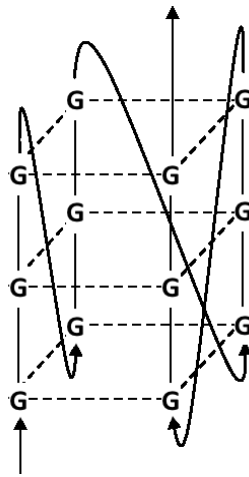


**Peptide Nucleic Acid conjugates regulate gene expression *in vitro*:
Mode of action and new strategies for improved cellular uptake**

Douglas Vair-Turnbull

Supervisor: Dr. Sylvain Ladame



Submitted in accordance with the requirements for the degree of
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOENGINEERING

IMPERIAL COLLEGE LONDON

APRIL 2018

Declaration of Originality

I declare that this thesis and the experiments described herein are my own work and that all relevant contributions have been referenced appropriately.

Copyright Declaration

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

Abstract

RNA G-quadruplexes (G4) are an important class of nucleic acid secondary structures that are involved in mRNA translation, alternative splicing, localisation and 3'-end processing. Putative RNA G4 forming sequences have been identified in the regulatory regions of many disease-related genes, in particular oncogenes, and as a result have been of increasing interest as therapeutic targets for chemical intervention. G4-binding small molecule ligands and antisense oligonucleotides have both been used to effectively regulate translation. However, the therapeutic potential of drugs targeting RNA G4 structures is limited by a lack of specificity in the presence of alternative RNA secondary structures (including other G4s). Herein, the efficacy of a new generation of "sequence + structure" specific RNA G4 ligands is tested against the well characterised *NRAS* G4, chosen as a model system. Ligands were designed to target both (i) G4-specific structural features using a flat aromatic and cationic NDI platform and (ii) the single-stranded G4 flanking regions using short complementary peptide nucleic acid (PNA) sequences. Using an *in vitro* translation assay these PNA-small molecule conjugates were shown to inhibit translation with a significantly lower IC₅₀ than the PNA or NDI alone. However, evidence suggested that the observed effect was mainly non-specific, which we suggest is due to electrostatic interactions mediated by the positively charged lysine residues added to the PNA-conjugates to improve solubility. In addition, a new 5'UTR G4 in the mRNA of the Aurora A kinase gene has been identified and characterised.

Therapeutic PNA oligomers, built around a neutral peptide backbone instead of the negatively-charged oligonucleotide's sugar-phosphate backbone, offer a number of advantages when compared to natural oligonucleotides, including resistance to proteases and higher binding affinity and specificity for complementary DNA (or RNA) sequences. However, PNA oligomers have very poor membrane permeability, which means that PNA conjugation is likely to significantly decrease their bioavailability. Conjugation of molecular transporters, such as cell penetrating peptides (or peptoids), is currently the most viable method to improve cellular uptake of therapeutic PNA oligomers. Here we report on the development of an *in vitro* assay and present initial results showing that sonoporation (formation of small pores in cell membranes by combined use of ultrasound and microbubbles) can be used to increase the cellular uptake of PNAs without the need for pre-conjugation to cell-penetrating peptides.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Sylvain Ladame for his help and guidance throughout my PhD. I would also like to thank him for always striving to provide opportunities for every member of our group to explore our interests and develop new skills. The supportive and inquisitive atmosphere this has encouraged in our group has been invaluable throughout my studies.

I would also like to express my sincere gratitude to the rest of the Ladame group, their support and guidance has been greatly appreciated. In particular, to Dr. Mazen Sleiman for always taking an interest in my progress and taking time to share his technical expertise and to Gavin Metcalf, Isobel Steer, Youngeun Choi and Dana Al Sulaiman for providing a ready ear to talk over new ideas or help troubleshoot experiments.

I am also thankful to the Department of Bioengineering for my funding and to all the members of the different labs in which I have worked over these three years for their assistance. In particular, I would like to thank all of the laboratory technicians who have worked extremely hard to keep our laboratories and equipment running smoothly.

I want to thank my family and close friends for their love and support. My parents have been a particular source of support during my thesis writing, providing proof-reading services and a supportive ear whenever needed.

Finally, special thanks go to Sophie Pirouet for being supportive and understanding through many late nights and long days and providing continued encouragement, which has contributed greatly towards the completion of this thesis.

Table of Contents

Chapter 1: Introduction to G-quadruplex	1
1.1. <i>Nucleic acid secondary structures</i>	1
1.1.1. Double helices	1
1.1.2. RNA Tetra-loops	4
1.1.3. RNA Pseudoknots	5
1.2. <i>DNA G4 structure</i>	7
1.2.1. G4 topology	8
1.2.2. Telomeric G4	10
1.2.3. C-MYC G4	13
1.2.4. BCL-2 G4	15
1.2.5. VEGF G4	15
1.2.6. Limitations of structural studies	16
1.3. <i>RNA G4 structures</i>	17
1.3.1. r(UGGGGU) G4	17
1.3.2. TERRA G4	19
1.3.3. Other RNA G4	20
1.4. <i>In vitro functional evidence for DNA G4</i>	21
1.4.1. Regulating telomere extension	21
1.4.2. Immunoglobulin class switching	22
1.4.3. Epigenetic regulation	22
1.4.4. Regulation of Transcription	22
1.5. <i>In vitro functional evidence for RNA G-Quadruplex</i>	23
1.5.1. Regulation of Translation	23
1.5.2. 3'-end processing and poly-adenylation	24
1.5.3. Alternative Splicing	25
1.5.4. Mitochondrial transcription termination	25
1.5.5. mRNA localisation	26
1.5.6. Telomere maintenance and heterochromatin formation	26
1.6. <i>Evidence for in vivo activity of RNA and DNA G4</i>	27
1.6.1. Genomic studies	27
1.6.2. <i>In vivo</i> imaging studies	27
1.6.3. <i>In vivo</i> binding and functional studies	29
1.7. <i>Targeting G-quadruplex</i>	31

1.7.1.	Therapeutic potential of G4 targeting	31
1.7.2.	Small molecule ligands	32
1.7.3.	Antisense strategies	39
1.8.	<i>Peptide Nucleic Acids (PNA)</i>	39
1.8.1.	Structure and Backbone modifications	39
1.8.2.	PNA:DNA and PNA:RNA interactions	40
1.8.3.	Therapeutic applications of PNA	44
1.8.4.	Bioavailability and cellular uptake of PNA	44
Chapter 2: Targeting of RNA G-quadruplexes with PNA-small molecule conjugates		45
2.1.	<i>Introduction</i>	45
2.2.	<i>Methods</i>	50
2.2.1.	Ligand Design and Synthesis	50
2.2.2.	Translation Assays	52
2.2.3.	Binding Assays	55
2.3.	<i>Results</i>	57
2.3.1.	PNA-NDI conjugates design and synthesis	57
2.3.2.	<i>In vitro</i> translation assay: Introduction	64
2.3.3.	<i>In vitro</i> translation assay: Optimisation	65
2.3.4.	<i>In vitro</i> translation assay: Results	70
2.3.5.	Characterisation of binding	79
2.4.	<i>Discussion</i>	88
2.4.1.	Summary	88
2.4.2.	Effect of increasing PNA length	90
2.4.3.	Effect of targeting 3' or 5' flanking region	91
2.4.4.	PNA-NDI binding mechanism	92
2.4.5.	Specificity of PNA-NDI conjugates	94
2.4.6.	Assay limitations	95
2.4.7.	Future work	96
Chapter 3: Improving the cellular uptake of PNA by sonoporation		98
3.1.	<i>Introduction</i>	98
3.1.1.	PNA backbone modifications	98
3.1.2.	Lipofection	99
3.1.3.	Cell Penetrating Peptides (CPP)	99
3.1.4.	Cell Penetrating Peptoids	102

3.1.5.	Sonoporation	102
3.2.	<i>Methods</i>	104
3.2.1.	Cell culture	104
3.2.2.	PNA and CPP synthesis	105
3.2.3.	Microbubble manufacturing	106
3.2.4.	Flow system	106
3.2.5.	Ultrasound targeting	106
3.2.6.	Sonoporation	106
3.2.7.	Cell staining	107
3.2.8.	Microscopy	107
3.2.9.	Image and Data Analysis	107
3.2.10.	Splice Correction Assay	108
3.3.	<i>Results</i>	108
3.3.1.	Sonoporation set-up	108
3.3.2.	Eliminating standing waves	109
3.3.3.	Introducing Flow	109
3.3.4.	Optimising assay endpoint analysis	111
3.3.5.	Optimisation of sonoporation parameters	114
3.3.6.	Comparison of optimised sonoporation parameters and CPP	120
3.3.7.	Bioavailability of transfected PNA	121
3.4.	<i>Discussion</i>	122
3.4.1.	Summary	122
3.4.2.	Comparison with previous studies	122
3.4.3.	Comparison with (R-Ahx-R) ₄	123
3.4.4.	Future work	123
Chapter 4: Identification and characterisation of RNA G4 as potential new therapeutic targets		125
4.1.	<i>Introduction</i>	125
4.1.1.	Identification of novel PQS	125
4.1.2.	Evidence for formation of two quartet G4	126
4.1.3.	Aurora A: <i>In vivo</i> function	130
4.1.4.	Aurora A: Current therapeutics targeting Aurora kinases	131
4.1.5.	Col1A2: <i>In vivo</i> function	135
4.2.	<i>Methods</i>	136
4.3.	<i>Results</i>	137

4.3.1. Evidence of G-quadruplex formation <i>in vitro</i>	137
4.3.2. Aurora A variants	139
4.3.3. Effect of ligand binding on G4 formation	141
4.4. Discussion	142
4.4.1. Formation of G4 by novel PQS	142
4.4.2. Functional role of <i>Aurora A</i> G4	142
4.4.3 <i>Aurora A</i> G4 as a biomarker	144
4.4.4. <i>Aurora A</i> G4 as a therapeutic target	144
Conclusion	146
References	148

Abbreviations

A β : Amyloid-beta Protein
aCGH: array Comparative Genomic Hybridization
Aeg: Aminoethylglyine
ALS: Amyotrophic Lateral Sclerosis
APA: Alternative Poly-Adenylation Signal
APP: Amyloid Precursor Protein
ASO: Antisense Oligonucleotide
aTNA: acyclic (L)-Threoninol Nucleic Acid
BACE-1: β -secretase 1
CaMKIIa: Ca21/Calmodulin-dependent Protein Kinase IIa
CDS: Coding Sequence
ChIP: Chromatin Immunoprecipitation
c-MYC: V-myc Avian Myelocytomatosis Viral Oncogene Homolog
CPP: Cell Penetrating Peptide
CPPo: Cell-Penetrating Peptoid
CSR: Class Switch Recombination
DMD: Duchenne Muscular Dystrophy
dNTP: Deoxyribose Nucleoside Triphosphate
EGF: Epidermal Growth Factor
EGFR: Epidermal Growth Factor Receptor
EMSA: Electrophoresis Mobility Shift Assay
FAM: Fluorescein Amidite
FBS: Fetal Bovine Serum
F-EMSA: Fluorescence-based Electrophoresis Mobility Shift Assay
Fmoc: Fluorenylmethyloxycarbonyl chloride
FMR1: Fragile-X Mental Retardation 1
FMRP: Fragile-X Mental Retardation Protein
FRET: Fluorescence Resonance Energy Transfer
FTD: Frontotemporal Dementia
FXR1: FMR1 Autosomal Homolog 1
G4: Guanine-Quadruplex(es)
GFP: Green Fluorescent Protein
GPNA: L- γ -Glutamyl-p-nitroanilide
HIF1 α : Hypoxia Inducible Factor 1 alpha

HP: Homing or Targeting Peptide

hnRNP H/F: Heterogeneous Nuclear Ribonucleoprotein H/F

hTERT: Human Telomerase Reverse Transcriptase

Ig: Immunoglobulin

IRES: Internal Ribosome Entry Site

LB: Lysogeny Broth

LNA: Locked Nucleic Acid

LRP5: Low-density Lipoprotein Receptor-related Protein 5

LTR: Long Terminal Repeat

NDI: Naphthalene Di-imide

NHE III: Nuclease Hypersensitive Element III

nt: Nucleotides

ORC: Origin Recognition Complex

ORF: Open Reading Frame

pA: Poly-adenylation

PAX9: Paired Box Gene 9

PBS: Phosphate Buffered Saline

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

PEG: Polyethylene Glycol

PMO: Phosphorodiamidate Morpholino Oligonucleotides

PNA: Peptide Nucleic Acid

PPG: Pyrazolo[3,4-d]Pyrimidine Guanine

PQS: Putative Quadruplex Sequence

PSD-95: Postsynaptic Density Protein 95

QFS: Quadruplex Forming Sequence

RFP: Red Fluorescent Protein

RRL: Rabbit Reticulocyte Lysate

RT-PCR: Reverse Transcription - Polymerase Chain Reaction

qSNP: Quadruplex Single Nucleotide Polymorphism

SNP: Single Nucleotide Polymorphism

SPR: Surface Plasmon Resonance

TERRA: Telomeric-repeat containing RNA

TFA: Trifluoroacetic Acid

TIS: Triisopropylsilane

TMPyP4: Tetra(N-methyl-4-pyridyl) Porphyrin

TRF2: Telomere Repeat Factor 2

UTR: Untranslated Region

VEGF: Vascular Endothelial Growth Factor

Document specific abbreviations

“PNA”: Used in this document to denote a PNA₁-alkyl-PNA₂ entity in ligand design

“NDI”: Used in this document to denote an NDI functionalised with two Lysine residues

“UTR”: Used in this document to denote NRAS 5'UTR

Chapter 1: Introduction to G-quadruplex

1.1. Nucleic acid secondary structures

Since the B-form double-helical structure of DNA was demonstrated by Watson and Crick in 1953¹ a number of additional secondary structures have been identified in both RNA and DNA, such as triplexes, hairpin loops, pseudoknots and Guanine-quadruplexes (G4). Some of the main functions of these secondary structures *in vivo* include regulating transcription and translation either directly or through recruitment of binding proteins such as DNA methylase or ribosome subunits, regulation of chromosomal stability and regulation of telomere extension.²⁻⁵ While this thesis is focused on G4, it is helpful to first briefly consider the discovery and function of other relevant secondary structures.

1.1.1. Double helices

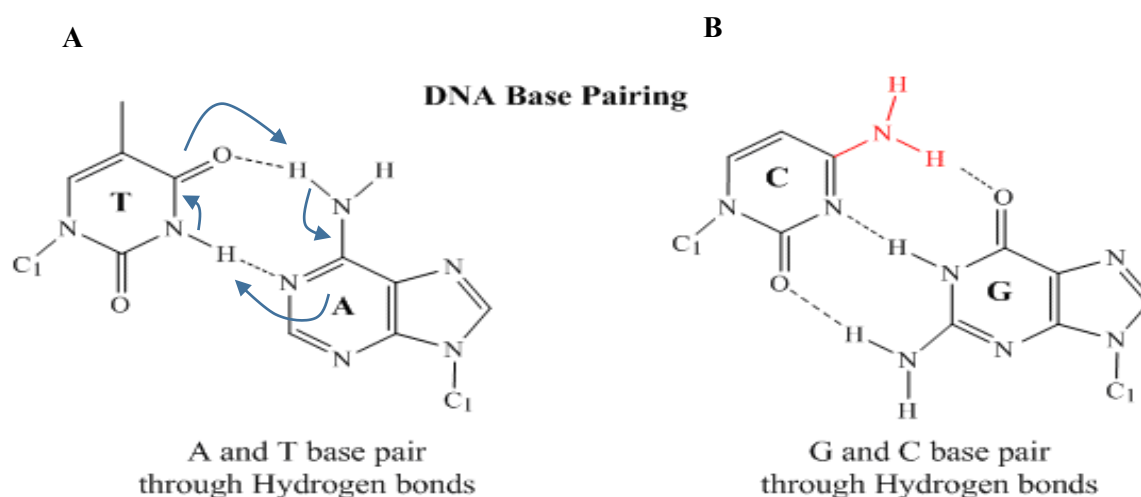


Figure 1: Watson-Crick base pairing interactions. A) Adenine-thymine (A-T) base pair. The bases interact through formation of two hydrogen bonds. Arrows show the movement of electrons during hydrogen bond formation and demonstrate the mutual stabilisation effect from having adjacent hydrogen bond acceptors and donors. B) Guanine-cytosine (G-C) base pair. The bases interact through a set of three hydrogen bonds, as a result C-G base pairs are more thermodynamically stable than A-T base pairs. Thus, the overall stability of double stranded nucleic acids is affected by the G-C ratio of the sequence, with strands that contain a higher proportion of G-C base pairs exhibiting greater stability.

In their seminal 1953 paper, Watson and Crick proposed a structure for double stranded DNA from which is derived much of our understanding of DNA secondary structure.¹ They proposed that a range of competing forces drive double-stranded DNA to fold into a defined secondary structure, now known as a B-form double helix. The first is Watson-Crick base pairing, which is the formation of hydrogen bonds between complementary bases, shown in more detail in Figure 1. As donor and acceptor groups are adjacent to each other in the nucleotide bases ring system, the hydrogen bonds are mutually

stabilising, i.e. the dipole formed by a proton donor group increases the electronegativity of the neighbouring acceptor group and thereby increases the stability of the hydrogen bond.

The second set of key forces are the Van-der-Waals stacking interactions between the base ring structures. A key reason for the stability of B-form DNA is that the bases are aligned parallel to each other and brought into close proximity, which results in strong Van der Waals interactions between the adjacent ring structures. The third key force is electrostatic repulsion between the negatively charged phosphate groups in the nucleic acid backbone. By destabilising the double helix, it reduces the energy required to separate DNA into single strands during transcription of specific genes or replication of the full DNA helix. In Figure 2, these initial three forces are shown in the context of a full double helical structure. A final factor that affects folding is the surrounding solvent environment. The backbone is highly hydrophilic and solvent (water) molecules play a key role in minimising electrostatic penalties and stabilising the overall structure. Surrounding water molecules form an ordered hydration shell around the helix which acts to stabilise the structure through temporary interactions with hydrophilic groups, in particular the phosphate groups.

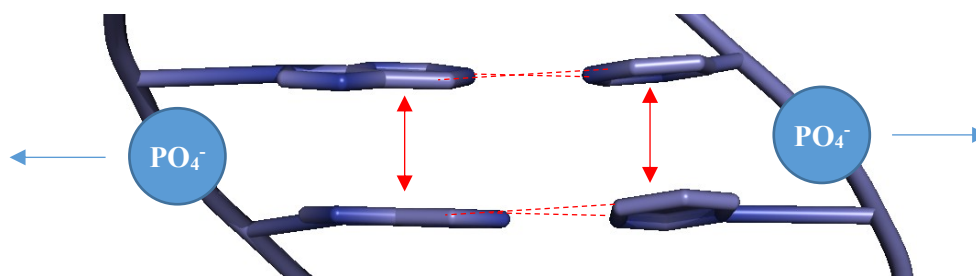


Figure 2: Representation of the key forces involved in DNA duplex formation. The nucleotide bases are involved in two key interactions, (i) Watson-Crick hydrogen bonding, shown with dashed red lines, and (ii) Base stacking interactions, or Van der Waals forces between ring systems of adjacent nucleotides, shown with red arrows. The distance dependence of Van der Waals forces is $1/r^6$, so structures that allow the bases to lie flat and interact as closely as possible across a wide area are favoured. (iii) The final force represented here is the electrostatic repulsion between the negatively charged phosphates on the nucleotide backbone, shown with blue arrows.

While double stranded DNA exists predominantly in the B-form under physiological conditions, under certain conditions it can form alternate helices, the most common of which is called A-form. The A- and B-forms differ in terms of the shape of the major and minor grooves, whether they are right or left handed, and the conformation of the bases (which is influenced by the sugar pucker of the ribose ring). Figure 3 shows the key differences between A- and B-form DNA. A more detailed description of the different types of DNA helices (including the Z-form helix which is beyond the scope of this work) can be found in a comprehensive review by Dickerson *et al.*⁶

RNA typically exists in a single stranded form but is also capable of forming double stranded structures. Three years after the determination of the structure of dsDNA, Rich and Davies solved the fibre X-ray diffraction structure for a double stranded RNA formed from the polymers poly(rA) and poly(rU),

demonstrating that it could form a helical double stranded structure analogous to that formed by DNA.⁷ This was the first demonstrated example of an RNA hybridisation reaction, and it was a vastly important

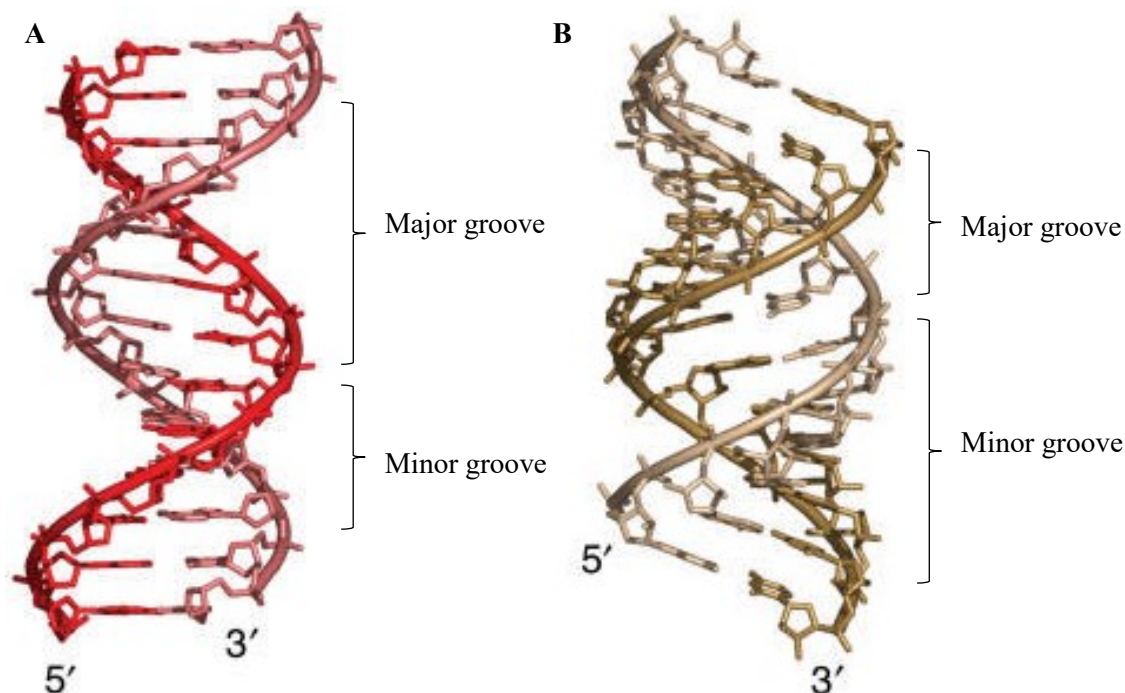


Figure 3: The structural differences between B-form and A-form double stranded nucleic acid helical structures. A) The B-form double helix. The helix is right-handed (the nucleic acid backbone rotates in a clockwise direction). The sugar pucker is in the C2'-*endo* conformation which induces limited rotation of the bases and forms a wide major groove and a narrow minor groove (approximately half as wide), labelled on the diagram. **B)** A-form double helix. The helix is also right-handed; however, the 2'-hydroxyl group pushes the sugar into the C3'-*endo* conformation. This results in a more pronounced rotation in the bases orientation with respect to the helical axis, producing a deep and narrow major groove and a wide, shallow minor groove. Figure reproduced from Reiter *et al.*⁸

step on the path to elucidating RNA's functions in cells. Additional experiments in 1960 demonstrated the first hybridisation reaction between DNA and RNA demonstrating the principle of how information could be transferred from one to the other.⁹ From this experiment was built our current day understanding of how single stranded mRNA acts to transmit genetic information from the nucleus out into the cell. At this point in time however, there was still no clarity on whether RNA could form any more complex structures. In 1973 Rich and co-workers published a further set of studies which led to the first single crystal structure of RNA helices.^{10, 11} From these studies the binding interactions and structural organisation of RNA helices were clarified, and dsRNA was shown to form a structure very similar to that of A-form DNA. The presence of the 2'-hydroxyl group on RNA restricts the sugar pucker to the C3'-*endo* conformation, thereby inducing formation of a more compact helix which is equivalent to A-DNA. We now know that there are a number of structures which build upon the RNA double helix and perform vital *in vivo* functions, in particular ribozymes and tRNA.¹² One of the most basic functions of the double helix in the formation of RNA secondary structures is in the formation of hairpins (for a schematic representation of an RNA hairpin see Figure 4).¹³ The hairpin was first identified in the eubacteria 16S-like ribosome in 1983 by Woese *et al.*, and is now known to be the most

common RNA structural motif.^{14, 15} The hairpin is the core element of a number of more complex secondary structures. In the next two sections, the structure and function of a selection of these elements, namely pseudoknots and tetraloops, will be described.

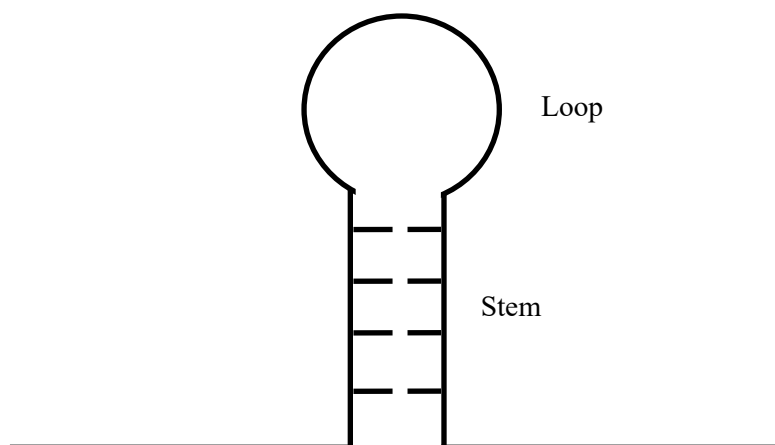


Figure 4: Schematic representation of an RNA hairpin structure. Hairpins are one of the most basic RNA structures, they are formed through formation of a double stranded RNA stem region with a connecting loop, both labelled.

1.1.2. RNA Tetra-loops

Tetraloops are a type of RNA hairpin with specific four-nucleotide loop regions. They are extremely common in RNA biology and are typically very stable compact structures. The shorter loop regions are able to form base-stacking, base-ribose and base-phosphate hydrogen bonds which stabilise the structure and minimise unfavourable base-solvent interactions (Figure 5).¹⁶ The most common (and most stable) variants of the structure are the UNCG (where N refers to A, U, C or G), GNRA (where R refers to A or G) and CUUG motifs. The sequence of the tetraloop can have a significant impact on its structural conformation and subsequent function, for example GNRA tetraloops have a defined structure with the guanine and adenine forming a sheared base pair and the remaining nucleotides forming a loop (Figure 5).¹⁷ This structure reveals the Watson-Crick edge of the guanine and adenine, which enables the loop to act as a binding site for other RNA or protein elements. A well-studied example is the GAAA tetraloop/11-nucleotide receptor interaction, where the tetraloop binds in the minor groove pocket of the 11-nucleotide double helix receptor. This interaction has been studied through structural analysis. For example the NMR structure of a 30kDa GAAA tetraloop-receptor complex was solved in 2005 by Davis *et al.*¹⁸, which built upon a previous structure by Butcher *et al.* in 1997.¹⁹ The interaction between the GAAA tetraloop and its receptor was shown to initiate tertiary structure formation in solution. Conserved tetraloop receptors have also been identified in group I and group II self-splicing introns, RNase P and rRNAs. Structural and functional studies have suggested that they also play roles in the folding of these RNA secondary structures.²⁰⁻²²

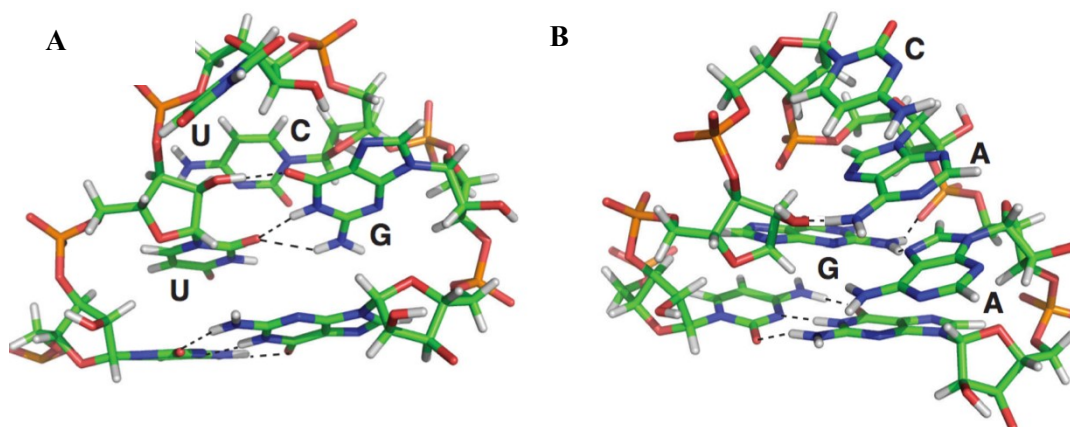


Figure 5: The structures of the UUCG (PDB ID: 2KOC) and GCAA (PDB ID: 1ZIH) tetraloops. **A)** The backbone rotation in the UUCG loop occurs between the first two U nucleotides and is stabilised by a base-phosphate hydrogen bond between the C3 amino group and the U2 5'-phosphate. The first and fourth nucleotides (U1 and G4) interact via a bifurcated hydrogen bond between U1 carbonyl and G4 imino and amino groups. This interaction stabilises the loop and provides a surface on which the C3 stacks. **B)** The structure of the GCAA tetraloop is similarly stabilised by interactions between the first and fourth nucleotides. In this case it is a sheared base pair interaction and the structure of the loop leaves the Watson-Crick faces of the nucleotides available for interactions with other tertiary structures including proteins and RNA structures. Hydrogen bonds are shown as dotted black lines, carbon atoms shown in black, phosphorus in purple, nitrogen in blue, oxygen in red and hydrogen in white. Figure reproduced from Cheong *et al.*²³

1.1.3. RNA Pseudoknots

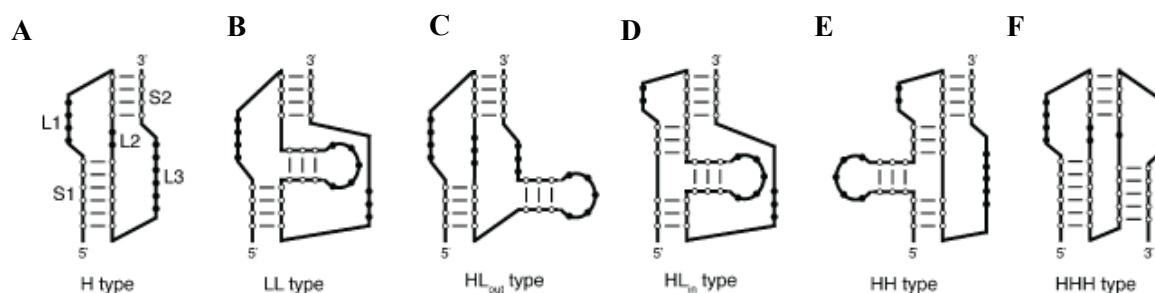


Figure 6: The six classes of intramolecular pseudoknot identified in the PseudoBase database.²⁴ **A)** The H type pseudoknot contains a hairpin with an atypical loop, which then binds to a complementary sequence outside of the hairpin. The result is a double hairpin structure with two stem regions (S1 and S2) which are linked by three linkers (L1-3). **B)** An LL type pseudoknot is similar to the H type but contains an extra internal hairpin within the loop region of the initial hairpin. **C)** Addition of an additional helix in the L3 region produces an HL_{out} type pseudoknot, where the extra helix sits downstream from the S1 element. The additional helix can also be placed within the S1 element. This can be either **D)** in the 3' region of the stem, producing a HL_{in} type knot or **E)** in the 5' region, producing an HH type knot. **F)** The HHH type knot is formed through complementary base pair interaction between the loop regions of two helices. H represents a helix structure, L a bulge, interior or multiple loop and S a stem structure. Figure reproduced from Han *et al.*²⁵

The pseudoknot is an RNA secondary structure, first identified in plant viral RNA in 1982, which is formed from a hairpin loop which interacts with a complementary sequence outside of the loop.²⁶ The structure can become increasingly complex with the addition of variable length loops and other RNA structural features, including bulges and non-canonical base pair interactions. A selection of pseudoknot structures is shown in Figure 6. At their most basic level, pseudoknots are structural building blocks that enable RNA to be built up into more complex tertiary structures, such as the ribosome. Pseudoknots have also been suggested to play a role in regulation of various cellular processes including

RNA processing and inactivation of toxins (for a more comprehensive description of the functions of pseudoknots see reviews by the Serganov and Pennel groups).²⁷⁻²⁹ Of particular interest for the work in this thesis is the evidence that pseudoknots in mRNA can play an important role in the regulation of gene expression, in particular through their involvement in the mechanisms of translation initiation and elongation.

In bacteria and viruses, pseudoknots can play a significant independent role in translation initiation. For example, upon binding of the cofactor S-adenosylmethionine-II (SAM-II), the SAM-II riboswitch forms an H-type pseudoknot which obstructs the Shine-Dalgarno sequence of the mRNA and thereby prevents translation of the downstream gene.³⁰ In addition, pseudoknots have also been described as important structural elements in internal ribosome entry site (IRES) formation and function. IRES enables cap-independent translation initiation by providing a binding site for the ribosome to begin assembly. The best characterised example of the role of a pseudoknot in IRES function is in the cricket paralysis virus gene (*CrPV*). This IRES is relatively short at 200nt, and its structure is dominated by three pseudoknots, which each sit within a separate domain (Figure 7). Pseudoknots form a key structural part of the IRES and form interactions with the ribosomal factors, which enables ribosomal assembly, as well as recruiting tRNA to the A-site of the ribosome.³¹

Pseudoknots in mRNA are also involved in translation elongation, in particular through the introduction of frameshifts, most typically programmed -1 ribosomal frameshifts. This is a mechanism utilised by viruses to produce two different proteins from a single mRNA and to control the mix of the two. One key element is a “slippery” surface which is where the frameshift itself occurs and the ribosome transitions to the -1 frame. The other core element is an adjacent structure to induce the frameshift to occur, this is most typically an H-type pseudoknot.³²

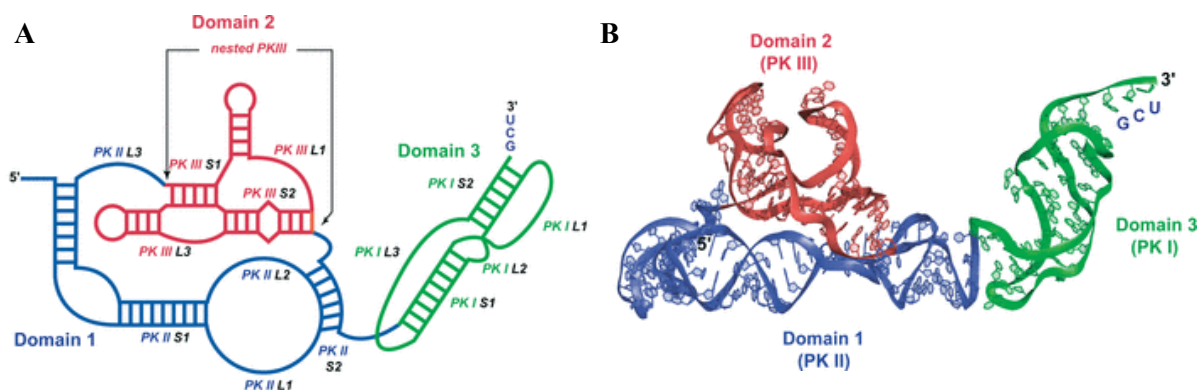


Figure 7: The *CrPV* IRES with the three domains identified, domain 1 (blue), domain 2 (red), domain 3 (green). Included is a 2D representation of the structure (A), and a 3D model (B), derived from Schuler *et al.*³¹ Within each domain the elements of the pseudoknot (PKI-III) are identified, the loop/linker regions within each pseudoknot are labelled as L1-3 and the stem regions labelled as S1-2. Domain 1 interacts with the 60S ribosome through the E- and P-sites. Domain 2 interacts with the 40S ribosomes via the E-site. Domain 3 is formed almost exclusively from PKI and acts as a P-site tRNA mimic when binding to the ribosome. In addition, the GCU (labelled) bases of domain 3 interact with tRNA^{Ala}, thereby recruiting it to the ribosome A-site. Figure reproduced from Brierley *et al.*²⁸

1.2. DNA G4 structure

A class of secondary structures that have been receiving a lot of attention in recent years are G-quadruplexes (or G4), which are formed through a range of specific interactions between guanine residues. The first indication that guanine-rich oligonucleotides could fold into unique secondary structures came in 1910 when it was reported by Ivar Bang that guanylic acid formed a gel at high concentrations in the presence of salts.³³ Guanylic acid, also known as guanosine monophosphate, is the monomeric form of the RNA base guanine, which requires a polymerisation reaction to form into an RNA polymer. The formation of a gel suggested that the guanosine monomers could interact with each other to form a more complex secondary structure. What became interesting in retrospect was that this binding occurred in the absence of what we now know as Watson-Crick hydrogen bonding. This physical property was only fully understood 50 years later when Gellert *et al.* used X-ray fibre diffraction to demonstrate that the guanylic acid molecules in the gel formed tetrameric structures called G-quartets (or tetrads of guanines).³⁴ These are cyclic co-planar arrangements of four guanines interacting with each other via Hoogsteen H-bonds between N₁-O₆ and N₂-N₇, further stabilised by a central monovalent metal ion (Figure 8A). The evidence that such highly organised structures could appear in DNA came more than 20 years later.

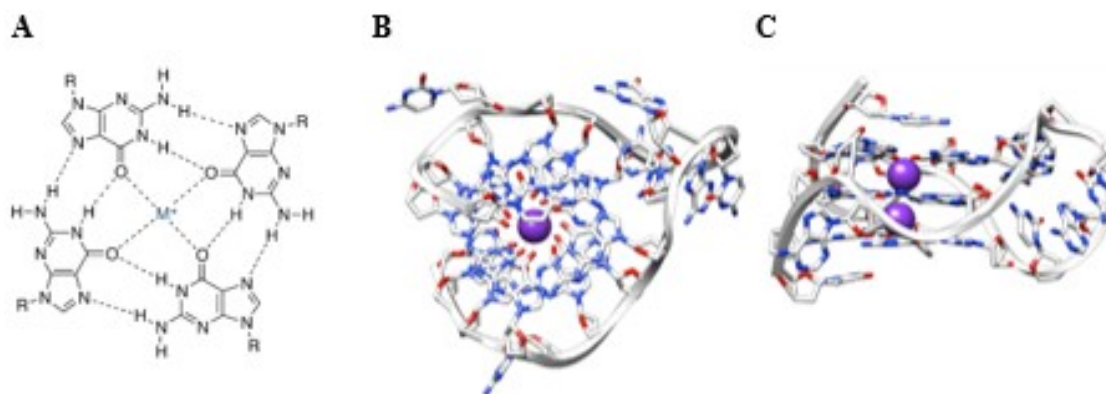


Figure 8: Chemical representation of a G-quartet and 3D structure of the *c-kit* DNA G4 solved by ¹H NMR. **A)** The structure of a G-quartet. The structure is stabilised by hydrogen bonds (dashed lines) between guanine residues (Hoogsteen hydrogen bonding) and between the C-6 carbonyl group and a monovalent cation (M⁺) in the centre of the structure. **B)** ¹H NMR structure (top view, down central cation channel) of parallel G4 in the *c-kit* oncogene. **C)** ¹H NMR structure (side view) of parallel G4 in *c-kit* oncogene.³⁵ In **B** and **C** the nitrogen (blue) and oxygen atoms (red) are highlighted, as well as the monovalent cations in the central channel (purple) (PDB Entry: 2KQH).

In 1988, Sen and Gilbert provided the first evidence of such G-rich structures through analysis of human immunoglobulin switch regions.³⁶ They found that the addition of cations to guanine-rich sequences prior to electrophoresis mobility experiments resulted in aggregation of the sequences, visible through reduced mobility. Further experiments investigating this effect revealed that sets of contiguous guanine residues were protected from methylation, suggesting that they were involved in stable secondary structures (Figure 9). This study was followed in 1989 by similar experiments carried out on telomeric DNA, demonstrating that the G-rich single-stranded 3'-overhang of telomeres formed secondary

structures that protected the guanines from methylation.³⁷ The authors postulated that the planar nature of G-quartets allowed them to stack atop each other to form a more complex structure called a Guanine-quadruplex (G-quadruplex or G4).

Oligo P: **GGG**ACCAGACCTAG**GCAGCTATGGGGGAGCTGGGG**A**AGGTGGG**AAT**GTGA**


Figure 9: Methylation patterns of oligonucleotide P from Immunoglobulin switch regions. Addition of dimethyl sulphate results in methylation of exposed guanines, highlighted in bold. The guanines involved in G4 formation are protected from methylation in Hoogsteen hydrogen bonding arrangements, underlined.

1.2.1. G4 topology

It is now commonly accepted that a unique oligonucleotide strand that contains four contiguous runs of at least two consecutive guanines has the potential to fold into a G4 structure, which consists of two or more stacked G-quartets directly linked by the DNA (or RNA) sugar phosphate backbone. As well as forming within one strand (intramolecular), G4 can also be formed through the interaction of G-rich regions on multiple strands (intermolecular)³⁸ including bi-molecular³⁹ and tetra-molecular quadruplexes (Figure 10).⁴⁰

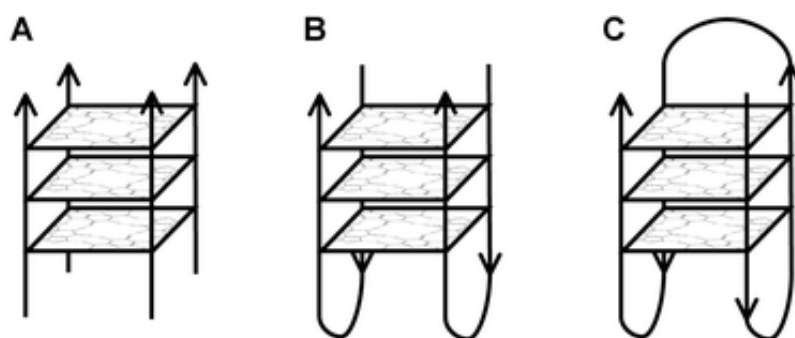


Figure 10: Schematic representation of G-quadruplexes of varying stoichiometries. A) A tetra-molecular G4 formed by bonding between contiguous guanine residues on four separate nucleic acid strands. B) A bi-molecular G4 formed as a result of bonding between two sets of contiguous residues on two separate strands. C) An intramolecular G4, in this case the G4 is formed from 4 separate runs of contiguous guanines within the same strand. Figure reproduced from Huppert.⁴¹

The backbone strands can be orientated in different directions and G4 can therefore be classified into a variety of topologies depending on whether the strands are aligned all-parallel, three parallel and one anti-parallel, adjacent parallel or alternating anti-parallel with respect to each other (Figure 11). G4 topology can be affected by a number of parameters, including loop length (i.e. number of nucleotides between G-tracts), cation and ligand binding.⁴² For example, the same G-rich DNA sequence can adopt different (potentially co-existing) G4 topologies depending on the nature of the central cation, i.e. with telomeric G4, which are discussed in more depth later in this section.⁴³ The glycosidic torsion angles of the guanine bases in the quadruplex are affected by the orientation of the backbone. All-parallel topologies require that guanines in all G-strands adopt the same glycosidic bond conformation, either

all guanines in the *syn* conformation or all in the *anti* conformation. By contrast, guanines of adjacent strands in the antiparallel orientation will have different glycosidic torsion angles (Figure 11). In addition, the relative orientation of the guanines in adjacent strands affects the size of the groove between strands. Adjacent parallel strands produce medium sized grooves, whereas adjacent anti-parallel strands will produce either wide or narrow grooves.

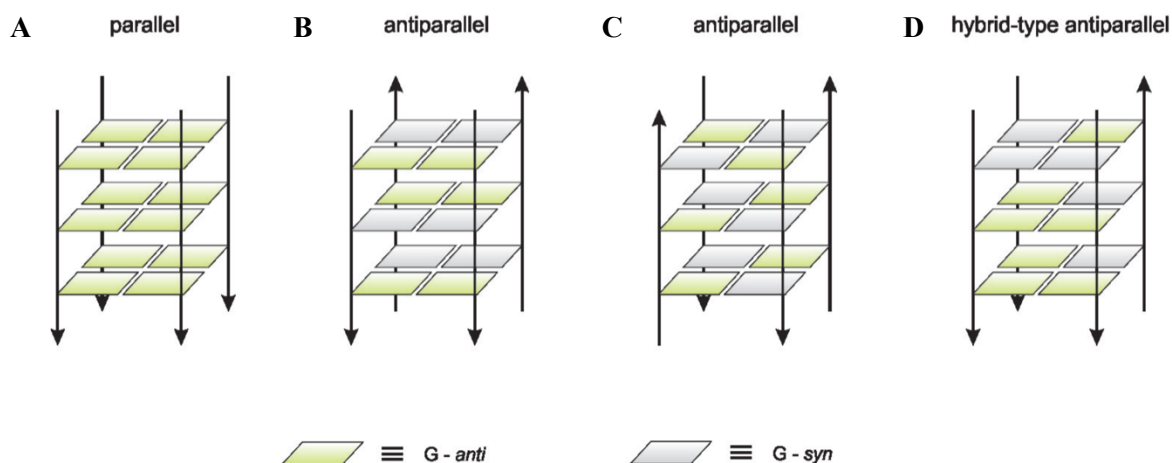


Figure 11: Schematic representation of a selection of different G-quadruplex topologies. Arrows indicate the strand polarities of the backbone strands and bases are coloured to represent the glycosidic bond angle as shown in the key. A) All-parallel quadruplexes have all backbone strands aligned in the same direction, as a result, all of the bond angles are the same, in this case they are shown as all *anti*, although all-*syn* parallel G4 are also possible. **B-C)** Anti-parallel G4 with two strands running in each direction have an equal split of bond angles within each tetrad. This means that within each tetrad, two guanines are in the *anti*-conformation and two in the *syn*-conformation. **D)** The final type of G4 topology is the hybrid-type antiparallel, also known as a 3+1 quadruplex, where a single strand is oriented in the opposite direction to the remaining three. In this case, within each tetrad 3 guanines adopt the same conformation (either *anti* or *syn*) and the other adopts the opposite conformation. Figure reproduced from Malgowska *et al.*⁴⁴

There are also a wide variety of conformations for the loops connecting the tetrads in intramolecular and bimolecular quadruplexes, which are affected by the sequence and topology of the quadruplex. The main three loop types are **(a)** diagonal loops connecting opposing anti-parallel strands and running diagonally across the terminal G-quartet, **(b)** edgewise loops (also called lateral loops) connecting adjacent antiparallel strands and **(c)** double-chain-reversal loops (also called propeller loops or side loops) connecting adjacent parallel strands.³⁸ There are also some common G4 types defined by their topology and loops, these include basket type G4 which have an adjacent parallel topology with a lateral-diagonal-lateral loop formation and propeller-type G4 which have an all-parallel topology with strands linked by propeller loops.^{39, 45} Several non-canonical G4 structures have also been uncovered, including long loop⁴⁶ and bulged structures.⁴⁷ A selection of the most common G4 topologies identified to date is shown in Figure 12.

The thermal stability of G4 structures can be influenced by a number of factors including ion concentration, loop length, number of G-tetrads, among others. Sequence variations such as smaller loops and longer G-tracts for instance were shown to be associated with an increase in thermal stability of G4.^{46, 48}

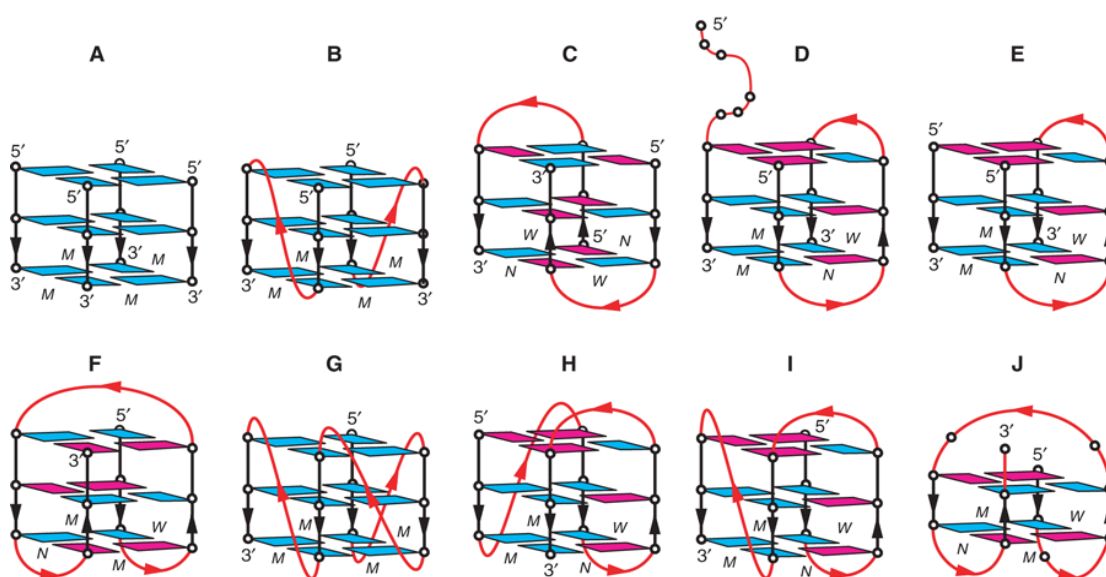


Figure 12: Schematic representation of human telomeric G4 topologies. **A)** Tetrameric parallel stranded G4,⁴⁰ **B)** Dimeric parallel G4 with propeller loops,⁴³ **C)** Dimeric anti-parallel G4 with edgewise/lateral loops,⁴³ **D)** Asymmetric dimeric (3 + 1) G-quadruplex with lateral loops,⁴⁹ **E)** Alternate asymmetric dimeric (3 + 1) G-quadruplex,⁴⁹ **F)** Basket-type G4 with lateral-diagonal-lateral loop structure,⁴⁵ **G)** Propeller-type all-parallel G4,³⁹ **H)** (3+1) G4 with propeller-lateral-lateral loop structure,⁵⁰ **I)** Alternative (3+1) G4 with lateral-lateral-propeller loop structure,⁵⁰ **J)** Two-quartet basket-type G4.⁵¹ *Anti* guanines are coloured cyan; *syn* guanines are coloured magenta; loops are coloured red. M, N and W represent medium, narrow and wide grooves, respectively. Figure reproduced from Phan.⁵²

1.2.2. Telomeric G4

Initial structural studies of G4 focused on those formed by the G-rich telomeric DNA, due to their involvement in the natural ageing process and their potential as therapeutic targets. The first atomic-resolution structures of G4 were co-published on March 12th, 1992. Smith and Feigon produced an NMR solution structure of the *Oxytricha nova* telomere repeats d(GGGGTTTTGGGG) and d(G₄T₄G₄T₄G₄T₄G₄).⁵³ They identified that both sequences formed symmetrical G4 structures with alternating parallel and antiparallel strands and *anti* and *syn* bond angles. At the same time Kang *et al.* released the first X-ray crystal structure of the d(GGGGTTTTGGGG) sequence and identified a heterodimeric quadruplex with the same alternating *anti* and *syn* bond angles.⁵⁴ Shortly after these studies, Aboul-ela *et al.* published the first structure of a parallel stranded quadruplex, formed with the d(TGGGG) sequence from *Tetrahymena*.⁵⁵ This study noted the similarity between the conformation of nucleotides in parallel-stranded quadruplexes and the conformation of the nucleotides in B-form DNA, namely all the glycosidic torsion angles in the structure were *anti*, as in right-handed helices. A higher resolution 0.95Å structure of the same sequence published by Phillips *et al.* in 1997 provided more clarity on the structural characteristics of parallel quadruplexes.⁵⁶ The higher resolution on the electron density maps from this study allowed researchers to confirm the stereochemistry and hydration pattern of parallel quadruplexes. The sugars were predominantly in the C2'-*endo* conformation as in B-

form DNA (Figure 13). Interestingly, some of the sugars in the G2 tetrads adopted a C3'-*exo* conformation, more commonly associated with A-form DNA, to avoid steric clashes with the hydroxyl groups. In addition, the high resolution allowed characterisation of the hydration pattern around the quadruplex. Interactions between water molecules and the grooves of the G4 are shown in Figure 14. The parallel topology of this quadruplex results in 4-fold symmetry and therefore relatively evenly sized grooves with similar hydration patterns.

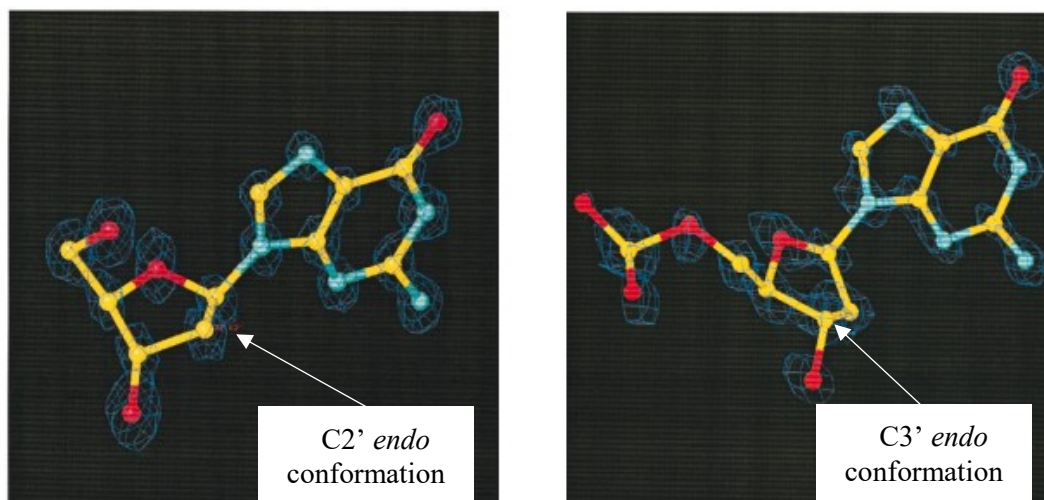


Figure 13: Representation of the two common sugar pucker conformations, C2'-endo (left-hand side) and C3'-endo (right-hand side). Oxygen-containing residues are highlighted in red and nitrogen-containing residues in teal. Figure reproduced from Phillips *et al.*⁵⁶

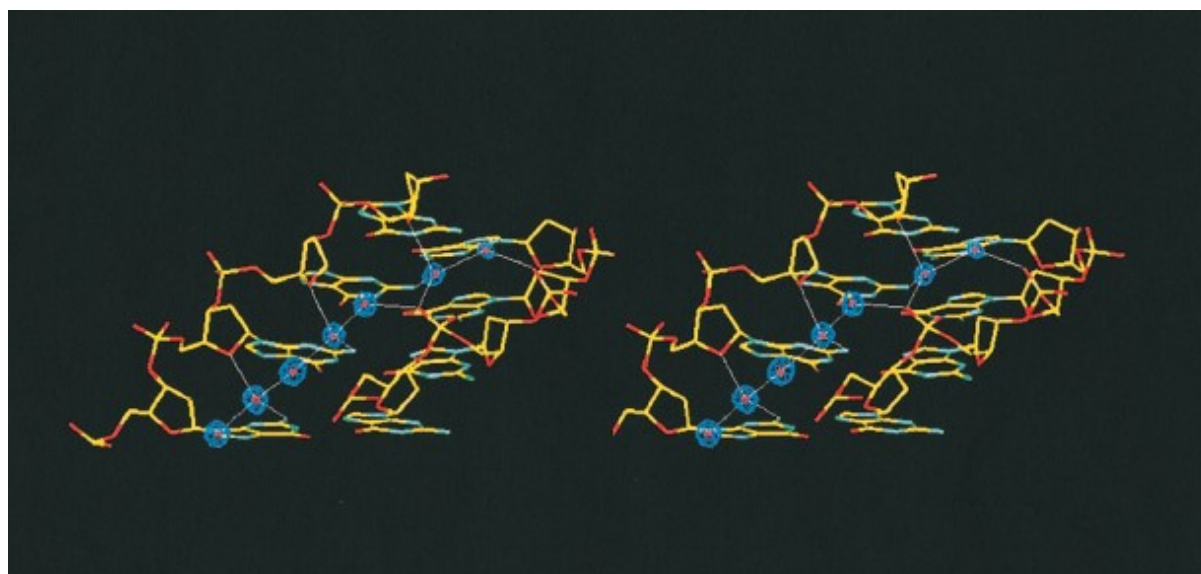


Figure 14: Interactions between water molecules and the grooves in the d(TGGGG) G4 crystal structure. Interactions are formed between water molecules and both the ribose group furanose oxygen, base N2 and N3 groups and phosphate hydroxyl groups. Water molecules are shown as red spheres with the electron density shown in blue. In the G4 structure the oxygen atoms are highlighted in red, and nitrogen in teal. Figure reproduced from Phillips *et al.*⁵⁶

The first structural study of a human G4 forming sequence was published in 1993, just a year after the first studies. An NMR solution structure of the human telomeric repeat d[AG₃(T₂AG₃)₃] G4 in the

presence of Na^+ was solved by Wang and Patel.⁴⁵ This revealed an intramolecular 3-tetrad G4 with *anti-anti-syn-syn* glycosidic bond angles and an adjacent parallel topology with the three TTA loops in sequential diagonal, edgewise and diagonal alignments (Figure 16A). A subsequent X-ray crystal structure from Parkinson *et al.*, almost ten years after the first human telomeric structure, showed a completely different structure, in this case a propeller-type, all parallel quadruplex with all *anti* glycosidic angles and double chain reversal loops (Figure 16B).³⁹ The structure also features a low electron density in the central channel, which was attributed to physiological monovalent cations such as K^+ and Na^+ , which intercalate between G-quartets and stabilise the structure through interactions with carbonyl oxygen atoms lining a central channel (Figure 15). The presence of monovalent cations is essential for G4 formation and the relative stability of G4 vary with different cations, in the order $\text{K}^+ > \text{Na}^+ \gg \text{Li}^+$.⁵⁷

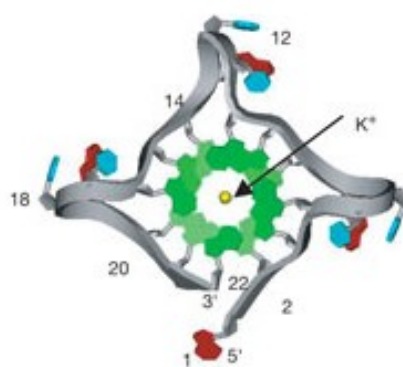


Figure 15: The central channel of the 22-mer telomeric G-quadruplex, viewed from the 3' end. The central channel forms an area of low electron density which was attributed to the presence of K^+ ions in the channel. The ion interacts with the negative carbonyls on the bases that line the channel. The backbone is shown as a grey ribbon, the central guanine bases are highlighted green and the loop thymine blue and adenines red. Figure reproduced from Parkinson *et al.*³⁹

The discrepancy between the NMR and X-ray crystal structures of the human telomeric repeat was clarified in a study from Phan and Patel in 2003 which provided clear evidence demonstrating that G4 can be polymorphic in solution.⁴³ The study used NMR and other biophysical and chemical methods to demonstrate that telomeric repeats formed both parallel and anti-parallel G4 in K^+ containing solution and that the two structures could co-exist and interconvert within solution. Schematic representations of these coexisting structures are shown in Figure 16C and 16D.

Since then, there has been a number of additional studies investigating DNA telomeric G-quadruplexes. One notable example was the study in 2005 from Zhang *et al.*, which was the first structural study that identified a 3+1 fold, where three tracts are parallel to each other with the final strand orientated in the anti-parallel direction.⁴⁹ In this case the sequence contained only three G-residues in each G-tract so the G4 was formed from a dimeric assembly. It was also the first structure demonstrating a hetero-dimeric quadruplex where the constituent strands are of significantly different lengths.

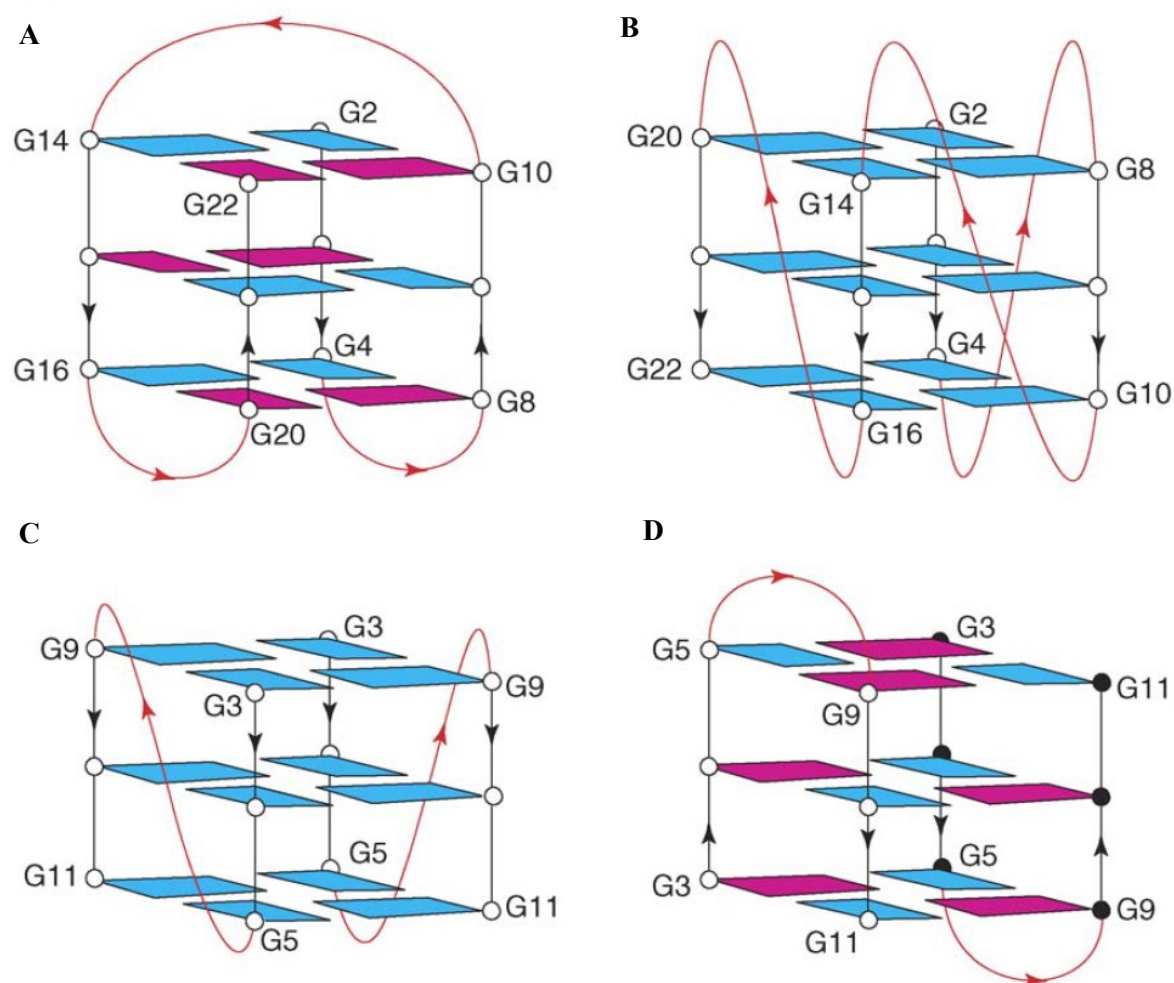


Figure 16: Alternate structures for human telomeric G4. The suggested structure for the d[AG₃(T₂AG₃)₃] G4 solved by **A)** NMR in Na⁺ by Wang and Patel⁴⁵ and **B)** through X-ray crystallography in K⁺ by Parkinson *et al.*³⁹ **C)** and **D)**, the two interconverting structures for d(TAGGGTTAGGGT) sequence in K⁺ solution solved by Phan and Patel.⁴³ Guanine residues are shown as squares coloured cyan and magenta to represent *anti* and *syn* bond angles respectively, loops are coloured red. Figure reproduced from Phan *et al.*⁵⁸

1.2.3. C-MYC G4

In the last decade and a half, the focus of G-quadruplex research has expanded out from focusing almost exclusively on telomeric quadruplexes to encompass G4 in numerous other genes. A particularly well-studied G4 has been that found in the *c-myc* promoter region. A number of insights into more general characteristics of G4 have come through study of this structure. In 2004, Phan *et al.* demonstrated that two sequences containing contiguous G-tracts from the *c-myc* promoter region (*myc*-2345, *myc*-1245) form a propeller-type intramolecular G4.⁵⁹ Notably, comparisons of the stability of the two sequences through melting studies suggested that the presence of shorter double chain reversal loops could increase G4 stability. A subsequent study by Seenisamy *et al.* showed that a four G-tract fragment from *c-myc* could form a range of parallel structures in solution, the ratio of which could be dynamically altered through addition of the G4 binding ligand TMPyP4.⁶⁰ This had significant implications for targeting G4, as well as providing a suggestion for how G4 formation may be modulated in cells.

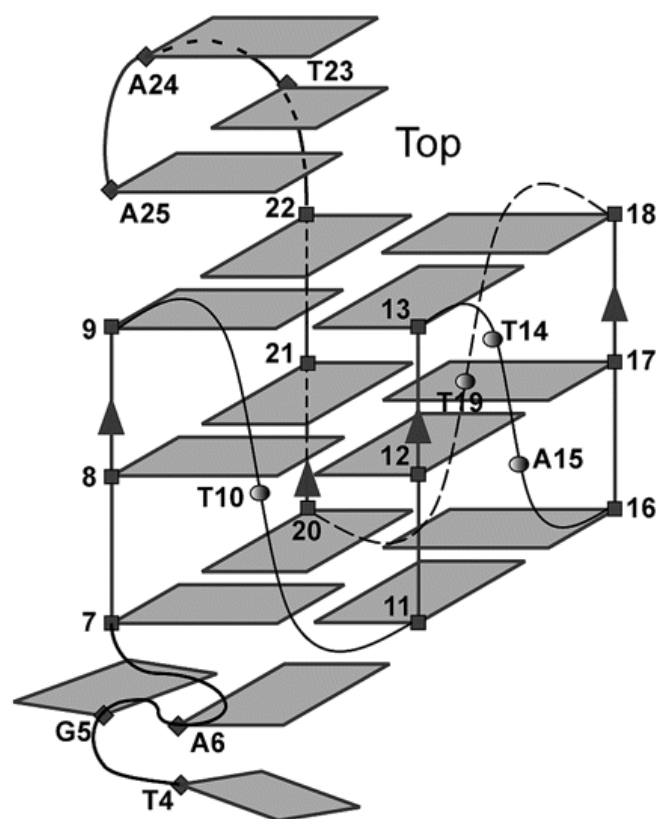


Figure 17: Proposed structure for *c-MYC* G4. The structure contains three stable, single nucleotide propeller loops. The 3' and 5' flanking regions are shown interacting with the terminal quartets through stacking interactions. Figure reproduced from Ambrus *et al.*⁶¹

In 2005 the first high resolution structure of the *c-MYC* G4 was published which provided more clarity on the topology and stability of the structure. The sequence forms an intramolecular parallel quadruplex with three stable loop regions.⁶¹ The stability of the one-nucleotide loops was consistent with the previous work from Phan *et al.* which identified that typically shorter loop regions correlate with greater G4 stability. The 5' and 3' flanking sequences fold back upon themselves to cap the terminal tetrads (Figure 17). This capping of the quadruplex with flanking nucleotides helps to explain the relatively high melting temperature of the quadruplex (>85°C).

A more recent study by Kim *et al.* has provided additional clarity on the role of cations in the stability of G4 structures. Through mutagenesis and kinetic studies, they demonstrated that previously identified stabilising cations, i.e. K^+ and Na^+ act as specifically bound ligands in the central channel of the quadruplex, stabilising the overall structure.⁶² However certain counter-ions, i.e. Cs^+ , which bind in a non-specific manner to the nucleotide backbone can have either a strong destabilising or stabilising impact on the structure. In the case of the MYC22-G14T mutant, the Cs^+ ions were found to stabilise the unfolded form of the sequence, reducing the melting temperature of the G4.

1.2.4. BCL-2 G4

Another well-studied G4 is found in the *BCL-2* promoter region. The Bcl-2 protein is a mitochondrial membrane protein that is an integral part of the apoptosis signalling pathway and is over-expressed in a range of human diseases, acting to increase cell survival.⁶³ In 2006 Dai *et al.* published an NMR solution structure of the *BCL-2* midG4, which forms a mixed parallel/antiparallel structure (Figure 18).^{64, 65} Agrawal *et al.* 2014 found a more stable alternative parallel quadruplex in the BCL-2 promoter. Interestingly this quadruplex includes an unusually long 13-nt loop, which would typically have been thought to reduce stability.⁶⁶ The 13-nt loop is the central loop in the sequence, with the other two loops both being 1-nt loops, as shown in Figure 18. This is notable as it is a fairly common arrangement in promoter G4.

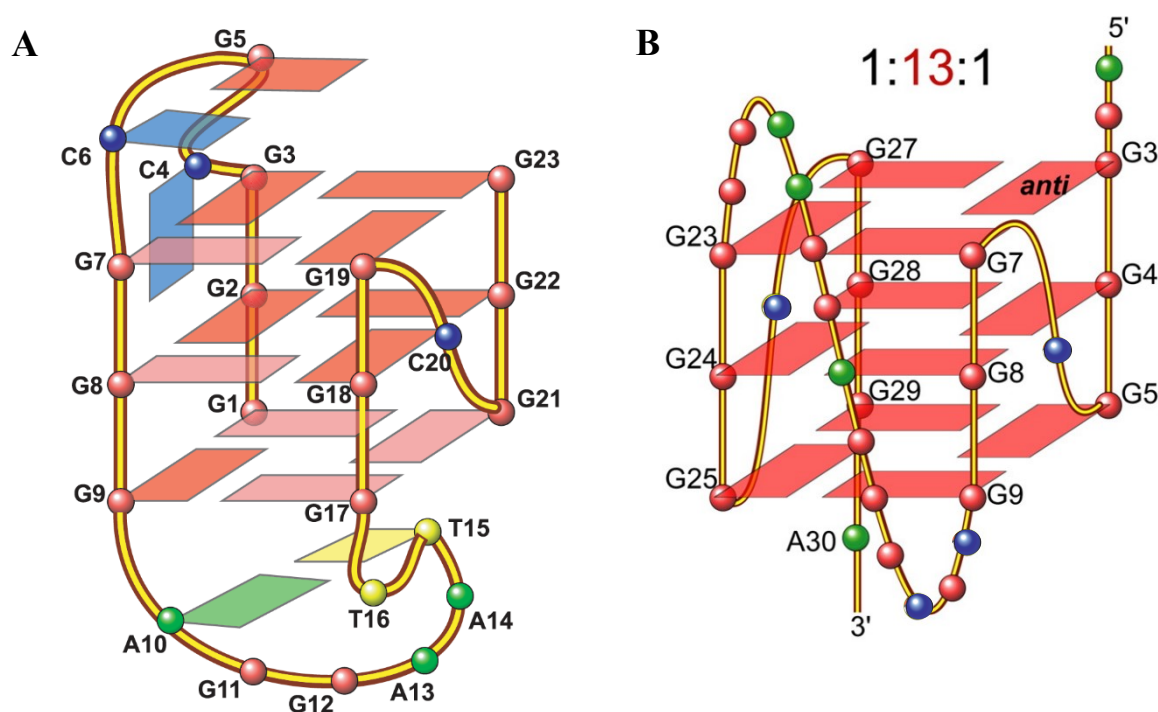


Figure 18: Proposed structures for *BCL-2* (A) MidG4 and (B) 1245G4. A) The *BCL-2* MidG4 structure which contains a one-nucleotide propeller-loop and two longer lateral loops which interact with the body of the G4 to stabilise the overall structure. (G=red, A=green, C=blue, T=yellow). Figure reproduced from Dai *et al.*⁶⁴ B) The *BCL-2* 1245 G4 structure contains a 5' and 3' one-nucleotide propeller loop as well as a significantly longer 13-nucleotide propeller mid-loop. (G=red, A=green, C=blue). Figure reproduced from Agrawal *et al.*⁶⁶

1.2.5. VEGF G4

In 2005 Sun *et al.* demonstrated that a G-rich sequence in the proximal promoter of the *VEGF* gene can form a G4 structure and subsequently showed that it could regulate transcription of the gene *in vivo*.⁶⁷ More recently, further investigation of the *VEGF* G4 by Fleming *et al.* has demonstrated a potential role for the fifth G-tracts that are often found next to the putative quadruplex sequence (PQS), shown in Figure 19.⁶⁸ Oxidative stress can increase instability in the genome, an effect which is particularly

notable in cancers. Reactive oxygen species can oxidise guanosines that form part of G4 structures preventing formation of the G4. However, Fleming and co-workers showed that the 5th G-tract can substitute in for the damaged G-tract, retaining G4 structure. This effect is illustrated in Figure 19.

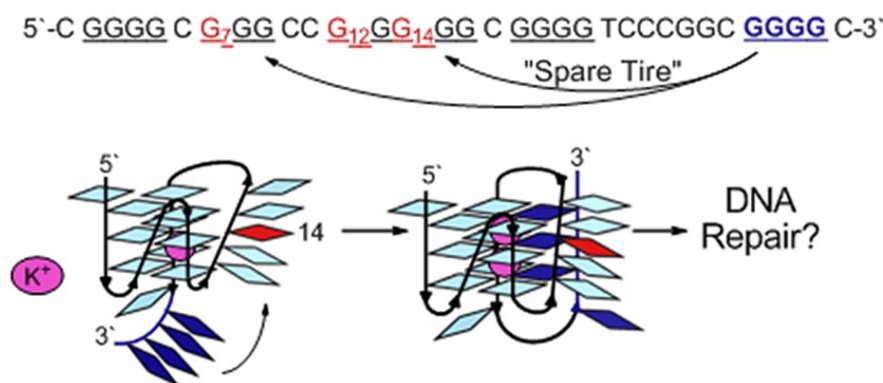


Figure 19: The “spare tire” mechanism to repair G4 following damage from reactive oxygen species. The 5th G-tract (coloured purple) can substitute for the G-tract containing a damaged guanine (red), this keeps the G4 structure intact until it can be repaired by the DNA repair machinery. The metal cation, in this case K⁺, is shown in pink, undamaged guanosines in the original four G-tracts are shown in teal. Figure reproduced from Fleming *et al.*⁶⁸

1.2.6. Limitations of structural studies

At this stage it is important to note that sources of atomic level 3D structural information about G4 are currently limited to X-ray crystallography and NMR spectroscopy. Both of these methods can provide a lot of useful structural information but they also have some limitations. In order to achieve sufficient resolution to provide accurate and detailed structure, both methods require that the quadruplex be formed from short nucleic acid strands. As a result, current structural information comes from synthetic sequences and do not represent the PQS in their full *in vivo* sequence context. In addition, both techniques are limited to being carried out *in vitro*, therefore the quadruplexes are not forming in a true *in vivo* environment. X-ray crystallography in particular requires specific environmental conditions, including pH and ion concentrations to allow for crystallisation. The crystallisation itself is another confounding factor, as the crystal structure may artificially stabilise structures that would not be able to form independently. NMR spectroscopy has the advantage that it allows interrogation of G4 structures under near physiological conditions. However, the samples still require to be pure and at high DNA strand concentrations (generally over 1 mM). In a cellular environment, nucleic acids are likely to be at significantly lower concentrations, which could impact on the overall stability of the structures. Furthermore, molecular crowding in a cellular environment can also have an impact on nucleic acid secondary structures. Interactions with proteins and other factors in the cell may impact the stability of G4 *in vivo* which would not be represented in structural studies. There are additional intracellular changes in cells (i.e. pH changes and ion concentration gradients) which will affect the stability of G4 and other secondary nucleic acid structures. Further development of tools to investigate G4 in its native environment are crucial to prove their *in vivo* stability and true structural conformation.

