

**Biophysical characterisation of a G-quadruplex in the Bcl-x pre-mRNA and the  
binding specificity of the ligand GQC-05**

Thesis submitted for the degree of  
Doctor of Philosophy  
at the University of Leicester

by

Mohammed Ali Hussayn Bhogadia BSc (Hons)  
Department of Molecular and Cell Biology  
University of Leicester

May 2021 (revised October 2021)

## Abstract

### **Biophysical characterisation of a G-quadruplex in the Bcl-x pre-mRNA and the binding specificity of the ligand GQC-05 – Mohammed Ali Hussayn Bhogadia**

Bcl-x is a member of the B-cell lymphoma 2 (Bcl-2) protein family. This protein family consists of both pro and anti-apoptotic proteins that regulate mitochondrial membrane permeability in cells. The two alternative 5' splice sites (5'ss) in exon 2 of Bcl-x gives rise to two antagonistic splice variants Bcl-X<sub>L</sub> and Bcl-X<sub>S</sub>, which are anti and pro-apoptotic respectively. In many cancers, activation of oncogenic pathways leads to overexpression of the Bcl-X<sub>L</sub> isoform, resulting in cancer cell survival and growth (Boucher et al., 2000). Increasing the amount of the X<sub>S</sub> isoform, thereby promoting apoptosis, is a novel way to kill cancer cells. Our recent studies have identified an RNA secondary structure known as a G-quadruplex (G4) that has the potential to form near both splice sites and alter the splicing pattern of Bcl-x (Weldon et al., 2017, 2018). The ellipticine derivative GQC-05, a previously identified DNA G4 specific ligand (Brown et al., 2011), has been shown to bind proximal to both 5'ss resulting in a 8-fold increase in X<sub>S</sub>/X<sub>L</sub> ratio, analogues of which showed considerably less extensive effects (Weldon et al., 2018). Using various biophysical techniques, we have characterised the RNA secondary structure element near the X<sub>S</sub> 5'ss and have postulated the possible effects of GQC-05 on these structural features and its role in altering splice site selection. Our results show that a stable G4 exists downstream of the X<sub>S</sub> 5'ss and this structure is stabilised in the presence of GQC-05. We also show that GQC-05 displays less G4 binding specificity in buffer, but shows greater G4 selectivity in the presence of a nuclear extract. Therefore, we aim to understand how these secondary structure elements stabilised by GQC-05 lead to splice site bias, enabling us to design more potent molecules as novel anti-cancer compounds.



## **COVID19 impact statement**

As my research was primarily lab based, it was essential for me to be able to access a laboratory with the necessary equipment for data collection. As such, data collection was put on hold during the initial lockdown period, consequently delaying achieving the objectives of my project. Additionally, due to occupancy restrictions when re-opened, I was only able to carry out my work in my allocated half of the day, significantly disrupting my workflow and as a consequence, slowed me down. Access to facilities such as the diamond light source were also put on hold due to the pandemic, delaying my progress with X-ray crystallography. Other objectives (which included in-depth NMR structural characterisations) were sadly not achievable due to the time lost during the lockdown.

## Acknowledgments

Firstly, I would like to thank my supervisor Dr Cyril Dominguez, for his help and advice throughout my PhD. Secondly, I would like to thank Professor Ian Eperon for his encouragement and guidance during our G4 meetings. I would also like to thank Professor Glenn Burley and his group who provided us the compounds for our biophysical experiments.

From the groups of Dr Cyril Dominguez and Professor Ian Eperon, I would particularly like to thank Dr Rafael Del Villar Guerra and Dr Sudipta Ghosh, who have given me invaluable advice with regards to my experiments as well as developing my knowledge and understanding of the techniques I was using. I very much enjoyed working with them and the discussions we had to understand the complex mechanisms behind GQC-05.

I would also like to thank: Dr Idir Malki, Dr Santosh Kumar Upadhyay, Dr Hesna Kara, Dr Mohamed Hassan, Dr Kayoko Tanaka, Dr Inara Liepina, Hatice Esenkaya, Hollie Watmuff, Mishal Tariq, Katie Berwick, Adam Lightfoot, Ayesha Mirza, Dr Beatriz Romartinez, Dr Vasileios Paschalis, Dr Chris Millard, Edward Brown, Dipti Vashi, Sharon Munday, Dr Thomas Schalch, Professor Peter Moody, Dr Olga Makarova, Dr Chinar Pathak, Mahdi Hussain, Dr Louise Fairall, Dr Siyu Wang, Dr Fred Muskett and Luke Bailey. It has been an absolute pleasure to know the aforementioned individuals, and I very much enjoyed our cross-discipline discussions.

I would like to thank my sister Fatema-Zahra Bhogadia for all of her help and support when proof reading my thesis. I would also like to thank my beautiful Grandmother, who sadly passed away during my PhD. She raised me when I was a boy and I would not be where I am today without her support and prayers. I therefore dedicate this thesis to her memory.

# Table of contents

ABSTRACT .....	2
COVID19 IMPACT STATEMENT .....	3
ACKNOWLEDGMENTS .....	4
TABLE OF CONTENTS .....	5
LIST OF FIGURES .....	10
LIST OF TABLES .....	14
LIST OF ABBREVIATIONS.....	15
CHAPTER 1: INTRODUCTION.....	19
1.1 CELL DEATH .....	20
1.1.1 THE IMPORTANCE OF CELL DEATH .....	20
1.1.2 APOPTOSIS AND NECROSIS .....	20
1.1.3 APOPTOSIS (PROGRAMMED CELL DEATH) .....	21
1.1.4 MECHANISMS OF APOPTOSIS.....	21
<b>THE EXTRINSIC PATHWAY</b> .....	23
<b>THE INTRINSIC PATHWAY</b> .....	24
1.1.5 THE BCL-2 FAMILY OF PRO AND ANTI-APOPTOTIC PROTEINS.....	25
1.1.6 TARGETING THE APOPTOTIC MACHINERY IN CANCER CELLS AS A POTENTIAL THERAPEUTIC STRATEGY .....	26
1.2 SPLICING .....	28
1.2.1 IMPORTANCE OF SPLICING.....	28
1.2.2 MECHANISM OF SPLICING .....	28
1.2.3 MECHANISM OF ALTERNATIVE SPLICING .....	33
1.2.4 FACTORS INFLUENCING SPLICE SITE SELECTION.....	35
1.2.4.1 SPLICING ENHANCERS AND SILENCERS.....	36
1.2.5 ALTERNATIVE SPLICING OF BCL-X.....	37
1.2.5.1 PROTEINS THAT REGULATE THE SPLICING OF BCL-X.....	37
1.2.6 TARGETING THE SPLICING MACHINERY AS A THERAPEUTIC STRATEGY .....	41
1.3 THE G-QUADRUPLEX (G4).....	43
1.4 G4S IN BIOLOGY.....	48
1.4.1 DNA G4s .....	48
1.4.2 RNA G4s.....	49
1.5 CHARACTERISATION OF G4 STRUCTURES IN NUCLEIC ACIDS.....	52
1.5.1 BIOINFORMATIC APPROACHES .....	52
1.5.2 BIOPHYSICAL APPROACHES.....	56
1.5.2.1 UV SPECTROSCOPY.....	56
1.5.2.2 CD SPECTROSCOPY .....	60
1.5.2.3 NMR SPECTROSCOPY .....	64
1.5.2.3.1 <i>IN VIVO</i> NMR DETECTION OF G4s.....	66
1.5.2.4 FLUORESCENCE DETECTION OF G4 STRUCTURES .....	71
1.5.3 BIOCHEMICAL AND FUNCTIONAL APPROACHES .....	77
1.6 G4 BINDING LIGANDS .....	83
1.6.1 PORPHYRINS .....	84

