓	Target	[100,102]
	c-kit2	AGGGCGGGCGCGAGGGAGGGG	70	Destabilizing ligands		↑		[101,103]
	c-kit*	GGCGAGGAGGGGCGTGGAAGG	55					[73]
c-myb		TTTCTCAGGAGAAAGAGGAGGAGGAGGTCACGGAGGAGGAG		Delete R1 or R3 or R1+R2		↑	Repressor	[104]
		GAGAAGGAGGAGGAGGAAACAGGTT		Delete R1+ R2+ R3		↓	/ Activator	
c-myc	Pu27	TGGGGAGGGTGGGGAGGGTGGGGAAGG	>80	G→A mutant, ↑3-fold	TMPyP4	↓	Repressor	[91]
HIF-1α		GCGCGGGGAGGGGAGAGGGGGCGGGAGCGCG		G→C mutant, ↓70%				[105]
HIV-1	LTR-II	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGTTTTT	49	G→T mutant, ↑3-fold	BRACO-19	↓	Repressor	[95]
	LTR-III	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGTTTTT	52					
HRAS	hras-1	TCGGGTTGCGGGCGCAGGGCACGGGCG	53, 67	G→T mutant, ↑4-fold	DIGP	↓	Repressor	[75]
	hras-2	CGAGGCCGGGGCGGGGCGGGGGCGGGGGCGCGCGGT	>90	G→T mutant, ↑5-fold	DIGP	↓	Repressor	

Table 1.1.1 (continued) Table of example characterized G-quadruplex sequences found in promoter regions in the literature.

Gene		Sequence	T_m(°C)	Observation in LRA	Ligand effect		Reg.	Ref
hTRET	I	TTTTTGGGGAGGGGCTGGGAGGGTTTTTTTT	68		BRACO-19	↓†		96,106
	FL	TTTTTGGGGAGGGGCTGGGAGGGCCCGGAGGGGGCTGGGCC G GGGACCCGGGAGGGGTCGGGACGGGGCGGGGTTTTTTTT						
KRAS		GGGAGGGAGGGAAGGAGGGAGGGAGGGA	68,80	G→T mutant,↓50%	DIGP	↑	Activator	77
PALB2		GGGACAGGGGTTGGTGGGGTAGGGGCAGGTTGGG	85	G→T mutant,↑ 38%			Repressor	94
VEGF		CGGGCGGGCCGGGGGCGGGT	77.9		Several	↓	Target	107,108

Melting values (T_m) are shown at 100 mM KCl, two values are presented for biphasic melting curves. Column "Observation in LRA=Luciferase reporter assay" shows the results seen when comparing the wild type sequence *versus* a mutant that disrupts G-quadruplex formation. Mutations may be several within a sequence. The "ligand effect" column shows the transcription observed for the wild type sequence using different G-quadruplex stabilizing ligands *versus* the untreated wild type. *Value extracted from the melting curves presented in the literature. [a]PD induces conformational changes in the G4-DNA, does not enhance its stability.†BRACO-19 is thought to promote targeted degradation.

Smestad and Maher[109] recently showed that the loss of RecQ helicase activity can be associated with the down-regulation of genes with abundance in putative G4-DNA sequences, supporting the idea that G-quadruplex formation in the promoter region suppresses transcription of genes. Interestingly, the authors of this study noticed that the genes that were up-regulated were low in G4-DNA, (however, they did have G-quadruplex forming potential as opposed to zero G-quadruplex forming potential). In other words, a high representation of potential G-quadruplex sequences was associated with down-regulation and low representation (but not zero representation) with up-regulation upon losing RecQ activity. They also noted that the effect in regulation of these structures was dependent on their position. Based on these observations, they proposed a model in which a G-quadruplex forming sequence located 140-270 bp downstream of the TSS of the sense strand have transcription activating effect. On the other hand, G-quadruplex sequences near the TSS are associated with decreased transcription. Their proposed mechanism is shown in Figure 1.1.10.

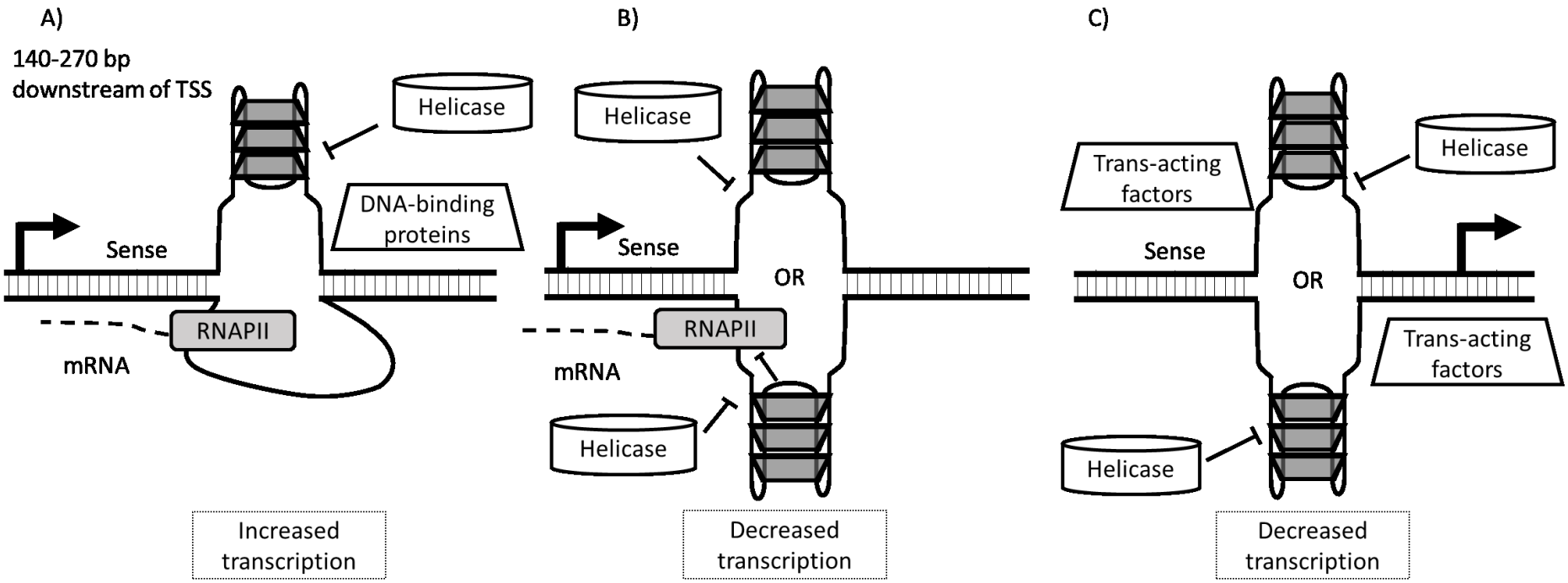


Figure 1.1.10 Schematic representation of the mechanism proposed by Smestad and Maher[109]. A) A putative quadruplex sequence (PQS) downstream the transcription start site (TSS) on the sense strand is correlated with increased transcription; this may be due to the antisense strand being available for RNA Polymerase II (RNAPII), aiding transcription. G-quadruplex formation may also recruit DNA-binding proteins. Helicases may attenuate this effect. B) A PQS that is positioned downstream the TSS on either the sense strand or the anti-sense strand can be correlated with decreased transcription because it blocks the activity of RNAPII, this effect may be attenuated by helicases. C) A PQS that is positioned upstream of the TSS on either the sense strand or the anti-sense strand can be associated with decreased transcription by recruitment of trans-acting factors that attenuate transcription, this effect may be alleviated by helicases. Figure adapted from the literature[109]. Cations have not been included for clarity. G-quadruplex formation is not implied in both strands simultaneously.

1.2 ZIP1 REGULATION

1.2.1 ZINC, ZIP1 AND PROSTATE CANCER

Zinc is found as a divalent cation, zinc(II), in biological systems and has no redox chemistry under physiological conditions. It has a flexible coordination geometry and is of labile nature.[110] It is known to have several essential biological roles and its deficiency is detrimental for human health.[111] Some well-known examples are the association of zinc intake with childhood growth[112] and with the immune system.[113] Deficiency consequences include growth retardation, altered immune response, skin changes and alopecia.[114] Its wide range of activity may be explained by its redox neutrality and coordination capacity which makes it a suitable cofactor in hundreds of enzymes as well as give it its central role in zinc finger proteins; in addition it is thought to be a key player in signalling pathways including as a neurotransmitter and a second messenger.[115] The only source of zinc is dietary and absorption is poor; the content of zinc in food is low in the most common zinc sources,[114] even in those that are considered zinc rich (meat, liver, egg) or have low bioavailability (whole wheat, whole corn) which means that zinc deficiency is common.[114]

Zinc concentration and distribution varies between different cell types, but in general total cell zinc concentration is in the micromolar range while cytosolic free zinc is in the picomolar range.[116] Given that zinc(II) is a Lewis acid it is not surprising that it has strong protein binding capacity and therefore its cytosolic concentration in eukaryotic cells is tightly regulated. Compartmentalisation, zinc transporters, and soluble zinc-binding proteins like metallothionein come together in a strict regulation of free cytosolic zinc concentration to counteract perturbations.[116]

It has long been established that prostate cells are able to accumulate much higher concentration of zinc than any other soft tissue, however cancerous prostate cells do not accumulate zinc and show levels similar to non-prostate cells.[117] Zinc levels have been proposed to be a useful measure for diagnostics, especially to distinguish prostate cancer from Benign Prostatic Hyperplasia (BPH), as well as a target for therapeutic approach.[118]

Zinc has been implicated in several metabolic pathways including the regulation of apoptosis, but the direction (inhibit or promote) seems to depend heavily on the cell type studied. In prostate cells it promotes apoptosis.[119] Mitochondrial zinc concentration in prostate epithelial cells is also higher than in non-prostate cells,[120] this high concentration has been shown to interfere with citrate oxidation and leads to the accumulation of citrate in prostate cells[121] by disrupting the Krebs cycle. The drop of zinc levels in prostatic cells occurs early in the malignancy development,[122] it allows citrate oxidation

changing the bioenergetics of the cell as well as diminishing the apoptogenic effect of zinc in the cells which in combination may contribute to the proliferation of the cancerous prostate cells.[123] Epidemiological studies that have used zinc supplementation to prevent prostate cancer have produced contradictory results which could be explained by the diminished expression of zinc uptake transporters.[124] It has been argued that decrease in zinc accumulation and malignancy development occurs due to a down-regulation of proteins involved in zinc uptake.[123,125,126]

Two families of zinc transporters regulate the intracellular concentration of this micronutrient in eukaryotic cells: ZnT (identified as SLC30A, its name comes from Zinc Transporter) and ZIP (identified as SLC39A, ZIP stands for Zrt/Irt-like protein where Irt refers to Iron regulated transporter). They have opposing functions as shown in Figure 1.2.1, ZnT proteins lower intracellular zinc concentration by pumping zinc outside of the cell or into intracellular vesicles (efflux) and ZIP proteins increase intracellular zinc concentration by taking zinc from the outside of the cell or from these vesicles (influx). The expression of members of both families have been shown to be regulated by zinc itself and hormones.[127,128]

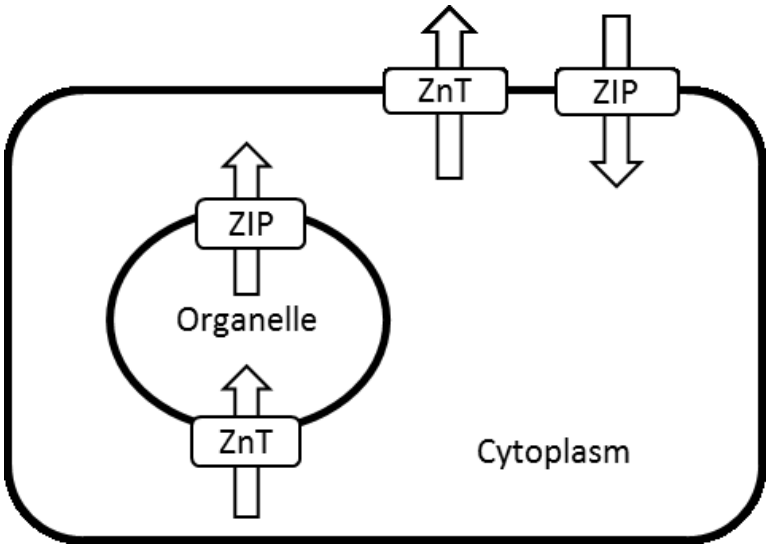


Figure 1.2.1 Schematic representation of zinc transport and the role of ZIP and ZnT transporters including direction of transportation. Image adapted from the literature.[129]

Since the first description of an influx zinc transporter in 1996,[130] the ZIP superfamily has been largely studied. It is found in a large variety of organisms including eubacteria, archae, fungi, insects, plants and mammals suggesting ancient origin.[131] There are fourteen members found in the human genome and have been named from ZIP1 to ZIP14. Most ZIP members have eight predicted transmembrane domains and a histidine-rich loop in the internal side of the membrane and both C and N termini on

the outside of the membrane as shown in Figure 1.2.2. They have been grouped in four subfamilies according to the similarity of their sequences.[131]

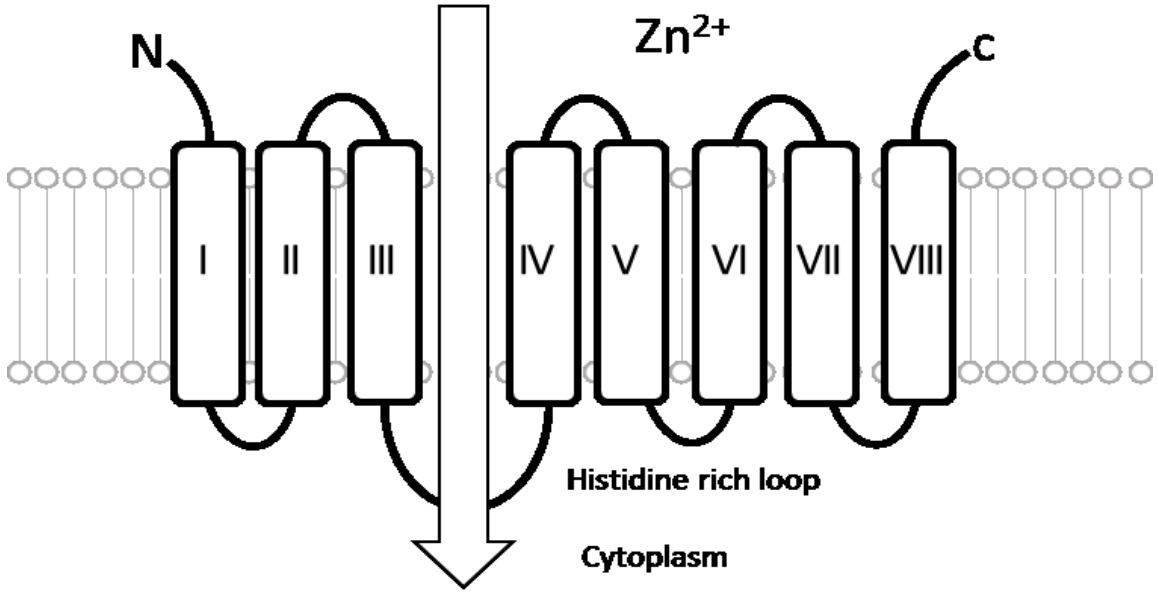


Figure 1.2.2 Schematic representation of a ZIP protein indicating the transmembrane domains and loops. Zinc direction of transportation is also indicated. Image adapted from the literature.[132]

Human ZIP1, member of ZIP subfamily II, is ubiquitously expressed in all human tissues although protein location varies depending on cell type.[133] It can be found located in the organelle membranes in zinc appropriate cultures but translocated to the cell surface when zinc is limited.[134] Changes in localisation have also been reported in gut epithelial cells according to differentiation state.[135] It was first experimentally shown to be an zinc transporter in K562 cells in 2001 by Gaither and Eide[136] and in prostate cells by Costello et al[137] in 2003. Its zinc uptake is energy-independent,[136] inhibited by copper, cadmium and iron.[136] It should be noted that despite the similarities of ZIP1 to iron transporters and the fact that zinc transport is inhibited by iron, it is suggested that ZIP1 is not an iron transporter based on experimental observations.[136] However, the ZIP1 mouse variant is only modestly inhibited by other metals.[138] ZIP1 capacity to uptake zinc depends on the ligands chelating it: while it will not take EDTA-bound zinc it will uptake citrate bound and free zinc.[137] A study using mutant ZIP1 concluded that the histidine-rich loop was key for zinc transport and suggested the mechanism for zinc transport would involve extracting the ion from chelating agents and binding to the loop.[139]

Knock-out studies in mice of ZIP1 and double knock-out of ZIP1,ZIP3 showed no obvious phenotypes changes when fed with appropriate levels of zinc, but did show abnormal embryonic development if fed zinc deficient diets, suggesting a key role during pregnancy of ZIP1.[140]

Evidence suggests that the loss of zinc accumulation in prostate cells is linked to a down-regulation of the *hZIP1* gene, this silencing occurs early in the malignancy development and that it happens at transcriptional level.[141] This has led to the proposal of labelling *hZIP1* as a tumour suppressor gene for prostate cancer and highlighted the importance of understanding the nature of its regulation.[142] Furthermore, there is evidence to suggest that overexpressing *hZIP1* inhibits prostate tumour growth by inhibiting transcription factor NF-κB activity and so the malignant pathways associated with it.[126]

Prostate cancer is not the only disease related to changes in *hZIP1* expression; it has been suggested that *hZIP1* functions as a tumour suppressor gene in clear cell renal cell carcinoma.[143] On the other hand, inhibiting a *Drosophila* Model for Alzheimer's disease orthologue of *hZIP1* showed amelioration of the symptoms suggesting that altering expression of this zinc transporter could be used as a treatment strategy[144] for this neurodegenerative disease.

1.2.2 Regulation of *hZIP1*

Prostate cancer cell models retain the potential to accumulate zinc, which suggests that the loss for zinc accumulation is not caused by a permanent mutation or deletion of the zinc transporter, but rather by silencing of the gene.[123,124] This section explores the known mechanisms that alter ZIP1 expression. Figure 1.2.3 shows a schematic representation of the regulation mechanisms discussed.

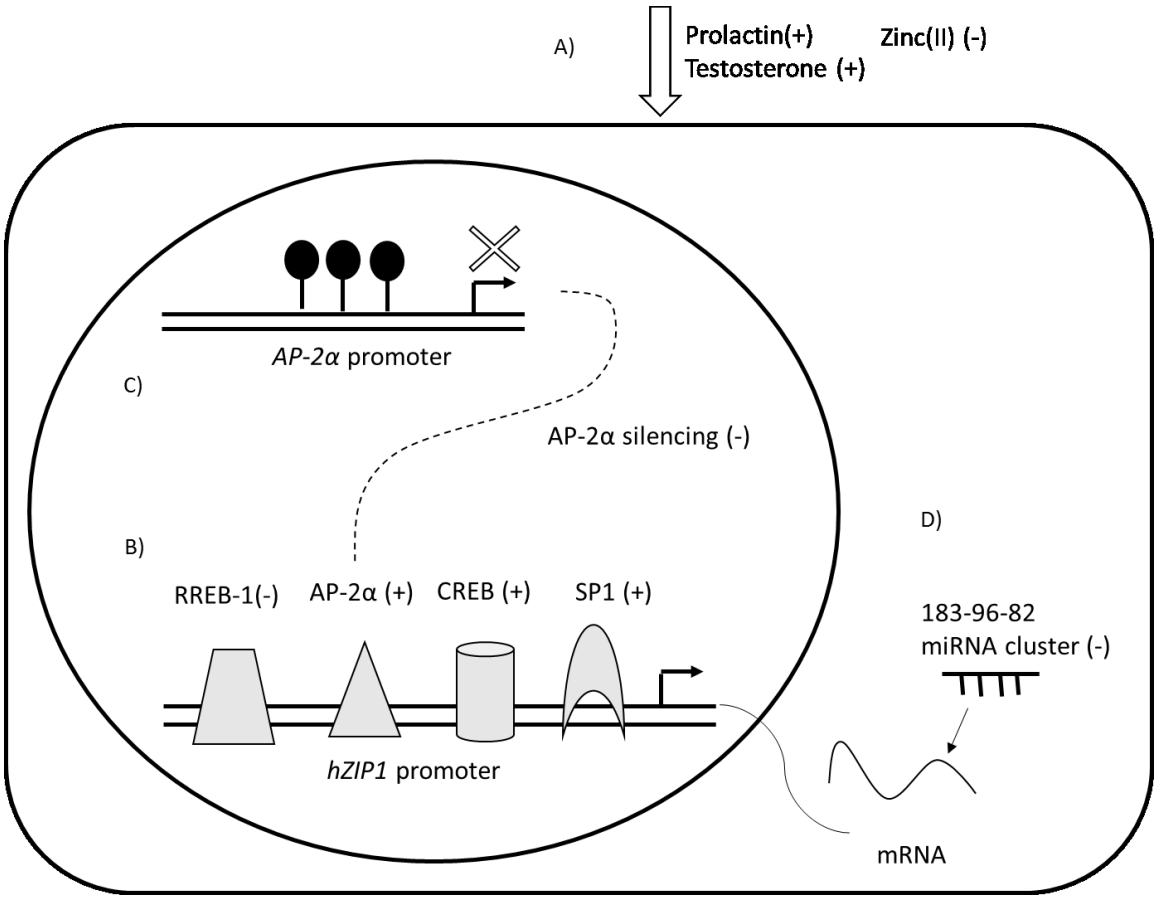


Figure 1.2.3 Summary of proposed regulation mechanisms for *hZIP1*. Up-regulators are marked with (+) while down-regulators are marked (-). A) Hormones testosterone and prolactin up-regulate ZIP1 while available zinc(II) down-regulates it. B) Transcription factors SP1, CREB and AP-2α are critical for transcription while RREB-1 acts as a repressor. C) Transcription factor AP-2α is down-regulated by methylation of its promoter which in turn down-regulates ZIP1. D) MicroRNA cluster 183-96-82 down-regulates ZIP1 post-transcription.

Hormonal and zinc regulation

Early studies by Costello *et al*[128] in 1999 showed that *hZIP1* is up-regulated by prolactin and testosterone as well as down-regulated by zinc *in cellulo*. Others were not able to replicate the prolactin result,[133] and argued it could be because of differences in protocol. A later review written by the authors of both studies argues in favour of the increase in zinc accumulation caused by hormonal signalling, possibly through ZIP1 up-regulation.[145]

The mechanism of zinc down-regulation has not been further explored, so the exact details of how it causes down-regulation of *hZIP1* are not clear. A general mechanism would be described as zinc(II) sensitive transcription factors that bind to the promoter depending on zinc(II) availability.[146] The detailed mechanism has been studied for other zinc transporters including some of the ZIP family.[146]

The zinc sensitive transcription factor MFT1 [(metal-response-element)-binding-transcription factor1] is involved in the transcriptional repression of ZIP10.[147,148] ZIP4 and ZIP5 transcription in mouse are not affected by zinc levels, however they are regulated post-transcriptionally; ZIP4 mRNA is not stable when zinc levels are adequate but is stabilized when levels are low and this leads to accumulation of the protein; ZIP5 mRNA in the other hand accumulates when the levels of zinc are adequate, but when it is deficient the translation is arrested.[149]

The promoter of *hZIP1* and transcription factors

In 2009, Makhov *et al*[150] reported a study on the TATA-less promoter of *hZIP1*, which they predicted using ElDorando and Gene2Promoter programs using data from Chromosome 1 (GenBank accession number NT_004487), and subsequently characterized experimentally using different techniques including luciferase reporter assay. The characterized transcription start site and core promoter around it is shown in Figure 1.2.4.

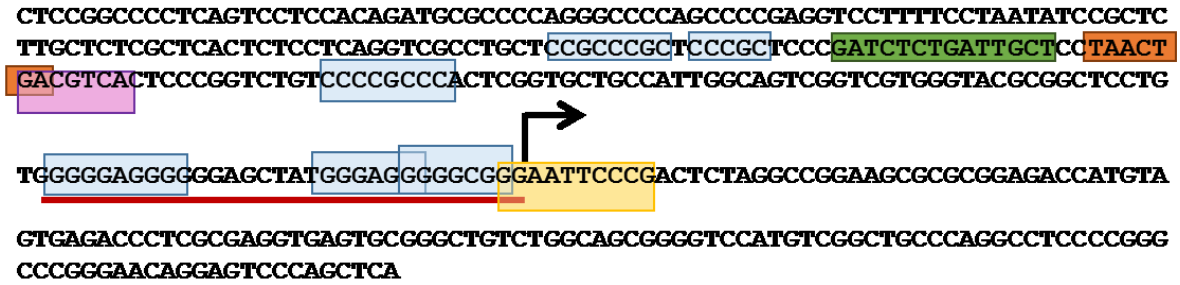


Figure 1.2.4 Promoter region of *hZIP1*. Image adapted from the literature[150] using promoter sequence downloaded from EPDNew. Transcription start site is marked by a black arrow. Blue boxes are predicted SP1 binding sites; the green box is a predicted GATA1 site; the red box is a predicted c-Myb binding site; the purple box is a predicted CREB binding site and the yellow box is a predicted NF-κB binding site. The G-rich section underlined is a putative G-quadruplex sequence.

They reported one transcription start site (TSS), a core promoter region with high transcriptional activity, and studied the effect of predicted binding sites and concluded that SP1 and CREB had critical roles in regulation (disrupting these binding sites dramatically decreases transcription). They could not directly disrupt NF-κB binding site because it overlaps the TSS, however they did not observe changes in transcription when using NF-κB stimulators or inhibitors, nor did they observe binding of the transcription factor to the sequence suggesting it may not be involved. Transcription factor AP-2α was also shown to be a critical transcription factor in a follow-up study by the same authors.[151]

Shortly afterwards, Milon *et al* [152] published a study in transcriptional regulation of *hZIP1* using data from NCBI Reference Sequence NM_014437.4 which led to the finding of a second TSS as well as a

confirmation of the site found described earlier. They noted that given the several different transcripts reported for this gene that have variable numbers of exons more than one promoter may be functional. As well as a good agreement on the TSS, the promoter region described is also in agreement with the core promoter reporter previously[150] described. Furthermore, they found that disrupting RREB-1 up-regulates *hZIP1* expression and therefore RREB-1 down regulates the expression of *hZip1*, this transcription factor is part of the RAS pathway that is up-regulated in prostate cancer. In a follow-up study, they confirmed their results by showing that overexpressing RREB-1 resulted in a decreased expression of *hZIP1.*[153]

Epigenetic mechanisms

Methylation of the GC-rich promoters is an epigenetic mechanism associated with the silencing of genes.[154] The possibility of the methylation of the *hZIP1* promoter was investigated by Makhov *et al*[151] in 2011 and they found that *hZIP1* promoter had not been methylated in the cancerous cell lines they studied. However they found that the promoter region of transcription factor AP-2α (which binds to the *hZIP1* promoter) was indeed hypermethylated, which helped explain the lower levels of this protein found in cancerous prostate cells. They found that inhibiting DNA methyltransferase, the enzyme that maintains methylation, the expression of ZIP1 was up-regulated. Furthermore, they found that the promoter region of this transcription factor was also methylated in prostate tissue of prostate cancer patients.

MicroRNA mechanisms

MicroRNAs (miRNA) are short non-coding RNA sequences that bind to target mRNA to either degrade it or arrest translation, and they are key players in gene regulation.[154] A study by Mihelich *et al*[155] in 2011 showed that miRNA down-regulates ZIP1. In particular, miR-182 was found to inversely correlate with ZIP1 mRNA levels in prostate cancer patients and was found to be a regulator of ZIP1 *in vitro*. The sequence miR-182 is expressed in a cluster overexpressed in prostate cancer, miR-183-96-182, which was found to suppress not only ZIP1 but other zinc transporters from both the ZIP and the ZnT families (ZIP3,7,9,10 and ZnT1,7,9).

Could a G-quadruplex be involved in regulation?

The promoter region of *hZIP1* described contains a G-rich section near the transcription starting site (Figure 1.2.4).This region has the potential to form a G-quadruplex. There are three SP1 binding sites overlapping the G-rich sequence, one of them has been shown to have an important effect in the regulation of the gene.[150]

The recent report of the experimental mapping of G-quadruplex of the entire genome using method G4-Seq published by Chambers *et al*[51] included a table (Supplementary Table 6) with all experimental observed G-quadruplexes in annotated genes. The results show that for SLC39A1, the number observed G-quadruplexes for the entire gene (length 8600 bases) are higher using both potassium and Pyridostatin than predicted using the Quadparser algorithm. These results are shown in Table 1.2.1, where genes MYC and KIT have also been included for reference. While the observed G-quadruplexes is not as high as for MYC or KIT, it is nevertheless higher than the predicted number. It should be noted that these results were published after the commencement of this research project, which aims is to investigate the effect of the formation of the potential G-quadruplex in the promoter of *hZIP1* near the transcription start site.

Table 1.2.1 Predicted G-quadruplexes versus experimentally found using G4-Seq. Data taken from the literature[51]. MYC and KIT are presented as reference.

G-quadruplex	SLC39A1	MYC	KIT
Number of Predicted Quadruplexes	5	4	7
Observed Quadruplexes with Pyridostatin	10	15	17
Observed Quadruplexes with potassium	6	11	17

1.3 OBJECTIVES

1.3.1 GENERAL AIMS

G-quadruplexes have been implicated in the regulation of several genes when formed in the promoter, and they have become attractive targets for therapeutic design of molecules. ZIP1 is a tumour suppressor protein that is down-regulated during the development of prostate cancer. Understanding the mechanisms of regulation of ZIP1 is an important step to disentangle the processes that govern malignancy progress. A potential G-quadruplex structure has been identified in the promoter region of *hZIP1* and given the importance of understanding the regulatory mechanisms, it is the aim of this project to study the effect of the potential G-quadruplex formation in the transcription of this gene and explore the effects of selective G-quadruplex binders.

1.3.2 SPECIFIC OBJECTIVES

In order to address the overall aims, a number of measurable objectives were identified:

Study the putative G-quadruplex sequence of the *hZIP1* gene to show if it can form an intramolecular G-quadruplex under physiological conditions *in vitro*. Study it using circular dichroism spectroscopy (CD), electrophoretic mobility shift assay (EMSA), thermal difference spectra (TDS), and polymerase stop assay (PSA).

Design relevant mutated sequences that disrupt the G-quadruplex formation (non-G4-forming mutants) as well as control mutated sequences that retain the G-quadruplex potential (G4-forming mutants). Study these sequences under physiological conditions *in vitro* and assess whether they can form stable intramolecular G-quadruplex structures.

Study the effect of such mutations in the stability of the G-quadruplex by studying thermodynamic and kinetic properties of the structures formed by hZIP1-1 and its mutants.

Study the effect of using stabilizing ligands on the thermodynamic and kinetic properties of said sequences.

Compare the mRNA levels produced by the wild type promoter to the mutated promoters *in vitro*. Compare the effect of G-quadruplex stabilizing ligands in the production of mRNA using wild type and mutated promoters *in vitro*. To do this, construct the relevant luciferase plasmids containing the mutations of interest in the promoter using site directed mutagenesis and adapt an *in vitro* transcription protocol for the study of G-quadruplexes in promoters.

Compare the activity of the wild type promoter *versus* mutated promoters *in cellulo*. Use luciferase reporter assay to assess the effect of G-quadruplex ligands using wild type and mutated promoters by quantifying expressed protein.

CHAPTER 2: CHARACTERIZATION OF HZIP1 G-QUADRUPLEX

The G-quadruplex forming potential of the G-rich sequence found in the promoter region of *hZIP1* is analysed in the first part of this chapter and suitable mutants are designed based on transcription factor analysis and G-quadruplex potential analysis. The selected sequences are studied using electrophoresis, circular dichroism spectroscopy, thermal difference spectra and polymerase stop assay under different annealing conditions. Conditions that promote unimolecular parallel G-quadruplex structures were identified. The sequences are further thermodynamically characterized using melting techniques like CD melting and FRET melting, and kinetically described using a reverse complementary strand competition.

2.1 BIOINFORMATICS ANALYSIS

2.1.1 G4-DNA SEQUENCE IN THE PROMOTER REGION OF *HZIP1*

To study the potential role of a G-quadruplex structure in the transcriptional regulation of *hZIP1*, different bioinformatics tools were used to analyse the promoter. As shown in Figure 1.2.4, the core promoter region of gene *hZIP1* located in chromosome 1, contains a putative G-quadruplex forming sequence near the experimentally validated transcription start site (TSS):[150] GGGGGAGGGGGGAGCTATGGGAGGGGGCGGG at position 153963551 . This promoter sequence was firstly analysed using the Quadparser online tool which at the time of writing is no longer available, and it was predicted to form a G-quadruplex. This sequence will be referred to as hZIP1-1 or WildType for the entire document.

Eukaryotic Promoter Database new (EPDNew)* is a database that contains information of experimentally characterized eukaryotic promoter regions and automatically compiled promoter list derived from next generation sequencing data from high-throughput promoter mapping.[156] It contains four human promoters for ZIP1 (Gene name SLC39A1), from which SLC39A1_3 agrees with the one described by Makhov *et al*[150], as well as two mouse promoters and one zebrafish promoter. When

* http://epd.vital-it.ch/

promoter SLC39A1_3 was analysed using QGRS Mapper[†], which is an online software that predicts G-quadruplex formation,[53]

with parameters set to search for a sequence with a minimal size of 3 guanines group (also referred in this work as "G-runs") and a loop size of maximum 15 bases, it finds 1 possible G-quadruplex sequence with 126 overlaps. The software assigns a G-score to each overlapping sequence, which is calculated by selecting the guanines involved in the G-tetrad, and suggests the more stable sequence based on shorter loops of roughly same size and greater number of tetrads formed, the higher the score the better the candidate.[53] When the entire promoter was analysed using QGRS Mapper, the sequence hZIP-1 was found as a putative G-quadruplex sequence and assigned it a G-score of 47 for GGGGGAGGGGGGAGCTATGGGAGGGGGCGGG where the underlined guanines are predicted to be part of the G-tetrad. For reference, the predicted G-score for *c-myc* Pu27 (one of the best characterised G-quadruplexes in a promoter) sequence is 41. This suggested that sequence hZIP1-1 was a good candidate sequence to study because it has a high G-quadruplex forming potential, and given it is located near the transcription starting site it may be involved in regulation. The mouse promoters also contained G-quadruplex folding sequences while the zebrafish promoter did not.

Basic Local Alignment Tool (BLAST) [‡] was used to search for similar sequences in the human genome. No highly similar sequences were found, therefore hZIP1-1 sequence is only found in the promoter region of SLC39A1. When this sequence is used to run a BLAST search against the SLC39A1 gene or predicted SLC39A1 genes of different animals, it is found in closely related genomes like chimpanzee, gibbon, orangutan and other primates.

The human promoter was aligned with the mouse, as well as the SLC39A1 upstream region for rat, chimpanzee and cow. These sequences show a good alignment, and interestingly the guanines predicted to be involved in the G-tetrads are well conserved among them as shown in Figure 2.1.1.

[†] http://bioinformatics.ramapo.edu/QGRS/index.php
[‡] http://blast.ncbi.nlm.nih.gov/Blast.cgi

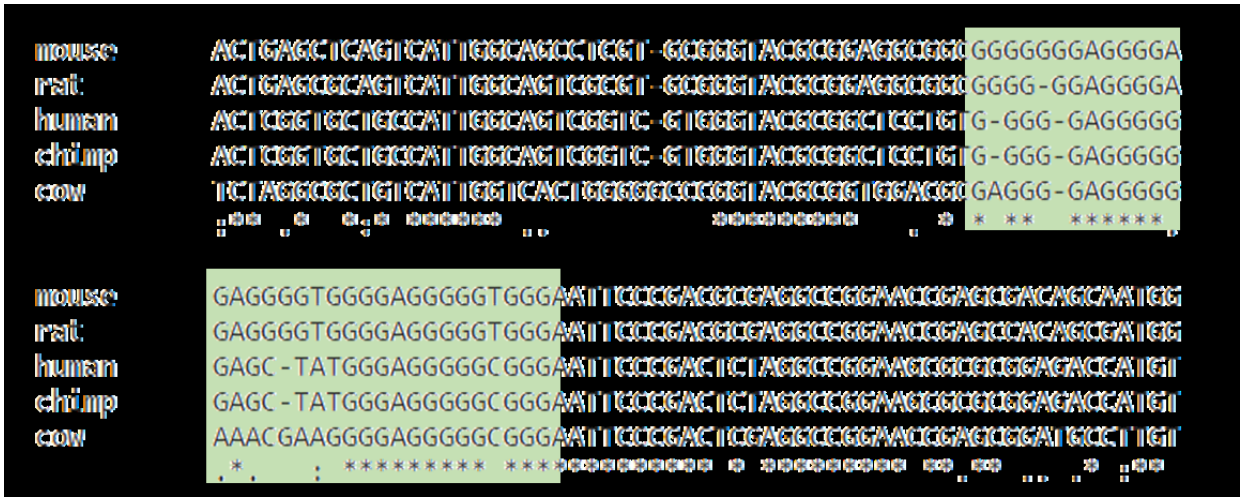


Figure 2.1.1. Alignment of promoter regions for different animals using Clustal Omega[§]. This tool can align several sequences of both DNA and proteins[157]. Conserved bases are shown with an asterisk. The G-quadruplex forming region found in human (hZIP1-1) and sequences aligned to it are highlighted in green.

It is perhaps not surprising to find this G-quadruplex forming sequence to be conserved in these animals given that it overlaps with important transcription factor binding sites. It has been reported that SP1 binding sites often overlaps G-quadruplex forming sites,[73] and indeed the sequence has been reported to overlap three SP1 binding sites, one which has been shown to have a critical role in regulation.[150] Figure 2.1.2 shows the SP1 binding sites that have been studied experimentally by Makhov *et al,*[150] from which SP1-3 site is known to be critical, while SP1-1 and SP1-2 sites were not considered as important. They used luciferase reporter assay (LRA) to determine that the double mutation of these two sites did not reduce promoter activity by much (approx. 10-20%). The mutations they used to disrupt the binding sites would still allow for G-quadruplex formation.



Figure 2.1.2 Identified SP1 binding sites that overlap hZIP1-1. It has been reported that SP1-3 is critical to transcription but SP1-1 and SP1-2 are not.[150]

[§] http://www.ebi.ac.uk/Tools/msa/clustalo/

2.1.2 MUTANT DESIGN FOR THE STUDY OF HZIP1-1

In order to better understand the role this putative G-quadruplex forming sequence can have in the promoter region of ZIP1, different mutants were designed to be used in biological experiments such as *in vitro* transcription and LRA. The sequences designed and used in this study are listed in Table 2.1.1. In order to avoid disrupting a critical SP1 binding site (SP1-3), all mutations were chosen in the first three G-runs. To fully disrupt G-quadruplex formation, a mutant was designed without enough G-runs to form three G-tetrads. This sequence, dubbed MutE, is made of several mutations that were split into MutB, a sequence that disrupts an unimportant SP1 binding site using the consensus SP1 mutation, MutG and MutF which do not contain the consensus mutation. MutH is also a G-quadruplex disrupting sequence with fewer mutations than MutE and does not contain the consensus mutation for SP1. MutF can potentially form a G-quadruplex but as opposed to MutB and MutG it is expected to have reduced stability as reflected in the calculated G-score because of the long loop located between the first and the second G-run.

Table 2.1.1 WildType sequence (hZIP1-1) and its mutants. Guanines underlined are predicted using QGRS Mapper to be part of tetrad. G-score was calculated using QGRS Mapper, the higher the value the more stable the G-quadruplex is expected to be. Mutations are shown in lower case and bold. Critical SP1 binding site, SP1-3, is highlighted in grey. Truncated versions of hZIP1-1 are also presented, deletions are marked with a dash (-).

Name	Sequence	G-score
hZIP1-1	TGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGA	47
MutB	TGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGA	47
MutE	TG**tttt**AGG**ttt**GAGCTATGGGAGGGGGCGGGA	No G4 found
MutF	TGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGA	37
MutG	TG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGA	44
MutH	TG**tt**GGAGG**ttt**GAGCTATGGGAGGGGGCGGGA	No G4 found
hZIP1-2	------AGGGGGGAGCTATGGGAGGGGGCGGGA	38
hZIP1-3	TGGGGGAGGGGGGAGCTATGGGAGGGGGC----	39

While MutE and MutH do not have enough G-runs to form G-quadruplex containing three tetrads, they could still potentially form two-tetrad structures. Biophysical characterization described in the next part of this chapter will prove they do not, in fact, form any G-quadruplex structure.

To ensure that no additional transcription factor binding site was disrupted or created in the design of these mutants, all sequences were analysed using the full promoter with appropriate mutations

with Nsite online program from Softberry[**] which works by searching statistically non-random regulatory elements using their consensus sequence.[158] A second analysis was done using Alibaba2.1[††] online software. Full results report for each promoter sequence can be found in the Appendix section A.1 for both tools, a summary of the results can be found on Table 2.1.2. Results are shown as either the binding site for a transcription factor (like SP1) or as a binding sequence (like GC-box).

When the WildType (hZIP1-1) was analysed using Nsite and AliBaba2.1 several regulatory sequences were found. These include the SP1 binding sites shown in Figure 2.1.2, plus transcription factors E2F-1, E2F-2, E2F-4, H4TF1, Krox20, and ETF. While SP1 binding sites have been studied[150] using a combination of consensus mutations in LRA and electrophoresis mobility super-shift assay to confirm binding, the others have not yet been studied. However, the disruptions used to study SP1 overlap many of these binding sites. MutB contains the consensus SP1 mutation for SP1-1 binding site used in the studies reported before[150] which showed a promoter activity of ≈80% with this mutation, therefore H4TF1 and the first Krox20 and ETF sites are not critical for transcription. Transcription factor binding sequences for the E2F family and the second Krox20 and ETF sites are not disrupted in any of the mutated sequences as they overlap the critical SP1-3 binding site and their relevance is not further studied. However, It should be noted that SP1, KROX and ETF binding sites distribution has been associated with putative G-quadruplex sequences.[74]

When MutB was analysed using Nsite the results show a loss of transcription binding sites at position that overlap the first two G-runs. MutE contains the SP1-1 binding disruption found in MutB plus other mutations, and when analysed it using Nsite it shows identical results to MutB. This is also true for all other mutants, including MutF and MutG that do not contain the SP1 consensus mutation. When the mutants are analysed using AliBaba2.1, MutG contains both SP1-1 and SP1-2, MutB contains SP1-2, MutF and MutH contain SP1-1, and MutE does not contain either SP1-1 or SP1-2. No new binding site is found in any of the mutants.

Given that SP1-1 and SP1-2 are not critical, it can be concluded that all mutants have predicted equivalent binding sites to each other. Therefore, their promoter activity is expected to be very similar between them, and just marginally worse than the wild type sequence. If MutE and MutH show a different activity, this may be attributable to the lack of G-quadruplex folding potential.

[**] http://www.softberry.com/berry.phtml?topic=nsite&group=programs&subgroup=promoter
[††] http://www.gene-regulation.com/

Table 2.1.2 Summary of results by Nsite and AliBaba2.1.

Name	Sequence	Transcription factor	
WildType	CCTGTGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGAATT		
Nsite	GGGGGAGGG	SP1-like	NB
	GGGGCGGAA	E2F-1,2,4	*
	GGGGGAGGG	H4TF1	NC
	GGGGCGGGA	SP1	*
AliBaba2.1	=CACCC-bi=	CACCC-box	NC
	======Sp1=====	Sp1	NB, NC
	==Krox-20=	Krox-20	NC
	====ETF===	ETF	NC
	====Sp1===	Sp1	NB
	=======Sp1=====	SP1	*
	==Krox-20=	Krox20	*
	====ETF===	ETF	*
MutB	CCTGTGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGAATT		
Nsite	GGCGGGAA	E2F-1,2,4	*
	GGgAGGGgGCGGGaA	GC-box	
	GGGGCGGGA	SP1	*
AliBaba2.1	====Sp1===	Sp1	NB
	=======Sp1=====	SP1	*
	==Krox-20=	Krox20	*
	====ETF===	ETF	*
MutE	CCTGTG**tttt**AGG**ttt**GAGCTATGGGAGGGGGCGGGAATT		
Nsite	GGCGGGAA	E2F-1,2,4	*
	GGgAGGGgGCGGGaA	GC-box	
	GGGGCGGGA	SP1	*
AliBaba2.1	=======Sp1=====	SP1	*
	==Krox-20=	Krox20	*
	====ETF===	ETF	*

Table 2.1.2 (continued) Summary of results by Nsite and AliBaba2.1

Name	Sequence	Transcription factor	
MutF	CCTGTGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGAATT		
Nsite	GGCGGGAA	E2F-1,2,4	*
	GGgAGGGgGCGGGaA	GC-box	
	GGGGCGGGA	SP1	*
AliBaba2.1	=CACCC-bi=	CACCC-box	NC
	=====Sp1====	SP1	NB, NC
	=======Sp1=====	SP1	*
	==Krox-20=	Krox20	*
	====ETF===	ETF	*
MutG	CCTGTG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGAATT		
Nsite	GGCGGGAA	E2F-1,2,4	*
	GGgAGGGgGCGGGaA	GC-box	
	GGGGCGGGA	SP1	*
AliBaba2.1	=======Sp1=====	SP1	NB, NC
	=====Sp1====	SP1	NB
	=======Sp1=====	SP1	*
	==Krox-20=	Krox20	*
	====ETF===	ETF	*
MutH	CCTGTG**tt**GGAGG**ttt**GAGCTATGGGAGGGGGCGGGAATT		
Nsite	GGCGGGAA	E2F-1,2,4	*
	GGgAGGGgGCGGGaA	GC-box	
	GGGGCGGGA	SP1	*
AliBaba2.1	====Sp1===	SP1	NB, NC
	=======Sp1=====	SP1	*
	==Krox-20=	Krox20	*
	====ETF===	ETF	*

Binding sites lost in MutB are marked with (NC), NC=non-critical. Sites shown not to bind experimentally are marked with (NB), NB=non-binding. Sites conserved in all mutants are marked with (*).

2.1.3 CONCLUDING REMARKS

A putative G-quadruplex sequence was found in the promoter region of *hZIP1*. This sequence, dubbed hZIP1-1 has a high G-quadruplex forming potential and it is localized near the transcription starting site. The sequence is also unique to this promoter. This makes hZIP1-1 an attractive sequence to study as a potential regulator of *hZIP1*. While the sequence is not conserved entirely between species, the guanines that are predicted to be the most relevant for G-quadruplex formation are indeed highly conserved.

In order to study the impact of the G-quadruplex *in vivo*, a series of mutants were designed to either disrupt G-quadruplex formation (MutE and MutH) or not disrupt it (MutB, MutF and MutG). The last two G-runs of hZIP1-1 were conserved among all mutants because they are reported to be critical for transcription. Software prediction of transcription factor binding based on consensus sequence analysis shows that the mutated sequences MutB, MutE, MutF, MutG and MutH will all have a similar binding, therefore a similar promoter activity can be expected. If there are differences in promoter activity they could be then attributed to their G-quadruplex forming potential. No new binding sites are predicted to have been formed by the designed mutations.

2.2 BIOPHYSICAL AND BIOCHEMICAL CHARACTERIZATION

2.2.1 ANNEALING CONDITIONS STUDIED BY ELECTROPHORETIC MOBILITY SHIFT ASSAY AND CIRCULAR DICHROISM SPECTROSCOPY

In order to study the characteristics of the G-quadruplex formed by hZIP1-1, it is necessary to assess if it forms said structure at physiological conditions of potassium and pH. The potential of hZIP1-1 to form an intramolecular G-quadruplexes was studied using a combination of methods. Because oligonuclotides G-rich sequences can be folded in a variety of ways there may be different G-quadruplex structures for the same given sequence. The number of possibilities is higher for sequences with a greater number of G-runs.[40] Because hZIP1-1 contains five G-runs polymorphism was expected, which could in turn complicate characterization. Additionally, the formation of tetrameric and dimeric structures would also need to be minimized in order to study the properties of the intramolecular structure. Polymorphism can be reduced by enriching one particular folding over the others, and this can be achieved by mutations , changing cations and adding flanking bases.[159] Of particular interest is the observation that using different annealing conditions such as temperature and concentration can enrich monomeric structures over higher order structures.[160]

Different annealing methods were therefore studied to fold hZIP1-1 into a monomeric G-quadruplex, since it was considered to be more biologically relevant. The resulting structure was studied using electrophoretic mobility shift assay (EMSA) and circular dichroism (CD) spectroscopy. EMSA can differentiate between high order and monomeric structures,[161] but cannot differentiate between different monomeric species. CD spectroscopy is a bulk measurement that helps us distinguish between known G-quadruplex topologies (parallel *versus* antiparallel), but tells us little of the contribution of the many possible structures that can exist simultaneously in solution.[160] CD is commonly used alongside information gathered from known G-quadruplex structural motifs: for all G-quadruplexes there is a positive peak near 210 nm; parallel structures will also show a negative band near 240 nm and a positive one near 265 nm; antiparallel structures show a small positive band near 240 nm, a negative band near 260 nm and a positive band near 295 nm; hybrid structures (like 3 +1) show a positive band both at 260 and 295 nm.[162] While this empirical information helps characterize new sequences by comparing them with the spectra of known structures, it is worth mentioning that the observed bands are explained by polarity as opposed to folding of the strands (the polarity in antiparallel conformation alternates while parallel structures do not alternate polarity).[27] All oligonucleotide sequences, annealing methods, protocols and materials are described in detail in the

Experimental Procedures; summaries with the key features of each annealing method are presented here.

METHOD 1 is based on typical DNA annealing: a single stranded DNA (ssDNA) solution in appropriate buffer is heated up to 95 °C and cooled down slowly to room temperature. This method aims to melt structures and anneal them in the most thermodynamically stable conformation. Figure 2.2.1 shows that the oligonucleotide sequence hZIP1-1 annealed by this method displays characteristic parallel G-quadruplex spectra in the presence of potassium.

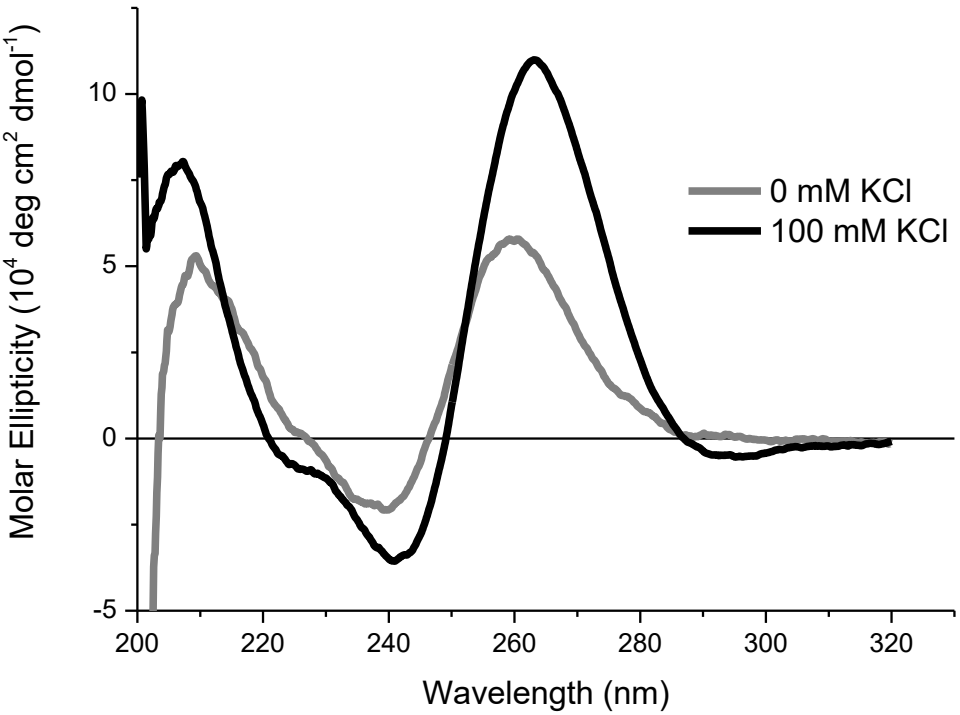


Figure 2.2.1, Circular dichroism spectra of hZIP1-1 DNA (5 µM) in 50 mM Tris buffer pH=7.4 at 20˚C annealed by METHOD 1 using different amounts of KCl.

METHOD 2 consists on heating up a ssDNA solution in buffer without cations to 95 °C and cooling it down rapidly by placing the sample in ice, after this the potassium or other cation is added and incubated for a few minutes, then brought to room temperature. This method relies in melting structures during the first step and then on the spontaneous formation of G-quadruplex when a G-rich sequence is in presence of stabilizing cations such as potassium. This method aims to favour rapidly folded structures. Changes in the CD spectrum of the sequence after each step are shown in Figure 2.2.2. It can be appreciated that the sequence at 95 °C is melted, after cooling it rapidly there seems to be some structure formation, and after adding potassium there is a clear G-quadruplex

structure with a small negative band near 240 nm, a strong positive peak near 260 nm and a small positive peak at 295 nm. The last peak suggests an antiparallel contribution (alternating polarity) that could be explained by the presence of a mixture of a parallel structures population and a population of antiparallel structures.

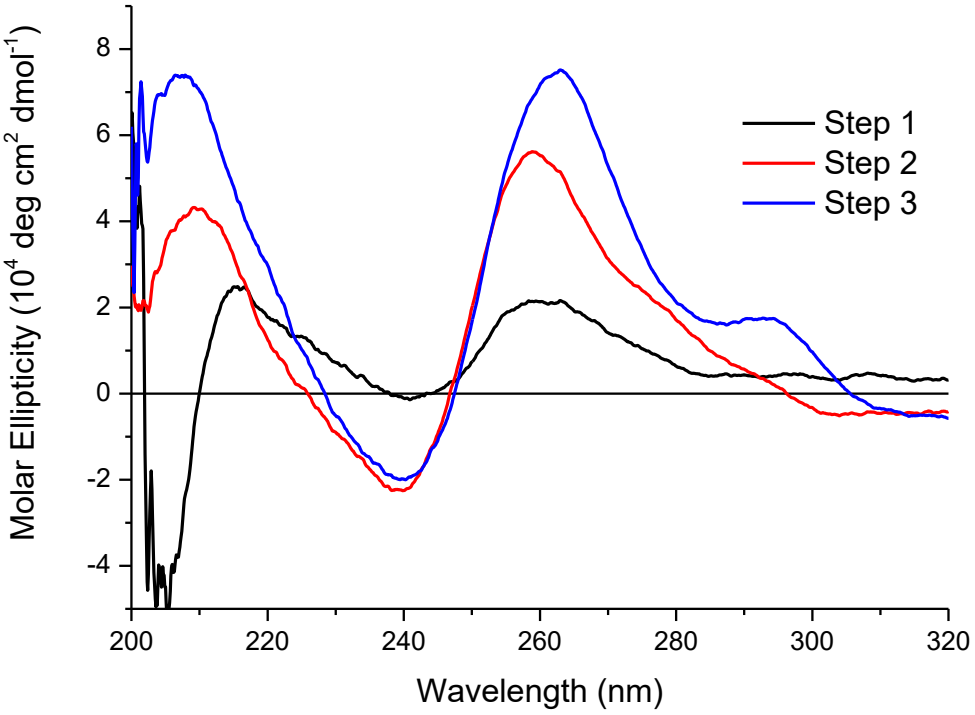


Figure 2.2.2, Circular dichroism spectra of hZIP1-1 DNA (5 µM) in 50 mM Tris buffer pH=7.4 during annealing by METHOD 2. Step 1: Sample heated up to 95 °C for 5 minutes with no added cations. Step 2: Sample cooled down rapidly in ice for 5 minutes with no salt. Step 3: Sample after adding 100 mM KCl.

