Investigating DNA's Mysterious Structures:

Development of Probes to Target the i-Motif Forming Region of the *MYC* Promoter

Summer Rosonovski

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

School of Pharmacy

November 2021

Declaration

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

Acknowledgements

My time at UEA would not have been the same without the past and present Waller and Morris groups who have shared in the highs and lows of lab-life. A special thanks to each of you that taught me biophysical or molecular techniques and for sharing your expert knowledge. An even bigger thanks for all the enjoyment you have all brought to my life, making work a place I looked forward to coming to and fun filled evenings and weekends exploring Norfolk, particularly the restaurants and bars.

Thank you to all my friends and family, in particular my mum and Atlanta, for their continued love and support through what has been the most challenging experience of my life. You've all given me words of encouragement, advice, and brought huge amounts of joy and fun to my life that have been essential to my success which I can never thank you all enough for. In particular to Grace and Conor for being my Norwich Family and being part of the tears, the triumphs, and the wine nights. Meeting you both was one of the most precious parts of this experience and I will always love you both dearly, even if you have left me for a warmer, more beautiful home. A special thanks to my Luton friends and family for being the perfect guinea pigs for my Sci Coms projects your feedback is always invaluable and I have loved being able to share my research with you all.

Abstract

Chapter 1 is a general introduction to DNA and the alternative secondary structures that it forms, leading into an introduction to *MYC* the oncogene and techniques used to determine probes that bind to the i-motif that forms in the *MYC* promoter region.

Chapter 2 explores the interactions of small molecules with the *MYC* i-motif DNA using a high-throughput biophysical screen with a library of 1596 compounds. The results are validated by further biophysical techniques with the most promising compounds continued into the next stages of development finishing with the beginning of cellular studies.

Chapter 3 explores the interactions of peptides with the *MYC* i-motif DNA using phage display with different phage libraries contain 10^9 different peptide sequences. The first section discusses method development of the phage display technique to achieve optimal conditions. The second uses biophysical techniques to demonstrate peptides binding to the *MYC* i-motif.

Chapter 4 discusses *Chapters 2* and *3* in the context of the current literature and explores the future work that this project could take.

Chapter 5 discusses the experimental procedures used in *Chapters 2* and *3*.

Table of Contents

List of Figures

Chapter 1: Introduction

Chapter 2: Compounds

Chapter 3: Peptides

Chapter 4: Discussions and Future Work

List of Tables

Chapter 1: Introduction

Chapter 2: Compounds

Chapter 3: Peptides

Chapter 5: Experimental

List of Abbreviations

A	Adenine
A549	Adenocarcinomic human alveolar basal epithial cells
ATXN2L	i-Motif formed in the promoter region of the gene that encodes Ataxcin-2-like protein
B95.8	Marmoset B-lymphoblastoid cell line
Bcl-2	i-Motif formed in the promoter region of the B-cell lymphoma-2 gene
BSA	Bovine serum albumin
°C	Degrees Celsius
C	Cytosine
CD	Circular Dichroism
CNBP	Cellular Nucleic-acid-Binding Protein
c-Myc	An oncogene promoter sequence containing secondary structures
c-MycC	The i-motif forming sequence in the *MYC* oncogene promoter region
c-MycC27	The 27 base pair sequence variation of the i-motif forming sequence in the *MYC* oncogene promoter region
c-MycC33	The 33 base pair sequence variation of the i-motif forming sequence in the *MYC* oncogene promoter region
c-MycC52	The 52 base pair sequence variation of the i-motif forming sequence in the *MYC* oncogene promoter region
c-MycG	The G-quadruplex forming sequence in the *MYC* oncogene promoter region
CTD	C-terminal domain
C-tract	Repeat sequence of cytosines hydrogen bonded to other cytosines

DAP	Death associated proteins
DMF	Dimethyl formamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribose nucleic acid
ELISA	Enzyme-linked immunosorbent assay
FBS	Foetal Bovine Serum
FID	Fluorescence Indicator Displacement
G	Guanine
G-quadruplex	Secondary DNA structure that forms in guanine rich sequences
G-quartet	Four guanines hydrogen bonded to each other in a circle
G-tract	Repeat sequence of guanines hydrogen bonded to other guanines
H-bonding	Hydrogen bonding
HeLa	Immortal cervical cancer cell line
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
hTeloC	i-Motif formed in the human telomeric region sequence
IC$_{50}$	Half maximal inhibitory concentration
ILPR	i-Motif formed in the insulin linked polymorphic region, tandem repeat sequence
KCl	Potassium chloride
KH	K homology
NaCaco	Sodium Cacodylate
NCI	National Cancer Institute
NHE III$_1$	Nuclease Hypersensitivity Element III$_1$ upstream of P1 promoter in *MYC* promoter region
NKE	Normal Kidney Epithelial Cell Line
NEB	New England BioLabs
NLS	Nuclear localisation sequence

NMR	Nuclear magnetic resonance spectroscopy
NSC	Compound identifier assigned by Developmental Therapeutics Program (DTP) to identify an agent, Cancer Chemotherapy National Service Centre number
NTD	N-terminal transactivation domain
MAX	MYC-associated protein X
MST	Microscale thermophoresis
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2h-tetrazolium bromide, a colorimetric assay for assessing metabolic activity
i-Motif	Secondary DNA structure that forms in cytosine rich sequences
IPGT	Isoproyl-β-D-thiogalactoside
ODN	Oligonucleotide sequences
PBS	Phosphate buffered saline buffer
PBSB	Phosphate buffered saline with BSA blocking buffer
PBST	Phosphate buffered saline with Tween20 washing buffer
PFU	Plaque forming units
pH	Power of hydrogen
Rb	Retinoblastoma gene
RET	Rearranged during transfection gene
RF	Replicated form
RMPI-1640	Roswell Park Memorial Institute Medium for culturing cells
RNA	Ribonucleic acid
RP-HPLC	Reverse phase high performance liquid chromatography
RT	Room temperature
RU	Resonance Unit

SM	A buffered used for phage manipulation and to stabilise phage for storage
SPR	Surface Plasmon Resonance
T	Thymine
T$_m$	Melting temperature
ΔT$_m$	Change in melting temperature
TBS	Tris-buffered saline
TO	Thiazole Orange
UV	Ultraviolet
VEGF	Vascular Endothelial Growth Factor gene
Xgal	5-Bromo-4-chloro-3-indolyl-ß-D-galactoside

Chapter 1: Introduction

1.1 Secondary DNA Structure

The story of DNA truly begins with Friedrich Miescher in 1869, a biochemist trying to elucidate the building blocks of life. His research began with isolating proteins from white blood cells, where he identified an unexpected substance with chemical properties differing from proteins. Miescher was fascinated with this novel substance and referred to it as 'nuclein'.[9,10] Unknown to Miescher, this was the moment that the molecule of life, DNA, was discovered.[11,12]

The next most notable discovery was by Albrecht Kossel in 1881 in identifying that 'nuclein' was a nucleic acid, it was this key discovery that gave the molecule of life its current name deoxyribonucleic acid (DNA).[13] Kossel also isolated all the purine and pyrimidine bases and identified that DNA is confined to the nucleus. Despite these early break throughs, the extent of DNA's significance was yet to be realised for many decades.

Views began to shift from the late 1940s following Erwin Chargaff's discovery that relative DNA base proportions were identical between individuals in the same species but differed between species. Chargaff also determined that molar ratios of A/T and C/G were essentially always one.[14,15] This paired with the discovery that DNA carries hereditary information brought DNA research centre stage.[16,17] This rediscovered curiosity in DNA led to the historical 1953 paper by James Watson and Francis Crick, based on Rosalind Franklin's, Maurice Wilkins, and Charegaff's work, to propose the double helix structure of DNA, known as B-DNA (*figure 1.1*).[18]
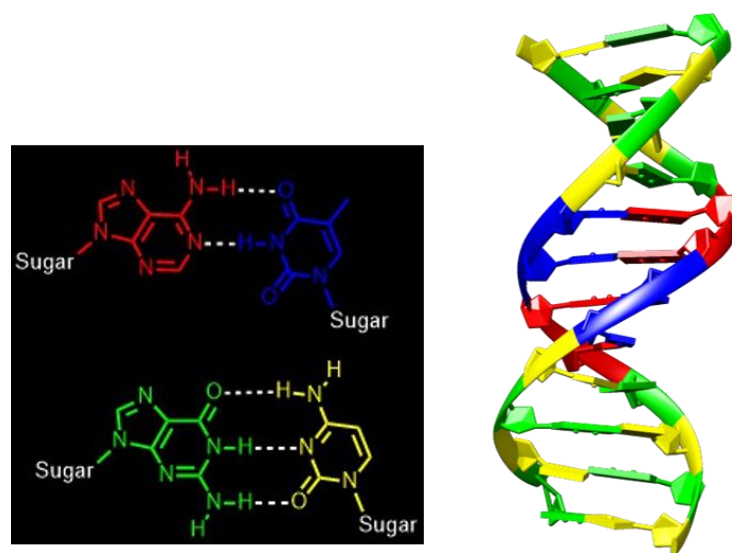
Figure 1.1.: Watson and Crick base pairing and the structure of B-form DNA (PDB ID: 1BNA). Colouring scheme: red is adenine, blue is thymine, green is guanine, yellow is cytosine.

This seemingly simple molecule has not stepped off the centre stage since 1953, due to its functionally important numerous structural polymorphisms (*figure 1.2*).[19–21] DNA can form three different double helical structures: B-DNA and A-DNA are both right-handed, antiparallel structures, where A is a shorter, wider version of B that requires a reduced water environment to form.[21] The third double helix structure is Z-DNA which is a left-handed helix that requires a non-physiological high-salt environment (*figure 1.2*). Certain other conditions permit other DNA structures to form including hairpins, triplexes, cruciforms, junctions, and quadruplexes (*figure 1.2*).[19,20]
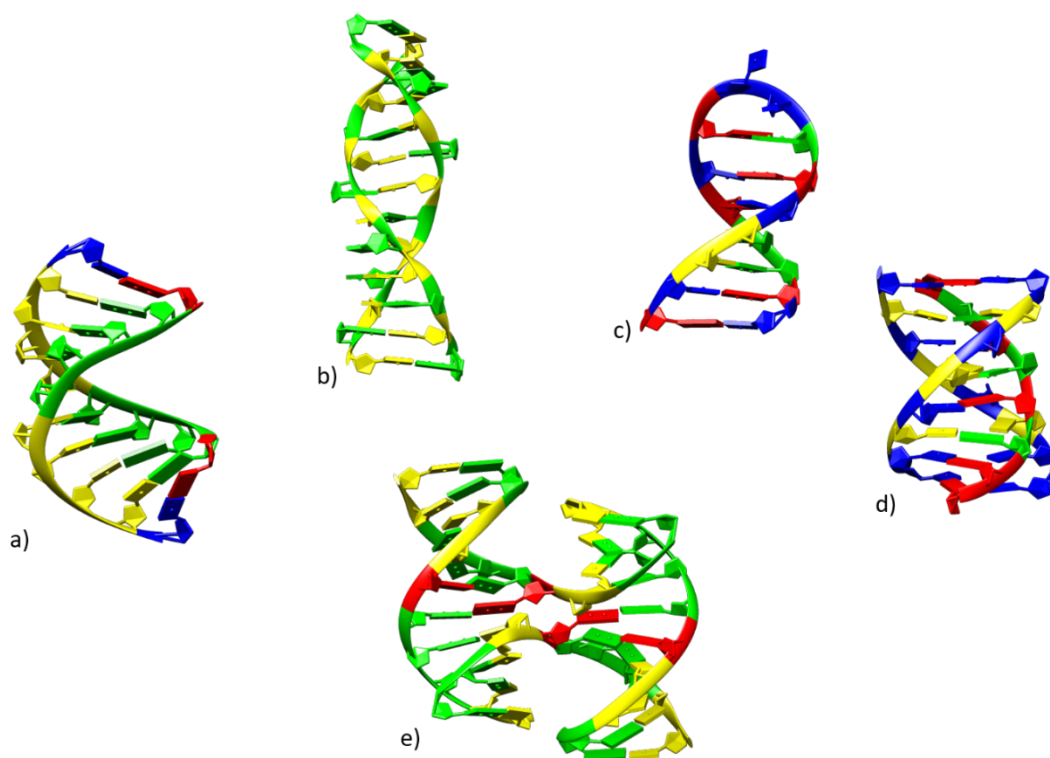
Figure 1.2: Schematic of some alternative DNA secondary structures. a) A-DNA (PBD ID: 440D) b) Z-DNA (PBD ID: 4OCB) c) Hairpin (PBD ID: 1AC7) d) triplex (PDB ID: 1D3X) e) Holiday Junction (PBD ID: 467D). Colouring scheme: red is adenine, blue is thymine, green is guanine, yellow is cytosine.

One of the most extensively studied alternative structures is a type of quadruplex, a four stranded structure, known as a G-quadruplex.[22] The formation of this structure is dependent on a guanine rich sequence and the presence of cations to form intramolecular/intermolecular and parallel/antiparallel structures. The characterisation of G-quadruplex's structure identified that Hoogsteen hydrogen bonds between four guanines gave rise to the formation of G-quartets, which stack on top of each other in a comparable manner to Watson and Crick base pairing.[20,23] Within each G-quartet each of the guanines act as hydrogen bond donors and acceptors.[24] The process of stacking of these G-quartets provides space where cations can localise to stabilise the G-quadruplex (*figure 1.3*).[20,23] These cations stabilise the electronegative core due to close proximities of the O_7s making them more electronically favourable.[25,26] Computational

analysis of the human genome identified enrichment for sequences capable of forming these structures at telomeres and gene promoters suggesting a potential significant biological function for G-quadruplexes.[27,28] Which has been further supported by, visualisation of G-quadruplexs *in vivo*,[29] visualisation of G-quadruplexes in live cells,[30] identification of G-quadruplex binding proteins,[31,32] mediation of protein transcription,[33] and further studies that identify G-quadruplex's as having a significant biological role in 'normal' and pathogenic states.[34,35]
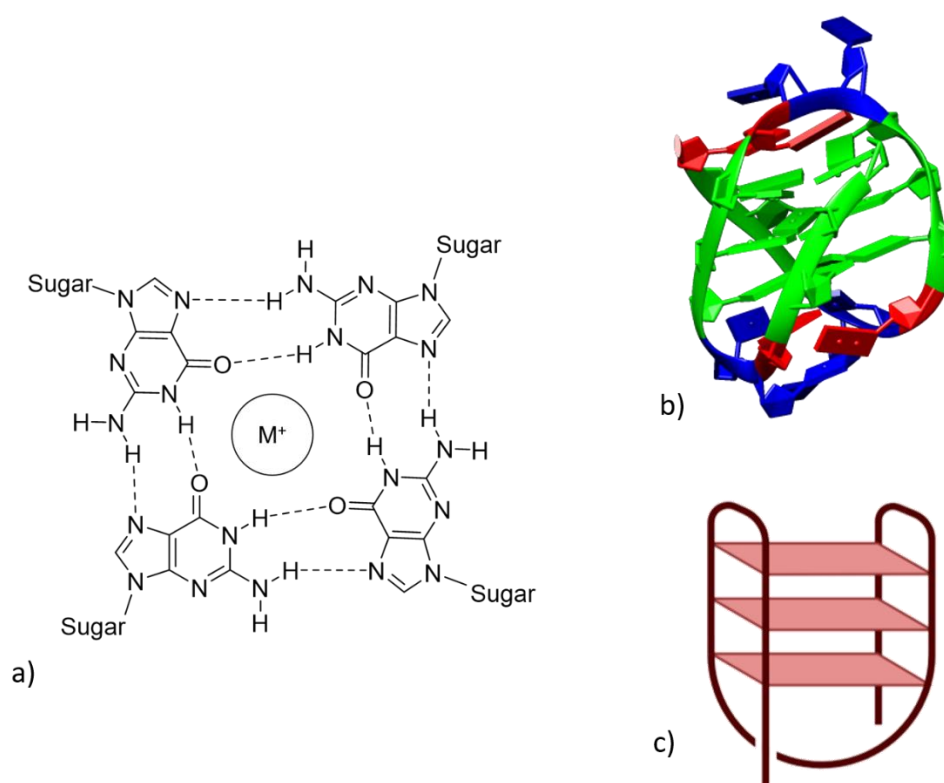
Figure 1.3: G-quadruplex DNA. a) Hoogsteen hydrogen bonding of G-quartets with stabilising cation in the core b) A structure of G-quadruplex DNA (PBD ID:143D) with the colouring scheme: red is adenine, blue is thymine, green is guanine, yellow is cytosine c) Schematic of G-quadruplex DNA created with Biorender.com

1.2 i-Motif DNA

The G-quadruplex has a less extensively studied sister quadruplex that forms on the complementary C-rich strand, known as the i-motif (*figure 1.4*). Since its first description in 1993, by Gehring, Leory and Guéron,[36] the i-motif field has made many advances investigating genome location, required environment, and biological function. Yet much is still left unanswered. One of the reasons for research in the i-motif field lagging behind its sister structure is due to the requirement for hemi-protonated cytosine base pairs to form the structure.[36] This formed the belief they could only form at acidic pH, thus causing great scepticism towards their biological relevance. Therefore, until recently, the majority of i-motif research has focused on nanotechnological applications.[37] In which their rapid reversible folding and unfolding in response to alternating pH made them an excellent platform for logistical application switches[38,39] and pH sensors in cells.[40,41]
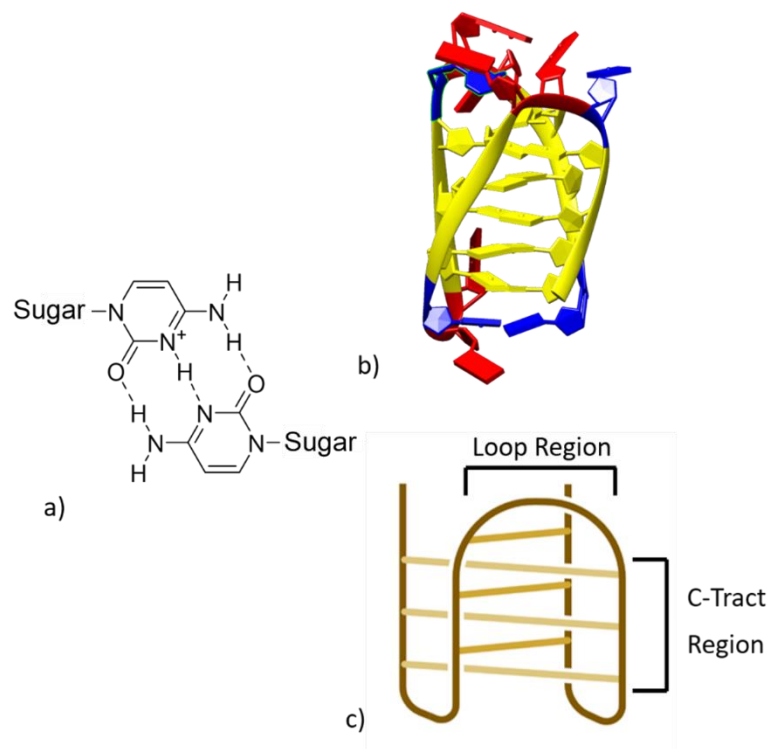
Figure 1.4: i-Motif DNA. a) Hoogsteen hydrogen bonding of cytosine-cytosine⁺ b) A structure of i-motif DNA (PBD ID: 1eln) with the colouring scheme: red is adenine, blue is thymine, green is guanine, yellow is cytosine c) Schematic of i-motif DNA with loop and C-tract regions labels, created with Biorender.com

1.2.1 Structure and Classifications of i-Motif DNA

i-Motif structures vary significantly and have multiple different ways to be classified. The first i-motif identified, by Gehring *et al.,* was an intermolecular, tetrameric structure consisting of two parallel duplexes in an anti-parallel orientation, were the duplexes are formed from cytosine tracts of two independent nucleic acid strands that are stabilised via intercalated C-C⁺ base pairs (*figure 1.5).*[36] Another possible intermolecular i-motif structure is achieved from the intercalation of a dimer consisting of two C-rich hairpin structures, stabilised in same manner as the tetrameric i-motif (*figure 1.5).*[42] Arguably, the most significant structure to be determined in the 1990's was the intramolecular structure, formed from a single strand (monomer) with a minimum of four intercalated cytosine

residues, that can form in natural sequences, and were detected at neutral pH (*figure 1.5*).[43] Determination of this structure provided a compelling argument for i-motifs biological relevance, but was still not widely accepted for some years.
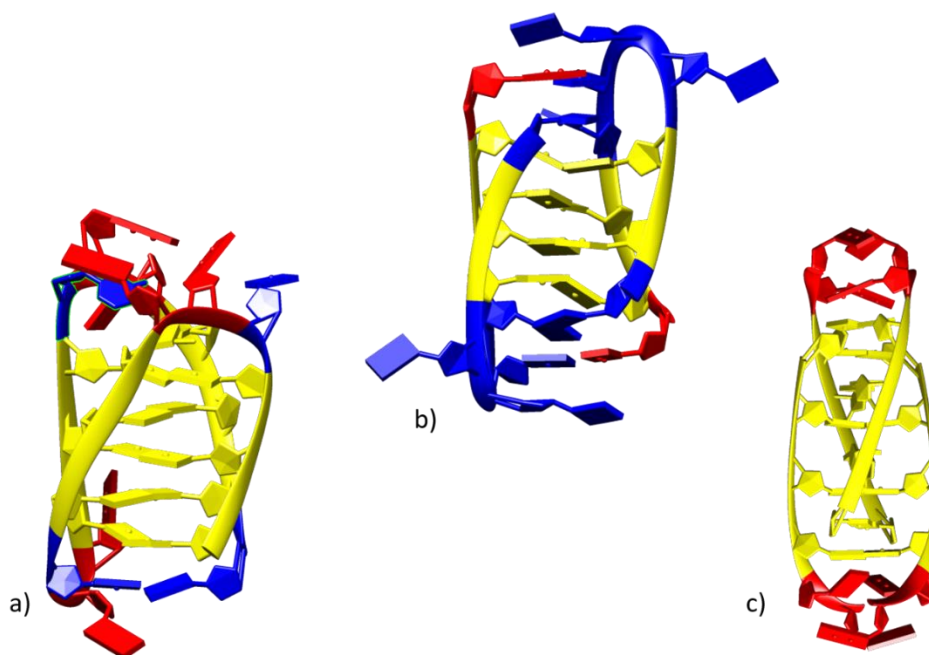
Figure 1.5: i-Motif DNA structures composed of different numbers of subunits. a) Monomeric i-motif (PBD ID: 1ELN) b) Dimeric i-motif (PBD ID: 2MRZ) c) Tetrameric i-motif structure (PBD ID: 1YBL). Colouring scheme: red is adenine, blue is thymine, green is guanine, yellow is cytosine.

An additional classification, described in 1998, was based upon i-motif topology, grouping structures via how intercalation occurs as either R-form or S-form, now more commonly referred to as 3′E or 5′E respectively. If the spatial arrangement of the terminal C-C$^+$ base pair is at the 3′ end or at the 5′ end then they are accordingly denoted as 3′E and 5′E (*figure 1.6*).[44,45] In 2001 a third topology, T-form, was identified where the terminal cytosines are not intercalated, subsequently this topology is less stable than 3′E and 5′E.[46] Notably, the three topologies have varied values of

enthalpy and entropy of formation and thus different conditions stabilise different topologies and thus shift the equilibrium between the topologies that exist for a particular i-motif forming sequence.
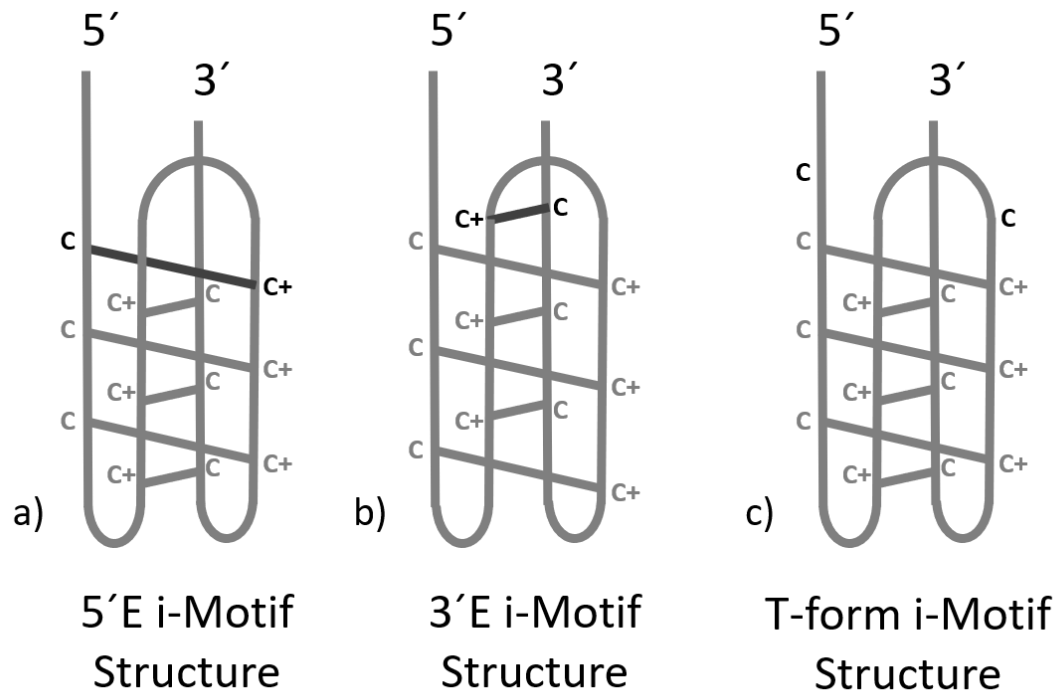
Figure 1.6: Classifications of i-Motifs by topology. The darker grey bonds and C's highlight the areas of interest. a) 3'E (R-form): the terminal C-C⁺ is at the 3' end b) 5'E (S-form): the terminal C-C⁺ is at the 5' end c) T-form: the terminal cytosines are not intercalated.

Intramolecular i-motifs have an added classification based on loop length that was described by Hurley and co-workers.[47] In which they defined class I (short loops) and class II (long loops) i-motifs (*figure 1.7*), as having the following loop sizes respectively: $5'$-(2 :3/4 :2)-$3'$ with four to six C-C⁺ base pairs and loop sizes: $5'$(6/8 :2/5 :6/7)-$3'$. In general, the transitional pH for class II i-motifs are less acidic in comparison to class I, due to enhanced stabilisation which was speculatively achieved by interactions between bases in the loop regions.[45]
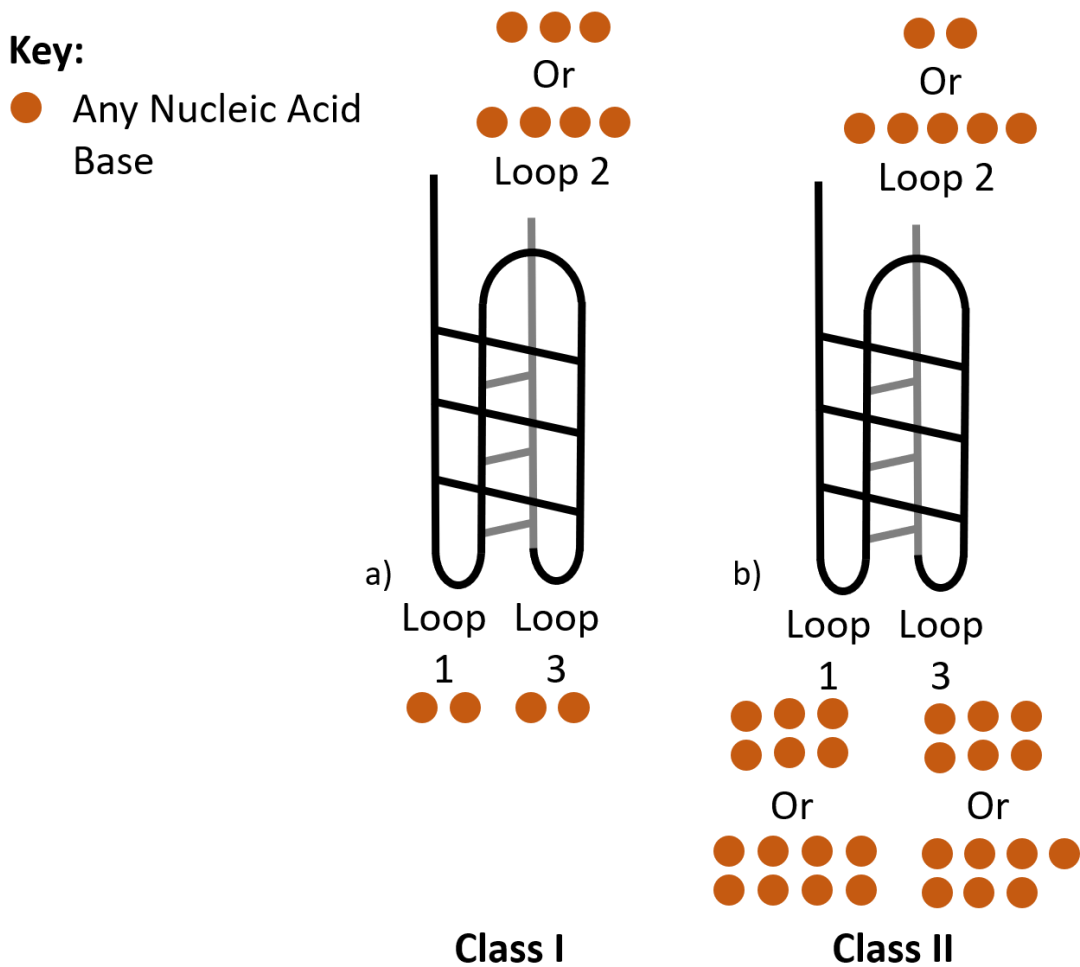
Figure 1.7: Classifications of i-Motif's by loop length, loops are labelled 1 to 3 and orange circles demonstrate any nucleic acid DNA bases where X circles equates to the number of bases in the loop to meet criteria of class I or class II. a) Class I: short loops with the following loop sizes for loops 1, 2, and 3 5'-(2 :3/4 :2)-3' b) Class II: long loops with the following loop sizes for loops 1, 2, and 3 5'-(6/8 :2/5 :6/7)-3'.

1.2.2 Stability of i-Motif DNA

The requirement of protonated cytosines for the formation of an i-motif means that pH is a fundamental factor to its formation. Although, the stability of an i-motif is multifactorial and the following also need to be taken into account: the phosphate backbone, loop length, loop nature, sequence length, c-tract length and environmental conditions such as: molecular crowding, temperature, and ligands (*figure 1.8*).[48]
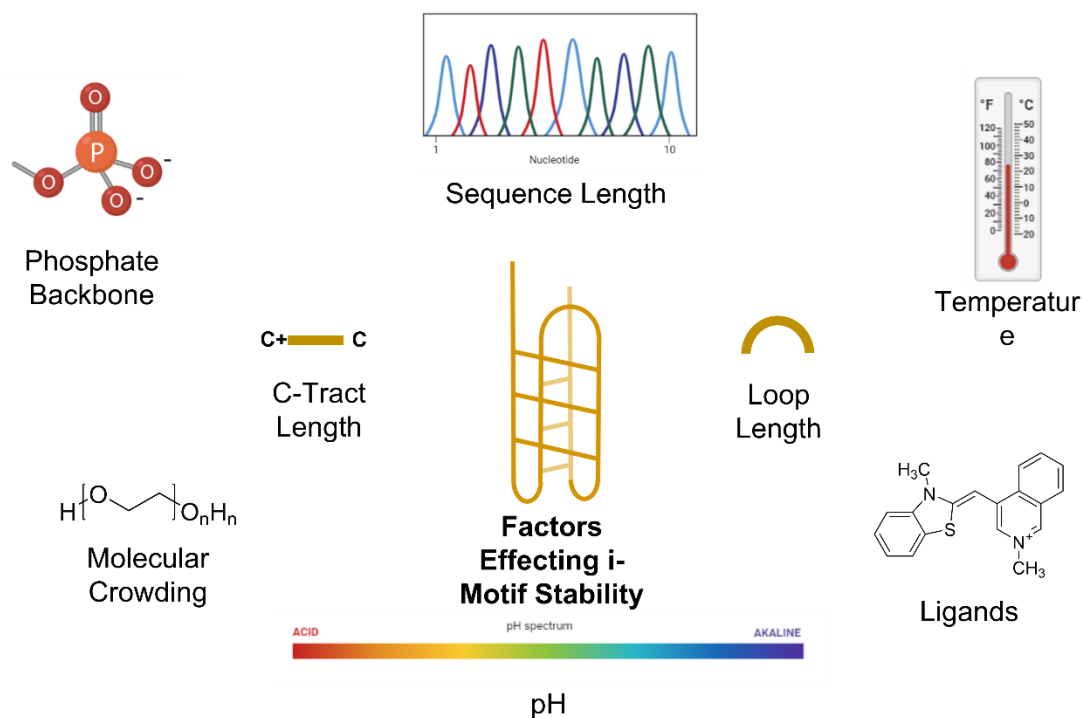
Figure 1.8: Figure summarising the factors that effect i-motif stability. Created with BioRender.com and ChemDraw 18.2.

1.2.2.1 The Phosphate Backbone and i-Motif Stability

To understand i-motif stability the phosphate backbone needs to be considered. The tetrameric i-motif has a 3.15 Å average rise per base pair and an 12-20° rotation per residue, both of which are smaller than observed in B-DNA (3.33 Å and 36.0 °).[49–51] Demonstrating that i-motif DNA has more compact stacking and a shallower right handed twist, forming a structure with two broad and two narrow grooves. In the narrow grooves the interstrand phosphates are as close 5.9 Å in comparison to 7 Å in B-DNA.[52,53] This causes electrostatic repulsion between the charged C-imino groups destabilising the structure. On the other hand, this small distance permits intermolecular CH---O hydrogen bonding between sugars, thus stabilising the i-motif structure. It is important to highlight that the free energy for this bond is 2.6 kJ mol^{-1}, this is significantly smaller than expected for a CH---O bond (7.5 kJ mol^{-1}).[54] Leroy *et al.*

suggested this was due to the H-bonding not being created but replacing H-bonds with surrounding water molecules.[55] Notably, 3′E topology is more stable than 5′E as 3′E has an extra sugar pair and therefore more van der waals interactions than 5′E.[55] Leroy and co-workers suggest the free energy difference this causes is small as both 3′E and 5′E exist. However, Malliavin *et al.* found in their work 5′E was unstable and didn't form suggesting that the free energy difference was significantly larger, due to other interactions, than for the i-motif Leroy investigated. Interestingly, it was originally suggested that these interactions caused the narrowing seen in the i-motif structure.[51] However, in 2001 it was determined that this was energetically untenable. Which formed the theory that the i-motif structure rotates as narrow as possible and repulsions between the strands components inhibit further narrowing, thus the structure is due to a compromise between narrowing and repulsion.[54]

1.2.2.2 Loop Length and i-Motif Stability

Previously, it was shown that the 3′E topology was the most stable conformation,[55] however this is not always the case. For example, it has been demonstrated in the monomeric, intramolecular hTeloC i-motif that the 5′E topology is favoured. Schwalbe's group determined this stabilisation of 5′E was due to a T-T pair between loop one and three.[56] Thus, highlighting the importance of loop interactions on the stability of the i-motif. Furthermore, it had been noticed that class II i-motifs, those with longer loops, generally had greater stability than those with shorter loops.[45] Which also highlighted the importance of loops in i-motif stability. Upon greater investigation by various groups it has been widely demonstrated that longer loops decrease the stability of the i-motif.[57,58] This attribute is likely due to the increased flexibility of longer loops which decreases stabilising interactions such as H-bonding.[57] It has been

demonstrated that bases present in the loops is of greater significance than the loop length itself. The first and last loop have the greatest significance on i-motif stability; having the same number of bases in these lateral loops is optimal for stability and the presence of G/T in these loops adjacent to the C-tract also provides stability from H-bond formation. Stacking within loops can also provide stability but has been deemed to have a lesser stabilising effect than H-bonding. Furthermore, additional C-C$^+$ base pairs also provides i-motif stability. Notably, the middle loop can accommodate a longer loop with minimal disruption to the structure, which could be beneficial for small molecule binding.[57,58] Therefore, the nature of the loops has a much greater significance on i-motif stability than loop length, especially in loops one and three.

1.2.2.3 C-Tract Length and i-Motif Stability

Another important factor for i-motif stability is of course the C-tract length. In general increasing the length increases the melting temperature, brings transitional pH (pH where you have 50% folded i-motif and 50% unfolded) closer to physiological, and increases the stability of the i-motif.[59,60] The work from Burrows group identified that this was only true for up to 30 C-bases, going beyond this caused destabilisation. They also identified that the most stable C-tract lengths followed the rule 4n-1 and that many of the strands where capable of forming i-motif at pH 7.0 at 37 ^{o}C.[60]

1.2.2.3 The Environment and i-Motif Stability

Unsurprisingly, the environment has been shown on multiple occasions to affect i-motif formation and stability. i-Motif formation has been shown to be temperature dependent. With one group identifying that the i-motif formed in the human telomeric region (hTeloC), the naturally occurring sequence, can form at physiological pH if at 4 °C.[61] Lowering the temperature favours i-motif formation potentially due a decreased amount of protons needed for formation.[62] Although, this is not to say that low temperatures are a requirement for i-motif formation near physiological pH.[59] Work within the Waller group interestingly highlighted that hTeloC rapidly forms the less stable, more kinetically favoured 3′E structure following the annealing process (5 minutes at 95 °C, then cooled to room temperature). However, if the sample is kept at 4 °C post-annealing the topology slowly converts to the more stable 5′E structure.[63] This, highlights that temperature needs to be considered when working with i-motifs and the treatment of samples needs to be applied consistently. Interestingly, by mimicking molecular crowding, similar to cellular conditions using molecular crowding agents, typically PEG, an increase in intramolecular i-motif stability has been observed in telomeric sequences and the i-motif forming region in the *MYC* promoter (c-MycC).[64,65] Molecular crowding has been shown to shift the pK_a towards physiological pH and favour quadruplex structures over B-DNA.[66,67]

Furthermore, negative superhelicity has also been shown to promote the formation of secondary DNA structures in the *MYC* promoter region.[68] It was found that c-MycC was capable of folding at neutral pH under negative superhelicity. Sun and Hurley suggested that the quadruplexes absorb torsional stress caused by negative superhelicity and that it is likely they are somehow involved in the transcription process. Another environmental

consideration is ionic strength. It has been demonstrated that increasing the ionic strength destabilises the i-motif,[43] however this destabilisation maxes out upon reaching a particular ionic strength (0-100 mM NaCl and 300 mM NaCl respectively). Furthermore, dehydration has been shown to stabilise i-motifs, due to the release of water molecules.[62,64]

Generally monovalent cations stabilise Watson-Crick and Hoogsteen base pairs by shielding the electrostatic repulsion between the negative charges associated with the phosphate groups of nucleic acids. As evident in the case of c-jun, a proto-oncogene, where different cations favour the formation of different DNA structures via stabilisation of base pair bonding. [69] Watson-Crick base pairs were stabilised by Na^+ and hence duplex formation was favoured, whereas K^+ stabilised Hoogsteen base pairing so G-quadruplex and i-motif structures were observed, and a combination of Mg^+ and Na^+ gave a mixture of duplex and quadruplex structures. Although, when investigating hTeloC in the absence of duplex forming sequences Na^+ and K^+ have a stabilising effect, [70] one could infer from this that Na^+ stabilises duplex DNA to a larger extent than i-motif, although these are different sequences examined under different conditions so they cannot be directly compared and further research would need to be carried out to begin to conclude this. Perhaps more interestingly it was shown that Li^+ ions have a non-stabilising effect on hTeloC.[70] It has been proposed that Li^+ fits between the two $C-C^+$ base pairs in space but cannot adequately fulfil the role of protons to mediate the hydrogen binding between the $C-C^+$ base pairs and instead disrupt it. Due to their larger radi, Na^+ and K^+ provide an adequate positive charge between the paired cytosines and thus stabilise the i-motif structure.[70] Hence, cations affect the stability of i-motifs as well as other DNA structures. Since monovalent cations are ubiquitous in the microenvironment of cells they need to be considered when exploring i-motifs, and suggest that these structures can form at physiological pH, especially as concentrations of Li^+ (which is an i-motif non-stabiliser) are significantly lower than Na^+ and K^+ (which stabilise the i-motif) (μM compared to ~100mM respectively).

Furthermore, some transition metal cations have also been investigated. It has been demonstrated that Hg^{2+}, Ag^+ and Cu^+ stabilise i-motif, the latter two at physiological pH.[71–73] Interestingly it was also shown that Cu^{2+} coverts the i-motif structure into a hairpin, further highlighting the importance of cations on DNA structure.[73] Therefore, it can be suggested that cations have the potential to act as a biological mechanism that control DNA polymorphisms such as i-motif formation.

Ultimately, these studies stress the importance of the sequence itself and the environment on i-motif structure and provide evidence that suggests i-motifs can form *in vivo*.

1.3 Ligand Interactions with i-Motif DNA

Evidence increasingly suggested that i-motif DNA is biologically significant. Therefore, the need for i-motif binding ligands intensified to enable investigation of biological functions and to identify potential i-motif interacting compounds. It needs to be considered when investigating ligands that conditions need to be consistent to compare across sequences as of the multiple factors that affect i-motif stability and the properties of compounds functional groups in varying conditions, for example the charge of groups at different pH values.

Currently, the research output for i-motif ligands lags behind its sister quadruplex. In an attempt to close this gap some groups have investigated known G-quadruplex ligands against i-motif sequences. The following G-quadruplex ligands interactions with i-motifs have been investigated: **Berberine**, **BRACO-19**, **Mitoxantrone**, **Phen-DC3**, **Pyridostatin**, **RHPS4**, and **TmPyP4**.[74] Randazzo *et al.* demonstrated that all of these listed compounds form interactions with hTeloC at pH 4.3 and 5.7 using FID, CD- and UV-melting studies. Although, compounds **BRACO-19**, **Mitoxantrone**, and **Phen-DC3** which are known to stabilise

G-quadruplexes, were found to destabilise hTeloC at both pH's using CD- and UV-melting experiments. Whereas, **Berberine**, **Pyridostatin**, and **RHPS4** had no effect on i-motif stability.[74] Similar results, by the Waller group, were seen investigating these compounds against two other i-motif sequences, known as ATXN2L and DAP, at pH 7.0 which generally caused destabilisation in CD- and UV-melting.[75] A greater extent of destabilisation was seen in respect to DAP than ATXN2L. Notably, different techniques within this study did produce contradicting data for **Mitoxantrone**. For example, CD-melting showed **Mitoxantrone** destabilised ATXN2L whereas UV-melting demonstrated it had a stabilising effect. Highlighting how different biophysical techniques can have conflicting results, which is one difficulty of discovering specific ligands for different DNA structures.

Other studies have investigated single-walled carbon nanotubes, SWCNTs, and demonstrated that they inhibit B-DNA formation and induce i-motif formation in hTeloC at physiological and even basic pH.[76] Further research has shown that SWCNTs can induce i-motif formation under crowded conditions which mimic cellular conditions.[77] It has even been demonstrated that telomerase activity can be inhibited by stabilising i-motif structure using SWCNTs and could be a powerful therapeutic route for cancer.[78] However, their use as therapeutics is controversial due to their toxicity to humans, including DNA damage and oxidative stress, and their poor metabolism leading to concerns about their accumulation along the food chain.[79,80]

A variety of other i-motif binding ligands have been reported including **bis-acridine**,[81] acridone derivatives,[82] **Crystal violet**,[83] **Coumarin 343**,[84] and **Thiazole orange**[85] (*figure 1.9*). Some of which are capable of inhibiting B-DNA formation and stabilising quadruplexes, others show selectivity for i-motif. An interesting study was on the Bcl2 i-motif where it was discovered that **IMC-48**, a cholestane derivative with a positively charged piperidine attached to it, binds and stabilises the i-motif structure

whereas **IMC-76**, a cholestane derivative, converts the i-motif to a hairpin structure. Which has led to interesting gene expression studies.[86] Additionally, a fluorescent indicator assay has allowed large screens for i-motif ligands and has identified the following known drugs interact with i-motif DNA: **Tilorone**, **Tobramycin**, **Phenazopyridine**, **Amodiaquine**, **Harmalol**, **Quinalizarin**, and **Minocycline tyrothricin**.[85] More recently, an *in situ* screen has been described that has identified specific probes for the BCL2 and *MYC* i-motif's: **3bm** and **3be** respectively which are triazole compounds.[6] A further two *MYC* i-motif probes have been identified, known as **B19** and **a9**, where **B19** is an acridone derivative and **a9** is a bisacridine derivative.[82,87]
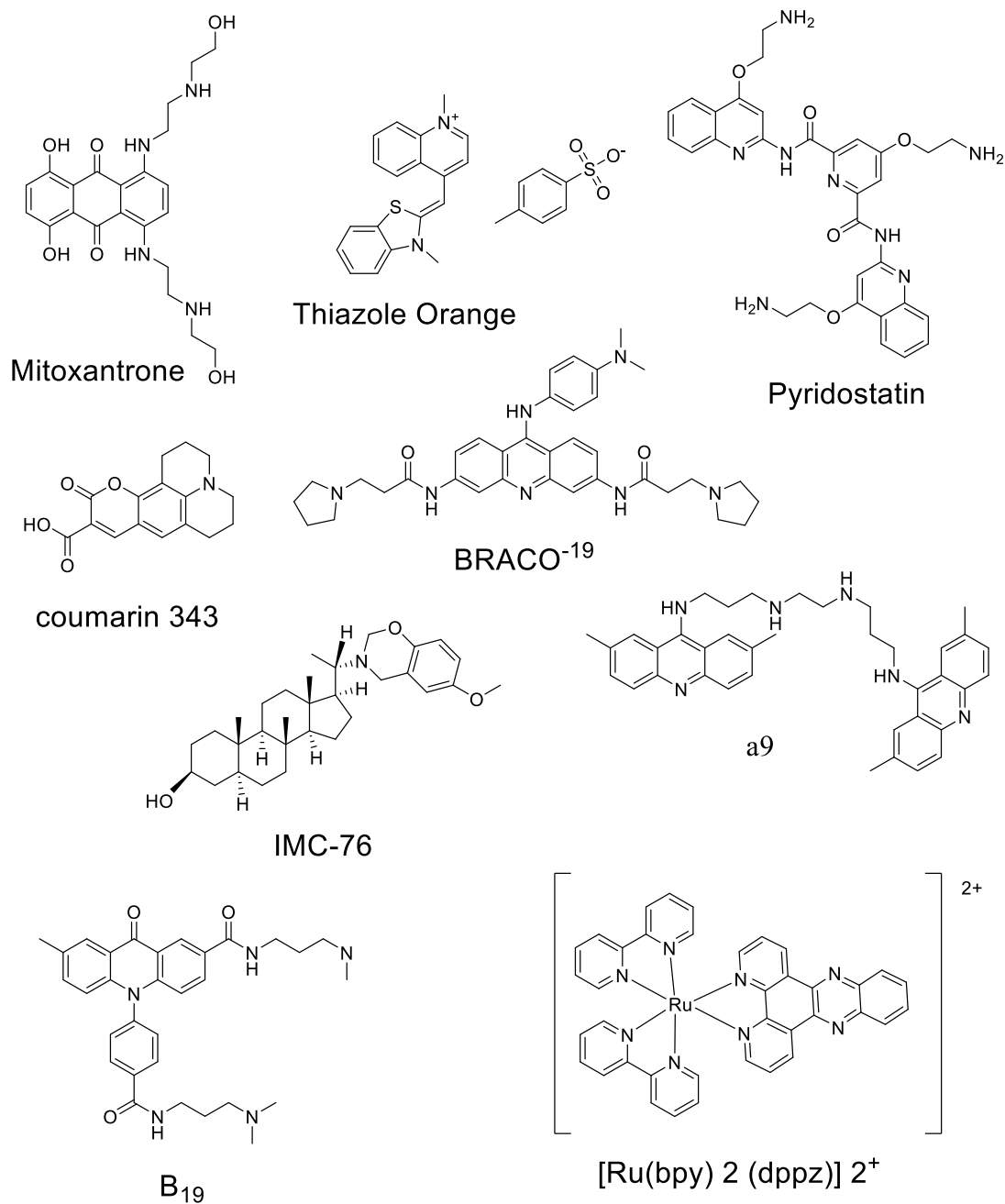
Figure 1.9: Figure showcasing some examples of different i-Motif ligands identified.

Furthermore, complexes containing central metal atoms are frequently studied as B-DNA binders, particularly ruthenium complexes due to their stability, affinity for B-DNA, and luminescence properties. Leading to an interest in their ability to bind i-motif DNA. Brazier's group identified that **[Ru(phen)₂(dppz)]²⁺** stabilises long-looped i-motif sequences. Highlighting i-motifs appeal as drug targets. The Waller group have also

explored ruthenium complexes and have identified further i-motif binders and even probes with a preference for i-motif DNA using emission lifetime measurements.[88]

Research into i-motif ligands has gained considerable interest over the recent years and the field continues to grow as the i-motif's biological significance gains increasing evidence and the prospect of use as therapeutics is explored. Yet, i-motif specific ligands are struggling to be unearthed and commonalities between known ligands is not clear cut (*figure 1.9*). Therefore, the hunt for i-motif specific ligands is still an essential area of research.

1.4 Biological Significance of i-Motif DNA

Initially researchers focused on investigating the biological function of the complementary strands which formed the sister quadruplex, known as the G-quadruplex. Although, i-motifs were not forgotten and the first review was published in 2000 which stated that interest was accumulating to investigate their biological significance.[89] Since, this area has continued to gain momentum and acquired increasing experimental support to prove their existence *in vivo* and determine their biological function (*figure 1.10*).
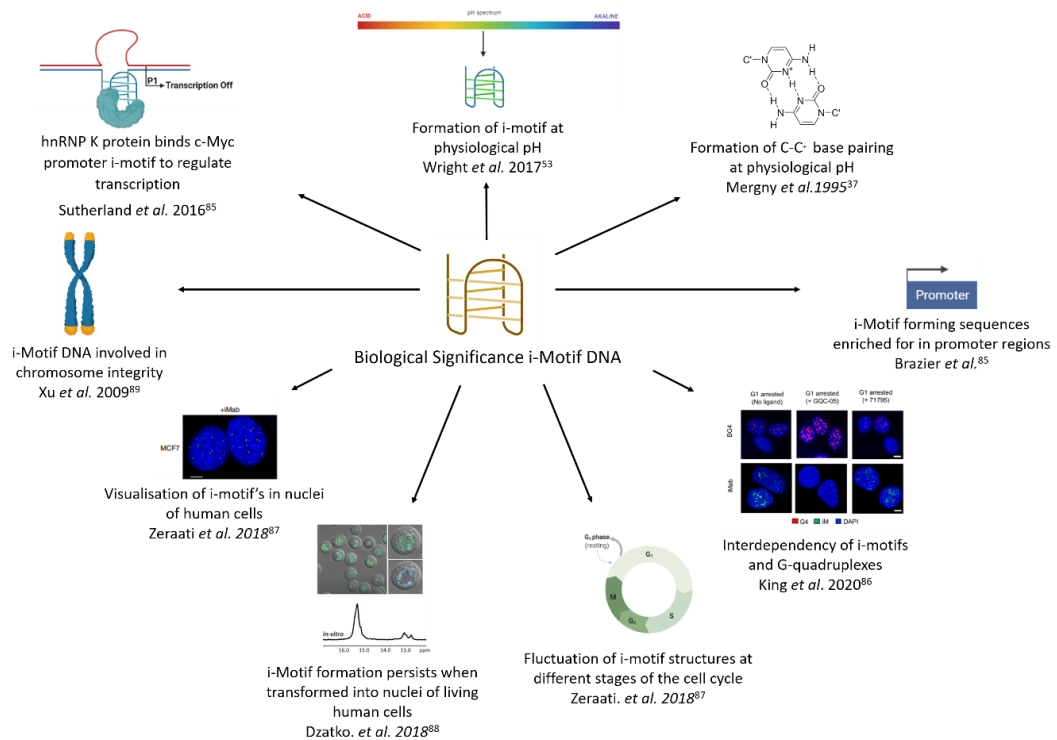
Figure 1.10: Figure summarising the biological significance of i-motif DNA. Created with BioRender.com and ChemDraw 18.2.

1.4.1 Formation of i-Motifs at Physiological pH

A major contributing factor to the scepticism of a natural biological function of i-motifs is that they were observed to form readily at acidic pH and their requirement for protonation suggested they would be unstable in a cellular environment at physiological pH. Although, there was evidence as early as 1995, just two years after discovery of the i-motif, that i-motifs could form at physiological pH by demonstrating that C-C$^+$ base pairing is detectable at physiological pH.[43]

Although, the stability was shown to be significantly higher at pH 5.0 and it was not investigated if the i-motif structure was preferred over B-form. But, this was significant in showing that i-motifs are capable of forming at physiological pH and that they could have a biological function. By 2010 there were minimal published studies showing evidence of i-motif

formation near physiological pH,[61] under molecular crowding conditions,[77] via ligand binding,[76] and at low temperatures.[61] The accumulating evidence clearly showed that i-motifs were capable of forming near physiological pH and identified that the environment played a key role in their stability and consequently ability to form. At this point there was a compelling case that i-motifs could form *in vivo,* suggesting a biological function. Yet, their biological significance was still argued and a larger volume of published research providing evidence of i-motif formation at physiological pH was required. The next few years saw an increase of research investigating i-motifs stability and their ability to form under physiological conditions: natural genomic sequences,[59,90] in the presence of ligands,[48,91] in the presence of cations,[72,92] and molecular crowding conditions.[67,92] Each year saw increasing evidence of *in vitro* evidence that i-motifs can form naturally under physiological conditions and serve a biological function. However compelling the building evidence was the field lacked *in vivo* evidence and the ability to investigate these structures in cells until 2018.

1.4.2 Formation of i-Motifs *in vivo*

2018 saw two crucial papers published that provided the first evidence of i-motif formation *in vivo*. The first was presented by Trantirek's group that demonstrated i-motif formation persisted following transformation into the nuclei of living human cells.[93] Arguably, the most convincing evidence for a biological function of i-motifs was by Zeraati *et al.* where i-motifs were visualised in the nuclei of human cells.[94] This study utilised the powerful molecular biology technique phage display to identify the antibody fragment named iMab, in a similar manner to how BG4, the G-quadruplex binding antibody, was established in 2013.[29] The visualisation of i-motifs in cells via a fluorescently tagged antibody probe allows further investigations into the biological significance of i-motif DNA. Zeraati *et al.*

and colleagues began to address this by investigating at which point i-motifs formed during the cell cycle, as was previously done with G-quadruplexes.[95] Interestingly, it was shown that the frequency of i-motifs is highest towards the end of G_1 (Gap 1) phase, which is responsible for preparing the cell for division metabolically by synthesising mRNA and proteins, and thus supports previous evidence that i-motifs are involved in transcription and cellular growth. However, G-quadruplexes highest formation frequency was during the S (synthesis) phase which is when DNA replicates.[29,95] Notably, there were G-quadruplexes observed at the end of G_1 and i-motifs in S phase. This suggests that these quadruplex structures can form in similar conditions but what is optimal for one is not for the other, supporting the theory that they are mutually exclusive.[96] The ability to visualise i-motifs in cells opens up exciting avenues for researchers in the field and will be an essential tool to further research investigating biological function. Which has led to a study by Smith *et al.* in 2020 which demonstrated fluctuating levels of G-quadruplexes and i-motif structures at different cell cycle stages agreeing with that seen in Zeraati *et al's* work.[97] Interestingly, the Smith *et al.* study used compounds to stabilise G-quadruplexes in human cells and observed destabilisation of the i-motif and vice versa. Highlighting that there is interdependency between the two structures and that the interplay of the structures needs to be considered when investigating them as gene regulatory switches. Thus, the field will benefit from the design of multiple probes which will allow appropriate probing properties to be chosen for specific needs to explore these structures as gene regulatory switches. Ideally designing probes that could allow discrimination not just between i-motifs and G-quadruplexes but go a step further and be capable of discriminating between different i-motif sequences and different G-quadruplexes.

1.4.3 Telomeric Regions and DNA Secondary Structures

A fundamental of chromosome integrity is protection of the ends by ribonucleoprotein structures known as telomeres. Telomeres act as a cap on the ends of chromosomes to avoid fusions and degradation.[98] They are characterised by GC-rich repetitive sequences; interestingly these regions can form secondary DNA structures such as G-quadruplexes and i-motifs.[99] Notably, during replication the C-rich strand and G-rich strand are respectively replicated by leading and lagging strand synthesis. Consequently the 3' strand has a 50-300nt G-overhang protecting the chromosome from degradation during cell replication and is known to form consecutive G-quadruplexes.[100,101] Each round of replication causes telomere shortening, thus the cell employs the enzyme telomerase to lengthen them as a control of chromosomal protection. As cells age the level of telomerase decreases, causing telomere shortening, consequently cells enter replicative senescence. However, many cancerous cells reactivate telomerase permitting the continuing lengthening of telomeres which subsequently delays cell senescence and enables immortality.[102]

Initially investigations exploring telomeres focused on the G-rich sequence as the discovery of the G-overhang was unique. The identification that these G-rich sequences formed G-quadruplexes added another layer of complexity to the regulation of telomeres and subsequently their effect became a research focus. Which produced compelling evidence that G-quadruplex formation uncaps telomeres and initiates the DNA damage-response mediated cell death. Consequently, multiple G-quadruplex ligands that inhibit telomerase are being investigated as drugs.[103] However, the formation of the G-quadruplex does not produce a full mechanism for telomere control. Another structure that adds to the complicity of this control mechanism is the tandem repeat human telomeric forming sequence denoted as hTeloC, with the sequence 5'd(CCCTAA)$_n$-3', which can form an i-motif structure. Although, a clear

biological mechanism of hTeloC is yet to be established there is evidence that it inhibits telomerase and initiates telomere uncapping which induces the DNA damage-response mediated cell death.[78] Consequently, hTeloC is frequently used for ligand studies.[48,85] Evidently, the i-motif and G-quadruplex appear to function in a similar manner in the case of telomeres and are both an interesting prospect as therapeutic targets.

1.4.4 i-Motif Binding Proteins

Poly(C)-binding proteins (PCBP) and their interactions with C-rich sequences are well documented, and so its unsurprising they have become an area of interest in the search for i-motif binding proteins.[104] Within the family of PCBP's are αCP1-4 and αCP1-KL which have been shown to bind to C-rich ssDNA and hnRNP K that binds to the C-rich sequence in the *MYC* promoter.[4,104,105] Although, it is not clear if these proteins bind to the i-motif or unfolded sequence. However, it has been demonstrated *in vitro* that hnRNP K and the splicing factor ASF/SF2 bind to the hTeloC i-motif.[106] hnRNP K has also been shown to bind to the *MYC* i-motif, which is explored in more detail in *Chapter 1.3.4*.[4] Additionally, the hnRNP LL protein binds to the lateral loops of the Bcl2 i-motif causing destabilisation of the i-motif and activating transcription.[107] Collectively these studies make a promising case that i-motifs have a biological function and are involved in transcriptional regulation. Nonetheless, it is clear that this exciting area needs further research.

1.4.5 Promoter Regions and DNA Secondary Structures

It is well known that gene transcription is highly regulated, one mechanism of control is by regulatory elements such as promoters.[108] Promoters are specific regions of DNA upstream of transcription start sites

and are recognised by transcription factors whose binding recruits RNA polymerase II to initiate transcription of target genes downstream.[108] Interestingly, i-motif forming sequences are enriched for in promoter regions suggesting an involvement in the regulation of gene transcription.[47,109]

One of the most widely studied promoter region capable of forming an i-motif is the B-cell lymphoma gene-2 (*Bcl2*) promoter region.[86] The *Bcl2* promoter is a known regulator of apoptosis which is frequently deregulated in cancers and is a common target of cancer therapies.[110] It is known that *Bcl2* interacts with a range of proteins which led to Hurley's group investigating the binding of small molecules to the C-rich sequence of the *Bcl2* promoter. They identified that the ligands preferentially bound to folded or unfolded i-motif forming sequences and proposed that proteins would also show structural selectively when binding to the sequence.[86] Specifically that **IMC-48** is selective for the i-motif structure, its binding constricts the lateral loops which allows hnRNP LL to bind which causes the structure to unfold and transcription is activated.[86,107] This proposes that formation of an i-motif structure could function as a control of gene expression. Interestingly, i-motifs regulatory role of transcription is further supported by studies investigating the *MYC* promoter, which curiously supresses *Bcl2* from activating apoptosis stimulating imortaility.[111,112]

Another promoter region of interest is the insulin-linked polymorphic region (*ILPR*), which is upstream of the human insulin gene.[42] *ILPR* has a C-rich strand that has been shown to form i-motif at acidic pH and a partial i-motif structure at pH 7.[42] Notably, *ILPR*'s length is polymorphic and it has been demonstrated an increase in sequence length causes an increase in stability. It has been suggested that the ILPR i-motif, similarly to the ILPR G-quadruplex, causes slippage during replication,[42,113] which is a form of mutation that causes expansion or contraction of the nucleotide sequence and thus can be linked to the length of the *ILPR* sequence.

Furthermore, the mechanical stabilities of both the i-motif and G-quadruplex meet the minimum required to stall RNA polymerases.[114] Notably, experimental data has shown that these two structures do not form at the same time,[114] although the conditions of this investigation were not replicating the *in vivo* environment so this may not be true for in cells. Nevertheless, one can postulate that the i-motif is functioning as a control mechanism for transcription of the human insulin gene, and thus *ILPR* could be a therapeutic target for insulin-dependent diabetes mellitus.

Some other mentionable promoters that form i-motifs include *c-Myc,*[115] *VEGF,*[116] *c-Kit,*[117] *Rb,*[118] *RET,*[119] and *KRAS.*[120] All of these promoters are associated with hallmarks of cancer and lend support to a biological function and cancer treatment targets. *MYC* is one of the best studied promoter i-motifs and will be explored extensively in *Chapter 1.5*.

1.5 The *MYC* Proto-Oncogene

MYC is a proto-oncogene that functions as a multifunctional transcription factor that targets an array of genes which coordinate a broad range of essential cellular functions, such as cell cycle progression regulation, proliferation, differentiation, and apoptosis (*figure 1.11*).[121,122] The biological significance of *MYC* is further highlighted by tight regulation by a variety of complex mechanisms Therefore, it is unsurprising that dysregulation of *MYC* is frequently recognised as a hallmark of cancer.
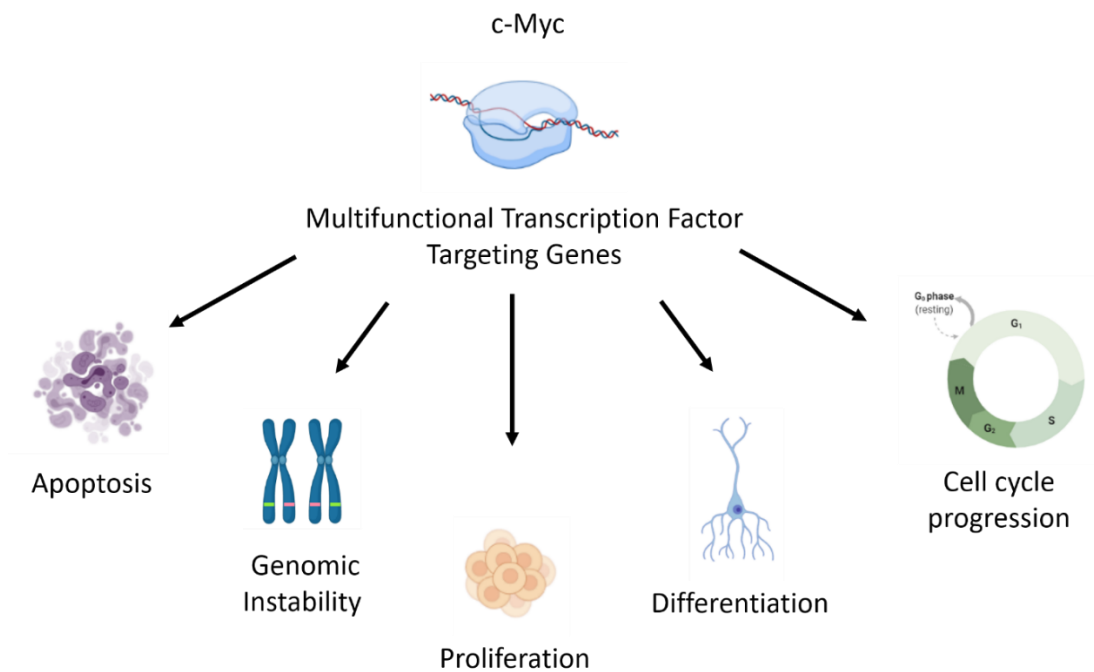
Figure 1.11: An example of cellular functions that the MYC gene regulates. Created with Biorender.com.

1.5.1 Structure and Functions of c-Myc, the Protein

MYC is a family of regulator genes, that are known oncogenes, which include three well defined members: *L-Myc*, *N-Myc* and *c-Myc*. The overall structure and organisation of these genes and their transcripts are very similar. They each contain 3 exons and the proteins encoded by these genes have significant homology.[123] *MYC* family proteins can be organised into an unstructured N-terminal transactivation domain (NTD), a central region, and a C-terminal domain (CTD) (*figure 1.12*).[124]
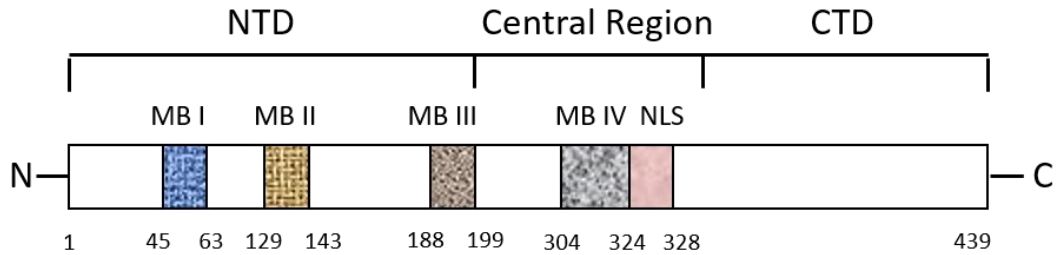
Figure 1.12: Structure of MYC family proteins. Where NTD is the N-terminal transaction domain, CTD is the C-terminal domain, NLS is the nuclear localisation sequence, and there are four MYC boxes (MB).

The NTD regulates transcription and transformation and this domain contains three MYC boxes (MB): MBI, MBII, and MB III (*figure 1.12*).[124] MB II is essential for cell cycle progression, apoptosis, and transformation, and although MB I reduces c-Myc activity it is not as significant as MB II in regards to its effect on c-Myc function.[125] MB III regulates transcriptional suppression and is a negative regulator of apoptosis.[126] Further downstream in the central region of the protein is the final *MYC* box, MB IV (*figure 1.11*). MB IV regulates c-Myc induced apoptosis, transformation, DNA binding, and G_2 arrest.[127] Notably, the central region is also important as it contains the nuclear localisation sequence (NLS) which tags c-Myc to be imported into the nucleus. The final domain is the CTD which is essential for protein and DNA interactions, it contains a basic region/helix-loop-helix/leucine zipper (bHLHLZ) which is mostly unstructured. This region forms a heterodimer with MAX, a protein that binds DNA, via its bHLHLZ site. The c-Myc and MAX heterodimer binds to specific E-boxes with the consensus sequence 5´-CACGTG-3´ which is typically found in c-Myc activated genes.[128,129] Research suggested for many years that MAX was essential for c-Myc's biological function, however in more recent times it has been demonstrated that c-Myc retains significant biological activity without MAX.[130] c-Myc is involved in many different biological functions and is essential to many processes some of which are briefly explored below.