1.6.2 QUINDOLINES .....	87
1.6.3 ELLIPTICINES .....	89
1.7 G4 LIGANDS THAT REGULATE BCL-X ALTERNATIVE SPLICING .....	93
CHAPTER 2: MATERIALS AND METHODS.....	96
2.1 RNA PREPARATION .....	97
2.1.1 SHORT OLIGONUCLEOTIDE PREPARATION .....	97
2.1.2 RNA ANNEALING PROTOCOL .....	97
2.1.3 PRODUCTION OF DNA TEMPLATES FOR <i>IN VITRO</i> TRANSCRIPTION .....	97
2.1.3.1 PLASMID DESIGN AND STOCK PREPARATION.....	97
2.1.3.2 TRANSFORMATION .....	97
2.1.3.3 PLASMID AMPLIFICATION .....	98
2.1.3.4 PLASMID PURIFICATION .....	98
2.1.3.5 PLASMID LINEARISATION.....	98
2.1.4 <i>IN VITRO</i> TRANSCRIPTION OF RNAs .....	99
2.1.4.1 TRANSCRIPTION OPTIMISATION .....	99
2.1.4.2 LARGE SCALE TRANSCRIPTIONS.....	99
2.1.5 PURIFICATION OF THE LARGE X <sub>S</sub> , SD, AND Q1 NON-RADIOLABELLED (COLD) <i>IN VITRO</i> TRANSCRIBED RNAs .....	99
2.1.6 PURIFICATION OF THE SMALL Q2 AND Q2.2 150-206 COLD <i>IN VITRO</i> TRANSCRIBED RNAs.....	100
2.1.7 SINGLE COMPONENT INJECTIONS OF THE TRANSCRIPTION REAGENTS.....	100
2.1.8 SYNTHESIS OF <sup>32</sup> P-RNA.....	100
2.1.8.1 <i>IN VITRO</i> TRANSCRIPTION OF <sup>32</sup> P-RNA .....	100
2.1.8.2 PURIFICATION OF <i>IN VITRO</i> TRANSCRIBED <sup>32</sup> P-RNAs .....	101
2.1.8.3 ETHANOL PRECIPITATION OF <sup>32</sup> P-RNA .....	101
2.1.8.4 <i>IN VITRO</i> TRANSCRIPTION OF 7-DEAZA-GTP <sup>32</sup> P-RNA.....	101
2.1.9 EXPRESSION AND PURIFICATION OF T7 RNA POLYMERASE (T7 RNAP).....	101
2.2 CIRCULAR DICHROISM (CD) AND ULTRA VIOLET (UV) SPECTROSCOPY MEASUREMENTS .....	103
2.2.1 ROOM TEMPERATURE SPECTRUM IN THE ABSENCE OF NUCLEAR EXTRACT .....	103
2.2.2 ROOM TEMPERATURE SPECTRUM IN THE PRESENCE OF NUCLEAR EXTRACT .....	103
2.2.3 INDUCED CD SPECTRUM RECORDINGS .....	103
2.2.4 CD MELTING EXPERIMENTS WITH GQC-05 .....	103
2.2.5 CD SPECTRA WITH NMM.....	104
2.3 NMR SPECTROSCOPY MEASUREMENTS .....	104
2.3.1 1D NMR SPECTRUM RECORDINGS OF WT AND MUTANT FRAGMENTS.....	104
2.3.2 KCL TITRATIONS.....	104
2.3.3 GQC-05 TITRATIONS.....	104
2.3.4 NMR KINETICS AND PSEUDO 2D RECORDINGS .....	104
2.4 FLUORESCENCE SPECTROSCOPY MEASUREMENTS.....	105
2.4.1 FLUORESCENCE EMISSION SPECTRA RECORDINGS WITH GQC-05 AND C-MYC DNA .....	105
2.4.2 FLUORESCENCE EMISSION SPECTRA WITH NMM AND THE BCL-X RNAs.....	105
2.4.3 FLUORESCENCE EMISSION SPECTRA OF RNA SECONDARY STRUCTURES .....	105
2.4.4 FLUORESCENCE TITRATIONS IN THE ABSENCE OF NUCLEAR EXTRACT .....	105
2.4.5 FLUORESCENCE TITRATIONS IN THE PRESENCE OF NUCLEAR EXTRACT .....	106
2.4.6 JOB PLOT ASSAYS.....	106
2.4.6.1 JOB PLOT WITH C-MYC DNA .....	106
2.4.6.2 JOB PLOT WITH Q2 RNA IN THE ABSENCE OF NUCLEAR EXTRACT.....	106
2.4.6.3 JOB PLOT WITH Q2 RNA IN THE PRESENCE OF NUCLEAR EXTRACT .....	106
2.5 PULL-DOWN ASSAYS.....	106
2.6 NATIVE GEL ASSAYS .....	107
2.6.1 Q2 KCL TITRATIONS .....	107

2.6.2 GQC-05 TITRATIONS.....	107
2.6.3 EMSA ASSAYS WITH THE U1 snRNA IN THE ABSENCE OF POLY T .....	107
2.6.4: MEASURING EFFECTS OF GQC-05 ON THE U1 MIGRATION .....	108
2.6.5: POLY T OPTIMISATIONS .....	108
2.6.6 EMSA WITH POLY T AND GQC-05 .....	108
2.7 X-RAY CRYSTALLOGRAPHY.....	108
CHAPTER 3: STRUCTURAL AND BIOPHYSICAL CHARACTERISATION OF THE PUTATIVE G4 REGION IN BCL-X (Q2) .....	110
3.1 INTRODUCTION .....	111
3.2 BIOINFORMATIC ANALYSIS OF THE Q2 FORMING REGION .....	111
3.3.1 CD SPECTROSCOPY OF Q2 AND MUTANTS.....	115
3.3.2 CD AND UV CHARACTERISATION OF THE Q2 G4 IN HELA NUCLEAR EXTRACT .....	117
3.4 USING NMR SPECTROSCOPY TO CHARACTERISE THE Q2 STRUCTURE .....	121
3.4.1 NMR SPECTROSCOPY ON THE Q2 AND MUTANT SEQUENCES .....	121
3.5 APPROACHES TO REDUCE THE STRUCTURAL HETEROGENEITY OF Q2.....	123
3.5.1 REDUCING SAMPLE HETEROGENEITY BY OPTIMISATION OF THE Q2 SEQUENCE .....	123
3.5.2 THE EFFECT OF K <sup>+</sup> CONCENTRATION IN THE STRUCTURAL HETEROGENEITY OF Q2.....	126
3.6 CHARACTERISATION OF THE G4 STRUCTURE IN THE ABSENCE OF SALT .....	129
3.6.1 CHARACTERISATION OF A G4 STRUCTURE USING NMM FLUORESCENCE.....	129
3.6.2 CD SPECTROSCOPY IN THE PRESENCE OF NMM .....	130
3.7 X-RAY CRYSTALLOGRAPHY OF Q2.....	131
3.8 DISCUSSION .....	133
CHAPTER 4: BINDING MODE AND SPECIFICITY OF GQC-05 FOR THE G4 REGION (Q2) IN BCL-X.....	135
4.1 INTRODUCTION .....	136
4.2 FLUORESCENCE SPECTROSCOPY TO STUDY THE STRUCTURAL SPECIFICITY OF GQC-05 .....	136
4.2.1 DETERMINING THE BINDING CONSTANT (K <sub>d</sub> ) OF GQC-05 TO THE C-MYC G4 USING FLUORESCENCE.....	136
4.2.2 STRUCTURAL VS SEQUENCE SPECIFICITY OF GQC-05 .....	138
4.2.3 DETERMINING THE BINDING CONSTANT OF GQC-05 FOR THE BCL-X Q2 SEQUENCE .....	139
4.2.4 DETERMINING THE BINDING CONSTANT OF GQC-05 IN CONDITIONS THAT SUPPORT <i>IN VITRO</i> SPLICING .....	141
4.2.5 BINDING STOICHIOMETRY OF GQC-05 TO THE C-MYC DNA AND Q2 RNA G4 SEQUENCES .....	144
4.3 NATIVE GEL ANALYSIS TO PROBE CONFORMATIONAL CHANGES OF Q2 IN THE PRESENCE OF GQC-05.....	147
4.4 CD SPECTROSCOPY TO STUDY THE BINDING MODE OF GQC-05 .....	148
4.4.1 INDUCED CD ANALYSIS OF GQC-05 .....	148
4.4.2 STRUCTURE-ACTIVITY PROFILING OF GQC-05 ANALOGUES BY ICD.....	150
4.4.3 THERMAL MELTING OF BCL-X Q2 AND ΔG4 RNA IN THE PRESENCE OF GQC-05 .....	155
4.5 NMR SPECTROSCOPY TO DETERMINE THE STRUCTURAL SPECIFICITY OF GQC-05 .....	156
4.5.1 1D NMR TITRATIONS WITH GQC-05 AND RNAs .....	156
4.5.2 COMPLEMENT TRAPPING TO DETERMINE STRUCTURAL SELECTIVITY OF GQC-05 .....	158
4.6 DISCUSSION .....	163
CHAPTER 5: STRUCTURAL CHARACTERISATION AND BINDING SPECIFICITY OF GQC-05 FOR LONGER RNAs SURROUNDING THE Q2 REGION .....	166
5.1 INTRODUCTION .....	167