METHOD 3 consists in heating up a ssDNA solution in buffer without cations to 95 °C and cooling it down to 60 °C; at this point potassium is added to the solution and it is incubated, then the sample is brought down to room temperature. METHOD 3 was designed after observing a structural transition at 60 °C in the melting curve of hZIP1-1 annealed by METHOD 2 as described in section 2.3.1 later in this chapter. This method aims to favour parallel over antiparallel G-quadruplex structures. Changes in the CD spectrum after each step are shown in Figure 2.2.3. It can be appreciated that the structure at both 95 °C and 60 °C is melted, and after adding potassium a G-quadruplex is formed. There is a small negative band near 240 nm and a strong positive band near 260 nm suggesting a parallel structure.

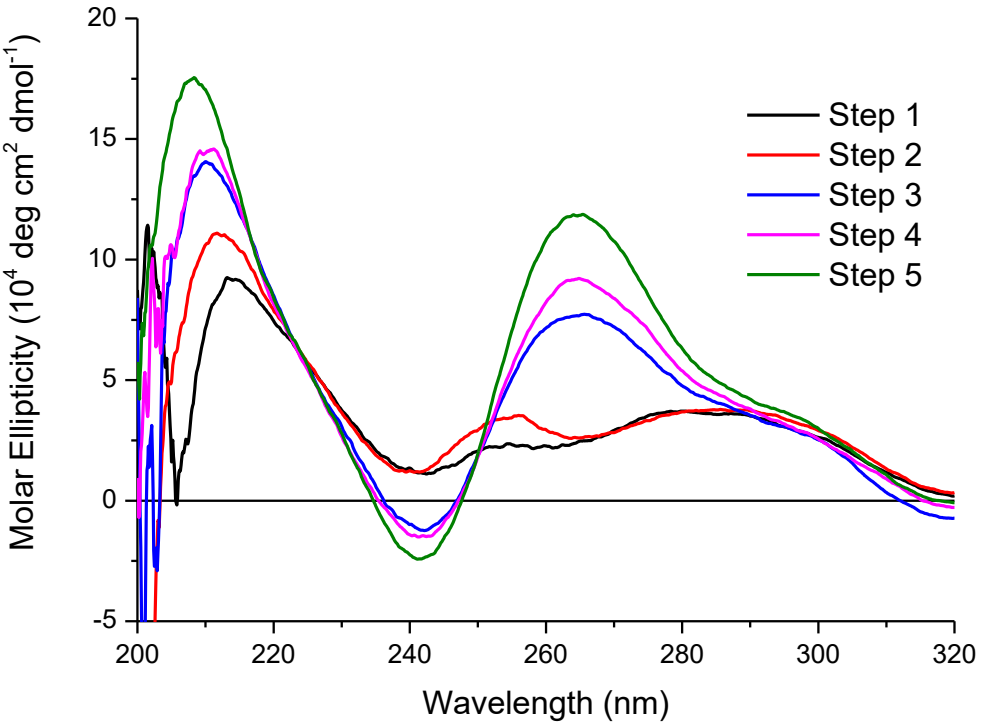


Figure 2.2.3, Circular dichroism spectra of hZIP1-1 DNA (5 µM) in 50 mM Tris buffer pH=7.4 during annealing by METHOD 3. Step 1: Sample at 95 °C with no added cations. Step 2: Sample at 60 °C with no added cations. Step 3: Sample at 60 °C immediately after adding 100 mM KCl. Step 4: Sample at 60 °C after 10 minutes. Step 5: Sample at 20 °C.

The CD spectra of hZIP1-1 annealed by the different methods are shown in Figure 2.2.4 and their respective EMSA are shown in Figure 2.2.5. When the sequence has been annealed in the presence of KCl it has a maximum at 263 nm for all methods. All gels were prepared and ran under the same conditions including 20 mM KCl in gel and running buffer, except gel A was prepared and ran without any added salt. A double stranded DNA (dsDNA) ladder was used for reference to compare different gels. As a negative control for G-quadruplex formation hZIP1-1 was annealed in 50 mM Tris buffer without added cations, heated up to 95 °C and brought back to room temperature by placing on the bench to cool down. It is shown as “no salt” in Figures 2.2.4 and 2.2.5A. The band in Figure 2.2.5A shows the migration of the unfolded sequence near the 25 bp mark.

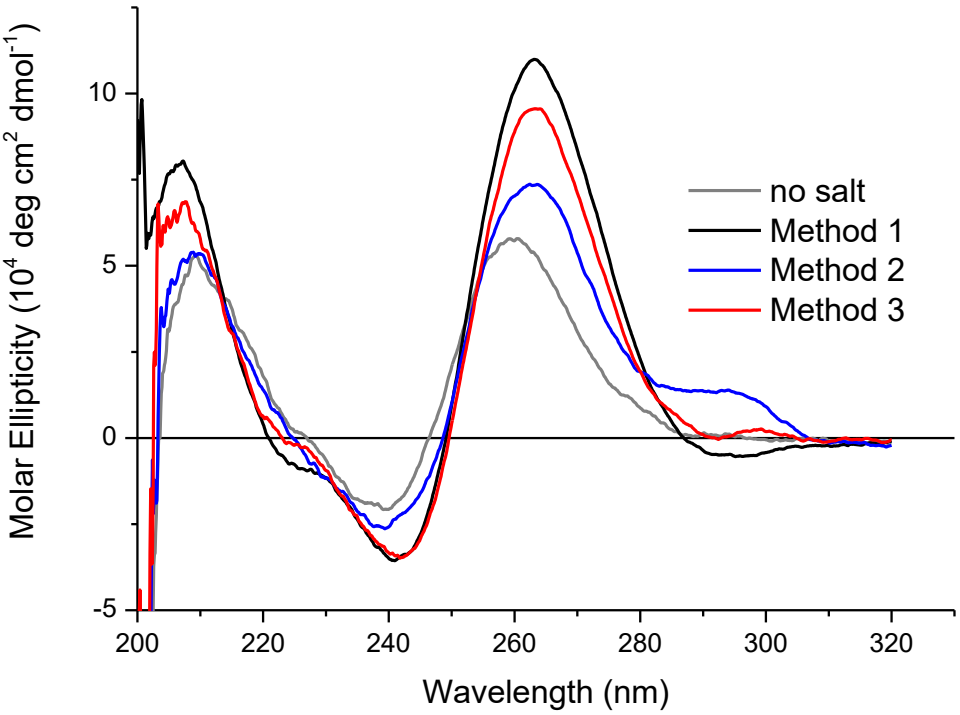


Figure 2.2.4 Circular dichroism spectra of hZIP1-1 DNA (5 µM) in 50 mM Tris buffer pH=7.4 at 20˚C annealed using different methods and 100 mM KCl except "no salt". This was prepared by that heating up the sample in 50 mM Tris buffer to 95 °C and brought back to room temperature.

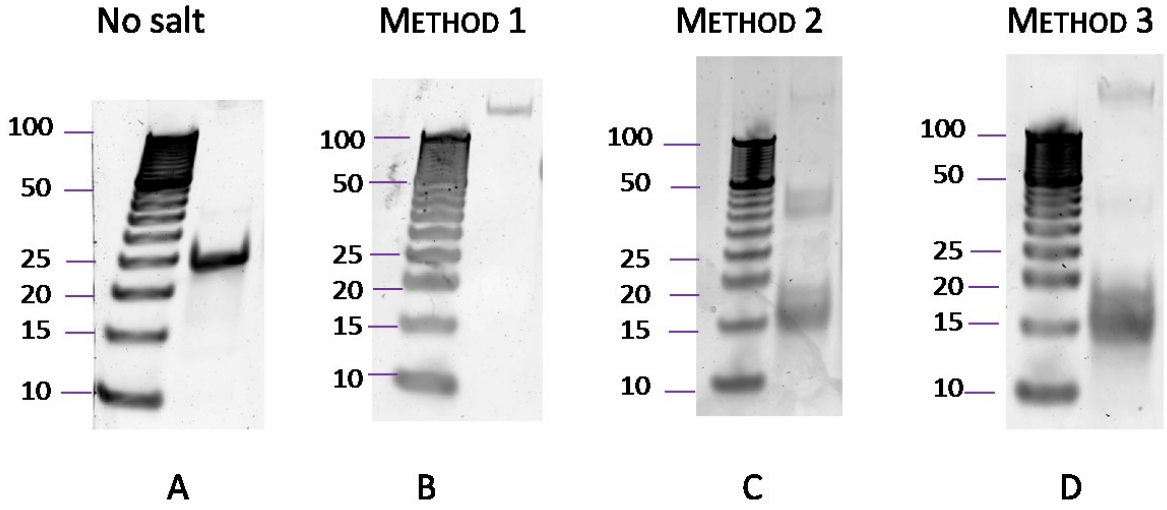


Figure 2.2.5 Native polyacrylamide (16%) with 20 mM KCl (except no salt) gels of hZIP1-1 annealed using different methods. All sequences were annealed at 5µM using Tris 50 mM buffer (pH=7.4) with 100 mM KCl except "no salt". A) hZIP1-1 was heated up to 95 °C and brought back to room temperature. B) hZIP1-1 was annealed using METHOD 1. C) hZIP1-1 was annealed using METHOD 2. D) hZIP1-1 was annealed using METHOD 3. All gels show a 5 bp DNA Ladder for reference.

Annealing hZIP1-1 using METHOD 1 gives a spectrum which exhibits a negative peak at 240 nm and a positive peak at 263 nm, and it shows the highest intensity of all samples. Figure 2.2.5B shows a very slow migrating band, which indicates the sequence is arranged in a multistranded fashion. A high intensity in the CD spectrum is in agreement with this observation.[160,163]

Figures 2.2.5C and 2.2.5D show that sequences annealed by METHOD 2 and METHOD 3 respectively have fast migrating bands near the 15 bp mark. These bands run faster that the band observed in the absence of potassium (Figure 2.2.5A) which indicates they correspond to intramolecular G-quadruplexes[161]. The samples were loaded without dyes that could affect migration[164], so the fact that more than one fast migrating band is observed is explained as two or more different intramolecular folding populations. Both annealing methods still produce a multistranded G-quadruplex population as evidenced by the slow migrating band; however the monomeric structure is considerably enriched. Additionally, there is a double band near 40 bp mark in the sample annealed by METHOD 2 which indicates the formation of dimers.

In Figure 2.2.4, the sample annealed by METHOD 2 shows a positive band near 263, a negative band near 240 nm and a small positive band near 290 nm, while the sample annealed by METHOD 3 shows only a positive band near 263 and a negative band near 240 nm. The intensity of the bands is not similar despite both being mostly monomeric, the sample annealed by METHOD 3 has a more positive band in 263 nm and a more negative band near 240 nm. The two positives bands described for the sequence annealed by METHOD 2 could be explain as either an antiparallel conformation (like 3+1), as a parallel dimer, or a mixture.[162,165] Given that the sequence is expected to be highly polymorphic and that it displays several bands when annealed using this method, several conformations contribute to the CD spectra. Figure 2.2.5C shows a dimeric band that is not present in Figure 2.2.5D which argues in favour of the 290 nm contribution being due to a parallel dimer in the population mixture. On the other hand, a sequence in the *hTERT* promoter has been shown to be able to form both parallel and antiparallel (3+1) intramolecular conformations, and the sum of the two conformations gives rise to a very similar spectrum as the one presented here.[26] Crucially, these two forms coexist in solution and display very similar melting curve behaviour of structure conversion as the one described for hZIP1-1 in section 2.3.1, which argues in favour of the coexistence of parallel and antiparallel intramolecular structures in hZIP1-1 when folded using METHOD 2. It is concluded that a parallel multistranded, parallel dimer, and parallel and antiparallel monomers coexist in solution when the sequence is annealed by this method.

The sequence annealed by METHOD 3 produce an enriched monomeric all parallel folding, therefore this annealing was chosen as the standard for characterization.

To further understand the sequence hZIP1-1, truncated versions missing either the first (hZIP1-2) or the last (hZIP1-3) G-runs, were also studied. The CD spectra recorded for the hZIP1-1, hZIP1-2, hZIP1-3, and different mutants annealed with different conditions can be found on the Appendix section A.3. The mutant sequences MutB, MutE, MutF, MutG and MutH that were designed for *in vitro* and *in cellulo* studies were annealed using METHOD 3 with 100 mM KCl and their electrophoretic migration and CD spectra are shown in Figure 2.2.6. It can be appreciated in Figure 2.2.6 (bottom) that all mutants display the typical parallel G-quadruplex bands. This is not expected neither for MutE nor MutH as they should fully disrupt G-quadruplex formation. While MutH shows a very similar spectrum to MutF, and expected G-quadruplex, MutE shows a slightly shifted main peak. Electrophoretic migration reveals that while MutB, MutF and MutG show a single fast migrating band near the 15 bp mark (like hZIP1-1), MutE and MutH show a slower migrating band near 20 bp mark (single stranded), as well as a second band near the 40 bp mark (dimeric). The CD spectra of MutE and MutH could then be explained as bimolecular G-quadruplexes. In the next section further evidence is provided that MutE and MutH cannot form intramolecular G4-DNA structures.

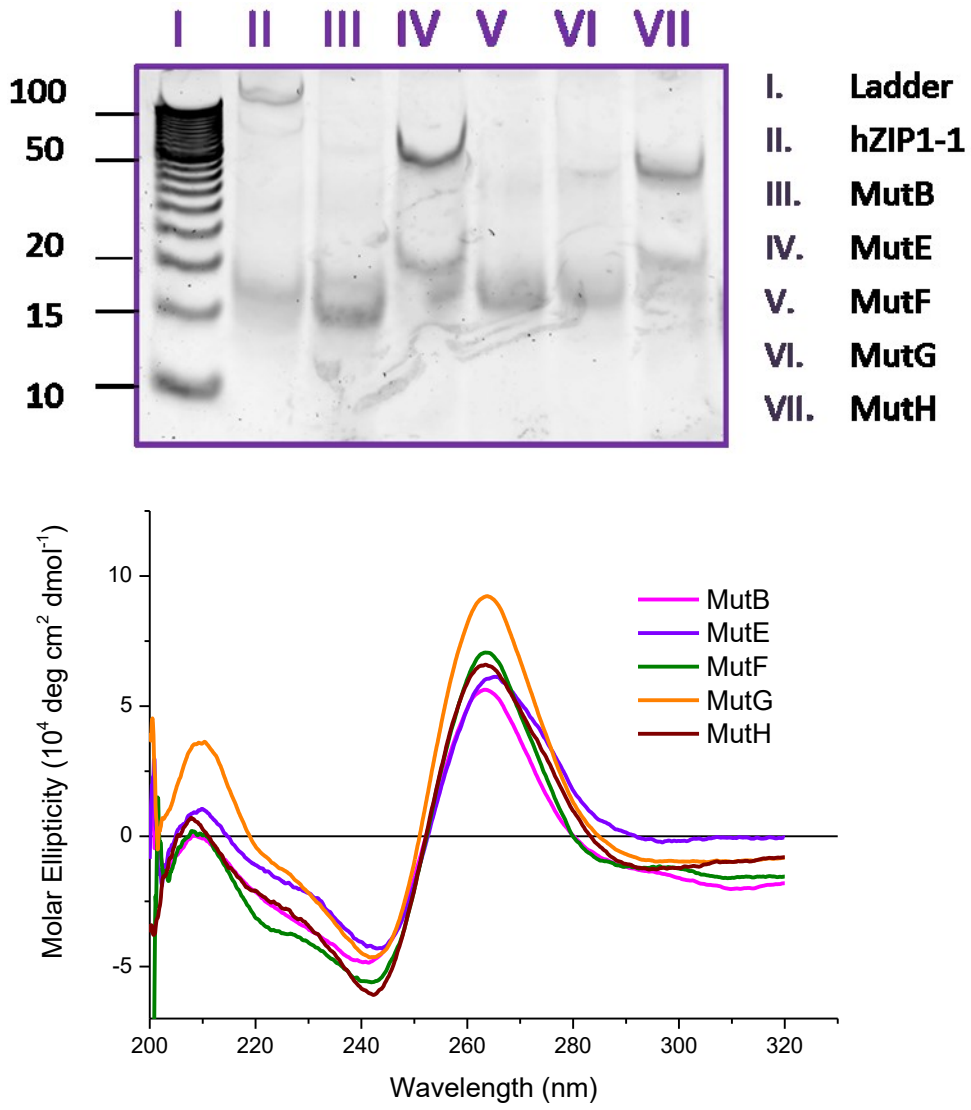


Figure 2.2.6, Electrophoretic mobility shift assay and circular dichroism spectra for MutB, MutE, MutF, MutG and MutH annealed by METHOD 3 in 50 mM Trist buffer (pH=7.4) with 100 mM KCl at 20˚C. Top) Native polyacrylamide electrophoretic gel (16%) with a 5 bp ladder as reference. Bottom) Circular dichroism spectra of mutants.

2.2.2 THERMAL DIFFERENCE SPECTRA AND POLYMERASE STOP ASSAY

To further characterize the potential of hZIP1-1 to form intramolecular G-quadruplexes and to assess the potential of the designed mutants to do so as well, thermal difference spectra (TDS) and polymerase stop assays (PSA) were used.

The difference between the UV absorbance spectra of a thermally unfolded and a folded sequence will give a thermal difference spectra (TDS), the shape of which indicates the structure of the folded sequence. The overall shape for a G-quadruplex is characterized by a major positive peak at 243 , another major positive peak at 273 and a negative peak at 295 nm.[166] This methodology is particularly

useful to help characterize G-quadruplex structures that can be melted under 90 °C, like monomeric structures. Higher-order structures will usually require harsher conditions to be melted.

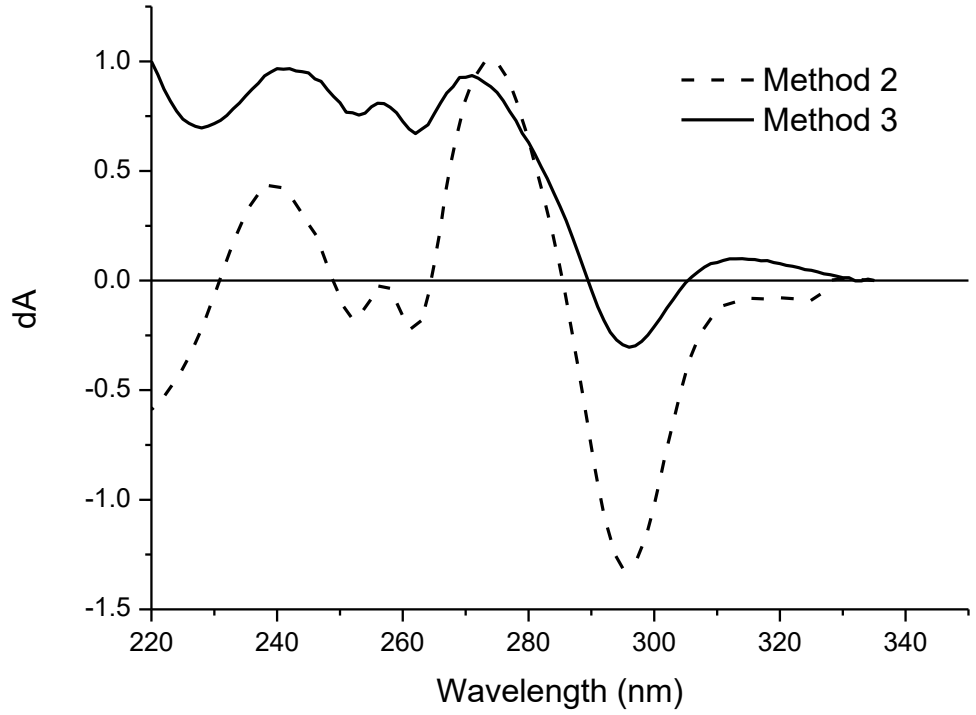


Figure 2.2.7, Thermal difference spectra of hZIP1-1 (5 µM) annealed by METHOD 2 and METHOD 3 in 10 mM Li cacodylate buffer (pH= 7.4) and 100 mM KCl.

Figure 2.2.7 shows the TDS obtained for hZIP1-1 by subtracting the UV absorption spectra recorded at 20 °C from the spectra recorded at 90 °C, the resulting graph displays the characteristic bands and overall shape for a G-quadruplex in the presence of potassium when annealed by both METHOD 2 and METHOD 3. The fact that the shapes are not identical is not surprising given that they are expected to yield different structures. When the sequence was annealed using lithium instead of potassium, the TDS obtained does not show the characteristic G-quadruplex bands. For reference, the TDS with other cations including lithium can be found in the Appendix section A.4.

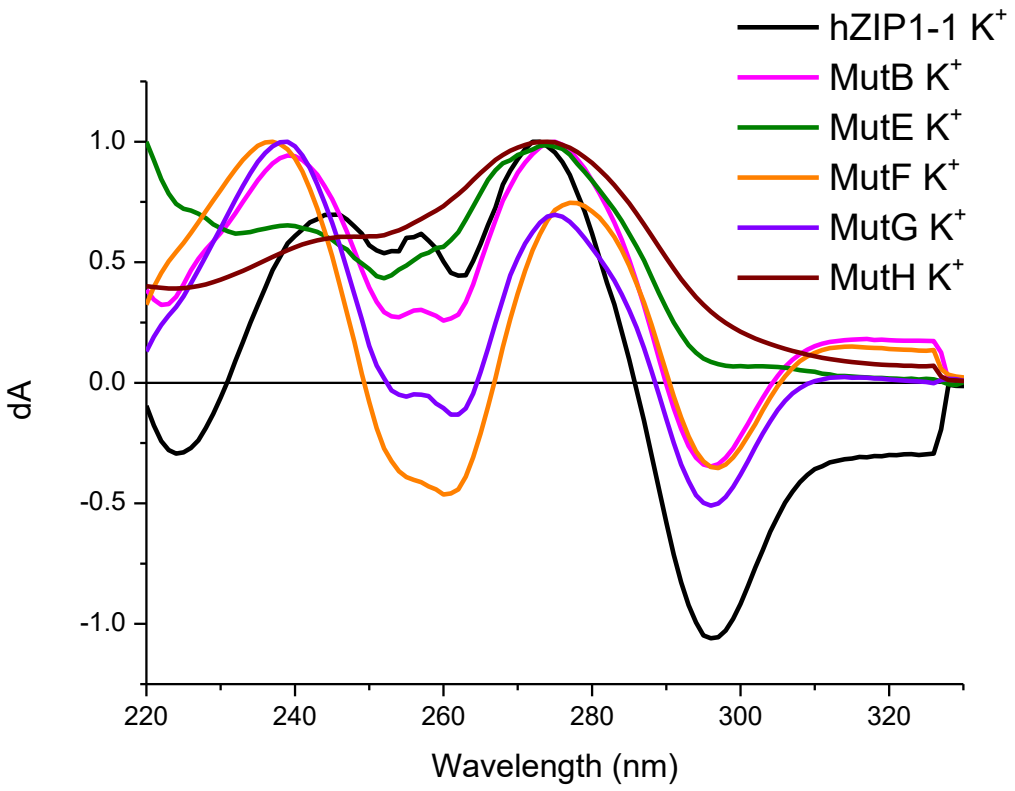


Figure 2.2.8, Thermal difference spectra of mutants MutB, MutE, MutF, MutG and MutH (all 5 µM) annealed by METHOD 2 in 10 mM Li cacodylate buffer (pH=7.4) and 100 mM KCl.

Figure 2.2.8 shows the TDS of the mutants designed for biological studies. It can be seen that MutB, MutF and MutG exhibit the characteristic shape for a G-quadruplex structure, while MutE and MutH do not. No sequence shows this characteristic shape when annealed in LiCl instead of KCl (data shown in the Appendix, Figure A.4.5). Closely related G-quadruplexes usually show similar TDS, which suggests MutB, MutF and MutG do not form structurally similar G-quadruplexes. This is not surprising given that they have major differences in the first two G-runs. For examples of closely related G-quadruplexes that show similar TDS, they can be found in the Appendix section A.4 (hZIP1-2 and Mut1, hZIP1-3 and Mut5, sequences are described in Experimental Procedures). This is evidence that MutE and MutH cannot form intramolecular G-quadruplexes, while MutB, MutF and MutG do.

The polymerase stop assay (PSA) is another established technique to analyse DNA structures *in vitro* and in particular monomeric G-quadruplexes. The technique works by using a small primer annealed to a template that contains the G-quadruplex forming sequence and using Taq DNA polymerase to extend the primer; since Taq DNA polymerase cannot resolve the G-quadruplex the extension will stop at this point giving a shorter sequence.[167]

Briefly, a long template containing either hZIP1-1 sequence (Wild type) or a mutant version was annealed with a FAM-labelled primer overnight in Tris buffer with or without 100 KCl. An extension reaction was set using Taq Polymerase and a buffer either containing KCl or NH_4Cl for half an hour at 37 °C. The reaction was stopped using formamide loading buffer and directly loaded to a denaturing polyacrylamide gel to analyse the products as shown in Figure 2.2.9. Further details on the experiment can be found in Experimental Procedures.

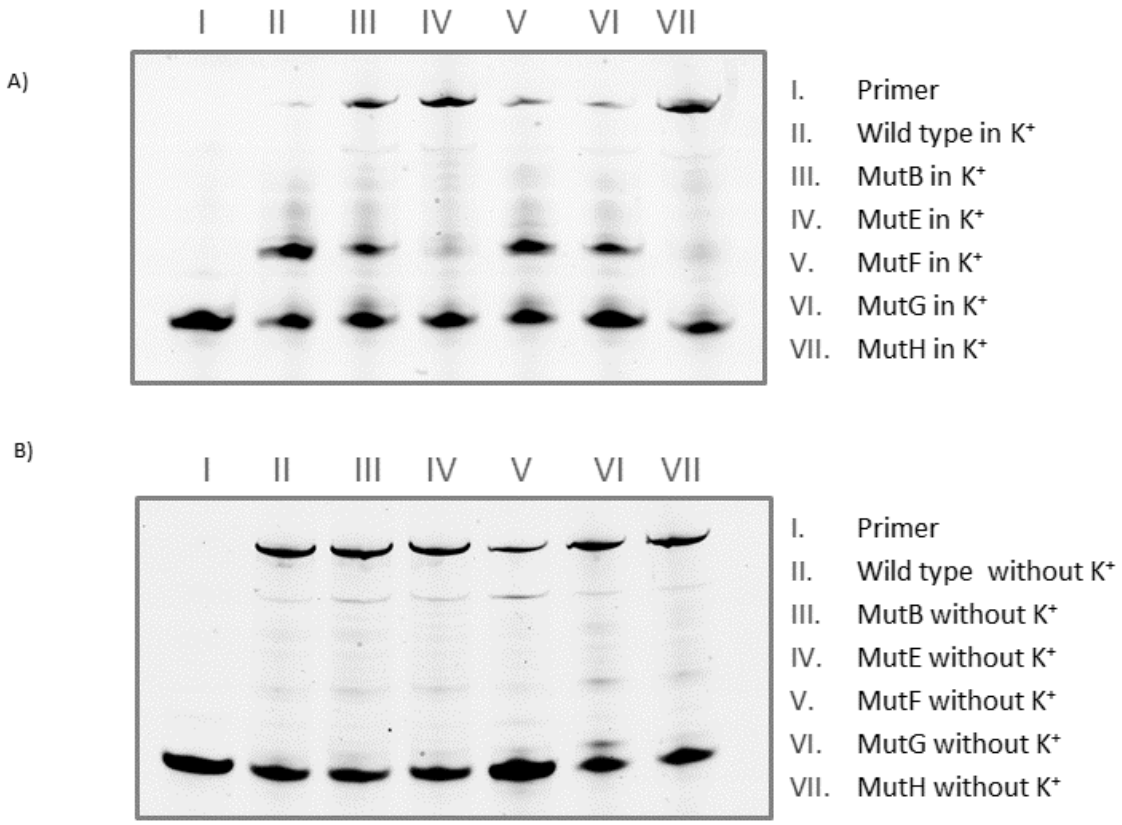


Figure 2.2.9, Denaturing 20% polyacrylamide gel showing the products of the polymerase stop assay for Wild Type (hZIP-1) and mutants MutB, MutE, MutF, MutG and MutH. A) Shows the results of the polymerase reaction in the presence of KCl and lanes II, III, V and VI show incomplete primer elongation. B) Shows the results of the polymerase reaction in the absence of KCl, all lanes show full primer elongation.

In the presence of potassium there is a clear polymerase arrest band for WildType (hZIP1-1), MutB, MutF and MutG but this band is not present for MutE and MutH. All sequences show a complete extension when the reaction is done in the absence of potassium, showing that the polymerase arrest is potassium dependent. This is in agreement with Wild Type, MutB, MutF and MutG forming a monomeric G-quadruplex and MutE and MutH not forming a monomeric G-quadruplex.

2.2.3 CONCLUDING REMARKS

Taking together the results from CD, EMSA, TDS and PSA it can be concluded that it is possible to fold hZIP1-1 into an intramolecular parallel G-quadruplex structure at pH 7.4 and 100 mM KCl by manipulating the annealing conditions to favour this conformation. Using these same conditions it is possible to fold the 4 G-run truncated versions, hZIP1-2 and hZIP1-3, as well as the relevant mutants MutB, MutF and MutG, into intramolecular G-quadruplexes. Meanwhile it is not possible to fold MutE and MutH into intramolecular G-quadruplexes, although *in vitro* they may form higher-order structures.

2.3 THERMODYNAMIC AND KINETIC CHARACTERISTICS

To understand the stability of the hZIP1-1 intramolecular G-quadruplex as well as to learn the effect of the mutations on stability, different techniques were used to study the thermodynamic and kinetic characteristics of hZIP1-1 and mutants MutB, MutF and MutG.

To study the stability of nucleic acids, UV/Vis spectroscopic melting experiments are frequently used by monitoring changes in the peak at 260 nm while increasing the temperature. However, monitoring these peaks is not appropriate for the study of G-quadruplexes given that the change in absorption during melting is rather small;[168] instead it is recommended to follow the 295 nm wavelength during thermal variation to extract thermodynamic parameters like changes in enthalpy (ΔH), entropy (ΔS) and free energy (ΔG).[36] It is also possible to follow the characteristic bands observed in circular dichroism (CD), typically 260 nm or 295 nm depending on the specific structure to extract valuable data, or follow fluorescence if the sequence has been labelled with a donor/acceptor pair relying on the fluorescence quench when the pair is close (fluorescence resonance energy transfer, FRET) due to the G-quadruplex folding, and the increase in fluorescence when the pair is separated when the sequence is melted[169] as shown in Figure 2.3.1.

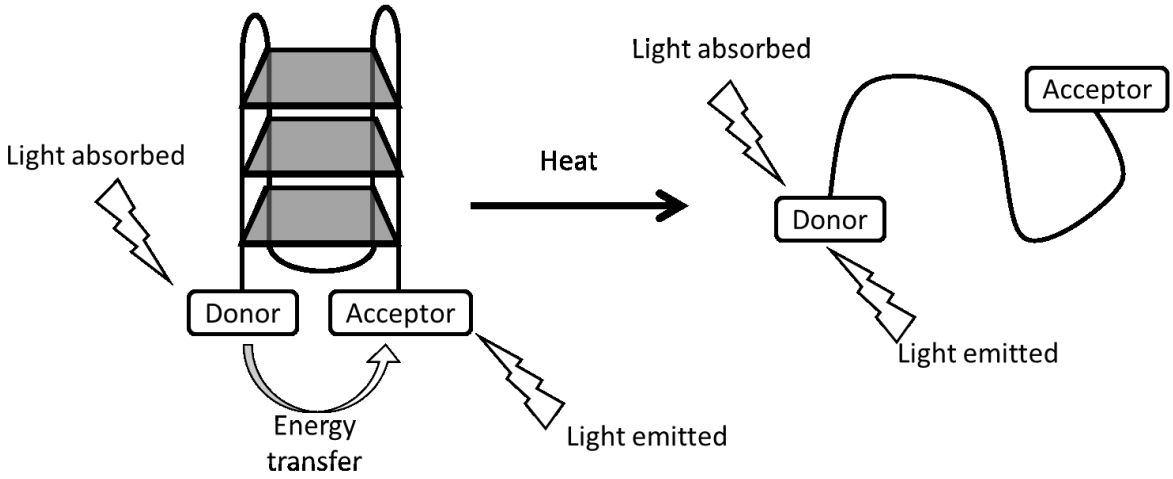


Figure 2.3.1 Schematic representation of a FRET-melting experiment. Initially, the structure is folded which brings the donor and acceptor pairs together, the donor absorbs light and the energy is transferred to the acceptor therefore there is little emission from the donor. Upon thermic denaturalization, there is greater mobility of the DNA strand decreasing the efficiency of the energy transfer between the donor and acceptor and the fluorescence observed from the donor.

Typical analysis of a melting curve would require defining the folded and unfolded baselines and use them to transform the response into a folded fraction, where the value of T_m (temperature of melting

or midstransition) can be then obtained and used for further calculations of thermodynamic properties. Equilibrium curves are reversible melting curves that show identical behaviour when following the changes by increasing the temperature (melting) or reducing it (annealing), and from them T_m can be extracted. A good melting curve will also require enough data points for a folded and unfolded fraction in order to establish good baselines. Non-equilibrium curves, on the other hand, show hysteresis: the curves obtained by melting and annealing are different and neither reflect a true equilibrium curve, therefore they are not reversible, and are caused by differences in the folding and unfolding of the structure which can happen if the rate of the folding/unfolding process is too different to the temperature gradient. This is not unusual for structures with high molecularity since the different folding steps may occur at very different rates.[170] Unimolecular G-quadruplexes often show reversible curves, but many do show hysteresis.[171] It should be noted that the temperature of midtranstion value is only valid when the melting curve is reversible, when the process of annealing and melting give rise to hysteresis it is only possible to obtain $T_{1/2}$ for each curve and its value will likely depend on the melting rate. Non-reversible melting curves have been observed in intermolecular and intramolecular G-quadruplexes, which has been explained as a slow rate in dissociation.[56] The melting temperature, T_m, of a structure gives insight into its stability and it is useful for analysis of different possible folding arrangements and topologies while $T_{1/2}$ gives only a qualitative sense of the stability of the structure. The latter is still useful for relative comparisons of increased or decreased stability due to external stabilizing factors, such as using different cations or stabilizing ligands. In some cases, it is not possible to melt the quadruplex.

2.3.1 MELTING AND ANNEALING CURVES

The sequence hZIP1-1 has been shown to produce higher-order structures when annealed using METHOD 1 and it is not unusual for this type of structures to have very high melting points. Indeed, as shown in Figure 2.3.2, there is very little change in the intensity of the 263 nm peak during a melting experiment. This is not surprising given that it had been observed that higher-order structures can still be seen after heating the sample with denaturing buffers (Figure A.2.1). What was indeed surprising was the melting profile obtained when using hZIP1-1 annealed using METHOD 2 (Figure 2.3.2). As discussed in section 2.2.1, this method yields an enriched monomeric fraction as can be appreciated by its electrophoretic migration and its CD spectrum (which shows a major peak near 263 nm). Interestingly, when this peak was monitored while increasing the temperature, an unusual melting profile was recorded: it showed an unexpected rise in intensity near 50 °C, reached a maximum near 60 °C before decreasing its intensity at higher temperatures. To understand what was causing the

"bump" in the melting profile a second experiment was performed recording a full spectrum at different temperatures throughout the melting experiment, the results can be seen in Figure 2.3.3.

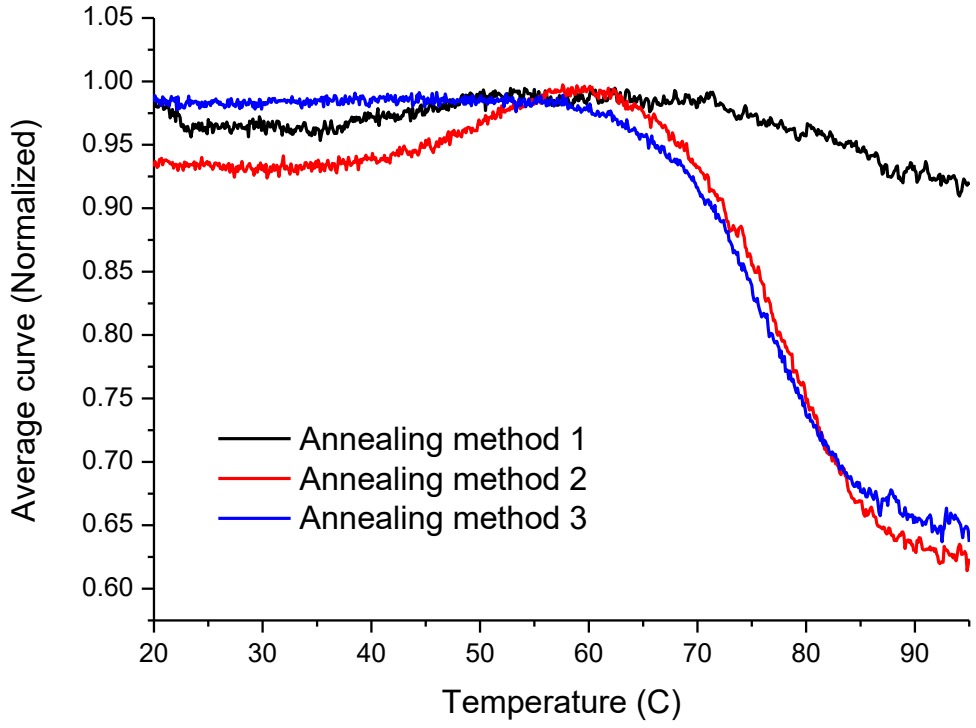


Figure 2.3.2, Circular dichroism melting curves for hZIP1-1 (5 µM) annealed by the three different methods discussed in 10 mM Li cacodylate pH=7.4 and 100 mM KCl. Melting rate is 1 °C/min. Note that the curves have been normalised by dividing by maximum number so they can be presented together.

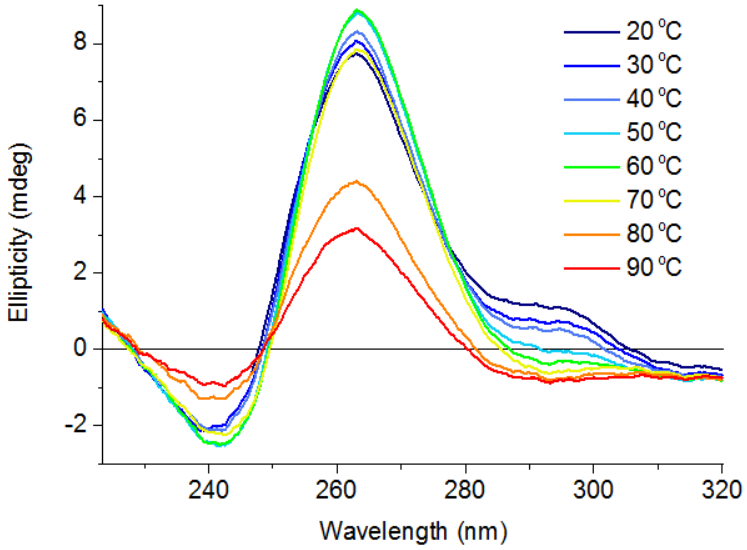


Figure 2.3.3 Circular dichroism spectra for hZIP1-1 (5 µM) annealed by METHOD 2 in 10 mM Li cacodylate pH=7.4 and 100 mM KCl at different temperatures.

In Figure 2.3.3 it is possible to appreciate that the band near 290 nm steadily decreases with increasing temperature. However, the band near 260 nm actually increases and the band at 240 nm decreases with temperature until it reaches 60˚C, giving an overall parallel structure without the 290 nm shoulder. At higher temperatures than 60˚C, the band near 260 nm begins to decrease while the band near 240 nm increases. This could be interpreted as the sequence undergoing a structural transition to a parallel structure at high temperatures. Given that METHOD 2 relies in heating up and cooling rapidly the sequence, it can be argued that the structures are formed under kinetic control and that by heating them up slowly they can transform into a thermodynamically favoured structure, namely a parallel structure. It has been shown by Xue *et al*[61] that the telomeric sequence $G_3(T_2AG_3)_3$ in K^+/PEG will first form an antiparallel structure (3+1) and will slowly transform into a parallel structure within hours. Additionally, the G-quadruplex sequence reported for the hTERT promoter exhibits a very similar CD melting curve,[26] which was explained as an interconversion of a antiparallel (3+1) to parallel structure at similar temperatures. The hypothesis that the sequence folds first into an antiparallel structure and then into a parallel structure is consistent with a multi-step kinetic folding mechanisms proposed in the literature for other G-quadruplexes.[172] If an intramolecular parallel structure is also more thermodynamically stable in the case of hZIP1-1 then it should be possible to anneal the sequence to favour such formation, but this must be done in such a way that it avoids the formation of intermolecular parallel structures that are even more stable as shown in Figure 2.3.2 with annealing METHOD 1. This observation was used to develop METHOD 3 as explained in the previous section, in which the formation of G-quadruplexes is promoted at 60 °C to enrich the parallel monomeric structure.

The melting curve of the sequence annealed by METHOD 3 is shown in Figure 2.3.2, where a typical melting curve is observed. The full spectra of the sequence annealed with this method were recorded at different temperatures (Figure 2.3.4), and there it can be observed that the peak at 263 nm does not increase, and decreases at temperatures above 60 °C. There is not an important band at 290 nm. METHOD 3 has been shown to give an enriched monomeric parallel structure for hZIP1-1 sequence with a typical melting curve, for this reason this annealing methodology was chosen as standard for the thermodynamic characterization of all sequences studied.

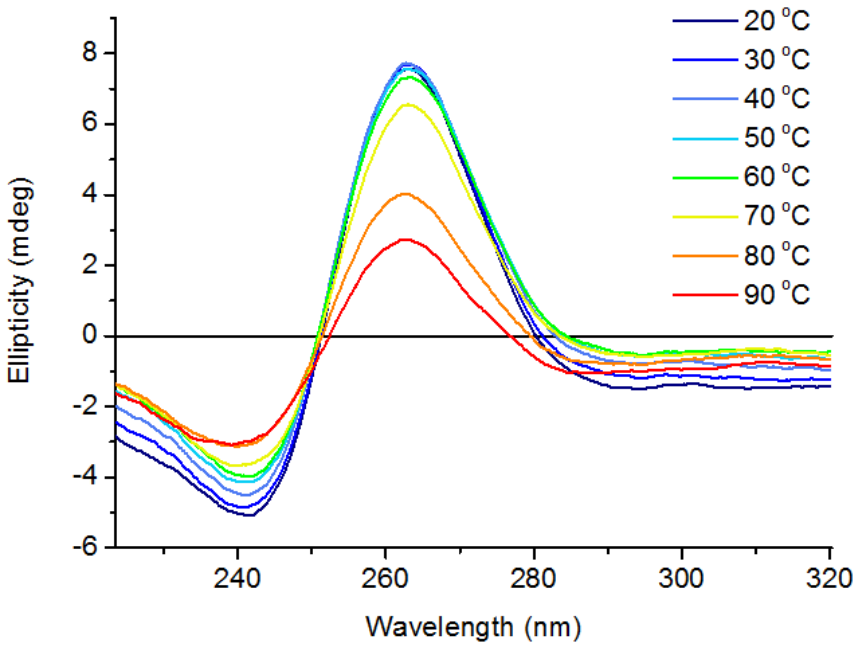


Figure 2.3.4 Circular dichroism spectra for hZIP1-1 (5 µM) annealed by METHOD 3 in 10 mM Li cacodylate pH=7.4 and 100 mM KCl at different temperatures.

The reversibility of the melting curve was studied by recording an annealing curve straight after the melting curve was recorded. The order of the recording is important given that is not possible to start with an annealing curve since it is necessary to fold the sequence in a controlled fashion. Figure 2.3.5 (top) shows the raw data obtained in melting and annealing experiments which seem to show hysteresis as they are not superimposable. Nevertheless, after adjusting the baselines for folded and unfolded stages and using them to convert the values into folded fraction (θ) following the protocol described by Mergny and Lacroix for G-quadruplexes[36](see Figure A.5.1 for baselines and equations), the resulting curves are closer together as shown Figure 2.3.5 (bottom). The temperature of midtransition is calculated as the value when the folded fraction is θ = 0.5. The values obtained for the transition using each gradient are similar, $T_{1/2}$(annealing) = 75.50 ± 0.74 °C and $T_{1/2}$(melting) = 75.09 ± 1.09 °C. This means the process is reversible and can be thermodynamically characterized.

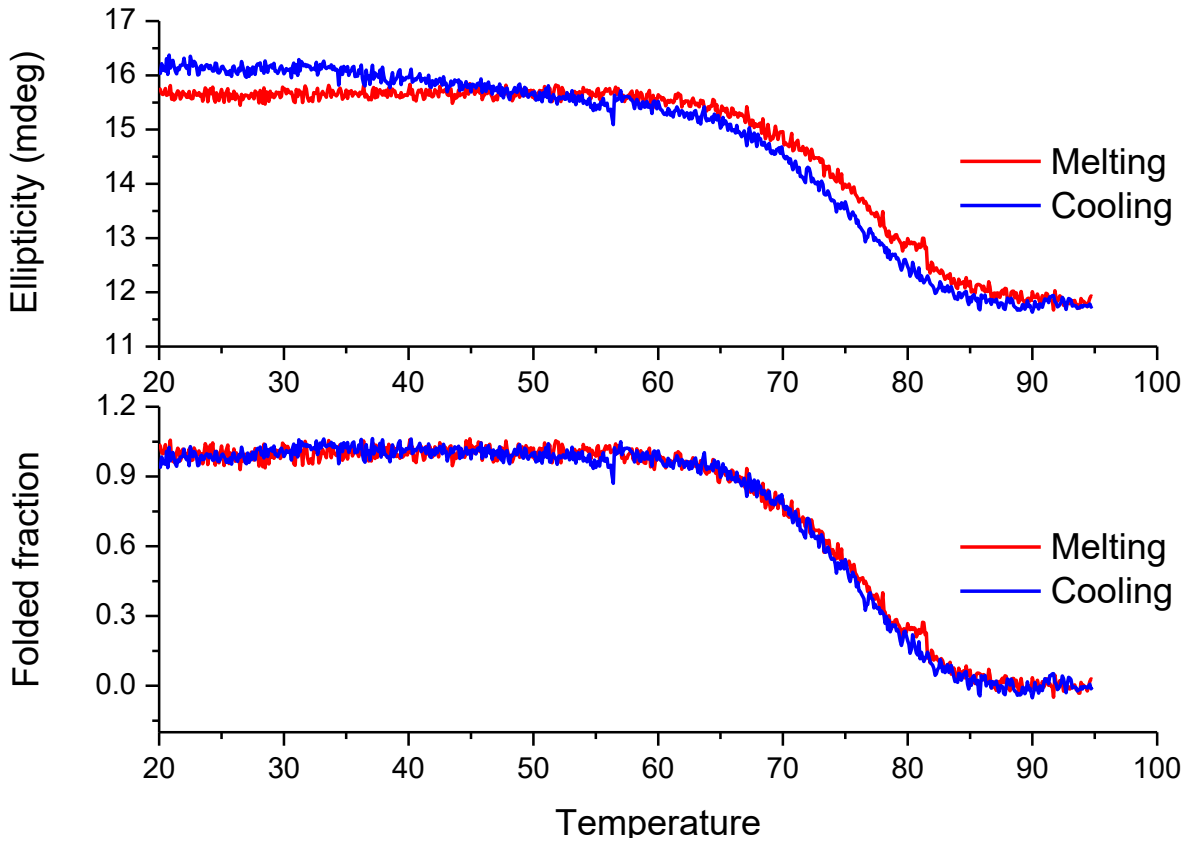


Figure 2.3.5, Circular dichroism melting and annealing curves monitoring the 263 nm peak for hZIP1-1 (5 µM) annealed by METHOD 3 in 10 mM Li cacodylate pH=7.4 and 100 mM KCl using a 1 °C/min temperature gradient. Top, using raw data, bottom, using data transformed to folded fraction.

Melting and annealing experiments using different temperature gradients (0.5 °C/min, 2 °C/min and 5 °C/min) can be found in the Appendix (Figure A.5.2 and Table A.5.1). Slower temperature rates do not show differences to the curves presented at 1 °C/min. Faster temperature rates, however, show hysteresis, with values for the melting curve increasing and for the annealing curve decreasing. This is in agreement with behaviour reported for intramolecular G-quadruplexes[169] which suggest slow kinetics of G-quadruplex association/dissociation.

Truncated versions of hZIP1-1 that have one G-run less on each side and the mutant sequences were also studied. It should be noted that mutants MutE and MutH are not included given that they do not form G-quadruplexes and therefore cannot be melted (data in Appendix, Figure A.5.3). Figure 2.3.6 shows the melting profiles for each of the studied sequences. Deletion or mutation of the first two G-runs has a profound impact on the stability of the structure, while the last G-run has less dramatic relevance when deleted.

hZIP1-1 or WildType	TGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGA
hZIP1-2	------AGGGGGGAGCTATGGGAGGGGGCGGGA
hZIP1-3	TGGGGGAGGGGGGAGCTATGGGAGGGGGC----
MutB	TGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGA
MutF	TGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGA
MutG	TG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGA

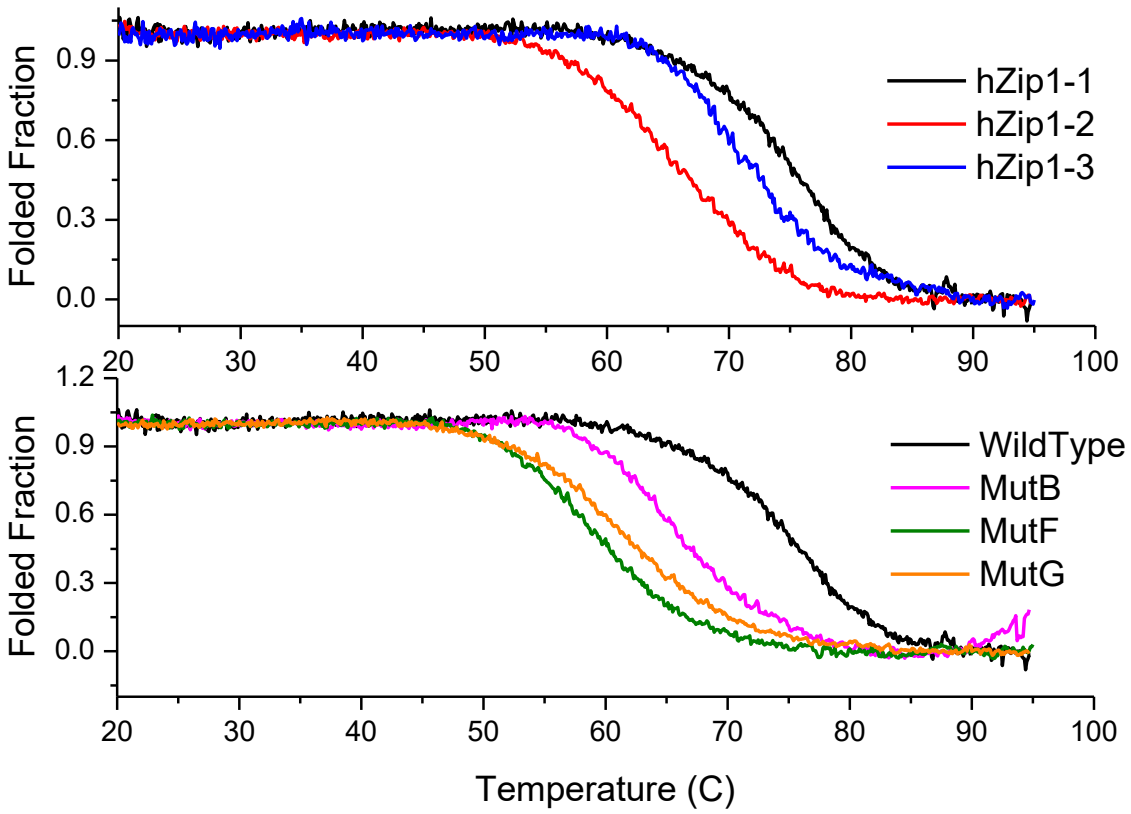


Figure 2.3.6, Circular dichroism melting of hZIP1-1 (WildType), truncated sequences hZIP1-2 and hZIP1-3, and mutants. All sequences (5 µM) were annealed by METHOD 3 in 10 mM Li cacodylate pH=7.4 and 100 mM KCl. Circular dichroism melting shows the monitoring of the 263 nm peak using a 1 °C/min temperature gradient. Top) Truncated sequences. Bottom) Mutants used in biological assays.

FRET melting curves are also widely used for the study of G-quadruplexes and can be recorded using a qPCR machine. A schematic representation of how the technique works can be found in Figure 2.3.1. These curves do not offer structural insight as observed using CD melting curves, however they allow for low concentration measurements, it can be done in high through-put fashion and are very useful for the study of ligand effects (see Chapter 3). Because it is known that the fluorescent labels can affect the stability of the sequence, it is expected that the sequences will show a decrease in the melting temperature value using this technique as compared to the value obtained with CD melting.

The effect of the labels is smaller than placing a flanking base, so the effect of a label next to the flanking base is expected to be small.[42]

Double labelled sequences (5'-FAM and 3'-TAMRA) were used to follow the melting of the structure using FRET. Samples were annealed at 0.4 µM using METHOD 3 and the FAM fluorescence was followed with a 1 °C/min temperature rate for both melting and then annealing. The baseline obtained is variable and therefore needs to be corrected in the same manner that it was done for the melting profiles obtained by CD spectroscopy. An example of baseline correction is shown in the Appendix (Figure A.6.1).

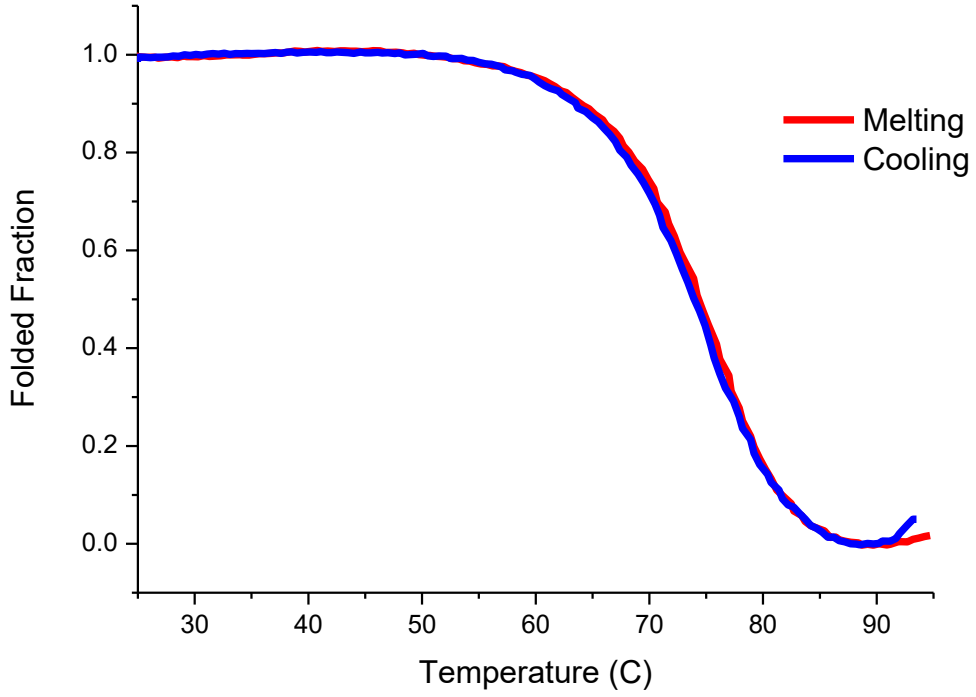


Figure 2.3.7 FRET melting and annealing of 0.4 µM hZIP1-1 annealed by METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl using a 1 °C/min temperature gradient. The data has been baseline corrected to obtain folded fraction.