MYC transcription is an essential early growth response gene for cells to enter the cell cycle.[131,132] c-Myc levels have been demonstrated throughout the cell cycle at differing levels (*figure 1.13*). Quiescent cells do not produce detectable amounts of c-Myc but upon entering the cell cycle c-Myc production rapidly increases (G_0/G_1 transition), following the surge c-Myc production falls to a rate that persists throughout the cell cycle.[2] Evidence suggest that c-Myc regulates cell cycle progression, and is required for G_0-S phase progression, via an array of genes including those involved in cell cycle checkpoints.[131,133,134] Multiple groups have demonstrated that c-Myc downregulation is essential for exiting the cell cycle and initiating differentiation.[132,135,136] Further research has also demonstrated that c-Myc has initiates cell differentiation by upregulating miRNAs that target differentiation genes.[137] Another process linked to c-Myc driving the cell cycle but also distinctly regulated is proliferation. c-Myc increases proliferation and cell size by providing the cell with the resources to grow and increasing rates of metabolism via a variety of targets.[138–141] Furthermore, c-Myc regulation has been shown as essential for genomic integrity and when this regulation is lost the gene is upregulated and genetic instability is observed. c-Myc appears to be one system that regulates cell cycle checkpoints this ensures that cells with damaged DNA cannot continue to divide.[142,143] c-Myc is one of the many agents that regulate apoptosis, in normal cells when c-Myc levels are elevated cellular feedback relays this information triggering response mechanisms that lead to apoptosis. c-Myc triggered apoptosis is related to cell type and physiological status and thus acts as a regulation to ensure the cell is in the correct environment.[144–146] It has even been suggested that apoptosis control is c-Myc's default function.[147] Research demonstrates c-Myc directly regulates angiogenesis, c-Myc expression is a requirement to coordinate expression of angiogenic regulators. Thus, c-Myc acts as a master regulator of angiogenesis.[148]
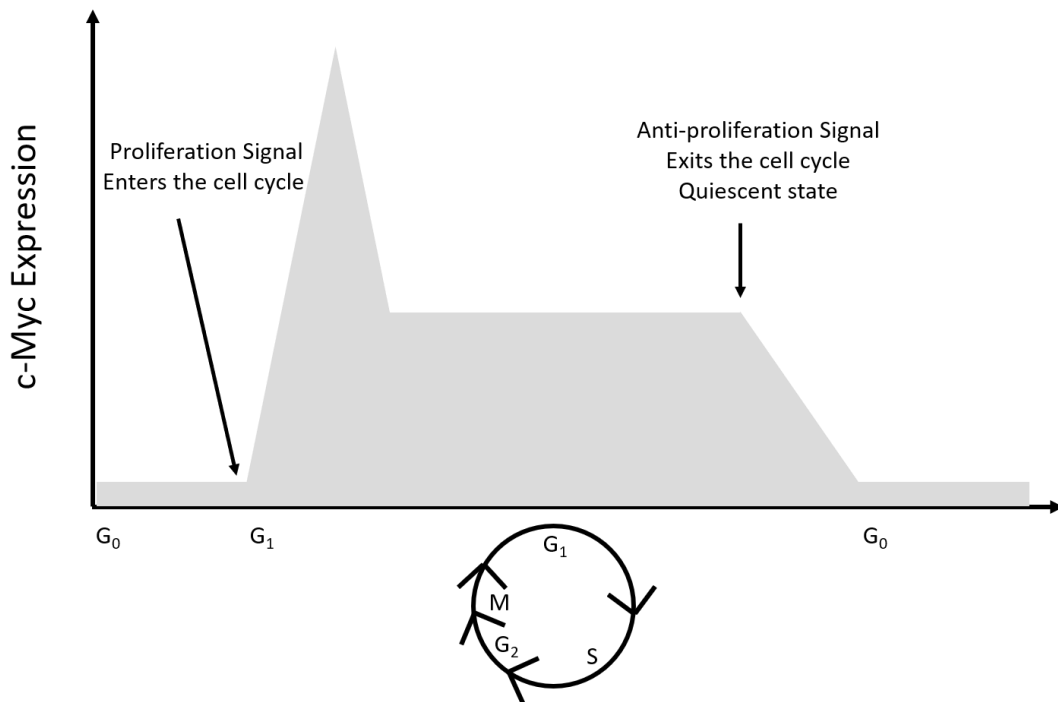
Figure 1.13: Schematic of c-Myc expression levels throughout the cell cycle. Adapted from Alves et al. (2008, Fig1)[1] and Lemaitre et al. (1996, Fig 8).[2]

Evidently, c-Myc is involved in multiple cellular processes and its regulation is essential for many cellular processes. To summarise c-Myc is a multifunctional transcription factor that regulates various cellular processes including, cell cycle progression,[149] cell proliferation,[150] cell differentiation,[131] genomic integrity,[151] apoptosis,[146,147] angiogenesis,[131] cell growth and metabolism[152] all of which can explored in much more detail in the referenced reviews (*figure 1.10*).

Even with the extensive research into c-Myc function the exact mechanisms of many of these regulatory pathways are still being explored. This exploration is essential to understanding normal cell functions and unsurprisingly how dysregulation leads to disastrous effects in the cell causing cancer.

1.5.2 c-Myc, the Protein, and the Hallmarks of Cancer

It's no surprise that c-Myc is tightly regulated in normal cells due to its essential central role in normal cell homeostasis and disruption of this pivotal cellular control has profound biological effects (*figure 1.14*).[153] Evident by >70% of cancers have some form of c-Myc dysregulation and this is associated with poor prognosis and low survival rates. c-Myc has been shown to not only play a role in tumor initiation but also in tumor maintenance showcasing the importance of its dysregulation in cancer.
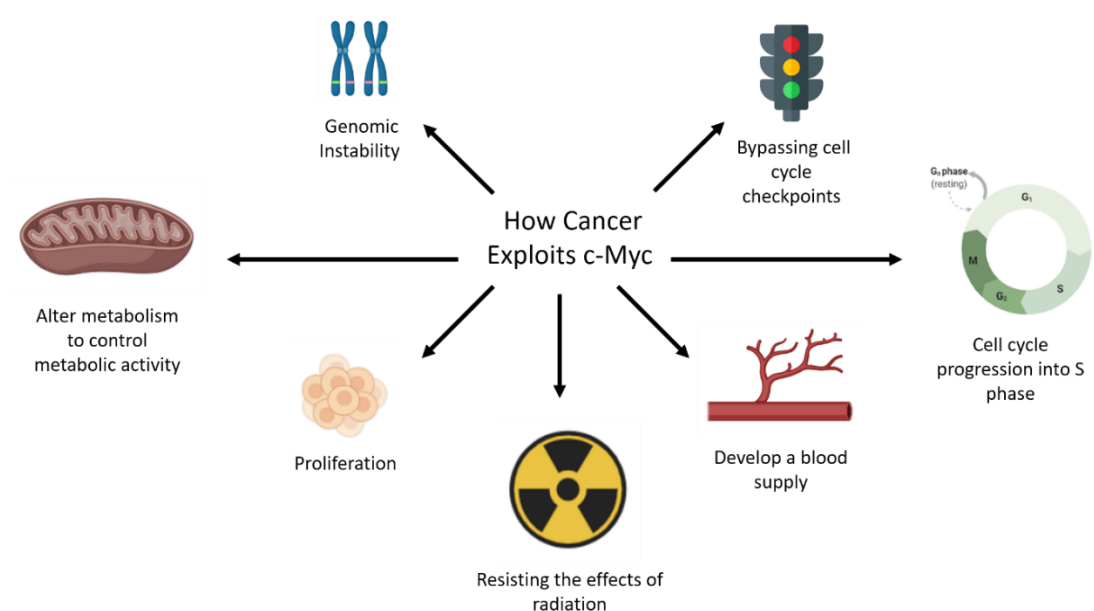
Figure 1.14: A summary of how cancer exploits c-Myc. Created with Biorender.com and freepik.com.

Genomic instability has been shown to be linked to overexpression of c-Myc by different pathways.[143] One way this happens is by cells bypassing check points in the cell cycle allowing damaged DNA to be passed onto daughter cells such as, gene amplification[142] and rearrangement.[143]

Much of c-Myc research has demonstrated a pivotal role in proliferation and apoptosis not only in normal cells but in cancer as well.[131] The

characteristic proliferation is attributed to promoting cells to advance from G_0 to S phase of the cell cycle.[133,134]

Multiple studies have documented that cells overexpressing c-Myc are metabolically addicted and c-Myc can manipulate multiple pathways that can feed this addiction.[154,155] It is essential that cancer cells can control metabolic activity to ensure a steady supply of building blocks and to assemble these blocks into macromolecules so that the tumour can continue to grow and multiply. c-Myc can alter metabolism via many routes including glutaminolysis, glycolysis and lipid synthesis.[156]

Vasculogenesis and angiogenesis are two processes that develop new blood cells, vasculogenesis forms entirely new blood vessels whereas angiogenesis forms from pre-formed vessels via sprouting and splitting.[157] These processes are essential for cancer growth and metastasis, supplying oxygen, nutrients and growth factors via ample blood supply.[157] c-Myc acts as a master regulator of vasculogenic and angiogenic development during tumor progression and loss of c-Myc impairs these processes in tumours.[148,158,159]

Finally, it has been shown that c-Myc affects resisting radiation-induced apoptosis, protection from radiation-induced DNA damages and even promote radiation-induced DNA repair.[160,161] Consequently, this contributes to the low survival rates correlated with dysregulation of c-Myc.[161]

The key point as demonstrated in *figure 1.14* is that c-Myc has profound biological effects in cancer. Evidently many cancers exploit c-Myc to support abnormal growth. Making c-Myc a highly valuable drug target for cancer treatments, which has led to decades of c-Myc research. One would have assumed c-Myc, being an oncogene, would be an ideal drug target because generally inhibiting excessive activity is an easier approach then restoring lost activity, meaning oncogenes are normally favoured drug targets. However, this area has been extremely challenging. One reason for this is because of the spectrum of cellular functions regulated by c-Myc

inactivation can negatively impact physiological functions and cause undesirable side effects in normal cells.[162] Although, there is evidence that these effects may be tolerable. Another blockade is the "undruggable" protein structure. The c-Myc protein lacks effective binding pockets due to the protein being highly disordered and functions via protein-protein or protein-DNA interactions which are difficult to disrupt with small molecules. Finally, much of the dysregulation seen in cancers is not due to the *c-Myc* gene but its upstream controls. Highlighting that the promoter region of *c-Myc* could be a powerful cancer therapeutic target.

1.5.3 Regulation of *MYC* Promoter Structure and Activity

MYC regulation needs to be stringent as it's a pivotal regulator of multiple cellular processes. One mechanism in which regulation is achieved is via four promoter regions: P_0, P_1, P_2, and P_3 *(figure 1.15)*.[163] Interestingly, P_1 and P_2 account for ~25% and ~75% of *MYC* transcription and thus have clear biological significance. Whereas, little is understood about the role of the P_0 and P_3 promoter regions. Another important regulatory component is the multiple cis-elements. Interestingly, during transcription RNA polymerase induces negative superhelicity which causes some cis elements to form non-canonical structures.[164,165] It is suggested their formation reduces the stress on the DNA structure and have proposed roles effecting transcription. For example, the far upstream element (FUSE) under torsional stress forms a single-stranded formation which allows FUSE binding protein (FBP) to bind and increase transcription. FBP-interacting repressor (FIR) can interact with FBP and FUSE causing a conformational change that reduces *MYC* expression to basal levels.[166–168] Thus, highlighting the importance of DNA structures in the promoter region of *MYC* in regulating it's transcriptional activity.
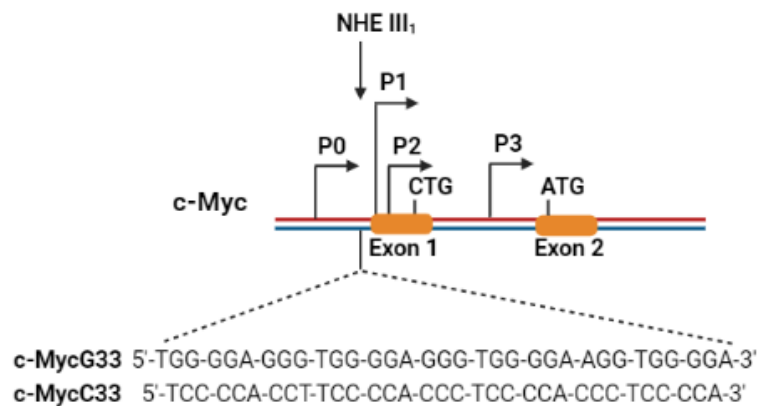
Figure 1.15: Promoter structure of the MYC gene and location of the nuclease hypersensitivity element (NHE). There are four promoter regions labelled P0-P3. The location of the nuclease hypersensitivity element (NHE) III$_1$ is shown by an arrow. The G-rich and C-rich strands. Created with Biorender.com.

In addition, the promoter region contains seven nuclease hypersensitive elements (NHEs). Arguably, the most crucial and complex is NHE III$_1$ (*figure 1.15*), regulating ~90% *MYC* transcription with multiple known transcription factors: hnRNP K, CNBP, Sp1, Sp3, MAZ, hnRNP A1, MAZi, TH-Zif-1, NM23-H2, and NSEP-1.[1] NHE III$_1$ is composed of C-rich coding strand capable of forming an i-motif structure and a G-rich noncoding strand that can fold into a G-quadruplex and it is known that transcription factors favour particular structures acting as additional regulatory element (*figure 1.15*).[169]

1.5.4 NHE III$_1$ Regulation by Nonconical DNA Structures

Leven suggested the NHE III$_1$ region may be the most complex of *c-Myc*, although is this unsurprising for a region that controls such a large percentage of *MYC* transcription.[163] If one thing is clear it is that such an

important transcriptional regulator requires multiple mechanisms to control transcription and ensure normal functioning. One mechanism of control gaining increasing amounts of interest and research, is the formation of the secondary DNA structures the G-quadruplex and i-motif.

Hurley and Yang have demonstrated that the NHE III$_1$ region can form two different G-quadruplex structures, importantly their mutational studies investigating effects on promoter activity via a luciferase assay demonstrated that only the intramolecular parallel G-quadruplex appears biologically relevant, which will be referred to as c-MycG from here on. Significantly, this study also demonstrated that destabilising c-MycG led to an increase of transcriptional activity. Suggesting that the formation of c-MycG acts as a repressor to *MYC* transcription and so is an interesting anti-cancer target.[170,171] Therefore, one could reasonably suggest that c-MycG stabilisers would supress *MYC* transcription. A well-known G-quadruplex stabiliser, TMPyP4, has been shown to supress *MYC* transcriptional activity.[171,172] This further establishes that the G-quadruplex structure needs to be unfolded for *MYC* to become transcriptionally active.

One proposed model of how this system works is that the unwinding of the double stranded DNA during the initiation phase of transcription causes negative superhelicity on the NHE III$_1$ region. This dynamic stress shifts the equilibrium towards formation of the non-canonical structures, the G-quadruplex on the purine-rich and i-motif on the pyrimidine-rich strand. Thus, transcriptional activators such as CNBP and hnRNP K, which are single-stranded DNA-binding proteins, are prevented from binding and consequently transcription is not activated.[170,171]

Recent studies have also identified proteins that bind to c-MycG to regulate *MYC* transcription. In 2009 it was shown that nucleolin binds, facilitates, and stabilises c-MycG *in vitro* with high specificity.[31] The same study also identified that nucleolin binds to the NHE III$_1$ region *in vivo* using ChIP assays and that overexpression of nucleolin repressed *MYC* activity in

luciferase assays. Therefore, it's reasonable to suggest from this evidence that nucleolin induces c-MycG formation within the NHE III$_1$ region which prevents transcriptional activators from binding and thus reduces *MYC* expression. This theory has been supported by further studies.[32] Another protein that has become of interest is NM23-H2, which was shown to bind the NHE III$_1$ region in 1993 and has since been found to bind to the purine and pyrimidine single strands.[173,174] Florescence energy transfer experiments have demonstrated that NM23-H2 binds and unfolds the G-quadruplex within the NHE III$_1$ region. This is further supported by luciferase reporter and ChIP assays which show *MYC* is activated by NM23-H2 via an interaction with the G-quadruplex.[175] Meaning it looks reasonable that NM23-H2 binds to c-MycG and destabilises the structure so it unfolds which leads to activation of *MYC* transcription. This is another example that appears to support the proposed model above and demonstrate that c-MycG acts as a transcriptional switch. Consequently, there is enormous interest in targeting c-MycG as powerful anti-cancer drugs.[32,176,177]

All of this evidence makes focusing on the G-quadruplex very appealing. However, *MYC* regulation is complex and to be able to have effective cancer treatments we need to understand how the whole regulatory system works including the i-motif capable of forming on pyrimidine strand, which will be referred to as c-MycC throughout this thesis.[115] It was known that different c-MycC variations could be formed within this region (*table 1.1*). Hurley and Sun identified that the major c-MycC form was a 6:2:6 loop isomer sequenced for the 33 base sequence, will be referred to as c-MycC33, 5′-TCC-CCA-CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA-3′. However, this loop formation for c-MycC was revised to 5:5:5 in 2016 with greater footprinting resolution using the c-MycC52 oligonucleotide (*figure 1.16*).[4] Notably, c-MycC33 can also form the 5:5:5 formation, the difference being the reduced lengths of the over hang on the 5′ becoming one nucleotide base and the 3′ having no hangover (*figure 1.16).*

Table 1.1: Sequences of c-MycC27 and c-MycC52 oligonucleotides. The underlined DNA bases in the c-MycC52 oligonucleotide shows the c-MycC27 sequence and the bold DNA bases show the c-MycC33 sequence. All of these sequences are different lengths of oligonucleotides taken from the MYC protomer region capable of forming an i-motif.

Name	Sequence 5'-3'
c-MycC27	CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA
c-MycC33	TCC-CCA-CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA
c-MycC52	CTT-C**TC-CCC-A**<u>CC-TTC-CCC-ACC-CTC-CCC-ACC-CTC-CCC-A</u>TA-AGC-GCC-CCT-CCC-G
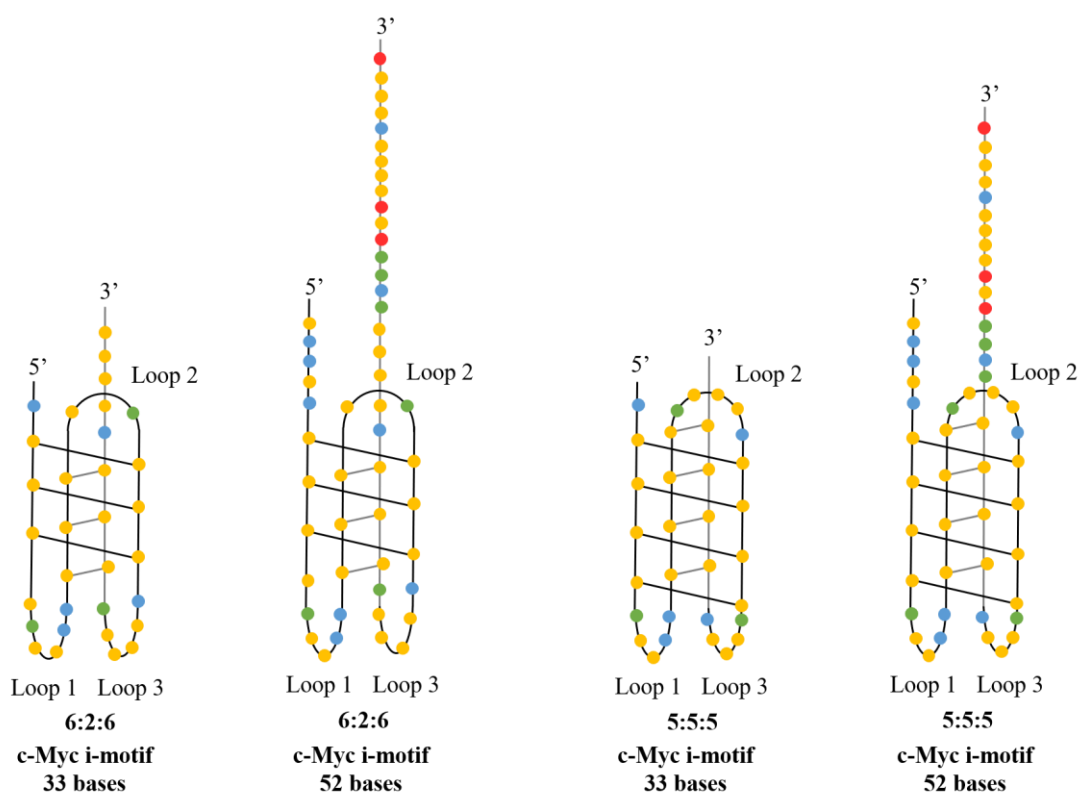
Figure 1.16: Proposed folding patterns of the MYC i-motif from 2010 (6:2:6) and the updated proposal in 2016 (5:5:5) for the oligonucleotide sequences that are 33 nucleotide bases and 52 nucleotide bases in length determined my Hurley et al.[3,4] Colouring scheme: red is adenine, blue is thymine, green is guanine, yellow is cytosine.

Researchers proposed c-MycC played a role in transcriptional regulation long before i-motifs were identified in cells.[178] Since there has been a surge of compelling evidence collaborating this view. Notably, that the *MYC* promoter is under superhelical stress *in vivo*, a condition in which c-MycC has been demonstrated to fold near physiological pH.[68,165] Both of these studies implied c-MycC could form *in vivo* and suggested a potential role in transcriptional regulation like c-MycG, with Leven's suggesting multiple DNA structures work in a dynamic equilibrium to regulate *MYC* transcription.[165] Interestingly, other studies had identified that *BCL2*'s transcription regulation involves an i-motif and hairpin complex providing support for i-motif's biological significance in promoter regions.[86,107] Significantly, it was demonstrated that hnRNP LL binds to the lateral loops of the BCL2 i-motif, unfolds it and initiates transcription, hnRNP LL is a member of the same family as the protein hnRNP K which is known to bind to the C-rich strand of NHE III$_1$ of the *MYC* promoter region. Although, it must be noted that hnRNP K is known to bind both ssDNA and RNA. Further suggesting it is plausible that c-MycC is involved in regulating *MYC* expression, and making it more likely that hnRNP K binds to c-MycC directly which was conjectured as early as 2000.[115] It took until 2016 to demonstrate that hnRNP K binds to the i-motif structure in the *MYC* promoter region.[4]

hnRNP K has three K homology (KH) domains which recognise the sequence consensus 5′-C-C-C/T-T-3′;[166] each of the three loops in the 5:5:5 formation of c-MycC contains 5′-CCTT-3′ in loop 1 and 5′-CCCT-3′ in loops 2 and 3 *(figure 1.16).* It has been demonstrated that two of the KH domains bind to two of the loops, with the following binding preference Loop3>loop2>loop1, leading to low levels of transcription activation. To achieve maximal activation it has been demonstrated that negative superhelicity and two additional cytosine tracts, located seven bases downstream of the i-motif forming region, are required, shown in the 52-base oligonucleotide sequence with the 5:5:5 loops in *figure 1.16*.[4] Which cooperates with a mutational study that identified all eight cytosine runs

are required for maximal transcriptional activation of P1 and P2.[7] This sequence will be referred to as c-MycC52, it is composed of 52 bases and can form a 5:5:5 loop isomer with a 16 base hangover on the 3′ end, 5′-CTT-CTC-CCC-ACC-TTC-CCC-ACC-CTC-CCC-ACC-CTC-CCC-ATA-AGC-GCC-CCT-CCC-G-3′. c-MycC52 achieves maximum transcriptional activation because hnRNP K recognises the 5′-C-C-C/T-T-3′ sequence in the loops and two of it's KH domains bind, and due to the i-motif being more conformationally constraining than ssDNA the final CT element can be accessed by the final KH domain (*figure 1.17*).[4]
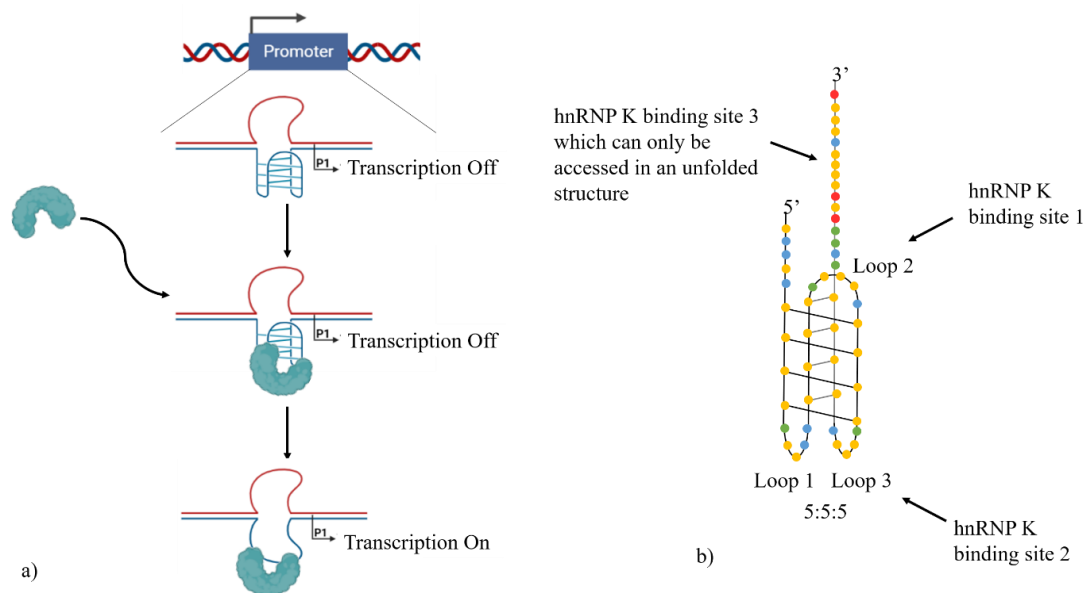
Figure 1.17: Proposed scheme for the control of MYC expression through the i-motif forming region in the NHE III₁. a) Proposed scheme for maximal transcriptional activation where hnRNP K binds to two of KH domains within the i-motif loops and due to conformational constraint the i-motif is unfolded and the additional KH domain is bound to forming a thermodynamically stable complex. b) An example schematic of the 52-base oligonucleotide sequence of the i-motif forming region within NHE III₁ and the KH domains, notably loop 1 is also be a KH domain but hnRNP K only binds to two i-motif KH domains within the loop region and one in the tail, it could bind loop 1 but this is the least favourable loop. Created with BioRender.com

This research highlights that c-MycG and c-MycC are mutually exclusive and the transcription factors that recognise these structures dictate activation or silencing of *MYC* transcription. Evidently these sister structures carry out regulatory roles in NHE III$_1$. Therefore, to fully understand the regulation by NHE III$_1$ the pyrimidine and purine strand needs to be considered. This dual-stranded system opens up the possibility of therapeutic avenues to explore targeting both structures. To begin to explore this a clearer understanding of how these regulatory structures function individually and as a system as a whole.

1.5.5 Targeting the i-Motif Structure in NHE III$_1$

The i-motif field have begun to explore targeting c-MycC as a prospective target for cancer therapeutics. One group has identified **B19**, an acridone derivative (*figure 1.18*), which selectively binds and stabilises c-MycC.[82] They further demonstrated that when **B19** is used in cells it can down-regulate *MYC* transcription and can cause apoptosis in cancer cells. This work is an interesting avenue for c-MycC anti-cancer drugs. However, this study used a 27-base oligonucleotide version of the i-motif forming region in the *MYC* promoter, which will be referred to throughout this thesis as c-MycC27, which cannot form the 5:5:5 loop structure that Hurley's group have demonstrated hnRNP K binds to.[4] Yet the i-motif structure can form a 5′-C-C-C/T-T-3′ sequence in loop 2, which has a similar loop to that of loop 3 in the c-MycC52 5:5:5 formation (*figure 1.19*). However, it has not been demonstrated where this compound binds and so the mechanism causing the downregulation of *MYC* is not determined.
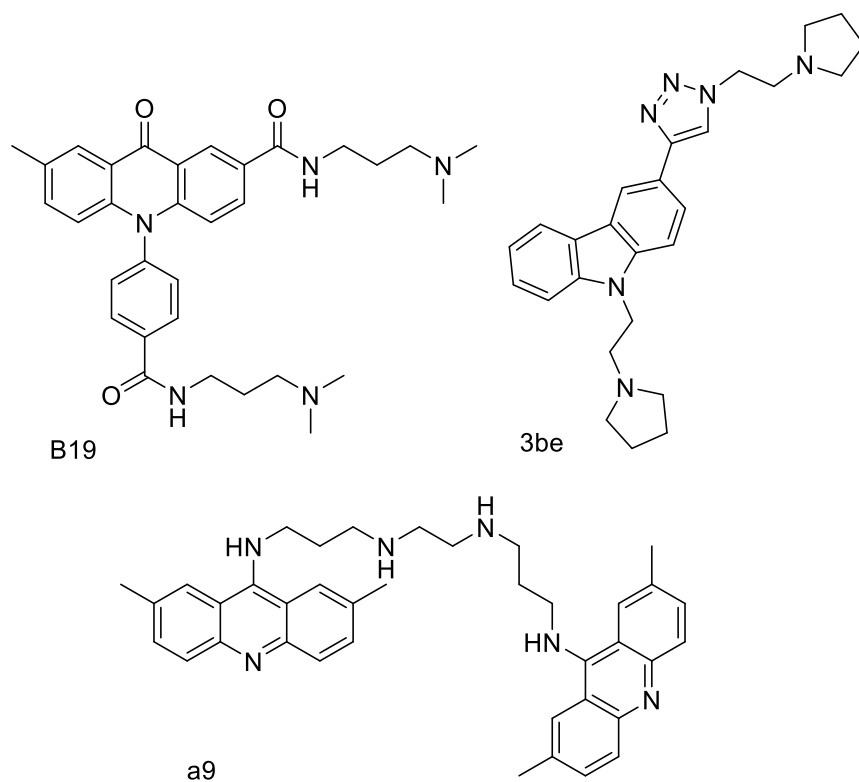
Figure 1.18: Structures of MYC i-motif ligands that affect MYC transcription in cells: B19 and 3be and a9.
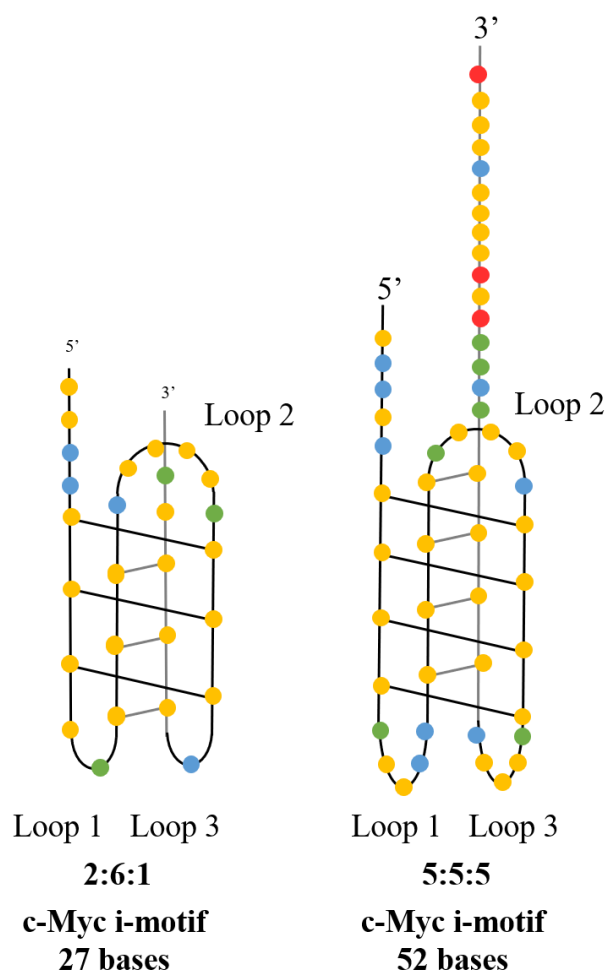
Figure 1.19: Proposed folding patterns of the MYC 52-base oligonucleotide sequence and a variation of of the MYC 27-base oligonucleotide variation with loops labelled and loop ratios determined. Colouring scheme: red is adenine, blue is thymine, green is guanine, yellow is cytosine.

Another group has also identified a c-MycC33 specific compound using an *in situ* approach using nanotemplates called **3be**, a triazole compound (*figure 1.18).*[6] The use of c-MycC33 as the target should translate well *in vivo* as this sequence forms the 5:5:5 structure that is required for maximal transcription (*figure 1.16*). Cell cytotoxicity studies identified that **3be** inhibits cell proliferation in the cancer cell lines (HeLa, A549, and B95.8) and is minimally toxic in the normal cell line NKE. It was also identified that **3be** enters the cell nucleus and causes apoptosis in a dose dependent manner in HeLa cells.[6]

A bisacridine derivative, **a9**, has been demonstrated to bind c-MycC and c-MycG and stabilises both structures (*figure 1.18*).[87] The cytotoxicity studies were carried out in multiple different cancer cells lines and demonstrated varied IC$_{50}$ (half maximal inhibitory concentration) values from 0.15 µM to 1.22 µM. It was further demonstrated that in SiHa cells **a9** caused a decrease of *MYC* transcription.[87]

Evidently, there is limited research on c-MycC specific ligands as drug targets, however studies so far and the successes of G-quadruplex ligands entering clinical trials make this exploration look promising.[179]

1.6 Techniques to Develop and Investigate i-Motif Binding Probes Used in this Thesis

Multiple techniques can be used in the process that identifies i-motif binding probes. The choice depends on the nature of the probe as well as the conditions required. Generally, it is best practice to use a multitude of techniques. The following provides a brief introduction to the procedures used in this thesis.

1.6.1 Phage Display Technology

Phage display technology utilises bacteriophages, commonly referred to as phage, as vectors to display a protein of interest on their surface (*figure 1.20*). Phage display was developed by George Smith in 1985[180] and the 2018 Nobel Prize in Chemistry was awarded jointly to George P. Smith and Sir Gregory P. Winter for the phage display of peptides and antibodies. Phage display was developed as a simple, efficient technology using filamentous phage to screen and identify specific ligand interactions with a molecular target. Phage display saw an eruption of scientific interest

after George Smith's 1988 publication establishing the use of phage to produce large libraries for exploration of protein-protein, protein-peptide, and protein-DNA interactions.[181] Phage display has multiple applications including drug discovery, cancer research, and identifying DNA binding ligands. Phage display has demonstrated its potential in discovering high affinity-binding ligands to molecular targets and is a continually advancing field.[182] Hence, this is a powerful technique to identify probes for secondary DNA structures.
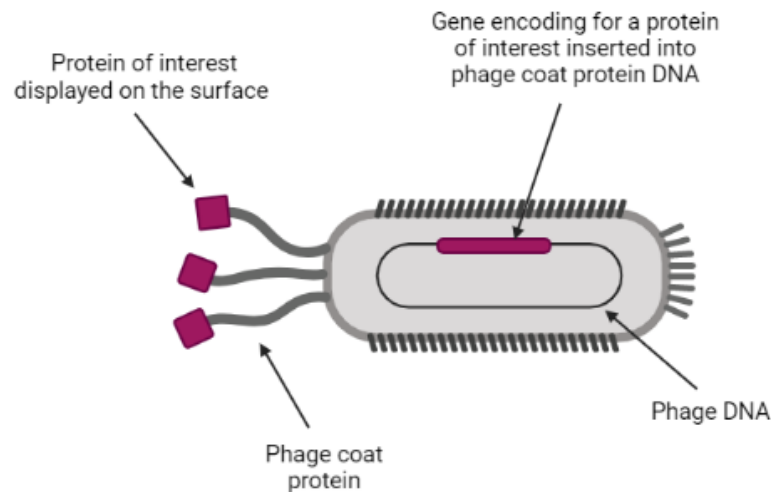
Figure 1.20: Illustration of an M13 phage that has the genetic information of a protein on interest, shown in pink, inserted into the phage DNA encoding for a coat protein. The protein of interest is shown as a pink square attached to an M13 coat protein. Created with BioRender.com.

1.6.1.1 Introduction to Bacteriophages

Phage display is a useful screening mechanism as the phenotype and genotype of the phage are physically linked (*figure 1.20*). Meaning that foreign DNA can be inserted into the ssDNA that makes up the phage

genome and the peptides or proteins coded for by the foreign DNA are then expressed in fusion with the phage coat protein.

There are 6 basic morphological phage groups (*figure 1.21*), in the interest of phage display the main phage of interest are DNA containing filamentous and hexagonal head-tail phage that can infect the model organism *E. coli*.[183–185] Those utilising phage display frequently use F-specific filamentous phage (M13, f1, and fd) although the following head-tail phage, T4, T7, and λ have also been used. Importantly, these phage exploit different reproductive cycles. Filamentous phage use the lysogenic life cycle which doesn't cause lysis or death of host cells whereas the head-tail phage reproduce via the lytic lifecycle and cause lysis and thus death of their hosts. The use of filamentous phages is more applicable because they don't cause lysis of host cells. For the research in this thesis the filamentous phage M13 was used as it is a better characterised phage display system and meets the criteria of the study.
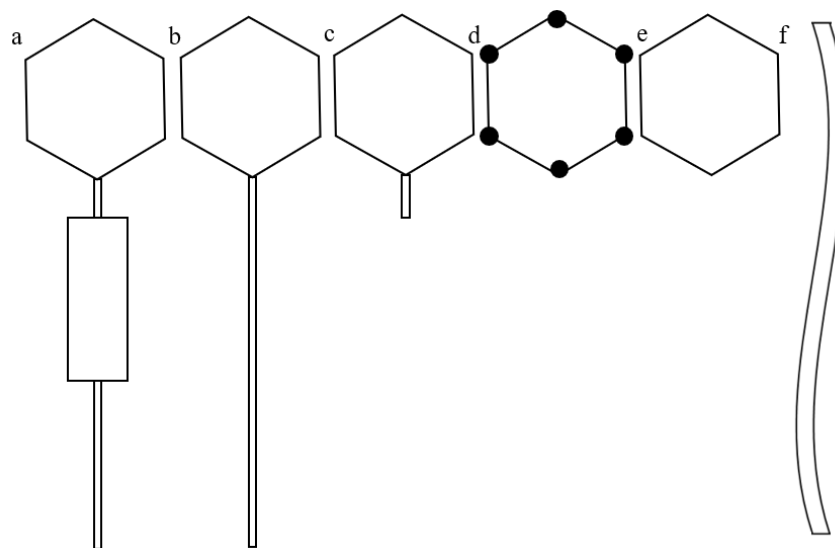
Figure 1.21: The 6 basic morphological types of bacteriophages. a) This is the most complex phage morphology. The shape is a hexagonal head with a tail and a contractile sheath that permits contraction of the tail. b) The shape is similar to a with hexagonal head and tail, however the tail in non-contractible. c) A hexagonal head with a short non-contractible tail d) A hexagonal head with either a large knob or capsomere on each apex. e) A hexagonal head. f) Long flexible filament with no attachments.

1.6.1.2 M13 Bacteriophage, Structure, and Life Cycle

M13 bacteriophage are rod shaped filamentous phage part of the Ff family (*figure 1.22*), which also includes f1 and fd phage.[186] All Ff phage are F (male) specific and infect gram negative bacteria such as *E. coli*. M13 are composed of a nucleoprotein that is ~900 nm in length and 6-7 nm width which encloses supercoiled, circular, ssDNA genome of 6407 bp that encodes for 11 proteins (*figure 1.22*). Where gene I (gI) encodes protein I (pI) and gII encode pII and so on. The proteins are organised into three classes: proteins involved in replication (pII, pV, and pX), morphogenic proteins required for assembly and secretion (pI, pIV, and pXI), and structural coat proteins (pIII, pVI, pVII, pVIII, and pIX).[186,187]
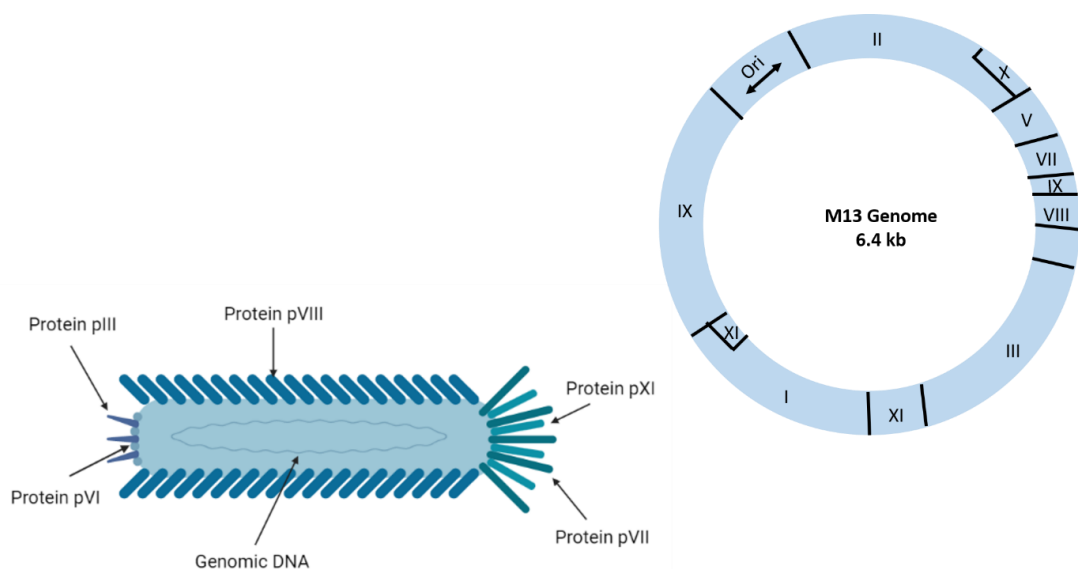
Figure 1.22: Filamentous phage M13 structure and genome. Created with BioRender.com.

The structural coat proteins are further grouped as the major coat protein, pVIII, and the minor coat proteins which are pIII, pVI, pVII, and pIX. There are 2700 copies of the major coat protein, pVIII, which is located along the lengths of the structure.[186,187] The coat proteins are largely

α-helical. The helices of pVIII extend in an axial direction and overlap producing a hollow rod that encloses the DNA. Amino acid residues that make up the α-helix are orientated so that the acidic residues extend in an axial direct and overlap interact with its external environment and the basic residues protrude internally neutralising the phosphate backbone of the DNA.[188–190] Whilst there are only 3 to 5 copies of each of the minor coat proteins that are located at the ends of M13. One end of the phage expresses pIII, and pVI which form an adsorption complex that exposes pIII on the phage exterior to recognise the F pilus.[186] Furthermore, alteration of pIII accommodates different sizes of DNA by increasing the length of M13 but not affecting the width, this contributes to the observed flexibility of M13.[191] On the other end of the phage structure is pVII, and pIX and this end initiates assembly of the phage in the membrane of the host.[186] The minor coat proteins are essential for initiating entry of M13 into the host, mediating phage assembly, and formation of pore complexes to export the phage from the host cell.[192] The reproductive cycle of M13 is known as the lysogenic cycle and do not lyse infected cells, but instead uses the host cell machinery to continually produce phage which are subsequently exported through the cell membrane. The reproductive cycle relies on pIII-F pilus binding through recombinant expression.[193,194]

1.6.1.3 Construction of Phage Display Libraries

Recombinant polypeptide libraries can be constructed by different approaches that either introduce random mutations or produce defined libraries by using synthetic DNA.[195] Oligonucleotide-directed mutagenesis is frequently used to introduce mutations into genes where the position is controlled.

There are 20 different natural amino acids so when randomising n amino acids the different possible peptide sequences are 20^n, thus one can produce

a very diverse library.[196] Generally solid phase DNA synthesis is used to produce randomised mutations, by using equimolar concentrations of the four nucleotides. However, this does cause a bias towards certain amino acids as the 64 natural codons code for the 20 amino acids and stop codons (there are 61 sense codons and three stop codons). Therefore, there is bias towards certain amino acids and stop codons prevent biosynthesis when inserted into the phage, subsequently the virion is uninfective. For randomised positions, relative representations of each amino acid can be improved by limiting the third position of each codon to G or T (to equal A or C on the synthetic library oligonucleotide).[197] Although this does not account for the disproportional abundance of amino acids seen as it is not proportional to the number of possible codons.[195,196]

1.6.1.4 Phage Display Method

Phage display is based on a phenotypic and genotypic link which leads to presentations of peptide/protein libraries on the phage surface.[198] This technology is used to create libraries with 10^6-10^{11} different ligand populations in a population of at least 10^{12} phage virions. Phage display is a biopanning method is based on repeating cycles of binding, washing, elution, and amplification as shown in *figure 1.23*.
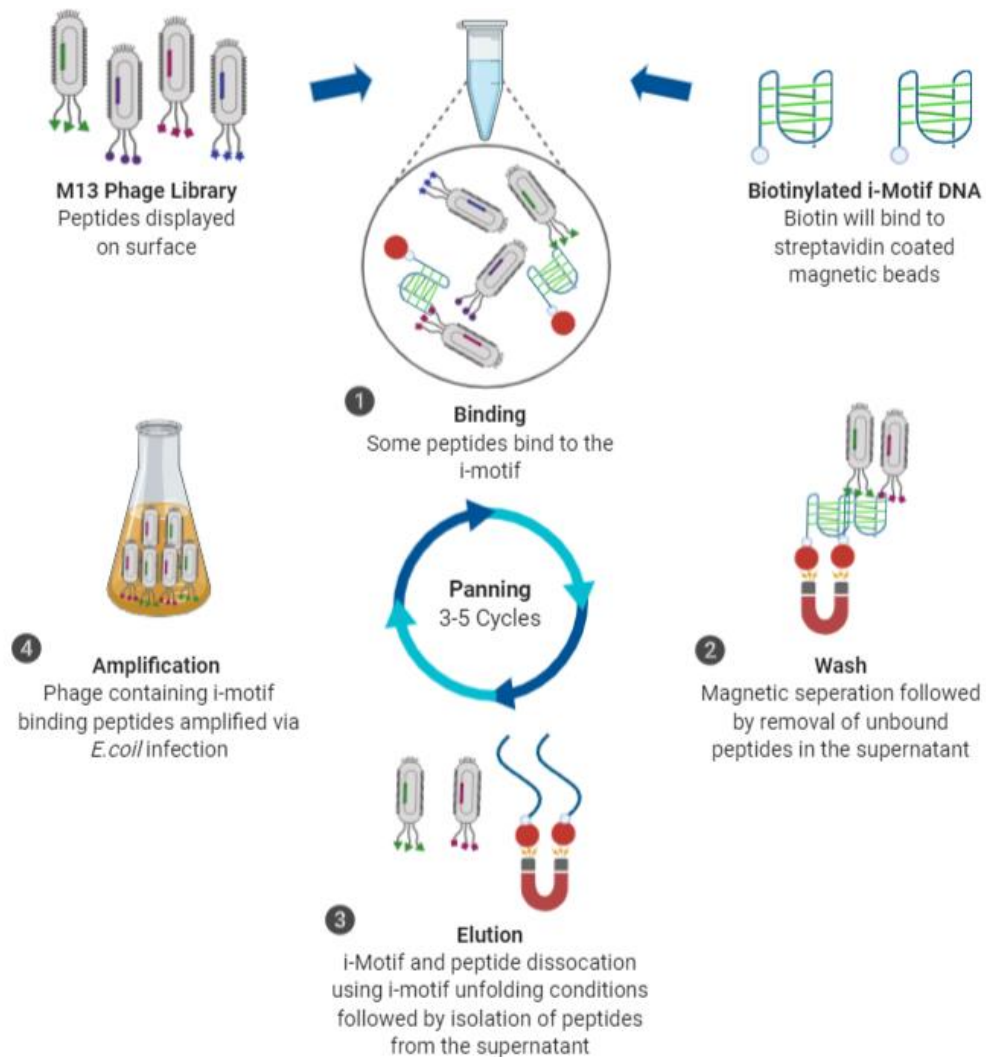
Figure 1.23: Generic phage display selection procedure where an i-motif is the target molecule of interest. The plate is coated with streptavidin, shown by grey crosses, so that the i-motif can bind via biotin, which is depicted as a blue ball. The first step is binding of the phage to the target molecule, followed by washing to remove non-binders and elution to obtain the phage with an affinity for the target molecule. The phage in the elution are amplified and then can be analysed or go into a further round of biopanning. Created with BioRender.com.

Phage display has multiple advantages in identifying target-binding peptides/proteins, the key being it can identify specific target-binding probes from a library of millions of different probes at once.[197] Making it possible to screen billions of probes weekly. Another advantage is that libraries can be created focusing on known epitopes that bind the target

molecule to identify better probes for example greater binding affinity or greater specificity.[198] Using peptides have even further advantages including ease of synthesis, structural simplicity, small molecular size, easy to amplify without replicative interferance, the process is easy, there are protocols in use in many forms, low cost, and more likely to readily diffuse such as into the nucleus than larger biomolecules.[199] The limitations of phage display include that probes identified can have low affinities, library construction is time consuming and commercial libraries are costly and not based off previously known binders so are likely to have lower binding affinities. Furthermore, this system has a replication bias which will affect the amino acids selected for the sequence making the process less random when constructing random libraries, and the prokaryotic machinery of the bacteria are unable to make crucial post-translational modifications.[197,200] To conclude, phage display can be a powerful technique to rapidly screen to identify specific interactions for a molecular target.

1.6.1.5 The Scope of Phage Display Applications

Phage are exploited for phage display for an array of reasons, including that their genotype and phenotype are linked, they are stable under a spectrum of conditions including across temperatures and pH, and loss of infectivity from insertion of foreign DNA can be avoided.[201] Filamentous phage are advantageous as cloning vectors and for the display of peptides/proteins compared to other phage. Including length flexibility that allows altered genome lengths, the N-terminus of pIII is easily accessible making the pIII gene an ideal target for insertion of peptides/proteins, insertions in non-essential regions do not disrupt packaging, and their genome can be either ssDNA or dsDNA.[201] Thus it is unsurprising phage display is utilised across multiple fields and is frequently used in drug discovery.[29,191,202]

1.6.1.6 Phage Display and Higher Order DNA Structures

As interest in higher order DNA structures and functions increases, the importance of specific probes grows. Once the probe is developed it can be exploited in various procedures including investigating where, when, and under what conditions these structures are forming in cells. Thus, providing potential biological functions that can be further investigated where opportunities for further use of the probes can arise.

At the time of writing there were limited published applications of phage display to identify probes specific for secondary DNA structures: Z-DNA,[203] G-quadruplexes,[29,204] and i-motif[94] structures. There were more instances of phage display being used to find ligands for DNA and RNA, although to a much smaller extent than in other fields such as drug discovery.[205–208] These studies used phage display to identify specific antibodies/proteins/peptides that bound to DNA/RNA for a variety of reasons. One of which was to investigate RNA-protein interactions that naturally occur whereas another was investigating specific antibody probes for RNA which could be developed as diagnostic tools or therapeutics to treat disease.[207] Even though the aims of these studies were different the principles and general method of phage display remained the same. Evidently it is plausible that this could be a powerful technique for identifying probes for secondary DNA structures.

The research that identified antibodies specific for Z-DNA were not actually looking for a probe to investigate Z-DNA, but rather a method to produce anti-Z-DNA antibodies for autoimmune studies.[203] Nevertheless, they demonstrated that using biotinylated Z-DNA as a target in phage display bound to particles being displayed on the anti-Z-DNA antibodies. Therefore, demonstrating that phage display can be used to identify amino acid residues interacting with nucleic acids.

The antibody probes for G-quadruplex and i-motif structures were achieved using phage display to identify the probes and further techniques to validate binding.[29,94] The antibody probe for G-quadruplexes, BG4, was one of the antibodies selected after 2 rounds of biopanning, then screened, and had their binding validated and affinity assessed by ELISA before cell culture experiments were carried out.[29] Identification of iMab, the antibody that binds to the i-motif, was identified in a similar manner but from three rounds of biopanning.[94] This study shows that phage display works at acidic pH and could be an excellent method to explore i-motif binding peptides.

Balasubramanian, whose lab identified the G-quadruplex antibodies,[29,209] has also investigated the binding of zinc fingers to G-quadruplex using phage display as a display strategy. Zinc fingers are a protein motif that resemble the shape of a fingers with a Zn^{2+} ion stabilising the folded secondary structure. Their biological function has a wide scope and their Cys_2His_2 domains are known to bind to DNA and RNA via recognition of specific nucleotide triplets.[210] Due to the ability of zinc fingers to recognise and bind specifically to DNA with high affinity and the success and promise of its use for dsDNA[211] it was investigated as a probe for G-quadruplexes in 2000.[204] Choo *et al.* demonstrated that zinc finger proteins can bind to G-quadruplexes, such as protein Gq1, and those that did had strong amino acid consensus suggesting similar binding modes. Their binding modes were explored using enzyme-linked immunosorbent assay (ELISA) and gel mobility shift assays and showed sequence and structure specificity,[204] meaning these proteins are excellent candidates as secondary DNA structure probes. Gq1 has been further investigated to identify how it recognises G-quadruplexes, which has identified which fingers of Gq1 can be modified without impaired recognition of G-quadruplexes or loss of discrimination of other DNA structures.[212] Therefore, investigating zinc fingers is a promising avenue for identifying probes that bind to DNA structures and could provide some structural

insights on which nucleic bases probes are important for specific interactions.

However, there are still limited probes available for higher order DNA structures and the probes available generally lack specificity. For example, it has recently been shown that the commercially used G4 small molecules also bind i-motifs.[75] Notably, the G-quadruplex binding proteins discussed above that were identified using phage display did not use i-motif DNA as competitor during the phage display rounds and data showing they don't bind i-motif was not presented.[94] Due to the binding of the "specific" G-quadruplex small molecules to i-motif DNA these proteins cannot be classified as G-quadruplex specific until tested against i-motif DNA. For example, the antibody fragment iMab was tested against G-quadruplexes as well as B-DNA to demonstrate that it selectively binds i-motif DNA.[94] However, iMab cannot differentiate between different i-motif sequences and specificity between i-motifs is generally lacked across the field. Highlighting the need for further probes which can be achieved by phage display.

1.6.2 Fluorescence Indicator Displacement (FID) Assay

Fluorescence indicator displacement (FID) assay is a biophysical assay that can detect binding of ligands to DNA structures. They fundamentally require an indicator which has optical properties that are modulated by reversible, non-covalent binding (*figure 1.24*). The indictor is initially unbound free in solution giving off no optical signal. When in the presence of the target the indicator-target complex causes the indictor to produce a fluorescent signal that is detected. Then addition of a competitor analyte is added to solution and if this analyte displaces the indicator then the indicator is free in solution and giving off no fluorescent signal. Therefore,

a reduction in fluorescent signal is directly related to displacement of the indictor from the target.
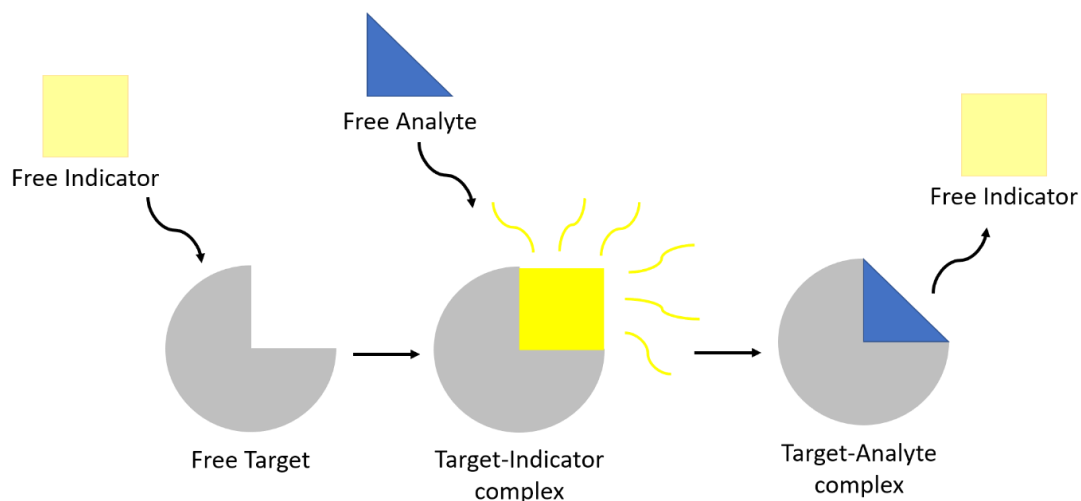
Figure 1.24: Schematic of fluorescence indicator displacement assay steps. Where there is a fluorescence molecule referred to as indictor which is only fluorescent upon binding to the target but not when free in solution. The target-indicator complex fluoresces but when a competitive analyte displaces the indictor to form a target-analyte complex the indicator is free in solution again and not fluorescence.

Thus, FID relies on the fluorescent enhancement of a probe upon binding a DNA structure and the selection of that probe is important for assay development. For i-motif studies **Thiazole orange** (**TO**) was identified as the best probe by the Waller group (*figure 1.25*). **TO** was selected as it is commercially available, can be used to assess binding against diverse DNA structures, gave significant fluorescence enhancement with the i-motif forming sequence in the human telomere (hTeloC) which this assay was developed with, **TO** demonstrated a strong enough binding to observe good fluorescence signal but weak enough to be displaced by a competitor analyte, low micromolar dissociation constant, could be used at acidic pH (essential for i-motif work as the structures are stabilised by acidic pH), structural circular dichroism studies showed that **TO** didn't affect the structure of hTeloC.[85] Since, **TO** has been used to carry out competitive

assays searching for i-motif ligands.[213] When developing this method it is essential that **TO** equilibrates with the DNA structure first so that the fluorescence is at a steady state to obtain maximum fluorescence. Once this is achieved competitor analytes can be added, where a decrease in fluorescence demonstrates **TO** displacement by competitor analytes.
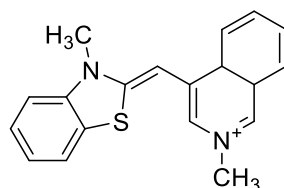
Figure 1.25: Structure of Thiazole orange.

FID assays are advantageous as they permit sensitive and quantitative detection, are relatively cheap and fast to perform. Furthermore, FID doesn't require site-specific labelling like other commonly used fluorescence-based methods in the non-canonical DNA structures field such as Förster resonance energy transfer (FRET). Additionally, it requires small quantities of material in comparison to other biophysical techniques used in the field including isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), and nuclear magnetic resonance (NMR) spectroscopy.[214] Thus, making FID an ideal technique for high throughput studies. While FID is an excellent choice for an initial high throughput study to evaluate binding of analytes to DNA structures obtained data is limited and there are drawbacks. One drawback of the technique is that for the binding of the competitor analyte to the DNA to be detected the competitor analyte must displace the indicator molecule. It is possible the competitor analyte binds at a different binding site on the target which doesn't cause displacement of the indicator molecule and thus these analytes, although binding, are not detectable with this assay. Additionally, whilst an excellent technique for rapidly collecting data the obtainable data is limited to percentage displacement, and the DC_{50}, the

concentration at which 50% of the indicator is displaced. Whereas other techniques enabling measuring of binding constants, binding epitopes, and kinetic parameters. Moreover, FID needs to be optimised and interpretation of results requires a specialist to ensure high quality and reliable method of screening.[214] Nonetheless, FID is an excellent high-throughput screening method to identify analytes that bind to your target from large libraries.

1.6.3 Ultraviolet (UV) Spectroscopy

Ultraviolet (UV) spectroscopy is a widely used analytical technique within multiple fields. Absorption of UV light by molecules causes electronic transitions. An energy transfer occurs that that excites electrons which causes those in lower energy orbitals excited into higher energy orbitals. The absorbance is then plotted against the wavelength of light to produce a spectrum.[215] UV spectroscopy is a useful way to determine DNA and peptide concentration. To do this the UV absorbance at a particular wavelength is converted to protein concentration using the Beer-Lambert law:

$$A = \epsilon\, cl$$

Equation 1.1: Beer-Lambert law where A is absorbance, ε the molar coefficient, $M^{-1}\, cm^{-1}$*, c is the concentration, and l is the path length of the cuvette in cm.*

Where A stands for absorbance, ε is the molar extinction coefficient, $M^{-1} cm^{-1}$, c is the concentration, and l is the path length of the cuvette in cm.[216] Peptide absorbance is most accurately read at 280 nm, which is the wavelength at which tryptophan and tyrosine absorb. It is commonly stated that peptides without tryptophan or tyrosine cannot have their concentration determined accurately by UV spectroscopy. The Nick Anthis' Protein Parameter Calculator permits determination of peptide

concentration using absorbance at 205 nm, although it will not be as accurate as using absorbance at 280 nm due to the added interference at the smaller wavelength.[217] This calculator also provides the molar extinction coefficient.

It should be mentioned that working within the i-motif field that UV spectroscopy is widely used to monitor nucleic acid folding and unfolding in solution although not utilised for this thesis. There are characteristic spectrums for different DNA structures due to hydrogen bonds and base stacking in the structure, although spectral changes are not fully understood they are sensitive enough to the base environment to determine changes in DNA structures.[218]

1.6.4 Circular Dichroism (CD) Spectroscopy

Circular dichroism (CD) is a light absorption spectroscopy based on the differential absorption of circularly polarised light which identifies characteristic signals related to the difference of absorption of left- and right-circular polarised light (*figure 1.26*). The main unit used for circular dichroism is ellipticity.[219]
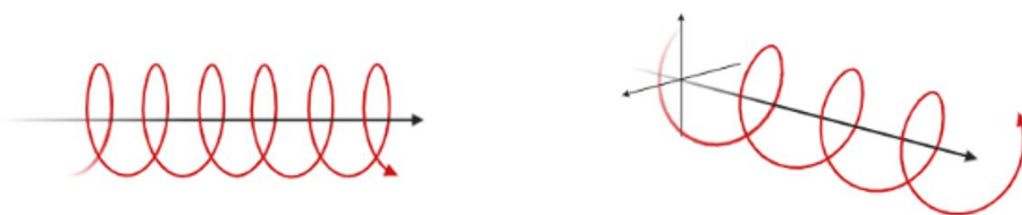
Figure 1.26: 2D (left) and 3D (right) schematics of circular polarised light. Created with BioRender.com.

CD uses a light source that is passed through a filter wheel to form monochromatic light which is light that is a single frequency.[220,221] The monochromatic light is passed through a plane polariser which only allows

light through which has waves with vibrations in only one direction, the same direction as the polarising filter, meaning there is one electromagnetic plane wave. This light is then passed through a circular polariser to make the light circularly polarised, meaning there is now two perpendicular electromagnetic plane waves which are 90° difference in phase but the same amplitude. The circular polarised light then passes through a sample and is absorbed by the sample and transmitted light is transmitted to the detector which creates a circular dichroism spectrum (figure 1.27). When the sample absorbs the left and right circularised light optically active molecules absorb the left- and right-circular polarised light differentially and this produces a CD spectrum
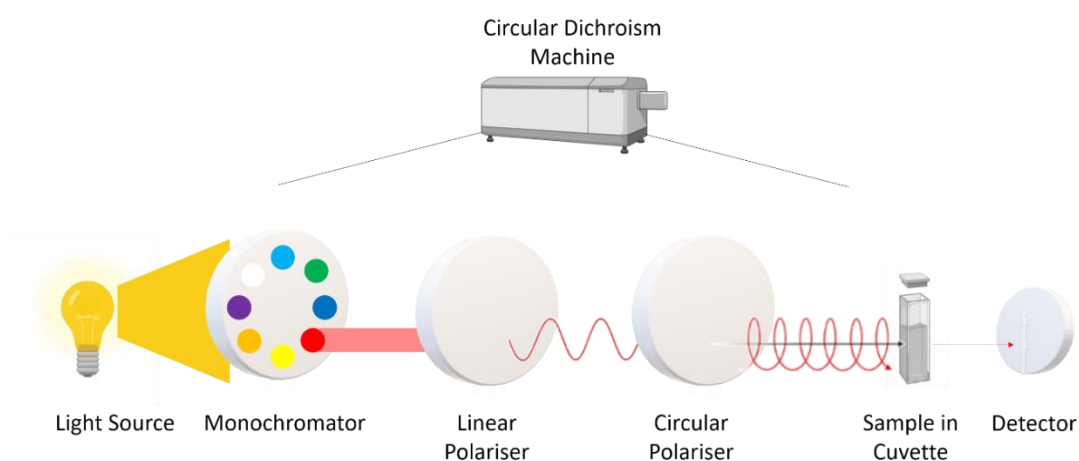
Figure 1.27: Schematic of the process of circular dichroism from the light source to the detector. Light leaves the light source and travels through a filter wheel to get monochromatic light. The monochromatic light is passed through a plane polariser to give one electromagnetic plane wave and is then passed through a circular polariser to make the light circular polarised, where there are two perpendicular electromagnetic plane waves that are 90° out of plane. The circular polarised light is absorbed by a sample and transmitted light is detected by the detector. Created with BioRender.com.

CD is used for rapid structural studies and has many applications that arise from its sensitivity to different conformational states of macromolecules such as proteins and DNA.[222,223] CD can determine effects

of mutations, environment such as temperature and pH, and the presence of ligands on macromolecules structures. Additionally, CD is also utilised to investigate macromolecule stability, folding, ligand binding, and characterisation studies.

The application of CD to monitor peptide structures is possible due to the optical activity of the C_α carbon chirality in all amino acids except glycine.[224] The spectral region 170-250 nm is dominated by the peptide backbone and is used to estimate protein secondary structure.[225] Therefore, the choice of buffers is limited due to the required lower wavelength generally potassium/sodium phosphate or fluoride buffers are used as they can go to 185 nm. Quartz glass is also essential to go down to the lower wavelength limit.[223] The accuracy of CD, compared to X-ray structures, is ~97% for α-helix structure, ~75% for a ß-sheet, and ~50% for ß-turns.[226] The characteristic spectra for a-helix, ß-sheet, and random coil are in (*figure 1.28*).[5] These reference spectrums are very useful for analysing protein structure CD spectrums however the exact reference spectra for different secondary structure types are not identical. Therefore, servers are available, such as DichroWeb, which analyse experimental data and provide a percentage of the different structures present based in a database of proteins known structures.[227–229]
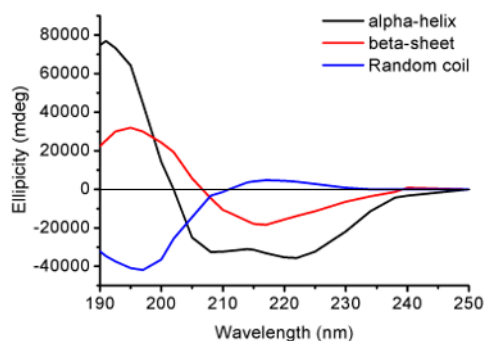
Secondary Structure Element	Signal	Position of minimum or maximum (nm)
α-helix	Positive	190-195
	Negative	208
	Negative	222
β-sheet	Positive	195-200
	Negative	215-220
Random coil	Negative	200
	Positive	220

Figure 1.28: Graph illustrating the reference circular dichroism spectrums of α-helix, β-sheet, and random coils (left) with a table listing the characteristic references (right). Data was obtained from Fasman et al.[5]

The application of CD spectroscopy to DNA structures is possible due to three different factors which are asymmetry of the sugar moity, relative base-stacking geometry, and helical positioning of the base pairs.[230] The spectral region 200-320 nm is used to monitor DNA structures. Therefore, there is a greater range of buffers that can be used for DNA studies than compared with reading peptide structures as the shorter wavelengths are not required. Different DNA structures have characteristic CD spectrums which can be used as reference spectrums to determine the DNA structure, although exact positions and amplitudes of CD bands to differ due to sequences even if the secondary structure is the same due to the three factors stated previously (*figure 1.29*). Where B-DNA has a strong positive band between 260-280 nm and a negative band at 245 nm,[231] the i-motif has a dominant positive band at 288 nm and a negative peak at 267 nm,[115] parallel G-quadruplex structures have a dominant positive peak at 260 nm, and a positive peak at 210 nm, and the anti-parallel G-quadruplex has a negative band at 260 nm, a positive band at 290 nm, and a positive band at 210 nm.[232] CD is commonly used within the alternative DNA structure

field to determine the structure of different oligonucleotides in different conditions and monitor the effects when ligands are bound.

Secondary DNA Structure	Signal	Position of minimum or maximum (nm)
B-DNA	Positive, Long	260-280
	Negative	245
i-Motif	Positive, dominant	290
	Negative	267
Parallel G-quadruplex	Positive, dominant	260
	Positive	210
Anti-parallel G-quadruplex	Negative	260
	Positive	290
	Positive	210
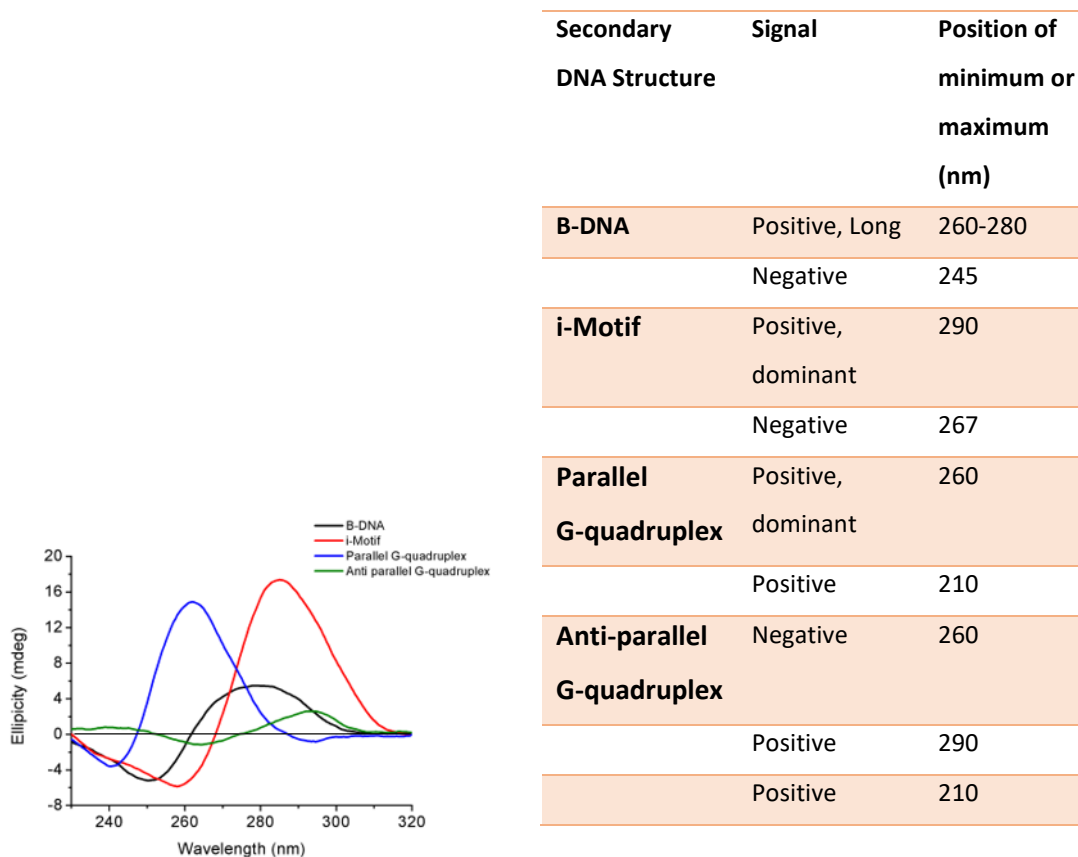
Figure 1.29: Graph illustrating the reference circular dichroism spectrums of B-DNA, i-motif, parallel G-quadruplex, and an anti-parallel G-quadruplex (left) with a table listing the characteristic references (right). Data for the anti-parallel G-quadruplex was provided by Carlos Gamez Alvarez, a colleague in the Waller lab.