1.3. RNA G4 structures

Compared to DNA G4, structural studies of RNA G4 have been relatively limited, with only 18 structures added to the PDB to date. However, these studies have highlighted a number of key features that differentiate RNA G4 from their DNA counterparts. In particular, it is notable that all RNA G4 structures are parallel stranded and relatively homogenous when compared to the highly polymorphic DNA quadruplexes. In addition, biophysical studies have demonstrated that RNA G4 are generally more thermodynamically stable than their DNA counterparts and as such, RNA G4 are generally thought to be more likely to form *in vivo* than DNA G4.^{69, 70} This is partially explained by the fact that RNA is typically single stranded, so quadruplex formation is not competing with hybridisation to a complementary strand as is the case in double-helical DNA. Investigation of atomic level structures of RNA G-quadruplexes have given researchers further structural insight into differences in stability between DNA and RNA G4.



Figure 20: NMR solution structure of r(UGGGGU) G-quadruplex (A) and representation of uridine-tetrad (U-tetrad) structure (B). In the r(UGGGGU) quadruplex (A), the G-tetrads are bookended by a pair of U-tetrads, which are formed from the uracil residues at either end of the sequences. U-tetrads are formed through a hydrogen bonding network between the N₃H imino hydrogen atoms of one uridine and the O4 carbonyl oxygen atom of another uridine (B). In (A) guanosine residues are represented in green and uridine residues are represented in purple. Figure reproduced from Malgowska *et al.*⁴⁴

1.3.1. r(UGGGGU) G4

The first atomic resolution structure of an RNA G-quadruplex was published in 1992 by Cheong and Moore, shortly after the first DNA G4 structures were reported.⁷¹ In the presence of 50 mM K⁺, r(UGGGGU) strands were investigated using NMR spectroscopy and shown to form a tetra-molecular, parallel stranded quadruplex (Figure 20). The guanosine residues in the structure were confined to the *anti*-conformation and the backbone strands aligned in a helical formation. In addition, the terminal uridine residues were involved in the formation of a U-tetrad. The extra hydroxyl group in RNA also constrains the conformation of RNA G4 by restricting the sugar pucker and rotation of the glycosidic bond linking the guanine bases and the sugar backbone (Figure 21). Steric hindrance of the hydroxyl group by the phosphate backbone restricts the sugar pucker to the C3'-*endo* conformation. This in turn prevents the base from attaining the *syn*-conformation required to form an anti-parallel quadruplex.

Instead the bases in RNA G4 are confined to the *anti*-conformation and the overall structure restricted to all-parallel topologies.⁷²

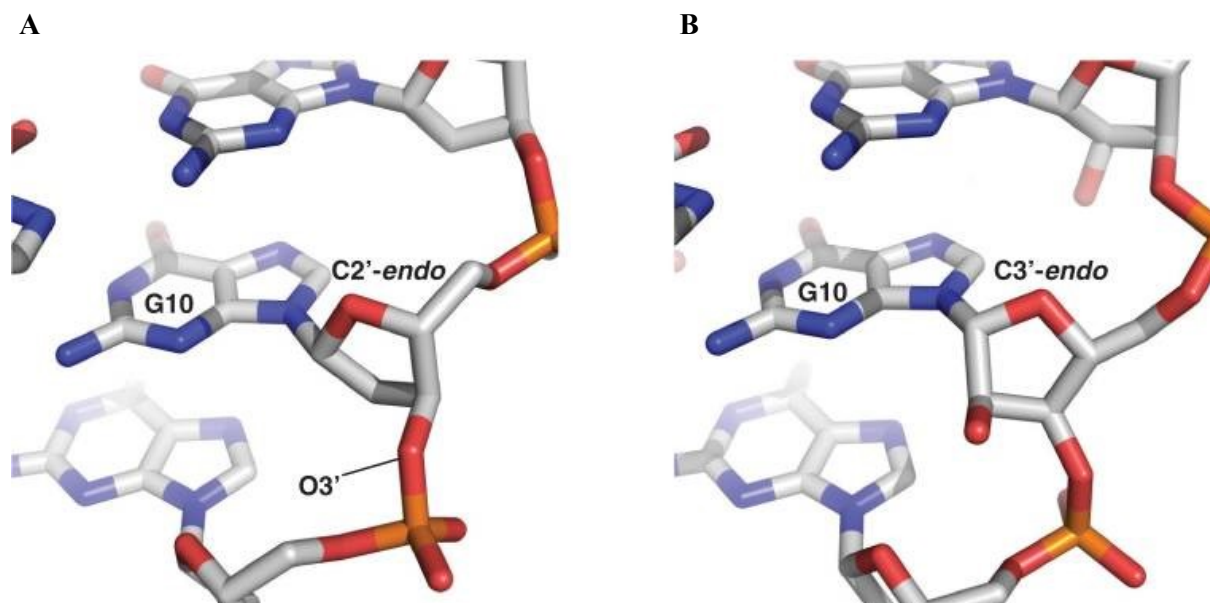


Figure 21: Differences in sugar pucker and dihedral angle of telomeric DNA (A) and RNA G4 (B). A) In telomeric DNA G4 the sugar adopts a C2'-endo conformation, which is possible because the deoxyribose doesn't clash with the O3'. B) In RNA G4 the hydroxyl group on the ribose sugar would clash with the O3' if it adopted C2'-endo conformation, instead the sugar adopts a C3'-endo conformation and therefore the dihedral angle of the guanine base is restricted to the *anti* conformation. Nitrogen atoms are labelled blue and oxygen atoms labelled red. Figure adapted from Collie *et al.*⁷³

The r(UGGGGU) quadruplex has been investigated in further structural studies which have provided additional clarity on the factors that influence RNA G4 conformations. In 2001, Deng *et al* reported a crystal structure for the r(UGGGGU) quadruplex with an Sr^+ cation.⁷⁴ As in the NMR structure, the oligonucleotides formed a parallel, tetra-molecular quadruplex with the guanosine bond angle in the *anti*-conformation. However, rather than forming a monomeric structure, the G4 were observed in an intertwined 5'-5' dimer structure. This illustrates one of the potential limitations of X-ray crystallography, which is that the crowded conditions in a crystal structure tend to favour intra-molecular interactions, increasing the prevalence of dimers and other higher-order structures.

A further study by Fyfe *et al.* in 2015 produced another X-ray structure of the r(UGGGGU) quadruplex in the presence of Sr^+ , supported by SAXS analysis, with the aim of further understanding the structural stability of RNA quadruplexes when compared to DNA quadruplexes.⁷⁵ A previous study by Zhang *et al* demonstrated that the 2'-hydroxyl group acts as a scaffold for ordered water molecules around the G4 groove, stabilising the C3'-endo pucker of the sugar backbone and locking it in the *anti*-conformation. This highly ordered hydration pattern provides a significant enthalpic contribution when compared to the hydration pattern of similar parallel DNA G4.⁷⁶ Similar hydration patterns were observed in the structure published by Fyfe *et al.* The r(UGGGGU) crystal structure was formed from octameric assemblies of the tetrameric G4 structures (Figure 22A). Water molecules in crystal structures

are typically relatively mobile and have short residency periods, which prevent them from being visualised through X-ray diffraction. The presence of visible water binding sites shows that water molecules are undergoing rapid exchange at a single site, suggesting a more significant structural role for the interaction at that site. The absence of the 2'-hydroxyl group on DNA oligonucleotides reduces the stability of the water lattice around DNA G4, which provides a partial explanation for their lower thermal stability.

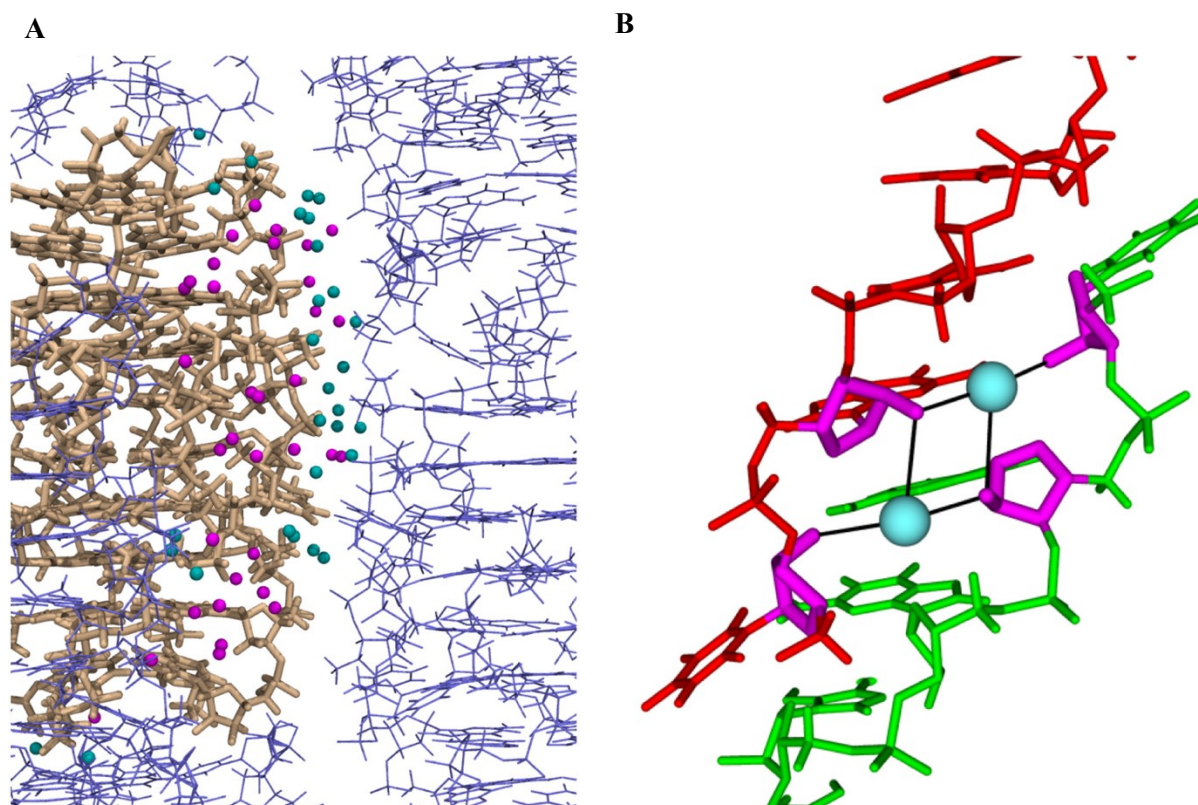


Figure 22: X-ray crystal structure for r(UGGGGU) octamer, highlighting interactions with external water molecules. **A)** The r(UGGGGU) octamer is shown in thick yellow lines. Adjacent octamers that interact with the central octamer via water linkages are shown in thin blue lines. The water molecules themselves are coloured according to the number of octamers with which they interact. Water molecules which interact with one octamer only are coloured magenta, those which interact with multiple octamers are coloured teal. **B)** Proposed hydrogen bonding pattern between surface water molecules and the O2' residues on the r(UGGGGU) quadruplex backbone. The ribose sugars involved in the interaction are coloured purple and water molecules blue, with the proposed hydrogen bond shown with a black line. The two nucleic acid strands are coloured green and red. Figure reproduced from Fyfe *et al.*⁷⁵

1.3.2. TERRA G4

Another RNA G-quadruplex that has been extensively studied in recent years is formed in long non-coding RNAs, known as TERRA RNAs, which are transcribed from human telomeric DNA. The telomeric DNA repeats were first shown to be transcribed to produce UUAGGG RNA repeats in 2007 by Azzalin and co-workers.⁷⁷ Shortly after, Xu *et al.* produced the first NMR solution structure demonstrating the potential for the TERRA sequence, in this case a double copy of the repeated motif r(UAGGGUUAGGGU), to form G-quadruplex in the presence of sodium ions.⁷⁸ This was swiftly followed by the first high resolution NMR structure of the TERRA G4 by Martadinata and Phan in

2009, in this instance in potassium solution.⁷⁹ Despite being conducted in different ionic solutions, the two structural studies were broadly in agreement, showing a propeller-type parallel stranded G4 formed with all guanosine in the *anti*-conformation, medium grooves and two double chain reversal UUA loops. The relative structural homogeneity of these TERRA quadruplexes is notable in comparison to the DNA telomeric quadruplexes, which formed a range of topologies in the presence of both potassium and sodium. This structural homogeneity was attributed to the presence of the 2'-hydroxyl group constraining the bond angle of the guanosine.

Subsequent X-ray crystallography studies of the TERRA G4 by Collie *et al.* in 2010 and 2011 provided further insight into their unexpectedly high thermal stability.^{73, 80} The first structure, r(^{Br}UAGGGUAGGGU) was a modified variant of the fragment used by Martadinata and Phan in their initial high-resolution structure. This structure also included several of the distinctive features identified in the initial study, i.e. the UUA double reversal loops. However, in contrast to the monomeric quadruplex observed in the NMR structures, the G4 was here shown to be a dimer. The ribose C2' hydroxyl group was shown to facilitate formation of an extended hydrogen bonding network including phosphate and oxygen backbone atoms, O4' sugar oxygen and the N2 acceptor groups on the quadruplex-forming guanines. The role of the C2'-hydroxyl group has been further verified in a study by Sacca *et al.* where systematic replacement of the hydroxyl group with different chemical analogues was shown to result in de-stabilisation of the quadruplex.⁶⁹

1.3.3. Other RNA G4

Further structural studies of RNA G4 are listed in Table 1, and for a more comprehensive review for their contribution towards our understanding of the RNA G4 topology see this excellent paper from Malgowska *et al.* in 2016.⁴⁴

Table 1: Published RNA G4 structures

Sequence	PDB ID	Topology	Technique	Reference
r(UGGGGU)	1RAU	Parallel	NMR	Cheong and Moore 1992 ⁷¹
r(UGGGGU)	1J8G	Parallel	X-ray	Deng <i>et al.</i> 2001 ⁷⁴
r(UGGGGU)	4RKV	Parallel	X-ray	Fyfe <i>et al.</i> 2015 ⁷⁵
	4RJ1			
	4RNE			
I(TGGGT)	1S9L	Parallel	NMR	Randazzo <i>et al.</i> 2004 ⁸¹

Sequence	PDB ID	Topology	Technique	Reference
I(TGGGT)	4LOA	Parallel	X-ray	Russo Krauss <i>et al.</i> 2014 ⁸²
r(UAGGGUUAGGGU)	2KPB	Parallel	NMR	Martadinata and Phan 2009 ⁷⁹
r(GGGUUAGGGU)	2M18	Parallel	NMR	Martadinata and Phan 2013 ⁸³
r(UAGGGUUAGGGU)	3MIJ	Parallel	X-ray	Collie <i>et al.</i> 2011 ⁸⁰
r(^{Br} UAGGGUUAGGGU)	3IBK	Parallel	X-ray	Collie <i>et al.</i> 2010 ⁷³
r(GGAGGUUUUGGAGG)	1MY9	Parallel	NMR	Liu <i>et al.</i> 2002 ⁸⁴
r(GGAGGAGGAGGA)	2RQJ	Parallel	NMR	Mashima <i>et al.</i> 2009 ⁸⁵
d(^{Br} U)r(GAGGU)	1J6S	Parallel	X-ray	Pan <i>et al.</i> 2003 ⁸⁶
d(^{Br} U)r(GIGGU)	2GRB	Parallel	X-ray	Pan <i>et al.</i> 2006b ⁸⁷
r(U)d(^{Br} G)r(AGGU)	1MDG	Parallel	X-ray	Pan <i>et al.</i> 2003a ⁸⁸
r(U)d(^{Br} G)r(GUGU)	2AWE	Parallel	X-ray	Pan <i>et al.</i> 2006a ⁸⁹
r(U)d(^{Br} G)r(UGGU)	1P79	Parallel	X-ray	Pan <i>et al.</i> 2003b ⁹⁰

1.4. *In vitro* functional evidence for DNA G4

1.4.1. Regulating telomere extension

Zahler *et al.* were first to demonstrate a correlation between G4 formation in the telomeric region of *Oxytricha nova* DNA and telomerase function *in vivo*.⁹¹ Formation of telomeric DNA G4 has also been shown to be regulated by telomere-end binding proteins, suggesting they play a significant biological role.⁹² Elongation of telomeric DNA by the enzyme telomerase requires hybridisation of the telomere single stranded 3'-overhang to the RNA template of telomerase. Trapping this G-rich single-stranded overhang into a quadruplex conformation (e.g. with small molecules) was suggested to prevent this key hybridisation step, thus preventing elongation of telomeric DNA. Considering that telomerase is found active in over 85% of tumour cells (while mostly inactive in somatic cells) and is responsible for their relative immortality, this has led to ever increasing interest in development of G4 stabilising ligands as an anti-cancer therapy (see section 1.6).⁹³ A potential limitation to the previously described strategy was exposed by a recent high profile paper demonstrating that parallel telomeric G4 act as a substrate for telomerase and that telomerase can catalyse their extension.⁹⁴ However, telomerase was also shown to localise to the same G4 structures, implying that they do have a biological function. It has also been suggested that telomeric DNA G-quartet structures are involved in mediation of chromosomal synapsis during meiosis.⁹⁵

1.4.2. Immunoglobulin class switching

The discovery of a novel G4-containing G-loop structure in immunoglobulin (Ig) switch regions in mature B-cells suggests a functional role for G4 in Ig class switch recombination (CSR).⁹⁶ The process of CSR involves switching the constant region of the Ig heavy chain to change the class of the antibody (i.e. from IgG to IgM) while the variable heavy chain region involved in antigen recognition remains the same, allowing modulation of the immune system response to recognition of a specific antigen.⁹⁷ This process is regulated by G-rich switch regions (S-regions) which act as binding sites for recombinant factors. Several of these factors have been shown to bind to S-region G4 and G-loops, including human exonuclease I and the mismatch repair heterodimer MutS α , suggesting that G4 recognition may be involved in the regulation of CSR.^{98, 99} A recent study also demonstrated stabilisation of S-region G4 by naphthalene and quinoline-based ligands with the future aim of developing ligands to restore functional CSR in selected immunodeficiency-related diseases.¹⁰⁰

1.4.3. Epigenetic regulation

G4 have also been suggested to play key roles in epigenetic regulation of gene expression through regulation of DNA methylation and nucleosome occupancy.^{101, 102} A genome wide study showed that CpGs within G4 motifs showed a much higher occurrence of hypo-methylation compared to regions devoid of G4 motifs, where a higher amount of hyper-methylation was observed.¹⁰³ Hypo-methylation and chromosomal instability are hallmarks of neoplastic transformations.¹⁰⁴ De *et al.* showed that regions of chromosomal instability were heavily hypo-methylated and contained a large number of G4 motifs, and that these regions were also strongly correlated with hotspots of somatic copy number alterations.¹⁰² This suggests roles for G4 in regulating chromosomal stability, possibly through known interactions with transcription factors, including Sp1, which regulate DNA methylation.

1.4.4. Regulation of Transcription

There is a high concentration of G4 motifs in the promoter regions of genes. Indeed, computational genomic analysis has shown that 50% of genes contain one or more G4 motifs less than 1000 nucleotides (nt) upstream of the transcription start sequence.¹⁰⁵ One of the best studied examples of the role of G4 motifs in transcription regulation is that of the *c-myc* gene, the product of which is a transcription factor involved in regulation of cell proliferation. Increased MYC expression has been observed in 80% of human cancer cells.^{106, 107} Nuclease hypersensitive element III (*NHE III*) controls >80% of *MYC* transcription and contains a G4 motif that has been shown to form a quadruplex structure

in vitro.¹⁰⁸ A comparison of gene expression under control of wild-type *NHE III* and mutated *NHE III* showed that the presence of the *c-myc* G4 motif acted to repress translation.¹⁰⁹ The exact mechanism for G4-mediated repression of transcription is not yet clear but it is believed to be through interaction with proteins such as nucleolin.¹¹⁰

1.5. *In vitro* functional evidence for RNA G-Quadruplex

Although the first indications that RNA G4 may have a functional role in human biology came in 1994 when a G4 was identified in Insulin-like growth factor II mRNA,¹¹¹ their biological functions and potential as therapeutic targets have only been explored relatively recently. The discovery in 2001 that Fragile-X mental retardation protein targeted G-quartets in mRNA responsible for loss of neuronal function in Fragile-X mental retardation syndrome prompted a flurry of interest, and in the last decade there has been substantial progress in understanding the role of RNA G4.¹¹²

Like their DNA counterparts, RNA G4 are mainly concentrated in several key regions of the human genome with distinct functional roles. There is on-going work mapping RNA G4 in the functionally relevant human transcripts using G4-mediated reverse transcriptase stalling integrated with hybridization based ligation-mediated PCR.¹¹³ If successfully completed, this work will provide a significant contribution towards the complete mapping of G4 prevalence within the human transcriptome. However, RNA G4 have already been linked with a number of different functional roles, a selection of which are presented below. (See Millevoi *et al.*¹¹⁴ and Agarwala *et al.*¹¹⁵ for recent reviews)

1.5.1. Regulation of Translation

A number of techniques have been developed to investigate the impact of G4 formation on translation. In particular, *in vitro* translation assays have been developed where 5'UTR containing a target G4 is added to a reporter gene. Expression of the reporter gene can be used to measure the impact of the G4 by comparing expression levels for mRNA containing the G4 and mRNA with a mutated sequence which prevents formation of the G-quadruplex, typically through replacing a G-tract with adenosines. In 2007 Kumari *et al.* provided the first evidence that formation of an RNA G4 structure can regulate translation *in vitro*.¹¹⁶ Biophysical techniques showed that a PQS in the 5'UTR of the human *NRAS* mRNA could fold into a highly stable G4 *in vitro*. Then through use of a cell-free translation reporter assay it was demonstrated that the presence of this quadruplex inhibits translation by ~80% compared to *NRAS* 5'UTR sequences lacking the PQS or with mutations preventing quadruplex formation. In a

separate study, inhibition of translation by a G4 structure in the 5'UTR of the mRNA coding for the human Zic-1 protein was confirmed in eukaryotic cells using a dual-luciferase assay.¹¹⁷ Subsequent studies have identified further quadruplex structures that act as repressors of translation, along with quadruplex structures located in internal ribosome entry sites (IRES) that promote translation.^{118, 119} The mechanism by which 5'UTR G4 repress translation is not yet confirmed, but it is generally thought to occur through blocking of initiation factor binding, resulting in dissociation of the ribosome complex (Figure 23), rather than by steric hindrance. This conclusion is supported by studies demonstrating that the translational effect of G4 is dependent on their proximity to the 5'-cap and that 5'UTR G4 forming sequences identified *in silico* exhibit a positional bias toward the 5'-end.^{120, 121} G4 in the coding regions have also been shown to affect translation. Constructs containing G4 in the open reading frame (ORF) can stall translation and cause an unusual ribosomal frameshift.^{122, 123}

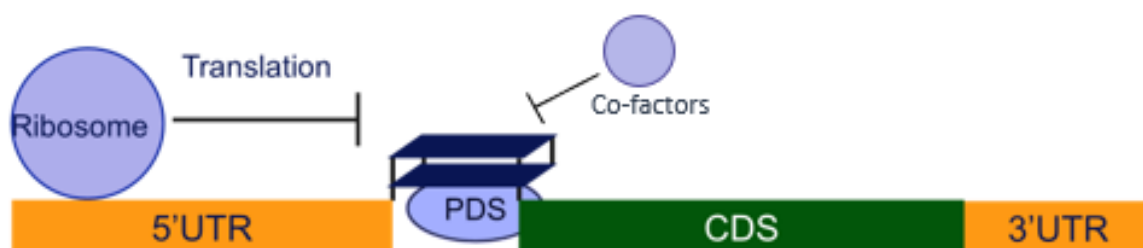


Figure 23: Representation of translation inhibition by a G4. The ribosome is assembled along the 5'UTR of the target mRNA, formation of a G4 in this region has been suggested to either conceal binding sites and result in disassembly of the ribosome or displace the ribosome from the mRNA. In either case G4 formation, aided in this example by binding of the G4 binding ligand pyridostatin (PDS), prevents translation of the mRNA.

Arora and Suess were first to demonstrate that 3'UTR G4 could also act as translational repressors. They used a dual luciferase assay to show that a stable quadruplex in the proto-oncogene *PIM1* repressed translation.¹²⁴ Translation of several other proteins has been shown to be under the control of 3'UTR G4, most notably amyloid precursor protein (APP) which plays a role in regulating A β production. Disruption of a G4 in the *APP* 3'UTR was shown to result in APP overexpression and a subsequent increase in A β levels, which has been associated with the onset of Alzheimer's disease.¹²⁵

1.5.2. 3'-end processing and poly-adenylation

RNA G4 in 3'UTRs have been demonstrated to be *cis*-acting regulatory elements that play a role in the post-transcriptional regulation of mRNA, in particular through regulation of 3'-end processing and poly-adenylation. The process of mRNA 3'-end formation, which involves cleavage of the pre-mRNA 3'-ends and addition of a poly-A tail, is a vital part of pre-mRNA processing and is required to produce functional mRNA. DNA damage induces a protective cellular response which involves down-regulation of mRNA synthesis and inhibition of 3'-end processing. However, during the DNA damage response there is increased expression of specific proteins involved in DNA damage repair, cell cycle arrest and apoptosis. The transcription factor p53 is an important regulator of this response, acting in a sequence

specific manner to upregulate DNA damage response genes.¹²⁶ The dysregulation of the p53 pathway can compromise the DNA damage response and is associated with a large number of cancers. Researchers recently demonstrated that an interaction between heterogeneous nuclear ribonucleoprotein H/F (hnRNP H/F) and a G4 in the *p53* poly-adenylation (pA) signal is responsible for maintaining 3'-end processing of *p53* while cell-wide transcription is repressed during the DNA damage response.¹²⁷ The use of alternative poly-adenylation signals (APA) in mRNA 3'UTR induces mRNA shortening through deletions in the 3'UTR. Beaudoin and Perrault showed that 3'UTR G4 in low-density lipoprotein receptor-related protein 5 (*LRP5*) and FMR1 autosomal homolog 1 (*FXR1*) mRNA improved the efficiencies of alternative polyadenylation sites, leading to shorter transcripts and an increase in transcription. The 3'-region downstream of the *FXR1* APA also contains a number of miRNA binding sites, the G4-induced deletion of which was shown to alter miRNA regulation of *FXR1*.¹²⁸

1.5.3. Alternative Splicing

RNA G4 in mRNA introns have been shown to be involved in regulating alternative splicing mechanisms. The first example of this was a G4 that was shown to be essential for splicing of the *scrNI* exon.¹²⁹ Subsequent studies have identified genes of several proteins of physiological importance that contain G4 in their mRNA and undergo alternative splicing, including B-tropomyosin,¹³⁰ human telomerase reverse transcriptase (*hTERT*),¹³¹ Fragile X mental retardation 1 (*FMRI*),¹³² p53^{127, 133} and Paired box gene 9 (*PAX9*).¹³⁴ An example with particular clinical interest is that of β -secretase 1 (BACE-1), which catalyses the first step in the breakdown of APP to amyloid-beta protein (A β). It has two isoforms which are regulated through alternative splicing, one active form and a shorter, essentially inactive form. Disruption of a G4 containing sequence within exon 3 resulted in increased use of the alternative splice site and production of the inactive isoform of BACE-1, reducing A β production.¹³⁵ Computational studies of intronic G4 have demonstrated that a large number are located in alternatively spliced genes or near RNA processing sites, suggesting this may be a relatively widespread regulatory mechanism.¹³⁶

1.5.4. Mitochondrial transcription termination

The mitochondria in human cells produce primers which are utilised in the initiation of leading-strand DNA replication. G4 structures in the mitochondrial mRNA have been shown to define the 3'-end of these primers by causing site specific termination of translation.¹³⁷ The G4 plays a similar role to the 3'-hairpin loop involved in Rho-independent transcription termination in bacterial cells.

1.5.5. mRNA localisation

3'UTR G4 in the mRNA sequence of post-synaptic proteins, such as postsynaptic density protein 95 (PSD-95) and Ca21/calmodulin-dependent protein kinase IIa (CaMKIIa), are necessary and sufficient for transport and localisation to the dendrites. The impact of the PSD-95 G4 on localisation of mRNA to neurites is shown in Figure 24. Over 30% of well-known dendritically localised proteins contain a similar G4 consensus sequence in their mRNA, suggesting that this may be a relatively common mechanism.¹³⁸

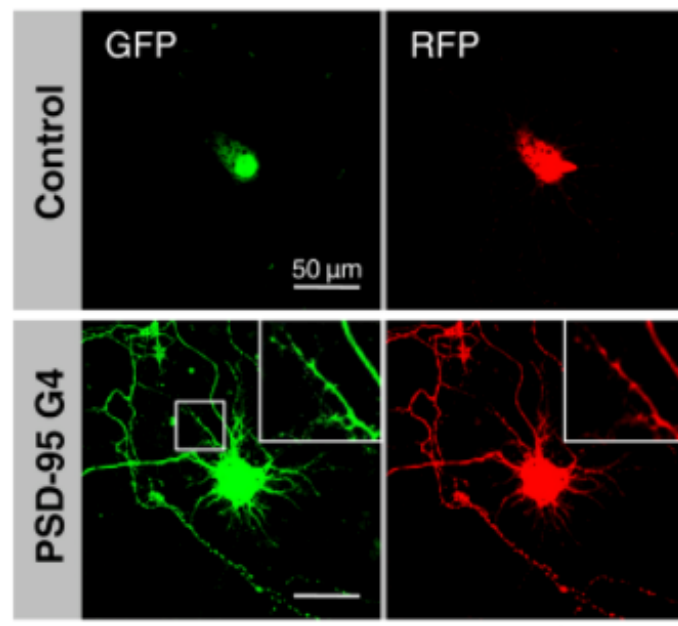


Figure 24: Localisation of mRNA by the native and mutated 3'UTR G4 from post-synaptic density protein 95 (PSD-95) in primary neuronal cells. The images show staining of the mRNA with antibody stains containing GFP (green) and RFP (red) respectively. The control mRNA doesn't contain the G4 sequence PQS and accumulation of the mRNA can be seen in the nucleus. In comparison, mRNA containing the PSD-95 G4 can be seen throughout the cell including in the neurites. Figure reproduced from Ishiguro *et al.*¹³⁹

1.5.6. Telomere maintenance and heterochromatin formation

Although previously considered to be transcriptionally silent, mammalian telomeres were shown in 2007 to transcribe telomeric-repeat containing RNA (TERRA) which consists of extended tandem repeats r(UUAGGG).^{77, 140} These G-rich repeat regions form stable propeller-type parallel G4 both *in vitro*^{73, 78} and *in vivo*,¹⁴¹ and stack together to form higher order structures.^{79, 142} TERRA RNA binds to telomere repeat factor 2 (TRF2), part of the Shelterin complex that recruits telomerase to telomere 3'-ends,¹⁴³ through formation of an intramolecular G4.¹⁴⁴ This suggests that the quadruplex structure is integral for localisation and regulatory function of TERRA. This interaction can occur alongside further interactions between TRF2 and DNA G4 and as part of a TERRA mediated ternary complex bringing together TRF2 and subunits of the origin recognition complex (ORC). This suggests that TERRA plays

a central role in regulating telomere structural maintenance and heterochromatin formation through mediating interactions between complexes, a conclusion supported by siRNA studies.¹⁴⁵ TERRA can also inhibit telomerase activity by acting as a competitive inhibitor to telomere DNA.^{146, 147}

1.6. Evidence for *in vivo* activity of RNA and DNA G4

1.6.1. Genomic studies

While it is still an unresolved question whether G4 have a significant functional role *in vivo*, there is a growing body of evidence that supports their importance. Initial computational/bioinformatics studies, using tools like Quadparser, had identified just over 370,000 putative quadruplex forming sequences (PQS) in the human genome.^{148, 149} The prevalence of PQS does imply a potential functional role but more significantly, PQS were found to be enriched within functionally significant regions of the genome. For example, 43% of human gene promoter regions were found to contain at least one PQS.^{48, 105} More recently, researchers in the Balasubramanian group developed G4-seq, a high throughput sequencing technique used to identify G4 structures in the human genome.¹⁵⁰ It combines elements of polymerase stop assays, used to identify G4 in a DNA template by detecting polymerase stalling,¹⁵¹ with Illumina next-generation sequencing¹⁵² to identify G4 structures in whole genomes. Researchers isolated and sequenced DNA from human primary B lymphocytes and identified 716,310 distinct G4. G4-seq results also showed a significant enrichment of G4 in oncogenes, tumour suppressors and somatic copy number alterations linked with cancer development.¹⁵³

The prevalence of PQS in functional regions of the genome suggests a functional role. If these sequences played no functional role in the genes, then it would be expected that they would undergo mutagenesis over time and would therefore vary between different species or even individuals. However, in 2005 Capra *et al.* ran a computational study of 6 yeast genomes and found that G4 sequences are highly evolutionarily conserved between species.¹⁵⁴ In addition, when analysed on a single nucleotide level, conservation was higher among nucleotides directly involved in G-quadruplex formation, suggesting that the evolutionary importance of the sequence is directly related to formation of a G4 structure.

1.6.2. *In vivo* imaging studies

Another key part of proving that G-quadruplexes are functional *in vivo* is demonstrating that PQS form G4 structures in cellular environments. In the last three years there have been significant breakthroughs with new technologies to visualise G4 *in vivo*. This is a vital step to prove that sequences shown to form

G4 *in vitro* over the past decades adopt the same conformation *in vivo*. The first evidence that G4 formed in cells came in 2001 in a study showing antibody staining of telomeric G4 in *Stylonychia lemnae* macronuclei.¹⁵⁵ More recent studies have used specific antibodies to visualise G4 in human cells, providing substantive evidence for their existence *in vivo*. The Balasubramanian group has developed the BG4 antibody which allowed specific visualisation of DNA¹⁵⁶ and RNA¹⁵⁷ G4 in human cells, (Figure 25). In a subsequent study, the Lansdorp group used the 1H6 antibody to visualise DNA G4 in both human and murine cells.¹⁵⁸

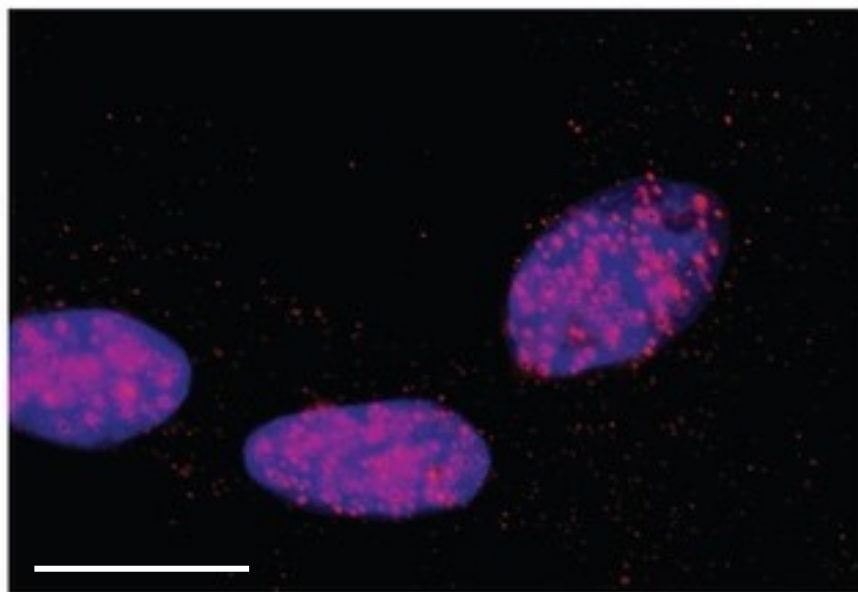


Figure 25: BG4 staining of SV40-transformed MRC5 fibroblasts. Nuclei are stained blue with DAPI and BG4 stained red through Alexafluor 594 fluorescence. Scale bar, 20 μm . The BG4 antibody staining is visible both within the nucleus, suggesting the presence of DNA G4 and in the cytoplasm of the cell suggesting the presence of RNA G4. Figure reproduced from Biffi *et al.*¹⁵⁷

However, a significant limitation of these antibody-based studies is the necessity for fixation and permeabilisation of the cells. To circumvent this limitation, cell-permeable fluorescent ligands have been developed that allow visualisation of G4 in live cells. In 2015 for example, the biomimetic ligand NaphthoTASQ was used to detect RNA G4 in the cytoplasm of live cells.¹⁵⁹ The ligand folds into a closed conformation upon specific binding to G4 resulting in a substantial increase in its fluorescent properties that can be detected using multiphoton microscopy. Moreover, a technique has recently been validated to visualise the formation of specific RNA G4 *in vivo* using bimolecular fluorescence complementation.¹⁶⁰ The quadruplex forming sequence of interest is expressed in a plasmid alongside an aptamer sequence. Ligands specific for the quadruplex and aptamer structures respectively are tagged with a split half of enhanced green fluorescent protein. Co-localisation of the ligands results in fusion of the two halves of green fluorescent protein and a fluorescent signal is produced, as shown in Figure 26. The current body of work demonstrating visualisation of G4 *in vivo* provides substantial evidence for the presence of these structures *in vivo*. However, it has not yet been conclusively proven that G4 formation was not induced by the ligands or antibodies used to visualise them.

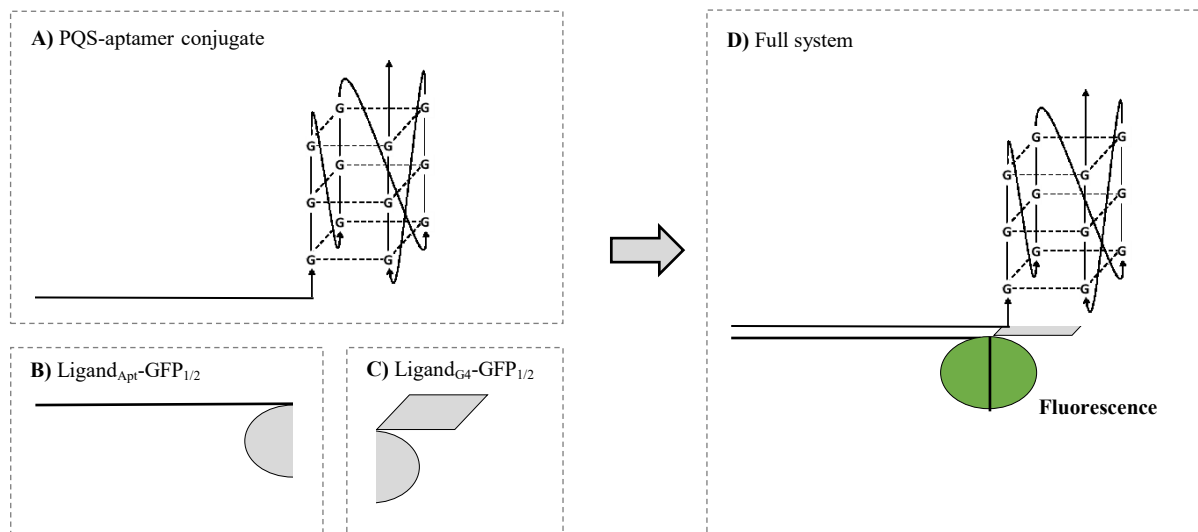


Figure 26: Mechanism of bimolecular fluorescence complementation. To test formation of a G4 from a PQS *in vivo*, three elements are introduced into cells, **A)** a conjugate of the PQS and an RNA aptamer (shown with a thick black line), **B)** The ligand for the aptamer (Ligand_{Apt}, shown with a thick black line) conjugated to one domain of the GFP protein and **C)** a G4 binding ligand (Ligand_{G4}) conjugated to the other domain of GFP. When the three elements are introduced into a cell and the PQS forms a G4 then **D)** the full system is constructed, where the two domains of the GFP are brought into close contact and can combine to produce fluorescence.

1.6.3. *In vivo* binding and functional studies

The most compelling evidence of G4 formation and functional significance would be evidence directly linking G4 formation to a functional result *in vivo*. As discussed in section 1.5, G4 have been linked to functional roles in many cellular processes including regulating transcription,¹⁰⁹ translation,¹¹⁶ telomere maintenance,¹⁶¹ RNA localisation,¹³⁸ epigenetic regulation^{102, 103, 162} and splicing.¹⁶³ However, the majority of functional experiments to date have been carried out *in vitro*, either in cell free systems, or in cell lines of primary cells in culture. These provide relatively strong evidence for G4 function but are only indirect evidence for an *in vivo* function. A number of studies have introduced mutations into genes for proteins shown to interact with G4 *in vitro* and investigated the subsequent impact on cellular processes. The major challenge with such experiments is proving that the functional impact of the mutation is mediated through disruption of the G4, rather than through an alternative effect of the mutation, i.e. disrupting a binding site for a protein or other co-factor.

In a study published by Cheung *et al.* in 2002, it was demonstrated that mutation of the gene for DOG-1 helicase results in widespread deletions at poly-guanine tracts, the suggested mechanism for the introduction of these deletions is shown in Figure 27.¹⁶⁴ The high prevalence of PQS among the deleted sites suggests that this effect is related to the formation of G4 by the G-rich sequences. From these results it was additionally suggested that DOG-1 is required to resolve G4 structures. A further study published in 2008 by Kruisselbrink *et al.* expanded on this work by studying transgenic animals lacking the *dog-1* gene.¹⁶⁵ The authors used unbiased aCGH analysis to demonstrate that the deletions in these

animals were restricted to PQS, providing strong evidence that G4 both form *in vivo* and can impact genetic replication. It is worth noting however, that it is not specifically proven in the experiments that G4 are formed by the PQS *in vivo*. Combining deletion of *dog-1* with visualisation of the G4 structures would provide stronger evidence that it is G4 formation that is the cause of the deletions.

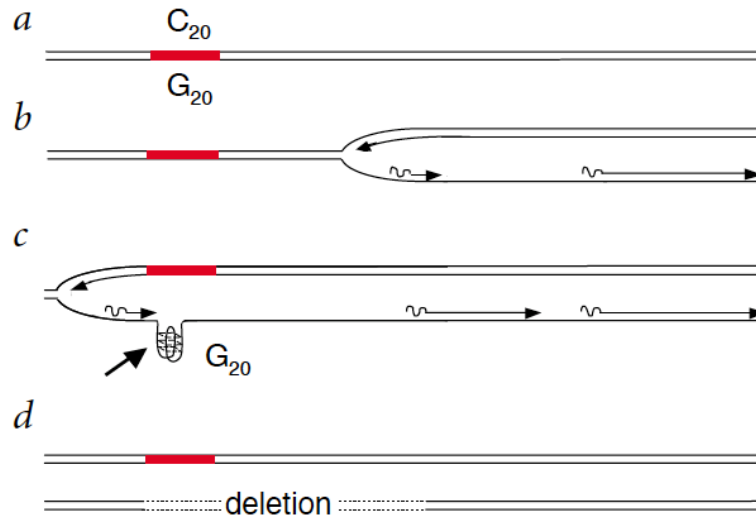


Figure 27: Proposed mechanism for introduction of deletions in *DOG-1* deficient cells. The replication fork travels along the double stranded DNA, with replication occurring along the lagging strand. The formation of G4 prevents replication and there is subsequently a deletion the size of the Okazaki fragment including the PQS. PQS highlighted in red. Figure reproduced from Cheung *et al.*¹⁶⁴

In 2011 researchers used chromatin immunoprecipitation to demonstrate that the Pif-1 helicase bound to G4 sites in the *S.cerevisae* genome.¹⁶⁶ The addition of G4-forming motifs slowed replication, and this effect was ameliorated through introduction of mutations in the G4 forming sequences. This provides significant evidence for the direct interaction between helicases and G4 *in vivo*. The fact that the effect of G4 sequences on helicase was eliminated through mutations that prevented G4 formation suggest that formation of a G4 structure is important for this effect, which was not clearly proven in the previous study by Cheung *et al.*

In addition, researchers have shown that important cellular proteins have strong affinity and selectivity for G4 structures, suggesting a significant role *in vivo*. In particular, it has been shown that nucleolin binds to the *c-myc* G4 and represses *c-myc* promoter activity.¹¹⁰ Nucleolin is a highly abundant nucleolar phosphoprotein which is involved in ribosome assembly and pre-rRNA transcription. If such an important protein has evolved to bind specifically to G4 structures, it suggests that they have a functional role *in vivo*.

1.7. Targeting G-quadruplex

1.7.1. Therapeutic potential of G4 targeting

Over the past ten years PQS have been identified in the RNA and DNA sequences of genes of clinical interest and have been suspected (and in some cases shown) to play a key role in the regulation of numerous disease processes. Fragile X syndrome is a genetic disorder that can result in developmental problems including learning disability.¹⁶⁷ The condition is caused by expansion of the CGG repeat in the 5'UTR of the *FMR1* gene, which can become hypermethylated, silencing the gene and resulting in the loss of the FMRP protein. This protein is responsible for regulating mRNA transport and protein synthesis at the synapses. Hyper-methylation is suggested to be caused by secondary structures forming in the expanded repeats. Fry and Loeb demonstrated that CGG repeats could form G4 in a cation solution.¹⁶⁸ Following this and further confirmatory studies,¹⁶⁹ the role of G4 in Fragile-X syndrome has been an on-going area of research.^{112, 170} In addition to Fragile-X syndrome, G4 have been linked with other neurological conditions, in particular Amyotrophic lateral sclerosis (ALS) and Frontotemporal dementia (FTD). Haeusler *et al.* demonstrated that the hexanucleotide repeat (GGGGCC)_n forms DNA and RNA G4 and promotes DNA:RNA hybrids (R-loops).¹⁷¹ These structures were then shown to bind to ribonucleoproteins and induce nucleolar stress and led researchers to suggest a mechanism by which the formation of G4 initiates molecular cascades leading to the ALS and FTD phenotypes. As previously mentioned, G4 have also been identified in numerous oncogenes and have been linked to a number of different cancer sub-types. In particular, G4 formed by G-rich telomeric repeats have been extensively investigated.¹⁷² In addition, G4 have also been linked with viruses, including Epstein Barr Virus¹⁷³ and HIV.^{174, 175} Amrane *et al.* identified G4 in the promoter region of the HIV provirus which provide a potential drug target to inhibit expression of the virus. The range of clinically relevant genes shown to form G4 has generated a lot of interest in investigating the therapeutic potential of G4 binding ligands.

In the early to mid-1990s, when the G4 field was focusing almost exclusively on the telomeric G4, the rationale of initial research into G4 binding ligands was to design ligands that could discriminate between G4 and duplex DNA with the hope that they would stabilise telomeric G4 and inhibit telomerase function in cancer cells. As more G4 structures have been discovered, the level of complexity has increased considerably and research into G4 ligands has expanded to include the targeting of other cancer related G4 targets, most notably G4 DNA in promoter regions to inhibit transcription and G4 RNA in 5'UTR to inhibit translation. There has also been investigation into G4 associated with other disease areas. An example of recent success is the repression of HIV transcription through ligand-induced stabilisation of G4 structures in the HIV-1 long terminal repeat.^{176, 177} As the

number of potential G4 targets increased there has been a growing need to develop ligands that not only differentiate between G4 and other secondary structures but also between different quadruplexes. As research moved towards more specific methods to target G4, two distinct approaches were followed. Researchers either developed small molecule ligands to target structural features of the quadruplex, or antisense strategies to target the quadruplex through sequence specificity.

The G4-targeting drug that has progressed the furthest to date is quarfloxin, which progressed through to Phase II clinical trials for the treatment of neuroendocrine/carcinoid tumours before being withdrawn due to issues with bioavailability.¹⁷⁸ It targets G4 structures in the non-template strand of GC-rich rDNA which form during transcription to prevent renaturation of template DNA and facilitate RNA Pol-I transcription of rRNA. Dis-regulation of RNA Pol-I in various cancers results in excessive rRNA biogenesis and the uncontrolled growth and proliferation of cells.¹⁷⁹ Nucleolin is an important nucleolar protein that is vital for RNA Pol-1 mediated rRNA transcription.^{180, 181} It is thought to facilitate transcription by stabilising rDNA G4, which it binds to with high affinity.¹⁸² Quarfloxin disrupts nucleolin-G4 interactions, inhibiting rRNA biosynthesis and resulting in apoptosis of cancer cells. However, ultimately the drug was withdrawn from trials due to issues with bioavailability.

1.7.2. Small molecule ligands

(i) Differentiating between G4 and duplex DNA

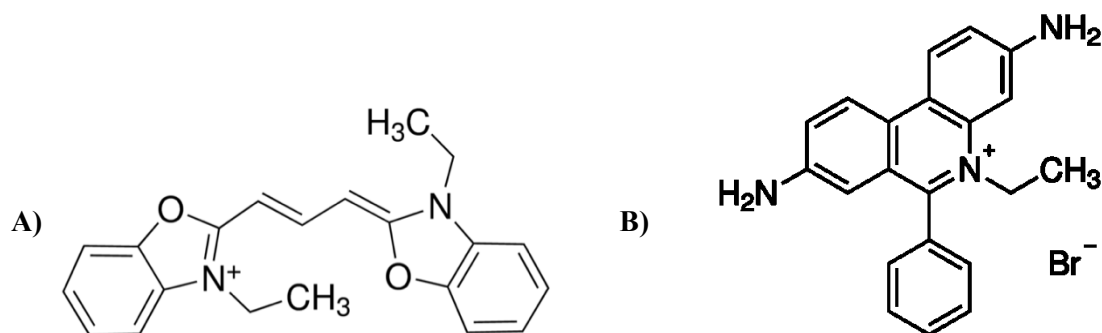


Figure 28: Structures of selected G4 binding dyes, A) 3,3'-diethyloxadicarbocyanine, B) Ethidium bromide.

A wide range of G4 binding ligands have been developed over the last two decades. (See Monchaud *et al.*^{183, 184} for reviews of the most representative classes of G4-specific ligands.) Initial G4 ligands were designed to discriminate between DNA G4 and other DNA secondary structures, mainly duplex DNA. The first ligands to be investigated were dye molecules already known to intercalate between Watson-Crick base-pairs in double-stranded DNA, for example ethidium bromide¹⁸⁵ and carbocyanine 3,3'-diethyloxadicarbocyanine¹⁸⁶ which were both shown to bind to synthetic G4 *in vitro* (see Figure 28 for structures).

The first evidence of telomerase inhibition by a G4 binding, non-nucleoside small molecule ligand was presented in 1997 by Neidle, Hurley and co-workers, who targeted telomeric G4 with bisamidoanthraquinone.⁹³ Quadruplex versus duplex discrimination can be achieved by targeting structural features that are specific to G4 structures including (a) the wide flat face of the terminal G-quartet, (b) the loops and (c) the grooves along the edge of the G4 structure. Targeting of each of these structures will be discussed in the upcoming section.

(ii) Stacking with terminal G-quartet

The planar surface of G-quartets provides a unique structural feature (absent in duplex DNA) that has been extensively exploited in the design of G4-specific binding ligands. These planar arrays of guanines are much larger than traditional Watson-Crick base-pairs found in duplex B-DNA and can accommodate ligands with much larger aromatic surfaces. The intercalation of ligands between G-quartets is not favoured as it has a destabilising effect on the quadruplex, displacing the monovalent cation and disrupting the stacking of the quartets.¹⁸⁷ By comparison, π - π stacking with the external face of the terminal G-quartet is by far the most common interaction utilised by G4 ligands. Ligands binding to the terminal quartet often include positively charged side chains that provide additional stabilisation of the ligand-DNA complex through electrostatic interactions with the negatively charged phosphate backbone. Such appendages were also successfully used to further enhance quadruplex versus duplex specificity (as detailed in the following sections). There are two main classes of terminal quartet targeting ligands, both of which interact via π -stacking with the exposed face of a G-tetrad. The first are small flat aromatic platforms, representative examples of which include: naphthalene di-imide (NDI) derivatives (Figure 29B),^{188, 189} pyridostatin (Figure 29C),¹⁹⁰ cationic porphyrins^{191, 192} and bisquinolinium derivatives.^{193, 194} These aromatic ring structures are commonly decorated with additional side chains of varying lengths. These side chains typically include hydrophilic moieties, which are incorporated to retain favourable solubility in water. The second class of ligands are macrocycles which are built from multiple, smaller, aromatic (and often heterocyclic) units linked in a cyclic structure. An interesting sub-section of these are a class of neutral macrocycle analogues based on the natural product telomestatin (Figure 29D). Shin-ya *et al.* isolated telomestatin from *Streptomyces annulatus* in 2001 and subsequent studies have confirmed it is a highly effective G4 ligand which can greatly stabilise G4 ($\Delta T_{1/2} = 24$ °C), exhibits little affinity for DNA duplex and inhibits telomerase activity with IC₅₀ values in the low nano-molar range.¹⁹⁵⁻¹⁹⁷ Other examples of macrocyclic G4-ligands are those built around quinacridine and acridine aromatic units, such as BOQ₁¹⁹⁸ and BisA¹⁹⁹, tri-oxazole macrocycles²⁰⁰ and the tetra-cationic porphyrin TMPyP4 (Figure 29E).²⁰¹

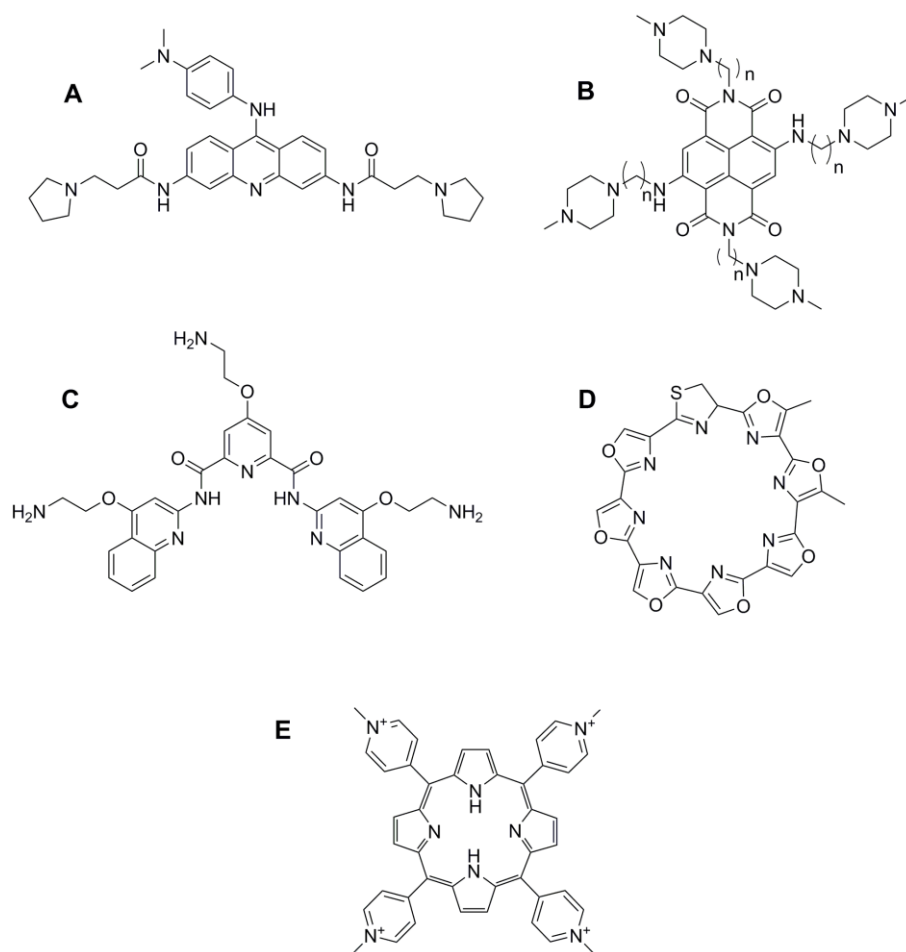


Figure 29: Structures of selected G4-binding ligands including flat aromatic molecules with positively charged side-chains: (A) trisubstituted acridine BRACO-19, (B) tetra-substituted NDI, (C) Pyridostatin, and two macro-heterocycles, (D) Telomestatin and (E) TMPyP4.

(iii) Groove binding

The grooves of the quadruplex also offer opportunities for specific binding and can be targeted in isolation or in conjugation (e.g. as appended side-chains) with end quartet bound ligands. A range of different DNA-duplex major and minor groove binding ligands have been tested for G4 binding and specificity.^{202, 203} The first ¹H-NMR structure of a groove binding ligand with a G4 was solved by Randazzo *et al.* in 2001 showing the duplex minor-groove binding ligand Distamycin-A bound to the parallel G4 d[**TGGGGT**]₄.²⁰⁴ Subsequent structural studies further clarified the binding mechanism, demonstrating that distamycin-A formed anti-parallel dimers which bound opposing grooves of the quadruplex.^{205, 206} It is worth noting that an NMR structure with an alternative quadruplex provided evidence of distamycin-A binding via π - π stacking with the terminal quartet, suggesting its mechanism may not be consistent between different quadruplex structures.²⁰⁷ A computational analysis of distamycin binding suggested that the central cation affects groove binding by modulating the

electrostatic potential of the middle and terminal guanines in the groove, thus driving ligand binding to the terminal end of the groove.²⁰⁸ Non-planar ligands including alkyls²⁰⁹ and steroids^{210, 211} have also been demonstrated to stabilise parallel and anti-parallel G4 through binding to the lateral grooves.

Although DNA duplexes contain only two grooves, so-called major and minor grooves, G4 structures have grooves in between each of the four backbone strands. As a result, the conjugation of three or more groove-binding side chains can improve the specificity of ligands for G4 over duplex DNA (see Figure 31 for binding modes of two tetra-substituted NDI derivative complexed with human telomeric G4). This was the rationale behind the design of BRACO-19, a 3,6,9-trisubstituted acridine which has demonstrated high specificity for G4 over duplex DNA as a result of side chain groove binding (Figure 29A).^{212, 213} Ladame *et al.* further validated this strategy, demonstrating that the addition of rationally designed groove binding tetrapeptide side chains to a telomeric G4 binding ligand (3,6-bis-[3-pyrrolidinopropionamide] acridine) significantly increased its specificity for G4 over duplex DNA.²¹⁴ The side chains of the substituted ligands have numerous binding modes in the grooves of the quadruplex, as shown in Figure 30.

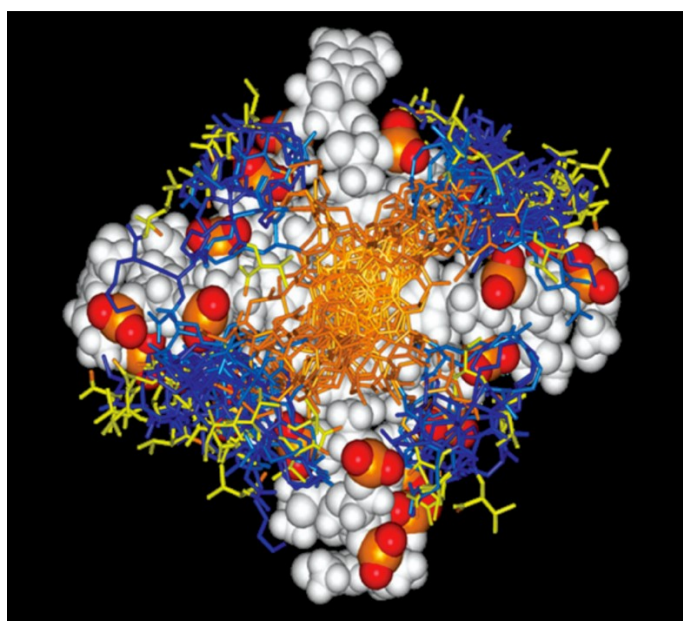


Figure 30: Multiple binding modes of 3,6-bisRKKV acridone 3a. The tricyclic acridone cores (orange) interact with the face of the terminal G-tetrad. Side chains bind in the major and minor G4 grooves. Arginine and lysine sidechains are coloured in blue and valine in yellow. The G4 structure is shown in grey for clarity, with the exception of the phosphate groups, where the phosphate atom is coloured orange and oxygen atoms coloured red. Figure reproduced from Ladame *et al.*²¹⁴

(iv) Loop binding

Loops typically make only a small contribution towards the overall stability of the quadruplex, so are not normally targeted in isolation. There are however many ligands that bind to both the face of the terminal G-quartet and to loop regions. The well characterised ligand tetra(N-methyl-4-pyridyl)

porphyrin (TMPyP4) (Figure 29E) has been shown to bind loops of telomeric and *c-kit* G4.²¹⁵ Targeting of loops is most typically used in antisense strategies or in conjunction with end-binding ligands to differentiate between individual G4, both of which are detailed in the corresponding sections of this report.

(v) Differentiating between individual G4

Once researchers discovered that there were significantly more potential G4 targets than those formed at telomere ends, developing ligands that could discriminate between individual G4 became an important but also increasingly challenging goal. Topological differences between DNA G4, for example in backbone strand orientation or in length and composition of loop regions, provide opportunities to design ligands that differentiate between individual quadruplexes. In 2001, Han *et al.*²⁰¹ demonstrated that the binding and stabilisation properties of three cationic porphyrins differed between parallel and antiparallel quadruplex structures. The authors showed that TMPyP3 has a greater inhibitory effect on unwinding of a parallel G4 by Sgs1 helicase than TMPyP4. In addition, this effect was reversed with an anti-parallel quadruplex, with TMPyP4 having a greater inhibitory effect and therefore suggesting that the two ligands bind to different G4 topologies with different specificity. In this assay the third ligand, TMPyP2, did not have any clear effect on helicase unwinding of either G4. This result was supported by further gel mobility experiments that showed the TMPyP3 specifically promotes the folding of DNA PQS into parallel structures. More recent examples include the oxazole-based ligand TOxPy and the bi-substituted NDI ligands ALI-C3, BBZ-ARO and BBZ-AROC₂.²¹⁶ ²¹⁷ The dimensions of G4-DNA grooves vary significantly between different DNA G4 types, which has concurrent effects on the pattern of available H-bonding sites for groove binding ligands.²¹⁸ For example, narrow minor grooves can prevent bulkier sidechains from accessing the hydrogen bonding sites on the RNA backbone whereas a wide major groove could enable binding of bulky sidechains. As such, groove binding, either in isolation or in co-operation with end terminal quartet binding, offers the potential to differentiate between individual quadruplex structures. Loop regions are also highly variable between quadruplexes and offer the opportunity for specific targeting, generally through the addition of side chains. X-ray crystallography and NMR studies of NDI ligands bound to telomeric G4 have demonstrated that the choice of side chain groups can impact groove and loop binding. The binding of tetra-substituted NDI derivatives to the telomeric end-quartet was shown to be stabilised by interactions between dimethylamine side chains and the flexible propeller-type TTA loop regions of the quadruplex (Figure 31).²¹⁹ In contrast, the positively charged side chains of a range of different NDI ligands were shown to form interactions with phosphate groups in the quadruplex grooves.¹⁸⁸ In 2015, NDI derivatives with laterally expanded aromatic cores demonstrated selective binding towards HIV long terminal repeat (LTR) G4 over telomeric G4. This selectivity was thought to be due to interactions between side chains and the long loop regions of the viral G4.¹⁷⁶

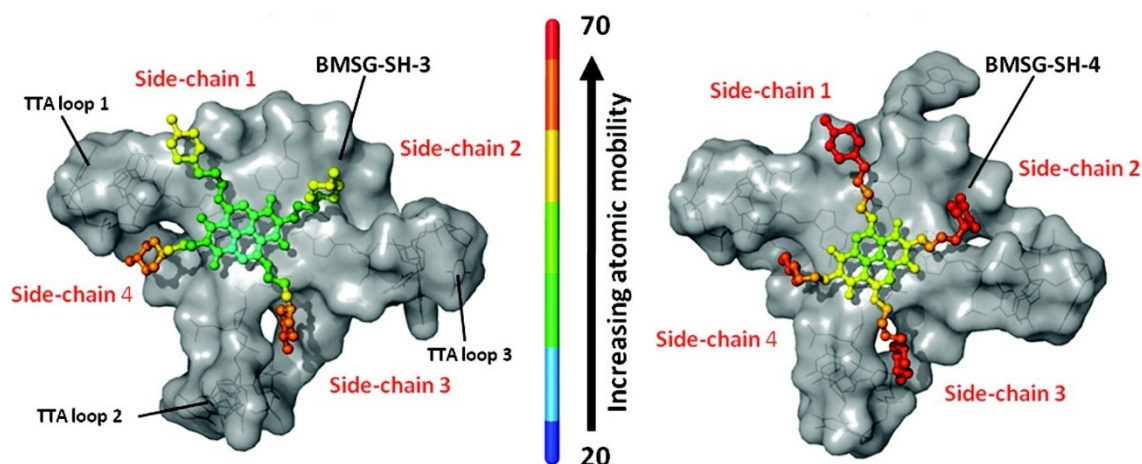


Figure 31: Representation of tetra-substituted NDI ligands BMSG-SH-3 (left) and BMSG-SH-4 (right) bound to the terminal G-quartet of the human telomeric DNA G4, Gtel22. Side chains can be seen interacting with the G4 grooves and TTA loops to enhance binding. The colouring of the ligands indicates their atomic mobility in the crystal lattice. Figure reproduced from Collie *et al.*²²⁰

(vi) RNA G4 specific ligands

Early studies targeting RNA G4 took small molecules that had demonstrated effective binding and stabilising properties with DNA G4 and produced variants with increased specificity for RNA G4. Following on from their initial work with the *NRAS* G4, the Balasubramanian group tested the effect of

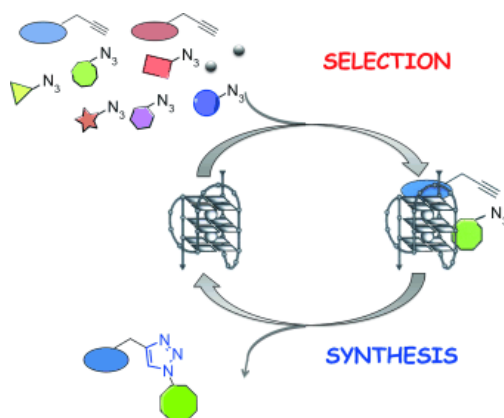


Figure 32: Selection and synthesis method utilised to identify and produce selective RNA binding ligands. Two G4 binding ligands, based on the ligand pyridostatin, were mixed with a selection of azides in the presence of a G4 structure and K^+ . Upon binding to the G4 the individual molecules would be brought into close contact and bind via an *in situ* Click reaction to generate a complex with strong G4 binding properties. Figure reproduced from Di Antonio *et al.*²²¹

a selection of G4 binding ligands on translation and identified a pyridine-2,6-bis-quinolino-dicarboxamide derivative, RR82, which showed specificity for RNA G4 in the presence of dsDNA duplex and RNA hairpin structures.¹⁹⁰ Gomez *et al.* targeted the *TRF2* RNA G4 with three bisquinolinium compounds and demonstrated that they inhibited translation *in vitro*.¹⁹⁴ In 2012, researchers in the Balasubramanian group used Click chemistry to test the binding affinities of

pyridostatin derivatives for telomeric RNA and DNA G4 (Figure 32). They identified one derivative that showed selective binding for RNA G4 in the presence of up to 100 mM of equivalent DNA G4. This was suggested to be driven by hydrogen bonding between the carboxylate moiety of the ligand and the 2'-OH group on the RNA G4.²²¹

Molecular dynamics simulations of the TERRA crystal structure complexed with three different NDI ligands suggest some mechanisms for selective binding of RNA over DNA G4 (Figure 33).⁷³ The 2'-OH groups of the RNA G4 form strong interactions with carbonyl groups on the central core ring of the ligand, as well as with nitrogen atoms on the side chains. Interactions with 2'-OH groups also stabilise the loops of the quadruplex in a more compact formation than the equivalent DNA G4. This means that the loop regions of RNA G4 are less able to accommodate ligands with bulky or inflexible side chains.

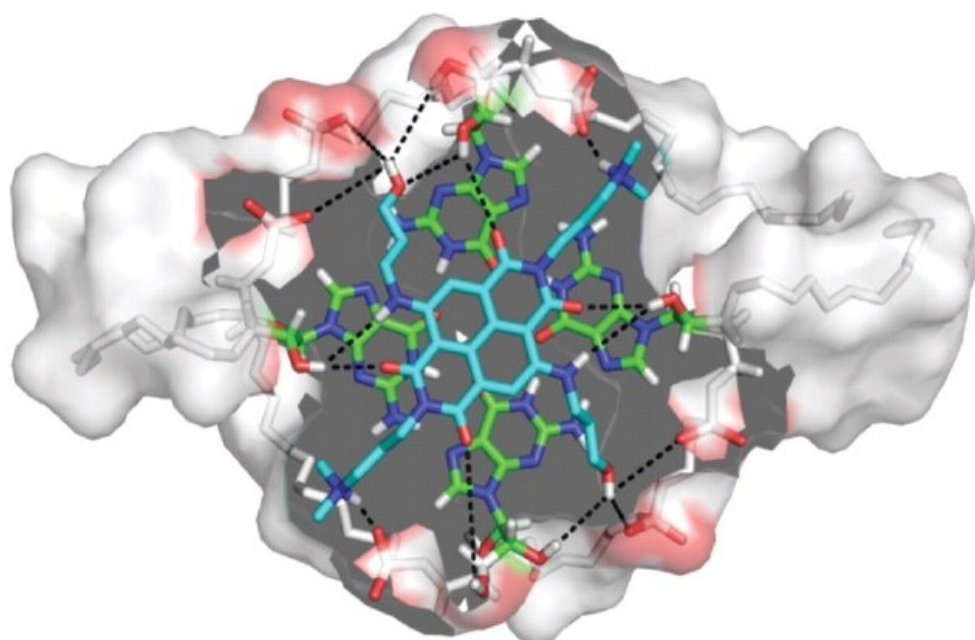


Figure 33: Molecular dynamics simulations of the binding of three tetra-substituted NDI ligands to TERRA G4. The ligand (cyan) is stacked onto the 3' G-quartet (carbons shown in green). Oxygen containing groups on the ligand and G4 are shown in red and nitrogen groups are shown in blue. Figure reproduced from Collie *et al.*⁷³

The stability and structural homogeneity of RNA G4 makes them an easier target for ligand binding than DNA quadruplexes. If the structure of the quadruplex can change depending on environmental conditions (as with many DNA G4) then ligand binding may become dependent on a number of different cellular processes, for example osmotic pressure which varies significantly during the cell cycle.²²² However, this same structural homogeneity also makes specific targeting of RNA quadruplexes in the presence of competing RNA quadruplex structures more challenging. The structural homogeneity of RNA G4 is offset by the diversity of alternative RNA secondary structures. Unlike DNA, which mainly exists as a double helical structure, RNA can form a range of different secondary

structures with diverse topologies. This introduces an additional challenge for specific targeting of RNA G4.

1.7.3. Antisense strategies

Antisense oligonucleotides (ASO) are short nucleic acid sequences used to target mRNA of specific genes with the aim of regulating their translation. Typically, binding of antisense oligonucleotides to target mRNA is designed to inhibit translation, either through steric hindrance of the ribosome or by blocking binding sites for translation initiation factors.²²³ ASO have also recently been used to upregulate translation of downstream primary ORF through binding to upstream ORF.²²⁴

Effective antisense activity *in vivo* requires high affinity for complementary oligonucleotides and resistance to nuclease degradation, which means that natural DNA and RNA oligonucleotides are relatively poorly suited to the task. As such, various backbone modifications and synthetic oligonucleotide analogues have been developed with significantly improved antisense properties. The most common of these include the 2'-O-Me backbone modified RNA, locked nucleic acids (LNA), phosphorodiamidate morpholino oligonucleotides (PMO) and peptide nucleic acids (PNA).²²⁵ The properties of PNA in particular will be discussed in further detail in section 1.8.

1.8. Peptide Nucleic Acids (PNA)

1.8.1. Structure and Backbone modifications

Peptide nucleic acids (PNA) are synthetic nucleic acid analogues that contain a charge-free peptide backbone instead of a negatively charged ribose-phosphate backbone. The original and most commonly used PNA backbone peptide is aminoethylglycine (aeg) (Figure 34). Aeg-PNAs form non-cyclic, flexible polymers, constrained by the two amide groups on the backbone.²²⁶ Alternative peptide backbones tend to result in lower affinity for complementary oligonucleotide strands making them less desirable, although with some notable exceptions. Substitution of the glycine backbone with the charged amino acids lysine (especially D-form lysine) or arginine have been shown to improve PNA-DNA duplex stability and endosomal uptake respectively.²²⁷⁻²³¹

Modifications of the aminoethyl group within the peptide backbone can also affect PNA activity. In particular, replacement of the aeg-ethyl group with a γ -methyl group in PNA (called γ -PNA) pre-organises the strand into a helical structure, which increases affinity for complementary

oligonucleotides.^{232, 233} There have also been attempts to generate cyclic backbones by replacing elements of the aeg group with cyclic derivatives to constrain the flexibility of the aegPNA polymeric structure. As with amino acid substitution, most cyclic derivatives have resulted in decreased hybridisation properties.²³⁴ However, cypPNA and pyrPNA, formed by the introduction of a *trans*-cyclopentyl linkage or alternating 2-aminocyclopentanoic acid and pyrrolidine-2-carboxylic acid respectively, have been shown to have a positive impact on the affinity and specificity of binding.²³⁵⁻²³⁷

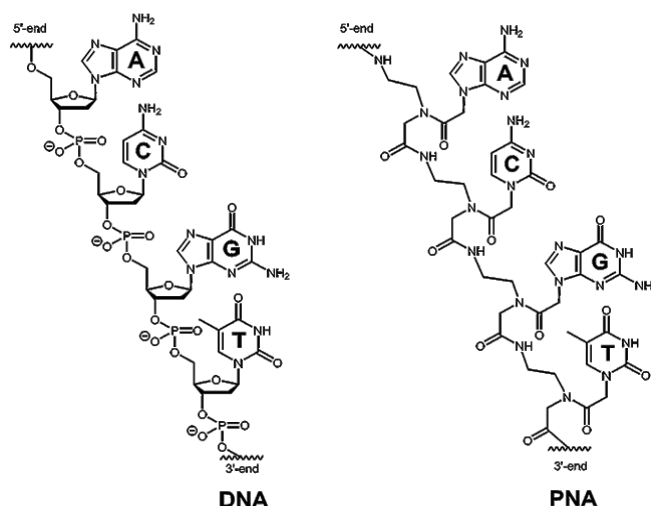


Figure 34: Structural representations of DNA and PNA. (left) DNA structure with a phosphate-ribose backbone. **(right)** PNA structure with aeg backbone.