Figure 2.3.7 above shows that the FRET system has a similar behaviour as the one observed through CD melting, the FRET melting and annealing curve are superimposable with values of $T_{1/2}$(annealing)= 74.27 ± 0.66 °C and $T_{1/2}$(melting)=75.19 ± 1.37 °C. Table 2.3.1 shows the transition temperature values for each curve. The value for melting temperature was calculated as an average of the values obtained for melting and annealing. The melting temperature difference between the value obtained using CD and the value obtained using FRET melting is small ($\Delta T_{1/2}$=0.6 °C), suggesting that there is not a strong destabilization caused by the fluorescent labels. This may be explained by the fact that the labels are not directly attached to the G-quadruplex forming sequence but to a flanking base, which itself may be causing destabilization, reducing the impact of the label.[42] It should be noted that the concentration

of the sample used for CD melting was 5 µM while the concentration used for FRET melting was 0.4 µM. This shows that the melting temperature is independent of concentration of the sample, confirming that there is no change of G-quadruplex molecularity at this concentration range, and the structure studied is monomeric.[36]

To obtain the thermodynamic parameters *ΔH* and *ΔS* using the two-state model, it is necessary to first calculate the equilibrium constant K_a which can be obtained from the folded fraction analysis $K_a=\theta/(1-\theta)$. A van't Hoff analysis is then carried out by plotting $\ln(K_a)$ versus 1/T (in Kelvin) as shown in Figure 2.3.8, the linear range is fitted to a curve and using the gas constant R (R= 8.314 J $K^{-1}mol^{-1}$) the values can be extracted according to:[36]

$$\ln(K_a)=(-\Delta H/R)*1/T+ \Delta S/R$$

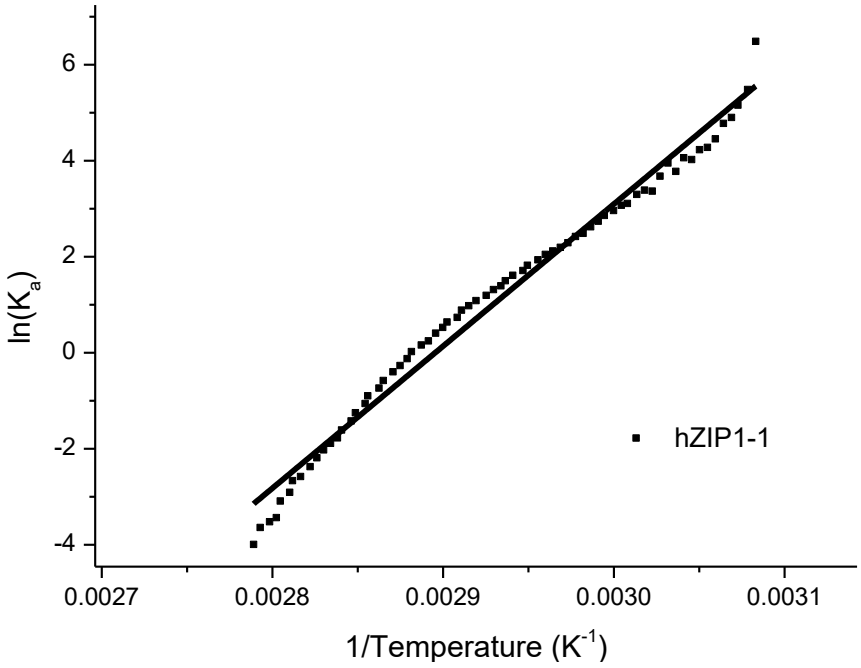


Figure 2.3.8 Arrhenius plot with a fitted linear curve for hZIP1-1 obtained from the FRET melting data shown above. Temperature values 56-84 °C to get linear range (r^2=0.99).

Using van't Hoff analysis the enthalpy values were calculated and entropy values were estimated using $\Delta H/T_m$ as described by Rachwal and Fox.[169] The values obtained are in agreement with those previously reported for other intramolecular G-quadruplex.[39,169]

Table 2.3.1 Temperature of midtransition for hZIP1-1 annealed by METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl and melted at a rate of 1 °C/min or cooled at a rate of -1 °C/min using variable temperature circular dichroism (CD) and FRET melting. Thermodynamic parameter ΔH calculated using van't Hoff analysis and ΔS as $\Delta H/T_m$.

Conditions		$T_{1/2}$ (°C)	T_m(°C)	ΔH (kJ/mol)	ΔH (kJ/mol)	ΔS (kJ/molK)	ΔS (kJ/molK)
CD	melting	75.1±1.1	75.3±0.9	-273.68±16.53	-260.83±24.36	-0.79±0.05	-0.75±0.07
	annealing	75.5±0.7		-247.97±26.74		-0.71±0.08	
FRET	melting	75.2±1.4	74.7±1.1	-236.64±17.49	-231.18±13.10	-0.68±0.05	-0.66±0.04
	annealing	74.3±0.7		-225.72±5.81		-0.65±0.02	

Sequences hZIP1-1, hZIP1-2, hZIP1-3, MutB, MutF and MutG were all annealed using METHOD 3 and 100 mM KCl and melted while following by fluorescence (FRET) and then cooled to obtain melting and annealing curves at 1 °C per minute, they all showed reversibility except for MutG for which hysteresis was observed. Curves can be found in the Appendix (Figure A.6.2). Figure 2.3.9 shows the comparison of the different melting curves for each studied sequence and are in agreement with the curves obtained by CD spectroscopy shown above in Figure 2.3.6.

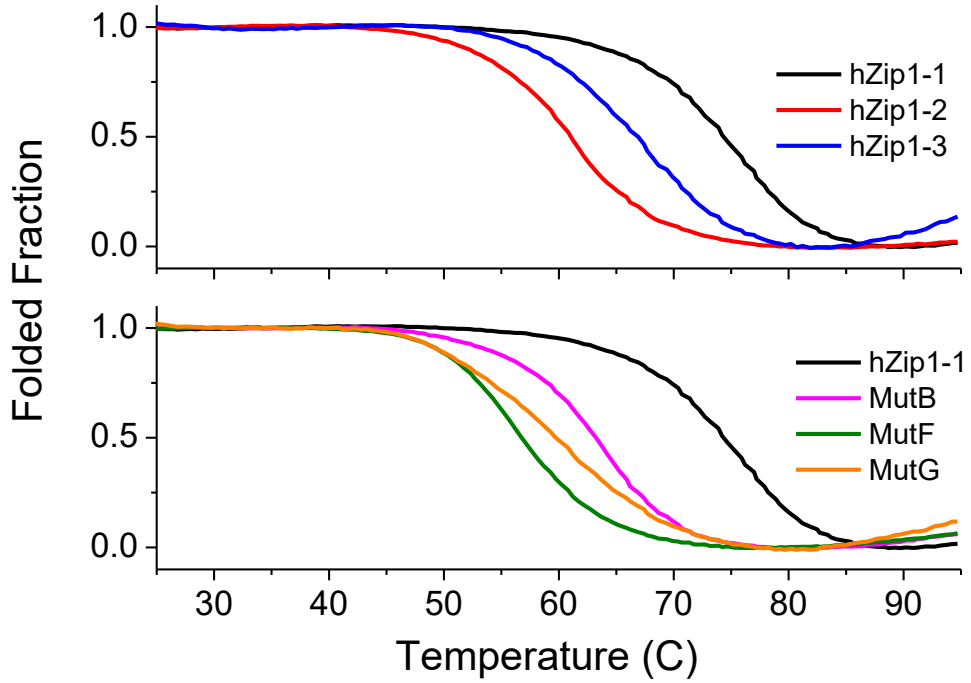


Figure 2.3.9 FRET melting of hZIP1-1 (WildType), truncated sequences hZIP1-2 and hZIP1-3, and mutants. All sequences were annealed at 0.4 µM using METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl. FRET melting shows the monitoring of FAM fluorescence using a 1 °C/min temperature gradient. Top) Truncated sequences. Bottom) Mutants used in biological assays.

The melting temperatures for truncated and mutated sequences are comparable for the melting temperatures obtained using CD spectroscopy, although in all cases they are slightly lower. Given that the behaviour and stability using FRET melting is very similar to the results observed using CD melting, it is concluded that the labelling modifications are not problematic and can be used for analysis of the sequences[38] including ligand effect (see Chapter 3). The melting values obtained are in agreement with reported values for intramolecular parallel structures found in other promoters as shown in Table 2.3.2. Significantly, all melting temperatures are above 37 °Cat physiological pH and potassium concentration.

Table 2.3.2 Different intramolecular parallel G-quadruplexes and the temperature of transition obtained via FRET melting or circular dichroism melting in 10 mM lithium cacodylate buffer (pH=7.2) and 100 mM KCl. Reference is included for values taken from the literature, where (*) indicates this work.

Name	Sequence	$T_{1/2}$ (°C)	Ref.
F-hZIP1-1-T	TGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGA	75.2±1.4	*
F-hZIP1-2-T	------AGGGGGGAGCTATGGGAGGGGGCGGGA	61.1±0.2	*
F-hZIP1-3-T	TGGGGGAGGGGGGAGCTATGGGAGGGGGC----	67.1±0.4	*
F-MutB-T	TGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGA	63.2±0.4	*
F-MutF-T	TGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGA	58.0±2.2	*
F-MutG-T	TG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGA	59.8±0.6	*
FmycT	TTGAGGGTGGGTAGGGTGGGTAA	79.5±0.3	[38]
Fkit1T	GGGAGGGCGCTGGGAGGAGGG	61±1	[38]
FMidG4T	CGGGCGCGGGAGGAAGGGGGCGGGAGC	48.5±0.7	[97]

Using van't Hoff analysis as described above, the thermodynamic parameters of all the sequences studied were estimated using FRET melting curves. A summary of the values calculated is presented in Table 2.3.3. Individual values for melting and annealing curves as well as values obtained using CD melting can be found in the Appendix (Table A.6.1). The values estimated for all sequences are fairly similar, which is perhaps not surprising given the sequences are related. Sequences hZIP1-3 and MutF show a large error. The free energy changes with temperature was calculated using $\Delta G=\Delta H(1-T/T_m)$ for 37°C using the melting temperature and enthalpy values extracted from the melting curves.[173]

Table 2.3.3 Thermodynamic parameters for hZIP1-1, hZIP1-2, hZIP1-3, MutB, MutF, MutG obtained using FRET melting.

Name	T_m(°C)	ΔH (kJ/mol)	ΔS (kJ/molK)	ΔG(37 °C) (kJ/mol)
hZIP1-1	74.7±1.1	-231.18±13.10	-0.66±0.04	-25.07
hZIP1-2	61.3±0.4	-240.98±4.34	-0.72±0.01	-17.54
hZIP1-3	66.4±1.3	-232.91±25.59	-0.69±0.07	-20.16
MutB	63.2±0.2	-238.90±16.06	-0.78±0.11	-18.63
MutF	58.1±1.8	-251.28±32.83	-0.76±0.10	-15.99
MutG	61.4±1.9	-221.97±12.08	-0.66±0.03	-16.17

The van't Hoff model for estimating thermodynamic parameters ΔH and ΔS not only assumes a two state model, it also assumes that the heat capacity of the system is the same both folded and unfolded states, which it has been argued may not be true.[40] The practice of correcting a sloping baseline has also been criticized as potentially oversimplifying the system and occluding small transitions that could happen previous to a complete unfolding of the structure, as well as implying the lack of intermediaries by using a two-state model.[40] It is important then to emphasise that for each sequence the baselines were adjusted by fitting a linear curve using data sufficiently apart from the transition temperature. In the case of hZIP1-1 the selected values that could be used for the folded baseline were from 25 to 50 °C which showed good linearity. Nevertheless, the temperature range for the unfolded baseline is very small given that it needs both to be apart from the transition temperature which is very high (74 °C) but not too close to the highest temperature values (95 °C) given that there may be some evaporation and fluorescence can exhibit temperature-dependence behaviour at very high temperatures. The data presented above is similar to values found in the literature using the same analysis, so while the model used may not be without drawbacks it is considered useful to compare the stability of the folded sequences among themselves and with other reported G-quadruplexes.

Given that the sequences under study form intramolecular G-quadruplexes, and that they contain an uneven number of guanines in each G-run, they are expected to be highly polymorphic. The annealing process aims to produce an enriched population of parallel folded G-quadruplexes as observed by CD spectroscopy, however there is still a significant amount of potential different parallel folding that can be adopted depending on which of the guanines are involved in the tetrad and which ones are involved in the loops. The distribution of these populations cannot be followed by CD or FRET, so the work has been done under the assumption that by annealing the sequences always under the same conditions

the distribution of the folding would be similar on each occasion. Some of the melting temperature values show a larger error, in particular MutF and MutG. It is possible that despite efforts to guarantee identical conditions during annealing, the folding distribution is not similar each time which may explain small differences of melting temperatures.

From the truncated sequences for hZIP1-1, it is observed that hZIP1-2 has the most different melting point showing an important destabilization (13°C) by eliminating the first G-run, while hZIP1-3 shows a less drastic decrease in stabilization (7.5°C) by eliminating the last G-run. This may be explained by the fact that hZIP1-3 contains larger G-runs than hZIP1-2. From the designed mutants, MutF shows the lowest melting point with a difference of 17°C between the mutant and the wild type sequence hZIP1-1. This is perhaps because of the large loop between the first and the second G-runs of 13 bases. The drop in melting temperature from hZIP1-1 to MutB and MutG is considerable (8°C and 13°C) but not as large as MutF. MutG can be thought of as hZIP1-2 with a large 5'- overhang which would usually destabilize the sequence,[37] nevertheless the decrease in melting temperature is small (1°C), which suggest the guanines in the overhang may be preventing this effect.

2.3.2 Double strand competition

If the studied sequence, hZIP1-1, was able to form a G-quadruplex in the promoter region of *hZIP1 in vivo*, then it could only possible to be folded as such when the double helix was opened either by enzymes or by negative supercoiling as discussed in section 1.1.3. The formation of the G-quadruplex would be there in direct competition with the formation of the double helix with opposite complementary strand. Therefore, understanding the kinetics of unfolding is important, especially to compare the kinetic stability of the different sequences to be used in the biological assays. Intramolecular G-quadruplexes are known to unfold slowly, although not as slowly as intermolecular G-quadruplexes, meaning that it is possible to follow the unfolding using methods such as FRET. This will not, however, give information regarding possible intermediary stages.

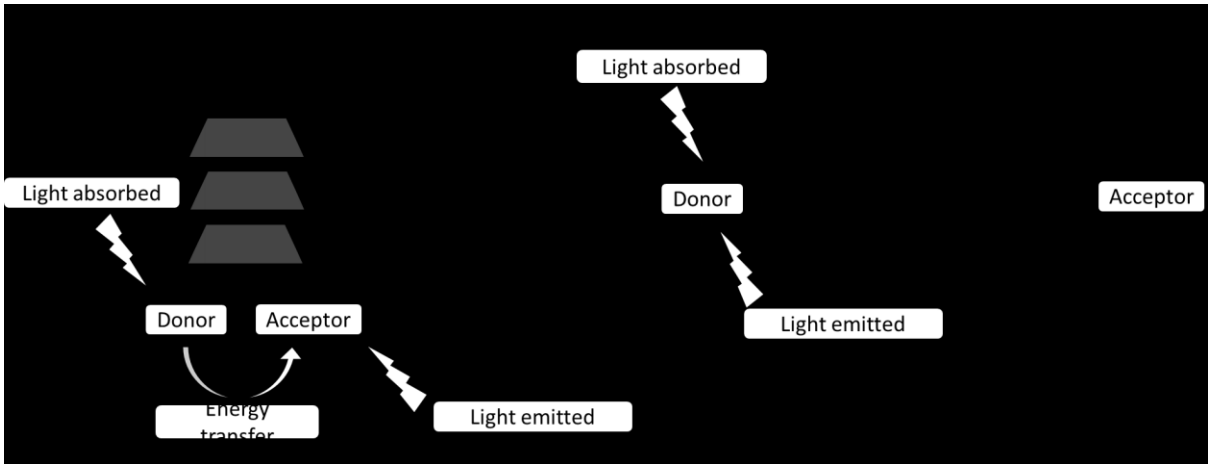


Figure 2.3.10 Schematic representation of the fluorescence change caused by the G-rich strand being trapped after the opening of the G-quadruplex. Light is absorbed by the donor and the energy is transferred to the acceptor and no fluorescence from the donor is observed. After the G-quadruplex is opened and trapped by the complementary strand, fluorescence from the donor can be observed.

To study the unfolding of the sequences, the double labelled sequences were annealed at a strand concentration of 0.4 µM with 100 mM KCl using METHOD 3. They were subsequently mixed with their reverse complement at a strand concentration of 40 µM and placed in the real-time PCR machine in order to follow the fluorescence change caused by the separation of the labels because of the double helix formation, as represented on Figure 2.3.10. When the G-quadruplex is folded, the fluorescence is quenched, so the reverse complementary strand is used to trap any unfolded sequence into a duplex where the donor and acceptor are separated. The concentration of the complementary strand is 100-fold higher than that of the G-quadruplex to make the unfolding an apparent first order process.[174] The unfolding was followed every 30 seconds for two hours. While the sequences were annealed with 100 mM KCl, the final concentration of unfolding reaction was 80 mM KCl. Results are shown in Figure 2.3.11.

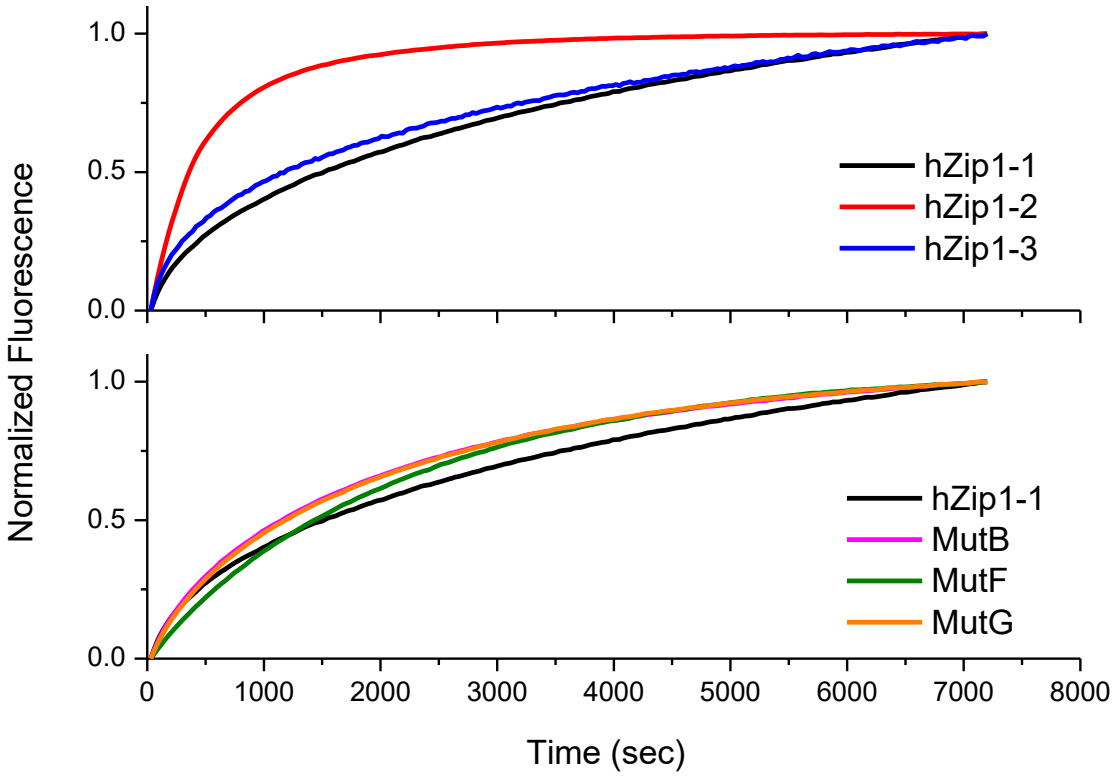


Figure 2.3.11 Unfolding of sequences hZIP1-1, hZIP1-2, hZIP1-3, MutB, MutF, MutG followed by FRET using competitor strand as trap. Samples (0.4 μM) were annealed METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl and mixed with complementary strand (40 μM). Final total potassium concentration of 80 mM.

Interestingly, hZIP1-2 unfolds much quicker than any of the other sequences studied. It is particularly interesting that MutG does not unfold as fast as or faster than hZIP1-2, given that MutG contains the same G-runs than hZIP1-2 plus an overhang that contains guanine. Studies using well known G-quadruplexes with or without an overhang have found that the presence of a single stranded tail next to a G-quadruplex accelerates the opening of the G-quadruplex,[43] which suggests that the overhang is not a single strand. This would be in agreement with the observation that the melting temperature for MutG is not significantly lower than for hZIP1-2, as expected for destabilizing overhangs. While all mutants seem to unfold more quickly than hZIP1-1, the difference is not very big.

The unfolding experiment was repeated with a smaller quantity of potassium, the samples were annealed at 10 mM potassium and 90 mM lithium, and after mixing with the reverse complementary sequence the cation concentration was kept constant. Results are shown in Figure 2.3.12.

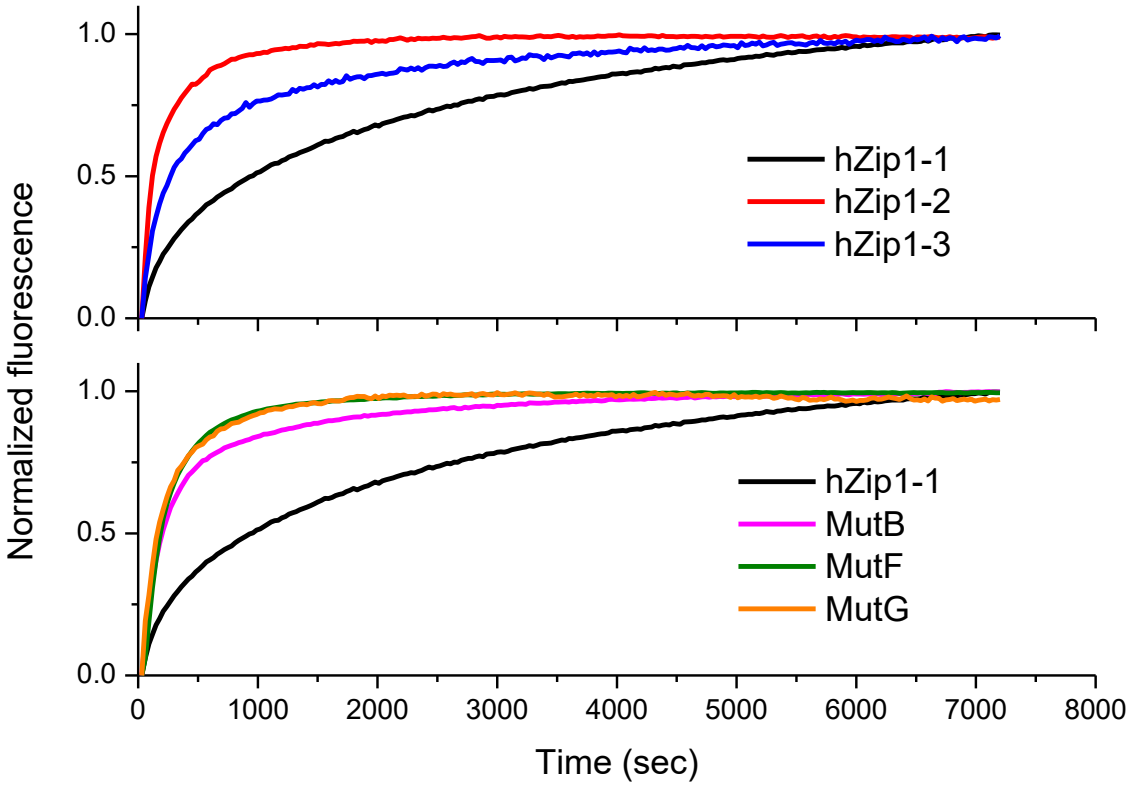


Figure 2.3.12 Unfolding of sequences hZIP1-1, hZIP1-2, hZIP1-3, MutB, MutF, MutG followed by FRET using competitor strand as trap. Samples (0.4 µM) were annealed METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 10 mM KCl and 90 mM LiCl, then mixed with complementary strand (40 µM). Final total potassium concentration of 10 mM.

In the reduced potassium conditions, the opening of the G-quadruplex is faster. The trend of the rate of unfolding is similar to the one observed by melting experiments.

While it is not possible to study the thermodynamic properties of the G-quadruplex of hZIP1-1 annealed by METHOD 2 because of its change of conformation at increased temperatures, it is nevertheless possible to use it for duplex competition. The unfolding of the G-quadruplex was monitored for both methods at high potassium concentrations shortly after annealing, and the result is shown in Figure 2.3.13.

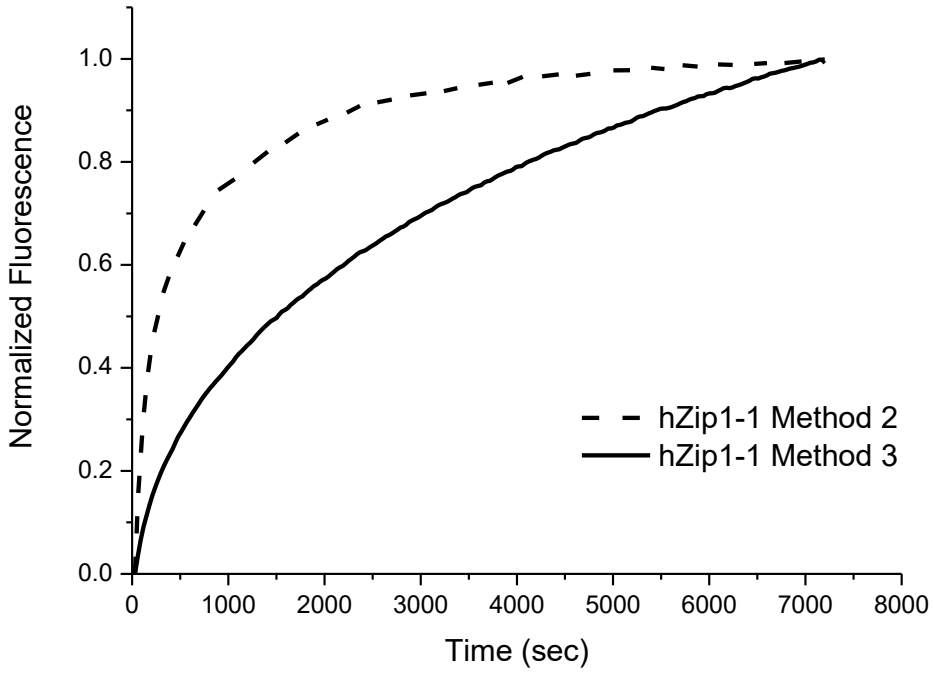


Figure 2.3.13 Unfolding of sequence hZIP1-1 (0.4 µM) when annealed using METHOD 2 or METHOD 3 in 10 mM Li cacodylate buffer pH=7.4, mixed with 40 µM complenetary strand. Final total potassium concentration of 80 mM.

Interestingly, it was observed that if the sample was used for the assay shortly after annealing using METHOD 2 or after a longer period at room temperature, the unfolding of the sequence was slower. The structure seems to then convert over time to a parallel structure in a similar fashion as it was observed during melting (Figure 2.3.3); to show this, a sample was prepared using METHOD 2 and followed by CD spectroscopy over time as shown in Figure 2.3.14.

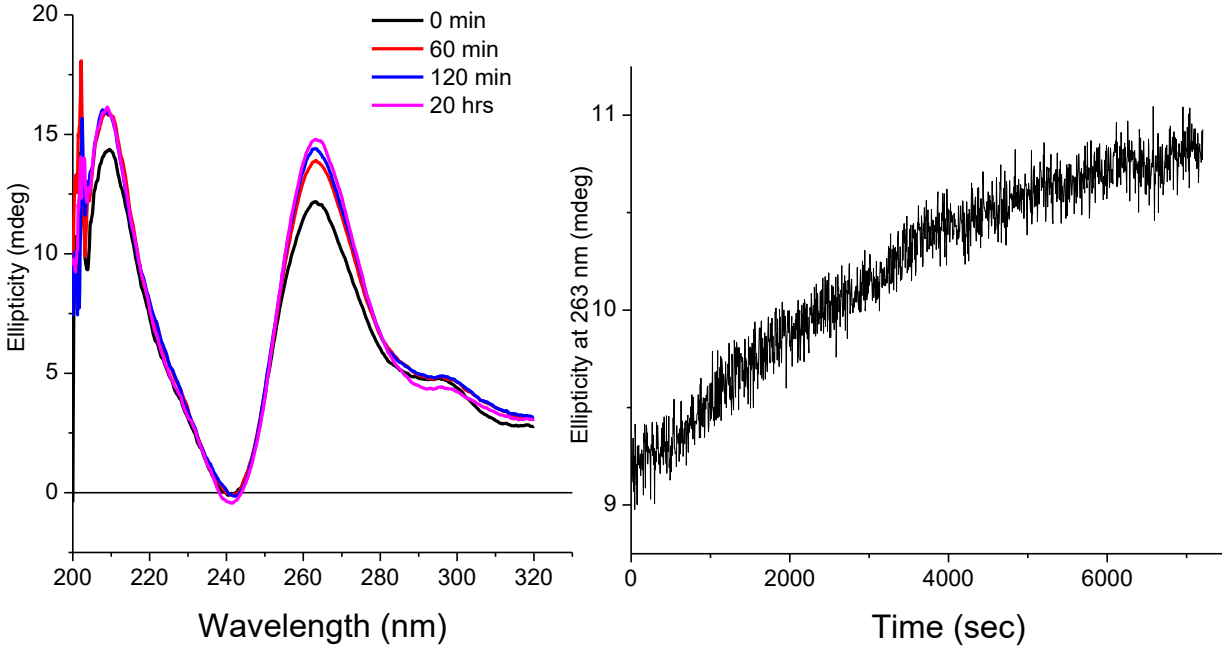


Figure 2.3.14 Left) Full circular dichroism spectra of a hZIP1-1 (5 µM) sample annealed by METHOD 2 in 50 mM Tris buffer pH=7.4 and 100 mM KCl kept at 37 °C at different time points. Note that time 0 refers to the start time it was brought to 37°C. Right) Monitoring the change in intensity of the 263 nm peak over two hours.

Indeed, it can be observed that the peak at 263 nm increases over time, while the peak at 240 nm slightly decreases. This could indicate the formation of a parallel structure. The interconversion of the antiparallel structure to a parallel structure over time and promoted by an increase of temperature is in agreement with the antiparallel structure having a faster association/dissociation rate than the parallel structure.

2.3.3 CONCLUDING REMARKS

Melting temperatures, enthalpy and entropy values obtained for all sequences are similar to the values found in the literature for other quadruplex-forming sequences. Using free energy (ΔG) at physiological temperature as a comparative value, the wild type sequence hZIP1-1 is the most stable of all sequences studied. Mutants MutB, MutF and MutG have very similar stability between them. The second more stable sequence is hZIP-3; while hZIP1-2 and the mutants show similar stability, meaning that alterations in the first and second G-run are the most detrimental to stability of the sequences analysed.

In agreement with the thermodynamic behaviour, hZIP1-1 unfolds the slowest from all sequences while MutB, MutF and MutG unfold at a similar rate. It can be concluded that by mutating the sequence there is a decrease in stability in the wild type intramolecular G-quadruplex; however, all the mutated versions have similar characteristics, therefore if one of the mutated sequences could form a G-quadruplex *in vivo* it is reasonable to expect all mutants can. Therefore, differences between promoter activities among these mutants cannot be explained in terms of G-quadruplex stability.

The data also suggests that not all possible conformations for hZIP1-1 are equally stable: the parallel intramolecular conformation is the most stable, and given enough temperature or time, other populations formed by kinetic trapping will rearrange into an intramolecular parallel topology.

2.4 SUMMARY, DISCUSSION AND CONCLUSIONS

There is a G-quadruplex potential sequence located in the promoter region of *hZIP1* which overlaps an important transcription factor (SP1) binding site as well as two non-important binding sites. This transcription factor is known to be able to recognize G-quadruplex sequences.[73] This sequence has been considered relevant to study in the context of G-quadruplex mediated gene expression and named hZIP1-1. In order to study the characteristics of the potential G-quadruplex formed by this sequence, it was necessary to assess if it forms said structure at physiological conditions of potassium and pH. Different annealing procedures were studied in order to fold the sequence of interest into a unimolecular G-quadruplex. Molecularity was followed using EMSA; a strong emphasis was put on preventing multimolecular G-quadruplex formation as they would not have relevant biological role in the promoter region of the gene.

In order to better understand the role of this structure in the promoter region, mutants were designed to disrupt G-quadruplex formation while disrupting only unimportant binding to the best of our knowledge, as well as control mutation that do not disrupt the formation of the G-quadruplex completely but would function as mutation controls for biological experiments. The studied sequences are shown in Table 2.4.1.

Table 2.4.1 Summary of the sequences hZIP1-1 and mutants studied. "G4-DNA?" refers to the ability of the sequence to form intramolecular G-quadruplexes.

Name	Sequence	G4-DNA?
hZIP1-1	TGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGA	Yes
MutB	TGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGA	Yes
MutE	TG**tttt**AGG**ttt**GAGCTATGGGAGGGGGCGGGA	No
MutF	TGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGA	Yes
MutG	TG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGA	Yes
MutH	TG**tt**GGAGG**ttt**GAGCTATGGGAGGGGGCGGGA	No

Using a standard annealing method (METHOD 1) of heating up the sample and cool down slowly overnight in the presence of potassium yields a large population of intermolecular G-quadruplexes, which are slow to fold but have a remarkable thermodynamic stability, as shown by the fact that no transition can be observed when the sequence is heated up and monitored by CD (Figure 2.3.2). In

order to enhance the formation of unimolecular structures, the sample was heated up and cooled down rapidly and afterwards the potassium was added (METHOD 2) in order to trap the unimolecular folding that would be kinetically favoured. This technique indeed favoured the formation of a unimolecular fraction as shown using EMSA, and gave a CD spectrum reminiscence of a parallel population with an antiparallel contribution (negative 240 nm, positive 260 nm and positive 290 nm) (Figures 2.2.4 and 2.2.5). This was not surprising given that the sequence has five G-runs of different sizes, so a high level of polymorphism was expected. However, when the sequence was heated up in order to melt it, it was observed that the peak near 290 nm decreased while the peak near 263 nm increased, suggesting an increase in the parallel contribution and a decrease in the antiparallel contribution. This could be explained as the temperature increases, the antiparallel structures present converted into a more stable parallel structure before melting (Figure 2.3.3). There are other promoter G-quadruplexes in the literature reported to show this kind of behaviour like hTERT[26] and PALB2[94] as well as the well-studied telomeric G-quadruplex.[25,61] Crucially, hTERT can form two solved structures, 3+1 and parallel, that when coexisting in solution they show a highly similar CD spectrum to the one reported here as well as a highly similar melting curve. A parallel bimolecular G-quadruplex could also explain the CD shape observed, as there is a dimeric band when the sample is analysed by EMSA. Bimolecular G-quadruplexes, however, are expected to have fast association (compared to tetramolecular) but very slow dissociation kinetics and higher melting points than their unimolecular counterparts[36,19] which means that while the dimeric structure may be part of the overall CD signature observed, it is unlikely to undergo a transition at 60 °C. Furthermore, when MutE is annealed using METHOD 3 it also shows an important dimeric band (Figure 2.2.6 top), however no transition is observed when the sample is heated up (Figure A.5.3). In conclusion, several topologies coexist in solution when hZIP1-1 is annealed using METHOD 2 including parallel and antiparallel unimolecular G-quadruplex populations. The unimolecular antiparallel fraction undergoes a structural transition when the temperature is increased and will re-fold into a parallel configuration. At lower temperatures, this rearrangement will also occur over long periods of time (several hours).

In order to avoid the structural interconversion, a third annealing method[175] was investigated (METHOD 3) in which the sequences was first heated up to 95°C and cooled down to 60°C, a temperature observed to have full parallel conversion before melting, and potassium is added at this point. This annealing method yield a unimolecular G-quadruplex as appreciated by EMSA with a CD spectrum similar to parallel structures, suggesting the annealing method was successful in folding the sequence into a parallel unimolecular G-quadruplex. When this sequence is heated up it shows high stability and melting temperature of 75°C. Using this annealing methodology, wild type sequence hZIP1-1 and all mutant sequences were folded into structures that show the characteristic CD peaks for parallel G-

quadruplex, including the sequences that not expected form intramolecular G-quadruplexes MutE and MutH, which are then explained as multimeric G4-DNA, possibly bimolecular (Figures 2.2.6). When all sequences were heated up and the characteristic peak 263 nm followed using CD spectroscopy, hZIP1-1, MutB, MutF and MutG all show a typical melting curve but MutE and MutH do not show any transition (Figure A.5.3). This further suggests that the observed peaks be may be caused by an intermolecular G-qudruplex that the sequences could form given that they are G-rich.

To further characterize that hZIP1-1 forms intramolecular G4-DNA under physiological conditions TDS and PSA were also used. The TDS obtained for any of the sequences studied when annealed using METHOD 3 in the presence of 100 mM LiCl does not resemble the fingerprint shape that identifies G-quadruplexes (Figure A.4.5). However, when this experiment was reapeated using 100 mM of KCl instead of LiCl for all sequences, the typical G-quadruplex fingerprint shape was observed for hZIP1-1, MutB, MutF and MutG and not for MutE and MutH (Figure 2.2.8). This is strong evidence that form G-quadruplexes while MutE and MutH do not. This observation is in agreement with MutE and MutH not showing a melting transition.

Finally, PSA was used to analyse the intramolecular forming potential of the sequences. This technique requires annealing a primer to the 3'- end for polymerase extension. The extension reaction is halted in the presence of secondary structures, which can be appreciated in the presence of potassium for hZIP1-1 (WildType), MutB, MutF and MutG but not for MutE or MutH (Figure 2.2.9). Furthermore, when the reaction was carried out in the absence of potassium, no arrest is observed for any of the sequences implying the secondary structure the polymerase cannot resolved is potassium dependent.

Considering all the evidence presented (CD, EMSA, TDS and PSA), it is concluded that the putative sequence found in the promoter region of *hZIP1*, hZIP1-1, can form parallel intramolecular G-quadruplexes at physiological conditions of pH and potassium concentration using annealing METHOD 3. Furthermore, the mutants designed as controls to not disrupt intramolecular G-quadruplex formation, MutB, MutF and MutG, do not indeed disrupt the G4-potential. Finally, the mutants designed to fully disrupt the G4-DNA potential, MutE and MutH, indeed are not able to form such structure.

It was predicted using transcription factor consensus binding sites analysis that the promoter activity of all mutants (MutB, MutE, MutF, MutG and MutH) will be similar because their mutations disrupt the same unimportant binding sites, while none of the mutations creates a new binding site. Therefore, if there is a promoter activity difference between MutE/MutH in comparison with MutB/MutF/MutG, it could be rationalize that the changes are due to their differences in G-

quadruplex forming potential. However, MutB, MutG and MutF contain different mutations expected to render structurally different G-quadruplexes (MutF contains a very long loop between the first and the second G-run, while MutG contains a long overhang before the first G-run). These differences could have a profound effect on the G-quadruplex stability and affect their relative viability *in vivo*.

To study the stability of the different G-quadruplex forming sequences, the response to increase temperature was followed by both CD and FRET. All sequences were annealed using METHOD 3, the melting and annealing was monitored and the curves were analysed. The transition temperature was extracted to be used as an estimation of the stability of each sequence. This showed that the stability of all mutants was lower than for hZIP1-1, and MutF was the least stable of all, presumably because of the long loop between the first and the second G-run. However, the thermodynamic stability of MutB, MutF and MutG among themselves was fairly similar.

The formation of structures *in vivo* is unlikely to be driven thermodynamically, so the kinetic stability of the sequences was also studied by following the unfolding of the G-quadruplex using FRET. A similar trend of kinetic stability was observed as with thermodynamic stability, with all mutants unfolding quicker than the WildType hZIP1-1, yet similarly among them.

Finally, the study of hZIP1-1 through different methodologies revealed that it can form both parallel and antiparallel intramolecular structures. The antiparallel structure, however, is much less stable that the parallel counterpart and will unfold quickly in the presence of the complementary strand. This means that if a parallel structure is formed in the promoter region of *hZIP1* it can be expected to be very stable, however should an antiparallel structure form first (as an intermediary to a parallel structure, for example), it would be unfavourable and it could be expected to be unstable.

CHAPTER 3: G-QUADRUPLEX STABILIZING LIGANDS AND THEIR EFFECT ON HZIP1

In this chapter, different well characterized G-quadruplex interacting ligands are introduced with emphasis on TMPyP4, BRACO-19, Pyridostatin and PhenDC3. Then these ligands are tested against the sequence of interest, hZIP1-1, and its mutants to understand the effect they have on its stability. These ligands are potential candidates to be used in further studies regarding the biological role of the G-quadruplex in the *hZIP1* promoter region. To ensure there is a good selectivity of the ligands for the sequences involved, a non-quadruplex sequence was used as control.

3.1 LIGAND STABILISATION

3.1.1 SMALL MOLECULES AND G-QUADRUPLEXES

G-quadruplex structures have captured a lot of attention as potential targets for new therapeutic strategies against different types of cancer, both in the telomere and the promoter region of key genes. Targeting and stabilizing G4-DNA with small molecule ligands has been a proposed strategy to interfere with biological pathways such as telomeric extension and silencing the transcription of oncogenes.[176] Ligands have been shown to often interact with a G-quadruplex structure by end-stacking and have an effect on the stability and kinetics of the structure; some have even been proposed to favour one topology over another.[31] Stacking occurs via π-π and electrostatic interactions between the ligand and the G-quadruplex structure. For this reason, stabilizing molecules often feature a large flat aromatic surface which improves selectivity over other DNA forms, as well as side arms chains that improve solubility and may interact with the negatively charged backbone. Often, researchers[177] may exploit the electrostatic interactions by: using amine arms that can be protonated *in situ* (like BRACO-19[178]), use N-methylated ligands that contain quaternary aromatic nitrogen (like TMPyP4[179] and PhenDC3[180]), or metal complexes (like Ni-salphen[181]). Successful ligands, however, may also be non-cationic (like telomestatin[182]). Interaction between the ligand and the DNA may also depend on the binding environment. While the ability of cationic ligands to inhibit telomerase activity has been shown to be diminished under molecular crowding conditions, the ability for anionic ligands was not affected.[183]

The neutral ligands include one of the most selective ligands for G-quadruplexes, telomestatin.[182] It is a natural product from *Streptomyces anulatus* with an 8-ring macroheterocycle structure as shown in

Figure 3.1.1, which has drawn a lot of attention due to its high affinity and selectivity for intramolecular G-quadruplexes[184] and its ability to inhibit telomerase.[182] It has shown great potential for its antiproliferative activity in cancerous cell lines while having little effect in normal cell lines.[185] However, the total synthesis of this molecule involves several steps, making its large scale production difficult.[186] One solution to circumvent this problem has been to synthetize chemically related compounds that have related scaffolds: they allow for further investigation of the activity of G-quadruplex ligands *in vivo* and are modifiable to increase water solubility.[185] One of these derivatives was used recently to investigate the interaction of this scaffold with telomeric G-quadruplex DNA in solution.[187]

Some examples of non-neutral ligands that interact with G-quadruplexes include berberine and N-methyl mesoporphyrin IX (NMM). Berberine is a cationic natural product that interacts with DNA, including G-quadruplexes;[188] NMM is a non-planar derivative of porphyrin negatively charged at physiological pH known to also interact with G4-DNA.[189] Their structures are shown in Figure 3.1.1. While berberine shows both low affinity and selectivity for G-quadruplexes, NMM shows low affinity but great selectivity for these structures[190] and can therefore selectively inhibit the unwinding of G4-DNA substrates by RecQ helicase, but not so for duplex substrates.[66] What is more, it has a good selectivity for the parallel conformation and will only bind to this topology.[191]

Metal complexes have also been used as G-quadruplex ligands, which usually use an aromatic planar scaffold with the added advantage of a cation in the centre.[192] The geometry of the metal complexes as well as nature of the substituents affect binding.[193] An example of this is the square planar Ni(II) salphen complex (Ni-Salphen), structure shown in Figure 3.1.1, which shows a very high stabilizing capacity for G4-DNA. Platinum(II) complexes are often emissive (especially when bound to DNA), which has been used as a tool to investigate the cellular uptake and localisation of the ligands.[193]

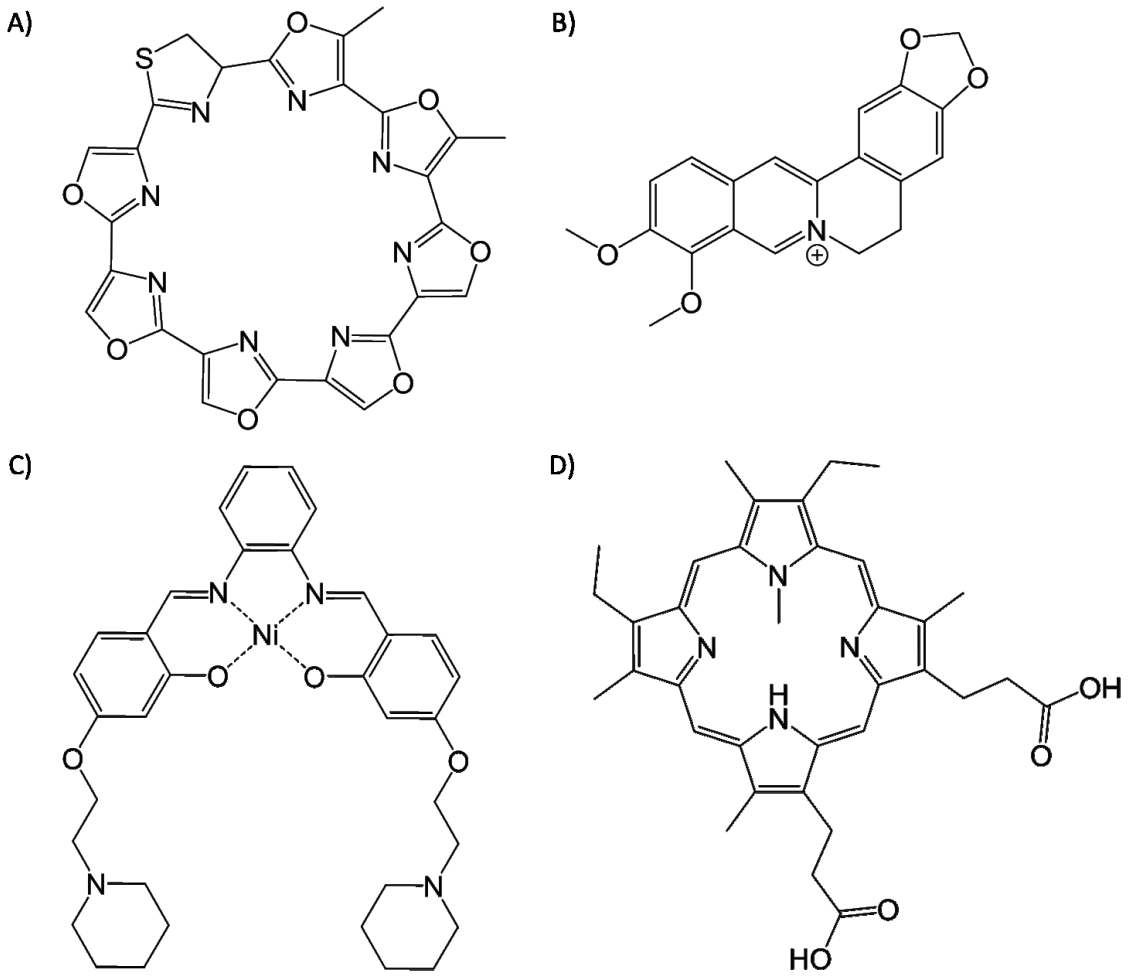


Figure 3.1.1 Molecular structure of telomestatin (A), berberine (B), nickel salphen (C) and NMM (D).

TMPyP4: A cationic porphyrin

A very well-known ligand used for G-quadruplex stabilization is 5,10,15,20-tetra-(N-methyl-4-pyridyl)porphine (TMPyP4), which has been used in several studies since it was first reported to bind to G4-DNA.[179] It is a cationic porphyrin, and the structure of this molecule is shown in Figure 3.1.2. This porphyrin macrocycle scaffold has been studied to develop other G-quadruplex ligands.[185] It has high affinity (for 1µM of compound: $\Delta T_{1/2}$=27˚C in 10 mM mM K^+, 110 mM total monovalent cations, for F21T sequence representing human telomeric repeat[194]) for G4-DNA but low selectivity over DNA duplex.[194] It has been shown to facilitate intermolecular G-quadruplex formation.[195] A crystal structure of bimolecular telomeric G-quadruplex with bound TMPyP4 has revealed that the binding mode of this compound does not involve direct interaction with the sequence's G-tetrads, which explains the poor selectivity of this ligand for G4-DNA in comparison to duplex.[196] Interestingly, while it seems to stabilize DNA G-quadruplex structures to a high degree, it has been reported that it can do the exact opposite to mRNA G-quadruplexes and induce unfolding[197] as well as destabilizing $(CGG)_n$ repeats[198].

It is often found in the literature as a molecular probe for G-quadruplex modulation, in tandem with the related molecule TMPyP2. While TMPyP2 has a very similar structure to TMPyP4, it does not interact with G-quadruplexes as well as TMPyP4 does.[199] The differences in biological activity have largely been explained by this difference in interaction.[200,91]

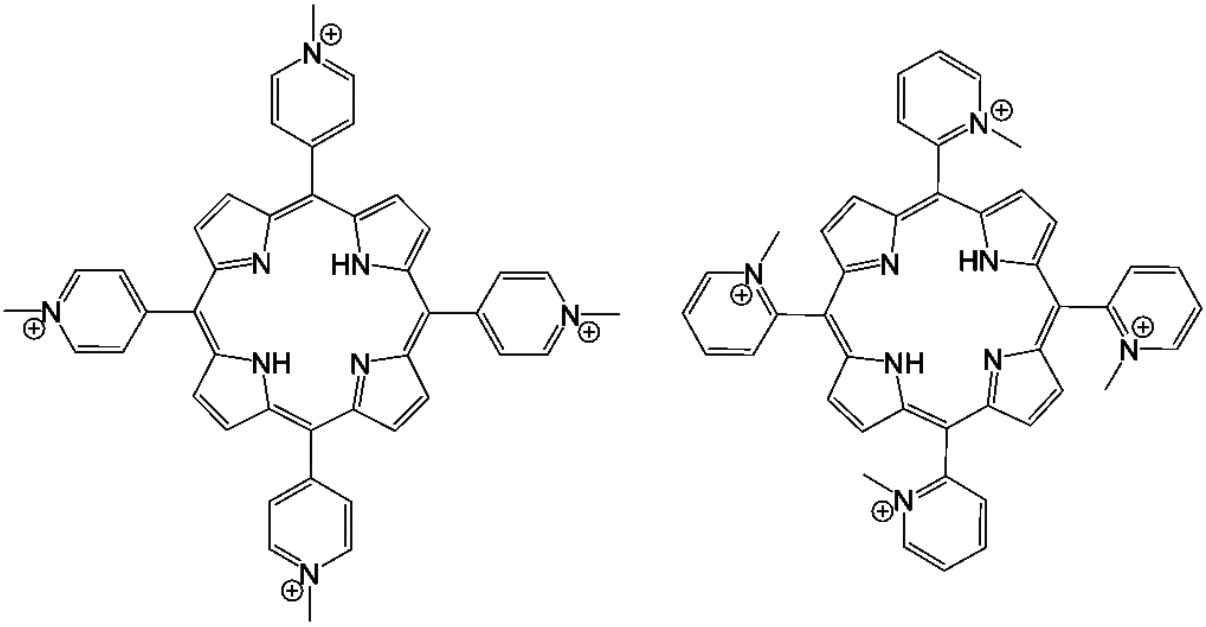


Figure 3.1.2. To the left: molecular structure of TMPyP4. To the right: molecular structure of TMPyP2.

Pyridostatin: a carboxamide with a pyridine core

Pyridostatin, structure shown in Figure 3.1.3, has been shown to have good selectivity for DNA G-quadruplexes over DNA duplexes and to strongly stabilize G4-DNA (for 0.18 μM of compound: $\Delta T_{1/2}$≈35˚C in 60 mM K^+, for a sequence representing human telomeric repeat).[201] It has been extensively used to study G-quadruplexes and to target them for medicinal purposes. Several Pyridostatin analogues have been reported in an effort to exploit the scaffold for versatile functionality.[202]

Pyridostatin was shown to activate DNA-damage response.[201] It stops the proliferation of cancerous cell lines through DNA-damage induce cell cycle arrest.[82] This has led to the idea of using it in tandem with an inhibitor of DNA-repair to increase the anticancerous activity of both molecules in a synergistic manner.[203]

Several Pyridostatin derivatives also show biological activity, which has allowed for investigation on the short term and long term effect of different substituents in the scaffold which could be due to differences in mechanism.[202] This molecule has been further modified to elucidate its biological

functionality. A biotinylated derivative was used for a pull-down assay which showed it could recognize G-quadruplex containing DNA fragments.[204] A fluorescently labelled derivative was used to investigate the target of the drug which was shown to be the non-teloremic region.[82] A Pyridostatine derivative that has RNA over DNA G-quadruplex selectivity[205] was used to visualise RNA G-quadruplexes in the cytoplasm of a cell.[32]

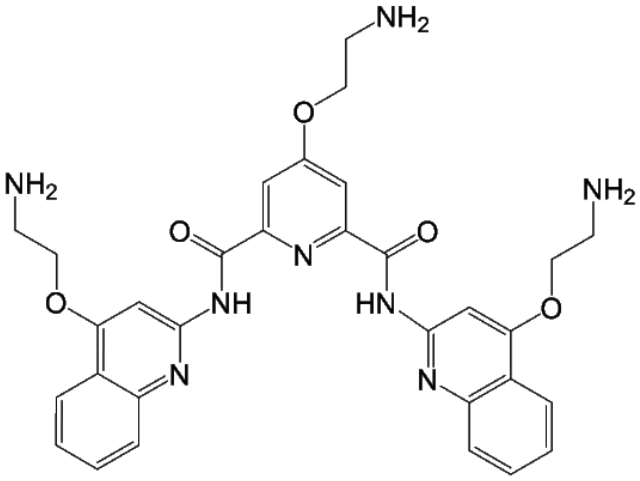


Figure 3.1.3 Molecular structure of Pyridostatin.