CD melting experiments can also be performed to gain further information on the thermal stability of the structure in various conditions including the presence of different ligands. CD melting is a method in which the temperature of a sample is gradually increased, and the CD spectrum is recorded at intervals throughout, at the lower temperature the DNA structure will be folded and at the higher temperature it will be melted and thus unfolded. The temperature at which this melting occurs depends on stability. Therefore, CD melting experiments can be used to monitor the

change in melting temperature (ΔTm) in different conditions and identify stabilising (positive ΔTm) or destabilising (negative ΔTm) conditions.

Using CD for peptide or DNA secondary structure analysis is advantageous as it is very sensitive which permits using low concentrations of DNA or peptides and can lengthen cuvette width to lower concentrations further. Which is particularly advantageous if studying samples with poor solubility such as long DNA structures or environmental conditions which tend to cause structures to aggregate for example high salt conditions. Additionally, it can be used to study a large variety of DNA or peptide lengths investigating short and long oligonucleotides and amino acid sequences and so it can be utilised for investigating structures of most sequences. Furthermore, CD can be used to explore the effect of titrating various agents and ligands into the sample of DNA or peptide to explore the effect of them on the structure, which is useful for understanding environmental and ligand induced changes. In comparison to other conformational analysis techniques CD is fast and relatively inexpensive once you have the equipment. Although, as with any technique there are limitations such as a lack of atomic level structure analysis. Finally, CD spectroscopy cannot be used to provide definitive structural data without complementary conformational analysis. Nonetheless, CD is an excellent technique for determining secondary structural elements of peptides and oligonucleotides and the effects of different conditions on these very quickly.

1.6.5 Surface Plasmon Resonance (SPR)

Surface plasmon resonance (SPR) is a powerful optical detection technique that allows real-time, label-free detection of biomolecular interactions.[233] This technology can be used to determine binding affinities and kinetics. SPR happens when polarised light strikes an electrical conductor, such as

gold, at the interface between two media. This causes electron charge density waves, known as plasmons, to resonance as the light is absorbed by delocalised electrons in the metal surface causing them to resonate, these resonating electrons are very sensitive to their environment which causes changes in the intensity of light reflected at the specific angle which is proportional to the mass on the metal surface (*figure 1.30*).
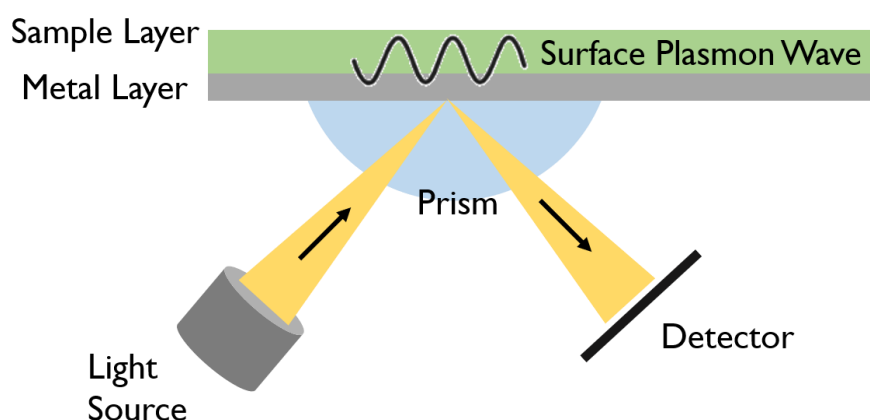
Figure 1.30: Schematic of how the optical technique surface plasmon resonance works. Where polarised light from a source strikes a mental conducting sensor, causing plasmons to resonant, causing a change in intensity of light reflected at a single angle.

The application of SPR to monitor DNA-ligand interactions is a popular method within the non-canonical DNA structures field.[234] The most frequently used immobilisation method of G-quadruplexes and i-motif DNA in the literature to evaluate interactions is by biotinylating the 5′-end of the oligonucleotide. This permits usual folding of the oligonucleotide structures, although care must still be taken as modifications can cause steric hinderance.[234,235] Therefore, as the target DNA will be biotinylated a streptavidin or neutravidin coated gold sensor chip is used for SPR studies to immobilise the DNA. The polarised light source passes through a prism and is reflected off the surface of the chip opposite the ligand immobilised (this is the glass side with no coatings) and into a detector as shown in *figure 1.31*. As ligands of interest (analytes) are passed over the surface of the chip in solution the analytes bind to the DNA immobilised

on the chip. Causing a change in the angle of reflected light, the change in this angle is proportional to the mass of analyte bound allowing biomolecular interactions to be recorded (*figure 1.31*).
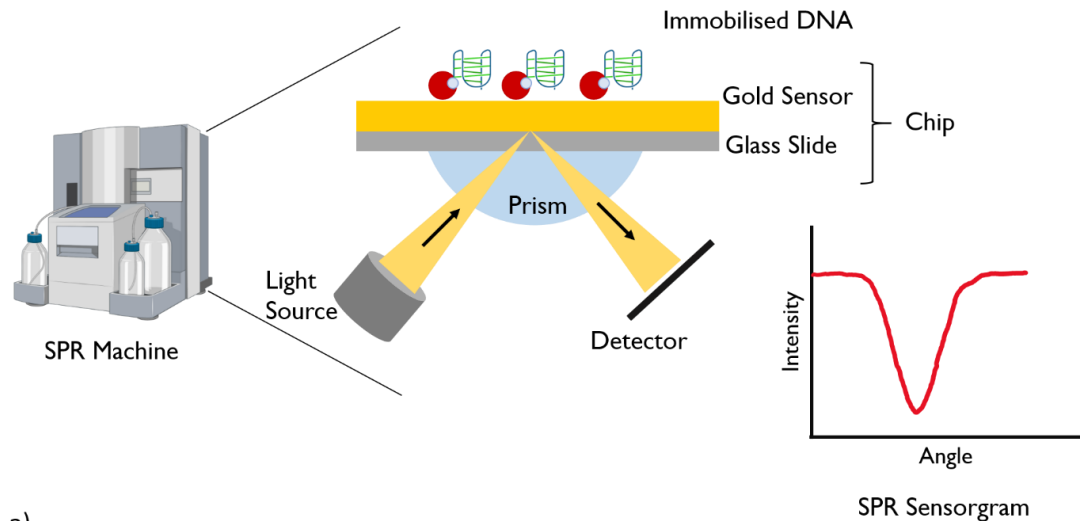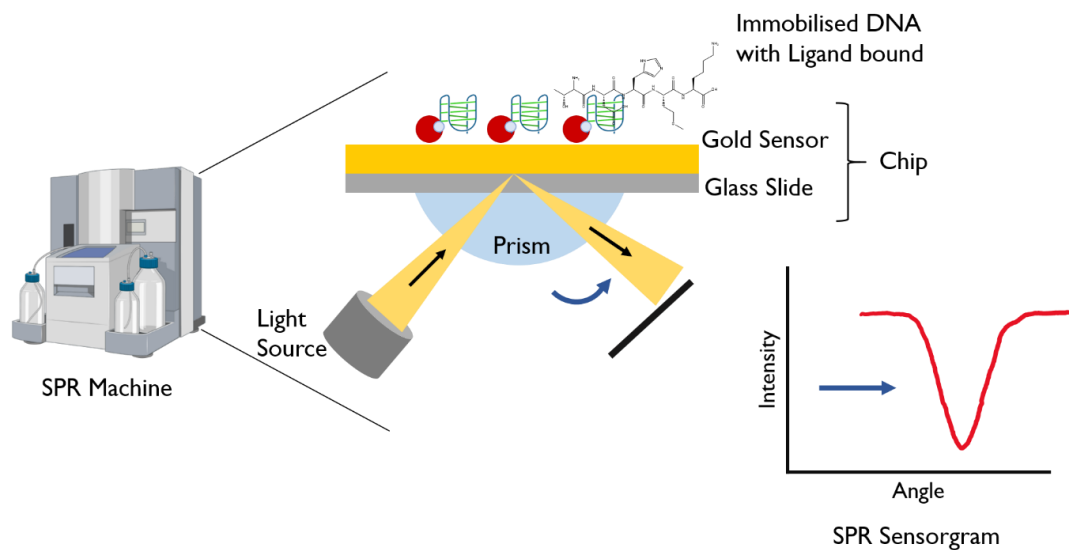
Figure 1.31: Schematic of how surface plasmon resonance works. Where polarised light from a source strikes a mental conducting sensor, causing plasmons to resonant, causing a change in intensity of light reflected at a single angle. Where the angle is directly linked to the mass of the surface of the chip. a) Demonstrates the angle of resonance and sensorgram when no ligand is bound. b) Demonstrates the angle of resonance and sensorgram when a ligand is bound, showing a change in resonance angle and a shift of the sensorgram indicated by blue arrows.

Utilising SPR technology for monitoring DNA-ligand interactions are advantageous as the it has high sensitivity permitting the use of low concentrations of DNA and ligand samples.[236] Which is especially advantageous for studying samples with poor solubility or environment conditions that cause aggregation as mentioned in *chapter 1.6.4*. In addition, SPR is time efficient, and data is collected in real-time. However, there are some disadvantages to be aware of including sophisticated and expensive equipment, use of labelling which can cause steric hinderance, immobilisation of the target which could disrupt interactions, and a high dependence on experimental conditions makes it difficult to compare results between different studies. Furthermore, maintaining different secondary DNA structures after immobilisation onto the chip is required and so storage conditions are extremely important.[236] Finally, there are drawbacks of using DMSO in the SPR system as it has a high-refractive index which can introduce significant error in SPR determinations and so calibrations are carried out to reduce this but DMSO free samples are preferred.[237] Overall, SPR is an excellent technique for obtaining binding affinity and kinetic data when studying DNA-ligand interactions.

1.6.6 Tetrazolium Dye Colorimetric Assay to Assess Cell Cytotoxicity

Tetrazolium dyes are colorimetric assays used to monitoring cell metabolic activity, reflecting the number of viable cells present.[238] Viable cells can reduce tetrazolium dyes using NAD(P)H-dependent oxidoreductase enzymes which causes a colour change that can be detected by a plate reader to determine absorbance at a particular wavelength. One common example is the MTT assay, where a yellow tetrazolium salt, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide known as MTT, is

reduced to an insoluble formazan which is a purple colour and quantified by measuring absorbance at 500-600 nm. The darker the purple coloured solution the greater number of viable, metabolically active cells are present (*figure 1.32*). MTT assays can be used for multiple purposes including quantification of cell growth and viability,[239] measuring cell proliferation,[238] measuring cytotoxicity,[240] and to study cell activation.[241]
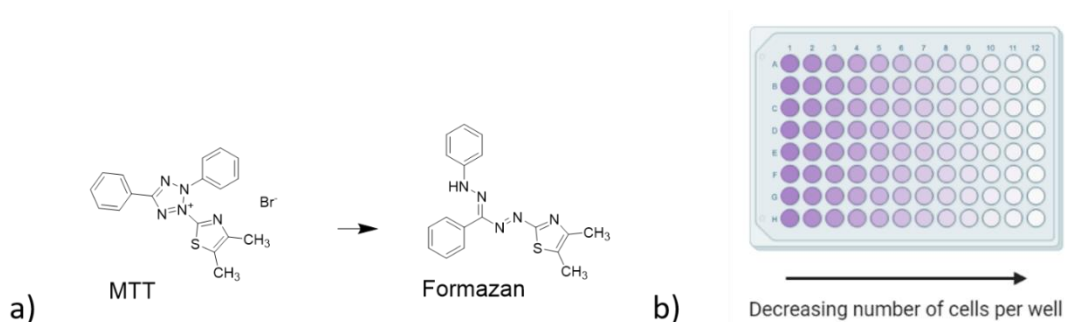
Figure 1.32. Metabolism of MTT to formazan salt by NAD(P)H-dependent oxidoreductase enzymes in viable cells. a) The chemical reaction b) The colour change seen in a 96 well plate with decreasing amounts of viable cells created with Biorender.com.

MTT is the gold standard for cytotoxicity testing, and it is suitable for high-throughput screening due to ease and rapidness of the assay which are advantageous.[242] Another advantage is that the protocol for MTT requires the compounds in media to be removed before adding the MTT to analyse the viability and thus coloured compounds will not affect the absorbance readings used to measure cell viability. However, the disadvantageous are conversion to formazan is dependent on metabolic rate and number of mitochondria resulting in multiple interferences and numerous wash steps.[242] Overall, this technique is excellent for determining what concentrations of ligands cause cell death.

1.7 Aims and Objectives

The overall aim of this project was to identify probes that bind to the i-motif that forms in the *MYC* promoter region. Work in *Chapter 2* aims to identify lead small compounds which can be further developed into c-MycC specific probes. The high-throughput method, fluorescence indictor displacement (FID) assay was used as a screening method followed up by further biophysical techniques. *Chapter 3* addresses this aim using the molecular biology technique phage display to identify peptides that bind c-MycC and uses biophysical data to further validate findings.

Chapter 2: Searching for Small Molecules to Target the i-Motif Forming Sequence in the Promoter Region of *c-Myc*

2.1 Introduction

MYC is a proto-oncogene which controls many cellular processes including cell cycle progression, proliferation, differentiation, and apoptosis.[121] Many cancers appear "hooked" on dysregulated *MYC* which isn't surprising considering its pivotal role in cellular control and involvement in not only tumour initiation but also tumour maintenance.[150] Unfortunately decades of research have found it extremely difficult to make advancements targeting c-Myc, partly due to our lack of understanding of the complexity of *MYC* regulation.[153] To begin to address this numerous studies have aimed to understand the dynamic nature of the nuclease hypersensitivity element (NHE) III_1 region as regulates ~90% of *MYC* transcription.[1] A clear understanding of how this region regulates transcription could be key to unearthing new effective cancer therapies.[153] This is a very real and attractive possibility. However, to understand the complexity of this regulation system an understanding of the following is required: the equilibrium of duplex, single-stranded, and non-canonical DNA structures that can form in the NHE III_1 region and the transcriptional factors controlling this process.

In an attempt to understand the full dynamics of NHE III_1 the following model was proposed by Hurley *et al* (*figure 2.1*).[3,120] The initiation of transcription induces negative superhelicity on the NHE III_1 region. To relieve this dynamic stress the single-stranded DNA forms a G-quadruplex on the purine-rich strand and i-motif on the pyrimidine-rich strand which will be respectively referred to as c-MycG and c-MycC throughout this

thesis. The formation of these non-canonical structures prevent transcriptional activators from binding and consequently transcription is not activated.[170,171]
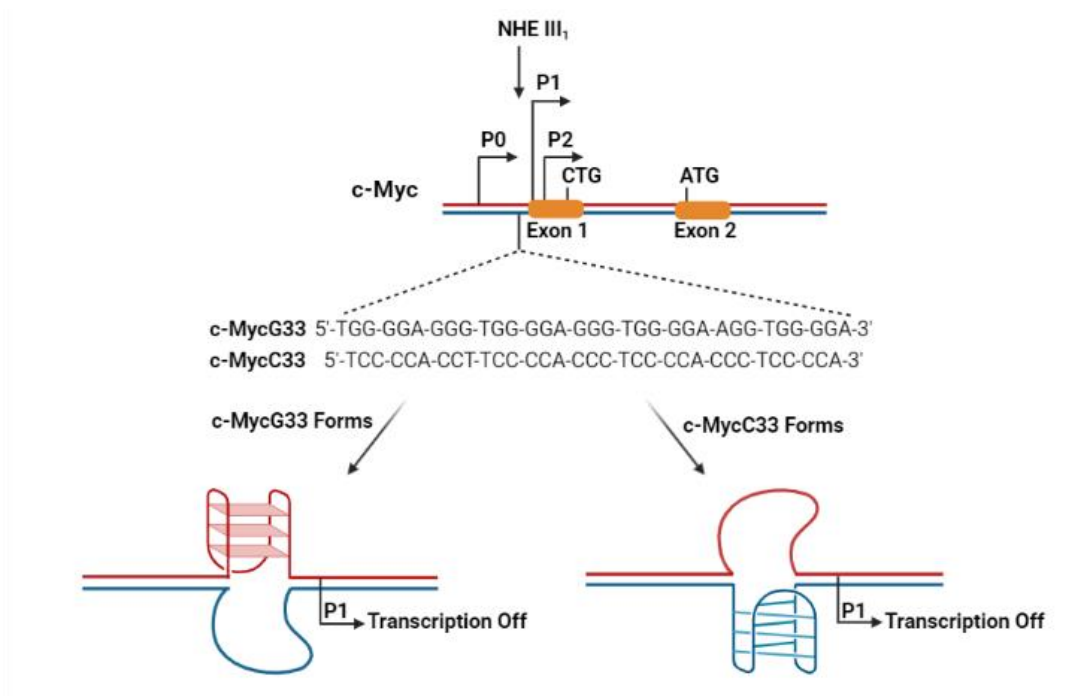
Figure 2.1: Proposed model for the effects of G-quadruplex and i-motif formation in the purine-rich and pyrimidine-rich strands respectively within the nuclease hypersensitivity element II₁ region of the MYC promoter. Created with BioRender.com adapted from Dash et al. 2020.[6]

This model is supported by a handful of studies that have shown that c-MycG destabilisation leads to an increase of *MYC* transcriptional activity.[170,171] Further studies into how c-MycG influences *MYC* activation have identified that nucleolin not only binds and stabilises c-MycG but can also facilitate its formation.[32] When nucleolin is overexpressed you see a decrease in *MYC* activation. It has also been identified that NM23-H2 binds and unfolds c-MycG causing an increase in MYC activation.[175] This model is supported by evidence from the i-motif field which has demonstrated decreased *MYC* transcription in tumor cells when using the ligand called **B19** (*figure 2.2)*, which shows preference for binding c-MycC and can induce its formation.[82] Another ligand of interest is **3be** (*figure*

1.18) which preferentially binds c-MycC over c-MycG and the i-motif that forms in the BCL2 promoter region whilst not appearing to bind to B-DNA.[6] Following this **3be**'s effects on 3 different cancer cell lines (HeLa, A549, B95.8) and a normal cell line (NKE) was evaluated. This study identified that **3be** inhibits cell proliferation in the cancer cell lines and also had low toxicity in the normal cell line.[6] It was also determined that **3be** induces apoptosis in HeLa cells in a dose-dependent manner. Even more significant was that **3be** was shown to be able to enter the cell nucleus and significantly repress transcriptional and translational levels of *MYC* expression in HeLa and B95.8 cell lines.[6] Therefore, this study also supports the NHE III_1 proposed model. A bisacridine compound, a compound class that are known DNA bis-intercalating agents, **a9** (*figure 1.18*), was demonstrated as a dual c-MycC and c-MycG binder that stabilises both quadruplex structures.[87] Furthermore, SPR data demonstrated that **a9** bound B-DNA but had greater binding affinity to c-MycC and c-MycG. Therefore, it would have been of interest to also explore how **a9** affected B-DNA stability however this was not determined. Interestingly, biological studies identified that **a9** downregulates MYC transcription and expression and induced apoptosis and cycle arrest in cancer cells suggesting it is an interesting prospect to be developed into a powerful therapeutic.[87] This study also provides further evidence that the proposed model is of G-quadruplex/i-motif interplay is correct.
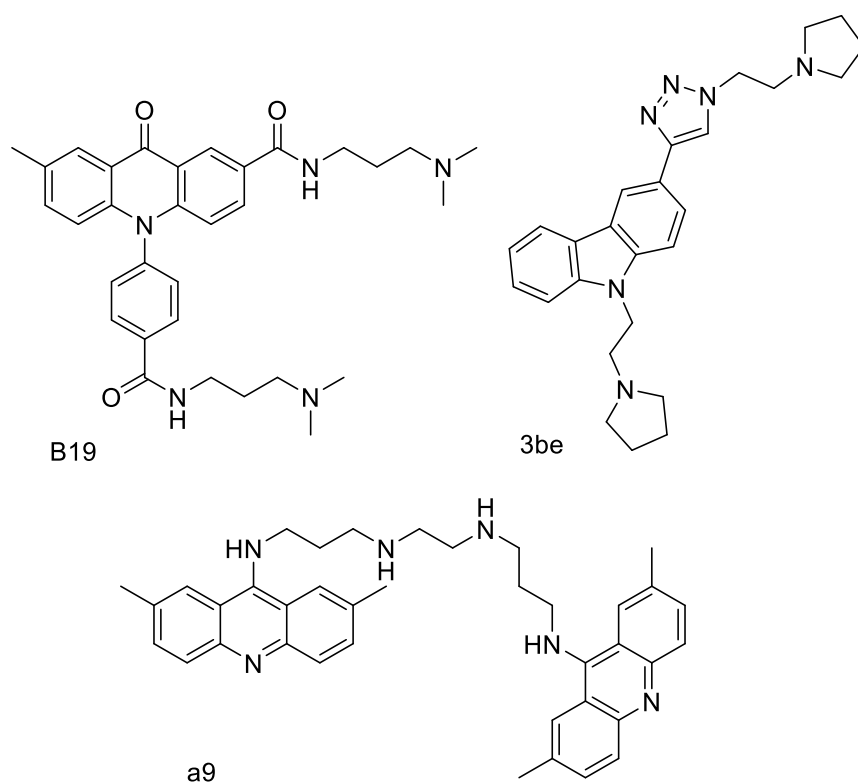
Figure 2.2: Structures of MYC i-motif specific ligands B19 and 3be and MYC dual-specific i-motif and G-quadruplex ligand a9.

Interestingly, **B19** has a acridone core, **3be** a carbazole core, and **a9** a bisacridine core which are similar structures (*figure 2.2*).[6,82,87] Notably, it was Dang Li's group that identified **B19** and then developed **a9** starting with the acridone core from **B19** to produce the ligand **a9**. Using acridone cores as a starting point for developing i-motif ligands is a popular route as the sister field has used them to develop G-quadruplex ligands as these are known to bind DNA structures.[213] Using carbazole cores is also a popular starting point as multiple have also been developed as G-quadruplex ligands.[243] The challenge is to develop specific probes based on carbazole and acridone cores is that they are DNA intercalators.[244,245] So far both the G-quadruplex and i-motif a have demonstrated carbazole and acridone core compounds binding to these two structures. Furthermore, compounds have been developed from these cores which showed a preference for a particular quadruplex structure over B-DNA.[6,87]

Interestingly, the G-quadruplex field have the addressed the challenge of specificity with acridine/acridone cores by using hybrid acridine/acridone-peptide molecules and find these hybrids to have the greater selectivity compared to the small molecules alone.[244] Overall screening carbazole and acridone libraries to gain a greater toolbox of quadruplex compounds which can be further developed into specific quadruplex probes is a popular route in the i-motif field.

2.2 Fluorescent Indicator Displacement Screening of National Cancer Institute IV Library

Despite the rapid growth of non-canonical DNA structures research, there are still limited reports in the literature of i-motif specific ligands and even fewer c-MycC specific ligands. And those that are documented generally show weak binding to other DNA structures and so are more selective rather than specific. Further investigation into compounds that bind to c-MycC is warranted. To identify lead small compounds which can be further developed into c-MycC specific probes a high-throughput technique that can screen large libraries of small compounds is needed.

As the literature is scarce for i-motif specific ligands let alone c-MycC specific ligands there is no clear concept of the structural features of a c-MycC specific compound and so high throughput screening of large libraries is appropriate to begin to build a database. Fluorescent indicator displacement (FID) was chosen to do this because it is cost-effective and fairly fast.

2.2.1 Fluorescent Indicator Displacement Method Development

In 2017 the Waller group developed an FID assay to identify new i-motif binding ligands.[85] FID works by using a compound that fluoresces when bound to DNA but not when it is free in solution. It was identified that the most suitable indicator molecule for FID was **thiazole orange (TO)** because it demonstrated the largest fluorescence enhancement upon binding hTeloC, compared to **ethidium bromide**, **acridine orange**, **crystal violet** and **pyrene derivative**. **TO** was the ideal indicator molecule because the large fluorescence enhancement was important as high levels of enhancement will improve signal:noise ratio, there were no fluorescent properties in the absence of DNA, which is important for having low background readings, and there was a low micromolar dissociation constant, which is required so that the fluorescent probe is easily displaced by other compounds binding. The Waller group used the FID assay as a high through put technique to determine new i-motif binding ligands. FID can be explained as three steps, 1) **TO** is in solution and there's no fluorescence, 2) DNA is added and **TO** fluoresces when it binds to the DNA, 3) a small molecule is added and if it displaces the **TO** so it is free in solution again there is a decrease in fluorescence (*figure 2.3).* This decrease in fluorescence can be calculated as percentage displacement and is used to determine if a molecule bind to the DNA, the higher the percentage displacement the greater amount of **TO** has been displaced and the better the relative binding affinity. Therefore, FID can determine if binding occurs but cannot provide an explanation of the interactions at the molecular level.
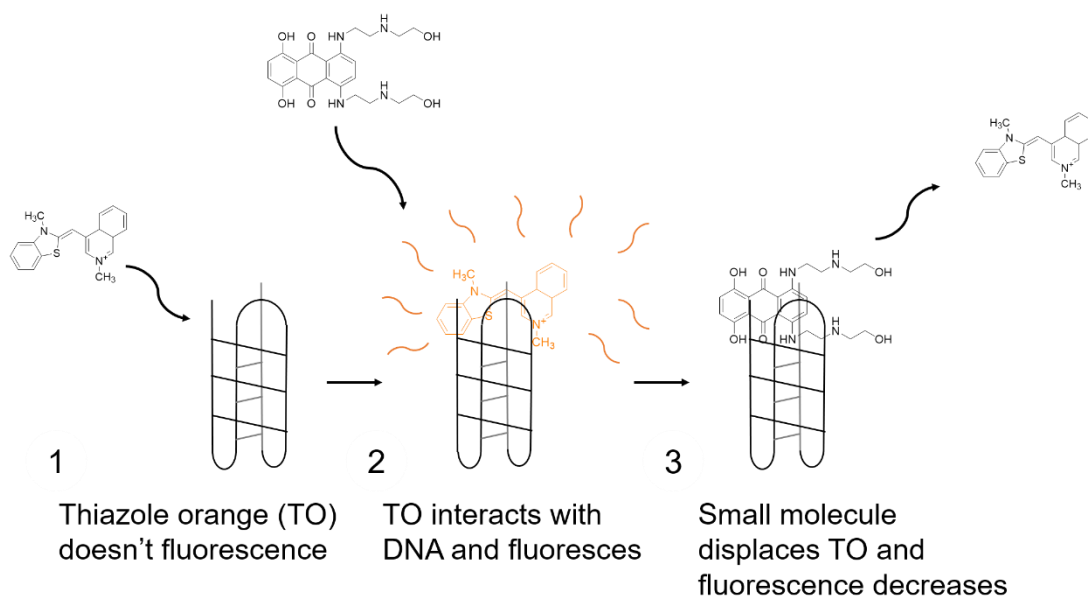
Figure 2.3: Schematic of fluorescence indicator displacement assay steps. 1) Thiazole orange is free in solution there is no fluorescence. 2) DNA is added to the solution the DNA-TO complex fluoresces. 3) A competitive molecule is added and if it displaces TO to form a DNA-molecule complex the indicator is free in solution again causing a decrease in fluorescence.

Prior to this förster resonance energy transfer (FRET) based melting was used as high-throughput screening method to identify alternative DNA structure ligands.[86,235] FID is advantageous over FRET because it is lower cost as labelled oligonucleotides are not required, the folding and stability of the structure is not affected by the label, and there are not restrictions on the lengths of oligonucleotides required due to spacing between donor and acceptor fluorophores.[85] Although, FID does not determine if a compound stabilises or destabilises a DNA structure and is based on the binding of a compound causing displacement of **TO**. This was not the first study to use an FID assay to determine compounds that bound to certain DNA structures, it had previously been used for G-quadruplexes,[246] triplexes,[247] hairpins,[248] and B-DNA,[249] but it was the first to use it for a high through-put screen against and i-motif. The study screened 960 different compounds and they investigated a handful of compounds as example hits using the FID screen, FRET, and SPR to validate the method.

This FID method was the basis for the FID studies carried out throughout this work.

To begin optimising the FID method, characterisation of c-MycC27 and c-MycC52 was carried out to identify the pH at which the experiment was to be performed. The method was to be performed at transitional pH, the pH at which you have a 50% folded (i-motif) and 50% unfolded (single-stranded) population. The transitional pH of c-MycC27 is well documented and known to be pH 6.6 from the literature.[250] The transitional pH of c-MycC52 was determined using circular dichroism (CD) as pH 6.64 (+/-0.04) as shown in *figure 2.4*. The CD spectrum for the oligonucleotide sequence is ran at a range of pH values usually pH 5.0-8.0 for i-motifs, where the sample is expected to be fully folded at pH 5.0 and fully unfolded at pH 8.0. The ellipiticity of the characteristic dominant positive i-motif peak ~290 nm is plotted at each pH value and fitted with a dose response curve of which the transitional pH can be extrapolated from the graph using finding x from y.[115] Therefore, the pH chosen was pH 6.6 to ensure both c-MycC strands were 50% folded.
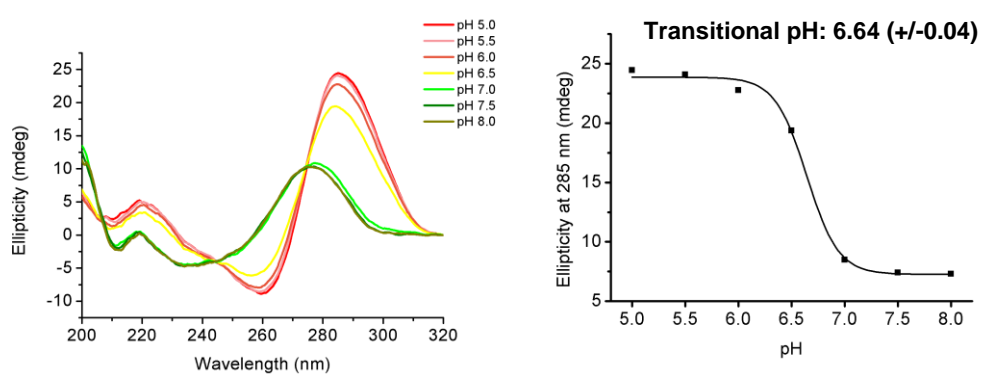
Figure 2.4: Circular dichroism characterisation of c-MycC52 in 10 mM NaCaco from pH 8.0 to 5.0 to determine the transitional pH, the pH at which the sequence is 50% folded (i-motif) and 50% unfolded (single-stranded). The graph of the ellipticity of the dominant characteristic MYC i-motif peak, 285 nm, against pH was fitted using a single sigmoidal fitting which was used to extrapolate the transitional pH in origin software to determine the transitional pH where the error is that of the line fitting.

The FID protocol also required optimisation on the duration of the equilibration time of the i-motif and **TO** binding, this is a crucial step because maximum fluorescence needs to be reached before addition of the ligang, otherwise you are unable to determine the change in fluorescence percentage. For example, if you added a ligand when only 50% of the **TO** was bound to the DNA more **TO**-DNA complexes could form and you may see an increase in relative fluorescence units (RFU). It is also possible you would see no change and determining no binding when actually the ligand has bound to the DNA in the free binding sites where **TO** has not bound. Therefore, this equilibration step is of importance and can vary between different structures. To determine how long the equilibration time required **TO** and DNA were mixed for 30 seconds and then the fluorescence was measured over 30-40 minutes, equilibrium time was determined as where RFU plateaued. The equilibrium time was determined as 10 minutes for both c-MycC27/52 (*figure 2.5*). Notably, c-MycC27 has a much lower fluorescent response compared to c-MycC52. For the more detailed method refer to *Section 5.1*.
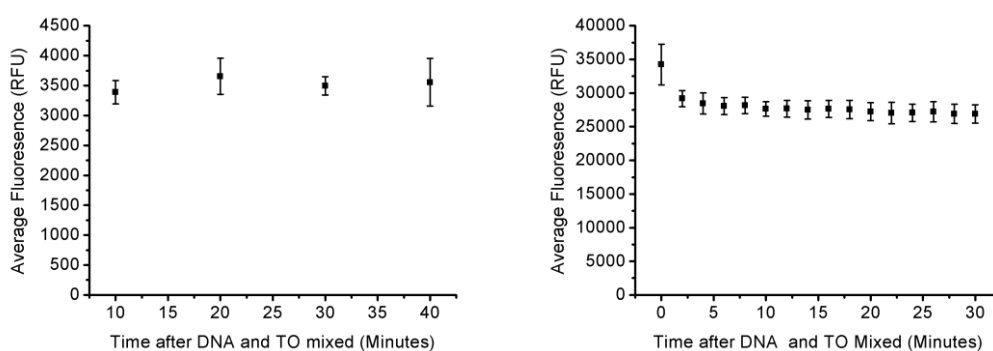
Figure 2.5: Monitoring relative fluorescence units over time after DNA and thiazole orange have been mixed for 30 seconds for c-MycC27 (left) and c-MycC52 (right) in 10 mM NaCaco pH 6.6.

2.2.2 Screening Compound Library Against c-MycC27 vs c-MycC52 with Fluorescent Indicator Displacement Assay

For this study the national cancer institute (NCI) IV Library Diversity Set was chosen as it represents maximal structural diversity and a wide range of potential pharmacophores containing 1,584 different known compounds.[251] This library was screened against c-MycC27 and c-MycC52 using 1 µM DNA, 2 µM TO and 2.5 µM of compound. In this work, we maintained >15% displacement as the hit threshold above which a compound is denoted as a hit to i-motifs (*figures 2.6 and 2.7*).[85] Furthermore, compounds that had <-15% displacement were also explored as potential hits, the rationale for this is explained and discussed below. Thus, either 15% or <-15% displacement demonstrated binding to i-motif DNA.
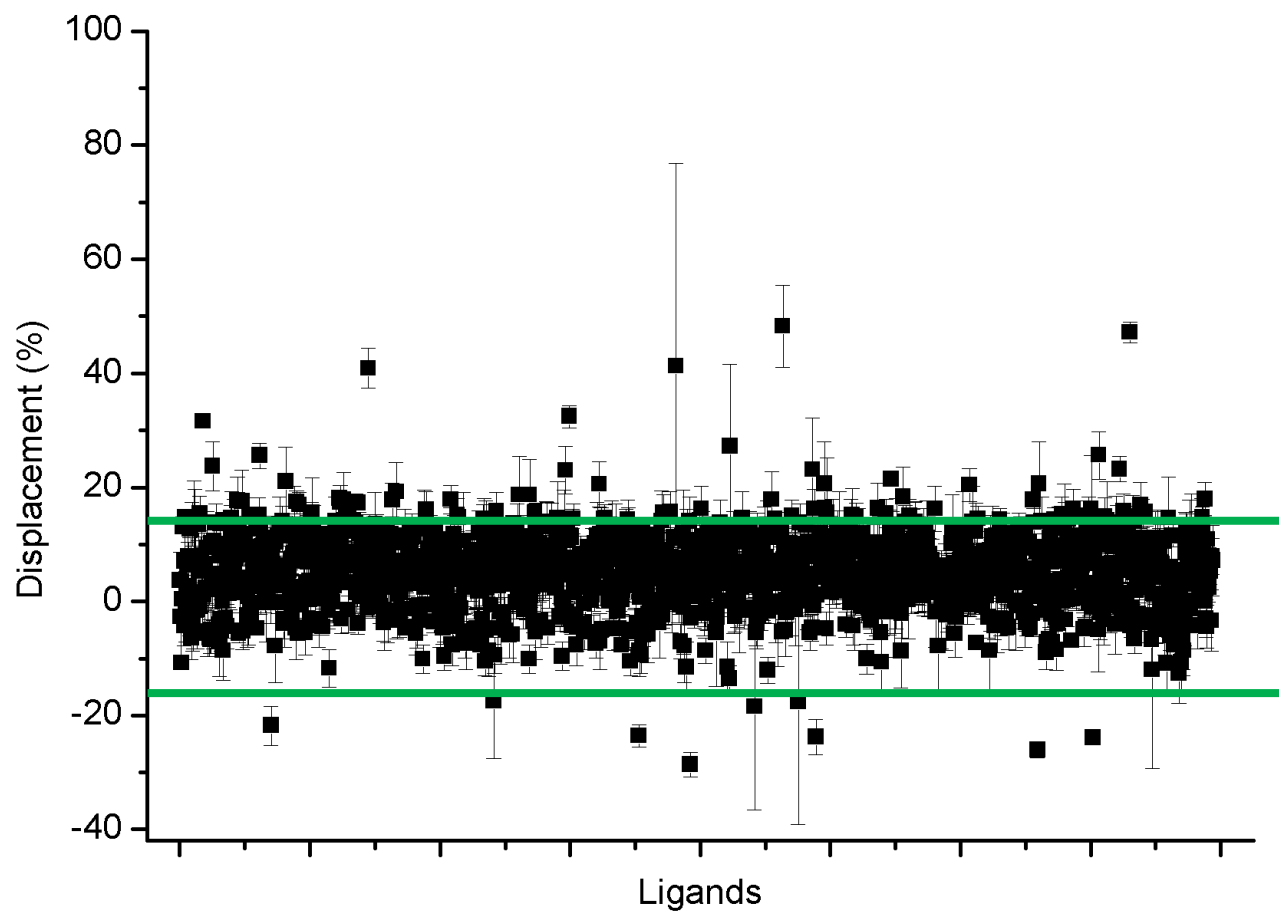
Figure 2.6: Fluorescent indicator displacement assay screen of National Cancer Institute Diversity Library IV against c-MycC27. The screen was ran using 1 µM DNA, 2 µM thiazole orange and 2.5 µM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15% these two thresholds are shown by the green lines. Error bars are standard deviation for ≥2 technical replicates.
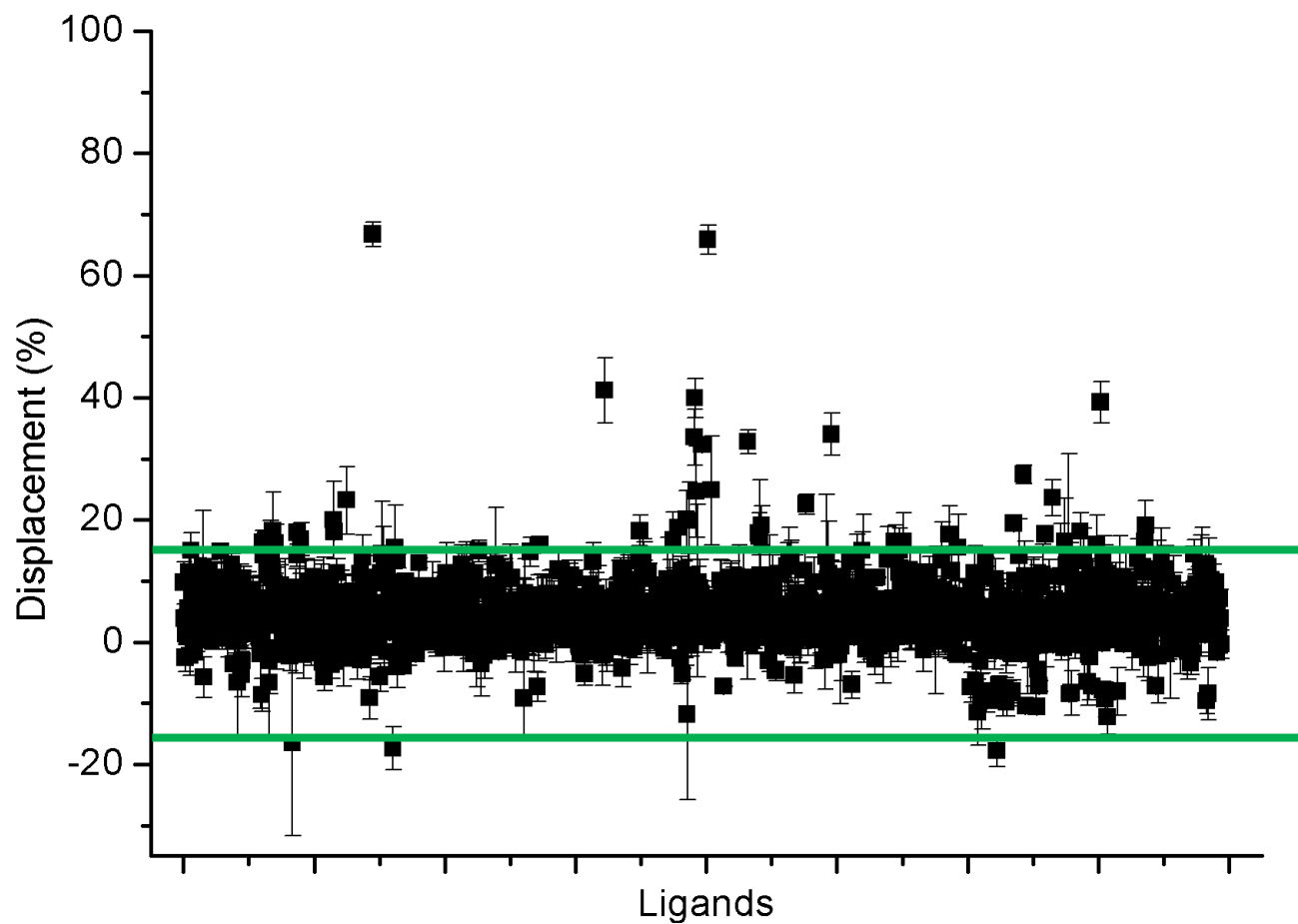
Figure 2.7: Fluorescent indicator displacement assay screen of National Cancer Institute Diversity Library IV against c-MycC52. The screen was ran using 1 μM DNA, 2 μM thiazole orange and 2.5 μM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15% these two thresholds are shown by the green lines. Error bars are standard deviation for ≥2 technical replicates.

Interestingly for c-MycC27, as shown in *figure 2.6,* there are multiple negative displacement values. One could expect to see some negative values near 0% and within error of 0% as no binding is taking place and this is error of the experiment, similarly to that seen for c-MycC52 shown in *figure 2.7*. However, c-MycC27 has multiple negative displacements smaller than -15%. As discussed previously in *Section 2.2* this could be due to relative fluorescence increases as equilibrium was not achieved, however *figure 2.5* shows this should not be the case. It is however, more likely that the compounds are also fluorescent at the same wavelength being read 450 nm and so there is an increase in fluorescence which would cause a negative displacement result. This highlights that this assay is not ideal for compounds which fluoresce at 450 nm as it affects the analysis, and one is unable to identify what is happening at the molecular level using FID to understand the cause of increased fluorescence. Which can be assessed by checking the compounds fluorescence. However, it is interesting that the Waller group had only observed this on rare occasions before during FID experiments screening the exact same library against other i-motif sequences. Which is further supported when looking at c-MycC52 (*figure 2.7*). Negative displacement could have indicated that the compound was shifting the equilibrium towards i-motif formation and thus there were more binding pockets available for **TO** to bind causing an increase in fluorescence. Therefore, these compounds showing a significant negative displacement (<-15%) were deemed of interest and a hit. Notably, NCS **129929**, *figure 2.8*, had a negative displacement of -152.2 (+/-8.7) this did not fit on the axis shown in *figure 2.6* and so is not shown. This displacement is much smaller than any of the other compounds and was found to be fluorescent. All of the other compounds with a displacement of <-15% were taken forward as hits for c-MycC27.
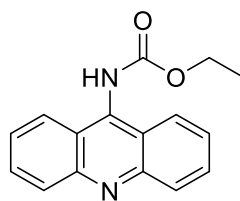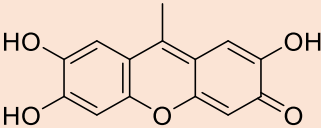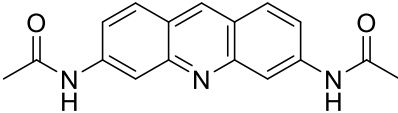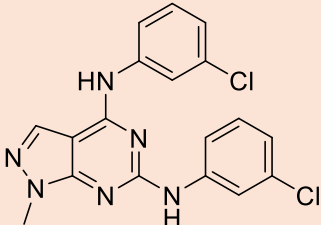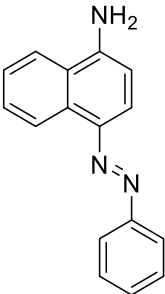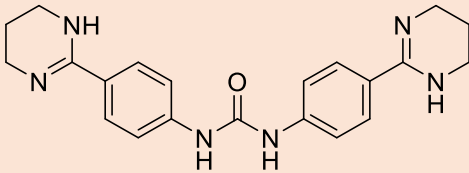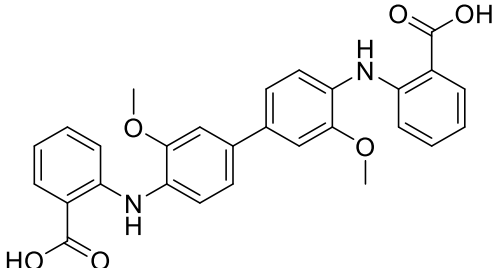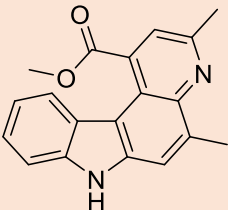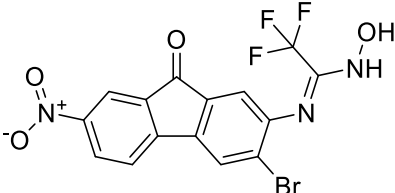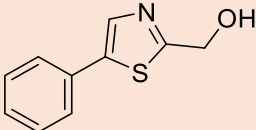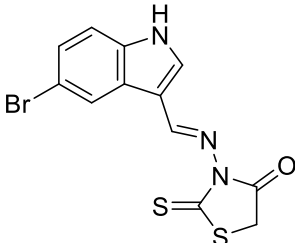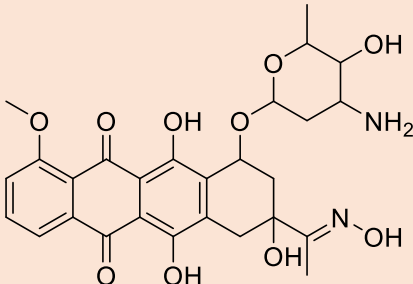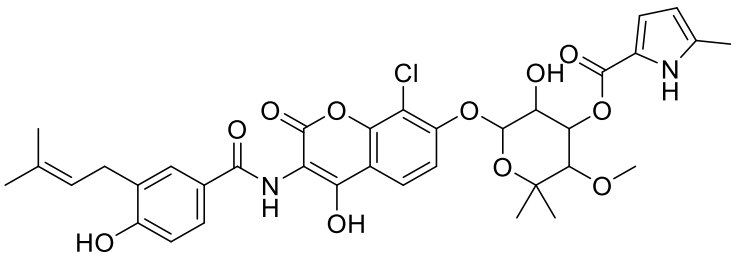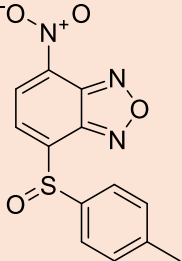
Figure 2.8: Structure of NSC 129929 compound.

Out of 1,584 different compounds screened 18 were identified as hits (>15% or <-15% displacement) binding to c-MycC27 and 20 to c-MycC52, four of which bound to both c-MycC structures. In total there were 34 different compounds identified as hits (*Table 2.1 and 2.2*). The four compounds that both c-MycC sequences are **7218**, **63680**, **143491**, and **309401**, where **143491** shows the greatest percentage displacement for both c-MycC sequences. The compounds **7218**, **63680**, and **309401** all show negative displacement against c-MycC27 and positive displacement against c-MycC52. The five ligands with the greatest percentage displacement values from highest to lowest are **143491**, **317003**, **309401**, **7218**, and **36758** for c-MycC52 and **345647**, **143491**, **260594**,**105781**, and **11668** for c-MycC27. The five ligands with the lowest percentage displacement values from lowest to highest are **63680**, **33353**, **30205**, **335504**, and **60339** for c-MycC52 and **309401**, **63680**, **7218**, **366289**, and **288387**. Notably, there are seven compounds that bind c-MycC52 with a displacement of over 30%, which is 35% of the compounds identified as c-MycC52 hits, whereas for c-MycC27 there are four compounds that displace 30% of TO which is 22% of compounds identified as c-MycC27 hits. Further, six of the c-MycC27 hits gave a negative displacement smaller than 15% which is a third of the compounds determined as c-MycC27 binders.

Table 2.1: Fluorescent indicator displacement assay screen hit results of National Cancer Institute Diversity Library IV against c-MycC27. The screen was ran using 1 μM DNA, 2 μM thiazole orange and 2.5 μM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15%. Error is standard deviation for ≥2 technical replicates.

Ligand (NSC)	Structure	c-MycC27 % Displacement
5426		20.4 ± 2.9
7218		-23.9 ± 1.2
11668		25.6 ± 2.2
13974		17.3 ± 1.3
63680		-26.1 ± 1.4
73735		25.6 ± 4.1

105781		31.6 ± 0.5
107582		23.7 ± 4.4
118832		-21.8 ± 3.4
124818		17.4 ± 2.0
143491		40.9 ± 3.5
227186		23.0 ± 4.2
228150		32.4 ± 2.0
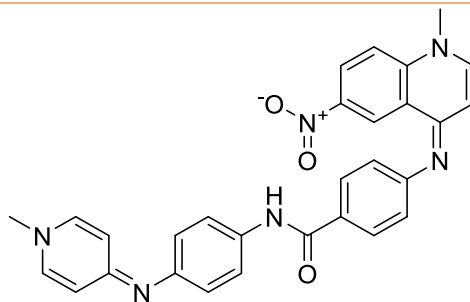
| 260594 | 20.5 ± 4.0 |

| 288387 | -23.6 ± 2.0 |

| 309401 | -28.6 ± 2.1 |

| 345647 | 48.3 ± 7.2 |

| 366289 | -23.8 ± 3.1 |
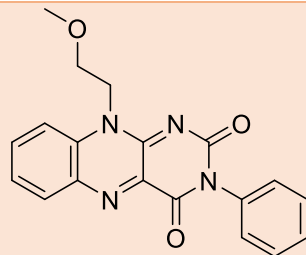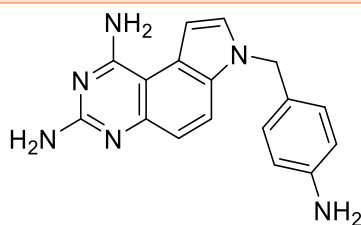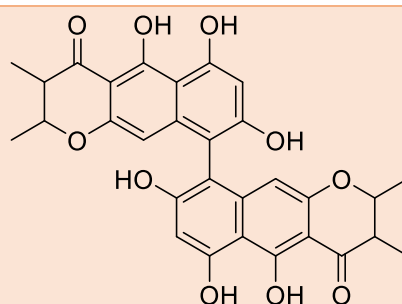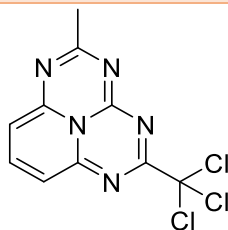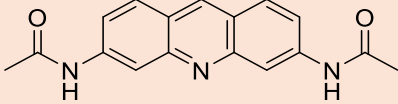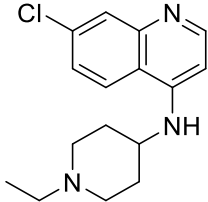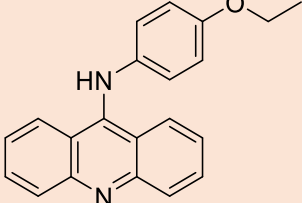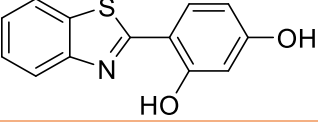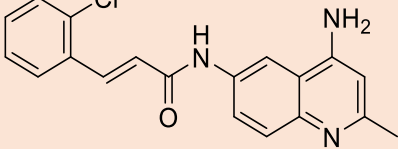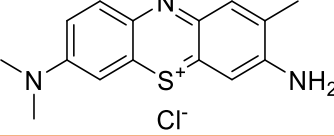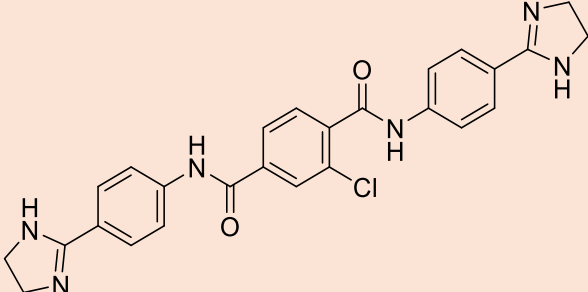
Table 2.2: *Fluorescent indicator displacement assay screen hit results of National Cancer Institute Diversity Library IV against c-MycC52. The screen was ran using 1 µM DNA, 2 µM thiazole orange and 2.5 µM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15%. Error is standard deviation for ≥2 technical replicates.*

Ligand (NSC)	Structure	c-MycC52 % Displacement
7218		39.3 ± 3.4
13248		20.0 ± 6.3
30205		18.7 ± 0.7
33005		32.8 ± 2.0
33353		17.7 ± 3.5
36758		34.0 ± 3.5
60339		19.4 ± 0.6

61642	27.5 ± 1.5
63680	17.6 ± 0.5
135168	23.2 ± 5.5
143491	66.7 ± 2.0
305780	20.1 ± 4.8
308848	39.9 ± 3.2
309401	24.6 ± 7.4
311153	32.3 ± 0.0
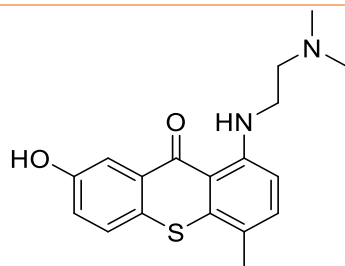
317003		65.9 ± 2.4
317605		24.8 ± 8.9
335504		19.1 ± 3.2
354844		22.5 ± 1.6
638432		23.6 ± 2.9
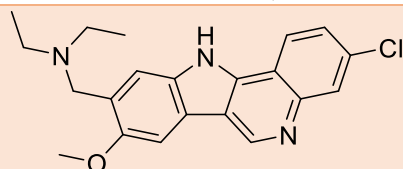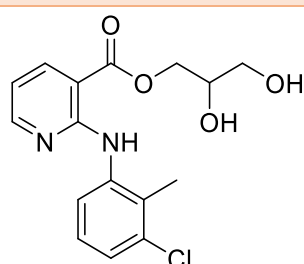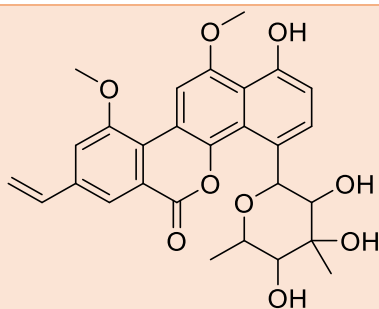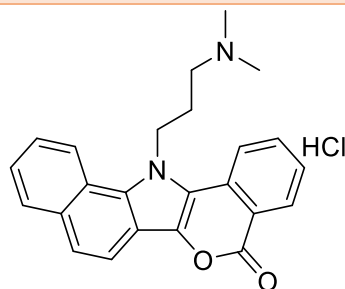
Before moving on one must explore reasons why there are significant (<-15%) negative displacements for c-MycC27. To begin to understand this the raw data for each stage must be analysed. The background fluorescence of this assay (**TO** and buffer only) is ~200 RFU, which is normal and the same as what was seen during assay development by the Waller lab.[85] The Waller group observed fluorescent enhancement of ~59 fold increase (~11800 RFU) when the **TO**-i-motif complex formed. The fluorescent enhancement seen when the **TO**-c-MycC27 complex forms is a modest ~17.5 fold (~3500 RFU) (*figure 2.5*). In contrast, c-MycC52 sees an enhancement of ~150 fold (~30000 RFU) when the **TO**-c-MycC52 complex forms, which is significantly more (*figure 2.5*). Therefore, for c-MycC27 any small fluorescence changes have a greater effect and thus greater error is expected than compared to the when the assay was developed. Whereas c-MycC52 will be impacted less by small fluorescent changes and thus the error of the experimental method itself will have a much lesser effect than compared to c-MycC27 and the assay development. Thus, it is reasonable that the compounds showing a displacement of <-15% for c-MycC27 and positive displacement values for c-MycC52 are indeed fluorescent but the RFU of that fluorescence is minimal compared to the ~30000 RFU observed for the **TO**-c-MycC52 complex and so does not affect the assay to the same extent. It must be noted here that fluorescent compounds will still affect this assay for example a compound could displace **TO** causing a 20% displacement but the ligand itself fluoresces at a -10% displacement and so the total displacement observed would be 10%, therefore a hit could be missed. Unfortunately, this is one downside of this screening technique as it is not ideal for compounds that fluoresce at 450 nm. It should also be highlighted that the low fluorescent enhancement from **TO** only to the **TO**-c-MycC27 complex has a low signal to noise ratio accounting for the large error bars seen for c-MycC27 data as demonstrated in *figure 2.5*. This led to challenges in large variations of percentage displacement shown as assay, machine, and human error had a more significant effect on RFU values causing large error bars as seen in *figure 2.6*. Overall, FID was a

useful tool for high throughput screening to identify compounds that bind to c-MycC.

2.2.3 Refining Screening Hits by Removing Double Stranded Binders

In total there were 34 different compounds identified as hits for c-MycC27 and c-MycC52 (*Table 2.1 and 2.2*). To reduce the number of compounds under investigation hits were screened against B-DNA to exclude non-specific DNA binders. The B-DNA sequenced used cannot form secondary structures and is in the double-stranded right-handed helical structure in the conditions used for the FID assay.

Before carrying out the FID assay against B-DNA an equilibration experiment of the TO-B-DNA complex was carried out and determined an equilibration time of 10 minutes was required (*figure 2.9*).
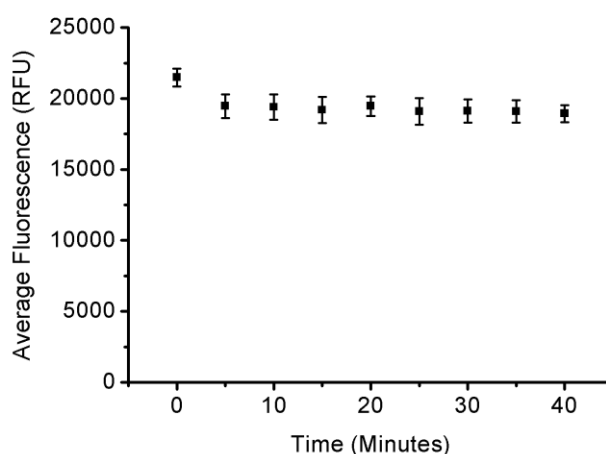
Figure 2.9: Monitoring relative fluorescence units over time after B-DNA and thiazole orange have been mixed for 30 seconds in 10 mM NaCaco pH 6.6.

To screen against B-DNA the FID assay was performed under the same assay conditions to allow valid dataset comparisons. Screening of the 34 compounds identified 11 compounds had >15% displacement and thus were determined as B-DNA binders *(figure 2.10)*. Therefore, these 11 compounds were disregarded to focus on specific c-MycC probes. This screening identified ~one third of compounds that bound to either c-MycC structure also bound B-DNA.
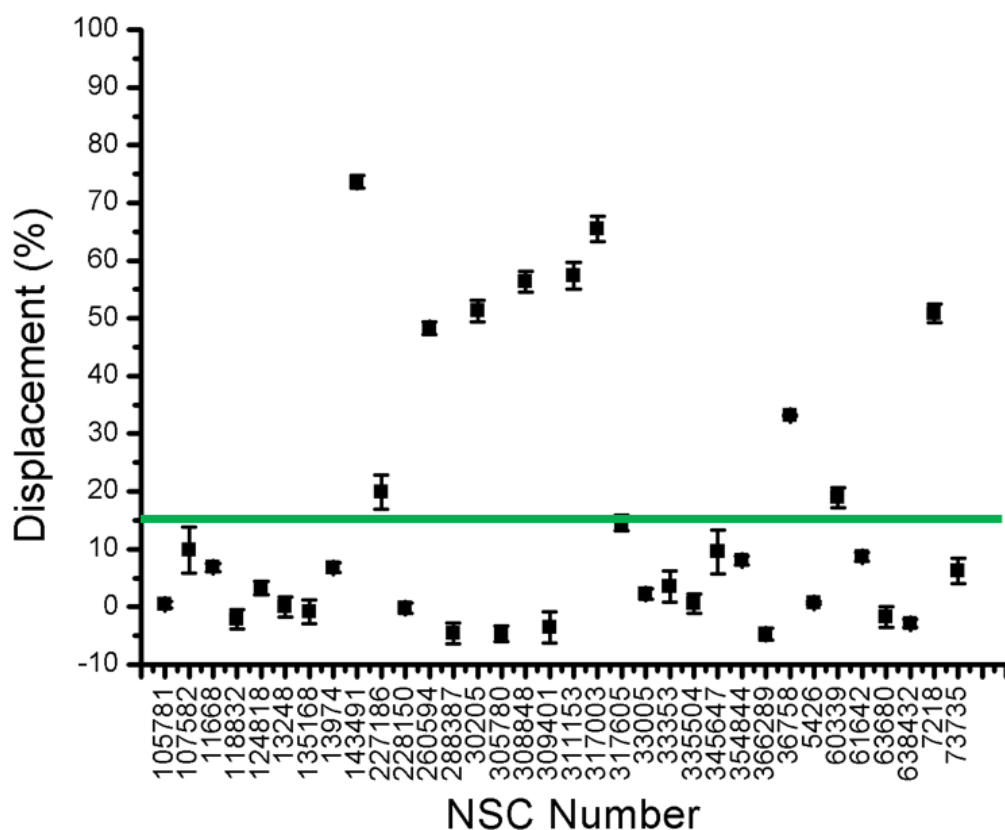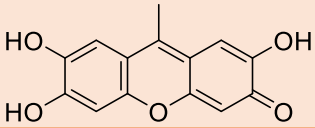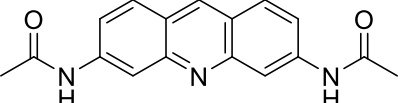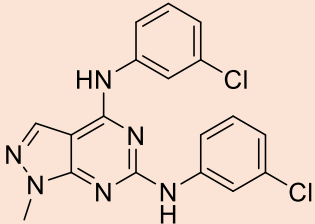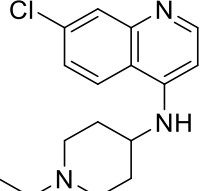
Figure 2.10: Fluorescent indicator displacement assay screen of National Cancer Institute Diversity Library IV of compounds c-MycC27 and c-MycC52 binding compounds against B-DNA. The screen was ran using 1 μM DNA, 2 μM thiazole orange and 2.5 μM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15% these two thresholds are shown by the green lines. Error bars are standard deviation for ≥2 technical replicates.
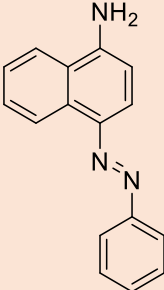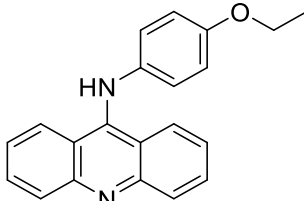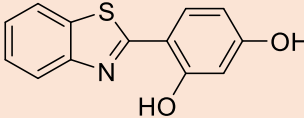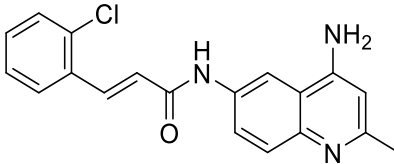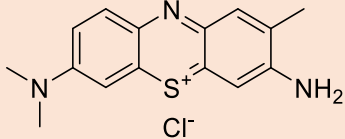
It's unsurprising that a ~one third of the compounds also bound to B-DNA as one of the problems the i-motif field face is identifying specific probes.[252]

This is not a unique to the i-motif field to have problems identifying specific probes it is also a challenge the G-quadruplex field face as many of the probes believed to be G-quadruplex specific have been shown to also bind i-motif DNA.[74,75] One reason it is hard to identify compounds that bind specifically to one DNA secondary structure is because they have the same four building blocks: Adenine, Thymine, Guanine, and Cytosine, attached to the phosphate backbone and so chemical structure wise there are similarities. Moreover, many B-DNA binders intercalate between the planar bases of the DNA,[253] a few well known DNA intercalators are **ethidium bromide**, **berberine**, and **doxorubicin**.[254,255] Yet B-DNA is not the only structure that has planar bases pairs. Other DNA secondary structures that have planar bases include G-quadruplex,[256] hairpins,[257] and even i-motif DNA.[258] Therefore, intercalators can bind non-specifically to different DNA structures. In addition, the i-motif C-C$^+$ base pairs are intercalated leading to a structure with major (wide) and minor (narrow) grooves. However, these differ from those seen in B-DNA which rise from the helical folding, the tertiary structure, causing the formation of two grooves, whereas for i-motif DNA these two grooves are due to the secondary structure.[259] Furthermore, the minor groove of the i-motif is extremely narrow, 3.1 Å, compared to the minor grooves in B-DNA, ~5.7 Å.[51] Moreover, the i-motif minor groove shows a high degree of hydrophobicity which reduces backbone repulsion to stabilise the structure, whereas ligands binding B-DNA are known to have cooperative effects with water.[259] This may suggest that ligands binding to the minor grooves of the i-motif or B-DNA are more likely to show DNA structure specificity compared to intercalators. The takeaway point is the interaction of ligands with i-motifs can be from intercalation, binding to the loops, and binding to the major or minor grooves. B-DNA can also interact with ligands via intercalation and groove binding, however the structural differences suggest specificity can be obtained for i-motif specific probes.
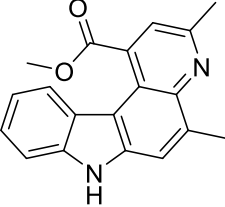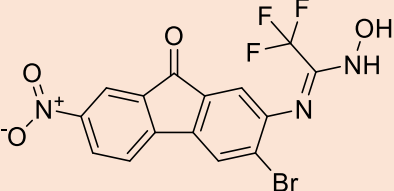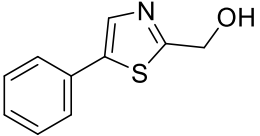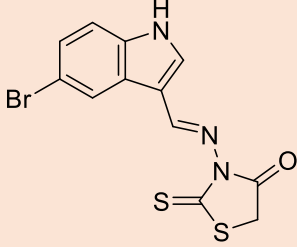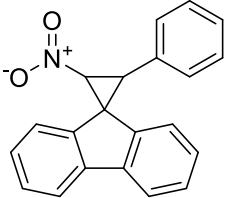
To summarise, 1,584 different NCI diversity set IV compounds were screened against the two different *MYC* i-motif sequences, c-MycC27 and c-MycC52, and 34 were identified as hits binding to either or both structures. FID hits studies were performed against B-DNA which revealed one third of the hits to be non-specific. The 11 compounds that were identified as binding B-DNA were: **7218**, **30205**, **33353**, **36758**, **60339**, **143491**, **227186**, **260594**, **308849**, **311153**, and **317003** (*table 2.3).* Leaving 23 hits to investigate further (*table 2.3*).
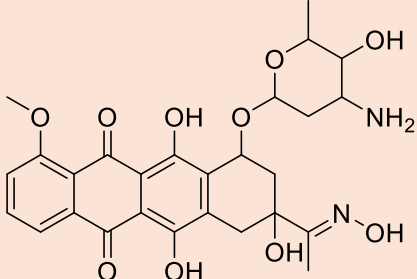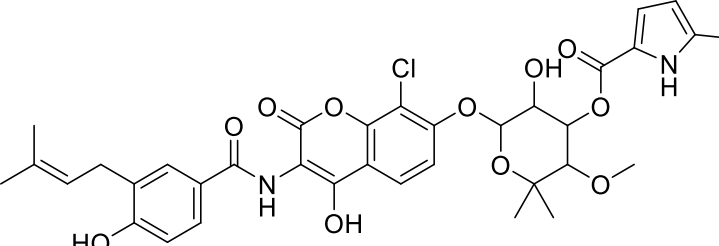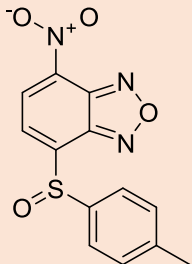
Table 2.3: Fluorescent indicator displacement assay screen of c-MycC National Cancer Institute Diversity Library IV hits against c-MycC27, c-MycC52, and B-DNA. The screen was ran using 1 μM DNA, 2 μM thiazole orange and 2.5 μM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15%. Error is standard deviation for ≥2 technical replicates.