Most PNAs are charge free and therefore not water soluble. For antisense applications this can be overcome by appending hydrophilic moieties to the PNA strand. These are commonly either charged amino acids (i.e. arginine, lysine) or water solubilising moieties (e.g., polyethylene glycol or PEG). The introduction of di-ethylene glycol at the gamma-backbone position showed significant promise, improving the water solubility and DNA binding specificity of PNA oligomers.²³⁸

1.8.2. PNA:DNA and PNA:RNA interactions

(i) PNA Duplexes

Replacing the negatively charged phosphate backbone with a charge free backbone removes the electrostatic repulsion between strands when binding/hybridising to complementary oligonucleotides. This confers a number of advantages for use in antisense applications. Early work by the Nielsen group demonstrated that PNA can bind to complementary DNA and form PNA:DNA hetero-duplexes with a higher thermal stability than equivalent DNA:DNA duplex. For instance, the melting temperature (T_m)

of PNA:DNA penta-decamer heteroduplex is 16 °C higher than that of comparable DNA:DNA homoduplex (69.5 °C versus 53.3 °C).²³⁹ The replacement of the oligonucleotide backbone also impacts recognition of PNA by nucleases. PNA was shown to be exceptionally stable in human serum and cellular extracts exhibiting reduced or no degradation under conditions that resulted in the total degradation of various control sequences.²⁴⁰ PNA exhibits much higher sequence specificity when binding complementary sequences than DNA or RNA. For instance, the introduction of single base pair mismatches into the middle of penta-decamer DNA:PNA duplex resulted in decreases in melting temperature (T_m) of between 8 °C and 20 °C.²³⁹

(ii) PNA Triplex formation

As well as hybridising to complementary DNA sequences to form PNA:DNA duplexes, pyrimidine-rich PNAs can also form a PNA₂:DNA triplex structure, where a second PNA strand binds to the DNA strand in the PNA:DNA duplex through Hoogsteen hydrogen bonding.²⁴¹ The same binding mechanism has been observed with PNA binding to complementary RNA sequences with similar T_m to equivalent PNA₂:DNA triplexes.²⁴² These structures are highly stable, which allows PNAs to displace strands in duplex DNA targets. To exploit this strong binding mode, bis-PNA have been developed to invade longer duplex DNA *in vivo* via triplex formation. These contain two PNA sequences joined by a flexible linker: the first binds to complementary antiparallel DNA via Watson-Crick base pairing and the second is a poly-pyrimidine strand which binds parallel to the target DNA and interacts via Hoogsteen bonds.²⁴³ Natural RNA triplexes are found in larger RNA secondary structures, such as RNA pseudoknots, riboswitches and the ribosome.^{30, 244, 245} Triplex-forming PNAs between 6-10 nucleotides in length directed against RNA duplexes have been shown to form stable, sequence-specific, major groove PNA:RNA₂ triplexes.²⁴⁶ PNA targeting of RNA duplexes has potential therapeutic applications such as targeting HIV viral REV protein binding or ribosomal frameshift-inducing RNA and unlike other nucleic acid analogues, PNAs form more stable triplex structures with RNA than DNA.^{247, 248} However, triplex formation by PNA oligomers containing natural cytosine residues is limited to low pH environments, as full protonation of the cytosine (pK_a of around 4.5) is required to form a C⁺.G-C base triple (Figure 35).^{243, 249} The addition of synthetic nucleotides has been explored to overcome the need for a low pH environment to activate the cytosine. Researchers have shown that modified PNA sequences can form triplexes with RNA hairpin structures at physiological pH and therefore inhibit translation both *in vitro* and in cells.²⁵⁰ Single stranded homo-purine RNA has also been shown to form triplexes with complementary nucleic acid analogues, as demonstrated with acyclic (L)-threoninol nucleic acid (aTNA).²⁵¹



Figure 35: Triplex formation by bis-PNA. The anti-parallel strand binds through complementary Watson-Crick base pair interactions, the second strand binds in the parallel orientation through Hoogsteen hydrogen bonding. The C.G interaction (highlighted in a grey box) is pH sensitive and requires low pH to form.

(iii) PNA interaction with G4

There have been a number of different approaches developed using antisense oligonucleotides to target G4 structures. Short oligonucleotide sequences can be used to either stabilise or induce formation of existing G4 structures and, as a result of this interaction, inhibit translation. This can be through **(1)** binding to G4 loops to stabilise an existing structure, **(2)** invasion of the DNA duplex and binding to the opposite strand to allow the quadruplex to fold, or **(3)** binding to the quadruplex-forming strand and forming hetero-quadruplexes with increased stability (illustrated in Figure 36).

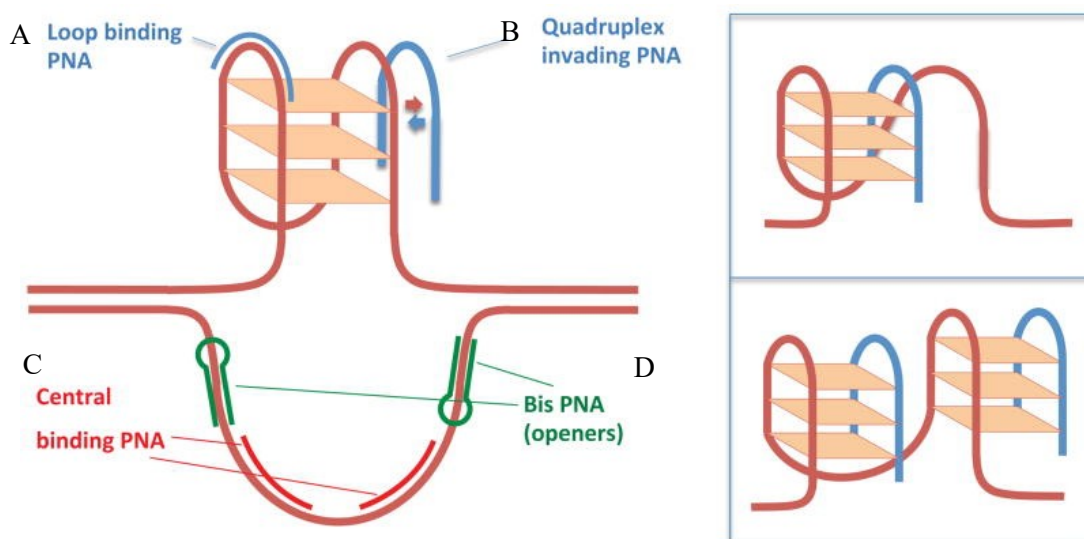


Figure 36: PNA targeting of DNA G4. There are several different methods to target G4 with PNA strands, including: **(A)** loop binding PNA which targets complementary loop regions via Watson-Crick base pairing, **(B)** quadruplex invading PNA that displaces the DNA guanines and binds via Hoogsteen hydrogen bonding, **(C)** central binding PNA that binds complementary regions on the non-coding strand to prevent duplex formation and **(D)** bis-PNA also targets the non-coding strand but via PNA₂-DNA triplex formation via complementary binding and Hoogsteen hydrogen bonding. Figure taken from Panyutin *et al.*²⁵²

Bimolecular hetero-quadruplexes can also be formed between guanine-rich oligonucleotides and DNA and RNA sequences without an existing PQS, but where there are two stretches of contiguous guanines. The use of antisense oligonucleotides to destabilise G4 has also been reported. This is typically still with the intent of inhibiting translation, either by destabilising a G4 in an IRES structure, or by invading an existing quadruplex structure and replacing it with a more stable oligonucleotide duplex.

In an initial *in vivo* study conducted in zebrafish, strand invasion and duplex formation induced by a short LNA strand targeting a G4 in the *c-myc* promoter region was shown to inhibit transcription.²⁵³ The Panyutin group have also conducted extensive work inducing DNA G4 formation through strand invasion with PNA oligonucleotides. In a series of studies, quadruplex specific invasion of the BCl-2 duplex by PNA targeted towards the complementary strand was shown to induce G4 formation.^{254, 255} A later study also demonstrated that replacing the guanine nucleobases in strand invading PNA sequences with the analogue pyrazolo[3,4-d]pyrimidine guanine (PPG) prevents PNA G4 formation, reducing aggregation and promoting strand invasion of target DNA.²⁵⁶ More recently, Rouleau *et al.* used LNA and 2'O-Me antisense oligonucleotides to both promote and inhibit RNA G4 formation and subsequently translation.²⁵⁷ Inhibition was achieved by targeting loop regions and the quartet forming guanines on either side in order to disrupt the quadruplex through formation of a competing duplex structure. In an opposing strategy, an ASO was designed to invade and disrupt an RNA hairpin structure concealing a potential quadruplex forming site in order to facilitate G4 formation.

Ito *et al.* first demonstrated the potential of short G-rich RNA sequences to form bi-molecular hetero-quadruplexes with G-rich tracts in 5'UTR mRNA and thereby inhibit translation.²⁵⁸ This concept was expanded in a later study where researchers designed slightly longer RNA which contained both a G-rich region to participate in hetero-quadruplex formation and a second region complementary to the flanking regions of specific G-tracts in the eIF-4E mRNA 5'UTR and coding regions, thereby inducing specific inhibition of eIF-4E translation.²⁵⁹ A study investigating the binding mechanism of hetero-quadruplex forming PNA found that they bind with a variety of complex conformations and varying stoichiometry, including quadruplex, duplex and triplex formation.²⁶⁰

PNA have been successfully used to control DNA G4 stability and subsequently transcription efficiency. (See Panyutin *et al.*²⁵² for a recent review.) An early study by the Armitage group demonstrated that short PNA sequences can invade stable, folded DNA quadruplex structures and form a PNA:DNA hetero-duplex.²⁶¹ Short PNAs have also been used to bind loop regions of G4 structures, invade the DNA duplex and bind the complementary strand to promote extrusion of the quadruplex-forming strand or bind to the quadruplex forming strand itself and form stable hetero-quadruplex structures (Figure 36).^{254, 262, 263}

Targeting of G4 with PNA oligomers has also demonstrated therapeutic potential. Induced PNA:DNA hetero-quadruplexes have been shown to be more effective inhibitors of polymerase η extension than DNA homo-quadruplexes.²⁶⁴ The IC₅₀ of hetero-quadruplex forming PNAs was also shown to be roughly equivalent to that of the G4-stabilising small molecule ligand BRACO-19. This provides strong evidence that targeting of G4 with PNA ASO could be used to inhibit DNA polymerase synthesis with potential applications in anti-cancer therapy. The potential for PNA-based targeting of quadruplex

structures extends to RNA and targeting translation of G4 containing oncogenes. In a very recent study the Armitage group showed that γ -PNA could inhibit translation by invading existing RNA quadruplex structures and replacing them with more stable duplex structures. The γ -PNA used in this study was also shown to bind with greater specificity than aeg-PNA or 2'-OMe RNA.²⁶⁵

1.8.3. Therapeutic applications of PNA

The most successful therapeutic application of PNAs to date is in correcting exon skipping in Duchenne muscular dystrophy (DMD). DMD is an X-linked muscular disorder characterised by a loss of functional dystrophin protein resulting in severe, progressive muscle wasting. It is caused by frameshift or non-sense mutations in the dystrophin gene. A major therapeutic goal in DMD treatment has been developing a method to correct the reading frame of the dystrophin gene and restore expression of functional dystrophin protein. Dunkley *et al.* were the first to demonstrate that 2'-O-Me ASO targeting splice sites in the dystrophin gene could be used to induce removal of the mutated exon 23 in cultured *mdx* myoblasts.²⁶⁶ However, there were concerns about observed toxic and non-specific effects mediated by the phosphorothioate backbone.²⁶⁷ Yin *et al.* demonstrated that PNA ASO could also be used to correct dystrophin expression and induce higher levels of exon skipping than 2'-O-Me ASO.²⁶⁸ They subsequently demonstrated that they could restore systemic functional dystrophin expression in *mdx* mice without detectable toxicity.²⁶⁹

1.8.4. Bioavailability and cellular uptake of PNA

A major disadvantage of PNAs for therapeutic purposes is their poor *in vivo* bioavailability. Researchers testing the effects of antisense PNAs injected in mice observed very little activity due to high excretion and low cellular uptake.²⁷⁰ The first problem is that, being large, hydrophobic molecules, PNA oligomers have a very short half-life in the blood stream and are rapidly excreted through the kidneys.²⁷¹ Improvements in half-life duration *in vivo* have been observed for PNAs containing tetralysine and oligoarginine.^{268, 270} Naked PNA oligomers are also known for their extremely poor cellular uptake, a number of potential solutions to which are discussed in chapter 3.

Chapter 2: Targeting of RNA G-quadruplexes with PNA-small molecule conjugates

2.1. Introduction

The main limitation of RNA quadruplexes as a target for therapeutic ligands remains the capacity for a small molecule to bind to the quadruplex of interest with a high specificity, and to do so despite the high number of potential alternative RNA targets (including other RNA quadruplexes) *in vivo*. To address this problem, we have designed a novel class of “sequence + structure”-specific ligands through the conjugation of short PNA strands to G4-binding small molecule ligands. These PNA-small molecule conjugates are designed to stabilise and/or induce the formation of RNA G4 by simultaneously targeting the quadruplex structure itself with a small molecule ligand (structural specificity) and the two single stranded RNA flanking regions with complementary PNA oligomers (sequence specificity), thereby binding with “sequence + structure”-specificity (Figure 37) to unique RNA G4 in the genomes (i.e. those located between the RNA sequences complementary to the PNAs). The potential benefits of this approach are two-fold. Firstly, there should be a synergistic effect on ligand affinity. Secondly, the addition of a complementary oligonucleotide provides sequence specificity to complement the structural specificity conferred by the small molecule ligand.

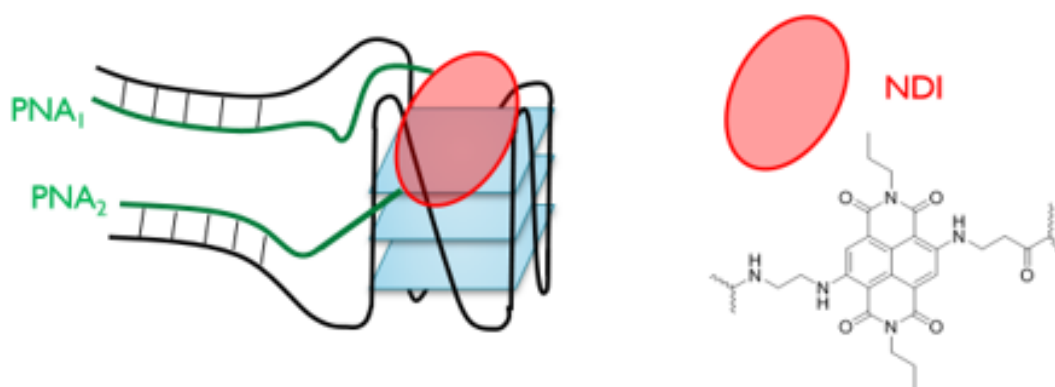


Figure 37: Proposed mechanism for "Sequence + Structure"-specific PNA-small molecule ligand binding to RNA G4. The central G4 binding ligand, in this case NDI, binds to the planar surface of the terminal G-quartet while PNA₁ and PNA₂ hybridise to the single stranded G4 flanking regions.

A number of previous studies have explored the use of oligonucleotide-small molecule conjugates to target DNA and RNA. In the 1980's the Hélène group published a series of studies investigating inhibition of translation by intercalating agents covalently bound to complementary oligonucleotides. In 1989 they demonstrated that conjugation of an acridine to a homopyrimidine oligonucleotide increased its binding affinity for target duplex DNA via formation of a DNA triplex.²⁷² Most

interestingly, sequence specific targeting was achieved via this strategy. Random intercalation of the acridine moiety was lost due to the repulsive effect of the negatively charged oligonucleotide tail. Instead, intercalation occurred only at those double-stranded sequences recognised by the acridine-DNA conjugate and remained even in the presence of an excess of alternative DNA duplexes (e.g. calf thymus DNA). Other representative examples based on this strategy are also discussed in a 2012 book chapter by Rusling and Fox.²⁷³

In 2003, the Nielsen group conducted a similar study looking at conjugation of acridine ligands to improve oligonucleotide binding which also looked at the specificity of such conjugates. However, in this case, acridine was conjugated to PNA rather than DNA.²⁷⁴ The rationale of the experiment was to reduce the ionic strength sensitivity of DNA duplex invasion. An electrophoresis mobility shift assay (EMSA) was used to measure the binding properties of mono- and bis-PNA conjugates. The binding affinity of a *bis*-PNA with three lysine charges was shown to decrease around 100-fold when the ionic concentration was changed from 10 mM Na⁺ to physiologically relevant ionic conditions of 150 mM K⁺/Na⁺. However, conjugation of an acridine moiety (Figure 38) significantly reduced the sensitivity of the *bis*-PNA to ionic strength, with only a 10-fold decrease in binding at 150 mM K⁺/Na⁺ when compared to 10 mM Na⁺. The authors subsequently investigated the specificity of the PNA-acridine conjugates by comparing the binding of *bis*-PNA with and without acridine conjugates to complementary DNA and to DNA with a single base mismatch. The authors reported that a comparison of the ratios of the C50 values for the complementary and mutated DNA showed that the acridine conjugates were consistently as specific or more specific when compared to equivalent conventional PNA (lacking the acridine). However, while this suggests that specificity was not affected by conjugation, it provides no clear evidence that conjugation improved specificity, and as these results were not shown it is difficult to draw further conclusions. In addition, the experiment was carried out at 10 mM Na⁺ which meant that the PNA conjugates and conventional PNA had different C50 values (0.004 and 0.02 μ M respectively), to get more directly comparable results the experiment should have been conducted at an Na⁺ concentration where the C50 values for the two PNA types was the same.

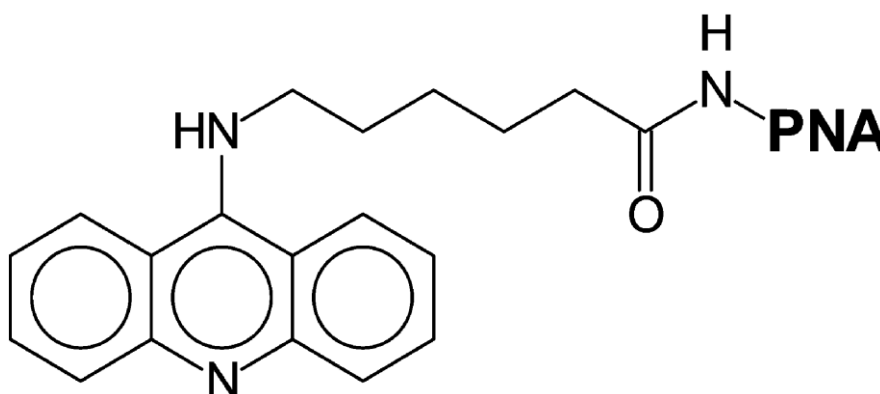


Figure 38: Structural representation of PNA-acridine conjugate. Figure reproduced from Bentin *et al.*²⁷⁴

The Nielsen group conducted a further study in 2005 which looked at the sequence specificity of PNA-minor groove binder conjugates.²⁷⁵ The experiment studied PNA-Hoechst conjugates to investigate whether the Hoechst would act to guide the PNA sequence to an adjacent complementary motif and thereby improve specificity. To test this, the PNA-Hoechst conjugates were combined with an equal mix of DNA-A, which contained Hoechst binding region proximal to a complementary PNA binding region (proposed binding mechanism shown in Figure 39), and DNA-B, which contained only the PNA binding region. The two complexes – PNA-Hoechst:DNA-A and PNA-Hoechst:DNA-B were separated using an electrophoresis mobility shift assay. The results showed that the PNA-Hoechst bound approximately 10-fold more DNA-A than DNA-B, whereas the PNA alone bound the two DNA types equally. This suggests that conjugation of a minor groove binder can enable specific targeting of PNA dependent on the context of a complementary sequence and in the presence of multiple complementary sequences. However, it is worth noting that the electrophoresis mobility shift assay could have impacted the results of this assay. For example, one cannot rule out a partial dissociation of some of the most-unstable complexes within the gel. Validation of these results using a different method, for example surface plasmon resonance, would help prove whether there was any experimental bias. The authors did take some steps to confirm the accuracy of the findings, demonstrating that the same trend was observed at a higher K^+ concentration.

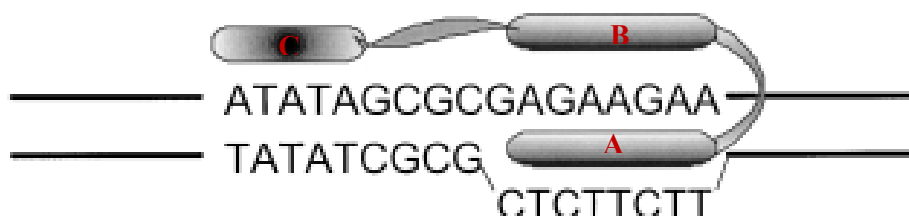


Figure 39: Suggested binding mechanism of PNA-Hoechst conjugate to target DNA duplex. The bis-PNA comprises of two strands, the first (A) invades the DNA strand and binds to an anti-parallel complementary sequence, the second (B) binds to the same sequence in parallel. The Hoechst (C) binds to the minor groove of an A-T region adjacent to the PNA binding site. Figure reproduced from Nielsen *et al.*²⁷⁵

A later study by Rapireddy *et al.* reported on DNA duplex strand invasion by γ -PNA-acridine conjugates.²³³ In this case the conjugation of a small intercalating molecule increased the affinity of the PNA significantly, allowing it to invade the DNA duplex (Figure 40). This paper was significant following on from the previous studies by the Nielsen group as it demonstrated that pre-organisation of the PNA can improve duplex invasion and that with a PNA analogue, binding can occur without triplex formation.

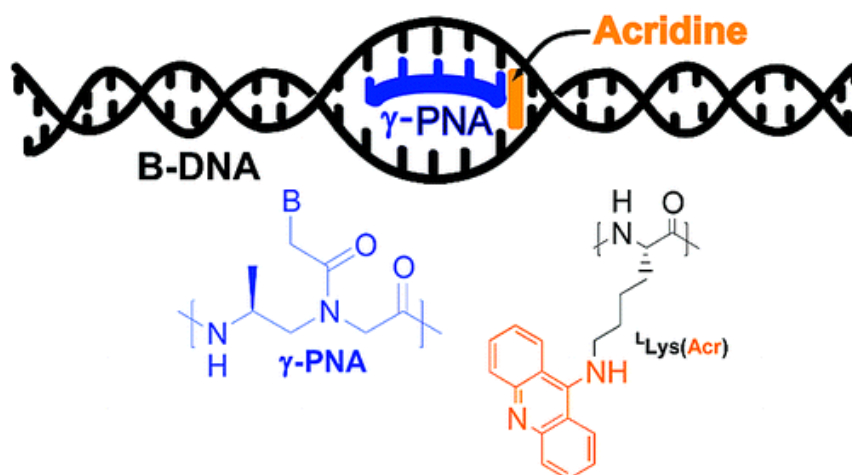


Figure 40: Strand invasion by PNA-acridine conjugate. The PNA (blue) invades the B-form DNA duplex, binding to the complementary strand via Watson-Crick base pairing. The mechanism by which the acridine (orange) stabilises the structure is not known; however, it is thought to be through interaction with the face of the exposed duplex base pair. Figure reproduced from Rapireddy *et al.*²³³

In 2003, Mokhir *et al.* developed PNA-naphthalene diimide (NDI) conjugates joined by polyamide linkers which were designed to improve binding to complementary oligonucleotide sequences.²⁷⁶ Upon binding to the target DNA, the planar, aromatic NDI interacts with and stabilises the PNA:DNA duplex via π -stacking interactions. The NDI-PNA:DNA duplexes were significantly more stable than the control PNA:DNA duplexes, with a maximum ΔT_m of 11.6 °C in 50 mM NaCl. The stabilising effect of the NDI was highest when it was attached to the N-terminus of the PNA with a linker of 5 atoms. It is notable that the NDI-PNA conjugates stabilised DNA with a single terminal mismatch to a similar degree as the fully complementary sequence, i.e. both had a melting temperature of ca. 68 °C. However, when bound to DNA with a mismatch in the centre of the binding sequence, the stability of the NDI-PNA:DNA complex dropped to as low as 25.9 °C. These results suggest that NDI-PNA conjugates can have non-specific binding activity with a terminal mismatch but are highly sensitive to mutations in the centre of the sequence.

Small molecule conjugates have also been used to target G4 structures. In 2006, Casals *et al.* generated a library of acridine-oligonucleotide conjugates specific for telomeric DNA G4 (example structure shown in Figure 41A).²⁷⁷ The conjugates were incubated with synthetic G4-forming oligonucleotides containing human telomeric repeat sequences (GGGTTA)_n and additional flanking nucleotides complementary to the oligonucleotide conjugates. The G4 forming oligonucleotides were synthesised with a dye and quencher attached, allowing the stability of the G4 in the presence of the acridine-oligonucleotide conjugates to be measured through melting studies (Figure 41B). The conjugates improved the stability of the G4, with a ΔT_m of up to 6.5 °C at a concentration of 10 μ M. However, this study did not compare the stabilising effects of the conjugates with those of the small molecules by themselves, so there is no clear evidence that the oligonucleotides improved the affinity of the ligands. In addition, the impact of oligonucleotide conjugation on the specificity of the ligands was not tested.

This would have required the ligands to be incubated with a quadruplex with a scrambled flanking region to identify whether the ligands had a greater effect on the stability of the G4 with a complementary flanking sequence when compared to the scrambled sequence.

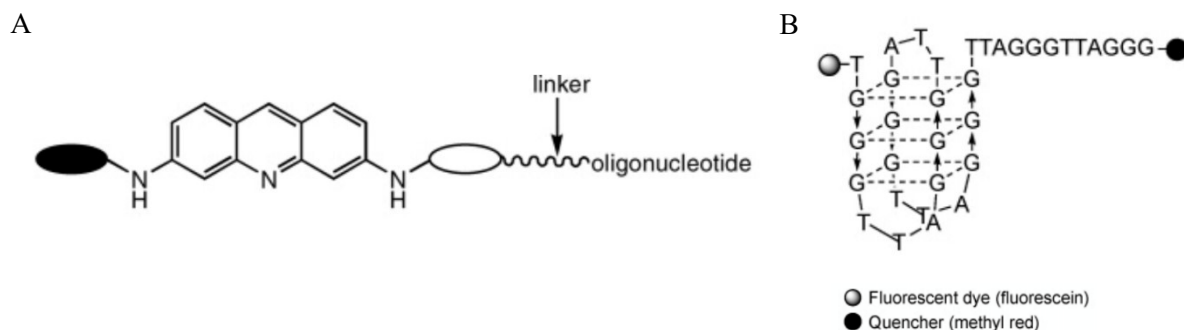


Figure 41: Example structure for acridine-oligonucleotide conjugates and G4 forming oligonucleotide. A) The acridine is conjugated to the oligonucleotide (complementary to the flanking region of the target quadruplex) with a flexible linker. B) G4-forming oligonucleotide formed from the (GGGTTA) repeat. Four repeats are required to form the G4, an additional repeat along with a final GGG tract is included to allow binding of the complementary oligonucleotide on the acridine-oligonucleotide conjugate. The G4-forming oligonucleotide includes a fluorescent dye and quencher which allow measurement of the melting temperature of the G4. When the G4 is formed the quencher and dye are brought into close proximity, quenching the fluorescence. At higher temperatures the G4 denatures and the dye and quencher move further apart, resulting in an increase in the fluorescent signal. Figure reproduced from Casals *et al.*²⁷⁷

There have been relatively few studies investigating targeting of RNA using oligonucleotide-small molecule conjugates when compared to targeting of DNA. Most of the studies that do exist focus on conjugation as a method to improve cellular uptake of the oligonucleotide. For example, Das *et al.* demonstrated improved bioavailability for PNA-aminosugar conjugates.²⁷⁸ They also studied the specificity of the PNA-aminosugar conjugates, however this was a relatively imprecise experiment as specificity was not compared to the PNA with the aminosugar conjugate but compared to a scrambled PNA-aminosugar sequence and tested on non-target RNA (featuring either a deletion removing the target region or a different piece of RNA). There are also a number of studies investigating the application of DNA-acridine conjugates for targeted cleavage of RNA strands, an example is a study by Kuzuya *et al.* in 2002.²⁷⁹ However, these focus on the effect of adding the DNA sequences to acridines on RNA cleavage and do not assess the impact of conjugation on binding.

In comparison to DNA G4, the structure and function of only very few examples of biologically relevant RNA G4 have been fully characterised in the literature.²⁸⁰ In order to validate our strategy of “sequence + structure” targeting with PNA-small molecule conjugates, the well-characterised RNA quadruplex located in the 5' UTR region of the *NRAS* gene was used as a model system.^{116, 190} *NRAS* has already been validated as a therapeutic target, as its suppression has been shown to induce apoptosis in human melanoma cells.²⁸¹ In this Chapter we describe how we have designed and synthesised a small library of PNA-NDI conjugates that are directed against the *NRAS* RNA quadruplex, and have assessed their effect on the translation of mRNA under the control of the *NRAS* 5'UTR *in vitro*. In addition, binding

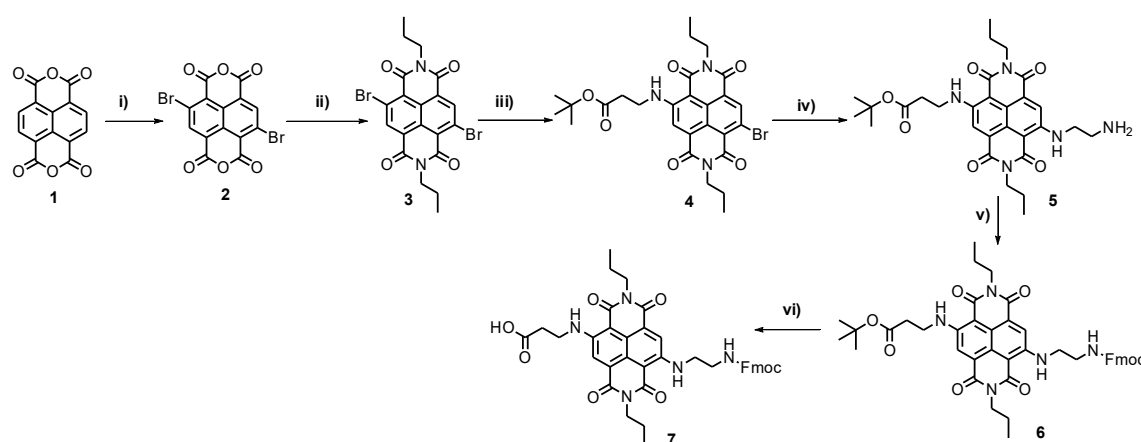
studies were undertaken to gain further insight into the specific and non-specific interactions between PNA-NDI ligands and their target RNA.

2.2. Methods

2.2.1. Ligand Design and Synthesis

Within all the quadruplex binding scaffolds available in the literature, we have chosen to use an NDI derivative since their quadruplex end-stacking binding mode has been clearly established with both DNA and RNA G4 (as previously explained in section 1.8.2). As part of a scientific collaboration with the group of Prof. Mauro Freccero (Pavia University, Italy) an Fmoc-protected NDI amino acid derivative was kindly synthesised by Dr Claudia Percivalle, (at that time, an ERASMUS-funded visiting PhD student from Pavia University). (See Figure 42 for synthetic scheme.)

General synthetic scheme



*i) dibromoisocyanuric acid (DBI), H_2SO_4 , reflux, 12 h, yield 93%; ii) glacial acetic acid, *n*-propylamine, reflux 1h, yield 45%; iii) *H*-β-Ala-OtBu-HCl, DCM, DIPEA, N_2 , 40°C, 40h, yield 70%; iv) ethylene diamine, $CHCl_3$, N_2 , 60°C, 24h, yield 68%; v) Fmoc-Cl, DCM, TEA, N_2 , 0°C (10 min) → rt (4h), 80% yield; vi) TFA, TIPS, DCM, 0°C (10 min) → rt (2h), yield 92%.*

Figure 42: General synthetic scheme for Fmoc protected NDI synthesis.

This molecule containing a free carboxylic acid and an Fmoc-protected primary amine, was incorporated into PNA oligomers as any other standard amino acid or PNA monomer using microwave-assisted solid-phase peptide synthesis and Fmoc Chemistry.

(i) PNA-NDI Synthesis

PNA monomers were purchased from Link Technologies Ltd (Scotland) and other Fmoc-protected amino acids from Merck-Millipore. PNA-NDI conjugates were synthesised on solid support using Polystyrene wang resin (0.43 mmol/g loading) (Merck Millipore). Microwave-assisted reactions on solid phase were performed on a Biotage initiator + SPWave Synthesiser using Fmoc chemistry. Typical PNA synthesis proceeds in two steps as described below: (1) amide coupling between amine functionalised resin and Fmoc-protected amino acid activated in solution, followed by (2) Fmoc deprotection to enable subsequent attachment of the next amino acid.

Coupling of Fmoc-protected amino acids and PNA monomers were achieved by adding a mixture of 5 equiv. of monomer, 5 equiv. of oxyma and 5 equiv. of DIC in DMF onto the resin and heating under vortex at 75 °C for 20 minutes. Coupling of the NDI derivative on resin was achieved by using HATU/DIPEA (5 eq: 5 eq: 10 eq) in DMF at room temperature over 45 minutes. Fmoc deprotection was done by mixing the resin with a solution of 20% Piperidine in DMF and vortexing for 10 minutes at room temperature. After each new reaction (coupling or Fmoc deprotection), the resin was washed with DMF (2 x 5 ml), MeOH (2 x 5 ml) and DCM (2 x 5 ml). After each stage the presence of free primary amino groups was assessed using the Kaiser test. A small number (<10) of beads were transferred to a glass vial and a drop of the following solutions added sequentially: (1) phenol, 80% in ethanol, (2) KCN in H₂O/pyridine and (3) ninhydrin, 6% in ethanol. The reaction of ninhydrin with free primary amide turns the solution and beads blue indicating successful deprotection. If the solution or beads turn blue after a coupling step, this indicates that free (uncoupled) amines remain, in which case a second coupling of the same amino acid is necessary.

(ii) Cleavage and purification

Functionalized PNAs were cleaved from their solid support by treatment with a solution of Trifluoroacetic acid (TFA): Triisopropylsilane (TIS): double distilled water (95: 2.5: 2.5, v/v) for 90 minutes at room temperature. Upon addition of diethyl ether, a precipitate formed which was collected by centrifugation and washed twice with diethyl ether. The solid was then solubilised in water and subsequently purified by reverse phase HPLC, using a gradient of water/acetonitrile both containing 0.1% of TFA with a Phenomenex C18 Gemini-NX 110 Å column (250 x 10.0 mm). Fractions containing the pure PNA conjugates were then lyophilized to afford the desired products as dried solids. The identity and purity of all PNA conjugates were checked by MALDI-TOF recorded on a MALDI micro MX instrument using sinapinic acid as a matrix and/or positive mode electrospray ionization mass spectrometry.

Structures of all the PNA-NDI conjugates synthesised are summarised in section 2.3.1. Example structures and sequence shown below in Figure 43 and Table 2.

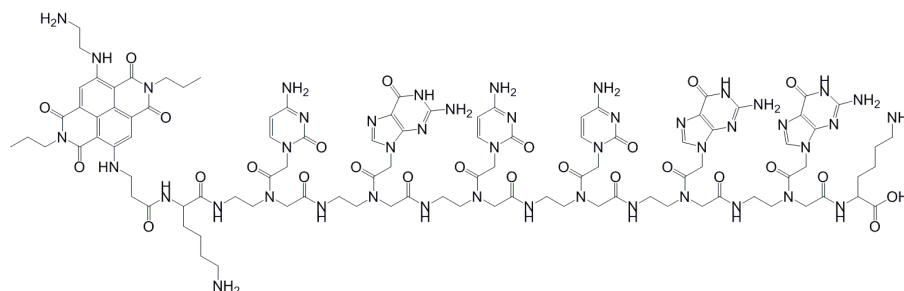


Figure 43: Chemical structure of 6LO_{HALF} (NDI-Lys-CGCCGG-Lys).

Table 2: Sequences of all PNA-conjugates synthesised and of their RNA target. mRNA is expressed from 5' to 3' and PNA sequences from N-terminus to C-terminus. A PNA-free NDI derivative was also synthesised and used as a control. All ligands are shown in the 5' to 3' orientation, PNA sequences complementary to the 5'-flanking region of the *NRAS* RNA are shown in blue, and PNA sequences complementary to the 3'-flanking region are shown in red.

Ligand	Sequence
<i>NRAS</i> RNA	5'- CCCGUG -UGGGAGGGGCGGGUCUGGGU- GCGGCC -3'
6LO	GGGCAC -Lys-NDI-Lys- CGCCGG -Lys
5LO	GGCAC -Lys-NDI-Lys- CGCCG -Lys
4LO	GCAC -Lys-NDI-Lys- CGCC -Lys
4LOa	Lys- CCGC -Lys-NDI-Lys- CACG
3LO	CAC -Lys-NDI-Lys- CGC -Lys
6LO Half	NDI-Lys- CGCCGG -Lys
“PNA”	CCGC -Lys-NH(CH ₂) ₅ CONH-Lys- CACG -Lys
“NDI”	Lys-NDI-Lys

2.2.2. Translation Assays

(i) Plasmid construction

pUTR plasmid was kindly provided by Prof. S. Balasubramanian (Cambridge University, UK). The pUTR plasmid is a pUC18 (Invitrogen) vector, containing the full human genomic *NRAS* 5'UTR (258bp) fused with the firefly luciferase gene and under the transcriptional control of the minimal T7 promoter (5'-AAGCTTTAATACGACTCACTATAGGGACGTCCCGTGTGGGAGGGGCG-3'). Mutations were introduced using the QuikChange Mutagenesis kit (Agilent) following the manufacturers protocol and primers were purchased from Invitrogen and designed using Quikchange primer design software (Agilent).

(ii) Primer Design

Mutation of the full 6-nucleotide flanking region with one primer required increased primer length and increased the likelihood of an incomplete mutation. Therefore, to improve primer accuracy, mutations in the flanking regions were introduced three nucleotides at a time. The primer sequences used to introduce the mutations are shown in Table 3. Primers Mut A and Mut B were designed to mutate the 5'-flanking region of the G4. The sense and antisense Mut A primers were applied first and introduced a GTG to ACA mutation in the final nucleotides of the sequence. The Mut B primers were designed to target this mutated sequence and introduce a second three nucleotide mutation, changing the first three nucleotides of the 5'-flanking region from CCC to TTT. Mut C and Mut D sense and antisense primers introduced sequential mutations in the 3'-flanking region in a similar manner, with Mut C introducing a GCG to ATA mutation in the first three nucleotides, then Mut D introducing a GCC to ATT mutation in the final three. The sense and antisense Mut G primers were then used independently on the native sequence to introduce a GGG to AAA mutation in the quadruplex forming region to prevent quadruplex formation.

Table 3: Primers for pUTR mutagenesis. The sequence of the native *NRAS* mRNA is included and the mutation sites are labelled and highlighted. Sequences for the sense and antisense (*) DNA primers are then listed. Quadruplex forming sequence is underlined and mutated nucleotides are highlighted in colour by site; site A (dark green), site B (light green), site C (dark blue), site D (light blue) and site G (red)

mRNA Sequences

Native NRAS mRNA sequence

5'...CCCGUGUGGGAGGGGCGGGUCUGGGUGCGGCC...-3'

NRAS mRNA mutation sites

Site: B A G C D

5'...TTTACA-UGGGAGGGGCAAAUCUGGGU-ATAATT...-3'

DNA Primer Sequences

Mut A	5'-CACTATAGAAACGTCCC <u>ACA</u> T <u>GGGAGGGGCGGGTCTG</u> -3'
Mut A*	5'- <u>CAGACCCGCCCTCCCA</u> <u>TGT</u> GGGACGTTTCTATAGTG-3'
Mut B	5'-GCTTTAATACGACTCACTATAGAAACGT <u>TTTACA</u> <u>TGGGAGGGGCGG</u> -3'
Mut B*	5'- <u>CCGCCCCCTCCCA</u> <u>TGTAAA</u> ACGTTTCTATAGTGAGTCGTATTAAAGC-3'
Mut C	5'- <u>GGGAGGGGCGGGTCTGGGT</u> <u>ATA</u> GCCTGCCGCA-3'
Mut C*	5'-TGCGGCAGGC <u>TAT</u> <u>ACCCAGACCCGCCCTCCC</u> -3'
Mut D	5'- <u>GAGGGGCGGGTCTGGGT</u> <u>TATAATT</u> TGCCGCATGACTC-3'
Mut D*	5'-GAGTCATGCGGCA <u>AATTAT</u> <u>ACCCAGACCCGCCCTC</u> -3'
Mut G	5'-CCGTGT <u>GGGAGGGGC</u> <u>AAA</u> <u>TCTGGGT</u> TGCGGCCTG-3'
Mut G*	5'-CAGGCCGCACCCAGA <u>TTT</u> GCCCCTCCACACGG-3'

(iii) PCR

PCR of the original plasmids in the presence of the mutated primers produced amplified products containing the desired mutations. Briefly, reaction mixes containing pUTR, dNTP mix, forward and reverse primers and reaction buffer were cycled in a PCR thermocycler with 1 µl *Pfu*Turbo DNA polymerase. Following PCR, the reaction mix was then incubated with 1 µl Dpn-1 restriction enzyme at 37 °C for 1 hour to digest parental (non-mutated) supercoiled dsDNA.

(iv) Plasmid selection and amplification

Mutated plasmids were amplified through transformation into bacterial cells. XL1-Blue super competent cells were transformed with the amplified plasmids on agar plates containing ampicillin and incubated at 37 °C and 5% CO₂ for >16 hours. Plasmids contained the ampicillin resistance gene to enable the selection of successfully transformed cells. To isolate the DNA, individual colonies were selected using a pipette tip then incubated with 1-5 ml of LB medium overnight at 37 °C and 5% CO₂ with shaking at 200-225 rpm. The bacterial culture was then pelleted by centrifugation at >8000 rpm (6800 x g) for 3 minutes at room temperature and the resulting pellet re-suspended, lysed and purified using a QIAprep Mini-Prep Kit (Qiagen). The purified DNA was finally eluted in 50 µl nuclease-free water (Thermo). After each set of mutations, the mutated plasmids were sequenced by GATC Biotech to confirm the mutagenesis was successful. Plasmids were then run on 1% agarose gels to check size, purity and integrity.

(v) Plasmid linearisation

Plasmids were linearised through incubation with 1 µl *Eco*R1 restriction enzyme (Invitrogen) at 37 °C for 1 hour. Linearised plasmids were purified using ethanol precipitation; 1/10th volume of sodium acetate, pH 5.2, was added to the mixture, followed by mixing and addition of 2-2.5 volumes of cold absolute ethanol. This mixture was stored at -20 °C for 20 minutes and then spun at maximum speed in a microcentrifuge for 10-15 minutes. The supernatant was removed and the pellet washed in 70% ethanol, air-dried, then re-suspended in nuclease free water.

(vi) *In vitro* transcription

5'-capped transcripts were synthesised from linearised plasmids *in vitro* using the TranscriptAid T7 High Yield Transcription Kit (ThermoScientific). Briefly, plasmids were incubated with Ribo m7G cap Analogue (Promega), nucleotide mixes and TranscriptAid enzyme mix and incubated at 37 °C for 2

hours. Transcripts were then treated with DNase A to digest the parental DNA and the RNA purified using a GeneJet RNA purification kit (ThermoScientific). RNA concentration was determined using a Nanodrop and the integrity and size of each transcript confirmed on 1% agarose gel.

(vii) *In vitro* translation

Translation of purified mRNA was carried out in a cell-free translation system consisting of nuclease-treated rabbit reticulocyte lysate (Promega). A 10 µl reaction volume was created containing 200 ng RNA, 70% (v/v) rabbit reticulocyte lysate, 10 µM amino acid mixture minus leucine, 10 µM amino acid mixture minus methionine and 2 units RNasin ribonuclease inhibitor and incubated at 30 °C for 90 minutes.

(viii) Luciferase activity

Following translation, *firefly* luciferase activity was measured on a BMG Biotech Fluorostar Omega microplate reader in luminescence mode. 50 µl luciferase assay reagent (Promega) was added to 4.5 µl sample using an autodispenser and the luciferase light intensity measured for 10 seconds following a 2 second delay. Each sample was measured in duplicate, and each experiment repeated at least three times across two or more days and with two or more separate batches of RNA.

(ix) Data analysis

Results are expressed as a percentage of the positive control (*in vitro* translation of RNA in the absence of ligand) and normalised using a negative control (translation mix in the absence of RNA). Graphs and statistical analyses were compiled in GraphPad Prism.

2.2.3. Binding Assays

(i) CD Spectroscopy

The following HPLC-purified RNA oligonucleotides were purchased lyophilised from Eurogentec and used without further purification. See below for sequences, mutated regions in the Mut G and Mut ABCD sequences are underlined.

UTR:	5'-CCCGUGUGGGAGGGGCGGGUCUGGGUGCGGCC-3'
Mut G:	5'-CCCGUGUGGGAGGGGCAAAUCUGGGUGCGGCC-3'
Mut ABCD:	5'-UUUACAUGGGAGGGGCGGGUCUGGGU <u>AUAAUU</u> -3'

RNA samples were prepared at 10 μM final strand concentration in 10 mM Tris-HCl, pH 7.4 buffer and varying concentrations of KCl, ranging from 0-100 mM. Prior to spectroscopy, RNA was folded by heating to 95 $^{\circ}\text{C}$ for 5 minutes and leaving to cool to room temperature overnight.

CD was conducted using a J-380 Spectrometer (Jasco). Unless otherwise stated, RNA samples were measured in cuvettes with a 2 mm optical path length and at a final strand concentration of 10 μM . Spectra were recorded over a wavelength range of 200-320 nm at a continuous scanning speed of 100 nm min^{-1} , with a response time of 4 seconds, 0.1 nm pitch and 2 nm bandwidth. For each sample, three scans were recorded and an average taken. Data was baseline corrected with a blank consisting of 10 mM Tris-HCl and the relevant KCl concentration in RNase-free water and the final spectra were normalised to zero at 320 nm. Graphs were produced in GraphPad Prism.

(ii) Fluorescence - Electrophoresis Mobility Shift Assay (F-EMSA)

HPLC-purified Fluorescently-labelled RNA oligonucleotides of sequences below were ordered from either Eurogentec or IBA and used without further purification, mutated regions in the Mut ABCD sequences are underlined.:

FAM-UTR:	(6-FAM)-CCCGUGUGGGAGGGGCGGGUCUGGGUGCGGCC-3'
Cy3-UTR:	(Cy3)-CCCGUGUGGGAGGGGCGGGUCUGGGUGCGGCC-3'
Cy5-UTR:	(Cy5)-CCCGUGUGGGAGGGGCGGGUCUGGGUGCGGCC-3'
Dy781-UTR:	(Dy781)-CCCGUGUGGGAGGGGCGGGUCUGGGUGCGGCC-3'
FAM-Mut ABCD:	(6-FAM)-UUUACAUGGGAGGGGCGGGUCUGGGU <u>AUAAUU</u> -3'
Cy3-Mut ABCD:	(Cy3)-UUUACAUGGGAGGGGCGGGUCUGGGU <u>AUAAUU</u> -3'
Dy781-Mut ABCD:	(Dy781)-UUUACAUGGGAGGGGCGGGUCUGGGU <u>AUAAUU</u> -3'
Cy5-Mut ABCD:	(Cy5)-UUUACAUGGGAGGGGCGGGUCUGGGU <u>AUAAUU</u> -3'

The concentration of RNA was measured on a NanoDrop 2000C (ThermoScientific) and integrity checked by running on a 1% agarose gel. To form the quadruplex, 10 μM RNA was heated to 95 $^{\circ}\text{C}$ for 5 minutes with 1 mM KCl and 10 mM Trizma (pH 7.4), left to cool overnight to room temperature and then stored at 4 $^{\circ}\text{C}$ until used.

Set ratios of ligand were added to pre-folded RNA (at a final concentration of 4 μM) and left to incubate for 1 hour at room temperature. 2.5 μl of 80% Glycerol was added to the RNA:ligand solution and the mixture run on a 1.1% agarose gel for 45 minutes at 100V in a Thermo Owl horizontal electrophoresis system or on a pre-formed 20% polyacrylamide gel for 90 minutes at 150V in a Thermo Xcell Sure-lock mini-cell electrophoresis tank. The gels were kept on ice to prevent degradation of the RNA due

to heating as the experiment progressed. Gels were imaged using a Typhoon FLA 9500 (GE Life Sciences). Images in TIFF format were analysed using gel analysis tools on FIJI (ImageJ) image analysis software and graphs produced using Prism (Graphpad).

(iii) Fluorescence spectroscopy

The emission spectra of 400 μ l of pre-folded, fluorescently-labelled 10 μ M *NRAS* RNA was measured on a Fluorolog fluorometer (Jobin Yvon Horiba Inc.). The respective excitation and emission wavelength settings for the different fluorophores were; fluorescein (450 nm, 465 nm-750 nm), Cy3 (540 nm, 550 nm-750 nm), Cy5, (640 nm, 650 nm-750 nm) and Dy-781 (760, 770 nm-850 nm). Fluorescence emission spectra were then recorded following addition of ligand, starting at a concentration of 0.0625 μ M and increasing sequentially up to 4 μ M. The solution was mixed and left at room temperature for 4 minutes following each addition of ligand before recording commenced. Data was plotted and analysed in Prism (GraphPad).

2.3. Results

2.3.1. PNA-NDI conjugates design and synthesis

PNA-NDI conjugates were designed in order to test the effect of PNA conjugation, PNA length and PNA strand orientation on ligand activity (i.e. ability to inhibit mRNA translation *in vitro*). A tetra-substituted NDI derivative was chosen as the central G4-stabilising small molecule.

Hybrid molecules of general structure PNA₁-Heterocycle-PNA₂ have been synthesised where Heterocycle and PNA_{1,2} are predicted to recognise the quadruplex structure and the two single-stranded flanking RNAs, respectively. These PNA-heterocycle conjugates will exploit the PNA-RNA sequence specific base-pairing association to direct the quadruplex binding platform towards the only quadruplex of interest i.e. that located in between the two single-stranded flanking regions.

(i) G4-binding heterocycle

Within all the quadruplex binding scaffolds available in the literature, we have chosen to use NDIs since their quadruplex-binding properties are well established. The DNA binding properties of NDI derivatives have been studied extensively during the past ten years and demonstrated the potential of this heterocyclic scaffold for targeting RNA quadruplexes,⁷³ as well as DNA quadruplexes.¹⁸⁸ Studies have also shown how the nature of side chains can influence binding of NDI derivatives to RNA G4.

For example, Collie *et al.* demonstrated that longer, bulkier side chains can interfere with the binding of NDI to RNA G4.⁷³ Based on the results of these studies and in the absence of further structural information (although modelling of PNA-NDI conjugates complexed to G4 RNA was considered as a potentially attractive predictive tool early on during this project, it was considered as extremely challenging by experts in the field), we have chosen an NDI functionalised with four short, non-bulky and flexible side chains. The shorter side chains will minimise the risk of steric hindrance interfering with PNA binding.

(ii) Water solubility

Lysine amino acids were introduced at the start and end of each PNA and also before and after the NDI to ensure these otherwise highly hydrophobic PNA-NDI conjugates are hydro-soluble. Their amino side-chains, protonated at physiological pH, could also participate in electrostatic interactions with the RNA phosphate backbone, thus providing additional stabilisation of the PNA:RNA complexes. The number of lysine residues was kept to a minimum to avoid non-specific electrostatic interactions with any negatively charged nucleic acid while ensuring the PNA-NDI conjugates were fully soluble in aqueous solutions.

(iii) Final ligand design and synthesis

Three classes of molecules have been synthesised of general structures PNA₁-NDI-PNA₂ (1) (Figure 44), PNA₁-alkyl-PNA₂ (2) and PNA₁-NDI (3). Molecules of type 1 were expected to target simultaneously the core structure of the RNA quadruplex as well as its 3' and 5' single stranded flanking arms. These should be the tightest and most specific ligands that can potentially create π - π stacking interactions (via the NDI element), electrostatic interactions (via the lysine side-chains) and favourable H-bond and hydrophobic interactions (via the anti-sense PNAs) with the RNA target. The type 2 molecule (so-called “PNA” throughout this chapter) was used as a control to evaluate the effect of PNA hybridisation on quadruplex formation and stability. Indeed, hybridisation of PNA₁ and PNA₂ to the two flanking regions of the RNA quadruplex will induce a significant “packing” of the RNA strand that could be sufficient to favour quadruplex folding, even in the absence of any G-tetrad binder. The type 3 molecule (so-called 6LO_{HALF}) was synthesised to investigate the necessity of targeting both flanking regions to induce quadruplex specific recognition or whether one PNA is sufficient to direct the heterocyclic platform to its target in a sequence specific manner. The length of each PNA was varied between 3 and 6 bases (i.e. between 6 and 12 bases long per molecule). The effect of PNA strand orientation was also investigated. Finally, an NDI functionalised with two Lysine residues (so-called “NDI” throughout this chapter) was synthesised and used as a control to assess the impact of the G4 binding ligand alone, without any PNA attached.

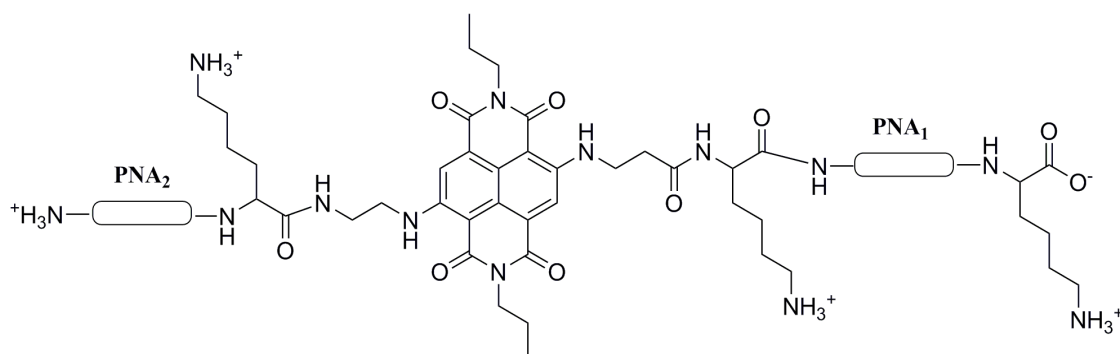


Figure 44: General chemical structure of PNA-NDI conjugates. The molecules consist of a central NDI ligand flanked by two PNA strands of variable length, labelled PNA₁ and PNA₂.

Following synthesis of the molecules (as described in the methods section), Mass Spectroscopy was run to confirm the mass of the products matched with the theoretical mass. The mass spec peaks identified for each ligand are summarised in Table 4. The structures of the ligands and mass spec results are shown in Figure 45 and Figure 46 respectively.

Table 4: Molecular Formulae and Mass Spectra of Ligands. The mass spec peaks identified for each PNA-NDI ligand, referenced against the masses calculated from the structures.

Ligand	Molecular Formula	[M + H] ⁺ calculated	[M + H] ⁺ found
3LO	C ₁₀₅ H ₁₄₅ N ₄₅ O ₂₆	2452.1	2452
4LO	C ₁₂₆ H ₁₇₁ N ₅₇ O ₃₂	2994.3	2994
5LO	C ₁₄₈ H ₁₉₇ N ₇₁ O ₃₈	3576.5	3580
6LO	C ₁₇₀ H ₂₂₂ N ₈₅ O ₄₄	4157.8	4157
4LOa	C ₁₂₆ H ₁₇₁ N ₅₇ O ₃₂	2994.3	2994
6LO _{HALF}	C ₁₀₀ H ₁₃₁ N ₄₅ O ₂₆	2380.4	2381
“PNA”	C ₁₀₇ H ₁₅₃ N ₅₃ O ₂₈	2630.2	2630
NDI	C ₃₇ H ₅₃ N ₉ O ₈	752.9	754

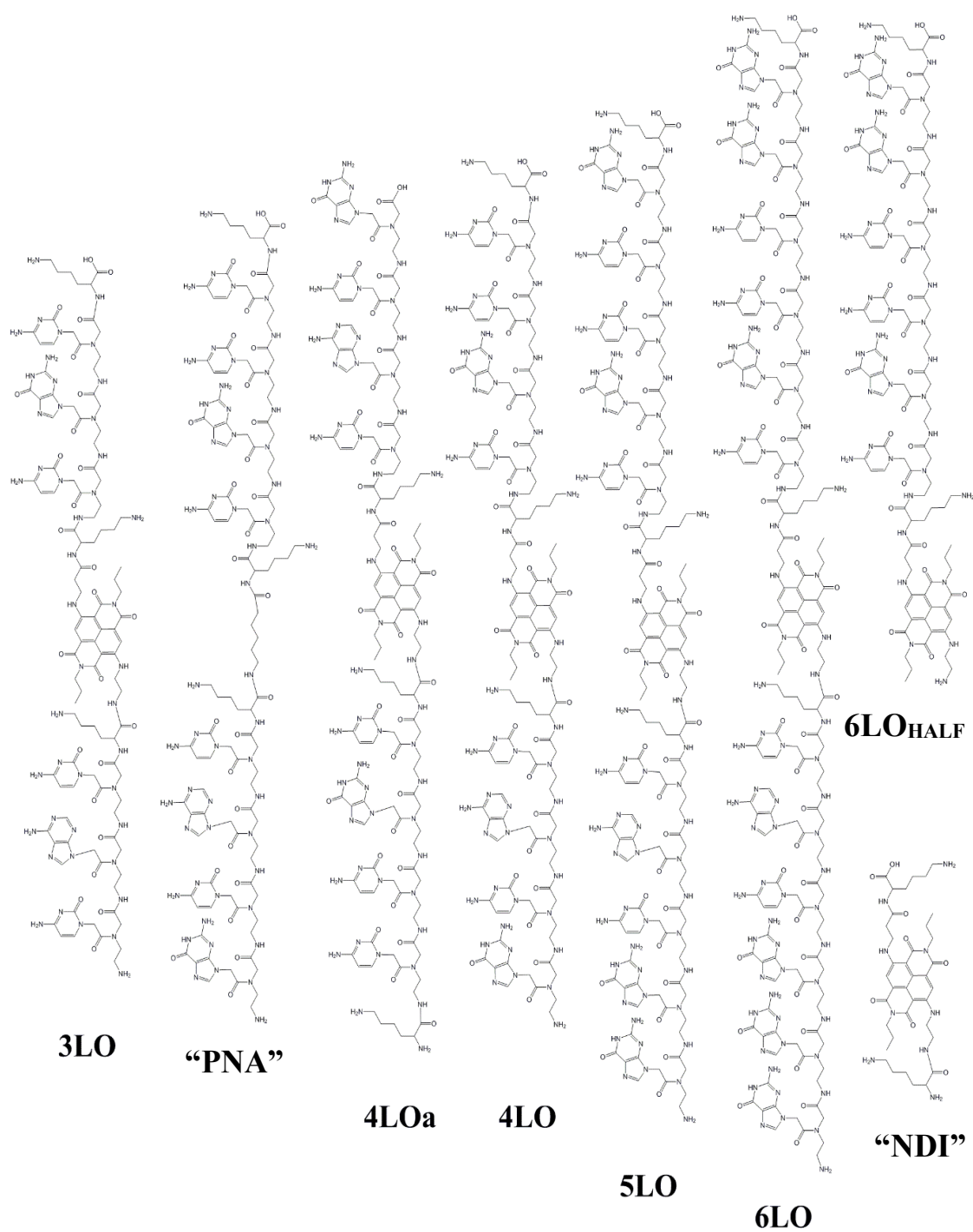
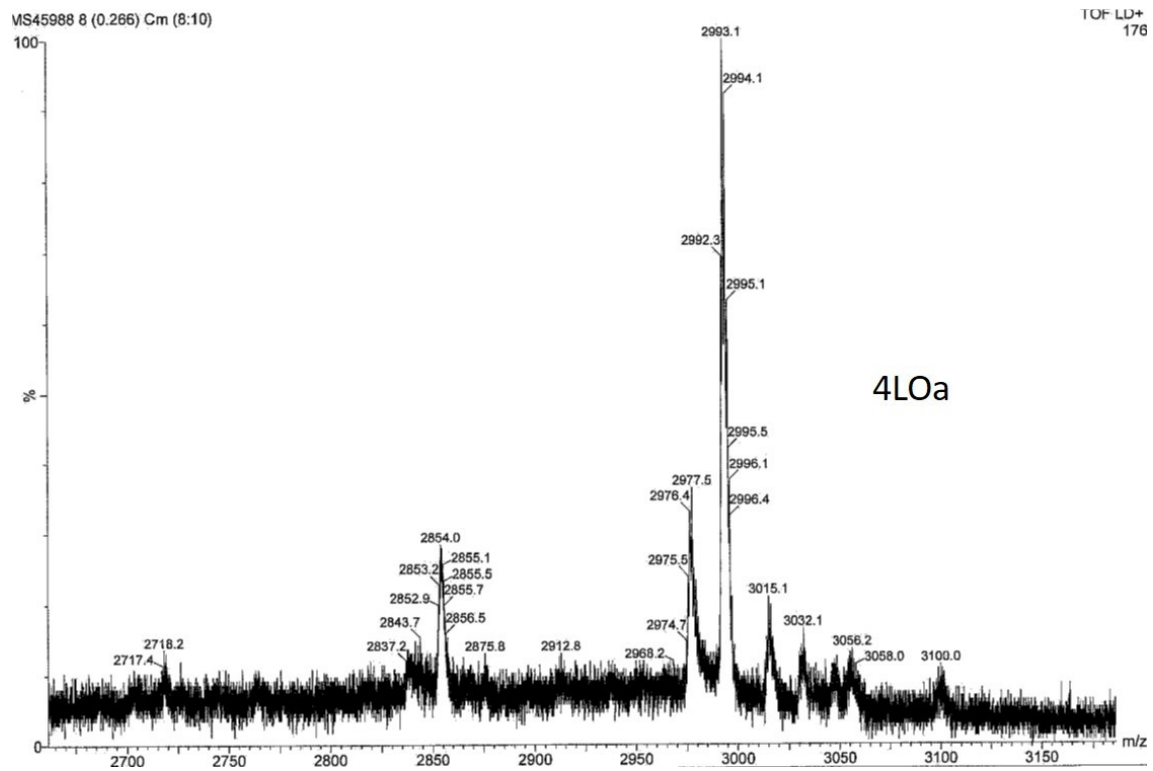
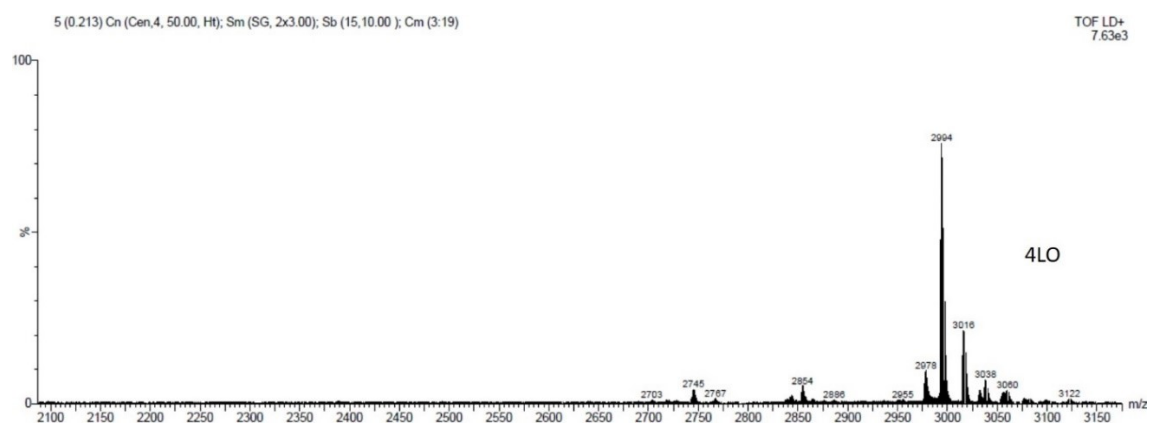
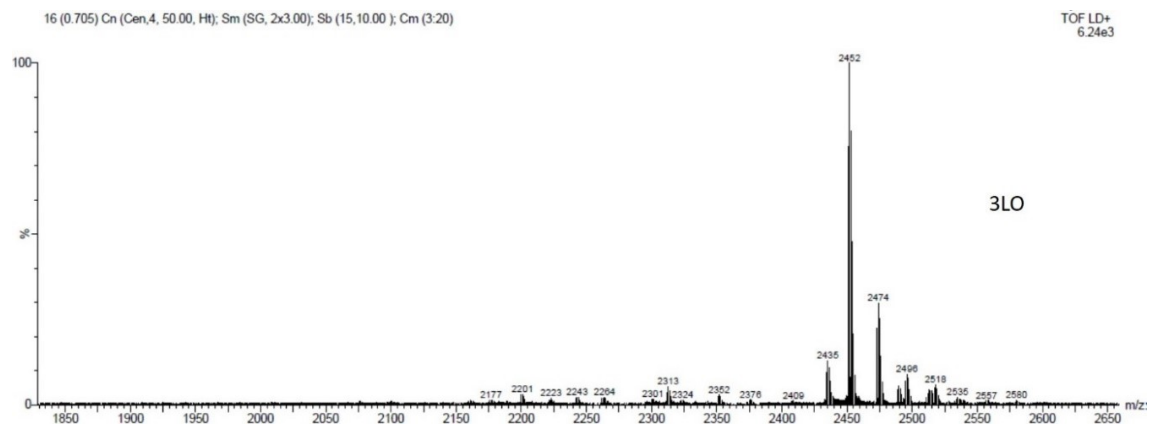
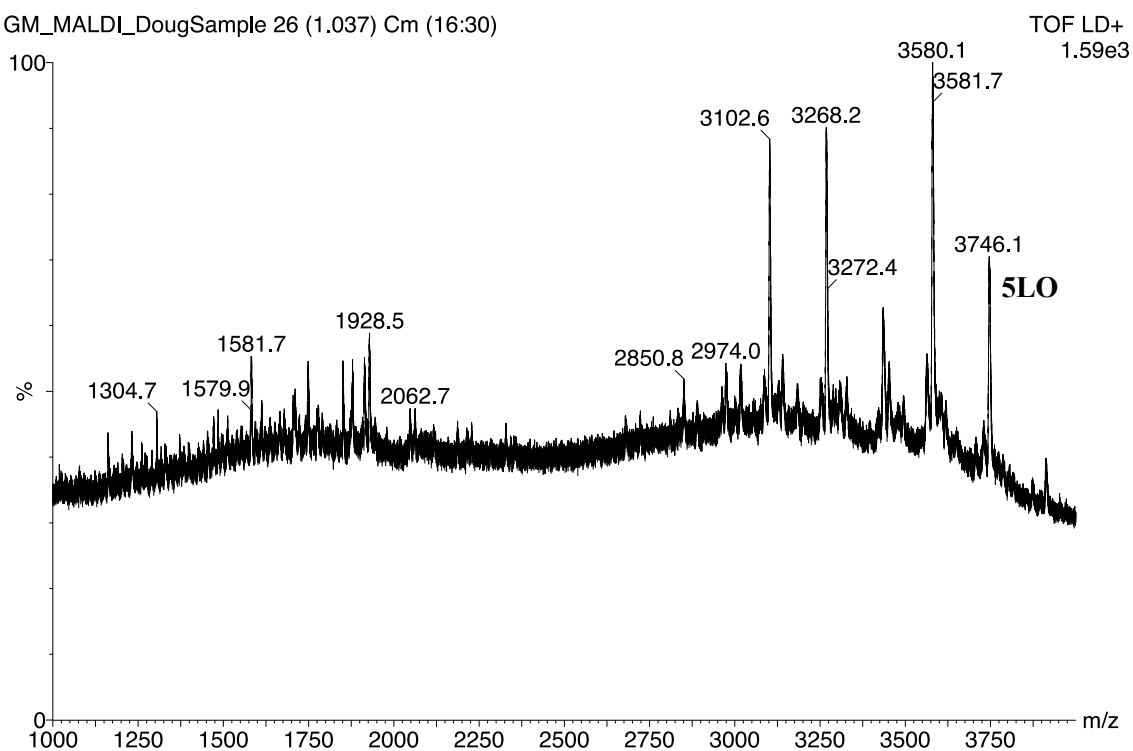


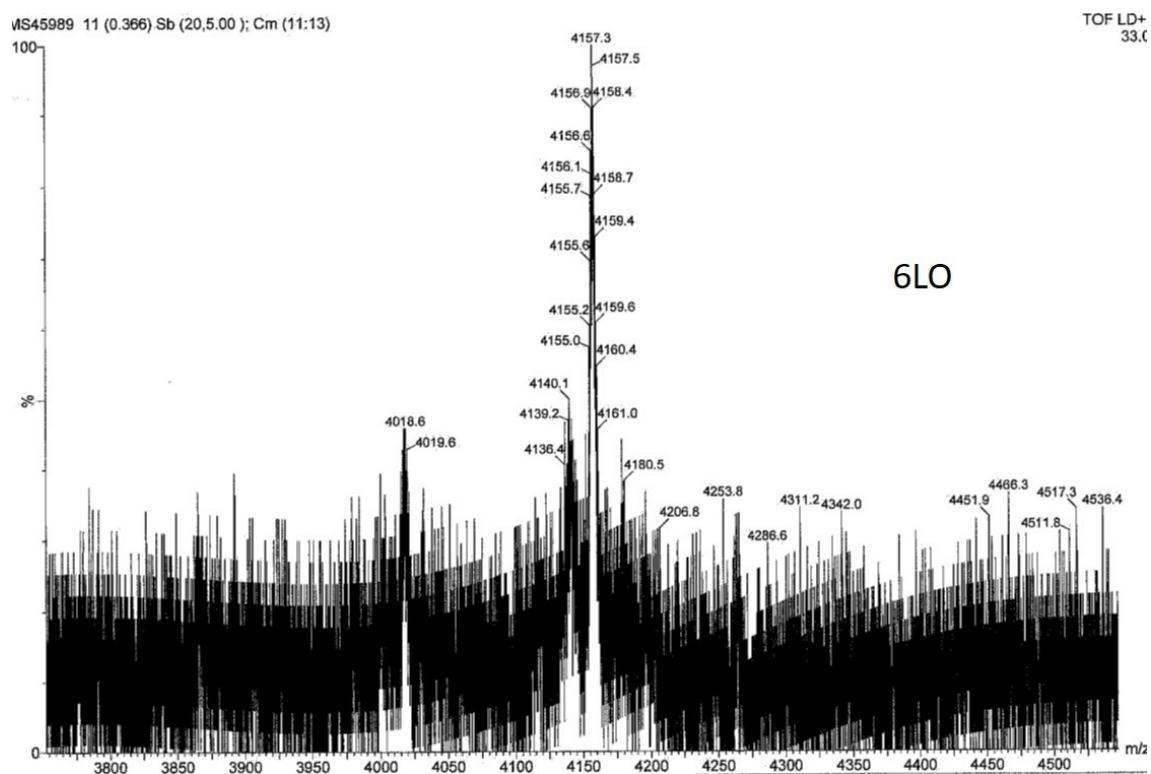
Figure 45: Structures of PNA small molecule conjugates.