BRACO-19: A substituted acridine

BRACO-19 is a trisubstituted acridine, as shown in Figure 3.1.4, designed as a telomerase inhibitor by G-quadruplex stabilization[178] (for 1 μM of compound: $\Delta T_{1/2}$=27.5 °C in 50 mM K^+ for a sequence representing human telomeric repeat).[206] This compound showed a good selectivity for G-quadruplex *versus* duplex which made it a good candidate for inhibiting telomerase,[178] and has been shown to have anti-cancer activity *in vivo.*[207] It has also been shown to interact with the G4-DNA found in HIV-1 and able to act as an anti-HIV-1 molecule.[208]

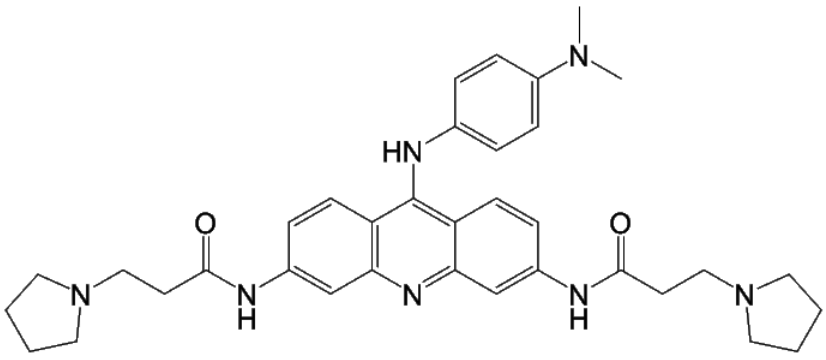


Figure 3.1.4 Molecular structure of BRACO-19.

PhenDC3: A bisquinolinium compound

Bisquinolinium compounds were shown to be highly selective for G-quadruplexes and exhibit a promising biological activity as antiproliferative compounds.[180] The scaffold was further explored using a phenanthroline core, as can be seen in Figure 3.1.5, and the resulting ligand named Phen-DC3 exhibited a very high affinity and selectivity for G-quadruplexes (for 1µM of compound: $\Delta T_{1/2}$≈30˚C in 100 mM Na^+, 110 mM total monovalent cations, for F21T sequence representing human telomeric repeat).[209]

Recently, solution structure of PhenDC3 bound to a G-quadruplex was reported[210] and shown to interact 1:1 via π-stacking, importantly the ligand interacts with all four guanines on the 5'-end surface tetrad, which may explain its strong binding affinity.

PhenDC3 was used in a genome wide study that showed it was able to induce changes in gene expression,[211] interestingly causing both up and down-regulation.

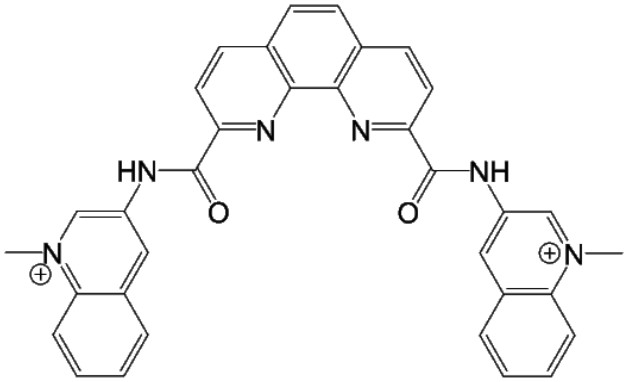


Figure 3.1.5 Molecular structure of Phen-DC3

3.1.2 FRET MELTING ASSAY

For the purpose of studying the stabilizing effect of ligands in the G-quadruplex sequence hZIP1-1, the four representative ligands have been chosen: Pyridostatin, BRACO-19, PhenDC3 and TMPyP4 because of their well-studied G4-DNA affinity (see section 3.1.1) and commercial availability.

Different methods have been designed to assess the effect of ligands in G-quadruplexes, including the FRET melting assay which is discussed in this section. Other methods to study ligand-DNA interaction include CD melting and fluorescent intercalator displacement (FID).

The FRET-melting assay to analyse the effect of ligands in the stability of G-quadruplexes works by labelling the sequence of interest with a fluorophore and a quencher, typically FAM and TAMRA, on

the 5' end and 3' end respectively as described previously on Chapter 2 section 2.3.1. When a G-quadruplex structure is stabilised by a small molecule its $T_{1/2}$ increases. The difference between $T_{1/2}$ values with and without the ligand, $\Delta T_{1/2}$, is indicative of the ligand-DNA interaction.[194] This method can be done using a thermocycler designed for qPCR in a high through-put fashion.

Ideally the G-quadruplex melting temperature in the absence of any ligand should not be too high in order to use this methodology, because then an increase in stability would be difficult to measure reliably. A common practice to decrease the value of $T_{1/2}$ is to lower the potassium concentration, which should then be complemented with lithium in order to keep the ionic strength constant. Figure 3.1.6 shows the different melting curves obtained using increasing amounts of potassium and annealing using METHOD 3. It was then concluded that the best conditions would be to use 10 mM Li cacodylate buffer with 10 mM of KCl and 90 mM of LiCl.

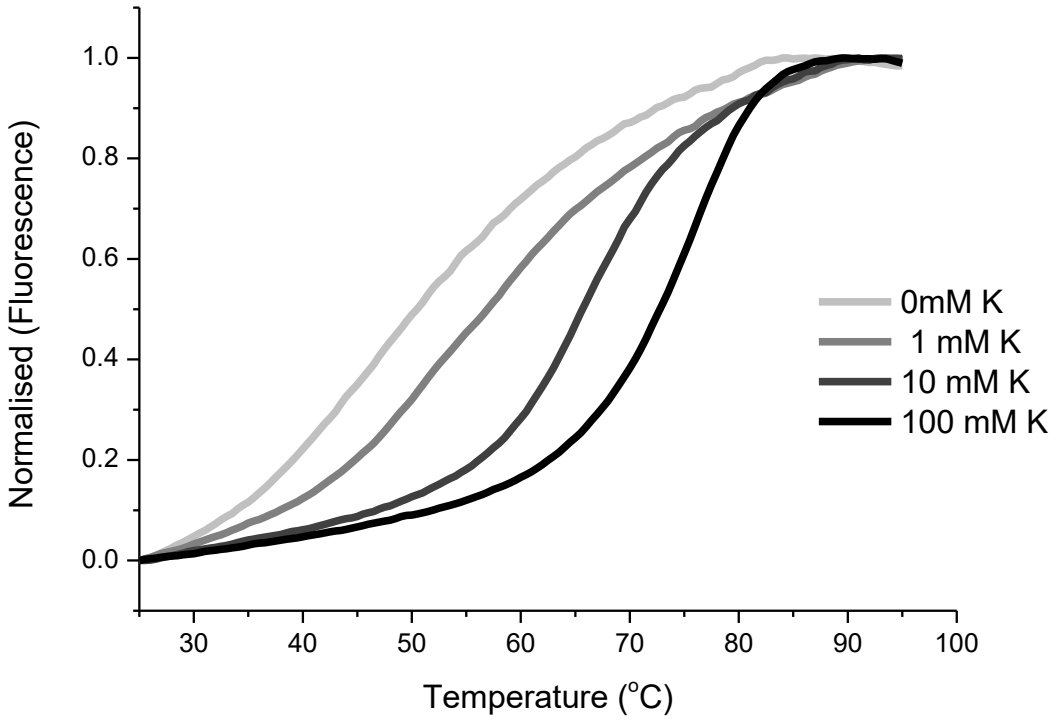


Figure 3.1.6 Plots showing the FRET melting of the double labelled (FAM/TAMRA) hZIP1-1 sequence annealed by METHOD 3 using different amounts of KCl in 10 mM Li cacodylate buffer pH=7.4. To maintain ionic strength, the annealing was complemented with LiCl to ensure a net 110 mM of monovalent ion.

To ensure results are comparable, all mutants were treated using the same annealing conditions even if some mutants showed sufficiently low melting temperatures at 100 mM KCl. A sequence that does not form G-quadruplex, MutE, was included in the screening as a control. Results are shown in Figure 3.1.7.

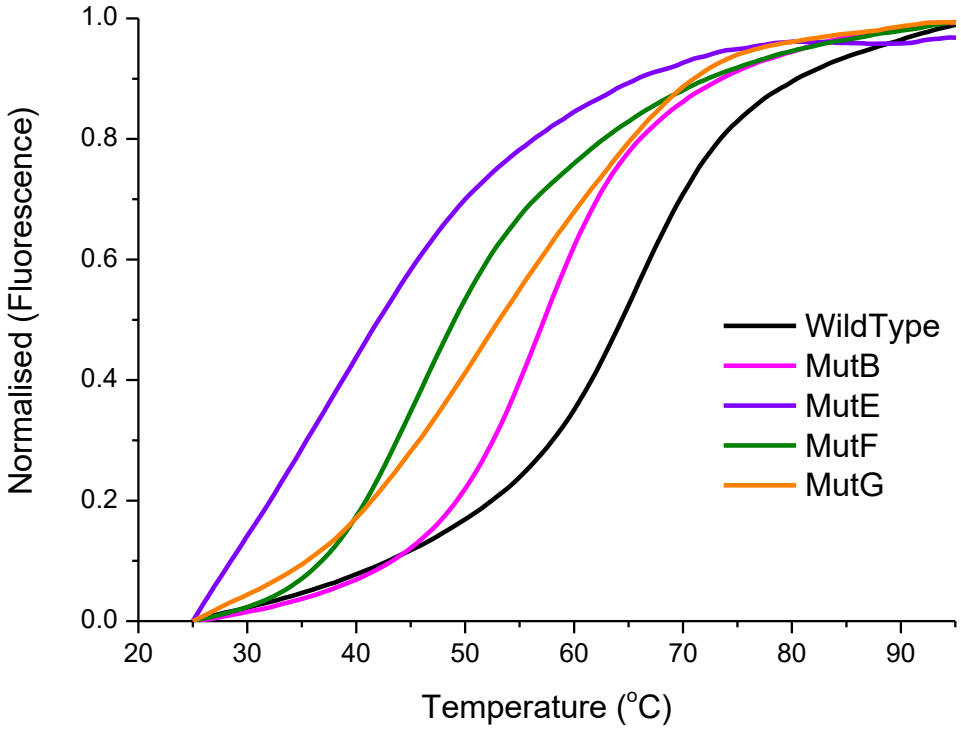


Figure 3.1.7 Plots showing the FRET melting of the double labelled (FAM/TAMRA) WildType (hZIP1-1), MutB, MutE, MutF and MutG sequences annealed by METHOD 3 using 10 mM KCl in 10 mM Li cacodylate buffer pH=7.4 and 90 mM LiCl to ensure a net 110 mM of monovalent ion. .

Figure 3.1.8 shows the effect on each sequence, hZip1-1 (wild type), MutB, MutF, MutG and MutE that is produced by 1 µM of G-quadruplex ligands BRACO-19, Pyridostatin (PDS), PhenDC3 and TMPyP4. Plots showing the same information but arranged by compound instead of sequence can be found in the Appendix (Figure A.6.4) as well as plots showing the effect of increasing concentration of each ligand in each sequence (Appendix, Figures A.6.5 to A.6.9). The values for $\Delta T_{1/2}$ for 1µM of each ligand can be found on Table 3.1.1.

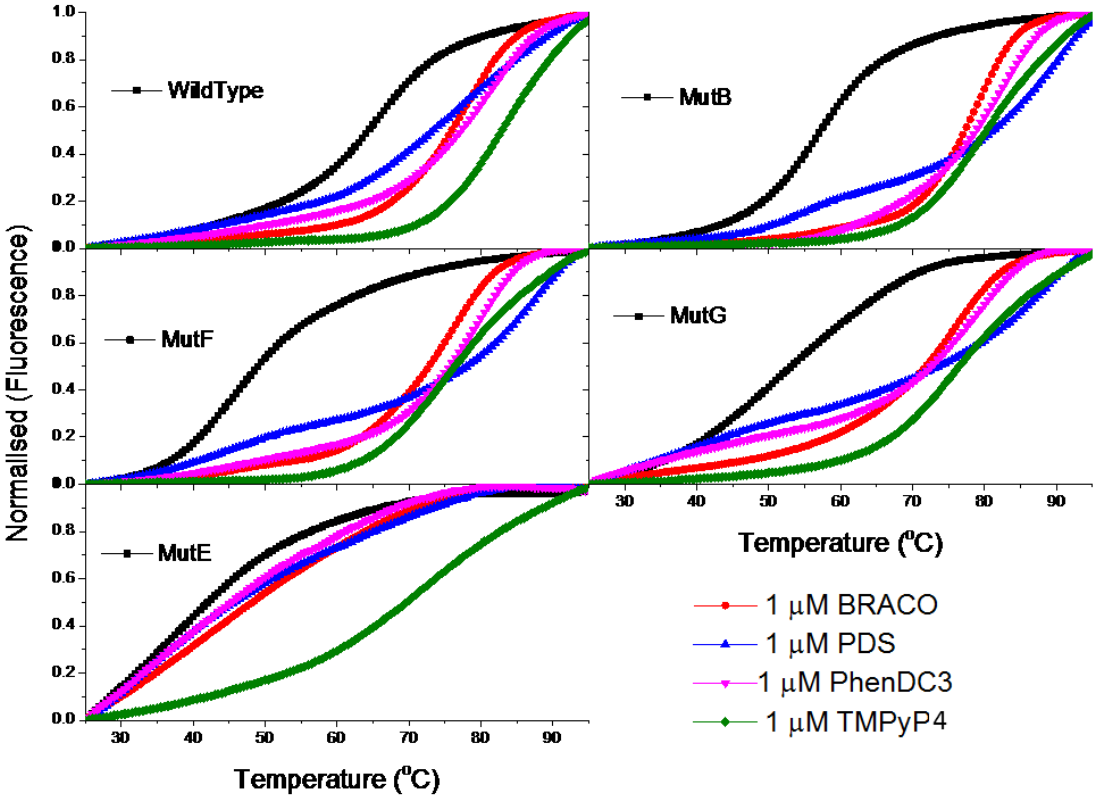


Figure 3.1.8 Increase in apparent melting temperature ($\Delta T_{1/2}$) caused by 1 µM of ligand in the different sequences studied. WildType (hZIP1-1), MutB, MutF and MutG form intramolecular G-quadruplexes while MutE does not. All sequences were annealed by METHOD 3 using 10 mM KCl in 10 mM Li cacodylate buffer pH=7.4 and 90 mM LiCl to ensure a net 110 mM of monovalent ion

While TMPyP4 shows the largest increase in melting temperature for all sequences, it does so as well for MutE, a non G-quadruplex forming sequence. This highlights the known lack of selectivity of this compound for different DNA structures. The increase in melting temperature could be explained by the promotion of intermolecular structures which this ligand is known to favour, and that are more difficult to melt.[195] On the other hand, BRACO-19, Phen-DC3 and Pyridostatin all showed a very high stabilization for all G-quadruplex forming sequences as well as low interaction with the non-quadruplex forming sequence (MutE).

Table 3.1.1 Stabilization ($\Delta T_{1/2}$) of studied G-quadruplex forming sequences (hZIP1-1 or WildType, MutB, MutF and MutG) as well as G-quadruplex negative control (MutE) by 1 µM of ligands Pyridostatin, BRACO-19, PhenDC4 and TMPyP4.

Sequence	$T_{1/2}$	G4 ligand	Δ $T_{1/2}$ (°C)
WildType	64.36±0.88	Pyridostatin	9.37±3.71
		BRACO19	10.51±0.77
		PhenDC3	13.24±1.52
		TMPyP4	17.62±0.47
MutB	57.25±0.41	Pyridostatin	23.43±4.97
		BRACO19	20.07±0.72
		PhenDC3	21.59±0.83
		TMPyP4	11.42±0.39
MutE	42.17 ±2.39	Pyridostatin	2.79±3.71
		BRACO19	6.47±2.08
		PhenDC3	3.06±2.34
		TMPyP4	27.73±1.32
MutF	49.03±0.49	Pyridostatin	27.82±5.73
		BRACO19	20.89±0.54
		PhenDC3	26.08±0.56
		TMPyP4	27.32±0.34
MutG	53.13±0.63	Pyridostatin	19.71±8.21
		BRACO19	14.59±1.75
		PhenDC3	16.40±0.78
		TMPyP4	20.47±0.83

3.1.3 CONCLUDING REMARKS

In agreement with their reported properties, ligands Pyridostatin, BRACO-19 and PhenDC3 were able to greatly stabilize the G-quadruplex structure and showed good selectivity over the non-G4 sequence. TMPyP4 did not show good selectivity, and while this molecule shows great stabilizing properties, it will not be appropriate for the study of the G-quadruplex formation *in vivo*.

3.2: DISCUSSION AND CONCLUSIONS

The use of G-quadruplex targeting ligands for therapeutic purposes has emerged as a novel anticancer strategy[176]. These small molecules have been designed to interact with G4-DNA over other forms of DNA and are thought to trap or promote their formation *in vivo*, exacerbating the effect of these structures in gene transcription and telomere function.

G-quadruplexes that may form in the promoter region would be in competition with their double stranded counterpart, so it could be rationalized that for labile structures that unfold easily their association/dissociation would be very dynamic. The use of small stabilizing molecules could trap the folded state, and would thus enhance the consequence of a G4-DNA that would have not otherwise shown a strong effect. On the other hand, if the G-quadruplex formed was very stable it would not spontaneously unfold and would require helicases to alleviate its formation.[109] The use of stabilizing ligands would prevent the unfolding by helicases,[66] meaning the G-quadruplex would remain folded and its effect would be enhanced. Finally, any binding sites that overlap G-quadruplex formation would be further occluded or enhanced by the interaction with G-quadruplex ligands.[77]

As described in Table 1.1.1, the most common observed effect of G-quadruplex stabilizing ligands is repression of gene expression. Interestingly, this is observed when the disruption of G-quadruplex formation increases transcription (for example HRAS[75]) as well as when it has little effect on transcription (for example AR[99]). This means that very often, even if the G4-DNA potential does not play a critical role (because the G-quadruplex is either labile or efficiently resolved) it can still be targeted to silence a gene. Because of the relationship between oncogenes and G-quadruplex formation, stabilizing these structures for silencing of oncogenes has been a hot topic in recent years.[176]

The promoter region of *hZIP1* may form a G-quadruplex *in vivo*, given that we have now confirmed the ability of the hZIP1-1 sequence to form G-quadruplex (see Chapter 2). This gene has been proposed as a tumour suppressor for prostate cancer, and the down-regulation of this gene has been associated with the development of malignancy.[141] The effect of well-known G4-DNA stabilizing ligands on the G-quadruplex sequence hZIP1-1 was studied in this chapter, and it was confirmed that Pyridostatin, BRACO-19 and PhenDC3 can stabilize this structure selectively over a related non-quadruplex forming sequence. It is then hypothesized that these molecules will exhibit an effect on *hZIP1* expression *in vivo*.

Intriguingly, a study found that promoters enriched with transcription binding sites for AP2 and SP1 (confirmed *hZIP1* regulators) as well as ETF and KROX (non-confirmed but predicted regulators) were differentially expressed upon treatment with TMPyP4.[74]

Trends reported in the literature suggest that a G-quadruplex forming sequences located upstream of the transcription start site could function as a gene repressor, and therefore down-regulation of the gene when treated with G4 stabilizing ligands could be expected for *hZIP1.* The strategy of using G-quadruplex ligands to treat other aberrantly expressed genes in prostate cancer[176] could then potentially lead to a decrease in the expression of hZIP1 which is a tumour suppressor. Therefore, the role of the potential G-quadruplex in gene regulation of *hZIP1* must be experimentally investigated and the effect of these small molecules in the promoter activity is an important question to be explored.

CHAPTER 4: THE EFFECT OF G4 IN THE PROMOTER REGION OF HZIP1

In this section, the biological relevance of the G-quadruplex promoter is explored using two different strategies: G-quadruplex disruption using mutations and G-quadruplex stabilization using ligands. The effects of these conditions are studied using two different platforms, *in vitro* transcription and luciferase reporter assay. The first one is used to quantify mRNA levels in a cell-free system and the second is used to quantify the gene product of the promoter *in cellulo*.

4.1 *IN VITRO* TRANSCRIPTION

4.1.1 PLASMID CONSTRUCTION

To study the effect of the G-quadruplex, plasmids containing G4-disrupting mutants were constructed based on a plasmid containing the core promoter. The design of the biological relevant mutants was discussed in Chapter 2. Based on transcription factor binding sites prediction, it is expected that all mutants (MutB, MutE, MutF, MutG and MutH) will have a similar promoter activity, which is expected to be lower than that of the WildType, but not by a lot (10-20% based on a similar mutation to MutB reported in the literature).[150] If there is a change in activity for MutE and MutH, it may be because they cannot fold into a G-quadruplex.

A common technique to study the promoter activity is to quantify the protein product that it controls. The firefly luciferase enzyme uses luciferin as a substrate to produce a luminescent signal that can be monitored using a luminometer, which makes it a suitable reporter for studying promoter activity.[90] Luciferase reporter vector pGL3 from Promega is designed to be a suitable tool for promoter study. The plasmid containing the core promoter region of *hZIP1* in a pGL3 vector was kindly supplied by Dr. Makhov[150] (Fox Chase Cancer Center, Philadelphia, USA) and used in these studies as the wild type promoter.

The plasmid containing the wild type sequence was then mutated to produce the desired mutants using site directed mutagenesis.[212] Full details describing the primers and PCR conditions used can be found in the Experimental Procedures section. In brief, the plasmid pGL3-hZIP1 was amplified and purified from *E. coli*, used as a template for PCR with primers containing the desired mutations, further treated with Dpn1 (a restriction enzyme that only cleaves methylated DNA) and used to transform *E. coli* to amplify it. Three different colonies were used to extract the DNA, purify it and sequence it in

order to screen for a successful mutant. All mutants were obtained in this way, some were used as templates for other mutations to decrease mismatches.

4.1.2 *IN VITRO* TRANSCRIPTION OF *HZIP1* PROMOTER

To study the effect of the G-quadruplex potential in the promoter region, a cell-free assay was adapted to understand the changes in promoter activity caused by the designed mutations. Nuclear extracts from HeLa cells are commercially available from Promega for *in vitro* transcription of mammalian promoters. A detailed protocol used for *in vitro* transcription and mRNA quantification can be found in the Experimental Procedures section. This was adapted from the protocol described to test for transcription inhibitors published by Voss *et al.*[213] Briefly, a reaction mixture containing the plasmid of interest and nuclear extracts is incubated to produce mRNA. The total RNA is purified by spin column and digested (Dpn1 and DNAseI) to remove plasmid DNA. The purified RNA is quantified in a single step RT-qPCR. The primers and probe used for RT-qPCR were taken from the literature and conditions optimized for the purposes of this study.[214] Levels of cDNA were quantified using a standard curve prepared using linear plasmid which can be found in the Appendix (Figure A.7.1) and total copy number was calculated. An exemplary amplification plot for an *in vitro* transcription reaction is shown in Figure 4.1.1.

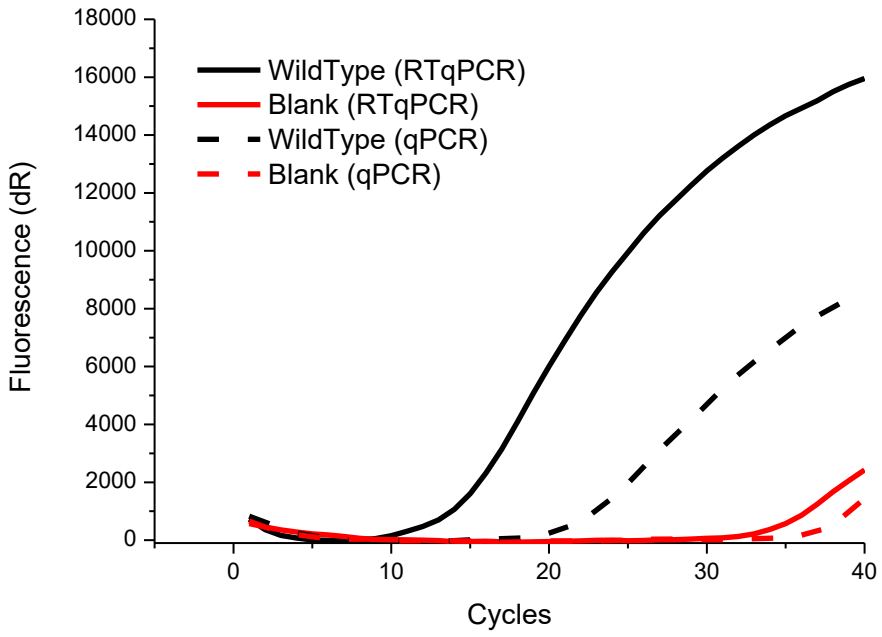


Figure 4.1.1 Representative amplification plot for WildType *in vitro* transcription reaction after total RNA purification. Reaction mixtures for amplification reaction differ from RTqPCR (solid line) and qPCR (dashed line) only in the addition of a reverse transcriptase (SuperScriptIII) for the RTqPCR reaction. As a negative control, water was used (blank, red line).

Despite DNAseI and Dpn1 treatment, complete digestion of the plasmid failed which can be appreciated by the amplification of the reverse transcriptase negative control (Figure 4.1.1). However, the difference in cycle threshold (C_T) which is used to calculate copy number shows that the remnant DNA is too small to significantly affect the measurement of the produced cDNA. A 10-fold decrease in DNA amount is on average equivalent to a 3.3 increase in C_T value (Figure A.7.1), therefore a larger difference between reaction and control (ΔC_T= 7) was considered sufficient. The effect of DNA contamination with a 100-fold lower concentration is very small and was considered negligible for the purposes of this study (Figure A.7.2).

Figure 4.1.2 shows the results for *in vitro* transcription, mRNA levels are shown normalized to the WildType promoter activity (100%). Individual repeats for these experiments can be found in the Appendix (Figure A.7.3). Results show important error bars, which are attributed to the differences in activity of the nuclear extracts. However, the error is significantly larger for experiments done using G-quadruplex stabilizing ligands, which may be due to small differences in ligand concentration. More measurements would be necessary to improve the quality of the data.

Non-G4-forming mutant (MutE) shows a decrease in transcription, and while it is not statistically significant, it is interesting because this is the observed effect using luciferase reporter assay discussed in section 4.21. There is also an average increase in mRNA levels for the plasmids treated with G4-stabilizing ligands, which is the observed effect using luciferase reporter assay discussed in section 4.2.2.

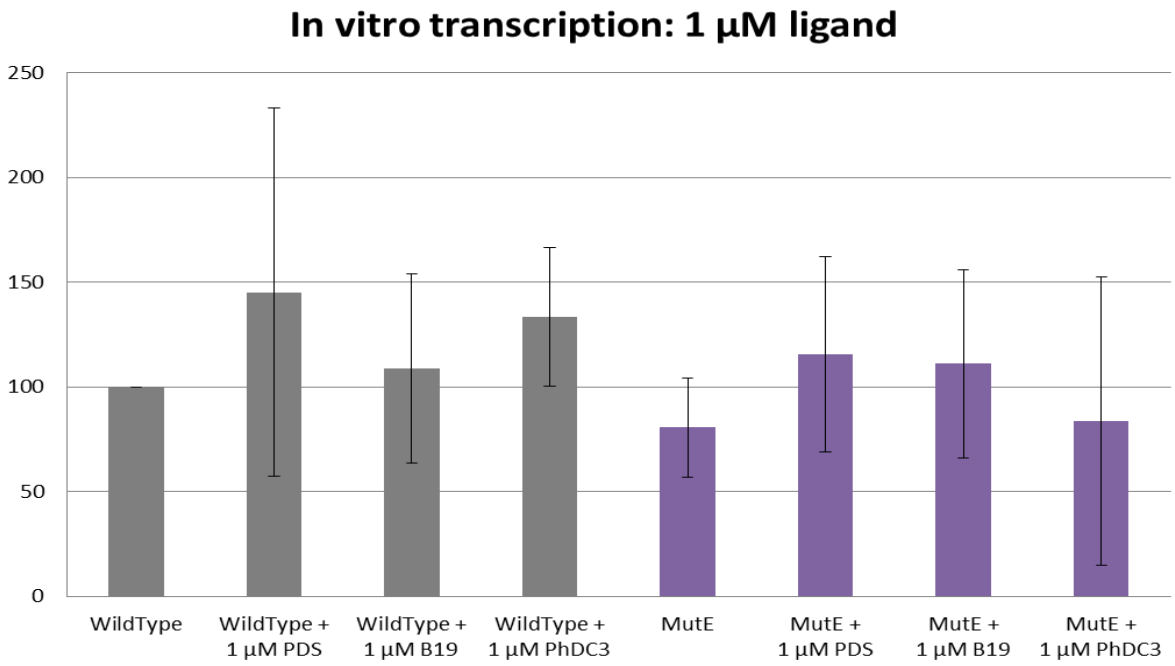


Figure 4.1.2 *In vitro* transcription results. *In vitro* transcription reactions were carried out using HeLa nuclear extracts with plasmids containing either the wild type promoter (WildType) or a mutant that cannot form G-quadruplexes (MutE). They were treated either with 1 µM G-quadruplex stabilizing ligand or with water as control. Total RNA was purified and mRNA levels were quantified using RTqPCR and normalised to the WildType promoter copy number. Errors bars show standard deviation and n=3.

4.2 LUCIFERASE REPORTER ASSAY

4.2.1 *IN VIVO* TRANSCRIPTION OF *HZIP1* PROMOTER

A popular methodology for the analysis of promoter activity is the use of reporter assays in which the promoter regulates the activity of a protein that is easy to quantify through its activity. The activity of the protein is then related to the activity of the promoter. Luciferase has been the reporter of choice for several G-quadruplex promoter studies. The details of the luciferase reporter assay (LRA) used can be found in the Experimental Procedures section.

In brief, the plasmid constructs described before were used to transfect cells along with a renilla luciferase plasmid that serves as transfection control. After a period of incubation, the cells were lysed and the luciferase activity was quantified. The firefly luciferase values were normalised against the renilla luciferase values. WildType activity was assigned to 100% and the mutants were normalised to this value.

To study the promoter activity of zinc transporter ZIP1, two different cell lines were used. While normal healthy prostate cells accumulate high levels of zinc, it is known that this ability is lost during the development of prostate cancer. In order for a model to be useful, the mechanisms for zinc accumulation must be in place and zinc accumulation should be possible. Cell line PC3 are prostate cancerous cells which have retained the capacity for accumulating zinc under *in vitro* conditions[123], and they have been shown to express *hZIP1.*[141] This cell line has been the model of choice for previous *hZIP1* transcriptional studies,[126,137,150,151,152] therefore it was chosen for the current work as a cancerous cell model. A second non-cancerous prostate cell line, PNT1A, was also selected. This cell line has been used for G-quadruplex related transcriptional studies.[99] The LRA results for PNT1A and PC3 48 hours after transfection are shown in Figure 4.2.1 and Figure 4.2.2 respectively. Table 4.2.1 summarizes the information.

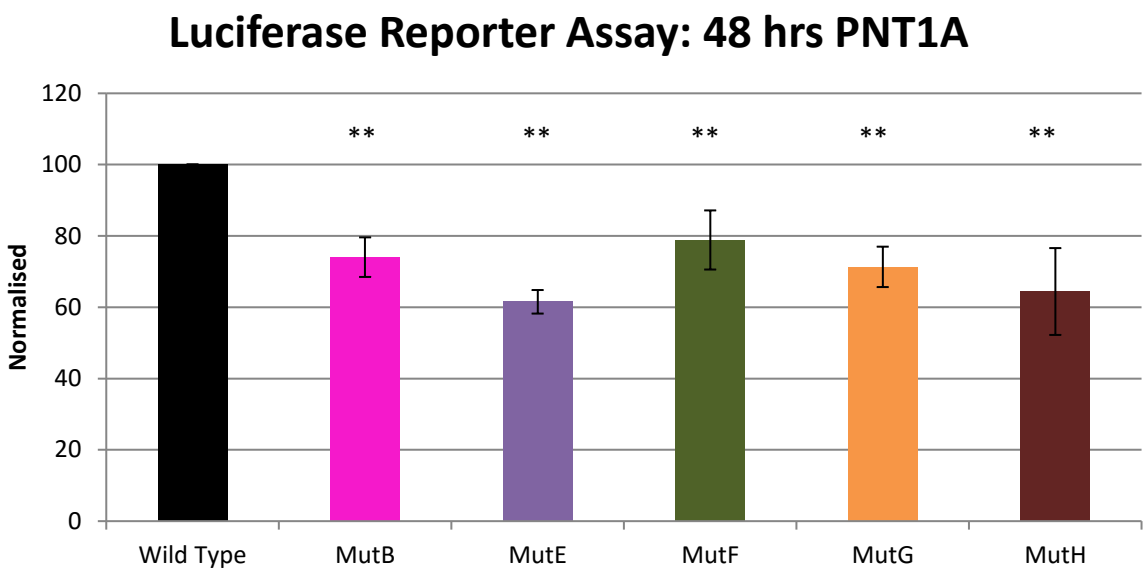


Figure 4.2.1 Luciferase reporter assay 48 hours after PNT1A cells were transfected with WildType plasmid (hZIP1) and its mutants. Firefly luciferase values were normalised to renilla luciferase values, and then all values were expressed as a percentage of normalised WildType activity. Error bars show standard deviation. Student's t-test was used to calculate p, * is used for $p<0.1$ and ** for $p<0.05$, and $n=4$.

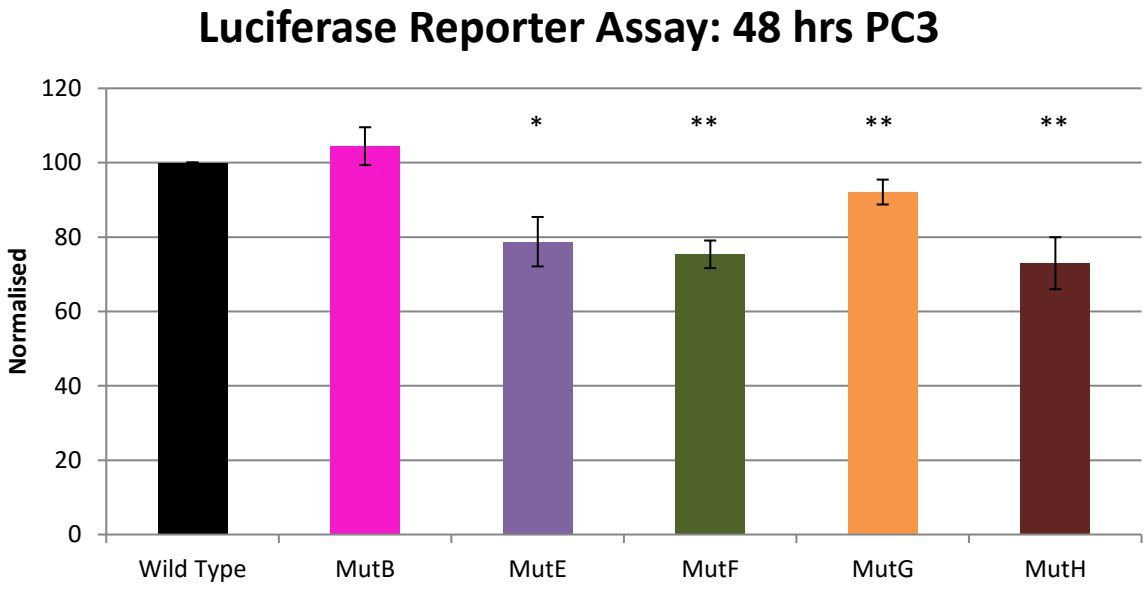


Figure 4.2.2 Luciferase reporter assay 48 hours after PC3 cells were transfected with WildType plasmid (hZIP1) and its mutants. Firefly luciferase values were normalised to renilla luciferase values, and then all values were expressed as a percentage of normalised WildType activity. Error bars show standard deviation. Student's t-test was used to calculate p, * is used for $p<0.1$ and ** for $p<0.05$ and $n=3$.

Table 2.1.1 Luciferase reporter assay (LRA) results for hZIP1 and its mutants in PC3 and PNT1A after 48 hours. "G4" column shows intramolecular G-quadruplex formation. Mutant firefly/renilla luciferase values were normalised to WildType values (100%). Experiments were repeated three times (PC3) and four times (PNT1A).

Name	Sequence	G4	LRA (PNT1A)	LRA (PC3)
WildType	TGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGA	Yes	100	100
MutB	TGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGA	Yes	74.05 ± 5.55	104.44 ± 5.09
MutE	TG**tttt**AGG**ttt**GAGCTATGGGAGGGGGCGGGA	No	61.52 ± 3.30	78.76 ± 6.65
MutF	TGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGA	Yes	78.87 ± 8.29	75.36 ± 3.71
MutG	TG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGA	Yes	71.32 ±5.67	92.10 ± 3.35
MutH	TG**tt**GGAGG**ttt**GAGCTATGGGAGGGGGCGGGA	No	64.41 ± 12.18	72.96 ± 7.00

As it was expected, MutB mutation had little effect in transcription when the activity was studied in PC3 cells in agreement with similar mutants reported elsewhere.[150] Interestingly, the effect was higher when the assay was done in PNT1A cells (25% drop in activity). This may be due to changes in the transcription machinery in the cancerous *vs*. the non-cancerous cell lines.

Promoter activity for MutB, MutF and MutG (G4 forming mutants) was expected to be very similar because they do not have significantly different transcription factor binding sites and because they exhibit similar G-quadruplex thermodynamic and kinetic characteristics. The promoter activity of MutE and MutH (non-G4 forming mutants, with also similar transcription binding prediction) was expected to be similar to that of the G4 forming mutants if the G-quadruplex formation is not important, or different if the G-quadruplex potential indeed affects transcription. To analyse the similarities of the mutants, all values obtained for G4 forming mutants were treated collectively and non-G4 forming mutants were in turn treated collectively. The individual responses for each repeat can be found in the Appendix (A.8.10 and A.8.11).

Table 2.1.2 Luciferase reporter assay (LRA) results for G4 forming mutant *versus* non-G4 forming mutants. p value was calculated using Student's test.

Cell line	Mutant	LRA	p value
PC3	G4 forming	90.63 ± 13.13	0.0258
	Non-G4 forming	75.86 ± 6.88	
PNT1A	G4 forming	74.75 ± 6.82	0.0028
	Non-G4 forming	62.97± 8.41	

This hypothesis is indeed confirmed that G-quadruplex formation in the promoter region of *hZIP1* has a role in regulation. However, when the G-quadruplex was disrupted transcription was reduced (40% in non-cancerous cell line PNT1A and 25% in cancerous cell line PC3) which suggests that the G-quadruplex does not act as a repressor, as it was previously hypothesised, but as an activator. This has been observed in the murine *KRAS* promoter.[77]

4.2.2 *IN VIVO* TRANSCRIPTION AND THE EFFECT OF G4 LIGANDS

To further understand the role of G-quadruplex formation in the promoter region of *hZIP1* and its potential as a drug target, the effect G-quadruplex stabilizing compounds on transcription was investigated.

Experimental details can be found the Experimental Procedures section. Briefly, PC-3 cells were transfected with a firefly plasmid containing either the WildType promoter, or a non-G4 forming mutant (MutE) and a renilla plasmid as transfection control. After 24 hours of transfection the culture media was replaced with media containing a G-quadruplex stabilizing molecule (i.e. Pyridostatin, PhenDC3 or BRACO-19) and incubated for a further 24 hours. 48 hours after transfection, cells were lysed and luciferase and renilla activities were quantified. Firefly activity was normalised to renilla. Values were further normalised to the untreated WildType control. Results are presented in Figure 4.2.3.

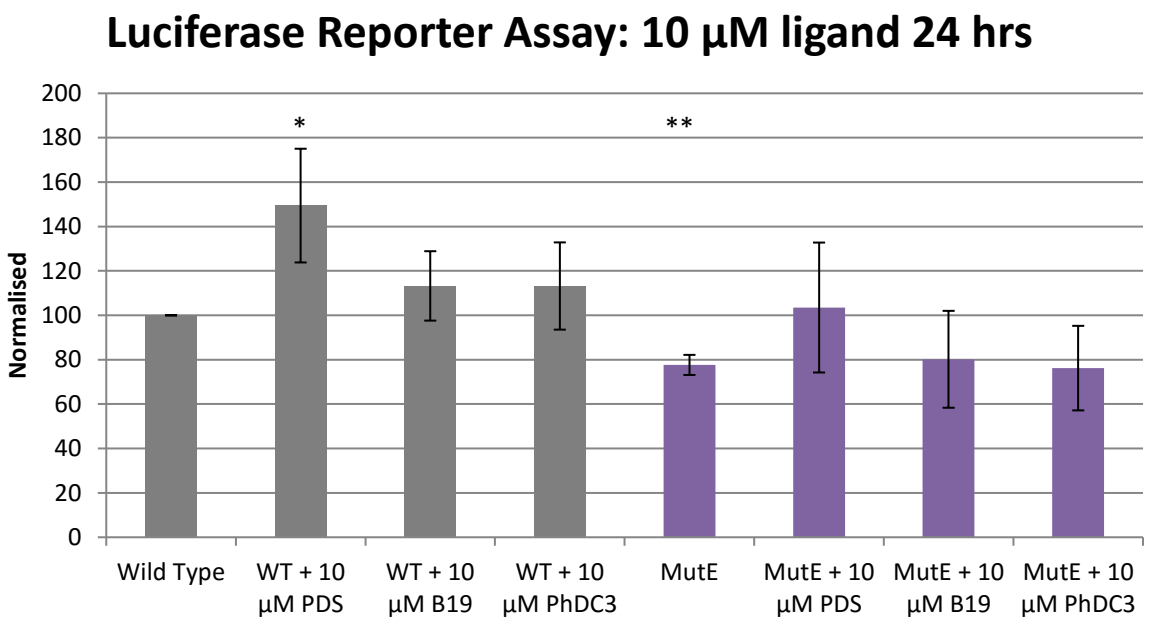


Figure 4.2.3, Luciferase reporter assay 48 hours after PC3 cells were transfected with WildType plasmid (hZIP1) and MutE, and 24 hours after adding 10 µM of ligand. Firefly luciferase values were normalised to renilla luciferase values, and then all values were expressed as a percentage of normalised WildType activity without ligand. Error bars show standard deviation. PDS= Pyridostatin, B19=BRACO-19, PhDC3=PhenDC3. Student's t-test was used to calculate p, * is used for $p<0.1$ and ** for $p<0.05$ and n=3.

Upon treatment of G-quadruplex stabilizing ligands, there is an increase of Luciferase activity in the G4 forming WildType sequence. This effect is particularly strong with Pyridostatin, however the effect for BRACO-19 and PhenDC3 is within error. As a control, treatment of a non-G4 sequence (MutE) was also included. While Pyridostatin seem to have an effect in this mutant it is not statistically significant. BRACO-19 and PhenDC3 do not show any effect at all on MutE. Lower ligand concentration and shorter incubation times did not show any significant effect on transcription (see Appendix Figure A.8.12 and Figure A.8.13).

4.3: DISCUSSION AND CONCLUSIONS

It had been hypothesised that the G-quadruplex forming sequence in the promoter region of *hZIP1* could be involved in gene regulation. To test this idea, two non-G4-forming mutants and three G4-forming mutants were designed as discussed in Chapter 2 and their G-quadruplex forming potential was confirmed.

Unimolecular parallel G-quadruplexes found in promoter regions have widely been reporter as repressors of gene regulation.[215] The position of the G-quadruplex near the transcription start site and the overlap with important transcription binding sites were indicators that this G-quadruplex may function as a repressor as well.[109,216] However, the opposite was observed. When the G-quadruplex potential was disrupted, gene expression dropped (40% for non-cancerous cells, 25% for cancerous cells), suggesting G4-DNA may be acting as an activator of expression. This has been observed before in other G-rich promoters like murine *KRAS,*[77] which is able to form two topologically distinct quadruplexes *in vitro*. It was shown that the transcription required zinc finger MAZ which in turn was able to bind and stabilize G4-DNA formation for this sequence.[77] Bcl-2 promoter contains the sequences dubbed "p32", which recently was also shown to down-regulate transcription when the G-quadruplex potential is disrupted[98] and yet the sequence can be targeted with G4-stabilizing ligands to decrease transcription. To explain these results, the authors argued that the G-quadruplex may be necessary for binding IGF-1 but at the time of writing have not published further experiments validating this hypothesis.

The results for non-cancerous cell line PNT1A show that the G4-forming mutations (MutB, MutF and MutG) had all similar gene expression between them and were clearly different from the non-G4-forming sequences (MutE and MutH) (see Figure 4.2.1). However, in the cancerous cell line PC3 the trend was different. While G4-forming mutants MutB and MutG indeed had similar expression, MutF did not. The expression observed for MutF was more similar to the observed expression in non-G4-forming mutants MutE and MutH (see Figure 4.2.2). MutF was shown in Chapter 2 to be the least stable of all sequences which was attributed to a long loop between the first and second G-run. However, kinetic and thermodynamic analysis revealed that this sequence had similar stability to MutB and MutG, therefore the drop in transcription cannot conclusively be attributed to structural differences. One possible explanation is that for the G-quadruplex to function as an enhancer of gene expression there is the need for stabilizing proteins that prevent the unfolding of the G-quadruplex. These proteins may be sensitive to structural differences, and may be expressed differentially in cancerous *versus* non-cancerous cells.

G-quadruplex stabilizing ligands have emerged as novel cancer therapeutic strategies, however they pose the drawback that they might interact with the G-rich promoter regions that are not overexpressed during cancer development. An example of this is *hZIP1*, which has been considered a tumour suppressor for prostate cancer. Three well-known G-quadruplex stabilizing ligands were used as probes for the effects these ligands may have in this promoter (see Chapter 3). Both BRACO-19 and PhenDC3 showed no significant change in the promoter activity, suggesting these ligands and their derivatives could be used to target G4-rich oncogenes in prostate cancer without compromising zinc uptake. Pyridostatin on the other hand increased the promoter activity, suggesting that if this molecule was used to target oncogenes it may have a synergistic effect of enhancing *hZIP1* expression. This would increase zinc(II) accumulation which in turn promotes apoptosis in prostate cells. Furthermore, none of the molecules had an effect in the transcription of the G4 negative control, MutE. This argues in favour of a G-quadruplex stabilizing mechanism.

Results for the *in vitro* transcription assay had very large errors, which do not allow for conclusive interpretation of the data. However, the average response of the results show a remarkable resemblance with the results obtained using LRA. In both platforms, MutE is down-regulated as compared to the WildType promoter, and in both cases Pyridostatin induces an increase of the WildType regulation. The procedure should be further optimized to increase reproducibility in order to develop a novel platform to test G-quadruplex stabilizing ligands.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

5.1 GENERAL DISCUSSION AND CONCLUSIONS

Prostate cells lose their ability to accumulate high levels of zinc in the early stages of malignancy due to the down-regulation of the zinc transporter ZIP1.[141] The regulatory mechanisms of this gene have been the focus of recent research, as the reversal of the down-regulation could be a therapeutic route.[126] Research has shown that the gene is up-regulated by testosterone,[128] prolactin,[128] and down-regulated by available zinc.[128] Transcription factors that are critical for its expression include SP1,[150] CREB[150] and AP-2α,[151] the latter has been shown to be down-regulated by methylation of its promoter in prostate cancer patients which in turns down-regulates *hZIP1.*[151] Other factors that contribute to the down-regulation of this gene in cancerous cells are the transcription factor RREB-1[152] and the microRNA cluster 183-96-182.[155] However, the role of non-canonical structures in the regulation of this gene had not been explored until now.

The promoter region of *hZIP1* contains a G-rich sequence (hZIP1-1) near one of its transcription start sites[150] that was predicted to form an intramolecular G-quadruplex in physiological conditions using online bioinformatics tools.[53] There is increasing evidence that G-quadruplexes form *in vivo,*[81] and that they play key roles in transcriptional control.[217] Additionally, *hZIP1* was found to contain G-quadruplexes in a genome-wide experimental search for these structures.[51] It was hypothesised that the formation of a G-quadruplex in the promoter region of *hZIP1* was involved in gene regulation.

The G-quadruplex formation of hZIP1-1 was investigated under physiological conditions of pH and potassium concentration. It was shown that by manipulating the annealing conditions it is possible to enrich the unimolecular fraction (favoured using annealing METHODS 2 and 3) over the more thermodynamically stable intermolecular structures (favoured using annealing METHOD 1). Of the intramolecular structures studied, it is clear that not all possible topologies of hZIP1-1 are equally stable: the parallel conformation is the most stable. Nevertheless, when the sequence is folded quickly using METHOD 2 there is a population of structures with an antiparallel contribution (Figure 2.2.4). Given enough temperature or time, the population of antiparallel topologies will rearrange into a parallel topology (Figure 2.3.2, Figure 2.3.13). Annealing METHOD 2 increases the population of kinetically favoured structures, which can be thought of as intermediaries to a more stable structure, in this case the intramolecular parallel. If the G-quadruplex was to form in the promoter during transcription, the topology would have an important impact on its stability.[40] Proteins that interact with the G-quadruplexes *in vivo* may also be sensitive to structural differences. Given that the

conformations that exhibit an antiparallel peak seem to fold quickly and then unfold quickly again at physiological temperature, they seem unlikely to stay formed *in vivo* unless the environmental conditions, such as crowding,[218] protein interaction[77] and supercoiling[57] further stabilize these topologies or help them rearrange into a parallel structure.

Small molecules have been designed to interact with G-quadruplexes and stabilize them *in vivo* in the hope of targeting G-rich promoters and alter gene expression.[176] Some ligands display selectivity for some structures over others,[38] so the effect of a G-quadruplex ligand will depend on the ability of the ligand to stabilize the structure formed *in vivo* which may differ from the one studied *in vitro*. Pyridostatin, PhenDC3 and BRACO-19 were found to stabilize *in vitro* the G4 parallel conformation of hZIP1-1 in a similar fashion. Their ability to stabilize the alternative antiparallel conformation could not be tested because this conformation cannot be enriched by altering the annealing conditions, so they can only be formed as a mixture of parallel and antiparallel. A possibility to study this in the future is to generate point mutations that drive the antiparallel configuration over the parallel configuration.[26]

The thermodynamic and kinetic study of the mutants designed as mutation controls that are G4 forming (MutB, MutF and MutG) showed that while they are all significantly less stable than the WildType sequence, they have similar stability between them. MutB is the most stable, followed by MutG and MutF, but the differences are small. These studies were done using modified short oligonucleotides, so it cannot be ruled out that under other conditions the difference in stability could increase, for example when the G-quadruplex is formed with a duplex overhang instead of single flanking bases, in crowding conditions, or in a supercoiled plasmid, which would be all closer conditions to the ones found *in vivo*. All of these conditions are likely to affect the stability of the G-quadruplexes[,37,58,59], so the question remains if the mutants would all be affected in a similar fashion and retain their similar stability in different conditions or if they will be affected differently because of their differences in G-run length and loop length. A suitable assay to further investigate the difference in kinetic stability of the wild type and the different mutants would be using a mini-circle platform recently developed in the group for single molecule spectroscopy that allows the study of the G-quadruplex formation surrounded by duplex DNA.[219]

The question of whether the mutants exhibit a similar stability *in vivo* is relevant in order to interpret the results obtained in the luciferase reporter assay. The experiment was designed to have three types of studied plasmids: first, the wild type sequence; second, two different mutants that would fully disrupt G-quadruplex formation, “non-G4 forming” (MutE and MutH); and third, three different mutants that share some of mutations as the “non-G4 forming” yet are able to form G-quadruplex

structures, “G4 forming” (MutB, MutF and MutG). The G-quadruplex disrupting sequences can be thought of as the sum of the non-disruptive sequences: MutE= MutB + MutG +MutF and MutH=MutG+MutF. The mutants were designed in order to avoid disruption of any relevant known transcription binding site, and as discussed in Chapter 2, are all expected to show similar promoter activity.

The LRA results show that MutB and MutG have a similar activity to the wild type promoter in PC3 cells (prostate cancer cells), which confirms that the mutations do not have a disrupting effect. However, MutE, MutF and MutH all show a diminished activity. All three mutants share the same three G→T mutations in the centre of the second G-run. This suggests that the diminished activity observed in MutE and MutH is due to the sequence mutation found in MutF. However, when the results are treated collectively for “G4 forming” and “non G4 forming”, a statistical difference between the two groups is observed. When a non-cancerous prostate cell line was used, PNT1A, the behaviour was as predicted, with “G4 forming sequences” showing very similar activities among them and statistically different than “non-G4 forming sequences”. In this case, the mutation found in MutF was not more or less detrimental that the mutations found in MutB and MutG, which meant the drop in transcription for MutE and MutH can be better explained in terms of G-quadruplex potential.

Consensus sequence analysis was used to predict the binding behaviour of the mutants, however this methodology has been criticized because it is overly simplistic.[220] In particular, that MutF does not disrupt a binding site could be challenged based on the observation that the guanines involved in the mutation are all conserved, as can be seen in Figure 2.1.1, which would deem them biologically relevant and likely to be a regulatory element. A crucial piece of information that consensus sequence analysis lacks, besides sequence variability, is structural information. Transcription factors have preferential structural binding, sequence logo analysis revealed for example that RapA binding site binds to distorted (non-B) DNA.[221] It has also been discussed that zinc fingers (consensus) binding sites are found near putative G-quadruplex structures.[74] More research between the relationship of transcription factor binding and structural characteristics of DNA will shed light into the implications of individual structurally altering mutations.

Considering the above, the differences between the PC3 and PNT1A cell lines may be due to the different protein profiles they express; perhaps the proteins that interact with the G-quadruplex sequence are found in different concentration in cancerous and non-cancerous cells, and these proteins could be sensitive to structural differences. Because of the long loop found between the first and second G-run in MutF, it is expected to be structurally different to MutB and MutG.

Finally, the ligand's effects on *hZIP1* promoter activity suggests that using G-quadruplex stabilizing compounds to target and increase the expression of *hZIP1* in prostate cancer is a possible therapeutic strategy. Selective ligands could be designed to stabilize hZIP1-1 and up-regulate ZIP1. The design of these ligands should explore the effect of stabilizing the parallel and the antiparallel conformation which would shed light into which of the two conformations, or both, can be formed and targeted *in vivo.* Furthermore, targeting the *hZIP1* promoter with G4-ligands is a compatible strategy with targeting oncogene promoters with general G4-ligands. Some of these drugs may not have an effect on the expression of *hZIP1*, but they will likely not have a detrimental effect either, as could be feared due to the G-quadruplex formation potential in the promoter while other drugs could have a synergistic effect of targeting both oncogenes to down-regulate and up-regulate this tumour suppressor. Of particular interest is the finding that the androgen receptor (AR), another important key player in prostate health and cancerous development, has recently been shown to have a drugable G-quadruplex putative structure in its promoter.[99]

In summary, the G-quadruplex putative sequence found near the transcription start site in the promoter of *hZIP1* forms stable intramolecular G-quadruplexes *in vitro,* under physiological conditions of pH and potassium concentration. Disrupting the potential of this sequence to form such structures has a detrimental effect in gene transcription, while the use of G-quadruplex stabilizing ligand Pyridostatin increases transcription. This suggests that using G4-ligands for up-regulation of tumour suppressor *hZIP1* could be a viable strategy for prostate cancer treatment, which is in turn compatible with the use of these small molecules targeting other genes simultaneously.

5.2 FUTURE WORK

It would be interesting to further study the effect of structurally selective (parallel *versus* antiparallel) G4-stabilization both *in vitro* as well as *in vivo*. To do so, the following experiments are proposed: explore the possibility of favouring the antiparallel conformation over the parallel using mutations, and use this mutant to screen for G4-ligands that show selectivity for one conformation over the other. LRA of the two conformations and G4-ligand treatment may help elucidate the relevance of each conformation. This would help clarify the relevance of kinetically or thermodynamically driven conformations for the G-quadruplex in *hZIP1*.

The results shown in section 4.2.1 suggest that using an *in vitro* transcription protocol could be a promising platform to study the effect of G4-stabilizing ligands in promoters. This was reflected in the similarity of the results in comparison to those obtained *in vivo*. However, conditions have to be further optimized to increase the reproducibility of the assay. It would be particularly interesting to use this platform with nuclear extracts of cancerous and non-cancerous cell lines, to help understand the effect of altered transcription machineries in the role of G-quadruplexes in promoter regions.