Ligand (NSC)	Structure	% Displacement		
		c-MycC27	c-MycC52	B-DNA
5426		20.4 ± 2.9	8.8 ± 1.4	0.7 ± 0.3
7218		-23.9 ± 1.2	39.3 ± 3.4	50.9 ± 1.6
11668		25.6 ± 2.2	14.2 ± 3.5	6.8 ± 0.7
13248		-11.7 ± 3.3	20.0 ± 6.3	0.0 ± 1.7
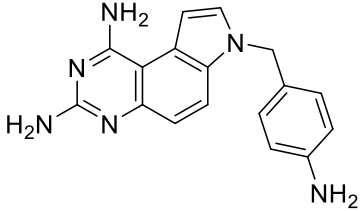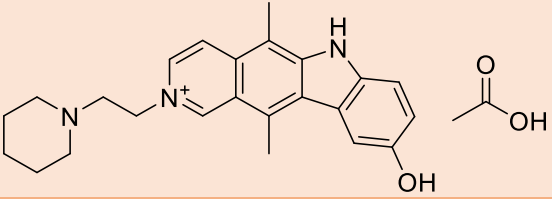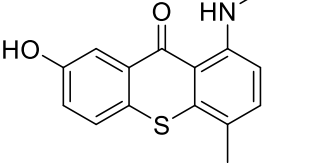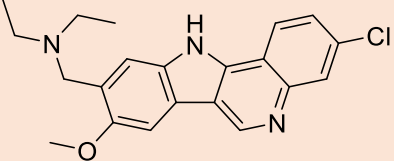
13974		17.3 ± 1.3	13.8 ± 3.7	6.8 ± 0.8
30205		-0.2 ± 1.6	18.7 ± 0.7	51.3 ± 1.9
33005		0.5 ± 2.0	32.8 ± 2.0	2.3 ± 0.9
33353		11.2 ± 0.3	17.7 ± 3.5	19.3 ± 3.3
36758		14.4 ± 1.3	34.0 ± 3.5	33.2 ± 0.2
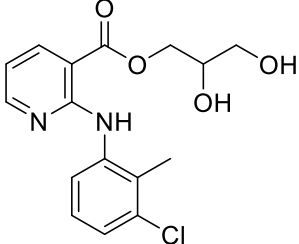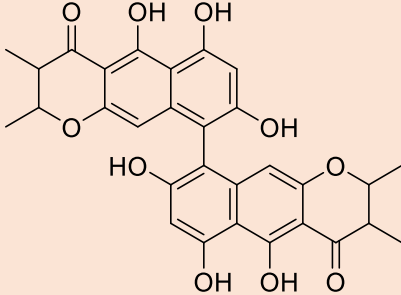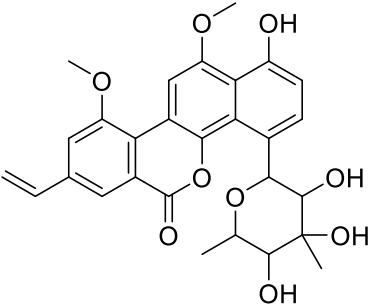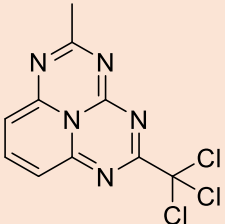
60339		-3.2 ± 2.7	19.4 ± 0.6	18.9 ± 1.7
61642		10.7 ± 3.4	27.5 ± 1.5	8.7 ± 0.8
63680		-26.1 ± 1.4	17.6 ± 0.5	-1.7 ± 1.8
73735		25.6 ± 4.1	6.1 ± 0.5	6.3 ± 2.2

105781		31.6 ± 0.5	10.6 ± 2.5	0.4 ± 0.6
107582		23.7 ± 4.4	9.6 ± 1.0	9.9 ± 3.9
118832		-21.8 ± 3.4	16.2 ± 3.0	-2.1 ± 1.6
124818		17.4 ± 2.0	16.8 ± 2.7	3.3 ± 1.2
135168		3.8 ± 0.2	23.2 ± 5.5	-0.9 ± 2.0

143491		40.9 ± 3.5	66.7 ± 2.0	73.7 ± 1.1
227186		23.0 ± 4.2	10.9 ± 0.6	19.9 ± 2.9
228150		32.4 ± 2.0	10.4 ± 2.9	-0.2 ± 0.9

260594		20.5 ± 4.0	13.6 ± 5.3	48.3 ± 1.1
288387		-23.6 ± 2.0	10.4 ± 3.0	-4.6 ± 1.8
305780		-7.1 ± 3.0	20.1 ± 4.8	-4.6 ± 1.3
308849		9.2 ± 2.7	39.9 ± 3.2	56.3 ± 1.8

309401	-28.6 ± 2.1	24.6 ± 7.4	-3.5 ± 2.7
311153	-9.0 ± 2.6	32.3 ± 0.0	57.4 ± 2.3
317003	4.1 ± 4.0	65.9 ± 2.4	65.5 ± 2.2
317605	-8.6 ± 2.2	24.8 ± 8.9	14.5 ± 1.3

335504		2.6 ± 1.9	19.1 ± 3.2	0.6 ± 1.6
345647		48.3 ± 7.2	13.6 ± 5.2	9.5 ± 3.8
354844		13.7 ± 3.4	22.5 ± 1.6	8.0 ± 0.8
366289		-23.8 ± 3.1	2.0 ± 1.1	-4.7 ± 1.0
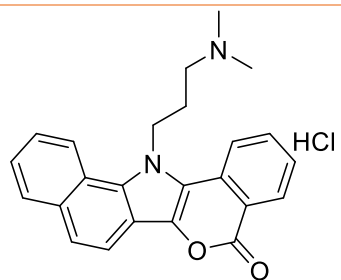
638432   HCl   -1.6 ± 6.2   23.6 ± 2.9   -2.9 ± 0.7
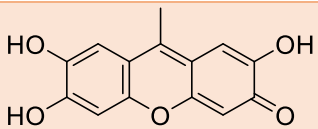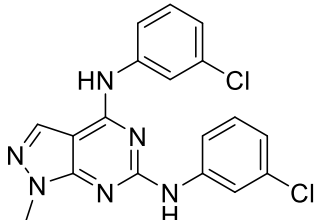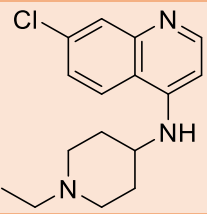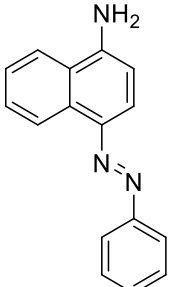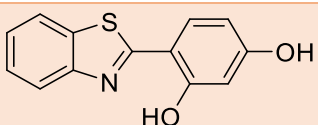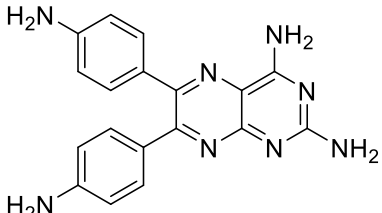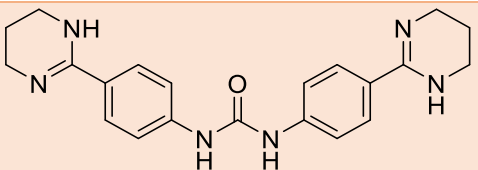
2.2.4 Validation of c-MycC27 and c-MycC52 Hits
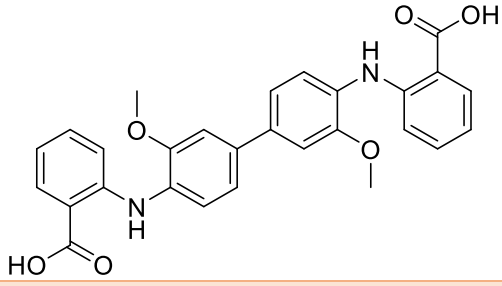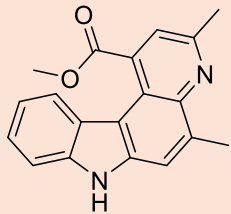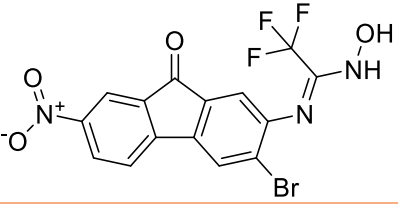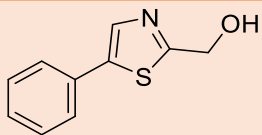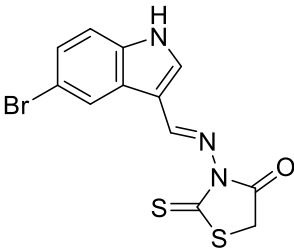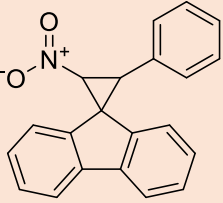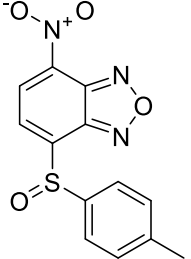
From the FID screen there were 23 identified hits which bound either c-MycC structure and did not bind to B-DNA (*table 2.3*).
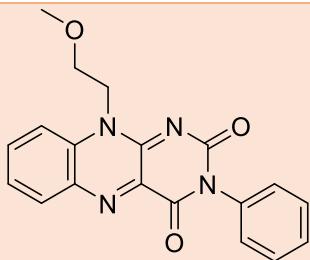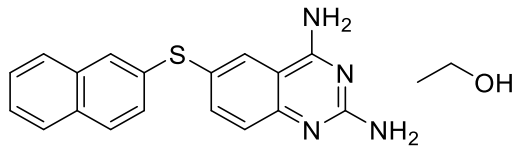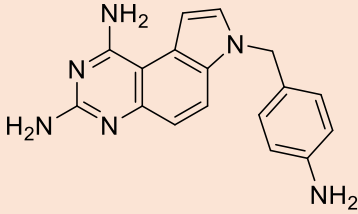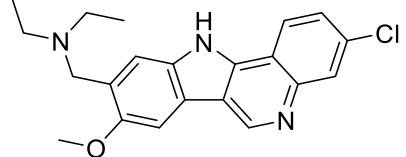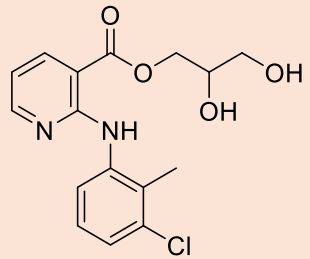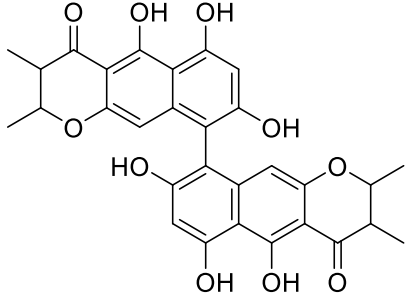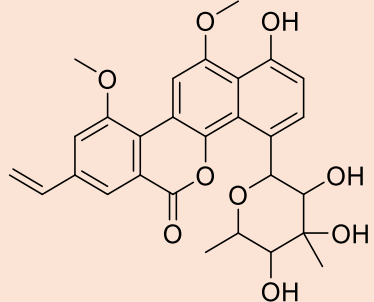
The FID assay was repeated for all 23 hit compounds using fresh batches of each compound against c-MycC27, c-MycC52, DS, and c-MycG. The rationale for repeating the FID against the c-MycC sequences was to ensure no issues from cross contamination or degradation.
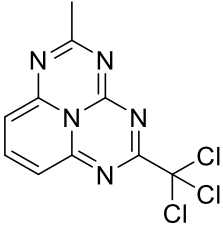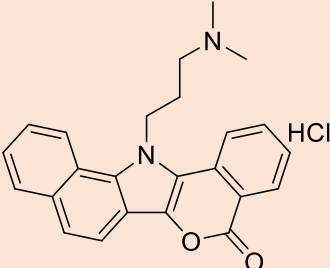
Following an FID assay carried out against c-MycC27 and c-MycC52 these 23 compounds that were initial hits were reduced to 10 in the repeated assay as they did not meet the hit criteria of >15% or <-15% (*table 2.4*). 13 compounds did not demonstrate >15% or <-15% displacement and were disregarded (*table 2.4*). The most likely reasons for this lack of reproducibility are: 1) the compound stocks in the screening plates were old and so the compounds may have degraded through multiple defrost cycles, although this is limited as much as possible, which can cause degradation 2) the storage conditions at -80 °C in 100% DMSO may not have been optimal for all the compounds and so degradation occurred 3) the plates wells may have become cross-contaminated 4) the plates were incorrectly labelled.

Table 2.4: Fluorescent indicator displacement assay screen of c-MycC National Cancer Institute Diversity Library IV hits validation against c-MycC27 and c-MycC52. The screen was ran using 1 µM DNA, 2 µM thiazole orange and 2.5 µM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15%. Error is standard deviation for ≥2 technical replicates.

Ligand (NSC)	Structure	% Displacement	
		c-MycC27	c-MycC52
5426		-13.2 ± 6.7	0.6 ± 0.9
11668		-3.0 ± 2.7	4.1 ± 1.0
13248		-98.5 ± 8.1	3.3 ± 2.1
13974		37.5 ± 3.9	25.6 ± 1.8
33005		-4.1 ± 4.6	4.8 ± 0.2
61642		-12.0 ± 1.7	4.9 ± 0.9
63680		-71.4 ± 22.1	0.5 ± 0.4

73735		19.8 ± 6.3	8.2 ± 1.1
105781		-16.5 ± 8.3	0.1 ± 2.9
107582		-9.1 ± 3.9	1.8 ± 1.7
118832		-14.2 ± 0.6	1.0 ± 1.8
124818		3.8 ± 7.2	5.7 ± 1.4
135168		-13.7 ± 7.5	0.8 ± 0.5
228150		-6.7 ± 9.2	-1.4 ± 4.0

288387		-101.4 ± 0.8	-3.5 ± 0.5
305780		-13.2 ± 7.0	0.8 ± 1.7
309401		-73.2 ± 9.4	7.0 ± 2.2
317605		-91.0 ± 14.3	26.4 ± 1.9
335504		-14.9 ± 15.3	-2.0 ± 2.9
345647		27.0 ± 10.5	15.6 ± 1.1
354844		15.2 ± 2.2	31.8 ± 0.7

| 366289 | | -24.1 ± 9.9 | -1.7 ± 1.4 |
| 638432 | HCl | -19.2 ± 4.8 | -0.8 ± 0.9 |

The 10 compounds identified as hits for either c-MycC oligonucleotide were: **13248**, **13974**, **63680**, **73735**, **105781**, **288387**, **309401**, **317605**, **345647**, and 354844 (*figure 2.11*). FID determined that out of the 10 compounds three bound to c-MycC52 and nine bound c-MycC27. Of these 10 compounds two, **13974**, and **317605**, were determined as hits for both c-MycC27 and c-MycC52. There were seven compounds determined as hits for c-MycC27 only, **13248**, **63680**, **73735**, **105781**, **288387**, **309401**, and **345647** and one compound was identified as a hit for c-MycC52 only, **354844**. In total there were three compounds determined as c-MycC52 binders and if ranking the ligands from best to worst, where the best has the highest percentage displacement and the worst has the lowest, for c-MycC52 the ranking is: 354884 (32% +/-1), **317605** (26% +/- 2), **13974** (26% +/-2) (*table 2.5*). If doing the same ranking system c-MycC27 for the compounds with positive percentage displacement values then the ranking is: **13974** (38 +/-4), **345647** (27 +/-11), and **73735** (20 +/-8) (*table 2.5*). Ranking the compounds with negative percentage displacement with the best having the smallest value and the worst the largest is: **288387** (-101 +/-1), **13248** (-99 +/-8), **317605** (-91 +/-14), **309401** (-73 +/-9), **63680** (-71 +/-22), and **105781** (-16.5 +/-8) (*table 2.5*). Compound **13974** is the best ranked ligand for c-MycC27 but the worst ranked for c-MycC52. Whereas the best ranked compound for c-MycC52 is **354844** and it was determined this

doesn't bind c-MycC27, although it was on the boarder of being determined as a hit (15% +/-2).
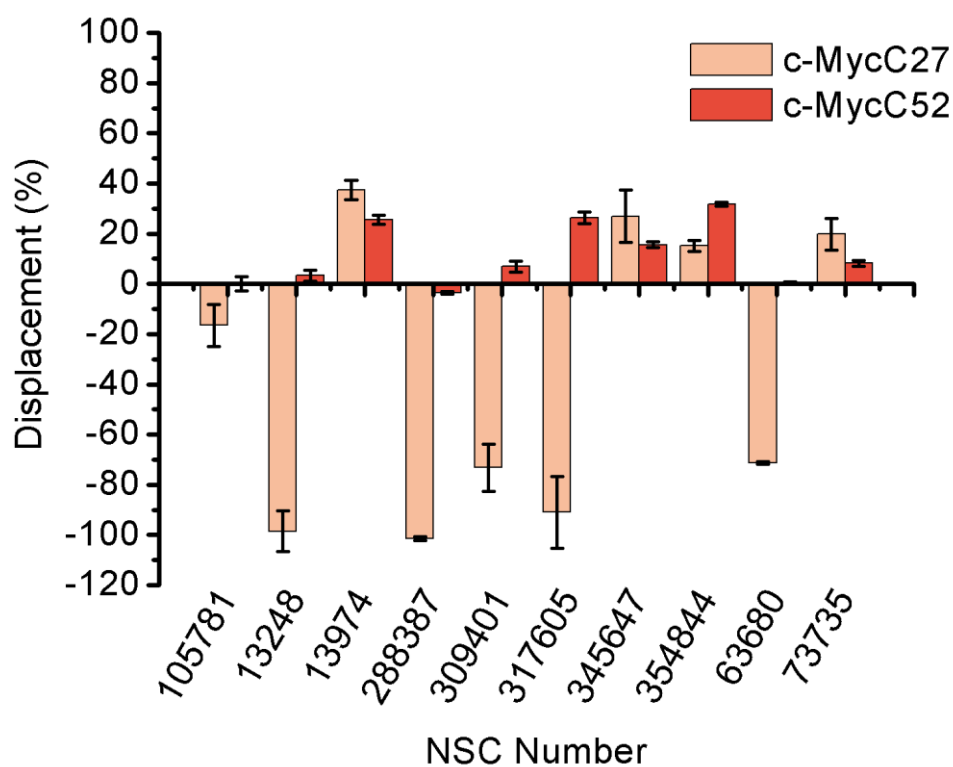
Figure 2.11: Fluorescent indicator displacement assay screen validation hits from the National Cancer Institute Diversity Library IV against c-MycC27 and c-MycC52. The screen was ran using 1 µM DNA, 2 µM thiazole orange and 2.5 µM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15%. Error bars are standard deviation for ≥2 technical replicates.

Table 2.5: Fluorescent indicator displacement assay screen validation hits summary from the National Cancer Institute Diversity Library IV against c-MycC27 and c-MycC52. The screen was ran using 1 µM DNA, 2 µM thiazole orange and 2.5 µM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15%. Error is standard deviation for ≥2 technical replicates. Error is standard deviation for ≥2 technical replicates.

NSC Number	% Displacement (error)	
	c-MycC27	c-MycC52
13248	-98.5 ± 8.1	3.3 ± 2.1
13974	37.5 ± 3.9	25.6 ± 1.8
63680	-71.4 ± 22.1	0.5 ± 0.4
73735	19.8 ± 6.3	8.2 ± 1.1
105781	-16.5 ± 8.3	0.1 ± 2.9
288387	-101.4 ± 0.8	-3.5 ± 0.5
309401	-73.2 ± 9.4	7.0 ± 2.2
317605	-91.0 ± 14.3	26.4 ± 1.9
345647	27.0 ± 10.5	15.6 ± 1.1
354844	15.2 ± 2.2	31.8 ± 0.7

Having identified the 10 compounds of interest further FID was carried out to identify which of the ligands had greater promise as specific c-MycC probes.
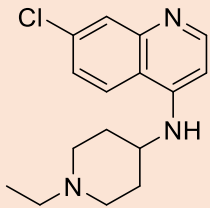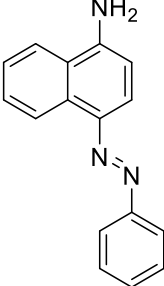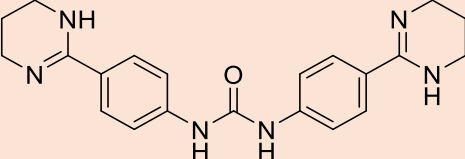
2.2.5 Testing the Specificity of c-MycC27 and c-MycC52 Hits using Fluorescent Indicator Displacement Assay

To investigate the specificity of the compounds FID was carried out to assess their binding to B-DNA and the G-quadruplex structure that forms in the *MYC* promoter region opposite c-MycC which will be referred to c-MycG. It is common within the i-motif field to test specificity against double stranded DNA (B-DNA) as this is the most common structural form of DNA in the cell and also against the G-quadruplex. The G-quadruplex these compounds were tested against is the one that forms on the opposite strand of the i-motif of interest in the *MYC* promoter region. This is

arguably the most important G-quadruplex to test as when investing compound effects on the *MYC* promoter in cellular conditions it needs to be known if any effects seen are due to binding to the i-motif or G-quadruplex and therefore biophysical data can provide evidence for suggesting how compounds are interacting and causing effects on *MYC* expression.

The 10 compounds of interest were tested against B-DNA using the FID assay in the exact same conditions used for FID experiments against both c-MycC sequences, 10 mM NaCaco pH 6.6. This ensures that conditions are identical expect the target DNA. Of the 10 compounds one, **13974**, was determined as a hit for B-DNA, >15% or <-15% displacement, (*table 2.6*). Therefore, **13974** was not pursued further as it was not c-MycC specific. Compound **13248** was also not taken further within this project due to the stock being finished and limited time, thus will be followed up by colleagues in the Waller lab.

Table 2.6: Fluorescent indicator displacement assay screen hits from the National Cancer Institute Diversity Library IV against B-DNA, c-MycG, c-MycC27 and c-MycC52. The screen was ran using 1 μM DNA, 2 μM thiazole orange and 2.5 μM of compound in 10 mM NaCaco pH 6.6. Compounds were determined as hits if the displacement was over 15% or under -15%. Error is standard deviation for ≥2 technical replicates. N/D means not determined.

Ligand (NSC)	Structure	% Displacement			
		c-MycC27	c-MycC52	B-DNA	c-MycG
13248		-98.5 ± 8.1	3.3 ± 2.1	-8.2 ± 1.4	N/D
13974		37.5 ± 3.9	25.6 ± 1.8	20.3 ± 0.6	N/D
63680		-71.4 ± 22.1	0.5 ± 0.4	-3.1 ± 6.2	-5.2 ± 3.9

73735	19.8 ± 6.3	8.2 ± 1.1	7.2 ± 0.9	5.5 ± 0.5
105781	-16.5 ± 8.3	0.1 ± 2.9	2.2 ± 0.7	6.5 ± 2.5
288387	-101.4 ± 0.8	-3.5 ± 0.5	-5.1 ± 3.1	-8.6 ± 2.1
309401	-73.2 ± 9.4	7.0 ± 2.2	-13.4 ± 6.0	9.3 ± 4.8

317605		-91 ± 14.3	26.4 ± 1.9	8.1 ± 1.4	19.7 ± 0.8
345647		27 ± 10.5	15.6 ± 1.1	12.1 ± 2.3	21.4 ± 1.2
354844		15.2 ± 2.2	31.8 ± 0.7	6.2 ± 1.2	21.6 ± 3.8

To further investigate the specificity of these compounds it was analysed if they bind c-MycG using the FID assay (*table 2.6*). Importantly, the buffering conditions were adjusted when testing against the G-quadruplex structure as they are stabilised by cations and so 100 mM KCl was added to the 10 mM NaCaco pH 6.6 to ensure that the G-quadruplex was folded.[260] The rest of the components of the buffer were identical to those used for the c-MycC and B-DNA sequences including the pH, this means the compounds will be the same charge. The data shows that **317605** (20 (+/-1), **345647** (21 (+/-1), and **354844** (22 +/- 4), have greater than 15% displacement with c-MycG and thus by the threshold set for this project (*table 2.6*). These compounds have the same displacement of **TO** within error. Furthermore, compound **354844** was determined by FID as only binding to c-MycC52 and c-MycG means that this compound could have dual specific binding of the *MYC* i-motif and G-quadruplex (*table 2.6*). To confirm this binding to a larger number of DNA sequences including multiple i-motif and G-quadruplex sequences need to be determined. Compound **317605** was identified as binding only to the i-motif and G-quadruplex structures and so this may have dual binding to i-motif and G-quadruplex structures whilst not binding to other DNA structures such as hairpins, and thus this needs to be tested against other DNA structures to explore this.

The other five compounds, **63680**, **73735**, **105781**, **288387** and **309401** were determined to only bind c-MycC27 (*table 2.6*). Suggesting that **105781**, **288387**, **309401**, **63680**, and **73735** are all specific to c-MycC27. Although, all of these compounds, except **73735**, gave negative percentage displacement and so other techniques will be required to confirm their binding which will inform if a negative displacement of <-15% will be regarded as a hit in the future in the Waller lab. The identification of specific compounds for c-MycC27 and none for c-MycC52 is surprising as in theory it would be more likely to find specific compounds for c-MycC52 as this i-motif structure has a greater number of longer loop regions in comparison to c-MycC27 (*figure 2.12*). Longer loop regions provide more

bases and so the sequence to which the compound binds could be longer and is less likely to appear at random as you have a greater number of bases in a sequence for recognition and binding. One could hypothesise that the loop regions are where specificity could arise between i-motif sequences as the centre of the structures are runs of three or more intercalated hemi-protonated cytosine bases and so a compound binding in that region is likely to bind to all i-motif sequences and if not other DNA structures especially if it intercalates. However, bromine footprinting carried out by Hurley and team in 2016 demonstrated that although c-MycC52 has an additional 25 bases it doesn't have the longest loop region.[4] The central loop of c-MycC27 is the longest if the morphology depicted in *figure 2.10* if it is in fact the one forming in solution. Yet c-MycC52's first and final loop are longer than the first and last loop of c-MycC27. Notably, c-MycC52 has a long tail on the 5' prime end and is 16 bases long and so one could have suggested this tail would have potentially allowed for more specificity than any other region of both structures (*figure 2.12)*. However, that is not what has been observed with the screen ran using the FID assay and NCI IV library, although that may not be the case with all investigations into ligands.
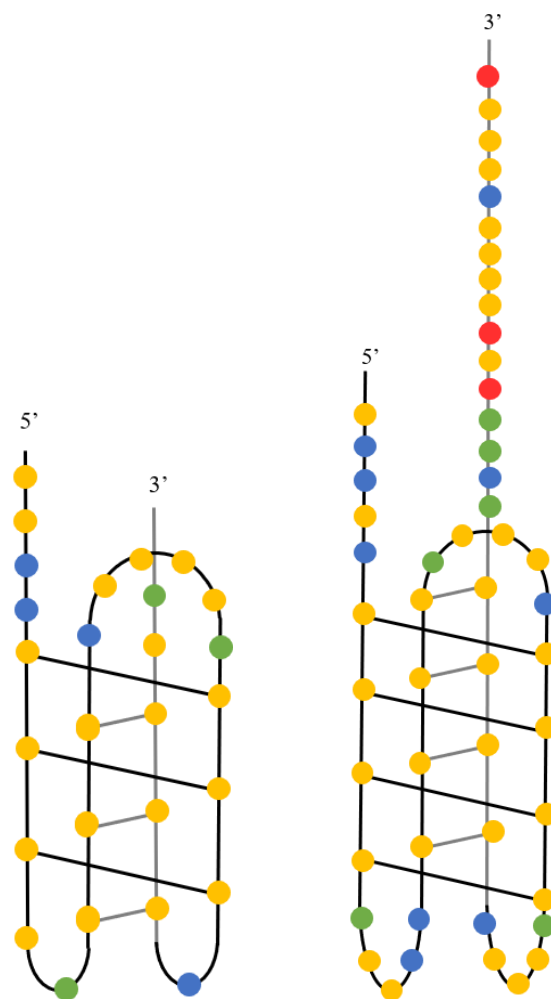
Figure 2.12: Illustrations of c-MycC27 (Left) and c-MycC52 (Right) likely i-motif morphologies formed. The morphology depicted for c-MycC52 is based on bromide footprinting by Hurley and team in 2016.[4,7] The circles demonstrate DNA bases and are colour coded as follows: Yellow-Cytosine, Red-Guanine, Blue-Thymine, Green-Adenine.

The FID experiments demonstrate that the aim to identify c-MycC small compounds that can be developed into c-MycC specific probes has been achieved. In addition, the FID experiments may have identified a potential dual specific probe for c-MycC52 and c-MycG, **354844**, a probe that may only bind to i-motif and G-quadruplex structures, **317605**, and potentially multiple c-MycC27 specific probes. The c-MycC52 probes are of particular interest as this is the biologically relevant c-MycC sequence and it has been identified that these 52 bases are needed for transcriptional activation in the cell.[4,7] According to the model proposed by Hurley and colleagues the

MYC promoters transcription is regulated by DNA structural changes and that i-motif or G-quadruplex formation both turn off transcription.[170,171] Therefore, exploring **354844** as a dual *MYC* i-motif and G-quadruplex specific probe is of particular interest in an aim to understand how this promoter region is functioning and the prospect of cancer therapeutics where *MYC* transcriptional activity is dysregulated. It was decided to progress all eight compounds for further biophysical testing: **63680**, **73735**, **105781**, **288387**, **309401**, **317605**, **345647**, and **354844**.

2.2.6 Determining DC$_{50}$ Values using Fluorescent Indicator Displacement Titrations

The compounds identified as c-MycC hits were further characterised using FID titrations. The parameters and protocol are identical to those of the FID assay except for the addition of ligand step, where 1 µM of ligand is titrated per 0.9 µL. For each compound the concentration range was 1-10 µM to obtain the DC$_{50}$, the concentration at which 50% of TO is displaced, 10 µM was not exceeded as this was 10 equivalents of the DNA concentration and to avoid precipitation of the compounds.

The eight compounds progressing forward from *Section 2.2.5* were **63680**, **73735**, **105781**, **288387**, **309401**, **317605**, **345647**, and **354844** (*figure 2.13)*. The compounds were only titrated against the c-MycC sequence they were shown to bind previously as determined in *section 2.2.4* shown in *table 2.5*. For c-MycC52 there were two compounds of interest: **317605** and **354844**. c-MycC27 had seven compounds of interest: **63680**, **73735**, **105781**, **288387**, **309401**, **317605**, and **345647**. The FID titration data is shown in *figure 2.14*. The graphs show that for each compound ran there is an increase of binding as the compound concentration is increased demonstrating concentration dependent binding. Furthermore, the curve shapes are all similar and appear to be asymptotic, where a plateau is not

reached. The asymptotic graph shape indicates non—specific binding is occurring between the compounds and c-MycC structures.
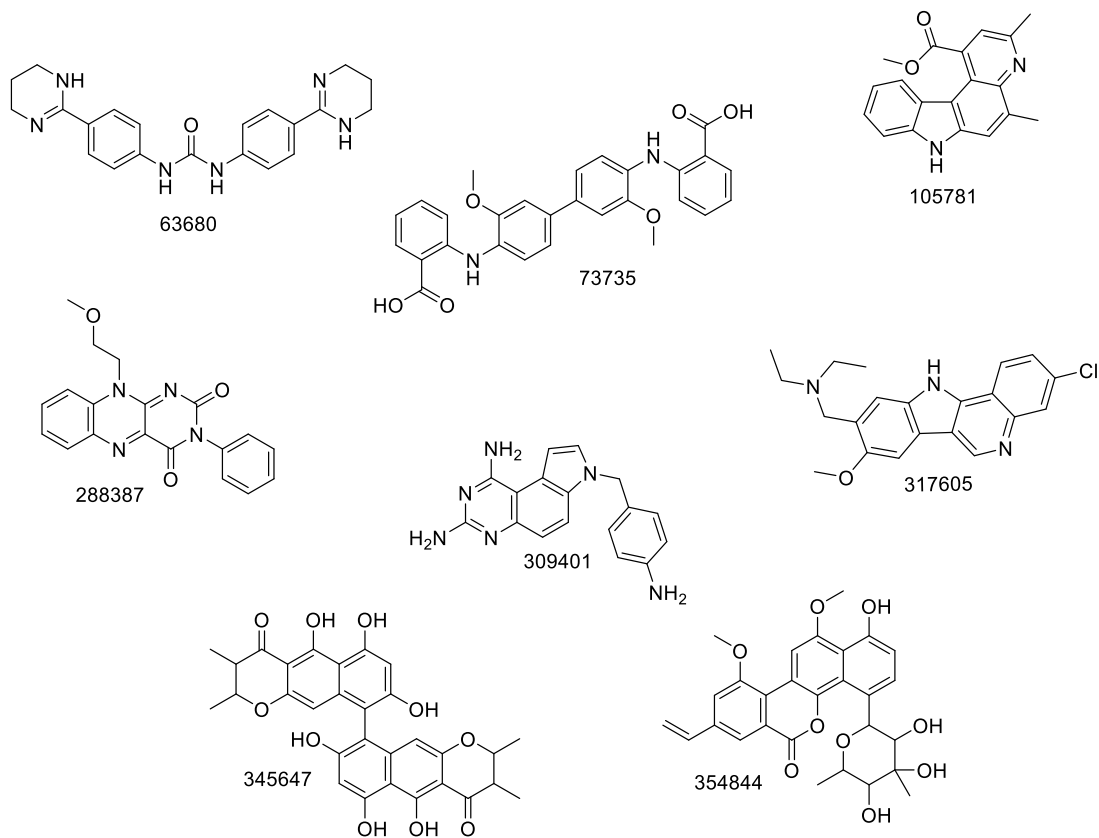
Figure 2.13: Structure of eight compounds of interest as c-MycC27 and/or c-MycC52 binders, NSC 63680, 73735, 105781, 288387, 309401, 317605, 345647, and 354844.
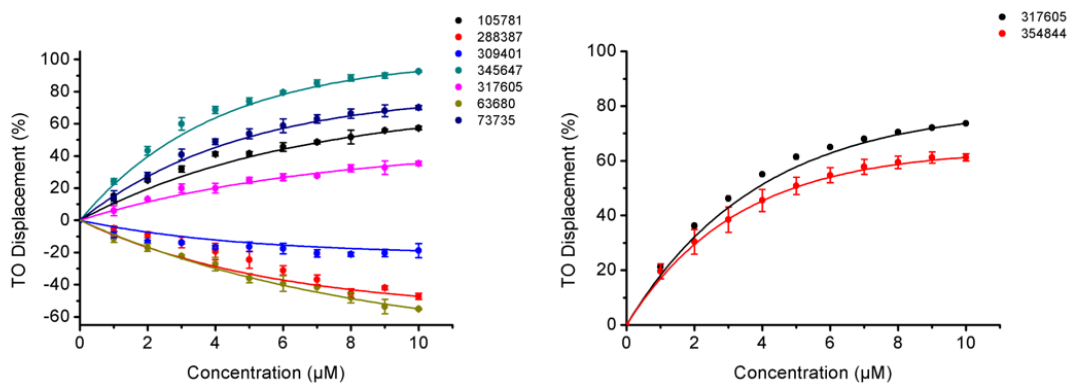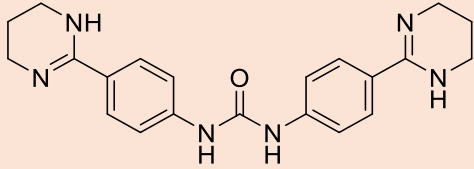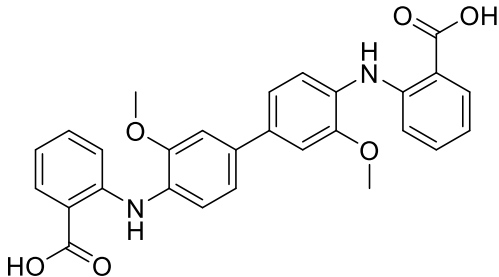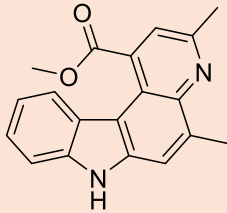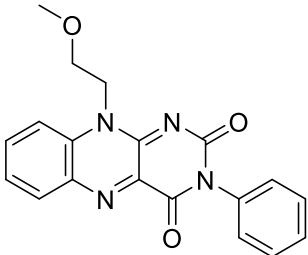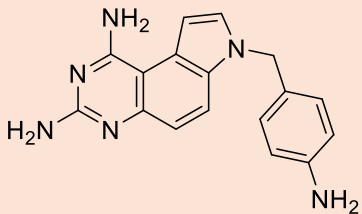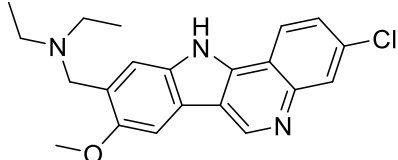
Figure 2.14: Fluorescence intercalator displacement assay titration of compounds from 1 equivalence to 10 equivalences to analyse binding to c-MycC27 (left) and c-MycC52 (right) in the buffering conditions 10 mM NaCaco pH 6.6. 1 µM DNA, 2 µM TO, and 1 to 10 µM of compound added in 1 µM additions.

The titration studies, *figure 2.14*, were used to obtain DC_{50} values (*table 2.7*). The lower the DC_{50} value the lower concentration of the compound required to displace 50% TO from binding to the DNA structure, and therefore a better binder. The DC_{50} was determined as >10 µM for **288387**, **309401**, and **317605** against c-MycC27 (*table 2.7*). Thus, these are weak binders. Ranking the compounds from best to worst binder according to DC_{50} for c-MycC27: **345647** (2 µM), **73735** (4 µM), **105781** (7 µM), **63680** (9 µM), and then the ligands with a DC_{50} of >10 µM which one cannot discriminate between. Notably, this ranking is a mix of compounds that gave positive and negative percentage displacements, if one was to rank the positive results best to worst it would be the following: **345647**, **73735**, and **105781**. The negative percentage displacements would rank as so: **63680**, and then **288387** and **309401**. The best method of ranking these compounds is uncertain due to the lack of clarity around what is happening at a molecular level with the negative percentage displacement results. Whereas the ranking for c-MycC52 is more straight forward as there is not a mix of positive and negative displacement values. The ranking for c-MycC52 best to worst binders are: **317605** (4 µM) and **354844** (5 µM).

Table 2.7: DC_{50} values, the concentration at which a compound displaces 50% of TO bound to the DNA sequence, were calculated using FID titration assay from 0 µM to 10 µM. Assay was performed in 10 mM NaCaco pH 6.6. N/D shows any compound where a DC_{50} was not determined as it wasn't tested and >10 µM demonstrates that the DC_{50} value was not reached and so a higher concentration is required.

NSC Number	Structure	DC_{50} (µM)	
		c-MycC27	c-MycC52
63680		8.9 ± 0.5	N/D
73735		4.3 ± 0.4	N/D
105781		6.9 ± 0.5	N/D
288387		>10	N/D
309401		>10	N/D
317605		>10	3.4 ± 0.1

345647	2.4 ± 0.2	N/D
354844	N/D	4.9 ± 0.8
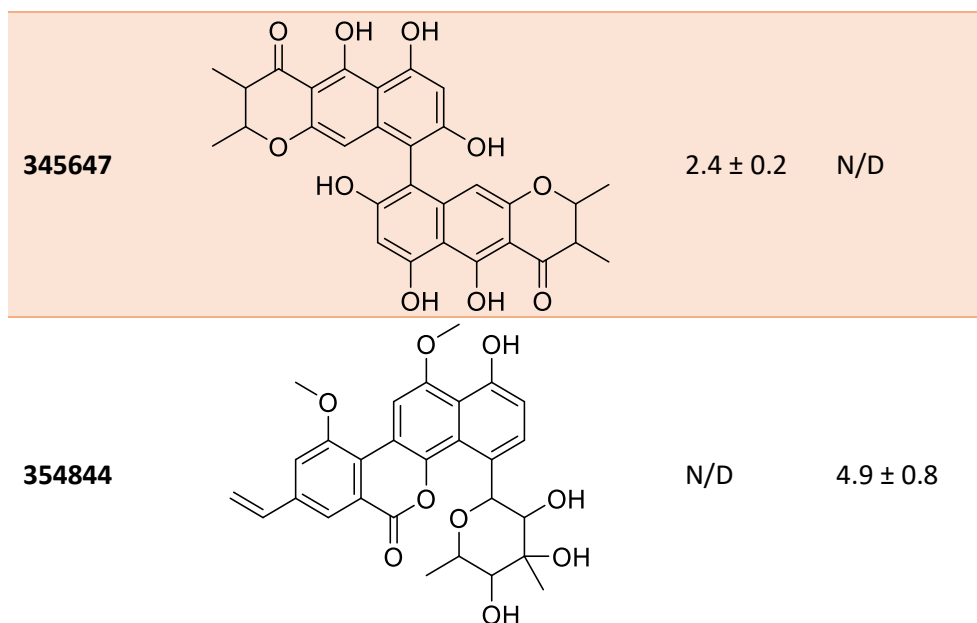
c-MycC52 rankings from the FID hits validation assay discussed in *Section 2.2.4* were **354844** (32% +/-1) and the **317605** (26 +/-2) (*table 2.5*) whereas using the titration data presented in *table 2.6* the ranking positions have swapped over with **317605** and **354844** using DC_{50} values, of 3 µM and 5 µM respectively.

Ranking of c-MycC27 compounds showing positive percentage displacement were **345647** (27 +/-11) and **73735** (20 +/-8) from the FID validation assay (*table 2.5*). Whereas analysis of the FID titration data shows that not only are these two compounds displaying positive percentage displacement values but so are **105781** and **317605**. Ranking these compounds using the DC_{50} values the best to worst binders were as follows: **345647** (2 µM), **73735** (4 µM), **105781** (7 µM), **317605** (>10 µM). Therefore, **345647** and **73735** ranked respectively 1st and 2nd best binders in both experimental approaches showing consistency of the FID method. To investigate the discrepancies between the two approaches for **105781** and **317605** the line fitting for the FID titration study, (*figure 2.11*), was used to get a percentage displacement value at 2.5 µM of compound which is the concentration the ligand was ran at in the FID validation assay, (*table 2.5*), to allow some form of a comparison. Previously **105781** showed low negative percentage displacement (-17% +/-8) and now is showing

positive percentage displacement (29% at 2.5 μM), and **317605** showed a large negative percentage displacement (-91 +/-14) and is now showing a low positive percentage displacement (15 at 2.5 μM). The biggest change in percentage displacement is from **317605**. These discrepancies suggest the compound is affecting the fluorescence reading at 450 nm, which is a limitation of this assay. It was decided not to repeat this data but to progress these compounds into further biophysical testing where fluorescence properties won't affect the results of binding data. Ranking of c-MycC27 compounds showing negative percentage displacement were: **288387** (-101 +/-1), **317605** (-91 +/-14), **309401** (-73 +/-9), **63680** (-71 +/-22), and **105781** (-16.5 +/-8) for the FID validation assay (*table 2.5)*. The FID titration data demonstrated that the ranking using DC_{50} values was: **63680** (9 μM) and then **288387** and **309401** cannot be discriminated between as they were both >10 μM and **317605** and **105781** were discussed above (*table 2.7*).

Overall, the FID assay was a useful tool for high throughput testing and screening large libraries against different DNA structures. However, it has been shown that it is not suitable for compounds that affect the fluorescence reading of **TO** at 450 nm as the analysis becomes complicated and like with most procedures needs other procedures to corroborate findings with. To investigate the binding of these compounds against different DNA structures further biophysical techniques were employed.

2.3 Biophysical Testing of Compounds of Interest

2.3.1 Ligand Binding Studied by Circular Dichroism

Circular polarised light can be utilised to observe characteristic spectrums of different DNA structures to determine what structure a sequence has formed. CD is able to differentiate between different structural

polymorphisms of DNA including B-DNA (positive strong band at 260-280 nm and a negative band at 245 nm) G-quadruplexes (parallel G-quadruplexes have a dominant positive peak at 260 nm and anti-parallel G-quadruplexes have a negative band at 260 nm and a positive band at 295 nm), and i-motifs (dominant positive peak at 288 nm and a negative band at 267 nm).[115,231] As CD determines DNA structure it can be used to monitor structural changes under the effects of changing conditions for example different cations or pH, and the addition of ligands. Therefore, it is possible to use CD to determine if a compound affects the stabilisation of a particular DNA structure or changes the equilibrium.

CD spectroscopy was used to further analyse the effects of the compounds of interests on different DNA structures. To do this, compounds were titrated into 10 µM of c-MycC52 in 1 µL additions of 10 µM up to 50 µM to see any changes to the CD spectrum up to five equivalents. After addition of 50 µM the sample was melted from 5 °C to 95 °C to identify the melting temperature, T_m, which is where the sample is 50% folded into i-motif and 50% random coil (*figure 2.15*). To analyse the effects of compounds on the DNA structure the T_m in the absence of ligand is compared to the T_m with 50 µM of the compound added, if the T_m is shifted to the lower temperature side the compound has destabilised the structure and if it has shifted to the higher temperature side it has stabilised.
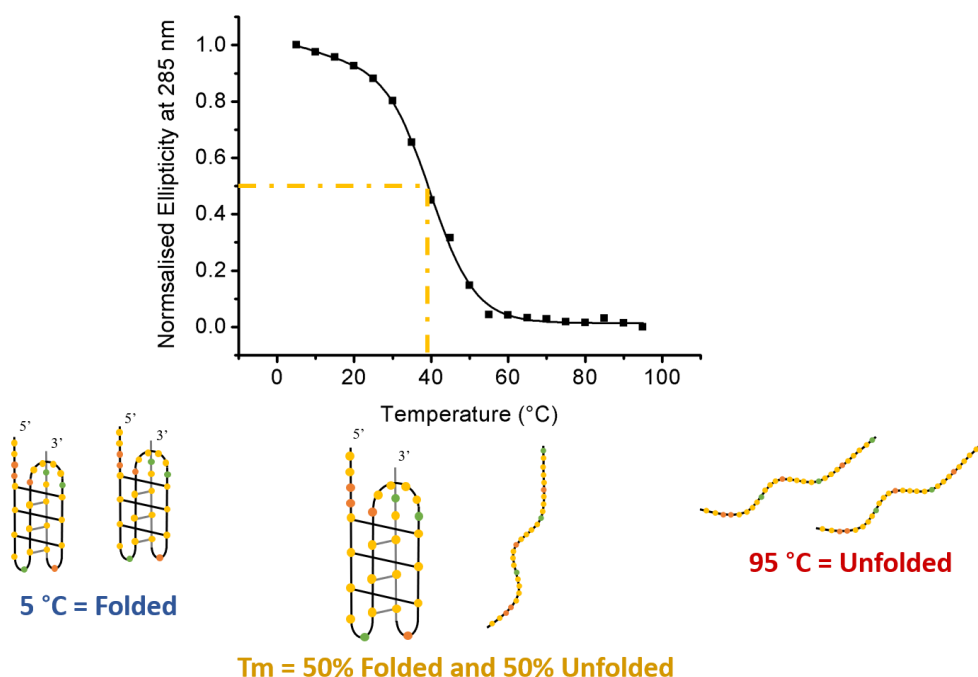
Figure 2.15: Schematic of CD normalised ellipticity showing fraction folded from 5 ᵒC to 95 ᵒC to determine the Tₘ, which is the temperature at which the structure is 50% folded and 50% unfolded show by the yellow lines on the graph. Data used for the CD graph was 10 μM i-motif, 52 base i-motif that forms in the *MYC* promoter region, in 10 mM NaCaco pH 6.6.

In the hunt for i-motif probes it is ideal to have a mixture of probes that cause different effects on stability. For example, a probe that stabilises the i-motif can be used to increase i-motif formation in cells and observe the effects of that in cellular conditions from reporter gene assays.[97] In addition, i-motifs are dynamic structures with several different conformations, making it difficult to get crystals composed of one conformation to allow the structure to be solved. Ligands can shift the equilibrium to one formation so that the structure can be solved. This could further our understandings of the mechanisms of action of i-motif structures. Probes that destabilise while not useful for crystallisation can be used to determine the effects of the i-motif in cells by observing the effects on transcription of decreasing i-Motif formation. Pairing that information with observations on the effects of a stabiliser will increase the

understanding of the function of the particular i-motif. Compounds that don't stabilise or destabilise i-motif structures will make ideal probes for investigating when they form, they may need fluorescent probes linked to them to do so, for example like the i-motif specific antibody iMab[94] and go further to compare how this differs in normal and diseased cells for i-motif structures in general and also for specific i-motif sequences like c-MycC which is the aim of this study.

To investigate the specificity of these compounds the melting temperatures were assessed against multiple DNA structures in the absence and presence of ligands using CD. The buffering conditions are kept identical for all DNA structures, 10 mM NaCaco pH 6.6, except for when using c-MycG as this is a G-quadruplex it needs cations to stabilise the structure. In the case of c-MycG it would be optimal to add 100 mM KCl to the buffer conditions, however this makes c-MycG so stable it doesn't melt by 95 ^{o}C, see *appendix A1*. It was demonstrated that the addition of 10 mM KCl the structure melts at ~60 ^{o}C, see *appendix A2*. Thus, for all CD experiments investigating the melting temperature of c-MycG the buffering conditions are 10 mM NaCaco 10 mM KCl at pH 6.6. This complicates the comparison of compounds effects on different DNA structures as buffer composition can play a role in this, and 10 mM KCl was not used to investigate i-motif sequences as it destabilises them and so a more acidic pH would be required to counteract this.[261] A more acidic pH is then even further from the pH in cellular conditions and could then affect the charges on the compounds which could affect binding. In this instance it was decided to only add KCl to the G-quadruplex experiments and be aware that the KCl could be having an involvement in the results.

The two compounds identified as c-MycC52 binders, **317605** and **345844**, demonstrate stabilisation of c-MycC52 in CD melting experiments by a small degree (*figure 2.16*). The data is fitted with a bi-dose response as although there is only one transition seen the bi-rose response fitted the data set better and c-MycC is known to have two transitions which can

clearly be seen in some environments due to different i-motif populations. The recently published paper by Smith *et al.* also analysed all c-MycC CD data with bi-dose response fittings even though the second transition was only seen in the presence of particular ligands.[97]

If ranking the two compounds by largest stabilisation value, compound **317605** would be first with a stabilisation of +2°C and **345844** as second with a stabilisation of +1°C and (*table 2.8*).
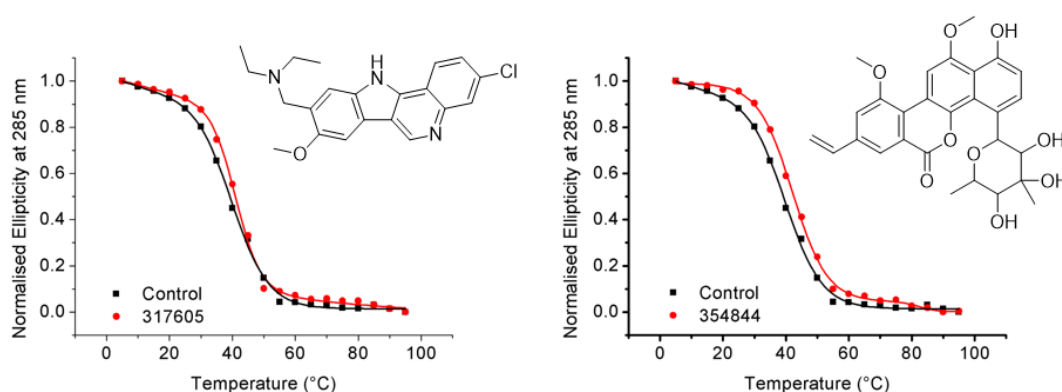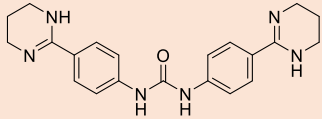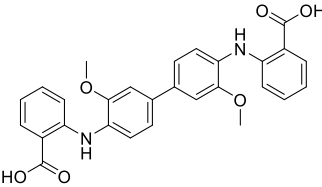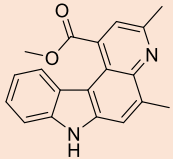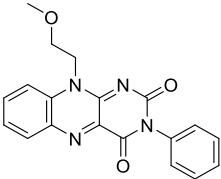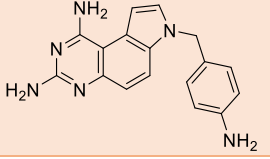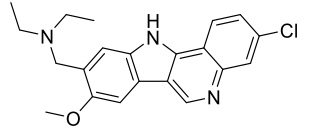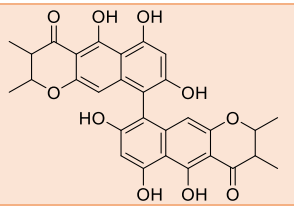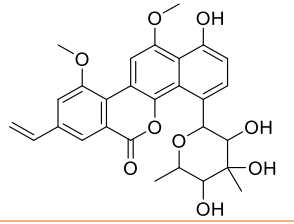
Figure 2.16: Circular Dichroism of normalised ellipiticity at 285 nm, the maximum positive characteristic i-motif peak for c-MycC, of 10 µM c-MycC52 in 10 mM NaCaco pH 6.6 with addition of 50 µM of a given compound. The control (black) is addition of 5 µL of 100% DMSO, this the vehicle control. Line fitting is bi-dose response.

Table 2.8: Change in T_m for c-MycC52 compounds of interest. Change in T_m is determined by the difference in the T_m with no compounds present compared to that with 50 μM of compound present against 10 μM DNA determined from circular dichroism melts from 5 ºC to 95 ºC in buffer conditions 10 mM NaCaco pH 6.6.

NSC Number	Structure	cMycC52		cMycC27			B-DNA		c-MycG	
		ΔTm 1	Fitting Error	ΔTm 1	ΔTm 2	Fitting Error	ΔTm 1	Fitting Error	ΔTm 1	Fitting Error
63680		N/D	N/D	-1.2	N/D	0.0031	-14.7	0.010	+16.5	0.009
73735		N/D	N/D	0.0	26.7	0.0002	-3.0	0.017	+3.2	0.005
105781		N/D	N/D	0.0	N/D	0.0057	-1.8	0.009	+1.6	0.008
288387		N/D	N/D	+0.3	N/D	0.0049	-4.6	0.018	+2.1	0.002

309401		N/D	N/D	+0.4	N/D	0.0035	-7.3	0.004	+6.5	0.007
317605		+1.3	0.00228	+0.3	+8.1	0.0077	-2.9	0.004	+20.5	0.008
345647		N/D	N/D	-0.5	+27.6	0.0046	-1.2	0.005	+13.5	0.014
354844		+2.3	0.00096	N/D	N/D	N/D	-2.9	0.013	+10.1	0.004

There were seven compounds identified as c-MycC27 binders of interest from FID experiments, **63680**, **73735**, **105781**, **288387**, **309401**, **317605**, and **345647**. CD melting experiments were used to determine if the compounds affected the stability of c-MycC27 (*figure 2.17*). The line fittings used were bi-dose response for all even though some did not show two transitions, such as 63680 (*figure 2.17*), as this fitting fitted better than the dose response as discussed for the c-MycC52 CD melts in *figure 2.16*.

The CD c-MycC27 melting experiments determined that most of the compounds did not affect the stability of c-MycC27 as one can see in *figure 2.17* and supported by the determination of the change in melting temperature of c-MycC27 in *table 2.8*. Compounds, **105781**, **288387**, and **309401** all have no effect on c-MycC27 structure stability (*table 2.8*). Whereas compound **63680** had a ΔTm value of -1 °C and thus is a weak destabiliser of c-MycC27, if rounding to one significant figure **345647** also has a ΔTm value of -1 °C however at two significant figures this value is -0.5 °C and so could be weakly destabilising c-MycC27 but is not destabilising c-MycC27 in these conditions (*table 2.8*). Interestingly, compounds **73735, 317605**, and **345647** had an additional transition compared to the control as shown in *figure 2.17*. A second transition is possible because the sequence can form more than one i-motif structure, therefore these three compounds shifted the equilibrium between 40 °C and 60 °C more towards the i-motif structure that causes the second transition seen in the CD compared to that of the control. There are examples of this in the literature, such as Smith *et al.* who demonstrated that ellipticine derivatives caused a second transition in c-MycC CD melts and that the higher concentration of the ligand was added the greater this stabilisation was providing evidence that this transition is real.[97] Thus, these compounds are stabilising the particular i-motif population that is represented by the second transition. Furthermore, none of those three compounds affected the stability of the i-motif that causes the first transition, but all stabilised the i-motif structure that causes the second transition. The biggest stabilisation was seen with **345647** and the

smallest with **317605**. Arguably, compounds **288387** and **309401** may demonstrate a second transition at a higher concentration as at 50 μM there is a small change where the second transition occurred for other compounds, such as **345647**, however it is not a transition at 50 μM.
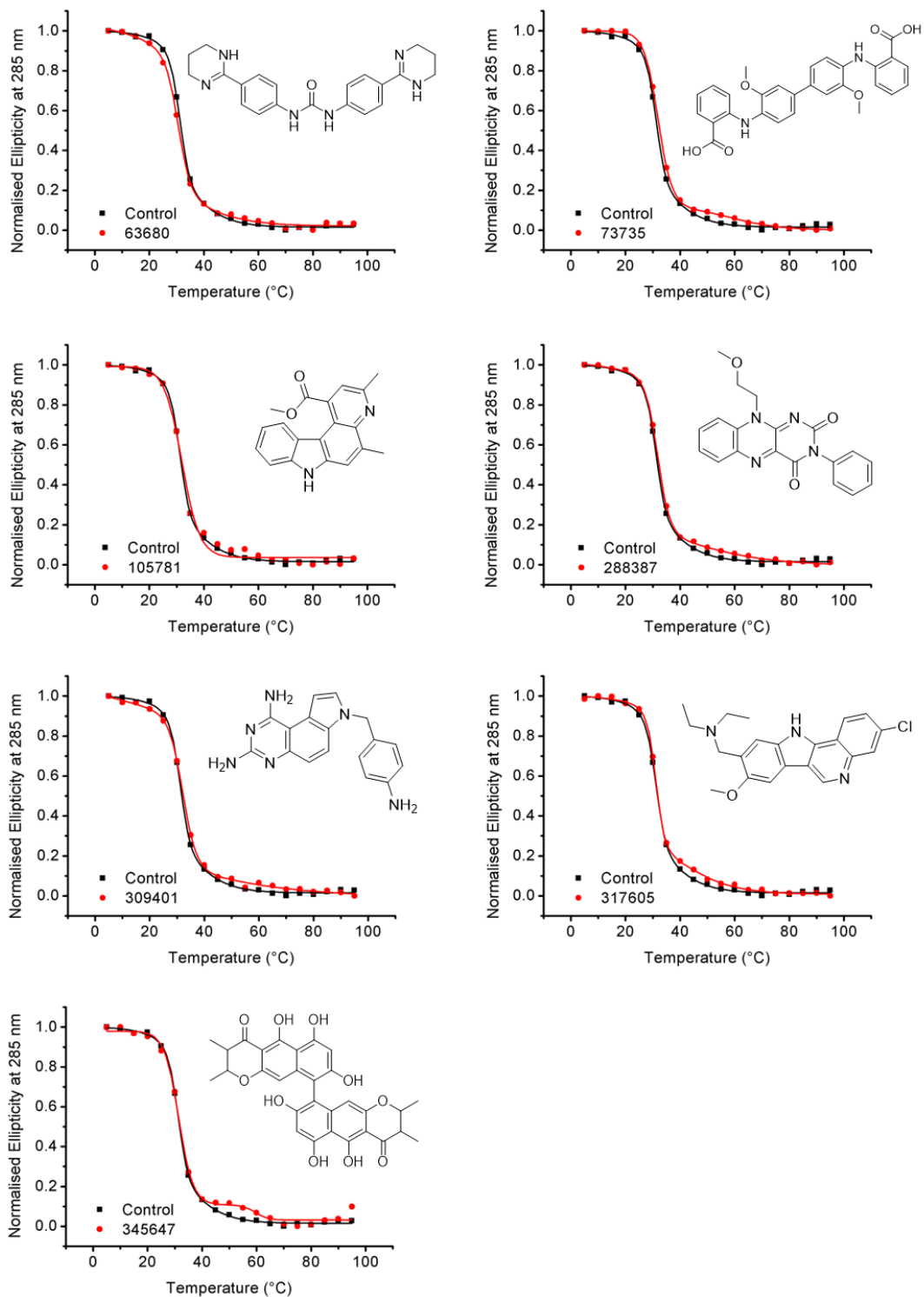
Figure 2.17: Circular Dichroism of normalised ellipiticity at 285 nm, the maximum positive characteristic i-motif peak for c-MycC, of 10 µM c-MycC27 in 10 mM NaCaco pH 6.6 with addition of 50 µM of a given compound. The control (black) is addition of 5 µL of 100% DMSO, this the vehicle control. Line fitting bi-dose response.

The effect of the eight compounds of interest on B-DNA stability was investigated using CD (*figure 2.18*). All eight compounds in 10 mM NaCaco pH 6.6 showed that they destabilised B-DNA and thus have a binding interaction with B-DNA. The majority of the compounds, all except **63680**, destabilised B-DNA by <10 °C and **63680** destabilised it by >10 °C (*table 2.8*). The FID experiments determined that these eight compounds did not bind B-DNA, yet the CD data disagrees and suggests the eight compounds are non-specific DNA binders. One could reason that the FID and CD came to different conclusions because they are testing different objectives. FID is testing for binding that displaces the binding of **TO** whereas CD determines changes in stability to the DNA structure via melting temperature and thus can be used to show binding if stability is affected. Furthermore, different variables were measured using FID and CD. Including, the compound concentration relative to B-DNA concentration was 2.5 equivalences in FID and 5 in the CD, the equivalence used in the FID may not have been sufficient to see >15% and be determined as binding, the compounds were binding to B-DNA without displacing **TO** and so were not detected as binding in the FID experiments, or a higher concentration of B-DNA was needed to displace **TO** in the **TO**-B-DNA complex. Thus, highlighting the importance of having the same variables across different methods to make comparisons between techniques more appropriate. Overall, CD confirmed that all eight compounds destabilise and thus bind B-DNA in the conditions the CD was ran under (*table 2.8*), and that the FID experiments should be designed using the same equivalences of DNA to compound as is used routinely in the CD to keep as many variables identical as possible. Colleagues following up on the work presented in this chapter are advised to repeat the FID experiments using 5 equivalences of DNA to compound against all DNA structures and then carry out comparisons to the data presented in this work.
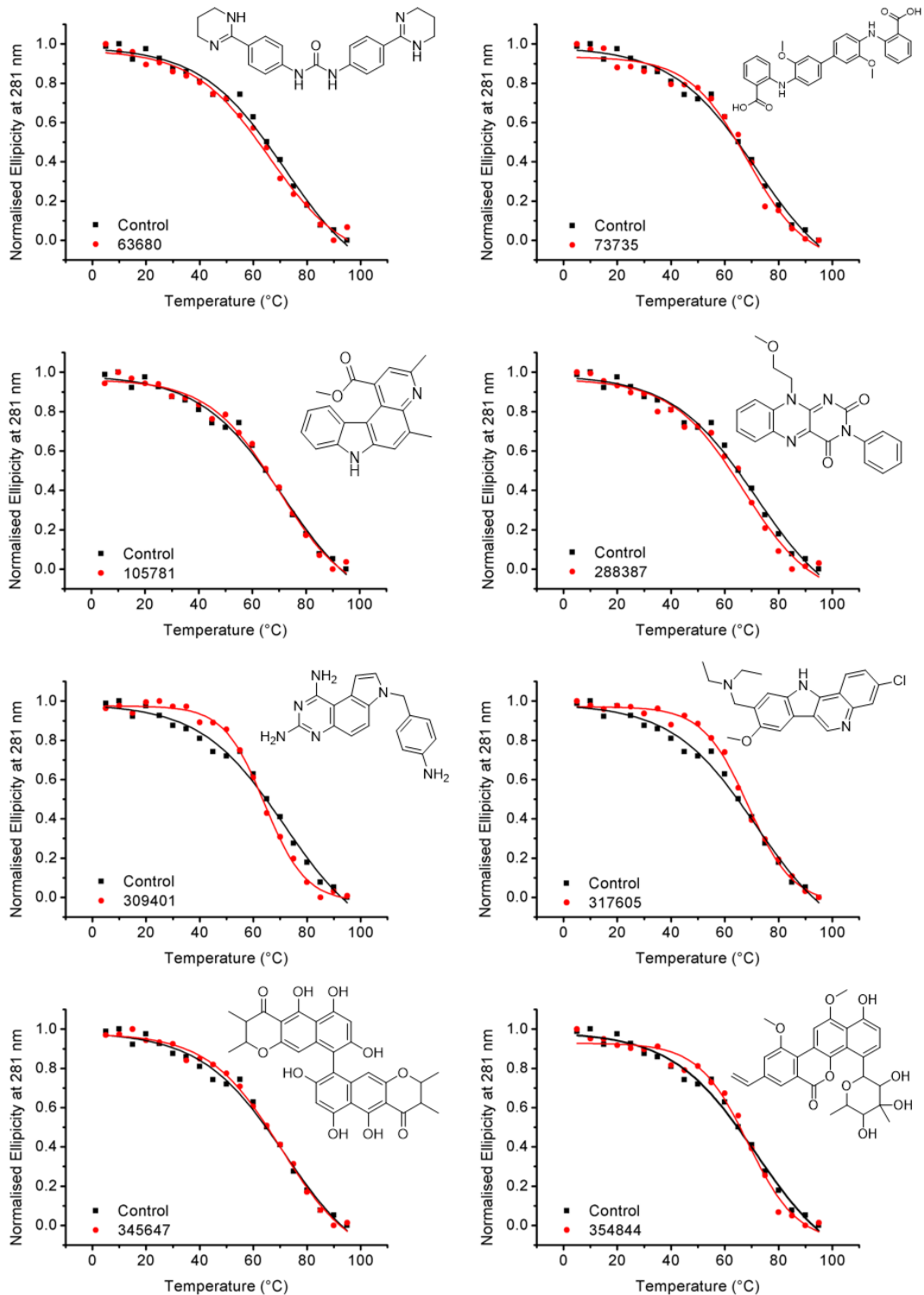
Figure 2.18: Circular Dichroism of normalised ellipiticity at 250 nm, the maximum positive characteristic B-DNA peak, of B-DNA in 10 mM NaCaco pH 6.6 with addition of 50 μM of a given compound. The control (black) is addition of 5 μL of 100% DMSO, this the vehicle control. Line fitting is dose response.

CD was used to assess the binding of the eight compounds of interest against c-MycG to test for specificity (*figure 2.19)*. Notably, the change in melting temperature reported for the i-motif sequences appear small in comparison compared to those reported for G-quadruplex DNA, this is typical of results reported for the i-motif field.[75] All of the eight compounds stabilised the G-quadruplex. Compounds **105781** (+2 °C), **288387** (+2 °C), **73735** (+3 °C), and **309401** (+6.5 °C) all have a ΔTm value of less than +10 °C (*table 2.8)*. Whereas **354844** (+ 10 °C), **345647** (+14 °C), **63680** (+ 17 °C), and **317605** (+ 21 °C) all have a ΔTm value of greater than +10 °C (*table 2.8)*. This data suggests that all eight compounds bind to c-MycG and thus are not specific c-MycC ligands. The data for CD may appear contradictory to the FID data which determined three compounds as binding to c-MycG, **317605**, **345647**, and **354844** (*table 2.6*). However, there were different variables being measured by the two methods as discussed above for B-DNA, as well as the use of different buffers used in FID and CD when exploring binding and effects on stability to c-MycG. For the FID screen for c-MycG experiments 10 mM NaCaco 100 mM KCl pH 6.6 was used and CD screen used 10 mM NaCaco 10 mM KCl pH 6.6. Therefore, the experiments in the CD had 10-fold less KCl, this was due to the c-MycG being too stable with 100 mM KCL to melt within the temperature range of the CD experiments, thus c-MycG is less stable in the CD experiments compared to the FID experiments is an added changed variable between the two methods. Based on this knowledge it is suggested that the FID is repeated as discussed above for B-DNA and also using the buffering conditions required for CD as the FID buffering conditions were not applicable for the CD method to keep as many variables the same between the different experimental methods. Overall CD has determined that all eight of the compounds caused increased stability and thus interact with c-MycG under the CD conditions (*table 2.8*).
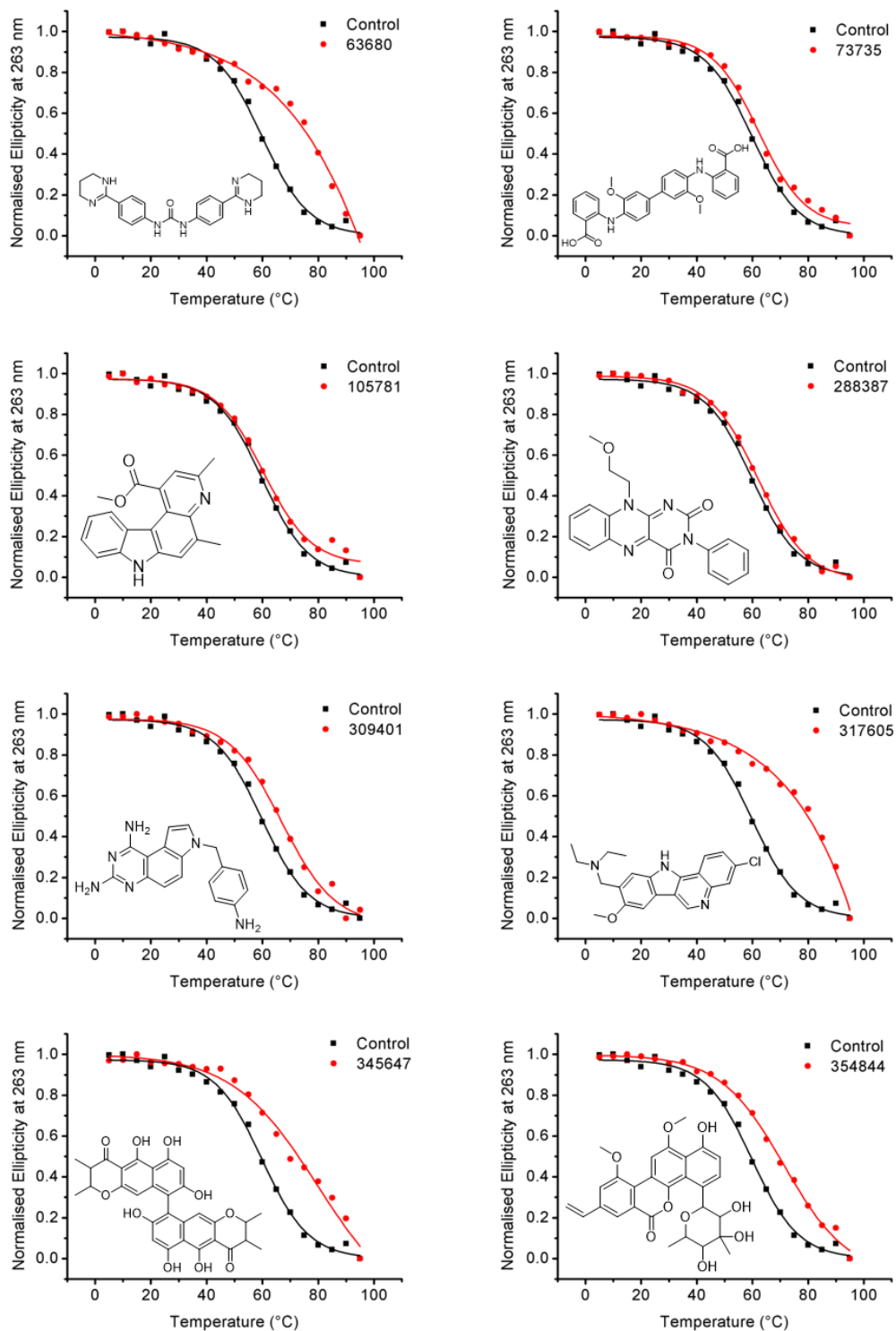
Figure 2.19: Circular Dichroism of normalised ellipiticity at 263 nm, the maximum positive characteristic G-quadruplex peak for c-MycG, of c-MycG in 10 mM NaCaco 10 mM KCl pH 6.6 with addition of 50 μM of a given compound. The control (black) is addition of 5 μL of 100% DMSO, this the vehicle control. Line fitting is dose response.