5.2 RATIONALE BEHIND THE DESIGN OF EACH OF THE LARGER BCL-X CONSTRUCTS.....	167
5.3 GENERATING TEMPLATE DNA FOR TRANSCRIPTION OF THE LONGER FRAGMENTS OF BCL-X .....	170
5.4 TRANSCRIPTION AND PURIFICATION OF THE LARGE RNAS .....	174
5.4.1 <i>IN VITRO</i> TRANSCRIPTIONS OF THE LARGE BCL-X RNAS.....	174
5.4.2 HPLC PURIFICATIONS OF THE 150-206 BCL-X RNAS.....	178
5.4.3 HPLC PURIFICATIONS OF X <sub>5</sub> Q1 AND SD BCL-X RNAS.....	181
5.5 G4 CHARACTERISATION ON THE LARGER BCL-X RNAS .....	183
5.5.1 FLUORESCENCE SPECTROSCOPY TO CHARACTERISE A G4 IN THE LONGER RNAS.....	183
5.5.2 CD SPECTROSCOPY TO CHARACTERISE A G4 IN THE LONGER RNAS.....	186
5.6 IMPORTANT STRUCTURAL ELEMENTS FOR THE SPECIFICITY OF THE GQC-05 INTERACTION.....	190
5.6.1 BIOTIN PULL-DOWN ASSAYS TO PROBE THE STRUCTURAL SPECIFICITY OF GQC-05 .....	190
5.6.2 FLUORESCENCE TITRATIONS TO IDENTIFY KEY RNA STRUCTURAL ELEMENTS IN THE GQC-05 INTERACTION.....	192
5.6.3 BINDING OF GQC-05 TO THE LARGER RNAS IN THE PRESENCE OF A NUCLEAR EXTRACT.....	194
5.7 EFFECTS OF GQC-05 IN THE STRUCTURAL REMODELLING OF THE X <sub>5</sub> DOMAIN .....	197
5.7.1 CD SPECTROSCOPY TO MONITOR THE CONFORMATIONAL CHANGES OCCURRING WITH GQC-05 .....	197
5.7.2 HARNESSING THE INTRINSIC FLUORESCENCE PROPERTIES OF G4S TO PROBE CONFORMATIONAL CHANGES UPON GQC-05 ADDITION .....	199
5.8 THE EFFECT OF GQC-05 IN ALTERING THE U1 ACCESSIBILITY AT THE X <sub>5</sub> 5'SS.....	201
5.8.1 USING EMSA TO MONITOR THE U1-X <sub>5</sub> INTERACTION.....	201
5.8.2 EMSA ASSAYS IN THE PRESENCE OF HIGHER CONCENTRATIONS OF GQC-05 .....	203
5.9 DISCUSSION .....	206
CHAPTER 6: CONCLUSIONS AND FUTURE PERSPECTIVES.....	209
6.1 INTRODUCTION .....	210
6.2 THE Q2 G-QUADRUPLEX .....	211
6.3 EFFECTS OF GQC-05 ON THE BCL-X MRNA .....	214
6.4 EFFECT OF GQC-05 ON THE ACCESSIBILITY OF THE X <sub>5</sub> 5'SS.....	215
6.5 FINAL CONCLUSIONS .....	216
CHAPTER 7: APPENDIX.....	218
7.1 PLASMID VECTOR USED FOR CLONING THE LARGER BCL-X FRAGMENTS .....	219
7.2 SEQUENCES OF THE LARGE BCL-X CONSTRUCTS .....	219
7.3 10 X TRIS-BORATE EDTA (TBE) .....	220
7.4 GEL RECIPES .....	220
7.4.1 AGAROSE GEL 1% .....	220
7.4.2 ACRYLAMIDE GELS .....	220
7.5 TOLUIDINE BLUE GEL STAIN .....	220
7.6 GEL LOADING DYES.....	221
7.6.1 DENATURING LOADING DYE (2 X).....	221
7.6.2 NATIVE LOADING DYE (5 X).....	221
7.7 TRANSCRIPTION BUFFER (TB 20X).....	221
7.8 2TY GROWTH MEDIA (1 LITRE) .....	222
BIBLIOGRAPHY .....	223



# List of figures

## Chapter 1

Figure 1.11 Overview of the apoptotic pathways. ....	22
Figure 1.21 The splicing reaction.....	29
Figure 1.22 Overview of the mechanism of alternative splicing .....	30
Figure 1.23 Initial stages of spliceosome assembly .....	32
Figure 1.24 Different patterns of alternative splicing .....	35
Figure 1.25 Mechanisms of Bcl-x alternative splicing.....	40
Figure 1.31 G-quadruplex and duplex DNA structures.....	44
Figure 1.32 Different possible strand stoichiometry of G4s.....	46
Figure 1.33 Strand polarities and loop orientations.....	47
Figure 1.41 Structure and function of the TERRA RNA G4 .....	51
Figure 1.512 Structure of a non-canonical G4s not predicted by QGRS mapper .....	54
Figure 1.514 UV characterization of DNA/RNA structures .....	57
Figure 1.515 TDS spectra of different DNA structures .....	59
Figure 1.516 CD spectra of the different G4 structures .....	62
Figure 1.517 CD spectra and melts of each PQS sequence of the Bcl-x pre-mRNA.....	63
Figure 1.518 Showing the expected chemical shift regions within the 1D spectrum for all of the exchangeable and non-exchangeable protons in nucleic acids .....	68
Figure 1.519 NMR characterization of G4s.....	69
Figure 1.521 Heterogeneity in the c-myc DNA sequence and how this relates to the 1D NMR spectrum.	70
Figure 1.522 <i>In vivo</i> NMR studies on the human telomere G4 DNA sequence: d(G <sub>3</sub> (TTAG <sub>3</sub> ) <sub>3</sub> T).	71
Figure 1.523 Principles of the fluorescence turn on assay .....	73
Figure 1.524 Structure and specificity of the porphyrin class molecule NMM .....	75



Figure 1.525 Structure and specificity of the Benzothiozole ligand ThT. ....	76
Figure 1.526 EMSA and foot-printing assays on the Bcl-x 681 construct .....	81
Figure 1.527 Significant changes occurring to the Bcl-x 681 RNA in the oligo accessibility RNase H assays. ....	83
Figure 1.61 Structure and complexes of the porphyrin class of molecules.....	86
Figure 1.62 Structure and complexes of the quindoline class of molecules .....	88
Figure 1.63 Structure and complexes of the Ellipticine class of molecules .....	91
Figure 1.64 Binding specificity of GQC-05 and proposed binding mode .....	92
Figure 1.71 Bcl-x <i>in vitro</i> splicing and proposed mechanism of action of GQC-05. ....	95

### Chapter 3

Figure 3.21 QGRS mapper analysis of Q2 and proposed model.....	113
Figure 3.22 RNA-fold of Q2 .....	114
Figure 3.31 CD spectroscopy of the Q2 fragment. ....	115
Figure 3.32 QGRS mapper comparison of Q2 and Q2.2 RNA. ....	116
Figure 3.33 RNA-fold of Q2.2 .....	116
Figure 3.34 CD spectroscopy of the Q2.2 fragment.....	117
Figure 3.35 RNA-fold of $\Delta$ G4 .....	118
Figure 3.36 CD data of Q2 RNA in the presence of 5% nuclear extract .....	119
Figure 3.37 CD data of $\Delta$ G4 RNA in the presence of 5% nuclear extract .....	120
Figure 3.38 UV spectrum comparison of the Q2 and $\Delta$ G4 RNAs in 5% HeLa NE .....	120
Figure 3.41 1D NMR spectroscopy of WT and mutant fragments.....	122
Figure 3.51 QGRS, RNA-fold output and NMR characterisation of Q2 C.....	124
Figure 3.52 CD characterisation of Q2 C.....	125
Figure 3.53 Native gel and NMR analysis with K <sup>+</sup> titrations.....	128
Figure 3.61 Fluorescence analysis of Q2 RNA in the presence of NMM ( $\lambda_{exc}$ = 393nm).....	130
Figure 3.62 CD spectrum of Q2 RNA in the presence of NMM .....	131
Figure 3.71 Crystallisation trials of Q2 in the Helix screen .....	132

## Chapter 4

Figure 4.21 GQC-05 and c-myc fluorescence emission spectra ( $\lambda_{exc} = 320\text{nm}$ ) .....	137
Figure 4.22 Fluorescence emission at 560nm of GQC-05 with the RNA AGGGAU ( $\lambda_{exc} = 320\text{nm}$ ) .....	138
Figure 4.23 Fluorescence emission at 560 nm of GQC-05 with different RNA fragments ( $\lambda_{exc} = 320\text{nm}$ ) .....	140
Figure 4.24 Fluorescence emission at 560nm of GQC-05 with different RNA fragments in the presence of 10% HeLa nuclear extract ( $\lambda_{exc} = 320\text{nm}$ ). .....	143
Figure 4.25 Job plots on the G4 forming DNA/RNA.....	146

Figure 4.31 Native PAGE of Q2 with increasing equivalents of GQC-05.....	147
--	-----

Figure 4.41 Induced CD spectrums of GQC-05 with different DNA/RNA sequences at 20 $\mu\text{M}$ .....	149
Figure 4.42 SAR profiling of GQC-05 analogues .....	151
Figure 4.43 Induced CD spectrums of GQC-05 and its analogues .....	154
Figure 4.44 CD thermal melts in the presence of GQC-05.....	155

Figure 4.51 NMR titrations with GQC-05 on Q2 and $\Delta\text{G4}$ RNA.....	157
Figure 4.52 Complement trapping theory and 1D NMR kinetics.....	160
Figure 4.53 Pseudo 2D traces in the absence and presence of GQC-05.....	161
Figure 4.54 Rate curves showing duplex formation in the absence and presence of GQC-05 .....	162

Figure 4.61 Schematic showing proposed mechanism of GQC-05 at the $X_s$ 5'ss .....	165
--	-----