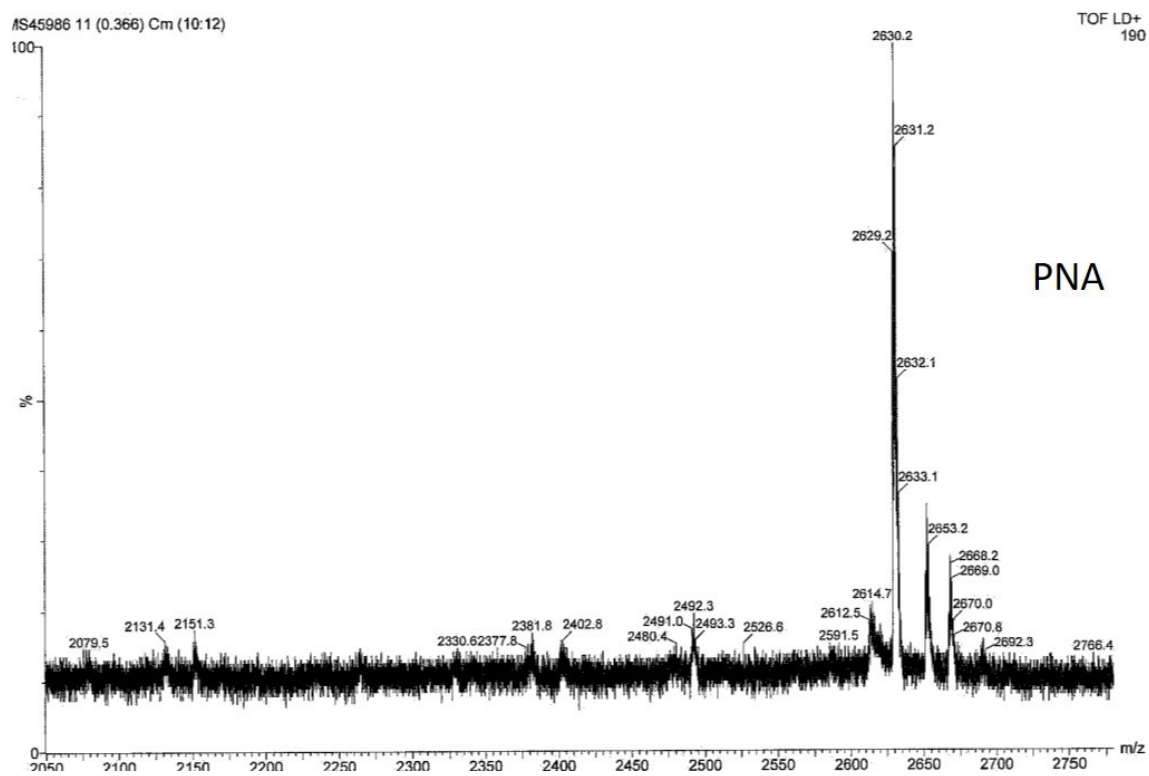
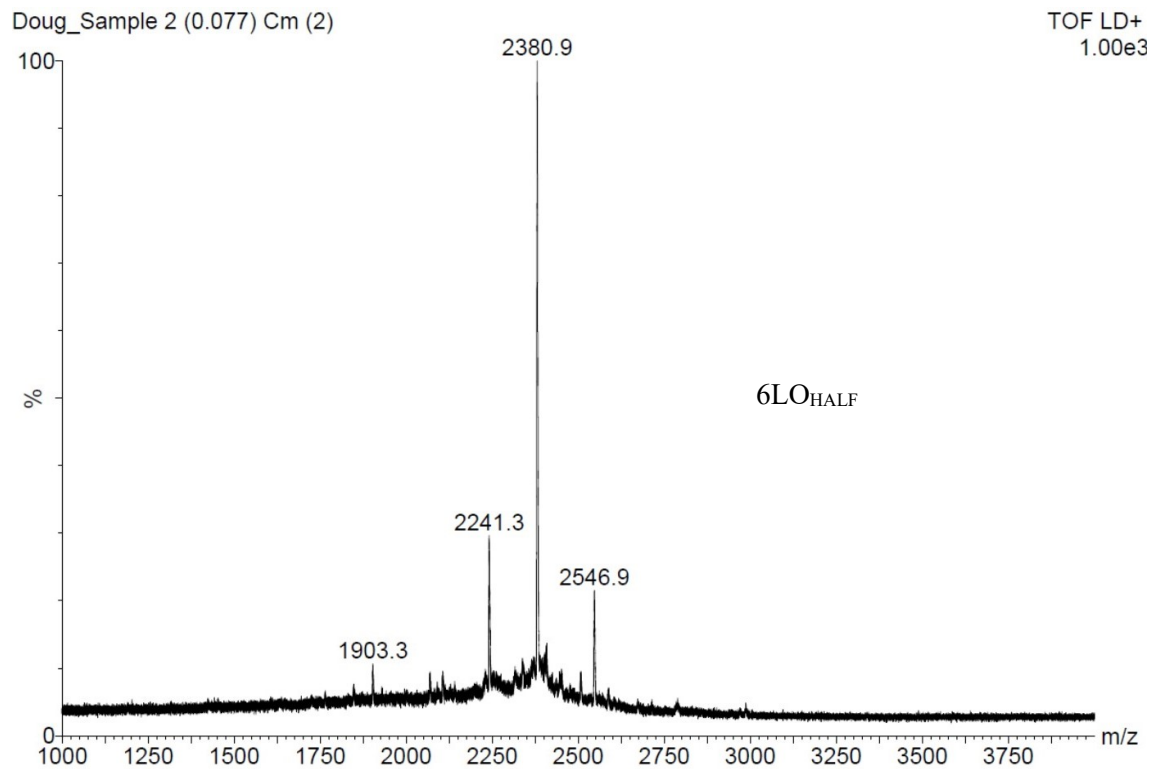


GM_MALDI_DougSample 26 (1.037) Cm (16:30)



MS45989 11 (0.366) Sb (20,5.00) ; Cm (11:13)





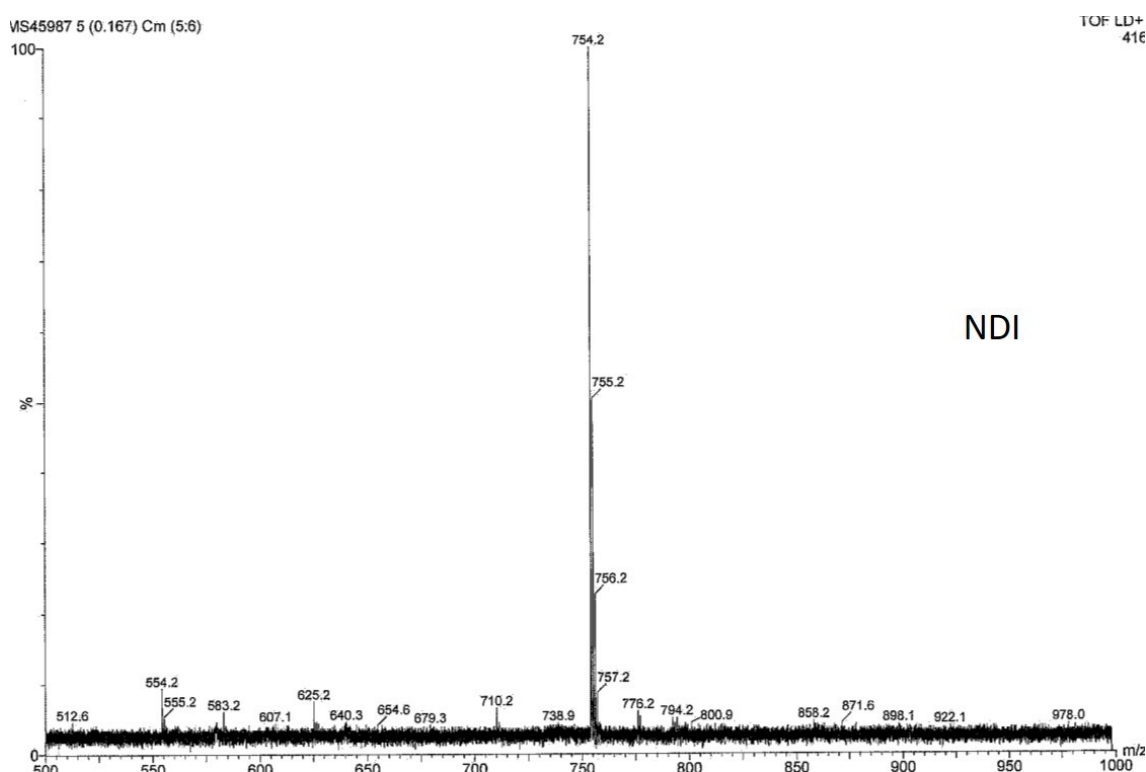


Figure 46: Mass spectroscopy (Maldi-TOF) characterisation of engineered PNA conjugates.

2.3.2. *In vitro* translation assay: Introduction

In vitro translation is a common method used to investigate the effect of 5'UTR G-quadruplexes on translation. The 5'UTR containing the quadruplex of interest, in this case the *NRAS* 5'UTR, is inserted into a plasmid upstream of a reporter gene and the mRNA expressed and purified. In a cell free system, translation of the reporter protein from the mRNA is facilitated by adding a cell lysate, typically rabbit reticulocyte lysate, and additional translation reagents. Expression of the reporter protein can then be used to quantify the effect of the quadruplex on translation.¹¹⁶ This technique can also be used to assess the effect of ligand-induced quadruplex stabilisation or destabilisation on translation by incubating the mRNA with ligand for half an hour before initiating translation.¹⁹⁰ The method for reporter protein quantification depends on the reporter gene chosen. The two most common reporter genes are luciferase and fluorescent proteins such as green or red fluorescent protein (GFP and RFP, respectively) which are quantified through measuring luminescence and fluorescence respectively (Figure 47).

To test the specificity and effectiveness of PNA-NDI conjugates as inhibitors of translation, the ligands were tested on mRNA constructs containing either the full native *NRAS* 5'UTR (referred to as *NRAS* 5'UTR throughout this chapter) or various modified constructs. A plasmid containing a GGG to AAA

mutation to prevent G4 formation (*NRAS* Mut G) was used to investigate the role of G4 formation in translation inhibition by the ligands (structure specificity). In a second plasmid, the G4 forming sequence remained intact but the flanking regions of the quadruplex were randomised to assess the impact of PNA-RNA duplex formation on ligand induced translational regulation (sequence specificity) (RNA from this plasmid is referred to as *NRAS* Mut ABCD)

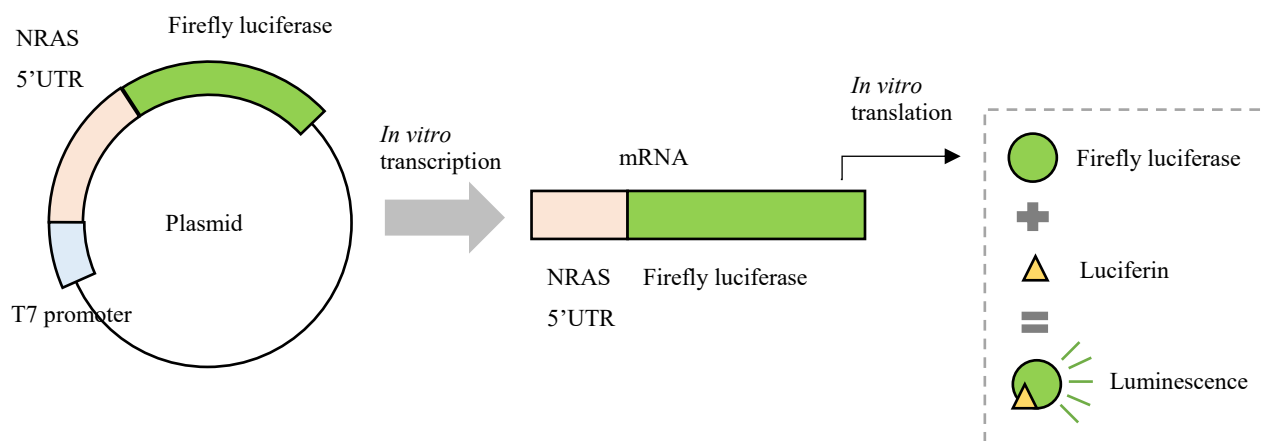


Figure 47: *In vitro* translation assay using firefly luciferase. A plasmid is designed incorporating the *NRAS* 5-UTR in front of the luciferase gene, under control of a T7 promoter. Following *in vitro* transcription, the purified mRNA undergoes *in vitro* translation in the presence of rabbit reticulocyte lysate to generate luciferase. The quantity of luciferase produced shows the effect of the 5'UTR on translation and is measured by mixing luciferin with the luciferase, thereby producing a luminescent signal which can be measured by a fluorometer.

2.3.3. *In vitro* translation assay: Optimisation

(i) Plasmid Mutagenesis

Following mutagenesis of the pUTR plasmid, the mutated plasmids were sent for sequencing to confirm successful incorporation of the mutation. An example chromatogram for the Mut A plasmid is shown in Figure 48. Electrophoresis was then used to confirm the molecular weight of the plasmids and check for successful linearisation (Figure 49). The linearised plasmids showed a band at around 4.5 kbp as expected. Circular plasmids are condensed and as such travel further on the gel, showing bands near the 3 kbp marker.

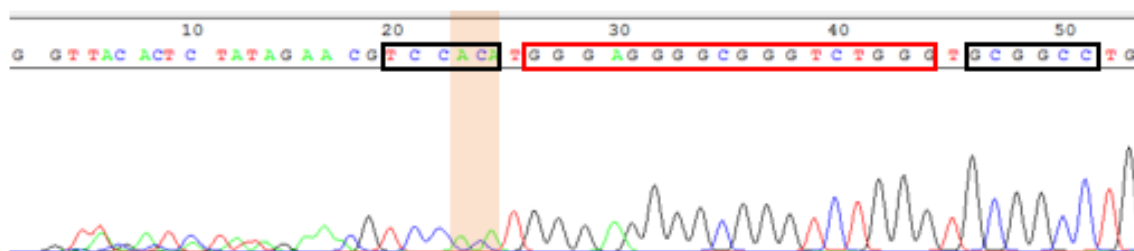


Figure 48: Excerpt from chromatogram showing sequenced Mut A plasmid following *in vitro* mutagenesis. Plasmid contains the *NRAS* 5'UTR G-quadruplex forming sequence (highlighted by red box) and the 5' and 3' flanking regions (highlighted by black boxes). The orange area shows the site of the successful mutation (GTG to ACA).

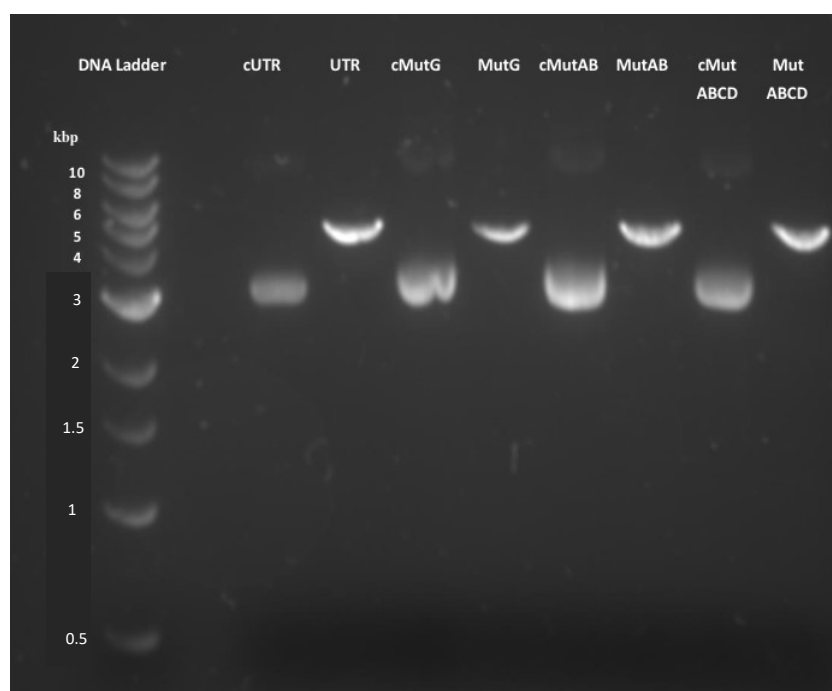


Figure 49: Electrophoresis of linearised (e.g. UTR) and circular (e.g. cUTR) plasmids following different stages of mutagenesis. Circularised plasmids formed less defined bands just behind 3kb marker. Linearised plasmids formed well defined bands of c.4.5kbp, demonstrating that there were no anomalous deletions or insertions during mutagenesis. The above gel shows checks for Mut G, Mut AB and Mut ABCD plasmids compared to the non-mutated UTR plasmid for reference. Samples were 5µl DNA mixed with 10µl of loading buffer and were electrophoresed on a 1% agarose gel alongside the DNA ladder.

During the process of introducing mutations, an inconsistency in the plasmid sequence map provided alongside the initial plasmids was uncovered. The sequence map contained a three-nucleotide variance when compared with the chromatograms produced following mutagenesis. This variance was in the T7 promoter region upstream of the G4 sequence. The impact of this variance was that the Mut A primers contained an extra mutation in the T7 region, resulting in reduced mutagenesis success rates. In addition, the mutated plasmids then contained an error in the T7 promoter which significantly reduced RNA translation. To correct this, new primers were ordered with the correct sequence and the mutagenesis process re-started.

(ii) Analysis of *in vitro* transcribed RNA

In previous studies researchers have purified RNA on a 1% agarose gel, using gel extraction kits to extract bands of the correct size.¹¹⁶ There are several concerns with this method. Firstly, extracting the RNA from the gel inevitably results in lower yields than other techniques. In addition, the number of additional steps in a gel mean that the likelihood of contamination is considerably greater, even when using RNAase free reagents. Finally, the process is very time consuming, which was a significant concern given the volume of RNA being produced for this study. In order to improve the process, a

number of alternative RNA purification assays were tested, including; phenol extraction and ethanol precipitation, GeneJet RNA purification kit (ThermoScientific) and RNeasy mini-kits. The simplicity of the ThermoScientific kit resulted in significantly greater yields than the other techniques, both through fewer incidences of contamination and reduced loss of RNA during the purification steps. This kit was therefore chosen for use throughout the experiments.

Following *in vitro* transcription of the plasmids, the size and integrity of the mRNA transcripts were verified on 1% agarose gel (Figure 50). Transcripts were only used if transcripts formed a single band around the 2000 nt marker as expected. This provides evidence that the transcripts were intact and of the correct length without any duplications or truncated regions.

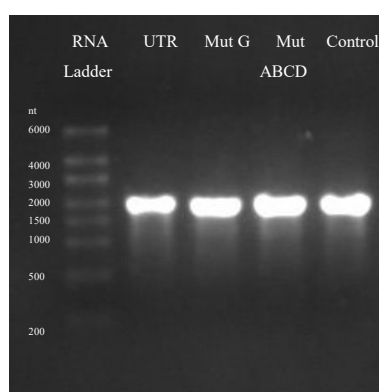


Figure 50: Electrophoresis of *in vitro* transcribed *NRAS* 5'-UTR, *NRAS* Mut G, *NRAS* Mut ABCD and control luciferase RNA. All of the transcripts form strong bands between the 1500 and 2000 nt markers as expected demonstrating that they are of the predicted length and have not produced truncated or elongated versions.

To test for G4 formation in the native and mutated *NRAS* 5'UTR, the CD spectra of short RNA oligonucleotides containing the *NRAS* quadruplex forming sequence (QFS) and flanking regions with the relevant mutations were assessed. CD spectra of both UTR and Mut ABCD RNA oligonucleotides folded in the presence of 100 mM KCl (optimal conditions for G4 formation) exhibited a clear positive peak centred at 263 nm and a negative peak centred at 240 nm which is the characteristic signature of an all-parallel G-quadruplex. In comparison, no such signature was observed in the spectrum for Mut G RNA confirming that the three G→A mutations do indeed successfully prevent folding into a G4 conformation. On the other hand, mutations of the flanking regions do not affect G4 formation; Mut ABCD still forms an all-parallel quadruplex under high salt concentrations (Figure 51). These results fit with previous work by Collie *et al.* which also concluded that the flanking sequence of 22-, 23- and 24-mer telomeric RNA G4 do not affect their topology.²⁸²

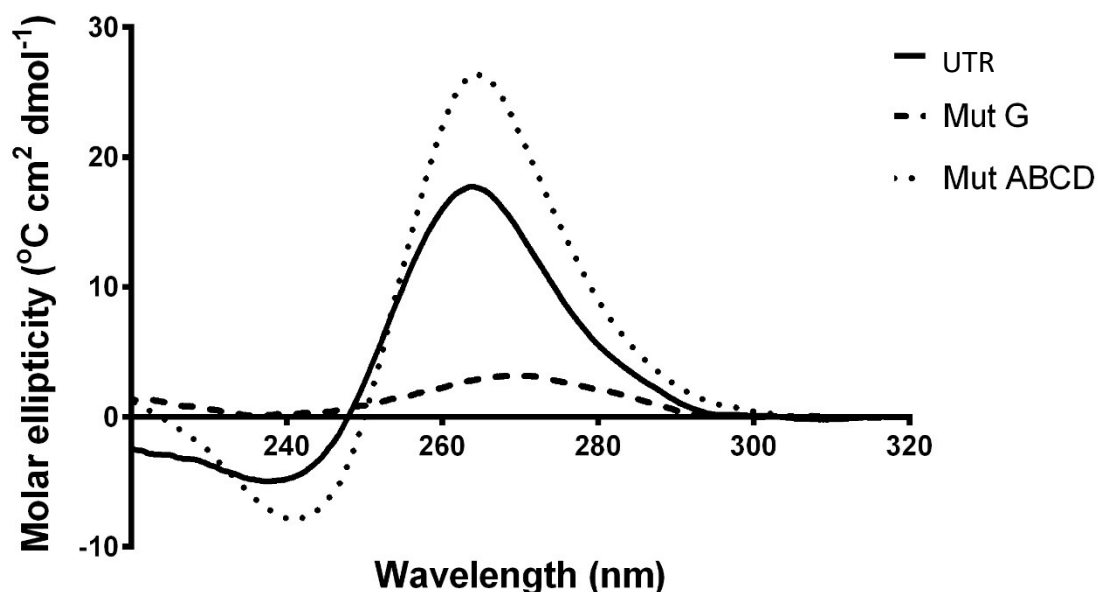


Figure 51: CD spectra of 10 μ M UTR, Mut G and Mut ABCD RNA folded at room temperature in the presence of 100 mM KCl. The spectrum of both the UTR and Mut ABCD mutants exhibit characteristic peaks at 263 nm and 240 nm, which is not clear in the mutated Mut G plasmid.

(iii) *In vitro* translation

The luminescence reader had to be calibrated to ensure that readings were consistent between separate experiments. There were three degrees of optimisation **(i)** choice of assay materials (i.e. multi-well plates) **(ii)** optimisation of the assay protocol (i.e. improving workflow and reagent quantities to make the assay repeatable) **(iii)** mechanical calibration of the luminescence reader (i.e. identifying optimal gain levels).

Assay materials: A key step in optimising the assay materials used was the choice of multi-well plate. For luminescence assays on a machine which reads from the top of the plate, the optimum plate is a solid bottomed white plate. For luminescence, white plates are preferred to reflect the light emitted and maximise the output signal. With fluorescence the signal is typically weaker compared to the background, so black-walled plates are preferred to reduce background and cross-talk. For this assay, Corning CoStar and Greiner Bio-one CELLSTAR plates were tested. There was no measurable difference in performance between the two plates as tested by comparing the signal produced by the luciferase produced from control RNA, so for practical reasons the Corning CoStar was chosen for use in all experiments.

Assay protocol: The initial set-up protocol for the assay is vital to ensure that results are comparable across different experiments and between repeats within the same experiment. Each separate experiment contained at least two internal repeats to ensure consistency and the experiment was repeated at least three times across separate days. Once added to the luciferase, the luciferase

assay reagent generates a signal that is nearly constant for around one minute. This creates a challenge when working with large numbers of repeats on one plate as it can take more than a minute for reagent to be added to all wells and a reading taken for all wells. In order to overcome this, an auto-dispenser was used, which allowed the reagent to be automatically added to each well and the reading taken at a precise interval after addition of the reagent, thereby ensuring consistency between different wells on the same plate.

Mechanical calibration: Once the assay protocol had been optimised the plate reader was tested to ensure that consistent readings were seen for control mixes both between different wells in the same plate and across different well plates. Using these results, the machine was calibrated. However, during the course of the experiments it was noted that there were some calibration issues and the machine had to be re-calibrated; these could have produced some additional variability between repeats on different days.

The relative luciferase expression levels (i.e. translation efficiency) of the new RNA constructs containing mutations in the sequence of the quadruplex itself (*NRAS* Mut G) or in its flanking regions (*NRAS* Mut ABCD) were compared with the expression level of *NRAS* 5'UTR (Figure 52). Disruption of the G-quadruplex resulted in a significant increase in translation ($p < 0.01$) of ~4x. This is a similar increase in translation to the 3.5x increase observed by Kumari *et al.*¹¹⁶

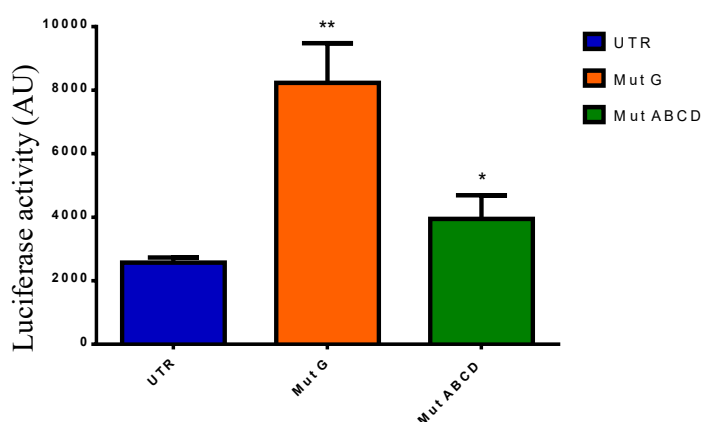


Figure 52: Luciferase activity of *NRAS* 5'UTR, *NRAS* Mut G and *NRAS* Mut ABCD RNA constructs. Results represent the absolute value of luminescence emitted during the assay with each transcript. The *NRAS* Mut G transcript emits a significantly high amount of luminescence suggesting that disrupting the quadruplex increases transcription. Error bars represent standard deviation for n=6 repeats, significance testing was via multiple t-tests, $p < 0.05$ (*), $p < 0.01$ ().**

Scrambling the flanking regions also resulted in a much smaller but significant increase in translation ($p < 0.05$). This could suggest that the flanking regions of the quadruplex play a small role in stabilising the quadruplex structure and that mutating these sequences upstream and downstream of the quadruplex could destabilise it, thus leading to a moderate increase in translation efficiency. However, we cannot rule out other hypotheses: for example, the flanking regions could be involved in binding of translation related proteins or factors which could be disrupted as a result of point mutations.

2.3.4. *In vitro* translation assay: Results

(i) Effect of PNA strand orientation on translation inhibition by PNA-NDI conjugates

The effect of PNA strand orientation on translation inhibition was investigated using 4LO (PNA-NDI conjugate with two 4-mer PNAs appended to the central NDI derivative) and its antiparallel analogue 4LOa. The orientation of the strands is defined with reference to their binding to complementary DNA. Antiparallel binding means PNA oligomers bind complementary oligonucleotides with their C-terminus facing the 5'-end of the complementary DNA (or RNA) and N-terminus facing the 3'-end. This is reversed for parallel binding as illustrated below (Figure 53).



Figure 53: schematic representation of parallel and antiparallel binding of a PNA oligomer (red) to a complementary RNA oligonucleotide (Blue).

Previous studies suggest that PNA binds to complementary antiparallel DNA with a higher affinity than to parallel sequences. For example, the T_m values of duplexes between parallel and antiparallel PNA and the penta-decamer H-TGTACGTCACAACTA-NH₂ were 56.1 °C and 69.5 °C respectively.²³⁹ Capillary-based electrophoresis also provided evidence that the kinetics of duplex formation is significantly quicker with anti-parallel PNA:DNA hybrids than with equivalent parallel hybrids.²⁸³ Similar effects have also been seen with PNA:RNA hybrids. Gokhale *et al.* reported higher melting temperatures for anti-parallel PNA:RNA duplexes versus parallel PNA:RNA duplexes with the same sequence.²⁸⁴ This suggested that the antiparallel PNA-NDI conjugates may have a greater effect on translation inhibition.

Unexpectedly, and although 4LO proved capable to inhibit *NRAS* 5'UTR translation with a low-micromolar IC₅₀ value, there was no significant difference between the effect of the parallel binding ligand 4LO and its antiparallel analogue 4LOa on translation inhibition (Figure 54 and Table 5). This indicates that the orientation of the PNA conjugates does not significantly affect the efficacy of translation inhibition. The difference between these results and those in the literature could be explained by the presence of the NDI conjugate and the G4 between the complementary flanking regions. The orientation of the PNA strands could affect the steric interactions between the PNA and the loop regions or edge of the grooves of the G4. If negative, this effect could counteract the positive impact seen from antiparallel orientation of the PNA strands. As a result, the arbitrary decision was made to synthesise PNA-conjugates that hybridise to their RNA target in a parallel orientation and all other PNA-NDI

conjugates tested were designed to bind in this specific orientation. However, further experiments would be needed to confirm if there were any differences in binding mode or affinity that were not reflected in the effect of the ligands on translation. In addition, melting studies using the individual PNA strands without the NDI would give an indication of whether the NDI was affecting the results.

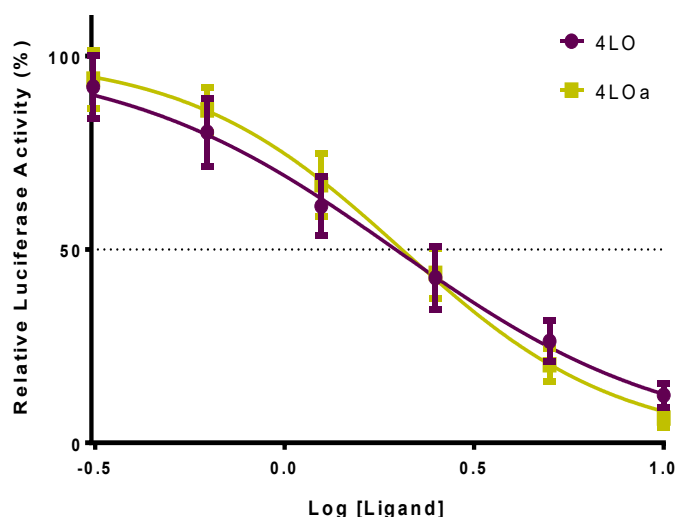


Figure 54: Effect of PNA-NDI conjugates on translation of *NRAS* 5'UTR RNA with PNAs designed to hybridise to their target RNA in a parallel (4LO) and antiparallel (4LOa) orientation. The amount of translation is tested using a luciferase assay and the results expressed as relative luciferase activity, where the level of activity with no ligand added is set at 100. In this study there is no clear difference in IC_{50} between the 4LO and 4LOa ligands. Lines of best fit generated using non-linear regression in Graphpad prism. Error bars represent standard deviation for $n=6$ (4LOa) and $n=17$ (4LO) repeats.

Table 5: IC_{50} values for the effect of 4LO and 4LOa ligands on translation of *NRAS* 5'UTR RNA

Ligand	IC_{50} (μM)	R^2	n
4LO	1.99 ± 0.48	0.9412	17
4LOa	2.03 ± 0.17	0.9700	6

(ii) The effect of PNA length on translational repression by PNA-NDI conjugates

To test the effect of PNA length on translation inhibition, we compared the activities of a small library of PNA-NDI conjugates where the same molecule of NDI was flanked by PNA sequences of increasing lengths (3LO, 4LO, 5LO and 6LO) on translation of *in vitro* transcribed RNA (Figure 55). The effect of a control ligand (so-called “PNA”), which featured two 4-mer PNA oligomers complementary to the G4 flanking regions and linked via a flexible alkyl (i.e. aminohexanoic acid) linker instead of the NDI derivative was also tested to identify whether the effects of the PNA-NDI conjugates were only

mediated by the PNA strands, or if there was a synergistic effect between the PNA and the NDI. A molecule of NDI flanked by only two lysine residues (so-called “NDI”) was also used as a control to further explore the effect of PNA conjugation on NDI activity. Interestingly, as shown on Figure 55, all PNA-NDI conjugates tested inhibited *NRAS* translation *in vitro*. In the following sections, we will present a Structure-Activity Relationship (SAR) analysis of this new family of PNA-NDI conjugates.

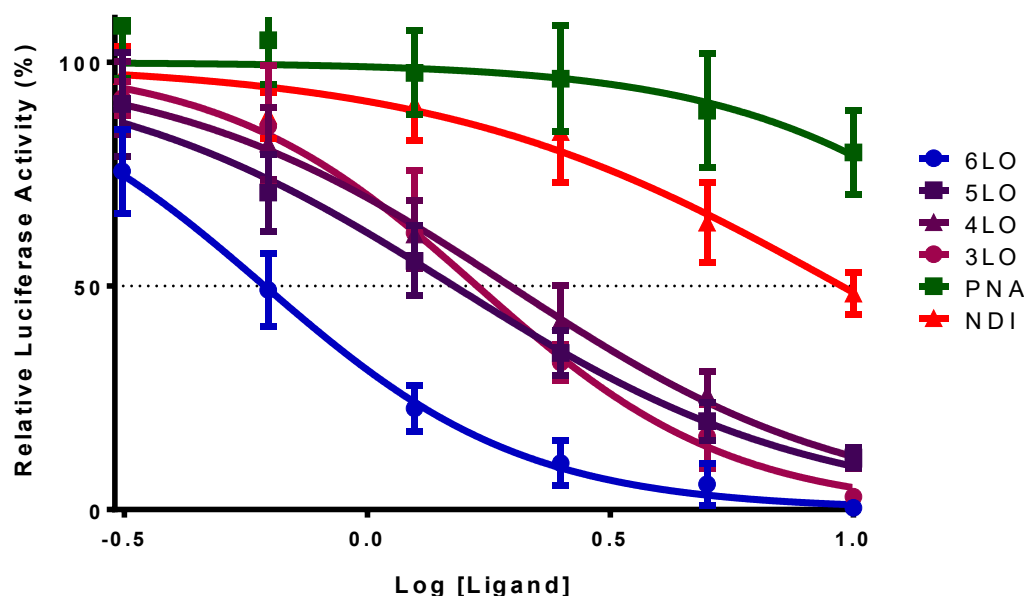


Figure 55: Effect of PNA length on activity of PNA-NDI conjugates. The effect of 6-mer (3LO), 8-mer (4LO), 10-mer (5LO) and 12-mer (6LO) PNA-NDI conjugates and “NDI” and “PNA” control ligands on translation of *NRAS* 5’UTR mRNA. The 4 PNA-NDI conjugate ligands (3LO, 4LO, 5LO, 6LO) inhibit translation at lower concentrations than either individual elements (“PNA”, NDI). In addition, there is a clear increase in inhibition with the change in PNA length from 5 nucleotides (5LO) to 6 nucleotides (6LO), however there is limited change when increasing from 3 to 5 nucleotides (3LO to 5LO). Lines of best fit generated using non-linear regression in Graphpad prism. Error bars represent standard deviation for n=6 (3LO, 5LO, “PNA”, NDI), n=10 (6LO) and n=17 (4LO) repeats.

(iii) Synergistic effect of PNA-NDI conjugation

Comparison between the PNA-NDI conjugates and both control ligands (“NDI” lacking the PNAs and “PNA” lacking the central NDI) was used to assess whether there was a synergistic effect between the PNA and the NDI. The results shown in Figure 55 indicate that conjugation of PNA of any length significantly increased the inhibitory effect of NDI ligands on translation of luciferase under the control of the *NRAS* 5’UTR. All the PNA-NDI conjugates demonstrated significantly lower IC_{50} values (see Table 6) ($p < 0.01$) than the “NDI” ligand as measured by multiple t-tests. In a similar way, Control “PNA” consisting of two 4-mer PNA strands but lacking the G4 binding platform had no significant effect on translation, even at high ligand concentration. The fact that PNA-NDI conjugates greatly outperform each component taken individually could suggest that there is a synergistic effect between the quadruplex binding ligand (NDI) and the PNA side chains, and that this synergy is responsible for increased translation inhibition. However, the lack of specificity in the results mean that the effect may not be due to G-quadruplex binding at all, and that rather than a synergistic effect on G4 binding, the

change in translation is due to a non-specific effect mediated by the PNA-conjugates. Further investigation is required to confirm this.

(iv) Effect of PNA-length

PNA-NDI conjugates with varying length PNA strands, including a 6-mer (3LO), 8-mer (4LO), 10-mer (5LO) and 12-mer (6LO) ligand, were compared to assess the effect of PNA length on translational inhibition by PNA-NDI conjugates. There was no significant difference in the effects of 3LO, 4LO and 5LO on translation. 6LO, however, proved a significantly ($p < 0.01$, measured using student's t-tests) more effective inhibitor (Figure 55), with an IC_{50} of $0.62 \mu M$, which compares to values of $1.5\text{--}2 \mu M$ for other PNA-NDI (Table 6).

Table 6: IC_{50} values for PNA-NDI conjugates with increasing PNA length.

Ligand	$IC_{50} (\mu M) \pm SD$	R^2	n
6LO	0.62 ± 0.12	0.9521	10
5LO	1.52 ± 0.40	0.9376	6
4LO	1.99 ± 0.48	0.9473	17
3LO	1.68 ± 0.43	0.9438	6
“PNA”	>10	-	6
NDI	9.28 ± 2.0	0.8185	6

It was expected that PNA conjugation would result in increased affinity relative to the length of the PNA, which would then translate to increases in translation inhibition efficacy. From these results it is unclear whether the observed increase in efficacy between 5LO and 6LO is due to a non-linear effect of PNA length on affinity, PNA-mediated antisense activity or a PNA-mediated increase in G4 stability.

(v) Effect of PNA targeting towards either 5'-flanking region or both 5' and 3'-flanking regions

To test the interaction between the length and position of the PNA strands and their effect on translation, 3LO, 6LO_{HALF} and 6LO were compared using the translation assay (Figure 56). 6LO_{HALF} contains the same number of PNA residues as 3LO (6 residues) but all are positioned on one side of the NDI so as to target only the 5'-flanking region of the RNA G4. 6LO_{HALF} exhibited a significantly reduced effect on translation inhibition when compared to 3LO (see Table 7). This suggests that PNA-NDI conjugates

targeting both the 5'- and 3'-flanking regions of the RNA G4 are more effective inhibitors of translation than ligands targeting just the 5'-flanking region.

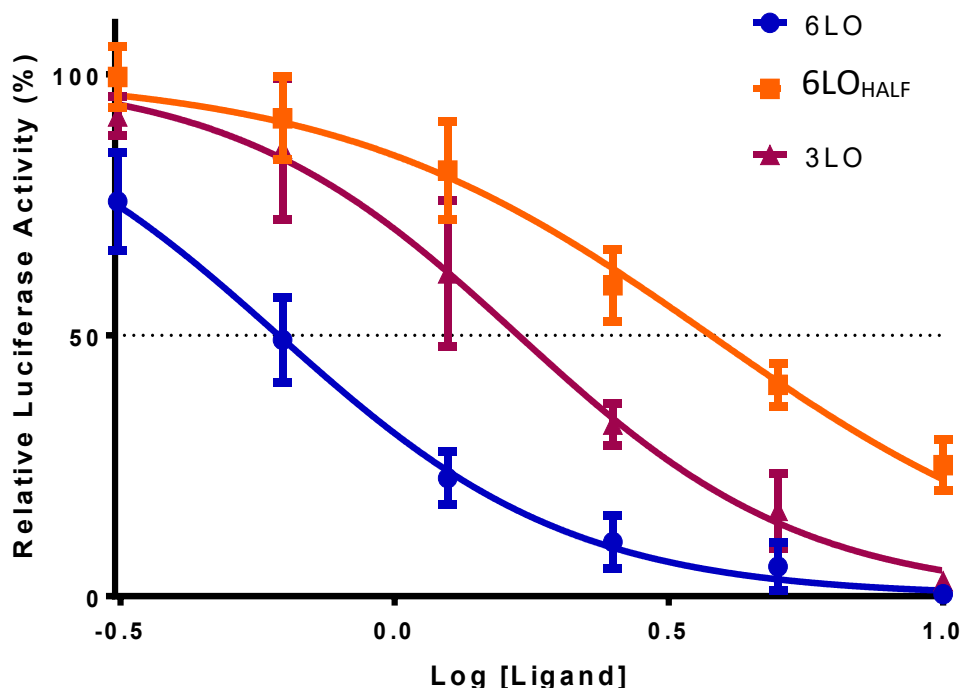


Figure 56: Effect of targeting one (6LO_{HALF}) or both (3LO, 6LO) flanking regions on translation inhibition of *NRAS* 5'UTR by PNA-NDI conjugates. This experiment suggests that the placement of the PNA flanking regions is important for inhibition of translation. The 6LO_{HALF} ligands contains 6 PNA nucleotides, the same number of nucleotides as 3LO, but aligned on one side of the NDI to bind to a single flanking region only. 3LO exhibits an inhibitory effect at lower concentrations suggesting that targeting both flanking regions is more effective than targeting a single region. [Lines of best fit generated using non-linear regression in Graphpad prism]. Error bars represent standard deviation for n=6 (3LO) and n=10 (6LO and 6LO_{HALF}) repeats.

Table 7: IC₅₀ values for PNA-NDI conjugates targeted towards either the 5'- or both the 5'- and 3'-flanking regions

Ligand	IC ₅₀ (μM) ± SD	R ²	n
6LO	0.62 ± 0.12	0.9521	10
6LO _{HALF}	3.74 ± 0.78	0.9387	10
3LO	1.68 ± 0.43	0.9438	6

(vi) Effect of PNA-conjugation on the specificity of NDI

The specificity of the PNA-NDI conjugates was assessed by comparing their IC₅₀ values when targeting the native *NRAS* 5'UTR mRNA and various mutated controls: *NRAS* Mut G (lacking the quadruplex), *NRAS* Mut ABCD (with scrambled G4 flanking sequences to prevent sequence-specific PNA-RNA hybridisation) and Control (positive control from the assay kit with no resemblance to the *NRAS* construct) mRNA.

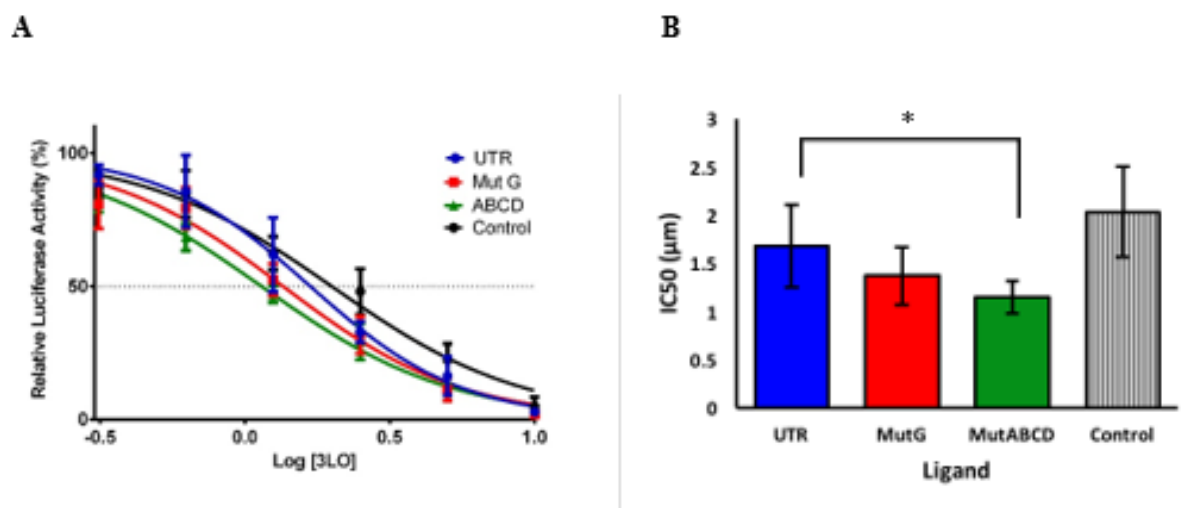


Figure 57: Inhibition in translation of *NRAS* 5'UTR (UTR), *NRAS* Mut G (Mut G), *NRAS* Mut ABCD (ABCD) and Control mRNA by 3LO PNA-conjugate ligand. (A) Relative luciferase activity following translation of luciferase mRNA in the presence of 3LO ligand. Translation of luciferase, as measured by luciferase activity, was shown to be inhibited in the presence of increasing concentrations of 3LO ligand. (B) IC₅₀ values for 3LO-mediated inhibition of translation of *NRAS* 5'UTR, Mut G, Mut ABCD and control mRNA, shown in blue, red, green and black respectively. There were not significant differences in the IC₅₀ of the ligand for *NRAS* 5'UTR when compared to control or Mut G mRNA. However, 3LO exhibited significantly higher inhibitory activity, and therefore a lower IC₅₀, for Mut ABCD mRNA. [Lines of best fit generated using non-linear regression in Graphpad prism]. Error bars represent standard deviation for n=6 repeats. Significance testing was via multiple t-tests, $p < 0.05$ (*).

The results from the translation assays with 3LO show limited evidence of an increase in specificity (Figure 57). There was no significant difference between the effect of the ligand on *NRAS* 5'-UTR mRNA and the control mRNA. This shows that there is a non-specific effect occurring, which is at least partially responsible for the inhibitory effect on translation.

In addition, although there was a significant difference between the effect of the ligand on *NRAS* 5'-UTR and *NRAS* Mut ABCD mRNA, this was contradictory to our expectations. The *NRAS* Mut ABCD mRNA contains mutations in the flanking regions of the *NRAS* PQS, meaning that they are no longer complementary to the PNA sequences on the PNA-conjugate. As such we were expecting the IC₅₀ to increase as PNA binding would be prevented. However, for 3LO the IC₅₀ was lower for *NRAS* Mut ABCD when compared to the *NRAS* 5'-UTR mRNA. It is possible that because the PNA can no longer bind to the intended target it shifts the binding equilibrium towards another binding interaction which is responsible for the non-specific inhibition. By contrast, with *NRAS* Mut G and *NRAS* 5'-UTR there is stronger binding between the PNA-conjugate and flanking regions of the *NRAS* PQS, which reduced the amount of ligand available to bind via the alternative route, thereby increasing the IC₅₀. However, it is unclear why this effect would only impact the 3LO ligand.

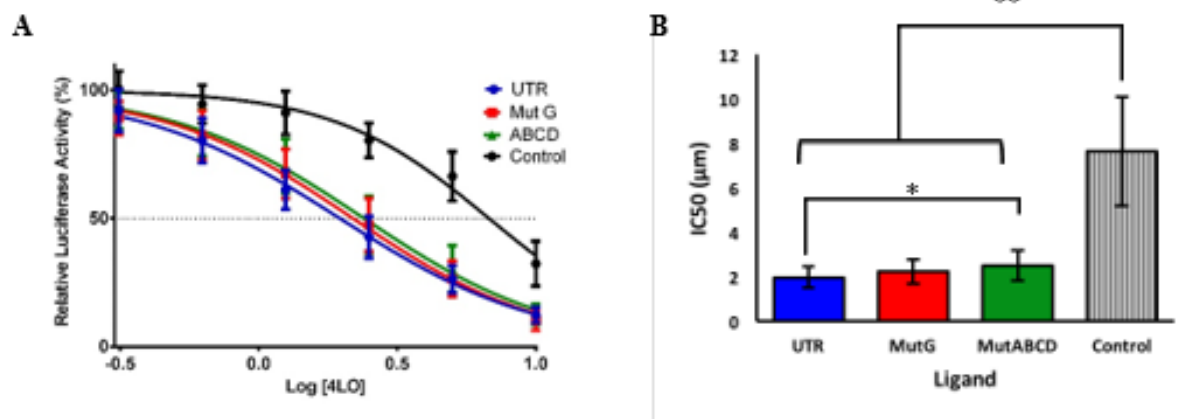


Figure 58: Inhibition in translation of *NRAS* 5'UTR (UTR), *NRAS* Mut G (Mut G), *NRAS* Mut ABCD (ABCD) and Control mRNA in the presence of 4LO PNA-conjugate ligand. (A) Relative luciferase activity following translation of luciferase mRNA in the presence of 4LO ligand. Translation of luciferase was shown to be inhibited in the presence of increasing concentrations of 4LO ligand. **(B)** IC₅₀ values for 4LO-mediated inhibition of translation of UTR, Mut G, Mut ABCD and control mRNA, shown in blue, red, green and black respectively. There was a significant increase in the inhibitory effect of 4LO on translation of UTR, Mut G and Mut ABCD mRNA when compared to control mRNA, as seen by the lower IC₅₀ values. In addition, 4LO had the highest inhibitory effect on UTR as opposed to the mutated forms, although there was only a significant difference between UTR and Mut ABCD. [Lines of best fit generated using non-linear regression in Graphpad prism]. Error bars represent standard deviation for n=17 repeats. Significance testing was via multiple t-tests, $p < 0.05$ (*) and $p < 0.01$ (**).

There is a significant difference between the effect of 4LO on translation of mRNA containing the *NRAS* 5'UTR sequence (UTR, Mut G and Mut ABCD) when compared to the control mRNA (Figure 58). However it is unclear to what degree this is related to the presence of the *NRAS* PQS sequence. There was no significant difference between the IC₅₀ of 4LO with *NRAS* 5'-UTR and *NRAS* Mut G mRNA. This suggests that the increased inhibition is not primarily driven by stabilisation of *NRAS* PQS to form a G-quadruplex, as the *NRAS* Mut G mRNA does not form a G4 (as shown in initial CD characterisation of the mRNA).

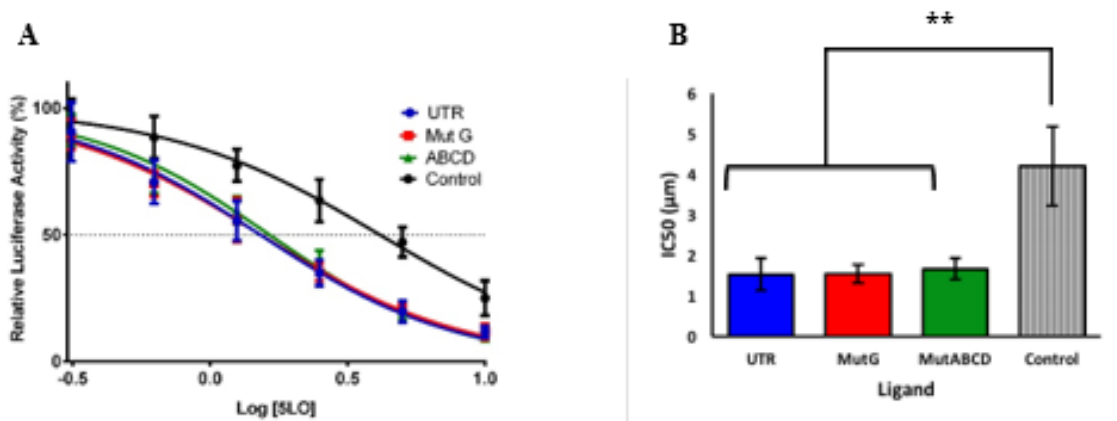


Figure 59: Inhibition in translation of *NRAS* 5'UTR (UTR), *NRAS* Mut G (Mut G), *NRAS* Mut ABCD (ABCD) and Control mRNA in the presence of 5LO PNA-conjugate ligand. (A) Relative luciferase activity following translation of luciferase mRNA in the presence of 5LO ligand. Translation of luciferase was shown to be inhibited in the presence of increasing concentrations of 5LO ligand. **(B)** IC₅₀ values for inhibition of translation of luciferase mRNA downstream of UTR, Mut G, Mut ABCD and control mRNA, shown in blue, red, green and black respectively, in the presence of 5LO. There was a significant increase in the inhibitory effect of 5LO on translation of UTR, Mut G and Mut ABCD mRNA compared to the control mRNA, as seen by the lower IC₅₀ values. There was no significant difference between the inhibitory effect of 5LO on the native *NRAS* 5'UTR PQS when compared to the mutated forms. [Lines of best fit generated using non-linear regression in Graphpad prism]. Error bars represent standard deviation for n=6 repeats. Significance testing was via multiple t-tests, $p < 0.05$ (*) and $p < 0.01$ (**).

There was a significant difference between the effect of 5LO (Figure 59) and 6LO (Figure 60) on control mRNA and the *NRAS* 5'UTR, *NRAS* Mut G and *NRAS* Mut ABCD mRNA, with significantly lower IC_{50} for the latter. However, there was no significant difference between the *NRAS* 5'UTR, *NRAS* Mut G and *NRAS* Mut ABCD mRNA for either ligand, suggesting that this is a non-specific effect.

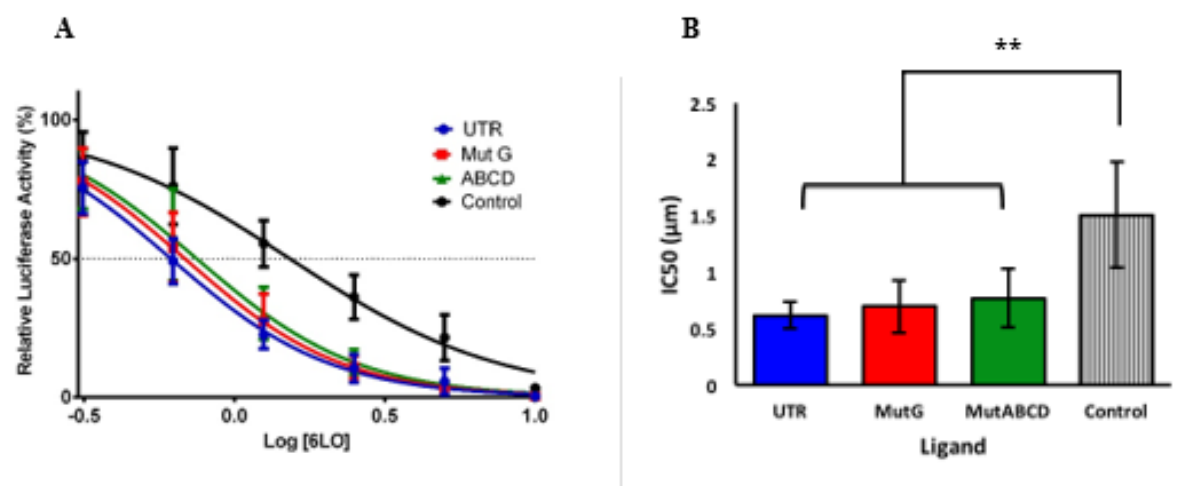


Figure 60: Inhibition in translation of *NRAS* 5'UTR (UTR), *NRAS* Mut G (Mut G), *NRAS* Mut ABCD (ABCD) and Control mRNA in the presence of 6LO PNA-conjugate ligand. (A) Relative luciferase activity following translation of luciferase mRNA in the presence of 6LO ligand. Translation of luciferase was shown to be inhibited in the presence of increasing concentrations of 6LO ligand. **(B)** IC_{50} values for 6LO-mediated inhibition of translation of luciferase mRNA downstream of UTR, Mut G, Mut ABCD and control, shown in blue, red, green and black respectively. There was a significant increase in the inhibitory effect of 6LO on translation UTR, Mut G and Mut ABCD mRNA compared to the control mRNA, as seen by the lower IC_{50} values. There was no significant difference between the inhibitory effect of 6LO on the native *NRAS* 5'UTR PQS when compared to the mutated forms. [Lines of best fit generated using non-linear regression in Graphpad prism]. Error bars represent standard deviation for n=6 repeats. Significance testing was via multiple t-tests, $p < 0.05$ (*) and $p < 0.01$ (**).

The relative IC_{50} of the ligands are summarised in Figure 61. There were significant differences between the effect of 6LO, 5LO, 4LO and the “NDI” ligand on the control mRNA. However, for most ligands there were no significant differences between their effects on *NRAS* 5'UTR, *NRAS* Mut G and *NRAS* Mut ABCD mRNA, except for 4LO where there was a significant decrease in IC_{50} when targeting *NRAS* 5'UTR when compared to *NRAS* Mut ABCD (see Table). These results do not demonstrate a clear or significant trend with PNA-length and specificity, but visual analysis suggests that while 4LO has the weakest activity it may have the highest specificity. This will be worth further investigation. In addition, comparison between the different ligands gives an indication of the effect of PNA length on the activity of the ligands. We expected increased PNA length to result in increased translation inhibition. However, while this is the case for 4LO, 5LO and 6LO, 3LO showed a lower IC_{50} inhibiting *NRAS* 5'UTR than 4LO. Further discussion of this effect is included in the discussion section at the end of this chapter.

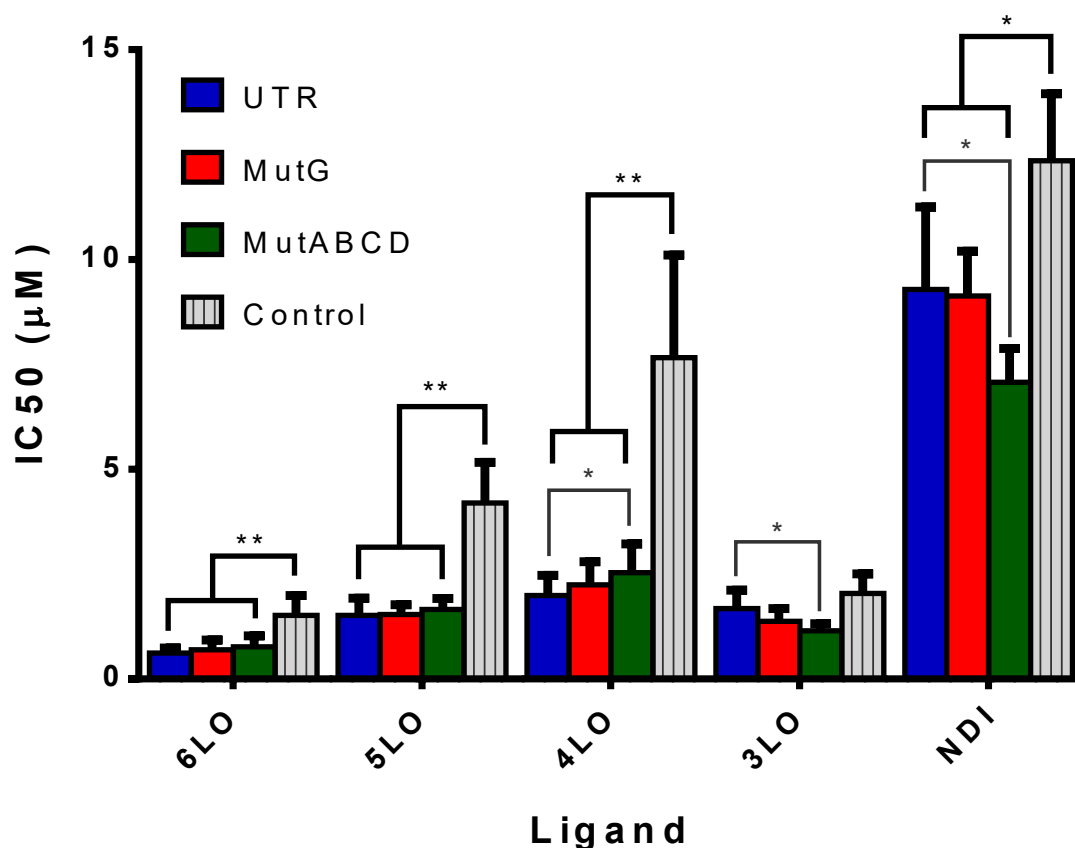


Figure 61: IC₅₀ values for ligand-mediated inhibition of translation. For each ligand, results are shown for the effect on translation of mRNA containing: the native *NRAS* 5'UTR including complete *NRAS* G4 (UTR), mutated *NRAS* 5'UTR with GG to AA mutation to prevent quadruplex folding (Mut G), mutated *NRAS* with scrambled flanking regions (Mut ABCD) and luciferase control mRNA without the *NRAS* 5'UTR. For all ligands except 3LO there was a significant difference between the control mRNA and the mRNA containing *NRAS* 5'UTR. However, the control mRNA was also affected by the ligands suggesting that there are non-specific effects, and translation of each of the mutated mRNA was also affected by each ligand, suggesting that PNA-conjugation does not add extra specificity. Error bars represent standard deviation for n=6 (3LO, 5LO, NDI), n=10 (6LO) and n=17 (4LO) repeats. Significance testing was via multiple t-tests, $p < 0.05$ (*) and $p < 0.01$ (**).

Table 7: IC₅₀ values for translation inhibition by PNA-NDI ligands.

Ligand	IC ₅₀ ± SD (μM)			
	UTR	Mut G	Mut ABCD	Control
6LO	0.62 ± 0.12	0.70 ± 0.23	0.77 ± 0.26	1.51 ± 0.47
5LO	1.52 ± 0.40	1.54 ± 0.23	1.66 ± 0.26	4.19 ± 0.97
4LO	1.99 ± 0.48	2.24 ± 0.55	2.53 ± 0.68	7.66 ± 2.45
3LO	1.68 ± 0.43	1.37 ± 0.30	1.15 ± 0.17	2.03 ± 0.47
“PNA”	18.8 ± 6.20	30.6 ± 12.29	34.4 ± 25.5	>100
NDI	9.28 ± 1.96	9.13 ± 1.07	7.07 ± 0.81	12.4 ± 1.6
6LO _{HALF}	3.74 ± 0.77	3.91 ± 0.79	3.61 ± 0.86	7.75 ± 2.78
4LOa	2.03 ± 0.17	2.25 ± 0.17	2.17 ± 0.11	2.06 ± 0.45

2.3.5. Characterisation of binding

(i) Circular Dichroism

Circular Dichroism (CD) spectroscopy was used to monitor quadruplex formation and to assess the effect of ligand binding on RNA folding. CD is a spectroscopic technique commonly used to identify secondary structures in nucleic acids and proteins. It works by detecting differences in their absorption of circularly polarised light. Chiral groups in secondary structures absorb left and right circularly polarised light to a different extent at certain wavelengths, depending on their environment.

CD has been used to assess the effect of sequence, cation concentration, chemical modification and ligand binding on the stability and topology of G4 structures.²⁸⁵ G4 have a characteristic CD signature which is determined by the orientation of glycosidic bonds within the structure. Depending on the topology of the quadruplex, glycosidic bonds may adopt either the *syn* or *anti* conformation in order to allow G-G stacking interactions. All-parallel G-quadruplexes exhibit a specific signature composed of a positive peak at 263 nm and a negative peak at 241 nm, which arises from all glycosidic bonds adopting the *anti* conformation.^{72, 286, 287} Conversely, within anti-parallel quadruplexes there is a mixture of bonds in the *syn* and *anti* conformations which produces a positive peak at 295 nm and negative peak at 240 nm. Interestingly, PNA-RNA duplexes have been shown to exhibit a similar spectrum with Gong and Desaulniers identifying negative peaks at 210 nm and 250 nm and a positive peak at 265 nm for such heteroduplexes.²⁸⁸ To understand the lack of specificity of the PNA-NDI conjugates in our translation assays, the ability of the conjugates to induce quadruplex formation in UTR, Mut G and Mut ABCD RNA was investigated using CD spectroscopy.

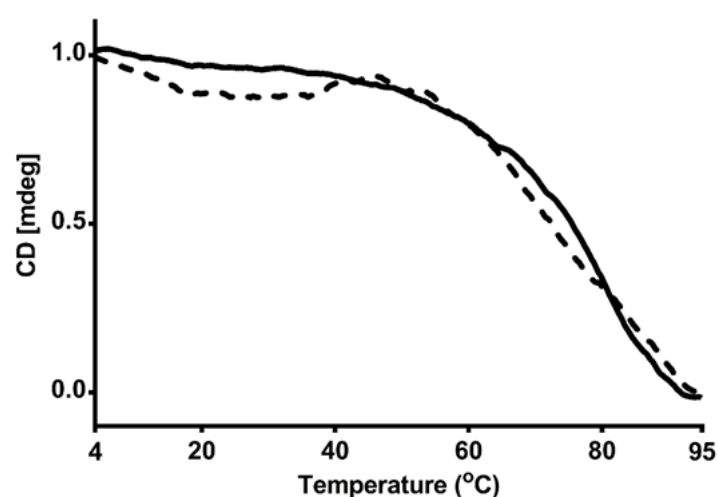


Figure 62: Melting curve of *NRAS* G4, with no added KCl and 0 (solid line) and 100 μ M (dashed line) 6LO PNA-conjugate ligand measured at 265 nm. The results show no significant difference between the melting temperature in the presence or absence of 6LO.

Measuring the effect of ligands on the CD spectra of quadruplexes in low levels of KCl can be used to indicate the ability of small molecules to induce quadruplex formation under otherwise highly unfavourable conditions. Initially, we ran a melting curve to identify whether the 6LO ligand stabilised the *NRAS* 5'UTR G-quadruplex. As the *NRAS* quadruplex is extremely stable when pre-folded in the presence of even low (e.g. 1 mM) KCl concentrations, the addition of small molecule ligands did not result in any significant change to the melting temperature, which would indicate structural changes (Figure 62). In addition, as the PNA-conjugates bind using multiple modes, i.e. via both a PNA:RNA duplex and NDI:G4 interaction, both of which have a CD signal around 265 nm, it is difficult to isolate the specific signal corresponding to G4 unfolding. Due to the overlap of multiple effects we concluded that CD melting experiments were not suitable for this type of binding analysis. Instead, the effect of ligand binding on the overall CD spectrum of the mRNA was tested at room temperature. To maximise the visible effects of ligand binding on the CD spectra, we tested the effect of 6LO (the most potent inhibitor of the series) and of the two controls “NDI” (lacking PNAs) and “PNA” (lacking NDI) on RNA folding in water, in the absence of any added KCl (Figure 63).

First, it is noteworthy that even in the absence of any added salt, both UTR and Mut ABCD RNAs showed a CD signature characteristic of a parallel-stranded G4, although of very weak intensity. This can be explained by the extremely high stability of this RNA G4 and by the likely presence of small concentrations of salt in the water used for these spectroscopic studies.

Increasing concentrations of “NDI” induced an increase in the positive peak at around 266 nm in UTR (**A**) and Mut ABCD (**B**) spectra but induced no significant changes in the Mut G spectrum (**C**). These results suggest that addition of “NDI” can induce the formation of a G4 structure in both UTR and Mut ABCD RNA whilst Mut G, lacking a stretch of 3 consecutive guanines, cannot form such a structure, even in the presence of high concentrations of NDI.

The addition of the “naked” “PNA” to UTR (**D**) and Mut ABCD (**E**) RNA oligonucleotides induced increases in the positive peaks at 222 nm and 265 nm, and an increase in the negative peak at 210 nm. At high ligand concentrations (above 50 μ M) the peaks centred at 240 nm and 265 nm were also red-shifted. Looking in particular at the UTR peaks, the shifts seem to match those observed in PNA:RNA duplexes.²⁸⁸ This could suggest that the effect observed on the CD spectrum is at least partially caused by the formation of a PNA:RNA duplex rather than through G4 formation induced by PNA binding. This is however very unlikely to be the case for the “PNA” binding to Mut ABCD as the sequence of the PNA is not complementary to that of the RNA. The changes in the CD spectrum of Mut ABCD may be caused by conformational changes induced by non-specific electrostatic interactions between the negatively charged RNA and positively charged PNA.

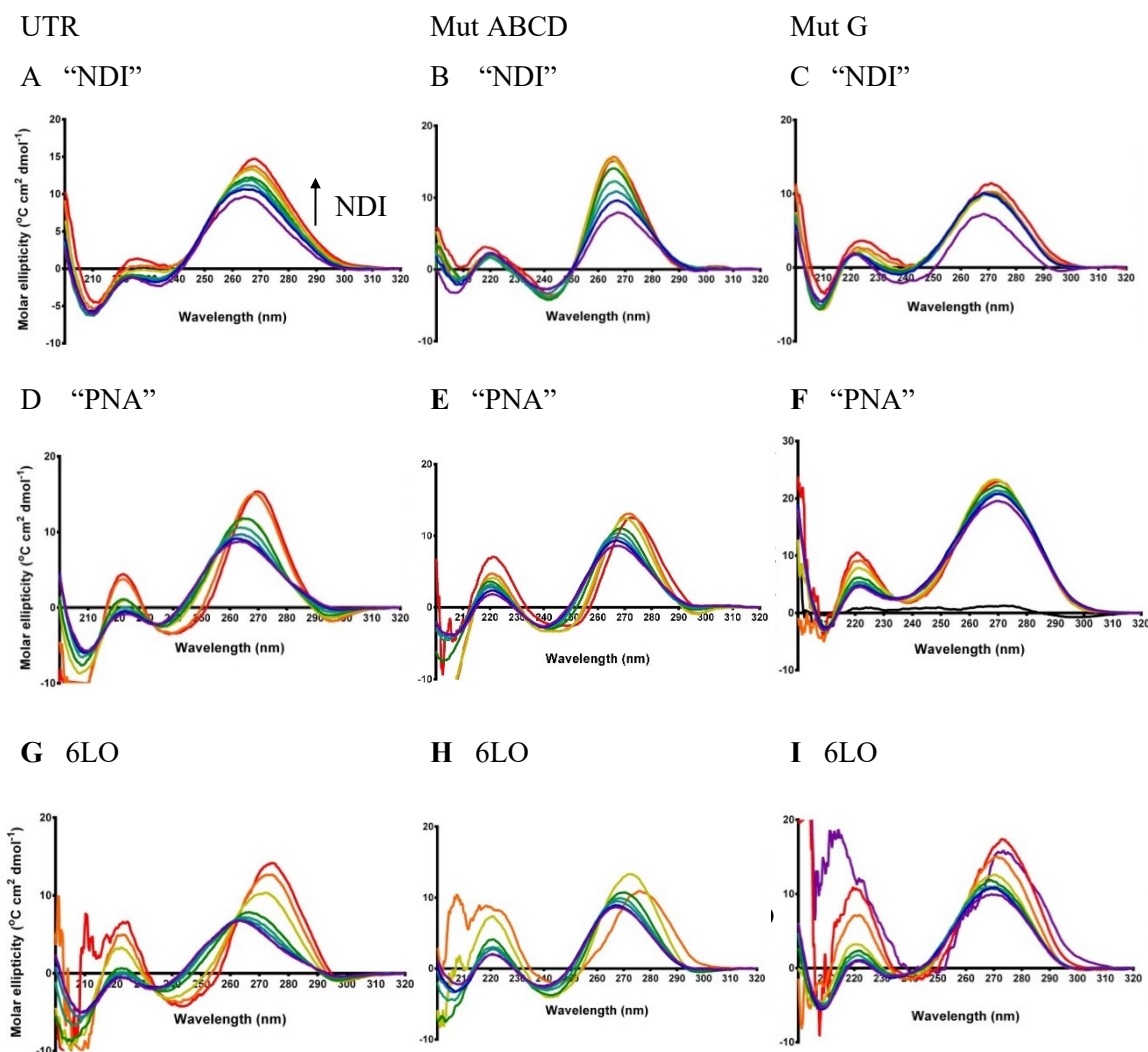


Figure 63: CD spectra of UTR Mut G and Mut ABCD RNA in the presence of increasing concentrations of ligand at room temperature. “NDI” (A, B and C), “PNA” (D, E and F) and 6LO (G, H and I) in the absence of added KCl. Colours represent the concentration of ligand added, with purple representing 0 μM of ligand and increasing to 100 μM in the order, blue, light blue, teal, green, yellow, orange and red.

In contrast, addition of “PNA” to Mut G (F) only induced an increase in the 222 nm peak. This suggests that “PNA” can’t bind UTR and Mut G the same way despite both RNA having flanking regions complementary to that of “PNA”.

The spectra for UTR (G) and Mut ABCD (H) in the presence of 6LO much more strongly resemble the spectra in the presence of “PNA” rather than the “NDI” spectra. This suggests that, at least in the absence of any added salt, PNA:RNA duplex formation has a much greater effect on the CD spectra than G4 formation induced by “NDI” binding. One significant difference is an increase in the levels of noise below 220 nm, which is consistent across all three spectra in the presence of 6LO but present in none of the “NDI” or “PNA” spectra. The spectra of Mut G in the presence of 6LO are interesting, as neither the “PNA” nor “NDI” control ligands affected the spectra. These results suggest that there may

be a synergistic effect on PNA-conjugate binding that is independent of G4 formation. The presence of a single crossing point in the spectra in **D**, **E** and **G**, suggests co-existence of two RNA conformations which most likely correspond to the free RNA and RNA complexed with 6LO.

(ii) Electrophoresis Mobility Shift Assay (EMSA)

Electrophoresis can be used to investigate the binding of ligands to nucleic acids through their effect on mobility in hydrogels. Ligand binding changes the size and charge of the complex, which alters the effect of the electric field and the gel structure on mobility. Typically, ligand bound nucleic acids travel through the gel at slower speeds, creating a separate band further up the gel. Quantifying the change in intensity of the bands with increased ligand concentration allows the binding constants to be calculated. First the background fluorescence of the gel is subtracted from the readings, then the relative intensities of the bands are used to calculate the bound RNA as a fraction of total RNA. This is then plotted against ligand concentration and fit with a non-linear regression curve using the equation below, which generates values for K_D and B_{max} .

$$Fraction\ Bound = \frac{Bound}{(Bound + Unbound)}$$

$$Fraction\ Bound = B_{max} \frac{[Ligand]}{K_D + [Ligand]}$$

There are several different methods that can be applied to visualise and quantify bands, using either labelled or unlabelled RNA. Firstly, unlabelled RNA can be visualised by soaking the gel with an intercalating dye such as ethidium bromide or SYBR safe. Alternatively, the RNA can be labelled either with a fluorescent dye or a tag containing a radioactive element, such as ^{32}P .²⁸⁹ A significant advantage of using labelled RNA is that it produces gels with much higher resolution and lower background noise, which is a particular advantage when running small RNAs at low concentrations. Fluorescent labelling has been used in previous studies to quantify the binding of ligands to RNA.^{290, 291} It avoids the safety concerns associated with the use of radioactive labelling and was therefore chosen for use in this study.

EMSA is also a relatively common technique used to study binding to G4.²⁹² It is of particular use in investigating interactions between proteins and G4,²⁹³ as their high molecular weight produces clear separation between the bound and unbound RNA even when run at low voltages. However, it has also been used in studies with small molecule ligands to investigate binding affinity.²⁹⁴

In early studies using FAM-labelled target RNA there was no clear appearing band produced by the RNA:ligand complex. This implied that the ligand:RNA complexes were either degrading, dispersed or that the dye was being quenched by the ligand.

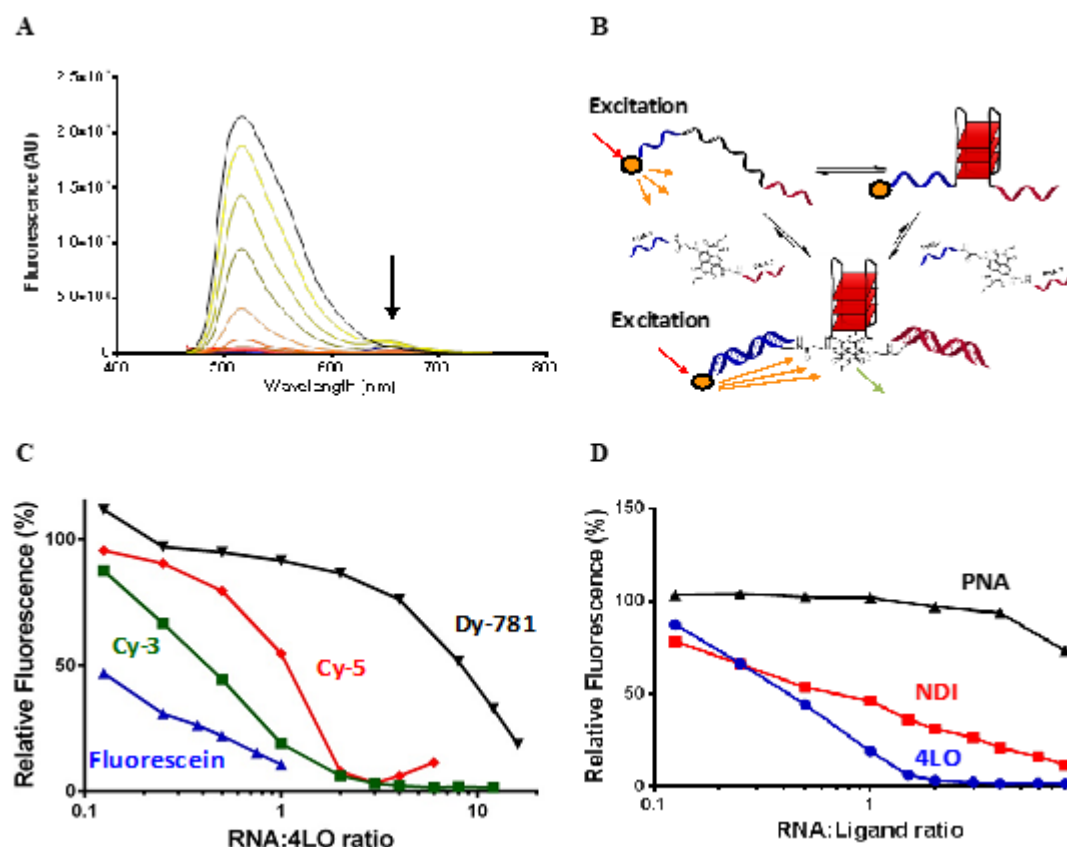


Figure 64: Dye selection and optimisation for EMSA assay. **A)** Example fluorescence emission spectra of fluorescein labelled-RNA with increasing concentrations of 4LO. The fluorescent peak of Fluorescein-labelled UTR RNA at (516 nm) decreases with increased 4LO concentration. **B)** Proposed mechanism for fluorescence quenching. When ligands bind, NDI absorbs the fluorescence from the dye labelled RNA. **C)** Comparison of the reduction in fluorescent peak intensity for RNA labelled with Fluorescein (green), Cy3 (blue), Cy5 (red) and Dy-781 (black) with increased RNA:4LO ratio. The peaks for the ligands were 516, 562, 665 and 800 nm respectively. **D)** Comparison of the effect of “PNA” (black), “NDI” (red) and 4LO (blue) on the fluorescent signal of fluorescein-labelled UTR RNA.