EXPERIMENTAL PROCEDURES

PROMOTER ANALYSIS AND MUTANT DESIGN

Promoter sequences for SLC39A1 human and mouse were downloaded from EPDNew (http://epd.vital-it.ch/) from -599 to 100 bases each.

Promoter sequences were analysed using QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/analyze.php) to find G-quadruplex structures with three tetrads and a maximum loop of 15 bases.

Human and mouse promoter sequences were used to search similar sequences using nucleotide BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) without filtering for low complexity regions. Good matches were all found in SLC39A1 predicted sequences for related species. When the human sequence was used to run a BLAST search against the predicted SLC39A1 genes of different animals, it was found highly conserved in closely related genomes like *Pan troglodyetes*, *Nomascus leucogenys*, *Mandrillus leucophaeus*, *Pongo abelii* and other primates. The mouse promoter sequence was highly conserved in *the Rattus norvegicus* promoter.

Relevant sequences were selected and downloaded: *Homo sapiens* promoter SLC39A1_3 from EPDNew*, Mus musculus* promoter Slc39a1_1 from EPDNew*,* as well as the SLC39A1 upstream region for *Rattus norvegicus, Pan troglodyetes, Bos Taurus*. They were then used for alignment using Cluster Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/), a multiple sequence alignment tool from the tool framework EMBL-EB.

The human promoter transcription factor binding sites were analysed using Nsite (http://www.softberry.com/berry.phtml?topic=nsite&group=programs&subgroup=promoter) and Alibaba2.1 (http://www.gene-regulation.com/index2). Mutations were designed to minimize transcription factor disruption while disrupting G-quadruplex formation. This was checked using QGRS Mapper, Nsite and Alibaba2.1.

BUFFER AND REAGENT PREPARATION

Li cacodylate (10 mM)	This buffer was used for all variable temperature experiments, and was prepared according to recommended methodology[36] to avoid potassium or sodium cations. 20 mM Cacodylic acid was dissolved in distilled water and pH was adjusted using 1 M LiOH. The solution was sterile filtered and kept in the refrigerator. Buffer is diluted to 10 mM according to needs.
Tris buffer	50 mM Tris buffer pH 7.4
TBE (10X)	1 M Tris base 1 M boric acid 0.02 M EDTA When running gels supplemented with salt, the same salt should be used to supplement the buffer.
TAE (50X)	2 M Tris base 1 M acetic acid 50 mM EDTA
Transcription buffer	20mM HEPES pH 7.9 100mM KCl 0.2mM EDTA 0.5mM DTT 20% glycerol
rNTP mix (6.25X)	2.5 mM rATP 2.5 mM rGTP 2.5 mM rCTP 2.5 mM rUTP
Alkaline loading buffer	80% formamide 10 mM NaOH
Non-denaturing loading buffer	10% glycerol 1x TBE

RT-qPCR mix and qPCR mix	SuperScript® III RT/Platinum Taq Mix (Invitrogen) (0.4 µL per 20 µL) or Platinum Taq (Invitrogen) only (0.08 µL per 20 µL, 2 u) 1X Reaction Mix (Invitrogen) Forward primer 1 µM Reverse primer 0.5 µM Probe 250 nM 0.5 nM $MgSO_4$ (on top of reaction mixture) 1 u RNase inhibitor (20 u/µL)

GEL PREPARATION

Type of Gel	Final concentration
Non-denaturing Polyacrylamide for EMSA	Main Gel 12% or 16% Acrylamide 1x TBE 20 mM MCl (M=Li, Na, K) APS and TEMED to polymerize
Denaturing Polyacrylamide for EMSA	Main Gel 20% Acrylamide 1x TBE 8 M urea APS and TEMED to polymerize
Plasmid purification/visualization	Agar 1%

DNA PREPARATION

All DNA sequences were purchased from Eurogentec (purified by RP-cartridge or RP-HPLC) and solubilized in filtered 10 mM Tris buffer for storage at 100 µM and kept at -20°C. For assays performed at room temperature or lower DNA was prepared in 50 mM Tris buffer pH=7.4, while for assays requiring variable temperature DNA was prepared in 10 mM Li cacodylate buffer pH=7.4.DNA was annealed using three different methods.

METHOD 1: DNA was prepared in buffer containing the desired salt (100 mM MCl, M= Li, Na or K) and heated up to 95°C for at least five minutes and then cooled down to room temperature slowly for at least 3 hours or overnight.

METHOD 2: DNA was prepared in buffer, heated up to 95°C for at least five minutes and then cooled down rapidly by placing the sample on ice for at least five minutes after which the desired salt was added (100 mM MCl, M= Li, Na or K).

METHOD 3: DNA was prepared in buffer, heated up to 95°C for at least five minutes and transferred to a 60°C bath where it was incubated for five minutes, the desired salt was added (100 mM MCl, M= Li, Na or K) and incubated for another five minutes and then it was cooled down to room temperature.

When a concentration of salt different to 100 mM was needed, for example 10 mM KCl, the buffers were supplemented with lithium salt to keep the ionic strength constant (for example, 10 mM KCl and 90 mM LiCl). Table 6.1 shows the unmodified DNA oligonucleotide sequences used in biophysical characterization.

Table 6.1: Sequences used for G-quadruplex characterization (electrophoretic mobility shift assay, circular dichroism spectroscopy, thermal difference spectra, and circular dichroism melting). Mutations are represented in lower case bold, deletions by a dash.

Name	Sequence
hZIP1-1	TGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGA
hZIP1-2	------AGGGGGGAGCTATGGGAGGGGGCGGGA
hZIP1-3	TGGGGGAGGGGGGAGCTATGGGAGGGGGC----
c1	TCCCGCCCCCTCCCATAGCTCCCCCCTCCCCCA
c2	TCCCGCCCCCTCCCATAGCTCCCCCCT------
c3	----GCCCCCTCCCATAGCTCCCCCCTCCCCCA
Mut1	T**ttttt**AGGGGGGAGCTATGGGAGGGGGCGGGA
Mut2	TGGGGGA**tttttt**AGCTATGGGAGGGGGCGGGA
Mut3	TGGGGGAGGGGGGAGCTAT**ttt**AGGGGGCGGGA
Mut4	TGGGGGAGGGGGGAGCTATGGGA**ttttt**CGGGA
Mut5	TGGGGGAGGGGGGAGCTATGGGAGGGGGC**ttt**A
MutB	TGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGA
MutC	GGGttAGGGGGGAGCTATGttAGGGGGCGGG
MutE	TG**tttt**AGG**ttt**GAGCTATGGGAGGGGGCGGGA
MutF	TGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGA
MutG	TG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGA
MutH	TG**tt**GGAGG**ttt**GAGCTATGGGAGGGGGCGGGA

Electrophoretic Mobility Shift Assay (EMSA)

12% or 16% polyacrylamide non-denaturing gels were prepared using TBE with or without 20 mM MCl (M= Li, Na or K) and pre-ran at 90 V for 30 minutes inside an ice box. Samples were loaded in a buffer lacking dyes (TBE with or without 20 mM MCl, where M=Li, Na or K and 10% glycerol) and ran in the corresponding buffer at 90 V for 80-120 minutes inside an ice box. Gels were stained using SYBR® Gold (Life Technologies) except when fluorescently labelled and imaged using GE Healthcare Life Sciences Ettan DIGE Imager. O'RangeRuler 5 bp DNA Ladder (ThermoFisher) was used for reference.

Circular Dichroism Spectroscopy (CD)

DNA samples were prepared at 5 µM in 50 mM Tris buffer with or without 100 mM KCl, to a final volume of 500 µL and annealed as previously discussed. The spectra were recorded using a Jasco J-715 spectropolarimeter with PTC-348W1 temperature control using a 0.2 cm length cuvette at 20°C;

data pitch of 0.1 nm, bandwidth of 2 nm, 4 seconds response, and a scanning speed of 100 nm/min and an accumulation of 3 in a 320-200 nm range. Spectra were recorded at least twice and a representative curve is presented. Data is shown in Ellipticity/ mdeg (millidegrees) or Molar Ellipticity/ deg cm^2 $dmol^{-1}$.

THERMAL DIFFERENCE SPECTRA (TDS)

This was done according to the protocol described by Mergny *et al* 2005.[166] DNA samples were prepared at 5 µM in 10 mM Li cacodylate buffer with 100 mM KCl (or NaCl or LiCl), to a final volume of 500 µL and annealed as previously discussed. The spectra were recorded using a UV-Vis Spectrophotometer (PerkinElmer) with temperature control at 20°C and 90°C, the first spectrum was subtracted from the second giving the thermal difference spectra which was then normalised by dividing the data by maximum value to give comparable results.

POLYMERASE STOP ASSAY (PSA)

This assay was done following the protocol described by Sun and Hurley 2010[161] with small modifications. In brief, a long sequence was designed such that it contained 11 flanking bases in the 5'-end, the G-quadruplex forming sequence in the middle and 33 flanking bases in the 3'-end, the last 22 of them being the reverse compliment of T7 EEV primer. This sequence was annealed with FAM-labelled T7 EEV primer (5'-FAM-AAGGCTAGAGTACTTAATACGA-3') in 50 mM Tris buffer the presence or absence of 100 mM KCl overnight. A 20 µL polymerase reaction was set using Taq buffer with KCl or $(NH_4)_2SO_4$ from Fermentas at 30°C for 30 minutes. The reaction was stopped with formamide loading buffer and loaded to a 20% polyacrylamide denaturing gel containing 8 M urea for analysis.

Table 6.2: G-quadruplex characterization using polymerase stop assay. Mutations are represented in lower case bold, italics indicate primer complementary sequence.

Name	Mod.5'-	Sequence
T7 EEV	FAM	AAGGCTAGAGTACTTAATACGA
PSA-hZIP1	None	GCGGCTCCTGTGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGAA TTCCCGAC*TTCGTATTAAGTACTCTAGCCTT*
PSA-MutB	None	GCGGCTCCTGTGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGAA TTCCCGAC*TTCGTATTAAGTACTCTAGCCTT*
PSA-MutE	None	GCGGCTCCTGTG**tttt**AGG**ttt**GAGCTATGGGAGGGGGCGGGAA TTCCCGAC*TTCGTATTAAGTACTCTAGCCTT*
PSA-MutF	None	GCGGCTCCTGTGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGAA TTCCCGAC*TTCGTATTAAGTACTCTAGCCTT*
PSA-MutG	None	GCGGCTCCTGTG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGAA TTCCCGAC*TTCGTATTAAGTACTCTAGCCTT*
PSA-MutH	None	GCGGCTCCTGTG**tt**GGAGG**ttt**GAGCTATGGGAGGGGGCGGGAA TTCCCGAC*TTCGTATTAAGTACTCTAGCCTT*

Melting and Annealing

For CD melting and annealing, DNA samples were prepared at 5 µM in 10 mM Li cacodylate and melted from 20°C to 95°C monitoring the 263 nm wavelength at a 1°C/min rate except where otherwise stated. For annealing experiments, measurements started at 95°C and were brought to 20°C using the same sample that had been melted. Melting and annealing curves were repeated twice or thrice.

For FRET melting and annealing, double labelled (FAM/TAMRA) DNA samples were prepared using Method 3 at 0.4 µM in 10 mM Li cacodylate with 100 mM KCl and melted from 25°C to 95°C monitoring FAM fluorescence with Stratagene Mx3005p thermocycler from Agilent every 30 seconds with a 1°C/min rate. For annealing experiments, measurements started at 95°C and were brought to 25°C using the same sample that had been melted. Measurements were repeated at least twice. The oligonucleotides used in this assay are found in Table 6.3.

For both methods, the baseline was corrected assuming a folded/unfolded model and the data is presented as folded fraction as described by Mergny and Lacroix[36] to calculate T_m or $T_{1/2}$.

Table 6.3: Labelled sequences for FRET-based ligand screening. Mutations are represented in lower case bold.

Name	Mod.5'-	Sequence	Mod. 3'-
F-hZIP1-T	FAM	TGGGGGAGGGGGGAGCTATGGGAGGGGGCGGGA	TAMRA
F-hZIP1-2-T	FAM	------AGGGGGGAGCTATGGGAGGGGGCGGGA	TAMRA
F-hZIP1-3-T	FAM	TGGGGGAGGGGGGAGCTATGGGAGGGGGC----	TAMRA
F-MutB-T	FAM	TGGG**tt**AGGGGGGAGCTATGGGAGGGGGCGGGA	TAMRA
F-MutE-T	FAM	TG**tttt**AGG**ttt**GAGCTATGGGAGGGGGCGGGA	TAMRA
F-MutF-T	FAM	TGGGGGAGG**ttt**GAGCTATGGGAGGGGGCGGGA	TAMRA
F-MutG-T	FAM	TG**tt**GGAGGGGGGAGCTATGGGAGGGGGCGGGA	TAMRA

Kinetic Stability Assay

Double labelled DNA samples were annealed at 0.4 µM by Method 3 discussed above in 10 mM Li cacodylate buffer with either 100 mM KCl or 10 mM KCl supplemented with 90 mM LiCl, and mixed with an equal volume of 40 µM of their reverse complementary strand containing either 60 mM KCl or 10 mM KCl supplemented with 90 mM LiCl to give a final concentration of 80 mM KCl or 10 mM KCl. FAM fluorescence was monitored using Stratagene Mx3005p thermocycler from Agilent at 37 °C every 30 seconds for two hours. Double labelled sequences used can be found in Table 6.3 and reverse complementary strands are found in Table 6.1.

LIGANDS FOR G-QUADRUPLEX STABILIZATION

Different ligands were tested against the sequence of interest. Pyridostatin and PhenDC3 were kind gifts from Dr. Marco Di Antonio (University of Cambridge, UK). BRACO-19 was purchased from Sigma Aldrich and TMPyP4 from Tocris. Ligands were prepared in 10 mM stock solutions in either water or DMSO and kept at -20 °C.

LIGAND STABILIZATION USING FRET MELTING

This assay was designed following Renciuk *et al* 2012[222] methodology to assess ligand interaction with the sequence of interest. Labelled (FAM/TAMRA) used can be found in Table 6.3, each sequence was annealed following METHOD 3 discussed above in 10 mM Li cacodylate buffer with 10 mM KCl and 90 mM LiCl with a DNA final concentration of 0.4 µM. Working solutions of stabilizing ligands were prepared at 2x desired concentrations in 10 mM Li cacodylate, 10 mM KCl, 90 mM LiCl buffer and added to the bottom of a clear 96-well plate, an equal volume of DNA was added to the ligand solution and mixed by pipetting. The mixture was incubated at room temperature for five minutes and then melted using Stratagene Mx3005p thermocycler from Agilent. FAM fluorescence was recorded every 0.5 °C from 25 °C to 95 °C with a 1 °C/min melting rate. Melting curves were recorded three times, normalised and averaged.

PLASMID CONSTRUCTION

The pGL3-hZIP1 plasmid was kindly provided by Dr. Peter Makhov (Fox Chase Cancer Center, Philadelphia, USA) containing the core promoter of *hZIP1,*[150] and is referred here as pGL3-hZIP1 or WildType (Wt).

The pGL3-hZIP1 plasmid was modified using site-directed mutagenesis to obtain all relevant mutations pGL3-MutB, pGL3-MutE, pGL3-MutF, pGL3-MutG and pGL3-MutH. Briefly, the plasmid was propagated using *E. coli* DH5α and purified by mini prep kit (Qiagen) following supplier's instructions. It was later amplified using PCR with mismatching primers containing the desired mutations using Pfu polymerase (Agilent) and the PCR product was treated with Dpn1 (NEB). For primers and PCR parameters used please see Table 6.4 and Table 6.5. *E. col*i was transformed with the digested PCR product, three colonies were selected and DNA was purified again by mini prep kit (Qiagen) and sequenced. The purified DNA that contained the desired mutations was later obtained in larger quantities using a maxi prep kit (Qiagen) following supplier's instructions.

Table 6.4: Plasmid construction by site directed mutagenesis

Name	Sequence
MutB FW	CGCGGCTCCTGTGGG**tt**AGGGGGGAGCTATGG
MutB RV	CCATAGCTCCCCCCT**aa**CCCACAGGAGCCGCG
MutE FW	GGTACGCGGCTCCTGTG**tttt**AGG**ttt**GAGCTATGGGAGGGGGC
MutE RV	GCCCCCTCCCATAGCTC**aaa**CCT**aaaa**CACAGGAGCCGCGTACC
MutF FW	CTCCTGTGGGGGAGG**ttt**GAGCTATGGGAGGGG
MutF RV	CCCCTCCCATAGCTC**aaa**CCTCCCCCACAGGAG
MutG FW	CGCGGCTCCTGTG**tt**GGAGGGGGGAGCT
MutG RV	AGCTCCCCCCTCC**aa**CACAGGAGCCGCG
MutH FW	GGTACGCGGCTCCTGTG**tt**GGAGG**ttt**GAGCTATGGGAGGGGGC
MutH RV	GCCCCCTCCCATAGCTC**aaa**CCTCC**aa**CACAGGAGCCGCGTACC

Table 6.5: Conditions for PCR to produce site-directed mutagenesis.

Sequence	Segment	Cycles	Temperature (°C)	Time (seconds)
MutB (from wild type)	1	1	95	30
	2	18	95	30
			50	60
			68	300
	3	1	4	Hold
MutE (from MutB)	1	1	95	30
MutG (from wild type)	2	30	95	30
MutH (from wild type)			68	60
			68	300
	3	1	68	300
	4	1	4	Hold
MutF (from wild type)	1	1	95	60
	2	30	95	60
			65	30
			74	300
	3	1	74	300
	4	1	4	Hold

IN VITRO TRANSCRIPTION, RNA PURIFICATION AND QUANTIFICATION

For *in vitro* transcription, HeLa Scribe kit (Promega) was used following supplier's instructions with some modifications by Voss *et al* 2014.[213] HeLa nuclear extracts were mixed with transcription buffer, 3 mM $MgCl_2$, 100 ng of circular plasmid, 1 µM of G4-stabilizing ligand (or water as control) and incubated in ice for five minutes. Reaction was activated adding rNTP mix (final concentration of 0.4 mM each), incubated for 30 minutes at 30˚C. Reaction was stopped using an equivalent volume lysis buffer with 2-mercaptoethanol (PureLink) for total RNA purification and frozen (-80˚C). Samples were thawed for processing and ethanol was added. The mixture was loaded to a purification spin column from PureLink and purified following the supplier's instructions. After RNA was eluted it was treated with Dpn1 (NEB) for 1 hour at 37˚C. The reaction was stopped using lysis buffer with mercapoethanol and ethanol (equivalent volumes) and loaded into a second PureLink spin column. The sample was

purified and treated using DNAse1 on-column treatment from PureLink. RNA was eluted using 30 µL of RNAse-free water and mixed with either pre-prepared RTqPCR or qPCR reaction mixtures (described in buffer and regent preparation section).

Table 6.6: *In vitro* transcription and RNA detection using RT-qPCR. Primers and probe sequences were taken from the literature[214].

Name	Mod.5'-	Sequence	Mod. 3'-
Forward	None	TGCTCCAACACCCCAACATC	None
Reverse	None	ACTTGACTGGCGACGTAATCC	None
Probe	FAM	ACGCAGGTGTCGCAGGTCTTCC	TAMRA

Table 6.7 RTqPCR thermocycler conditions

Segment	Cycles	Temperature (°C)	Time
1	1	50	15 min
2	1	95	2 min
3	40	95	30 sec
		60	1 min (record fluorescence)

CELL CULTURE CONDITIONS

Human prostate cancer cells PC3 (obtained from ECACC) were cultured in F-12K media supplemented with 10% Fetal Bovine Serum and subcultured according to the supplier's advice.

Human prostate cells PNT1A (obtained from ECACC) were cultured in RPMI 1640 media supplemented with 10% Fetal Bovine Serum and subcultured according to the supplier's advice.

Mouse fibroblast embryonic cells NIH 3T3 (obtained from ECACC) were cultured in DMEM supplemented with 10% Calf Serum and subcultured according to the supplier's advice.

LUCIFERASE REPORTER ASSAY

Cells were seeded in 24-well plates at $6x10^4$ cells per well and incubated for 24 hours at 37°C and 5% CO_2, and then co-transfected with 0.5 µg of pGL3 construct and 0.01 µg of pRL-TK using TransIT-LT1 transfection reagent (Mirus) (ratio of 3 µL of transfection reagent per 1 µg of DNA) and following manufacturer's instructions. Cells were lysed at 24 or 48 hours after transfection. Firefly and renilla luciferase luminescence values were recorded with Promega GloMax 20/20 Luminometer using Dual-

Luciferase reporter assay System (Promega) according to manufacturer's instructions except using 50 µL of LARII and Stop&Glo per reaction.

For cells treated using a G4 ligands, the media was removed 18 or 24 hours after transfection and washed with PBS, and new complete growth media was added containing 1 µM or 10 µM of either Pyridostatin, BRACO-19 or PhenDC3. Cells were subsequently lysed at either 24 hours (6 hours of ligand exposure) or at 48 hours (24 hours exposure) and luminescence was recorded as before.

REFERENCES

1. Watson, J. & Crick, F. H. F. Molecular structure of nucleic acids. Nature 171, 737–7388 (1953).

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

3. Watson, J. D. & Crick, F. H. Genetical implications of the structure of deoxyribonucleic acid. Nature 171, 964–967 (1953).

4. 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).

5. Gao, Y. G., Robinson, H. & Wang, A. H. High-resolution A-DNA crystal structures of d(AGGGGCCCCT). An A-DNA model of poly(dG) x poly(dC). Eur. J. Biochem. 261, 413–420 (1999).

6. Drew, H. R. et al. Structure of a B-DNA dodecamer: conformation and dynamics. Proc. Natl. Acad. Sci. U. S. A. 78, 2179–2183 (1981).

7. Brzezinski, K. et al. High regularity of Z-DNA revealed by ultra high-resolution crystal structure at 0.55 A. Nucleic Acids Res. 39, 6238–6248 (2011).

8. Tarköy, M., Phipps, A. K., Schultze, P. & Feigon, J. Solution structure of an intramolecular DNA triplex linked by hexakis(ethylene glycol) units: d(AGAGAGAA-(EG)6-TTCTCTCT-(EG)6-TCTCTCTT). Biochemistry 37, 5810–5819 (1998).

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

10. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic. Acid. Proc. Natl. Acad. Sci. U. S. A. 48, 2013–2018 (1962).

11. Rich, A. The molecular structure of polyinosinic acid. Biochim. Biophys. Acta 29, 502–509 (1958).

12. Thiele, D. & Guschlbauer, W. The structures of polyinosinic acid. Biophysik 9, 261–277 (1973).

13. Arnott, S., Chandrasekaran, R. & Marttila, C. M. Structures for polyinosinic acid and polyguanylic acid. Biochem. J. 141, 537–543 (1974).

14. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I. & Blackburn, E. H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. Cell 51, 899–908 (1987).

15. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature 334, 364–366 (1988).

16. Hoogsteen, K. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. Acta Cryst 12, 822–823 (1959).

17. Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature 342, 825–829 (1989).

18. Panyutin, I. G. et al. G-DNA: a twice-folded DNA structure adopted by single-stranded oligo(dG) and its implications for telomeres. Proc. Natl. Acad. Sci. 87, 867–870 (1990).

19. Parkinson, G. N. in Quadruplex Nucleic Acids (eds. Neidle, S. & Balasubramanian, S.) 1–30 (Royal Society of Chemistry, 2006).

20. Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M. & Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 A resolution. J. Mol. Biol. 273, 171–182 (1997).

21. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. Crystal structure of four-stranded Oxytricha telomeric DNA. Nature 356, 126–131 (1992).

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

23. 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).

24. Marathias, V. M. & Bolton, P. H. Structures of the potassium-saturated, 2:1, and intermediate, 1:1, forms of a quadruplex DNA. Nucleic Acids Res. 28, 1969–1977 (2000).

25. Hu, L., Lim, K. W., Bouaziz, S. & Phan, A. T. Giardia telomeric sequence D(TAGGG)4 forms two intramolecular G-quadruplexes in K+ solution: Effect of loop length and sequence on the folding topology. J. Am. Chem. Soc. 131, 16824–16831 (2009).

26. Lim, K. W. et al. Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. J. Am. Chem. Soc. 132, 12331–12342 (2010).

27. Masiero, S. et al. A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. Org. Biomol. Chem. 8, 2683–2692 (2010).

28. Sen, D. & Gilbert, W. A sodium-potassium switch in the formation of four-stranded G4-DNA. Nature 344, 410–414 (1990).

29. Miyoshi, D., Nakao, a & Sugimoto, N. Structural transition of d(G4T4G4) from antiparallel to parallel G-quartet induced by divalent cations. Nucleic Acids Res. Suppl. 2, 259–260 (2001).

30. Miyoshi, D., Nakao, A. & Sugimoto, N. Molecular crowding regulates the structural switch of the DNA G-quadruplex. Biochemistry 41, 15017–15024 (2002).

31. Mergny, J.-L. et al. in Quadruplex Nucleic Acids (eds. Neidle, S. & Balasubramanian, S.) 31–80 (Royal Society of Chemistry, 2006).

32. Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat. Chem. 6, 75–80 (2014).

33. Travascio, P., Li, Y. & Sen, D. DNA-enhanced peroxidase activity of a DNA aptamer-hemin complex. Chem. Biol.5, 505-517, (1998).

34. Ruttkay-Nedecky, B. et al. G-Quadruplexes as Sensing Probes. Molecules 18, 14760–14779 (2013).

35. Lane, A. N. The stability of intramolecular DNA G-quadruplexes compared with other macromolecules. Biochimie 94, 277–286 (2012).

36. Mergny, J.-L. & Lacroix, L. UV Melting of G-Quadruplexes. Curr. Protoc. Nucleic Acid Chem. 17, Unit 17.1 (2009).

37. 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, 1-8,(2010).

38. De Rache, A. & Mergny, J.-L. Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. Biochimie 115, 194–202 (2015).

39. Rachwal, P. A., Brown, T. & Fox, K. R. Sequence effects of single base loops in intramolecular quadruplex DNA. FEBS Lett. 581, 1657–1660 (2007).

40. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. Nucleic Acids Res. 36, 5482–515 (2008).

41. Zhang, A. Y. Q. & Balasubramanian, S. The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. J. Am. Chem. Soc. 134, 19297–19308 (2012).

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

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

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

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

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

47. Zhou, W. et al. Possible regulatory roles of promoter g-quadruplexes in cardiac function-related genes - human TnIc as a model. PLoS One 8, e53137 (2013).

48. Andorf, C. M. et al. G-Quadruplex (G4) Motifs in the Maize (Zea mays L.) genome are enriched at specific locations in thousands of genes coupled to energy status, hypoxia, low sugar, and nutrient deprivation. J. Genet. Genomics 41, 627–647 (2014).

49. Rawal, P. et al. Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. Genome Res. 16, 644–655 (2006).

50. Lim, K. W. et al. Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. Nucleic Acids Res. 43, 5630–5646 (2015).

51. Chambers, V. S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. advance on, 1–7 (2015).

52. Menendez, C., Frees, S. & Bagga, P. S. QGRS-H Predictor: a web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. Nucleic Acids Res. 40, W96–W103 (2012).

53. Kikin, O., D'Antonio, L. & Bagga, P. S. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 34, W676–W682 (2006).

54. Yadav, V. K., Abraham, J. K., Mani, P., Kulshrestha, R. & Chowdhury, S. QuadBase: Genome-wide database of G4 DNA - Occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. Nucleic Acids Res. 36, 381–385 (2008).

55. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res. 34, 3887–3896 (2006).

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

57. Brooks, T. a & Hurley, L. H. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. Nat. Rev. Cancer 9, 849–861 (2009).

58. 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).

59. Doghaei, a V., Housaindokht, M. R. & Bozorgmehr, M. R. Molecular crowding effects on conformation and stability of G-quadruplex DNA structure: Insights from molecular dynamics simulation. J. Theor. Biol. 364, 103–112 (2014).

60. Miyoshi, D., Karimata, H. & Sugimoto, N. Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions. J. Am. Chem. Soc. 128, 7957–7963 (2006).

61. Xue, Y. et al. Kinetic and thermodynamic control of G-quadruplex folding. Angew. Chem. Int. Ed. Engl. 50, 8046–8050 (2011).

62. Charames, G. S. & Bapat, B. Genomic instability and cancer. Curr. Mol. Med. 3, 589–596 (2003).

63. Tarsounas, M. & Tijsterman, M. Genomes and G-quadruplexes: for better or for worse. J. Mol. Biol. 425, 4782–4789 (2013).

64. Paeschke, K. et al. Pif1 family helicases suppress genome instability at G-quadruplex motifs. Nature 497, 458–462 (2013).

65. Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. J. Biol. Chem. 273, 27587–27592 (1998).

66. Wu, X. & Maizels, N. Substrate-specific inhibition of RecQ helicase. Nucleic Acids Res. 29, 1765–1771 (2001).

67. Masuda-Sasa, T., Polaczek, P., Peng, X. P., Chen, L. & Campbell, J. L. Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. J. Biol. Chem. 283, 24359–24373 (2008).

68. Giraldo, R. & Rhodes, D. The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. EMBO J. 13, 2411–2420 (1994).

69. Giraldo, R., Suzuki, M., Chapman, L. & Rhodes, D. Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. Proc. Natl. Acad. Sci. U. S. A. 91, 7658–7662 (1994).

70. Frantz, J. D. & Gilbert, W. A yeast gene product, G4p2, with a specific affinity for quadruplex nucleic acids. J. Biol. Chem. 270, 9413–9419 (1995).

71. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. Nat. Struct. Mol. Biol. 12, 847–854 (2005).

72. Brázda, V., Hároníková, L., Liao, J. & Fojta, M. DNA and RNA Quadruplex-Binding Proteins. Int. J. Mol. Sci. 15, 17493–17517 (2014).

73. Raiber, E.-A., Kranaster, R., Lam, E., Nikan, M. & Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. Nucleic Acids Res. 40, 1–10 (2011).

74. Kumar, P. et al. Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. Nucleic Acids Res. 39, 8005–8016 (2011).

75. Membrino, A., Cogoi, S., Pedersen, E. B. & Xodo, L. E. G4-DNA formation in the HRAS promoter and rational design of decoy oligonucleotides for cancer therapy. PLoS One 6, e24421 (2011).

76. Cogoi, S., Shchekotikhin, A. E. & Xodo, L. E. HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. Nucleic Acids Res. 42, 8379-8388 (2014).

77. Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K. K. & Xodo, L. E. The KRAS promoter

responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. J. Biol. Chem. 285, 22003–22016 (2010).

78. Schaffitzel, C. et al. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proc. Natl. Acad. Sci. U. S. A. 98, 8572–8577 (2001).

79. Fernando, H., Rodriguez, R. & Balasubramanian, S. Selective recognition of a DNA G-quadruplex by an engineered antibody. Biochemistry 47, 9365–9371 (2008).

80. Lam, E. Y. N., Beraldi, D., Tannahill, D. & Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. Nat. Commun. 4, 1796 (2013).

81. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. 5, 182-186 (2013).

82. Rodriguez, R. et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. 8, 301–310 (2012).

83. Shivalingam, A. et al. The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. Nat. Commun. 6, 8178 (2015).

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

85. Kim, N. W. et al. Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. Science (80-.). 266, 2011–2015 (1994).

86. Agarwala, P., Pandey, S. & Maiti, S. The tale of RNA G-quadruplex. Org. Biomol. Chem. 13, 5570–5585 (2015).

87. Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. Biochemistry 44, 2048–2058 (2005).

88. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S. & Patel, D. J. Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. J. Am. Chem. Soc. 129, 4386–4392 (2007).

89. Dai, J., Chen, D., Jones, R. A., Hurley, L. H. & Yang, D. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. Nucleic Acids Res. 34, 5133–5144 (2006).

90. de Wet, J. R., Wood, K. V, DeLuca, M., Helinski, D. R. & Subramani, S. Firefly luciferase gene: structure and expression in mammalian cells. Mol. Cell. Biol. 7, 725–737 (1987).

91. 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. 99, 11593–11598 (2002).

92. Fekete, A. et al. The Guanine-Quadruplex Structure in the Human c-myc Gene's Promoter Is Converted into B-DNA Form by the Human Poly(ADP-Ribose)Polymerase-1. PLoS One 7, e42690 (2012).

93. Brown, R. V., Gaerig, V. C., Simmons, T. & Brooks, T. A. Helping eve overcome ADAM: G-quadruplexes in the ADAM-15 promoter as new molecular targets for breast cancer therapeutics. Molecules 18, 15019–15034 (2013).

94. Li, F., Zhou, J., Xu, M. & Yuan, G. Investigation on the formation, conversion and bioactivity of a G-quadruplex structure in the PALB2 gene. Int. J. Biol. Macromol. 83, 242–248 (2016).

95. Perrone, R. et al. A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. J. Med. Chem. 56, 6521–6530 (2013).

96. Palumbo, S. L., Ebbinghaus, S. W. & Hurley, L. H. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. J. Am. Chem. Soc. 131, 10878–10891 (2009).

97. Wang, X. D. et al. Turning off transcription of the bcl-2 gene by stabilizing the bcl-2 promoter quadruplex with quindoline derivatives. J. Med. Chem. 53, 4390–4398 (2010).

98. Sun, H. et al. A newly identified G-quadruplex as a potential target regulating Bcl-2 expression. Biochim. Biophys. Acta 1840, 3052–3057 (2014).

99. Mitchell, T. et al. Downregulation of Androgen Receptor transcription by promoter g-quadruplex stabilization as a potential alternative treatment for castrate-resistant prostate cancer. Biochemistry 52, 1429–1436 (2013).

100. Rankin, S. et al. Putative DNA quadruplex formation within the human c-kit oncogene. J. Am. Chem. Soc. 127, 10584–10589 (2005).

101. Fernando, H. et al. A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. Biochemistry 45, 7854–7860 (2006).

102. McLuckie, K. I. E. et al. G-quadruplex-binding benzo[a]phenoxazines down-regulate c-KIT expression in human gastric carcinoma cells. J. Am. Chem. Soc. 133, 2658–2663 (2011).

103. 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. 131, 12628–12633 (2009).

104. Palumbo, S. L. et al. A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. Nucleic Acids Res. 36, 1755–1769 (2008).

105. De Armond, R., Wood, S., Sun, D., Hurley, L. H. & Ebbinghaus, S. W. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the Hypoxia Inducible Factor 1α promoter. Biochemistry 44, 16341–16350 (2005).

106. Burger, A. M. et al. The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. Cancer Res. 65, 1489–1496 (2005).

107. Sun, D. et al. The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. Mol. Cancer Ther. 7, 880–889 (2008).

108. Wu, Y. et al. Stabilization of VEGF G-quadruplex and inhibition of angiogenesis by quindoline derivatives. Biochim. Biophys. Acta 1840, 2970–2977 (2014).

109. Smestad, J. A. & Maher, L. J. Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription. BMC Med. Genet. 16, 91 (2015).

110. Frausto da Silva, J. J. R. & Williams, R. J. P. The biological chemistry of the elements: The inorganic chemistry of life. (Clarendon Press, 1994).

111. Prasad, A. S. Discovery of human zinc deficiency: 50 years later. J. Trace Elem. Med. Biol. 26, 66–69 (2012).

112. Brandão-Neto, J., Stefan, V. & Mendonça, B. The essential role of zinc in growth. Nutr. Res. 15, 335–358 (1995).

113. Rink, L. Zinc and the immune system. Proc. Nutr. Soc. 59, 541–552 (2000).

114. Salgueiro, M., Zubillaga, M. & Lysionek, A. Zinc as an essential micronutrient: a review. Nutr. Res. 20, 737–755 (2000).

115. Fukada, T., Yamasaki, S., Nishida, K., Murakami, M. & Hirano, T. Zinc homeostasis and signaling in health and diseases: Zinc signaling. J. Biol. Inorg. Chem. 16, 1123–1134 (2011).

116. Colvin, R. a, Holmes, W. R., Fontaine, C. P. & Maret, W. Cytosolic zinc buffering and muffling: their role in intracellular zinc homeostasis. Metallomics 2, 306–317 (2010).

117. Györkey, F., Min, K., Huff, J. A., And & Phyllis, G. Zinc and magnesium in human prostate gland :normal, hypoplastic, and neoplastic. Cancer Res. 27, 1348–1353 (1967).

118. Costello, L. C. & Franklin, R. B. Novel role of zinc in the regulation of prostate citrate metabolism and its implications in prostate cancer. Prostate 35, 285–296 (1998).

119. Franklin, R. B. & Costello, L. C. The important role of the apoptotic effects of zinc in the development of cancers. J. Cell. Biochem. 106, 750–757 (2009).

120. Liu, Y., Franklin, R. B. & Costello, L. C. Prolactin and testosterone regulation of mitochondrial zinc in prostate epithelial cells. Prostate 30, 26–32 (1997).

121. Costello, L. C., Liu, Y., Franklin, R. B. & Kennedy, M. C. Zinc inhibition of mitochondrial aconitase and its importance in citrate metabolism of prostate epithelial cells. J. Biol. Chem. 272, 28875–28881 (1997).

122. Johnson, L. a, Kanak, M. a, Kajdacsy-Balla, A., Pestaner, J. P. & Bagasra, O. Differential zinc accumulation and expression of human zinc transporter 1 (hZIP1) in prostate glands. Methods 52, 316–321 (2010).

123. Costello, L. C., Feng, P., Milon, B., Tan, M. & Franklin, R. B. Role of zinc in the pathogenesis and treatment of prostate cancer: critical issues to resolve. Prostate Cancer Prostatic Dis. 7, 111–117 (2004).

124. Kolenko, V., Teper, E., Kutikov, A. & Uzzo, R. Zinc and zinc transporters in prostate carcinogenesis. Nat. Rev. Urol. 10, 219–226 (2013).

125. Huang, L., Kirschke, C. P. & Zhang, Y. Decreased intracellular zinc in human tumorigenic prostate epithelial cells: a possible role in prostate cancer progression. Cancer Cell Int. 6, 10 (2006).

126. Golovine, K. V et al. Overexpression of the zinc uptake transporter hZIP1 inhibits nuclear factor-kappaB and reduces the malignant potential of prostate cancer cells *in vitro* and *in vivo*. Clin. Cancer Res. 14, 5376–5384 (2008).

127. Cousins, R. J., Liuzzi, J. P. & Lichten, L. a. Mammalian zinc transport, trafficking, and signals. J. Biol. Chem. 281, 24085–24089 (2006).

128. Costello, L. C., Liu, Y., Zou, J. & Franklin, R. B. Evidence for a zinc uptake transporter in human prostate cancer cells which is regulated by prolactin and testosterone. J. Biol. Chem. 274, 17499–17504 (1999).

129. Liuzzi, J. P. & Cousins, R. J. Mammalian zinc transporters. Annu. Rev. Nutr. 24, 151–172 (2004).

130. Zhao, H. & Eide, D. J. The yeast ZRT1 gene encodes the zinc transporter protein of a high-affinity uptake system induced by zinc limitation. Proc. Natl. Acad. Sci. U. S. A. 93, 2454–2458 (1996).

131. Gaither, L. A. & Eide, D. J. Eukaryotic zinc transporters and their regulation. BioMetals 14, 251–270 (2001).

132. Eide, D. J. The SLC39 family of metal ion transporters. Pflugers Arch. 447, 796–800 (2004).

133. Milon, B., Dhermy, D., Pountney, D., Bourgeois, M. & Beaumont, C. Differential subcellular localization of hZip1 in adherent and non-adherent cells. FEBS Lett. 507, 241–246 (2001).

134. Jeong, J. & Eide, D. J. The SLC39 family of zinc transporters. Mol. Aspects Med. 34, 612–619 (2013).

135. Michalczyk, A. a & Ackland, M. L. hZip1 (hSLC39A1) regulates zinc homoeostasis in gut epithelial cells. Genes Nutr. 8, 475–486 (2013).

136. Gaither, L. a & Eide, D. J. The human ZIP1 transporter mediates zinc uptake in human K562 erythroleukemia cells. J. Biol. Chem. 276, 22258–22264 (2001).

137. Franklin, R. B. et al. Human ZIP1 is a major zinc uptake transporter for the accumulation of zinc in prostate cells. J. Inorg. Biochem. 96, 435–442 (2003).

138. Dufner-Beattie, J., Langmade, S. J., Wang, F., Eide, D. J. & Andrews, G. K. Structure, function, and regulation of a subfamily of mouse zinc transporter genes. J. Biol. Chem. 278, 50142–50150 (2003).

139. Milon, B., Wu, Q., Zou, J., Costello, L. C. & Franklin, R. B. Histidine residues in the region between transmembrane domains III and IV of hZip1 are required for zinc transport across the plasma membrane in PC-3 cells. Biochim. Biophys. Acta 1758, 1696–1701 (2006).

140. Showell, C. & Conlon, F. L. Mouse ZIP1 and ZIP3 Genes Together Are Essential for Adaptation

to Dietary Zinc Deficiency During Pregnancy. Genesis 45, 418–426 (2007).

141. Franklin, R. B. et al. hZIP1 zinc uptake transporter down regulation and zinc depletion in prostate cancer. Mol. Cancer 4, 32 (2005).

142. Costello, L. C. & Franklin, R. B. The clinical relevance of the metabolism of prostate cancer; zinc and tumor suppression: connecting the dots. Mol. Cancer 5, 17 (2006).

143. Dong, X. et al. hZIP1 that is down-regulated in clear cell renal cell carcinoma is negatively associated with the malignant potential of the tumor. Urol. Oncol. 32, 885–892 (2014).

144. Lang, M. et al. Genetic inhibition of solute-linked carrier 39 family transporter 1 ameliorates aβ pathology in a Drosophila model of Alzheimer's disease. PLoS Genet. 8, e1002683 (2012).

145. Franklin, R. B., Milon, B., Feng, P. & Costello, L. C. Zinc and zinc transporters in normal prostate function and the pathogenesis of prostate cancer. Front. Biosci. 10, 2230 (2005).

146. Jackson, K. a, Valentine, R. a, Coneyworth, L. J., Mathers, J. C. & Ford, D. Mechanisms of mammalian zinc-regulated gene expression. Biochem. Soc. Trans. 36, 1262–1266 (2008).

147. Wimmer, U., Wang, Y., Georgiev, O. & Schaffner, W. Two major branches of anti-cadmium defense in the mouse: MTF-1/metallothioneins and glutathione. Nucleic Acids Res. 33, 5715–5727 (2005).

148. Zheng, D., Feeney, G. P., Kille, P. & Hogstrand, C. Regulation of ZIP and ZnT zinc transporters in zebrafish gill: zinc repression of ZIP10 transcription by an intronic MRE cluster. Physiol. Genomics 34, 205–214 (2008).

149. Weaver, B. P., Dufner-Beattie, J., Kambe, T. & Andrews, G. K. Novel zinc-responsive post-transcriptional mechanisms reciprocally regulate expression of the mouse Slc39a4 and Slc39a5 zinc transporters (Zip4 and Zip5). Biol. Chem. 388, 1301–1312 (2007).

150. Makhov, P. B. et al. Transcriptional regulation of the major zinc uptake protein hZip1 in prostate cancer cells. Gene 431, 39–46 (2009).

151. Makhov, P. B. et al. Reversal of epigenetic silencing of AP-2alpha results in increased zinc uptake in DU-145 and LNCaP prostate cancer cells. Carcinogenesis 32, 1773–1781 (2011).

152. Milon, B., Agyapong, A., Bautista, R., Costello, L. C. & Franklin, R. B. Ras responsive element binding protein-1 (RREB-1) down-regulates hZIP1 expression in prostate cancer cells. Prostate 70, 288–296 (2010).

153. Zou, J., Milon, B., Desouki, M. M., Costello, L. C. & Franklin, R. B. hZIP1 zinc transporter down-regulation in prostate cancer involves the overexpression of ras responsive element binding protein-1 (RREB-1). Prostate 1524, 1518–1524 (2011).

154. Magklara, A. & Lomvardas, S. in Gene Regulatory Sequences and Human Disease (ed. Ahituv, N.) 253–279 (Springer New York, 2012).

155. Mihelich, B. L. et al. miR-183-96-182 cluster is overexpressed in prostate tissue and regulates zinc homeostasis in prostate cells. J. Biol. Chem. 286, 44503–445011 (2011).

156. Dreos, R., Ambrosini, G., Périer, R. C. & Bucher, P. The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. Nucleic Acids Res. 43, D92–6 (2015).

157. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539–539 (2014).

158. Solovyev, V. V., Shahmuradov, I. A. & Salamov, A. A. in Computational Biology of Transcription Factor Binding (ed. Ladunga, I.) 674, 57–83 (Humana Press, 2010).

159. Dai, J., Carver, M. & Yang, D. Polymorphism of human telomeric quadruplex structures. Biochimie 90, 1172–1183 (2008).

160. Le, H. T. et al. Not all G-quadruplexes are created equally: an investigation of the structural polymorphism of the c-Myc G-quadruplex-forming sequence and its interaction with the porphyrin TMPyP4. Org. Biomol. Chem. 10, 9393–9404 (2012).

161. Sun, D. & Hurley, L. H. in G-Quadruplex DNA: Methods and Protocols (ed. Baumann, P.) 608, 65–79 (Humana Press, 2010).

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

163. Petraccone, L. et al. Structure and stability of higher-order human telomeric quadruplexes. J. Am. Chem. Soc. 133, 20951–20961 (2011).

164. Carisson, C., Johnson, M. & Åkerman, B. Double bands in DNA gel elctrophoresis caused by bis-intercalating dyes. Nucleic Acids Res. 23, 2413–2420 (1995).

165. 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).

166. 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, e138 (2005).

167. Weitzmann, M. N., Woodford, K. J. & Usdin, K. The Development and Use of a DNA Polymerase Arrest Assay for the Evaluation of Parameters Affecting Intrastrand Tetraplex Formation. J. Biol. Chem. 271, 20958–20964 (1996).

168. Mergny, J.-L., Phan, a T. & Lacroix, L. Following G-quartet formation by UV-spectroscopy. FEBS Lett. 435, 74–78 (1998).

169. Rachwal, P. a & Fox, K. R. Quadruplex melting. Methods 43, 291–301 (2007).

170. Mergny, J.-L. & Lacroix, L. Analysis of thermal melting curves. Oligonucleotides 537, 515–537 (2003).

171. 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).

172. Gray, R. D., Trent, J. O. & Chaires, J. B. Folding and unfolding pathways of the human telomeric G-quadruplex. J. Mol. Biol. 426, 1629–1650 (2014).

173. Olsen, C. M. & Marky, L. A. in G-Quadruplex DNA: Methods and Protocols (ed. Baumann, P.) 608, 147–158 (2010).

174. Raghuraman, M. K. & Cech, T. R. Effect of monovalent cation-induced telomeric DNA structure on the binding of Oxytricha telomeric protein. Nucleic Acids Res. 18, 4543–4552 (1990).

175. Shivalingam, A. Small Molecule Probes for the Elucidation of G-quadruplex Function. (Imperial College London, 2015).

176. 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).

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

178. Read, M. et al. Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. Proc. Natl. Acad. Sci. U. S. A. 98, 4844–4849 (2001).

179. Wheelhouse, R. T., Sun, D., Han, H., Han, F. X. & Hurley, L. H. Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(N -methyl-4-pyridyl)porphine with Quadruplex DNA. J. Am. Chem. Soc. 120, 3261–3262 (1998).

180. Pennarun, G. et al. Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. Oncogene 24, 2917–2928 (2005).

181. Reed, J. E., Arnal, A. A., Neidle, S. & Vilar, R. Stabilization of G-quadruplex DNA and inhibition of telomerase activity by square-planar nickel(II) complexes. J. Am. Chem. Soc. 128, 5992–5993 (2006).

182. Shin-ya, K. et al. Telomestatin, a Novel Telomerase Inhibitor from Streptomyces anulatus. J. Am. Chem. Soc. 123, 1262–1263 (2001).

183. Yaku, H. et al. Study on effects of molecular crowding on G-quadruplex-ligand binding and ligand-mediated telomerase inhibition. Methods 64, 19–27 (2013).

184. 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).

185. Monchaud, D. et al. 'One ring to bind them all'-part I: the efficiency of the macrocyclic scaffold for g-quadruplex DNA recognition. J. Nucleic Acids 2010, 1–21 (2010).

186. Doi, T., Yoshida, M., Shin-Ya, K. & Takahashi, T. Total synthesis of (R)-telomestatin. Org. Lett. 8, 4165–4167 (2006).

187. Chung, W. J. et al. Solution Structure of an Intramolecular (3 + 1) Human Telomeric G-Quadruplex Bound to a Telomestatin Derivative. J. Am. Chem. Soc. 135, 13495–13501 (2013).

188. Cui, X., Lin, S. & Yuan, G. Spectroscopic probing of recognition of the G-quadruplex in c-kit promoter by small-molecule natural products. Int. J. Biol. Macromol. 50, 996–1001 (2012).

189. Li, Y., Geyer, C. R. & Sen, D. Recognition of anionic porphyrins by DNA aptamers. Biochemistry 35, 6911–6922 (1996).

190. Ren, J. & Chaires, J. B. Sequence and structural selectivity of nucleic acid binding ligands. Biochemistry 38, 16067–16075 (1999).

191. Nicoludis, J. M., Barrett, S. P., Mergny, J.-L. & Yatsunyk, L. a. Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. Nucleic Acids Res. 40, 5432–5447 (2012).

192. Georgiades, S. N., Abd Karim, N. H., Suntharalingam, K. & Vilar, R. Interaction of metal complexes with G-quadruplex DNA. Angew. Chem. Int. Ed. Engl. 49, 4020–4034 (2010).

193. Abd Karim, N. H. et al. Salphen metal complexes as tunable G-quadruplex binders and optical probes. RSC Adv. 4, 3355 (2014).

194. De Cian, A., Guittat, L., Shin-Ya, K., Riou, J.-F. & Mergny, J.-L. Affinity and selectivity of G4 ligands measured by FRET. Nucleic Acids Symp. Ser. 49, 235–236 (2005).

195. Kim, M., Gleason-Guzman, M., Izbicka, E., Nishioka, D. & Hurley, L. H. The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. Cancer Res. 63, 3247–3256 (2003).

196. Parkinson, G. N., Ghosh, R. & Neidle, S. Structural basis for binding of porphyrin to human telomeres. Biochemistry 46, 2390–2397 (2007).

197. 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).

198. 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).

199. Han, F. X., Wheelhouse, R. T. & Hurley, L. H. Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. J. Am. Chem. Soc. 121, 3561–3570 (1999).

200. Grand, C. L., Han, H. & Weitman, S. Cationic porphyrin tmpyp4 down-regulates c-myc and human telomerase reverse transcriptase expression and inhibits tumor growth *in vivo*. Mol. Cancer Ther. 1, 565–573 (2002).

201. Rodriguez, R. et al. A novel small molecule that alters shelterin integrity and triggers a dna-damage response at telomeres. J. Am. Chem. Soc. 130, 15758–15759 (2008).

202. Müller, S. et al. Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. Org. Biomol. Chem. 10, 6537–6546 (2012).

203. McLuckie, K. I. E. et al. G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. J. Am. Chem. Soc. 135, 9640–9643 (2013).

204. Müller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated G-quadruplex isolation from human cells. Nat. Chem. 2, 1095–1098 (2010).

205. Di Antonio, M. et al. Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry.

Angew. Chem. Int. Ed. Engl. 51, 11073–8 (2012).

206. Schultes, C. M., Guyen, B., Cuesta, J. & Neidle, S. Synthesis, biophysical and biological evaluation of 3,6-bis-amidoacridines with extended 9-anilino substituents as potent G-quadruplex-binding telomerase inhibitors. Bioorganic Med. Chem. Lett. 14, 4347–4351 (2004).

207. Gowan, S. M. et al. A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and *in vivo* antitumor activity. Mol. Pharmacol. 61, 1154–1162 (2002).

208. Perrone, R. et al. Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. J. Antimicrob. Chemother. 69, 3248–3258 (2014).

209. De Cian, A., DeLemos, E., Mergny, J.-L., Teulade-Fichou, M. P. & Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. J. Am. Chem. Soc. 129, 1856–1857 (2007).

210. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P. & Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. Angew. Chemie - Int. Ed. 53, 999–1002 (2014).

211. Halder, R., Riou, J.-F., Teulade-Fichou, M. P., Frickey, T. & Hartig, J. S. Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. BMC Res. Notes 5, 138 (2012).

212. Carter, P. Site-directed mutagenesis. Biochem. J. 237, 1–7 (1986).

213. Voss, C. et al. A novel, non-radioactive eukaryotic *in vitro* transcription assay for sensitive quantification of RNA polymerase II activity. BMC Mol. Biol. 15, 7 (2014).

214. Kines, K. J. et al. Electroporation facilitates introduction of reporter transgenes and virions into schistosome eggs. PLoS Negl. Trop. Dis. 4, e593, (2010).

215. Brooks, T. a., Kendrick, S. & Hurley, L. H. Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J. 277, 3459–3469 (2010).

216. González, V. & Hurley, L. H. The c- MYC NHE III 1 : Function and Regulation. Annu. Rev. Pharmacol. Toxicol. 50, 111–129 (2010).

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

218. Mittal, S., Chowhan, R. K. & Singh, L. R. Macromolecular crowding: Macromolecules friend or foe. Biochim. Biophys. Acta - Gen. Subj. 1850, 1822–1831 (2015).

219. Klejevskaja, B. Biophysical and functional studies of potential g-quadruplex motifs from the proximal gata4 gene promoter. (Imperial College London, 2014).

220. Schneider, T. D. Consensus sequence Zen. Appl. Bioinformatics 1, 111–119 (2002).

221. Schneider, T. D. Strong minor groove base conservation in sequence logos implies DNA distortion or base flipping during replication and transcription initiation. Nucleic Acids Res. 29, 4881–4891 (2001).

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

APPENDIX

A.1 TRANSCRIPTION FACTOR BINDING PREDICTION

A.1.1 WILD TYPE PROMOTER

A.1.1.1 Promoter sequence WildType

>WildType

GGTACCCCTC AGTCCTCCAC AGATGCGCCC CAGGGCCCCA GCCCCGAGGT CCTTTTCCTA ATATCCGCTC TTGCTCTCGC TCACTCTCCT CAGGTCGCCT GCTCCGCCCG CTCCCGCTCC CGATCTCTGA CTGCTCCTAA CTGACGTCAC TCCCGGTCTG TCCCCGCCCA CTCGGTGCTG CCATTGGCAG TCGGTCGTGG GTACGCGGCT CCTGTGGGGG AGGGGGGAGC TATGGGAGGG GGCGGGAATT CCCGACTCTA GGCCGGAAGC GCGCGGAGAC CATGTAGTGA GACCCTCGCG AGGTGAGTGC GGGCTGTCTG GCAGCGGGGT CCCTCGAG

A.1.1.2 Nsite results for Wild Type

Program Nsite | Version 5.2013

Search for motifs of 1500 Regulatory Elements (REs)

SET of REs: RegsiteAN DB: 1500 human/animal Transcription Regulatory Elements (real or consensus sites) [Last update: 10.03.2006; Softberry Inc]

Search PARAMETRS:

Expected Mean Number : 0.0100000

Statistical Siginicance Level : 0.9500000

Level of homology between known RE and motif: 80%

Variation of Distance between RE Blocks : 20%

NOTE: RE - Regulatory Element/Consensus | AC - Accession No of RE in a given DB

OS - Organism/Species | BF - Binding Factor or One of them

Mism. - Mismatches | Mean. Exp. Number - Mean Expected Number | Up.Conf.Int. - Upper Confidence Interval

==========

QUERY: >WildType

Length of Query Sequence: 338 bp | Nucleotide Frequencies: A - 0.13 G - 0.32 T - 0.20 C - 0.36

.....