## Chapter 5

Figure 5.21 Larger Bcl-x constructs around the Q2 region generated for G4 characterization and GQC-05 binding .....	169
---	-----

Figure 5.31 Stages in T7 RNAP transcription .....	171
Figure 5.32 Scheme in generating the larger Bcl-x DNA templates.....	173

Figure 5.41 Transcription optimization gels for the Q2 150-206 fragment.....	176
Figure 5.42 Transcription optimization gels for the $X_s$ 38-206 fragment .....	177
Figure 5.43 Analytical anion exchange chromatography purification optimisation on Q2 150-206.....	180
Figure 5.44 Preparative anion exchange chromatography purifications of $X_s$ , SD and Q1 ..	182

Figure 5.51 NMM emission spectra on and around the $X_s$ domain.....	185
--	-----

Figure 5.52 CD spectra and thermal melts of the 150-206 fragments .....	187
Figure 5.53 CD spectra and thermal melts of the X <sub>S</sub> 38-206 domain .....	189
Figure 5.61 Biotin GQC-05 pull downs of the X <sub>S</sub> domain .....	191
Figure 5.62 : Fluorescence emission titrations at 560nm of each of the indicated Bcl-x constructs with GQC-05.....	193
Figure 5.63 The effect of nuclear extract on the binding of GQC-05 and the formation of a G4 in X <sub>S</sub> and Q2 150-206.....	196
Figure 5.71 CD spectra and thermal melts of X <sub>S</sub> in the presence of GQC-05 .....	198
Figure 5.72 The intrinsic fluorescence of G4s and how this is used to monitor the conformational changes observed with GQC-05 on X <sub>S</sub> RNA. ....	200
Figure 5.81 Using EMSA to examine the U1 accessibility to the X <sub>S</sub> 5'ss .....	202
Figure 5.82 EMSA of U1 to the X <sub>S</sub> RNA with 200 $\mu$ M GQC-05 and 70 $\mu$ M T-50 .....	205
<b>Chapter 6</b>	
Figure 6.21 Example of a non-canonical G4 topology of the KRAS promotor G4. ....	211
Figure 6.22 1D NMR spectra in the imino region of G4 RNA sequences displaying higher order structures.....	213

## List of tables

### Chapter 1

Table 1.511 Features of known G4 prediction bioinformatic tools.....	53
--	----

Table 1.513 QGRS and G4 hunter score outputs for the six PQS regions of the Bcl-x pre-mRNA.....	56
---	----

Table 1.520 Expected NMR chemical shifts.....	70
---	----

### Chapter 3

Table 3.1 RNA oligonucleotide fragments. ....	111
---	-----

### Chapter 5

Table 5.41 .....	176
------------------	-----

Table 5.42 .....	177
------------------	-----

## List of abbreviations

APS – Ammonium persulphate

AS – Alternative splicing

ATP – Adenosine-5'-triphosphate

Bcl-X – B-cell lymphoma extra

Bcl-X<sub>L</sub> – B-cell lymphoma extra large

Bcl-X<sub>s</sub> – B-cell lymphoma extra small

BH – Bcl-2 homology

bp – base pair

BPS – Branch point sequence

BSA – Bovine Serum Albumin

CD – Circular dichroism

c-MYC – cellular myc

c.p.s – counts per second

C.V – Column volumes

1D- 1 dimensional

2D – 2 dimensional

7-deaza-GTP – 7-deazaguanosine-5'-triphosphate

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic Acid

dNTPs – Deoxyribonucleotide triphosphates

DTT – Dithiothreitol

EDTA – Ethylenediamine – N,N,N,N' -tetraacetic acid

EMSA – Electrophoretic mobility shift assay

ESE – Exon splicing enhancer

ESS – Exon splicing silencer

EtBr – Ethidium bromide

FMRP – Fragile X mental retardation protein

FT – Fourier transformation

G4 – G-quadruplex

GTP – Guanosine 5' -triphosphate

HAT – Histone acetyltransferase

HDAC – Histone deacetylases

HeLa – Henrietta Lacks (cells)

h – hour

hnRNP – Heterogeneous nuclear ribonucleoprotein

HPLC – High performance liquid chromatography

ICD – Induced CD

IPTG – Isopropyl- $\beta$ -D-Thiogalactopyranoside

ISE – Intronic splicing enhancer

ISS – Intronic splicing silencer

Kb – Kilobase

K<sub>d</sub> – Dissociation constant

kDa – kiloDalton

L – Litre

m – milli

M – Molar

Min – Minute(s)

mRNA – Messenger ribonucleic acid

n – Nano

NaAc – Sodium acetate

NE – Nuclear extract

NMR – Nuclear magnetic resonance

NOESY – Nuclear Overhauser effect

nt – Nucleotide

OD – Optical density

Oligo – Oligonucleotide

2'OMe – 2' O-methyl

p – pico

PAGE – Polyacrylamide gel electrophoresis

PBS – Phosphate buffered saline

PCD – programmed cell death

PDB – Protein database

PEG – Poly(ethylene glycol)

Pre-mRNA – Precursor messenger RNA

ppm – Parts per million

PTB – Polypyrimidine tract binding protein

PQS – Putative quadruplex sequence

QGRS – Quadruplex forming G-rich sequences

RNA – Ribonucleic acid

RNAP – RNA polymerase

RNase – Ribonuclease

rNTPs – Ribonucleotide triphosphates

rpm – rotations per minute

RRM – RNA recognition motif

RT – room temperature

Sam68 – Src-associated in mitosis 68 kDa protein  
SDS – Sodium Dodecyl Sulphate  
sec – second(s)  
SHAPE – Selective 2'hydroxyl acylation and primer extension  
siRNA – Small interfering RNA  
snRNA – Small nuclear ribonucleic acid  
snRNP – Small nuclear ribonucleoprotein  
SR – Serine- Arginine  
SRE – Splicing regulator element  
SRSF – Serine/ Arginine-rich splicing factor  
ss – splice site  
SSOs – Splice switching oligonucleotides  
STAR - Signal transduction and activation of RNA metabolism

TBE – Tris-borate-EDTA  
TDS – Thermal difference spectrum  
TE.1 – Tris-EDTA  
TEMED - N,N,N',N'-tetramethylethylenediamine  
ThT – Thioflavin T  
TMPyP4 - 5, 10, 15, 20-tetra-(N-methyl-4-pyridyl)porphyrin  
tRNA – transfer ribonucleic acid

U snRNP – Uridine-rich Small Nuclear Ribonucleoprotein particle  
U2AF – U2 snRNP auxiliary factor  
UTR – Untranslated region  
UV – Ultraviolet

V – Volts

WT – Wild-type

Zn – DIGP - Tetrakis-(diisopropyl-guanidino) zinc phthalocyanine



## **CHAPTER 1: Introduction**

**1.1 Cell death**

**1.2 Splicing**

**1.3 The G-quadruplex (G4)**

**1.4 Structure and function of biologically relevant G4s**

**1.5 Characterisation of G4 structures in nucleic acids**

**1.6 G4 binding ligands**

**1.7 G4 ligands that regulate Bcl-x alternative splicing**

## **1.1 Cell death**

### **1.1.1 The importance of cell death**

Cell death is an essential mechanism within the body to keep a fine balance between dying and newly formed cells, thereby controlling total cell number (Renehan et al., 2001). In an adult, there are around 10 billion cells that die daily to achieve such balance (Renehan et al., 2001). It is therefore crucial that such process is tightly regulated in order to only remove those cells that are damaged and not required, rather than those that are healthy and provide important bodily functions (such as undamaged cardiomyocytes in the contraction of the heart muscle). Fortunately, there are many regulatory elements and pathways involved in controlling the cell death response, in which malfunctions have been shown to cause many different types of diseases. For example, overactive cell death pathways leads to many kinds of neurodegenerative diseases such as Alzheimer's and Parkinson's disease, whereas an attenuation in this pathway is linked to autoimmune disorders and cancer (Renehan et al., 2001; Tatton and Olanow, 1999). Therefore, understanding these pathways at the molecular level is crucial to develop treatments that can be used to intervene in key areas in patients that have a defective cell death response.