We therefore decided to monitor the interaction between PNA-NDI conjugates and various RNA oligonucleotides labelled with fluorescent dyes emitting at different wavelengths, including; Fluorescein (exc. 494 nm, em. 512 nm), Cy3 (exc. 550 nm, em. 570 nm), Cy5 (exc. 650 nm, em. 670 nm), Dy781 (exc. 783 nm, em. 800 nm) and unlabelled RNA with SYBR Green (exc. 497 nm, em. 520 nm). The fluorescence emission spectra of the labelled RNA in the presence of increasing amounts of PNA-NDI conjugates were recorded using a fluorescence spectrometer. A typical example of fluorescence titration experiment using the effect of 4LO on fluorescein-labelled RNA is shown in Figure 64A.

For all labelled RNA tested, addition of increasing concentrations of PNA-NDI conjugates resulted in a reduction in emitted fluorescence intensity. This suggests that the PNA-NDI conjugates are capable of quenching the fluorescent label, causing a reduction in intensity of the appearing band corresponding to the complex formed between the ligand and the RNA at higher ligand concentrations. The fluorescence of the dyes with lower excitation and emission frequencies (Cy5 and Dy-781) dropped to

a lesser degree (Figure 64C). The fluorescence of Dy-781 was least affected by the quenching and as such was chosen for use in further EMSA experiments. Comparison of the effect of “PNA”, “NDI” and 4LO on fluorescence showed that the effect is limited to ligands containing the NDI group (Figure 64D). This implies that the NDI is the molecule responsible for the fluorescence quenching, most likely through fluorescence resonance energy transfer (FRET) or direct interaction between the dye and the NDI, or both.

(iii) FRET

Binding of the NDI in close proximity of the dye, not necessarily with any direct (e.g. π - π stacking) interaction, could lead to quenching through FRET. NDI derivatives are characterised by a very large absorption band between 470 nm and 650 nm and maximally emit in the red/near infrared (emission band centred at around 650 nm).²⁹⁵ In a typical titration experiment of 4LO into a solution of FAM-labelled RNA, increasing PNA-NDI concentration resulted in a decrease of the characteristic emission band of fluorescein (centred at 520 nm) and the appearance of a weaker emissive band centred at 654 nm (shown by the black arrow in Figure 64A). This can be explained by a strong FRET between fluorescein (acting as a donor) and the NDI (acting as an acceptor). The emission spectra of SYBR green, fluorescein and Cy3 do indeed strongly overlap with the broad absorption spectrum of the NDI, which enables the FRET effect to occur. A potential mechanism for this quenching effect is shown in Figure 64B. Since the emission spectra of Cy5 and Dy-781 do not overlap with the absorption spectrum of the NDI, they are not affected by FRET and are therefore much better suited for our binding assays. However, it is noteworthy that there is still a moderate quenching effect with these dyes (which emit at much higher wavelengths) at high ligand concentration which may be caused by a separate quenching mechanism (possibly direct interaction between the dyes at high ligand concentration).

A direct interaction between the NDI and the dye could induce a conformational change that reduces the fluorescent properties of the dyes. All the dyes tested are indeed flat aromatic molecules that could therefore interact with the NDI through π -stacking. This is likely to be a relatively weak interaction, which is why it only occurs at much higher ligand concentrations.

Experiments were initially conducted using agarose gels and these results were compared with identical experiments conducted on polyacrylamide gels. As there were no significant differences between results observed with both types of gel, agarose gels were used for the bulk of experiments. Polyacrylamide gels were used when greater resolution bands were needed.

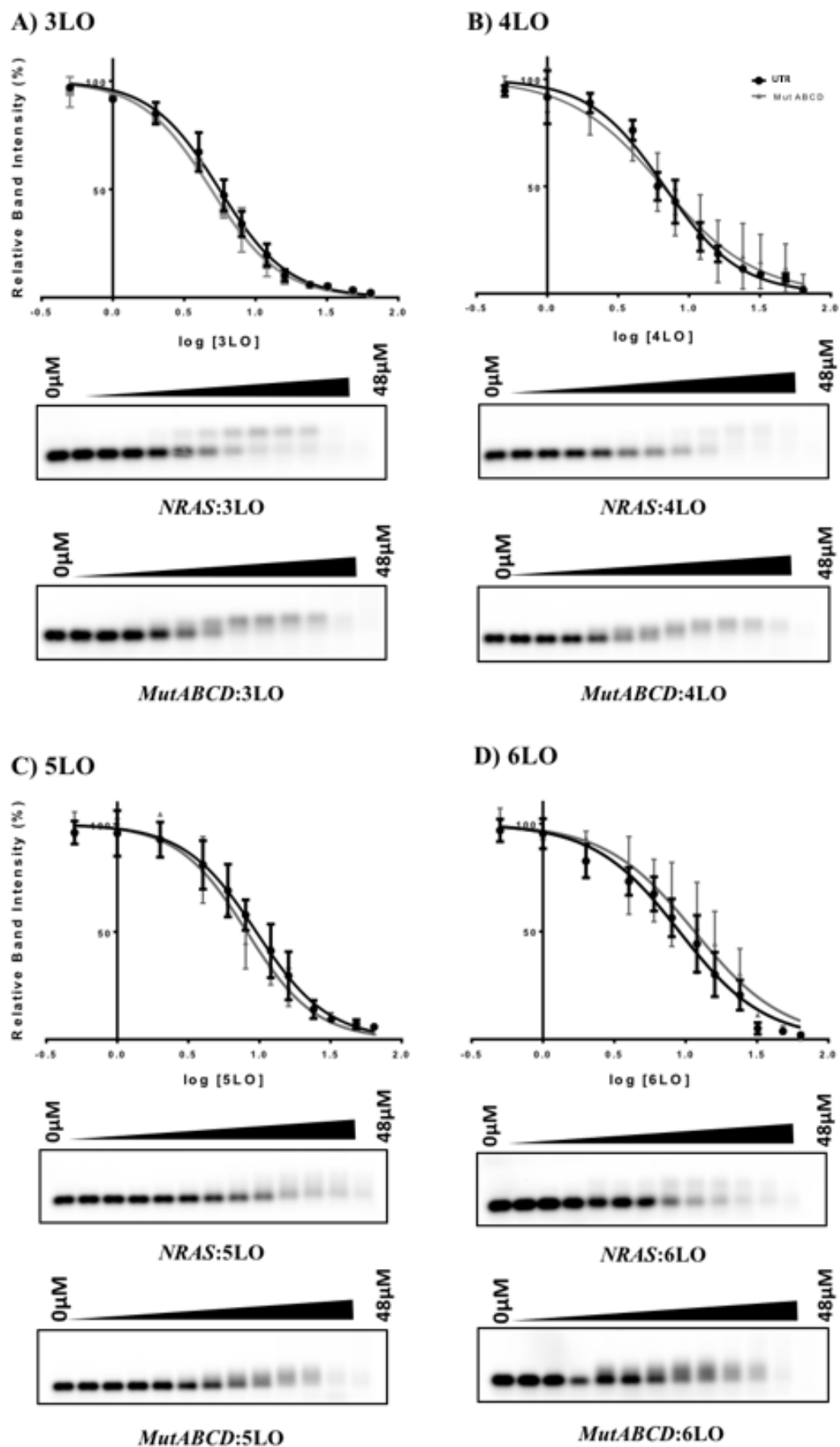
The attenuation of the appearing band as a result of the quenching effects discussed above makes estimation of the fraction of RNA bound, as described earlier in this chapter, highly inaccurate. As the

appearing band attenuates it will skew the results, artificially decreasing the fraction bound. To account for this, the apparent K_d was estimated from the disappearing band alone, assuming a B_{max} of 1.0. This is a reasonable assumption as the disappearing band appears to tend towards zero. As this band would not be affected by quenching, it is safe to assume that this represents near to 100% of the ligand binding.

To better understand the results observed in the translation assays, the binding of PNA-conjugates (6LO, 5LO, 4LO, and 3LO) and the “PNA” control ligand to target UTR mRNA and flanking regions was assessed with an EMSA (Figure 65). In particular, the differences in binding affinity of the ligands was analysed to understand whether they correlated with the effect of the same ligands on translation inhibition. In addition, the binding of the above ligands to Mut ABCD mRNA was tested to understand their lack of specificity observed in the *in vitro* translation assays. In particular, we wanted to understand how PNA-NDI conjugates could bind to non-complementary RNA oligonucleotides.

All PNA-NDI conjugates bound to target UTR mRNA with a significantly higher affinity than the control “PNA”, demonstrating that the presence of the “NDI” ligand has a significant impact on PNA binding. The effect of PNA length on binding affinity was less clear. There was no clear significant trend with increased PNA length and the binding of 6LO was not significantly higher, suggesting that its significantly increased translational inhibition is not due to increased binding affinity.

Results of the *in vitro* translation assays revealed no significant difference in IC_{50} for PNA-NDI conjugates binding to UTR or Mut ABCD mRNA. However, gel images showed two distinct appearing bands at higher ligand concentrations with 3LO, 4LO, 5LO and 6LO binding to Mut ABCD RNA (for example with 3LO, as shown in Figure 66), while only one appearing band was visible for the same ligands binding to the UTR RNA. This suggests that there is at least one different non-specific binding mechanism involved. The first appearing band is present at lower ligand:RNA ratios, first visible at ligand concentrations of 4 μ M and with maximum intensity at concentrations of around 16 μ M. The second appearing band is a lower mobility band that is visible from 9 μ M ligand concentrations with maximum intensity at 32 μ M. The fact that the second band becomes visible at higher concentrations, is less mobile than the first disappearing band and is the same distance from the first band as the first band is from the disappearing band suggests that it may be formed by two molecules of ligand binding to the same RNA. This means that there could be two separate binding modes available to the ligands: one specific binding via complementary base pairing between PNA strands and the flanking regions of the quadruplex; and one non-specific binding via another mechanism, potentially an electrostatic interaction between the positively charged lysine residues in the ligand and the negatively charged backbone of the RNA.



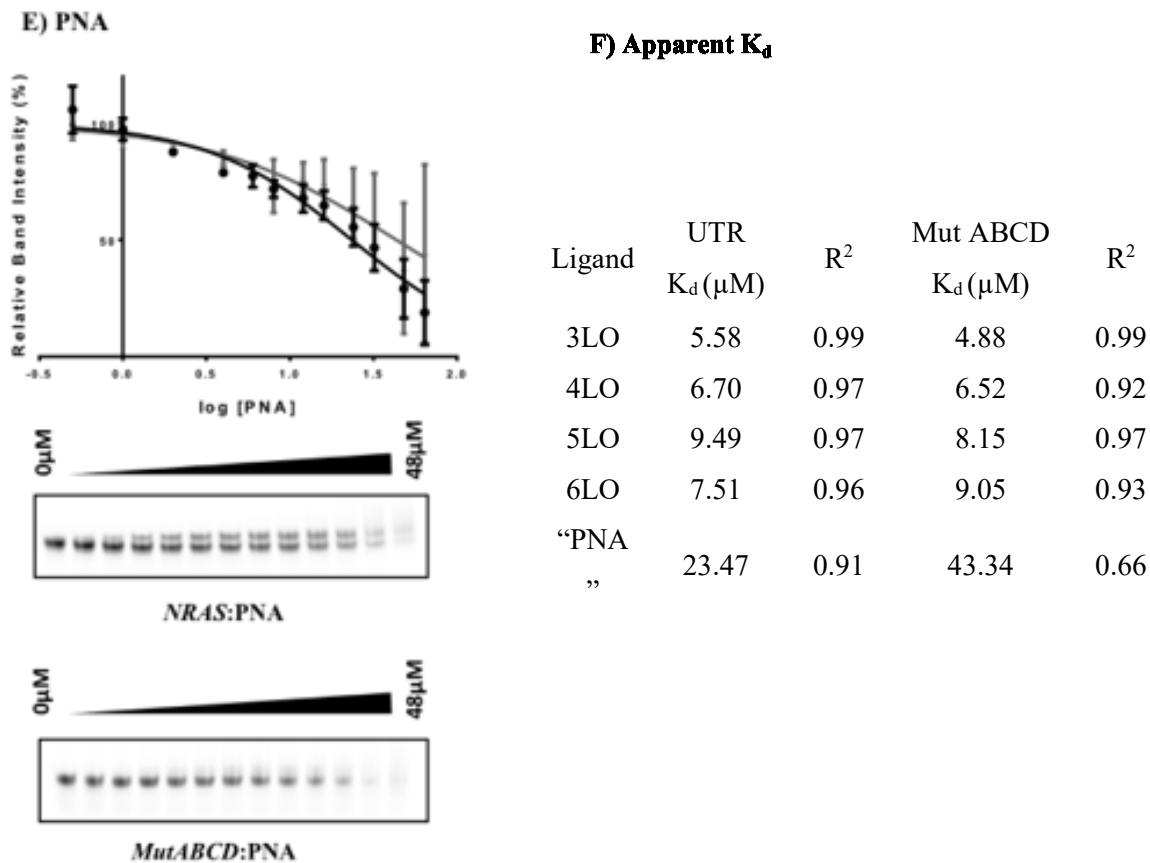


Figure 65: Binding of PNA-conjugates to UTR and Mut ABCD mRNA, measured using electrophoresis mobility shift assay (EMSA). Panels A-E show the results of the EMSA – the graph on top plots the relative intensity of the unbound bands (normalised so the intensity of the band with 0 μ M ligand is equal to 100%), results for the UTR mRNA are shown circles and the line of best fit shown as a black line, results for Mut ABCD are plotted as grey circles with the line of best fit shown as a grey line. Underneath the chart are shown the original images of the gel, with an example of the UTR mRNA gel on top and the Mut ABCD gel underneath. The table in panel F shows the apparent K_d calculated from the lines of best fit in panels A-E for both UTR and Mut ABCD mRNA along with the R^2 value for the K_d . The results do not show a clear difference between binding to Mut ABCD and UTR across the ligands suggesting that there is some non-specific binding. The only ligands where there appeared to be stronger binding for UTR than Mut ABCD were 6LO and “PNA”. Lines of best fit generated using non-linear regression in Graphpad prism. Error bars represent standard deviation for $n>3$ repeats.

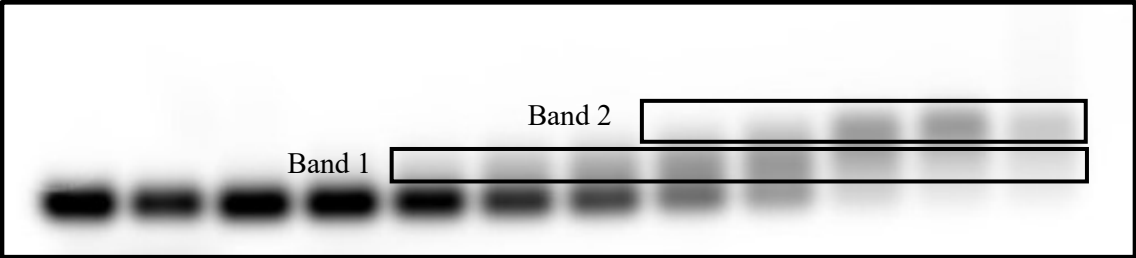


Figure 66: Binding of 3LO to Mut ABCD mRNA. Dual appearing bands on binding of 3LO to Mut ABCD RNA in polyacrylamide gel run at 120V for 120 minutes. These bands suggest that 3LO may have multiple binding modes when binding to Mut ABCD mRNA.

2.4. Discussion

2.4.1. Summary

This chapter set out to demonstrate that conjugation of PNA sequences complementary to the flanking regions of RNA G4 can increase the inhibitory effect of G4 stabilising small molecule ligands on translation. Results demonstrated that PNA-NDI conjugates exhibited significantly lower IC_{50} than either “NDI” (9.5 μ M) or “PNA” (>10 μ M) alone, with 6LO in particular demonstrating around a 10-fold increase in activity compared to “PNA” or “NDI” binding alone. However, there was no clear evidence to show that this inhibitory effect was due to stabilisation of the *NRAS* PQS. There were no consistent, significant differences between the effect of the PNA-conjugates on the *NRAS* 5'UTR mRNA which contained the native PQS and the *NRAS* Mut G mRNA which contained a mutation preventing G4 formation. Initial CD spectroscopy studies showed that the mutation had been effective at preventing G4 formation, so it therefore suggests that the PNA-conjugates inhibited translation through a G4 independent mechanism. In addition, scrambling the flanking region of the PQS which is complementary to the PNA strands in the ligand also did not result in consistent, significant reduction in translation inhibition. This suggests that the effect was also independent of specific PNA conjugation to the *NRAS* PQS sequence. As a result, we suggest that there is an independent non-specific interaction between the PNA-conjugates and a critical component in the translation process, either within the mRNA sequence or within the ribosomal RNA.

Comparison with previous studies investigating oligonucleotide small molecule conjugates is difficult as a number of studies have looked at the challenge from the opposite angle, i.e. thinking about how a small molecule can improve oligonucleotide binding rather than how oligonucleotide binding can improve small molecule binding. The most relevant existing study in the literature to compare these results with is that of the Helene group, where the effect of oligonucleotide conjugation on the specificity of an intercalating ligand was specifically tested.²⁷² Binding of an intercalator (9-aminoacridine) linker-oligonucleotide conjugate and an intercalator linker conjugate were compared. Both conjugates were bound to a target DNA strand which contained a complementary sequence to the oligonucleotide. When bound to the target sequence the conjugates emitted fluorescence, however the fluorescence was quenched by non-specific binding. To test the relative specificity of the conjugates, calf thymus DNA was titrated into the solution. The intercalator binds in a non-specific manner and the addition of the oligonucleotide was theorised to increase specificity as, in addition to affinity for the complementary sequence, the negative charge of the oligonucleotide was expected to inhibit non-specific binding. This proved to be accurate, as titration of calf thymus DNA resulted in quenching of the intercalator conjugate without an oligonucleotide sequence, whereas the fluorescence emitted by

the intercalator-oligonucleotide conjugate remained constant, showing that the conjugate remained bound to its target DNA. The notable difference between this study and our work is that the oligonucleotide was negatively charged and therefore there was electrostatic repulsion between the oligonucleotide and the target DNA. Our ligands were solubilised with positively charged lysine groups, which could result in non-specific binding, as discussed on page 94. This could explain why this study showed an increase in specificity for conjugates, whereas we did not identify any additional specificity from the conjugates.

Another study which showed an increase in specificity in oligonucleotide conjugates was the 2005 study by Nielsen *et al.* investigating the binding properties of PNA-Hoechst conjugates.²⁷⁵ The experiments showed that conjugation of a minor groove binding ligand could be used to target a PNA sequence towards a specific complementary sequence located next to an A-T region and that it would preferentially bind that sequence over identical complementary sequences that are not located next to an A-T region. An initial key difference to note is in the experimental approach used. Both this study and the previous study mentioned (that of the Helene group) tested the specificity of the conjugates by mixing them with target and non-target DNA and then identifying which one it preferentially bound to, either through fluorescence or through separation using gel electrophoresis. This enables a direct comparison of the affinity of the ligand for the different targets and can avoid the difficulties of comparing between separate experiments. This would be a worthwhile approach to use in future studies. The experiment conducted by Nielsen *et al.* investigates binding of a PNA-small molecule conjugate to an oligonucleotide sequence containing a complementary sequence to the PNA only and an oligonucleotide containing a complementary sequence and a proximal binding site for the small molecule ligand. In this sense, it is most directly comparable to binding of our PNA-small molecule conjugates to the Mut G mRNA (containing a complementary sequence, but no small molecule binding site) and the UTR sequence (both a complementary sequence and small molecule binding site). Unfortunately, the affinity of Mut G was not tested using the same technique (EMSA). It is notable however that the methodologies used for EMSA in this experiment were different from our own. The gels were run at 37 °C for 20 hours rather than on ice at 100V for 2 hrs and imaged using ³²P labelling rather than fluorescence which avoids issues with quenching. The reason for the experimental differences is that RNA is significantly easier to degrade than DNA (used in the Nielsen experiment) and considering these additional challenges, it may be useful for experiments using RNA to be conducted using different techniques that do not expose the RNA to potentially harmful environments. Moving on to compare the results, Nielsen *et al.* found that conjugation improved specificity, which does not match with the results of our translation assay, this may be partially due to the differences in experimental approach as mentioned already. It may also be partially due to the nature of the small molecule. The Hoechst is a minor groove binding ligand which can bind to the A-T sequence in the target DNA in its duplex form. However, in our experiments the binding of the NDI is dependent on

the formation of the G4. If our experimental conditions prevented the formation of the quadruplex, then there would be no difference in the binding affinity of the PNA-conjugates towards the UTR or Mut G mRNA.

However, earlier work by the Nielsen group investigating PNA-acridine conjugates showed similar results to our work in terms of specificity.²⁷⁴ In these experiments, unpublished data compared the impact of a mutation in the target DNA on the binding affinity of PNA and PNA-acridine conjugates. The authors reported that the PNA conjugates were consistently “as specific or superior” to the PNA alone. Compared to other reports which suggested a significant increase in specificity, this is more similar to our findings which suggested limited difference in specificity. Like our work, and unlike the other studies mentioned above, in this experiment the binding of the conjugates to target and mutated DNA was not tested by incubating all three together in the same experiment.

In summary, work in the literature suggests that interestingly, the inhibitory activity of the ligands does not increase linearly with PNA length, with no significant difference in IC₅₀ between the 6-mer ligand 3LO (1.7 μ M) and the 10-mer 5LO (1.5 μ M). However, the 12-mer ligand 6LO showed a significantly lower IC₅₀ value (0.6 μ M), which we propose is due to increased antisense activity by the PNA strands. This is compared in more detail to previous studies on PNA length below. In addition, there are limitations with the assays/techniques used which could account for our inability to identify any increased specificity. There are also a number of modifications that can be made to the ligands to improve their binding specificity in the future, as discussed below.

2.4.2. Effect of increasing PNA length

Our initial hypothesis was that ligand affinity would increase in relatively linear fashion with PNA length, resulting in greater inhibition of translation. However, the apparent K_D values of ligands as calculated in EMSA studies appeared to be unaffected by PNA-length. Although, as previously mentioned, there was a significant non-linear effect of PNA length on translation inhibition, with 6LO showing a roughly 2-fold increase in activity when compared to the other PNA-conjugates.

Previous studies investigating the effect of oligonucleotide length on antisense activity have also observed similar non-linear effects. Johansson *et al.* tested the effect of length on the translational activity of 2'-O-allyl modified RNA oligonucleotides targeting the 5'UTR of chloramphenicol acetyltransferase mRNA in a cell free translation assay.²⁹⁶ Translation of target mRNA in RRL in the presence of a 40-fold molar excess of ASO was reduced by 96-99% after addition of 16-mer and 14-mer oligonucleotides, by around 80% for 12-mer oligonucleotides and by only 46% for 8-mers. A subsequent study by Flanagan *et al.* with C-5 propyne-modified ASO targeting the coding region of

luciferase mRNA found that a 15-mer induced a 59% reduction in luciferase expression at a concentration of 25 nM.²⁹⁷ Sequential shortening of the oligonucleotide from the 5' or 3' end reduced the effect on translation to 54% and 48% for 13-mers, 39% and 33% for 11-mers and to 10% and 17% for 9-mers. Most relevantly, Doyle *et al.* investigated the effect of length on PNA antisense activity in the 5'UTR in *in vitro* cells for 10-18 nt PNA sequences.²⁹⁸ The full 18-mer sequence inhibited luciferase translation by 88% and shortening the sequence to produce a 15-mer had only a minimal impact, resulting in 80% inhibition. In contrast, reduction to a 13-mer and 11-mer respectively produced 60% and 43% inhibition, and an alternative 11-mer and a 10-mer only reduced luciferase translation by 3% and 9% respectively. There is a close match between the sudden drop in efficacy between 6LO (12-mer) and 5LO (10-mer) and the results of these studies.

These studies could suggest that increased inhibitory activity of 6LO is a result of an increase in antisense activity by the PNA strands and that increasing the length of the PNA further could result in further step increases in inhibitory behaviour in cell free assays. However, such an increase in cell free assays would only translate to *in vivo* efficacy if coupled with an effective transfection technique. In a study investigating CPP-PNA targeting of *E.coli* translation the optimum length for PNA antisense activity was found to be 9-12 nucleotides, with the inhibitory activity of longer ASO limited by reduced cellular uptake.²⁹⁹

Another potential explanation for the increased efficacy of 6LO is that it binds in a more favourable position in the 5'UTR than the other ligands. Endoh *et al.* demonstrated that the inhibitory effects of G4 in the coding strand are affected by the position of the G4 with single nucleotide sensitivity. Repeated fluctuations in inhibitory effect were observed in repeatable 3-nucleotide patterns, suggesting that it is related to the reading frame of the ribosome.³⁰⁰ However, in the same study they observed no positional effect in the 5'UTR on the single nucleotide level, so it is unlikely to be affecting the activity of our ligands. Testing the activity of a wider range of PNA lengths could help identify if there is a pattern in inhibitory activity that fluctuates every 3 nt which would suggest a reading frame dependent effect. The efficacy of ASO and siRNA have also shown dependence on their position in the coding region, but not in the 5'UTR. However, there have not been any studies to our knowledge investigating the effect of PNA or ASO positioning in the 5'UTR on a single nucleotide level, so a single nucleotide positioning effect can't be ruled out.

2.4.3. Effect of targeting 3' or 5' flanking region

Targeting of just the 5'-flanking region with the 6-mer ligand 6LO_{HALF} resulted in a significant 2-fold increase in IC₅₀ when compared to the equivalent 6-mer targeting both the 5'- and 3'-flanking regions (3LO). This suggests that the strategy of targeting both flanking regions is more effective. However,

further experiments with a ligand targeting only the 3'-flanking region would confirm whether it was targeting of the 3'-flanking region itself that resulted in this increase, or the effect of targeting both sides of the G4. However, there is no evidence in the literature to suggest that targeting of the 3'-flanking region should be any more effective than targeting the 5'-region.

2.4.4. PNA-NDI binding mechanism

EMSA and CD results suggest that the PNA-NDI ligands have at least two separate binding modes. There are inherent limitations in the binding studies used which limit how much detail we can infer about binding mechanisms, but there are still several key characteristics of PNA-NDI binding that can be inferred:

- 1) There are multiple binding modes for PNA-NDI conjugates that involve different stoichiometries.
- 2) PNA can form non-specific interactions with non-complementary RNA.
- 3) Addition of NDI in PNA sequences can induce an alternative binding mechanism that is independent of G4 formation.

A complete understanding of the molecular basis of these effects could help explain the binding activity of PNA-NDI ligands and the lack of specificity observed in translation studies. FRET studies can provide some information on G4 formation upon ligand binding but provides only limited information.²⁹⁵ A full structural characterisation of the binding interaction via crystallographic studies is one option, although weaker non-specific interactions are unlikely to be preserved in the crystal structure. ¹H NMR is likely to provide the most relevant assessment of ligand binding mechanisms. Titration of ligand into the G-quadruplex can result in a broadening and/or shift of the peaks of protons adjacent to the ligand binding site. Well-resolved protons (i.e. imino protons) next to the binding site can be used to identify the location of these sites. However, there are multiple problems that limit the efficacy of NMR for studying ligand quadruplex interactions including; molecular aggregation, lack of intermolecular NOEs and poor definition of complexes.³⁰¹

1) Stoichiometry

The two separate ligand:RNA complexes observed in the EMSA of Mut ABCD RNA were visualised as two equally spaced bands, suggesting that they could represent complexes with a 1:1 and 2:1 ligand:RNA stoichiometry. There are a number of alternative techniques that could be used to confirm the stoichiometry of the bound ligand:RNA complexes. Electrospray ionisation (ESI) mass spectrometry for instance has been used to demonstrate affinity, stoichiometry and sequence selectivity of groove binders to duplex DNA.³⁰² Surface plasmon resonance (SPR) can also be used to measure binding affinity (see example³⁰³) and also to approximate stoichiometry via the equation:

$$\frac{RU_{DNA_{immob}}}{MW_{DNA_{immob}}} = \frac{RU_{max\ PNA}}{x(MW_{PNA})}$$

where: $RU_{DNA_{immob}}$ = the response units of immobilized DNA

$RU_{max\ PNA}$ = saturating response units obtained during PNA titration

MW = molecular weight

x = number of PNA strands hybridised to each DNA strand.

However, this would be a relatively complex experiment to undertake. Optimising surface binding is a difficult procedure and the different structures formed by the G4 and flanking regions would make the results difficult to interpret. Repeating the EMSA experiments using radiolabelled ligands could also provide additional information on the two observed binding modes. Radioactive-labelling should be unaffected by the binding mode and would allow accurate quantification of the appearing bands (corresponding to the RNA:ligand complexes) and more accurate calculation of binding constants for the different complexes.

2) Non-specific PNA interactions

The presence of non-specific interactions between the PNA-NDI ligands and the UTR mRNA can be inferred from the translation assays, where there was no significant difference between the effect of ligands on UTR and Mut ABCD RNA. This is supported by the results of our binding studies, where interactions were observed between both “PNA” and PNA-NDI ligands and Mut ABCD. Potential non-specific binding interactions include: **a)** electrostatic interactions, **b)** mis-pair binding between “PNA” and Mut ABCD flanking regions **c)** PNA:G4 interactions.

- a) **Electrostatic interactions:** PNA-NDI conjugates are positively charged due to the lysine groups introduced to improve solubility in water. These have the potential to form electrostatic interactions with the negatively charged RNA backbone. This effect has been observed in previous studies, with charged bis-PNA showing increased binding to non-specific DNA, especially at low salt concentrations.³⁰⁴
- b) **Mis-pair binding:** In the design of the Mut ABCD mutation the pyrimidines (C, T) were switched to the opposing pyrimidine (i.e. C was changed to T) and the same with purines (A, G). This was to prevent Watson-Crick base pairing interactions while maintaining the same purine:pyrimidine ratio. However, the sequence was not fully scrambled, which with regular nucleotides would allow the possibility for G/T and A/C mis-pair binding. However, as previously mentioned, the stability of PNA:RNA duplexes is significantly reduced by even single base mismatches, which makes it unlikely that any significant stable binding occurred with 6-12 mis-paired nucleotides.

- c) **PNA:G4 Interactions:** There is the potential for interaction between the PNA strands and the grooves or loop regions of the quadruplex, for example, the 3'-PNA sequence CCGC could bind in the major groove of the G4 through Hoogsteen hydrogen bonding. However, the pH dependence of C-G hydrogen bonding and mismatched guanine in the PNA sequence would make this a relatively unstable binding mode.

3) **Non-G4 mediated effect of NDI**

The effect of 6LO on the CD signal of Mut G in the absence of any effect in the presence of “NDI” or “PNA” ligands individually shows that there is an effect that is independent of G4 formation. This particular effect is difficult to rationalise but could simply be down to differences in the size or flexibility of the NDI group when compared to the linker used in the control “PNA”. Alternatively it could be mediated by non-G4 NDI interactions, for example NDI ligands have been shown to bind to duplex DNA and could potentially bind to PNA:RNA duplexes through a similar mechanism.³⁰⁵ Further investigation would be needed to ascertain which, if any, of these mechanisms are involved.

2.4.5. Specificity of PNA-NDI conjugates

Any of the “PNA” or NDI-mediated non-specific interactions discussed above could contribute towards the lack of specificity of the PNA-NDI conjugates in the translation assays, although electrostatic interactions are likely to have the largest effect. There are a number of different modifications to the PNA-NDI conjugates that are likely to improve their sensitivity and specificity and are therefore worth investigating in future studies, including modifications in; **1)** solubilising groups, **2)** PNA backbone, **3)** G4 binding ligand, **4)** spacing groups and **5)** PNA strand orientation.

1) **Solubilising groups**

In order to improve the solubility of the PNA-NDI conjugates, positively charged lysine residues were added to the PNA. These positively charged residues form electrostatic interactions with the negatively charged phosphate groups on the target RNA. Charge-free polyethylene glycol (PEG) groups can also be used to improve solubility while avoiding unspecific electrostatic interactions with negatively charged nucleic acids.³⁰⁶

2) **PNA Backbone**

Changing the PNA backbone could also improve the chemical properties of the PNA conjugates. For example, gamma-peptides have demonstrated improved cellular uptake and higher affinity and sequence specificity.²⁶⁵

3) G4 binding ligand

The NDI derivative designed for this study does not possess intrinsic specificity for RNA G4. Replacing or modifying the NDI could prevent any NDI-mediated non-specific binding. For example, a simple modification to improve the specificity of NDI ligands was identified by Collie *et al.*, who found that replacing the –NH₂ side chain group with an –OH group improved specificity for RNA G4 over DNA G4 and duplex DNA.²⁸² Alternative G4 binding platforms could also be considered.

4) Spacing group

In the initial design of the ligands the optimal length for the linkers between the PNA and NDI were estimated from the sequence. Modelling studies to predict the structure of the PNA-NDI:G4 complex and rationally design the nature and length of the linkers between the NDI and PNA required for optimal binding (e.g. identify any steric clashes with the spacing group) could help increase the binding affinity of the ligands. However, modelling of the G4 RNA with flanking regions is very complex due to the relative flexibility of the flanking regions and although considered at the beginning of this study, modelling of PNA:RNA complexes was deemed too challenging.

5) PNA strand orientation

Our initial studies suggested that strand orientation did not affect translational inhibition of the ligands and did not give a clear indication of its effect on non-specific binding. Further studies to investigate the effect of strand orientation on non-specific binding by PNA-NDI conjugates could give further insight.

2.4.6. Assay limitations

The lack of specificity observed in the translation and binding assays may also reflect limitations of the assays, as well as non-specific binding by the ligands. The potential limitations of each assay type are discussed below.

1) Translation assays

The different *in vitro* transcribed RNAs tested in the translation assays have significantly different base expression levels, the relative luciferase expression of Mut G RNA being around 4x higher than that of UTR RNA. This makes direct comparisons between the effects of the ligands on the different RNA less accurate. In addition, the effect of the ligands on the positive control suggests that there may be artefacts produced by the assay conditions. The effect of weak ligand interactions on translation is likely to be attenuated in cells and *in vivo* where additional regulatory mechanisms

take place. The mammalian dual-luciferase assay, or a GFP-based mammalian cell reporter assay, are alternatives to quantify the effect of ligands on translation.³⁰⁷ They have the advantage of being in a human cellular environment, and GFP assays have the additional advantage of not requiring lysing of the cells before quantification.

2) Binding Assays

Several common techniques to assess the relative stability of ligand:G4 complexes were unsuitable for use with our ligands. UV melting is commonly used to assess the stability of quadruplex structures and ligand binding can be tracked through its effect on the melting point of a quadruplex measured at 295 nm.³⁰⁸ However, the NDI absorbs in the same area of the spectrum and would parasite the results. CD melting was also attempted, but accurate quantification was difficult due to the appearance of a band at 270 nm (next to that specific to G4 at 264 nm) as the RNA unfolds.

3) EMSA

The binding affinities of ligands in a gel environment do not necessarily reflect their binding in *in vitro* translation assays or *in vivo*. Differences in the buffer conditions and the stresses induced by movement through the gel structure can both influence binding. In addition, the complexes are not at a chemical equilibrium while the gel is being run, as once the complex dissociates the two molecules can be physically separated by their different gel mobility. Finally, the mobility of the complexes is not directly related to molecular weight, which means that identifying different complexes and binding modes is not always possible.^{309, 310}

4) CD

CD has been used to identify ligand binding to G4 through its effect on the CD spectra, however there is not yet a validated approach to directly infer the changes to the structure that produce these spectral changes.²⁸⁵ As such, CD spectroscopy can provide qualitative information about the effect of ligands on G4 stability or formation, but must be cautiously interpreted when trying to identify binding mechanisms.

2.4.7. Future work

The next steps for this study would be to re-assess the design of the PNA conjugate ligands and the techniques used in the study. From comparison with previous studies in the literature I would hypothesise that the non-specific binding effects are likely to be due to the positively charged solubilising groups added to the PNA. The first step will be to test this theory. Other modifications mentioned above such as use of alternative PNA backbone, optimisation of spacing group and G4 binding ligand can also be tested. In terms of experimental approach, the approach used elsewhere in

the literature of testing specificity by including the PNA-conjugate with both the target and non-target oligonucleotide should be tested. Different methods of separating the two complexes can also be tested, i.e. creating differing length target and non-target DNA and separating on a gel or attaching different binding groups (i.e. biotin) to the target and non-target oligonucleotides and separating on a chromatography column, then quantifying the relative amounts through mass spectroscopy.

Chapter 3: Improving the cellular uptake of PNA by sonoporation

3.1. Introduction

A significant disadvantage of PNA as therapeutic agents is their limited cellular uptake.³¹¹ There has been significant work over the last two decades developing different techniques to improve PNA uptake *in vivo*. The most successful methods include backbone modifications, lipofection and attachment of molecular transporters and are discussed in more detail below, with particular attention paid to transporters belonging to the family of so-called cell penetrating peptides (CPP). While many of these techniques have been shown to effectively improve uptake *in vivo*, limitations including cost, complexity of synthesis, toxicity and lack of bioavailability have limited their efficacy in clinical settings. A method that has received growing attention over the last few years is sonoporation, an ultrasound-based method to improve cellular delivery. It requires no covalent attachment to the cargo and shows promising safety results in exploratory *in vivo* studies.³¹² In this chapter, an *in vitro* assay to assess the sonoporation-mediated uptake of PNA is developed and transfection efficiency achieved via sonoporation is compared with that obtained by covalent attachment to one of the most efficient CPP to date (i.e. (R-Ahx-R)₄).

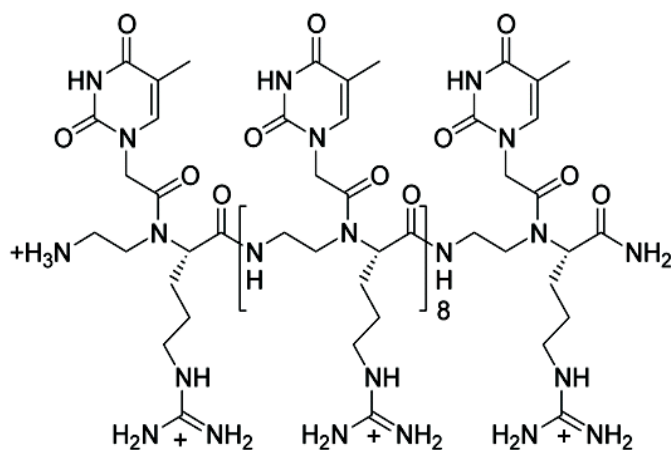


Figure 67: Structure of guanidine PNA (G-PNA). A single monomer is delineated using square brackets.

3.1.1. PNA backbone modifications

There are a small number of PNA backbone modifications which have been shown to improve the cellular uptake of PNA oligonucleotides. In particular, guanidine-based PNA (G-PNA) translocate across the cell membrane at similar efficiencies to the TAT CPP, a peptide derived from the HIV-1 protein.³¹³ The backbone of these G-PNAs consists of arginine amino acids linked together with the guanidine group extended as a side-chain (Figure 67). Other backbone modifications have shown more

modest improvements in uptake, for example the introduction of fluorine into the backbone of *apg*-PNA increases lipophilicity and was shown to improve the cellular uptake of the PNA by 2-3 fold.³¹⁴

3.1.2. Lipofection

Lipofection has emerged as an efficient alternative for the transfection of DNA and siRNA. This technique utilises positively charged transfection reagents, such as cationic lipids and polymers, which spontaneously self-assemble with negatively charged nucleic acids and enable translocation across biological membranes. However, these reagents do not interact with charge-free PNA. There are a number of PNA backbone modifications which have been suggested to allow conjugation of cationic carriers. These have shown some success in improving cellular uptake but often reduce hybridisation efficacy and specificity and increase the cost and complexity of synthesis.³¹⁵⁻³¹⁷ Other techniques, including conjugation of lipophilic tails to the PNA³¹⁸ and binding of partially complementary charged oligonucleotides to allow lipid binding³¹⁹ have also demonstrated efficacy at improving cellular uptake, although with similar drawbacks.

3.1.3. Cell Penetrating Peptides (CPP)

Covalent attachment of molecular transporters is currently the most promising method to improve transfection of PNA molecules. The most common transporters are cell penetrating peptides, which are short peptide sequences that induce cellular uptake. Initial CPPs were identified from natural sequences such as the HIV-1 Tat protein, which was first identified to have cell penetrating properties in 1998 and was subsequently found to effectively transport conjugated proteins into cells.³²⁰⁻³²² Other natural CPPs have been identified and provided the template for subsequent rationally designed CPP. There is high structural diversity among CPP, but they can be broadly classified as either cationic, amphipathic or hydrophobic depending on the prevalence and distribution of charged amino acids within the sequence.³²³

An important milestone in the application of CPP to improve PNA uptake was the first example of effective *in vivo* antisense activity by a CPP-PNA conjugate, which was demonstrated in the rat spinal cord.³²⁴ Conjugation of CPPs to PNA is relatively simple (although time-consuming and expensive) and does not require any alteration of the PNA backbone, although covalent attachment can have an effect on the activity of the PNA molecule. The optimal CPP sequence for use with a specific PNA is affected by its sequence and length, meaning that there is no universal CPP sequence optimal for all PNAs. Non-covalent attachment mechanisms have been very successful in improving uptake of oligonucleotides³⁸ but can only be applied to PNA with charged backbone modifications.^{325, 326}

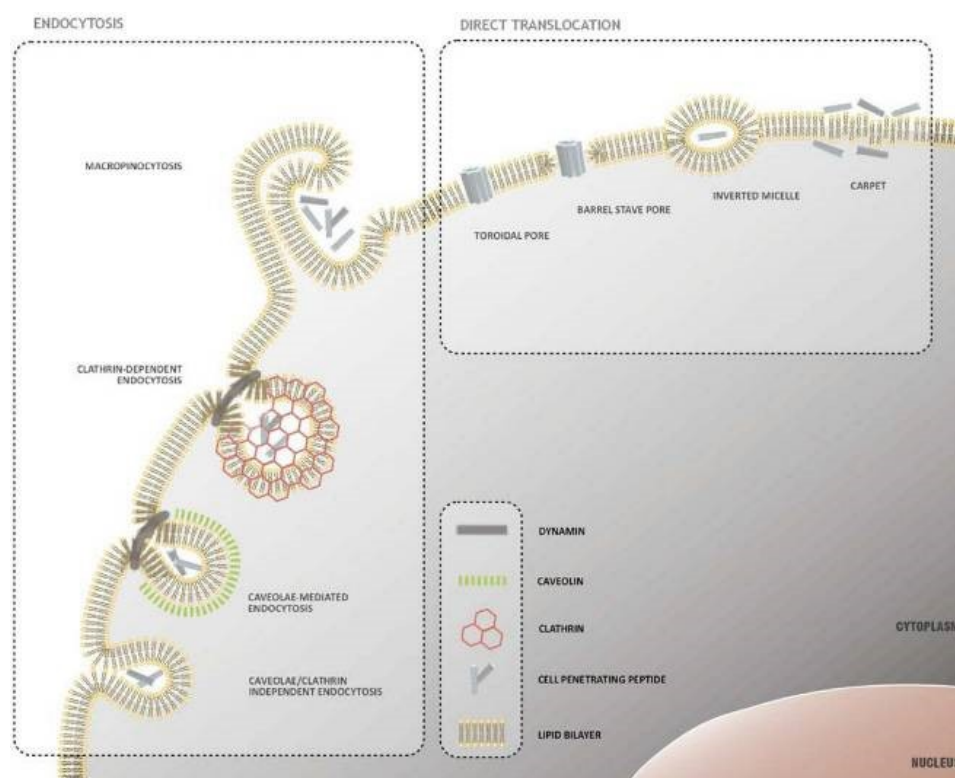


Figure 68: Suggested mechanisms for CPP-mediated uptake. The different mechanisms that have been suggested for uptake of CPP can be broadly split into two categories, endocytosis and direct translocation. Endocytosis can then be subdivided according to the mechanism of cargo encapsulation, including interactions with cell surface proteins such as clathrin and caveolin, and micropinocytosis (left-hand side). Direct translocation can be characterised as either translocation via pores or localised disruption of the cell membrane (right hand side). Figure reproduced from Trabulo *et al.*³²⁷

(i) Mechanism of CPP uptake

The broad consensus is that cellular uptake of CPPs is mediated by endocytosis under most experimental conditions, although this is somewhat controversial and it is likely that other mechanisms, such as direct translocation, are involved under certain conditions³²⁸ (for a schematic representation of different uptake mechanisms see Figure 68). A concern with uptake via an endosomal route is reduced bioavailability of the transfected cargo due to endosomal entrapment.^{329, 330} However, two families of arginine-rich CPPs, named (R-Ahx-R)₄AhxB and R₆Pen (where Ahx = amino hexanoic acid, B = beta-alanine and Pen = penetratin peptide), have been shown to allow efficient nuclear delivery of splice correcting PNAs and PMOs at micromolar concentrations in the absence of endosomolytic agents.³³¹ The efficiency of (R-Ahx-R)₄AhxB has been demonstrated *in vivo* in murine Duchenne muscular dystrophy models.³³² Furthermore, Deuss *et al.* used splicing redirection assays to demonstrate the relative bioavailability of a library of CPP-PNA705 cargo.³³³ They found that the bioavailability of CPP-PNA was directly correlated with the number of positive charges in the CPP sequence, particularly with arginine. However, these arginine-rich CPPs are relatively long sequences (6 and 12 residues) which drastically increases the cost associated with the synthesis and purification of the corresponding PNA-CPP conjugates.

(ii) Systemic delivery of CPP-PNA

A major limitation for *in vivo* applications of CPP is their vulnerability to proteolytic degradation,³³⁴ although some CPPs incorporating modified peptides, for example β -peptides,³³⁵ oligocarbamates³³⁶ and proline-rich peptidomimetics³³⁷ have shown increased stability. A further concern is that CPP delivery is not typically tissue specific, which increases the required dosage and can be a safety concern when delivering non-specific drugs. There are various methods to add tissue specificity to CPP, however these often involve conjugation of targeting molecules and chemical linkers, which further increase the complexity and cost of synthesis and require modification to adapt to different targets.³³⁸ Examples of two representative methods are the addition of cell-specific targeting peptides³³⁹ and stimuli-initiated activation.³⁴⁰

Targeting peptides, also called homing peptides (HP), are peptides that localise to specific cell types through interaction with cell surface receptors or proteins. The first HP targeted towards tumours (RAFT-RGD) was developed to recognise integrin $\alpha_v\beta_3$, an extracellular matrix receptor overexpressed on neoendothelial cells and many cancer cell sub-types.³⁴¹ Some HP can also improve cellular uptake of cargo through receptor-mediated endocytosis, for example Pep42 which binds glucose-regulated protein 78.³⁴² Other homing proteins have been conjugated to CPP to successfully deliver cargo *in vitro*.^{343, 344} However, HP-CPP delivery systems have so far shown limited efficacy *in vivo* and in the clinic. In several studies HP have been observed to interfere with the activity of both CPPs and cargos, which could explain this lack of efficacy.^{345, 346}

Stimuli-initiated activation typically involves the addition of inactivating groups to the CPP, such as chemical groups to conceal the basic residues on the CPP or PEG moieties to mask the overall CPP structure, using cleavable linkers. The choice of linker affects the activating stimuli, which can be chosen to target different tissues of disease phenotypes. Examples of disease phenotypes suitable as activating stimuli include: high H_2O_2 in areas of oxidative stress,³⁴⁷ low pH in acidified tumour environments^{348, 349} and matrix metalloproteinase-2 overexpression for tumour targeting.³⁵⁰ External stimuli can also be used, for example UV, which allows laser targeted release.³⁵¹ A particularly interesting example is delivery of PNA oligomers with microbiota activated CPP, which facilitated the specific targeting of colonic epithelial cells.³⁵² Cleavable linkages have also been used to release the CPP from its cargo upon cellular uptake, for example the reductive environment within the cell can cleave disulphide linkages, releasing the cargo from the CPP.

3.1.4. Cell Penetrating Peptoids

An interesting class of CPP mimetics named cell-penetrating peptoids (CPPo) are also worthy of mention. CPPo are oligomers of *N*-substituted glycines that exhibit similar cell penetrating properties to CPPs but are highly resistant to enzymatic degradation.³⁵³ The *N*-substituted glycine backbone is achiral and does not contain any donor groups which can participate in hydrogen bonding. This means that peptoids do not form any major secondary structures. They have been demonstrated to provide effective delivery of siRNA³⁵⁴ and PNA³⁵⁵ and can be modified with cationic side chains to escape endosomal entrapment and target specific cellular organelles.³⁵⁶ As such, they provide a potential alternative to CPPs as a cellular delivery mechanism.

3.1.5. Sonoporation

Sonoporation is a transfection technique that has been used to increase the cellular uptake of fluorescent dyes,³⁵⁷ DNA plasmids³⁵⁸ and antisense nucleotides such as siRNA.³⁵⁹ It involves temporary permeabilisation of cell membranes through the co-application of ultrasound and microbubble contrast agents. Microbubbles are gas-filled macromolecules stabilised by a lipid, peptide or polymeric shell. Due to their gas cores microbubbles are echogenic, which facilitates their use as contrast agents, and also oscillate when subjected to ultrasound. At lower intensities the oscillation of microbubbles is inversely proportional to ultrasound intensity and is known as stable cavitation.³⁶⁰ At higher pressures microbubbles respond in a non-linear fashion as the decreasing integrity of the shell makes them more resistant to compression than expansion. This process is known as inertial cavitation.³⁶¹

(i) Pore Formation

When microbubble cavitation occurs near to the cell membrane it can increase the uptake of local non-permeable molecules into the cell. There are a number of different mechanisms that have been suggested for microbubble-mediated pore formation (Figure 69A-E). Firstly, that the microbubble cavitation results in a push or pull effect which distorts the membrane and ruptures it, forming temporary pores which increase the permeability of the membrane.³⁶² In addition, the fluid streaming around an oscillating bubble near the cell membrane can produce shear stress of sufficient intensity to rupture the membrane and form a pore.³⁶² Lipid encapsulated microbubbles have been shown to fuse or translate through the cell membrane.³⁶³ At relatively high pressures, during inertial cavitation, microbubbles can undergo asymmetric collapse. This can result in microbubble jetting, where the force of this collapse creates a funnel shaped “jet” that can disrupt the cell membrane.³⁶⁴

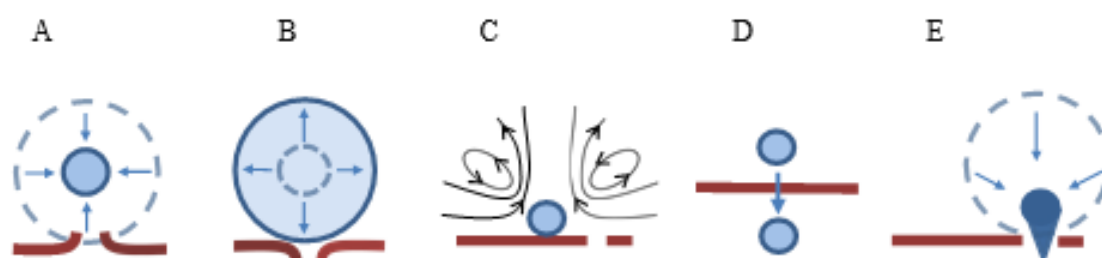


Figure 69: Suggested methods for microbubble-mediated pore formation in sonoporated cells. (A) Pull effect where contraction of the microbubble creates a force that splits the cell membrane, (B) Push effect where expansion of microbubbles creates a pore in the cell membrane, (C) Fluid streaming induced shear stress, whereby the currents induced by expansion and contraction of the microbubble creates a pore in the membrane, (D) Translocation whereby the microbubble moves through the cell membrane, (E) Inertial cavitation where the microbubble over expands, losing its elasticity, and collapses in on itself to create a strong force.

(ii) Mechanism of sonoporation-induced cellular uptake

Cellular uptake following sonoporation has been suggested to occur through both endocytosis and pore formation, with different mechanisms accounting for transport of different sized cargo. Sonoporation was used to deliver dextran molecules from 4-500 kDa in size into primary bovine aortic endothelial cells with and without inhibitors of clathrin and caveolin-1 mediated endocytosis and micropinocytosis.³⁶⁵ The results suggested that pore formation at least partially accounted for transfection of smaller molecules (4-70 kDa) but that larger molecules (155-500 kDa) were transported via a combination of different endocytic pathways. The results were further replicated *in vivo*, demonstrating the role of endocytosis in ultrasound mediated uptake of 500 kDa dextran in the endothelial cells of the rat femoral artery. Another study in a human prostate cancer cell line demonstrated that sonoporation could induce the uptake and localisation of plasmid DNA (pDNA) to the nucleus.³⁶⁶ Shortly after sonoporation around 30% of transfected pDNA was visible in the nucleus and remained over 24 hours. The remaining pDNA was initially split between autophagosomes/autophagolysosomes and cytoplasm but was cleared over the following 24 hours. There was no visible co-localisation of endosomal stains and the Cy3 labelled pDNA, suggesting, in contrast to the previously mentioned study, that endosomes did not play a role in uptake. The molecular weight of the plasmid was not specified, but even a small plasmid of 1000 nt is likely to be around 600 kDa. The different transfection efficiencies and optimum settings found by these individual studies demonstrate the effect that variations in cargo and experimental setup can have on transfection results.

(iii) Disruption of endothelial barriers

As well as inducing cellular uptake of cargos, co-application of ultrasound and microbubbles can also be used to disrupt endothelial barriers and allow drugs and nanoparticles entry into the surrounding tissues. A notable recent success in this area was the application of focused ultrasound in the presence of microbubbles to enable liposomes to cross the blood brain barrier (BBB).³⁶⁷ Researchers are currently

exploring potential therapeutic applications of microbubble-mediated crossing of the BBB in humans.³⁶⁸

(iv) Advantages of sonoporation

A major advantage of sonoporation for intracellular delivery is that it requires no chemical modification of the molecule to be delivered. If applied to PNA delivery, this means that it potentially offers a number of advantages when compared to more traditional bio-conjugation techniques (e.g., bio-conjugation to CPP and CPPo). No chemical modification of the PNAs is required prior to transfection, which means lower cost and increased simplicity and versatility. This also means that the integrity of the PNA is preserved and as a result its DNA (or RNA) binding properties will also be preserved. There are several types of microbubble contrast agents already in clinical use in humans for medical imaging, for example Sonovue³⁶⁹ and Optison,^{370, 371} which already have regulatory approval. Ultrasound is applied externally in a non-invasive manner and can also be spatiotemporally controlled, providing a unique opportunity to selectively target transfection to specific tissues. Microbubbles can even be designed which incorporate ligands or antibodies³⁷² to target them to specific cell types.³⁷³

(v) Limitations of sonoporation

The main concern with using microbubbles for gene therapy-type applications is their inability to cross the endothelium, which means that only endothelial cells can be transfected via systemic delivery. This implies that for *in vivo* applications to non-endothelial cells direct injection is required. However, nanobubbles and nanodroplets are currently being developed which can overcome this limitation, in particular when targeting tumours. The fenestrated epithelia in the vasculature of tumours and reduced lymphatic drainage produce an enhanced permeability and retention (EPR) effect which results in accumulation of nanoparticles in tumour tissue.³⁷⁴ Nanodroplets have been shown to accumulate in tumours and coalesce to form microbubbles.³⁷⁵ Other drawbacks include the ability of sonoporation to non-specifically permeabilise the cells. This results in membrane hyperpolarisation and disruption of ion gradients, both of which place the cell under stress and can eventually lead to apoptosis. The membrane can also permanently rupture leading to cell death.

3.2. Methods

3.2.1. Cell culture

HeLa pLuc705 cells were kindly donated by Prof. Michael Gait (MRC LMB, Cambridge, UK). HeLa and HeLa pLuc 705 cells were seeded in T25 flasks, with D-MEM containing 10% FBS, 1% glutamine,

1% Penicillin/Streptomycin and incubated at 37 °C and 5% CO₂. Media was replaced every 2-3 days and cells were passaged at 80-90% confluence using Trypsin/EDTA pre-warmed to 37 °C.

3.2.2. PNA and CPP synthesis

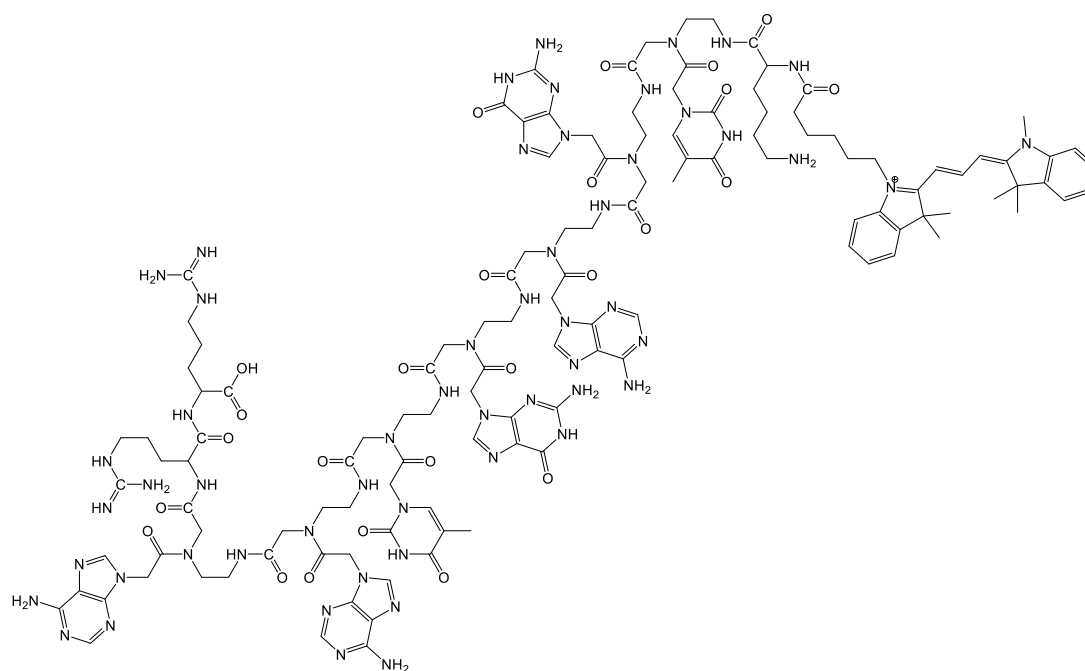
PNA-Cy3 (see Figure 70 for structure and characterisation by MALDI) and CPP-PNA were synthesised using standard solid phase peptide chemistry as described in Chapter 2. PNA-705 was purchased from Panagene Inc. (South Korea).

PNA-Cy3: Arg-Arg-AAT-GAG-T-Lys-Cy3

PNA-705: Lys-CCT-CTT-ACC-TCA-GTT-ACA-Lys

CPP-PNA: (R-Ahx-R)₄-AAT-GAG-T-Cy3 Where: Ahx = 6-aminohexanoic acid

A



B

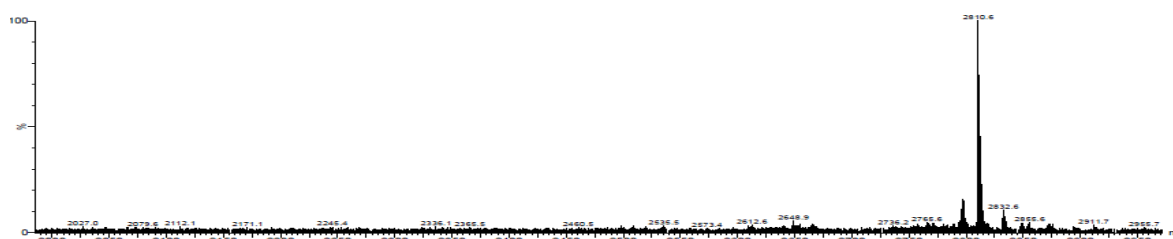


Figure 70: Chemical structure and mass spectroscopy (MALDI-TOF) characterisation of PNA-Cy3. A) Molecular structure of PNA-Cy3. **B)** Mass spectroscopy characterisation of PNA-Cy3.

3.2.3. Microbubble manufacturing

Microbubble solutions were kindly provided by Dr James Choi's group and manufactured as detailed in previous protocols.³⁷⁶ The solution consists of three commercially available lipids (Avanti Polar Lipids Inc.): Dipalmitoylphosphatidylcholine (DPPC), Dipalmitoylphosphatidic acid (DPPA), and Dipalmitoylphosphatidylethanolamine-PEG5000 (DPPE-PEG 5000) which were mixed and diluted with glycerol (5% v/v) and saline (80% v/v). Microbubbles were formed through activation of the solution immediately before each sonoporation experiment. Activation was achieved by introducing perfluorobutane gas into the container and shaking with an amalgamator (FluoroMed L.P) for 40 seconds. The average microbubble concentration in an activated microbubble solution was $7.3 \pm 4 \times 10^9$ microbubbles/ml.

3.2.4. Flow system

Confluent cells were seeded in collagen IV coated μ -Slide^{0.6} Leuer flow chambers (Ibidi) at a concentration of 2.5×10^5 cells/ml and incubated for >4 hours at 37 °C and 5% CO₂ to allow the cells to attach to the surface of the flow chamber. The chamber was then connected to the flow system using silicone and polypropylene tube adapters (Tube Adapter Set sterile, Ibidi GmbH Ltd., Martinsried, Germany) and silicone elastomer tubes (Saint-Gobain Performance Plastics, Paris, France), both of which had been autoclaved and filled with culture medium. The peristaltic pump (Harvard MPII Mini-Peristaltic Pump, Harvard Apparatus) was used to produce a constant flow rate of 24.5 ml/min and the cells were left to equilibrate at a flow rate of 24.5 ml/min for 2 hours at 37 °C.

3.2.5. Ultrasound targeting

Before sonoporation a 35 MHz ultrasonic pulser/receiver (DRP300, JSR Ultrasonics, Imaginant Inc., Pittsford, NY, USA) and picoscope (Picoscope PC Oscilloscope 3205A, Pico Technology Ltd., Texas, USA) were used to target the focal area of the transducer to the centre of an empty flow chamber. The flow chamber was set to a 65 μ s delay between the trigger pulse and the reflected ultrasound signal and centred so that the ultrasound focal area was in the middle of the chamber.

3.2.6. Sonoporation

Ultrasound equipment was set up as shown in Figure 73 unless otherwise stated. The flow chamber was secured in a water tank containing degassed water preheated at 37 °C and microbubbles were added to the media at a concentration of 2×10^6 microbubbles/ml and PNA-Cy3 at a concentration of 10 μ M.

The flow of 24.5 ml / min was left running for 2-4 minutes after addition of microbubbles with a stirrer running to ensure even distribution of microbubbles throughout the flow system. The ultrasound pulse signal was produced by a function generator (Agilent technologies) which was connected to a 50 dB amplifier and used to drive a 1 MHz transducer to produce the ultrasound wave. Ultrasound was applied for between 400 and 1600 seconds at a frequency of 1 MHz, Burst period of 1 second, Peak negative pressure (PNP) of 100-1400 kPa and cycle number of 50-1000. After the application of ultrasound, flow chambers were incubated for 30 minutes in the sonoporation medium at 37 °C and 5% CO₂.

3.2.7. Cell staining

Following post-sonoporation incubation, cells were stained with the nuclear intercalating dye Hoechst 33372 (Sigma-Aldrich) at a concentration of 1 µg/ml for 5 minutes. The chambers were then washed three times with 1xPBS pre-warmed to 37 °C. When applicable, Calcein-AM (Sigma-Aldrich) was added at a concentration of 1 µM for 30 minutes, then removed with three washes of 1xPBS pre-warmed to 37 °C.

3.2.8. Microscopy

Brightfield images of the flow cells were taken immediately before sonoporation, and after the 30 minute post-sonoporation incubation period, with a Zeiss PALM MicroBeam Laser capture microdissection microscope. The automated stage was used to ensure a consistent field of view for the images. A wide field Zeiss Axiovert 200 inverted fluorescence microscope (Carl Zeiss, Gottingen, Germany) and a Hamamatsu OCRA-ER camera (Hamamatsu Photonics Ltd, Tokyo, Japan) were used to take images of fluorescence.

Confocal microscopy was conducted using a Leica TCS SP5 spectral confocal microscope. Flow chambers were imaged using the DAPI, TRITC and FITC channels to image Hoechst, Cy3 and calcein respectively.

3.2.9. Image and Data Analysis

Images in TIFF format were analysed using FIJI (ImageJ). The number of cells was calculated from either brightfield images or from Hoechst staining using the automated cell counter. These results were expressed as the percentage of cells post-sonoporation compared to the number before sonoporation. Results from the sonicated focal area were normalised using images of the inlet region of the same flow

cell which was not subject to ultrasound. This provided a measure of the effect of ultrasound on cell viability while accounting for individual variations in viability between experiments.

Transfection efficiency was calculated by analysing the brightfield and Hoechst images to segment individual cells. The intensity of staining on the Cy3 channel was then used to identify cells that had been successfully transfected. Variation between the intensity of staining in individual experiments made it impractical to set a consistent threshold, so results were normalised to the transfection efficiency in the inlet which did not experience ultrasound. Graphs were created in Prism (GraphPad).

3.2.10. Splice Correction Assay

A solution of PNA 705 was prepared in OptiMEM media at a final PNA concentration of 10 μ M. HeLa pLuc 705 cells were seeded in a flow chamber at 1.5×10^5 cells/ml, sonoporated with PNA 705 as previously described and incubated for a further 24 hours. Cells were then washed twice in ice cold 1xPBS and lysed with passive lysis buffer (Promega). The total protein concentration in the cells was then determined using a BCA protein assay kit (Thermo, #23227) following the manufacturer's protocol. Luminescence was measured by mixing 50 μ l of cell lysate with 100 μ l Luciferase Assay Substrate (Promega) and scanning on a fluorescence plate reader. Expression of luciferase was counted as relative luminescence units per microgram of protein and shown as the fold increase in luminescence compared to negative control.

3.3. Results

3.3.1. Sonoporation set-up

Early work in our group utilised a relatively simple set-up to demonstrate that sonoporation could improve cellular uptake of PNAs *in vitro* (Figure 71 and Figure 72). The ultrasound apparatus consisted of an ultrasound therapeutic sonar (Electromedical Supplies Ltd.). Cells were plated at approximately 3.2×10^5 cells per well in clear 6 well plates and fixed approximately 1 cm from the transducer. The optimum settings for these experiments were found to be an acoustic pressure of 0.25 W/cm², 10% duty cycle and frequency of 1 MHz. However, there were a number of significant limitations to this set-up which produced conditions that would be very difficult to replicate *in vivo* as explained below. A new, optimised experimental set-up was designed that would eliminate these concerns and produce results that are more biologically relevant.

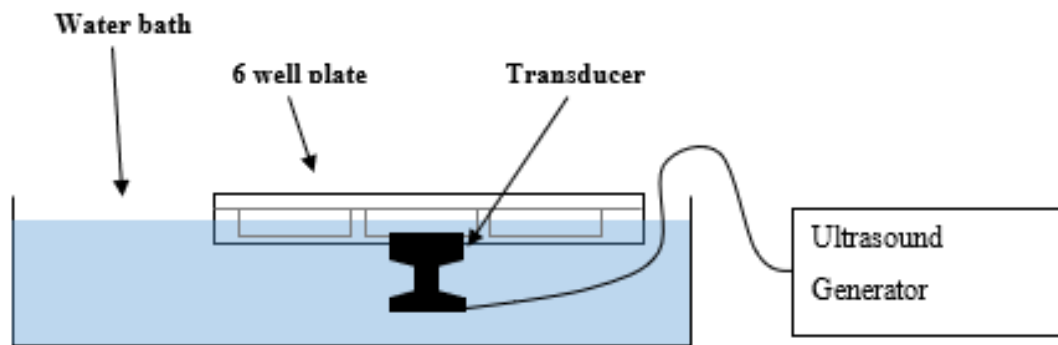


Figure 71: Initial ultrasound set-up. Cells were grown in a 6 well plate which was suspended at the surface of a water bath for the duration of the sonoporation process. Microbubbles were added directly to the media in the well plate. The transducer was suspended c.1 cm below the 6 well plate in the water bath. The ultrasound generator drives the transducer to produce ultrasound waves.

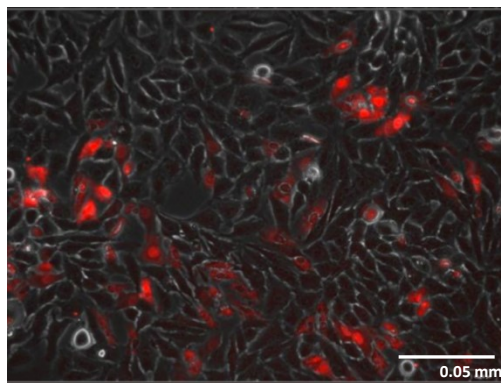


Figure 72: Uptake of PNA-Cy3 (red) by HeLa cells following sonoporation at 1 MHz, 0.5 W/cm² ultrasound for 1 minute, with a 20% duty cycle. The cells were exposed to ultrasound in the presence of 10 μ M PNA and with a 10% Optison microbubble concentration. n=1, scale bar represents 0.05 mm.

3.3.2. Eliminating standing waves

Transfection efficiency has been shown to be significantly affected by the formation of standing waves between the ultrasound transducer and air-media interfaces. Kinoshita and Hynynen found that eliminating standing waves in their experimental set-up resulted in a drop in the percentage of transfected cells from 30% to below 10% at an acoustic power of 4 W/cm².³⁷⁷ This is useful for *in vitro* sonoporation but would be difficult to replicate *in vivo*. Our original set-up contained an air-water interface directly above the cellular monolayer which would allow standing waves to form, so was changed by fully immersing the cell vessel in a temperature-controlled water tank. The 6 well plate was also replaced with a flow chamber, which doesn't contain any air.

3.3.3. Introducing Flow

Sonoporation is most easily applicable *in vivo* for the delivery of cargo to endothelial cells, for example in delivering antisense oligonucleotides to atherosclerotic plaques. As such, it is important to

demonstrate that microbubble-mediated cellular uptake of PNA is possible not just in static systems, but also in physiologically relevant flow systems. A further advantage is that microbubbles are circulated, which minimises inertial cavitation events which are more likely to occur when microbubbles are static. The introduction of flow *in vitro* presented a number of challenges including a much higher dead volume and high shear stress at physiological flow rates which can detach cells from the flow chamber. The experimental set-up was adjusted to minimise the dead volume and reduce the quantity of reagents used in each experiment.

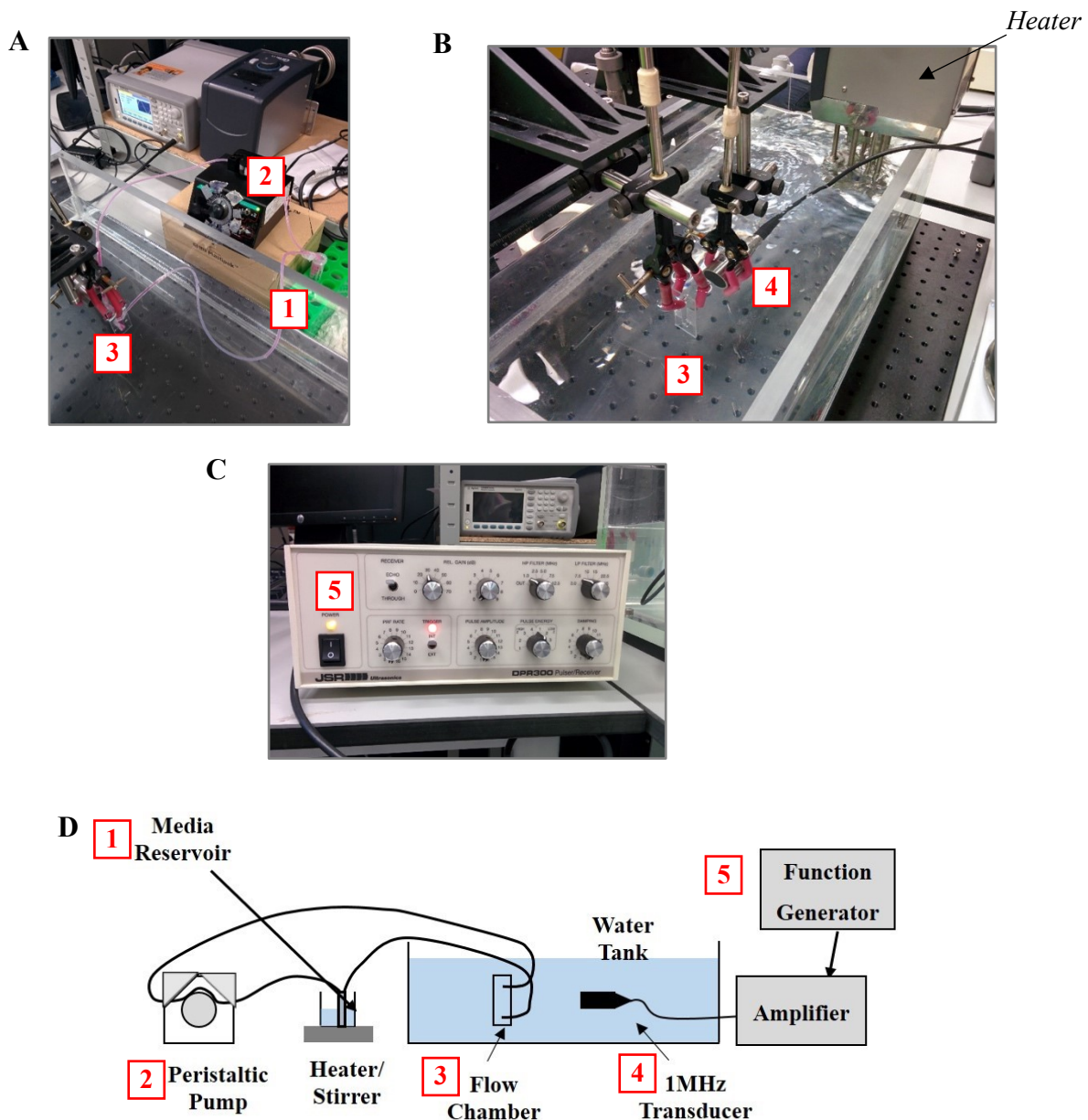


Figure 73: Optimised ultrasound setup. **A)** Flow set-up with flow chamber (3) suspended in water bath and connected via tubing to media reservoir (1) which contains media mixed with microbubbles and the peristaltic pump (2) which recreates the flow in blood vessels *in vivo*. **B)** Flow chamber (3) and transducer (4) suspended in heated (heater in view in top right of photo), degassed water bath at 37°C. **C)** View of the function generator (5) used. **D)** Full optimised set-up. The function generator creates the signal, which is amplified and then converted to ultrasound by the transducer in the water bath. The flow chamber is set a specific distance away and linked to a flow system controlled by the peristaltic pump. A water heater for the waterbath and heater stirrer for the reservoir are used to ensure that the media remains at 37°C.

The tubing and flow chamber size were also optimised to reduce flow related cell detachment, chambers with a height of 0.6 mm and 8 mm tubing (Ibidi) were chosen as they produced the low levels of cell detachment while minimising total reagent volume. A fully sealed system with an injector port (Ibidi) was tested, however it caused a significant increase in cell detachment with HeLa cells, so a media reservoir was created which was loosely sealed to prevent pressure build-up in the flow system (Figure 73).