RE: 39. AC: RSA00039//OS: human /GENE: FMR1/RE: Sp1-like /BF: Sp family

Motifs on "+" Strand: Mean Exp. Number 0.00423 Up.Conf.Int. 1 Found 1

216 GGGGGAGGG 224 (Mism.= 0)

.....

RE: 41. AC: RSA00041//OS: mouse (Mus musculus domesticus) /GENE: FMR1/RE: Sp1-like /BF: Sp family

Motifs on "+" Strand: Mean Exp. Number 0.00423 Up.Conf.Int. 1 Found 1

217 GGGGAGGGG 225 (Mism.= 0)

.....

RE: 170. AC: RSA00170//OS: human /GENE: beta-IFN/RE: PRD II /BF: NF-kappaB

Motifs on "-" Strand: Mean Exp. Number 0.00795 Up.Conf.Int. 1 Found 1

256 GTcGGgAATTCC 245 (Mism.= 2)

.....

RE: 234. AC: RSA00234//OS: rat /GENE: VMAT2/RE: CRE /BF: CREB; CBP/p300

Motifs on "+" Strand: Mean Exp. Number 0.00274 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00274 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 289. AC: RSA00289//OS: human /GENE: c-myc/RE: E2F BS /BF: E2F-1; E2F-2; E2F-4

Motifs on "+" Strand: Mean Exp. Number 0.00605 Up.Conf.Int. 1 Found 1

241 GGCGGGAA 248 (Mism.= 0)

.....

RE: 350. AC: RSA00350//OS: rat /GENE: spi 2.1/RE: C-II core /BF: C/EBPalpha

Motifs on "-" Strand: Mean Exp. Number 0.00577 Up.Conf.Int. 1 Found 1

61 TTAGGAAAa 53 (Mism.= 1)

.....

RE: 566. AC: RSA00566//OS: mouse /GENE: alphaA-crystallin/RE: alphaA-CRYBP1 BS /BF: alphaA-CRYBP1

Motifs on "+" Strand: Mean Exp. Number 0.00725 Up.Conf.Int. 1 Found 1

244 GGGAAtTCCC 253 (Mism.= 1)

.....

RE: 573. AC: RSA00573//OS: mouse /GENE: MHCIIAalpha/RE: CRE /BF: HB16

Motifs on "+" Strand: Mean Exp. Number 0.00274 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00274 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 621. AC: RSA00621//OS: adenovirus (Ad5) /GENE: E4/RE: CRE /BF: ETF-A

Motifs on "+" Strand: Mean Exp. Number 0.00574 Up.Conf.Int. 1 Found 1

142 TGACGTCAC 150 (Mism.= 1)

.....

RE: 632. AC: RSA00632//OS: human /GENE: H4/RE: GC-rich /BF: H4TF1

Motifs on "+" Strand: Mean Exp. Number 0.00423 Up.Conf.Int. 1 Found 1

216 GGGGGAGGG 224 (Mism.= 0)

.....

RE: 745. AC: RSA00745//OS: human /GENE: H3 (ST519)/RE: GC-box /BF:

Motifs on "+" Strand: Mean Exp. Number 0.00085 Up.Conf.Int. 1 Found 1

265 GGAAGcGCGCGGaGA 279 (Mism.= 2)

.....

RE: 838. AC: RSA00838//OS: human /GENE: endoglin/RE: Sp1 /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00945 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 976. AC: RSA00976//OS: human /GENE: NET/RE: Sp1 /BF: Sp1

Motifs on "+" Strand: Mean Exp. Number 0.00945 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 1121. AC: RSA01121//OS: mouse /GENE: ETF/Tead2/RE: proximal GA element /BF: GBF

Motifs on "+" Strand: Mean Exp. Number 0.00134 Up.Conf.Int. 1 Found 1

216 GGGGGAGGGG 225 (Mism.= 0)

.....

RE: 1163. AC: RSA01163//OS: human /GENE: GSTP1/RE: Sp1-like element II /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00475 Up.Conf.Int. 1 Found 1

239 GGGGCGGGA 247 (Mism.= 0)

.....

Totally 17 motifs of 15 different REs have been found

© 1999 - 2016 www.softberry.com

A.1.1.3 AliBaba2.1 results for Wild Type

```
================================================================================
seq(    0..   59)     ggtacccctcagtcctccacagatgcgccccagggccccagccccgaggtccttttcctа
Segments:
2.3.1.0     24   36                              ======Sp1====
9.9.539     25   34                               ====NF-1==
1.6.1.0     28   37                                  =AP-2alph=
9.9.539     32   41                                      ====NF-1==
2.3.1.0     33   45                                       ======Sp1====
9.9.270     35   44                                         ====ETF===
9.9.701     47   56                                                  =PTF1-beta
2.3.1.0     49   58                                                    ====Sp1===
1.1.3.0     52   61                                                       =C/EBPal
4.3.1.3     53   62                                                        ====MCM
================================================================================
seq(   60..  119)     atatccgctcttgctctcgctcactctcctcaggtcgcctgctccgcccgctcccgctcc
Segments:
1.1.3.0     52   61     p=
4.3.1.3     53   62     1==
1.3.1.2     78   87                        ====USF===
2.1.2.2     87   96                             ==RXR-beta
2.3.1.0     100  114                                      =======Sp1=====
2.3.1.0     110  119                                           ====Sp1===
1.1.3.0     118  127                                              =C
================================================================================
seq(  120..  179)     cgatctctgactgctcctaactgacgtcactcccggtctgtccccgcccactcggtgctg
Segments:
1.1.3.0     118  127    /EBPalp=
2.3.3.0     126  135          =CPE_bind=
1.1.1.1     135  149                   ======JunD=====
9.9.51      139  148                       ====ATF===
0.5.1.0     140  149                        =E1A_12S_=
1.1.1.6     140  149                        ==CRE-BP1=
1.1.2.0     140  149                        ====CREB==
1.1.3.0     140  149                        =C/EBPalp=
2.1.2.2     140  149                        ==RXR-beta
2.3.3.0     141  152                         =CPE_bindin=
2.3.1.0     157  171                                    =======Sp1=====
2.3.2.1     161  170                                        ===Egr-1==
9.9.173     178  187                                                  ==
================================================================================
seq(  180..  239)     ccattggcagtcggtcgtgggtacgcggctcctgtgggggaggggggagctatgggaggg
Segments:
9.9.173     178  187    ==CTF===
1.6.1.0     206  215                              ====AP-2==
9.9.77      213  222                                   =CACCC-bi=
2.3.1.0     214  227                                    ======Sp1=====
2.3.2.1     216  225                                      ==Krox-20=
9.9.270     217  226                                       ====ETF===
2.3.1.0     220  229                                          ====Sp1===
2.3.1.0     232  246                                                  =======S
2.3.2.1     234  243                                                    ==Krox
9.9.270     235  244                                                     ====E
================================================================================
```

```
seq(  240..  299)     ggcgggaattcccgactctaggccggaagcgcgcggagaccatgtagtgagaccctcgcg
Segments:
2.3.1.0    232  246   p1=====
2.3.2.1    234  243   -20=
9.9.270    235  244   TF===
2.3.4.0    243  252      ===MBP-2==
4.1.1.0    243  252      ===c-Rel==
9.9.588    243  252      =NF-kappaB
9.9.590    243  252      =NF-kappaB
9.9.592    243  252      =NF-kappa=
9.9.594    243  252      ====RelA==
3.5.2.0    260  269                    =c-Ets-1_=
2.3.1.0    268  277                           ====Sp1===
================================================================================
seq(  300..  359)     aggtgagtgcgggctgtctggcagcggggtccctcgag
Segments:
2.3.1.0    302  315     ======Sp1=====
9.9.539    313  322                ====NF-1==
2.3.1.0    319  332                     ======Sp1=====

47 segments in this sequence identified as potential binding sites
```

A.1.2 MUTB PROMOTER

A.1.2.1 Promoter sequence MutB

>MutB

GGTACCCCTC AGTCCTCCAC AGATGCGCCC CAGGGCCCCA GCCCCGAGGT CCTTTTCCTA ATATCCGCTC TTGCTCTCGC TCACTCTCCT CAGGTCGCCT GCTCCGCCCG CTCCCGCTCC CGATCTCTGA CTGCTCCTAA CTGACGTCAC TCCCGGTCTG TCCCCGCCCA CTCGGTGCTG CCATTGGCAG TCGGTCGTGG GTACGCGGCT CCTGTGGGTT AGGGGGGAGC TATGGGAGGG GGCGGGAATT CCCGACTCTA GGCCGGAAGC GCGCGGAGAC CATGTAGTGA GACCCTCGCG AGGTGAGTGC GGGCTGTCTG GCAGCGGGGT CCCTCGAG

A.1.2.2 Nsite results for MutB

Program Nsite | Version 5.2013

Search for motifs of 1500 Regulatory Elements (REs)

SET of REs: RegsiteAN DB: 1500 human/animal Transcription Regulatory Elements (real or consensus sites) [Last update: 10.03.2006; Softberry Inc]

Search PARAMETRS:

Expected Mean Number : 0.0100000

Statistical Siginicance Level : 0.9500000

Level of homology between known RE and motif: 80%

Variation of Distance between RE Blocks : 20%

NOTE: RE - Regulatory Element/Consensus | AC - Accession No of RE in a given DB

OS - Organism/Species | BF - Binding Factor or One of them

Mism. - Mismatches | Mean. Exp. Number - Mean Expected Number | Up.Conf.Int. - Upper Confidence Interval

==========

QUERY: >MutB

Length of Query Sequence: 338 bp | Nucleotide Frequencies: A - 0.13 G - 0.31 T - 0.21 C - 0.36

.....

RE: 170. AC: RSA00170//OS: human /GENE: beta-IFN/RE: PRD II /BF: NF-kappaB

Motifs on "-" Strand: Mean Exp. Number 0.00824 Up.Conf.Int. 1 Found 1

256 GTcGGgAATTCC 245 (Mism.= 2)

.....

RE: 234. AC: RSA00234//OS: rat /GENE: VMAT2/RE: CRE /BF: CREB; CBP/p300

Motifs on "+" Strand: Mean Exp. Number 0.00279 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00279 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 289. AC: RSA00289//OS: human /GENE: c-myc/RE: E2F BS /BF: E2F-1; E2F-2; E2F-4

Motifs on "+" Strand: Mean Exp. Number 0.00550 Up.Conf.Int. 1 Found 1

241 GGCGGGAA 248 (Mism.= 0)

.....

RE: 350. AC: RSA00350//OS: rat /GENE: spi 2.1/RE: C-II core /BF: C/EBPalpha

Motifs on "-" Strand: Mean Exp. Number 0.00639 Up.Conf.Int. 1 Found 1

61 TTAGGAAAa 53 (Mism.= 1)

.....

RE: 566. AC: RSA00566//OS: mouse /GENE: alphaA-crystallin/RE: alphaA-CRYBP1 BS /BF: alphaA-CRYBP1

Motifs on "+" Strand: Mean Exp. Number 0.00706 Up.Conf.Int. 1 Found 1

244 GGGAAtTCCC 253 (Mism.= 1)

.....

RE: 573. AC: RSA00573//OS: mouse /GENE: MHCIIAalpha/RE: CRE /BF: HB16

Motifs on "+" Strand: Mean Exp. Number 0.00279 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00279 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 621. AC: RSA00621//OS: adenovirus (Ad5) /GENE: E4/RE: CRE /BF: ETF-A

Motifs on "+" Strand: Mean Exp. Number 0.00586 Up.Conf.Int. 1 Found 1

142 TGACGTCAC 150 (Mism.= 1)

.....

RE: 745. AC: RSA00745//OS: human /GENE: H3 (ST519)/RE: GC-box /BF:

Motifs on "+" Strand: Mean Exp. Number 0.00948 Up.Conf.Int. 1 Found 2

234 GGgAGGGgGCGGGaA 248 (Mism.= 3)

265 GGAAGcGCGCGGaGA 279 (Mism.= 2)

.....

RE: 838. AC: RSA00838//OS: human /GENE: endoglin/RE: Sp1 /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00927 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 976. AC: RSA00976//OS: human /GENE: NET/RE: Sp1 /BF: Sp1

Motifs on "+" Strand: Mean Exp. Number 0.00927 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 1163. AC: RSA01163//OS: human /GENE: GSTP1/RE: Sp1-like element II /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00416 Up.Conf.Int. 1 Found 1

239 GGGGCGGGA 247 (Mism.= 0)

.....

Totally 14 motifs of 11 different REs have been found

© 1999 - 2016 www.softberry.com

A.1.2.3 AliBaba2.1 results for MutB

```
================================================================================
seq(    0..   59)     ggtacccctcagtcctccacagatgcgccccagggccccagccccgaggtccttttcctа
Segments:
2.3.1.0     24   36                            ======Sp1====
9.9.539     25   34                             ====NF-1==
1.6.1.0     28   37                                =AP-2alph=
9.9.539     32   41                                    ====NF-1==
2.3.1.0     33   45                                     ======Sp1====
9.9.270     35   44                                       ====ETF===
9.9.701     47   56                                                 =PTF1-beta
2.3.1.0     49   58                                                   ====Sp1===
1.1.3.0     52   61                                                      =C/EBPal
4.3.1.3     53   62                                                       ====MCM
================================================================================
seq(   60..  119)     atatccgctcttgctctcgctcactctcctcaggtcgcctgctccgcccgctcccgctcc
Segments:
1.1.3.0     52   61     p=
4.3.1.3     53   62     1==
1.3.1.2     78   87                        ====USF===
2.1.2.2     87   96                            ==RXR-beta
2.3.1.0     100  114                                  =======Sp1=====
2.3.1.0     110  119                                        ====Sp1===
1.1.3.0     118  127                                             =C
================================================================================
seq(  120..  179)     cgatctctgactgctcctaactgacgtcactcccggtctgtccccgcccactcggtgctg
Segments:
1.1.3.0     118  127   /EBPalp=
2.3.3.0     126  135         =CPE_bind=
1.1.1.1     135  149                  ======JunD=====
9.9.51      139  148                      ====ATF===
0.5.1.0     140  149                       =E1A_12S_=
1.1.1.6     140  149                       ==CRE-BP1=
1.1.2.0     140  149                       ====CREB==
1.1.3.0     140  149                       =C/EBPalp=
2.1.2.2     140  149                       ==RXR-beta
2.3.3.0     141  152                        =CPE_bindin=
2.3.1.0     157  171                                  =======Sp1=====
2.3.2.1     161  170                                     ===Egr-1==
9.9.173     178  187                                                ==
================================================================================
seq(  180..  239)     ccattggcagtcggtcgtgggtacgcggctcctgtgggttaggggggagctatgggaggg
Segments:
9.9.173     178  187   ==CTF===
1.6.1.0     206  215                          ====AP-2==
1.3.1.2     209  218                             ====USF===
2.3.1.0     221  230                                  ====Sp1===
2.3.1.0     232  246                                       =======S
2.3.2.1     234  243                                         ==Krox
9.9.270     235  244                                          ====E
================================================================================
```

```
seq(  240..  299)     ggcgggaattcccgactctaggccggaagcgcgcggagaccatgtagtgagaccctcgcg
Segments:
2.3.1.0    232  246   p1=====
2.3.2.1    234  243   -20=
9.9.270    235  244   TF===
2.3.4.0    243  252      ===MBP-2==
4.1.1.0    243  252      ===c-Rel==
9.9.588    243  252      =NF-kappaB
9.9.590    243  252      =NF-kappaB
9.9.592    243  252      =NF-kappa=
9.9.594    243  252      ====RelA==
3.5.2.0    260  269                      =c-Ets-1_=
2.3.1.0    268  277                             ====Sp1===
================================================================================
seq(  300..  359)     aggtgagtgcgggctgtctggcagcggggtccctcgag
Segments:
2.3.1.0    302  315     ======Sp1=====
9.9.539    313  322                ====NF-1==
2.3.1.0    319  332                      ======Sp1=====

44 segments in this sequence identified as potential binding sites
```

A.1.3 MutE Promoter

A.1.3.1 Promoter sequence MutE

>MutE

GGTACCCCTC AGTCCTCCAC AGATGCGCCC CAGGGCCCCA GCCCCGAGGT CCTTTTCCTA ATATCCGCTC TTGCTCTCGC TCACTCTCCT CAGGTCGCCT GCTCCGCCCG CTCCCGCTCC CGATCTCTGA CTGCTCCTAA CTGACGTCAC TCCCGGTCTG TCCCCGCCCA CTCGGTGCTG CCATTGGCAG TCGGTCGTGG GTACGCGGCT CCTGTGTTTT AGGTTTGAGC TATGGGAGGG GGCGGGAATT CCCGACTCTA GGCCGGAAGC GCGCGGAGAC CATGTAGTGA GACCCTCGCG AGGTGAGTGC GGGCTGTCTG GCAGCGGGGT CCCTCGAG

A.1.3.2 Nsite results for MutE

Program Nsite | Version 5.2013

Search for motifs of 1500 Regulatory Elements (REs)

SET of REs: RegsiteAN DB: 1500 human/animal Transcription Regulatory Elements (real or consensus sites) [Last update: 10.03.2006; Softberry Inc]

Search PARAMETRS:

Expected Mean Number : 0.0100000

Statistical Siginicance Level : 0.9500000

Level of homology between known RE and motif: 80%

Variation of Distance between RE Blocks : 20%

NOTE: RE - Regulatory Element/Consensus | AC - Accession No of RE in a given DB

OS - Organism/Species | BF - Binding Factor or One of them

Mism. - Mismatches | Mean. Exp. Number - Mean Expected Number | Up.Conf.Int. - Upper Confidence Interval

==========

QUERY: >MutE

Length of Query Sequence: 338 bp | Nucleotide Frequencies: A - 0.13 G - 0.30 T - 0.22 C - 0.36

.....

RE: 170. AC: RSA00170//OS: human /GENE: beta-IFN/RE: PRD II /BF: NF-kappaB

Motifs on "-" Strand: Mean Exp. Number 0.00893 Up.Conf.Int. 1 Found 1

256 GTcGGgAATTCC 245 (Mism.= 2)

.....

RE: 234. AC: RSA00234//OS: rat /GENE: VMAT2/RE: CRE /BF: CREB; CBP/p300

Motifs on "+" Strand: Mean Exp. Number 0.00291 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00291 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 289. AC: RSA00289//OS: human /GENE: c-myc/RE: E2F BS /BF: E2F-1; E2F-2; E2F-4

Motifs on "+" Strand: Mean Exp. Number 0.00431 Up.Conf.Int. 1 Found 1

241 GGCGGGAA 248 (Mism.= 0)

.....

RE: 350. AC: RSA00350//OS: rat /GENE: spi 2.1/RE: C-II core /BF: C/EBPalpha

Motifs on "-" Strand: Mean Exp. Number 0.00815 Up.Conf.Int. 1 Found 1

61 TTAGGAAAa 53 (Mism.= 1)

.....

RE: 566. AC: RSA00566//OS: mouse /GENE: alphaA-crystallin/RE: alphaA-CRYBP1 BS /BF: alphaA-CRYBP1

Motifs on "+" Strand: Mean Exp. Number 0.00656 Up.Conf.Int. 1 Found 1

244 GGGAAtTCCC 253 (Mism.= 1)

.....

RE: 573. AC: RSA00573//OS: mouse /GENE: MHCIIAalpha/RE: CRE /BF: HB16

Motifs on "+" Strand: Mean Exp. Number 0.00291 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00291 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 621. AC: RSA00621//OS: adenovirus (Ad5) /GENE: E4/RE: CRE /BF: ETF-A

Motifs on "+" Strand: Mean Exp. Number 0.00610 Up.Conf.Int. 1 Found 1

142 TGACGTCAC 150 (Mism.= 1)

.....

RE: 745. AC: RSA00745//OS: human /GENE: H3 (ST519)/RE: GC-box /BF:

Motifs on "+" Strand: Mean Exp. Number 0.00649 Up.Conf.Int. 1 Found 2

234 GGgAGGGgGCGGGaA 248 (Mism.= 3)

265 GGAAGcGCGCGGaGA 279 (Mism.= 2)

.....

RE: 838. AC: RSA00838//OS: human /GENE: endoglin/RE: Sp1 /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00883 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 976. AC: RSA00976//OS: human /GENE: NET/RE: Sp1 /BF: Sp1

Motifs on "+" Strand: Mean Exp. Number 0.00883 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 1163. AC: RSA01163//OS: human /GENE: GSTP1/RE: Sp1-like element II /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00296 Up.Conf.Int. 1 Found 1

239 GGGGCGGGA 247 (Mism.= 0)

.....

Totally 14 motifs of 11 different REs have been found

© 1999 - 2016 www.softberry.com

A.1.3.3 Alibaba2.1 results for MutE

```
================================================================================
seq(    0..   59)     ggtacccctcagtcctccacagatgcgccccagggccccagccccgaggtccttttccta
Segments:
2.3.1.0    24   36                          ======Sp1====
9.9.539    25   34                           ====NF-1==
1.6.1.0    28   37                              =AP-2alph=
9.9.539    32   41                                  ====NF-1==
2.3.1.0    33   45                                   ======Sp1====
9.9.270    35   44                                     ====ETF===
9.9.701    47   56                                                =PTF1-beta
2.3.1.0    49   58                                                  ====Sp1===
1.1.3.0    52   61                                                     =C/EBPal
4.3.1.3    53   62                                                      ====MCM
================================================================================
seq(   60..  119)     atatccgctcttgctctcgctcactctcctcaggtcgcctgctccgcccgctcccgctcc
Segments:
1.1.3.0    52   61    p=
4.3.1.3    53   62    1==
1.3.1.2    78   87                    ====USF===
2.1.2.2    87   96                        ==RXR-beta
2.3.1.0   100  114                                 =======Sp1=====
2.3.1.0   110  119                                       ====Sp1===
1.1.3.0   118  127                                             =C
================================================================================
seq(  120..  179)     cgatctctgactgctcctaactgacgtcactcccggtctgtccccgcccactcggtgctg
Segments:
1.1.3.0   118  127   /EBPalp=
2.3.3.0   126  135         =CPE_bind=
1.1.1.1   135  149                 ======JunD=====
9.9.51    139  148                     ====ATF===
0.5.1.0   140  149                      =E1A_12S_=
1.1.1.6   140  149                      ==CRE-BP1=
1.1.2.0   140  149                      ====CREB==
1.1.3.0   140  149                      =C/EBPalp=
2.1.2.2   140  149                      ==RXR-beta
2.3.3.0   141  152                       =CPE_bindin=
2.3.1.0   157  171                                   =======Sp1=====
2.3.2.1   161  170                                      ===Egr-1==
9.9.173   178  187                                                ==
================================================================================
seq(  180..  239)     ccattggcagtcggtcgtgggtacgcggctcctgtgttttaggtttgagctatgggaggg
Segments:
9.9.173   178  187   ==CTF===
1.6.1.0   206  215                        ====AP-2==
2.3.1.0   232  246                                              =======S
2.3.2.1   234  243                                               ==Krox
9.9.270   235  244                                                ====E
================================================================================
```

```
seq(  240..  299)     ggcgggaattcccgactctaggccggaagcgcgcggagaccatgtagtgagaccctcgcg
Segments:
2.3.1.0    232  246   p1=====
2.3.2.1    234  243   -20=
9.9.270    235  244   TF===
2.3.4.0    243  252      ===MBP-2==
4.1.1.0    243  252      ===c-Rel==
9.9.588    243  252      =NF-kappaB
9.9.590    243  252      =NF-kappaB
9.9.592    243  252      =NF-kappa=
9.9.594    243  252      ====RelA==
3.5.2.0    260  269                    =c-Ets-1_=
2.3.1.0    268  277                           ====Sp1===
================================================================================
seq(  300..  359)     aggtgagtgcgggctgtctggcagcggggtccctcgag
Segments:
2.3.1.0    302  315     ======Sp1=====
9.9.539    313  322               ====NF-1==
2.3.1.0    319  332                     ======Sp1=====

42 segments in this sequence identified as potential binding sites
```

A.1.4 MutF promoter

A.1.4.1 Promoter sequence MutF

>MutF

GGTACCCCTC AGTCCTCCAC AGATGCGCCC CAGGGCCCCA GCCCCGAGGT CCTTTTCCTA ATATCCGCTC TTGCTCTCGC TCACTCTCCT CAGGTCGCCT GCTCCGCCCG CTCCCGCTCC CGATCTCTGA CTGCTCCTAA CTGACGTCAC TCCCGGTCTG TCCCCGCCCA CTCGGTGCTG CCATTGGCAG TCGGTCGTGG GTACGCGGCT CCTGTGGGGG AGGTTTGAGC TATGGGAGGG GGCGGGAATT CCCGACTCTA GGCCGGAAGC GCGCGGAGAC CATGTAGTGA GACCCTCGCG AGGTGAGTGC GGGCTGTCTG GCAGCGGGGT CCCTCGAG

A.1.4.2 Nsite results for MutF

Program Nsite | Version 5.2013

Search for motifs of 1500 Regulatory Elements (REs)

SET of REs: RegsiteAN DB: 1500 human/animal Transcription Regulatory Elements (real or consensus sites) [Last update: 10.03.2006; Softberry Inc]

Search PARAMETRS:

Expected Mean Number : 0.0100000

Statistical Siginicance Level : 0.9500000

Level of homology between known RE and motif: 80%

Variation of Distance between RE Blocks : 20%

NOTE: RE - Regulatory Element/Consensus | AC - Accession No of RE in a given DB

OS - Organism/Species | BF - Binding Factor or One of them

Mism. - Mismatches | Mean. Exp. Number - Mean Expected Number | Up.Conf.Int. - Upper Confidence Interval

==========

QUERY: >MutF

Length of Query Sequence: 338 bp | Nucleotide Frequencies: A - 0.13 G - 0.31 T - 0.21 C - 0.36

.....

RE: 170. AC: RSA00170//OS: human /GENE: beta-IFN/RE: PRD II /BF: NF-kappaB

Motifs on "-" Strand: Mean Exp. Number 0.00838 Up.Conf.Int. 1 Found 1

256 GTcGGgAATTCC 245 (Mism.= 2)

.....

RE: 234. AC: RSA00234//OS: rat /GENE: VMAT2/RE: CRE /BF: CREB; CBP/p300

Motifs on "+" Strand: Mean Exp. Number 0.00282 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00282 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 289. AC: RSA00289//OS: human /GENE: c-myc/RE: E2F BS /BF: E2F-1; E2F-2; E2F-4

Motifs on "+" Strand: Mean Exp. Number 0.00525 Up.Conf.Int. 1 Found 1

241 GGCGGGAA 248 (Mism.= 0)

.....

RE: 350. AC: RSA00350//OS: rat /GENE: spi 2.1/RE: C-II core /BF: C/EBPalpha

Motifs on "-" Strand: Mean Exp. Number 0.00671 Up.Conf.Int. 1 Found 1

61 TTAGGAAAa 53 (Mism.= 1)

.....

RE: 566. AC: RSA00566//OS: mouse /GENE: alphaA-crystallin/RE: alphaA-CRYBP1 BS /BF: alphaA-CRYBP1

Motifs on "+" Strand: Mean Exp. Number 0.00696 Up.Conf.Int. 1 Found 1

244 GGGAAtTCCC 253 (Mism.= 1)

.....

RE: 573. AC: RSA00573//OS: mouse /GENE: MHCIIAalpha/RE: CRE /BF: HB16

Motifs on "+" Strand: Mean Exp. Number 0.00282 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00282 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 621. AC: RSA00621//OS: adenovirus (Ad5) /GENE: E4/RE: CRE /BF: ETF-A

Motifs on "+" Strand: Mean Exp. Number 0.00592 Up.Conf.Int. 1 Found 1

142 TGACGTCAC 150 (Mism.= 1)

.....

RE: 745. AC: RSA00745//OS: human /GENE: H3 (ST519)/RE: GC-box /BF:

Motifs on "+" Strand: Mean Exp. Number 0.00880 Up.Conf.Int. 1 Found 2

234 GGgAGGGgGCGGGaA 248 (Mism.= 3)

265 GGAAGcGCGCGGaGA 279 (Mism.= 2)

.....

RE: 838. AC: RSA00838//OS: human /GENE: endoglin/RE: Sp1 /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00918 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 976. AC: RSA00976//OS: human /GENE: NET/RE: Sp1 /BF: Sp1

Motifs on "+" Strand: Mean Exp. Number 0.00918 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 1163. AC: RSA01163//OS: human /GENE: GSTP1/RE: Sp1-like element II /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00389 Up.Conf.Int. 1 Found 1

239 GGGGCGGGA 247 (Mism.= 0)

.....

Totally 14 motifs of 11 different REs have been found

© 1999 - 2016 www.softberry.com

A1.4.3 AliBaba2.1 results for MutF

```
================================================================================
seq(    0..   59)     ggtacccctcagtcctccacagatgcgccccagggccccagccccgaggtccttttcctа
Segments:
2.3.1.0     24   36                              ======Sp1====
9.9.539     25   34                               ====NF-1==
1.6.1.0     28   37                                  =AP-2alph=
9.9.539     32   41                                      ====NF-1==
2.3.1.0     33   45                                       ======Sp1====
9.9.270     35   44                                         ====ETF===
9.9.701     47   56                                                    =PTF1-beta
2.3.1.0     49   58                                                     ====Sp1===
1.1.3.0     52   61                                                        =C/EBPal
4.3.1.3     53   62                                                         ====MCM
================================================================================
seq(   60..  119)     atatccgctcttgctctcgctcactctcctcaggtcgcctgctccgcccgctcccgctcc
Segments:
1.1.3.0     52   61     p=
4.3.1.3     53   62     1==
1.3.1.2     78   87                        ====USF===
2.1.2.2     87   96                                ==RXR-beta
2.3.1.0    100  114                                        =======Sp1=====
2.3.1.0    110  119                                                ====Sp1===
1.1.3.0    118  127                                                     =C
================================================================================
seq(  120..  179)     cgatctctgactgctcctaactgacgtcactcccggtctgtccccgcccactcggtgctg
Segments:
1.1.3.0     118  127   /EBPalp=
2.3.3.0     126  135         =CPE_bind=
1.1.1.1     135  149                  ======JunD=====
9.9.51      139  148                      ====ATF===
0.5.1.0     140  149                       =E1A_12S_=
1.1.1.6     140  149                       ==CRE-BP1=
1.1.2.0     140  149                       ====CREB==
1.1.3.0     140  149                       =C/EBPalp=
2.1.2.2     140  149                       ==RXR-beta
2.3.3.0     141  152                        =CPE_bindin=
2.3.1.0     157  171                                    =======Sp1=====
2.3.2.1     161  170                                        ===Egr-1==
9.9.173     178  187                                                  ==
================================================================================
seq(  180..  239)     ccattggcagtcggtcgtgggtacgcggctcctgtgggggaggtttgagctatgggaggg
Segments:
9.9.173     178  187   ==CTF===
1.6.1.0     206  215                           ====AP-2==
9.9.77      213  222                                 =CACCC-bi=
2.3.1.0     213  224                                 =====Sp1====
2.3.1.0     232  246                                              =======S
2.3.2.1     234  243                                               ==Krox
9.9.270     235  244                                                ====E
```

```
seq(  240..  299)     ggcgggaattcccgactctaggccggaagcgcgcggagaccatgtagtgagaccctcgcg
Segments:
2.3.1.0    232  246   p1=====
2.3.2.1    234  243   -20=
9.9.270    235  244   TF===
2.3.4.0    243  252      ===MBP-2==
4.1.1.0    243  252      ===c-Rel==
9.9.588    243  252      =NF-kappaB
9.9.590    243  252      =NF-kappaB
9.9.592    243  252      =NF-kappa=
9.9.594    243  252      ====RelA==
3.5.2.0    260  269                        =c-Ets-1_=
2.3.1.0    268  277                               ====Sp1===
================================================================================
seq(  300..  359)     aggtgagtgcgggctgtctggcagcggggtccctcgag
Segments:
2.3.1.0    302  315     ======Sp1=====
9.9.539    313  322                 ====NF-1==
2.3.1.0    319  332                       ======Sp1=====

44 segments in this sequence identified as potential binding sites
```

A.1.5 MutG Promoter

A.1.5.1 Promoter sequence MutG

>MutG

GGTACCCCTC AGTCCTCCAC AGATGCGCCC CAGGGCCCCA GCCCCGAGGT CCTTTTCCTA ATATCCGCTC TTGCTCTCGC TCACTCTCCT CAGGTCGCCT GCTCCGCCCG CTCCCGCTCC CGATCTCTGA CTGCTCCTAA CTGACGTCAC TCCCGGTCTG TCCCCGCCCA CTCGGTGCTG CCATTGGCAG TCGGTCGTGG GTACGCGGCT CCTGTGTTGG AGGGGGGAGC TATGGGAGGG GGCGGGAATT CCCGACTCTA GGCCGGAAGC GCGCGGAGAC CATGTAGTGA GACCCTCGCG AGGTGAGTGC GGGCTGTCTG GCAGCGGGGT CCCTCGAG

A.1.5.2 Nsite results for MutG

Program Nsite | Version 5.2013

Search for motifs of 1500 Regulatory Elements (REs)

SET of REs: RegsiteAN DB: 1500 human/animal Transcription Regulatory Elements (real or consensus sites) [Last update: 10.03.2006; Softberry Inc]

Search PARAMETRS:

Expected Mean Number : 0.0100000

Statistical Siginicance Level : 0.9500000

Level of homology between known RE and motif: 80%

Variation of Distance between RE Blocks : 20%

NOTE: RE - Regulatory Element/Consensus | AC - Accession No of RE in a given DB

OS - Organism/Species | BF - Binding Factor or One of them

Mism. - Mismatches | Mean. Exp. Number - Mean Expected Number | Up.Conf.Int. - Upper Confidence Interval

==========

QUERY: >MutG

Length of Query Sequence: 338 bp | Nucleotide Frequencies: A - 0.13 G - 0.31 T - 0.21 C - 0.36

.....

RE: 170. AC: RSA00170//OS: human /GENE: beta-IFN/RE: PRD II /BF: NF-kappaB

Motifs on "-" Strand: Mean Exp. Number 0.00824 Up.Conf.Int. 1 Found 1

256 GTcGGgAATTCC 245 (Mism.= 2)

.....

RE: 234. AC: RSA00234//OS: rat /GENE: VMAT2/RE: CRE /BF: CREB; CBP/p300

Motifs on "+" Strand: Mean Exp. Number 0.00279 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00279 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 289. AC: RSA00289//OS: human /GENE: c-myc/RE: E2F BS /BF: E2F-1; E2F-2; E2F-4

Motifs on "+" Strand: Mean Exp. Number 0.00550 Up.Conf.Int. 1 Found 1

241 GGCGGGAA 248 (Mism.= 0)

.....

RE: 350. AC: RSA00350//OS: rat /GENE: spi 2.1/RE: C-II core /BF: C/EBPalpha

Motifs on "-" Strand: Mean Exp. Number 0.00639 Up.Conf.Int. 1 Found 1

61 TTAGGAAAa 53 (Mism.= 1)

.....

RE: 566. AC: RSA00566//OS: mouse /GENE: alphaA-crystallin/RE: alphaA-CRYBP1 BS /BF: alphaA-CRYBP1

Motifs on "+" Strand: Mean Exp. Number 0.00706 Up.Conf.Int. 1 Found 1

244 GGGAAtTCCC 253 (Mism.= 1)

.....

RE: 573. AC: RSA00573//OS: mouse /GENE: MHCIIAalpha/RE: CRE /BF: HB16

Motifs on "+" Strand: Mean Exp. Number 0.00279 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00279 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 621. AC: RSA00621//OS: adenovirus (Ad5) /GENE: E4/RE: CRE /BF: ETF-A

Motifs on "+" Strand: Mean Exp. Number 0.00586 Up.Conf.Int. 1 Found 1

142 TGACGTCAC 150 (Mism.= 1)

.....

RE: 745. AC: RSA00745//OS: human /GENE: H3 (ST519)/RE: GC-box /BF:

Motifs on "+" Strand: Mean Exp. Number 0.00948 Up.Conf.Int. 1 Found 2

234 GGgAGGGgGCGGGaA 248 (Mism.= 3)

265 GGAAGcGCGCGGaGA 279 (Mism.= 2)

.....

RE: 838. AC: RSA00838//OS: human /GENE: endoglin/RE: Sp1 /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00927 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 976. AC: RSA00976//OS: human /GENE: NET/RE: Sp1 /BF: Sp1

Motifs on "+" Strand: Mean Exp. Number 0.00927 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 1163. AC: RSA01163//OS: human /GENE: GSTP1/RE: Sp1-like element II /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00416 Up.Conf.Int. 1 Found 1

239 GGGGCGGGA 247 (Mism.= 0)

.....

Totally 14 motifs of 11 different REs have been found

© 1999 - 2016 www.softberry.com

A.1.5.3 AliBaba2.1 results for MutG

```
================================================================================
seq(    0..   59)     ggtacccctcagtcctccacagatgcgccccagggccccagccccgaggtccttttcctа
Segments:
2.3.1.0     24   36                            ======Sp1====
9.9.539     25   34                             ====NF-1==
1.6.1.0     28   37                                =AP-2alph=
9.9.539     32   41                                    ====NF-1==
2.3.1.0     33   45                                     ======Sp1====
9.9.270     35   44                                       ====ETF===
9.9.701     47   56                                                 =PTF1-beta
2.3.1.0     49   58                                                   ====Sp1===
1.1.3.0     52   61                                                      =C/EBPal
4.3.1.3     53   62                                                       ====MCM
================================================================================
seq(   60..  119)     atatccgctcttgctctcgctcactctcctcaggtcgcctgctccgcccgctcccgctcc
Segments:
1.1.3.0     52   61     p=
4.3.1.3     53   62     1==
1.3.1.2     78   87                        ====USF===
2.1.2.2     87   96                                ==RXR-beta
2.3.1.0     100  114                                       =======Sp1=====
2.3.1.0     110  119                                             ====Sp1===
1.1.3.0     118  127                                                   =C
================================================================================
seq(  120..  179)     cgatctctgactgctcctaactgacgtcactcccggtctgtccccgcccactcggtgctg
Segments:
1.1.3.0     118  127   /EBPalp=
2.3.3.0     126  135         =CPE_bind=
1.1.1.1     135  149                  ======JunD=====
9.9.51      139  148                      ====ATF===
0.5.1.0     140  149                       =E1A_12S_=
1.1.1.6     140  149                       ==CRE-BP1=
1.1.2.0     140  149                       ====CREB==
1.1.3.0     140  149                       =C/EBPalp=
2.1.2.2     140  149                       ==RXR-beta
2.3.3.0     141  152                        =CPE_bindin=
2.3.1.0     157  171                                    =======Sp1=====
2.3.2.1     161  170                                       ===Egr-1==
9.9.173     178  187                                                  ==
================================================================================
seq(  180..  239)     ccattggcagtcggtcgtgggtacgcggctcctgtgttggaggggggagctatgggaggg
Segments:
9.9.173     178  187   ==CTF===
1.6.1.0     206  215                          ====AP-2==
2.3.1.0     213  227                                =======Sp1=====
2.3.1.0     219  230                                      =====Sp1====
2.3.1.0     232  246                                                  =======S
2.3.2.1     234  243                                                    ==Krox
9.9.270     235  244                                                     ====E
================================================================================
```

```
seq(  240..  299)     ggcgggaattcccgactctaggccggaagcgcgcggagaccatgtagtgagaccctcgcg
Segments:
2.3.1.0    232  246   p1=====
2.3.2.1    234  243   -20=
9.9.270    235  244   TF===
2.3.4.0    243  252      ===MBP-2==
4.1.1.0    243  252      ===c-Rel==
9.9.588    243  252      =NF-kappaB
9.9.590    243  252      =NF-kappaB
9.9.592    243  252      =NF-kappa=
9.9.594    243  252      ====RelA==
3.5.2.0    260  269                      =c-Ets-1_=
2.3.1.0    268  277                             ====Sp1===
================================================================================
seq(  300..  359)     aggtgagtgcgggctgtctggcagcggggtccctcgag
Segments:
2.3.1.0    302  315     ======Sp1=====
9.9.539    313  322                 ====NF-1==
2.3.1.0    319  332                       ======Sp1=====

44 segments in this sequence identified as potential binding sites
```

A.1.6 MutH Promoter

A.1.6.1 Promoter sequence MutH

>MutH

GGTACCCCTC AGTCCTCCAC AGATGCGCCC CAGGGCCCCA GCCCCGAGGT CCTTTTCCTA ATATCCGCTC TTGCTCTCGC TCACTCTCCT CAGGTCGCCT GCTCCGCCCG CTCCCGCTCC CGATCTCTGA CTGCTCCTAA CTGACGTCAC TCCCGGTCTG TCCCCGCCCA CTCGGTGCTG CCATTGGCAG TCGGTCGTGG GTACGCGGCT CCTGTGTTGG AGGTTTGAGC TATGGGAGGG GGCGGGAATT CCCGACTCTA GGCCGGAAGC GCGCGGAGAC CATGTAGTGA GACCCTCGCG AGGTGAGTGC GGGCTGTCTG GCAGCGGGGT CCCTCGAG

A.1.6.2 Nsite results for MutH

Program Nsite | Version 5.2013

Search for motifs of 1500 Regulatory Elements (REs)

SET of REs: RegsiteAN DB: 1500 human/animal Transcription Regulatory Elements (real or consensus sites) [Last update: 10.03.2006; Softberry Inc]

Search PARAMETRS:

Expected Mean Number : 0.0100000

Statistical Siginicance Level : 0.9500000

Level of homology between known RE and motif: 80%

Variation of Distance between RE Blocks : 20%

NOTE: RE - Regulatory Element/Consensus | AC - Accession No of RE in a given DB

OS - Organism/Species | BF - Binding Factor or One of them

Mism. - Mismatches | Mean. Exp. Number - Mean Expected Number | Up.Conf.Int. - Upper Confidence Interval

==========

QUERY: >MutH

Length of Query Sequence: 338 bp | Nucleotide Frequencies: A - 0.13 G - 0.30 T - 0.22 C - 0.36

.....

RE: 170. AC: RSA00170//OS: human /GENE: beta-IFN/RE: PRD II /BF: NF-kappaB

Motifs on "-" Strand: Mean Exp. Number 0.00866 Up.Conf.Int. 1 Found 1

256 GTcGGgAATTCC 245 (Mism.= 2)

.....

RE: 234. AC: RSA00234//OS: rat /GENE: VMAT2/RE: CRE /BF: CREB; CBP/p300

Motifs on "+" Strand: Mean Exp. Number 0.00287 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00287 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 289. AC: RSA00289//OS: human /GENE: c-myc/RE: E2F BS /BF: E2F-1; E2F-2; E2F-4

Motifs on "+" Strand: Mean Exp. Number 0.00476 Up.Conf.Int. 1 Found 1

241 GGCGGGAA 248 (Mism.= 0)

.....

RE: 350. AC: RSA00350//OS: rat /GENE: spi 2.1/RE: C-II core /BF: C/EBPalpha

Motifs on "-" Strand: Mean Exp. Number 0.00740 Up.Conf.Int. 1 Found 1

61 TTAGGAAAa 53 (Mism.= 1)

.....

RE: 566. AC: RSA00566//OS: mouse /GENE: alphaA-crystallin/RE: alphaA-CRYBP1 BS /BF: alphaA-CRYBP1

Motifs on "+" Strand: Mean Exp. Number 0.00677 Up.Conf.Int. 1 Found 1

244 GGGAAtTCCC 253 (Mism.= 1)

.....

RE: 573. AC: RSA00573//OS: mouse /GENE: MHCIIAalpha/RE: CRE /BF: HB16

Motifs on "+" Strand: Mean Exp. Number 0.00287 Up.Conf.Int. 1 Found 1

142 TGACGTCA 149 (Mism.= 0)

Motifs on "-" Strand: Mean Exp. Number 0.00287 Up.Conf.Int. 1 Found 1

149 TGACGTCA 142 (Mism.= 0)

.....

RE: 621. AC: RSA00621//OS: adenovirus (Ad5) /GENE: E4/RE: CRE /BF: ETF-A

Motifs on "+" Strand: Mean Exp. Number 0.00602 Up.Conf.Int. 1 Found 1

142 TGACGTCAC 150 (Mism.= 1)

.....

RE: 745. AC: RSA00745//OS: human /GENE: H3 (ST519)/RE: GC-box /BF:

Motifs on "+" Strand: Mean Exp. Number 0.00757 Up.Conf.Int. 1 Found 2

234 GGgAGGGgGCGGGaA 248 (Mism.= 3)

265 GGAAGcGCGCGGaGA 279 (Mism.= 2)

.....

RE: 838. AC: RSA00838//OS: human /GENE: endoglin/RE: Sp1 /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00901 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 976. AC: RSA00976//OS: human /GENE: NET/RE: Sp1 /BF: Sp1

Motifs on "+" Strand: Mean Exp. Number 0.00901 Up.Conf.Int. 1 Found 1

37 CCCAGCCCC 45 (Mism.= 0)

.....

RE: 1163. AC: RSA01163//OS: human /GENE: GSTP1/RE: Sp1-like element II /BF: SP1

Motifs on "+" Strand: Mean Exp. Number 0.00340 Up.Conf.Int. 1 Found 1

239 GGGGCGGGA 247 (Mism.= 0)

.....

Totally 14 motifs of 11 different REs have been found

© 1999 - 2016 www.softberry.com

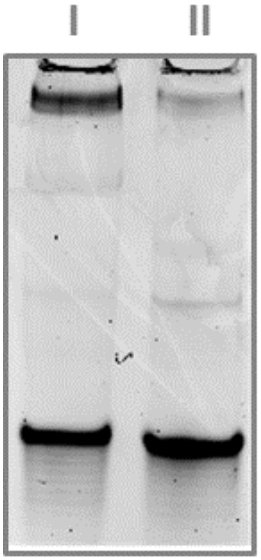
A.1.6.3 AliBaba2.1 results for MutH

```
================================================================================
seq(    0..   59)     ggtacccctcagtcctccacagatgcgccccagggccccagccccgaggtccttttcctа
Segments:
2.3.1.0     24   36                              ======Sp1====
9.9.539     25   34                               ====NF-1==
1.6.1.0     28   37                                  =AP-2alph=
9.9.539     32   41                                      ====NF-1==
2.3.1.0     33   45                                       ======Sp1====
9.9.270     35   44                                         ====ETF===
9.9.701     47   56                                                    =PTF1-beta
2.3.1.0     49   58                                                      ====Sp1===
1.1.3.0     52   61                                                         =C/EBPal
4.3.1.3     53   62                                                          ====MCM
================================================================================
seq(   60..  119)     atatccgctcttgctctcgctcactctcctcaggtcgcctgctccgcccgctcccgctcc
Segments:
1.1.3.0     52   61     p=
4.3.1.3     53   62     1==
1.3.1.2     78   87                        ====USF===
2.1.2.2     87   96                                ==RXR-beta
2.3.1.0    100  114                                        =======Sp1=====
2.3.1.0    110  119                                              ====Sp1===
1.1.3.0    118  127                                                    =C
================================================================================
seq(  120..  179)     cgatctctgactgctcctaactgacgtcactcccggtctgtccccgcccactcggtgctg
Segments:
1.1.3.0    118  127   /EBPalp=
2.3.3.0    126  135         =CPE_bind=
1.1.1.1    135  149                  ======JunD=====
9.9.51     139  148                      ====ATF===
0.5.1.0    140  149                       =E1A_12S_=
1.1.1.6    140  149                       ==CRE-BP1=
1.1.2.0    140  149                       ====CREB==
1.1.3.0    140  149                       =C/EBPalp=
2.1.2.2    140  149                       ==RXR-beta
2.3.3.0    141  152                        =CPE_bindin=
2.3.1.0    157  171                                    =======Sp1=====
2.3.2.1    161  170                                        ===Egr-1==
9.9.173    178  187                                                  ==
================================================================================
seq(  180..  239)     ccattggcagtcggtcgtgggtacgcggctcctgtgttggaggtttgagctatgggaggg
Segments:
9.9.173    178  187   ==CTF===
1.6.1.0    206  215                           ====AP-2==
2.3.1.0    213  222                                 ====Sp1===
2.3.1.0    232  246                                                 =======S
2.3.2.1    234  243                                                   ==Krox
9.9.270    235  244                                                    ====E
================================================================================
```

```
seq(  240..  299)     ggcgggaattcccgactctaggccggaagcgcgcggagaccatgtagtgagaccctcgcg
Segments:
2.3.1.0    232  246   p1=====
2.3.2.1    234  243   -20=
9.9.270    235  244   TF===
2.3.4.0    243  252      ===MBP-2==
4.1.1.0    243  252      ===c-Rel==
9.9.588    243  252      =NF-kappaB
9.9.590    243  252      =NF-kappaB
9.9.592    243  252      =NF-kappa=
9.9.594    243  252      ====RelA==
3.5.2.0    260  269                     =c-Ets-1_=
2.3.1.0    268  277                           ====Sp1===
================================================================================
seq(  300..  359)     aggtgagtgcgggctgtctggcagcggggtccctcgag
Segments:
2.3.1.0    302  315     ======Sp1=====
9.9.539    313  322               ====NF-1==
2.3.1.0    319  332                     ======Sp1=====

43 segments in this sequence identified as potential binding sites
```

A.2 ELECTROPHORETIC GELS



I. hZip1-1 loaded using formamide buffer
II. hZip1-1 loaded using urea buffer

Figure A.2.1, Denaturing 20% polyacrylamide electrophoretic gel showing a sample of hZIP1-1 treated with different loading buffers and heated up to 95 °C for five minutes before loading.

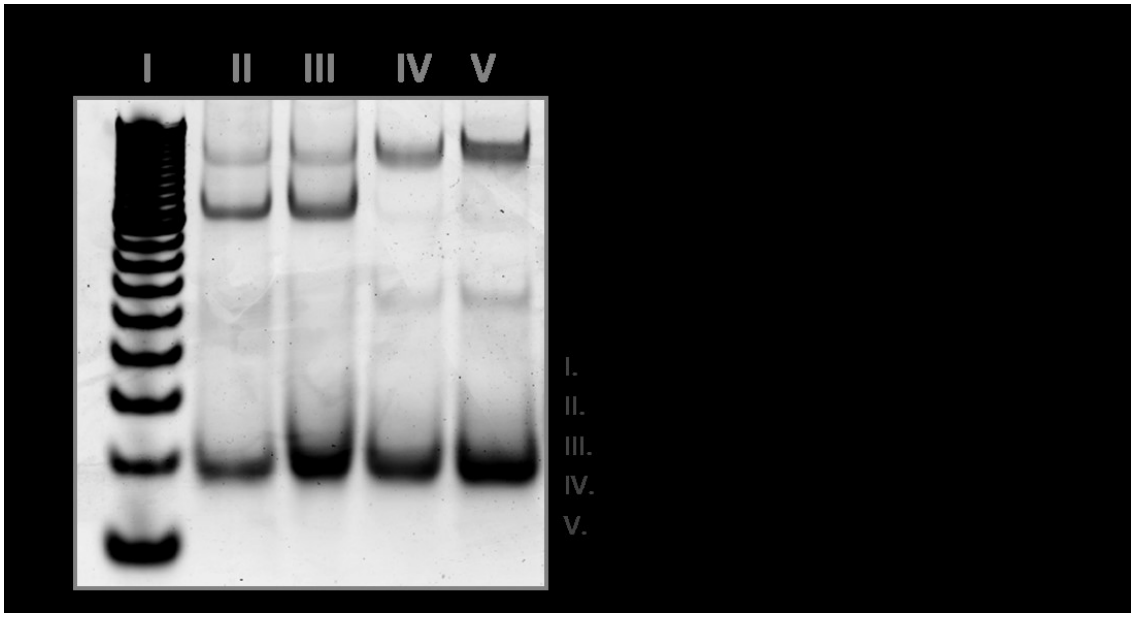


Figure A.2.2, Non-denaturing 12% polyacrylamide electrophoretic gel showing a sample of hZIP1-1 annealed with different DNA concentration and different annealing methods using NaCl.

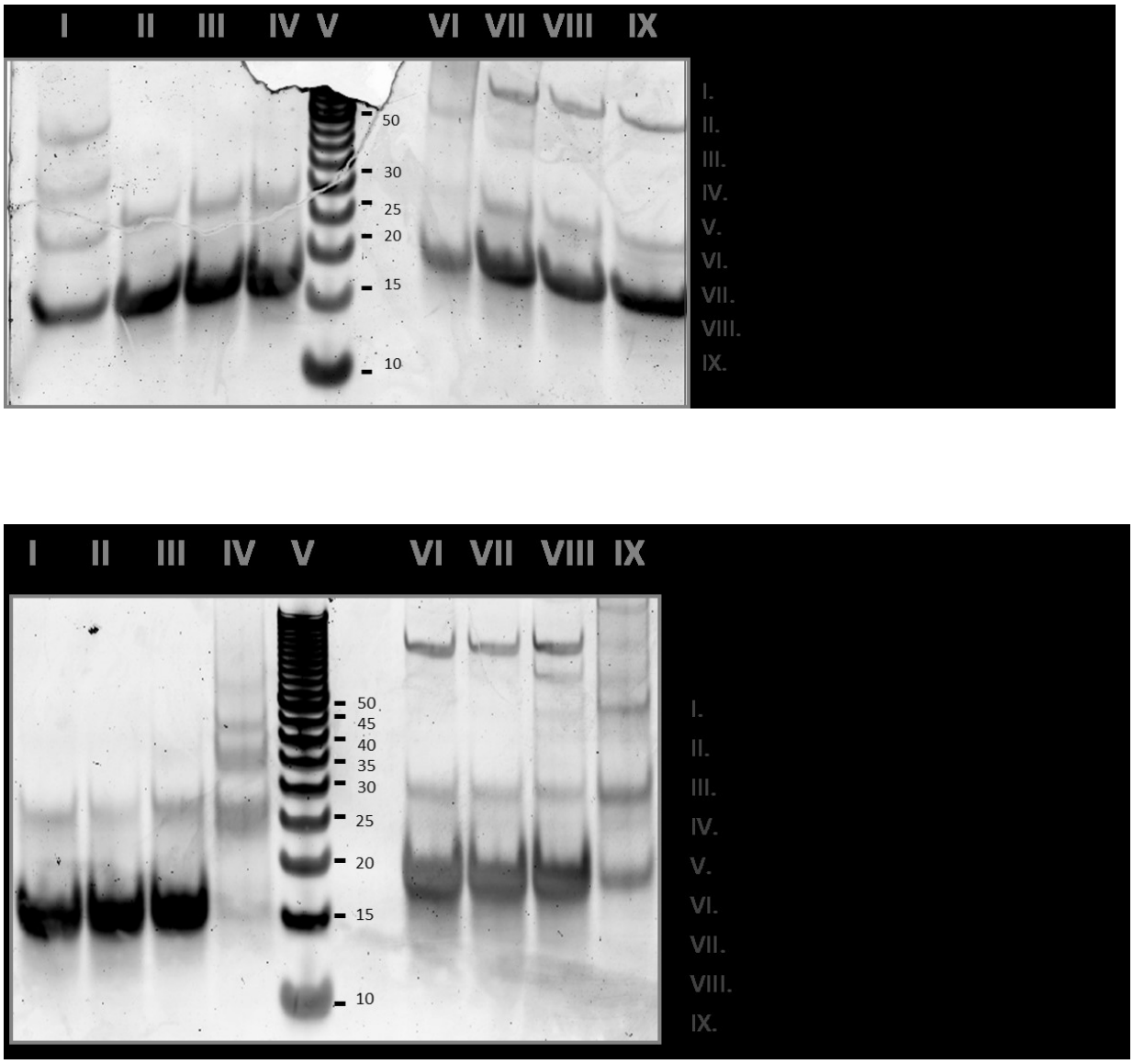


Figure A.2.3, Non-denaturing 12% polyacrylamide electrophoretic gels showing hZIP1-2 and hZIP1-3 annealed with different methods and salts. Samples were not desalted previous to experiments, so may contain salt from manufacturing. METHODS 1 and 2 are not appropriate to destroy secondary structures.