### **1.1.2 Apoptosis and necrosis**

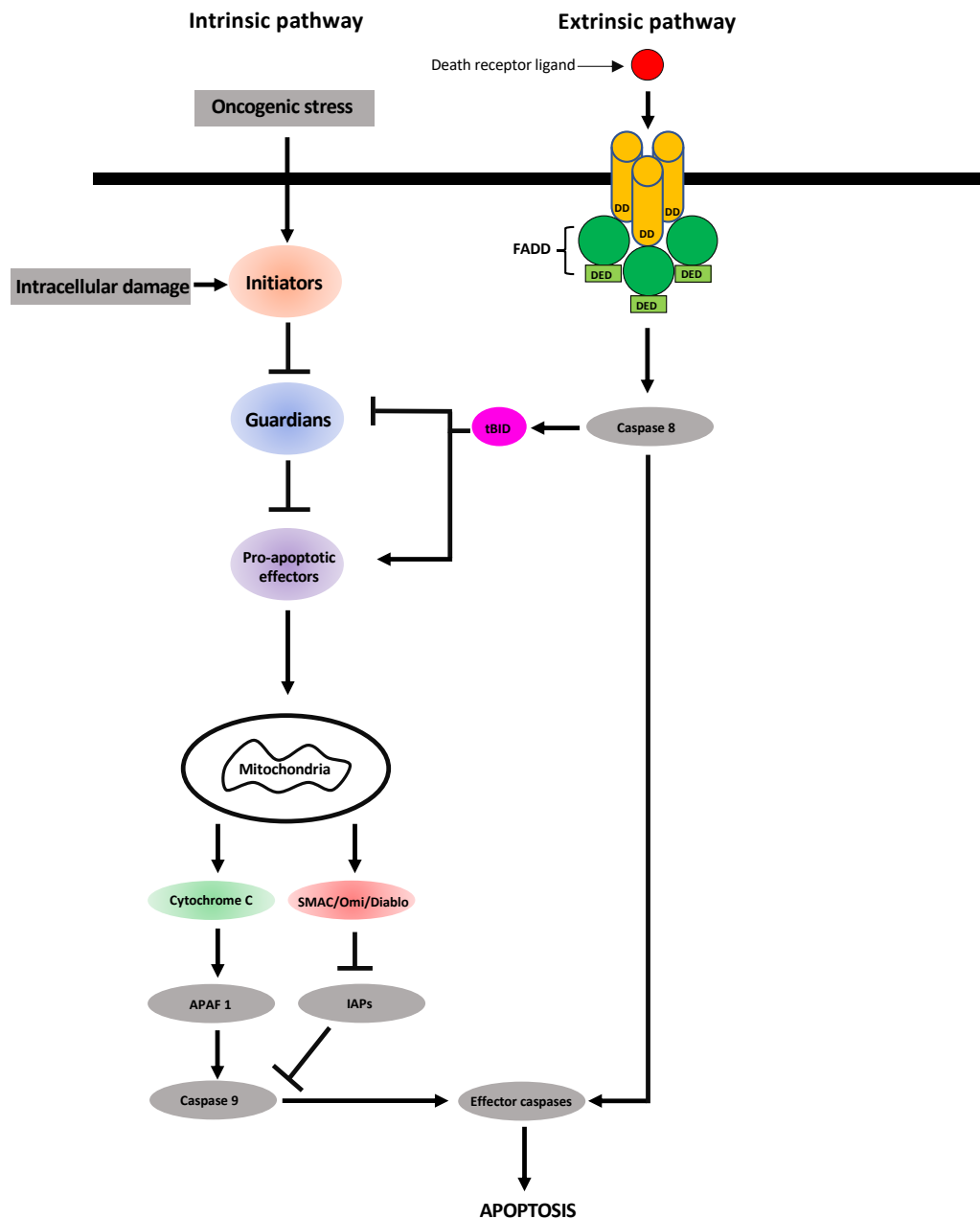
There are two different mechanisms that govern cell death, apoptosis and necrosis . Apoptosis is known as 'death by suicide', and is a clean and efficient way to eliminate unwanted cells. However, necrosis arises as a result of cell injury or mechanical strain, inducing rupture of the cell membrane and release of cytoplasmic content into the organism, causing a local inflammatory response. Failure to clear apoptotic cells may also result in a necrotic response, and has been shown to be one of the major causes of various autoimmune diseases such as systemic lupus erythematosus (Tas et al., 2006). In this section, we will be focussing on the apoptotic pathway, as malfunctions in this response have been linked to many types of malignancies.

### **1.1.3 Apoptosis (programmed cell death)**

Apoptosis (type 1 programmed cell death) is an essential process starting at the very early stages of human development, such as the formation of digits and organ sculpture (Hernández-Martínez and Covarrubias, 2011). Apoptosis was first described in 1972 by John Kerr, who observed distinct morphological changes of apoptotic cells such as cytoplasmic condensation and membrane blebbing (Kerr, 1972). Apoptosis is a natural homeostatic mechanism to control cell number in metazoans, such that overgrowth of damaged cells which threaten the organism can be prevented (Ashkenazi and Dixit, 1998). More recently, the evasion of apoptosis and replication of damaged cells were shown to be two of the major hallmarks of cancer cell initiation and progression (Hanahan and Weinberg, 2011).

### **1.1.4 Mechanisms of apoptosis**

In order to prevent replication and growth of damaged cells, the pro-apoptotic pathway is activated in response to oncogenic stress such as DNA damage. Figure 1.11 shows an overview of the two main apoptotic pathways that occur in cells as a result of such stimuli: the mitochondrial intrinsic pathway and the death receptor pathway (also known as the extrinsic pathway). Both pathways are involved in increasing the mitochondrial membrane permeability and converge in the release of cytochrome c from the mitochondria and activation of caspase 9. Caspase 9 then cleaves and activates effector caspases 3, 6 and 7, which are involved in degrading cellular components such as the nuclear lamina, resulting in the apoptotic response (Slee et al., 2001). These two pathways are discussed in more detail below.



**Figure 1.11 Overview of the apoptotic pathways.**

Showing both the intrinsic and extrinsic mechanisms of the apoptotic response (adapted from Czabotar et al, 2014).

## **The Extrinsic pathway**

One method of regulating apoptosis in cells is through extracellular ligands binding to death receptors on the plasma membrane of cells. These death receptors belong to the tumour necrosis factor (TNF) receptor family as they all include death domains in the cytoplasmic side of the receptor (Elmore, 2007). There are various ligands that activate apoptosis in this way, such as Fas and TRAIL, which are protein ligands expressed in a many human tissues, all of which have a common mechanism in activating the apoptotic response (Nagata and Golstein, 1995; Wiley et al., 1995).

Binding of the ligand to its cognate receptor results in receptor trimerisation. The death domains (DD) on the cytoplasmic side of the receptor aggregate, resulting in the recruitment of adaptor proteins such as FADD (figure 1.11). The death effector domain (DED) of FADD recruits caspases 8 via interactions with the DED domain of the caspase, bringing them in close proximity for cross cleavage and hence activation (figure 1.11) (Jin and El-Deiry, 2005). Unlike the intrinsic pathway, this particular pathway does not involve the Bcl-2 family of proteins nor regulates mitochondrial membrane permeability directly, but does have an influence on both of these in a more discrete manner. Activation of caspase 8 via the extrinsic pathway results in the cleavage of the BH3 only protein Bid, which is normally inactive, to a truncated Bid (tBid). The active tBid translocates to the mitochondrial outer membrane and inhibits the anti-apoptotic family of Bcl-2 proteins (guardians), increasing mitochondrial membrane permeability (figure 1.11) (Jin and El-Deiry, 2005; Kantari and Walczak, 2011). Active caspase 8 can also activate effector caspases to trigger the apoptotic response (Czabotar et al, 2014).

### **The intrinsic pathway**

Another method of activating caspases is through regulating mitochondrial membrane permeability, known as intrinsic activation (figure 1.11).

The intrinsic pathway is controlled by BH3 only initiator proteins such as BIM and BAD. When cells are subjected to oncogenic stress, initiator proteins bind and inhibit the 'guardians' of mitochondrial membrane permeability proteins (anti-apoptotic factors), resulting in the activation and multimerization of pro-apoptotic effectors such as BAX, forming a pore, increasing mitochondrial membrane permeability to cytochrome c (Czabotar et al., 2014), resulting in its release in the cytoplasm and binding to APAF1. Binding induces a conformational change in APAF1, resulting in the recruitment the protein pro-caspase 9 (a zymogen). When two pro-caspase 9 enzymes are brought into close proximity, they cleave each other *in trans*, producing an active tetrameric form of caspase 9, which can now act as an initiator caspase for the activation of effector caspases, which the action of the latter ultimately results in apoptosis (Cain et al., 2002).

In addition to cytochrome c, molecules which inhibit the inhibitor of apoptotic proteins (IAPs) are also released from the mitochondria. Examples include Omi, Smac and Diablo, which inhibit the inhibitor of caspases, ensuring full caspase activation (Cain et al., 2002).

### **1.1.5 The Bcl-2 family of pro and anti-apoptotic proteins**

The anti-apoptotic Bcl-2 protein was originally found to be overexpressed in lymphoma as a result of a translocation of the Bcl-2 gene with the strong promotor of immunoglobulins, resulting in overexpression of the Bcl-2 protein (Monni et al., 1997). Since then, many homologues of the Bcl-2 proteins have been identified, some of which have both pro and anti-apoptotic functions.