3.3.4. Optimising assay endpoint analysis

(i) Cell viability

An important element to include in transfection assays is a measurement of cell viability. This is both to highlight the effect of the transfection technique on cell viability and to avoid false positives, which can result from uptake of cargo by apoptotic cells as a result of increased membrane permeability rather than the effect of the transfection technique being tested.³⁷⁸ It is also useful to differentiate between cell death induced by sonoporation (sonoporation-specific) and that induced by other factors (assay-induced). In this assay there are various steps in the protocol with the potential to induce cell death or detachment such as the addition of flow, washing steps and transfers between different lab spaces. Initial experiments in the absence of ultrasound showed that there was indeed a level of assay-induced cell death in the absence of ultrasound (Figure 74). Throughout subsequent experiments the level of assay-induced cell death broadly decreased as a result of increased technical proficiency, however there was some variation between days depending on availability of equipment which affected the length of some procedures and environmental conditions, such as temperature.

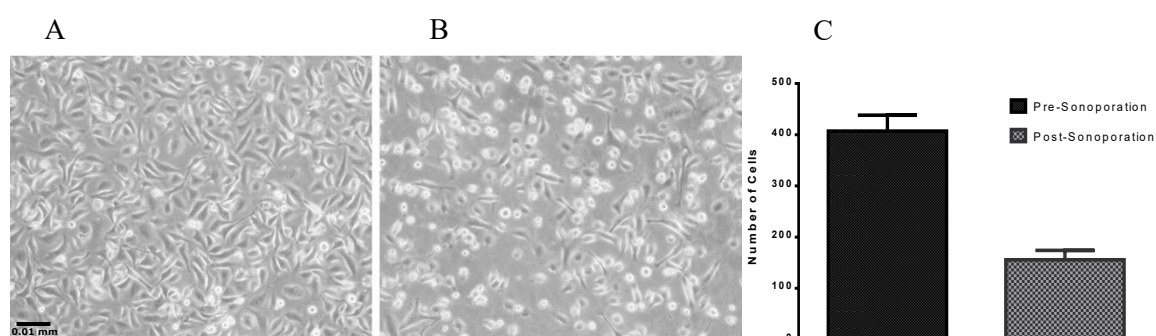


Figure 74: Cell detachment following full sonoporation protocol under sub-optimal flow conditions and in the absence of ultrasound. Brightfield image of HeLa cells **(A)** before sonoporation assay, in the absence of microbubbles and before washing steps and **(B)** after complete sonoporation protocol, including addition of 2×10^6 microbubbles/ml, flow rates of 24.5 ml/min, and washing steps, but without the addition of PNA-conjugates or application of ultrasound. **(C)** Quantification of cell numbers before and after sonoporation assay. Error bars represent standard deviation for n=3 repeats, scale bar represents 0.01 mm.

To quantify sonoporation-specific cell detachment, the number of cells before and after sonoporation were measured in the focal area of the flow chamber and in control regions near the inlet and outlet which were not exposed to ultrasound. The percentage of cell death/detachment in the control region was then used to normalise the results from the focal area, giving an estimate of the effect of sonoporation on viability.

(ii) Calcein staining

The live-cell indicator Calcein-AM was tested for inclusion in the assay in order to differentiate between live transfected cells and cells entering apoptosis. In live cells the acetoxymethyl ester (AM) group of calcein-AM is cleaved by intracellular esterases. Removal of the AM group restores calcein fluorescence but removes its ability to cross the cell membrane, resulting in the accumulation of fluorescent molecules in the cell.³⁷⁹ Analysis of initial images showed no overlap between cells with increased transfection and calcein staining (Figure 75). However, the transfected cells exhibited healthy morphologies and no detachment several hours after sonoporation. It is therefore possible that sonoporation acts to prevent calcein-AM cleavage or accumulation through a mechanism other than induction of apoptosis. In particular, extended pore formation could prevent accumulation of calcein. To test this theory, the duration of pore formation was investigated.

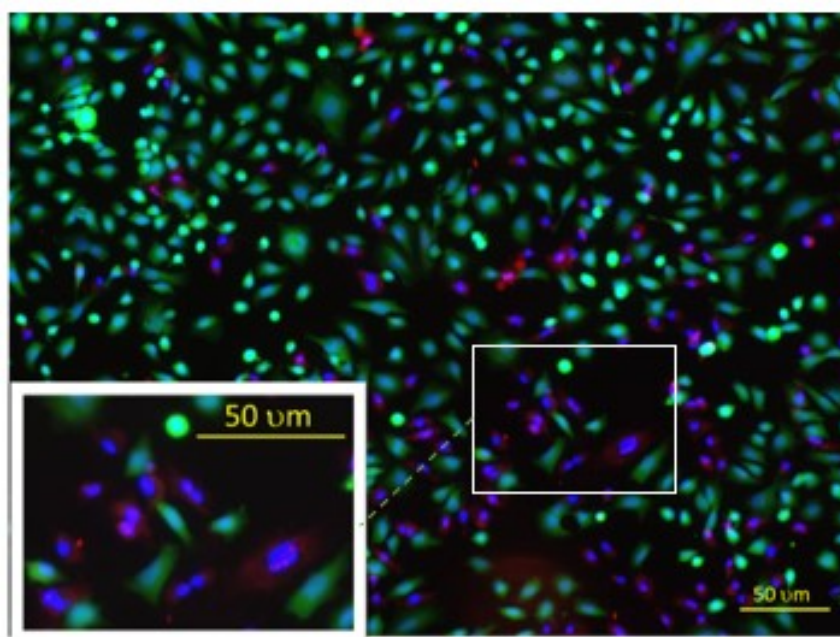


Figure 75: Transfection of HeLa cells following sonoporation at 1000 kPa, 1000 cycles for 400 seconds. Cells underwent sonoporation in the presence of 2×10^6 microbubbles/ml and $10 \mu\text{M}$, and stained with Hoechst 33372 (blue), Calcein (green) and PNA-Cy3 (red). Scale bar represents 0.05 mm.

(iii) Duration of pore formation

The incubation of cells in the transfection medium (containing microbubbles and PNA-Cy3) for 30 minutes after sonoporation resulted in a significant increase ($p < 0.01$) in transfection efficiency from $3.28 \pm 1.91\%$ to $20.94 \pm 1.20\%$, suggesting that membrane poration continues for at least 30 minutes post-sonoporation (Figure 76). Indeed, other studies have indicated that ultrasound-induced membrane permeability continues for 1-3 hours after termination of sonoporation.³⁸⁰ This window was consistent across a range of different cell types and ultrasound acoustic pressures. This explains our results showing that there was increased transfection when the PNA-Cy3 was incubated with the cells for 30 minutes after sonoporation and suggest that stronger staining could be achieved through longer incubation time. It could also explain the lack of calcein staining in transfected cells.

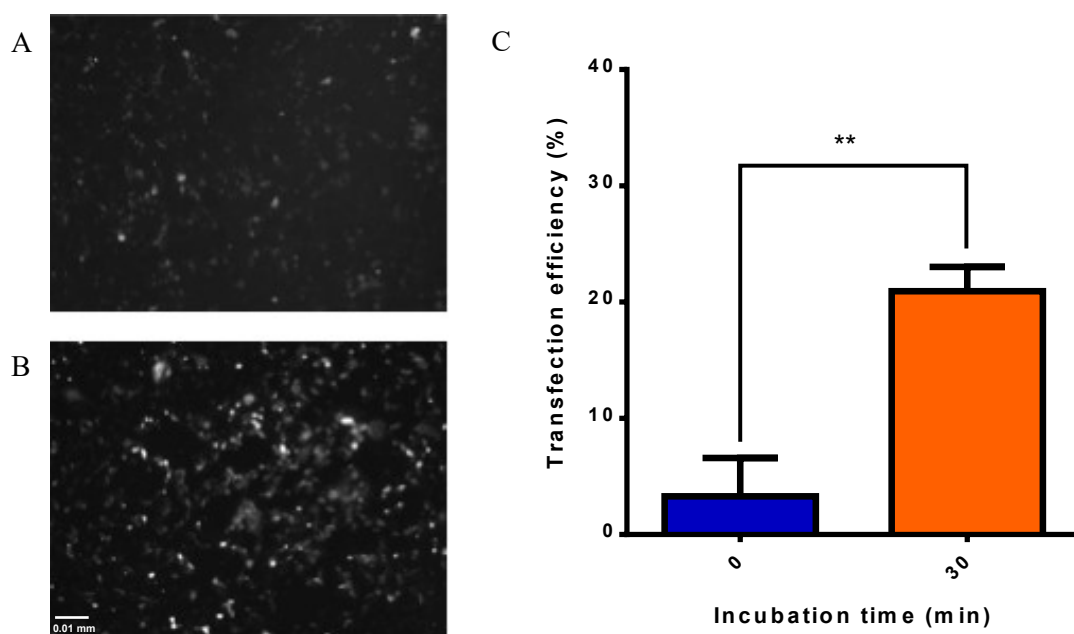


Figure 76: Effect of post-sonoporation incubation with PNA-Cy3 on transfection efficiency following exposure to ultrasound at 1000 kPa and 500 cycles for 400 seconds in the presence of 2×10^6 microbubbles/ml. A) Image showing PNA-Cy3 staining of HeLa cells immediately after microbubble-mediated transfection. **B)** PNA-Cy3 staining of HeLa cells after 30 minutes of incubation. **C)** Quantification of transfection efficiencies with and without post-sonoporation incubation. Error bars represent standard deviation for $n=3$ repeats, significance tested using a student t-test (** represents $p < 0.01$), scale bar represents 0.01 mm.

(iv) Image analysis

The intensity of the staining in transfected cells varied across individual experiments and repeats. For example, see Figure 77 which shows the mean intensity in pixels of the transfected cells tested at different cycle numbers. There was a step change in intensity at higher cycle numbers with higher average intensity at 700, 1000, and 1300 cycles. This created difficulties with analysis as a consistent threshold could not be applied when segmenting transfected cells. Without consistent thresholds the

number of false positives detected by the sorting algorithm is variable between experiments. To adjust for this effect, transfection efficiencies in the focal region were normalised using the inlet control regions.

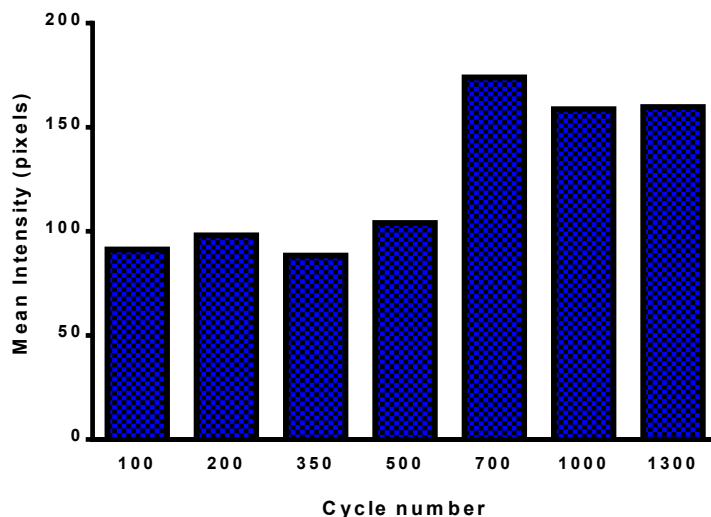


Figure 77: Mean intensity of PNA-Cy3 staining in HeLa cells following sonoporation. Mean intensity calculated as the average pixels per cell for transfected cells. Experiments conducted with a PNP of 1000 kPa for 400 seconds in the presence of 2×10^6 microbubbles/ml with variable cycle numbers (x-axis).

3.3.5. Optimisation of sonoporation parameters

Table 8: Ultrasound parameters used in subsequent experiments.

Peak Negative	500 kPa	700 kPa	1000 kPa	1300 kPa
Pressure (PNP)	(144 mVpp)	(204 mVpp)	(293 mVpp)	(383 mVpp)
Number of cycles	100, 200, 350, 500, 700, 1000, 1300			
Pulse Centre	1 MHz			
Frequency (PCF)	(pulse duration = No. of cycles/PCF)			
Pulse repetition	1 Hz			
frequency (PRF)	(burst period=1/PRF=1 s; duty cycle=pulse duration/burst period)			
Number of pulses	400, 800, 1600			
	(insonation time = No. of pulses x burst period = 400 s, 800 s, 1600 s)			

To identify the optimum ultrasound parameters for sonoporation, we tested the effect of Peak Negative Pressure (PNP), number of cycles and number of pulses (see Table 8) on both transfection efficiency and cell viability.

PNP is controlled by the amplitude of the ultrasound wave (measured in millivolts peak to peak, or mVpp) experienced by the microbubbles as an increase in pressure (kPa) increases the force of the cavitation effect. The number of cycles refers to the number of individual waves produced within each separate pulse and together with the Pulse Centre Frequency (PCF), which is typically determined by the frequency of the transducer used, it determines pulse duration. The final adjustable parameter is the burst period, which is the sum of the pulse duration and the listening period (so called as it is used for detecting the reflected signal during ultrasound imaging) (Figure 78). This parameter subsequently determines the Pulse Repetition Frequency (PRF) and along with the number of pulses determines the total insonation time. When the burst period is set to 1 second, the insonation time in seconds is the same as the number of pulses.

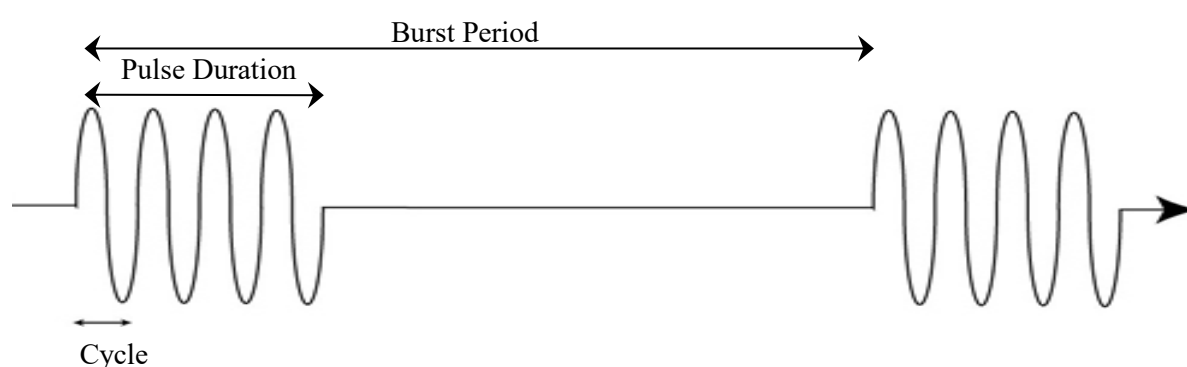


Figure 78: Characteristic parameters of ultrasound waves. Pulse duration represents the on time for the ultrasound. Burst period is the total time from the start of one burst the start of the next. Cycles represent a full wave, i.e. the energy moving through both positive and negative peaks of a sine wave and returning to the same point as it started.

(i) Effect of peak negative pressure on sonoporation efficiency and cell viability

Increasing PNP from 500 kPa to 1300 kPa in the presence of microbubbles resulted in an increase in transfection efficiency (measured by uptake of PNA-Cy3) from 1% to 28% and a decrease in sonoporation-specific cell viability from 100% to 65% (Figure 79). Results were normalised using the inlet region which exhibited an average apparent transfection efficiency of $9.23\% \pm 3.64$. The trends were expected, as PNP directly affects the force of cavitation, which will both increase pore formation and increase the likelihood of inertial cavitation, which is more likely to damage the cells. Application of ultrasound at a PNP of 1000 kPa resulted in around 16% transfection efficiency with 85% cell viability, which was within our acceptable level of cell viability ($>80\%$). It was therefore chosen for use in future experiments where other settings were modified to try and increase the transfection efficiency without resulting in any further decrease in viability. This experiment was repeated multiple times, however issues with cell detachment rendered other repeats un-suitable for use, therefore only one repeat is shown here.

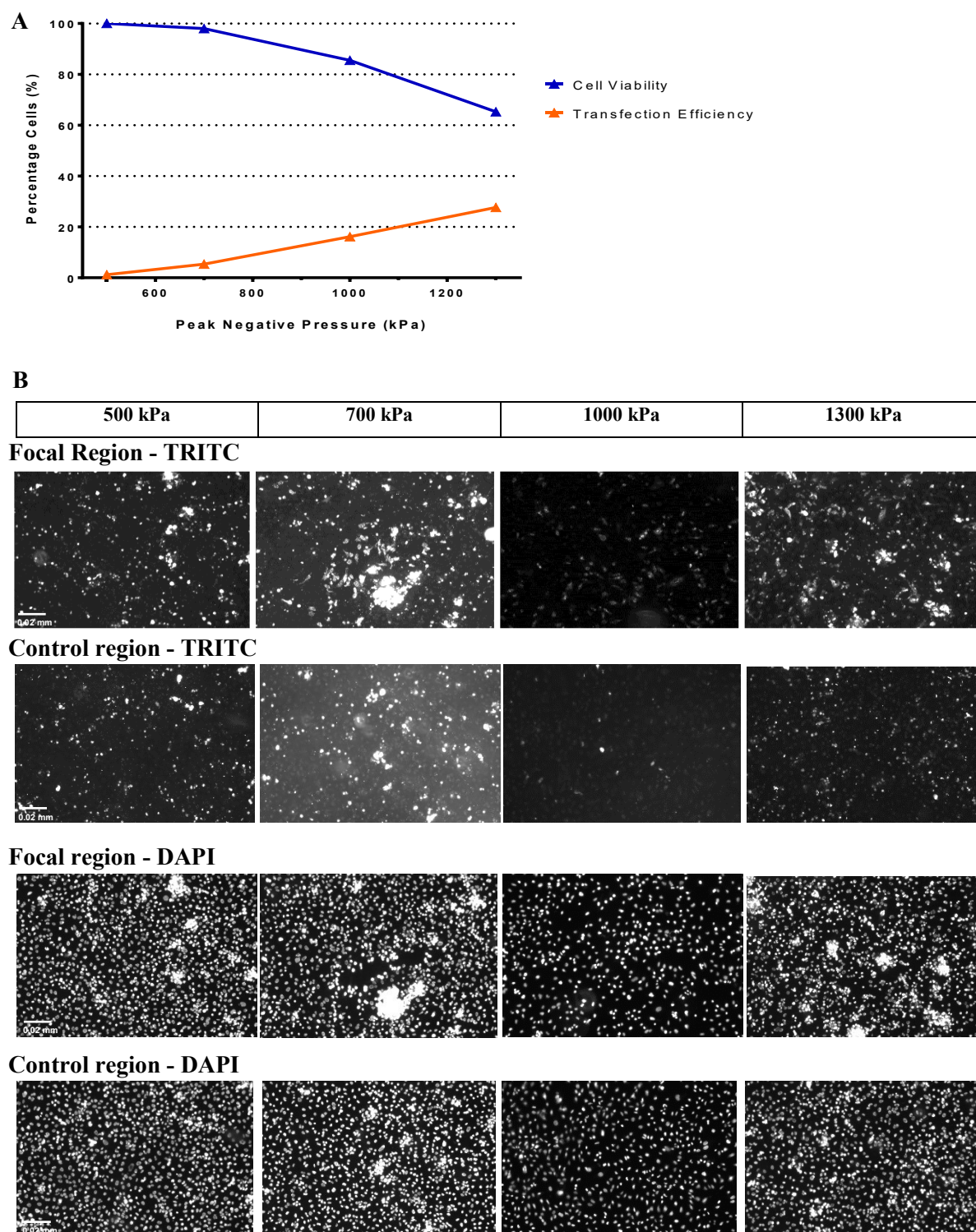


Figure 79: Effect of peak negative pressure (PNP) on microbubble-mediated PNA-Cy3 uptake and viability of HeLa cells following sonoporation. Ultrasound was applied to the cells with a Peak negative pressure of 500, 700, 1000 or 1300 kPa, a cycle number of 200 and insonation times of 400 seconds, in the presence of 2×10^6 microbubbles/ml. **A)** Analysis of images from the focal area following sonoporation, normalised using readings from the control area. The total percentage of transfected cells, as calculated by the presence of staining by PNA-Cy3 within the cells boundaries, increased with peak negative pressure, however, cell viability decreased. Results shown as the percentage of cells with PNA-Cy3 staining (transfection efficiency) or the percentage of cells surviving following sonoporation (cell viability). **B)** Images of PNA-Cy3 staining of cells in the focal area (in the centre of the flow chamber and therefore exposed to ultrasound) and the control region (near the inlet of the flow chamber and not exposed to ultrasound) following sonoporation, taken using the TRITC channel (showing PNA-Cy3 staining) or the DAPI channel (Showing Hoechst staining). Experiments repeated once ($n=1$), scale bar represents 0.02 mm.

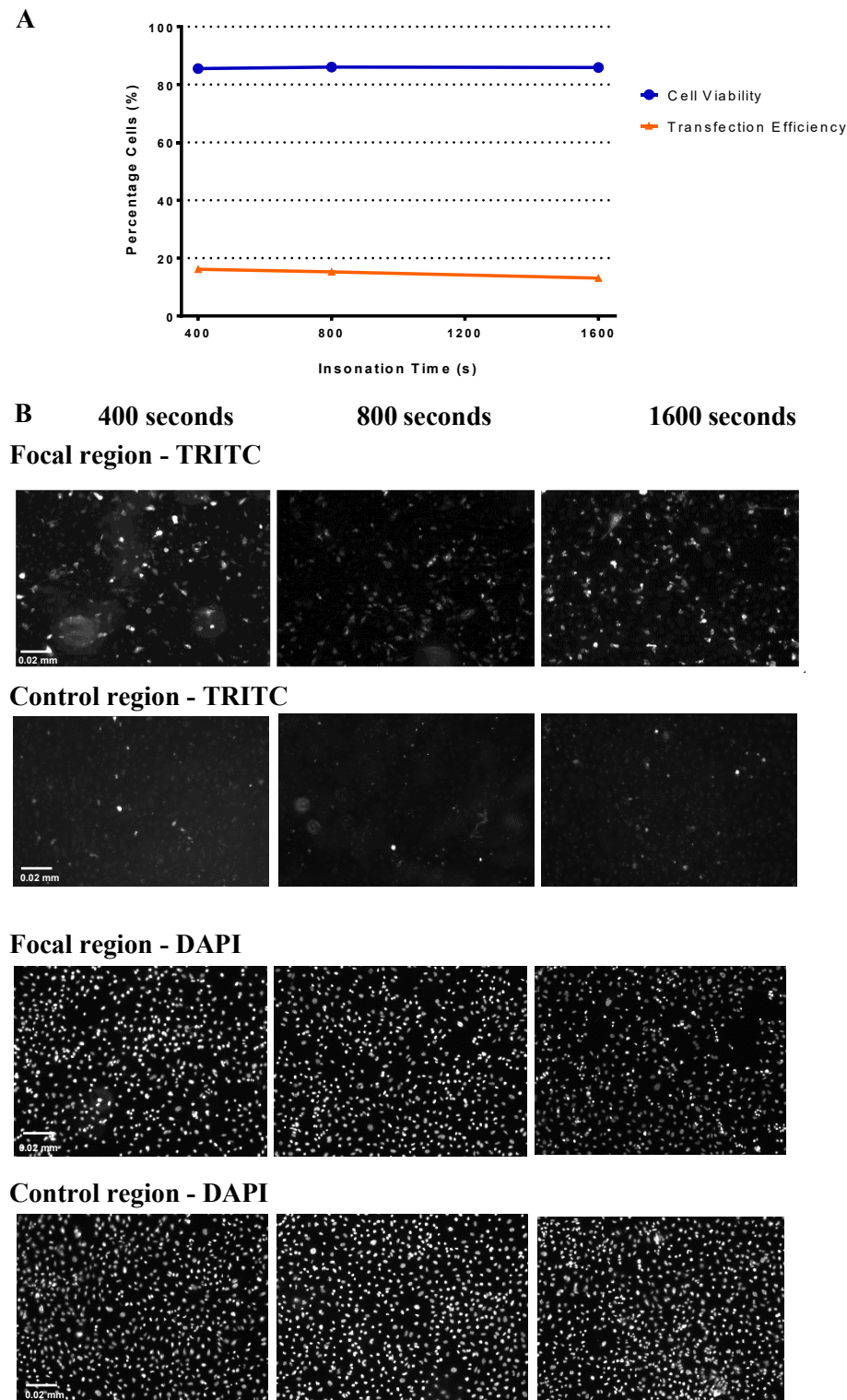


Figure 80: Effect of insosation time on microbubble-mediated PNA-Cy3 uptake and viability of HeLa cells following sonoporation. Ultrasound was applied to the cells with a Peak negative pressure of 1000 kPa and a cycle number of 200 in the presence of 2×10^6 microbubbles/ml for 400, 800 or 1600 seconds. **A)** Analysis of images from the focal area following sonoporation, normalised using readings from the control area. The total percentage of transfected cells, as calculated by the presence of staining by PNA-Cy3 within the cells boundaries, as well as the viability of the cells remained constant between experiments over different times. Results shown as the percentage of cells with PNA-Cy3 staining (transfection efficiency) or the percentage of cells surviving following sonoporation (cell viability). **B)** Images of PNA-Cy3 staining of cells in the focal area (in the centre of the flow chamber and therefore exposed to ultrasound) and the control region (near the inlet of the flow chamber and not exposed to ultrasound) following sonoporation, taken using the TRITC channel (PNA-Cy3 staining) or the DAPI channel (Hoechst staining). Experiments repeated once ($n=1$), scale bar represents 0.02 mm.

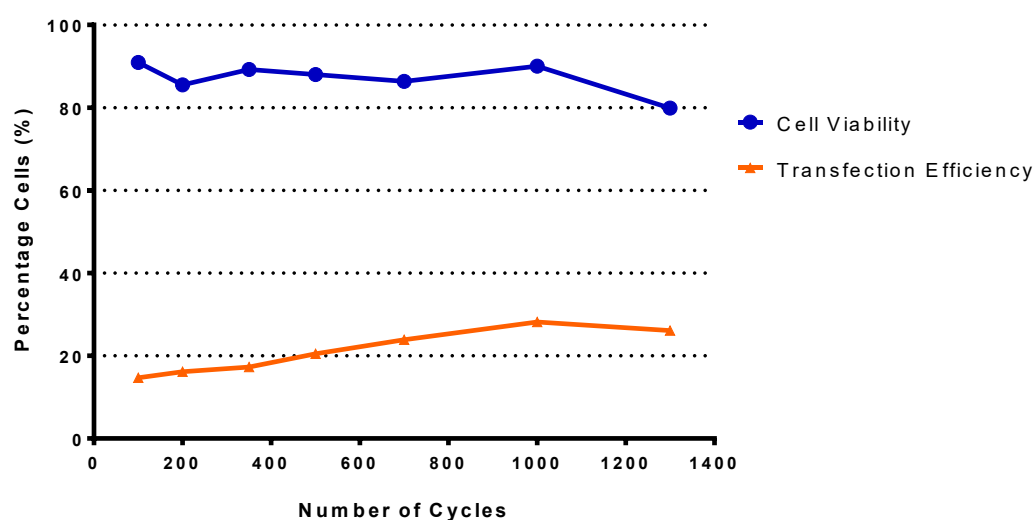
(ii) Effect of insonation time on transfection efficiency and viability

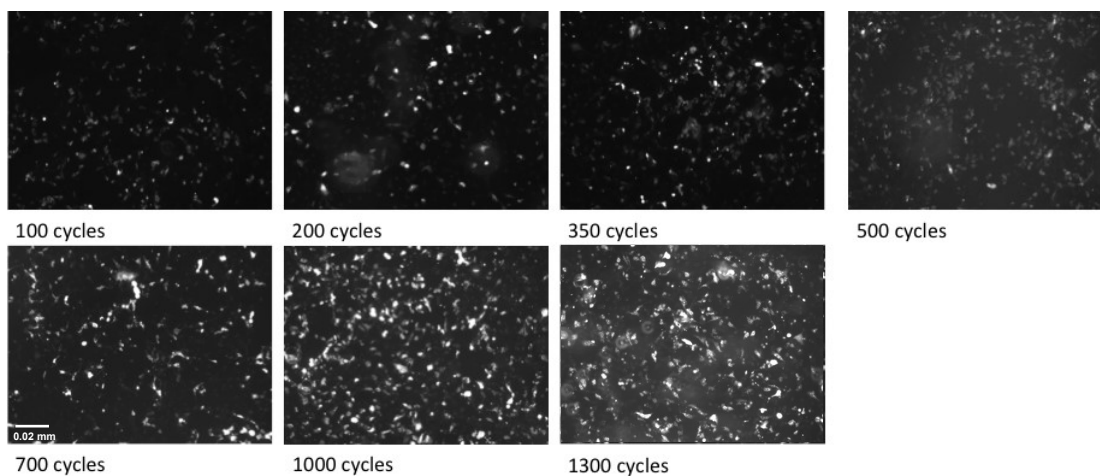
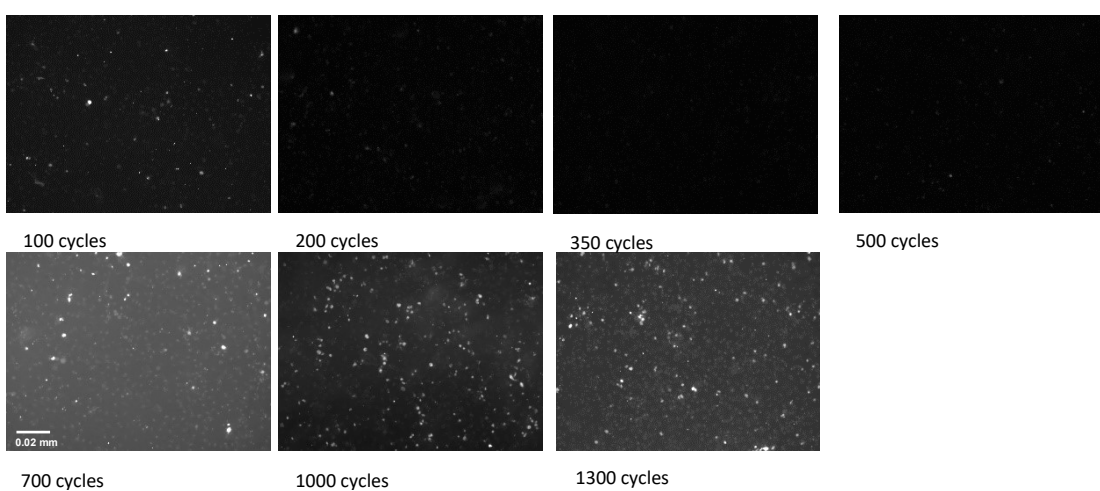
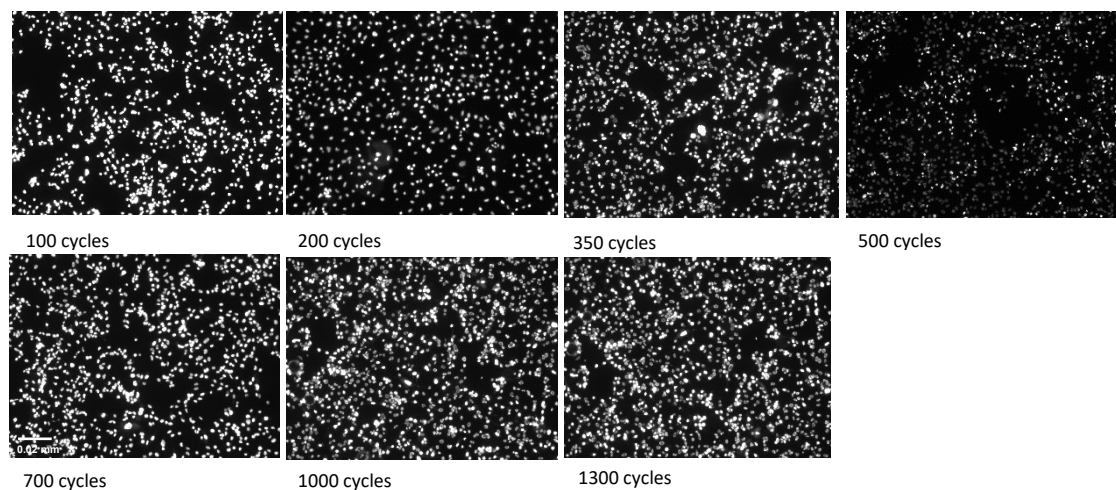
Increasing the insonation time did not result in any change of either transfection efficiency or cell viability. The results were normalised using readings from the inlet region, the average of which were a transfection efficiency of $2.80\% \pm 1.08$ (Figure 80). The experiments were carried out in the presence of microbubbles and transfection efficiency measured through PNA-Cy3 uptake. As there was no benefit to longer insonation times, the minimum time tested (400 seconds) was chosen for use in future experiments. The fact that there is no reduced cell viability with longer insonation times suggests that the negative effects of ultrasound and microbubble exposure on cells occurs in the first few minutes, with no significant increased damage caused after that point. This set of experiments produced only one set of consistent results to compare, this was due to issues with detachment of cells in the un-used results.

(iii) Effect of cycle number on sonoporation efficiency and cell viability

The effect of cycle number on sonoporation-induced transfection of PNA-Cy3 in the presence of microbubbles was also tested. There appears to be a trend of increased transfection efficiency with increased cycle number. There is no clear related decrease in cell viability, which remains at 85-90%, until 1300 cycles where viability drops to 80%. Results were normalised using inlet results, the average of which were a transfection efficiency of $3.99\% \pm 2.09$. 1000 cycles appeared to be the optimum number to generate an increase in transfection without a concurrent decrease in cell viability (Figure 81). This set of experiments produced only one set of consistent results to compare, this was due to issues with detachment of cells in the un-used results.

A



B**Focal region - TRITC****Control region - TRITC****Focal region - DAPI**

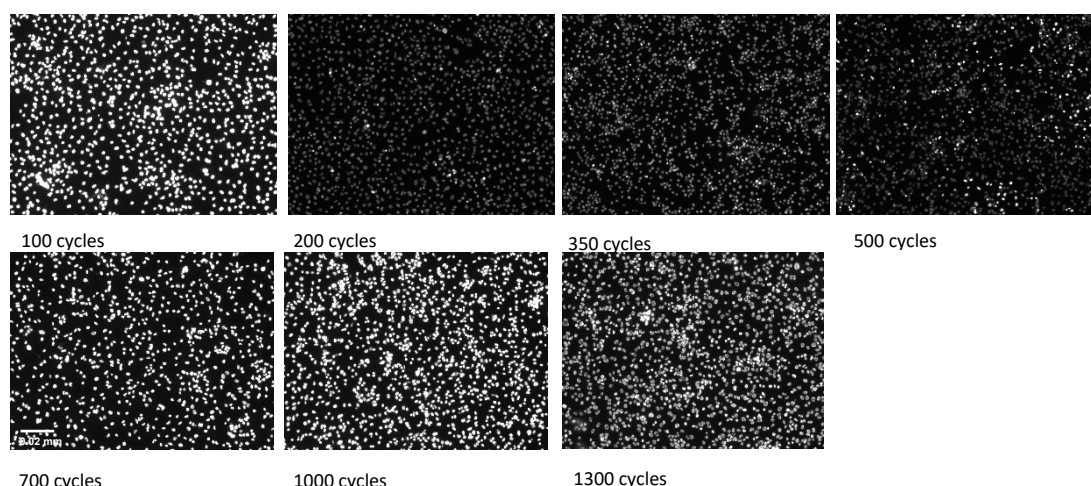
Control region – DAPI

Figure 81: Effect of cycle number on microbubble-mediated PNA-Cy3 uptake and viability of HeLa cells following sonoporation. Ultrasound was applied to the cells with a Peak negative pressure of 1000 kPa and a cycle number of 100, 200, 350, 500, 700, 1000 or 1300, in the presence of 2×10^6 microbubbles/ml for 400 seconds. **A)** Analysis of images from the focal area following sonoporation, normalised using readings from the control area. The total percentage of transfected cells, as calculated by the presence of staining by PNA-Cy3 within the cells boundaries, increased as the cycle number increased. However, at the highest cycle number the viability also began to decrease. Results shown as the percentage of cells with PNA-Cy3 staining (transfection efficiency) or the percentage of cells surviving following sonoporation (cell viability). **B)** Images of PNA-Cy3 staining of cells in the focal area (in the centre of the flow chamber and therefore exposed to ultrasound) and the control region (near the inlet of the flow chamber and not exposed to ultrasound) following sonoporation, taken using the TRITC channel (PNA-Cy3 staining) or DAPI channel (Hoechst staining). Experiments repeated once ($n=1$), scale bar represents 0.02 mm.

3.3.6. Comparison of optimised sonoporation parameters and CPP

The conjugation of CPP is currently the most effective method to improve the cellular uptake of PNA. The microbubble-mediated cellular uptake of PNA-Cy3 with our optimised set-up and parameters (PNP: 1000 kPa, Cycles: 200, Time: 400s) was compared to cellular uptake of (R-Ahx-R)₄-PNA-Cy3. Images were taken of HeLa cells transfected with PNA-Cy3 conjugates using each method and used to calculate the relative transfection efficiencies. (R-Ahx-R)₄ conjugation produced significantly higher transfection efficiencies ($p < 0.005$) than sonoporation (Figure 82).

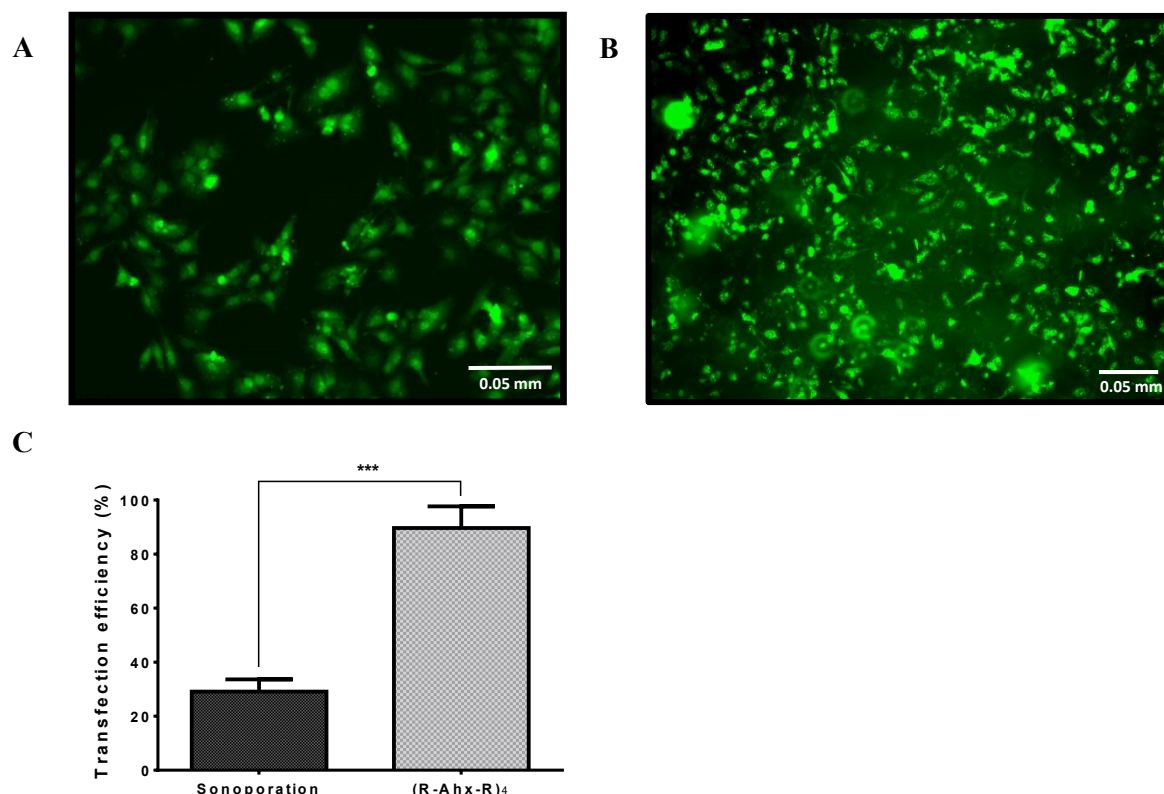


Figure 82: Comparison of the effectiveness of conjugation to a CPP and sonoporation for transfection of PNA cargo. A) HeLa cells following transfection with (R-Ahx-R)₄-PNA-Cy3 (green). B) HeLa cells following transfection with PNA-Cy3 (green) using optimised sonoporation parameters, PNP: 1000 kPa, Cycles: 200, Time: 400s in the presence of 2×10^6 microbubbles/ml. C) Analysis of the transfection efficiency of sonoporation and CPP conjugation. Error bars represent standard deviation for n=2 repeats, significance tested using a student's t-test (***= $p < 0.005$), scale bar represents 0.05 mm.

3.3.7. Bioavailability of transfected PNA

A common problem with most common transfection techniques, including cell penetrating peptides, is that of bioavailability. Transfected nucleic acids can be sequestered within endosomes, preventing them from binding to target sequences. The most effective method to test bioavailability is the splice correction assay first reported by the Kole group in 1998.³⁸¹ The assay utilises cells transfected with the pLuc705 plasmid, which contains the *firefly* luciferase reporter gene interrupted by a mutated β -globin intron 2. The mutation in nucleotide 705 of the intron creates an additional 5' splice site. The presence of this site activates an upstream cryptic-splice site and prevents full excision of the intron. The remaining fragment contains a stop codon and therefore prevents translation of a functioning luciferase protein. Successful transfection and binding of a complementary 705 oligonucleotide corrects the RNA splicing and restores luciferase expression. Therefore, measuring luciferase activity in HeLa pLuc705 cells following transfection of PNA705 provides an effective method to test the efficacy of transfection techniques in term of delivering bioavailable cargo. The technique has previously been used with CPPs, for example (R-x-R)₄-PMO705 conjugates were shown to allow efficient splicing correction in the absence of endosomolytic agents.³³¹

Optimised sonoporation parameters were used to transfect 10 μ M PNA705 into HeLa pLuc705 cells to test the bioavailability of sonoporated PNA. There was no significant increase seen in the preliminary experiments below (Figure 83), however this is likely to be due to experimental limitations as discussed in section 3.4.3.

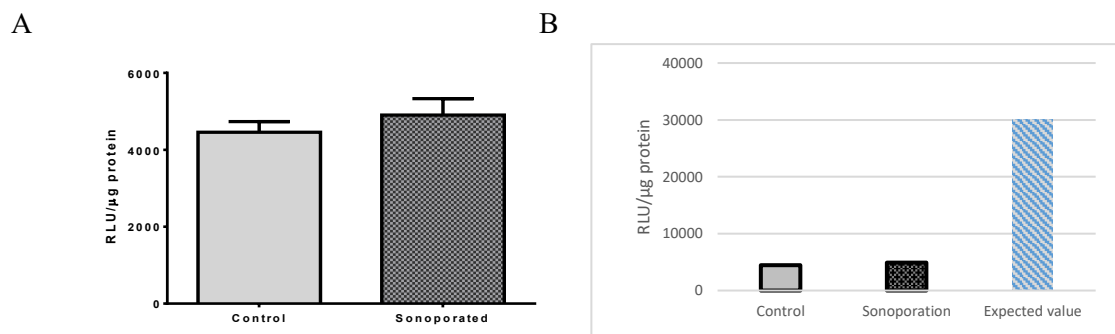


Figure 83: Relative luciferase activity in HeLa pLuc 705 cells following incubation with PNA705 with and without sonoporation. (A) Actual results from splice correction assay showing the values obtained, error bars represent standard deviation for n=2 repeats. **(B)** Comparison of actual results shown in A with expected results, which are indicative and based on the results in previous experiments, for example Abes³³¹ where conjugation of (R-x-R)₄ produced a reading of over 30,000 at 1 μ M.

3.4. Discussion

3.4.1. Summary

This study demonstrates the potential of sonoporation to improve the cellular uptake of PNA *in vitro*. The optimal settings identified in this study were: **PNP:** 1000 kPa, **Number of Cycles:** 1000, **Burst Period:** 1 second, **Insonation time:** 400 seconds. These produced a maximum transfection efficiency of 28% with cell viability (measured by cell number) of 90%. Ideally, before aiming to improve transfection efficiency further the complexity of the assay system needs to be reduced. A number of technical difficulties, such as the introduction of bubbles into the flow system, transition of the flow system between laboratories etc. require high dexterity and experience from researchers. In addition, a cell viability assay needs to be validated and adapted to fit with the protocol.

3.4.2. Comparison with previous studies

The trends we identified from adjusting the ultrasound parameters broadly match those reported in previous studies for DNA or RNA delivery, although they do also highlight how different experimental conditions and cell types can affect results, making direct comparisons of results difficult. Karshafian *et al.* also observed no significant change in transfection of 70 kDa FITC-Dextran or viability with increased insonation time over 60 seconds and saw similar increases in transfection efficiency and

decreases in viability with increased PNP up to 570 kPa.³⁸² In contrast, experiments using sonoporation to transfect luciferase plasmid DNA in vivo in mice showed maximum luciferase activity at 200 kPa.³⁸³ In the same experiment peak luciferase activity was seen with 2 minutes insonation time and co-injection of 5×10^5 microbubbles with plasmid DNA. Differences in cargo, cell type and experimental set-up could all contribute towards the variety in responses to ultrasound parameters.

In an experiment with a very similar set-up, although with HUVEC cells and the dye propidium iodide (PI) as cargo, Shamout *et al.* identified very different optimal parameters.³⁵⁷ As discussed above, this could in part be due to different uptake mechanisms, as the PNA sequence tested is a much larger molecule than PI. Additionally, in this study a 5 MHz transducer was used rather than a 1 MHz transducer. It is therefore possible that a higher frequency transducer enables transfection at lower PNP.

3.4.3. Comparison with (R-Ahx-R)₄

(R-Ahx-R)₄ conjugation is broadly considered the most effective method to improve cellular uptake of PNA. Our preliminary results suggest that it produces both higher transfection efficiencies and better bioavailability than sonoporation. However, there were major potential limitations to these preliminary experiments. The main limitation to the splice correction assay is that only a small proportion of the cells in the flow cells are exposed to ultrasound during sonoporation, as a result of the small focal area of the transducer. As luciferase activity is measure relative to total protein, the presence of a significant number of un-transfected cells will obscure the activity in transfected cells. Introducing fluorescence assisted cell sorting (FACS) to select transfected cells before testing for luciferase activity could allow for more accurate quantification.

3.4.4. Future work

(i) Intensity of transfected cells

An interesting result identified in this chapter is the change in intensity of Cy3 staining between experimental conditions. As this was an exploratory study, the percentage of transfected cells (transfection efficiency) was chosen as an assay endpoint. However, higher intensity staining in transfected cells could represent a different uptake mechanism, or improved uptake, and as such would be a useful additional endpoint for the assay. In particular, combining analysis of the intensity of staining with a functional assay (i.e. the pLuc705 splice correction assay) could reveal settings that improve the efficacy of cargo in transfected cells.

(ii) Uptake mechanism

Improving our understanding of the mechanism for PNA uptake would also be beneficial in future studies with this assay. Imaging of the microbubbles during sonoporation along with analysis of the reflected ultrasound wave can be used to identify the type of cavitation microbubbles undergo. Combined with confocal microscopy, this could help identify formation of pores or endosomes.³⁸⁴ In addition, co-localisation of endosomal antibody stains such as EEA-1 (BD Bioscience, San Jose, CA) or CD63 (Developmental Studies Hybridoma Bank, Iowa City, IA) and fluorescent cargo following sonoporation could be used to track endosomes.³⁶⁶

(iii) Cell viability

There were no visible phenotypic changes in transfected cells within the assay timeframe, however, following induction of apoptosis through UV-C exposure, changes in HeLa cell morphology were not observed for over 8 hours,³⁸⁵ so inclusion of a well characterised viability assay is vital. Other studies have observed similar calcein staining effects, i.e. no staining in transfected cells, and have been able to modulate ultrasound parameters to produce a sub-set of transfected cells that also show calcein staining.³⁵⁷ If pore formation affects calcein staining as we have suggested, then this subset of cells could represent an alternative uptake mechanism, i.e. endocytosis.

Chapter 4: Identification and characterisation of RNA G4 as potential new therapeutic targets

4.1. Introduction

Thus far, only a relatively modest number of RNA G4 have been demonstrated to modulate translation *in vivo*, most of these being located in the 5'- and 3'-UTR.^{115, 280} In this chapter, two new PQS are identified in therapeutically relevant genes and shown to form G4 *in vitro*. Targeting of the *Aurora A* G4 by small molecules is also explored to assess the potential for therapeutic modulation of G4 stability.

4.1.1. Identification of novel PQS

Historically the search for PQS have been run on a genome wide scale utilising programs such as Quadparser, however similar studies of RNA quadruplexes have been more limited. In this chapter a different approach was taken, focussing first on genes of clinical interest and then undertaking further close analysis of the functional regions of the gene to identify PQS of potential clinical interest that could be missed in less targeted transcriptome-wide searches. First a literature review was undertaken to identify genes of therapeutic interest that had not yet been investigated in previous RNA G4 studies. Following the target identification stage, an initial analysis was undertaken of mRNA sequences in the NCBI database, focusing on two genes of clinical interest; known oncogenes *Aurora A* and *Collagen type I alpha 2 (Col1A2)*. The clinical relevance of these two oncogenes and their potential as viable therapeutic targets will be discussed in more detail in sections 4.1.3, 4.1.4 and 4.1.5 below. A short program was created in Python to identify sequences containing two or more contiguous guanines and sets of four such clusters within a defined parameter (less than 30 nucleotides long) were flagged as potential PQS. All search results were manually processed and those in regions of interest (5'UTR, CDS) flagged. Two G4 of particular interest were noted and chosen for further study (reasons discussed below), the sequence of these are shown below, with the guanine tracts of the PQS highlighted in bold and (where applicable) the start codon highlighted in red:

Col1A2: 5'- AAAGGGCGGGGAGGGCGGGAGG -3'

Aurora A: 5'- UUCGGAGGCCGAGGCAUC**AUG**GACC -3'

The *Aurora A* PQS is located at the end of the 5'UTR and overlaps the translation start codon. Therefore, its stabilisation could have a significant impact on translation (e.g. masking the start codon could potentially alter translation initiation), making it an interesting potential therapeutic target. It was initially hypothesised that, as with all RNA G4 reported to date, this PQS would form a parallel-stranded

structure with compact propeller-type loop regions (Figure 84B). The fact that it contained only two G-quartets also made us predict that the G4 would be relatively unstable and unlikely to form *in vivo* unless stabilised by a protein or small molecule.

The *CollA2* quadruplex contains longer three-nucleotide G-tracts and should therefore form a more stable G4 structure than the *Aurora A* G4 (Figure 84A).³⁸⁶ The PQS is located 29 nucleotides downstream from the 5'-cap and may therefore be optimally positioned to regulate translation. Indeed, a previous study demonstrated that inhibitory effect of the *NRAS* G4 was removed when the G4 was moved over 50 nt from the 5'-cap.¹²⁰

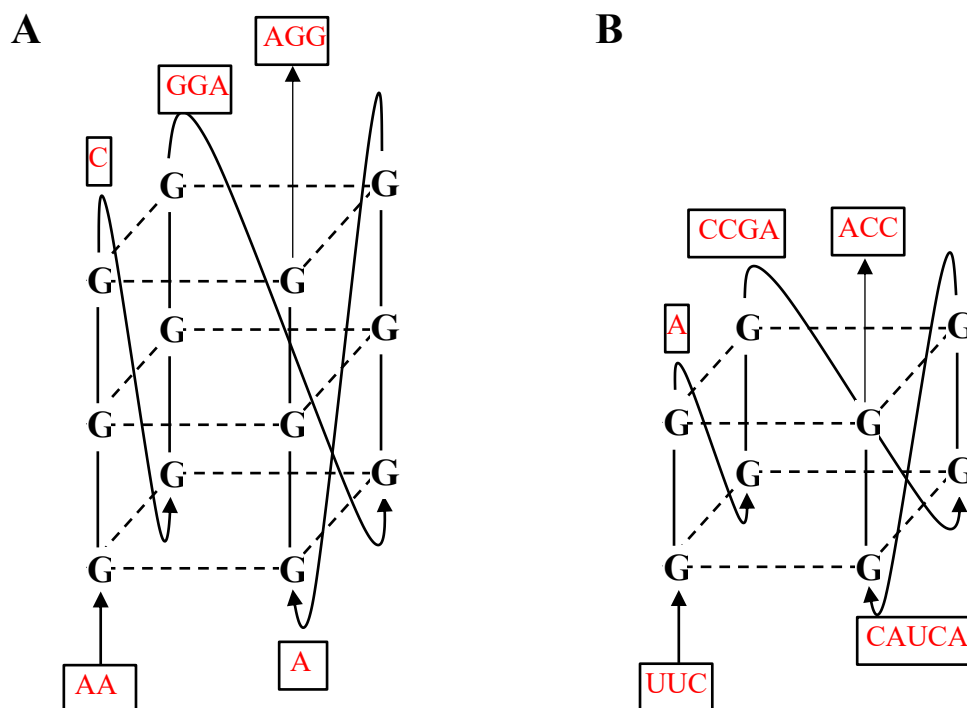


Figure 84: Predicted structures for the *CollA2* (A) and *Aurora A* (B) G4. Nucleotides in loop regions are shown in red lettering.

4.1.2. Evidence for formation of two quartet G4

Many original bioinformatics studies were run using the consensus sequence $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$,^{48, 149} which does not include putative G4 with only two guanine quartets, so there are relatively few well-characterised examples when compared to three-quartet G4. However, in 2015 Qin *et al.* reported a bioinformatical and biophysical study investigating the prevalence and stability of two quartet DNA G4 (G₂4). The study identified c.8.5m putative G₂4 sequences, at a density of 2.9 per kb, compared to 0.1 per kb for quadruplexes with three or more quartets.³⁸⁶ In addition, the study demonstrated using NMR and CD spectroscopy that aptamers and selected native putative G₂ sequences formed stable G4 structures *in vitro*. The stability of these structures was highly dependent on loop

length, with sequences containing loops of 1 or 3 nucleotides typically forming structures with higher thermal stability.

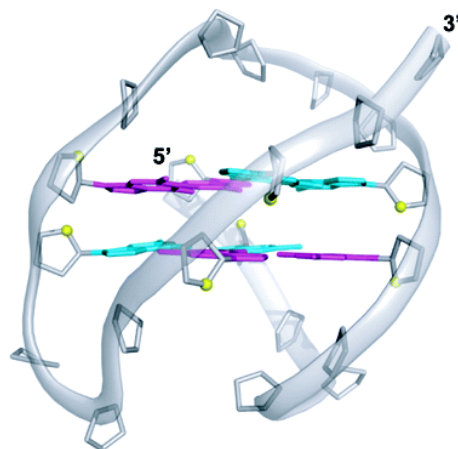


Figure 85: Structure of human telomeric DNA form 3 [GGGTTAGGGTTATGGTTAGGGT] in K^+ solution. In this ribbon view of a representative structure, *anti* and *syn* guanines are coloured cyan and magenta, respectively; backbone, grey; O4' atoms, yellow. Figure reproduced from Lim *et al.*⁵¹

Lim *et al.* reported the NMR structure of a stable two-quartet basket-type G4 formed by human telomeric DNA (Figure 85).⁵¹ Interestingly this structure was shown to be more stable than the three-G-tetrad intramolecular G4 previously reported for human telomeric DNA in K^+ , with a melting temperature of up to 57°C depending on conformational factors and solvent.

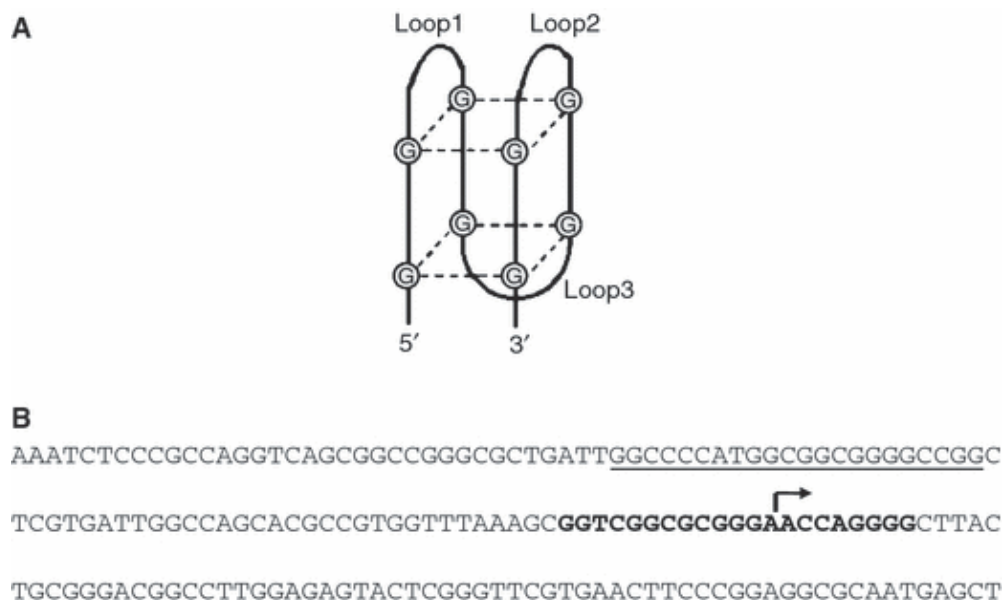


Figure 86: Two quartet putative quadruplex forming sequences in the TK1 promoter region. (A) example structure of a 2-quartet quadruplex. (B) TFD promoter sequence with PQS highlighted, TKQ1 shown in bold and TKQ2 underlined. In addition, the TSS is shown with an arrow. Figure reproduced from Basundra *et al.*³⁸⁷

In 2010, Basundra *et al.* investigated two putative G₂4 sequences in the promoter region of human thymidine kinase 1 (*TK1*) referred to as TKQ1 and TKQ2.³⁸⁷ Of particular interest, the first of the sequences (TKQ1, shown in bold in Figure 86B) overlaps the transcriptional start site (TSS) in a similar manner to the PQS found in the Aurora A kinase 5'UTR. *In vitro* biophysical studies demonstrated that truncated sequences containing the PQS showed structural characteristics of G4 in the presence of K⁺, e.g. characteristic CD peaks. In addition, G to A mutations in the G₂4 sequence resulted in increased transcription in *in vitro* dual luciferase assays, suggesting that folding of the quadruplex can partially inhibit transcription. This suggests that the G4 could play a role *in vivo* and furthermore provides some supporting evidence that G4 can form over start sites.

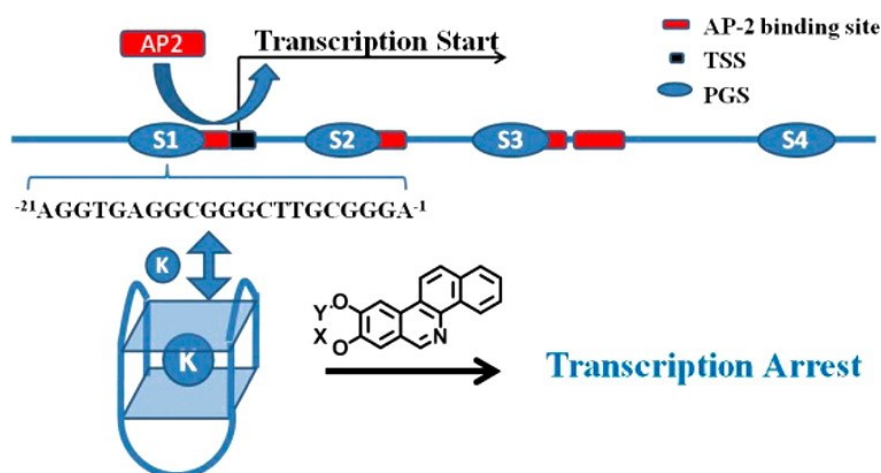


Figure 87: Transcriptional arrest of HIF1 α following stabilisation of S1 G₂4 by benzo[c]phenanthridine derivatives. Stabilisation of the S1 putative G-quadruplex forming site (labelled as PGS in this figure, but more commonly referred to as a PQS throughout this report), this blocks the binding site for the transcriptional protein AP2. Figure reproduced from Chen *et al.*³⁸⁸

Additional studies have demonstrated functional roles for DNA G₂4, including a 2014 study by Chen *et al.* investigating G4 structures in the *HIF1 α* promoter.³⁸⁸ The region contains four PQS, including S1 which forms a G₂4 and is located within the binding site for the transcriptional protein AP2 (Figure 87). The authors demonstrated using a dual luciferase assay that mutation of the PQS increased transcription of *HIF1 α* . In addition, a conformational change of the S1 quadruplex from anti-parallel formation to parallel following binding of benzo[c]phenanthridine derivatives was shown to block the binding site for AP2, preventing transcription of the gene. This was evidenced through ChIP binding assays to show the relative binding of AP2, biophysical studies investigating the conformation of bound and unbound S1 and RT-PCR assays to confirm the impact of binding on transcription. While this study is focused on a DNA G4 it provides evidence for G₂4 playing a functional role in cells. However further *in vivo* studies are needed to confirm that these more transient structures do form *in vivo*, as studies to date have been *in vitro*.

Two-quartet G4 have also been reported in mRNA sequences. For example, formation of a two quartet G4 has been demonstrated for the RNA aptamer *Spinach* by X-ray crystallography.³⁸⁹ The *Spinach* aptamer binds to a GFP-like molecule called DFHBI, activating its fluorescence, which gives it potential applications in imaging and labelling. The spinach aptamer consists of two helical domains separated by a bulge, which folds into a two-quartet G4 (Figure 88).

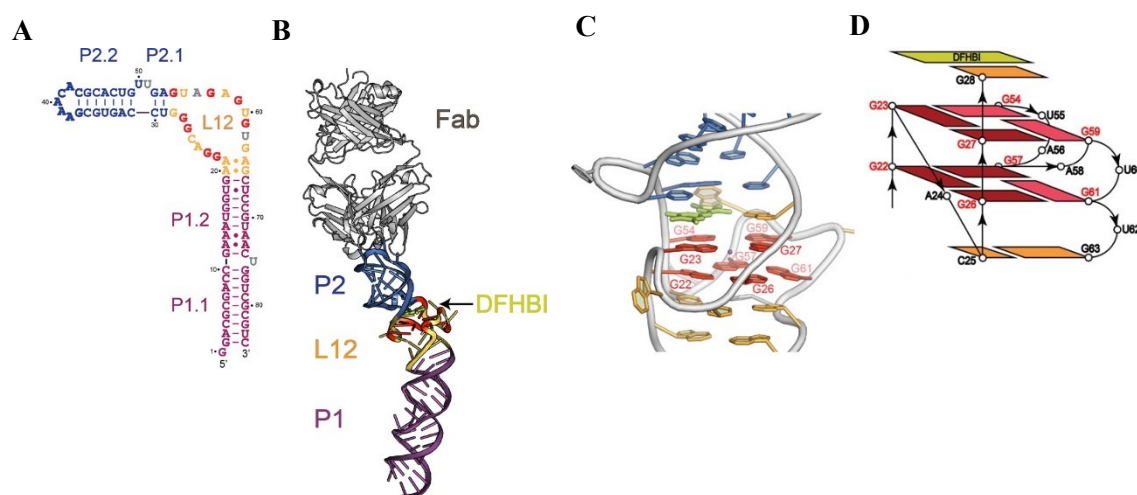


Figure 88: Structure of *spinach* RNA aptamer, containing two-quartet G4. (A) Sequence structure of *spinach* aptamer. The aptamer forms two separate domains P1 (shown in purple) and P2 (shown in blue) which is connected by a loop region L12 (yellow, with G4 forming guanine residues shown in red) which forms a two-quartet G4. (B) Structure of the *spinach* aptamer bound to Fab antibody. The L12 G4 forms a platform on which the DFHBI ligand (lime green) can stack. (C) Structure of loop region of *spinach* aptamer. The G4 structure shown in red is stabilized by a potassium ion (purple) which sits in the central channel. (D) Structure of *spinach* RNA G4, with anti-glycosidic guanines highlighted in dark red. Figure reproduced from Huang *et al.*³⁸⁹

The first study specifically investigating the structural characteristics of two-quartet RNA G4 was published in 2012 by Mullen *et al.*³⁹⁰ The authors investigated the K^+ cooperativity of folding for 2 and 3 quartet G4 with loop sequences based on sequences found in the *Arabidopsis* genome. The results of the study demonstrated that the two quartet G4 had a strong dependence on K^+ concentration and folded with positive cooperativity, as shown by the Hill coefficients (1.7-2.7). This was in contrast to the generally more stable three quartet G4, which demonstrated limited cooperativity and would fold in much lower K^+ concentrations. These results suggest that there are likely to be distinct functional roles for two and three quartet G4. The relative stability of the three quartet G4 means that they are more likely to remain folded under unstressed conditions and play a role in the everyday function of the cell. By contrast, two-quartet G4 are more likely to require additional co-factors or ligand binding to form *in vivo* and/or to require higher salt concentrations. This suggests a role in response pathways where the G4 are required to respond rapidly to changing conditions. This could be particularly relevant for plants where salt concentrations change rapidly when the plant is under stress, thereby taking advantage of the cooperativity of two-quartet G4 folding.

4.1.3. Aurora A: *In vivo* function

Aurora A is a serine/threonine kinase (STK) which is upregulated during G2 in the cell cycle and localised to the centrosomes, spindle poles and spindles. It is an important regulator of mitosis, in particular performing critical roles in controlling mitotic entry, chromosome maturation and spindle formation.³⁹¹ (Figure 89)

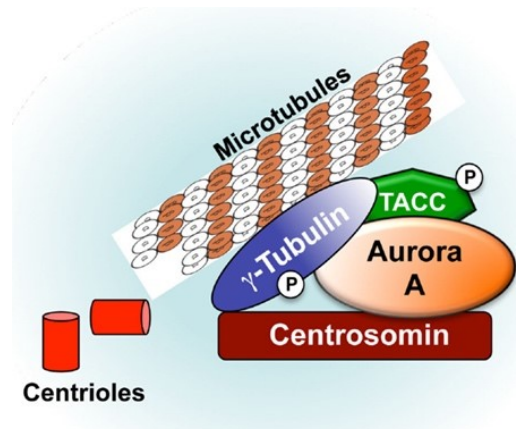


Figure 89: Aurora A complex during mitosis. During mitosis Aurora A localises to the mitotic poles and adjacent spindle microtubules. Here it forms a complex with other mitotic proteins including gamma-tubulin, TACC and centrosomin. This complex plays a significant role in spindle formation and centrosome assembly. Figure reproduced from Bavetsias *et al.*³⁹²

Inhibition of Aurora A prevents correct spindle formation within the cell, resulting in a range of abnormal spindle types, including; monopolar spindles, multipolar spindles and monoastral bipolar spindles.³⁹³ These spindle defects trigger the mitotic checkpoint, which arrests the cell in mitosis. However, within a few hours cells will escape mitosis and these tetraploid cells then face the “post-mitotic check point” which is p53 dependent. In the presence of inactive p53 the cells will enter S-phase and then undergo further mitosis with multipolar spindles. If active p53 is present, then the cells will enter G1 arrest and can either directly enter the apoptotic pathway or move through S-phase and additional mitosis cycles (Figure 90).^{394, 395}

Overexpression of *Aurora A* has been studied in numerous cell lines. In HeLa cells abnormally high levels of Aurora A resulted in polyploidy in cells and this effect was induced following overexpression of both active and inactive forms of the kinase.³⁹⁶ By contrast, in primary mouse embryonic fibroblasts this effect is only observed with active Aurora A kinase.³⁹⁷ The explanation for this could be the inhibitory effect that active Aurora A exerts on p53 which can enable the cell to escape the post-mitotic G1 checkpoint. Aurora A has been shown to phosphorylate key residues on p53, namely Ser215 and Ser315 which respectively inactivate and trigger degradation of p53.³⁹⁸

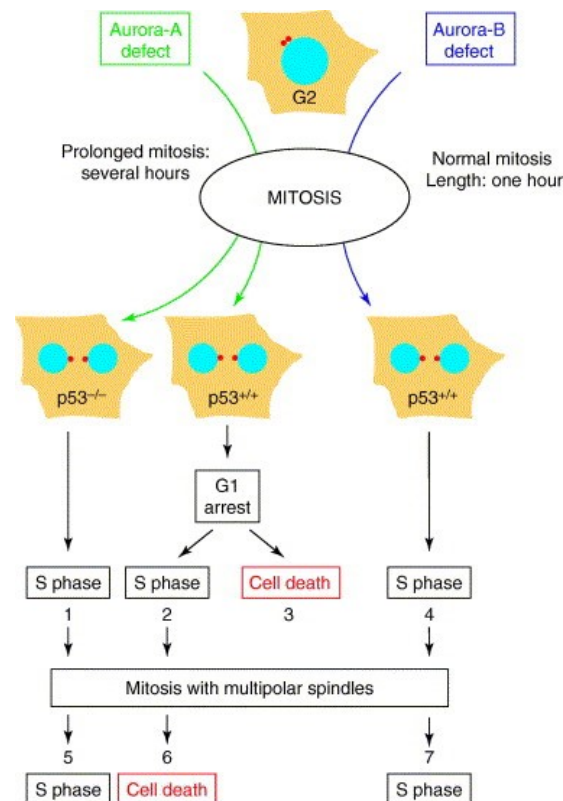


Figure 90: Mitotic pathways following inhibition of Aurora A or Aurora B. In the absence of Aurora A kinase, or presence of a defective form of Aurora A kinase, mitosis is prolonged to several hours. The cell eventually escapes from mitosis at which point the process becomes p53 dependent. Cells with inactive p53 will progress through to S-phase (1) and then complete mitosis with multipolar spindles (5). In cells with active p53 there are two potential pathways which are decided depending on the level of damage sustained during and length of mitosis. Cells with more limited damage will typically progress through S-phase (2) and repeat mitosis resulting in cell death (6). Following sustained mitotic arrest and more significant chromosomal defects the cells will enter the apoptotic pathway following arrest at G1 (3). Following Aurora B defects the spindle checkpoint becomes defective, meaning that cells progress through mitosis at the normal rate. Cells are then able to progress through to S-phase as the G1 post-mitotic checkpoint has not been activated by prolonged mitosis and the cells survive, even in the presence of active p53. Figure reproduced from Giet *et al.*³⁹⁹

4.1.4. Aurora A: Current therapeutics targeting Aurora kinases

Aurora A has been designated as an oncogene since 1998, when it was demonstrated in two independent studies that overexpression of Aurora A was sufficient to transform NIH3T3 and rat1 cells.^{400, 401} Overexpression of *Aurora A* has been reported in, among others; breast,⁴⁰² lung,⁴⁰³ ovarian and endometrial,^{404, 405} prostate⁴⁰⁶ and colorectal cancers.⁴⁰⁷ A number of different meta-analyses, for example Qin *et al.* 2015, have linked polymorphisms in the *Aurora A* gene to over-expression in different cancer sub-types.^{72, 408} For example, polymorphisms in the *Aurora A* gene have been linked to increased lung cancer risk.⁴⁰⁹ De Castro *et al.* demonstrated through genetic ablation studies that *Aurora A* expression is required for progression of mammary-gland and skin tumours.⁴¹⁰ It is hypothesised that *Aurora A* overexpression contributes towards tumour progression by promoting cell cycle progression and therefore enabling increased proliferation.

Aurora A specific inhibitors have been shown to have an anti-proliferative effect on tumours *in vivo* and there are a number of inhibitors that have entered clinical trials for a variety of tumour subtypes (see Table 8). The most well studied inhibitor is Alisertib,⁴¹¹ which has reached phase II/III trials for the treatment of a variety of tumours.⁴¹² Alisertib has a benzazepine core scaffold with an aryl carboxylic acid and a fused amino pyrimidine ring (Figure 91). Initial studies demonstrated that Alisertib is a selective inhibitor of Aurora A kinase, with an IC_{50} of 1.2 nM, compared to 396.5 nM for Aurora-B.⁴¹³ Its effects as a tumour suppressor have been tested in a range of preclinical cell lines and *in vivo* models, as well as in a variety of phase I-III clinical trials against several different applications (Table 8). In preclinical trials, not only did Alisertib display anti-proliferative activity against cell lines including for multiple cancer subtypes,⁴¹³⁻⁴¹⁵ but also showed evidence that it acted in a cooperative manner with nilotinib in the treatment of chronic myeloid leukaemia.⁴¹⁶ The drug also exhibits positive pharmacokinetic properties and is able to be dosed in an oral form. However, in clinical trials there have been some recent late stage failures. In May 2015, Takeda terminated its phase III trial for patients with relapsed or refractory peripheral T-cell lymphoma following an interim analysis that suggested that the drug would not be able to meet pre-defined criteria indicating a superior progression free survival rate compared to the current standard of care. However, trials for a number of other applications, including small cell lung cancer, continue and currently appear promising.⁴¹⁷

This provides a mixed picture for *Aurora A* as a relevant cancer target. There appears to be potential for *Aurora A* inhibitors as a therapeutic agent for treating certain cancers. However, their efficacy can be limited by a number of factors, including the presence of p53 mutations, and even successful therapeutics have thus far failed to produce significant advancements compared to the standard of care. As such, when developing potential Aurora inhibitors, the mutagenic profile of the end therapeutic applications should be considered along with the relative efficacy of any potential drug targets. Drug candidates with lower binding activity may also need to be cut out at an early stage.