Overall, CD analysis confirmed that **317605** and **354844** bind to c-Myc52 and determined that they weakly stabilise the structure. Interestingly, these two structures also stabilised c-MycG and destabilised B-DNA, showing that these compounds lack specificity. Therefore, these two compounds could be interesting to explore as although they show non-specific binding, they destabilise B-DNA and stabilise the *MYC* promoter region i-motif and G-quadruplex and thus how these two compounds affect transcription of the *MYC* promoter, and so could be worth investigating. Although, in cells there's a higher percentage of B-DNA and so the destabilisation effects of these compounds on B-DNA are likely to cause multiple consequences if their effects are not isolated to the *MYC* promoter. Yet, these compounds could be used for investigations aiming to understand the role of these two secondary DNA structures on *MYC* transcription if experimental design was considered carefully. Interestingly, **63680** was the best B-DNA destabiliser and the best c-MycG stabiliser, it did slightly destabilise c-MycC27, and wasn't ran against c-MycC52 due to showing no binding in FID. This compound could be of interest as a c-MycG probe. The other compound of interest following the CD results was **345627**, which stabilised c-MycG by >10 °C, destabilised B-DNA and destabilised the first melting point of c-MycC27 but stabilised the second melting point by >25 °C and was not tested against c-MycC52 as FID showed it did not bind. Therefore, this compound may be of interest as a c-MycC or c-MycG probe. All eight compounds were progressed into further studies due to varying results from FID and CD experimental data.

2.3.2 Ligand Binding Studied by Surface Plasmon Resonance

Surface plasmon resonance (SPR), introduced in *Section 1.5.5*, is a technique used to determine binding affinity and kinetics data between targets and ligands. SPR can also be used in a high throughput manner to

detect binding, which is how it was employed for this study. Binding events were determined between the eight hit compounds and eight different 5′-biotinylated DNA structures immobilised on a streptavidin coated chip: B-DNA, c-MycC27, c-MycC52, c-MycG, DAP (i-motif), ATXN2L (i-motif), ILPR (i-motif), and hTeloC (i-motif), refer to *table 5.2* for sequences. All SPR work was performed collaboratively with Dr Clare Stevenson from the John Innes Centre. In SPR, the targets are immobilised onto the chips surface, which is, in this case it is biotinylated DNA, and the ligands are passed over the chip, in this case the hit compounds. SPR determines the amount of analyte bound to the ligand by monitoring the reflected light intensity or tracking resonance angle shifts, allowing for real-time analysis. Light is shone through a prism and glass slide until it reaches the gold sensor which has the ligand immobilised on the surface. This light is reflected into a detector and the intensity and angle are recorded. This initial recording is the baseline in the absence of ligands. The ligands are passed over the chip in buffer and if they bind to the immobilised DNA they cause a molecular weight change, causing a change in intensity and/or angle causing a change in response units (RU) (*figure 1.32*). The SPR experiments used an SA (streptavidin coated) chip with sixteen different flow cells. Eight flow cells contain immobilised DNA and eight are reference cells with nothing immobilised onto the streptavidin coat. Reference cells are the controls which identify how much background binding occurs to the chip itself. This background binding is removed from the binding to the immobilised ligand to ensure the binding read is real.

SPR experiments were performed with each of the eight compounds of interest at a low (10 µM) and high (100 µM) concentration, in the buffering conditions 10 mM NaCaco 10 mM KCl 0.05% Tween20 pH 6.6 1% DMSO. The KCl was added to fold c-MycG into the G-quadruplex structure. This made it increasingly difficult to make direct comparisons to the FID and CD data for the i-motif and B-DNA experiments as the variable of buffering conditions are different. DMSO was required in the buffer as DMSO has a large reflective index and so needs to be in the buffer at the same

percentage it is in the samples to account for this as well as using DMSO correction curves. The results should show a relationship between the compound concentration injected and the size of the relative response observed for those that bind.

SPR determined that compounds **73735**, **288387** and **345647** do not bind c-MycC27 where all other compounds showed binding (*figure 2.20*). Compounds **105781** and **345844** are weak binders with <50 RU at the high concentration (100 µM) and **63680**, **309401**, and **317605** are strong binders with >50 RU at the low concentration (10 µM). The ranking of the c-MycC27 binders from highest relative response units to lowest at 100 µM was **309401**, **317605**, **63680**, **107581** and **354844** (*figure 2.20*).
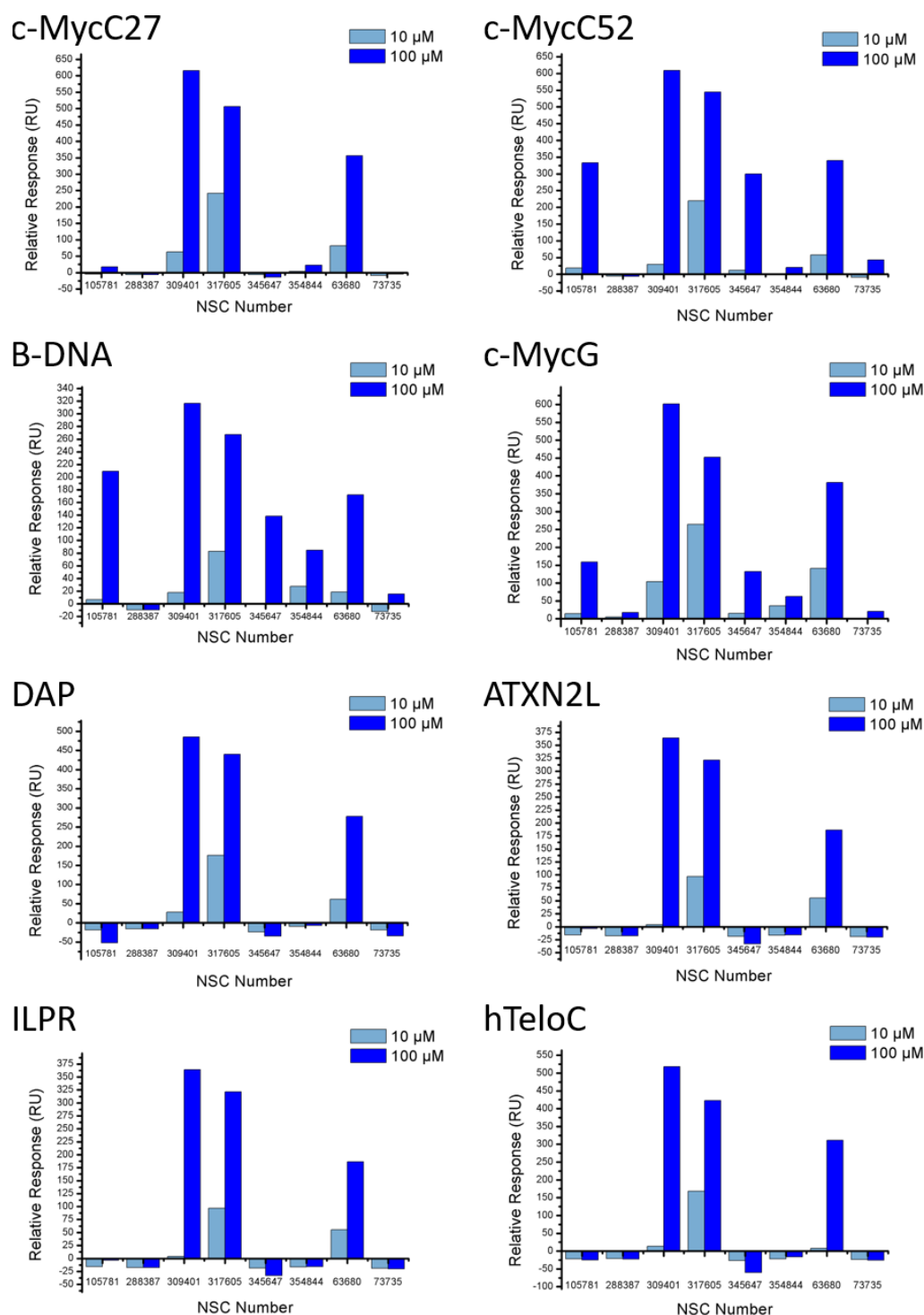
Figure 2.20: Surface plasmon resonance response units vs ligand concentrations at 10 µM and 100 µM of ligands against eight different DNA structures: c-MycC27, c-MycC52, B-DNA, c-MycG, DAP, ATXN2L, ILPR, and hTeloC. DNA immobilised on the chip was ~500-600 RU, compounds 100 µM, in 10 mM NaCaco 10 mM KCl 0.05% Tween20 1% DMSO pH 6.6. Where c-MycC27, c-MycC52, DAP, ATXN2L, ILPR, and hTeloC are i-motif sequences and c-MycG is a G-quadruplex.

Analysing the SPR data for the target, c-MycC52, seven out of eight compounds bound (*figure 2.20*). **288387** did not bind c-MycC52 in the SPR experiment. Compounds **73735** and **354844** were weak binders (<50 RU at 100 µM), **309401** and **317605** were strong binders (>50 RU at 10 µM), and **63680**, **105781**, **345647**, and 354647 showed medium binding (>50 RU at 100 µM). Ranking the compounds from highest relative response units to lowest at 100 µM was **309401**, **317605**, **63680**, **105781**, **345647**, **354647**, **73735**, and **354844**.

Investigating B-DNA as the target by SPR identified that all compounds except **288387** bound to the double stranded structure in the conditions ran, suggesting that these compounds do not show selectivity over B-DNA. There was one weak binder, **73735** (<50 RU at 100 µM). There were four compounds that showed medium binding, **63680**, **105781**, **309401**, **345647**, and **354844** (>50 RU at 100 µM) and only **317605** demonstrated strong binding (>50 RU at 10 µM). Ranking from highest relative response units to lowest at 100 µM for binding to B-DNA was determined as: **309401**, **317605**, **63680**, **105781**, **345647**, **354844**, **73735**.

SPR experiments identified, similarly to c-MycC52 and B-DNA, that all compounds except **288387** bound to c-MycG under these conditions. Two weak binders (<50 RU at 100 µM), **354844** and **73735** were identified, four medium binders (>50 RU at 100 µM), **63680**, **105781**, **309401**, and **345647**, and one strong binder (>50 RU at 10 µM), **317605**. The ranking from highest relative response units to lowest at 100 µM against B-DNA was determined as, **309401**, **317605**, **63680**, **105781**, **345647**, **73735**, and **354844**.

The SPR data demonstrated that for the target sequences DAP, ATXN2L, and hTeloC there were three compounds out of eight identified as binding: **63680**, **309401**, and **317605**. ILPR was similar and had four out of eight compounds identified as binding in the SPR experiments which were: **63680**, **309401**, **317605**, and **345647**. For DAP and ATXN2L **63680** is a

medium binder (>50 RU at 100 μM). and **309401** and **317605** are strong binders (>50 RU at 10 μM). For hTeloC **63680**, **309401**, and **317605** were all medium binders (>50 RU at 100 μM). Analysis of ILPR data identified **354844** as a medium binder (>50 RU at 100 μM) and **63680**, **309401**, and **317605** were all strong binders (>50 RU at 10 μM). The ranking from highest relative response units to lowest at 100 μM against DAP, ATXN2L, ILPR, and hTeloC was determined as **309401**, **317605**, **63680** and then **354844** for ILPR.

To summarise the SPR experiment determined that **63680**, **309401**, and **317605** bound all DNA structures and **288387** bound no structures tested under these conditions (*table 2.9*). Compounds **73735** and **345647** only bound c-MycC52, c-MycG, and B-DNA (*table 2.9*). Whereas compound **105781** bound c-MycC27, c-MycC52, c-MycG, and B-DNA (*table 2.9*). Finally, compound **354844** bound c-MycC27, c-MycC52, c-MycG, B-DNA, and ILPR (*table 2.9*). Compound **73735** bound only to c-MycC52, c-MycG, and B-DNA all with low response and so this ligand is not a good choice for developing into a more specific ligand as it has no preference between those three structures. Furthermore, **354844** had the greatest RU for B-DNA, suggesting a preference for B-DNA and thus this ligand would not be ideal to be take further. Compounds **63680**, **309401**, and **317605** showed the greatest RU for c-MycC27, c-MycC52, and c-MycG and the lowest for B-DNA, suggesting that the weakest binding is with B-DNA and the strongest with the i-motif and G-quadruplex forming structures in the *MYC* promoter region. This suggests they show selectivity over B-DNA for the *MYC* G-quadruplex and i-motif. Furthermore, compounds **105781** and **345647** show the highest RU for c-MycC52 and c-MycG and the lowest for B-DNA. Suggesting stronger binding to the i-motif sequence that Hurley *et al.* determined is required for full transcriptional activation of the *MYC* promoter and the G-quadruplex that forms in this region.[4] Therefore, **63680**, **105781**, **309401**, **317605**, and **345647** are of interest as dual binders to the i-motif and G-quadruplex forming in the *MYC* promoter region with a preference over other DNA structures.
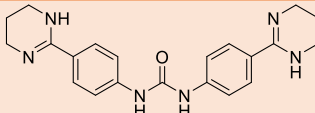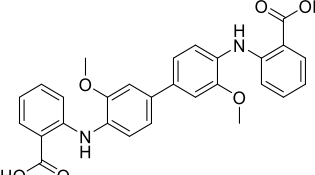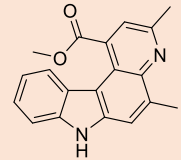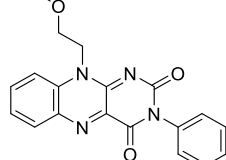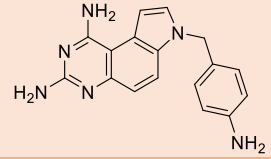
This SPR data should be discussed with the data reported from the CD melting experiments. Notably, the CD melting experiments determined if a ligand caused stabilisation or destabilisation of a DNA structure, not binding, although if stabilisation or destabilisation was observed then one can conclude binding occurred however no effect on stability does not conclude no binding occurred. Whereas SPR determines if binding occurs. Another point to remember is that the buffers differed for CD and SPR. The buffer used in CD melting experiments were buffer 10 mM NaCaco at pH 6.6 for c-MycC27, c-MycC52, and B-DNA, and with the addition of 10 mM KCL for c-MycG. The buffer used for SPR was 10 mM NaCaco 10 mM KCL 0.05% Tween20 1% DMSO at pH 6.6. Therefore, the c-MycC DNA strands are less than 50% folded into the i-motif structures in the SPR conditions due to destabilisation by KCL, compared to there being 50% folded i-motif structures for the c-MycC sequences in the CD conditions. Thus, these add differing variables between the experiments to be considered when comparing.
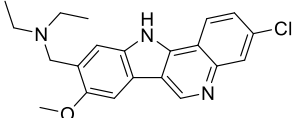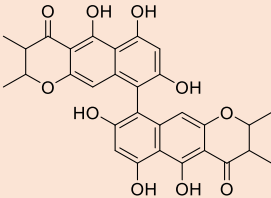
The CD melting experiments identified that **317605** and **354844** caused a change in stability of c-MycC52, c-MycG, and B-DNA and thus bind. This collaborates with the SPR data. Furthermore, compounds **33680**, **73735**, and **345647** were shown to effect stability and thus bind to c-MycC27, B-DNA, and c-MycG, which is supported by the SPR data. Compounds **105781** and **309401** were shown to affect stability and thus bind to B-DNA and c-MycG according to CD data. No change in stability was observed against c-MycC27 and thus it could not be determined if binding occurred. This collaborates with the SPR data which showed both bound to B-DNA and c-MycG and provided further information that they both bound c-MycC27. There was one compound where the CD and SPR data differed. This was for **288387**, where the CD data showed it had an effect on stability and thus bound to B-DNA and c-MycG, but no effect on c-MycC27 stability and thus it could not be determined if binding occurred or not. However, SPR data showed that **288387** bound to no DNA structures tested. This could be due to different variables including buffer differences

or that **288387** precipitated in the SPR experiment and thus showed no binding, or that the changes of stability seen in the CD were not real or within error. A repeat of the experiments would be required to explore the differences seen.

Overall, the SPR data determined that **63680**, **105781**, **309401**, and **317605** are probes that could be developed further with the aim of achieving dual *MYC* i-motif and G-quadruplex specific ligands or specificity to one of those secondary structures.

Table 2.9: Table summarising which DNA structures different compounds bind to, determined from surface plasmon resonance, a tick indicates binding and a cross indicates no binding.

NSC Number	Structure	DNA							
		c-MycC27	c-MycC52	B-DNA	c-MycG	DAP	ATXN2L	ILPR	hTeloC
63680		✓	✓	✓	✓	✓	✓	✓	✓
73735		X	✓	✓	✓	X	X	X	X
105781		✓	✓	✓	✓	X	X	X	X
288387		X	X	X	X	X	X	X	X
309401		✓	✓	✓	✓	✓	✓	✓	✓

317605		✓	✓	✓	✓	✓	✓	✓	✓
345647		X	✓	✓	✓	X	X	X	X
354844		✓	✓	✓	✓	X	X	✓	X

2.4 Cellular Studies

2.4.1 Cytotoxicity Studies

Following biophysical techniques which showed varying results but did identify compounds that bound c-MycC and c-MycG, with the potential to be further developed into c-MycC and/or c-MycG specific probes. Some compounds appeared to show a preference for binding to c-MycC and/or c-MycG and so an investigation into their effects in a cellular environment could be interest, such as a luciferase reporter assay. Before exploring cellular studies, a concentration range where cell death dose not occur is required, and thus cytotoxicity of the compounds was investigated. The colourimetric MTT assay was used. The model cell line of choice was HEK293 cells which are immortalised human embryonic kidney cells.[262] This cell line has multiple advantages including being low-maintenance, rapid dividing doubling every 36 hours, easily transfected and are a well characterised cell line.[263]

The cytotoxicity experiments were carried out on five different compounds, three of interest from this project and two compounds from the literature which are known to interact with i-motif or G-quadruplexes. Of the eight compounds of interest there were three that stood out as most intriguing following CD and SPR analysis. The CD analysis identified that **317605** and **354844** stabilise c-MycC, stabilise c-MycG, and destabilise B-DNA and thus could be interesting to explore their effects on transcription in cellular assays. SPR identified that compounds **317605** and **345647** were of interest as dual *MYC* i-motif and G-quadruplex binding compounds as they had the highest RU for c-MycC52 and c-MycG and the lowest for B-DNA. The other two compounds selected were **Pyridostatin** and **Ellipticine** as they will likely be used as controls in future cellular experiments. This is because **Pyridostatin** is a gold standard G-quadruplex ligand that

stabilises G-quadruplex and destabilises i-motif, and **Ellipticine** is known to stabilise i-motif without causing a direct effect on the G-quadruplex.[97,264]
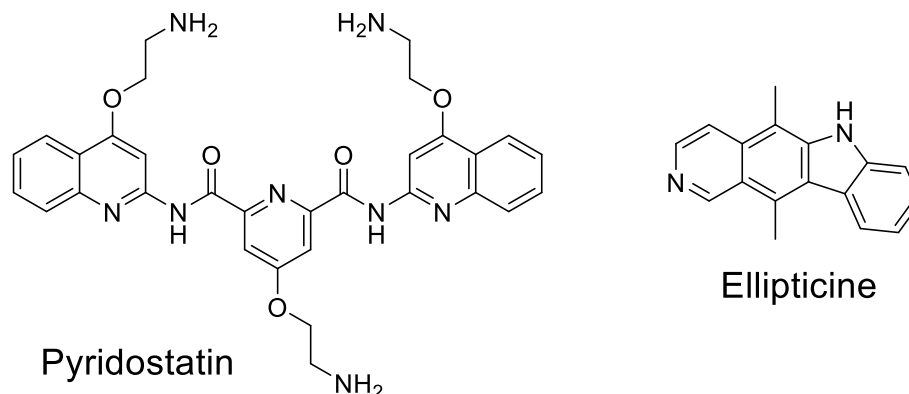
Figure 2.21: Structure of Pyridostatin and Ellipticine.

The cytotoxicity experiments were performed in the presence of 1% DMSO as the vehicle control which was shown to have a minimal effect on cell viability (see *appendix A3*). Cell viability was assessed after 24 hours, which is a typical timescale for luciferase reporter assays investigating *MYC* transcription when assessing the effect of c-MycG binding compounds, as this is the next planned experiment.[171] The average of three biological replicates were used for each concentration. The concentration was logged to produce a sigmoidal shaped curve and was used to determine the IC_{50} value using fitted curves in OriginPro8 software.

Compounds **317605**, **345647**, and **Ellipticine** showed a typical sigmoidal relationship between concentration and cell viability (*figure 2.22*). Compound **354844** had a hyperbolic curve and demonstrated high cytotoxicity levels. Compound **317605** has a more gradual cytotoxicity in the concentration range used than **345647** or **Ellipticine**. Compound **345647** showed the steeper curve than **317605** demonstrating the greater cytotoxicity. These observations were reflected by the IC_{50} values in *table 2.10*, where compound **354844** had the greatest cytotoxicity (DC_{50} 0.2 µM), then **345647** (DC_{50} 2.4 µM), **Ellipticine** (DC_{50} 11.5 µM), and **317605** the least cytotoxicity (DC_{50} 37.5 µM). **Pyridostatin** did not significantly

reduce cell viability below 80% at any tested concentration, this is because the **Pyridostatin** stock was a 10-fold lower concentration in DMSO than all of the other compound stocks due to insoulublity of the particular form of **Pyridostatin**, leading to a 10-fold lower max concentration being tested in the assay (*figure 2.22*). Using higher concentrations in the assay from that stock would have a higher percentage of DMSO than 1% and thus causing cell death, therefore this positive control requires repeating at an identical concentration range to the other compounds in 1% DMSO where one would then expect to see cell death caused my **Pyridostatin** in line with the literature.[265] Therefore, **Pyridostatin** could not be ranked as further concentrations were required to do so.
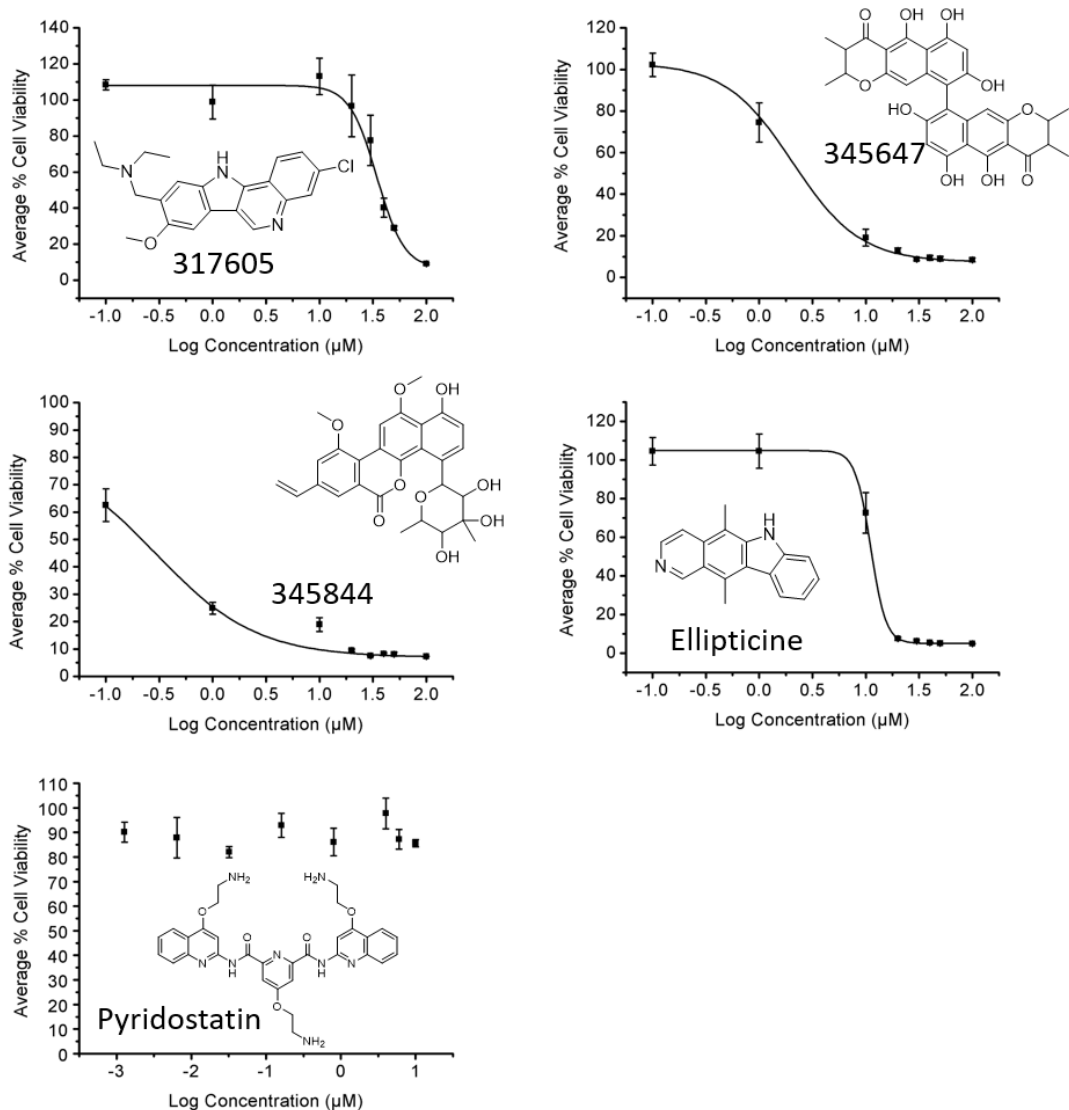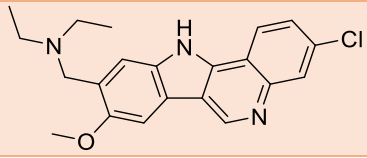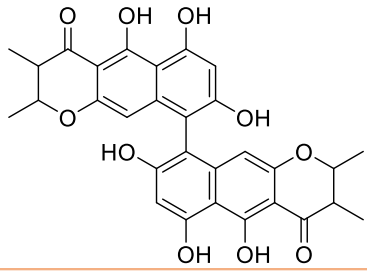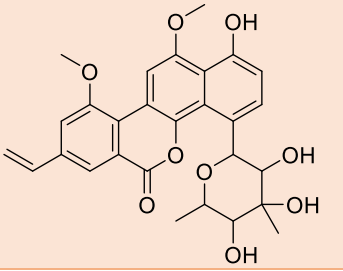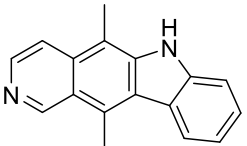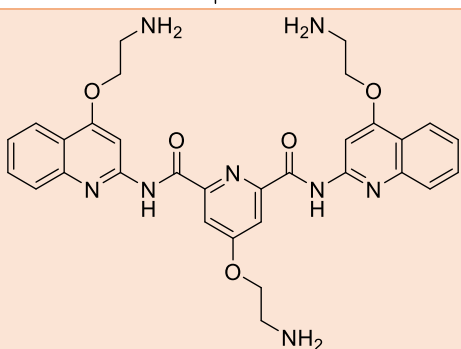
Figure 2.22: Average percentage cell viability of HEK293 cells after 24 h treatment with indicated compounds. Viability is normalised to 1% DMSO vehicle control. Error bars are 3 replicates of biological controls. The best fit lines are sigmoidal dose-response relationships plotted in OriginPro8.

Table 2.10: Summary of the IC$_{50}$ values determined from MTT cytotoxicity experiments in HEK293 cells after 24 h treatment with indicated compounds.

Compound	Structure	IC$_{50}$ (μM)	Fitting Error
317605		37.5	0.00276
345647		2.42	0.02455
354844		0.22	0.23108
Ellipticine		11.5	0.02319
Pyridostatin		>10.0	N/D

2.5 Discussion and Further Work

Chapter 2 explored identifying lead small compounds that interacted with c-MycC that could be later developed into c-MycC specific probes. To achieve this aim, this project tested the compounds from the National Cancer Institute Diversity set IV library which is comprised of pharmacophores against two variations of the i-motif sequence that forms in the *MYC* promoter region. c-MycC27 which was commonly used in the i-motif field and c-MycC52 which was determined to be required for full transcriptional activation of *c-Myc*.[4,7] To screen this library the high throughput method FID assay was used to identify compounds that bound to either of the targets, c-MycC27 or c-MycC52. This was followed up by further FID assays to investigate specificity, FID validation assays, FID titrations to determine DC_{50} values, CD melting studies, SPR binding studies, and MTT assays to investigate cytotoxicity of compounds of interest (*figure 2.23*)
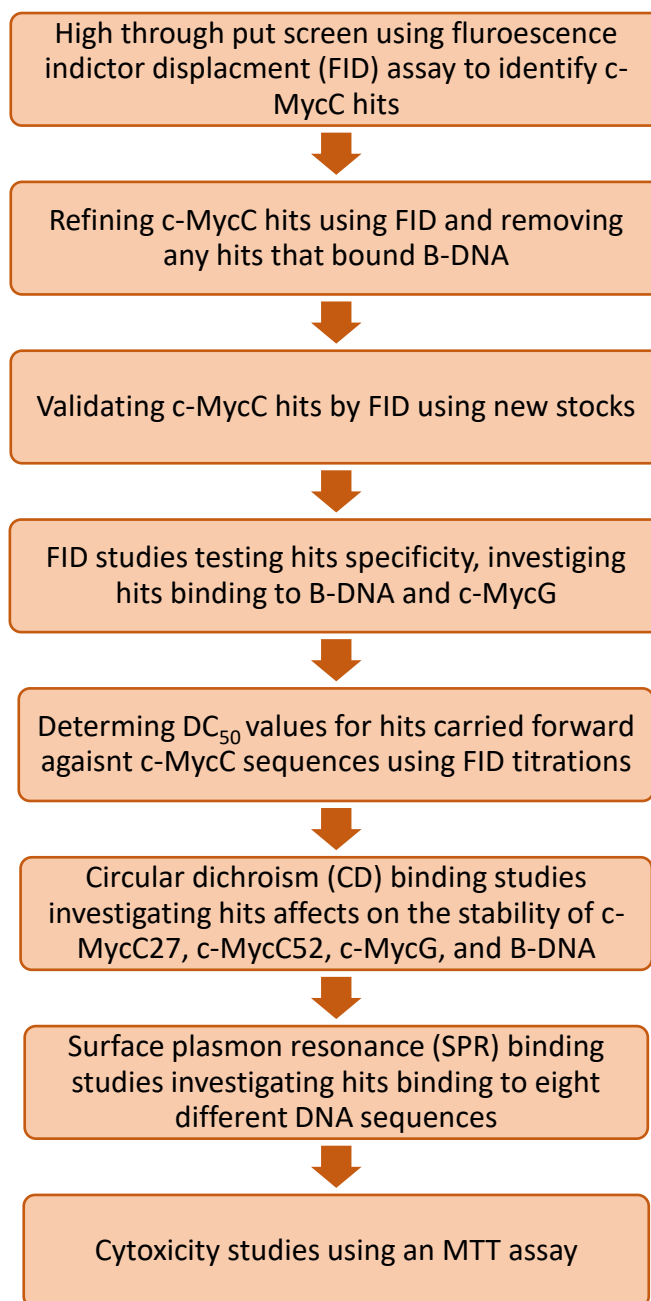
Figure 2.23: Flow chart detailing the process of screening compounds that bound to the i-motif forming region in the MYC promoter.

The high throughput FID screen of >1500 compounds identified that 34 compounds bound to either/both c-MycC sequences investigated. These compounds of interest, **63680**, **73735**, **105781**, **288387**, **309401**, **317605**, **345647**, and **354844** (*figure 2.24*). The binding of these compounds was confirmed using CD and SPR techniques where all eight showed binding to either c-MycC sequence in either or both techniques. Therefore, this

project achieved identifying c-MycC interacting ligands and explored the specificity of these compounds identifying novel compounds.
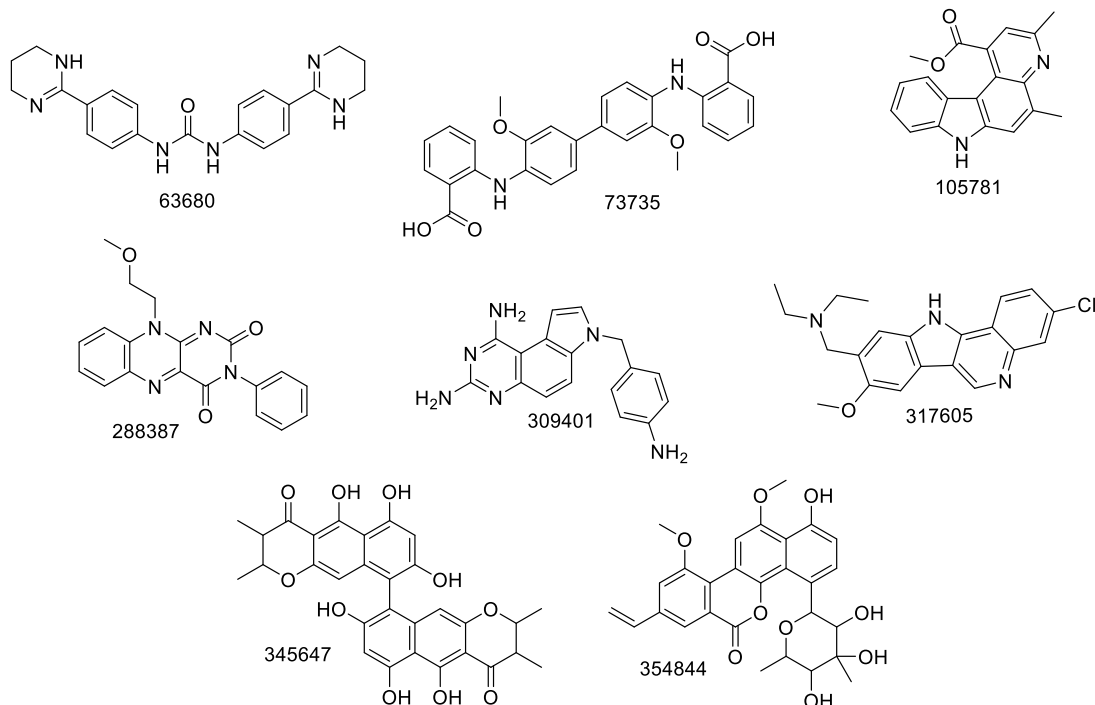
Figure 2.24: Schematic of compounds with the following NSC numbers 63680, 73735, 105781, 288387, 309401, 317605, 345647, and 354844.

The i-motif field has a limited number of known i-motif interacting compounds, many of which have been inspired by compounds that interact with its sister structure, the G-quadruplex.[48,252] For example, the first known i-motif interacting compound, reported in 2000 by Hurley and colleagues, was TMPyP4, a known G-quadruplex binding compound.[266] Naturally, the i-motif field has continued the search for i-motif interacting compounds using G-quadruplex ligands characteristic cores as a starting point, such as acridine cores.[213] This is proving to be successful in finding i-motif binding ligands, such as **B19** (*figure 2.2*),[82] which was shown not to interact with c-MycG or B-DNA but did stabilise c-MycC and had a repressive effect on the *MYC* promoter. Interestingly, the biophysical data for **B19** used c-MycC27, not c-MycC52 which Hurley's team identified as

being required for full biological activity yet it has an effect on *MYC* transcription.[4,7] From the work carried out in this project it is clear that compounds are likely capable of binding to all i-motif sequences, such as **63680** (*table 2,9*), and other compounds show a preference to certain i-motif structures, such as **354844** (*table 2.9*). Therefore, it would be interesting to see if **B19** also binds to c-MycC52 sequence, and other i-motif sequences. Consequently, **B19** could be an i-motif specific sequence rather than c-MycC specific which could account for the biophysical data and biological data seen.

Additionally, there have been multiple i-motif binding ligands identified that were not inspired by G-quadruplex interacting compounds. Including mitoxantrone by the Waller group from screening a compound library.[235] Similarly, the *Bcl2* i-motif binding ligand IMC-48 was also identified through screening of a chemical library.[86] Highlighting, that using compound libraries that contain varied chemical compounds is an excellent method used to increase the number of i-motif binding compounds and also in identifying characteristic cores of i-motif binders that may differ from those of known G-quadruplex binders. This project used a diverse compound library and identified new leads to be developed as c-MycC specific probes, **63680**, **73735**, **105781**, **288387**, **309401**, **317605**, **345647**, **354844** (*figure 2.24*).

This project has identified three novel probes of particular interest for further development, **317605**, **345647**, and **354844** (*figure 2.24*). Overall, the biggest challenge was getting selectivity over B-DNA for c-MycC, this is unsurprising as this is the currently challenge of the i-motif field. There are multiple reasons that could account for the lack of selectivity such as, B-DNA and i-motif being composed of the four nucleic acid bases, both having a major and minor groove, and both being able to bind to intercalators. Its not uncommon for compounds binding i-motif DNA to promiscuous and bind other DNA structures, in particular B-DNA.

Compound **317605** was identified by FID as biding to c-MycC52 and not to B-DNA (*tables 2.3* and *2*.4). Further studies identified that **317605** stabilises c-MycC52, stabilises c-MycG and destabilises B-DNA as shown by CD melting studies (*table 2.8*). Suggesting a lack of specificity over B-DNA. SPR identified that this compound, **317605,** demonstrated binding to all target sequences, which is not unsurprising as this compound is an indoloquinoline, which are known to interact with different DNA structures including B-DNA and G-quadruplexes.[267] Interestingly, **317605** had the highest response for c-MycC and c-MycG and the lowest for B-DNA (*table 2.9* and *figure 2.20*). Suggesting, that **317605** has a binding preference for c-MycC52 and c-MycG over other B-DNA and even other i-motifs. It has been found that indoloquinoline derivatives have high affinity for G-quadruplexes and so it is likely that this compound has a preference for i-motif and G-quadruplex over B-DNA. This is of particular interest as this unique property that could be further developed to make this compound a dual specific *MYC* i-motif and G-quadruplex probe. In the literature there is one example of a compound that selectively binds to c-MycC and c-MycG and stabilises both structures, **a9** *(figure 2.*2), similarly to the results identified in this project for **317605**.[87] However, the stabilising was determined from FRET and CD titrations which monitored the intensity of characteristic peaks with and without compound present but the gold standard CD melts were not ran and the effect of **a9**, which was shown by SPR to bind B-DNA, on B-DNA's stability was also not determined. As **a9** is a bisacridine derivative and bisacridines are known bisintercalators that interact with multiple DNA structures such as B-DNA and holiday junctions, thus further experiments testing binding against holiday junctions would be of interest.[268–270] Therefore, **317605** is novel as the data suggest it preferentially binds c-MycC and c-MycG, stabilises c-MycC, stabilises c-MycG and destabilises B-DNA. Therefore, this is an excellent starting point for developing a specific dual c-MycC, c-MycG probe.

Compound **345647** is another novel compound identified that has potential to be further developed into a dual *MYC* i-motif and G-quadruplex probe (*figure 2.24*). **345647** was highlighted as of particular interest as SPR experiments determined that it only bound to c-MycC, c-MycG, and the response for B-DNA was ~1/2 of that seen for c-MycC and c-MycG (*figure 2.20*). This was further supported by CD experiments which determined **345647** destabilises B-DNA, stabilises c-MycG, no effect on c-MycC27, and wasn't tested against c-MycC52. The third novel compound of particular interest was **354844** (*figure 2.24*). CD determined it stabilised c-MycC52 and was the best c-MycC52 stabiliser of the compounds investigated in this project, stabilised c-MycG by >10 °C (ranked in the high stabiliser group), and destabilised B-DNA (*table 2.8*). Thus, this is of interest to be further developed to explore as a therapeutic that could regulate transcription of *MYC*.

To further progress this work the two compounds that appeared to have a binding preference for c-MycC and c-MycG, **317605** and **345647**, need to have their binding constant, K_D, determined. To do this SPR should be used at a range of concentrations form 0 µM to 100 µM with at least 10 concentration points. If that data showed preferential binding SPR could be used to explore a wider number of DNA structures, such as more G-quadruplexes, and triplexes especially as **317605** is a indoloquinoline derivatives which are known to have a high affinity for triplexes.[267] Additionally, the broader context of this project is to produce probes that can be further developed and explored as therapeutics and thus cellular studies must be carried out. Compounds **317605**, **345647**, and **354844** were identified of particular interest and their cytotoxicity after a 24 h incubation in HEK293 cells was determined. Using those results concentrations where the compounds are not cytotoxic can be selected for essential cellular assays to study the compounds effects on transcription. One cellular assay which has been used to study the effect of G-quadruplex binding ligands on *MYC* transcription is a luciferase assay and so this

could be carried out to explore the effects of these compounds on *MYC* transcription.

In addition, there is future work that can be explored using all eight of compounds identified as c-MycC hits that would benefit the fields understanding of i-motif binding ligands. One is what epitopes of different compounds bind to the i-motif to develop the field understanding of what i-motif ligands look like and important chemical properties. Saturation transfer difference (STD) NMR would be used to determine the binding epitope of the ligand when bound to c-MycC. The second would be exploration of the effects of buffer conditions on binding to different DNA structures could be an interesting avenue, especially as in the SPR conditions (10 mM NaCaco 10 mM KCl 0.05% Tween20 1% DMSO pH 6.6) **288387** showed no binding to any DNA structures, yet in the CD under similar conditions (10 mM NaCaco pH 6.6 or 10 mM NaCaco 10 mM KCl pH 6.6) it showed destabilisation of B-DNA and stabilisation of c-MycG, and thus must interact with these structures to change their stability. Furthermore, another binding technique that can also use varying conditions whilst investigating binding may need to be used such as isothermal titration calorimetry (ITC).

Furthermore, synthesis of analogues of these eight compounds or the three of particular interest aiming to improve affinity and specificity to c-MycC for c-MycC probes or c-MycC and c-MycG if working to obtain a dual probe. Followed by biophysical techniques used throughout this chapter to determine if the aims were achieved. The compound with increased affinity or specificity for the desired DNA targets would be progressed into cellular experiments to determine effects on *MYC* transcription aiming to gain a better understanding of the mechanism by which the *MYC* promoter is regulated by the i-motif and to be a starting point for developing powerful cancer treatments that target *MYC* regulation.

Chapter 3: Searching for Peptides to Target the i-Motif forming sequences in the Promoter Region of c-*Myc*

3.1 Introduction

The 'master regulator', *MYC*, is known to control many aspects of cellular growth and cellular metabolism, making *MYC* regulation essential for normal cell functioning.[121,271] c-Myc takes a central role in sustaining higher levels of metabolism such as that seen in cancer cells. Unfortunately, despite efforts to target the c-Myc protein directly, it has proven challenging and largely unsuccessful.[272,273]

A possibility being explored is targeting the promoter region itself, particularly the NHE III$_1$ region as it is capable of forming alternative DNA structures, the G-quadruplex and i-motif, which are gaining more and more evidence showcasing their ability to regulate transcription[170]. In fact, stabilising these structures has been demonstrated to reduce *MYC* expression.[87,274] Thus, they could be targets for powerful therapeutics. The current published studies that have identified probes that preferentially bind c-MycC were small molecules,[6,82,87] similarly to the approach used in *Chapter 2*. One rationale for this may be that small compounds are more likely to be able to pass through the nuclear pore as well as being a classic route for drug discovery. Although, an increasingly popular route is therapeutic peptides.[275] In general, this is advantageous due to high target specific and selectivity, fewer side effects, high potency of action, and low toxicity to name a few. However, they do have some disadvantages which include metabolic instability, poor membrane permeability, poor solubility and higher manufacturing costs.[273] Although, peptides, <~50 kDa, may be able to pass through the nuclear envelope freely.

Broadly speaking, it is ideal for the i-motif field to have a library of different i-motif specific binders made up of small molecules, peptides, and proteins which have the capacity to stabilise, destabilise, and have no effect on stability. This would allow the field to investigate the understanding and function of i-motifs using multiple different techniques and manipulate them in a multitude of ways. The same will be required for the *MYC* i-motif found in the *MYC* promoter region (c-MycC). Specific probes would enable a fuller understanding of i-motif biology and hopefully pave the way for some new therapeutics that could revolutionise cancer treatments.[276]

To date, the only i-motif specific protein identified is the antibody fragment iMab.[94] This discovery was one of the most significant for providing evidence that i-motifs formed in cells and are physiologically relevant structures. Christ's team utilised antibody phage display to identify iMab.[94] Phage display is a molecular biology technique used to study protein-protein, protein-peptide, protein-DNA, and peptide-DNA interactions, in this case peptide-DNA interactions. Its discovery by George Smith and adaptation to the antibody display format by Greg Winter was recognised in the 2018 Nobel Prize in Chemistry.[277] The direct link between phage genotype and phenotype and the ease of using phage vectors are two key reasons why this is an excellent screening technique.[278]

This chapter questioned whether phage display could be used to find a peptide that was c-MycC specific.[94] A peptide was chosen over a protein in this instance as the aim of this project was to identify a probe that could be used to analyse the effects of c-MycC on expression and therefore a smaller molecule is more optimal as entry into the nucleus may be achievable freely using a small peptide, whereas a protein must use active transport.[273]

3.2 Phage Display Screening of Ph.D.12 and Ph.D.-C7C Libraries

To address the aim of identifying peptides that bind c-MycC this chapter used two different length variations of c-MycC. These two oligonucleotides were 27 and 52 base pairs long, referred to, respectively, as c-MycC27 and c-MycC52. It was decided that two commercially available libraries would be tested: Ph.D.™12, a linear randomised 12-mer library, and Ph.D.™-C7C, a cyclic 7-mer library where the 7-mer randomised sequence is flanked by a pair of cysteine residues. Both libraries consist of circa 10^9 randomised peptide sequences with approximately 100 copies of each present in the library stock. It was hypothesised that the cyclic library had a lower entropic penalty for binding due to the structural constraint by the disulfide compared to the linear library. There was concern that a 7-mer sequence may be too short to provide specific binding and so a 12-mer library was also used (*figure 3.1*).
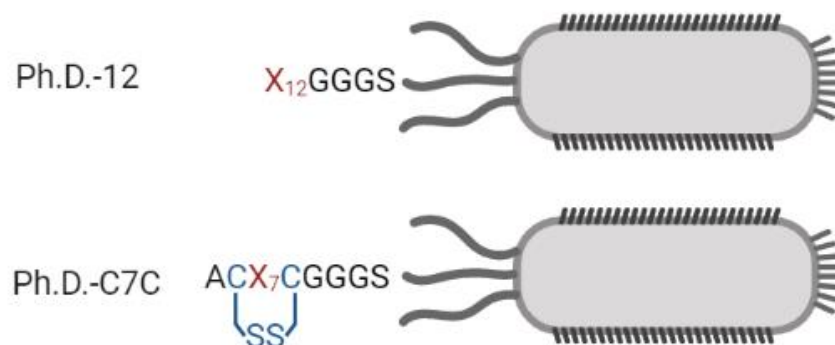
Figure 3.1: Peptide libraries used from New England BioLabs. The peptide sequence is displayed on the pIII minor coat protein on the distal tip of the M13 phage and contains a linker sequence GGGS shown in black. In the cyclic library, the flanking cytosine residues are shown in blue, the X demonstrates randomised amino acids, and the subscript details the number of randomised amino acids in each sequence. Created in BioRender.com.

3.2.1 Phage Display Screening Method Development and Results Targeting the 27bp i-Motif Forming Sequence in the Promoter Region of *MYC*

3.2.1.1 Phage Display Screening Method 1 and Results Targeting the 27 bp i-Motif Forming Sequence in the Promoter Region of *MYC*

The method was developed from the Ph.D.™ Phage Display Libraries instruction manual from New England BioLabs (NEB), see Surface Panning Procedure, for further details see *section 5.1.2*.[197] The buffer suggested in this section is 0.1 M NaHCO$_3$ pH 8.6, however this is not appropriate for studies i-motif interactions as acidic pH is required to stabilise the i-motif structure,[36] and the buffering range of bicarbonate is pH 8.9-10.8[279] making this buffer unsuitable. NEB suggests using a buffer to stabilise the target containing metal ions and of a similar ionic strength. Therefore, PBS was chosen as it has a buffering range of pH 5.8-8.0, which spans the pH range typically used for testing i-motif folding.[280] The blocking buffer suggested was 0.1 M NaHCO$_3$ pH 6.8 with 5 mg/ml BSA, this was changed to PBS at acidic pH 5 mg/ml BSA and will be referred to PBSB. The wash buffer suggested in the manual was TBST (TBS + Tween20), again this buffer was unsuitable as it's good buffering range is pH 7-9.2 and therefore PBS was used with an addition of Tween 20, and will be referred to as PBST.[281] The Tween 20 percentage was adjusted throughout the method development process to find the optimal percentage to reduce background interactions. The final buffer adjusted was the elution buffer, this is used to elute bound phage from the target and so this buffer needs to disrupt the interaction between the target and phage. In this case the best approach to this was to continue using PBS and to increase the pH to 7.4.[280] As shown in *figure 3.2* pH 7.4 causes both i-motif sequences to unfold and therefore the peptide binding sites will be

deformed and the phage will unbind. Further support for using PBS comes from the phage display conditions used by Christ's group when identifying iMab as they also used PBST and PBSB.[94]
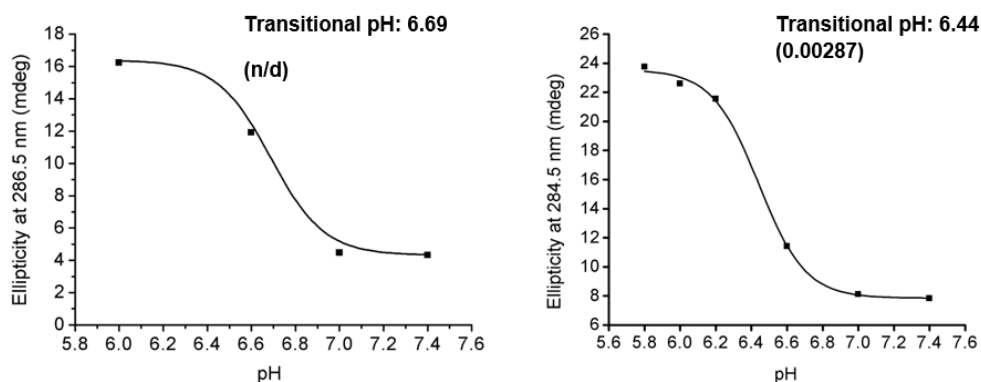
Figure 3.2: CD characterisation of c-MycC27 (left) and c-MycC52 (right) in PBS from pH 7.4 to 6.0 to determine the transitional pH, the pH at which the sequence is 50% folded (i-motif) and 50% unfolded (random coil). The graph was fitted with a dose response curve, the error of this fitting is shown in brackets on the graph, which was used to determine transitional pH in Origin software.

To determine the optimal pH to use for the phage display selection, the 27 base pair variant of the i-motif that forms in the promoter region, c-MycC27, was characterised in PBS using CD. Which determined the transitional pH as pH 6.69 (*figure 3.2*), therefore pH 6.6 was decided as the pH for PBS buffers.

Phage display method 1, shown in detail in *table 5.4* and summarised in *table 3.1*, aimed to find any peptides within the New England Biolabs (NEB) Ph.D.-C7C library that bound to c-MycC27. Round 1 used c-MycC27 as target DNA to identify any peptides in the library that bound c-MycC27, for details see *table 5.4*. For round 2 the selection pressure was increased by adding competitive non-biotinylated DNA/RNA into the mix at 100-fold higher concentration than c-MycC27, refer to *table 5.4*. These DNA structures included multiple different possible secondary structures other than an i-motif structure and were used to increase the selection pressure.

The competitive DNA was added at a higher concentration with the aim of removing any peptides that bind c-MycC27 but also other DNA structures. Even if the peptides had a slightly stronger affinity to c-MycC27 than the other structures, the higher concentration of the competitive sequences should mean they are still removed. The selection pressure was increased again during round 3 using other i-motif sequences as competitors, refer to *table 3.1* for conditions and *table 5.1* for DNA sequences.

Table 3.1: Phage display method 1 for panning against the target DNA sequence c-MycC27.

Round	Plate or Beads	Phage Library	Phage	Target DNA	Blocking Buffer	Wash Buffer	Competitors added to Phage Mixture
1st	Beads	Ph.D.-C7C	1x10^11 Phage Library, Ph.D.-C7C, in 500 μL PBS pH 6.6	10 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	None
2nd	Beads	Ph.D.-C7C	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 μL PBS pH 6.6	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 01% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-Hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU
3rd	Beads	Ph.D.-C7C	1x10^11 Amplified Phage Selection 2 Elution 3 in 500 μL PBS pH 6.6	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP

Sequencing of 10 samples obtained from round 3 determined two instertless sequences (*table 3.2*). Indicating that there was a lack of strong selection pressure for c-MycC27 binders. Insertless phage should have been removed in selection round one. If insertless phage are present during the amplification round it is possible they will amplify at a faster rate than the recombinant phage due to the absence of the displayed peptides that might interfere with the interaction of pIII minor coat proteins and the F-pilus during *E. coli* infection.[282] Overall, the presence of insertless phage suggests that the method needed further development.

Table 3.2: Sequencing of peptides using one letter code identified as binding to c-MycC27. Identified from phage display method 1, round 3, selecting against other DNA structures.

Sample	Peptide Sequence
1	MAPDSRV
2	VQMPAHS
3	EFSKFRS
4	Insertless
5	TQRSHSS
6	DLLHRGA
7	LAQSHPL
8	DGHDQSL
9	DTSTKYL
10	Insertless

3.2.1.2 Phage Display Screening Method 2 and Results Targeting the 27 bp i-Motif Forming Sequence in the Promoter Region of *MYC*

One way to optimise phage display selection is by increasing the selection pressure in each round and by performing negative selection round in which the library is depleted of peptides that bind to an undesirable structure related to the target. Method 2, shown in detail in *table 5.4* and summarised in *table 3.3,* was performed similarly to method 1, *table 3.*1. The main difference was that round 2 was a negative selection procedure against single stranded c-MycC27, *table 3.3*, the selection round was performed in PBS at pH 7.4 so that c-MycC27 was unfolded. Therefore, the washes from round 2 should have contained peptides that did not bind the unfolded c-MycC27 but bound the folded c-MycC27 i-motif structure. The peptides in round 2 washes were progressed into round 3. Round 3 was

performed identically to round 1 using less washes, *table 5.4.* Method 2 round 4 was performed identically to method 1 round 3, *tables 3.1* and *3.3.* Following the final selection round 14 different sequences were sequenced shown in *table 3.4.* Unfortunately, 11 of these sequences were determined insertless and the five sequences identified had no consensus between the sequences and all differed from those identified from method 1. This suggested further method development was required as the phage display system appeared not to be able to identify strong c-MycC27 binders using the methods so far discussed.

Table 3.3: Phage display method 2 for panning against the target DNA sequence c-MycC27.

Round	Plate or Beads	Phage Library	Phage	Target DNA	Blocking Buffer	Wash Buffer	Competitors added to Phage Mixture
1st	Beads	Ph.D.-C7C	1x10^11 Phage Library, Ph.D.-C7C, in 500 μL PBS pH 6.6	10 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	None
2nd	Beads	Ph.D.-C7C	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 μL PBS pH 7.4	10 pmol c-MycC27	PBS pH 7.4, 5% BSA	PBS pH 7.4, 0.1.% Tween 20	None
3rd	Beads	Ph.D.-C7C	1x10^11 Amplified Phage Selection 2 Wash in 500 μL PBS pH 6.6	10 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	None
4th	Beads	Ph.D.-C7C	1x10^11 Amplified Phage Selection 3 Elution 3 in 500 μL PBS pH 6.6	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-Hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU

Table 3.4: Sequencing of peptides using one letter code identified as binding to c-MycC27. Identified from phage display method 2, round 4, selecting against other DNA structures.

Sample	Peptide Sequence
1	WKHGNFN
2	Insertless
3	FPNSMYQ
4	Insertless
5	Insertless
6	Insertless
7	Insertless
8	Insertless
9	Insertless
10	Insertless
11	Insertless
12	TNLSPRA
13	HTEALHA
14	QPDLPGV
15	Insertless

3.2.1.3 Phage Display Screening Method 3 and Results Targeting the 27 bp i-Motif Forming Sequence in the Promoter Region of *MYC*

Following the disappointing results from method 1 and 2 it was decided to investigate pH as one of the most important factors related to i-motif formation.[36] The pH was decided using the data presented in *figure 3.2,* which determined the transitional pH as 6.69 for c-MycC27 in PBS. The Waller lab usually works with c-MycC27 in 10 mM NaCaco at the transitional pH 6.6.[250] To compare the CD spectrum of c-MycC27 in these two buffers were overlayed (*figure 3.3*). It can clearly be seen that PBS at pH 6.6 gave a lower ellipicity than in 10 mM NaCaco at pH 6.6 and so PBS pH 6.0 was also overlayed (*figure 3.3*). *Figure 3.3* demonstrates that the amount of c-MycC27 formed in 10 mM NaCaco pH 6.6 is more similar to PBS pH 6.0 (*figure 3.3*). Therefore, the initial phage display screens may have contained less than 50% folded c-MycC27 target, which would have reduced the selection pressure for i-motif binders. This could explain why

there was an increase in insertless sequences identified when unfolded c-MycC27 was negatively selected against in method 2, compared to method 1. Furthermore, it must be considered that the PBS buffer would be autoclaved before use in phage display and that autoclaving is known to adjust pH for multiple different reasons, one is due to the expulsion of CO_2 during the sterilisation process making the buffer more alkaline. This is because the solubility of gases, CO_2 in this case, decreases as the temperature increases, this is based on Henry's law.[283] Usually this is not a significant change for most biological studies, however when working with a highly pH-sensitive i-motif this can affect the percentage of the sequence that is folded. This feature has been exploited by researchers aiming to use i-motif structures to monitor pH changes.[284] Unfortunately, the buffer cannot be pH corrected after autoclaving as the pH facilities are unsterile. Therefore, to account for this and the based on the data presented in *figure 3.3* the slightly more acidic pH, pH 6.0, was used for the next method.
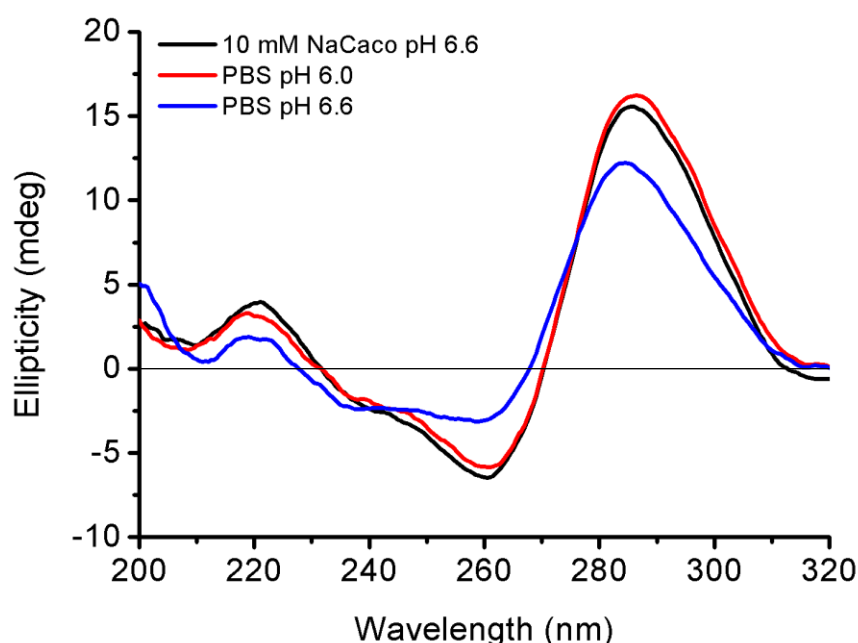
Figure 3.3: CD comparison of 10 μM c-MycC27 in buffers 10 mM NaCaco pH 6.6, PBS pH 6.0, and PBS pH 6.6.

The other factor changed for method 3 was the library being used, (*table 3.5*). New England BioLabs are unable to predict which library will be the most suited to your target without knowledge of the target-ligand interaction. The Ph.D.-C7C library was originally chosen as the starting point as it was structurally constrained which may be more optimal for binding to surface loops in proteins and therefore could be better at binding the loops in i-motif structures. However, after three of four rounds of selections with no consensus sequence identified it could suggest that the library does not bind with sufficient affinity to be selected and so another library may be a better option. Therefore, the Ph.D.-12 library has an advantage over the Ph.D.-C7C library as it could be used for targets that required 7 or fewer residues for binding that could not be contained within the 7-residue library.[285] As an example, if the amino acids ASEGP were a consensus sequence but they needed to be separated by four different amino acid residues, ASEXXXXGP, this would have required nine amino acids and so cannot be found in the 7-mer library but can be in the 12-mer library. Furthermore, the Ph.D.-12 library peptides are long enough that they can begin to present structural elements which could be useful assuming structural elements are needed when biding to c-MycC27.

Table 3.5: Phage display method 3 for panning against the target DNA sequence c-MycC27.

Round	Plate or Beads	Phage Library	Phage	Target DNA	Blocking Buffer	Wash Buffer	Competitors added to Phage Mixture
1st	Beads	Ph.D.-12	1x10^11 Phage Library, Ph.D.-12, in 500 µL PBS pH 6.0	10 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None
2nd	Beads	Ph.D.-12	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.0, 0.1% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-Hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU
3rd	Beads	Ph.D.-12	1x10^11 Amplified Phage Selection 2 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.0, 0.1% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP Hif1α

Following these discussions on pH and library selection for method 3 the pH used for buffers was pH 6.0, other than the elution PBS buffer which was kept that pH 7.4 and the library being screened was Ph.D.-12 (*table 3.5*). Everything else about method 3 was kept identical to method 1. Sequencing of 10 different sequences revealed that 40% insertless sequences, 20% Pep-SQD (sample 2 and 10), 30% Pep-MHP (sample 3, 7 and 9), and 10% Pep-SGV (sample 6) (*table 3.6*). The sequences were analysed for a consensus using multiple sequence alignment Clustal Omega.[286,287] The is a similarity between Pep-SQD and Pep-SGV which is $X^1XXXXWX^1$ where X is any amino acid and X^1 is an amino with a property of basic/amine, refer to *figure 3.4*. There is also a similarity between Pep-SGV and Pep-MHP $X^1X^1X^2XX^3$ where X^2 is an amino acid that is small and X^3 is basic (*figure 3.4*). A consensus is also found between Pep-SQD and Pep-MHP which is $X^2X^3XXXGXXS$, refer to *figure 3.4*. There is not a clear consensus that occurs in all three peptides, but there are clearly some consensus' that are shared between two of the peptides at a time.

Furthermore, using PepCalc it was determined that Pep-SQD, Pep-MHP, and Pep-SGV all have a positive charge at pH 6.0 which was the pH conditions for the panning experiments (*table 3.7*).[288]

Table 3.6: Sequencing of peptides using one letter code identified as binding to c-MycC27. Identified from phage display method 3, round 3, selecting against other DNA structures.

Sample	Peptide Sequence
1	Insertless
2	SQDIRTWNGTRS
3	MHPNAGHGSLMR
4	Insertless
5	Insertless
6	SGVYKVAYDWQH
7	MHPNAGHGSLMR
8	Insertless
9	MHPNAGHGSLMR
10	SQDIRTWNGTRS
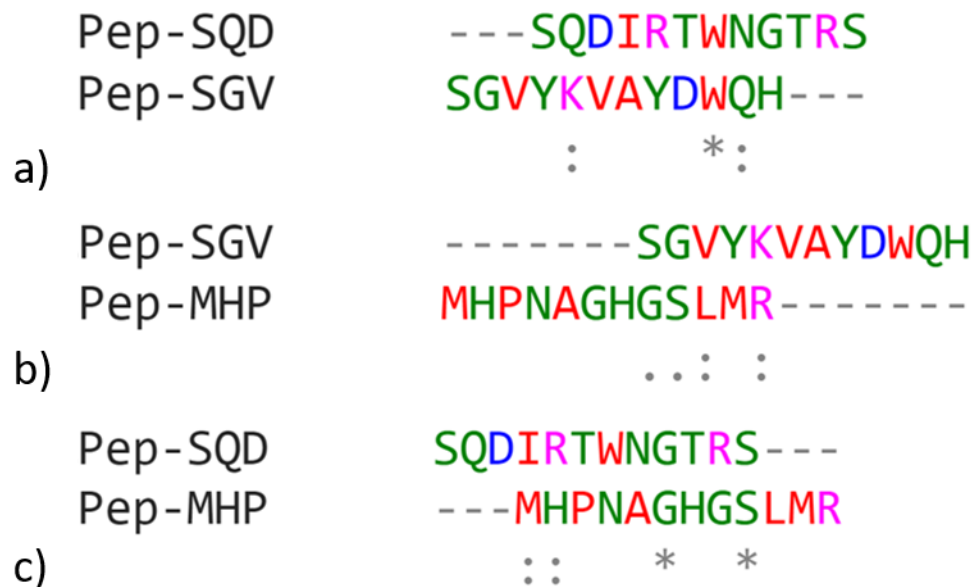
Figure 3.4: Multiple sequence alignment using Clustal Omega for a) Pep-SQD and Pep-SGV b) Pep-SQD and Pep-MHP c) Pep-SQD and Pep-MHP. Where the symbols "*" indicates perfect alignment, ":" indicates a site belonging to a group exhibiting strong similarity, and "." indicates a site belonging to a group exhibiting weak similarity. Where the colours are used to group properties: red-small, blue-acidic, pink-basic, green-hydroxyl/sulfhydryl/amine/G grey-unusual amino/imino acids.

Table 3.7: Peptides, Pep-SQD, Pep-MHP, and Pep-SGV, charge at pH 6.0 as determined using PepCalc.

Peptide	Charge at pH 6.0
Pep-SQD	+1
Pep-MHP	+1.2
Pep-SGV	+0.3

3.2.1.4 Phage Display Screening Methods 4-5 and Results Targeting the 27 bp i-Motif Forming Sequence in the Promoter Region of *MYC*

Although method 3 showed some peptides that bind c-MycC27, due to the large percentage of insertless phage being sequenced it highlighted either there are not strong binders within the libraries, or the method needed refinement. This was the same problem faced in methods 1 and 2, therefore using a 12-mer library did not eliminate this problem. Therefore, methods 4 and 5 used streptavidin coated plates rather than beads to see if this reduced non-specific binding, refer to *table 5.4* for the full method and the summarised method is shown in *table 3.8*. Method 4 used the Ph.D.-C7C library and method 5 used the Ph.D.-12 library, these were screened at the same time to increase productivity. Round 1 was sent for sequencing to analyse if the changed variables removed insertless sequences. *Tables 3.9* and *3.10* demonstrate that insertless sequences were still being identified and so these methods were not progressed past selection 1.

Table 3.8: Phage display methods 4 and 5 for panning against the target DNA sequence c-MycC27.

Phage Display Method	Round	Plate or Beads	Phage Library	Phage	Target DNA	Blocking Buffer	Wash Buffer	Competitors added to Phage Mixture
Method 4	1st	Plate	Ph.D.-C7C	1x10^11 Phage Library, Ph.D.-C7C, in 500 μL PBS pH 6.0	10 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None
Method 5	1st	Plate	Ph.D.-12	1x10^11 Phage Library, Ph.D.-12, in 500 μL PBS pH 6.0	10 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None

Table 3.9: Sequencing of peptides using one letter code identified as binding to c-MycC27. Identified from phage display method 4, round 1. N/D shows the sequence was not determined. A – means that amino acid was not determined.

Sample	Peptide Sequence
1	KPYTVSN
2	HVARLD–
3	TPHAVTQ
4	Insertless
5	ETKTSKT
6	Insertless
7	Insertless
8	TSYTGPH
9	QPEYQGH
10	N/D

Table 3.10: Sequencing of peptides using one letter code identified as binding to c-MycC27. Identified from phage display method 5, round 1.