The pro-survival guardian proteins, such as Bcl-X<sub>L</sub>, consist of 9 alpha helices forming four Bcl-2 homology (BH) domains. They also have a hydrophobic transmembrane domain, enabling membrane insertion into organelles such as the mitochondria, hence they are mainly found within the mitochondria outer membrane (Czabotar et al., 2014). These pro-survival proteins exert their effect by binding to pro-apoptotic effector proteins such as BAD, thereby neutralising their pro-apoptotic functions (Muchmore et al., 1996; Petros et al., 2000). Glycine-138 within Bcl-X<sub>L</sub> found within the binding cleft was shown to be a critical amino acid in the interaction with pro-apoptotic proteins due to its proton only side chain. Changing to a bulkier amino acid abolished binding to pro-apoptotic effectors, rendering it inactive (Aritomi et al., 1997; Muchmore et al., 1996)

### **1.1.6 Targeting the apoptotic machinery in cancer cells as a potential therapeutic strategy**

In 1971, Knudson described cancer as a multi-hit model, where multiple mutations are needed in order for a cell to become cancerous (Knudson, 1971). However, if there was a functional apoptotic machinery present in the cell, every mutation that results in the development of a cancer cell can be inhibited, as many malignancies, such as breast cancers, display overexpression of the anti-apoptotic factor Bcl-X<sub>L</sub> (España et al., 2004). Therefore, there is extensive research into trying to promote cancer cells towards the apoptotic response as this can be used in a variety of cancer treatments.

P53 is the most extensively studied protein and is the major tumour suppressor in cells. P53 contains a DNA binding domain and is a transcription factor responsible for the expression of a variety of genes, including pro-apoptotic genes such as BAX (Fischer, 2017; Zilfou and Lowe, 2009). Another target of p53 is the protein MDM2, a ubiquitin ligase that results in p53 degradation (Freedman et al., 1999). Therefore, cancer cells that overexpress MDM2 would increase p53 degradation and hence the expression of pro-apoptotic genes would be reduced. Around 20% of soft tissue tumours, such as Ewing's sarcoma, exhibit abnormally high amounts of MDM2 expression (Momand et al., 1998). Therefore, one strategy in ensuring the apoptotic response is by inhibiting the p53-MDM2 interaction, thereby increasing the levels of p53 and hence pro-apoptotic genes such as BAX. One such small molecule that has been developed is nutlin-2, which has 2 bromo-phenyl residues able to interact with the hydrophobic pockets on the surface of MDM2, thereby mimicking the p53 interaction (Shangary and Wang, 2009). However, this compound is only useful for cancer cells that overexpress MDM2 but still have wild type p53. Other treatment strategies are required for cancers that have mutant p53.



Another strategy to target apoptosis is by inhibiting the binding of BH3 only proteins to anti-apoptotic factors, thereby increasing the amounts of pro-apoptotic proteins such as BIM at the mitochondrial outer membrane. An example of a small molecule inhibitor is ABT-737, which binds via its chloro-biphenyl and thio-phenyl functional groups to the binding pockets on the surface of Bcl-X<sub>L</sub>, thereby preventing the binding of BH3 only proteins such as BIM (Lee et al., 2007). ABT-737 has been shown to reduce tumour volume in most small cell lung cancer (SCLC) and multiple myeloma cell lines (Hann et al., 2008; Kline et al., 2007).

The extrinsic pathway can also be a target to promote cells towards apoptosis. The TRAIL ligand that binds to its death receptors to activate the extrinsic pathway has a very short half-life (around 23 – 31 minutes in non-human primates), and so is rapidly cleared from the plasma (Kelley et al., 2001). However, there has been extensive research into developing anti-TRAIL receptor antibodies which mimic the ligand to cause receptor trimerisation, some of which are in phase II clinical trials (Von Pawel et al., 2014). The effect of such antibodies is to over-stimulate the extrinsic pathway to increase the apoptotic response.

In summary, targeting the apoptotic machinery has been shown to have advantageous therapeutic potential. However, rather than intervening with the apoptotic pathway directly, an alternative strategy is to target the post transcriptional processing (pre-mRNA splicing) of Bcl-2 family members in order to favour the expression of pro-apoptotic factors and inhibit the expression of anti-apoptotic proteins as a means to induce apoptosis in cancer cells, and this will be the aim of the project.

## **1.2 Splicing**

### **1.2.1 Importance of splicing**

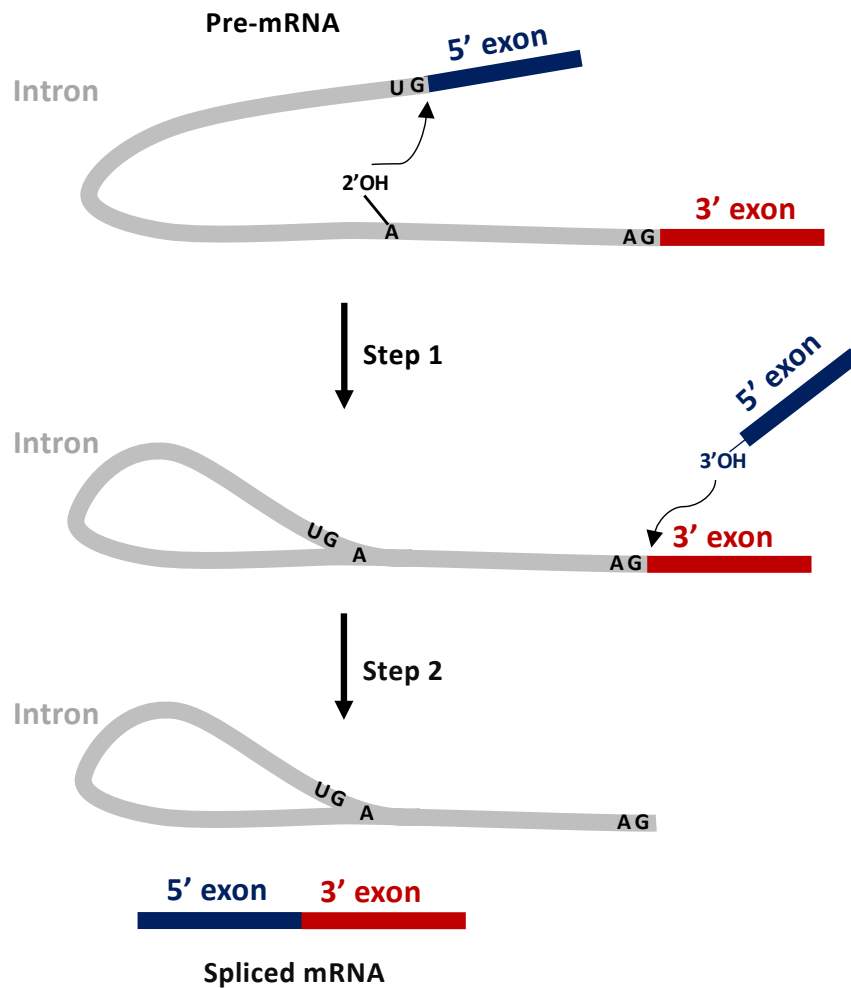
Alternative splicing (AS) is a key biological phenomenon first observed in 1977, which occurs in almost all eukaryotic organisms, providing both diversity to the proteome as well as a regulatory mechanism for the control of gene expression (Chow et al., 1977; Greenberg and Soreq, 2013). This process involves the removal of intronic sequences and the joining together of exons. The pattern by which these introns are removed and exons joined together can be very diverse, meaning that a single unspliced pre-mRNA can give rise to multiple spliced mRNA products, which in turn can lead to a whole array of proteins with differing functions, some of which can have antagonistic functions even when derived from the same pre-mRNA precursor. Before the discovery of AS, and due to the diverse array of proteins found in human cells, the initial consensus was that the human genome is very complex containing over 150,000 genes (Modrek and Lee, 2002). However, the discovery of only 22,000 genes in the human genome suggested that protein diversity must be occurring post-transcriptionally, and that a single pre-mRNA should give rise to multiple different proteins, suggesting that AS has a much stronger role to play in ensuring protein diversity than originally thought (Lander et al., 2001). Therefore, modulating the pattern of splicing is a key regulatory step in determining the pattern of gene expression, and will be discussed below.

### **1.2.2 Mechanism of splicing**

For mRNA splicing to take place, specific sequences on the pre-mRNA act as signals for the assembly of the spliceosomes and initiation of splicing. The three most important sequences are the 5' splice site (5'ss), which mark the exon/intron boundary at the 5' end of the exon, the 3' splice site (3'ss) which marks the intron/exon boundary at the 3' end of the exon, and the branched site. The 5'ss contains a highly conserved GU, followed by an A or G and in some rare instances a U (Reyes et al., 1996; Zhang et al., 2007). The 3'ss has a fully conserved AG and the intronic branched site generally consists of adenosine (Zhang et al., 2007).

Splicing is a 2-step transesterification reaction (figure 1.21). In the first step, the 2' OH of the branch point adenosine attacks a phosphate group at the 5'ss, resulting in a 5'exon fragment with a free 3'OH and a branched lariat. In the second step, the 3'OH of the 5' exon attacks the

3'ss, pushing off the intron and resulting in a fully spliced mature mRNA product (Moore et al., 1993; Staley and Guthrie, 1998).