Table 8: Current Aurora A or pan-Aurora inhibitors in clinical trials, as of 16-06-2017. Data reproduced from Borisa *et al.*⁴¹⁷ and originally retrieved from <http://clinicaltrials.gov>

Name	Clinical trial	Disease condition	Starting date	Current status
Alisertib (MLN8237)	Phase 3 NCT01482962	Relapsed or refractory peripheral T-Cell lymphoma	June 2012	Terminated (May 2015)
	Phase 2 NCT02560025	Induction Chemotherapy in patients with High-risk acute myeloid Leukaemia	December 2015	Ongoing

Name	Clinical trial	Disease condition	Starting date	Current status
	Phase 2 NCT01316692	Patients with unresectable stage iii-iv melanoma	October 2011	Terminated (April 2016)
	Phase 2 NCT02038647	Small-cell-lung Cancer (SCLC) treatment in combination with paclitaxel	February 2014	Ongoing
	Phase 1 NCT02551055	Neoplasms, Advanced or Metastatic	October 2015	Ongoing
	Phase 1 NCT02214147	Advanced solid tumours or Relapsed/Refractory lymphoma	September 2014	Completed
Tozasertib (VX-680, MK0547)	Phase 2 NCT00290550	Non-small-cell lung carcinoma	June 2006	Terminated
	Phase 2 NCT00405054	Leukaemia	December 2006	Terminated
	Phase 1 NCT00500006	Chronic myelogenous leukaemia, lymphoblastic, acute, philadelphia-positive	October 2007	Terminated
	Phase 1 NCT00111683	Chronic Myelogenous leukaemia in blast crisis, lymphocytic leukaemia, B Cell acute and chronic myelogenous leukaemia	June 2005	Completed August 2015
	Phase 1 NCT02532868	Advanced solid tumours	May 2005	Terminated
	Phase 1 NCT00099346	Colorectal cancer	January 2005	Terminated
MLN8054	Phase 1 NCT00249301	Advanced solid tumours	October 2005	Terminated
	Phase 1 NCT00652158	Advanced malignancies	April 2006	Terminated
Danusertib (PHA739358)	Phase 2 NCT00872300	Multiple myeloma	October 2008	Completed May 2014
	Phase 2 NCT00766324	Hormone refractory prostate cancer	September 2007	Completed May 2014
AMG-900	Phase 1 NCT00858377	Advanced solid tumours	August 2009	Ongoing

Name	Clinical trial	Disease condition	Starting date	Current status
	Phase 1 NCT01380756	Acute myeloid leukaemia	July 2011	Completed February 2015
VX-689 (MK-5108)	Phase 1 NCT00543387	Advanced and/or Refractory solid tumours	March 2008	Completed April 2011
SNS-314 Mesylate	Phase 1 NCT00519662	Advanced solid tumours	August 2007	Completed March 2009
TAK-901	Phase 1 NCT00807677	Advanced hematologic malignancies	March 2009	Completed March 2013
	Phase 1 NCT00935844	Advanced solid tumours or lymphoma	October 2009	Completed November 2011
CYC116	Phase 1 NCT00560716	Advanced solid tumours	June 2007	Terminated
PF03814735	Phase 1 NCT00424632	Advanced solid tumours	November 2006	Completed June 2009
ENMD-2076	Phase 2 NCT01914510	Ovarian clear cell cancers	September 2013	Ongoing
	Phase 2 NCT01639248	Previously treated locally advanced + metastatic tnbc	July 2012	Ongoing
	Phase 2 NCT02234986	Advanced fibrolamellar carcinoma	October 2015	Ongoing
	Phase 2 NCT01104675	Patients with ovarian cancer	April 2010	Completed December 2012
	Phase 1 NCT00658671	Advanced cancer	April 2008	Completed June 2012
	Phase 1 NCT00806065	Multiple myeloma	December 2008	Completed January 2012

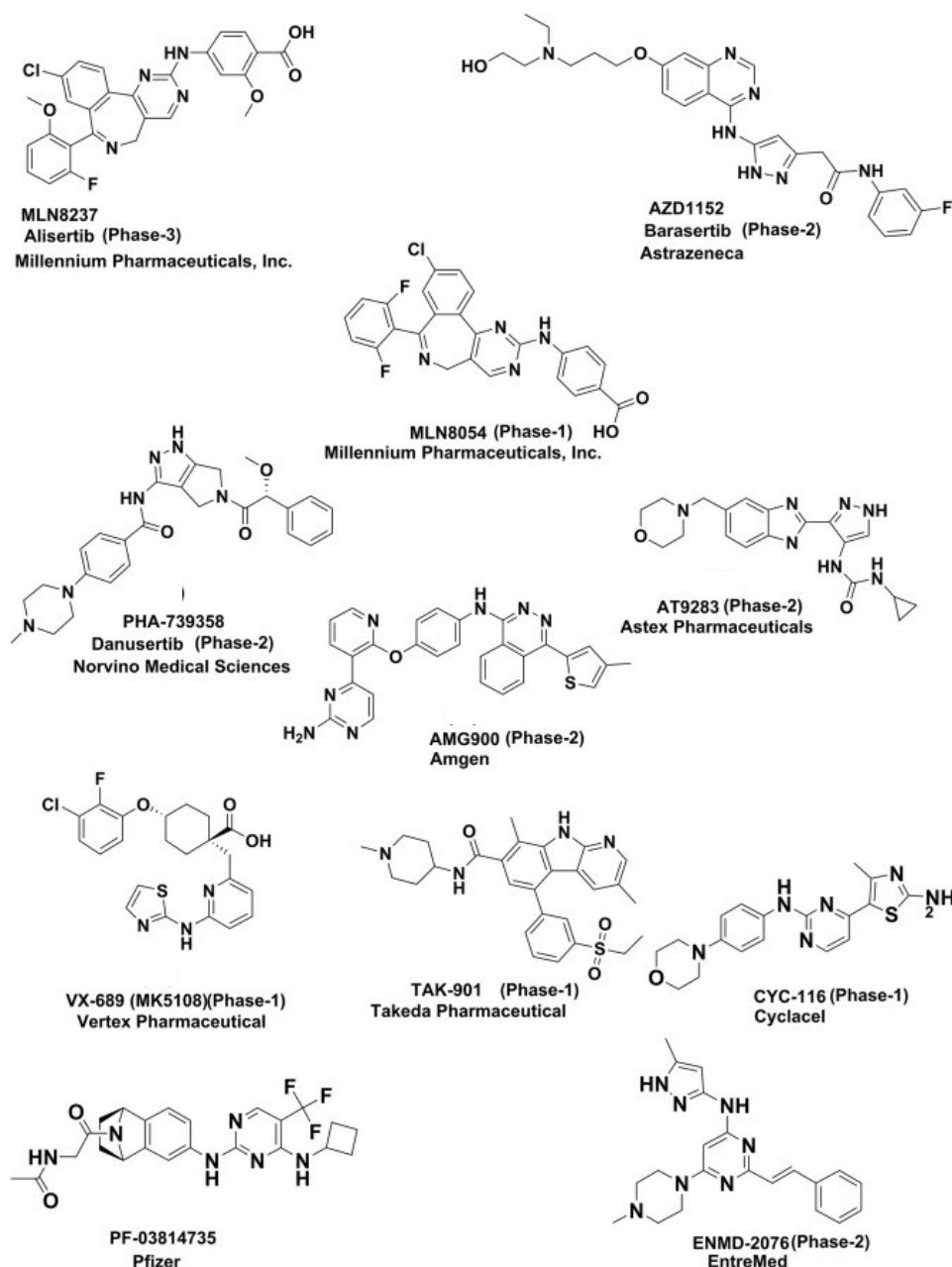


Figure 91: Chemical structures of Aurora Kinase inhibitors in clinical trials. Companies sponsoring the relevant trials are noted along with the current phase. Note: figure includes Aurora B inhibitors Barasertib and AT9283. Figure reproduced from Borisa *et al.*⁴¹⁷

4.1.5. Col1A2: *In vivo* function

The *Col1A2* gene encodes the α -2 chain of type I collagen, a major extracellular matrix (ECM) protein involved in skin and bone formation. The α -2 chain binds with two α -1 chains to form a triple helix, which is known as mature type I collagen (Figure 92). The expression of *Col1A2* is controlled by a range of pro and anti-fibrotic transcription factors (Figure 93), with TGF β a key factor stimulating increased transcription.

Mutations in the *Col1A2* gene can result in bone conditions including osteoporosis and osteogenesis imperfecta and skin conditions such as Ehlers-Dalos syndrome.⁴¹⁸ A polymorphism in the *Col1A2* 3'UTR has been associated with increased risk for hepatocellular carcinoma.⁴¹⁹ In addition, CpG hypermethylation of the *Col1A2* gene in human bladder cancer has been linked to increases in cell proliferation and tumour migration⁴²⁰ and overexpression of *Col1A2* and type I collagen observed in medulloblastoma,⁴²¹ (the most common malignant brain tumour in children) and in human breast cancer.⁴²²

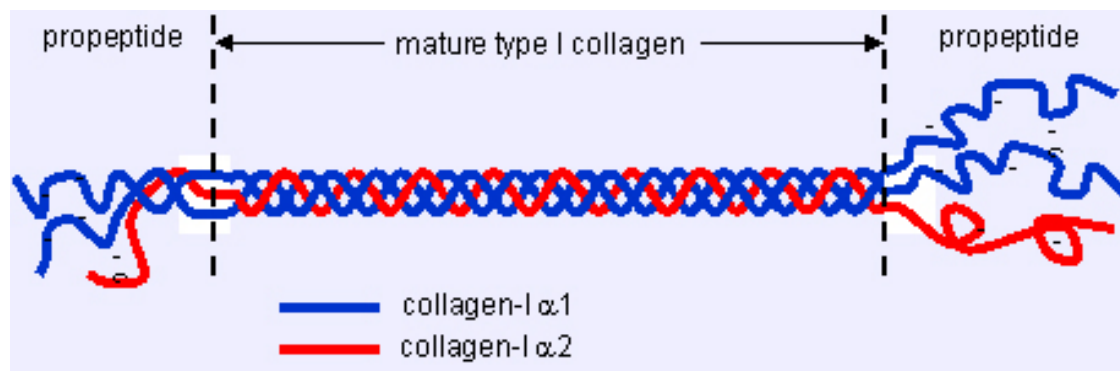


Figure 92: Structure of type I collagen. Two strands of collagen-1 α -1 bind with one strand of collagen-1 α -2 to form a triple helix structure. Figure reproduced from Mottes *et al.*⁴²³

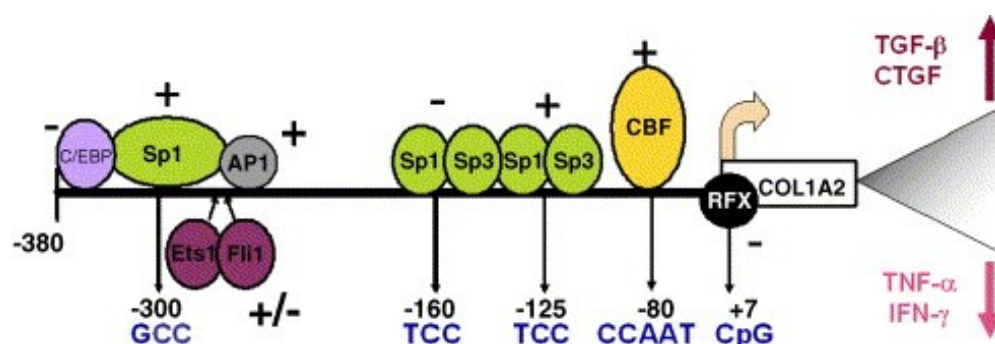


Figure 93: Representation of the cis and trans acting regulators of *Col1A2* transcription. Figure reproduced from Ramirez *et al.*⁴²⁴

4.2. Methods

RNA sample preparation

HPLC-purified Aurora A and Aurora A Mut RNA oligonucleotides were purchased from Eurogentec. The sequences are shown below, with G4 forming guanines shown in bold and mutated nucleotides underlined.

Aurora A: 5'-**GGAGGCCGAGGCAUCAUGG**-3'

Aurora A Mut: 5'-**GGAA**ACCGAGGCAUCAUGG-3'

CollA2: 5'-GGGCGGGGGAGGGCGGG-3'

CollA2 Mut: 5'-GGGCGGGGGAAACGGG-3'

RNA samples were prepared at 10 μ m final strand concentration in RNase-free water with 10 mM Tris-HCl, pH 7.4 buffer and varying concentrations of KCl, ranging from 0-100 mM. Prior to spectroscopy, RNA was folded by heating to 95°C for 5 minutes and leaving to slow-cool to room temperature overnight.

CD spectroscopy

CD spectra were produced as previously described in Chapter 2. CD spectra were normalised to 0 at 320 nm and graphs created in Prism (GraphPad).

CD thermal denaturation studies

Melting curves were obtained by measuring the CD value of 10 μ m *Aurora A* G4 RNA at 263 nm every 0.1°C using a J-380 spectropolarimeter (Jasco) with a Peltier temperature controller adjusting the temperature from 4°C to 80°C (at a gradient of 1°C per minute). When indicated, ligands were added at a final concentration of 20 μ m. Results were plotted in Prism (GraphPad) and melting temperatures calculated from the peak of the derivative of the CD spectra.

4.3. Results

4.3.1. Evidence of G-quadruplex formation *in vitro*

- *Aurora A* G4

The *in vitro* structure of the *Aurora A* PQS was investigated using CD spectroscopy and the effect of temperature on the CD signal at 263 nm used to determine its stability. The CD spectra of the *Aurora A* in 200 mM KCl exhibits the characteristic signature of a parallel-stranded G4. This signature is both temperature-dependent (i.e., disappears when increasing temperature up to 90°C), K⁺ sensitive (i.e. is attenuated at lower KCl concentrations) and sequence-dependent (i.e., mutation of one of the four guanine-tracts to adenines results in the disappearance of the negative peak and a positive shift in the positive peak to 273 nm) (Figure 94A and Figure 94B). These results suggest that the *Aurora A* PQS can adopt a G4 conformation *in vitro*. Variable-temperature CD at 263 nm indicated a melting temperature of 56.7°C (Figure 94C and Figure 94D). As expected considering that the *Aurora A* G4 can only form two tetrads, this is very low when compared to other RNA G4 previously reported in the literature but that can form three G-tetrads. For example, significantly higher T_m values were reported

for *NRAS* (69°C), *MT3* (71°C) and *TRF2* (77°C) even in the presence of much lower salt concentrations (5 mM K⁺).⁴²⁵

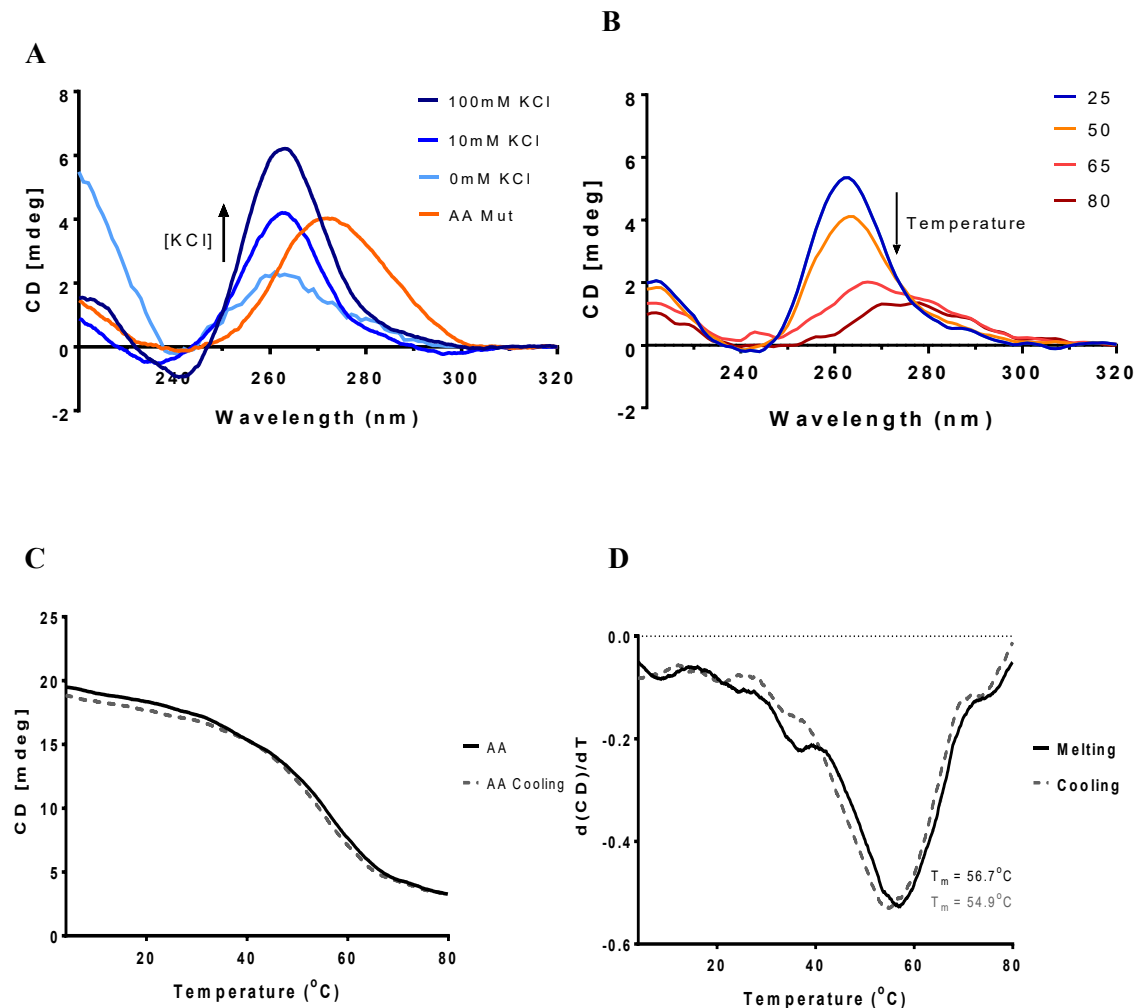


Figure 94: CD spectra of Aurora A PQS. A) CD spectra of Aurora A and Aurora A Mut mRNA in 100 mM KCl. **B)** CD spectra of Aurora A mRNA in 100 mM KCl at 25 °C, 50 °C, 65 °C and 80 °C. **(C)** Melting and cooling curves for Aurora A mRNA, measured at 263 nm. **(D)** First derivative of Aurora A melting and cooling curves, with derived T_m values of 56.7 °C and 54.9 °C respectively.

- **Col1A2 G4**

As anticipated from the primary sequence, the *Col1A2* RNA oligonucleotide, with a high density of guanines connected with very short single nucleotide loops was also shown to form a parallel stranded G4 structure in vitro, even at low K⁺ concentration (see Figure 95 for a CD spectrum of *Col1A2* RNA in a solution containing only 10 mM KCl). Attempts to fully melt this prefolded G4 proved unsuccessful, further demonstrating the extremely high thermal stability of this structure (T_m > 95°C at [KCl] > 10 mM).

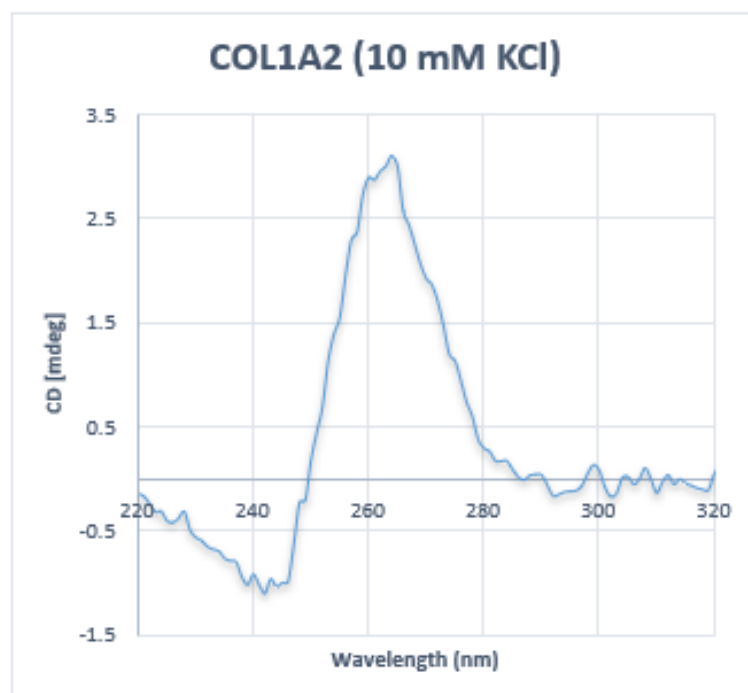


Figure 95: CD spectrum of Col1A2 PQS in 10 mM KCl solution.

Because of the intrinsically high stability of this *Col1A2* G4, it would have been extremely difficult to assess experimentally the stabilising potential of small molecules. For this reason and because Aurora kinase A was considered a potentially much more interesting therapeutic target than Col1A2, no further studies were carried out on the *Col1A2* quadruplex and we decided to focus our attention on the *Aurora A* G4, as described below.

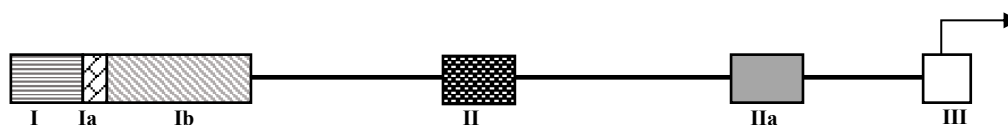
4.3.2. Aurora A variants

A search through the NCBI nucleotide database identified 11 different transcriptional variants of human *Aurora A* mRNA, including two predicted variants. The *Aurora A* G4 sequence is found in four of these transcripts: predicted transcript variant X1 (XM_017028034.1) and transcript variants 3 (NM_198434.2), 5 (NM_198436.2) and 7 (NM_001323303.1). The first section of the sequence is found at the end of *exon II* and the second section at the beginning of *exon III*. As a result, in the full 5'UTR the sequence is split by the presence of the 98bp *exon IIa*. Excision of *exon IIa* brings together the two sections of the G4, as demonstrated in Figure 96.

Changes in expression of *Aurora A* splice variants have been demonstrated in both breast and colorectal cancer. Shin *et al.* found three different splice variants, which showed differential expression in breast cancer cell lines and primary tumour cells, with the longer variants much more common in primary tumours.⁴²⁶ In addition, an analysis of the *Aurora A* mRNA in a sample of high-risk breast cancer

patients identified no major spliced mRNA.⁴²⁷ The lack of splicing in breast cancer cells could prevent regulation of *Aurora A* expression. The fact that the *Aurora A* G4 sequence is created through alternative splicing would support its role as part of this splicing dependent regulatory effect.

A



B

Transcript Variant 9 (NM_001323305.1)

241 gtggctcatg cccgtaatcc cagcattttc ggaggccgag gctggagtgc aatggtgtga
 301 tctcagctca ctgcaacctc tgcttctgg gttaagtga ttctctgcc tcagctccc
 361 gagtagctgg gattacaggc atcatgga gacc gatctaaaga aaactgcatt tcaggacctg

Key: Exon IIa

Transcript Variant 5 (NM_198436.2)

241 gtggctcatg cccgtaatcc cagcattttc ggaggccgag gcatcatgga ccgatctaaa

Figure 96: PQS formation through alternative splicing of *Aurora A* transcripts. (A) Map of exon sites in *Aurora A* 5'UTR, intronic regions shown as a line. (B) Excerpts from the 5'UTR sequences of *Aurora A* transcript variants 5 and 9, demonstrating G4 sequence formation upon excision of Exon IIa. Nucleotides in the PQS (underlined), quadruplex forming guanines (bold) and the translation start codon (red) are all highlighted.

Researchers investigating EGFR-overexpressed cancers investigated the expression transcript variants 1-6 and found that variants 3 and 5, which contain the complete *Aurora A* G4 sequence, were dominantly expressed in human cancer tissues and showed increased ribosome association and translation efficiency following EGF stimulation.⁴²⁸ The same study found that *Aurora A* and EGF were co-expressed in 61.54% of colorectal cancer samples. This study provides strong evidence that G4-containing transcripts play a role in *Aurora A* overexpression in colorectal cancers. Targeting the *Aurora A* G4 could therefore allow specific targeting of *Aurora A* over-expression in tumour cells. There are two potential mechanisms to target the G4. The first assumes EGFR-related overexpression of transcripts 3 and 5 is independent of G4 formation, in which case ligands can be designed to stabilise the quadruplex which will conceal the start codon and inhibit translation of these transcripts. However, the G4 may be involved in protein recruitment and the design of PNA sequences to destabilise the quadruplex would be the better therapeutic strategy.

4.3.3. Effect of ligand binding on G4 formation

We then investigated the ability of two well-characterised G4-stabilising small molecule ligands, PDS and NDI (Figure 97A and 97B), to stabilise the otherwise highly unstable *Aurora A* RNA G4. The addition of 2-equivalents of NDI had little effect on the G4 melting temperature ($T_m = 56.1^\circ\text{C}$). However, addition of PDS resulted in an increase in T_m from 56.7°C to 63.0°C , suggesting a ligand-induced stabilisation effect (Figure 97C and 97D). To gain more information on the nature of the interaction between PDS and the *Aurora A* G4, CD spectra of the *Aurora A* G4-forming sequence were recorded with increasing concentrations of PDS (up to $50\mu\text{M}$) (Figure 98). Addition of PDS induced an attenuation of the positive peak at 265 nm, suggesting that PDS does not stabilise the G4 structure and may instead destabilise it.

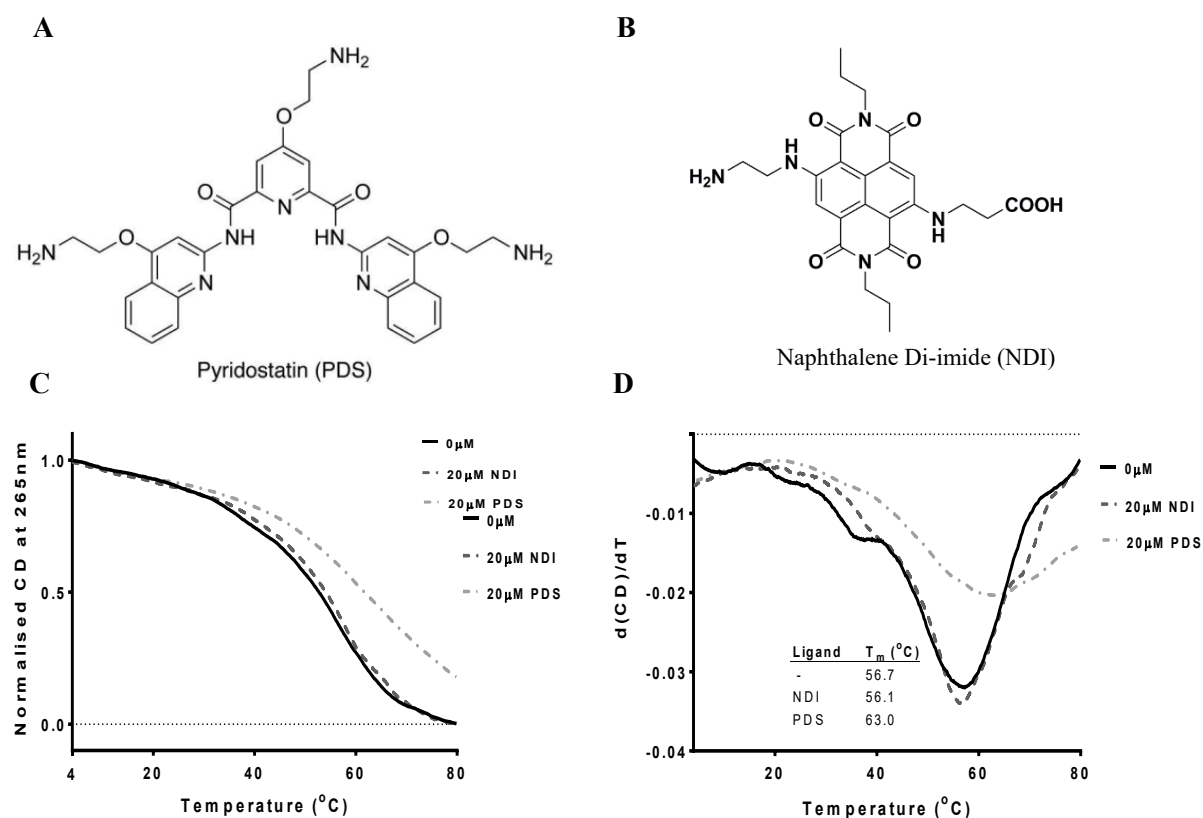


Figure 97: Effect of ligand binding on T_m of *Aurora A* G4. (A-B) Structures of the G4 binding small molecules PDS (A) and NDI (B). (C) Melting curves of *Aurora A* RNA without ligand (black, solid line), with 20 μM NDI (dark grey, dashed line) and 20 μM PDS (light grey, uneven dashed line), measured at 263 nm. (D) First derivative of melting curves and derived melting temperatures T_m of 56.7°C (no ligand), 56.1°C (NDI) and 63.0°C (PDS).

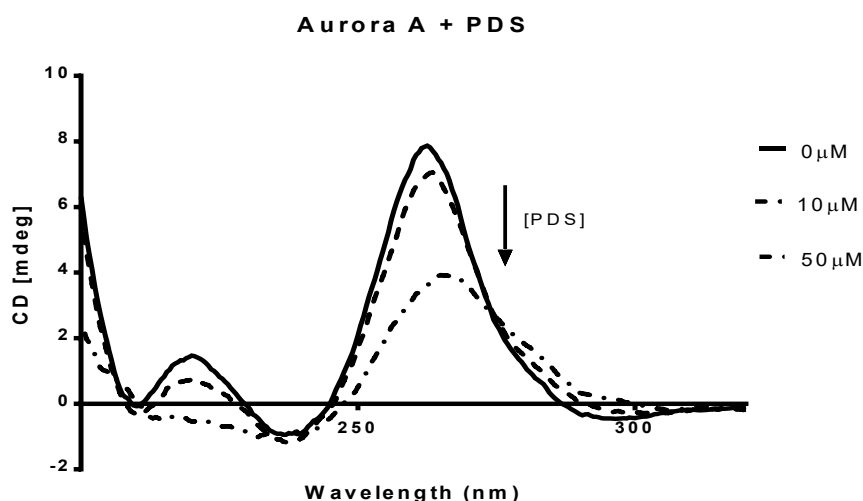


Figure 98: CD spectra of *Aurora A* G4 with 0 μM (solid line), 10 μM (dashed line) and 50 μM (uneven dashed line) PDS.

4.4. Discussion

4.4.1. Formation of G4 by novel PQS

This work has identified two novel PQS in therapeutically relevant *Aurora A* and *Col1A2* transcripts. Using CD spectroscopy, the ability of these PQS to form a G4 structure in vitro was investigated. Both PQS were shown to form a parallel stranded G4 in buffered solutions containing various concentrations of KCl. However, as expected based on their primary sequences, these two G4 showed contrasting thermal stabilities. Whilst the *Aurora* kinase A G4 proved highly unstable ($T_m = 56.7^\circ\text{C}$ at 100 mM KCl), the *Col1A2* G4 proved too stable to melt (even at low, non-physiological salt concentrations), thus preventing us from determining a T_m value. The stability of the *Col1A2* G-quadruplex fits with that of other three-quartet quadruplexes such as NRAS, which did not melt even at 95°C when tested at 100 mM KCl, and retained a melting temperature of $63 \pm 1^\circ\text{C}$ at 1 mM KCl.¹¹⁶ The low stability of the *Aurora* Kinase A G4 is consistent with previous research which has suggested that G4 formed from 2 G-quartets are less stable than those formed from 3 G-quartets.³⁹⁰ Further spectroscopic studies (e.g. ^1H NMR) to characterise *Aurora A* PQS will provide a more conclusive body of evidence for *Aurora A* G4 formation

4.4.2. Functional role of *Aurora A* G4

The most likely functional role for the *Aurora A* PQS would be to regulate translation, which would match the functional roles suggested for other confirmed RNA G4. However, there are several potential mechanisms by which it could enact this. The first key point to discuss is the potential structure of the G4 itself. While the PQS appears capable of forming a G4 structure, it is possible that *in vivo* a G4 is

formed using nucleotides from outside the core PQS sequence. Studies by Phan *et al.* investigating DNA G4 have identified G4 structures which involve isolated guanine residues, even when a set of four contiguous guanines are present in close proximity.^{429, 430} As a result it is possible that *in vivo* a more stable G4 is formed using elements of the PQS and isolated guanines or guanine tracts. In addition, it is also possible that the G4 forms as part of a larger structure, which would help to stabilise the G4. For example, G4 have been shown to form a part of larger IRES structures.⁴³¹ The *Aurora A* G4 could potentially form part of a larger structure such as a pseudoknot, which could act as a roadblock to translation machinery. A simple modelling study could identify whether the sequence around the PQS is likely to form any additional structures.

It would also be interesting to determine whether the quadruplex forms *in vivo* and if it is involved in regulating the EGFR-mediated increases in translation of exon II (+), IIa (-) transcripts. This could be ascertained with a battery of studies including the design of bimolecular fluorescent probes¹⁶⁰ to assess its formation *in vivo*, translation assays to demonstrate whether mutation of the G4 sequence prevents EGFR-mediated up-regulation and translation pull-down assays and binding studies to identify any EGFR-related proteins that bind directly to the G4. It is also worth investigating the other impacts that the splice variant can have on the expression of the gene. While the changes in splicing affect the PQS, they also affect the Kozak sequence around the translation start codon. The Kozak sequence is a consensus sequence which includes the translation start codon and plays a significant role in the initiation of translation.⁴³² It is possible that the changes in sequence occur to alter the Kozak sequence rather than to specifically create the PQS.

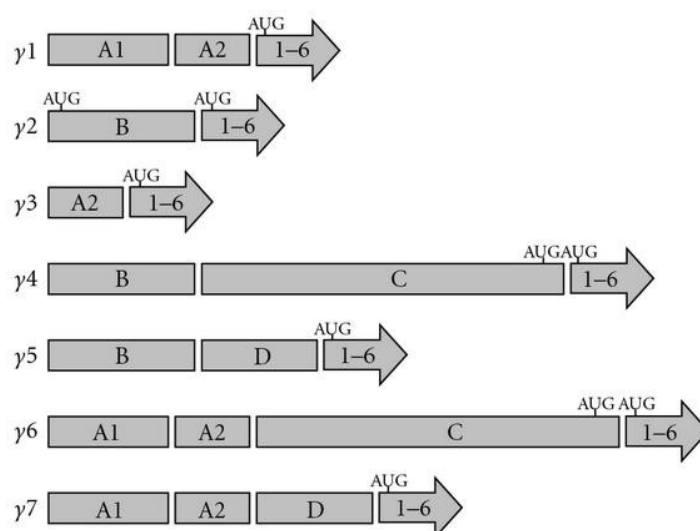


Figure 99: Structure of 5'UTR of PPAR gamma splice variants. The exons in the 5'UTR of the 7 PPAR gamma splice variants. The start codon for translation is marked with AUG, in addition, short open reading frames within the 5'UTR are marked with another AUG upstream of the correct start site. Figure reproduced from McClelland *et al.*⁴³³

There have been previous examples of changes in splicing affecting the translation of mRNA. McClelland *et al.* investigated translation of splice variants of PPAR gamma splice variants (Figure 99), demonstrating that variants 1, 2 and 5 are efficiently translated, whereas variants 4 and 7 show low levels of translation.⁴³³ The authors investigated a number of different factors which influenced the transcription of the different splice transcripts, these included; (i) formation of stable secondary structures in the exons, (ii) introduction of base pairing of nucleotides in the Kozak sequence, (iii) the number of start codons in the 5'UTR and (iv) the overall length of the 5'UTR. It is points (i) and (ii) that bear the most relevance to the *Aurora A* PQS, as the formation of a G4 both creates a secondary structure which can slow progress of the ribosome and introduces base pairing into the Kozak sequence.

4.4.3 *Aurora A* G4 as a biomarker

As the G4-forming sequence is only formed in transcripts upregulated in colorectal cancer it provides a potential biomarker with which to study transcription levels in this cancer sub-type. PNA-conjugation to a fluorescent G4-specific probe such as PyroTASQ⁴³⁴ could provide a method to probe the function of *Aurora A* in tumour progression and identify tumours with *Aurora A* upregulation.

4.4.4. *Aurora A* G4 as a therapeutic target

The first question to consider when thinking about the potential of the *Aurora A* G4 is whether the G4 is present in a relevant therapeutic setting and whether the end condition requires new treatment methods. The up-regulation of the G4 containing transcripts (3 and 5) in EGFR-overexpressed colorectal cancer make this the clearest potential application and for the remainder of this section discussion will focus on the potential of the *Aurora A* G4 to treat colorectal cancer. However, it is worth noting that *Aurora A* is over-expressed in a much larger range of cancers and that as experiments progress the potential impact on these should be considered further.

Currently, early stage colorectal cancer is typically treated with surgery, and with therapeutic treatments only used for stage IV and sometimes stage III cancers. In this case surgery is normally combined with chemotherapy, radiotherapy and / or therapeutic treatments to try and improve outcomes. Current drugs approved for the treatment of colorectal cancer include VEGF inhibitors; Bevacizumab (Avastin)⁴³⁵, Ramucirumab (Cyramza)⁴³⁶ and Ziv-aflibercept (Zaltrap)⁴³⁷, and EGFR inhibitors; Cetuximab (Erbix)⁴³⁸ and Panitumumab (Vectibix).⁴³⁹ A key issue with therapeutics for EGFR inhibitors is that all patients eventually develop resistance to the treatment.⁴⁴⁰ This should be considered in two respects when working on potential therapeutics; firstly any potential drug should be tested in combination with EGFR-inhibitors to identify whether there is any complementary effect limiting the build-up of

resistance and second the mechanism of resistance for EGFR targeting drugs needs to be considered to ensure that a similar effect isn't seen with a drug targeting *Aurora A*.

An additional consideration that should be taken into account is that therapeutics targeting Aurora A kinase have typically shown greater success against hematologic tumours rather than solid tumours.³⁹² This could be due to the fact that, as a result of targeting the mitotic cycle, the cell has to go through multiple rounds of replication before the drug has an impact. This naturally means that the impact will be greater on hematologic cells which have faster replication rates, allowing the drugs to work at a lower concentration, with the additional benefit of reduced side effects.

The next step with this work is to demonstrate whether the *Aurora A* G4 can inhibit translation *in vitro* using cell-free and mammalian cell reporter assays. This will provide evidence for the therapeutic potential of the G4. Following on from this a panel of G4 binding ligands should be tested to identify the impact on *Aurora A* G4 mediated translation. Ligands that show a significant impact on translation can then be progressed to further studies. Although initial ligand screening proved unsuccessful, broader screening may enable identification of a ligand capable of stabilising the structure.

Conclusion

The overall aim of this research was to develop a method to target specific/unique RNA G4 in the presence of alternative G4 and secondary structures. To achieve this, a novel class of PNA-small molecule conjugates was produced to target both the flanking sequences and the core structure of RNA G4, thereby increasing specificity and inhibitory activity. Results have demonstrated that PNA conjugation increases the inhibitory effect of NDI ligands on mRNA translation. In-vitro assays showed that PNA-NDI conjugates could inhibit translation with an IC_{50} circa one order of magnitude lower than that of the PNA sequence or the NDI alone. However, there were significant levels of non-specific binding to RNA oligonucleotides with mutations in the G4 forming sequence (to prevent G4 formation) and in the G4 flanking regions (to prevent sequence-specific PNA:RNA hybridisation). Although more structural studies are necessary to explain this lack of specificity, this is most likely the result of electrostatic interactions between the negatively charged RNA phosphate backbone and the multiple positively charged lysine residues on the ligand. The length of PNA (from 6-mer to 10-mer) did not appear to significantly affect the activity of PNA-conjugates in the translation assay, although a significant decrease in the IC_{50} was observed for the longest (12-mer) ligand 6LO. This non-linear increase in activity with PNA length is similar to that seen with antisense PNA, suggesting that this effect is mediated by increased antisense activity of the PNA strands.

For the next generation of PNA-small molecule conjugates it could therefore be beneficial to replace the lysine residues with PEG moieties to reduce non-specific electrostatic interactions and prevent non-specific PNA-RNA binding. There are also a number of other modifications that would be worth investigating to improve the properties of the ligands, including NDI side chain modifications to improve specificity and modification of the PNA backbone (e.g. GPNA) to improve cellular uptake and binding specificity. Further experiments testing the effect of PNA strand orientation on binding specificity would also be interesting (especially with longer, more active PNA oligomers), even though comparison of the effects of parallel (4LO) and antiparallel (4LOa) ligands on translation inhibition showed no significant differences in IC_{50} values. The *in vitro* translation assay also has intrinsic limitations which could be overcome by switching to a mammalian cell GFP reporter assay. In addition, it would be worth exploring further binding assays under conditions that better mimic what happens *in vivo*. This includes, for example, ^1H NMR to provide more detailed structural information and SPR for more quantitative analyses, including binding affinities and binding stoichiometry. However, for SPR in particular, care would have to be taken during optimisation and analysis as inclusion of the flanking regions drastically increases the complexity of the assay.

There is a clear need for additional work focused on assessing sonoporation as a versatile method to improve the cellular uptake of PNA. The development of an effective method to improve PNA uptake is vital to the success of therapeutic PNAs (including our class of PNA-conjugates) as the poor bioavailability of PNA would significantly limit *in vivo* efficacy. An *in vitro* sonoporation assay was developed, with modifications to improve physiological relevance including the addition of flow and elimination of standing waves. Initial results indicated optimal settings of PNP: 1000 kPa, 1000 cycles, burst period: 1 second, insonation time: 400 seconds, which produced a transfection efficiency of 28% with a cell viability of 90%. However, viability was only measured through cell attachment (detached cells were removed by flow) and there was no consistent assessment of whether the attached transfected cells had entered apoptosis. To fully validate these results, it is vital that a viability assay is optimised and included in the assay. There are also further adjustments that can be made to the parameters to improve transfection and viability. In particular, the use of a higher frequency transducer could allow efficient transfection at lower PNP, which could reduce inertial cavitation and improve cell viability. While there is more work to be done, our promising early results justify further investigation into sonoporation as a method to improve the cellular uptake of PNA.

A final goal of this work was to identify and characterise new therapeutically relevant RNA G4 targets. A PQS was identified in the 5'UTR sequence of *Aurora A* mRNA and CD was used to confirm that it could form a G4 structure *in vitro*. This *Aurora A* G4 is of particular interest due to its unique position encompassing the start codon. The PQS is split between two exonic regions and the full sequence is only formed in a subset of alternatively spliced transcripts. Notably, the transcripts containing the complete G4 sequence have been shown to be directly involved in EGF-mediated upregulation of *Aurora A* expression. This suggests that the G4 may be involved in mediating EGF-induced *Aurora A* upregulation, which is worthy of further investigation. Further structural characterisation of this new G4 is required (e.g. by ^1H NMR), along with an assessment of its potential effect on translation (cell free or mammalian cell translation assays). Finally, further ligand screening should be used to identify ligands that can bind and stabilise this structure and can also be used to assess the viability of ligand-induced G4-formation as an efficient method to inhibit translation.

References

- [1] Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature* 171, 737-738.
- [2] Chen, J. L., and Greider, C. W. (2005) Functional analysis of the pseudoknot structure in human telomerase RNA, *Proceedings of the National Academy of Sciences of the United States of America* 102, 8080-8085.
- [3] Moore, P. B. (1999) Structural motifs in RNA, *Annual Review of Biochemistry* 68, 287-300.
- [4] Greenbaum, J. A., Parker, S. C. J., and Tullius, T. D. (2007) Detection of DNA structural motifs in functional genomic elements, *Genome Research* 17, 940-946.
- [5] Bacolla, A., and Wells, R. D. (2004) Non-B DNA conformations, genomic rearrangements, and human disease, *Journal of Biological Chemistry* 279, 47411-47414.
- [6] Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., and Kopka, M. L. (1982) The anatomy of A-DNA, B-DNA, and Z-DNA, *Science* 216, 475-485.
- [7] Rich, A., and Davies, D. (1956) A new two-stranded helical structure - polyadenylic acid and polyuridylic acid, *Journal of the American Chemical Society* 78, 3548-3549.
- [8] Reiter, N. J., Maher, L. J., and Butcher, S. E. (2008) DNA mimicry by a high-affinity anti-NF-kappa B RNA aptamer, *Nucleic Acids Research* 36, 1227-1236.
- [9] Rich, A. (1960) A hybrid helix containing both deoxyribose and ribose polynucleotides and its relation to the transfer of information between nucleic acids, *Proceedings of the National Academy of Sciences USA* 46, 1044-1053.
- [10] Day, R. O., Seeman, N. C., Rosenber.Jm, and Rich, A. (1973) Crystalline fragment of double helix - structure of dinucleoside phosphate guanylyl-3',5'-cytidine, *Proceedings of the National Academy of Sciences of the United States of America* 70, 849-853.
- [11] Rosenberg, J. M., Seeman, N. C., Japarkki.J, Suddath, F. L., Nicholas, H. B., and Rich, A. (1973) Double helix at atomic resolution, *Nature* 243, 150-154.
- [12] Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) self-splicing RNA - auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of tetrahymena, *Cell* 31, 147-157.
- [13] Svoboda, P., and Di Cara, A. (2006) Hairpin RNA: a secondary structure of primary importance, *Cellular and Molecular Life Sciences* 63, 901-918.
- [14] Woese, C. R., Winker, S., and Gutell, R. R. (1990) Architecture of ribosomal-RNA - constraints on the sequence of tetra-loops, *Proceedings of the National Academy of Sciences of the United States of America* 87, 8467-8471.
- [15] Woese, C. R., Gutell, R., Gupta, R., and Noller, H. F. (1983) Detailed analysis of the higher-order structure of 16s-like ribosomal ribonucleic-acids, *Microbiological Reviews* 47, 621-&.
- [16] Varani, G. (1995) Exceptionally stable nucleic-acid hairpins, *Annual Review of Biophysics and Biomolecular Structure* 24, 379-404.
- [17] Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops, *Journal of Molecular Biology* 264, 968-980.
- [18] Davis, J. H., Tonelli, M., Scott, L. G., Jaeger, L., Williamson, J. R., and Butcher, S. E. (2005) RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex, *Journal of Molecular Biology* 351, 371-382.
- [19] Butcher, S. E., Dieckmann, T., and Feigon, J. (1997) Solution structure of a GAAA tetraloop receptor RNA, *Embo Journal* 16, 7490-7499.
- [20] Michel, F., and Westhof, E. (1990) Modeling of the 3-dimensional architecture of group-I catalytic introns based on comparative sequence-analysis, *Journal of Molecular Biology* 216, 585-610.
- [21] Geary, C., Baudrey, S., and Jaeger, L. (2008) Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors, *Nucleic Acids Research* 36, 1138-1152.
- [22] Tanner, M. A., and Cech, T. R. (1995) An important RNA tertiary interaction of group-I and group-II introns is implicated in gram-positive RNase-P RNAs, *Rna-a Publication of the Rna Society* 1, 349-350.

-
- [23] Cheong, H.-K., Kim, N.-K., and Cheong, C. (2015) RNA Structure: Tetraloops, *Encyclopaedia of Life Sciences*.
- [24] van Batenburg, F. H. D., Gulyaev, A. P., and Pleij, C. W. A. (2001) PseudoBase: structural information on RNA pseudoknots, *Nucleic Acids Research* 29, 194-195.
- [25] Han, K. S., and Byun, Y. (2003) PseudoViewer2: visualization of RNA pseudoknots of any type, *Nucleic Acids Research* 31, 3432-3440.
- [26] Rietveld, K., Van Poelgeest, R., Pleij, C. W., Van Boom, J. H., and Bosch, L. (1982) The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA.
- [27] Brierley, I., Pennell, S., and Gilbert, R. J. C. (2007) Viral RNA pseudoknots: versatile motifs in gene expression and replication, *Nature Reviews Microbiology* 5, 598-610.
- [28] Brierley, I., Gilbert, R. J. C., and Pennell, S. (2008) RNA pseudoknots and the regulation of protein synthesis, *Biochemical Society Transactions* 36, 684-689.
- [29] Peselis, A., and Serganov, A. (2014) Structure and function of pseudoknots involved in gene expression control, *Wiley Interdisciplinary Reviews-Rna* 5, 803-822.
- [30] Gilbert, S. D., Rambo, R. P., Van Tyne, D., and Batey, R. T. (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine, *Nature Structural & Molecular Biology* 15, 177-182.
- [31] Schuler, M., Connell, S. R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroeer, B., Mielke, T., Penczek, P. A., Westhof, E., and Spahn, C. M. T. (2006) Structure of the ribosome-bound cricket paralysis virus IRES RNA, *Nature Structural & Molecular Biology* 13, 1092-1096.
- [32] Giedroc, D. P., and Cornish, P. V. (2009) Frameshifting RNA pseudoknots: Structure and mechanism, *Virus Research* 139, 193-208.
- [33] Bang, I. (1910) Untersuchungen über die Guanynsäure, pp 293-311, *Biochemische Zeitschrift*.
- [34] Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) Helix formation by guanylic acid, *Proceedings of the National Academy of Sciences of the United States of America* 48, 2013-2018.
- [35] Hsu, S. T. D., Varnai, P., Bugaut, A., Reszka, A. P., Neidle, S., and Balasubramanian, S. (2009) A G-Rich Sequence within the c-kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics, *Journal of the American Chemical Society* 131, 13399-13409.
- [36] Sen, D., and Gilbert, W. (1988) Formation of parallel 4-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, *Nature* 334, 364-366.
- [37] Sundquist, W. I., and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops, *Nature* 342, 825-829.
- [38] Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure, *Nucleic Acids Research* 34, 5402-5415.
- [39] Parkinson, G. N., Lee, M. P. H., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature* 417, 876-880.
- [40] Wang, Y., and Patel, D. J. (1992) Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized g-quadruplexes with antiglycosidic torsion angles in solution, *Biochemistry* 31, 8112-8119.
- [41] Huppert, J. L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes, *Chemical Society Reviews* 37, 1375-1384.
- [42] Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012) DNA secondary structures: stability and function of G-quadruplex structures, *Nature Reviews Genetics* 13, 770-780.
- [43] Phan, A. T., and Patel, D. J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties, and folding/unfolding kinetics, *Journal of the American Chemical Society* 125, 15021-15027.
- [44] Malgowska, M., Czajczynska, K., Gudanis, D., Tworak, A., and Gdaniec, Z. (2016) Overview of the RNA G-quadruplex structures, *Acta Biochimica Polonica* 63, 609-621.
- [45] Wang, Y., and Patel, D. J. (1993) Solution structure of the human telomeric repeat dAG(3)(T(2)AG(3))₃ G-tetraplex, *Structure* 1, 263-282.
- [46] Guedin, A., Gros, J., Alberti, P., and Mergny, J. L. (2010) How long is too long? Effects of loop size on G-quadruplex stability, *Nucleic Acids Research* 38, 7858-7868.

-
- [47] Mukundan, V. T., and Phan, A. T. (2013) Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences, *Journal of the American Chemical Society* 135, 5017-5028.
- [48] Huppert, J. L. (2010) Structure, location and interactions of G-quadruplexes, *Febs Journal* 277, 3452-3458.
- [49] Zhang, N., Phan, A. T., and Patel, D. J. (2005) (3+1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex, *Journal of the American Chemical Society* 127, 17277-17285.
- [50] Phan, A. T., Luu, K. N., and Patel, D. J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K⁺ solution, *Nucleic Acids Research* 34, 5715-5719.
- [51] Lim, K. W., Amrane, S., Bouaziz, S., Xu, W. X., Mu, Y. G., Patel, D. J., Luu, K. N., and Phan, A. T. (2009) Structure of the Human Telomere in K⁺ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers, *Journal of the American Chemical Society* 131, 4301-4309.
- [52] Phan, A. T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences, *Febs Journal* 277, 1107-1117.
- [53] Smith, F. W., and Feigon, J. (1992) Quadruplex structure of oxytricha telomeric DNA oligonucleotides, *Nature* 356, 164-168.
- [54] Kang, C., Zhang, X. H., Ratliff, R., Moyzis, R., and Rich, A. (1992) Crystal-structure of 4-stranded oxytricha telomeric DNA, *Nature* 356, 126-131.
- [55] Aboulela, F., Murchie, A. I. H., and Lilley, D. M. J. (1992) NMR-study of parallel-stranded tetraplex formation by the hexadeoxynucleotide D(TG4T), *Nature* 360, 280-282.
- [56] Phillips, K., Dauter, Z., Murchie, A. I. H., Lilley, D. M. J., and Luisi, B. (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 angstrom resolution, *Journal of Molecular Biology* 273, 171-182.
- [57] Hardin, C. C., Watson, T., Corregan, M., and Bailey, C. (1992) Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG), *Biochemistry* 31, 833-841.
- [58] Phan, A. T., Kuryavyi, V., and Patel, D. J. (2006) DNA architecture: from G to Z, *Current Opinion in Structural Biology* 16, 288-298.
- [59] Phan, A. T., Modi, Y. S., and Patel, D. J. (2004) Propeller-type parallel-stranded g-quadruplexes in the human c-myc promoter, *Journal of the American Chemical Society* 126, 8710-8716.
- [60] Seenisamy, J., Rezler, E. M., Powell, T. J., Tye, D., Gokhale, V., Joshi, C. S., Siddiqui-Jain, A., and Hurley, L. H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4, *Journal of the American Chemical Society* 126, 8702-8709.
- [61] Ambrus, A., Chen, D., Dai, J. X., Jones, R. A., and Yang, D. Z. (2005) Solution structure of the biologically relevant g-quadruplex element in the human c-MYC promoter. implications for g-quadruplex stabilization, *Biochemistry* 44, 2048-2058.
- [62] Kim, B. G., Evans, H. M., Dubins, D. N., and Chalikian, T. V. (2015) Effects of Salt on the Stability of a G-Quadruplex from the Human c-MYC Promoter, *Biochemistry* 54, 3420-3430.
- [63] Adams, J. M., and Cory, S. (1998) The Bcl-2 protein family: Arbiters of cell survival, *Science* 281, 1322-1326.
- [64] Dai, J. X., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., and Yang, D. Z. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution, *Journal of the American Chemical Society* 128, 1096-1098.
- [65] Dai, J. X., Chen, D., Jones, R. A., Hurley, L. H., and Yang, D. Z. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region, *Nucleic Acids Research* 34, 5133-5144.
- [66] Agrawal, P., Lin, C., Mathad, R. I., Carver, M., and Yang, D. Z. (2014) The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-nt Loop in K⁺ Solution, *Journal of the American Chemical Society* 136, 1750-1753.
- [67] Sun, D. Y., Guo, K. X., Rusche, J. J., and Hurley, L. H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human

- VEGF gene by the presence of potassium and G-quadruplex-interactive agents, *Nucleic Acids Research* 33, 6070-6080.
- [68] Fleming, A. M., Zhou, J., Wallace, S. S., and Burrows, C. J. (2015) A Role for the Fifth G-Track in G-Quadruplex Forming Oncogene Promoter Sequences during Oxidative Stress: Do These "Spare Tires" Have an Evolved Function?, *Acs Central Science* 1, 226-233.
- [69] Sacca, B., Lacroix, L., and Mergny, J. L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides, *Nucleic Acids Research* 33, 1182-1192.
- [70] Joachimi, A., Benz, A., and Hartig, J. S. (2009) A comparison of DNA and RNA quadruplex structures and stabilities, *Bioorganic & Medicinal Chemistry* 17, 6811-6815.
- [71] Cheong, C. J., and Moore, P. B. (1992) Solution structure of an unusually stable RNA tetraplex containing G-quartet and U-quartet structures, *Biochemistry* 31, 8406-8414.
- [72] Tang, C. F., and Shafer, R. H. (2006) Engineering the quadruplex fold: Nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes, *Journal of the American Chemical Society* 128, 5966-5973.
- [73] Collie, G. W., Haider, S. M., Neidle, S., and Parkinson, G. N. (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex, *Nucleic Acids Research* 38, 5569-5580.
- [74] Deng, J. P., Xiong, Y., and Sundaralingam, M. (2001) X-ray analysis of an RNA tetraplex (UGGGGU)(4) with divalent Sr²⁺ ions at subatomic resolution (0.61 angstrom), *Proceedings of the National Academy of Sciences of the United States of America* 98, 13665-13670.
- [75] Fyfe, A. C., Dunten, P. W., Martick, M. M., and Scott, W. G. (2015) Structural Variations and Solvent Structure of r(UGGGGU) Quadruplexes Stabilized by Sr²⁺ Ions, *Journal of Molecular Biology* 427, 2205-2219.
- [76] Zhang, D. H., Fujimoto, T., Saxena, S., Yu, H. Q., Miyoshi, D., and Sugimoto, N. (2010) Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors, *Biochemistry* 49, 4554-4563.
- [77] Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007) Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends, *Science* 318, 798-801.
- [78] Xu, Y., Kaminaga, K., and Komiyama, M. (2008) G-quadruplex formation by human telomeric repeats-containing RNA in Na(+) solution, *Journal of the American Chemical Society* 130, 11179-11184.
- [79] Martadinata, H., and Phan, A. T. (2009) Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric RNA Sequences in K⁺ Solution, *Journal of the American Chemical Society* 131, 2570-2578.
- [80] Collie, G. W., Sparapani, S., Parkinson, G. N., and Neidle, S. (2011) Structural Basis of Telomeric RNA Quadruplex-Acridine Ligand Recognition, *Journal of the American Chemical Society* 133, 2721-2728.
- [81] Randazzo, A., Esposito, V., Ohlenschlager, O., Ramachandran, R., and Mayol, L. (2004) NMR solution structure of a parallel LNA quadruplex, *Nucleic Acids Research* 32, 3083-3092.
- [82] Russo Krauss, I., Parkinson, G. N., Merlino, A., Mattia, C. A., Randazzo, A., Novellino, E., Mazzarella, L., and Sica, F. (2014) A regular thymine tetrad and a peculiar supramolecular assembly in the first crystal structure of an all-LNA G-quadruplex, *Acta Crystallographica Section D-Biological Crystallography* 70, 362-370.
- [83] Martadinata, H., and Phan, A. T. (2013) Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K⁺ Solution, *Biochemistry* 52, 2176-2183.
- [84] Liu, H., Matsugami, A., Katahira, M., and Uesugi, S. (2002) A dimeric RNA quadruplex architecture comprised of two G : G(: A) : G : G(: A) hexads, G : G : G : G tetrads and UUUU loops, *Journal of Molecular Biology* 322, 955-970.
- [85] Mashima, T., Matsugami, A., Nishikawa, F., Nishikawa, S., and Katahira, M. (2009) Unique quadruplex structure and interaction of an RNA aptamer against bovine prion protein, *Nucleic Acids Research* 37, 6249-6258.

-
- [86] Pan, B. C., Xiong, Y., Shi, K., Deng, J. P., and Sundaralingam, M. (2003) Crystal structure of an RNA purine-rich tetraplex containing adenine tetrads: Implications for specific binding in RNA tetraplexes, *Structure* 11, 815-823.
- [87] Pan, B. C., Shi, K., and Sundaralingam, M. (2006) Crystal structure of an RNA quadruplex containing inosine tetrad: Implications for the roles of NH₂ group in purine tetrads, *Journal of Molecular Biology* 363, 451-459.
- [88] Pan, B. C., Xiong, Y., Shi, K., and Sundaralingam, M. (2003) An eight-stranded helical fragment in RNA crystal structure: Implications for tetraplex interaction, *Structure* 11, 825-831.
- [89] Pan, B., Shi, K., and Sundaralingam, M. (2006) Base-tetrad swapping results in dimerization of RNA quadruplexes: Implications for formation of the i-motif RNA octaplex, *Proceedings of the National Academy of Sciences of the United States of America* 103, 3130-3134.
- [90] Pan, B. C., Xiong, Y., Shi, K., and Sundaralingam, M. (2003) Crystal structure of a bulged RNA tetraplex at 1.1 angstrom resolution: Implications for a novel binding site in RNA tetraplex, *Structure* 11, 1423-1430.
- [91] Zahler, A. M., Williamson, J. R., Cech, T. R., and Prescott, D. M. (1991) Inhibition of telomerase by G-quartet DNA structures, *Nature* 350, 718-720.
- [92] Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H. J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo, *Nature Structural & Molecular Biology* 12, 847-854.
- [93] Sun, D. Y., Thompson, B., Cathers, B. E., Salazar, M., Kerwin, S. M., Trent, J. O., Jenkins, T. C., Neidle, S., and Hurley, L. H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound, *Journal of Medicinal Chemistry* 40, 2113-2116.
- [94] Moye, A. L., Porter, K. C., Cohen, S. B., Phan, T., Zyner, K. G., Sasaki, N., Lovrecz, G. O., Beck, J. L., and Bryan, T. M. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase, *Nature Communications* 6, 12.
- [95] Anuradha, S., and Muniyappa, K. (2004) Meiosis-specific yeast Hop1 protein promotes synapsis of double-stranded DNA helices via the formation of guanine quartets, *Nucleic Acids Research* 32, 2378-2385.
- [96] Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA, *Genes & Development* 18, 1618-1629.
- [97] Stavnezer, J., Guikema, J. E. J., and Schrader, C. E. (2008) Mechanism and regulation of class switch recombination, In *Annual Review of Immunology*, pp 261-292, Annual Reviews, Palo Alto.
- [98] Vallur, A. C., and Maizels, N. (2008) Activities of human exonuclease 1 that promote cleavage of transcribed immunoglobulin switch regions, *Proceedings of the National Academy of Sciences of the United States of America* 105, 16508-16512.
- [99] Larson, E. D., Duquette, M. L., Cummings, W. J., Streiff, R. J., and Maizels, N. (2005) MutS alpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions, *Current Biology* 15, 470-474.
- [100] Carvalho, J., Ferreira, J., Pereira, P., Coutinho, E., Guedin, A., Nottelet, P., Salgado, G. E., Mergny, J. L., Queiroz, J. A., Sousa, F., Cabrita, E. J., and Cruz, C. (2016) Stabilization of novel immunoglobulin switch regions G-quadruplexes by naphthalene and quinoline-based ligands, *Tetrahedron* 72, 1229-1237.
- [101] Hansel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., Tannahill, D., and Balasubramanian, S. (2016) G-quadruplex structures mark human regulatory chromatin, *Nature Genetics* 48, 1267-1272.
- [102] De, S., and Michor, F. (2011) DNA secondary structures and epigenetic determinants of cancer genome evolution, *Nature Structural & Molecular Biology* 18, 950-U115.
- [103] Halder, R., Halder, K., Sharma, P., Garg, G., Sengupta, S., and Chowdhury, S. (2010) Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide, *Molecular Biosystems* 6, 2439-2447.
- [104] Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of Cancer: The Next Generation, *Cell* 144, 646-674.

-
- [105] Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome, *Nucleic Acids Research* 35, 406-413.
 - [106] Marcu, K. B., Bossone, S. A., and Patel, A. J. (1992) Myc function and regulation, *Annual Review of Biochemistry* 61, 809-860.
 - [107] Pelengaris, S., Khan, M., and Evan, G. (2002) c-MYC: More than just a matter of life and death, *Nature Reviews Cancer* 2, 764-776.
 - [108] Simonsson, T., Pecinka, P., and Kubista, M. (1998) DNA tetraplex formation in the control region of c-myc, *Nucleic Acids Research* 26, 1167-1172.
 - [109] Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, *Proceedings of the National Academy of Sciences of the United States of America* 99, 11593-11598.
 - [110] Gonzalez, V., and Hurley, L. H. (2010) The c-MYC NHE III1: Function and Regulation, In *Annual Review of Pharmacology and Toxicology*, pp 111-129, Annual Reviews, Palo Alto.
 - [111] Christiansen, J., Kofod, M., and Nielsen, F. C. (1994) A guanosine quadruplex and 2 stable hairpins flank a major cleavage site in insulin-like growth-factor-II messenger-RNA, *Nucleic Acids Research* 22, 5709-5716.
 - [112] Darnell, J. C., Jensen, K. B., Jin, P., Brown, V., Warren, S. T., and Darnell, R. B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function, *Cell* 107, 489-499.
 - [113] Kwok, C. K., and Balasubramanian, S. (2015) Targeted Detection of G-Quadruplexes in Cellular RNAs, *Angewandte Chemie-International Edition* 54, 6751-6754.
 - [114] Millevoi, S., Moine, H., and Vagner, S. (2012) G-quadruplexes in RNA biology, *Wiley Interdisciplinary Reviews-Rna* 3, 495-507.
 - [115] Agarwala, P., Pandey, S., and Maiti, S. (2015) The tale of RNA G-quadruplex, *Organic & Biomolecular Chemistry* 13, 5570-5585.
 - [116] Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation, *Nature Chemical Biology* 3, 218-221.
 - [117] Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., and Kurreck, J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif, *Rna-a Publication of the Rna Society* 14, 1290-1296.
 - [118] Morris, M. J., Negishi, Y., Pazsint, C., Schonhoft, J. D., and Basu, S. (2010) An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES, *Journal of the American Chemical Society* 132, 17831-17839.
 - [119] Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A. C., and Vagner, S. (2003) A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons, *Journal of Biological Chemistry* 278, 39330-39336.
 - [120] Kumari, S., Bugaut, A., and Balasubramanian, S. (2008) Position and Stability Are Determining Factors for Translation Repression by an RNA G-Quadruplex-Forming Sequence within the 5' UTR of the NRAS Proto-oncogene, *Biochemistry* 47, 12664-12669.
 - [121] Huppert, J. L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008) G-quadruplexes: the beginning and end of UTRs, *Nucleic Acids Research* 36, 6260-6268.
 - [122] Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013) Translational halt during elongation caused by G-quadruplex formed by mRNA, *Methods* 64, 73-78.
 - [123] Endoh, T., and Sugimoto, N. (2013) Unusual-1 Ribosomal Frameshift Caused by Stable RNA G-Quadruplex in Open Reading Frame, *Analytical Chemistry* 85, 11435-11439.
 - [124] Arora, A., and Suess, B. (2011) An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation, *Rna Biology* 8, 802-805.
 - [125] Crenshaw, E., Leung, B. P., Kwok, C. K., Sharoni, M., Olson, K., Sebastian, N. P., Ansaloni, S., Schweitzer-Stenner, R., Akins, M. R., Bevilacqua, P. C., and Saunders, A. J. (2015) Amyloid Precursor Protein Translation Is Regulated by a 3' UTR Guanine Quadruplex, *Plos One* 10, 18.
 - [126] Levine, A. J., and Oren, M. (2009) The first 30 years of p53: growing ever more complex, *Nature Reviews Cancer* 9, 749-758.

-
- [127] Decorsiere, A., Cayrel, A., Vagner, S., and Millevoi, S. (2011) Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage, *Genes & Development* 25, 220-225.
 - [128] Beaudoin, J. D., and Perreault, J. P. (2013) Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening, *Nucleic Acids Research* 41, 5898-5911.
 - [129] Min, H. S., Chan, R. C., and Black, D. L. (1995) The generally expressed hnRNP-F is involved in a neural-specific pre-messenger-RNA splicing event, *Genes & Development* 9, 2659-2671.
 - [130] Sirandpugnet, P., Durosay, P., Brody, E., and Marie, J. (1995) An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken beta-tropomyosin pre-messenger-RNA, *Nucleic Acids Research* 23, 3501-3507.
 - [131] Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J. L., and Riou, J. F. (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing, *Nucleic Acids Research* 32, 371-379.
 - [132] Didiot, M. C., Tian, Z. X., Schaeffer, C., Subramanian, M., Mandel, J. L., and Moine, H. (2008) The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer, *Nucleic Acids Research* 36, 4902-4912.
 - [133] Marcel, V., Tran, P. L. T., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M. P., Hall, J., Mergny, J. L., Hainaut, P., and Van Dyck, E. (2011) G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms, *Carcinogenesis* 32, 271-278.
 - [134] Ribeiro, M. M., Teixeira, G. S., Martins, L., Marques, M. R., de Souza, A. P., and Line, S. R. P. (2015) G-quadruplex formation enhances splicing efficiency of PAX9 intron 1, *Human Genetics* 134, 37-44.
 - [135] Fisette, J. F., Montagna, D. R., Mihailescu, M. R., and Wolfe, M. S. (2012) A G-Rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing, *Journal of Neurochemistry* 121, 763-773.
 - [136] Kostadinov, R., Malhotra, N., Viotti, M., Shine, R., D'Antonio, L., and Bagga, P. (2006) GRSDb: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences, *Nucleic Acids Research* 34, D119-D124.
 - [137] Wanrooij, P. H., Uhler, J. P., Simonsson, T., Falkenberg, M., and Gustafsson, C. M. (2010) G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation, *Proceedings of the National Academy of Sciences of the United States of America* 107, 16072-16077.
 - [138] Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J. L., and Moine, H. (2011) G-quadruplex RNA structure as a signal for neurite mRNA targeting, *Embo Reports* 12, 697-704.
 - [139] Ishiguro, A., Kimura, N., Watanabe, Y., Watanabe, S., and Ishihama, A. (2016) TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation, *Genes to Cells* 21, 466-481.
 - [140] Schoeftner, S., and Blasco, M. A. (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II, *Nature Cell Biology* 10, 228-U106.
 - [141] Xu, Y., Suzuki, Y., Ito, K., and Komiyama, M. (2010) Telomeric repeat-containing RNA structure in living cells, *Proceedings of the National Academy of Sciences of the United States of America* 107, 14579-14584.
 - [142] Martadinata, H., Heddi, B., Lim, K. W., and Phan, A. T. (2011) Structure of Long Human Telomeric RNA (TERRA): G-Quadruplexes Formed by Four and Eight UUAGGG Repeats Are Stable Building Blocks, *Biochemistry* 50, 6455-6461.
 - [143] de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres, *Genes & Development* 19, 2100-2110.
 - [144] Biffi, G., Tannahill, D., and Balasubramanian, S. (2012) An Intramolecular G-Quadruplex Structure Is Required for Binding of Telomeric Repeat-Containing RNA to the Telomeric Protein TRF2, *Journal of the American Chemical Society* 134, 11974-11976.

-
- [145] Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., and Lieberman, P. M. (2009) TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres, *Molecular Cell* 35, 403-413.
- [146] Redon, S., Reichenbach, P., and Lingner, J. (2010) The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase, *Nucleic Acids Research* 38, 5797-5806.
- [147] Azhibek, D., Skvortsov, D., Andreeva, A., Zatsepin, T., Arutyunyan, A., Zvereva, M., and Dontsova, O. (2016) TERRA mimicking ssRNAs prevail over the DNA substrate for telomerase in vitro due to interactions with the alternative binding site, *Journal of Molecular Recognition* 29, 242-247.
- [148] Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA, *Nucleic Acids Research* 33, 2901-2907.
- [149] Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome, *Nucleic Acids Research* 33, 2908-2916.
- [150] Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome, *Nature Biotechnology* 33, 877-+.
- [151] Weitzmann, M. N., Woodford, K. J., and Usdin, K. (1996) The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation, *Journal of Biological Chemistry* 271, 20958-20964.
- [152] Bentley, D. R., and Balasubramanian, S., and Swerdlow, H. P., and Smith, G. P., and Milton, J., and Brown, C. G., and Hall, K. P., and Evers, D. J., and Barnes, C. L., and Bignell, H. R., and Boutell, J. M., and Bryant, J., and Carter, R. J., and Cheetham, R. K., and Cox, A. J., and Ellis, D. J., and Flatbush, M. R., and Gormley, N. A., and Humphray, S. J., and Irving, L. J., and Karbelashvili, M. S., and Kirk, S. M., and Li, H., and Liu, X. H., and Maisinger, K. S., and Murray, L. J., and Obradovic, B., and Ost, T., and Parkinson, M. L., and Pratt, M. R., and Rasolonjatovo, I. M. J., and Reed, M. T., and Rigatti, R., and Rodighiero, C., and Ross, M. T., and Sabot, A., and Sankar, S. V., and Scally, A., and Schroth, G. P., and Smith, M. E., and Smith, V. P., and Spiridou, A., and Torrance, P. E., and Tzonev, S. S., and Vermaas, E. H., and Walter, K., and Wu, X. L., and Zhang, L., and Alam, M. D., and Anastasi, C., and Aniebo, I. C., and Bailey, D. M. D., and Bancarz, I. R., and Banerjee, S., and Barbour, S. G., and Baybayan, P. A., and Benoit, V. A., and Benson, K. F., and Bevis, C., and Black, P. J., and Boodhun, A., and Brennan, J. S., and Bridgham, J. A., and Brown, R. C., and Brown, A. A., and Buermann, D. H., and Bundu, A. A., and Burrows, J. C., and Carter, N. P., and Castillo, N., and Catenazzi, M. C. E., and Chang, S., and Cooley, R. N., and Crake, N. R., and Dada, O. O., and Diakoumakos, K. D., and Dominguez-Fernandez, B., and Earnshaw, D. J., and Egbujor, U. C., and Elmore, D. W., and Etchin, S. S., and Ewan, M. R., and Fedurco, M., and Fraser, L. J., and Fajardo, K. V. F., and Furey, W. S., and George, D., and Gietzen, K. J., and Goddard, C. P., and Golda, G. S., and Granieri, P. A., and Green, D. E., and Gustafson, D. L., and Hansen, N. F., and Harnish, K., and Haudenschild, C. D., and Heyer, N. I., and Hims, M. M., and Ho, J. T., and Horgan, A. M., and Hoschler, K., and Hurwitz, S., and Ivanov, D. V., and Johnson, M. Q., and James, T., and Jones, T. A. H., and Kang, G. D., and Kerelska, T. H., and Kersey, A. D., and Khrebtukova, I., and Kindwall, A. P., and Kingsbury, Z., and Kokko-Gonzales, P. I., and Kumar, A., and Laurent, M. A., and Lawley, C. T., and Lee, S. E., and Lee, X., and Liao, A. K., and Loch, J. A., and Lok, M., and Luo, S. J., and Mammen, R. M., and Martin, J. W., and McCauley, P. G., and McNitt, P., and Mehta, P., and Moon, K. W., and Mullens, J. W., and Newton, T., and Ning, Z. M., and Ng, B. L., and Novo, S. M., and O'Neill, M. J., and Osborne, M. A., and Osnowski, A., and Ostadan, O., and Paraschos, L. L., and Pickering, L., and Pike, A. C., and Pinkard, D. C., and Pliskin, D. P., and Podhasky, J., and Quijano, V. J., and Racz, C., and Rae, V. H., and Rawlings, S. R., and Rodriguez, A. C., and Roe, P. M., and Rogers, J., and Bacigalupo, M. C. R., and Romanov, N., and Romieu, A., and Roth, R. K., and Rourke, N. J., and Ruediger, S. T., and Rusman, E., and Sanches-Kuiper, R. M., and Schenker, M. R., and Seoane, J. M., and Shaw, R. J., and Shiver, M. K., and Short, S. W., and Sizto, N. L., and Sluis, J. P., and Smith, M. A., and Sohna, J. E. S., and Spence, E. J., and Stevens, K., and Sutton, N., and Szajkowski, L., and Tregidgo, C. L., and Turcatti, G., and vandeVondele, S., and Verhovsky, Y., and Virk, S. M., and Wakelin, S., and Walcott, G. C., and Wang, J. W.,

- and Worsley, G. J., and Yan, J. Y., and Yau, L., and Zuerlein, M., and Mullikin, J. C., and Hurles, M. E., and McCooke, N. J., and West, J. S., and Oaks, F. L., and Lundberg, P. L., and Klenerman, D., and Durbin, R., and Smith, A. J. (2008) Accurate whole human genome sequencing using reversible terminator chemistry, *Nature* 456, 53-59.
- [153] Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., Lawrence, M. S., Zhang, C. Z., Wala, J., Mermel, C. H., Sougnez, C., Gabriel, S. B., Hernandez, B., Shen, H., Laird, P. W., Getz, G., Meyerson, M., and Beroukhi, R. (2013) Pan-cancer patterns of somatic copy number alteration, *Nature Genetics* 45, 1134-U1257.
- [154] Capra, J. A., Paeschke, K., Singh, M., and Zakian, V. A. (2010) G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in *Saccharomyces cerevisiae*, *Plos Computational Biology* 6, 13.
- [155] Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei, *Proceedings of the National Academy of Sciences of the United States of America* 98, 8572-8577.
- [156] Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells, *Nature Chemistry* 5, 182-186.
- [157] Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells, *Nature Chemistry* 6, 75-80.
- [158] Henderson, A., Wu, Y. L., Huang, Y. C., Chavez, E. A., Platt, J., Johnson, F. B., Brosh, R. M., Sen, D., and Lansdorp, P. M. (2014) Detection of G-quadruplex DNA in mammalian cells, *Nucleic Acids Research* 42, 860-869.
- [159] Laguerre, A., Hukezalie, K., Winckler, P., Katranji, F., Chanteloup, G., Pirrotta, M., Perrier-Cornet, J. M., Wong, J. M. Y., and Monchaud, D. (2015) Visualization of RNA-Quadruplexes in Live Cells, *Journal of the American Chemical Society* 137, 8521-8525.
- [160] Liu, H. H., Zheng, K. W., He, Y. D., Chen, Q., Hao, Y. H., and Tan, Z. (2016) RNA G-quadruplex formation in defined sequence in living cells detected by bimolecular fluorescence complementation, *Chemical Science* 7, 4573-4581.
- [161] Neidle, S. (2010) Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer, *Febs Journal* 277, 1118-1125.
- [162] Baral, A., Kumar, P., Halder, R., Mani, P., Yadav, V. K., Singh, A., Das, S. K., and Chowdhury, S. (2012) Quadruplex-single nucleotide polymorphisms (Quad-SNP) influence gene expression difference among individuals, *Nucleic Acids Research* 40, 3800-3811.
- [163] Huang, H. L., Zhang, J., Harvey, S. E., Hu, X. H., and Cheng, C. H. (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF, *Genes & Development* 31, 2296-2309.
- [164] Cheung, I., Schertzer, M., Rose, A., and Lansdorp, P. M. (2002) Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA, *Nature Genetics* 31, 405-409.
- [165] Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D. B., Cuppen, E., and Tijsterman, M. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C-elegans*, *Current Biology* 18, 900-905.
- [166] Paeschke, K., Capra, J. A., and Zakian, V. A. (2011) DNA Replication through G-Quadruplex Motifs Is Promoted by the *Saccharomyces cerevisiae* Pif1 DNA Helicase, *Cell* 145, 678-691.
- [167] Crawford, D. C., Acuna, J. M., and Sherman, S. L. (2001) FMR1 and the fragile X syndrome: Human genome epidemiology review, *Genetics in Medicine* 3, 359-371.
- [168] Fry, M., and Loeb, L. A. (1994) The fragile-X syndrome d(CGG)(n) nucleotide repeats form a stable tetrahelical structure, *Proceedings of the National Academy of Sciences of the United States of America* 91, 4950-4954.
- [169] Kettani, A., Kumar, R. A., and Patel, D. J. (1995) Solution structure of a DNA quadruplex containing the fragile-X syndrome triplet repeat, *Journal of Molecular Biology* 254, 638-656.
- [170] Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif, *Embo Journal* 20, 4803-4813.