Sample	Peptide Sequence
1	HLAKELSFMARP
2	Insertless
3	NLPHQLNYDYRT
4	VISTTLPPQTPA
5	Insertless
6	THWQGMSSLRFS
7	VTSRAPDPNSAR
8	GTFIAIATDTWG
9	NVSLNVLHKTVD
10	TYPPPLQSRMHA

3.2.1.5 Phage Display Screening Method 6 and Results Targeting the 27 bp i-Motif Forming Sequence in the Promoter Region of *MYC*

Methods 1-5 had an abundance of insertless phage. The optimisation carried out for methods 1-5 included 12-mer linear peptides vs 7-mer cyclic libraries and using streptavidin coated beads compared to streptavidin coated plates. Suggesting that non-specific binding was occurring, therefore the Tween 20 percentage was increased to reduce non-specific binding. Furthermore, reviewing the method using the beads it was decided to carry out further optimisation. When pulling the magnetic beads out of solution for methods 1-3 they were pulled to the bottom of the tube and the supernatant is removed (*figure 3.5*). This is similar to how the streptavidin plate works in that the supernatant during washes is removed from the bottom of the well, but in the case of a plate the streptavidin is attached to the plate. However, with beads when removing washes supernatant it is likely that some beads will be sucked into the pipette and lost. Meaning some binders end up in the washes, or that some of the wash is left in the tube when trying to avoid sucking up beads. Leftover wash in the tube could be the source of insertless phage. Therefore, the design of pulling the magnetic beads was changed so that they are pulled to the side

of the tube (*figure 3.5*). This allows for all of the supernatant during washes to be removed. These two adjustments were used for method 6 using the Ph.D.-12 library, refer to *table 5.4* for full details and *table 3.11* for the summarised version. Only the Ph.D.-12 library was explored as the Ph.D.-C7C stock was empty.
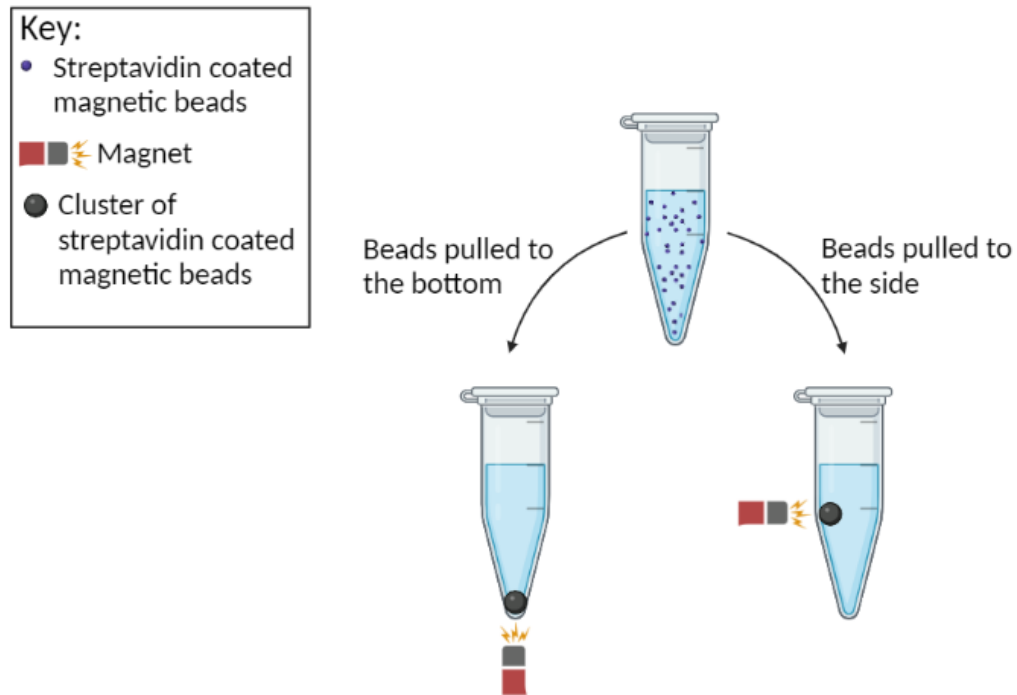
Figure 3.5: Schematic demonstrating two different methods, pulled to the bottom, and pulled to the side, on how streptavidin coated magnetic beads can be pulled out of solution in a tube using a magnet. Created with BioRender.com.

Table 3.11: Phage display method 6 for panning against the target DNA sequence c-MycC27.

Round	Plate or Beads	Phage Library	Phage	Target DNA	Blocking Buffer	Wash Buffer	Competitors added to Phage Mixture
1st	Plate	Ph.D.-12	1x10^11 Phage Library, Ph.D.-12, in 500 µL PBS pH 6.0	10 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None
2nd	Plate	Ph.D.-12	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-Hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU
3rd	Plate	Ph.D.-12	1x10^11 Amplified Phage Selection 2 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP Hif1α

The sequencing results for method 6 round 3, shown in *table 3.12*, identified 40% insertless sequences and six different sequences. Sequences were analysed using multiple sequence alignment Clustal Omega identified that Pep-TWP and Pep-TMA show a similarity $X^1XXXX^1X^2XXXXXX^2$.[286,287] There is also a similarity between Pep-SAN and Pep-SLD which is $X^1X^2XXX^2WT$. Where X is any amino acid, X^1 is an amino with a property of basic/amine, and X^2 is an amino acid that is small. Although, there is no clear consensus between more than two peptides PepCalc shows that at pH 6.0/6.6 the following peptides are all positively charged: Pep-TWP, Pep-SAN, Pep-MFK, and Pep-DGS. Whereas, Pep-SLD and Pep-CPG have no net charge and Pep-TMA has a negative charge.[288] Overall the most of the seven peptides identified as binding to c-MycC27 using method 6 are positively charged at the pH the phage display was performed at (pH 6.0).

Table 3.12: Sequencing of peptides using one letter code identified as binding to c-MycC27. Identified from phage display method 6, round 3.

Sample	Peptide Sequence
1	SLDWNWTTSAVK
2	TWPNNLKPRLMF
3	CPGTCSRSDIPP
4	SANYNVQAGWTH
5	TMAYNVNTDSNF
6	MFKNNQVGMAVV
7	DGSMLNRMRGFS
8	Insertless
9	Insertless
10	Insertless
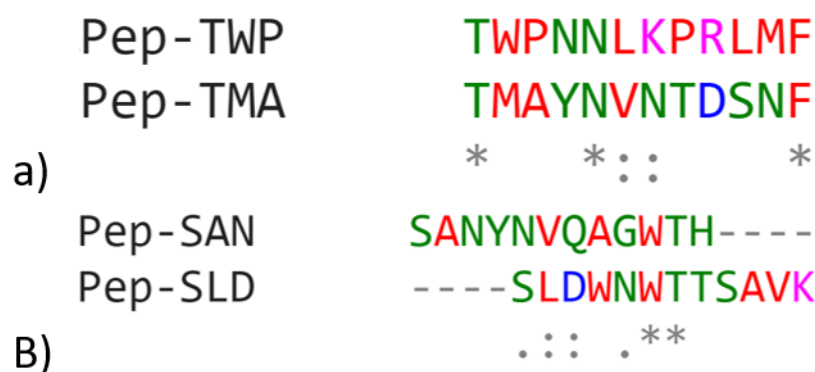
Figure 3.6: Multiple sequence alignment using Clustal Omega for a) Pep-TWP and Pep-TMA b) Pep-SAN and Pep-SLD. Where the symbols "*" indicates perfect alignment, ":" indicates a site belonging to a group exhibiting strong similarity, and "." indicates a site belonging to a group exhibiting weak similarity. Where the colours are used to group properties: red-small, blue-acidic, pink-basic, green-hydroxyl/sulfhydryl/amine/G grey-unusual amino/imino acids.

Overall phage display was unsuccessful in the search for c-MycC27 peptide probes. Some peptides were identified from sequencing following multiple phage display methods, however these were identified alongside insertless sequences. Furthermore, all sequences identified from methods 1-6, 37 different sequences, were analysed for similarities by multiple sequence

alignment (*figure 3.7*). The multiple sequence alignment did not identify any clear alignment. Providing evidence that the libraries Ph.D.-C7C and Ph.D.-12 did not contain any clones that bind tight enough to the target to be selected.[197] Therefore, none of the peptides were taken forward for further analysis as the phage display was unsuccessful in identifying peptides that bound c-MycC27.

```
Pep-QPD    ---------------QPDLPGV---------    7
Pep-QPE    ---------------QPEYQGH---------    7
Pep-NLP    ------------------NLPHQLNYDYRT--   12
Pep-ETK    ---------ETKTSKT--------------    7
Pep-TSY    ------------TSYTGPH-----------    7
Pep-HVA    ----------HVARLD--------------    6
Pep-SGV    --------SGVYKVAYDWQH----------   12
Pep-SAN    --------SANYNVQAGWTH----------   12
Pep-DTS    ---------DTSTKYL--------------    7
Pep-DGH    ---------DGHDQSL--------------    7
Pep-VQM    -------------VQMPAHS----------    7
Pep-VIS    ----------VISTTLPPQTPA--------   12
Pep-LAQ    --------------LAQSHPL---------    7
Pep-NVS    ------------------NVSLNVLHKTVD   12
Pep-DLL    ------------------DLLHRGA-------    7
Pep-TYP    ---------------TYPPPLQSRMHA------   12
Pep-HLA    ------------------HLAKELSFMARP--   12
Pep-TNL    ------------------TNLSPRA--------    7
Pep-TWP    ---------------TWPNNLKPRLMF------   12
Pep-TMA    --------TMAYNVNTDSNF----------   12
Pep-KPY    ---------KPYTV-SN-------------    7
Pep-CPG    ---------CPGTC-SRSDIPP--------   12
Pep-SQD    -SQDIR--TWNGTRS---------------   12
Pep-TPH    -----------TPHAVTQ-----------    7
Pep-SLD    ----SL--DWNWTTSAVK------------   12
Pep-EFS    -----------EFSKFRS------------    7
Pep-TQR    --------------TQRSHSS----------    7
Pep-THW    -----T--HWQGMSSLRFS-----------   12
Pep-WKH    ----WKHGNFN-------------------    7
Pep-HTE    -HTEALHA---------------------    7
Pep-MHP    MHPNAGHGSLMR-----------------   12
Pep-DGS    ------DGSMLNRMRGFS-----------   12
Pep-MFK    --------MFKNNQVGMAVV----------   12
Pep-GTF    ----------GTFIAIATDTWG--------   12
Pep-FPN    ------------------FPNSM---YQ---    7
Pep-MAP    -----------------MAPDSRV---------    7
Pep-VTS    -------------VTSRAPDPNSA---R----   12
```

Figure 3.7: Multiple sequence alignment using Clustal Omega for peptides identified as binding to c-MycC27. Where the symbols "*" indicates perfect alignment, ":" indicates a site belonging to a group exhibiting strong similarity, and "." indicates a site belonging to a group exhibiting weak similarity. Where the colours are used to group properties: red-small, blue-acidic, pink-basic, green-hydroxyl/sulfhydryl/amine/G grey-unusual amino/imino acids.

3.2.2 Phage Display Screening Method Development and Results Targeting the 52bp i-Motif Forming Sequence in the Promoter Region of

The method development of the phage display protocol used for c-MycC27 in *Section 3.2.1* was used to inform the method development for c-MycC52. As c-MycC52 is longer there may be a greater chance of finding peptides that bind with a stronger affinity and so no identification of insertless phage. As carried out in c-MycC27 method 6 the beads were pulled to the side of the tube for c-MycC52 method 1. The pH for the buffers used was pH 6.0, because the transitional pH for c-MycC52 was determined at pH 6.4, refer to *figure 3.2*, and autoclaving causing adjustments to pH required greater acidity to ensure at least 50% folded i-motif, as discussed in Section 3.2.1.[283] Thus, the buffers were kept identical to those used in method 6, refer to *table 3.11* and *table 3.13*. Using the same pH values as for c-MycC27 also makes comparison between peptides identified as c-MycC27 and c-MycC52 more easily comparable.

Table 3.13: Phage display methods 1 and 2 for panning against the target DNA sequence c-MycC52.

Phage Display Method	Round	Plate or Beads	Phage Library	Phage	Blocking Buffer	Wash Buffer	Competitors added to Phage Mixture
Method 1							
	1st	Plate	Ph.D-12	1×10^{11} Phage Library, Ph.D.-12, in 500 µL PBS pH 6.0	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None
	2nd	Plate	Ph.D-12	1×10^{11} Amplified Phage Selection 1 Elution 3 in 500 µL PBS pH 6.0	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU
	3rd	Plate	Ph.D-12	1×10^{11} Amplified Phage Selection 2 Elution 3 in 500 µL PBS pH 6.0	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP Hif1α
Method 2							
	1st	Plate	Ph.D-C7C	1×10^{11} Phage Library, Ph.D.-C7C, in 500 µL PBS pH 6.0	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None
	2nd	Plate	Ph.D-C7C	1×10^{11} Amplified Phage Selection 1 Elution 3 in 500 µL PBS pH 6.0	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU
	3rd	Plate	Ph.D-C7C	1×10^{11} Amplified Phage Selection 2 Elution 3 in 500 µL PBS pH 6.0	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP Hif1α

The sequencing results for method 1 are shown in *table 3.14*, originally 10 sequences were sequenced and as 9/10 were Pep-EIE a further 10 sequences were analysed. From 20 sequences it was identified that 16/20 were Pep-EIE and 1/20 for the following peptides: Pep-RVS, Pep-SLC,

Pep-VSE, and Pep-PTN. Clustal Omega was used to determine any consensus between the sequences multiple sequence alignment. When all five peptides were put through at once there was no consensus determined by Clustal Omega software (*figure 3.7*).[286,287] However, if looking at the sequences four of the five sequences show some similarities. Pep-PTN, Pep-SLC, Pep-VSE, Pep-EIE, and Pep-RVS have a similarity of D/EXX^{2}X^3 (*figure 3.7*). Where X is any amino acid, X^2 is small amino acids, and X^3 are basic. Clustal Omega is designed to align sequences to produce alignments of divergent samples and usually for longer sequences, whereas for these results similarities need to be identified.[289] Therefore, the software didn't find the similarities identified above as meaningful but by looking at the alignment you can determine them. Notably, there is a five amino acid consensus between Pep-EIE and Pep-RVS, TDHMK, interestingly the DHMK is part of the similarity sequence D/EXX^{2}X^3 determined for four of the peptides (*figure 3.8*). Using PepCalc it was determined that at pH 6.0 Pep-RVS, Pep-SLC, and Pep-PTN are all positively charged, Pep-EIE has a negative charge, and Pep-VSE has no charge (*table 3.15*).[288] This determined that the peptides have different overall charges. Overall, this method was successful at achieving the aim to identify peptides that bind c-MycC52.

Table 3.14: Sequencing of peptides using one letter code identified as binding to c-MycC52.
Identified from phage display method 1, round 3.

Sample	Peptide Sequence
1-7, 9-15, 18-19	EIEYTDHMKELG
8	RVSTDHMKGRGG
16	SLCDIIRIEKVR
17	VSEAWKEVKGFF
20	PTNVSGRNYLFC

```
Pep-PTN          -------PTNVSGRNYLFC
Pep-SLC          SLCDIIRIEKVR-------
Pep-VSE          ------VSEAWKEVKGFF-
Pep-EIE          ---EIEYTDHMKELG----
Pep-RVS          ----RVSTDHMKGRGG---
```

Figure 3.8: Multiple sequence alignment using Clustal Omega for Pep-PTN, Pep-SLC, Pep-VSE, Pep-EIE, and Pep-RVS. Where the symbols "" indicates perfect alignment, ":" indicates a site belonging to a group exhibiting strong similarity, and "." indicates a site belonging to a group exhibiting weak similarity. Where the colours are used to group properties: red-small, blue-acidic, pink-basic, green-hydroxyl/sulfhydryl/amine/G grey-unusual amino/imino acids.*

Table 3.15: Peptides, Pep-PTN, Pep-SLC, Pep-VSE, Pep-RVS, and Pep-EIE, charge at pH 6.0 as determined using PepCalc.

Peptide	Charge at pH 6.0
Pep-PTN	+1
Pep-SLC	+0.9
Pep-VSE	0
Pep-RVS	+2.3
Pep-EIE	-2.7

Sequencing results for method 2, refer to *table 3.*13, are shown in *table 3.16*, and 5/10 were identified as insertless and there were no clear consensus or similarities identified using multiple sequence alignment Clustal Omega (*figure 3.9*).[286,287] Using the PepCalc tool it was identified that at pH 6.0 Pep-SLK and Pep-TTV are positively charged, Pep-HLI and Pep-IFP are neutral, and Pep-QLV is negatively charged (*table 3.17*).[288] Showing the peptides had differing overall charges at pH 6.0.

Table 3.16: Sequencing of peptides using one letter code identified as binding to c-MycC52 Identified from phage display method 2, round 3.

Sample	Peptide Sequence
1	Insertless
2	Insertless
3	SLKYDRF
4	Insertless
5	HLIDSSH
6	Insertless
7	Insertless
8	IFPGMQP
9	QLVSLHD
10	TTVDKRS

```
Pep-TTV        TTVD----KRS
Pep-HLI        HLIDSSH----
Pep-SLK        ----SLKYDRF
Pep-QLV        -QLVSLHD---
Pep-IFP        -IFPGMQP---
```

Figure 3.9: Multiple sequence alignment using Clustal Omega for Pep-TTV, Pep-HLI, Pep-SLK, Pep-QLV, and Pep-IFP. Where the symbols "*" indicates perfect alignment, ":" indicates a site belonging to a group exhibiting strong similarity, and "." indicates a site belonging to a group exhibiting weak similarity. Where the colours are used to group properties: red-small, blue-acidic, pink-basic, green-hydroxyl/sulfhydryl/amine/G grey-unusual amino/imino acids.

Table 3.17: Peptides, Pep-TTV, Pep-HLI, Pep-SLK, Pep-QLV, and Pep-IFP, charge at pH 6.0 as determined using PepCalc.

Peptide	Charge at pH 6.0
Pep-SLK	+1
Pep-TTV	+1
Pep-HLI	0
Pep-IFP	0
Pep-QLV	-0.3

Overall methods 1 using the Ph.D.-12 library was successful in identifying peptides that bind to c-MycC52, whereas method 2 using the Ph.D.-C7C library identified insertless phage similarly to the results seen for c-MycC27 in *Section 3.2.1*. In comparison the library Ph.D.-12 had no insertless sequences identified from double the amount of sequencing compared to Ph.D.-C7C. Therefore, it appears that the Ph.D.-12 library has stronger binders. Furthermore, this suggests that a 7-mer sequence may not be long enough or due to structural constraints the confirmation required for target binding was unachievable and thus there was not enough confirmational diversity in the Ph.D.-C7C library.[197] Nevertheless, all the sequences identified as c-MycC52 binders, 10 different sequences, were aligned in *figure 3.10* to determine similarities. Analysis did not determine a clear similarity between all the sequences or between the Ph.D.-12 or Ph.D.-C7C libraries. Based on method 2 being determined as unsuccessful in determining c-MycC52 peptide binders, due to insertless sequences and a lack of consensus, the five peptides successfully identified as c-MycC52 binders in method 1 were taken forward for further testing (*figure 3.11*).

```
Pep-IFP        ------------IFPGMQP        7
Pep-EIE        EIEYTDHMKELG--------        12
Pep-RVS        -RVSTDHMKGRGG-------        12
Pep-PTN        ----PTNVSGRNYLFC----        12
Pep-TTV        -----TTVDKRS--------        7
Pep-HLI        -----HLIDSSH--------        7
Pep-VSE        ---VSEAWKEVKGFF-----        12
Pep-SLC        --SLCDIIRIEKVR------        12
Pep-QLV        -----QLVSLHD--------        7
Pep-SLK        -------SLKYDRF-----        7
```

Figure 3.10: Multiple sequence alignment using Clustal Omega for Omega for peptides identified as binding to c-MycC52. Where the symbols "*" indicates perfect alignment, ":" indicates a site belonging to a group exhibiting strong similarity, and "." indicates a site belonging to a group exhibiting weak similarity. Where the colours are used to group properties: red-small, blue-acidic, pink-basic, green-hydroxyl/sulfhydryl/amine/G grey-unusual amino/imino acids.
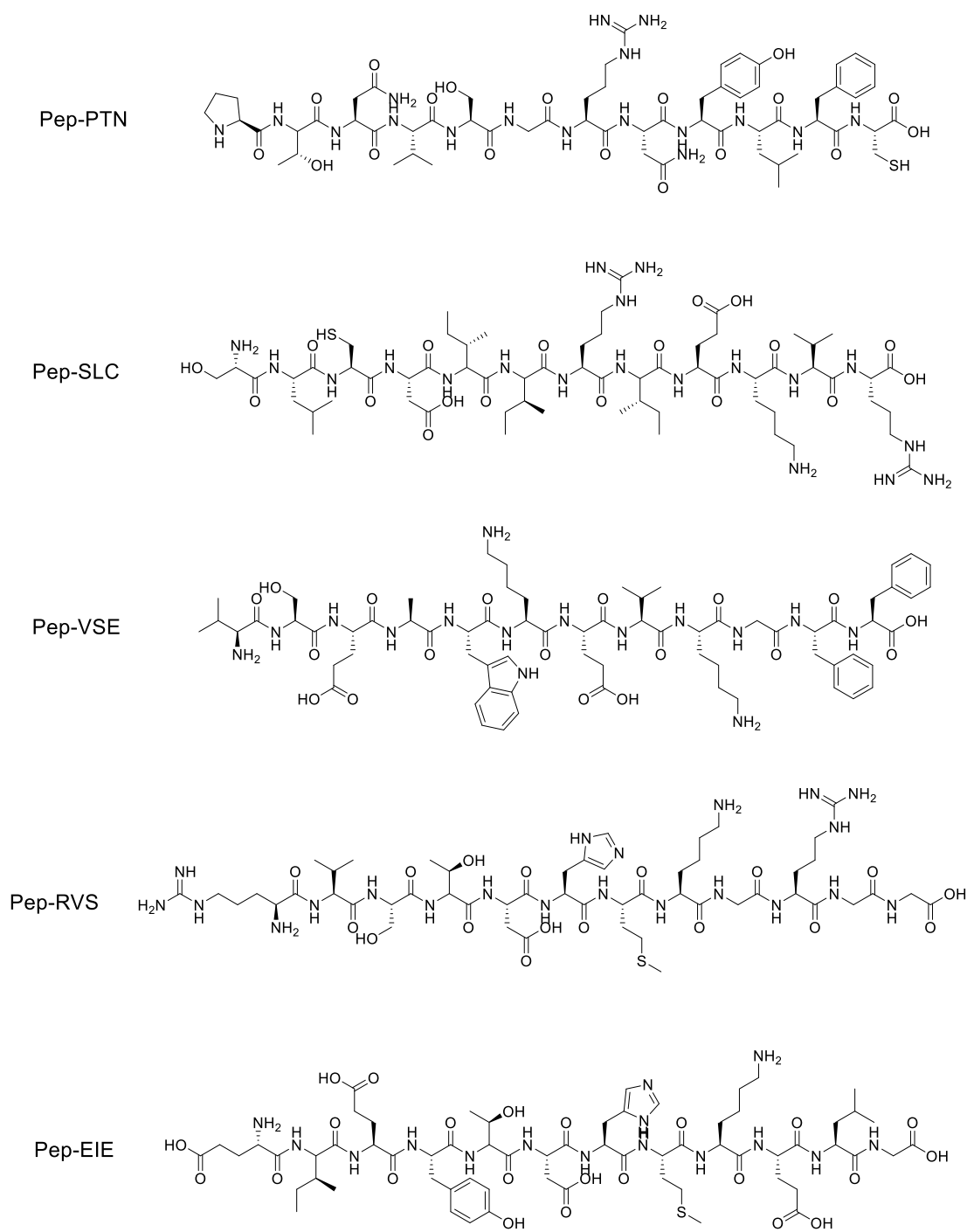
Figure 3.11: Structures of five different 12-mer peptides of interest: Pep-PTN, Pep-SLC, Pep-VSE, Pep-RVS, and Pep-EIE.

3.3 Biophysical Investigation of Binding

To determine further binding data multiple biophysical techniques were chosen to assess the peptides, which are commonly used in the non-canonical DNA structures field.

3.3.1 Fluorescent Indicator Displacement Assay

Originally, the Waller group developed the fluorescent indicator displacement assay (FID) to determine if small molecules bound to i-motif DNA.[85] This method uses an indicator molecule called **thiazole orange** (**TO**) which does not have fluorescent properties in the absence of DNA and upon binding to DNA it fluoresces, see *Section 2.2.1*. This is a useful property as if a competitor compound is added into solution with the DNA-**TO** complex and the competitor displaces the **TO** then the **TO** is free in solution and not fluorescent. Thus, if a compound binds to the DNA in a way that displaces **TO** the fluorescence's decreases which can be measured to determine if a compound binds to the DNA (*figure 2.2).*

This method was designed to assess the binding of small compounds and so there was speculation to if this could be used to assess peptides binding to the i-motif. The FID method optimised for c-MycC52 in *section 2.2.1* was used, 1 μM:2 μM DNA:TO in 10 mM NaCaco pH 6.6. The peptides were originally titrated as 0.5 μM increments from 0 μM to 5 μM and 1 μM increments from 0 μM to 10 μM (*appendix A4*). However, none of the peptides were identified as a hit (achieving over 15% TO displacement). These ranges are the ranges the Waller lab would typically look at small molecules, and this range is not necessarily applicable to peptides. If the peptides have a low binding affinity for c-MycC52 then this concentration range may not be high enough to see binding using FID. Thus, the concentration range titrating every 25 μM from 0 μM to 200 μM was used

(*figure 3.12*). Which demonstrated that all the peptides except Pep-RVS are hits, achieved over 15% TO displacement, within the range of 0 µM to 200 µM. This supports that the aim of the study to identify i-motif binding peptides was accomplished. Pep-RVS had 14 % (+/-1) displacement at 200 µM and so a higher concentration may have been required to show binding as an increase in percentage displacement is viewed which does appear concentration dependent, it didn't reach the usual criteria to be considered binding to c-MycC52, although this criterion was set for small molecules not peptides. Notably, Pep-SLC shows the greatest percentage of displacement and the strongest binding by the shape of the curve.
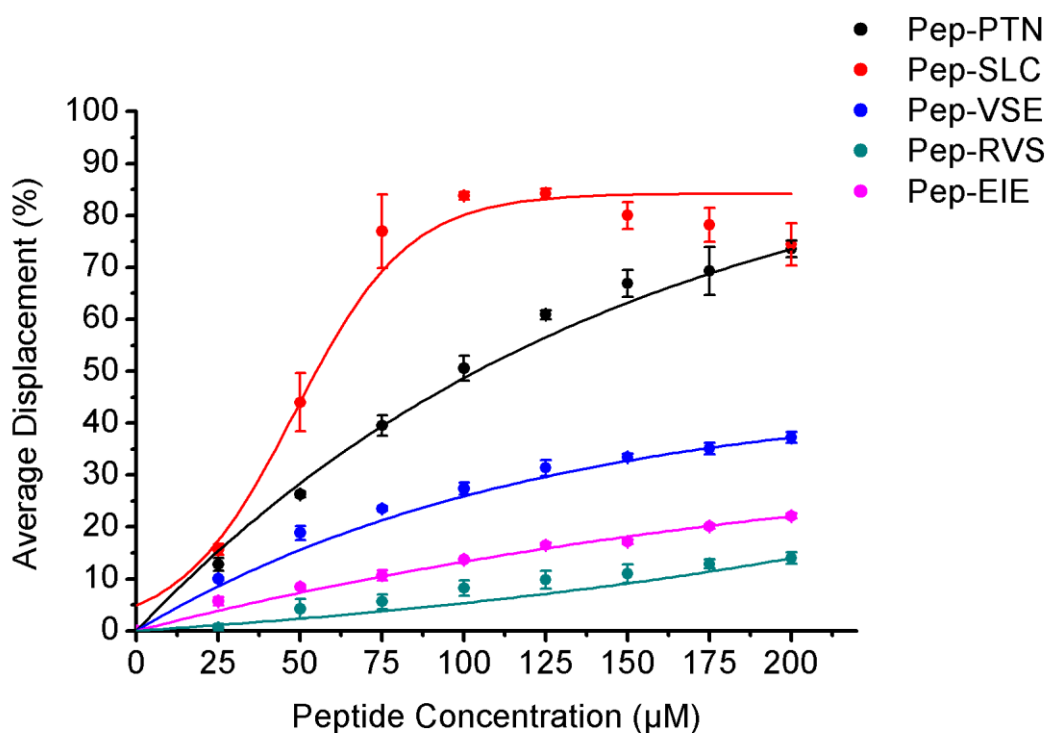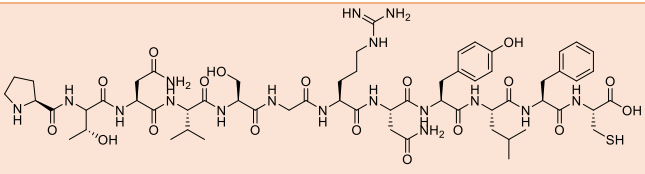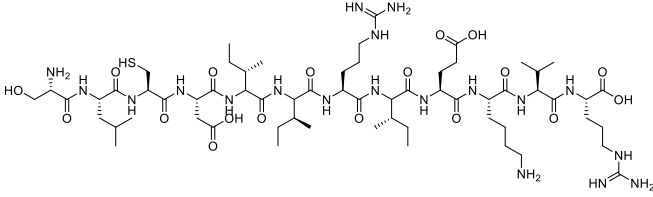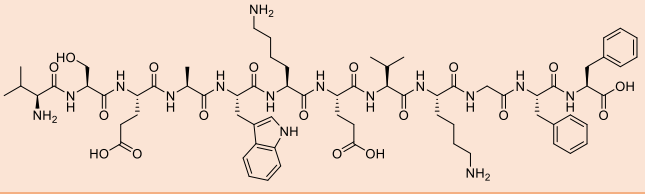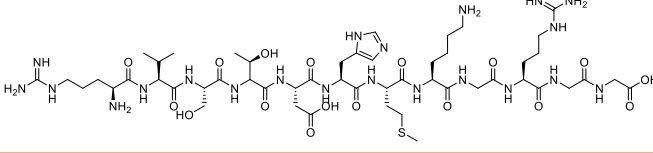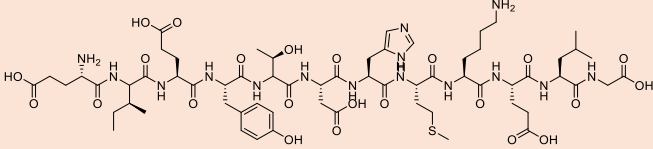
Figure 3.12: Fluorescent indicator displacement assay in the following conditions: 1 µM:2 µM c-MycC52:To in 10 mM NaCaco pH 6.6 (where TO is thiazole orange). Five different peptides of interest were titrated every 25 µM from 0 µM to 200 µM and fitted with a dose response (Pep-SLC and Pep-PTN) or exponential curve (Pep-VSE, Pep-RVS, and Pep-EIE) using OriginPro 8 software.

For each peptide the DC_{50} value, the concentration at which 50% of **TO** is displaced, was determined in *table 3.18*. Pep-VSE, Pep-RVS, and Pep-EIE did not reach 50% **TO** displacement and so the DC_{50} value is over 200 μM. The DC_{50} values for Pep-PTN and Pep-SLC were 100.5 μM and 51.4 μM respectively. Pep-SLC has a lower DC_{50} value and so a lower concentration of Pep-SLC displaces 50% of **TO** compared to all the other peptides, this suggests Pep-SLC may have greater binding affinity to c-MycC52 and could be the most promising peptide to investigate. Further biophysical data is required to provide information on the binding affinities and specificity to determine which peptide(s) are most promising.

Table 3.18: DC$_{50}$ values, the concentration at which a compound displaces 50% of thiazole orange (TO) bound to the DNA sequence, were calculated using FID titration assay from 0 µM to 200 µM in the following conditions: 1 µM:2 µM c-MycC52:TO in 10 mM NaCaco pH 6.6 (where TO is thiazole orange). >200 µM demonstrates that the DC$_{50}$ value was not reached and so a higher concentration is required, N/D stands for not determined. The DC$_{50}$ value was determined from dose response (Pep-SLC and Pep-PTN) or exponential curve (Pep-VSE, Pep-RVS, and Pep-EIE) line fittings of the titration data using OriginPro 8 software.

Peptide	Structure	DC$_{50}$ (µM)	Fitting Error
Pep-PTN		100.5	N/D
Pep-SLC		51.4	N/D
Pep-VSE		>200	N/D
Pep-RVS		>200	N/D
Pep-EIE		>200	N/D

3.3.2 Circular Dichroism DNA Melting Studies

Circular dichroism (CD) was used to investigate the effects of peptide binding on the stability of c-MycC52. CD provides an i-motif "fingerprint" one can monitor structural changes under changing conditions using CD melting, *figure 2.15*. CD melting experiments were used to explore if the

peptides caused stabilisation or destabilisation of c-MycC52 (*figure 3.13*). Melting experiments used relevant water or **DMSO** controls to account for solvent dilution effects, as described in *figure 3.13*.

The fittings for the graphs fit better with a bi-dose sigmoidal curve fit rather than dose response for all graphs except the water control and Pep-EIE. This is due to there being different i-motif populations within the sample and those different i-motif formations have differing stabilities causing multiple melting points, and thus multiple transitions. Some graphs, such as that for Pep-SLC and Pep-VSE had a clear second transition after 60 °C (*figure 3.13*). Whereas, other graphs, such as the water or **DMSO** control do not show a clear second transition. Interestingly, although neither control showed a clear second transition the bi-dose response fitting, which is used for data with two transitions, fitted the **DMSO** control data much better than the dose-response. This has been seen in the literature before for c-MycC, most recently with the Smith *et al.* paper investigating ellipticine derivatives effects on i-motif sequences including c-MycC.[97] In this case all c-MycC normalised ellipticity CD melts were fitted with bi-dose response as this fitted the data better than a dose response fitting, even if a clear second transition could not be seen. The data in *figure 3.13* in general found the same, for example the **DMSO** control and Pep-RVS do not clearly show second transition but the data was found to have a better fitting using a bi-dose response fit rather than a dose-response fit. The bi-dose response fitting did fit the water control better before 30 °C and from 60 °C onwards. But it had poor fitting around the middle temperatures ran, especially at ~50 °C and did not show a smooth sigmoidal curve (*appendix A5*). Therefore, it was decided that the dose response fitting was more appropriate due to the better sigmoidal shape which more appropriately represented the data shown in this instance.
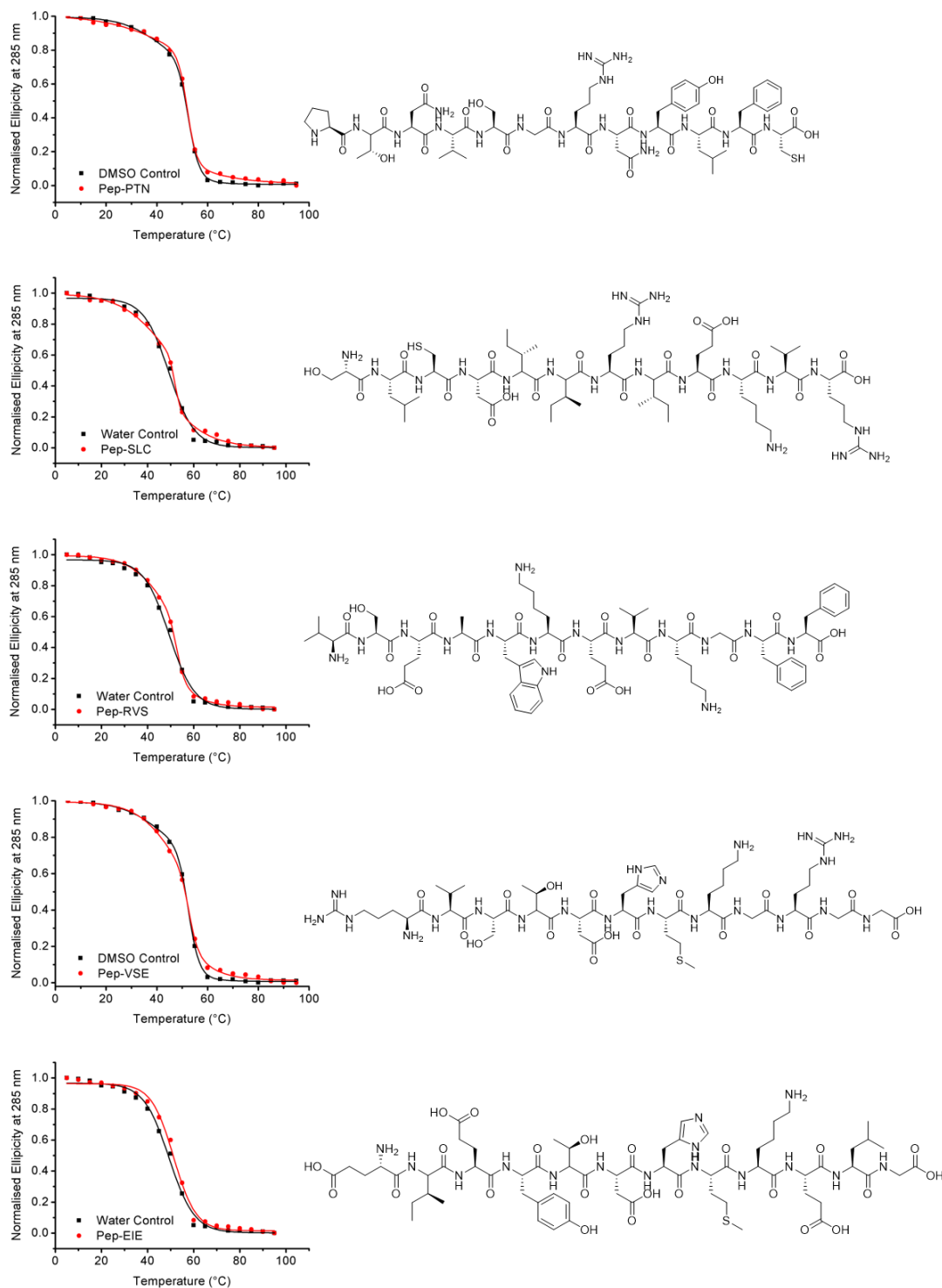
Figure 3.13: Circular dichroism of normalised ellipiticity at 285 nm, the maximum positive characteristic peak for c-MycC52, of 10 µM c-MycC52 in 10 mM NaCaco pH 6.6 in the absence and presence of peptides at 10 equivalence. The control (black) is addition of 5 µL of water for Pep-SLC, Pep-RVS, and Pep-EIE and 5 µL of 100% DMSO for Pep-PTN and Pep-VSE as the appropriate vehicle controls. The data analysis was carried out in OriginPro 8 and all graphs were fitted with a bi-dose response fitting except the water control which was fitted with dose response fitting.

The fittings shown in *figure 3.13* were used to determine the melting temperature(s) (T_m) of the i-motif structure(s) in the presence of 100 μM of peptide compared to the appropriate control. The change in T_m is recorded in *table 3.19*. Looking at *figure 3.13* it was determined that c-MycC52 has one transition in the presence of either control, Pep-RVS, and Pep-EIE and thus only one melting temperature was determined. Both Pep-RVS and Pep-EIE were determined to stabilise c-MycC52 (*table 3.19*).

Interestingly, *figure 3.13* demonstrated that in the presence of Pep-PTN, Pep-SLC, and Pep-VSE there is a second transition between 60 °C to 80 °C. These three peptides had no effect on the stability of the first transition (all <0.5 °C) and stabilised the i-motif population that causes the second transition seen in c-MycC. This is a real transition as there are published examples of this in the literature, for example, Smith *et al.* demonstrated in the presence of ellipticine derivatives c-MycC has a second transition and saw a direct correlation between concentration of the ligand and stabilisation of this second transition,[97] clearly demonstrating it is a true transition. When analysing the transitions of c-MycC52 in the presence of Pep-PTN, Pep-SLC, and Pep-VSE the OriginPro 8 software using the bi-dose response fitting determined two transitions however they were both under 60 °C. The software was taking the transitions from the top of the sigmoidal curve ~40 °C and at the first transition ~50 °C, whereas one can clearly see from *figure 3.13* there is a second transition after 60 °C. The reasoning for the software not determining the second transition is due to the fitting of the data. Therefore, to overcome this I split the data for each peptide into two sets, transition one using the temperature range 5 °C to 70 °C and transition two using the temperature range 60 °C to 95 °C for Pep-SLC and 55 °C to 95 °C for Pep-PTN and Pep-VSE. All these data sets were fitted using dose response in OriginPro 8 as there is only one transition in each of the data sets (*appendix A6*). However, this approach only worked for Pep-SLC, this is because the first and second transition

data sets for Pep-SLC formed sigmoidal shaped curves (*appendix A6*). Whereas, for Pep-PTN and Pep-VSE this method worked for determining the first transition as there was sigmoidal data however the second transition did not give a clear sigmoidal shaped curve (*appendix A6*) and so the second transition could not be determined. Nonetheless, *figure 3.13* demonstrated that Pep-PTN and Pep-VSE had second transitions but a larger concentration of peptide may be required to be able to quantify the second transition.

Overall Pep-RVS and Pep-EIE were the only peptides that affected the stability of the first transition i-motif population, where Pep-RVS demonstrated the greatest stabilisation. However, neither of these two peptides had any effect on the second transition i-motif population which were stabilised by Pep-PTN, Pep-SLC, and Pep-VSE, which was only quantifiable for Pep-SLC which had a large stabilisation effect (*table 3.19*). This supports that all the peptides bind to c-MycC52 as to cause a change in structural stability they are binding to c-MycC52, supporting that this work has achieved its aim to identify c-MycC binding peptides.

Table 3.19: Melting temperature(s) (T_m) and change in T_m for c-MycC52 with peptides presence at 10 equivalence compared to no peptides present. Change in T_m is determined by the difference in the T_m with no compounds present compared to that with 100 μM of peptide present against 10 μM DNA determined from circular dichroism melts from 5 ^{o}C to 95 ^{o}C in buffer conditions 10 mM NaCaco pH 6.6. The data analysis was carried out in OriginPro 8 and the T_m was determined from dose or bi-dose response fitting of ellipticity (mdeg) VS temperature (^{o}C). DMSO is the control for Pep-PTN and Pep-VSE, water is the control for Pep-SLC, Pep-RVS, and Pep-EIE. N/A means not applicable as there was no 2nd transition to provide a second ΔT_m and N/D means not determined.

Sample	T_m (°C)	2 nd T_m (°C)	ΔT_m (°C)	2nd ΔT_m (°C)	Fitting Error
DMSO	52.4	N/A	N/A	N/A	0.0003
Water	49.3	N/A	N/A	N/A	0.0012
Pep-PTN	52.0	N/D	-0.4	N/D	0.0007
Pep-SLC	49.4	73.0	+0.1	+23.7	0.0018
Pep-VSE	52.6	N/D	+0.2	N/D	0.0096
Pep-RVS	52.6	N/A	+3.3	N/A	0.0096
Pep-EIE	51.3	N/A	+2.0	N/A	0.0044

3.3.3 Surface Plasmon Resonance Binding Studies

Surface plasmon resonance (SPR) is a powerful spectroscopy technique that allows real time monitoring of noncovalent molecular interactions to determine binding affinity and kinetic data between targets and ligands, *figure 1.32*. As the mass of the peptides are higher than that of the small compounds that were screened using SPR in *Section 2.3.2* there should be greater molar responses and thus better resolution.

The SPR studies aimed to corroborate with other biophysical data that the peptides bind to c-MycC52, whilst also determining affinity constants from kinetic data. Another objective was to assess the specificity of these peptides by providing kinetic data which was unable to be determined from other biophysical techniques such as CD and test a greater number of different oligonucleotide sequences. The five compounds were analysed at 10 different concentrations in duplicate from 0 μM to 100 μM using a 2-fold

series dilution. Buffering conditions were kept identical to those ran for the small molecules in *Section 2.3.2*, 10 mM NaCaco 10 mM KCl 0.05% Tween 20 at pH 6.6, peptides dissolved in DMSO also had 2% DMSO in the buffer to avoid mismatching between the samples and running buffer. Under these buffer conditions the i-motif and G-quadruplex structures are folded.

A schematic model sensorgram produced from SPR is shown in *figure 3.14*, a sensorgram has five phases. The initial phase is the baseline, the running buffer is flowed over the chip and this baseline has to be flat to show the system is functioning correctly. Phase two is known as the association phase, this is where ligands are passed over the chip and binding causes an initial sharp rise in the sensorgram signal, this curve is used to determine the K_{on}. As the sensorgram levels off this is the steady-state phase where no more ligand is binding to the target. Once the ligand solution being passed over the chip is replaced with running buffer the dissociation phase begins and the specific bonds between the ligands and targets on the chip are broken, this causes a downwards slope curve and is how the K_{off} is determined. The final stage is the regeneration phase which is used to ensure any ligand is removed from the chip and to establish a steady baseline identical to the signal at the start of the experiment. Once the data has a suitable binding model fitted the dissociation constant (K_D) can be determined from the ratio of K_{off}/K_{on}, which is determined from the steady-state position on the sensorgram.
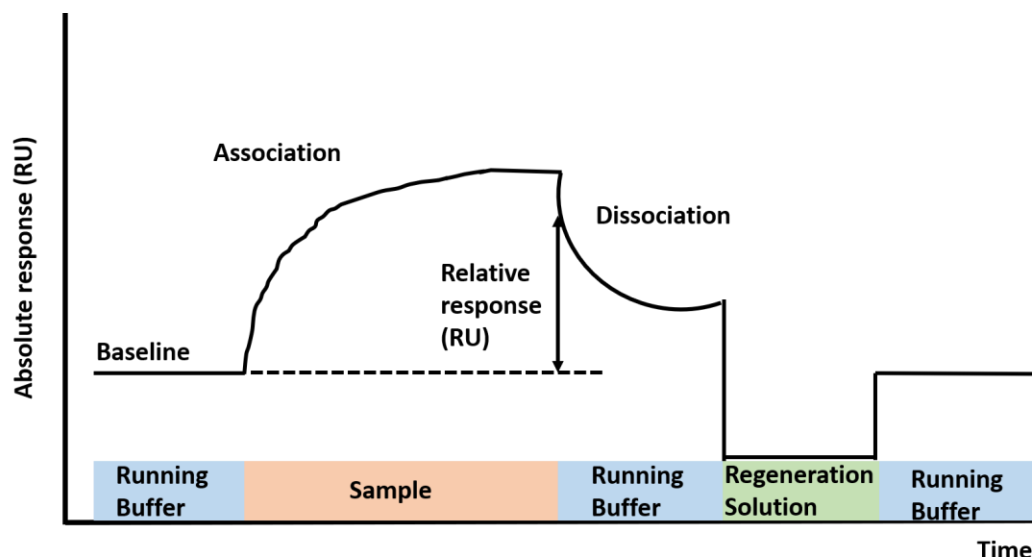
Figure 3.14: A labelled schematic illustration of a typical sensorgram produced from surface plasmon resonance.

The binding events between the peptides was analysed against eight different DNA structures B-DNA, c-MycC27, c-MycC52, c-MycG, DAP (i-motif), ATXN2L (i-motif), ILPR (i-motif), and hTeloC (i-motif) see *table 5.4* in *Chapter 5 experimental*. Two example sensorgrams demonstrating response vs time, a) for Pep-SLC and c-MycC52 and b) Pep-PTN and B-DNA, are shown in *figure 3.15*. There are two replicates for each concentration, shown by the same-coloured line, there is some variation shown between each replication. The 100 µM concentration for panel a), orange line, has only one replicate included as one replicate lost the usual shape of the sensorgram as shown in grey, it is likely this sample has precipitation. The rest of the sensorgrams shown in a) are of good quality, clearly showing all the phases described above and shown in the *figure 3.15*. The shape of the sensorgrams shown suggest a fast on and fast off rates of Pep-SLC binding to c-MycC52. On the other hand, *figure 3.15 b)* didn't show the typical sensorgram shape, there were no clear phases, the sensorgrams are jumping around the baseline other than the blue line which showed negative response units. All of these sensorgrams aren't smooth lines and demonstrate no binding conversely to the sensorgram

shown in *figure 3.15 a)*. To view the rest of the sensorgrams produced see *appendix A7-11*.

a)

b)
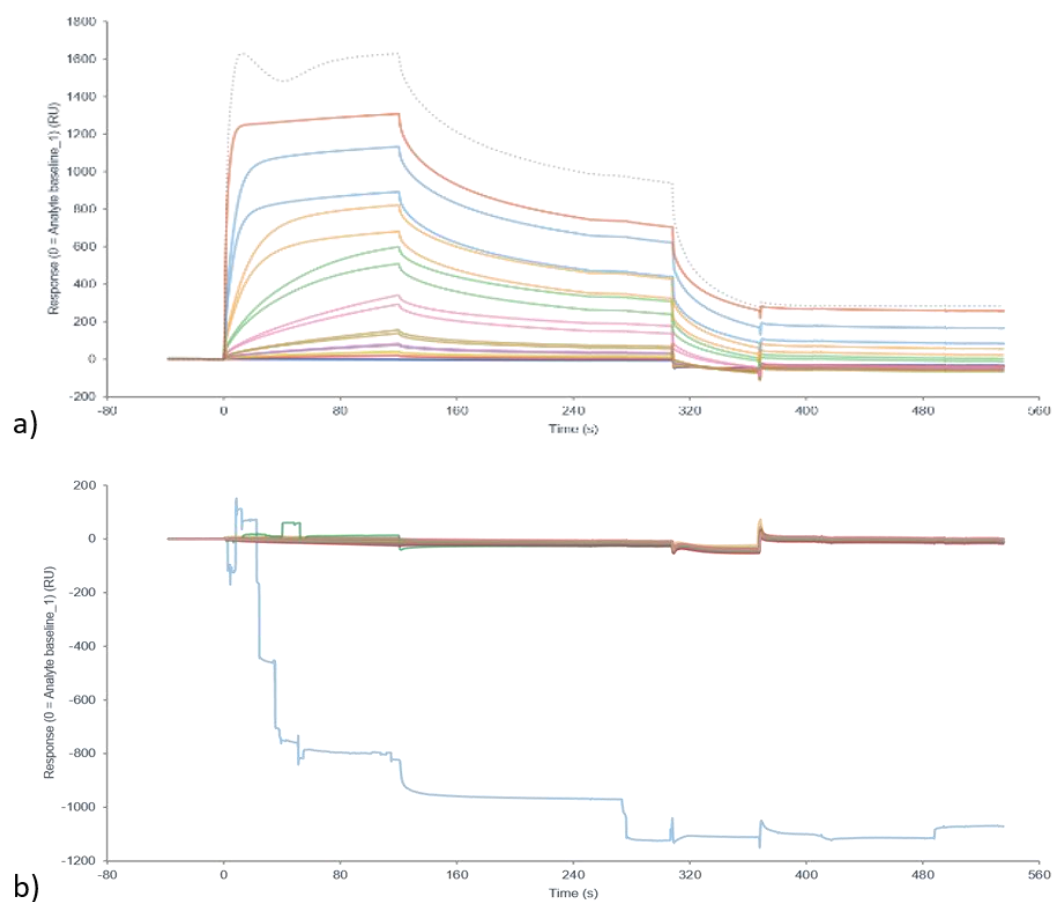
Figure 3.15: Surface plasmon resonance sensorgram of response vs time plots at concentrations from 0 μM to 100 μM in 2-fold intervals in the buffer 10 mM NaCaco 10 mM KCl 0.05% Tween 20 pH 6.6. There are two replicates for every concentration shown by the same colour. a) Sensorgram plot of Pep-SLC with c-MycC52, the top sensorgram is greyed out as the typical shape has been lost and so is removed. b) Sensorgram plot of Pep-PTN with B-DNA, the buffer has an addition of 2% DMSO. SPR experiments used Biacore SPR software.

The data obtained from the sensorgrams was plotted as saturation curves to evaluate the affinity of the peptides to DNA structures, plotting concentration (μM) vs Response Units (RU) shown in *figure 3.16*. The

graph labelled as *a)* shows the data for Pep-PTN against every DNA structure, many of the saturation curves demonstrate large error. The error is specifically at the top concentration which is likely due to precipitation and large negative responses, which can be for a range of reasons which is discussed in the *Section discussion and future work 3.6.* Due to the poor quality of this data when plotting all the DNA on one graph the initially more promising looking data for c-MycC52 is dwarfed and so it is plotted alone as graph *b)*. However, even with this data plotted separately the RU is low compared to that seen in graphs *c)* and *e)*, there's large error, the line of best fit is poor, and the graph didn't saturate suggesting non-specific binding. SPR analysis of Pep-VSE also had poor quality data. The only data that the software could fit a line of best fit was for c-MycC52. However, the data is poorly fitted, and saturation is not achieved suggesting unspecific binding to the chip. Pep-SLC and Pep-RVS, respectively *c)* and *e)*, demonstrated good quality SPR data and clearly showed binding to all DNA structures. Pep-SLC demonstrated saturation in this range with all DNA except hTeloC and Pep-RVS doesn't appear to have reached saturated binding with any of the DNA structures. SPR analysis demonstrated weak binding of Pep-EIE to some DNA structures, *f)*. *Pep-EIE* was shown to bind c-MycC27, c-MycC52, c-MycG, and ILPR, for the rest of the DNA structures the graph suggests no binding. However, the shape of that data is similar to that for c-MycC27, c-MycC52, c-MycG, and ILPR but the response units were too low to fit a line. The shape of the data curves, low RU, and saturation not being achieved for Pep-EIE suggests non-specific binding. It could be worth repeating the SPR for Pep-EIE at a higher concentration range to make more definitive conclusions based on the data. All of the binding curves not achieving saturation suggests unspecific binding to the chip is occurring. This is supported by *figure 3.15 b)* sensorgram of Pep-PTN, which did not produce typical sensorgrams as shown in *figure 3.14*, further suggesting that unspecific binding to the chip is likely.

Interestingly, Pep-PTN and Pep-VSE both had poorly fitted saturation curves and produced negative signals in some of the results against different DNA structures. Negative SPR signals can be caused from an array of reasons including buffer mismatch, different behaviour of the reference and flow cell to the ligand, non-specific matrix interactions and non-specific reference interactions. For the peptides producing negative signal results the SPR will require repeating after some optimisation to eliminate these negative results seen. Interestingly only the peptides dissolved in DMSO demonstrate negative responses, and DMSO is known to affect the refraction in SPR. Therefore, a likely reason for this is a buffer mismatch due to DMSO, the samples dry stocks were dissolved in DMSO a ~year before the SPR experiments were ran, even though the same stock of DMSO was used in the running buffer this could be a reason for buffer mismatch. Further support for this rationale is that when using DMSO in the SPR a calibration curve is used to correct for the solvent, unfortunately the DMSO correction for this experiment failed and so could not be used to correct the graphs producing sub-optimal saturation curve fittings. Consequently, this data cannot be further analysed or commented on as they are not properly corrected and thus the data cannot be interpreted. To overcome this problem fresh dry stocks, running buffer, and calibration curve samples should be prepared and ran to effectively evaluate any binding. Therefore, these experiments will require repeating with a new optimised SPR protocol to confirm if they are indeed specific.
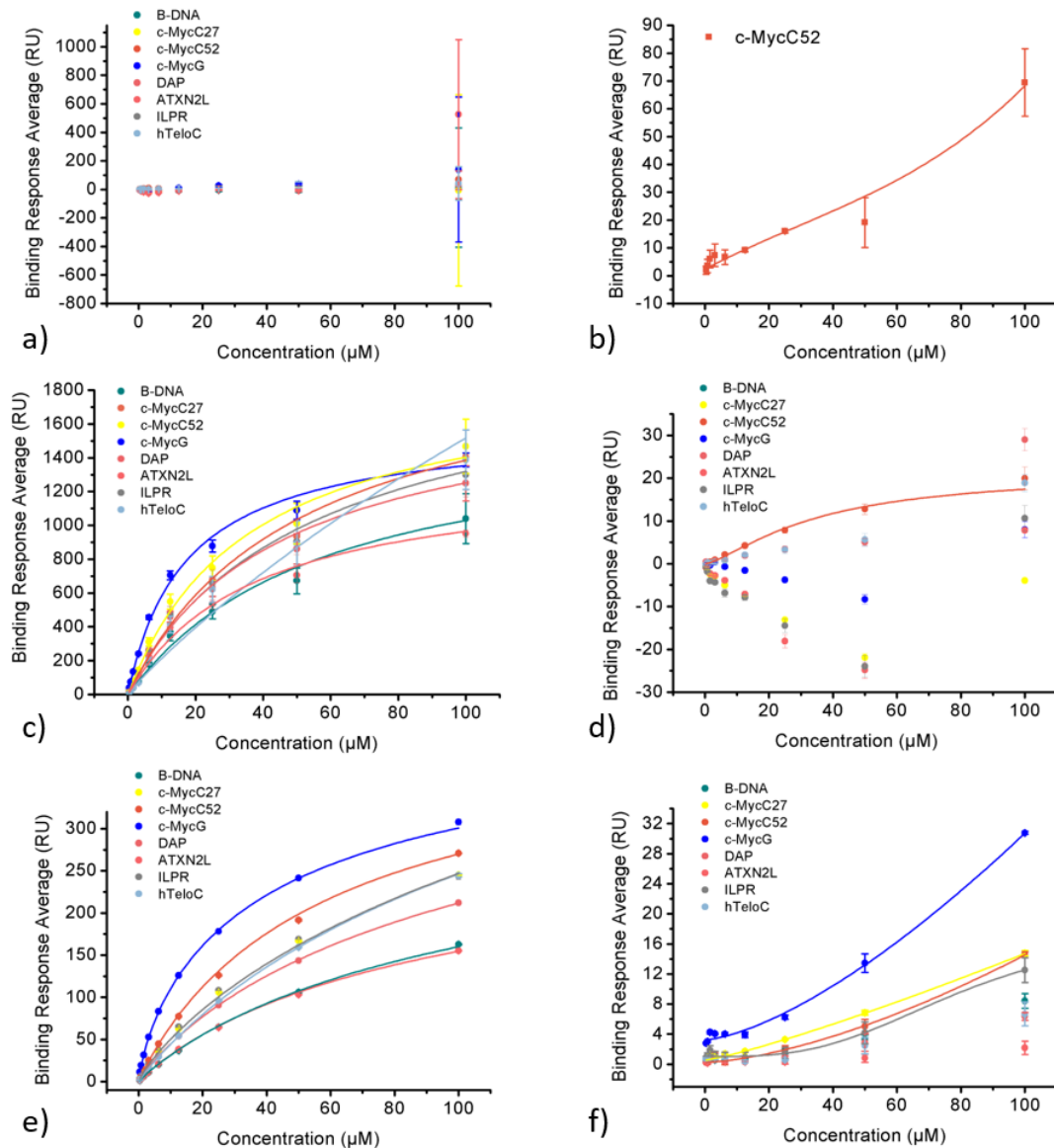
Figure 3.16: Saturation curves of five different peptides against eight different DNA structures using surface plasmon resonance in the conditions 10 mM NaCaco 10 mM KCl 0.05% Tween 20 pH 6.6 for experiments b), d), and e), the experiments a) and d) had 2% DMSO added to the conditions stated. a) Pep-PTN b) Pep-PTN and c-MycC52 plot only c) Pep-SLC d) Pep-VSE e) Pep-RVS f) Pep-EIE g) Structures of the peptides. The data analysis was carried out in OriginPro 8 and fitted using appropriate growth fittings dependant on curve shape. Where c-MycG is a G-quadruplex, and all the rest of the sequences except B-DNA are i-motif forming sequences.

Overall, of the three peptides which the saturation data can be interpreted for they all bound c-MycC52. Highlighting, that the phage display described in *Section 2.2* achieved the aims of this study to identify *MYC* i-motif binding peptides. Pep-SLC and Pep-RVS bind to all the sequences tested, showing a lack of specificity for c-MycC52. Pep-EIE appears to show specificity for certain i-motif and G-quadruplex structures but not B-DNA. Although, as discussed previously this data needs repeating at higher concentrations to make scientifically accurate observations.

To further analyse the SPR data K_D values were determined. To identify the K_D values for each peptide-DNA pair a kinetics model fitting needs to be applied to the data, usually 1:1 binding is assumed to do this, and in using this model a theoretical R_{max} can be calculated using the following equation 3.1:

$$R_{max} = \frac{Mw_{Ligand}}{Mw_{DNA}} \times R_{DNA}$$

Equation 3.1: Where R_{max} is maximal response, R_{DNA} is the response units measured upon immobilisation of a given DNA structure onto the surface of the chip, MW_{Ligand} is the molecular weight of the ligand and MW_{DNA} is the molecular weight of the DNA.

The R_{max} can be calculated as the response units in SPR that are proportional to mass of the ligand that binds to the target on the chip. The theoretical R_{max} using a 1:1 binding model for each peptide-DNA pair is shown in *table 3.20*. The R_{max} for the Pep-SLC and c-MycC52 pair was determined as 56.4 RU in *table 3.20* yet the actual response determined in *figure 3.16* is above this value for the concentration 1.56 µM and thus Pep-SLC doesn't bind to c-MycC52 via a 1:1 binding model.

Table 3.20: R$_{max}$, the maximal response, for each peptide-DNA pair assuming a 1:1 binding model for surface plasmon resonance. Where c-MycG is a G-quadruplex, and all the rest of the sequences except B-DNA are i-motif forming sequences.

DNA Sequence	Max Response (RU)				
	Pep-PTN	Pep-SLC	Pep-VSE	Pep-RVS	Pep-EIE
B-DNA	52.8	55.6	54.9	50.2	56.4
c-MycC27	89.3	94.0	92.9	84.9	95.3
c-MycC52	53.5	56.4	55.7	50.9	57.1
c-MycG	73.4	77.3	76.3	69.8	78.3
DAP	89.1	93.8	92.6	84.7	95.0
ATXN2L	97.3	102.4	101.1	92.4	103.7
ILPR	78.9	83.1	82.1	75.0	84.2
hTeloC	110.6	116.4	115.0	105.1	118.0

Therefore, a 2:1 binding model was used, as the fitting was more accurate than 1:1, to work out the K_D values using SOLVER function in excel as described in *Section 5.3.3*, *equation 5.1*. Only the peptide-DNA pairs that have fitted curves in *figure 3.16* were assessed for K_D values, except for Pep-PTN and Pep-VSE as it was determined the curve fittings were poor and taking a K_D value from them would be scientifically inaccurate. The K_D values were reported in *table 3.21*. Pep-SLC and Pep-RVS were shown to bind all DNA structures and were found to have the lowest K_D values with c-MycG and the largest with hTeloC and B-DNA. Demonstrating the strongest binding affinity with c-MycG and weakest with hTeloC and B-DNA. The K_D for Pep-SLC and Pep-RVS binding B-DNA were respectively 2-fold and 2.5-fold higher than that for c-MycG. Pep-EIE demonstrated differing response units against different DNA structures. However, Pep-EIE showed no true difference in K_D values for DNA structures it was determined as binding. Furthermore, Pep-EIE demonstrated the weakest binding affinities of any of the peptide-DNA interactions. This cooperates with the saturation curves not reaching saturation unlike what was seen with Pep-SLC and Pep-RVS.

Overall, the SPR data demonstrates phage display found peptides that bind c-MycC. Although, the peptides are not specific to c-MycC and further development is needed to achieve specificity.

Table 3.21: K_D values calculated for each peptide-DNA pair in 10 mM NaCaco 10 mM KCl 0.05% Tween 20 pH 6.6 (with 2% DMSO for Pep-PTN and Pep-VSE). N/D stands for not determined. Where c-MycG is a G-quadruplex, and all the rest of the sequences except B-DNA are i-motif forming sequences.

Peptide	K_D (µM) (SD)							
	B-DNA	c-MycC27	c-MycC52	c-MycG	DAP	ATXN2L	ILPR	hTeloC
Pep-PTN	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D
Pep-SLC	32.6 ± 2.4	32.6 ± 2.3	26.8 ± 1.0	14.4 ± 2.1	27.8 ± 0.4	22.4 ± 3.2	27.2 ± 2.2	40.5 ± 0.0
Pep-VSE	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D
Pep-RVS	47.1 ± 0.1	40.1 ± 0.2	33.9 ± 0.3	17.8 ± 0.6	39.8 ± 0.3	43.6 ± 0.2	36.6 ± 1.1	48.9 ± 1.3
Pep-EIE	58.0 ± 4.1	70.8 ± 10.3	62.4 ± 17.9	52.6 ± 5.6	20.8 ± 2.5	N/D	58.4 ± 13.9	N/D

3.4 Characterisation of Peptide Structure

The 12-mer peptides that were identified as c-MycC binders have the possibility to form some secondary structures due to their length. Understanding the structures of the peptides may be useful in identifying what makes a good i-motif binding peptide and highlight a consensus confirmation between structures. To do this far-UV circular dichroism (CD) was used. CD was chosen because it is excellent for rapid determination of secondary structures and once these are determined can go a step further to see how the structure is affected by the addition of the DNA structures of interest. The information obtained from the CD will not be as detailed as that from X-ray crystallography or NMR but it is faster and uses lower concentrations of the peptides.

To investigate peptide structure using CD the far UV region is used, 190 nm to 240 nm is the standard range used. Due to the far UV region being the shorter wavelengths these are more readily absorbed than the longer wavelengths and thus there are limitations on which buffers can be used. The phage display that these peptides were identified from was carried out in PBS pH 6.0 as discussed in *Section 3.3.1*, and biophysical analysis is carried out in NaCaco pH 6.0/6.6. Unfortunately, both buffers absorb in the far UV region and so the peptides structures could not be determined in these buffers. Therefore, 10 mM NaH_2PO_4 in the absence of NaCl was used in the far UV region.[223] Notably, the characteristic peaks for an α-helix are negative bands at 222 nm and 208 nm and a positive band at 193 nm,[290] whereas β-sheets have a negative band at 218 nm and a positive band at 195 nm,[5] and random coils have low ellipticity above 210 nm and negative bands near 195 nm (*figure 1.28*).[291]

Peptides Pep-SLC, Pep-RVS, and Pep-EIE were all dissolved in water. Whereas Pep-PTN and Pep-VSE have solubility issues and were dissolved in 100% DMSO for binding experiments. However, DMSO reads in the far UV region and so a sample of both dry stocks were dissolved in 50:50

MeCN:Water as this doesn't absorb significantly in the far UV region to allow the structure of the peptides to be determined. The sequence and structures were determined in *figure 3.10* and *figure 3.11*. If one compares Pep-SLC, Pep-RVS, and Pep-EIE in 10 mM NaCaco at pH 6.0 and pH 6.6 you can see the general shape is the same and the spectrum is dominantly negative, although the broadness and exact positions of the peaks dose change (*figure 3.17*). This demonstrates that the structure of these three peptides does not significantly change at the two different pH values. Pep-RVS and Pep-EIE have very similar spectrum traces, notably these two peptides both have an amino acid consensus sequence of TDHMK which may be contributing to the similar structures. Pep-PTN and Pep-VSE were analysed in 50:50 MeCN:water due solubility and have different characteristics compared to the peptides in the 10 mM NaCaco buffering conditions, but this is likely due to the buffer. Comparing the spectra for the five peptides against the reference spectra, (*figure 1.28*). All of the peptides CD spectrums bear some of the hallmarks for alpha-helix and random coil and therefore are a mix.
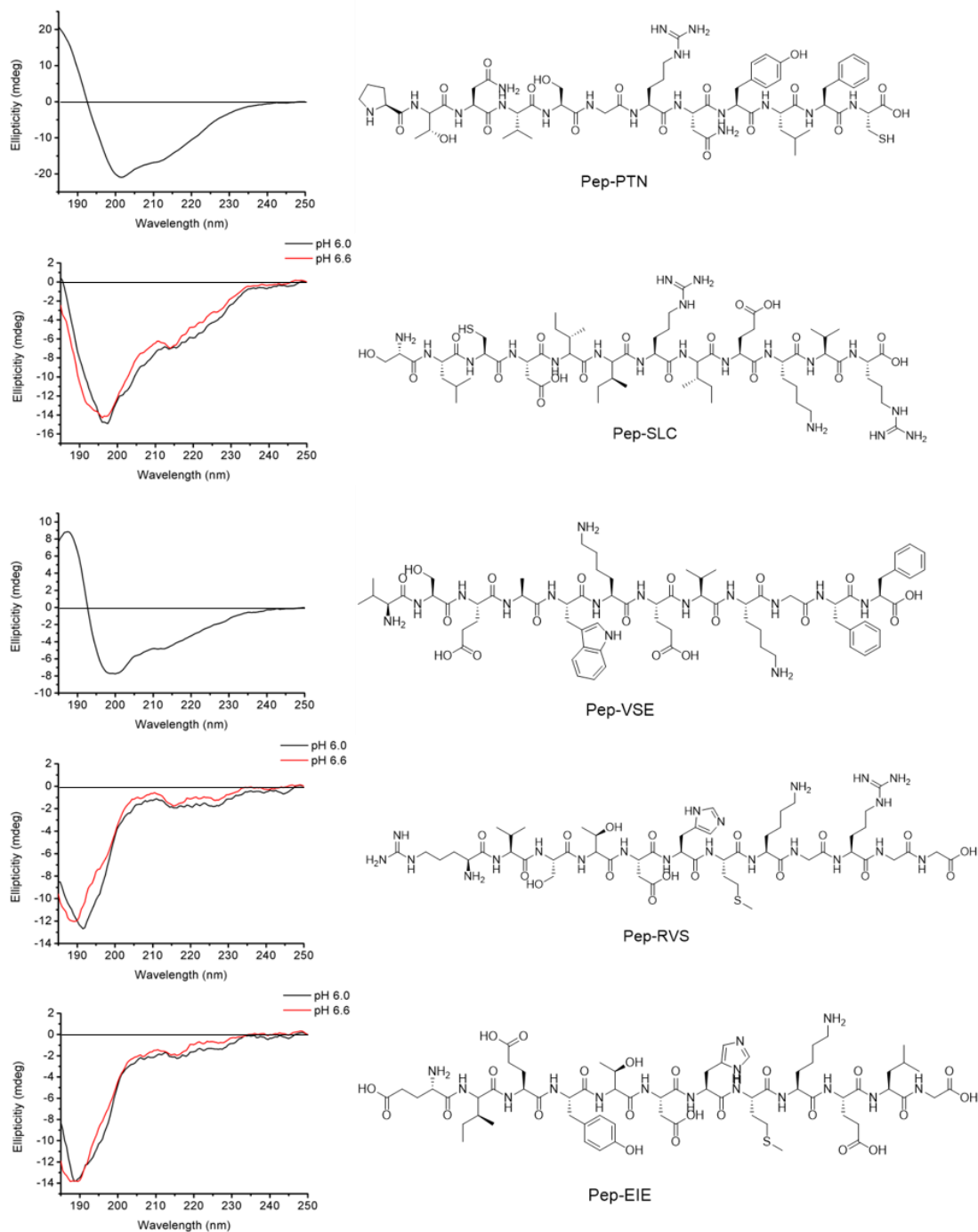
Figure 3.17: Circular dichroism characterisation 12-mer peptides of interest secondary structures: 0.1 mg/mL Pep-SLC, 0.1 mg/mL Pep-RVS, and 0.01 mg/mL Pep-EIE in 10 mM NaH₂PO₄ at pH 6.0 and pH 6.6 and 0.1 mg/mL Pep-PTN and 0.1 mg/mL Pep-VSE in 50:50 MeCN:Water.