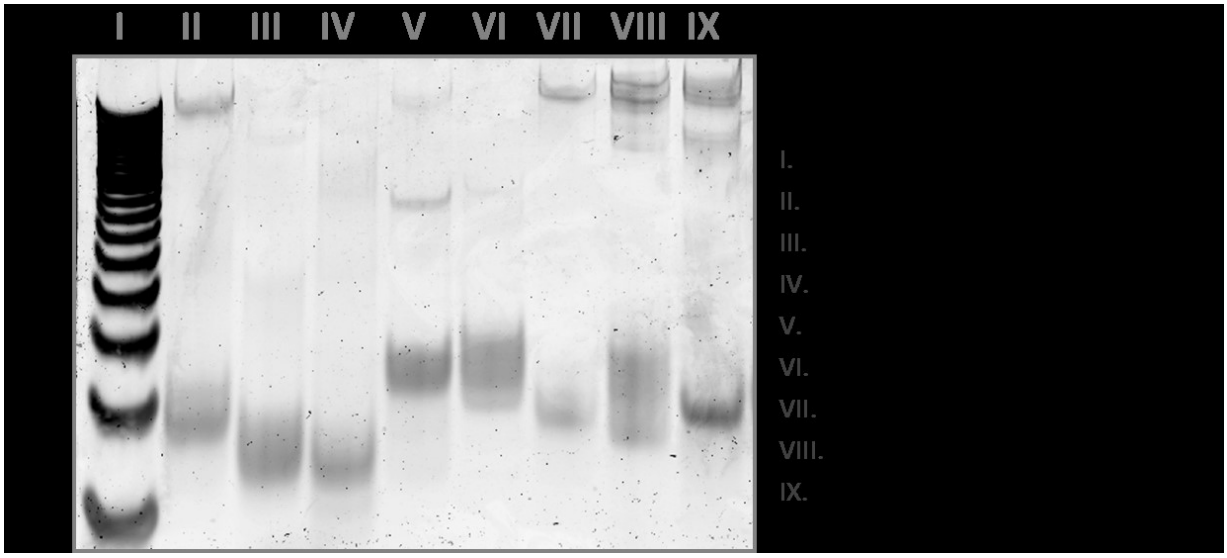
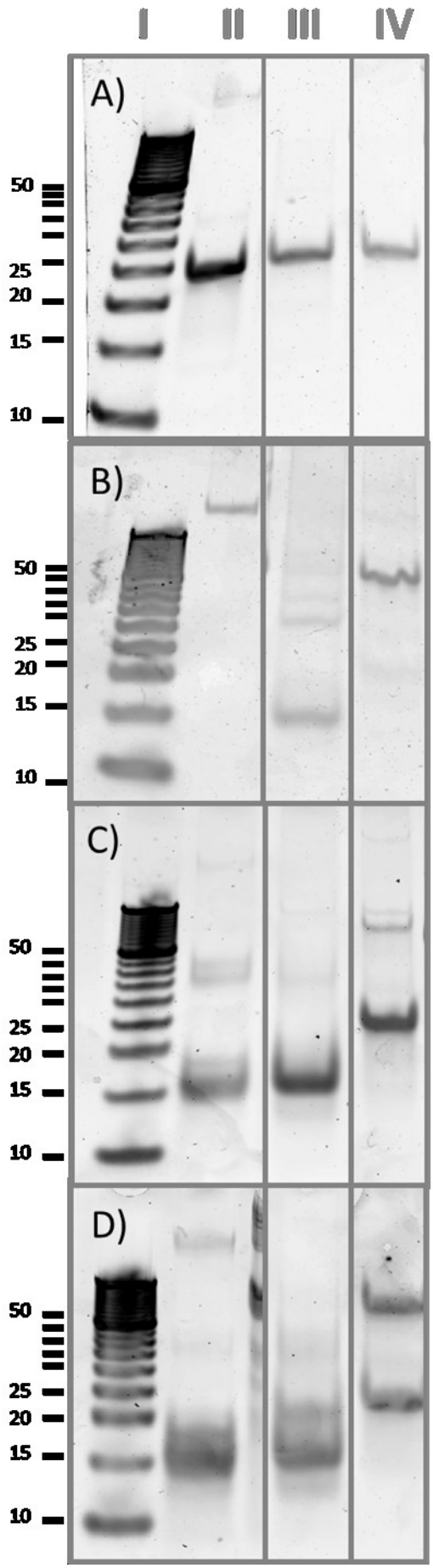
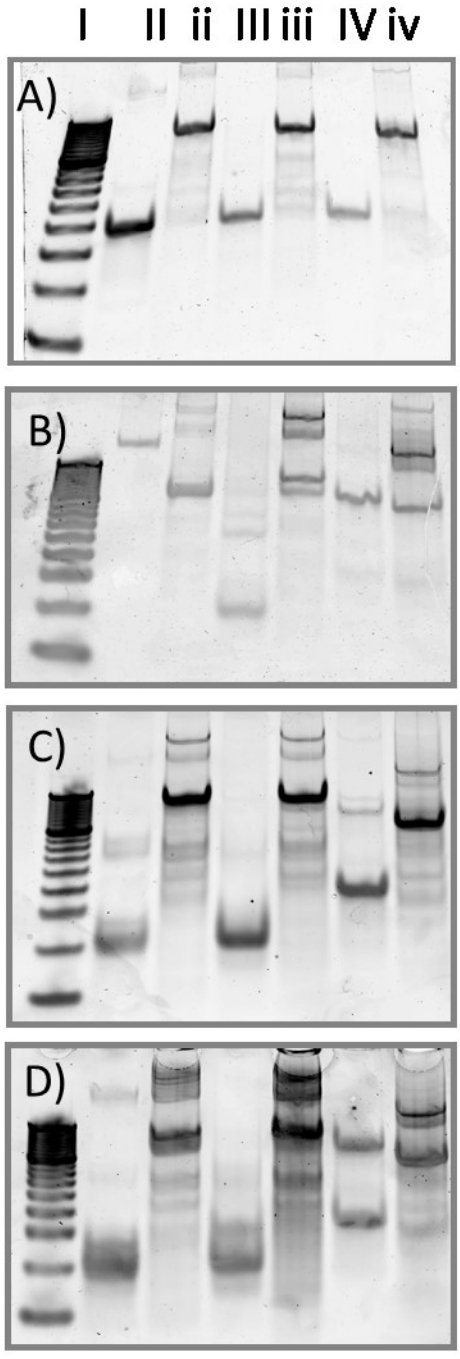


Figure A.2.4, Non-denaturing 16% polyacrylamide electrophoretic gels showing different DNA sequences annealed with 100 mM KCl using METHOD 2. Mutants 1 to 5 are all the same length than hZIP1-1. The long smeared bands suggest several populations that do not resolve properly. This pattern was obtained in several independent trials.

Figure A.2.5 (next page) Non-denaturing 16% polyacrylamide electrophoretic gels showing different DNA sequences annealed using different techniques. Except for gel A, all gels were run with 20 mM KCl in gel and running buffer. A) Samples were heated up in Tris buffer to 95 °C and cool down to room temperature without salt to destroy secondary structures. Sequences hZIP1-1, MutB and MutE all migrate similarly because they are all the same length. B) Samples were annealed using METHOD 1 and 100 mM KCl. They all show different migration patterns. hZIP1-1 shows very slow migrating band, possibly a tetrameric G-quadruplex, MutB shows several bands, MutE shows a dimer. C) Samples were annealed using Method 2 and 100 mM KCl, hZIP1-1 and MutB migrate similarly, with an intramolecular G-quadruplex and a dimer. MutE also shows two bands, an unstructured band and a dimer. Many of the bands are doubled. D) Samples were annealed using METHOD 3 and 100 mM KCl, hZIP1-1 and MutB migrate similarly as intramolecular G-quadruplexes with double bands. MutE shows an unstructured band and a dimer.





I. Ladder (5bp)
II. hZIP1-1
ii. PSA-hZIP1
III. MutB
iii. PSA-MutB
IV. MutE
iv. PSA-MutE

A.3 Circular Dichroism

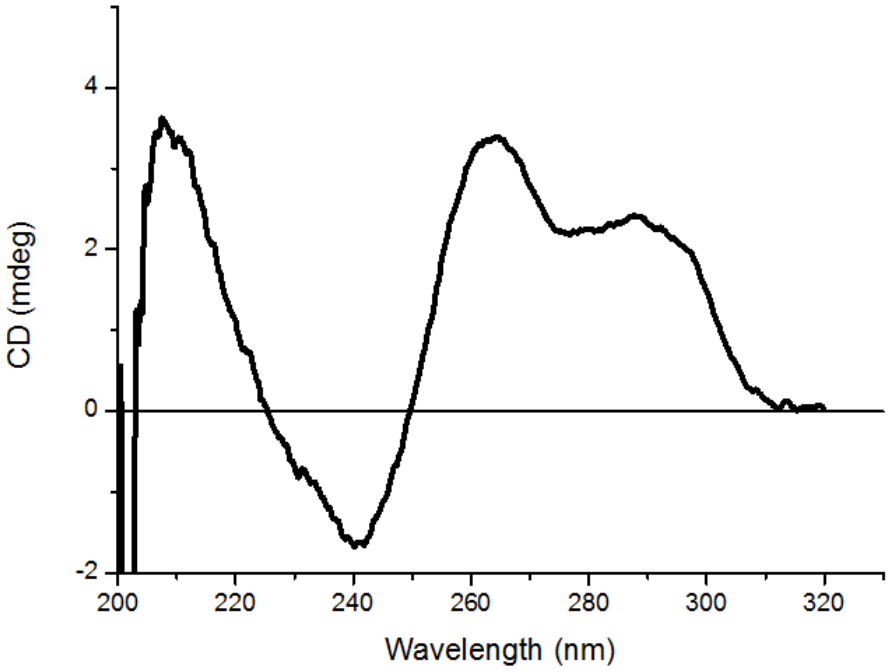


Figure A.3.1 Circular dichroism spectrum of MutC (5 µM) annealed by Method 2 in 100 mM KCl and 20 mM Li cacodylate buffer pH=7.4.

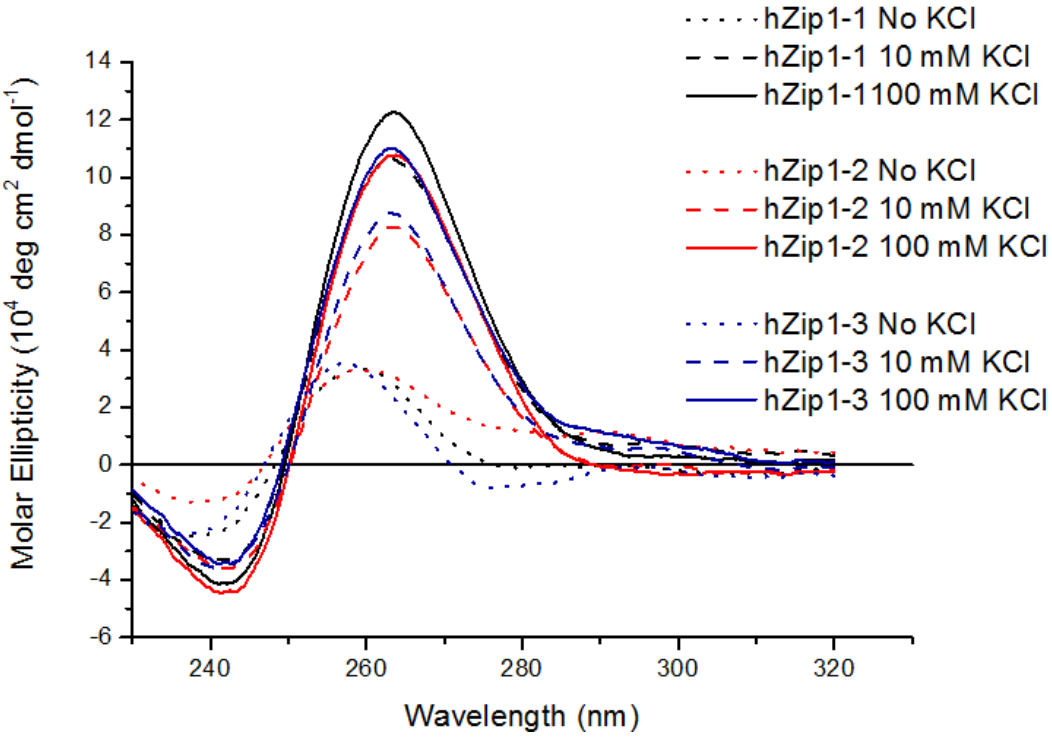


Figure A.3.2 Circular dichroism spectra of hZIP1-1, hZIP1-2 and hZIP1-3 MutC (5 µM) annealed by Method 1 using different amounts of KCl in 50 mM Tris buffer pH=7.4.

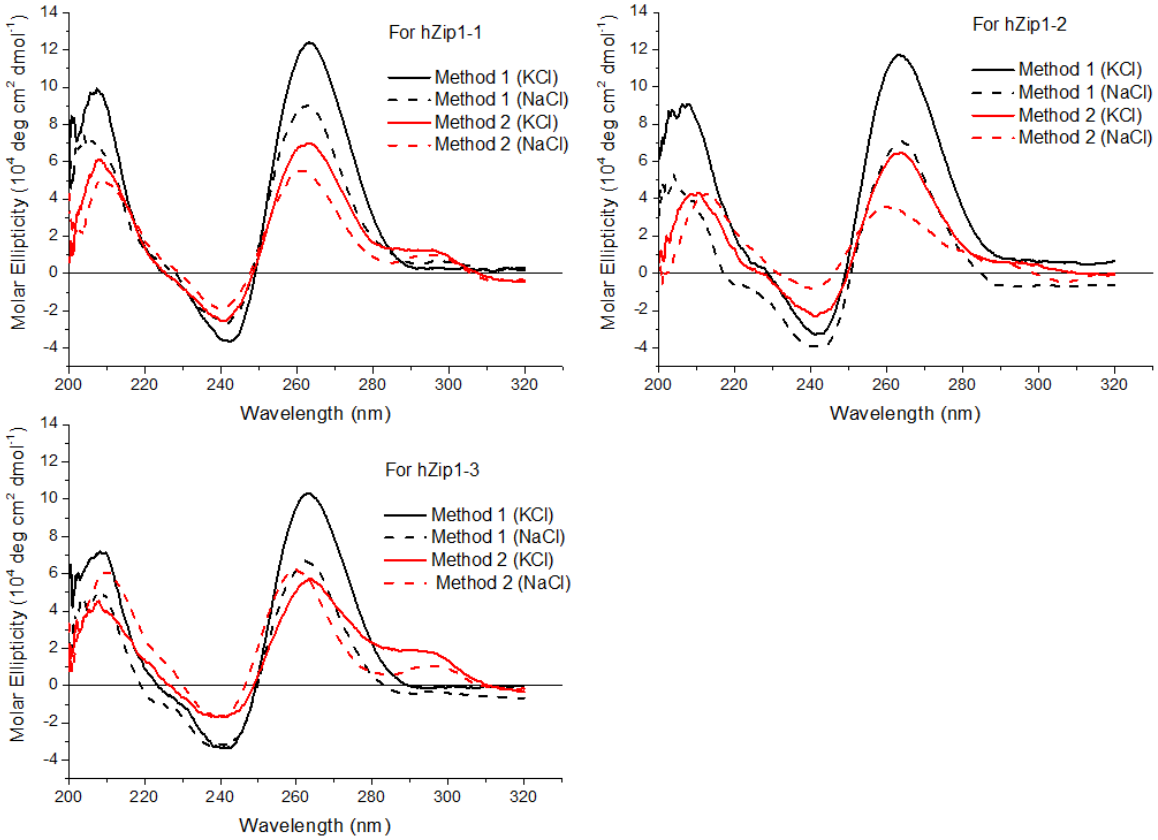


Figure A.3.3 Sequences hZIP1-1, hZIP1-2 and hZIP1-3 (5 µM) annealed by METHOD 1 (black line) or METHOD 2 (red line) with either 100 mM KCl (solid line) or 100 mM NaCl (dashed line) in 50 mM Tris buffer pH=7.4.

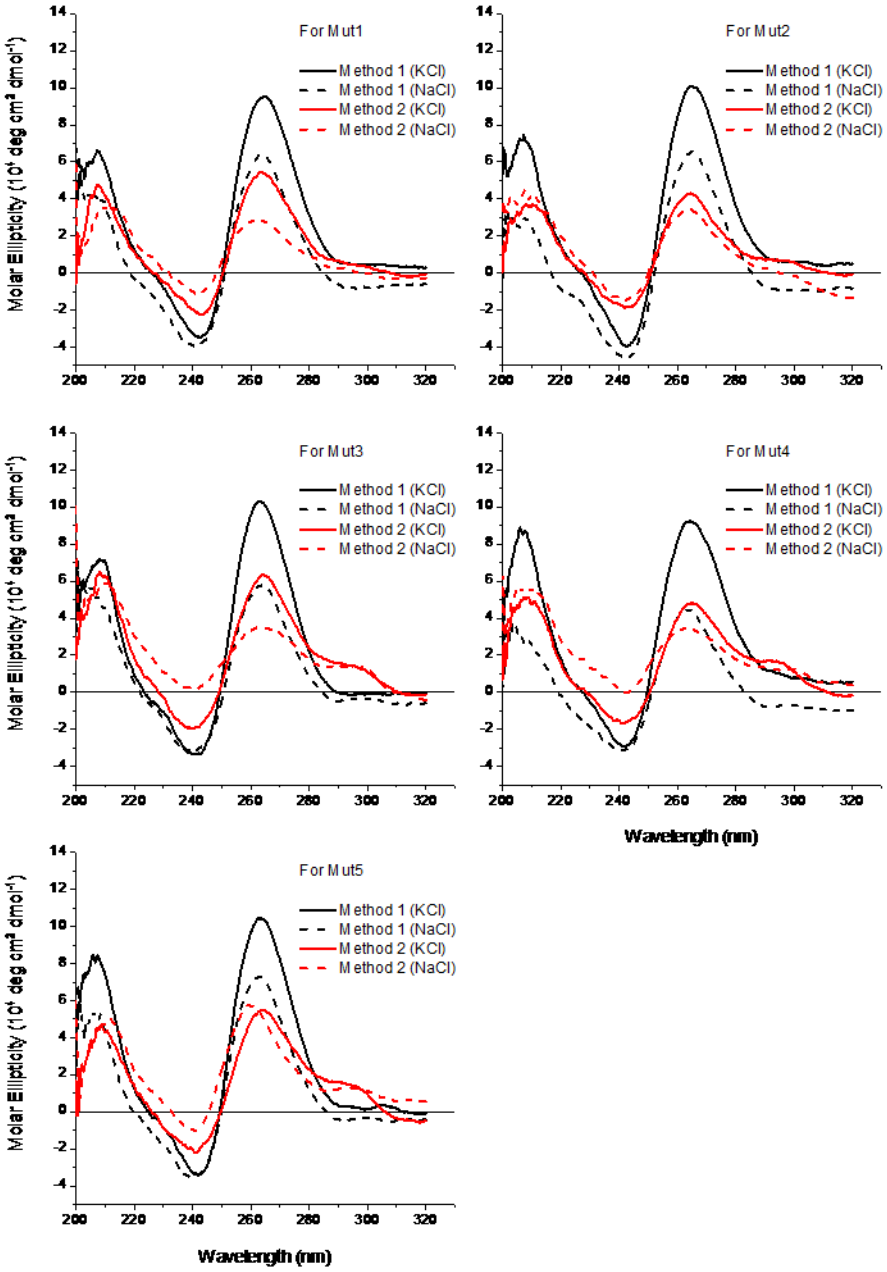


Figure A.3.4 Mutants Mut1, Mut2, Mut3, Mut4 and Mut5 (5 µM) annealed by METHOD 1 (black line) or METHOD 2 (red line) with either 100 mM KCl (solid line) or 100 mM NaCl (dashed line) in 50 mM Tris buffer pH=7.4.

A.4 Thermal Difference Spectra

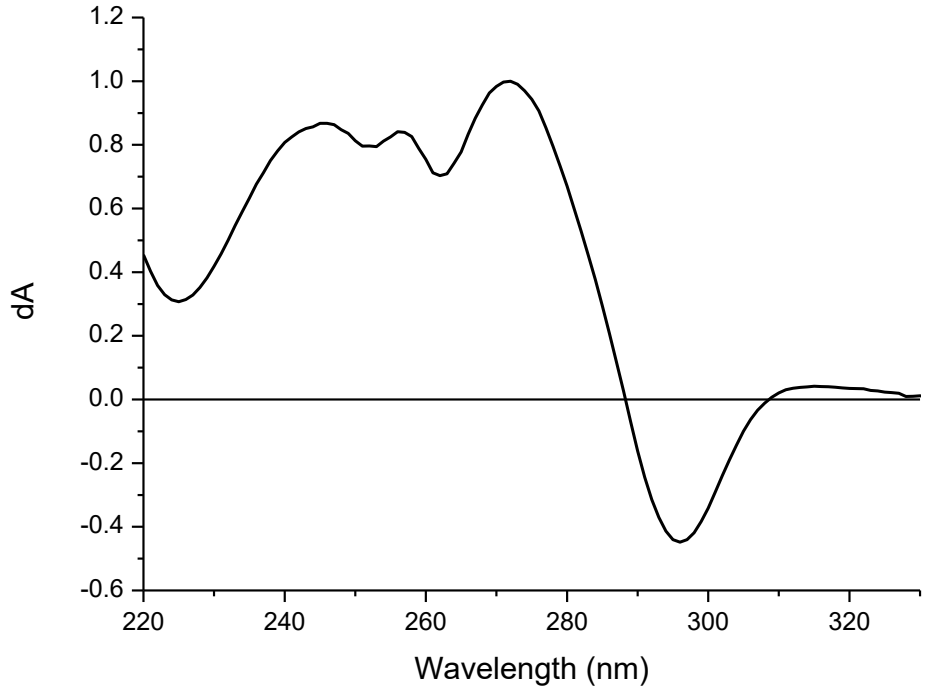


Figure A.4.1 Thermal difference spectra of MutC (5 µM) annealed by Method 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl.

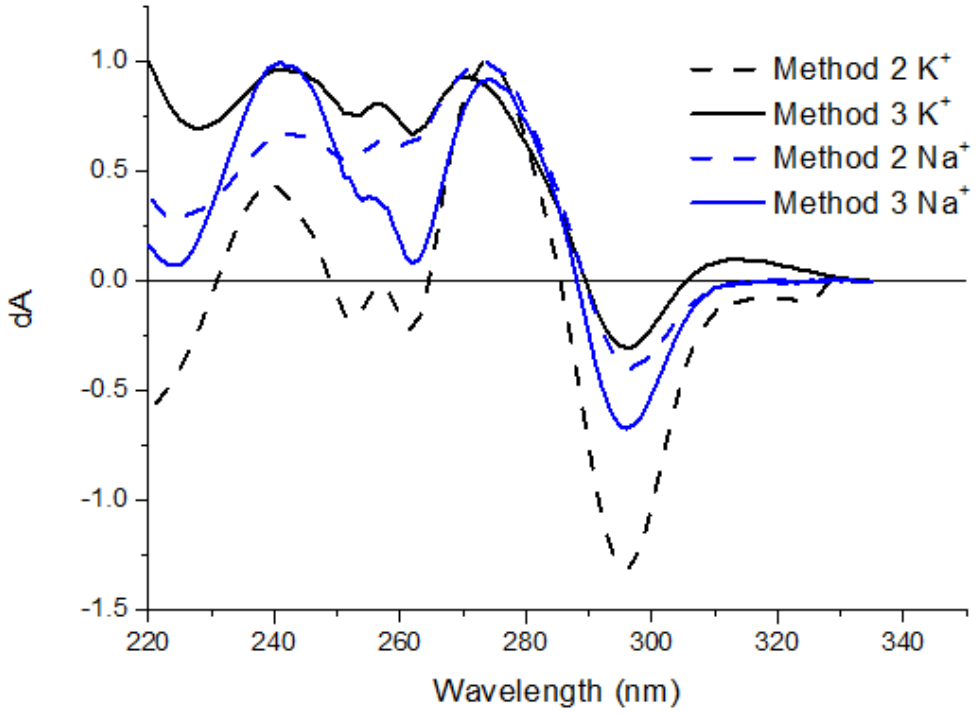


Figure A.4.2 Thermal difference spectra of hZIP1-1 (5 µM) annealed by Method 2 or Method 3 in 10 mM Li cacodylate buffer pH=7.4 with either 100 mM KCl or 100 mM NaCl.

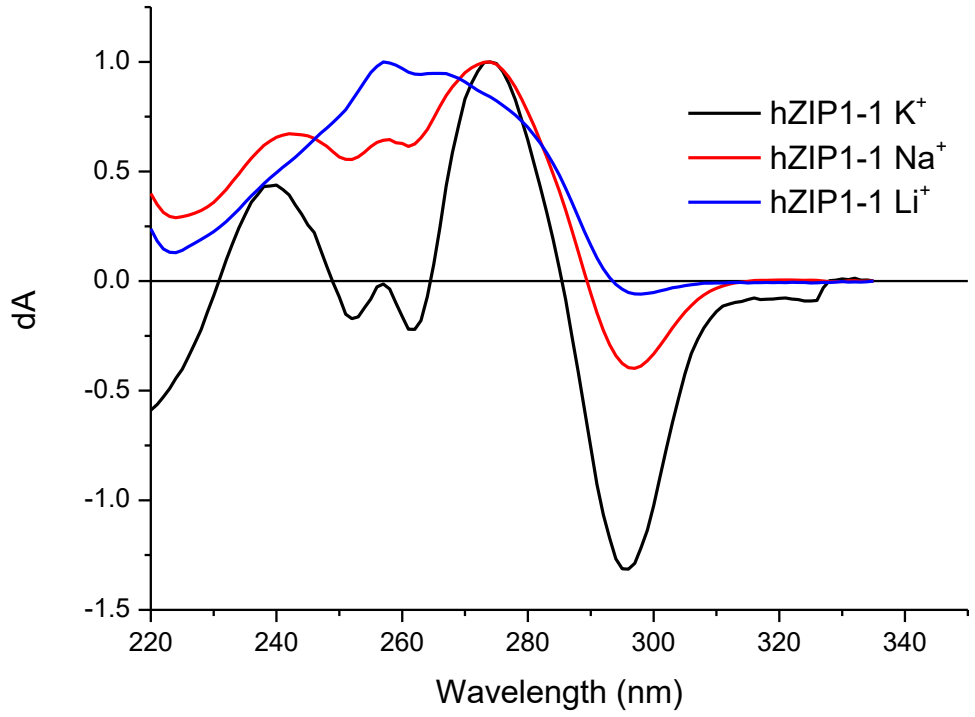


Figure A.4.3 Termal difference spectra of hZIP1-1 (5 µM) annealed by METHOD 2 in 10 mM Li cacodylate buffer pH=7.4 with either 100 mM of KCl, NaCl or LiCl.

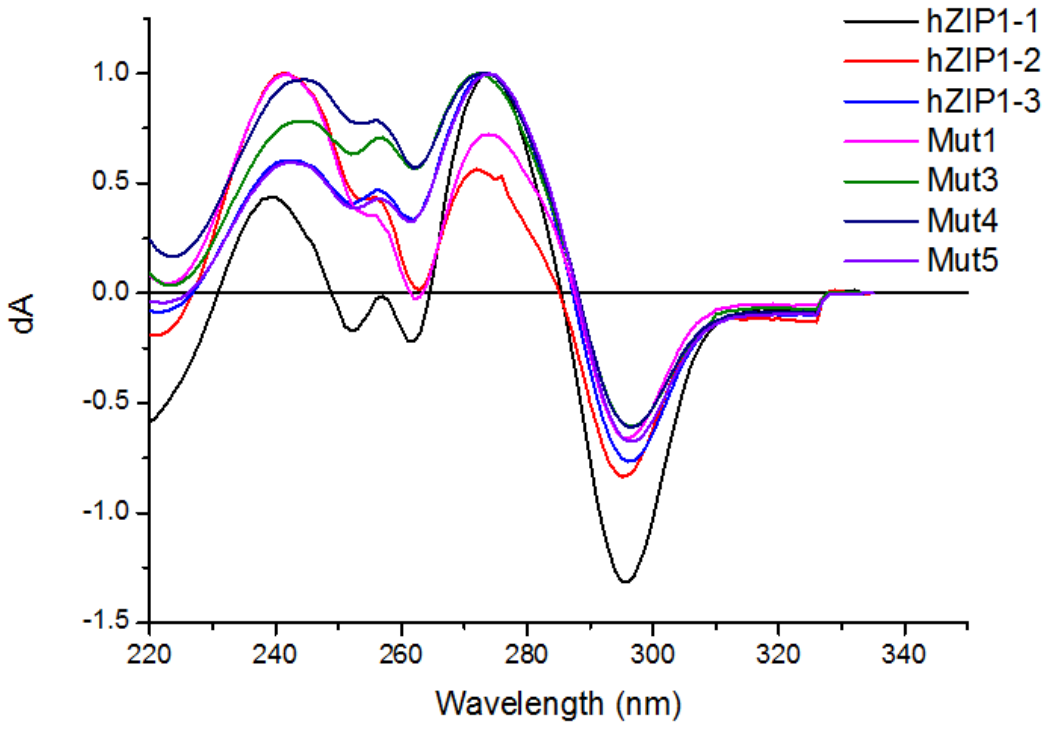


Figure A.4.4 Thermal difference spectra of truncated sequences and mutants (5 µM) annealed by METHOD 2 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl.

.

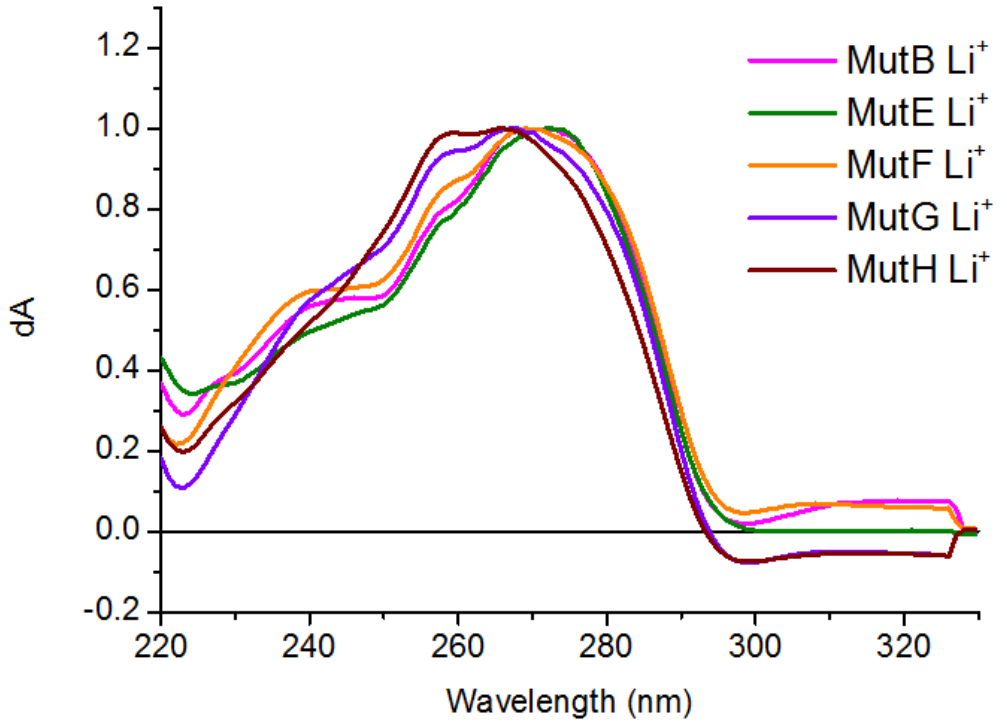


Figure A.4.5 Thermal difference spectra of mutants MutB, MutE, MutF, MutG and MutH (5 µM) annealed by METHOD 2 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM LiCl.

.

A.5 MELTING AND ANNEALING CURVES

Adjusting baselines and transforming to folded fraction as described by Mergny and Lacroix[36] to calculate T_m:

Fit a linear baseline for the folded fraction: $A_F(T) = A_F^0 + m_F T$

Fit a linear baseline for the unfolded fraction: $A_U(T) = A_U^0 + m_U T$

The baselines are used to calculate the median and to convert ellipticity A(T) to folded fraction θ(T):

$$Median(T) = \frac{A_U(T) + A_F(T)}{2}$$

$$\theta(T) = \frac{A(T) - A_U(T)}{A_F(T) + A_U(T)}$$

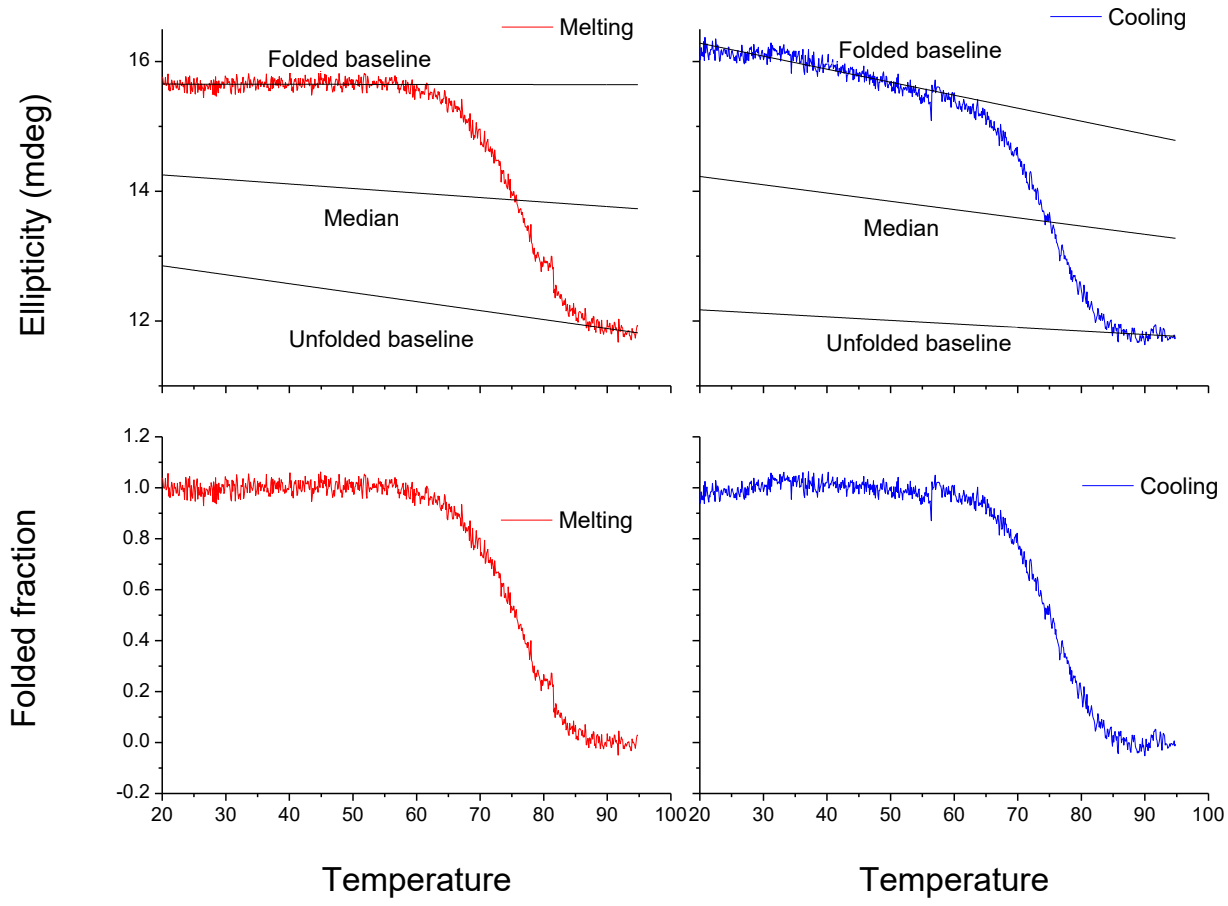


Figure A.5.1 Circular dichroism melting and annealing curves for hZIP1-1 (5 µM) annealed by METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl using a 1 °C/min temperature gradient. On top, data is shown raw with fitted upper and lower baselines and calculated median. On bottom, data is shown after being transformed into folded fraction. When θ=0.5 then T=T_m. Alternatively, when the median intersects the melting curve then T=T_m.

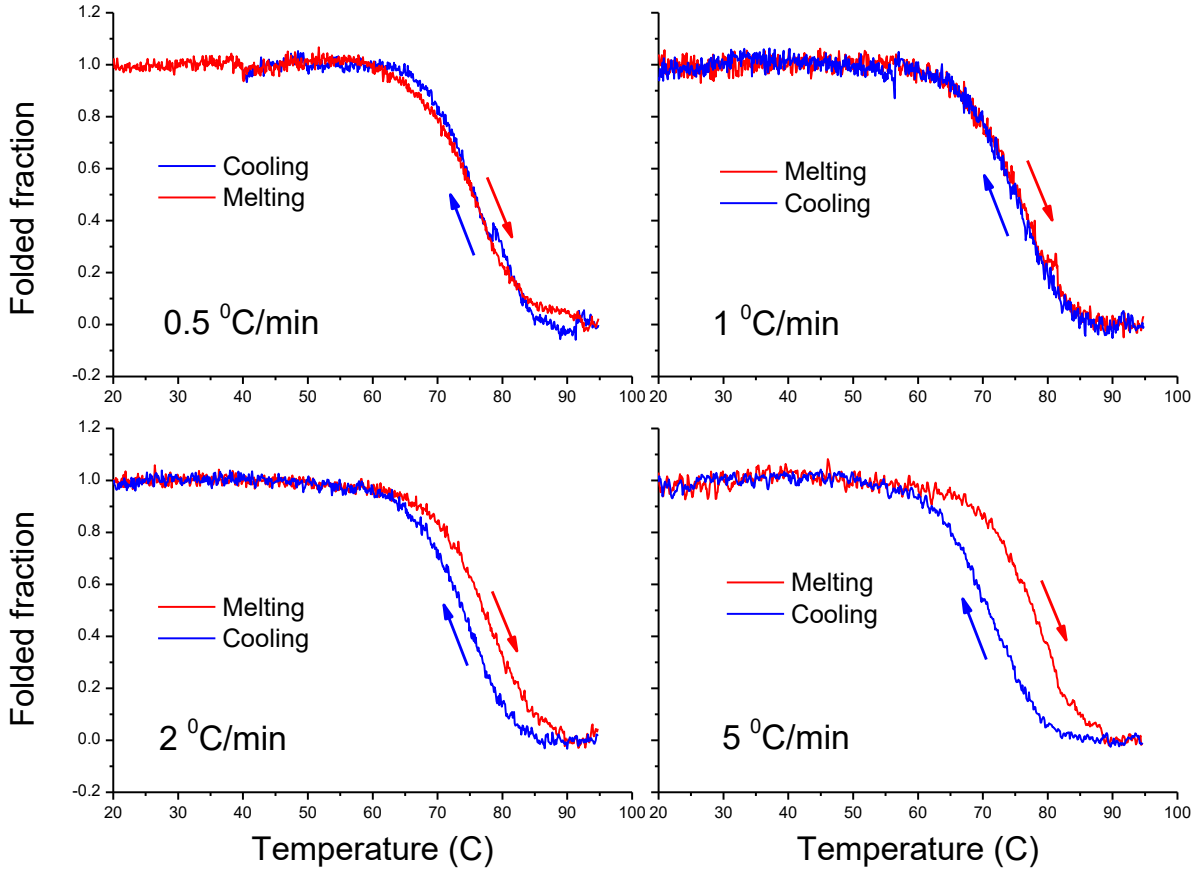


Figure A.5.2 hZIP1-1 (5 µM) annealed by METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl, followed circular dichroism 265 nm peak during melting and annealing at different rates 0.5, 1, 2 and 5 °C/min.

Table A.5.1 Temperature of transition (folded fraction = 0.5) values for hZIP1-1 (5 µM) annealed by METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl, by monitoring 265 nm peak by circular dichroism during melting and annealing at different rates 0.5, 1, 2 and 5 °C/min.

Rate	$T_{1/2}$ (°C)
Melting (0.5 °C/min)	75.58 ± 1.89
Annealing (0.5 °C/min)	75.71 ± 0.35
Melting (1 °C/min)	75.09 ±1.09
Annealing (1 °C/min)	75.05 ± 0.74
Melting (2 °C/min)	76.96 ± 0.14
Annealing (2 °C/min)	73.93 ±0.54
Melting (5/min)	77.88 ± 0.62
Annealing (5/min)	70.58 ±0.38

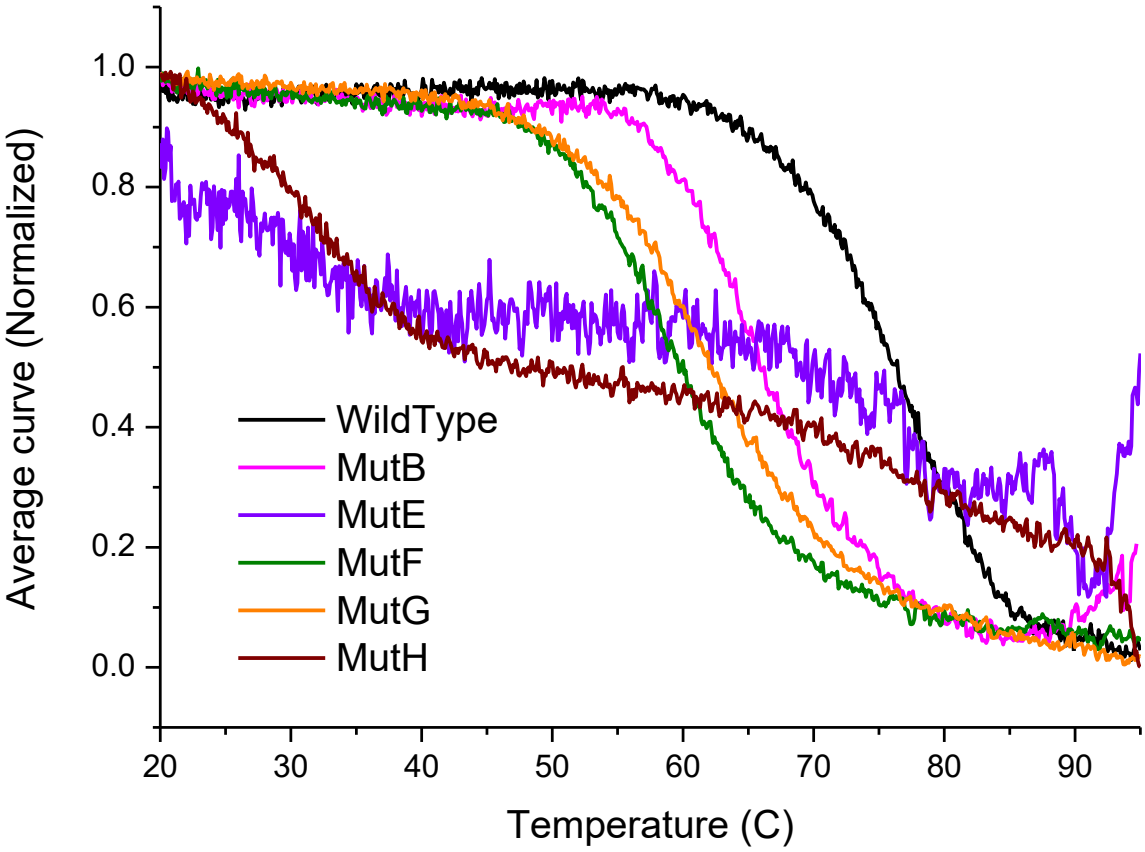


Figure A.5.3 Melting curve of hZIP1-1 (Wild Type), MutB, MutE, MutF, MutG and MutH (5 µM) annealed by METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl, monitored 265 nm peak during temperature increase at 1°C/min. Data is presented normalised (divided by maximum).

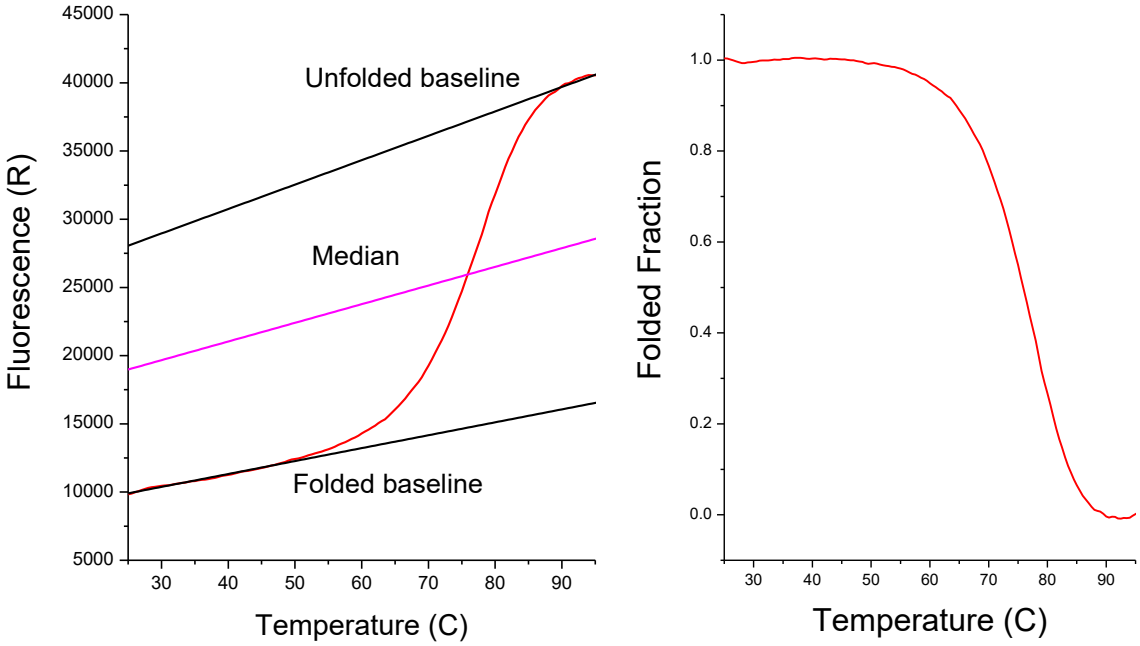


Figure A.6.1 FRET melting curve for FAM-hZIP1-1-TAMRA annealed by METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl monitoring the fluorescence. Raw data is shown to the left with the adjusted baselines, and to the right the data has been processed into folded fraction.

A.6 FRET MELTING

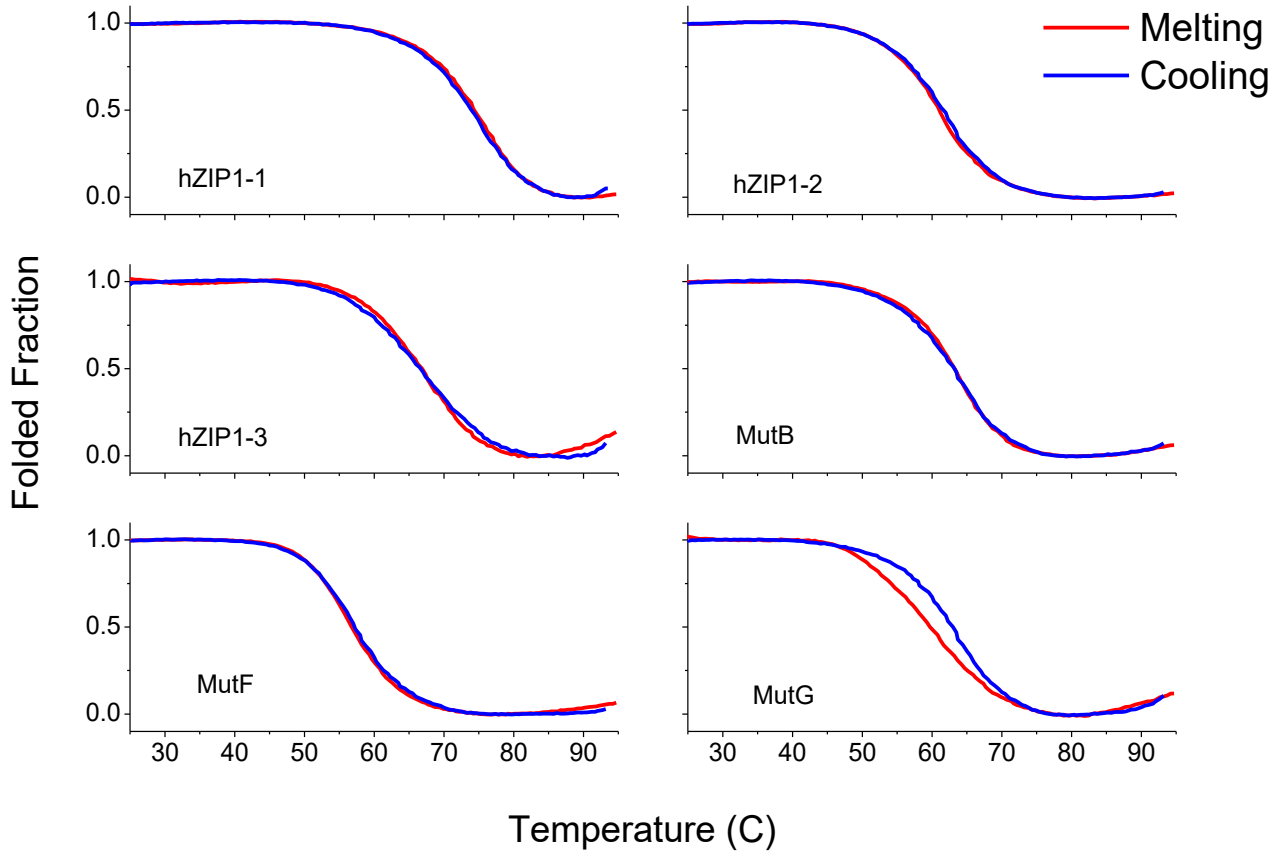


Figure A.6.2 Sequences hZIP1-1 (Wild Type), hZIP1-2, hZIP1-3, MutB, MutF, and MutG annealed by METHOD 3 in 10 mM Li cacodylate buffer pH=7.4 with 100 mM KCl, monitored by FRET during melting and annealing at a rate of 1°C per min

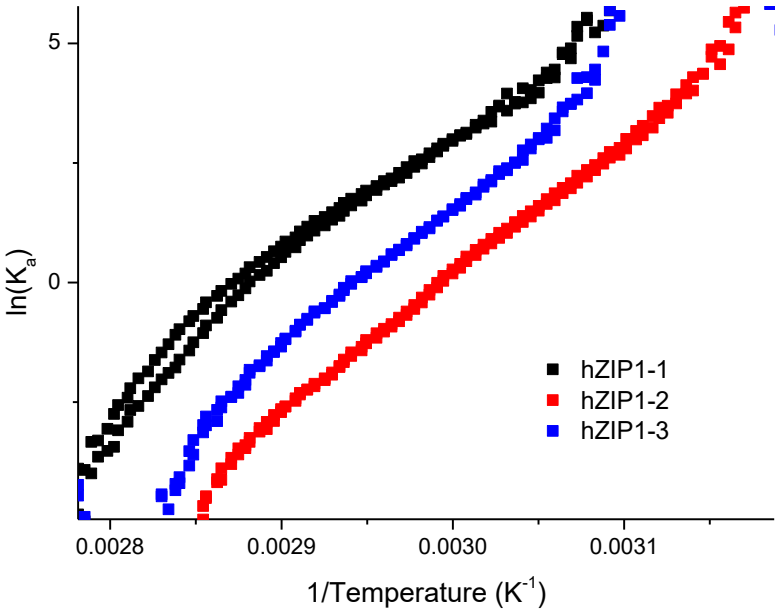


Figure A.6.3 Arrhenius plot for van't Hoff analysis for sequences hZIP-1, hZIP-2, hZIP-3 used to calculate enthalpy values.

Table A.6.1 Thermodynamic data for circular dichroism (CD) and FRET melting and annealing curves for sequences hZIP-1, hZIP-2, hZIP-3,MutB, MutF and MutG.

Name	Conditions		$T_{1/2}$ (°C)	T_m(°C)	ΔH (kJ/mol)	ΔH (kJ/mol)	ΔS (kJ/molK)	ΔS (kJ/molK)
hZIP1-1	CD	melting	75.09±1.09		-273.68±16.53		-0.79±0.05	
		annealing	75.50±0.74	75.30±0.86	-247.97±26.74	-260.83±24.36	-0.71±0.08	-0.75±0.07
	FRET	melting	74.53±0.76		-237.10±18.11		-0.68±0.05	
		annealing	73.97±0.53	74.25±0.63	-241.42±0.57	-239.26±10.76	-0.70±0.003	-0.69±0.03
hZIP1-2	CD	melting	65.60±0.62	65.60±0.62	-235.13±0.02	-235.13±0.02	-0.78±0.12	-0.78±0.12
	FRET	melting	61.05±0.21		-242.92±4.65		-0.73±0.01	
		annealing	61.64±0.10	61.34±0.37	-239.05±4.47	-240.98±4.34	-0.71±0.01	-0.72±0.01
hZIP1-3	CD	melting	71.60±1.63	71.60±1.63	-277.97±84.66	-277.97±84.66	-0.81±0.25	-0.81±0.25
	FRET	melting	66.88±0.17		-258.49±5.49		0.76±0.02	
		annealing	66.58±0.76	66.73±0.48	-213.12±25.59	-235.81±30.24	-0.63±0.08	-0.69±0.09
MutB	CD	melting	66.34±0.83	66.34±0.83	-271.43±44.39	-271.43±44.39	-0.80±0.13	-0.80±0.13
	FRET	melting	63.23±0.37		-250±12.86		-0.74±0.04	
		annealing	63.11±0.7	63.23±0.22	-227.39±8.93	-238.90±16.06	-0.68±0.03	-0.71±0.5
MutF	CD	melting	59.46±0.57	59.46±0.57	-239.55±5.09	-239.55±5.09	-0.72±0.01	-0.72±0.01
	FRET	melting	57.04±2.04	57.18±1.36	-272.29±46.29	-261.07±33.75	-0.83±0.15	-0.79±0.11
		melting	57.32±1.15		-249.85±27.75		-0.76±0.09	
MutG	CD	melting	61.77±1.81	61.77±1.81	-211.45±32.97	-211.45±32.97	-0.63±0.1	-0.63±0.1
	FRET	melting	59.82±0.64		-213.80±12.86		-0.64±0.04	
		annealing	62.92±0.88	61.37±1.90	-230.14±2.30	-221.97±12.08	-0.68±0.01	-0.66±0.03

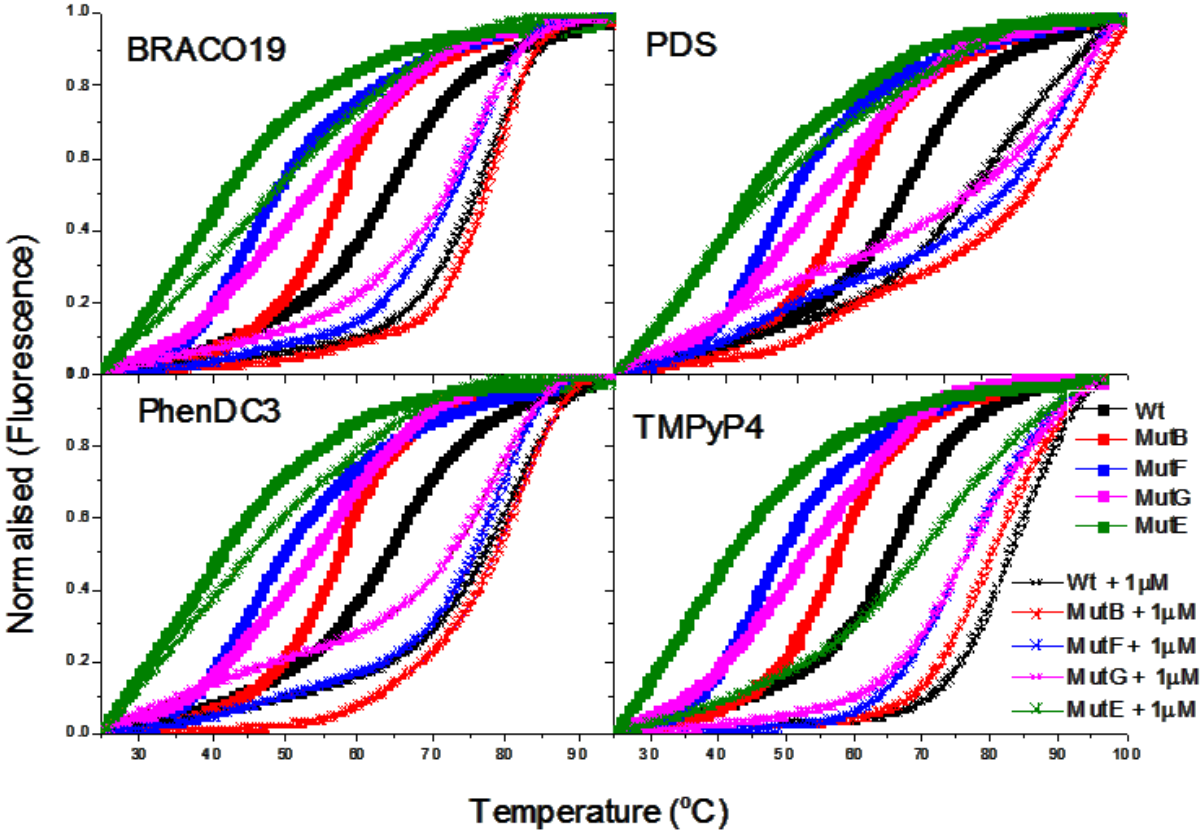


Figure A.6.4 Sequences hZIP1-1 (Wild Type), MutB, MutE, MutF, MutG and MutG FRET melting at 1°C/min stabilised by 1 µM of BRACO-19, Pyridostatin (PDS), PhenDC3 and TMPyP4. The results are shown arranged by ligand.