-
- [171] Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A. J., Shaw, P. G., Kim, M. S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., and Wang, J. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease, *Nature* 507, 195-+.
- [172] Balasubramanian, S., Hurley, L. H., and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?, *Nature Reviews Drug Discovery* 10, 261-275.
- [173] Murat, P., Zhong, J., Lekieffre, L., Cowieson, N. P., Clancy, J. L., Preiss, T., Balasubramanian, S., Khanna, R., and Tellam, J. (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation, *Nature Chemical Biology* 10, 358-NIL_364.
- [174] Do, N. Q., Lim, K. W., Teo, M. H., Heddi, B., and Phan, A. T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity, *Nucleic Acids Research* 39, 9448-9457.
- [175] Amrane, S., Kerkour, A., Bedrat, A., Vialet, B., Andreola, M. L., and Mergny, J. L. (2014) Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development, *Journal of the American Chemical Society* 136, 5249-5252.
- [176] Perrone, R., Doria, F., Butovskaya, E., Frasson, I., Botti, S., Scalabrin, M., Lago, S., Grande, V., Nadai, M., Freccero, M., and Richter, S. N. (2015) Synthesis, Binding and Antiviral Properties of Potent Core-Extended Naphthalene Diimides Targeting the HIV-1 Long Terminal Repeat Promoter G-Quadruplexes, *Journal of Medicinal Chemistry* 58, 9639-9652.
- [177] Perrone, R., Butovskaya, E., Daelemans, D., Palu, G., Pannecouque, C., and Richter, S. N. (2014) Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19, *Journal of Antimicrobial Chemotherapy* 69, 3248-3258.
- [178] Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C. B., Proffitt, C., Trent, K., Whitten, J. P., Lim, J. K. C., Von Hoff, D., Anderes, K., and Rice, W. G. (2009) Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis, *Cancer Research* 69, 7653-7661.
- [179] Arabi, A., Wu, S. Q., Ridderstrale, K., Bierhoff, H., Shiue, C., Fatyol, K., Fahlen, S., Hydbring, P., Soderberg, O., Grummt, I., Larsson, L. G., and Wright, A. P. H. (2005) c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription, *Nature Cell Biology* 7, 303-310.
- [180] Rickards, B., Flint, S. J., Cole, M. D., and LeRoy, G. (2007) Nucleolin is required for RNA polymerase I transcription in vivo, *Molecular and Cellular Biology* 27, 937-948.
- [181] Storck, S., Thiry, M., and Bouvet, P. (2009) Conditional knockout of nucleolin in DT40 cells reveals the functional redundancy of its RNA-binding domains, *Biology of the Cell* 101, 153-167.
- [182] Hanakahi, L. A., Sun, H., and Maizels, N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA, *Journal of Biological Chemistry* 274, 15908-15912.
- [183] Monchaud, D., and Teulade-Fichou, M. P. (2008) A hitchhiker's guide to G-quadruplex ligands, *Organic & Biomolecular Chemistry* 6, 627-636.
- [184] Muller, S., and Rodriguez, R. (2014) G-quadruplex interacting small molecules and drugs: from bench toward bedside, *Expert Review of Clinical Pharmacology* 7, 663-679.
- [185] Guo, Q., Lu, M., Marky, L. A., and Kallenbach, N. R. (1992) Interaction of the dye ethidium-bromide with DNA containing guanine repeats, *Biochemistry* 31, 2451-2455.
- [186] Chen, Q., Kuntz, I. D., and Shafer, R. H. (1996) Spectroscopic recognition of guanine dimeric hairpin quadruplexes by a carbocyanine dye, *Proceedings of the National Academy of Sciences of the United States of America* 93, 2635-2639.
- [187] Searle, M. S., and Balkwill, G. D. (2006) DNA Quadruplex-Ligand Recognition: Structure and Dynamics, In *Quadruplex Nucleic Acids* (Neidle, S., and Balasubramanian, S., Eds.), pp 131-153, The Royal Society of Chemistry.
- [188] Collie, G. W., Promontorio, R., Hampel, S. M., Micco, M., Neidle, S., and Parkinson, G. N. (2012) Structural Basis for Telomeric G-Quadruplex Targeting by Naphthalene Diimide Ligands, *Journal of the American Chemical Society* 134, 2723-2731.
- [189] Hampel, S. M., Sidibe, A., Gunaratnam, M., Riou, J. F., and Neidle, S. (2010) Tetrasubstituted naphthalene diimide ligands with selectivity for telomeric G-quadruplexes and cancer cells, *Bioorganic & Medicinal Chemistry Letters* 20, 6459-6463.

- [190] Bugaut, A., Rodriguez, R., Kumari, S., Hsu, S. T. D., and Balasubramanian, S. (2010) Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex, *Organic & Biomolecular Chemistry* 8, 2771-2776.
- [191] Morris, M. J., Wingate, K. L., Silwal, J., Leeper, T. C., and Basu, S. (2012) The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells, *Nucleic Acids Research* 40, 4137-4145.
- [192] Faudale, M., Cogoi, S., and Xodo, L. E. (2012) Photoactivated cationic alkyl-substituted porphyrin binding to g4-RNA in the 5' -UTR of KRAS oncogene represses translation, *Chemical Communications* 48, 874-876.
- [193] Halder, K., Largy, E., Benzler, M., Teulade-Fichou, M. P., and Hartig, J. S. (2011) Efficient Suppression of Gene Expression by Targeting 5' -UTR-Based RNA Quadruplexes with Bisquinolinium Compounds, *Chembiochem* 12, 1663-1668.
- [194] Gomez, D., Guedin, A., Mergny, J. L., Salles, B., Riou, J. F., Teulade-Fichou, M. P., and Calsou, P. (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells, *Nucleic Acids Research* 38, 7187-7198.
- [195] Shin-ya, K., Wierzbza, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Hayakawa, Y., and Seto, H. (2001) Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*, *Journal of the American Chemical Society* 123, 1262-1263.
- [196] Kim, M. Y., Vankayalapati, H., Kazuo, S., Wierzbza, K., and Hurley, L. H. (2002) Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex, *Journal of the American Chemical Society* 124, 2098-2099.
- [197] Tauchi, T., Shin-ya, K., Sashida, G., Sumi, M., Okabe, S., Ohyashiki, J. H., and Ohyashiki, K. (2006) Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia, *Oncogene* 25, 5719-5725.
- [198] Teulade-Fichou, M. P., Carrasco, C., Guittat, L., Bailly, C., Alberti, P., Mergny, J. L., David, A., Lehn, J. M., and Wilson, W. D. (2003) Selective recognition of G-quadruplex telomeric DNA by a bis(quinacridine) macrocycle, *Journal of the American Chemical Society* 125, 4732-4740.
- [199] Alberti, P., Ren, J. S., Teulade-Fichou, M. P., Guittat, L., Riou, J. F., Chaires, J. B., Helene, C., Vigneron, J. P., Lehn, J. M., and Mergny, J. L. (2001) Interaction of an acridine dimer with DNA quadruplex structures, *Journal of Biomolecular Structure & Dynamics* 19, 505-513.
- [200] Jantos, K., Rodriguez, R., Ladame, S., Shirude, P. S., and Balasubramanian, S. (2006) Oxazole-based peptide macrocycles: A new class of G-quadruplex binding ligands, *Journal of the American Chemical Society* 128, 13662-13663.
- [201] Han, H. Y., Langley, D. R., Rangan, A., and Hurley, L. H. (2001) Selective interactions of cationic porphyrins with G-quadruplex structures, *Journal of the American Chemical Society* 123, 8902-8913.
- [202] Jain, A. K., and Bhattacharya, S. (2011) Interaction of G-Quadruplexes with Nonintercalating Duplex-DNA Minor Groove Binding Ligands, *Bioconjugate Chemistry* 22, 2355-2368.
- [203] Ilyinsky, N. S., Varizhuk, A. M., Beniaminov, A. D., Puzanov, M. A., Shcheyolkina, A. K., and Kaluzhny, D. N. (2014) G-quadruplex ligands: Mechanisms of anticancer action and target binding, *Molecular Biology* 48, 778-794.
- [204] Randazzo, A., Galeone, A., and Mayol, L. (2001) H-1-NMR study of the interaction of distamycin A and netropsin with the parallel stranded tetraplex d(TGGGGT) (4), *Chemical Communications*, 1030-1031.
- [205] Cosconati, S., Marinelli, L., Trotta, R., Virno, A., De Tito, S., Romagnoli, R., Pagano, B., Limongelli, V., Giancola, C., Baraldi, P. G., Mayol, L., Novellino, E., and Randazzo, A. (2010) Structural and Conformational Requisites in DNA Quadruplex Groove Binding: Another Piece to the Puzzle, *Journal of the American Chemical Society* 132, 6425-6433.
- [206] Martino, L., Virno, A., Pagano, B., Virgilio, A., Di Micco, S., Galeone, A., Giancola, C., Bifulco, G., Mayol, L., and Randazzo, A. (2007) Structural and thermodynamic studies of the interaction of distamycin a with the parallel quadruplex structure d(TGGGGT) (4), *Journal of the American Chemical Society* 129, 16048-16056.
- [207] Cocco, M. J., Hanakahi, L. A., Huber, M. D., and Maizels, N. (2003) Specific interactions of distamycin with G-quadruplex DNA, *Nucleic Acids Research* 31, 2944-2951.

-
- [208] Li, J. L., Fu, J., Wang, J. P., Hu, D. H., Su, Z. M., and Jin, X. Q. (2012) Molecular dynamic and quantum mechanics study of drug recognition for the extremity of DNA G-quadruplex groove, *Medicinal Chemistry Research* 21, 4010-4016.
- [209] Li, Q., Xiang, J. F., Li, X. D., Chen, L. R., Xu, X. J., Tang, Y. L., Zhou, Q. J., Li, L., Zhang, H., Sun, H. X., Guan, A. J., Yang, Q. F., Yang, S., and Xu, G. Z. (2009) Stabilizing parallel G-quadruplex DNA by a new class of ligands: Two non-planar alkaloids through interaction in lateral grooves, *Biochimie* 91, 811-819.
- [210] Li, J. L., Jin, X. Q., Hu, L. H., Wang, J. P., and Su, Z. M. (2011) Identification of nonplanar small molecule for G-quadruplex grooves: Molecular docking and molecular dynamic study, *Bioorganic & Medicinal Chemistry Letters* 21, 6969-6972.
- [211] Brassart, B., Gomez, D., De Cian, A., Paterski, R., Montagnac, A., Qui, K. H., Temime-Smaali, N., Trentesaux, C., Mergny, J. L., Gueritte, F. O., and Riou, J. F. (2007) A new steroid derivative stabilizes g-quadruplexes and induces telomere uncapping in human tumor cells, *Molecular Pharmacology* 72, 631-640.
- [212] White, E. W., Tanious, F., Ismail, M. A., Reszka, A. P., Neidle, S., Boykin, D. W., and Wilson, W. D. (2007) Structure-specific recognition of quadruplex DNA by organic cations: Influence of shape, substituents and charge, *Biophysical Chemistry* 126, 140-153.
- [213] Campbell, N. H., Parkinson, G. N., Reszka, A. P., and Neidle, S. (2008) Structural basis of DNA quadruplex recognition by an acridine drug, *Journal of the American Chemical Society* 130, 6722-+.
- [214] Ladame, S., Schouten, J. A., Stuart, J., Roldan, J., Neidle, S., and Balasubramanian, S. (2004) Tetrapeptides induce selective recognition for G-quadruplexes when conjugated to a DNA-binding platform, *Organic & Biomolecular Chemistry* 2, 2925-2931.
- [215] Parkinson, G. N., Ghosh, R., and Neidle, S. (2007) Structural basis for binding of porphyrin to human telomeres, *Biochemistry* 46, 2390-2397.
- [216] Hamon, F., Largy, E., Guedin-Beaurepaire, A., Rouchon-Dagois, M., Sidibe, A., Monchaud, D., Mergny, J. L., Riou, J. F., Nguyen, C. H., and Teulade-Fichou, M. P. (2011) An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-Quadruplex Topologies, *Angewandte Chemie-International Edition* 50, 8745-8749.
- [217] Tandon, V. (2015) Novel smart C2-symmetrical NDI derivatives as G-quadruplex stabilizing ligand with a potential to differentiate between topological structures, *Journal of Biomolecular Structure & Dynamics* 33, 79-79.
- [218] Kerwin, S. M. (2000) G-quadruplex DNA as a target for drug design, *Current Pharmaceutical Design* 6, 441-471.
- [219] Parkinson, G. N., Cuenca, F., and Neidle, S. (2008) Topology conservation and loop flexibility in quadruplex-drug recognition: Crystal structures of inter- and intramolecular telomeric DNA quadruplex-drug complexes, *Journal of Molecular Biology* 381, 1145-1156.
- [220] Collie, G. W., Parkinson, G. N., Neidle, S., Rosu, F., De Pauw, E., and Gabelica, V. (2010) Electrospray Mass Spectrometry of Telomeric RNA (TERRA) Reveals the Formation of Stable Multimeric G-Quadruplex Structures, *Journal of the American Chemical Society* 132, 9328-9334.
- [221] Di Antonio, M., Biffi, G., Mariani, A., Raiber, E. A., Rodriguez, R., and Balasubramanian, S. (2012) Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry, *Angewandte Chemie-International Edition* 51, 11073-11078.
- [222] Zimmerman, S. B., and Minton, A. P. (1993) Macromolecular crowding - biochemical, biophysical, and physiological consequences, *Annual Review of Biophysics and Biomolecular Structure* 22, 27-65.
- [223] Kurreck, J. (2003) Antisense technologies - Improvement through novel chemical modifications, *European Journal of Biochemistry* 270, 1628-1644.
- [224] Liang, X.-H., Shen, W., Sun, H., Migawa, M. T., Vickers, T. A., and Croke, S. T. (2016) Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames, pp 875-880, *Nature Biotechnology*.
- [225] Bonifacio, L., Church, F. C., and Jarstfer, M. B. (2008) Effect of locked-nucleic acid on a biologically active G-quadruplex. A structure-activity relationship of the thrombin aptamer, *International Journal of Molecular Sciences* 9, 422-433.

- [226] Nielsen, P. E., Egholm, M., Berg, R. H., and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide, *Science* 254, 1497-1500.
- [227] Haaime, G., Lohse, A., Buchardt, O., and Nielsen, P. E. (1996) Peptide nucleic acids (PNAs) containing thymine monomers derived from chiral amino acids: Hybridization and solubility properties of D-lysine PNA, *Angewandte Chemie-International Edition* 35, 1939-1942.
- [228] Sforza, S., Corradini, R., Ghirardi, S., Dossena, A., and Marchelli, R. (2000) DNA binding of a D-lysine-based chiral PNA: Direction control and mismatch recognition, *European Journal of Organic Chemistry*, 2905-2913.
- [229] Menchise, V., De Simone, G., Tedeschi, T., Corradini, R., Sforza, S., Marchelli, R., Capasso, D., Saviano, M., and Pedone, C. (2003) Insights into peptide nucleic acid (PNA) structural features: The crystal structure of a D-lysine-based chiral PNA-DNA duplex, *Proceedings of the National Academy of Sciences of the United States of America* 100, 12021-12026.
- [230] Dragulescu-Andrasi, A., Zhou, P., He, G. F., and Ly, D. H. (2005) Cell-permeable GPNA with appropriate backbone stereochemistry and spacing binds sequence-specifically to RNA, *Chemical Communications*, 244-246.
- [231] Zhou, P., Dragulescu-Andrasi, A., Bhattacharya, B., O'Keefe, H., Vatta, P., Hyldig-Nielsen, J. J., and Ly, D. H. (2006) Synthesis of cell-permeable peptide nucleic acids and characterization of their hybridization and uptake properties, *Bioorganic & Medicinal Chemistry Letters* 16, 4931-4935.
- [232] Dragulescu-Andrasi, A., Rapireddy, S., Frezza, B. M., Gayathri, C., Gil, R. R., and Ly, D. H. (2006) A simple gamma-backbone modification preorganizes peptide nucleic acid into a helical structure, *Journal of the American Chemical Society* 128, 10258-10267.
- [233] Rapireddy, S., He, G., Roy, S., Armitage, B. A., and Ly, D. H. (2007) Strand invasion of mixed-sequence B-DNA by acridine-linked, gamma-peptide nucleic acid (gamma-PNA), *Journal of the American Chemical Society* 129, 15596-15600.
- [234] Ganesh, K. N., and Nielsen, P. E. (2000) Peptide nucleic acids: Analogs and derivatives, *Current Organic Chemistry* 4, 931-943.
- [235] Pokorski, J. K., Witschi, M. A., Purnell, B. L., and Appella, D. H. (2004) (S,S)-trans-cyclopentane-constrained peptide nucleic acids. A general backbone modification that improves binding affinity and sequence specificity, *Journal of the American Chemical Society* 126, 15067-15073.
- [236] Vilaivan, T., and Srisuwannaket, C. (2006) Hybridization of pyrrolidinyl peptide nucleic acids and DNA: Selectivity, base-pairing specificity, and direction of binding, *Organic Letters* 8, 1897-1900.
- [237] Dekoning, H., and Pandit, U. K. (1971) Unconventional nucleotide analogues .6. synthesis of purinyl-peptides and pyrimidinyl-peptides, *Recueil Des Travaux Chimiques Des Pays-Bas* 90, 1069-&.
- [238] Sahu, B., Sacui, I., Rapireddy, S., Zanolli, K. J., Bahal, R., Armitage, B. A., and Ly, D. H. (2011) Synthesis and Characterization of Conformationally Preorganized, (R)-Diethylene Glycol-Containing gamma-Peptide Nucleic Acids with Superior Hybridization Properties and Water Solubility, *Journal of Organic Chemistry* 76, 5614-5627.
- [239] Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B., and Nielsen, P. E. (1993) PNA hybridizes to complementary oligonucleotides obeying the watson-crick hydrogen-bonding rules, *Nature* 365, 566-568.
- [240] Demidov, V. V., Potaman, V. N., Frankkamenetskii, M. D., Egholm, M., Buchard, O., Sonnichsen, S. H., and Nielsen, P. E. (1994) Stability of peptide nucleic-acids in human serum and cellular-extracts, *Biochemical Pharmacology* 48, 1310-1313.
- [241] Egholm, M., Nielsen, P. E., Buchardt, O., and Berg, R. H. (1992) Recognition of guanine and adenine in DNA by cytosine and thymine containing peptide nucleic-acids (PNA), *Journal of the American Chemical Society* 114, 9677-9678.
- [242] Knudsen, H., and Nielsen, P. E. (1996) Antisense properties of duplex- and triplex-forming PNAs, *Nucleic Acids Research* 24, 494-500.

-
- [243] Egholm, M., Christensen, L., Dueholm, K. L., Buchardt, O., Coull, J., and Nielsen, P. E. (1995) Efficient pH-independent sequence-specific DNA-binding by pseudoisocytosine-containing bis-PNA, *Nucleic Acids Research* 23, 217-222.
- [244] Kim, N. K., Zhang, Q., Zhou, J., Theimer, C. A., Peterson, R. D., and Feigon, J. (2008) Solution Structure and Dynamics of the Wild-type Pseudoknot of Human Telomerase RNA, *Journal of Molecular Biology* 384, 1249-1261.
- [245] Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: The A-minor motif, *Proceedings of the National Academy of Sciences of the United States of America* 98, 4899-4903.
- [246] Devi, G., Zhou, Y., Zhong, Z. S., Toh, D. F. K., and Chen, G. (2015) RNA triplexes: from structural principles to biological and biotech applications, *Wiley Interdisciplinary Reviews-Rna* 6, 111-128.
- [247] Moses, A. C., Huang, S. W., and Schepartz, A. (1997) Inhibition of ReV center dot RRE complexation by triplex tethered oligonucleotide probes, *Bioorganic & Medicinal Chemistry* 5, 1123-1129.
- [248] Devi, G., Yuan, Z., Lu, Y. P., Zhao, Y. L., and Chen, G. (2014) Incorporation of thio-pseudoisocytosine into triplex-forming peptide nucleic acids for enhanced recognition of RNA duplexes, *Nucleic Acids Research* 42, 4008-4018.
- [249] Escude, C., Francois, J. C., Sun, J. S., Ott, G., Sprinzl, M., Garestier, T., and Helene, C. (1993) Stability of triple helices containing RNA and DNA strands - experimental and molecular modelling studies, *Nucleic Acids Research* 21, 5547-5553.
- [250] Endoh, T., Hnedzko, D., Rozners, E., and Sugimoto, N. (2016) Nucleobase-Modified PNA Suppresses Translation by Forming a Triple Helix with a Hairpin Structure in mRNA InVitro and in Cells, *Angewandte Chemie-International Edition* 55, 899-903.
- [251] Kumar, V., Kesavan, V., and Gothelf, K. V. (2015) Highly stable triple helix formation by homopyrimidine (L)-acyclic threoninol nucleic acids with single stranded DNA and RNA, *Organic & Biomolecular Chemistry* 13, 2366-2374.
- [252] Panyutin, I. G., Onyshchenko, M. I., Englund, E. A., Appella, D. H., and Neumann, R. D. (2012) Targeting DNA G-Quadruplex Structures with Peptide Nucleic Acids, *Current Pharmaceutical Design* 18, 1984-1991.
- [253] Kumar, N., Patowary, A., Sivasubbu, S., Petersen, M., and Maiti, S. (2008) Silencing c-MYC Expression by Targeting Quadruplex in P1 Promoter Using Locked Nucleic Acid Trap, *Biochemistry* 47, 13179-13188.
- [254] Onyshchenko, M. I., Gaynutdinov, T. I., Englund, E. A., Appella, D. H., Neumann, R. D., and Panyutin, I. G. (2009) Stabilization of G-quadruplex in the BCL2 promoter region in double-stranded DNA by invading short PNAs, *Nucleic Acids Research* 37, 7570-7580.
- [255] Onyshchenko, M. I., Gaynutdinov, T. I., Englund, E. A., Appella, D. H., Neumann, R. D., and Panyutin, I. G. (2011) Quadruplex formation is necessary for stable PNA invasion into duplex DNA of BCL2 promoter region, *Nucleic Acids Research* 39, 7114-7123.
- [256] Englund, E. A., Gupta, P., Micklitsch, C. M., Onyshchenko, M. I., Remeeva, E., Neumann, R. D., Panyutin, I. G., and Appella, D. H. (2014) PPG Peptide Nucleic Acids that Promote DNA Guanine Quadruplexes, *Chembiochem* 15, 1887-1890.
- [257] Rouleau, S. G., Beaudoin, J. D., Bisailon, M., and Perreault, J. P. (2015) Small antisense oligonucleotides against G-quadruplexes: specific mRNA translational switches, *Nucleic Acids Research* 43, 595-606.
- [258] Ito, K., Go, S., Komiyama, M., and Xu, Y. (2011) Inhibition of Translation by Small RNA-Stabilized mRNA Structures in Human Cells, *Journal of the American Chemical Society* 133, 19153-19159.
- [259] Bhattacharyya, D., Nguyen, K., and Basu, S. (2014) Rationally Induced RNA:DNA G-Quadruplex Structures Elicit an Anticancer Effect by Inhibiting Endogenous eIF-4E Expression, *Biochemistry* 53, 5461-5470.
- [260] Gaynutdinov, T. I., Englund, E. A., Appella, D. H., Onyshchenko, M. I., Neumann, R. D., and Panyutin, I. G. (2015) G-Quadruplex Formation Between G-Rich PNA and Homologous Sequences in Oligonucleotides and Supercoiled Plasmid DNA, *Nucleic Acid Therapeutics* 25, 78-84.

-
- [261] Datta, B., and Armitage, B. A. (2001) Hybridization of PNA to structured DNA targets: Quadruplex invasion and the overhang effect, *Journal of the American Chemical Society* 123, 9612-9619.
- [262] Amato, J., Pagano, B., Borbone, N., Oliviero, G., Gabelica, V., De Pauw, E., D'Errico, S., Piccialli, V., Varra, M., Giancola, C., Piccialli, G., and Mayol, L. (2011) Targeting G-Quadruplex Structure in the Human c-Kit Promoter with Short PNA Sequences, *Bioconjugate Chemistry* 22, 654-663.
- [263] Lusvardi, S., Murphy, C. T., Roy, S., Tanious, F. A., Sacui, I., Wilson, W. D., Ly, D. H., and Armitage, B. A. (2009) Loop and Backbone Modifications of Peptide Nucleic Acid Improve G-Quadruplex Binding Selectivity, *Journal of the American Chemical Society* 131, 18415-18424.
- [264] Murphy, C. T., Gupta, A., Armitage, B. A., and Opresko, P. L. (2014) Hybridization of G-Quadruplex-Forming Peptide Nucleic Acids to Guanine-Rich DNA Templates Inhibits DNA Polymerase α Extension, *Biochemistry* 53, 5315-5322.
- [265] Oyaghire, S. N., Cherubim, C. J., Telmer, C. A., Martinez, J. A., Bruchez, M. P., and Armitage, B. A. (2016) RNA G-Quadruplex Invasion and Translation Inhibition by Antisense gamma-Peptide Nucleic Acid Oligomers, *Biochemistry* 55, 1977-1988.
- [266] Duncley, M. G., Manoharan, M., Villiet, P., Eperon, I. C., and Dickson, G. (1998) Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides, *Human Molecular Genetics* 7, 1083-1090.
- [267] Lebedeva, I., and Stein, C. A. (2001) Antisense oligonucleotides: Promise and reality, *Annual Review of Pharmacology and Toxicology* 41, 403-419.
- [268] Yin, H. F., Lu, Q. L., and Wood, M. (2008) Effective exon skipping and restoration of dystrophin expression by peptide nucleic acid antisense oligonucleotides in mdx mice, *Molecular Therapy* 16, 38-45.
- [269] Gao, X. J., Shen, X. Y., Dong, X., Ran, N., Han, G., Cao, L. M., Gu, B., and Yin, H. F. (2015) Peptide Nucleic Acid Promotes Systemic Dystrophin Expression and Functional Rescue in Dystrophin-deficient mdx Mice, *Molecular Therapy-Nucleic Acids* 4, 8.
- [270] Sazani, P., Gemignani, F., Kang, S. H., Maier, M. A., Manoharan, M., Persmark, M., Bortner, D., and Kole, R. (2002) Systemically delivered antisense oligomers upregulate gene expression in mouse tissues, *Nature Biotechnology* 20, 1228-1233.
- [271] Hamzavi, R., Dolle, F., Tavitian, B., Dahl, O., and Nielsen, P. E. (2003) Modulation of the pharmacokinetic properties of PNA: Preparation of galactosyl, mannosyl, fucosyl, N-acetylgalactosaminyl, and N-acetylglucosaminyl derivatives of aminoethylglycine peptide nucleic acid monomers and their incorporation into PNA oligomers, *Bioconjugate Chemistry* 14, 941-954.
- [272] Sun, J. S., Francois, J. C., Montenaygarestier, T., Saisonbehmoaras, T., Roig, V., Thuong, N. T., and Helene, C. (1989) Sequence-specific intercalating agents - intercalation at specific sequences on duplex DNA via major groove recognition by oligonucleotide intercalator conjugates, *Proceedings of the National Academy of Sciences of the United States of America* 86, 9198-9202.
- [273] Rusling, D. A., and Fox, K. R. (2012) Chapter 3 Small Molecule-Oligonucleotide Conjugates, In *DNA Conjugates and Sensors*, pp 75-102, The Royal Society of Chemistry.
- [274] Bentin, T., and Nielsen, P. E. (2003) Superior duplex DNA strand invasion by acridine conjugated peptide nucleic acids, *Journal of the American Chemical Society* 125, 6378-6379.
- [275] Nielsen, P. E., Frederiksen, K., and Behrens, C. (2005) Extended target sequence specificity of PNA-minor-groove binder conjugates, *Chembiochem* 6, 66-68.
- [276] Mokhir, A. A., and Kraemer, R. (2003) Conjugates of PNA with naphthalene diimide derivatives having a broad range of DNA affinities, *Bioconjugate Chemistry* 14, 877-883.
- [277] Casals, J., Debethune, L., Alvarez, K., Risitano, A., Fox, K. R., Grandas, A., and Pedroso, E. (2006) Directing quadruplex-stabilizing drugs to the telomere: Synthesis and properties of acridine - Oligonucleotide conjugates, *Bioconjugate Chemistry* 17, 1351-1359.
- [278] Das, I., Desire, J., Manvar, D., Baussanne, I., Pandey, V. N., and Decout, J. L. (2012) A Peptide Nucleic Acid-Aminosugar Conjugate Targeting Transactivation Response Element of HIV-1 RNA Genome Shows a High Bioavailability in Human Cells and Strongly Inhibits Tat-

- Mediated Transactivation of HIV-1 Transcription, *Journal of Medicinal Chemistry* 55, 6021-6032.
- [279] Kuzuya, A., Machida, K., Mizoguchi, R., and Komiyama, M. (2002) Conjugation of various acridines to DNA for site-selective RNA scission by lanthanide ion, *Bioconjugate Chemistry* 13, 365-369.
- [280] Wang, S. K., Wu, Y., and Ou, T. M. (2015) RNA G-Quadruplex: The New Potential Targets for Therapy, *Current Topics in Medicinal Chemistry* 15, 1947-1956.
- [281] Eskandarpour, M., Kiaii, S., Zhu, C. Y., Castro, J., Sakko, A. J., and Hansson, J. (2005) Suppression of oncogenic NRAS by RNA interference induces apoptosis of human melanoma cells, *International Journal of Cancer* 115, 65-73.
- [282] Collie, G., Reszka, A. P., Haider, S. M., Gabelica, V., Parkinson, G. N., and Neidle, S. (2009) Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes, *Chemical Communications*, 7482-7484.
- [283] Rose, D. J. (1993) Characterization of antisense binding-properties of peptide nucleic-acids by capillary gel-electrophoresis, *Analytical Chemistry* 65, 3545-3549.
- [284] Gokhale, S. S., and Kumar, V. A. (2010) Amino/guanidino-functionalized N-(pyrrolidin-2-ethyl)glycine-based pet-PNA: Design, synthesis and binding with DNA/RNA, *Organic & Biomolecular Chemistry* 8, 3742-3750.
- [285] Paramasivan, S., Rujan, I., and Bolton, P. H. (2007) Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding, *Methods* 43, 324-331.
- [286] Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., and Sasisekharan, V. (1992) Hairpin and parallel quartet structures for telomeric sequences, *Nucleic Acids Research* 20, 4061-4067.
- [287] Hardin, C. C., Perry, A. G., and White, K. (2001) Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids, *Biopolymers* 56, 147-194.
- [288] Gong, W., and Desaulniers, J. P. (2012) Synthesis and properties of RNAs that contain a PNA-RNA dimer, *Nucleosides Nucleotides & Nucleic Acids* 31, 389-400.
- [289] Igarashi, P. (1999) Radioactive Labeling of DNA and RNA Probes, In *Techniques in Molecular Medicine* (Hildebrandt, F., and Igarashi, P., Eds.), pp 114-125, Springer Berlin Heidelberg, Berlin, Heidelberg.
- [290] Pagano, J. M., Clingman, C. C., and Ryder, S. P. (2011) Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes, *Rna-a Publication of the Rna Society* 17, 14-20.
- [291] Kaymak, E., and Ryder, S. P. (2013) RNA Recognition by the Caenorhabditis elegans Oocyte Maturation Determinant OMA-1, *Journal of Biological Chemistry* 288, 30463-30472.
- [292] Wei, Q., Qu, F., and Gong, R. (2015) Research Progresses on Methods for Studying The Interactions Between G-Quadruplex and Its Ligands, *Progress in Biochemistry and Biophysics* 42, 624-638.
- [293] Chen, S. Q., Su, L. J., Qiu, J., Xiao, N. N., Lin, J., Tan, J. H., Ou, T. M., Gu, L. Q., Huang, Z. S., and Li, D. (2013) Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription, *Biochimica Et Biophysica Acta-General Subjects* 1830, 4769-4777.
- [294] Pivetta, C., Lucatello, L., Krapcho, A. P., Gatto, B., Palumbo, M., and Sissi, C. (2008) Perylene side chains modulate G-quadruplex conformation in biologically relevant DNA sequences, *Bioorganic & Medicinal Chemistry* 16, 9331-9339.
- [295] Doria, F., Oppi, A., Manoli, F., Botti, S., Kandoth, N., Grande, V., Manet, I., and Freccero, M. (2015) A naphthalene diimide dyad for fluorescence switch-on detection of G-quadruplexes, *Chemical Communications* 51, 9105-9108.
- [296] Johansson, H. E., Belsham, G. J., Sproat, B. S., and Hentze, M. W. (1994) Target-specific arrest of messenger-RNA translation by antisense 2'-o-alkyloligoribonucleotides, *Nucleic Acids Research* 22, 4591-4598.
- [297] Flanagan, W. M., Kothavale, A., and Wagner, R. W. (1996) Effects of oligonucleotide length, mismatches and mRNA levels on C-5 propyne-modified antisense potency, *Nucleic Acids Research* 24, 2936-2941.

-
- [298] Doyle, D. F., Braasch, D. A., Simmons, C. G., Janowski, B. A., and Corey, D. R. (2001) Inhibition of gene expression inside cells by peptide nucleic acids: Effect of mRNA target sequence, mismatched bases, and PNA length, *Biochemistry* 40, 53-64.
- [299] Good, L., Awasthi, S. K., Dryselius, R., Larsson, O., and Nielsen, P. E. (2001) Bactericidal antisense effects of peptide-PNA conjugates, *Nature Biotechnology* 19, 360-364.
- [300] Endoh, T., and Sugimoto, N. (2016) Mechanical insights into ribosomal progression overcoming RNA G-quadruplex from periodical translation suppression in cells, *Scientific Reports* 6, 8.
- [301] Adrian, M., Heddi, B., and Phan, A. T. (2012) NMR spectroscopy of G-quadruplexes, *Methods* 57, 11-24.
- [302] Rosu, F., Gabelica, V., Houssier, C., and De Pauw, E. (2002) Determination of affinity, stoichiometry and sequence selectivity of minor groove binder complexes with double-stranded oligodeoxynucleotides by electrospray ionization mass spectrometry, *Nucleic Acids Research* 30, 9.
- [303] Kormuth, K. A., Woolford, J. L., and Armitage, B. A. (2016) Homologous PNA Hybridization to Noncanonical DNA G-Quadruplexes, *Biochemistry* 55, 1749-1757.
- [304] Abibi, A., Protozanova, E., Demidov, V. V., and Frank-Kamenetskii, M. D. (2004) Specific versus nonspecific binding of cationic PNAs to duplex DNA, *Biophysical Journal* 86, 3070-3078.
- [305] Yamashita, K., Sato, S., Takamiya, H., Takagi, M., and Takenaka, S. (2001) Analysis of the complex of oligonucleotide duplexes with ligands by MALDI-TOF mass spectroscopy, *Chemistry Letters*, 680-681.
- [306] Bonora, G. M., Drioli, S., Ballico, M., Faccini, A., Corradini, R., Cogoi, S., and Xodo, L. (2007) PNA conjugated to high-molecular weight poly(ethylene glycol): Synthesis and properties, *Nucleosides Nucleotides & Nucleic Acids* 26, 661-664.
- [307] Bugaut, A., and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting, *Nucleic Acids Research* 40, 4727-4741.
- [308] Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008) Stability and kinetics of G-quadruplex structures, *Nucleic Acids Research* 36, 5482-5515.
- [309] Hellman, L. M., and Fried, M. G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions, *Nature Protocols* 2, 1849-1861.
- [310] Holden, N. S., and Tacon, C. E. (2011) Principles and problems of the electrophoretic mobility shift assay, *Journal of Pharmacological and Toxicological Methods* 63, 7-14.
- [311] Nielsen, P. E. (2004) PNA technology, *Molecular Biotechnology* 26, 233-248.
- [312] Escoffre, J. M., Zeghimi, A., Novell, A., and Bouakaz, A. (2013) In-Vivo Gene Delivery by Sonoporation: Recent Progress and Prospects, *Current Gene Therapy* 13, 2-14.
- [313] Zhou, P., Wang, M. M., Du, L., Fisher, G. W., Waggoner, A., and Ly, D. H. (2003) Novel binding and efficient cellular uptake of guanidine-based peptide nucleic acids (GPNA), *Journal of the American Chemical Society* 125, 6878-6879.
- [314] Ellipilli, S., and Ganesh, K. N. (2015) Fluorous Peptide Nucleic Acids: PNA Analogues with Fluorine in Backbone (gamma-CF₂-apg-PNA) Enhance Cellular Uptake, *Journal of Organic Chemistry* 80, 9185-9191.
- [315] Uhlmann, E. (1998) Peptide nucleic acids (PNA) and PNA-DNA chimeras: From high binding affinity towards biological function, *Biological Chemistry* 379, 1045-1052.
- [316] Wickstrom, E., Choob, M., Urtishak, K. A., Tian, X., Sternheim, N., Talbot, S., Archdeacon, J., Efimov, V. A., and Farber, S. A. (2004) Sequence specificity of alternating hydroxypropyl/phosphono peptide nucleic acids against zebrafish embryo mRNAs, *Journal of Drug Targeting* 12, 363-372.
- [317] Shiraishi, T., Bendifallah, N., and Nielsen, P. E. (2006) Cellular delivery of polyheteroaromatic-peptide nucleic acid conjugates mediated by cationic lipids, *Bioconjugate Chemistry* 17, 189-194.
- [318] Ljungstrom, T., Knudsen, H., and Nielsen, P. E. (1999) Cellular uptake of adamantyl conjugated peptide nucleic acids, *Bioconjugate Chemistry* 10, 965-972.
- [319] Hamilton, S. E., Simmons, C. G., Kathiriya, I. S., and Corey, D. R. (1999) Cellular delivery of peptide nucleic acids and inhibition of human telomerase, *Chemistry & Biology* 6, 343-351.

-
- [320] Green, M., and Loewenstein, P. M. (1988) Autonomous functional domains of chemically synthesized human immunodeficiency virus TAT trans-activator protein, *Cell* 55, 1179-1188.
- [321] Frankel, A. D., and Pabo, C. O. (1988) Cellular uptake of the TAT protein from human immunodeficiency virus, *Cell* 55, 1189-1193.
- [322] Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L. L., Pepinsky, B., and Barsoum, J. (1994) TAT-mediated delivery of heterologous proteins into cells, *Proceedings of the National Academy of Sciences of the United States of America* 91, 664-668.
- [323] Wang, F. H., Wang, Y., Zhang, X., Zhang, W. J., Guo, S. R., and Jin, F. (2014) Recent progress of cell-penetrating peptides as new carriers for intracellular cargo delivery, *Journal of Controlled Release* 174, 126-136.
- [324] Pooga, M., Soomets, U., Hallbrink, M., Valkna, A., Saar, K., Rezaei, K., Kahl, U., Hao, J. X., Xu, X. J., Wiesenfeld-Hallin, Z., Hokfelt, T., Bartfai, T., and Langel, U. (1998) Cell penetrating PNA constructs regulate galanin receptor levels and modify pain transmission in vivo, *Nature Biotechnology* 16, 857-861.
- [325] Margus, H., Arukuusk, P., Langel, U., and Pooga, M. (2016) Characteristics of Cell-Penetrating Peptide/Nucleic Acid Nanoparticles, *Molecular Pharmaceutics* 13, 172-179.
- [326] Morris, M. C., Gros, E., Aldrian-Herrada, G., Choob, M., Archdeacon, J., Heitz, F., and Divita, G. (2007) A non-covalent peptide-based carrier for in vivo delivery of DNA mimics, *Nucleic Acids Research* 35, 10.
- [327] Trabulo, S., Cardoso, A. L., Mano, M., and de Lima, M. C. P. (2010) Cell-Penetrating Peptides—Mechanisms of Cellular Uptake and Generation of Delivery Systems, *Pharmaceutics* 3, 961-993.
- [328] Madani, F., Lindberg, S., Langel, Ü., Futaki, S., and Gräslund, A. (2011) Mechanisms of Cellular Uptake of Cell-Penetrating Peptides, *Journal of Biophysics* 2011, 414729.
- [329] Koppelhus, U., Awasthi, S. K., Zachar, V., Holst, H. U., Ebbesen, P., and Nielsen, P. E. (2002) Cell-dependent differential cellular uptake of PNA, peptides, and PNA-peptide conjugates, *Antisense & Nucleic Acid Drug Development* 12, 51-63.
- [330] Abes, S., Williams, D., Prevot, P., Thierry, A., Gait, M. J., and Lebleu, B. (2006) Endosome trapping limits the efficiency of splicing correction by PNA-oligolysine conjugates, *Journal of Controlled Release* 110, 595-604.
- [331] Abes, S., Moulton, H. M., Clair, P., Prevot, P., Youngblood, D. S., Wu, R. P., Iversen, P. L., and Lebleu, B. (2006) Vectorization of morpholino oligomers by the (R-Ahx-R)(4) peptide allows efficient splicing correction in the absence of endosomolytic agents, *Journal of Controlled Release* 116, 304-313.
- [332] Moulton, H. M., Fletcher, S., Neuman, B. W., McClorey, G., Stein, D. A., Abes, S., Wilton, S. D., Buchmeier, M. J., Lebleu, B., and Iversen, P. L. (2007) Cell-penetrating peptide-morpholino conjugates alter pre-mRNA splicing of DMD (Duchenne muscular dystrophy) and inhibit murine coronavirus replication in vivo, *Biochemical Society Transactions* 35, 826-828.
- [333] Deuss, P. J., Arzumanov, A. A., Williams, D. L., and Gait, M. J. (2013) Parallel synthesis and splicing redirection activity of cell-penetrating peptide conjugate libraries of a PNA cargo, *Organic & Biomolecular Chemistry* 11, 7621-7630.
- [334] Palm, C., Jayamanne, M., Kjellander, M., and Hallbrink, M. (2007) Peptide degradation is a critical determinant for cell-penetrating peptide uptake, *Biochimica Et Biophysica Acta-Biomembranes* 1768, 1769-1776.
- [335] Rueping, M., Mahajan, Y., Sauer, M., and Seebach, D. (2002) Cellular uptake studies with beta-peptides, *Chembiochem* 3, 257-259.
- [336] Wender, P. A., Rothbard, J. B., Jessop, T. C., Kreider, E. L., and Wylie, B. L. (2002) Oligocarbamate molecular transporters: Design, synthesis, and biological evaluation of a new class of transporters for drug delivery, *Journal of the American Chemical Society* 124, 13382-13383.
- [337] Farrera-Sinfreu, J., Giralt, E., Royo, M., and Albericio, F. (2007) Cell-Penetrating Proline-Rich Peptidomimetics, In *Peptide Characterization and Application Protocols* (Fields, G. B., Ed.), pp 241-267, Humana Press, Totowa, NJ.
- [338] Martín, I., Teixidó, M., and Giralt, E. (2010) Building Cell Selectivity into CPP-Mediated Strategies, *Pharmaceutics* 3.

-
- [339] Kumar, P., Wu, H. Q., McBride, J. L., Jung, K. E., Kim, M. H., Davidson, B. L., Lee, S. K., Shankar, P., and Manjunath, N. (2007) Transvascular delivery of small interfering RNA to the central nervous system, *Nature* 448, 39-43.
- [340] Sawant, R. M., Hurley, J. P., Salmaso, S., Kale, A., Tolcheva, E., Levchenko, T. S., and Torchilin, V. P. (2006) "SMART" drug delivery systems: Double-targeted pH-responsive pharmaceutical nanocarriers, *Bioconjugate Chemistry* 17, 943-949.
- [341] Sancey, L., Garanger, E., Foillard, S., Schoehn, G., Hurbin, A., Albiges-Rizo, C., Boturyn, D., Souchier, C., Grichine, A., Dumy, P., and Coll, J. L. (2009) Clustering and Internalization of Integrin $\alpha(v)\beta(3)$ With a Tetrameric RGD-synthetic Peptide, *Molecular Therapy* 17, 837-843.
- [342] Liu, Y., Steiniger, S. C. J., Kim, Y., Kaufmann, G. F., Felding-Habermann, B., and Janda, K. D. (2007) Mechanistic studies of a peptidic GRP78 ligand for cancer cell-specific drug delivery, *Molecular Pharmaceutics* 4, 435-447.
- [343] Svensen, N., Walton, J. G. A., and Bradley, M. (2012) Peptides for cell-selective drug delivery, *Trends in Pharmacological Sciences* 33, 186-192.
- [344] Nguyen, L. T., Yang, X. Z., Du, X., Wang, J. W., Zhang, R., Zhao, J., Wang, F. J., Dong, Y., and Li, P. F. (2015) Enhancing tumor-specific intracellular delivering efficiency of cell-penetrating peptide by fusion with a peptide targeting to EGFR, *Amino Acids* 47, 997-1006.
- [345] Mae, M., Myrberg, H., El-Andaloussi, S., and Langel, U. (2009) Design of a Tumor Homing Cell-Penetrating Peptide for Drug Delivery, *International Journal of Peptide Research and Therapeutics* 15, 11-15.
- [346] Saleh, A. F. A., Aojula, H. S., and Pluen, A. (2008) Enhancement of gene transfer using YIGSR analog of Tat-derived peptide, *Biopolymers* 89, 62-71.
- [347] Weinstein, R., Sayariar, E. N., Felsen, C. N., and Tsien, R. Y. (2014) In Vivo Targeting of Hydrogen Peroxide by Activatable Cell-Penetrating Peptides, *Journal of the American Chemical Society* 136, 874-877.
- [348] Kale, A. A., and Torchilin, V. P. (2007) "Smart" drug carriers: PEGylated TATp-Modified pH-Sensitive Liposomes, *Journal of Liposome Research* 17, 197-203.
- [349] Jin, E. L., Zhang, B., Sun, X. R., Zhou, Z. X., Ma, X. P., Sun, Q. H., Tang, J. B., Shen, Y. Q., Van Kirk, E., Murdoch, W. J., and Radosz, M. (2013) Acid-Active Cell-Penetrating Peptides for in Vivo Tumor-Targeted Drug Delivery, *Journal of the American Chemical Society* 135, 933-940.
- [350] Shi, N. Q., Gao, W., Xiang, B., and Qi, X. R. (2012) Enhancing cellular uptake of activable cell-penetrating peptide-doxorubicin conjugate by enzymatic cleavage, *International Journal of Nanomedicine* 7, 1613-1621.
- [351] Hansen, M. B., van Gaal, E., Minten, I., Storm, G., van Hest, J. C. M., and Lowik, D. (2012) Constrained and UV-activatable cell-penetrating peptides for intracellular delivery of liposomes, *Journal of Controlled Release* 164, 87-94.
- [352] Lee, S. H., Moroz, E., Castagner, B., and Leroux, J. C. (2014) Activatable Cell Penetrating Peptide-Peptide Nucleic Acid Conjugate via Reduction of Azobenzene PEG Chains, *Journal of the American Chemical Society* 136, 12868-12871.
- [353] Miller, S. M., Simon, R. J., Ng, S., Zuckermann, R. N., Kerr, J. M., and Moos, W. H. (1994) Proteolytic studies of homologous peptide and n-substituted glycine peptoid oligomers, *Bioorganic & Medicinal Chemistry Letters* 4, 2657-2662.
- [354] Utku, Y., Dehan, E., Ouerfelli, O., Piano, F., Zuckermann, R. N., Pagano, M., and Kirshenbaum, K. (2006) A peptidomimetic siRNA transfection reagent for highly effective gene silencing, *Molecular Biosystems* 2, 312-317.
- [355] Fara, M. A. (2008) Cell penetrating peptoid carriers : microwave assisted synthesis and bioactive molecules delivery, University of Edinburgh.
- [356] Kolmel, D. K., Horner, A., Ronicke, F., Nieger, M., Schepers, U., and Brase, S. (2014) Cell-penetrating peptoids: Introduction of novel cationic side chains, *European Journal of Medicinal Chemistry* 79, 231-243.
- [357] Shamout, F. E., Pouliopoulos, A. N., Lee, P., Bonaccorsi, S., Towhidi, L., Krams, R., and Choi, J. J. (2015) Enhancement of non-invasive trans-membrane drug delivery using ultrasound and

- microbubbles during physiologically relevant flow, *Ultrasound in Medicine and Biology* 41, 2435-2448.
- [358] Tlaxca, J. L., Anderson, C. R., Klibanov, A. L., Lowrey, B., Hossack, J. A., Alexander, J. S., Lawrence, M. B., and Rychak, J. J. (2010) Analysis of in vitro transfection by sonoporation using cationic and neutral microbubbles, *Ultrasound in Medicine and Biology* 36, 1907-1918.
- [359] Burgess, M. T., and Porter, T. M. (2015) Acoustic cavitation-mediated delivery of small interfering ribonucleic acids with phase-shift nano-emulsions, *Ultrasound in Medicine and Biology* 41, 2191-2201.
- [360] Quaia, E. (2005) Physical Basis and Principles of Action of Microbubble-based Contrast Agents, In *Contrast Media in Ultrasonography: Basic Principles and Clinical Applications* (Quaia, E., Ed.), pp 15-30, Springer Berlin Heidelberg, Berlin, Heidelberg.
- [361] Sboros, V. (2008) Response of contrast agents to ultrasound, *Advanced Drug Delivery Reviews* 60, 1117-1136.
- [362] Marmottant, P., and Hilgenfeldt, S. (2003) Controlled vesicle deformation and lysis by single oscillating bubbles, *Nature* 423, 153-156.
- [363] Delalande, A., Bouakaz, A., Renault, G., Tabareau, F., Kotopoulis, S., Midoux, P., Arbeille, B., Uzbekov, R., Chakravarti, S., Postema, M., and Pichon, C. (2011) Ultrasound and microbubble-assisted gene delivery in Achilles tendons: Long lasting gene expression and restoration of fibromodulin KO phenotype, *Journal of Controlled Release* 156, 223-230.
- [364] Prentice, P., Cuschierp, A., Dholakia, K., Prausnitz, M., and Campbell, P. (2005) Membrane disruption by optically controlled microbubble cavitation, *Nature Physics* 1, 107-110.
- [365] Meijering, B. D. M., Juffermans, L. J. M., van Wamel, A., Henning, R. H., Zuhorn, I. S., Emmer, M., Versteilen, A. M. G., Paulus, W. J., van Gilst, W. H., Kooiman, K., de Jong, N., Musters, R. J. P., Deelman, L. E., and Kamp, O. (2009) Ultrasound and Microbubble-Targeted Delivery of Macromolecules Is Regulated by Induction of Endocytosis and Pore Formation, *Circulation Research* 104, 679-U226.
- [366] Liu, Y., Yan, J., Santangelo, P. J., and Prausnitz, M. R. (2016) DNA uptake, intracellular trafficking and gene transfection after ultrasound exposure, *Journal of Controlled Release* 234, 1-9.
- [367] Shen, Y. Y., Guo, J. X., Chen, G. S., Chin, C. T., Chen, X., Chen, J., Wang, F., Chen, S. G., and Dan, G. (2016) Delivery of liposomes with different sizes to mice brain after sonication by focused ultrasound in the presence of microbubbles, *Ultrasound in Medicine and Biology* 42, 1499-1511.
- [368] Burgess, A., and Hynynen, K. (2016) Microbubble-Assisted Ultrasound for Drug Delivery in the Brain and Central Nervous System, In *Therapeutic Ultrasound* (Escoffre, J. M., and Bouakaz, A., Eds.), pp 293-308, Springer Int Publishing Ag, Cham.
- [369] Piscaglia, F., Bolondi, L., Italian Soc Ultrasound Med Biol, S. I., and Study Grp Ultrasound Contrast, A. (2006) The safety of Sonovue (R) in abdominal applications: Retrospective analysis of 23188 investigations, *Ultrasound in Medicine and Biology* 32, 1369-1375.
- [370] Li, T. L., Tachibana, K., and Kuroki, M. (2003) Gene transfer with echo-enhanced contrast agents: Comparison between Albunex, Optison, and Levovist in mice - Initial results, *Radiology* 229, 423-428.
- [371] Cohen, J. L., Cheirif, J., Segar, D. S., Gillam, L. D., Gottdiener, J. S., Hausnerova, E., and Bruns, D. E. (1998) Improved left ventricular endocardial border delineation and opacification with OPTISON (FS069), a new echocardiographic contrast agent - Results of a phase III multicenter trial, *Journal of the American College of Cardiology* 32, 746-752.
- [372] Yan, P., Chen, K. J., Wu, J., Sun, L., Sung, H. W., Weisel, R. D., Xie, J., and Li, R. K. (2014) The use of MMP2 antibody-conjugated cationic microbubble to target the ischemic myocardium, enhance Timp3 gene transfection and improve cardiac function, *Biomaterials* 35, 1063-1073.
- [373] Klibanov, A. L. (2008) Targeted Microbubbles: Ultrasound Contrast Agents for Molecular Imaging, In *Nanoparticles in Biomedical Imaging: Emerging Technologies and Applications* (Bulte, J. W. M., and Mado, M. M. J., Eds.), pp 327-341, Springer New York, New York, NY.
- [374] Maeda, H. (2010) Tumor-Selective Delivery of Macromolecular Drugs via the EPR Effect: Background and Future Prospects, *Bioconjugate Chemistry* 21, 797-802.

-
- [375] Rapoport, N., Gao, Z. G., and Kennedy, A. (2007) Multifunctional nanoparticles for combining ultrasonic tumor imaging and targeted chemotherapy, *Journal of the National Cancer Institute* 99, 1095-1106.
 - [376] Pouliopoulos, A. N., Bonaccorsi, S., and Choi, J. J. (2014) Exploiting flow to control the in vitro spatiotemporal distribution of microbubble-seeded acoustic cavitation activity in ultrasound therapy, *Physics in Medicine and Biology* 59, 6941-6957.
 - [377] Kinoshita, M., and Hynynen, K. (2007) Key factors that affect sonoporation efficiency in in vitro settings: The importance of standing wave in sonoporation, *Biochemical and Biophysical Research Communications* 359, 860-865.
 - [378] Bailey, R. W., Nguyen, T., Robertson, L., Gibbons, E., Nelson, J., Christensen, R. E., Bell, J. P., Judd, A. M., and Bell, J. D. (2009) Sequence of Physical Changes to the Cell Membrane During Glucocorticoid-Induced Apoptosis in S49 Lymphoma Cells, *Biophysical Journal* 96, 2709-2718.
 - [379] Hollo, Z., Homolya, L., Davis, C. W., and Sarkadi, B. (1994) Calcein accumulation as a fluorometric functional assay of the multidrug transporter, *Biochimica Et Biophysica Acta-Biomembranes* 1191, 384-388.
 - [380] Lammertink, B., Deckers, R., Storm, G., Moonen, C., and Bos, C. (2015) Duration of ultrasound-mediated enhanced plasma membrane permeability, *International Journal of Pharmaceutics* 482, 92-98.
 - [381] Kang, S. H., Cho, M. J., and Kole, R. (1998) Up-regulation of luciferase gene expression with antisense oligonucleotides: Implications and applications in functional assay developments, *Biochemistry* 37, 6235-6239.
 - [382] Karshafian, R., Bevan, P. D., Williams, R., Samac, S., and Burns, P. N. (2009) Sonoporation by ultrasound-activated microbubble contrast agents: effect of acoustic exposure parameters on cell membrane permeability and cell viability, *Ultrasound in Medicine and Biology* 35, 847-860.
 - [383] Shapiro, G., Wong, A. W., Bez, M., Yang, F., Tam, S., Even, L., Sheyn, D., Ben-David, S., Tawackoli, W., Pelled, G., Ferrara, K. W., and Gazit, D. (2016) Multiparameter evaluation of in vivo gene delivery using ultrasound-guided, microbubble-enhanced sonoporation, *Journal of Controlled Release* 223, 157-164.
 - [384] Helfield, B., Chen, X., Watkins, S. C., and Villanueva, F. S. (2016) Biophysical insight into mechanisms of sonoporation, *Proceedings of the National Academy of Sciences*.
 - [385] Miller, M. L., Andringa, A., Elliott, J., Conwell, K., Dixon, K., and Carty, M. P. (1998) The morphological and spectral phenotype of apoptosis in HeLa cells varies following exposure to UV-C and the addition of inhibitors of ICE and CPP32, *Cell Proliferation* 31, 17-33.
 - [386] Qin, M. Y., Chen, Z. X., Luo, Q. C., Wen, Y., Zhang, N. X., Jiang, H. L., and Yang, H. Y. (2015) Two-Quartet G-Quadruplexes Formed by DNA Sequences Containing Four Contiguous GG Runs, *Journal of Physical Chemistry B* 119, 3706-3713.
 - [387] Basundra, R., Kumar, A., Amrane, S., Verma, A., Phan, A. T., and Chowdhury, S. (2010) A novel G-quadruplex motif modulates promoter activity of human thymidine kinase 1, *Febs Journal* 277, 4254-4264.
 - [388] Chen, H., Long, H. T., Cui, X. J., Zhou, J., Xu, M., and Yuan, G. (2014) Exploring the Formation and Recognition of an Important G-Quadruplex in a HIF1 alpha Promoter and Its Transcriptional Inhibition by a Benzo c phenanthridine Derivative, *Journal of the American Chemical Society* 136, 2583-2591.
 - [389] Huang, H., Suslov, N. B., Li, N.-S., Shelke, S. A., Evans, M. E., Koldobskaya, Y., Rice, P. A., and Piccirilli, J. A. (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore, *Nature chemical biology* 10, 686-691.
 - [390] Mullen, M. A., Assmann, S. M., and Bevilacqua, P. C. (2012) Toward a Digital Gene Response: RNA G-Quadruplexes with Fewer Quartets Fold with Higher Cooperativity, *Journal of the American Chemical Society* 134, 812-815.
 - [391] Carmena, M., and Earnshaw, W. C. (2003) The cellular geography of aurora kinases, *Nature Reviews Molecular Cell Biology* 4, 842-854.
 - [392] Bavetsias, V., and Linardopoulos, S. (2015) Aurora Kinase inhibitors: Current Status and Outlook, *Frontiers in Oncology* 5, 10.

-
- [393] Giet, R., McLean, D., Descamps, S., Lee, M. J., Raff, J. W., Prigent, C., and Glover, D. M. (2002) Drosophila Aurora A kinase is required to localize D-TACC to centrosomes and to regulate astral microtubules, *Journal of Cell Biology* 156, 437-451.
 - [394] Di Leonardo, A., Khan, S. H., Linke, S. P., Greco, V., Seidita, G., and Wahl, G. M. (1997) DNA rereplication in the presence of mitotic spindle inhibitors in human and mouse fibroblasts lacking either p53 or pRb function, *Cancer Research* 57, 1013-1019.
 - [395] Lanni, J. S., and Jacks, T. (1998) Characterization of the p53-dependent postmitotic checkpoint following spindle disruption, *Molecular and Cellular Biology* 18, 1055-1064.
 - [396] Meraldi, P., Honda, R., and Nigg, E. A. (2002) Aurora-A overexpression reveals tetraploidization as a major route to centrosome amplification in p53(-/-) cells, *Embo Journal* 21, 483-492.
 - [397] Anand, S., Penrhyn-Lowe, S., and Venkitaraman, A. R. (2003) AURORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol, *Cancer Cell* 3, 51-62.
 - [398] Katayama, H., Sasai, K., Kawai, H., Yuan, Z. M., Bondaruk, J., Suzuki, F., Fujii, S., Arlinghaus, R. B., Czerniak, B. A., and Sen, S. (2004) Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53, *Nature Genetics* 36, 55-62.
 - [399] Giet, R., Petretti, C., and Prigent, C. (2005) Aurora kinases, aneuploidy and cancer, a coincidence or a real link?, *Trends in Cell Biology* 15, 241-250.
 - [400] Bischoff, J. R., Anderson, L., Zhu, Y. F., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., Chan, C. S. M., Novotny, M., Slamon, D. J., and Plowman, G. D. (1998) A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers, *Embo Journal* 17, 3052-3065.
 - [401] Zhou, H. Y., Kuang, J., Zhong, L., Kuo, W. L., Gray, J. W., Sahin, A., Brinkley, B. R., and Sen, S. (1998) Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation, *Nature Genetics* 20, 189-193.
 - [402] Siggelkow, W., Boehm, D., Gebhard, S., Battista, M., Sicking, I., Lebrecht, A., Solbach, C., Hellwig, B., Rahnenfuhrer, J., Koelbl, H., Gehrman, M., Marchan, R., Cadenas, C., Hengstler, J. G., and Schmidt, M. (2012) Expression of aurora kinase A is associated with metastasis-free survival in node-negative breast cancer patients, *Bmc Cancer* 12, 11.
 - [403] Lo Iacono, M., Monica, V., Saviozzi, S., Ceppi, P., Bracco, E., Papotti, M., and Scagliotti, G. V. (2011) Aurora Kinase A expression is associated with lung cancer histological-subtypes and with tumor de-differentiation, *Journal of Translational Medicine* 9, 6.
 - [404] Gritsko, T. M., Coppola, D., Paciga, J. E., Yang, L., Sun, M., Shelley, S. A., Fiorica, J. V., Nicosia, S. V., and Cheng, J. Q. (2003) Activation and overexpression of centrosome kinase BTAK/Aurora-A in human ovarian cancer, *Clinical Cancer Research* 9, 1420-1426.
 - [405] Kurai, M., Shiozawa, T., Shih, H. C., Miyamoto, T., Feng, Y. Z., Kashima, H., Suzuki, A., and Konishi, I. (2005) Expression of Aurora kinases A and B in normal, hyperplastic, and malignant human endometrium: Aurora B as a predictor for poor prognosis in endometrial carcinoma, *Human Pathology* 36, 1281-1288.
 - [406] Bar-Shira, A., Pinthus, J. H., Rozovsky, U., Goldstein, M., Sellers, W. R., Yaron, Y., Eshhar, Z., and Orr-Urtreger, A. (2002) Multiple genes in human 20q13 chromosomal region are involved in an advanced prostate cancer xenograft, *Cancer Research* 62, 6803-6807.
 - [407] Goos, J., Coupe, V. M. H., Diosdado, B., Diemen, P., Karga, C., Belien, J. A. M., Carvalho, B., van den Tol, M. P., Verheul, H. M. W., Geldof, A. A., Meijer, G. A., Hoekstra, O. S., Fijneman, R. J. A., and DeCoDe, P. E. T. G. (2013) Aurora kinase A (AURKA) expression in colorectal cancer liver metastasis is associated with poor prognosis, *British Journal of Cancer* 109, 2445-2452.
 - [408] Qin, J., He, X. F., Wei, W., Liu, Z. Z., Xie, J. J., Wang, W., Du, Y. P., Chen, Y., and Si, H. Q. (2015) Association between the STK15 polymorphisms and risk of cancer: a meta-analysis, *Molecular Genetics and Genomics* 290, 97-114.
 - [409] Gu, J., Gong, Y. B., Huang, M. S., Lu, C., Spitz, M. R., and Wu, X. F. (2007) Polymorphisms of STK15 (Aurora-A) gene and lung cancer risk in Caucasians, *Carcinogenesis* 28, 350-355.
 - [410] de Castro, I. P., Aguirre-Portoles, C., Fernandez-Miranda, G., Canamero, M., Cowley, D. O., Van Dyke, T., and Malumbres, M. (2013) Requirements for Aurora-A in Tissue Regeneration and Tumor Development in Adult Mammals, *Cancer Research* 73, 6804-6815.

-
- [411] Manfredi, M. G., Ecsedy, J. A., Meetze, K. A., Balani, S. K., Burenkova, O., Chen, W., Galvin, K. M., Hoar, K. M., Huck, J. J., LeRoy, P. J., Ray, E. T., Sells, T. B., Stringer, B., Stroud, S. G., Vos, T. J., Weatherhead, G. S., Wysong, D. R., Zhang, M. K., Bolen, J. B., and Claiborne, C. F. (2007) Antitumor activity of MLN8054, an orally active small-molecule inhibitor of Aurora A kinase, *Proceedings of the National Academy of Sciences of the United States of America* 104, 4106-4111.
- [412] Melichar, B., Adenis, A., Lockhart, A. C., Bennouna, J., Dees, E. C., Kayaleh, O., Obermannova, R., DeMichele, A., Zatloukal, P., Zhang, B., Ullmann, C. D., and Schusterbauer, C. (2015) Safety and activity of alisertib, an investigational aurora kinase A inhibitor, in patients with breast cancer, small-cell lung cancer, non-small-cell lung cancer, head and neck squamous-cell carcinoma, and gastro-oesophageal adenocarcinoma: a five-arm phase 2 study, *Lancet Oncology* 16, 395-405.
- [413] Manfredi, M. G., Ecsedy, J. A., Chakravarty, A., Silverman, L., Zhang, M. K., Hoar, K. M., Stroud, S. G., Chen, W., Shinde, V., Huck, J. J., Wysong, D. R., Janowick, D. A., Hyer, M. L., Leroy, P. J., Gershman, R. E., Silva, M. D., Germanos, M. S., Bolen, J. B., Claiborne, C. F., and Sells, T. B. (2011) Characterization of Alisertib (MLN8237), an Investigational Small-Molecule Inhibitor of Aurora A Kinase Using Novel In Vivo Pharmacodynamic Assays, *Clinical Cancer Research* 17, 7614-7624.
- [414] Carol, H., Boehm, I., Reynolds, C. P., Kang, M. H., Maris, J. M., Morton, C. L., Gorlick, R., Kolb, E. A., Keir, S. T., Wu, J. R., Wozniak, A. E., Yang, Y., Manfredi, M., Ecsedy, J., Wang, J. M., Neale, G., Houghton, P. J., Smith, M. A., and Lock, R. B. (2011) Efficacy and pharmacokinetic/pharmacodynamic evaluation of the Aurora kinase A inhibitor MLN8237 against preclinical models of pediatric cancer, *Cancer Chemotherapy and Pharmacology* 68, 1291-1304.
- [415] Kelly, K. R., Nawrocki, S. T., Espitia, C. M., Zhang, M. K., Yang, J. J., Padmanabhan, S., Ecsedy, J., Giles, F. J., and Carew, J. S. (2012) Targeting Aurora A kinase activity with the investigational agent alisertib increases the efficacy of cytarabine through a FOXO-dependent mechanism, *International Journal of Cancer* 131, 2693-2703.
- [416] Kelly, K. R., Ecsedy, J., Medina, E., Mahalingam, D., Padmanabhan, S., Nawrocki, S. T., Giles, F. J., and Carew, J. S. (2011) The novel Aurora A kinase inhibitor MLN8237 is active in resistant chronic myeloid leukaemia and significantly increases the efficacy of nilotinib, *Journal of Cellular and Molecular Medicine* 15, 2057-2070.
- [417] Borisa, A. C., and Bhatt, H. G. (2017) A comprehensive review on Aurora kinase: Small molecule inhibitors and clinical trial studies, *European Journal of Medicinal Chemistry* 140, 1-19.
- [418] Kuivaniemi, H., Tromp, G., and Prockop, D. J. (1991) Mutations in collagen genes - causes of rare and some common diseases in humans, *Faseb Journal* 5, 2052-2060.
- [419] Zhu, Z. S., Jiang, Y. T., Chen, S. G., Jia, S. S., Gao, X. R., Dong, D., and Gao, Y. Z. (2011) An insertion/deletion polymorphism in the 3' untranslated region of type I collagen $\alpha 2$ (COL1A2) is associated with susceptibility for hepatocellular carcinoma in a Chinese population, *Cancer Genetics* 204, 265-269.
- [420] Mori, K., Enokida, H., Kagara, I., Kawakami, K., Chiyomaru, T., Tatarano, S., Kawahara, K., Nishiyama, K., Seki, N., and Nakagawa, M. (2009) CpG hypermethylation of collagen type I α 2 contributes to proliferation and migration activity of human bladder cancer, *International Journal of Oncology* 34, 1593-1602.
- [421] Liang, Y., Diehn, M., Bollen, A. W., Israel, M. A., and Gupta, N. (2008) Type I collagen is overexpressed in medulloblastoma as a component of tumor microenvironment, *Journal of Neuro-Oncology* 86, 133-141.
- [422] Kauppila, S., Stenback, F., Risteli, J., Jukkola, A., and Risteli, L. (1998) Aberrant type I and type III collagen gene expression in human breast cancer in vivo, *Journal of Pathology* 186, 262-268.
- [423] Mottes, M., and Marie-Jeanne Lievens, P. (2016) Molecular Defects and Cellular Dysfunctions in Restricted Growth Conditions, InTech, Restricted Growth - Clinical, Genetic and Molecular Aspects.

-
- [424] Ramirez, F., Tanaka, S., and Bou-Gharios, G. (2006) Transcriptional regulation of the human alpha 2(I) collagen gene (COL1A2), an informative model system to study fibrotic diseases, *Matrix Biology* 25, 365-372.
- [425] Zhang, A. Y. Q., Bugaut, A., and Balasubramanian, S. (2011) A Sequence-Independent Analysis of the Loop Length Dependence of Intramolecular RNA G-Quadruplex Stability and Topology, *Biochemistry* 50, 7251-7258.
- [426] Shin, S. O., Lee, K. H., Kim, J. H., Baek, S. H., Park, J. W., Gabrielson, E. W., and Kwon, T. K. (2000) Alternative splicing in 5'-untranslational region of STK-15 gene, encoding centrosome associated kinase, in breast cancer cell lines, *Experimental and Molecular Medicine* 32, 193-196.
- [427] Guenard, F., Labrie, Y., Ouellette, G., Beauparlant, C. J., Durocher, F., and BRCAs, I. (2009) Genetic sequence variations of BRCA1-interacting genes AURKA, BAP1, BARD1 and DHX9 in French Canadian Families with high risk of breast cancer, *Journal of Human Genetics* 54, 152-161.
- [428] Lai, C. H., Tseng, J. T., Lee, Y. C., Chen, Y. J., Lee, J. C., Lin, B. W., Huang, T. C., Liu, Y. W., Leu, T. H., Chen, Y. P., Chang, W. C., and Hung, L. Y. (2010) Translational up-regulation of Aurora-A in EGFR-overexpressed cancer, *Journal of Cellular and Molecular Medicine* 14, 1520-1531.
- [429] Phan, A. T., Kuryavyi, V., Gaw, H. Y., and Patel, D. J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter, *Nature Chemical Biology* 1, 167-173.
- [430] Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter, *Journal of the American Chemical Society* 129, 4386-4392.
- [431] Cammas, A., Dubrac, A., Morel, B., Lamaa, A., Touriol, C., Teulade-Fichou, M.-P., Prats, H., and Millevoi, S. (2015) Stabilization of the G-quadruplex at the VEGF IRES represses cap-independent translation, *RNA biology* 12, 320-329.
- [432] Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger-RNAs, *Nucleic Acids Research* 15, 8125-8148.
- [433] McClelland, S., Shrivastava, R., and Medh, J. D. (2009) Regulation of Translational Efficiency by Disparate 5'-UTRs of PPAR gamma Splice Variants, *Ppar Research*, 8.
- [434] Laguerre, A., Stefan, L., Larrouy, M., Genest, D., Novotna, J., Pirrotta, M., and Monchaud, D. (2014) A Twice-As-Smart Synthetic G-Quartet: PyroTASQ Is Both a Smart Quadruplex Ligand and a Smart Fluorescent Probe, *Journal of the American Chemical Society* 136, 12406-12414.
- [435] Pinter, T., Klippel, Z., Cesas, A., Croitoru, A., Decaestecker, J., Gibbs, P., Hotko, Y., Jassem, J., Kurteva, G., Novotny, J., O'Reilly, S., Salek, T., Reiner, M., Morrow, P. K., Choi, M. R., Whittaker, S., and Blanke, C. (2017) A Phase III, Randomized, Double-Blind, Placebo-Controlled Trial of Pegfilgrastim in Patients Receiving First-Line FOLFOX/Bevacizumab or FOLFIRI/Bevacizumab for Locally Advanced or Metastatic Colorectal Cancer: Final Results of the Pegfilgrastim and Anti-VEGF Evaluation Study (PAVES), *Clinical Colorectal Cancer* 16, 103-+.
- [436] Osawa, H., Ding, Y., Minakata, K., Goto, H., and Handa, N. (2017) The efficacy and safety of Ramucirumab and FOLFIRI for the treatment of advanced/ recurred colorectal cancer in late line, *Annals of Oncology* 28, 1.
- [437] Strickler, J. H., Rangwala, F. A., Rushing, C., Niedzwiecki, D., Altomare, I., Uronis, H. E., Hsu, S. D., Zafar, Y., Morse, M., Chang, D. Z., Wells, J. L., Blackwell, K. L., Marcom, P. K., Webb, A. R., Dropkin, E., Arrowood, C., and Hurwitz, H. (2016) X-TRAP: Phase I/II study of capecitabine (X) plus ziv-aflibercept (TRAP) in metastatic colorectal cancer (mCRC), *Journal of Clinical Oncology* 34, 1.
- [438] Messersmith, W. A., and Ahnen, D. J. (2008) Targeting EGFR in colorectal cancer, *New England Journal of Medicine* 359, 1834-1836.
- [439] Battaglin, F., Dadduzio, V., Bergamo, F., Manai, C., Schirripa, M., Lonardi, S., Zagonel, V., and Loupakis, F. (2017) Anti-EGFR monoclonal antibody panitumumab for the treatment of

- patients with metastatic colorectal cancer: an overview of current practice and future perspectives, *Expert Opinion on Biological Therapy* 17, 1297-1308.
- [440] Chong, C. R., and Janne, P. A. (2013) The quest to overcome resistance to EGFR-targeted therapies in cancer, *Nature Medicine* 19, 1389-1400.