3.5 Discussion and Future Work

The sequences identified from phage display using any of the different conditions explored in this chapter did not produce a clear consensus when selecting for c-MycC27 binders. This paired with 20%-60% frequency of insertless phage being identified for each method used suggested that either there were no c-MycC27 binders under the conditions used or the binders have a very weak affinity.

However, the peptide sequence MHPNAGHGSLMR (Pep-MHP) was identified three times in method three and SQDIRTWNGTRS (Pep-SQD) twice in method three which leads to the question was this significant. The answer to this is not straight forward as 40% of the sequences identified from this method were insertless and these two peptides do not show sequence similarity. Furthermore, the other sequence identified using this method was SGVYKVAYDWQH (Pep-SGV), and both Pep-MHP and Pep-SGV have been identified as fibronectin domain binders in another study.[292] This implies that these peptides are enriched in the library or have favourable amplification in *E. coli*. Although, proteins with these domains are usually in the extracellular matrix not the nucleus where the i-motif would be located. Therefore, they could still hold promise as c-MycC27 binders. However, based on the percentage of insertless sequences and identification in another study it is predicted they are poor binders as this is likely an artefact from weak binding pressure. Out of all the sequences identified for c-MycC27 there were no amino acid consensus sequences identified, further suggesting that phage display was unsuccessful in identifying c-MycC27 peptide binders.

There were 10 sequences identified as c-MycC52 binders, five from the Ph.D.-C7C library and five from the Ph.D.-12 library. Interestingly, none of these sequences had been identified in the c-MycC27 screen, *figures 3*.7 and *3.10*. The sequencing following the screening of the Ph.D.-C7C library identified 50% insertless sequences and the peptides identified shared no

consensus or clear similarities. This is of interest as in the sister field phage display was explored by Balasubramanian *et al.* to identify bicyclic peptides that bound G-quadruplexes.[293] These bicyclic peptides showed submicromolar affinity to G-quadruplex structures and selectivity over B-DNA. Notably, they did discuss that whilst the phage display identified bicyclic peptides from diverse libraries can bind G-quadruplexes they did have weak to modest binding affinities and that optimisation with greater selection pressures were required to achieve greater binding affinities. To do this they added genomic DNA in 100-fold excess, which was done in round 2 of the methods described in this project and so exploring bicyclic libraries could be an avenue explored to further build on the work in this project.

The Ph.D.-12 library had no insertless sequences found from 20 different samples, 80% of these sequences were Pep-EIE, and this peptide had a consensus with Pep-RVS which was TDHMK. These results suggested that the Ph.D.-12 library produced stronger binders as no insertless sequences where identified. Thus, the five peptides identified in c-MycC52 method one was taken forward as peptides of interest.

Comparing the results from FID, CD melt, and SPR, these studies corroborated that Pep-SLC and Pep-EIE bind c-MycC52. Notably, for Pep-EIE the SPR data could be repeated due to multiple factors yet it does still suggest weak binding. In the case of Pep-RVS the CD melt and SPR corroborated binding to c-MycC52. Whereas using FID the >15% displacement criteria to be determined a hit was not met. Although at a higher concentration it is possible it will be determined a hit as at 200 μM the displacement was 14%. It also must be considered that FID can miss ligands that bind to the structure of interest if they bind and do not cause the displacement of **TO**, where the ligand binds at a different site that **TO** to the target which does not cause structural changes that displace **TO** form its binding site. Which is a drawback of this method. Furthermore, FID is a method for analysing small compound binding and may require

further optimisation for peptides or should not be used at all. For peptides Pep-PTN and Pep-VSE the FID determined binding to c-MycC52. Whereas, the SPR data had poor saturation curve data and fittings and so suggested no binding, although the DMSO correction was not applied and so the data cannot be analysed or used to determine if binding occurred. Furthermore, the CD melt data could not be used to determine if either compound bound to c-MycC52 as neither affected stability, and thus binding cannot be determined as a peptide could bind and not have an effect on stability.

Interestingly, the FID titration demonstrated that Pep-SLC had the lowest DC_{50} value, the concentration at which 50% of **TO** is displaced. Suggesting Pep-SLC had the greatest binding affinity to c-MycC52. Supported by SPR data that determined Pep-SLC had the smallest K_D value compared to the other peptides for binding to c-MycC52. CD data cannot support binding affinity, however it is interesting that it caused the greatest change in stability, although this was for the second i-motif population and had no effect on the first i-motif population, Thus, the FID titration and SPR experiments suggest Pep-SLC as the peptide with the strongest binding affinity for c-MycC52. Notably, the SPR determined that the peptide with the weakest affinity for c-MycC52 was Pep-EIE, which corroborates with FID data which demonstrated Pep-EIE as the second worst ligand. Ranking the peptides from best to worst c-MycC52 binders for each technique are (where for FID titration and SPR the strongest affinity and for CD the greatest change in melting temperature were deemed best ranking):

FID Titration: Pep-SLC>Pep-PTN>Pep-VSE>Pep-EIE>Pep-RVS

CD: Pep-SLC>Pep-RVS>Pep-EIE>Pep-PTN>Pep-VSE

SPR: Pep-SLC>Pep-RVS>Pep-EIE

The three different biophysical approaches did agree that Pep-SLC, Pep-RVS, and Pep-EIE all bind c-MycC52. Interestingly, all three

techniques determined that Pep-SLC is the best c-MycC52 binder. In the interest to the i-motif field, to date, there are no currently published peptides that bind i-motif highlighting at least three, if not all five peptides are novel. Although, the sister field has had some recent successes when exploring peptides as G-quadruplex binders. Dash *et al.* in 2018 explored thiazole peptides targeting c-MycG. They identified a thiazole peptide, TH$_3$ (*figure 3.18*),[8] which selectively bound c-MycG over B-DNA and other G-quadruplexes and was shown to inhibit MYC transcription which was determined to be by a G-quadruplex dependent mechanism. However, binding against i-motifs was not explored and so without that data one cannot rule out that this thiazole peptide also binds to c-MycC and the effects on *MYC* transcription are not due to a dual interaction with c-MycG and c-MycC. Additional, bicyclic peptides have been identified as G-quadruplex ligands from phage display which were confirmed using FRET melting, fluorescence-quench equilibrium binding and computational modelling.[293] It looks as though these peptides bind G-quadruplexes however their binding to i-motifs was not explored and thus it would be interesting to see if these did bind i-motif. Highlighting, that the peptides identified in this project are novel and the first peptides shown to bind i-motif DNA and corroborate with the G-quadruplex field that peptides are an excellent route to explore to identify G-quadruplex and i-motif binding ligands.
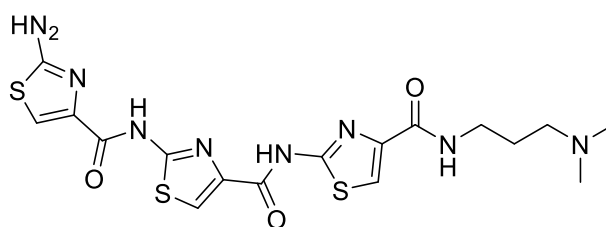
Figure 3.18: Structure of TH$_3$, a thiazole peptide that targets the G-quadruplex that forms in the MYC promoter.[8]

Currently, in the literature there are no known i-motif binding peptides, only small compounds.[252] Thus, this project presents novel data showing the first identified i-motif binding peptides.

The future work of this project would include repeating the SPR experiments Pep-PTN and Pep-VSE as previously mentioned. This will require repeating with a new optimised SPR protocol with an optimal DMSO correction to confirm if they bind any DNA structures. Alternatively, in the future isothermal titration calorimetry (ITC) could be used to study the binding and determine binding affinities. This is advantageous because no calibration for DMSO is required as DMSO has no effect on how the technique obtains affinity data and requires no immobilisation or labelling. Although, one must note it uses larger amounts of sample and is more time consuming and labour intensive.

In addition, CD determined that the secondary structures of the peptides appeared a mix of alpha-helixes and random coils, *figure 3.17*. Therefore, future work could investigate this further using spectral analysis to estimate structural abundance and search for any structural similarities. This would aid the fields understanding of what structural elements are important in binding c-MycC. Furthermore, to deepen the fields understanding of what amino acids bind c-MycC the binding epitopes of the peptides need to be identified, this can be done using saturation transfer difference (STD) NMR. Another approach to further the fields understanding of the binding activities between c-MycC52 and the peptides would be to use the CD and titrate the DNA into the peptide whilst monitoring the peptide and DNA structure to give insights about binding, this can also be carried out where the peptide is titrated into the DNA. Notably, the usual i-motif buffer of NaCaco would not be suitable as it absorbs strongly in the far UV region, and therefore the DNA would need to be characterised in a suitable buffer such as 10 mM NaH_2PO_4.[223] However, this means that the buffers used for biophysical analysis would not be compatible with this experiment and arguably the insight gained on

structural changes is not relevant due to differing buffering conditions. Therefore, the use of STD NMR is preferable. Further support for using STD NMR is that the G-quadruplex field have successfully used it to further understand the interactions between the DNA and ligands. For example, Randazzo's group used it to investigate the interaction of distamycin A and a G-quadruplex.[294] If one wanted even more detail X-ray crystallography using ligand soaking or crystallisation could be used to determine the structures of the protein-DNA complex, however this is more complicated and time consuming than the STD NMR. Although, that has not stopped the G-quadruplex field from benefiting from the structural and spatial characterisation that can be determined from using X-ray crystallography. For example, Gratteri's group used it to investigate the binding interactions between the human telomeric G-quadruplex and the well-established G-quadruplex binding ligand, berberine.[295] Thus, there is strong evidence that STD NMR and X-ray crystallography would be ideal future experiments to gain structural and spatial characterisation of i-motif-ligand interactions. Although, X-ray crystallography, due to the time consuming and klarge expense to run, may be better used later when the field has peptides with stronger binding affinities or even show specificity.

After an improved understanding of what amino acids and structural features bind to c-MycC the next logical step would be to produce further phage libraries using the binding epitopes with randomised amino acids. Followed by biophysical analysis such as SPR to identify peptides with stronger binding affinities than first discovered and identify any specific peptides. This approach can be used to improve the initial weak binding affinities. Further studies into structural features and binding epitopes of importance as described above would be beneficial after identifying peptides with greater binding affinities and/or showing specificity. These structural features and epitopes can then be compared to those identified for the weaker binders.

Once i-motif or c-MycC selective/specific peptides were identified the work should be progressed further using cellular based assays assessing the peptides effects on transcription, such as luciferase assays.

To summarise, this project aimed to identify c-MycC binding peptides using the Nobel prize-winning technique, phage display and has been successful in doing so. Although, these peptides have shown a lack of specificity over other DNA structures including B-DNA. Nonetheless, this opens new and exciting avenues for the next stages of this work including structural and spatial characterisation to determine the residues involved in the biding interaction, designing peptides with improved binding affinities, and identifying the improved peptides effects on transcription.

Chapter 4: Discussion and Future Work

The aim of this PhD research was to develop novel probes that target the i-motif forming region of the *MYC* promoter (c-MycC). Ultimately, this aim was achieved using two different approaches, one targeting c-MycC with small compounds and the other with peptides.

4.1 Discussion

Chapter 2 explored compounds from the National Cancer Institute (NCI) Diversity Set IV plated library, which contained 1596 different compounds representing maximal structural diversity with a wide range of pharmacophores. This project aimed to identify c-MycC binding small molecules to expand the current set of i-motif binding probes, specifically those for c-MycC. Currently there are a range of known i-motif ligands that are small molecules and this is continually growing.[252] As far as the literature for c-MycC binding small molecules there are three, two that preferentially bind c-MycC[6,82] and one that preferentially binds either c-MycC or c-MycG (*figure 4.1*).[87] Using biophysical studies including FID, CD, and SPR the NCI library was narrowed down from 1596 to eight small molecules of interest. This research also set out to compare small molecules identified as c-MycC27 probes to those determined as c-MycC52 probes. This investigation was important to the field as much of the i-motif field work, especially before 2016, used the c-MycC27 oligonucleotide yet in 2016 Hurley showed that this sequence length doesn't form the i-motif theorised to form in cells leading to questions about the biological relevance of testing against c-MycC27. The sequence c-MycC52 was identified as being needed in cells to be transcriptionally active and these two sequences form different i-motif formations, *figure 2.10*.[170,171] Interestingly, the biophysical techniques did find different compounds

bound c-MycC27 and c-MycC52 and in some instances select compounds bound both. Given the current hunt for specific i-motif probes, this work poses the consideration of which c-MycC sequence to use. Based on the hypothesis of the longer sequence being required for transcriptional activation, those looking to use the probes in cells or to develop drug therapies should highly consider using c-MycC52.

Figure 4.1: Structures of MYC i-motif ligands. Made in ChemDraw 18.2.

Peptide probes were explored in *Chapter 3* to build on the limited number of c-MycC known probes. Of these known probes all are small compounds and have a preference for c-MycC or dual preference for c-MycC and c-MycG over B-DNA and other i-motif and G-quadruplex structures.[6,82,87] The i-motif field as a whole has a greater-known number of binders but none are peptides.[252] Therefore, identifying peptides that bind to i-motifs is novel to the field. In general peptides have advantages over small molecules and could offer higher target specificity and selectivity, high

potency of action, and lower toxicity.[296] In general, it is a good idea to have a wealth of different probes that bind c-MycC with different properties as different experimental procedures and exploration of cancer therapies will work optimally with different types of probes. The experimental approach to finding peptide probes was phage display, which had been successful in identifying an i-motif specific antibody fragment.[94] Furthermore, the sister field, G-quadruplex field, also identified success using phage display to identify G-quadruplex binding peptides,[293] the bicyclic peptides identified showed a selectivity for G-quadruplexes over B-DNA but were not tested against i-motif DNA. Furthermore, phage display was successful in identifying c-MycC binding peptides, five different c-MycC52 binding peptides were determined by FID, and further confirmed for three using CD melting and SPR. From these peptides two had a consensus sequence of TDHMK, Pep-RVS and Pep-EIE. The sister field, G-quadruplex field, has achieved more than this project as it has identified a cell penetrating thiazole peptide, TH$_3$ (*figure 2.18*), with a preference for c-MycG over other G-quadruplexes and B-DNA. Although, TH$_3$ binding against i-motif was not determined. This should be completed as many G-quadruplex ligands that were believed to be specific have been shown as also being able to bind i-motif,[75] there is a compound that has a preference for c-MycC and c-MycG which causes downregulation of MYC transcription.[87] The identification of peptides that bind c-MycC can be used as the starting point to benefit the field to further our understanding of the *MYC* promoter's complicated regulation. This also opens an interesting avenue for cancer therapeutics.

The main conclusions that can be made from the results obtained is that c-MycC probes have successfully identified and can be further developed for better binding affinity and specificity. Two compounds were identified, **317605** and **345647**, with the highest RU determined from SPR for c-MycC and c-MycG compared to other structures, and compounds **317605** and **354844** destabilised B-DNA and stabilised c-MycC and c-MycG in CD experiments. Furthermore, multiple novel i-motif binding peptides were identified. These probes are promising for further development to find

compounds with greater selectivity to enable exploration of effects on *MYC* transcription.

4.2 Future Work

It is crucial for the i-motif field to continue its quest to identify i-motif binding probes, particularly specific ones, and even more so ones that bind specifically to one i-motif. This will propel the field forward and allow an increased understanding of the different roles i-motifs are performing as well as being able to explore manipulating them as therapeutic treatments.

The literature and work carried out in *Chapter 2* highlighted that different techniques can generate data that doesn't appear to corroborate, however often these techniques are exploring different aims and are producing different pieces to the puzzle. For example, SPR is identifying binding affinities whereas CD is determining effects on the stability of structures. Furthermore, variations in buffers have been shown to cause differing results for example difference between using 10 mM NaCaco pH 6.6 and 10 mM NaCaco 10 Mm KCl pH 6.6. Thus, making it difficult to make accurate assumptions of what will occur *in vitro* or *in vivo*. It has been observed in the non-canonical DNA structures field that the ligands determined to be the most promising in biophysical experiments don't necessarily translate into transcriptional changes *in vitro*. The literature is naturally skewed with positive hits of compounds, some of which will not be able to get into cells or have poor stability such as BRACO-19.[297] Therefore, the next logical steps would be to test the effects of the eight compounds explored in *Section* 2.3 using SPR *in vitro*. This work was started using MTT assays to determine the cytotoxicity of the three compounds deemed of most promise. But should be expanded to all eight compounds of interest taking in consideration all of the biophysical data and that the ligands identified as most promising biophysically may not

necessarily correlate to cellular experiments. Cytotoxicity assays will establish the concentration range that each compound can be used at for further cellular assays. To increase the understanding of the mechanism of interaction between c-MycC or c-MycG with compounds their effects on *MYC* expression needs to be determined. Similar studies within the non-canonical DNA structures field researching *MYC* have used luciferase assays to assess different compounds effects on *MYC* transcription.[6,82,87,171] The best approach for this will be to use a non-cancerous cell line, such as NKE,[298] a normal cell line and a cancerous cell line where *MYC* expression is dysregulated (e.g. amplified), such as A549,[299] and compare the effects of each compound in these different cellular environments. This work will further our understanding of the functioning's of the *MYC* promoter in 'normal' and cancerous states to learn more about how *MYC* expression can be controlled in the two states and begin to explore the starting points for cancer therapies targeting the non-canonical structures forming in the *MYC* promoter. Although, whilst carrying out these experiments the challenge associated, for both the G-quadruplex and i-motif field, of positive controls for up and down regulation and disentangle the effects on the G-quadruplex and i-motif. As many compounds that bind to G-quadruplex or i-motif structures are not specific but preferentially bind.

However, Smith *et al.* determined that GQC-05 and 71795 respectively increased the number of G-quadruplexes and i-motif structures in cells. Therefore, these could be useful positive controls for downregulation of *MYC* and produce further support for the current hypotheses that both c-MycC and c-MycG cause downregulation of *MYC*. Although, finding a control for upregulation of *MYC* may be more difficult as it needs to destabilise the G-quadruplex and i-motif or stabilise the B-DNA. Furthermore, to disentangle if the effect on *MYC* transcription is due to binding to c-MycC or c-MycG could be determined using the CA46 exon-specific assay.[300] This has previously been used to discriminate between decreasing c-Myc levels through binding of c-MycG or a secondary effect. This assay worked by separating mRNA productions from exon 1,

the non-translocated exon which maintains c-MycG-maintaining MYC gene, and exon 2, the translocated exon with the c-MycG-lost chromosome. This determines if the effect of compounds is mediated by c-MycG providing greater understanding of the compounds mechanism of binding.[300] If the assay was carried out for c-MycC and in an identical manner for c-MycG one could determine if the compound is having effects through c-MycC, c-MycG, another binding method, or a mix of these options.

Following cellular studies and identification of the ligands that cause an effect on *MYC* expression the next steps would be to synthesis derivatives of these compounds of interest to improve their properties such as specificity, binding affinity, cytotoxicity, and effects on *MYC* expression. Following synthesis biophysical experiments would be carried out including those used in this project and then cellular experiments to determine effects on *MYC* transcription. This would be a critical step when exploring using ligands that bind c-MycC as a potential avenue for cancer therapeutics.

Chapter 3 identified peptides that bind to c-MycC. Notably, the SPR experiment for Pep-PTN and Pep-VSE needs to be repeated before any other work is completed. Due to the DMSO calibration failing and thus the data was not optimally corrected for DMSO, and thus the data could not be analysed or have conclusions drawn from it.

An interesting avenue following the SPR is to explore the peptides structural features and epitopes. The far-UV CD determined each peptide beard some of the hallmarks for alpha-helix and random coils and therefore were determined to be a mix. Currently in the literature there was no known i-motif binding peptides. Therefore, it would be interesting to investigate the structures of these peptides further to help identify what structural elements if any are shared to direct further investigations looking for i-motif binding peptides. To do this spectral analysis should be used to estimate structural abundance and to search for structural

similarities. Furthermore, the structures of the peptides can be monitored using CD whilst titrating in i-motif. Following spectral analysis of that data it could be identified if upon binding to i-motif there is a structural change providing greater understanding of how the i-motif-peptide interaction occurs.

Moreover, a consensus sequence, TDHMK, was identified from sequencing of hit peptides. Suggesting that this consensus is important in binding to c-MycC52. To further investigate this consensus, sequence the binding interactions between c-MycC52 and the peptides with this consensus, Pep-RVS and Pep-EIE, need to be investigated to confirm if this consensus is involved in the binding interaction. There are current plans do this using STD NMR in collaboration with Dr Jesus Angulo which can be used to identify binding epitopes. STD NMR should also be used for all the peptides binding to provide greater information on the binding interactions. Another method that can be used to support this work is mutational studies of the peptides, changing any amino acid/multiple amino acids to an alanine, followed by biophysical assessments using FID, CD, and SPR to compare the effects seen on binding.

Another approach is to use alanine screening to determine the contribution of a specific amino acid to the peptides binding ability to c-MycC. Alanine is used for the substitution as it eliminates the side chain beyond the β carbon but dose not alter main-chain conformation like glycine does and can reveal functional hotspots.[301,302] Biophysical and cellular assays can then be carried out and results compared to identify which amino acids are most important in binding to c-MycC and the effects of substituting each amino acid can be used to build a greater understanding of how the c-MycC and the peptides interact. This knowledge could be used to inform construction of phage display libraries based on the binding epitopes followed by further phage display to identify peptides with greater binding affinities or further selectivity, proving the field with better i-motif/c-MycC probes. To further build on the fields understanding of how non-canonical

structures affect the regulation of the *MYC* promoter cellular studies will be required as described for the compounds above.

An alternative avenue to increase specificity and binding of the compounds and peptides identified in *Chapter 2* and *Chapter 3* is to use a linker to design peptides conjugated with small molecules. The integration of these utilise the advantages of peptide-based pharmacology with traditional medicinal chemistry. By offering the ability to enhance pharmacological effects whilst minimising dose-dependent toxicology.[303] Furthermore, they have a wide spectrum of biological targets and thus this could be an excellent avenue for the i-motif and G-quadruplex field to explore and could help the field find the specific probes they are both struggling to determine. This would be followed by biophysical techniques to determine binding affinity and specificity and progressed into cellular studies to explore the effects on *MYC* transcription.

Both the compound and peptide work have similar experimental "next steps" and both could also be explored using *in vitro* biophysical techniques such as immunofluorescence microscopy to determine the effects of the probes on i-motif and G-quadruplex stability using the iMab[94] and BG4[29] antibody/antibody fragment which selectively bind respectively to i-motif and G-quadruplex structures. This approach has only been possible for the i-motif since 2018 when iMab was first reported. It has been used to successfully further our understanding of different compounds effects on the stability of these "sister" structures *in vitro* and would provide greater information on the mechanisms involving the non-canonical structures of the *MYC* promoter.[97,304] For example, a similar study to that carried out by Smith *et al.*[97] could be ran using iMab[94] and BG4[29] to detect i-motif and G-quadruplex DNA respectively in human cells during the different stages of the cell cycle in the absence and presence of the probes of most interest from this project. This would allow exploration of the effects of the probes on i-motif and G-quadruplex structures at different stages of the cell cycle and in different cell lines such as a normal cell line, NKE,[298] and a

cancerous cell line, A549.[299] Pairing this data with results from other cellular assays exploring the probes effect on *MYC* transcription and exploring which DNA structure(s) the probes bind to in cellular conditions could provide valuable information on the mechanisms of the *MYC* promoter.

In addition, the work carried out in *Chapter 2* and *Chapter 3* both showed interesting CD melting experiments where some small molecules caused a second transition, highlighting that they were stabilising a second i-motif population. This opened an avenue for exploration of the effects of the small molecules on the i-motif's formation. One way to explore i-motif structural formation is using bromine footprinting which can be used to identify the cytosines involved in the $C-C^+$ base pairing and thus work out the bases in the loop regions.[305] Hurley *et al.* have used this technique to compare the formation of c-MycC in the presence and absence of hnRNP K,[4] therefore a similar approach could be taken to investigate what i-motif formation the small molecules identified in this project bind to. This would allow the field to begin to produce a database consisting of small molecules known to bind i-motif DNA and where they bind so that small molecules can be grouped by where they bind and their properties. Hopefully, this would uncover that small molecules with certain properties bind to one region and others to another and allow the field to uncover key properties for interacting with i-motifs to make discovering more probes easier. This could then be added to G4LBD 2.2,[306] the new version of G4LDB,[307] database. G4LDB 2.2 is a database containing data sets covering the physical properties and structure of G-quadruplex and i-motif ligands, provided web-based tools to assist in G-quadruplex and i-motif ligand design based on the information in the database, and facilitate the exploration of the prospect of G-quadruplex and i-motif probes as therapeutics for various diseases such as cancer.[306]

Overall, the future work for this project will build on the biophysical data investigating the binding between the probes and DNA structures

increasing the fields understanding of the properties of a c-MycC probe. Whilst the biological data explores the probes effects on *MYC* expression and the ability of c-MycC or c-MycG to be targeted as cancer therapies in the future.

Chapter 5: Experimental

5.1 General Experimental

Chemical reagents were used as supplied, unless otherwise stated, and were of general-purpose grade unless otherwise stated. All were purchased from Sigma-Aldrich, Thermo Fisher, or Alfa Aesar.

Oligonucleotides both modified and unmodified (*tables 5.1 and 5.2*) were purchased from Eurogentec, reverse phase-high performance liquid chromatography (RP-HPLC) and supplied dry. Dry DNA samples were dissolved in ultrapure water to make 1 mM stock solutions. Stock concentrations were confirmed using the extinction coefficients provided by the manufacturer and their UV absorbance at 260 nm with a Nanodrop ND-1000 spectrophotometer for each sequence.

Table 5.1: Oligonucleotide sequences used throughout this research. When text refers to B-DNA it consists of the B-DNA and B-DNA compliment annealed together, SS refers to the B-DNA sequence alone, and holiday junction is made up of Hjb, Hjh, Hjr, and Hjx.

Name	Sequence 5'-3'
ATXN2L	CCC-CCC-CCC-CCC-CCC-CCC-CCC-CCC
c-MycC27	CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA
c-MycC52	CTT-CTC-CCC-ACC-TTC-CCC-ACC-CTC-CCC-ACC-CTC-CCC-ATA-AGC-GCC-CCT-CCC-G
DAP	CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC
hTeloC	TAA-CCC-TAA-CCC-TAA-CCC-TAA-CCC
B-DNA	GGC-ATA-GTG-CGT-GGG-CGT-TAG-C
B-DNA Complement	GCT-AAC-GCC-CAC-GCA-CTA-TGC-C
C-hairpin	CTC-TCT-TCT-CTT-CAT-TTT-TCA-ACA-CAA-CAC-AC
c-MycG	TGG-GGA-GGG-TGG-GGA-GGG-TGG-GGA-AGG-TGG-GGA
hTeloG	GGG-TTA-GGG-TTA-GGG-TTA-GGG
NasC	GGG-CGG-GCT-GGG-CAT-TGC-GGG
NasT	GGG-AGC-GGG-ACG-GGG-GCC-GGG
Hjb	CGG-TAG-CAG-TAC-CGT-TGG-TGG-C
Hjh	GCC-TAG-CAT-GAT-ACT-GCT-ACC-G
Hjr	GCC-ACC-ACC-GGC-GTC-AAC-TGC-C
Hjx	GGC-AGT-TGA-CGT-CAT-GCT-AGG-C
RNA C1UUU	CUU-UCU-UUC-UUU-CUU-UC

Table 5.2: Modified oligonucleotide sequences used throughout this research. When text refers to B-DNA$_{Biotin}$ it consists of the B-DNA$_{Biotin}$ and B-DNA compliment annealed together. Biotinylated sequences were modified at the 5' end using Biotin dR 521.57 and were used throughout this research for phage display and SPR experiments.

Name	Sequence modification-5'-3'
B-DNA$_{Biotin}$	Biotin-GGC-ATA-GTG-CGT-GGG-CGT-TAG-C
c-MycC27$_{Biotin}$	Biotin- CCT-TCC-CCA-CCC-TCC-CCA-CCC-TCC-CCA
c-MycC52$_{Biotin}$	Biotin- CTT-CTC-CCC-ACC-TTC-CCC-ACC-CTC-CCC-ACC-CTC-CCC-ATA-AGC-GCC-CCT-CCC-G
c-MycG$_{Biotin}$	Biotin- TGG-GGA-GGG-TGG-GGA-GGG-TGG-GGA-AGG-TGG-GGA
DAP$_{Biotin}$	Biotin-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CCG-CCC-CC
ATXN2L$_{Biotin}$	Biotin-CCC-CCC-CCC-CCC-CCC-CCC-CCC-CCC
ILPR$_{Biotin}$	Biotin-TGT-CCC-CAC-ACC-CCT-GTC-CCC-ACA-CCC-CTG-T
hTeloC$_{Biotin}$	Biotin-TAA-CCC-TAA-CCC-TAA-CCC-TAA-CCC

DNA annealing, unless stated otherwise, was performed in an Applied Biosystems Veriti 96 well thermal cycler. Samples were heated to 95 °C and held at this temperature for 5 minutes, then returned to 20 ˚C at a rate of 1 ˚C/min and left at 20 ˚C overnight (a minimum of 12 hours at 20 ˚C due

to the time it takes for c-MycC52 to fold, all DNA was annealed in this manner to keep conditions identical).

Circular dichroism (CD) experiments were performed using a Jasco J-810 spectropolarimeter using a 1 mm path length quartz cuvette. Samples were scanned using a scanning speed of 200 nm/min, response time of 1 s, 0.5 nm data pitch, a 2 nm bandwidth, and accumulation of 4 scans. Data was zeroed at 320 nm unless otherwise stated, and, if required, corrected for compounds that produce a CD spectrum in the regions of interest where relevant. Data was analysed in OriginPro 8. CD melting experiments use the same measurement parameters, the sample is cooled brought down from room temperature to 5 °C and held at this temperature for 10 minutes, then heated at a rate of 1 °C/min up to 95 °C with measurements taken at 5 °C intervals. OriginPro 8 data analysis software was used to plot normalised ellipticity at the wavelength with the largest positive ellipticity, 285 nm for c-MycC27/52, against temperature to calculate the melting temperature. The Tm was calculated using the fitting from the sigmoidal dose-response curves.

Fluorescent Indicator Displacement (FID) experiments were performed on a BMG CLARIOstar plate reader using BRAND 96-well, solid black flat bottom plates. A 10 mM **TO** stock was prepared in **DMSO** and diluted in the appropriate buffer to 2 µM. Each well had 9 µL of the 2 µM **TO** solution added and was excited at 430 nm with fluorescence emission at 450 nm measured; this was normalised to 0% to account for background fluorescence. 1 µL of 90 µM DNA is added to each well and shaken in the plate reader for 30 seconds using double orbital shaking at 700 rpm and left for 10 minutes to equilibrate. Following equilibration, fluorescence emission was measured and normalised to 100% depicting maximum fluorescence from **TO** binding to the DNA. 2.5 µM of ligand was added into each well (in triplicate) in 2.25 µL and measure as stated above.

Fluorescence emission for each well was normalised between 0 and 100% which was taken away from 100 to give the percentage of displacement; ligands that averaged ≥15% displacement was considered hits and investigated further. FID titration studies were carried out on ligands considered hits, to identify the concentration at which each ligand displaces 50% of **TO** known as a DC_{50} value. Titration experiments were performed as described above until addition of the ligand, which is adjusted to X μM ligand in 0.9 μL of each ligand (in triplicate) 10 times to each well and measured as above. Fluorescent measurements after each ligand addition were analysed as described above. This data was analysed in Origin data analysis software and fitted with hyperbolic dose-response curves where the equations of the curves were solved for y = 50 to determine the DC_{50} values.

Surface Plasmon Resonance (SPR) measurements were performed using an eight-channel GE Healthcare Biacore with a series S streptavidin coated gold chip (Biacore SA-chip). On a single chip there are 8-channels available for immobilisation of DNA, each with their own reference flow cell. Immobilisation of biotinylated DNA was carried out in 0.01 M HEPES, 0.15 M NaCl pH 7.4, only DS was annealed prior to immobilisation as secondary structures, particularly the i-motif, won't form in these conditions. Immobilisation aimed to achieve 500 RU (*table 5.3*). The chip was then primed with 10 mM NaCacco, 10 mM KCl pH 6.6 and left overnight at 4 °C in this buffer to allow secondary DNA structures to fold, this buffer is a compromise between what is ideal conditions for i-motif (10 mM NaCaco, no KCl, acidic pH) and G-quadruplexes (10 mM NaCaco, 100 mM KCl). Notably the immobilisation of the DNA onto the chip could not be performed in 10 mM NaCacco, 10 mM KCl pH 6.6 because cacodylate results in reduced immobilisation and therefore 0.01 M HEPES, 0.15 M NaCl pH 7.4, a generic immobilisation buffer, was used. DNA binding experiments were performed at 25 °C, 120 s contact time, 120 s dissociation time, and a flow rate of 30 μL/min. A 10 mM NaCaco 10 mM KCl 0.05% Tween 20 pH 6.6 running buffer was used with the required percentage of

DMSO as stated *Section 5.2.2* and *Section 5.3.3*. Each injection was performed in duplicate and the average response at equilibrium was used for analysis. See *Section 5.2.2* and *Section 5.3.3* for further information on analysis methods used for compound and peptide experiments respectively.

Table 5.3: Oligonucleotide sequences immobilised onto the Biacore SA-chip, their flow cell location and response units of immobilisation. When text refers to B-DNA$_{Biotin}$ it consists of the B-DNA$_{Biotin}$ and B-DNA compliment annealed together.

DNA	Flow Cell	Response Units (RU)
B-DNA$_{Biotin}$	1	519.4
MycC27$_{Biotin}$	2	531.6
MycC52$_{Biotin}$	3	603.0
MycG$_{Biotin}$	4	572.3
DAP$_{Biotin}$	5	568.3
ATXN2L$_{Biotin}$	6	509.9
ILPR$_{Biotin}$	7	544.0
HTeloC$_{Biotin}$	8	559.6

5.2 General Experimental for Chapter 2

5.2.1 Compounds

Plated compounds were obtained from the National Cancer Institute (NCI). The plate set used was NCI Diversity Set IV with ~1500 compounds (20 plates), received at 20 µL/10 mM in 100% **DMSO**. The set was derived from ~140,000 compounds available for distribution from the Developmental Therapeutics program repository. The final set was elected using the programs Chem-X (Oxford Molecular Group) and Catalyst (Accelrys, Inc.). This defined pharmacophoric centres and defined distance intervals to create a finite set of three dimensional, 3-point pharmacophores, resulting in over 1,000,000 possible pharmacophores for

the set. To generate the diversity set, the pharmacophores for any candidate compound were compared to the set of all pharmacophores found in structures already in the set, if they differed tehy were added to the set. An additional aim was to create a diverse set of compounds that were amendable to forming structure-based hypotheses.[251] These plates were diluted to 1 mM in 100% DMSO. The compounds were >90% purity and in general were checked by NMR and LC-MS before delivery. The diversity set represents the maximal structural diversity with a wide range of pharmacophores.

Compounds of interest were obtained from the NCI Developmental Therapeutic Program's Open Compound Repository in larger quantities, 5 mg, as dry stocks in individual vials. Each compound was made up to 1 mM stock in 100% DMSO.

5.2.2 Surface Plasmon Resonance Experiments

SPR experiments were performed as described in *Section 5.1*. This SPR was ran as a high throughput screen to identify which compounds were of the most interest. The compound experiments were performed at a high and low concentration, 10 μM and 100 μM and diluted into the running buffer. The compounds stocks were dissolved in 100% DMSO and following the dilution to 100 μM there was 1 % DMSO, the 10 μM sample also maintained a 1% DMSO concentration and the running buffer had 1% DMSO. Additionally, a solvent correction was carried out using four DMSO concentrations: 1.8%, 1.3%, 0.9%, and 0.5%.

SPR analysis plotted the response at equilibrium (RU) against the concentration (μM) of compound to identify if binding occurred and if it appeared concentration dependent.

5.3 Experimental for Chapter 3

5.3.1 General Procedure for Identification of Peptides Using Phage Display

Phage Display experiments followed the solution-phase panning with a biotinylated target and streptavidin plate capture procedure outlined in the New England Biolabs Ph.D. Phage Display Libraries Instruction Manual.[197] Except for the buffers used which were buffer PBS pH 6.6 or 6.0, wash buffer PBS with 0.01-0.05% Tween 20 pH 6.6 or 6.0 referred to as PBST, blocking buffer PBS with 5 mg/mL BSA pH 6.6 or 6.0 referred to as PBSB, elution buffer PBS pH 7.4. For the conditions of the phage display targeting c-MycC27 see *Table 5.4* and targeting c-MycC52 see *Table 5.5*.

Table 5.4: Phage display methods for panning against the target DNA sequence c-MycC27, the i-motif forming region in the MYC promoter that is 27 bases long.

Phage Display Method	Selection Round	Plate Coating	Beads Pull down	Phage Library	Phage	Target DNA	Blocking Buffer	Wash Buffer	Competitors added to Phage Mixture	Beads	Blocking	Washing
Method 1												
	1st	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-C7C	1x10^11 Phage Library, Ph.D.-C7C, in 500 μL PBS pH 6.6	10 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	None	13 μL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.6
	2nd	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-C7C	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 μL PBS pH 6.6	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 01% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-Hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU	1.3 μL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.6
	3rd	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-C7C	1x10^11 Amplified Phage Selection 2 Elution 3 in 500 μL PBS pH 6.6	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP	1.3 μL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.6
Method 2												
	1st	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-C7C	1x10^11 Phage Library, Ph.D.-	10 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	None	13 μL Blocked B-PBS	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.6

	Round	Target	Bead Capture	Library	Selection	Competitor	Block	Wash	Competitors	Bead Prep	Incubation	Final Wash
					C7C, in 500 µL PBS pH 6.6					Wash PBST x3		
	2nd	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-C7C	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 µL PBS pH 7.4	10 pmol c-MycC27	PBS pH 7.4, 5% BSA	PBS pH 7.4, .% Tween 20	None	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 7.4
	3rd	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-C7C	1x10^11 Amplified Phage Selection 2 Wash in 500 µL PBS pH 6.6	10 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	None	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	3 x PBST pH 6.6
	4th	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-C7C	1x10^11 Amplified Phage Selection 3 Elution 3 in 500 µL PBS pH 6.6	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.6, 0.1% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-Hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.6
Method 3												
	1st	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-12	1x10^11 Phage Library, Ph.D.-12, in 500 µL PBS pH 6.0	10 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.0

2nd	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-12	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.0, 0.1% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-Hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU	1.3 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.0
3rd	Streptavidin	Beads pulled to the bottom of the tube 5 mins	Ph.D-12	1x10^11 Amplified Phage Selection 2 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC27	PBS pH 6.6, 5% BSA	PBS pH 6.0, 0.1% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP Hif1α	1.3 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.0
Method 4											
1st	Streptavidin	No beads, used a streptavidin plate	Ph.D-C7C	1x10^11 Phage Library, Ph.D.-C7C, in 500 µL PBS pH 6.0	10 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.0
Method 5											
1st	Streptavidin	No beads, used a streptavidin plate	Ph.D-12	1x10^11 Phage Library, Ph.D.-12, in 500 µL PBS pH 6.0	10 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.0
Method 6											
1st	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-12	1x10^11 Phage Library, Ph.D.-12, in 500 µL	10 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None	13 µL Blocked B-PBS	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.0

				PBS pH 6.0					Wash PBST x3		
2nd	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-12	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-Hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU	1.3 µL	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.0
3rd	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-12	1x10^11 Amplified Phage Selection 2 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC27	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP Hif1α	1.3 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST pH 6.0

Table 5.5: Phage display methods for panning against the target DNA sequence c-MycC52, the i-motif forming region in the MYC promoter that is 52 bases long.

Phage Display Method	Selection Round	Plate Coating	Beads Pull down	Phage Library	Phage	Target DNA	Blocking Buffer	Wash Buffer	Competitors added to Phage Mixture	Beads	Blocking	Washing
Method 1												
	1st	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-12	1x10^11 Phage Library, Ph.D.-12, in 500 μL PBS pH 6.0	10 pmol c-MycC52	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None	13 μL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST
	2nd	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-12	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 μL PBS pH 6.0	1 pmol c-MycC52	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU	1.3 μL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST
	3rd	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-12	1x10^11 Amplified Phage Selection 2 Elution 3 in 500 μL PBS pH 6.0	1 pmol c-MycC52	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP Hif1α	1.3 μL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST

Method 2											
	1st	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-C7C	1x10^11 Phage Library, Ph.D.-C7C, in 500 µL PBS pH 6.0	10 pmol c-MycC52	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST
	2nd	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-C7C	1x10^11 Amplified Phage Selection 1 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC52	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor NasT NasC hTeloG c-hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU	1.3 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST
	3rd	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-C7C	1x10^11 Amplified Phage Selection 2 Elution 3 in 500 µL PBS pH 6.0	1 pmol c-MycC52	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.5% Tween 20	100 pmol of each competitor hTeloC ILPR ATXN2L DAP Hif1α	1.3 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST
Method 3											
	1st	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-12	1x10^11 Phage Library, Ph.D.-C7C, in 500 µL PBS pH 6.0	10 pmol B-DNA	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST
	2nd	Streptavidin	Beads pulled to the side of the tube 5 mins	Ph.D-12	1x10^11 amplified combined washes from selection 1 in 500 µL PBS pH 6.0	1 0 pmol c-MycC52	PBS pH 6.0, 5% BSA	PBS pH 6.0, 0.1% Tween 20	None	13 µL Blocked B-PBS Wash PBST x3	Eppendorf tubes end-over-end-1 hour	10 x PBST

| 3rd | Streptavidin | Beads pulled to the side of the tube 5 m ins | Ph.D-12 | 1x10^11 Amplified Phage Selection 2 Elution 3 in 500 µL PBS pH 6.0 | 1 pmol c-MycC52 | PBS pH 6.0, 5% BSA | PBS pH 6.0, 0.5% Tween 20 | 100 pmol of each competitor NasT NasC hTeloG c-hairpin Holiday Junction B-DNA Calf Thymus RNA C1UUU | 1.3 µL Blocked B-PBS Wash PBST x3 | Eppendorf tubes end-over-end-1 hour | 10 x PBST |

These experiments were performed using either High Capacity Streptavidin Magnetic Beads. The phage libraries used were the Ph.D.-12 or Ph.D.-C7C Phage Display Peptide Library Kits from New England BioLabs. These kits are random combinational libraries where the peptide is fused to pIII, a minor coat protein, and displayed on the N-terminus of pIII of the M13 phage. The peptides are displayed in the form X_{12}-GGGS or AC-X_7-CGGGS respectively for the 12-mer and cyclic-7-mer where GGGS is a short linker sequence between pIII and the displayed peptide, and the flanking C residues in the Ph.D-C7C kit form a disulfide bond. Each library contains ~10^9 different electroporated (unique) sequences and 100 copies of each in 10 μL.[308,309]

The target for all experiments was the i-motif forming region in the *MYC* promoter, two variations of this sequence were panned against c-MycC27 and c-MycC52 which are respectively 27 and 52 base pairs long. For phage display these sequences were biotinylated on the 5′ end and for rounds with no competitors the concentration was 10 pmol, however in the presence of competitor sequences this was reduced 10-fold to 1 pmol. The competitor sequences were not biotinylated and were used at 100-fold higher concentration than the target sequence. The competitor sequences were split into two different groups: non-i-motif competitors and i-motif competitors. These two groups were used in different selection rounds to increase the selection pressure as additional phage selection rounds were carried out, generally a total of three selection rounds were carried out per method. The non-i-motif competitors were, B-DNA (double stranded, right-handed helical DNA, composed of B-DNA and B-DNA complement), hTeloG (G-quadruplex), NasC (G-quadruplex), NasT (G-quadruplex), holiday junction (composed of Hjb, Hjh, Hjr, and Hjx), c-hairpin, RNA C1UUU, and calf thymus (native DNA) as shown in *table 5.1*. The i-motif competitors used were hTeloC, ILPR, ATXN2L, DAP, and Hif1α as shown in *table 5.1*.

Host strain, as recommended by the New England Biolabs Ph.D. Phage Display Libraries Instruction manual, was an *E. coli* strain known as ER2738.[197] ER2738 was streaked from the glycerol stock onto LB agar plates with 1 mL/L of tetracycline stock (20 mg/ml in 1:1 ethanol:water) and incubated at 37 °C overnight and then stored covered with parafilm at 4 °C with a one month expiry date to ensure bacteria cell viablilty.[197] ER2738 cultures for infection were grown in LB only at 37 °C.

Phage tittering was used to determine the concentration of phage in a given sample in plaque forming unit (pfu) per 10 μL. This was carried out as per NEB's phage manual.[197] Except step one where 20 mL of LB was inoculated with ER2738 from a plate and incubated at 37 °C overnight. The phage- ER2738-Top Agar mixture was poured onto a pre-warmed LB Agar with 1 mL/L of IPGT/Xgal stock (1.25 g IPGT and 1 g Xgal in 25 mL DMF) plate over incubate overnight at 37 °C. The following morning blue plaques are counted and multiplied by their dilution factor to determining phage concentration, pfu per 10 μL.

Amplification of Phage is carried out as described in the New England Biolabs Ph.D. Phage Display Libraries Instruction manual page 16.[197] When phage were left to precipitate the first time in 20% PEG/2.5 M NaCl at 4 °C overnight. Amplified phage were stored in sterile conditions in blocked tubes (5 mg/mL BSA PBS pH 7.4) in PBS pH 7.4 at 4 °C.

Sequencing of Phage DNA were performed by Eurofin Genomics using their Mix2Seq overnight kit according to manufacturer's instructions using -96 gIII sequencing primer provided by NEB. Phage sent for sequencing were isolated from IPGT/Xgal plates from the phage titering after a round of panning and individually amplified using ER2738. Following *E. coli* removal phage DNA was extracted using ethanol precipitation.

5.3.2 Peptide Sequences

Peptides were purchased from Cellmano Biotech, RP-HPLC and supplied dry with N-terminal acetylation and a C-terminal amide (*Table 5.6*). Dry peptide samples were dissolved in ultrapure water or **DMSO** if insoluble in water to make 5 mg/mL stocks unless overwise stated. Stock concentrations were confirmed using the extinction coefficients of each peptide and their UV absorbance on an Agilent Technologies Cary 4000 UV-VIS spectrophotometer. The extinction coefficient was determined using Nick Anthis' Protein Parameter Calculator either at 280 nm if the sequence contained any tryptophan or tyrosine residues or 205 nm if the sequence didn't contain either of these aromatic residues.[217]

Table 5.6: Peptide sequences, 12-mers, shown N-terminus to C-terminus, with N-terminal acetylation and a C-terminal amide. Purity determined by HPLC analysis by the manufacturer.

Name	Sequence 1 Letter Code	Purity %
Pep-PTN	PTNVSGRNYLFC	95.09
Pep-SLC	SLCDIIRIEKVR	95.45
Pep-VSE	VSEAWKEVKGFF	98.49
Pep-RVS	RVSTDHMKGRGG	95.70
Pep-EIE	EIEYTDHMKELG	96.10

5.3.3 Surface Plasmon Resonance Experiments

SPR experiments were performed as described in *Section 5*.1. The peptides were tested at 10 different concentrations from 100 μM to 0 μM using a 2-fold dilution series, in running buffer. Testing of Pep-SLC, Pep-RVS, and Pep-EIE stocks were in water so no **DMSO** was added to the running buffer. For Pep-PTN and Pep-VSE their stocks were in 100% **DMSO** and following dilution to 100 μM the **DMSO** % was 2% **DMSO**,

further serial dilutions were performed to maintain the concentration of **DMSO** at 2% for all concentrations and the running buffer had 2% **DMSO**. Additionally, a solvent correction was carried out using four **DMSO** concentrations: 2.8%, 2.3%, 1.9%, and 1.5%.

SPR analysis plotted the response at equilibrium (RU) against the concentration (μM) of peptide to generate a hyperbolic binding curve. For all peptides the dissociation constants were determined using a 2:1 binding model using the *equation 5.1*.[310] Each experiment was repeated in duplicate and the average of the two K_D values was reported, the error represents the variance between the two values.

$$r = \frac{RU}{RU_{max}} = \frac{K_1 C + 2K_1 K_2 C^2}{1 + K_1 C + K_1 K_2 C^2}$$

Equation 5.1: Where r is the number of moles of analyte bound per molecule of DNA, C is the free analyte concentration, RU is the steady state response, RU_{max} is the maximum response per molecule bound of analyte, K_1 and K_2 are the association constants for the first and second binding sites.

References

1.   Wierstra, I. & Alves, J. The c-myc Promoter: Still MysterY and Challenge. *Adv. Cancer Res.* **99**, 113–333 (2008).

2.   Lemaitre, J. M., Buckle, R. S. & Méchali, M. c-Myc in the Control of Cell Proliferation and Embryonic Development. *Adv. Cancer Res.* **70**, 95–144 (1996).

3.   Brooks, T. A. & Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **1**, 641–649 (2010).

4.   Sutherland, C., Cui, Y., Mao, H. & Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III 1. *J. Am. Chem. Soc.* **138**, 14138–14151 (2016).

5.   Greenfield, N. J. & Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116 (1969).

6.   Saha, P., Panda, D., Müller, D., Maity, A., Schwalbe, H. & Dash, J. In situ formation of transcriptional modulators using non-canonical DNA i-motifs. *Chem. Sci.* **11**, 2058–2067 (2020).

7.   Desjardins1, E. & Hayl, N. *Repeated CT Elements Bound by Zinc Finger Proteins Control the Absolute and Relative Activities of the Two Principal Human c-myc Promoterst. MOLECULAR AND CELLULAR BIOLOGY* vol. 13 http://mcb.asm.org/ (1993).

8.   Dutta, D., Debnath, M., Müller, D., Paul, R., Das, T., Bessi, I., Schwalbe, H. & Dash, J. Cell penetrating thiazole peptides inhibit c-MYC expression via site-specific targeting of c-MYC G-quadruplex. *Nucleic Acids Res.* **46**, 5355–5365 (2018).

9.   Miescher, F. Miescher F (1869a) Letter I: to Wilhelm His; Tübingen, February 26th, 1869. In Die histochemischen und physiologischen arbeiten von Friedrich Miescher. vol. 1 33–38 (1869).

10. Miescher, F. l= U"\berdie Chemische Zusammen-setzung der Eiterzellen. *Hoppe-Seylers Med Chem Untersuchgn, Berlin* 441 (1870).

11. Dahm, R. Friedrich Miescher and the Discovery of DNA. *Dev. Biol.*

278, 274–288 (2005).

12. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet* **122**, 565–581 (2008).

13. The discovery of DNA. *yourgenome.org* yourgenome.org https://www.yourgenome.org/stories/the-discovery-of-dna (2021).

14. Chargaff, E. & Vischer, E. The composition of the desoxypentose nucleic acids of thymus and spleen. *J. Biol. Chem.* **177**, 405–416 (1949).

15. Chargaff, E. Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10**, 654–9 (1951).

16. Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *J. Exp. Med.* **79**, 137–158 (1944).

17. Hershey, A. D. & Chase, M. Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage. *J Gen Physiol* **36**, 39–56 (1952).

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

19. Choi, J. & Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **40**, 5893–5909 (2011).

20. Kaushik, M., Kaushik, S., Roy, K., Singh, A., Mahendru, S., Kumar, M., Chaudhary, S., Ahmed, S. & Kukreti, S. A bouquet of DNA structures: Emerging diversity. *Biochem. Biophys. Reports* **5**, 388–395 (2016).

21. Potaman, V. N. & Sinden, R. R. DNA: Alternative Conformations and Biology. *In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience* https://www.ncbi.nlm.nih.gov/books/NBK6545/.

22. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci.* **48**, 2013–2018 (1962).

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.   FW, S. & J, F. Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature* **356**, 164–168 (1992).

25.   Bergues-Pupo, A. E., Arias-Gonzalez, J. R., Morón, M. C., Fiasconaro, A. & Falo, F. Role of the central cations in the mechanical unfolding of DNA and RNA G-quadruplexes. *Nucleic Acids Res.* **43**, 7638–7647 (2015).

26.   Chowdhury, S. & Bansal, M. G-quadruplex structure can be stable with only some coordination sites being occupied by cations: A six-nanosecond molecular dynamics study. *J. Phys. Chem. B* **105**, 7572–7578 (2001).

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

28.   HB, C., H, B. & TZ, S. G4Boost: a machine learning-based tool for quadruplex identification and stability prediction. *BMC Bioinformatics* **23**, 240 (2022).

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

30.   Summers, P. A., Lewis, B. W., Gonzalez-Garcia, J., Porreca, R. M., Lim, A. H. M., Cadinu, P., Martin-Pintado, N., Mann, D. J., Edel, J. B., Vannier, J. B., Kuimova, M. K. & Vilar, R. Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat. Commun.* **12**, 1–11 (2021).

31.   González, V., Guo, K., Hurley, L. & Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* **284**, 23622–23635 (2009).

32.   Meier-Stephenson, V. G4-quadruplex-binding proteins: review and insights into selectivity. *Springer Link* **14**, 635–654 (2022).

33.   Kim, N. The Interplay between G-quadruplex and Transcription. *Curr. Med. Chem.* **26**, 2898–2917 (2019).

34.   Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D. & Antonio, M. Di. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nuclecic Acids Res.* **49**, 8419–8431 (2021).

35. Teng, F.-Y., Jiang, Z.-Z., Guo, M., Tan, X.-Z., Chen, F., Xu, X.-G. X. & Yong. G-quadruplex DNA: a novel target for drug design. *Cell. Mol. Life Sci.* **78**, 6557–6583 (2021).

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

37. Dembska, A. The analytical and biomedical potential of cytosine-rich oligonucleotides: A review. *Anal. Chim. Acta* **930**, 1–12 (2016).

38. Heinen, L. & Walther, A. Temporal control of i-motif switch lifetimes for autonomous operation of transient DNA nanostructures. *Chem. Sci.* **8**, 4100–4107 (2017).

39. Joo, H. N., Van Thi Nguyen, T., Chae, H. K. & Seo, Y. J. pH-Dependant fluorescence switching of an i-motif structure incorporating an isomeric azobenzene/pyrene fluorophore. *Bioorg. Med. Chem. Lett.* **27**, 2415–2419 (2017).

40. Dembska, A., Bielecka, P. & Juskowiak, B. pH-Sensing fluorescence oligonucleotide probes based on an i-motif scaffold: a review. *Anal. Methods* **9**, 6092–6106 (2017).

41. Sharma, J., Chhabra, R., Yan, H. & Liu, Y. pH-driven conformational switch of "i-motif" DNA for the reversible assembly of gold nanoparticles. *Chem. Commun.* **0**, 477–479 (2007).

42. Catasti, P., Chen, X., Deaven, L. L., Moyzis, R. K., Bradbury, E. M. & Gupta, G. Cystosine-rich strands of the insulin minisatellite adopt hairpins with intercalated cytosine+·cytosine pairs. *J. Mol. Biol.* **272**, 369–382 (1997).

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

44. Kanaori, K., Maeda, A., Kanehara, H., Tajima, K. & Makino, K. 1H Nuclear Magnetic Resonance Study on Equilibrium between Two Four-Stranded Solution Conformations of Short d(C n T). *Biochemistry* **37**, 12979–12986 (1998).

45. Benabou, S., Aviñó, A., Eritja, R., González, C. & Gargallo, R. Fundamental aspects of the nucleic acid i-motif structures. *RSC Adv.* **4**, 26956–26980 (2014).

46. Kanaori, K., Shibayama, N., Gohda, K., Tajima, K. & Makino, K. Multiple four-stranded conformations of human telomere sequence d(CCCTAA) in solution. *Nucleic Acids Res.* **29**, 831–840 (2001).

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

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

49. Kang, C., Berger, I., Lockshin, C., Ratliff4, R., Moyzisl, R. & Rich, A. *Crystal structure of intercalated four-stranded d(C3T) at 1.4 A resolution (CC+ base plrs/hydratlo/telomeres/base-stack ng/parafl-stranded duplex)*. *Chemistry* vol. 91 (1994).

50. Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J. & Dickerson, R. E. The crystal structure of the trigonal decamer C-G-A-T-C-G-6meA-T-C-G: A B-DNA helix with 10.6 base-pairs per turn. *J. Mol. Biol.* **231**, 768–784 (1993).

51. Berger, I., Egli, M., Rich, A. & Qu, X. Inter-strand C-H...O hydrogen bonds stabilizing four-stranded intercalated molecules: stereoelectronic effects of O4' in cytosine-rich DNA. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12116–12121 (1996).

52. Berger, I., Kang, C., Fredian, A., Ratliff, R., Moyzis, R. & Rich, A. Extension of the four-stranded intercalated cytosine motif by adenine•adenine base pairing in the crystal structure of d(CCCAAT). *Nat. Struct. Biol.* **2**, 416–425 (1995).

53. Rich, A. The rise of single-molecule DNA biochemistry. *Proceedings of the National Academy of Sciences of the United States of America* vol. 95 13999–14000 (1998).

54. Leroy, J.-L., Snoussi, K. & Guéron, M. Investigation of the energetics of C–H⋯O hydrogen bonds in the DNA i-motifvia the equilibrium between alternative intercalation topologies. *Magn. Reson. Chem.* **39**, S171–S176 (2001).

55. Malliavin, T. E., Gau, J., Snoussi, K. & Leroy, J. L. Stability of the i-motif structure is related to the interactions between phosphodiester backbones. *Biophys. J.* **84**, 3838–3847 (2003).

56. Lieblein, A. L., Buck, J., Schlepckow, K., Fürtig, B. & Schwalbe, H. Time-resolved NMR spectroscopic studies of DNA i-motif folding reveal kinetic partitioning. *Angew. Chemie - Int. Ed.* **51**, 250–253 (2012).

57. Gurung, S. P., Schwarz, C., Hall, J. P., Cardin, C. J. & Brazier, J. A. The importance of loop length on the stability of i-motif structures. *Chem. Commun* **51**, 5630–2 (2015).

58. Fujii, T. & Sugimoto, N. Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. *Phys. Chem. Chem. Phys.* **17**, 16719–16722 (2015).

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

60. Fleming, A. M., Ding, Y., Rogers, R. A., Zhu, J., Zhu, J., Burton, A. D., Carlisle, C. B. & Burrows, C. J. 4n−1 Is a " Sweet Spot " in DNA i-Motif Folding of 2'-Deoxycytidine Homopolymers. *J. Am. Chem. Soc* **139**, 4682–4689 (2017).

61. Zhou, J., Wei, C., Jia, G., Wang, X., Feng, Z. & Li, C. Formation of i-motif structure at neutral and slightly alkaline pH. *Mol. Biosyst.* **6**, 580–560 (2010).

62. Nguyen, T., Fraire, C. & Sheardy, R. D. Linking pH, Temperature, and K + Concentration for DNA i-Motif Formation. *J. Phys. Chem. B* **121**, 7872–7877 (2017).

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

64. Bhavsar-Jog, Y. P., Van Dornshuld, E., Brooks, T. A., Tschumper, G. S. & Wadkins, R. M. Epigenetic modification, dehydration, and molecular crowding effects on the thermodynamics of i-motif structure formation from C-rich DNA. *Biochemistry* **53**, 1586–94 (2014).

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

66. Miyoshi, D., Matsumura, S., Nakano, S. I. & Sugimoto, N. Duplex Dissociation of Telomere DNAs Induced by Molecular Crowding. *J.*

Am. Chem. Soc. **126**, 165–169 (2004).

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

68. 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**, 2863–74 (2009).

69. Saxena, S., Bansal, A. & Kukreti, S. Structural polymorphism exhibited by a homopurinehomopyrimidine sequence found at the right end of human c-jun protooncogene. *Arch. Biochem. Biophys.* **471**, 95–108 (2008).

70. Kim, S. E., Lee, I.-B. B., Hyeon, C. & Hong, S.-C. C. Destabilization of i-Motif by Submolar Concentrations of a Monovalent Cation. *J. Phys. Chem. B* **118**, 4753–4760 (2014).

71. Wang, Z. G., Elbaz, J. & Willner, I. DNA machines: Bipedal walker and stepper. *Nano Lett.* **11**, 304–309 (2011).

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

73. Abdelhamid, M. A., Fábián, L., MacDonald, C. J., Cheesman, M. R., Gates, A. J. & Waller, Z. A. Redox-dependent control of i-Motif DNA structure using copper cations. *Nucleic Acids Res.* **46**, 5886–5893 (2018).

74. Pagano, A., Iaccarino, N., Abdelhamid, M. A. S., Brancaccio, D., Garzarella, E. U., Di Porzio, A., Novellino, E., Waller, Z. A. E., Pagano, B., Amato, J. & Randazzo, A. Common G-Quadruplex Binding Agents Found to Interact With i-Motif-Forming DNA: Unexpected Multi-Target-Directed Compounds. *Front Chem.* **6**, 281 (2018).

75. Abdelhamid, M. A. S., Gates, A. J. & Waller, Z. A. E. Destabilization of i-Motif DNA at Neutral pH by G-Quadruplex Ligands. *Biochemistry* **58**, 245–249 (2018).

76. Li, X., Peng, Y., Ren, J. & Qu, X. Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation. *Proc. Natl. Acad. Sci.* **103**, 19658–19663 (2006).

77. Zhao, C., Ren, J. & Qu, X. Single-Walled Carbon Nanotubes Binding to Human Telomeric i-Motif DNA Under Molecular-Crowding Conditions: More Water Molecules Released. *Chem. - A Eur. J.* **14**, 5435–5439 (2008).

78. Chen, Y., Qu, K., Zhao, C., Wu, L., Ren, J., Wang, J. & Qu, X. Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres. *Nat. Commun.* **3**, 1074 (2012).

79. Jiang, T., Amadei, C. A., Gou, N., Lin, Y., Lan, J., Vecitis, C. D. & Gu, A. Z. Toxicity of single-walled carbon nanotubes (SWCNTs): effect of lengths, functional groups and electronic structures revealed by a quantitative toxicogenomics assay. *Environ. Sci. Nano* **7**, 1348–1364 (2020).

80. Helland, A., Wick, P., Koehler, A., Schmid, K. & Som, C. Reviewing the Environmental and Human Health Knowledge Base of Carbon Nanotubes. *Environ. Health Perspect.* **115**, 1125 (2007).

81. Alberti, P., Ren, J., Teulade-Fichou, M. P., Guittat, L., Riou, J. F., Chaires, J. B., Hélène, C., Vigneron, J. P., Lehn, J. M. & Mergny, J. L. Interaction of an acridine dimer with dna quadruplex structures. *J. Biomol. Struct. Dyn.* **19**, 505–513 (2001).

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

83. Ma, D. L., Kwan, M. H. T., Chan, D. S. H., Lee, P., Yang, H., Ma, V. P. Y., Bai, L. P., Jiang, Z. H. & Leung, C. H. Crystal violet as a fluorescent switch-on probe for i-motif: Label-free DNA-based logic gate. *Analyst* **136**, 2692–2696 (2011).

84. Satpathi, S., Sappati, S., Das, K. & Hazra, P. Structural characteristics requisite for the ligand-based selective detection of i-motif DNA. *Org. Biomol. Chem.* **17**, 5392–5399 (2019).

85. Sheng, Q., Neaverson, J. C., Mahmoud, T., Stevenson, C. E. M., Matthews, S. E. & Waller, Z. A. E. Identification of new DNA i-motif binding ligands through a fluorescent intercalator displacement assay. *Org. Biomol. Chem.* **15**, 5669–5673 (2017).

86. Kendrick, S., Kang, H. J., Alam, M. P., Madathil, M. M., Agrawal, P., Gokhale, V., Yang, D., Hecht, S. M. & Hurley, L. H. The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J. Am. Chem. Soc.* (2014) doi:10.1021/ja410934b.

87. Kuang, G., Zhang, M., Kang, S., Hu, D., Li, X., Wei, Z., Gong, X., An, L.-K., Huang, Z.-S. S., Shu, B. & Li, D. Syntheses and Evaluation of New Bisacridine Derivatives for Dual Binding of G-Quadruplex and i-Motif in Regulating Oncogene c-myc Expression. *J. Med. Chem.* **63**, 9136–9153 (2020).

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

89. Guéron, M. & Leroy, J.-L. The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.* **10**, 326–331 (2000).

90. Mir, B., Serrano, I., Buitrago, D., Orozco, M., Escaja, N. & Gonzaíez, C. Prevalent Sequences in the Human Genome Can Form Mini i-Motif Structures at Physiological pH. *Am. Chem. Soc.* **139**, 13985–13988 (2017).

91. Tsvetkov, V. B., Zatsepin, T. S., Belyaev, E. S., Kostyukevich, Y. I., Shpakovski, G. V, Podgorsky, V. V, Pozmogova, G. E., Varizhuk, A. M. & Aralov, A. V. i-Clamp phenoxazine for the fine tuning of DNA i-motif stability. *Nucleic Acids Res.* **46**, 2751–2764 (2018).

92. Saxena, S., Joshi, S., Shankaraswamy, J., Tyagi, S. & Kukreti, S. Magnesium and molecular crowding of the cosolutes stabilize the i-motif structure at physiological pH. *Biopolymers* **107**, (2017).

93. Dzatko, S., Krafcikova, M., Hänsel-Hertsch, R., Fessl, T., Fiala, R., Loja, T., Krafcik, D., Mergny, J.-L., Foldynova-Trantirkova, S. & Trantirek, L. Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chemie Int. Ed.* **57**, 2165–2169 (2018).

94. Zeraati, M., Langley, D. B., Schofield, P., Moye, A. L., Rouet, R., Hughes, W. E., Bryan, T. M., Dinger, M. E. & Christ, D. I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **10**, 631–637 (2018).

95. Schafer, K, A. The Cell Cycle: A Review. *Vet. Pathol.* **35**, 461–478 (1998).

96. Cui, Y., Kong, D., Ghimire, C., Xu, C. & Mao, H. Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **55**, 2291–2299 (2016).

97. King, J. J., Irving, K. L., Evans, C. W., Chikhale, R. V., Becker, R., Morris, C. J., Peña Martinez, C. D., Schofield, P., Christ, D., Hurley, L. H., Waller, Z. A. E., Iyer, K. S. & Smith, N. M. DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *J. Am. Chem. Soc.* **142**, 20600–20604 (2020).

98. Ankur, P., Grishma, J., Yunke+, F., Sherwin, S. & Angelo, M. Telomere elongation and Telomerase activity in Normal and Cancer cell lines: HEK-293, HeLa and A549. doi:10.1101/100446.

99. O'Sullivan, R. J. & Karlseder, J. Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **11**, 171–81 (2010).

100. Xu, Y., Ishizuka, T., Kurabayashi, K. & Komiyama, M. Consecutive Formation of G-Quadruplexes in Human Telomeric-Overhang DNA: A Protective Capping Structure for Telomere Ends. *Angew. Chemie Int. Ed.* **48**, 7833–7836 (2009).

101. Cimino-Reale, G., Pascale, E., Battiloro, E., Starace, G., Verna, R. & D'Ambrosio, E. The length of telomeric G-rich strand 3'-overhang measured by oligonucleotide ligation assay. *Nucleic Acids Res.* **29**, E35 (2001).

102. Shay, J. W. & Wright, W. E. Role of telomeres and telomerase in cancer. *Semin. Cancer Biol.* **21**, 349–53 (2011).

103. Neidle, S. Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **277**, 1118–1125 (2010).

104. Yoga, Y. M. K., Traore, D. A. K., Sidiqi, M., Szeto, C., Pendini, N. R., Barker, A., Leedman, P. J., Wilce, J. A. & Wilce, M. C. J. Contribution of the first K-homology domain of poly(C)-binding protein 1 to its affinity and specificity for C-rich oligonucleotides. *Nucleic Acids Res.* **40**, 5101–5114 (2012).

105. Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I. & Levens, D.

Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell. Biol.* **16**, 2350–2360 (1996).

106. Lacroix, L., Liénard, H., Labourier, E., Djavaheri-Mergny, M., Lacoste, J., Leffers, H., Tazi, J., Hélène, C. & Mergnya, J.-L. Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro. *Nucleic Acids Res.* **28**, 1564–1575 (2000).

107. Kang, H.-J., Kendrick, S., Hecht, S. M. & Hurley, L. H. The Transcriptional Complex Between the BCL2 i-Motif and hnRNP LL Is a Molecular Switch for Control of Gene Expression That Can Be Modulated by Small Molecules. *J. Am. Chem. Soc.* **136**, 4172–4185 (2014).

108. Ayoubi, T. A. & Van De Ven, W. J. Regulation of gene expression by alternative promoters. *FASEB J.* **10**, 453–60 (1996).

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

110. Frenzel, A., Grespi, F., Chmelewskij, W. & Villunger, A. Bcl2 family proteins in carcinogenesis and the treatment of cancer. *Apoptosis* **14**, 584–96 (2009).

111. Eischen, C. M., Packham, G., Nip, J., Fee, B. E., Hiebert, S. W., Zambetti, G. P. & Cleveland, J. L. Bcl-2 is an apoptotic target suppressed by both c-Myc and E2F-1. *Oncogene* **20**, 6983–6993 (2001).

112. Vaux, D. L., Cory, S. & Adams, J. M. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. *Nature* **335**, 440–442 (1988).

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

114. Dhakal, S., Yu, Z., Konik, R., Cui, Y., Koirala, D. & Mao, H. G-quadruplex and i-motif are mutually exclusive in ILPR double-stranded DNA. *Biophys. J.* **102**, 2575–84 (2012).