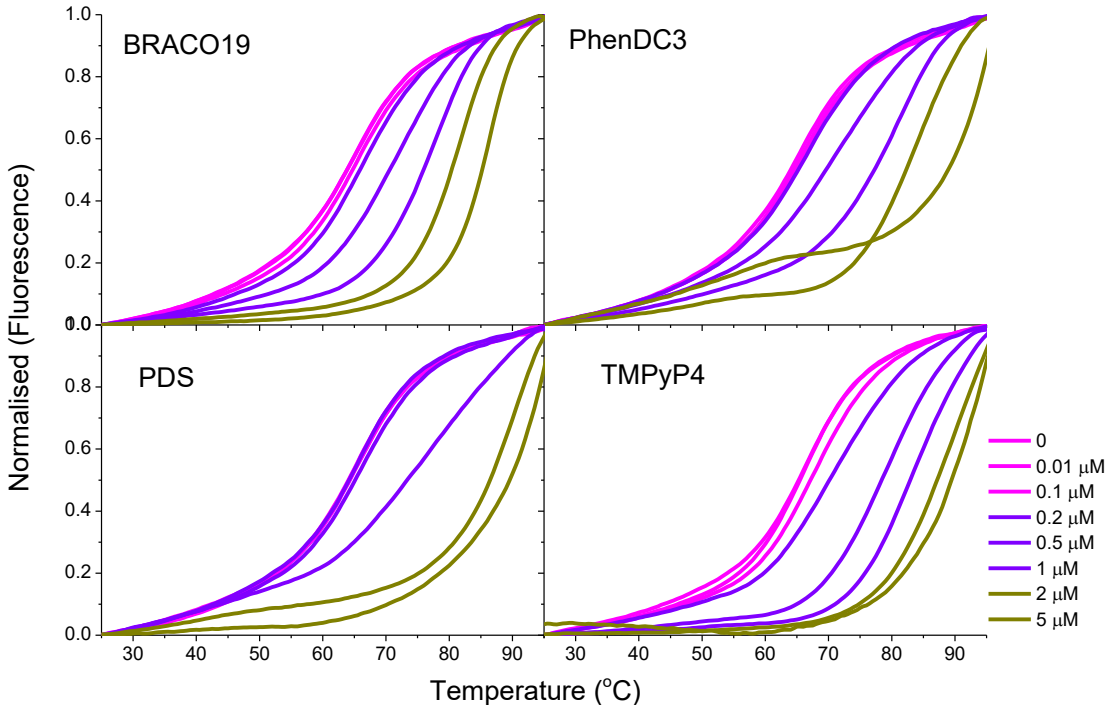


Figure A.6.5 Sequence hZIP1-1 (Wild Type) FRET melting at 1°C/min stabilised by increasing concentrations of BRACO-19, Pyridostatin (PDS), PhenDC3 and TMPyP4. The results are shown arranged by ligand.

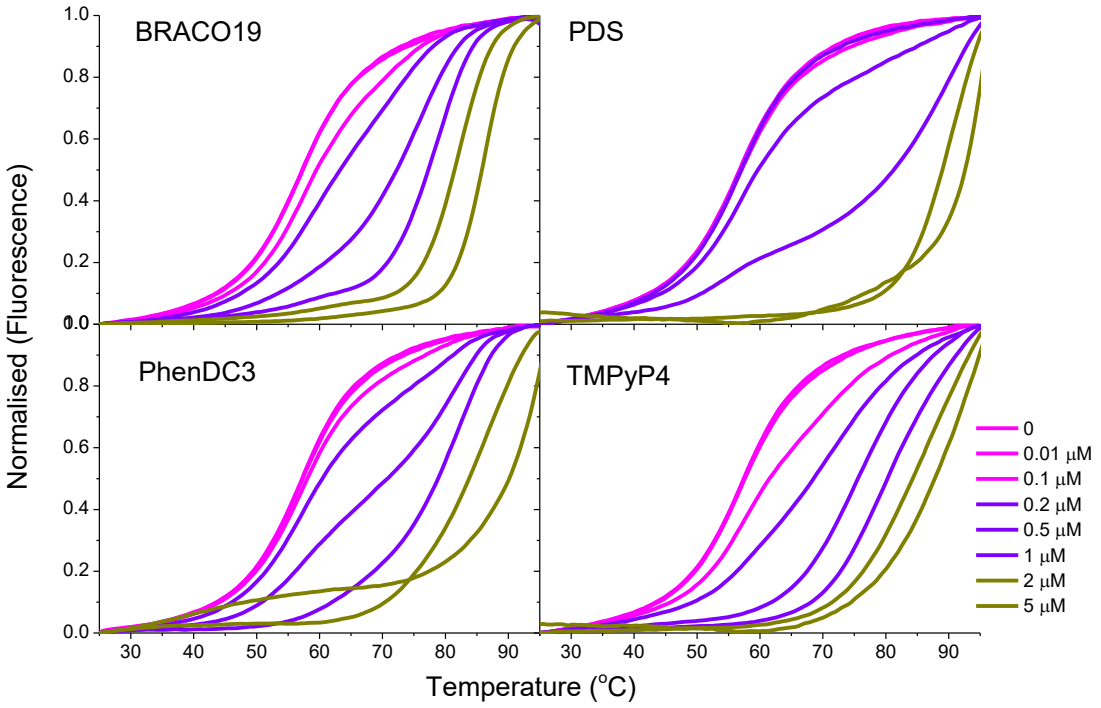


Figure A.6.6 Sequence MutB FRET melting at 1°C/min stabilised by increasing concentrations of BRACO-19, Pyridostatin (PDS), PhenDC3 and TMPyP4. The results are shown arranged by ligand.

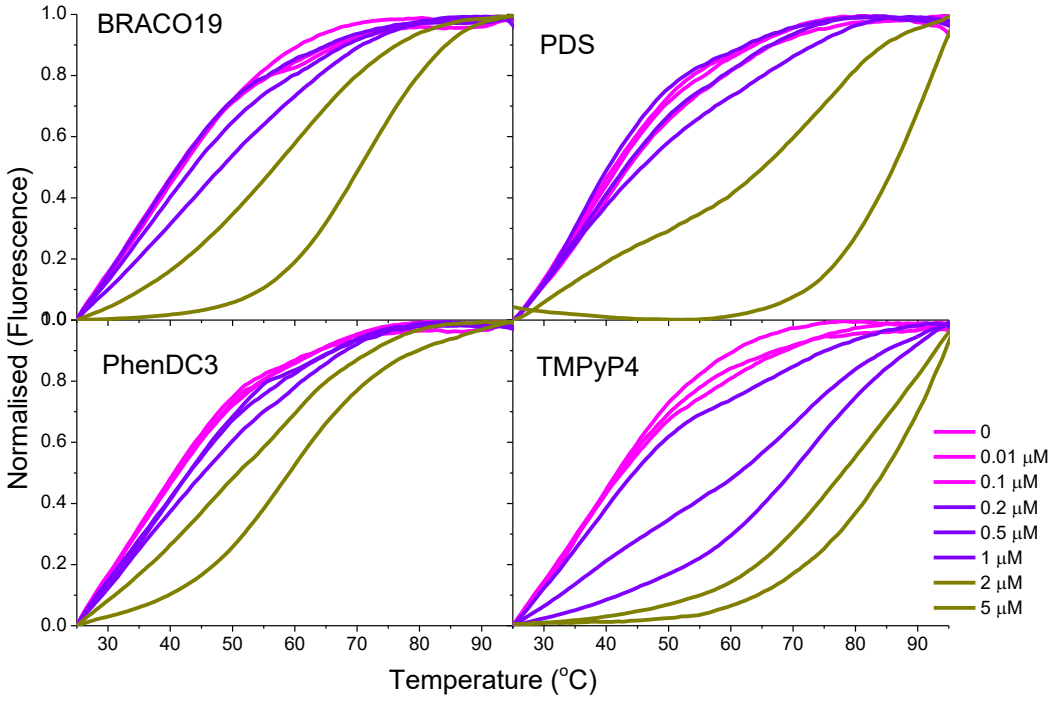


Figure A.6.7 Sequence MutE FRET melting at 1°C/min stabilised by increasing concentrations of BRACO-19, Pyridostatin (PDS), PhenDC3 and TMPyP4. The results are shown arranged by ligand.

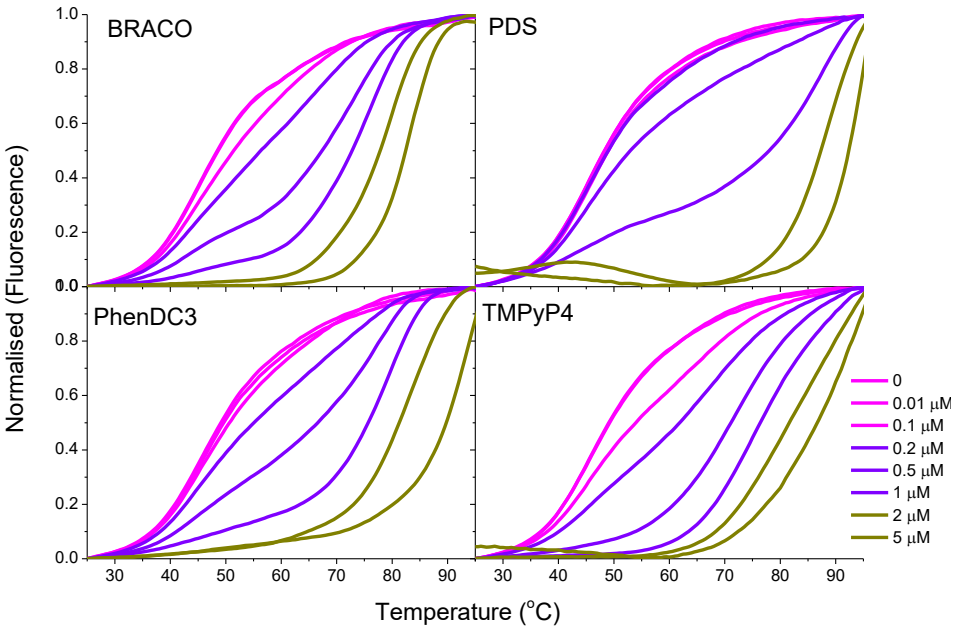


Figure A.6.8 Sequence MutF FRET melting at 1°C/min stabilised by increasing concentrations of BRACO-19, Pyridostatin (PDS), PhenDC3 and TMPyP4. The results are shown arranged by ligand.

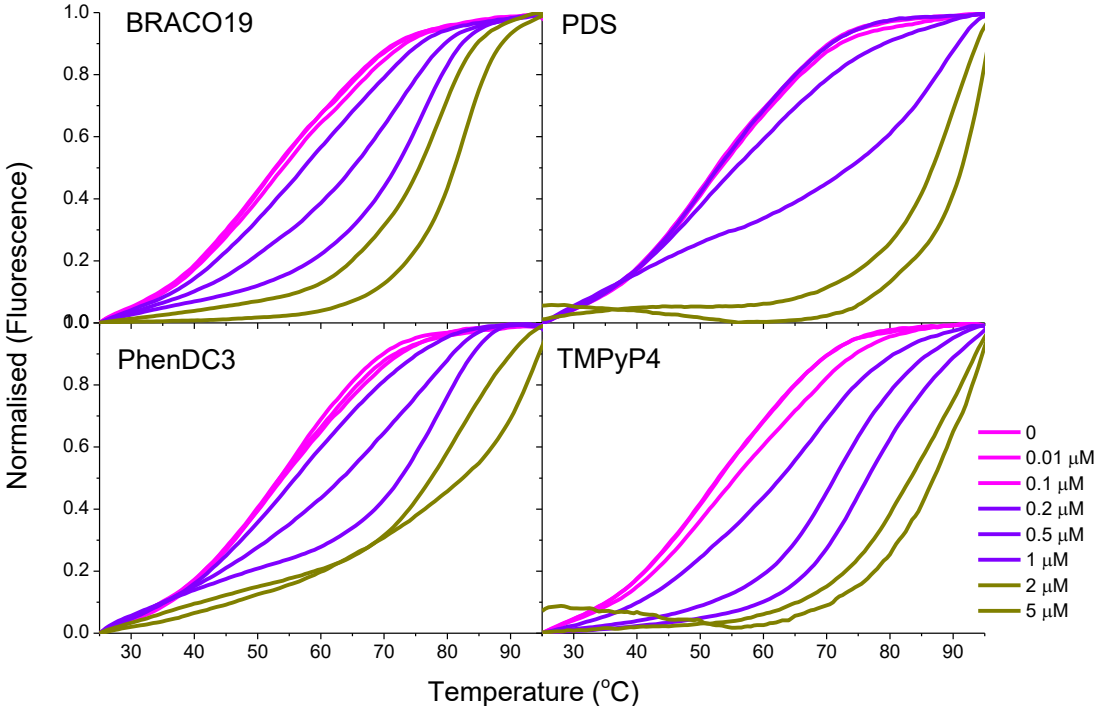
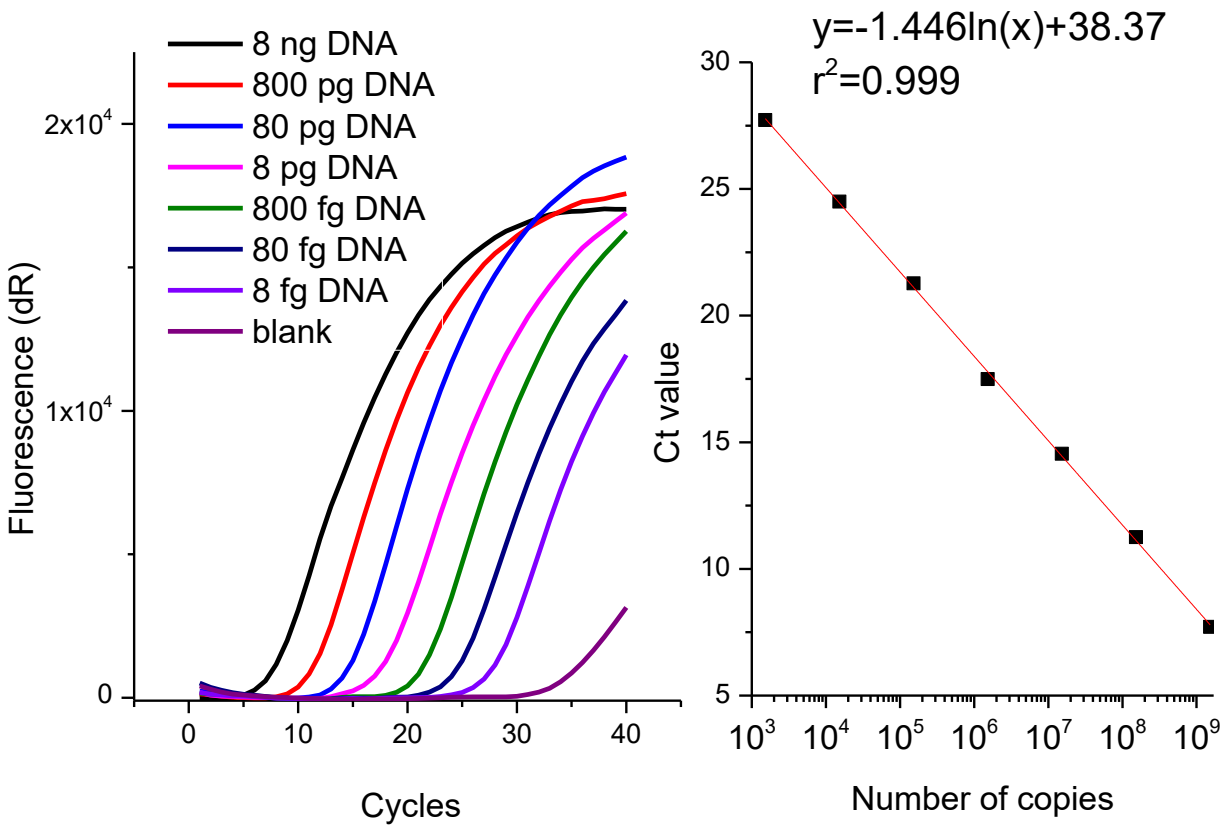


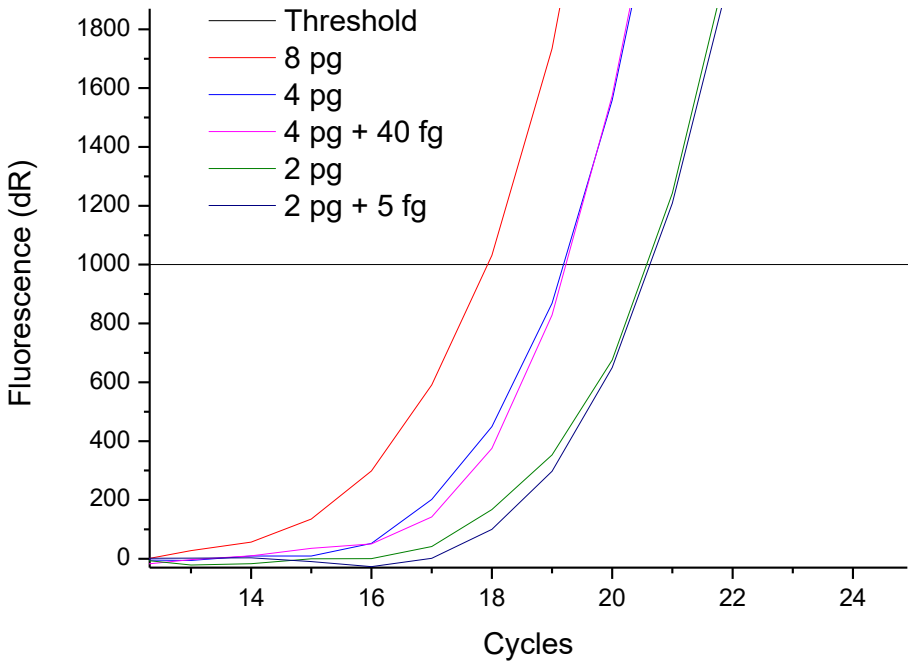
Figure A.6.9 Sequence MutG FRET melting at 1°C/min stabilised by increasing concentrations of BRACO-19, Pyridostatin (PDS), PhenDC3 and TMPyP4. The results are shown arranged by ligand.

A.7 *IN VITRO* TRANSCRIPTION

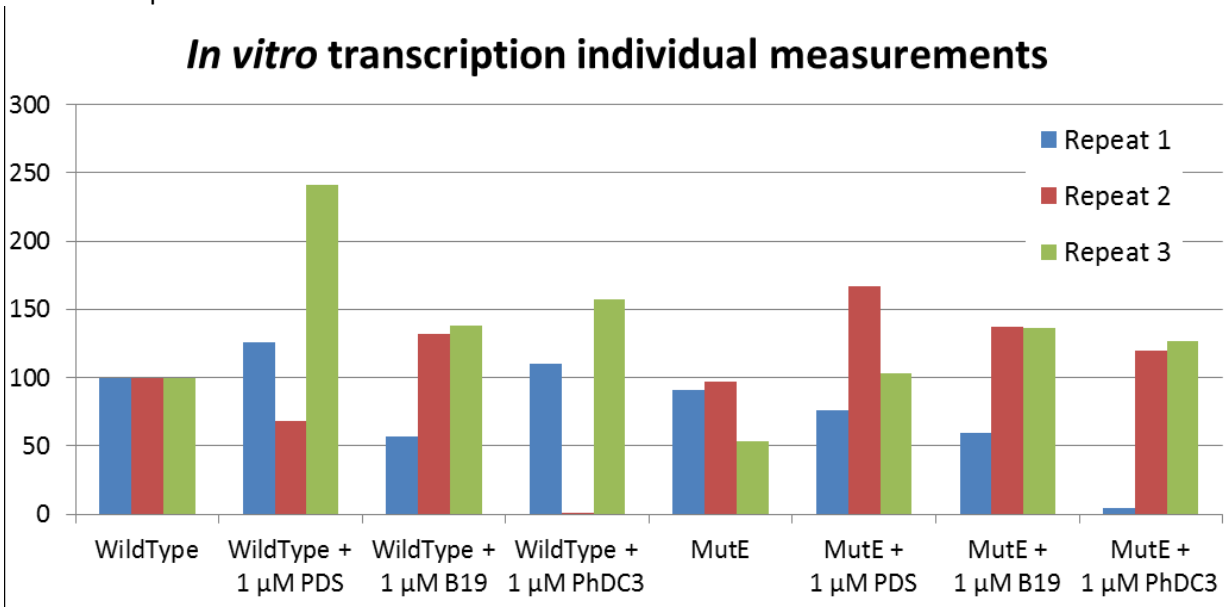


DNA	Copy number	C_T
8 ng	1.52x10^9	7.71 ± 0.43
800 pg	1.52 x10^8	11.25 ± 0.22
80 pg	1.52 x10^7	14.55 ± 0.59
8 pg	1.52 x10^6	17.49 ± 0.11
800 fg	1.52 x10^5	21.27 ± 0.64
80 fg	1.52 x10^4	24.50 ± 0.26
8 fg	1.52 x10^3	27.72 ± 0.32
blank	0	36.16 ± 2.50

A.7.1 Calibration curve for mRNA quantification using RTqPCR. Standards were prepared with linear plasmid.



A.7.2 Effect on amplification of diluted DNA in amplification of a concentrated sample.



A.7.3 Individual measurements for *in vitro* transcription reactions. mRNA levels were quantified using RTqPCR and normalised to the WildType mRNA leves of each repeat.

A.8 Luciferase Reporter Assay

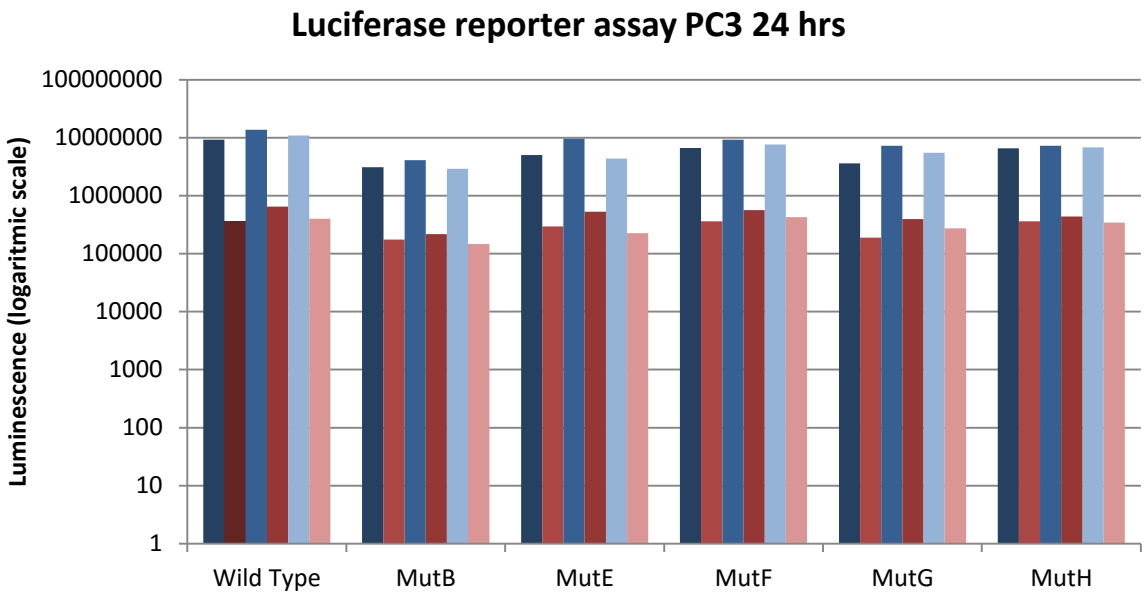


Figure A.8.1 Raw luminescence values for each repeat obtained for firefly (blue) and renilla (red) luciferases before normalisation for luciferase reporter assay, PC3 cell line, 24 hours after transfection.

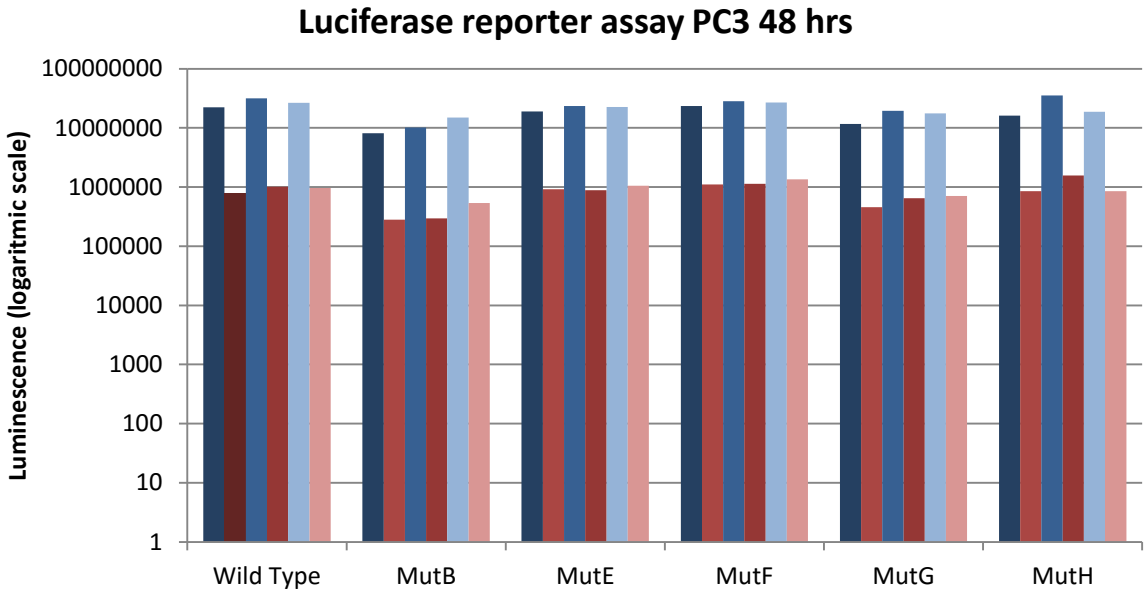


Figure A.8.2 Raw luminescence values for each repeat obtained for firefly (blue) and renilla (red) luciferases before normalisation for luciferase reporter assay, PC3 cell line, 48 hours after transfection.

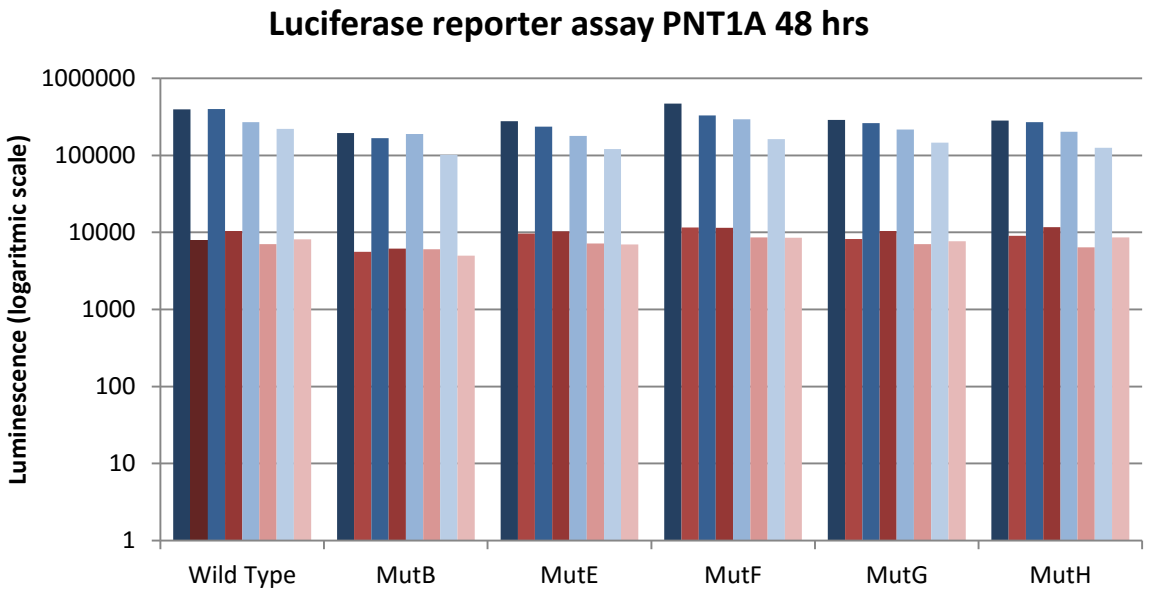


Figure A.8.3 Raw luminescence values for each repeat obtained for firefly (blue) and renilla (red) luciferases before normalisation for luciferase reporter assay, PNT1A cell line, 48 hours after transfection.

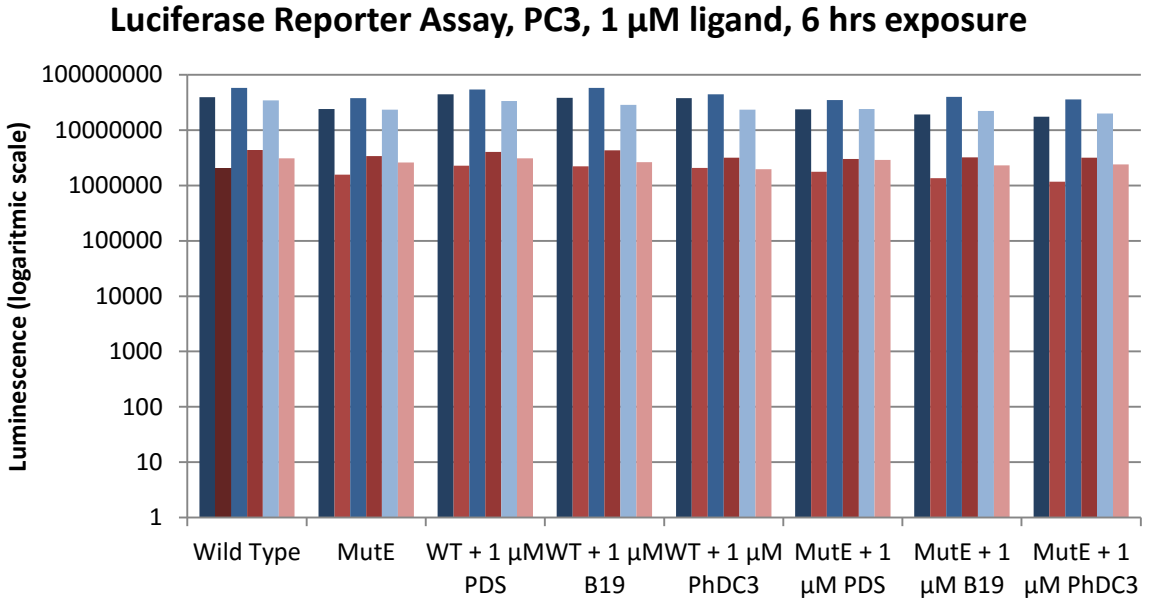


Figure A.8.4 Raw luminescence values for each repeat obtained for firefly (blue) and renilla (red) luciferases before normalisation for luciferase reporter assay, PC3 cell line, 24 hours after transfection and 6 hours after exposure to 1 µM ligand.

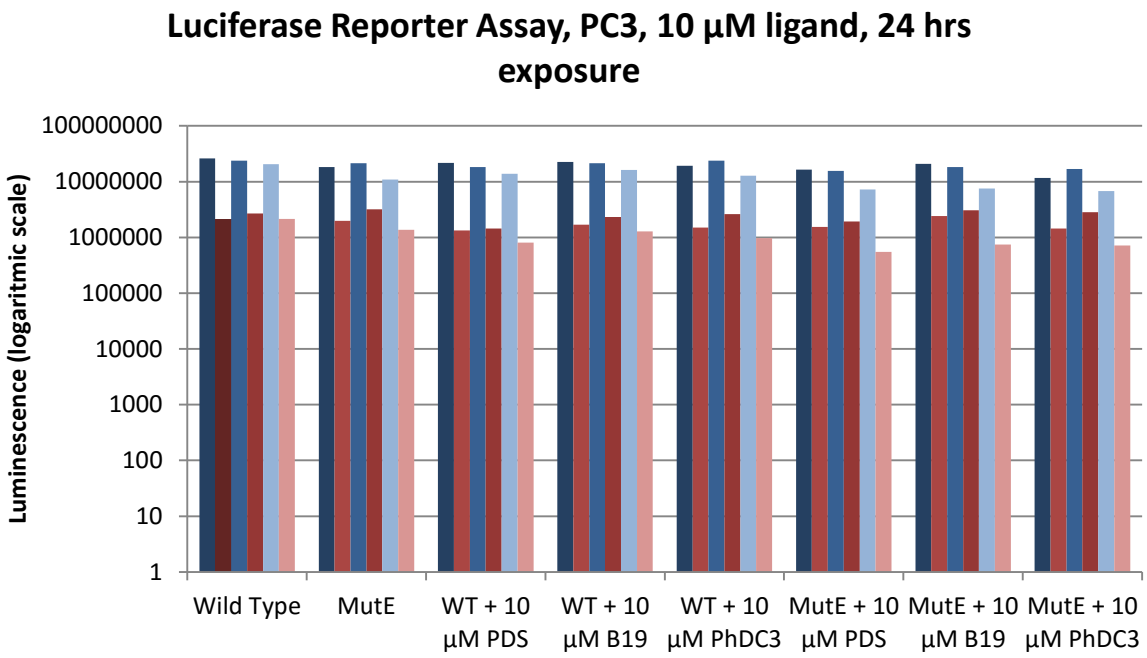


Figure A.8.5 Raw luminescence values for each repeat obtained for firefly (blue) and renilla (red) luciferases before normalisation for luciferase reporter assay, PC3 cell line, 48 hours after transfection and 24 hours after exposure to 10 µM ligand.

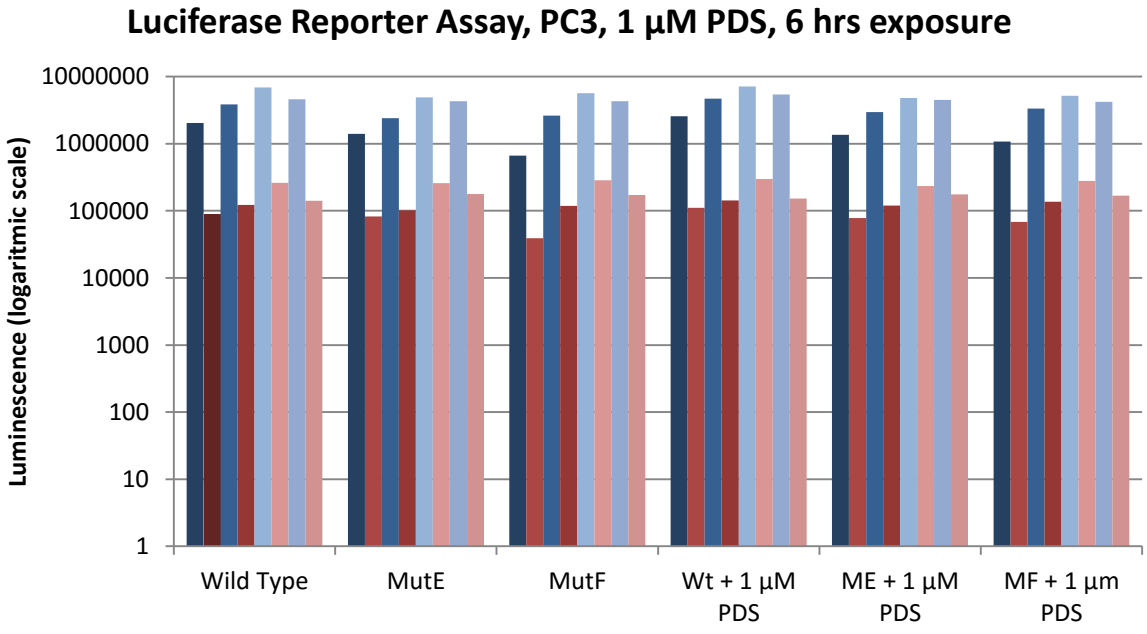


Figure A.8.6 Raw luminescence values for each repeat obtained for firefly (blue) and renilla (red) luciferases before normalisation luciferase reporter assay, PC3 cell line, 24 hours after transfection and 6 hours after exposure to 1 µM Pyridostatin.

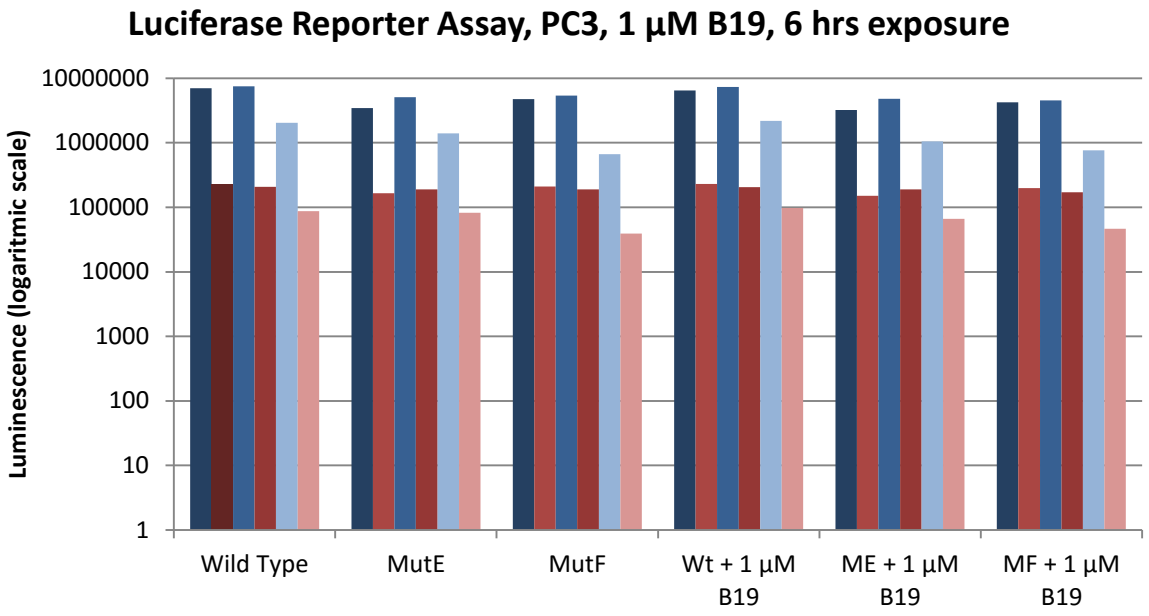


Figure A.8.7 Raw luminescence values for each repeat obtained for firefly (blue) and renilla (red) luciferases before normalisation for luciferase reporter assay, PC3 cell line, 24 hours after transfection and 6 hours after exposure to 1 µM BRACO-19.

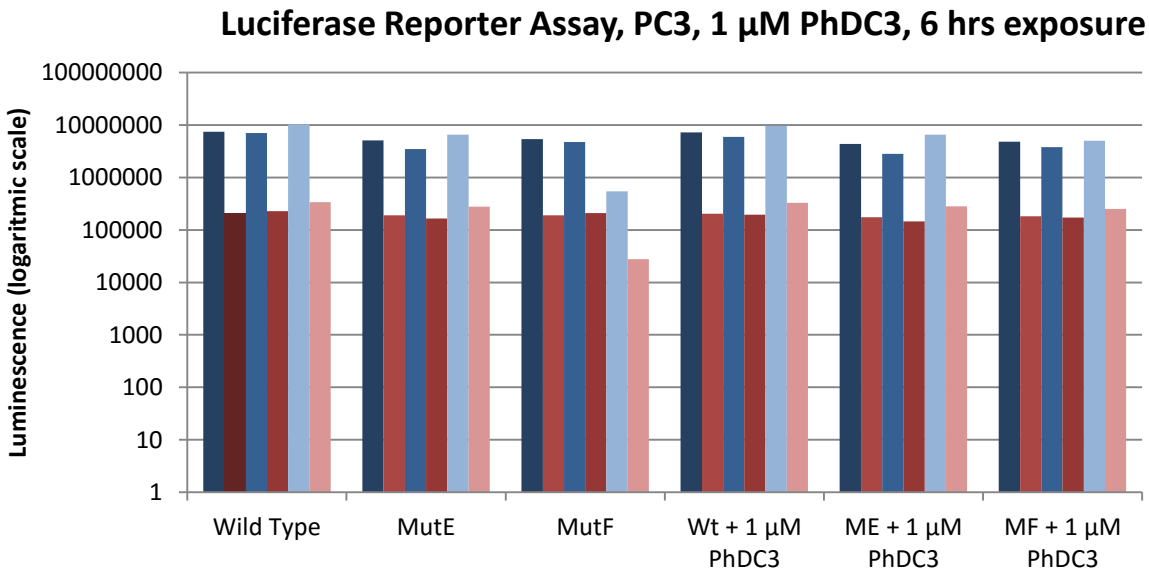


Figure A.8.8 Raw luminescence values for each repeat obtained for firefly (blue) and renilla (red) luciferases before normalisation for luciferase reporter assay, PC3 cell line, 24 hours after transfection and 6 hours after exposure to 1 µM PhenDC3

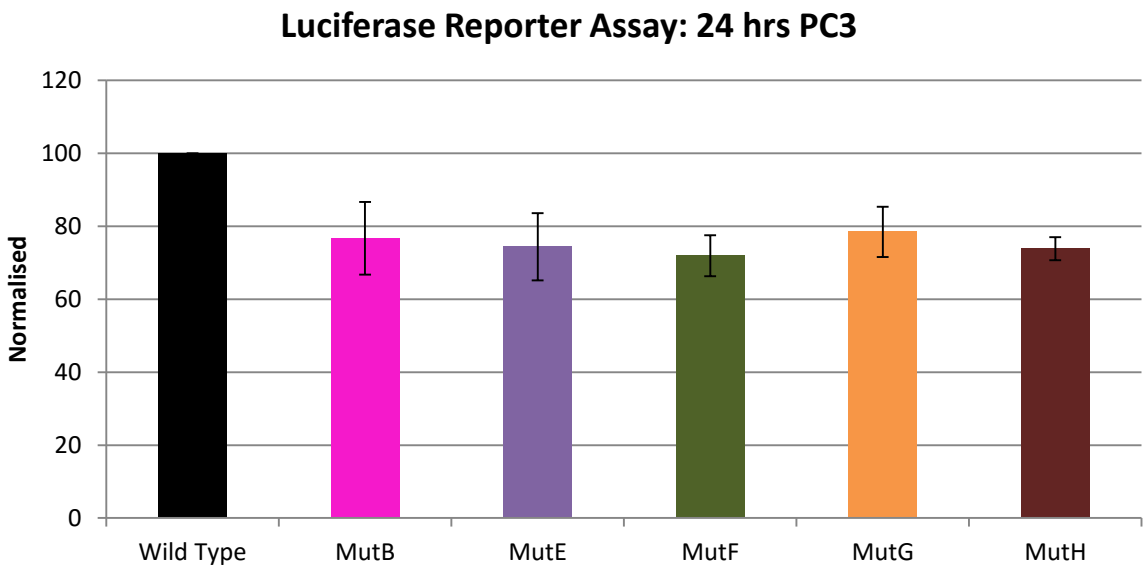


Figure A.8.9 Luciferase reporter assay 24 hours after PC3 cells were transfected with WildType plasmid (hZIP1) and its mutants. Firefly luciferase values were normalised to renilla luciferase values, and then all values were expressed as a percentage of normalised WildType activity. Error bars show standard deviation.

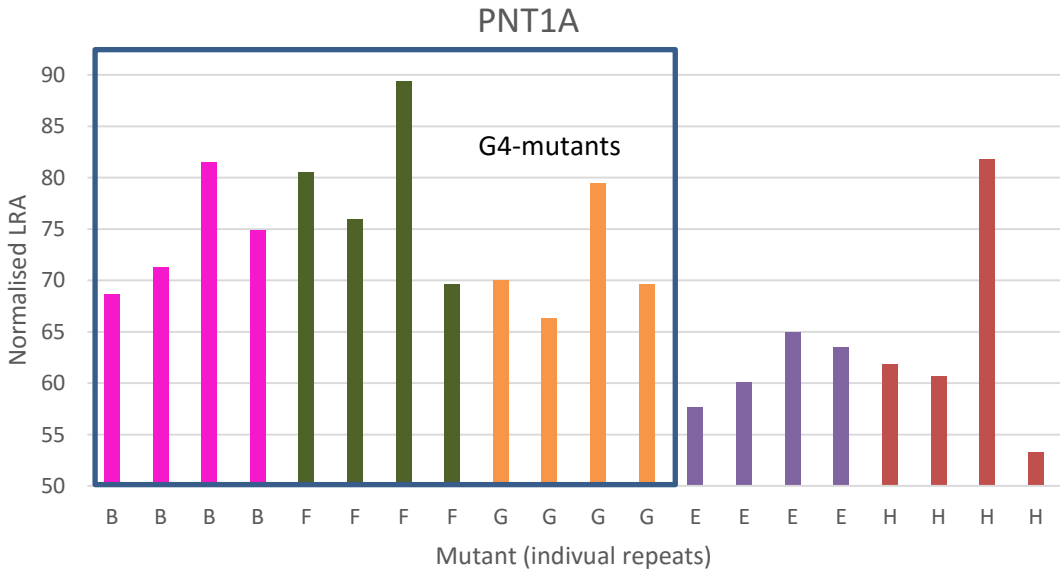


Figure A.8.10 Luciferase reporter assay (LRA) 48 hours after PNT1A cells were transfected with each mutant. Box shows G4-forming mutants. Firefly luciferase values were normalised to renilla luciferase values, and then all values were expressed as a percentage of normalised WildType activity. Each value is shown individually.

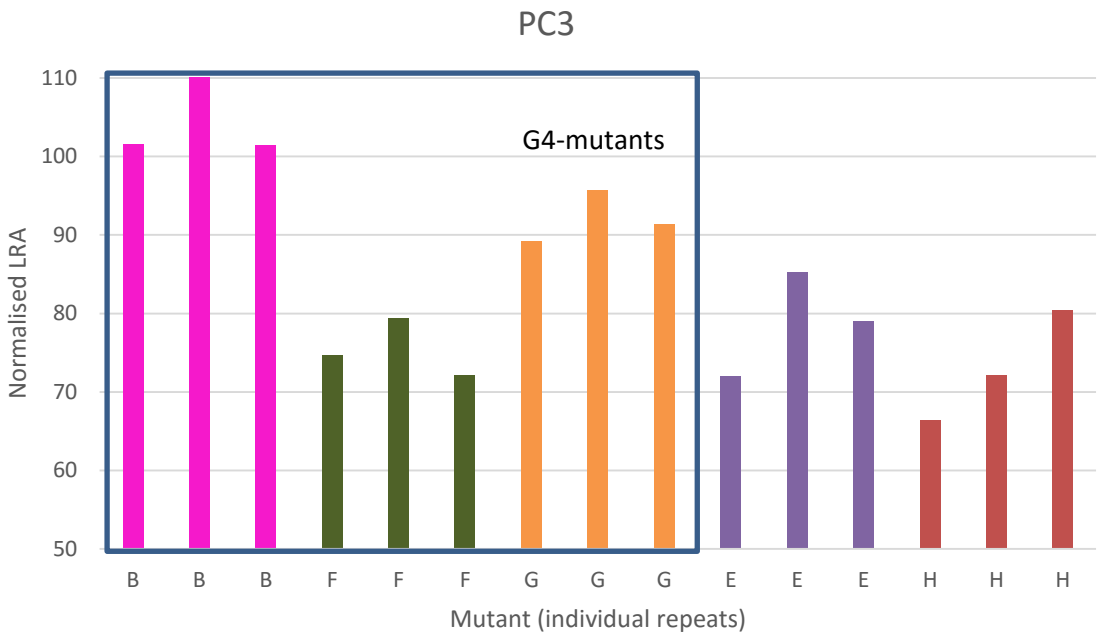


Figure A.8.11 Luciferase reporter assay (LRA) 48 hours after PC3 cells were transfected with each mutant. Box shows G4-forming mutants. Firefly luciferase values were normalised to renilla luciferase values, and then all values were expressed as a percentage of normalised WildType activity. Each value is shown individually.

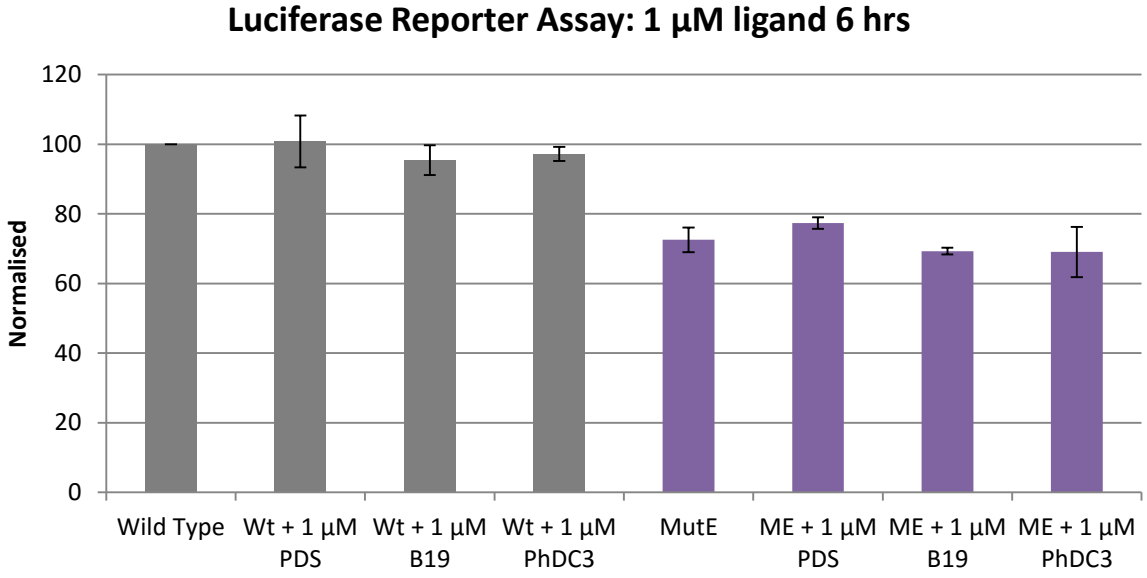


Figure A.8.12 Luciferase reporter assay normalised, 24 hours after PC3 cells were transfected with WildType plasmid (hZIP1) and MutE, and 6 hours after adding 1 µM of ligand. Firefly luciferase values were normalised to renilla luciferase values, and then all values were expressed as a percentage of normalised WildType activity. Error bars show standard deviation.

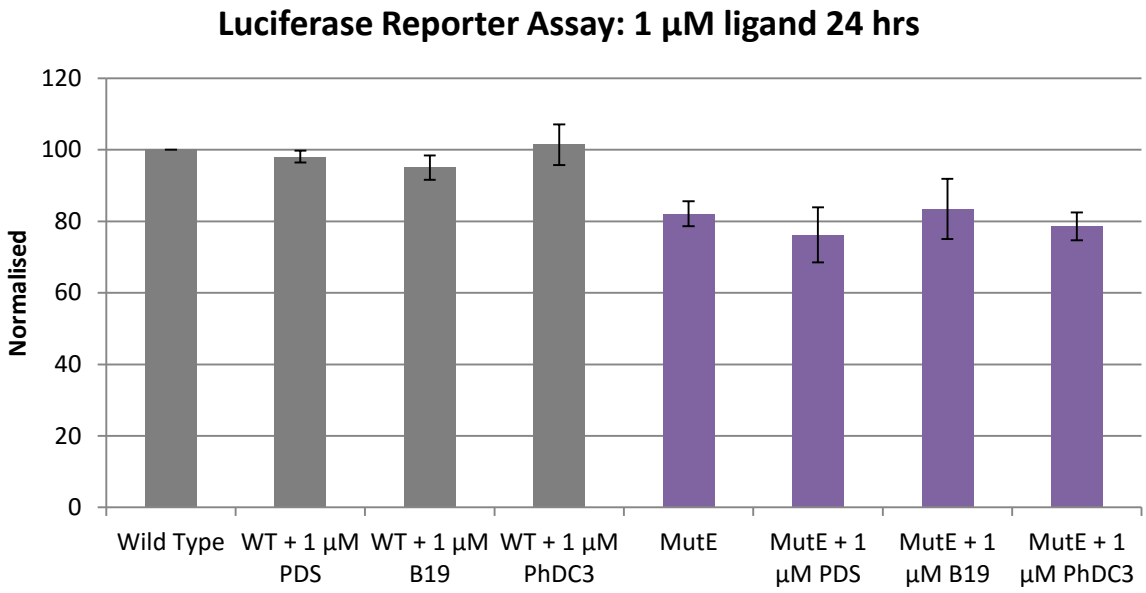


Figure A.8.13 Luciferase reporter assay 48 hours after PC3 cells were transfected with WildType plasmid (hZIP1) and MutE, and 24 hours after adding 1 µM of ligand. Firefly luciferase values were normalised to renilla luciferase values, and then all values were expressed as a percentage of normalised WildType activity. Error bars show standard deviation