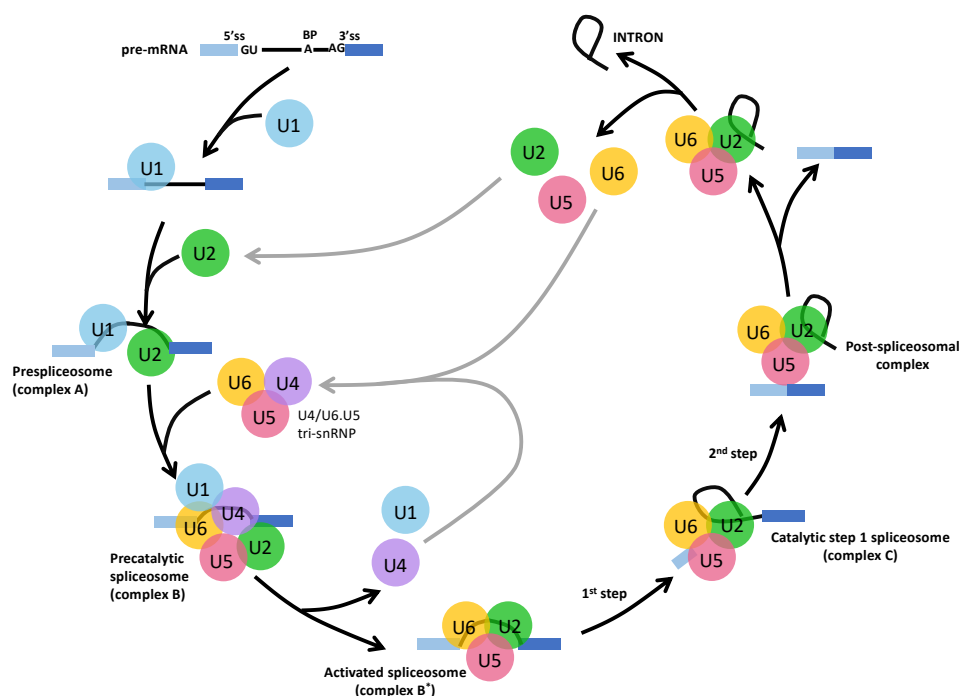


**Figure 1.21 The splicing reaction.**

The 2 step trans-esterification reaction of alternative splicing (adapted from Staley and Guthrie, 1998).

The splicing reaction described above is carried out by a large macromolecular complex called the spliceosome, which is composed of 5 non-coding small nuclear RNAs (snRNAs) (U1,U2,U4,U5 and U6 snRNAs) and hundreds of proteins (see later). However, the question arises as to why such complex spliceosome proteins are required to achieve the above process, if splicing can simply proceed by the RNA alone? Splicing requires nucleotide precision as to the point of intron excision and exon ligation, otherwise there would be catastrophic effects leading to frame shifts, and ultimately the formation of potentially lethal or inactive proteins. Therefore, the role of the spliceosome machinery is to ensure the precision and accuracy of the splicing reaction (Staley and Guthrie, 1998).

In the mid 80s, the incredibly large size of the spliceosome complex was discovered, after observing that the yeast pre-spliced mRNA was being associated in a 40S complex, which was necessary for the splicing reaction (Brody and Abelson, 1985). The spliceosome consists of more than 100 different proteins and 5 small nuclear RNAs (snRNAs), making 5 essential small nuclear ribonucleoproteins (snRNPs) termed U1, U2, U4, U5 and U6, which interact with the pre-mRNA at specific sequences, such as the 5'ss, the 3'ss and the branched site (Will and Lührmann, 1997). The sequence of events in their assembly is illustrated in figure 1.22 and described below.



**Figure 1.22 Overview of the mechanism of alternative splicing.**

The sequence of events during spliceosome assembly to achieve alternative splicing (adapted from Wahl et al., 2009).

The very first stage in this assembly is the base pairing of the U1 snRNP to the 5'ss. The U1 snRNP is a 165 nucleotide RNA consisting of four stem loops, 2 of which are bound by the 70K and U1-A proteins (figure 1.23 A) (Stark et al., 2001). Another U1 binding protein, termed U1-C, was shown not to interact with the U1 RNA directly, but rather the N-terminal portion of the 70K protein (Nelissen et al., 1994). The 5' end of the snRNA is perfectly complementary to a consensus 5'ss, although it has been shown that the base pairing does not need be 100% complementary for the splice site to be used (Roca et al., 2013). In order to stabilise the base pairing between the U1 RNA and the candidate 5'ss, the U1-C protein has a zinc finger which stabilises the interaction of the 5' end of U1 with the pre-mRNA. The 70K protein (to which the U1-C is attached) provides further support to the U1-C interaction to the base paired region, by holding up U1-C protein in place, ensuring a stable interaction of the U1 snRNP to the 5'ss (figure 1.23 B) (Pomeranz Krummel et al., 2009). At the 3'ss, the U2 auxiliary factor proteins (U2AF35 and U2AF65) form a heterodimer, with the zinc finger motif of U2AF35 interacting preferentially with a UAGG motif at the 3'ss, and two RNA recognition motifs (RRMs) of U2AF65 binding to a polypyrimidine rich sequence just upstream of the 3'ss (Corsini et al., 2007; Shao et al., 2014; Yoshida et al., 2015). The branch point adenosine forms specific hydrogen bonds to the protein SF1 (Liu et al., 2001). Altogether, at this stage of splicing, both the 5'ss and the 3'ss have been defined and is known as complex E (figure 1.23 C).

Progression to the pre-spliceosome complex A requires the activity of the ATP dependant RNA helicase UAP56, which displaces the SF1 protein allowing the U2 snRNA to base pair at the branch site (Kistler and Guthrie, 2001). Base pairing of U2 causes the branch point adenosine to bulge out and interact with the protein p14 and subsequently SF3B1 (Parker et al., 1987; Wu and Manley, 1989). U2AF65 at the 3'ss makes protein interactions with the U2 component SF3B1 (Chen et al., 2017) (figure 1.23 D). Another U2 protein, SF3A1, interacts with a stem-loop of U1 (Sharma et al., 2014).



The transition from complex A to the pre-catalytic spliceosome (complex B), involves the recruitment of the U4/U6.U5 tri-snRNP, in which the U4 and U6 snRNA are strongly base paired to one another. The 3' end of the U6 snRNA is single stranded and base pairs with U2 (Anokhina et al., 2013).

Progression from complex B to the activated complex B<sup>act</sup> involves the dissociation of the U1 snRNP at the 5'ss and the separation of U4 from U6, both of which are dependent on RNA helicases (Mozaffari-Jovin et al., 2012). With U4 dissociating from U6, and U1 being displaced, U6 base pairs to the 5'ss (Anokhina et al., 2013). This results in a complex in which U6 is pulling U2 towards the 5'ss, so that the 5'ss can come into contact with the branched site for step 1 of the splicing reaction. However, protein SF3B1 is still attached to the branch point adenosine which needs to be removed.

To progress from complex B<sup>act</sup> to the catalytically activated complex B\* requires the dissociation of SF3B1 by the RNA helicase prp2, making the branch point adenosine accessible to attack the 5'ss to achieve step one (Yan et al., 2017). Once step one is achieved, the formation of complex C is obtained which is ready for the second step of the transesterification reaction (Yan et al., 2017). At the end of the reaction, all of the splicing factors dissociate and are recycled for subsequent splicing reactions (Wahl et al., 2009).

### **1.2.3 Mechanism of alternative splicing**

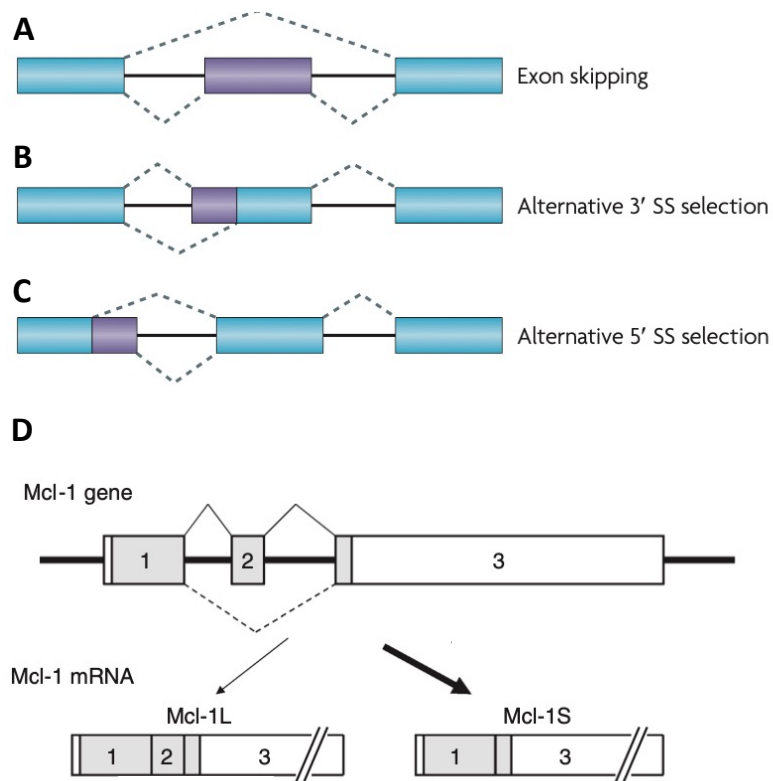
Alternative splicing describes the process in which splicing is regulated to generate multiple mRNAs from a single pre-mRNA. There are four main patterns of alternative splicing that can take place: Exon skipping, alternative 5' splice site (ss) usage, alternative 3'ss usage and intron retention, all of which are summarised below.