115. Simonsson, T., Pribylova, M. & Vorlickova, M. A Nuclease Hypersensitive Element in the Human c-myc Promoter Adopts

Several Distinct i-Tetraplex Structures. *Biochem. Biophys. Res. Commun.* **278**, 158–166 (2000).

116. Carmeliet, P. VEGF as a Key Mediator of Angiogenesis in Cancer. *Oncology* **69**, 4–10 (2005).

117. Bucek, P., Jaumot, J., Aviñó, A., Eritja, R. & Gargallo, R. pH-Modulated Watson-Crick Duple-Quadruplex Equilibria of Guanine-Rich and Cytosine-Rich DNA Sequences 140 Base Pairs Upstream of the c-kit Transcription Initiation Site. *Chem. - A Eur. J.* **15**, 12663–12671 (2009).

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

119. K, G., A, P., K, B.-R., V, G., D, S. & LH, H. Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J. Am. Chem. Soc.* **129**, 10220–10228 (2007).

120. Kaiser, C. E., Van Ert, N. A., Agrawal, P., Chawla, R., Yang, D. & Hurley, L. H. Insight into the Complexity of the i-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J. Am. Chem. Soc.* **139**, 8522–8536 (2017).

121. Miller, D. M., Thomas, S. D., Islam, A., Muench, D. & Sedoris, K. c-Myc and cancer metabolism. *Clin. Cancer Res.* **18**, 5546–53 (2012).

122. Zeller, K. I., Jegga, A. G., Aronow, B. J., O'Donnell, K. A. & Dang, C. V. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol.* **4**, R69 (2003).

123. DePinho, R., Mitsock, L., Hatton, K., Ferrier, P., Zimmerman, K., Legouy, E., Tesfaye, A., Collum, R., Yancopoulos, G. & P Nisen. Myc family of cellular oncogenes. *J. Cell. Biochem.* **33**, 257–266 (1987).

124. Elbadawy, M., Usui, T., Yamawaki, H. & Sasaki, K. Emerging Roles of C-Myc in Cancer Stem Cell-Related Signaling and Resistance to Cancer Chemotherapy: A Potential Therapeutic Target Against Colorectal Cancer. *Int. J. Mol. Sci.* **20**, 2340 (2019).

125. Conzen, S. D., Gottlob, K., Kandel, E. S., Khanduri, P., Wagner, A. J., O'Leary, M. & Hay, N. Induction of Cell Cycle Progression and

Acceleration of Apoptosis Are Two Separable Functions of c-Myc: Transrepression Correlates with Acceleration of Apoptosis. *Mol. Cell. Biol.* **20**, 6008–6018 (2000).

126. Herbst, A., Hemann, M. T., Tworkowski, K. A., Salghetti, S. E., Lowe, S. W. & Tansey, W. P. A conserved element in Myc that negatively regulates its proapoptotic activity. *EMBO Rep.* **6**, 177–183 (2005).

127. Cowling, V. H., Chandriani, S., Whitfield, M. L. & Cole, M. D. A Conserved Myc Protein Domain, MBIV, Regulates DNA Binding, Apoptosis, Transformation, and G2 Arrest. *Mol. Cell. Biol.* **26**, 4226–4239 (2006).

128. Blackwood, E. M. & Eisenman, R. N. Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science (80-.).* **251**, 1211–1217 (1991).

129. Blackwood, E. M., Liischer, B. & Eisenman, R. N. Myc and Max associate in vivo. *Genes Dev.* **6**, 71–80 (1992).

130. Gallant, P. & Steiger, D. Myc's secret life without Max. *Cell Cycle* **8**, 3848–3853 (2009).

131. Oster, S. K., Ho, C. S. W., Soucie, E. L. & Penn, L. Z. The myc Oncogene: MarvelouslY Complex. *Adv. Cancer Res.* **84**, 81–154 (2002).

132. Hölzel, M., Kohlhuber, F., Schlosser, I., Hölzel, D., Lüscher, B. & Eick, D. Myc/Max/Mad regulate the frequency but not the duration of productive cell cycles. *EMBO Rep.* **2**, 1132 (2001).

133. Mateyak MK1, Obaya AJ, Adachi S & Sedivy JM. Phenotypes of c-Myc-deficient rat fibroblasts isolated by targeted homologous recombination. *Cell Growth Differ.* **8**, 1039–1048 (1997).

134. Shichiri, M., Hanson, K. D. & Sedivy, J. M. Effects of c-myc expression on proliferation, quiescence, and the G0 to G1 transition in nontransformed cells. *Cell Growth Differ.* **4**, 93–104 (1993).

135. Chang DH, Angelin-Duclos C & Calame K. BLIMP-1: trigger for differentiation of myeloid lineage - PubMed - NCBI. *Nat. Immunol.* **1**, 169–176 (2000).

136. Smith, K. N., Singh, A. M. & Dalton, S. Myc represses primitive

endoderm differentiation in pluripotent stem cells. *Cell Stem Cell* **7**, 343–354 (2010).

137. Lin, C. H., Jackson, A. L., Guo, J., Linsley, P. S. & Eisenman, R. N. Myc-regulated microRNAs attenuate embryonic stem cell differentiation. *EMBO J.* **28**, 3157–3170 (2009).

138. Iritani, B. M. & Eisenman, R. N. c-Myc enhances protein synthesis and cell size during B lymphocyte development. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13180–13185 (1999).

139. Schuhmacher, M., Staege, M. S., Pajic, A., Polack, A., Weidle, U. H., Bornkamm, G. W., Eick, D. & Kohlhuber, F. Control of cell growth by c-Myc in the absence of cell division. *Curr. Biol.* **9**, 1255–1258 (1999).

140. Beier, R., Bürgin, A., Kiermaier, A., Fero, M., Karsunky, H., Saffrich, R., Möröy, T., Ansorge, W., Roberts, J. & Eilers, M. Induction of cyclin E–cdk2 kinase activity, E2F-dependent transcription and cell growth by Myc are genetically separable events. *EMBO J.* **19**, 5813–5823 (2000).

141. Melnik, S., Werth, N., Boeuf, S., Hahn, E. M., Gotterbarm, T., Anton, M. & Richter, W. Impact of c-MYC expression on proliferation, differentiation, and risk of neoplastic transformation of human mesenchymal stromal cells. *Stem Cell Res. Ther.* **10**, (2019).

142. Mai, S. Overexpression of c-myc precedes amplification of the gene encoding dihydrofolate reductase. *Gene* **148**, 253–260 (1994).

143. Rockwood, L. D., Torrey, T. A., Kim, J. S., Coleman, A. E., Kovalchuk, A. L., Xiang, S., Ried, T., Morse, H. C. & Janz, S. Genomic instability in mouse Burkitt lymphoma is dominated by illegitimate genetic recombinations, not point mutations. *Oncogene* **21**, 7235–7240 (2002).

144. Askew, D. S., Ashmun, R. A., Simmons, B. C. & Cleveland, J. L. Constitutive c-myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis. *Oncogene* **6**, 1915–1922 (1991).

145. Askew DS, Ihle JN & Cleveland JL. Activation of apoptosis associated with enforced myc expression in myeloid progenitor

cells is dominant to the suppression of apoptosis by interleukin-3. *Blood* **82**, 2079–2087 (1993).

146. Hoffman, B. & Liebermann, D. A. Apoptotic signaling by c-MYC. *Oncogene* **27**, 6462–6472 (2008).

147. McMahon, S. B. MYC and the control of apoptosis. *Cold Spring Harb. Perspect. Med.* **4**, a014407 (2014).

148. Baudino, T. A., McKay, C., Pendeville-Samain, H., Nilsson, J. A., Maclean, K. H., White, E. L., Davis, A. C., Ihle, J. N. & Cleveland, J. L. c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression. *Genes Dev.* **16**, 2530–2543 (2002).

149. Bretones, G., Delgado, M. D. & León, J. Myc and cell cycle control. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1849**, 506–516 (2015).

150. Bouchard, C., Staller, P. & Eilers, M. Control of cell proliferation by Myc. *Trends Cell Biol.* **8**, 202–206 (1998).

151. Kuzyk, A. & Mai, S. c-MYC-induced genomic instability. *Cold Spring Harb. Perspect. Med.* **4**, (2014).

152. Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nat. Rev. Cancer* **8**, 976–990 (2008).

153. Dang, C. V. MYC on the path to cancer. *Cell* **149**, 22–35 (2012).

154. Stine, Z. E., Walton, Z. E., Altman, B. J., Hsieh, A. L. & Dang, C. V. MYC, Metabolism, and Cancer. *Cancer Discov.* **5**, 1024–1039 (2015).

155. Camarda, R., Williams, J. & Goga, A. In vivo Reprogramming of Cancer Metabolism by MYC. *Front. cell Dev. Biol.* **5**, 35 (2017).

156. Goetzman, E. S. & Prochownik, E. V. The role for myc in coordinating glycolysis, oxidative phosphorylation, glutaminolysis, and fatty acid metabolism in normal and neoplastic tissues. *Front. Endocrinol. (Lausanne).* **9**, (2018).

157. Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**, 353–364 (1996).

158. Kerbel, R. S., Viloria-Petit, A., Okada, F. & Rak, J. Establishing a link between oncogenes and tumor angiogenesis. *Mol. Med.* **4**, 286

(1998).

159. Tikhonenko, A. T., Black, D. J. & Linial, M. L. Viral Myc oncoproteins in infected fibroblasts down-modulate thrombospondin-1, a possible tumor suppressor gene. *J. Biol. Chem.* **271**, 30741–30747 (1996).

160. Gravina, G., Festuccia, C., Popov, V., Di Rocco, A., Colapietro, A., Sanità, P., Monache, S., Musio, D., De Felice, F., Di Cesare, E., Tombolini, V. & Marampon, F. c-Myc Sustains Transformed Phenotype and Promotes Radioresistance of Embryonal Rhabdomyosarcoma Cell Lines. *Radiat. Res.* **185**, 411–422 (2016).

161. Wang, W. J., Wu, S. P., Liu, J. Bin, Shi, Y. S., Huang, X., Zhang, Q. B. & Yao, K. T. MYC regulation of CHK1 and CHK2 promotes radioresistance in a stem cell-like population of nasopharyngeal carcinoma cells. *Cancer Res.* **73**, 1219–1231 (2013).

162. Soucek, L., Whitfield, J., Martins, C. P., Finch, A. J., Murphy, D. J., Sodir, N. M., Karnezis, A. N., Swigart, L. B., Nasi, S. & Evan, G. I. Modelling Myc inhibition as a cancer therapy. *Nature* **455**, 679–683 (2008).

163. Levens, D. You Don't Muck with MYC. *Genes Cancer* **1**, 547–554 (2010).

164. Michelotti, G. A., Michelotti, E. F., Pullner, A., Duncan, R. C., Eick, D. & Levens, D. Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Mol. Cell. Biol.* **16**, 2656–2669 (1996).

165. Kouzine, F., Sanford, S., Elisha-Feil, Z. & Levens, D. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol.* **15**, 146–154 (2008).

166. Braddock, D. T., Baber, J. L., Levens, D. & Clore, G. M. Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: Solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.* **21**, 3476–3485 (2002).

167. Crichlow, G. V., Zhou, H., Hsiao, H. H., Frederick, K. B., Debrosse, M., Yang, Y., Folta-Stogniew, E. J., Chung, H. J., Fan, C., De La Cruz, E. M., Levens, D., Lolis, E. & Braddock, D. Dimerization of FIR upon FUSE DNA binding suggests a mechanism of c-myc inhibition. *EMBO J.* **27**, 277–289 (2008).

168. Levens, D. How the c-myc Promoter Works and Why It Sometimes Does Not. *J. Natl. Cancer Inst. Monogr.* **39**, 43 (2008).

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

170. Yang, D. & Hurley, L. Structure of the Biologically Relevant G-quadruplex in the c-MYC Promoter. *Nucleosides, Nucleotides and Nucleic Acids* **25**, 951–968 (2006).

171. 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).

172. Grand, C. L., Han, H., Muñoz, R. M., Weitman, S., Von Hoff, D. D., Laurence, H. H. & Bearss, D. J. The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo 1 This research was supported by grants from the NIH and the Arizona Disease Control Research Commission. *Mol. Cancer Ther.* **1**, 565–573 (2002).

173. Postel, E. H., Berberich, S. J., Flint, S. J. & Ferrone, C. A. Human c-myc transcription factor PuF identified as nm23-H2 nucleoside diphosphate kinase, a candidate suppressor of tumor metastasis. *Science (80-.).* **261**, 478–480 (1993).

174. Dexheimer, T. S., Carey, S. S., Zuohe, S., Gokhale, V. M., Hu, X., Murata, L. B., Maes, E. M., Weichsel, A., Sun, D., Meuillet, E. J., Montfort, W. R. & Hurley, L. H. NM23-H2 may play an indirect role in transcriptional activation of c-myc gene expression but does not cleave the nuclease hypersensitive element III 1. *Mol. Cancer Ther.* **8**, 1363–1377 (2009).

175. Thakur, R. K., Kumar, P., Halder, K., Verma, A., Kar, A., Parent, J. L., Basundra, R., Kumar, A. & Chowdhury, S. Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.* **37**, 172–183 (2009).

176. Deng, N., Wickstrom, L., Cieplak, P., Lin, C. & Yang, D. Resolving the Ligand-Binding Specificity in c-MYC G-Quadruplex DNA: Absolute Binding Free Energy Calculations and SPR Experiment. *J. Phys. Chem. B* **121**, 10484–10497 (2017).

177. Calabrese, D. R., Chen, X., Leon, E. C., Gaikwad, S. M., Phyo, Z., Hewitt, W. M., Alden, S., Hilimire, T. A., He, F., Michalowski, A. M., Simmons, J. K., Saunders, L. B., Zhang, S., Connors, D., Walters, K. J., Mock, B. A. & Schneekloth, J. S. Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex. *Nat. Commun.* **9**, 1–15 (2018).

178. Mathur, V., Verma, A., Maiti, S. & Chowdhury, S. Thermodynamics of i-tetraplex formation in the nuclease hypersensitive element of human c-myc promoter. *Biochem. Biophys. Res. Commun.* **320**, 1220–1227 (2004).

179. Asamitsu, S., Obata, S., Yu, Z., Bando, T. & Sugiyama, H. Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy. *Molecules* **24**, (2019).

180. Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–7 (1985).

181. Parmley, S. F. & Smith, G. P. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **73**, 305–318 (1988).

182. Omidfar, K. & Daneshpour, M. Advances in phage display technology for drug discovery. *Expert Opin. Drug Discov.* **10**, 651–669 (2015).

183. Takakusagi, Y., Kuramochi, K., Takagi, M., Kusayanagi, T., Manita, D., Ozawa, H., Iwakiri, K., Takakusagi, K., Miyano, Y., Nakazaki, A., Kobayashi, S., Sugawara, F. & Sakaguchi, K. Efficient one-cycle affinity selection of binding proteins or peptides specific for a small-molecule using a T7 phage display pool. *Bioorg. Med. Chem.* **16**, 9837–9846 (2008).

184. Sternberg, N. & Hoess, R. H. Display of peptides and proteins on the surface of bacteriophage lambda. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1609–13 (1995).

185. Efimov, V. P., Nepluev, I. V & Mesyanzhinov, V. V. Bacteriophage T4 as a surface display vector. *Virus Genes* **10**, 173–7 (1995).

186. Endemann, H. & Model, P. Location of Filamentous Phage Minor Coat Proteins in Phage and in Infected Cells. *J. Mol. Biol* **250**, 496–506 (1995).

187. Henry, T. J. & Pratt, D. THE PROTEINS OF BACTERIOPHAGE M13. (1968).

188. Marvin, D. A. & Wachtel, E. J. Structure and assembly of filamentous bacterial viruses. *Nature* **253**, 19–23 (1975).

189. Marvin, D. A., Symmons, M. F. & Straus, S. K. Structure and assembly of filamentous bacteriophages. *Prog. Biophys. Mol. Biol.* **114**, 80–122 (2014).

190. Marvin, D. A., Welsh, L. C., Symmons, M. F., Scott, W. R. P. & Straus, S. K. Molecular Structure of fd (f1, M13) Filamentous Bacteriophage Refined with Respect to X-ray Fibre Diffraction and Solid-state NMR Data Supports Specific Models of Phage Assembly at the Bacterial Membrane. *J. Mol. Biol.* **355**, 294–309 (2006).

191. Karimi, M., Mirshekari, H., Moosavi Basri, S. M., Bahrami, S., Moghoofei, M. & Hamblin, M. R. Bacteriophages and phage-inspired nanocarriers for targeted delivery of therapeutic cargos. *Adv. Drug Deliv. Rev.* **106**, 45–62 (2016).

192. Bennett, N. J., Gagic, D., Sutherland-Smith, A. J. & Rakonjac, J. Characterization of a Dual-Function Domain That Mediates Membrane Insertion and Excision of Ff Filamentous Bacteriophage. *J. Mol. Biol.* **411**, 972–985 (2011).

193. Riechmann, L. & Holliger, P. The C-Terminal Domain of TolA Is the Coreceptor for Filamentous Phage Infection of E. coli. *Cell* **90**, 351–360 (1997).

194. Click, E. M. & Webster, R. E. The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage f1 during infection. *J. Bacteriol.* **180**, 1723–8 (1998).

195. Sidhu, S. S. Phage display in pharmaceutical biotechnology. *Curr. Opin. Biotechnol.* **11**, 610–6 (2000).

196. Lindner, T., Kolmar, H., Haberkorn, U. & Mier, W. DNA Libraries for the Construction of Phage Libraries: Statistical and Structural Requirements and Synthetic Methods. *Molecules* **16**, 1625–1641 (2011).

197. BioLabs, N. E. Ph.D.™ Phage Display Libraries: Instruction Manual Version 2.1 7/16. 1–41 https://www.neb.com/-

/media/nebus/files/manuals/manuale8102.pdf?rev=a207dd0c588
9476d8b41aec9b3029d49&hash=1A7896F17EB4E7A546C3E7D32
CBACFC3 (2017).

198. Bazan, J., Całkosiński, I. & Gamian, A. Phage display—A powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. *Hum. Vaccin. Immunother.* **8**, 1817 (2012).

199. Yang, W., Yoon, A., Lee, S., Kim, S., Han, J. & Chung, J. Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. *Exp. Mol. Med.* **49**, e308–e308 (2017).

200. Patel, C. A., Wang, J., Wang, X., Dong, F., Zhong, P., Luo, P. P. & Wang, K. C. Parallel selection of antibody libraries on phage and yeast surfaces via a cross-species display. *Protein Eng. Des. Sel.* **24**, 711–719 (2011).

201. Pande, J., Szewczyk, M. M. & Grover, A. K. Phage display: Concept, innovations, applications and future. *Biotechnol. Adv.* **28**, 849–858 (2010).

202. Paschke, M. Phage display systems and their applications. *Appl. Microbiol. Biotechnol.* **70**, 2–11 (2006).

203. Maranhão, A. Q., Brígido, M. M. M. & Brígido, C. M. M. Expression of anti-Z-DNA single chain antibody variable fragment on the filamentous phage surface. **33**, 569–579 (2000).

204. Isalan, M., Patel, D. S., Balasubramanian, S. & Choo, Y. Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. *Biochemistry* **40**, 830–6 (2001).

205. Wölcke, J. & Weinhold, E. A DNA-BINDING PEPTIDE FROM A PHAGE DISPLAY LIBRARY. *Nucleosides, Nucleotides and Nucleic Acids* **20**, 1239–1241 (2001).

206. Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C. & Rani, V. DNA–protein interactions: methods for detection and analysis. *Mol. Cell. Biochem.* **365**, 279–299 (2012).

207. Sherman, E. M., Holmes, S. & Ye, J.-D. Specific RNA-Binding Antibodies with a Four-Amino-Acid Code. *J. Mol. Biol.* **426**, 2145–2157 (2014).

208. Laird-Offringa, I. A. In Vitro Genetic Analysis of RNA-Binding Proteins Using Phage Display. in *RNA-Protein Interaction Protocols* 189–216 (Humana Press, 1999). doi:10.1385/1-59259-676-2:189.

209. Fernando, H., Rodriguez, R. & Balasubramanian, S. Selective Recognition of a DNA G-Quadruplex by an Engineered Antibody [†]. *Biochemistry* **47**, 9365–9371 (2008).

210. Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**, 39–46 (2001).

211. Choo, Y. & Klug, A. Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.* **7**, 117–125 (1997).

212. Ladame, S., Schouten, J. A., Roldan, J., Redman, J. E., Neidle, S. & Balasubramanian, S. Exploring the recognition of quadruplex DNA by an engineered Cys2-His2 zinc finger protein. *Biochemistry* **45**, 1393–9 (2006).

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

214. Wicks, S. L. & Hargrove, A. E. Fluorescent Indicator Displacement Assays to Identify and Characterize Small Molecule Interactions with RNA. *Methods* **167**, 3 (2019).

215. Permyakov, E. A. The Use of UV–Vis Absorption Spectroscopy for Studies of Natively Disordered Proteins. *Methods Mol. Biol.* **895**, 421–433 (2012).

216. How to determine protein concentration | Abcam. https://www.abcam.com/help/how-do-i-determine-protein-concentration.

217. Anthis, N. J. & Clore, G. M. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci.* **22**, 851–8 (2013).

218. Rosu, F., Gabelica, V., De Pauw, E., Antoine, R., Broyer, M. & Dugourd, P. UV spectroscopy of DNA duplex and quadruplex structures in the gas phase. *J. Phys. Chem. A* **116**, 5383–5391 (2012).

219. Ranjbar, B. & Gill, P. Circular Dichroism Techniques: Biomolecular

and Nanostructural Analyses- A Review. *Chem. Biol. Drug Des.* **74**, 101–120 (2009).

220. Hammes, G. G. *Spectroscopy for the biological sciences*. (Wiley-Interscience, 2005).

221. BEROVA, N., NAKANISHI, K. & WOODY, R. W. *Circular Dichroism: Principles and Applications, 2nd Edition*. (USA: Plenum Publishing Corp, 200AD).

222. Fasman, G. D. *Circular Dichroism and the Conformatiobal Analysis of Biomolecules*. (Springer US, 1996). doi:10.1007/978-1-4757-2508-7.

223. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–90 (2006).

224. 3.6: Optical Activity - Chemistry LibreTexts. https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Book%3A_Organic_Chemistry_with_a_Biological_Emphasis_v2.0_(Soderberg)/03%3A_Conformations_and_Stereochemistry/3.06%3A_Optical_Activity.

225. Sreerama, N. & Woody, R. W. Computation and Analysis of Protein Circular Dichroism Spectra. *Methods Enzymol.* **383**, 318–351 (2004).

226. P, M. & WC, J. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **167**, 76–85 (1987).

227. Miles, A. J., Ramalli, S. G. & Wallace, B. A. DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. *Protein Sci.* (2021) doi:10.1002/PRO.4153.

228. L, W. & BA, W. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392–400 (2008).

229. L, W. & BA, W. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, (2004).

230. Nucleic acids and chromatin: 3.1 The helical structure of DNA - OpenLearn - Open University - S377_1. https://www.open.edu/openlearn/science-maths-

technology/science/biology/nucleic-acids-and-
chromatin/content-section-3.1.

231. Kypr, J., Kejnovská, I., Renčiuk, D. & Vorlíčková, M. Circular
dichroism and conformational polymorphism of DNA. *Nuclecic
Acids Res.* **37**, 1713–1725 (2009).

232. Maizels, N. & Gray, L. T. The G4 Genome. *PLoS Genet.* (2013)
doi:10.1371/journal.pgen.1003468.

233. Surface plasmon resonance | Cytiva.
https://www.cytivalifesciences.com/en/us/solutions/protein-
research/knowledge-center/surface-plasmon-resonance/surface-
plasmon-resonance.

234. Santos, T., Salgado, G. F., Cabrita, E. J. & Cruz, C. G-Quadruplexes
and Their Ligands: Biophysical Methods to Unravel G-
Quadruplex/Ligand Interactions. *Pharm. 2021, Vol. 14, Page 769*
14, 769 (2021).

235. Wright, E. P., Day, H. A., Ibrahim, A. M., Kumar, J., Boswell, L. J. E.
E., Huguin, C., Stevenson, C. E. M. M., Pors, K. & Waller, Z. A. E. E.
Mitoxantrone and Analogues Bind and Stabilize i-Motif Forming
DNA Sequences. *Sci. Rep.* **6**, 39456 (2016).

236. E, P., L, B., H, B., T, L., A, V. der H., G, P., J, D. & E, D. Influence of
the SPR Experimental Conditions on the G-Quadruplex DNA
Recognition by Porphyrin Derivatives. *Langmuir* **34**, 13057–13064
(2018).

237. Morcos, E. F., Kussrow, A., Enders, C. & Bornhop, D. Free-Solution
Interaction Assay of Carbonic Anhydrase to its Inhibitors Using
Back-scattering Interferometry. *Electrophoresis* **31**, 3691 (2010).

238. Mosmann, T. Rapid colorimetric assay for cellular growth and
survival: Application to proliferation and cytotoxicity assays. *J.
Immunol. Methods* **65**, 55–63 (1983).

239. Denizot, F. & Lang, R. Rapid colorimetric assay for cell growth and
survival: Modifications to the tetrazolium dye procedure giving
improved sensitivity and reliability. *J. Immunol. Methods* **89**, 271–
277 (1986).

240. Cole, S. P. C. Rapid chemosensitivity testing of human lung tumor
cells using the MTT assay. *Cancer Chemother. Pharmacol. 1986
173* **17**, 259–263 (1986).

241. Gerlier, D. & Thomasset, N. Use of MTT colorimetric assay to measure cell activation. *J. Immunol. Methods* **94**, 57–63 (1986).

242. Tonder, A. van, Joubert, A. M. & Cromarty, A. D. Limitations of the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) assay when compared to three commonly used cell enumeration assays. *BMC Res. Notes* **8**, (2015).

243. Głuszyńska, A., Juskowiak, B., Kuta-Siejkowska, M., Hoffmann, M. & Haider, S. Carbazole ligands as c-myc G-quadruplex binders. *Int J Biol Macromol .* **15**, 479–490 (2018).

244. Prasher, P. & Sharma, M. Medicinal chemistry of acridine and its analogues. *Medchemcomm.* **9**, 1589–1618 (2018).

245. Hamilton, P. L. & Arya, D. P. Natural product DNA major groove binders. *Nat Prod Rep* **29**, 134–43 (2012).

246. D, M., C, A. & MP, T.-F. Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. *Bioorg. Med. Chem. Lett.* **16**, 4842–4845 (2006).

247. Yeung, B. K. S., Tse, W. C. & Boger, D. L. Determination of binding affinities of triplex forming oligonucleotides using a fluorescent intercalator displacement (FID) assay. *Bioorg. Med. Chem. Lett.* **13**, 3801–3804 (2003).

248. DL, B. & WC, T. Thiazole orange as the fluorescent intercalator in a high resolution fid assay for determining DNA binding affinity and sequence selectivity of small molecules. *Bioorg. Med. Chem.* **9**, 2511–2518 (2001).

249. AR, M., JS, L., DE, P., NL, M. & DH, E. Review: ethidium fluorescence assays. Part 1. Physicochemical studies. *Nucleic Acids Res.* **7**, 547–565 (1979).

250. Dai, J., Hatzakis, E., Hurley, L. H. & Yang, D. I-Motif Structures Formed in the Human c-MYC Promoter Are Highly Dynamic–Insights into Sequence Redundancy and I-Motif Stability. *PLoS One* **5**, e11647 (2010).

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

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

253. Lerman, L. S. Structural considerations in the interaction of DNA and acridines. *J. Mol. Biol.* **3**, 18-IN14 (1961).

254. Soni, A., Khurana, P., Singh, T. & Jayaram, B. A DNA intercalation methodology for an efficient prediction of ligand binding pose and energetics. *Bioinformatics* **33**, 1488–1496 (2017).

255. Koch, H. P. & Czejka, M. J. Evidence for the Intercalation of Thalidomide into DNA: Clue to the Molecular Mechanism of Thalidomide Teratogenicity? *Zeitschrift fur Naturforsch. - Sect. C J. Biosci.* **41**, 1057–1061 (1986).

256. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **6180**, 364–6 (1988).

257. Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K. & Dickerson, R. E. X-ray structure of a DNA hairpin molecule. *Nature* **334**, 175–179 (1988).

258. Phan, A. T., Guéron, M. & Leroy, J. L. The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.* **299**, 123–144 (2000).

259. Singh, R. P., Blossey, R. & Cleri, F. Structure and mechanical characterization of DNA i-motif nanowires by molecular dynamics simulation. *Biophys. J.* **105**, 2820–31 (2013).

260. D, S. & W, G. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **344**, 410–414 (1990).

261. Day, H. A. Investigating the Effect of Small Molecule Ligands and Cations on i-Motif DNA. (University of East Anglia, 2015).

262. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. *J. Gen. Virol.* **36**, 59–72 (1977).

263. Thomas, P. & Smart, T. G. HEK293 cell line: A vehicle for the expression of recombinant proteins. *J. Pharmacol. Toxicol. Methods* **51**, 187–200 (2005).

264. E, M., IM, A., B, B., BL, V. & A, P. Major Achievements in the

Design of Quadruplex-Interactive Small Molecules. *Pharmaceuticals* **15**, 300 (2022).

265. Giulia Miglietta, Marco Russo, Renée C Duardo, G. C. G-quadruplex binders as cytostatic modulators of innate immune genes in cancer cells. *Nucleic Acids Res.* **49**, 6673–6686 (2021).

266. Fedoroff, O. Y., Rangan, A., Chemeris, V. V. & Hurley, L. H. Cationic porphyrins promote the formation of i-motif DNA and bind peripherally by a nonintercalative mechanism. *Biochemistry* **39**, 15083–90 (2000).

267. Riechert-Krause, F. & Weisz, K. Indoloquinolines as DNA binding ligands. *Heterocycl. Commun.* **19**, 145–166 (2013).

268. Brogden, A. L., Hopcroft, N. H., Searcey, M. & Cardin, C. J. Ligand bridging of the DNA Holliday junction: molecular recognition of a stacked-X four-way junction by a small molecule. *Angew Chem Int Ed Engl* **46**, 3850–4 (2007).

269. Rackham, B. D., Howell, L. A., Round, A. N. & Searcey, M. Non-covalent duplex to duplex crosslinking of DNA in solution revealed by single molecule force spectroscopy. *Org. Biomol. Chem.* **11**, 8340–8347 (2013).

270. Marín, M. J., Rackham, B. D., Round, A. N., Howell, L. A., Russell, D. A. & Searcey, M. A rapid screen for molecules that form duplex to duplex crosslinks in DNA. *Chem. Commun.* **49**, 9113–9115 (2013).

271. Gabay, M., Li, Y. & Felsher, D. W. MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harb. Perspect. Med.* **4**, (2014).

272. Chen, H., Liu, H. & Qing, G. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct. Target. Ther.* **3**, 5 (2018).

273. Cooper, G. M. The Nuclear Envelope and Traffic between the Nucleus and Cytoplasm. in *The Cell: A Molecular Approach.* (Sinauer Associates, 2000).

274. González, V. & Hurley, L. H. The C-Terminus of Nucleolin Promotes the Formation of the c-MYC G-Quadruplex and Inhibits c-MYC Promoter Activity. *Biochemistry* **49**, 9706–9714 (2010).

275. Marqus, S., Pirogova, E. & Piva, T. J. Evaluation of the use of

therapeutic peptides for cancer treatment. *J. Biomed. Sci. 2017 241* **24**, 1–15 (2017).

276. Brown, S. L. & Kendrick, S. The i-Motif as a Molecular Target: More Than a Complementary DNA Secondary Structure. *Pharm. 2021, Vol. 14, Page 96* **14**, 96 (2021).

277. MLA style: Advanced information. NobelPrize.org. Nobel Prize Outreach AB 2021. https://www.nobelprize.org/prizes/chemistry/2018/advanced-information/ (2021).

278. Chung, J., Lee, D.-Y., Yoo, S. Y., Chung, W.-J., Lee, D.-Y. & Yoo, S. Y. Chemical modulation of M13 bacteriophage and its functional opportunities for nanomedicine. *Int. J. Nanomedicine* **9**, 5825–5836 (2014).

279. E Delory, B. G., Infirmary, R. & King, E. J. A Sodium Carbonate-bicarbonate Buffer for Alkaline Phosphatases. **39**,.

280. Buffer tables. https://microscopy.berkeley.edu/Resources/instruction/buffers.html.

281. Tris-Buffered Saline (TBS) | Protocols Online. *Protocols Online* https://www.protocolsonline.com/recipes/buffers/tris-buffered-saline-tbs/ (2012).

282. Derda, R., Tang, S. K. Y., Li, S. C., Ng, S., Matochko, W. & Jafari, M. R. Diversity of Phage-Displayed Libraries of Peptides during Panning and Amplification. *Molecules.* **16**, 1776–1803 (2011).

283. Vallero, D. Air Pollutant Kinetics and Equilibrium. *Fundam. Air Pollut.* 437–473 (2014) doi:10.1016/B978-0-12-401733-7.00018-9.

284. Bielecka, P., Dembska, A. & Juskowiak, B. Monitoring of pH Using an i-Motif-Forming Sequence Containing a Fluorescent Cytosine Analogue, tC. *Molecules* **24**, (2019).

285. Which of the three ready-made libraries should I choose? | NEB. https://international.neb.com/faqs/2013/09/03/phage-library-choice.

286. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. Fast, scalable generation of high-quality protein

multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).

287. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, (2010).

288. PepCalc.com - Peptide calculator. https://pepcalc.com/.

289. Duggan, K. Clustal Omega. https://www.ebi.ac.uk/seqdb/confluence/display/THD/Clustal+Omega (2018).

290. Holzwarth, G. & Doty, P. The Ultraviolet Circular Dichroism of Polypeptides1. *J. Am. Chem. Soc.* **87**, 218–228 (2002).

291. None, V. Sy., IA, B., ZM, S., CS, W. & JT, Y. Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set. *Anal. Biochem.* **214**, 17–24 (1993).

292. Shoari, A., Rasaee, M. J., Kanavi, M. R. & Daraei, B. Functional mimetic peptide discovery isolated by phage display interacts selectively to fibronectin domain and inhibits gelatinase. *J. Cell. Biochem.* **120**, 19699–19711 (2019).

293. Liu, K. C., Röder, K., Mayer, C., Adhikari, S., Wales, D. J. & Balasubramanian, S. Affinity-Selected Bicyclic Peptide G-Quadruplex Ligands Mimic a Protein-like Binding Mechanism. *Am. Chem. Soc.* **142**, 8367–8373 (2020).

294. Martino, L., Virno, A., Pagano, B., Virgilio, A., Micco, S. Di, Galeone, A., Giancola, C., Bifulco, G., Mayol, L. & Randazzo, A. Structural and Thermodynamic Studies of the Interaction of Distamycin A with the Parallel Quadruplex Structure [d(TGGGGT)]4. *J. Am. Chem. Soc.* **129**, 16048–16056 (2007).

295. Bazzicalupi, C., Ferraroni, M., Bilia, A. R., Scheggi, F. & Gratteri, P. The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Res.* **41**, 632 (2013).

296. Davenport, A. P., Scully, C. C. G., Graaf, C. de, Brown, A. J. H. & Maguire, J. J. Advances in therapeutic peptides targeting G protein-coupled receptors. *Nat. Rev. Drug Discov. Vol.* **19**, 389–413 (2020).

297. Taetz, S., Baldes, C., Mürdter, T., Kleideiter, E., Piotrowska, K., Bock, U., Haltner-Ukomadu, E Mueller, J., Huwer, H., Schaefer, U., Klotz, U. & Lehr, C. Biopharmaceutical characterization of the telomerase inhibitor BRACO19. *Pharm Res.* **23**, 1031–7 (2006).

298. Carabet, L. A., Rennie, P. S. & Cherkasov, A. Therapeutic Inhibition of Myc in Cancer. Structural Bases and Computer-Aided Drug Discovery Approaches. *Int J Mol Sci.* **20**, 120 (2018).

299. FUKAZAWA, T., MAEDA, Y., MATSUOKA, J., YAMATSUJI, T., SHIGEMITSU, K., MORITA, I., FAIOLA, F., DURBIN, M. L., SOUCEK, L. & NAOMOTO, Y. Inhibition of Myc Effectively Targets KRAS Mutation-positive Lung Cancer Expressing High Levels of Myc. *Anticancer Res.* **30**, 4193–4200 (2010).

300. Boddupally, P. V. L., Hahn, S., Beman, C., De, B., Brooks, T. A., Gokhale, V. & Hurley, L. H. The anticancer activity and cellular repression of c-MYC by the G-quadruplex-stabilizing 11-piperazinyl quindoline is not dependent on direct targeting of the G-quadruplex in the c-MYC promoter. *J. Med. Chem.* **55**, 6076 (2012).

301. Lefèvre, F., Rémy, M.-H. & Masson, J.-M. Alanine-stretch scanning mutagenesis: a simple and efficient method to probe protein structure and function. *Nucleic Acids Res.* **25**, 447–448 (1997).

302. Bromberg, Y. & Rost, B. Comprehensive in silico mutagenesis highlights functionally important residues in proteins. *Bioinformatics* **24**, 207 (2008).

303. Brown, R. V *et al.* The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor β Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of. *J Am Chem Soc* **139**, 7456–7475 (2017).

304. Amato, J., Miglietta, G., Morigi, R., Iaccarino, N., Locatelli, A., Leoni, A., Novellino, E., Pagano, B., Capranico, G. & Randazzo, A. Monohydrazone BasedG-Quadruplex SelectiveLigands Induce DNA Damage and Genome Instability in Human Cancer Cells. *J. Med. Chem.* **63**, 3090 (2020).

305. Ross, S. A. & Burrows, C. J. Cytosine-specific chemical probing of DNA using bromide and monoperoxysulfate. *Nucleic Acids Res.* **24**,

5062–5063 (1996).

306. Wang, Y.-H., Yang, Q.-F., Lin, X., Chen, D., Wang, Z.-Y., Chen, B., Han, H.-Y., Chen, H.-D., Cai, K.-C., Li, Q., Yang, S., Tang, Y.-L. & Li, F. G4LDB 2.2: a database for discovering and studying G-quadruplex and i-Motif ligands. *Nucleic Acids Res.* (2021) doi:10.1093/NAR/GKAB952.

307. Li, Q., Xiang, J. F., Yang, Q. F., Sun, H. X., Guan, A. J. & Tang, Y. L. G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* **41**, (2013).

308. New England BioLabs. Ph.D.[TM]-12 Phage Display Peptide Library Kit. https://international.neb.com/products/e8110-phd-12-phage-display-peptide-library-kit#Product Information.

309. New England BioLabs. Ph.D.[TM]-C7C Phage Display Peptide Library. https://international.neb.com/products/e8121-phd-c7c-phage-display-peptide-library#Product Information.

310. Redman, J. E. Surface plasmon resonance for probing quadruplex foldingand interactions with proteins and small molecules. *Methods.* **43**, 302–12 (2007).

Appendix

A1. Circular Dichroism spectra of 10 µM c-MycG in a) 10 mM NaCaco 100 mM KCL pH 6.6. b) 10 mM NaCaco 10 mM KCL pH 6.6.
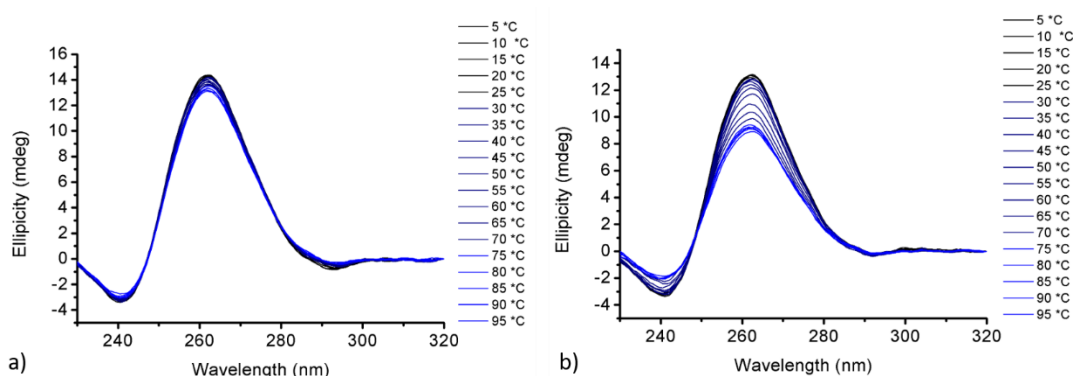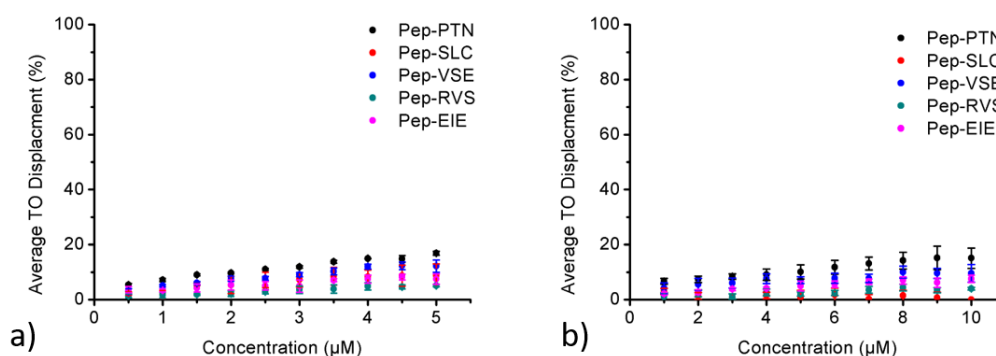
A2. Circular Dichroism spectra of normalised ellipiticity at 260 nm, the maximum positive characterisation parallel G-quadruplex peak for c-MycG. 10 µM c-MycG in 10 mM NaCaco 10 mM KCL pH 6.6. Line fitting is dose response.
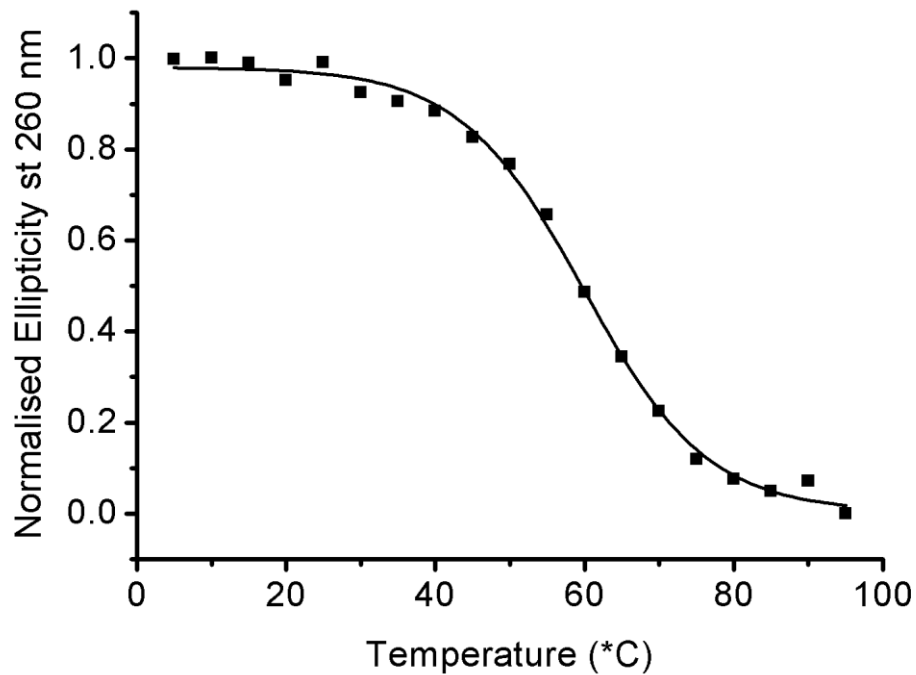
A3. Cell viability of HEK293 cells after 24 hr treatment with 1% DMSO (V:V), % cell viability was determined by (absorbance_{treated cells}/absorbance_{no treatment})*100.
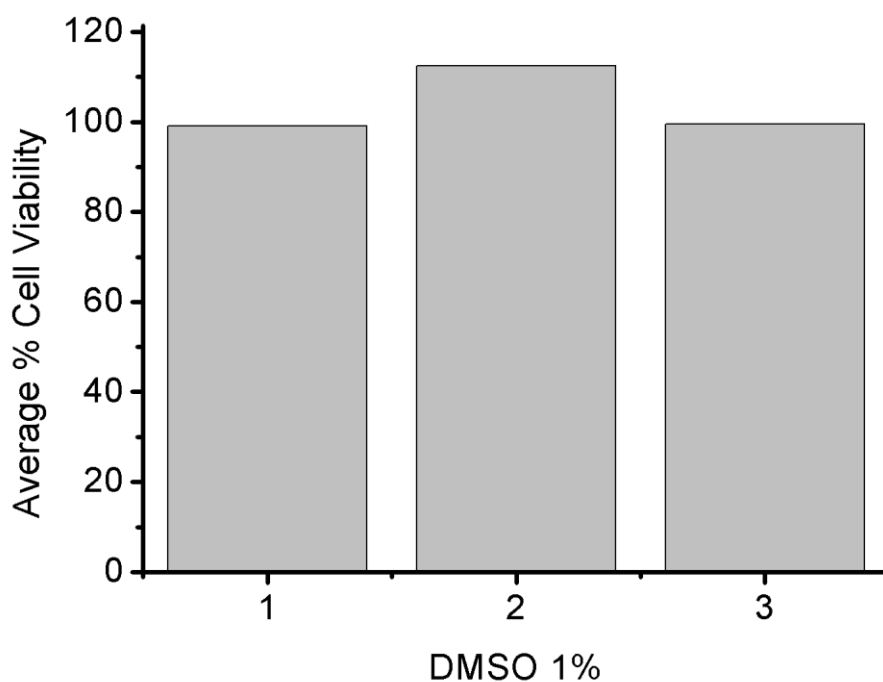
A4. Fluorescent indicator displacement assay in the following conditions: 1 μM:2 μM c-MycC52:To in 10 mM NaCaco pH 6.6. Five different peptides of interest were titrated every a) 0.5 μM from 0 μM to 5 μM and b) 1 μM from 0 μM to 10 μM. OriginPro 8 software was used for analysis.


A5. Circular dichroism of normalised ellipticity at 285 nm, the maximum positive characteristic peak for c-MycC52, of 10 μM c-MycC52 in 10 mM NaCaco pH 6.6 with the addition of 5 μL of water added. The data analysis

was carried out in OriginPro 8. a) The graph was fitted with a dose response fitting b) the graph was fitted with a bi-dose response fitting.


a)

b)
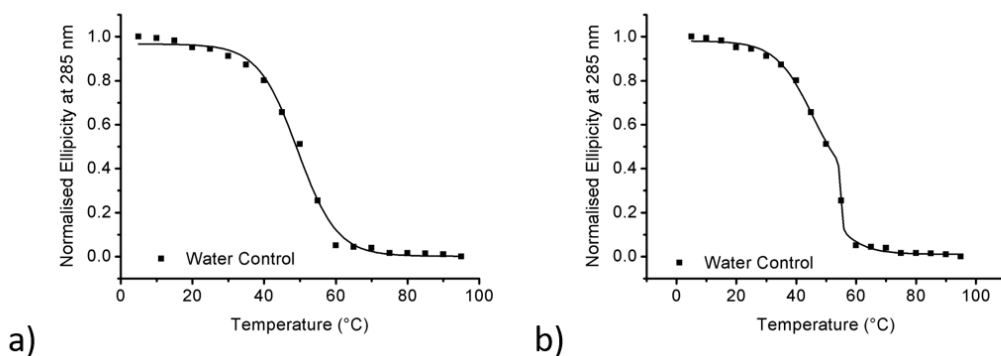
A6. Circular dichroism of normalised ellipticity at 285 nm, the maximum positive characteristic peak for c-MycC52, of 10 µM c-MycC52 in 10 mM NaCaco pH 6.6 with the addition of 10 equivalence of peptide. The data analysis was carried out in OriginPro 8. a) Pep-PTN at temperatures 5 °C

to 70 °C fitted with a dose response fitting b) Pep-PTN at temperatures 60 °C to 95 °C fitted with a dose response fitting c) Pep-SLC at temperatures 5 °C to 70 °C fitted with a dose response fitting d) Pep-SLC at temperatures 60 °C to 95 °C fitted with a dose response fitting e) Pep-VSE at temperatures 5 °C to 70 °C fitted with a dose response fitting f) Pep-VSE at temperatures 60 °C to 95 °C fitted with a dose response fitting.
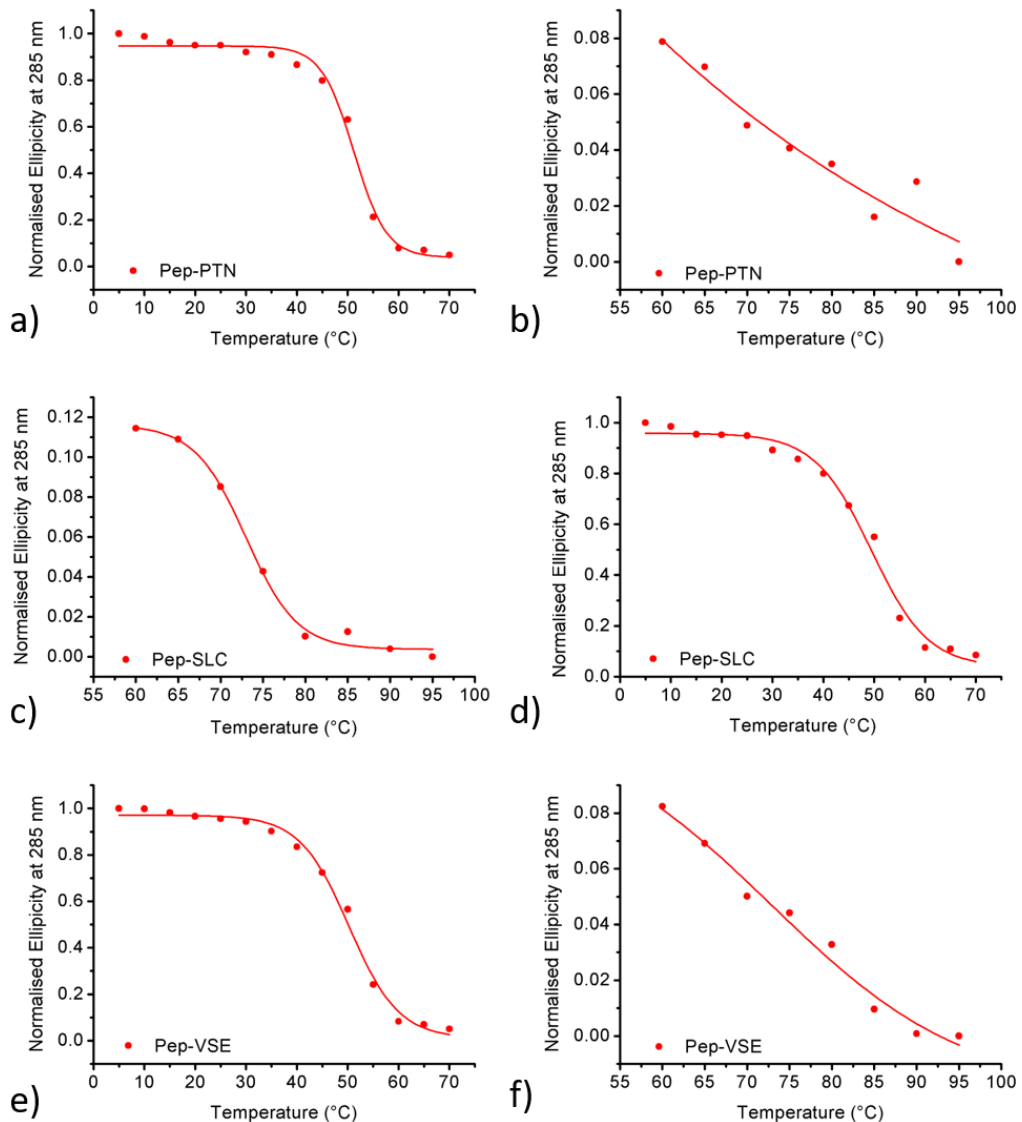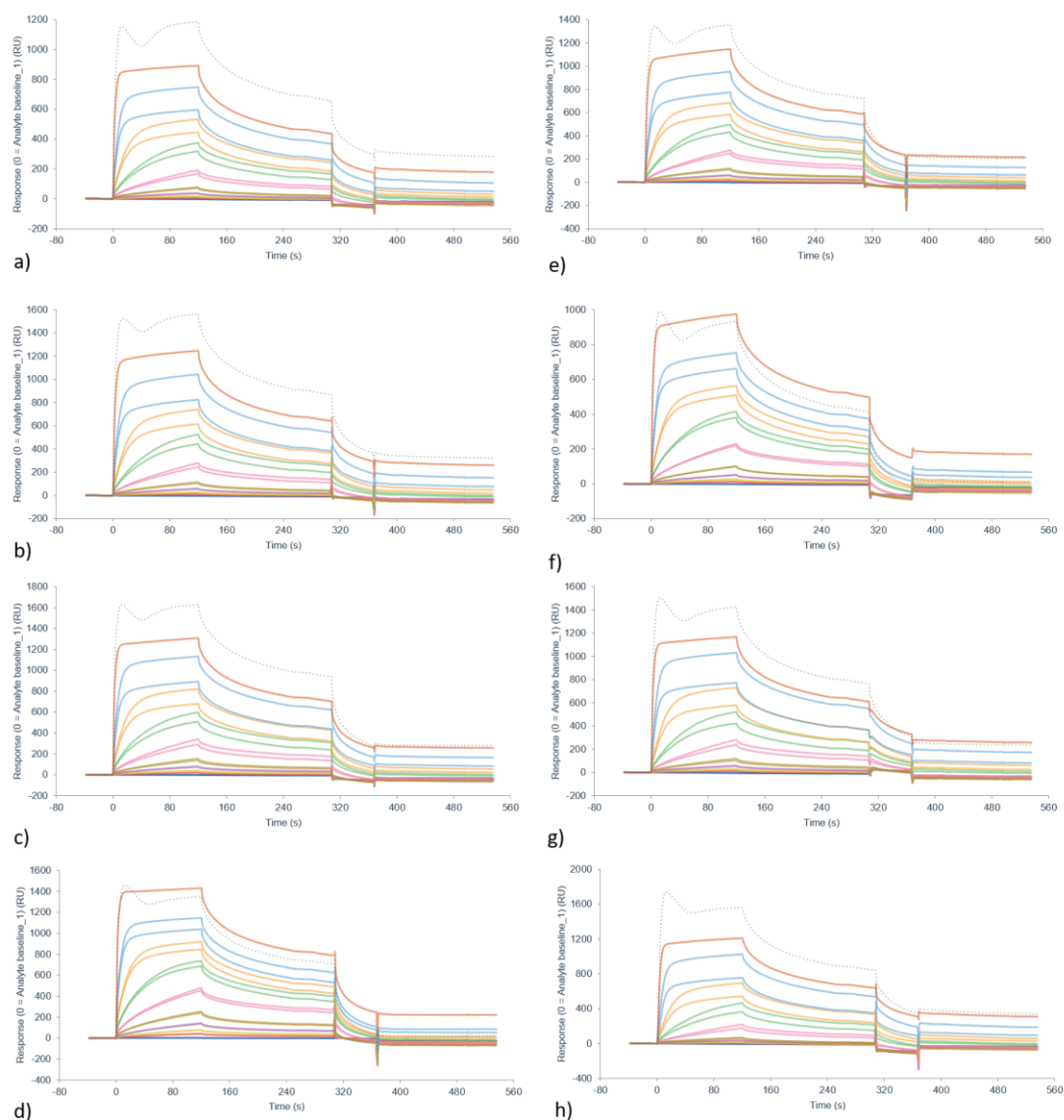

A7. Surface plasmon resonance sensorgrams for Pep-PTN of response vs time plots at concentrations from 0 μM to 100 μM in 2-fold intervals in the buffer 10 mM NaCaco 10 mM KCl 0.05% Tween20 pH 6.6. There are two replicates for every concentration shown by the same colour. a) B-DNA b) c-MycC27 c) c-MycC52 d) c-MycG e) DAP f) ATXN2L g) ILPR h) hTeloC

SPR experiments used Biacore SPR software. c-MycG is a G-quadruplex, all the other structures, other than B-DNA, are different i-motif forming sequences.
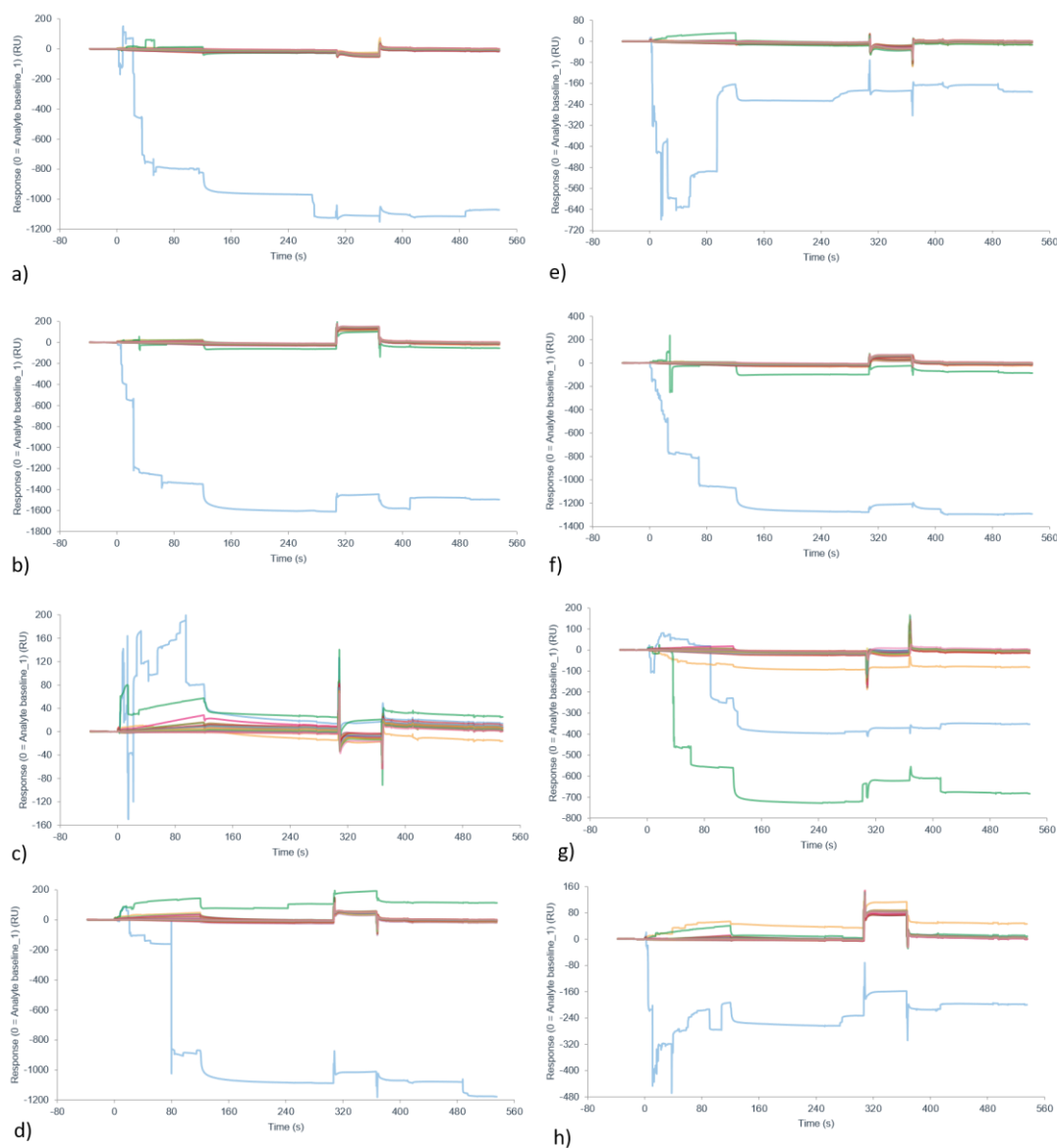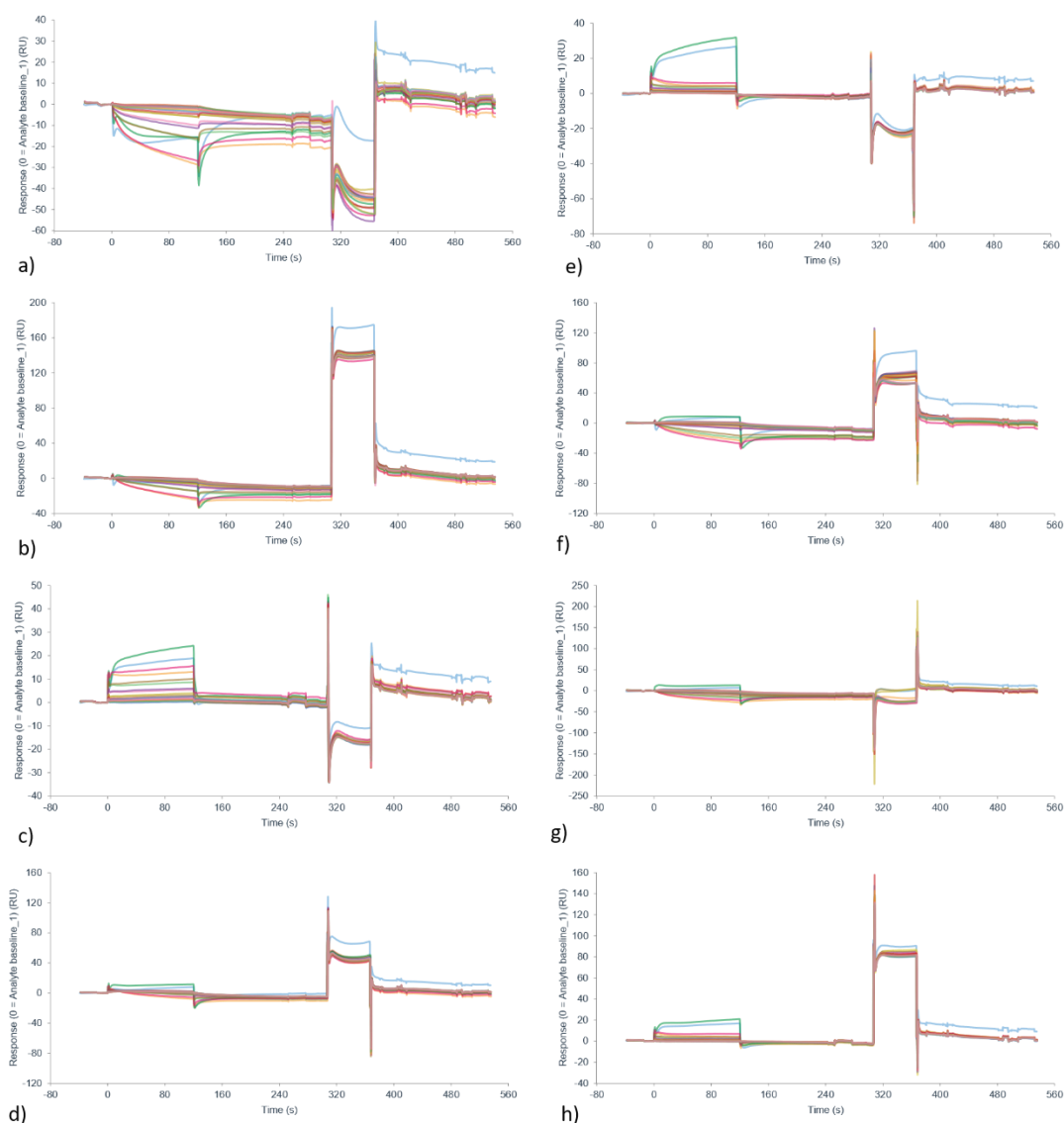

A8. Surface plasmon resonance sensorgrams for Pep-SLC of response vs time plots at concentrations from 0 μM to 100 μM in 2-fold intervals in the buffer 10 mM NaCaco 10 mM KCl 0.05% Tween20 pH 6.6. There are two replicates for every concentration shown by the same colour. a) B-DNA b) c-MycC27 c) c-MycC52 d) c-MycG e) DAP f) ATXN2L g) ILPR h) hTeloC SPR experiments used Biacore SPR software. c-MycG is a G-quadruplex,

all the other structures, other than B-DNA, are different i-motif forming sequences.
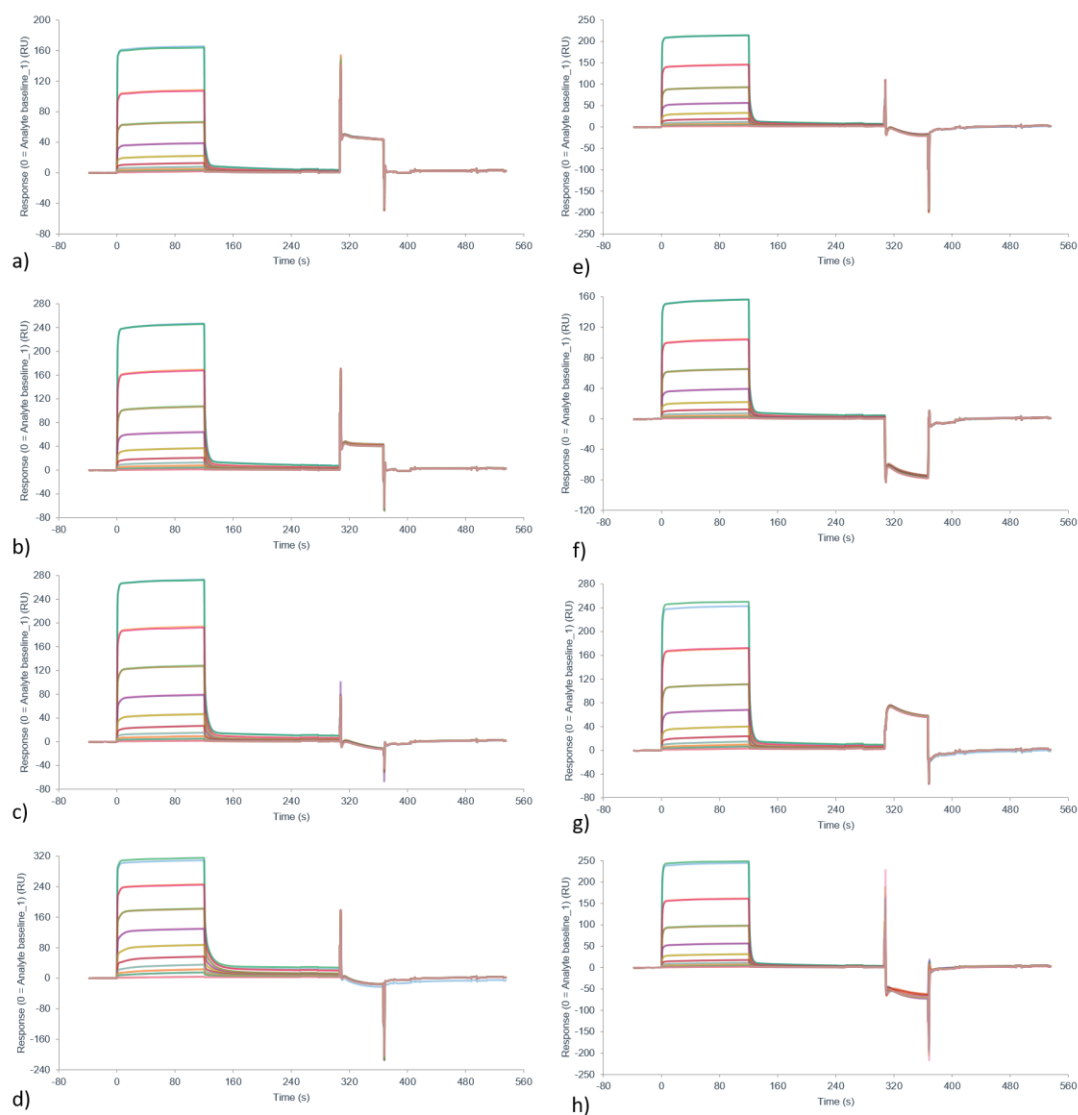

A9. Surface plasmon resonance sensorgrams for Pep-VSE of response vs time plots at concentrations from 0 μM to 100 μM in 2-fold intervals in the buffer 10 mM NaCaco 10 mM KCl 0.05% Tween20 pH 6.6. There are two replicates for every concentration shown by the same colour. a) B-DNA b) c-MycC27 c) c-MycC52 d) c-MycG e) DAP f) ATXN2L g) ILPR h) hTeloC SPR experiments used Biacore SPR software. c-MycG is a G-quadruplex,

all the other structures, other than B-DNA, are different i-motif forming sequences.

A10. Surface plasmon resonance sensorgrams for Pep-RVS of response vs time plots at concentrations from 0 µM to 100 µM in 2-fold intervals in the buffer 10 mM NaCaco 10 mM KCl 0.05% Tween20 pH 6.6. There are two replicates for every concentration shown by the same colour. a) B-DNA b) c-MycC27 c) c-MycC52 d) c-MycG e) DAP f) ATXN2L g) ILPR h) hTeloC SPR experiments used Biacore SPR software. c-MycG is a G-quadruplex,

all the other structures, other than B-DNA, are different i-motif forming sequences.

A11. Surface plasmon resonance sensorgrams for Pep-EIE of response vs time plots at concentrations from 0 μM to 100 μM in 2-fold intervals in the buffer 10 mM NaCaco 10 mM KCl 0.05% Tween20 pH 6.6. There are two replicates for every concentration shown by the same colour. a) B-DNA b) c-MycC27 c) c-MycC52 d) c-MycG e) DAP f) ATXN2L g) ILPR h) hTeloC SPR experiments used Biacore SPR software. c-MycG is a G-quadruplex,

all the other structures, other than B-DNA, are different i-motif forming sequences.