Exon skipping (figure 1.24 A) is a common pattern of splicing, accounting for about 40% of total splicing events in eukaryotes (Keren et al., 2010). This occurs when an entire exon is spliced out of the mRNA together with the surrounding intronic sequences. An example of this pattern of splicing can be observed in the protein myeloid cell leukaemia (Mcl-1), which belongs to the Bcl-2 family of proteins. Skipping of exon 2 results in the formation of the shorter pro-apoptotic splice variant Mcl-1<sub>s</sub>, whereas inclusion of this exon gives rise to the longer anti-apoptotic isoform (figure 1.24 D) (Shieh et al., 2009).

Alternative 5' (donor) and 3' (acceptor) ss usage are other possible patterns of AS (figure 1.24 B and C), accounting for around 8% and 18% of total splicing events in humans respectively (Koren et al., 2007). This occurs when there are multiple splice site signals present at either end of the exon which can be used (Keren et al., 2010; Kim et al., 2008). An example of a gene which has alternative 5'ss is Bcl-x, which has two antagonistic splice variants depending on which 5'ss is utilized (discussed in more detail later) (Stevens and Oltean, 2019).

Intron retention is the least common form of alternative splicing (accounting for less than 5% of total splicing events), and is when the intron is retained in the mature mRNA transcript (Keren et al., 2010). This is thought of as a failure of the splicing machinery to define the exon-intron boundary at weaker splice sites, resulting in intron inclusion (Kim et al., 2008). Other studies have also shown that intron retention is more common in genes containing short introns, are highly expressed, and contain a low density of both exonic splicing silencers and intronic splicing enhancers (see later) (Sakabe and de Souza, 2007).





**Figure 1.24 Different patterns of alternative splicing.**

(A) exon skipping. (B and C) alternative 3'ss and 5'ss usage respectively. (D) Pattern of splicing observed with the Mcl-1 pre-mRNA (Keren et al., 2010; adapted from Shieh et al., 2009).

#### 1.2.4 Factors influencing splice site selection

The mechanism of alternative 5'ss selection has been widely studied, as the very first stage in splicing is the base pairing of U1 to the 5'ss which commits a particular splice site to be used in splicing. An initial explanation of such a phenomenon is the ability of the U1 to base pair in the first instance, such that a splice site that most closely resembles the consensus will be biased over the other (Rogers and Wall, 1980). However, after sequencing of the human genome, it was demonstrated that more than 9000 sequences that resemble the 5'ss were not used in the splicing reaction, even with some having an optimum consensus sequence for base pairing to U1 (Roca et al., 2013). Therefore, the question arises as to what silences/enhances a particular splice site if base pairing of U1 is not the crucial step.

#### 1.2.4.1 Splicing enhancers and silencers

Splicing factors are proteins that regulate alternative splicing but are not part of the core spliceosome complex described previously (Fu and Ares, 2014). The first two family of proteins that were shown to regulate alternative splicing were the serine-arginine (SR) rich proteins that act as splicing enhancers, and the heterogeneous nuclear ribonucleoprotein (hnRNPs) that often act as splicing silencers (Fu and Ares, 2014). These proteins bind to *cis* acting sequences in the pre-mRNA known as splicing response elements (SREs). There are different types of SREs within the transcript: intron and exon splicing enhancers (ISEs and ESEs) and silencers (ISSs and ESSs) (Matlin et al., 2005). *Trans* acting factors, such as the SRs and hnRNPs, bind to these SREs and either activate or repress splicing. A single SRE can be bound by multiple different splicing factors, meaning that there is competition between the different trans-acting elements. Depending on which factor is bound would ultimately determine the nature of the splicing reaction. The most extensively studied repressors are hnRNP A/B and hnRNP I (also known as PTB) which antagonises binding of SR proteins (Fu and Ares, 2014).

These *trans*-acting elements have also been shown to interfere with the core spliceosome components to affect splicing. For example, it has been shown that hnRNP A1 competes with the U1 snRNP for binding to the 5'ss, whereas the SR protein SRSF1 competes and displaces hnRNP A1, enhancing U1 binding to the 5'ss (Eperon et al., 2000).

Secondary structure elements that form near splice sites have also been shown to influence splice site selection. For example, the tau protein is a key player in the pathogenesis of Alzheimer disease resulting in neurodegeneration, and it has been shown that retention of exon 10 in the tau mRNA leads to the disease phenotype (Grover et al., 1999). Using RNA structural analysis, it was shown that a stem loop forms near the 5'ss of exon 10. Mutations that remove this stem loop showed retention of exon 10, most probably due to the fact that the 5'ss is made available to the U1 snRNP, resulting in an altered tau protein causing Alzheimer disease. This shows that the stem loop in tau mRNA is providing a protective phenotype and is influencing the splicing behaviour of the gene (Grover et al, 1999). Earlier studies have also highlighted the relevance of RNA secondary structures in splice site selection and usage (Eperon et al., 1988, 1986).

### 1.2.5 Alternative splicing of Bcl-x

Bcl-x is an example of a gene in which alternative splicing has a dramatic influence on its function. Bcl-x pre-mRNA consists of three exons, and within exon 2 lies two alternative 5'ss which have the possibility to give rise to 2 different antagonistic splice variants (figure 1.25 A) (Boise et al., 1993). Use of the upstream site gives rise to the shorter pro-apoptotic Bcl-X<sub>s</sub> isoform, whereas using the downstream site results in the longer anti-apoptotic Bcl-X<sub>L</sub> isoform (figure 1.25 A) (Boise et al., 1993). Therefore, interfering with the mechanism of Bcl-x mRNA splicing by shifting the bias towards either splice variant could determine the fate of a cell. For example, the use of antisense oligonucleotides against the Bcl-X<sub>L</sub> 5'ss shift splicing in favour of the X<sub>s</sub> splice variant, and lead to significant levels of apoptosis in prostate cancer cell lines (Mercatante et al., 2001). This suggests that interfering with the splicing process of Bcl-x may prove to be an advantageous therapeutic strategy to treat cancers where the anti-apoptotic variant is overexpressed. Research into Bcl-x splicing goes beyond cancer therapy, as it was shown that overexpression of the X<sub>s</sub> splice variant resulted in apoptosis of pancreatic  $\beta$ -cells leading to diabetes (Barbour et al., 2015). Therefore, understanding the proteins that regulate the splicing of Bcl-x would help us understand the mechanism of its splice site selection and then derive potential therapeutic molecules to treat such diseases.

#### 1.2.5.1 Proteins that regulate the splicing of Bcl-x

There are many different proteins that regulate the splicing of Bcl-x which include: the SR proteins (SRSF1, SRSF2, SRSF3, SRSF7 SRSF9 and SRSF10); the hnRNP family of proteins (hnRNP A1, PTBP1, hnRNP K and hnRNP F/H); as well as other RNA binding proteins (Sam68, SF3B1, RBM4, RBM11, RBM25 and RBM10), some of which are summarised in figure 1.25 B and are discussed below (Bielli et al., 2014; Cloutier et al., 2008; Dominguez et al., 2010; Garneau et al., 2005; Inoue et al., 2014; Massiello et al., 2006; Merdzhanova et al., 2008; Paronetto et al., 2007; Pedrotti et al., 2012; Revil et al., 2009; Shkreta et al., 2016; Wang et al., 2014; Zhou et al., 2008).

One example is the RNA binding protein Sam68, which belongs to the signal transduction and activation of RNA metabolism (STAR) family of RNA binding proteins (Frisone et al., 2015). The C-terminus of Sam68 was shown to bind to hnRNP A1, and together they activate the upstream 5'ss of Bcl-x, resulting in production of the pro-apoptotic X<sub>s</sub> isoform (Paronetto et

al., 2007). The regulation of Sam68 is heavily dependent on its phosphorylation status by Fyn Kinase, as it was shown that tyrosine phosphorylation of Sam68 reduces its ability to interact with hnRNP A1, hence favouring the use of the downstream  $X_L$  site (Paronetto et al., 2007). SRSF1 was also shown to regulate the activity of Sam68 by inhibiting its action in activating the upstream 5'ss (Paronetto et al., 2007). Earlier studies showed that SRSF1 and hnRNP A1 are in direct competition in binding to the pre-mRNA, and because Sam68 function is dependent on its interaction with hnRNP A1, it is not surprising to see that its effects are neutralised in the presence of SRSF1 (Eperon et al., 2000). In addition, as SRSF1 increases the binding affinity of the U1 snRNP at both sites, due to the double occupancy model (which occurs when both alternative 5'ss are simultaneously occupied by the U1 snRNP), the downstream site is favoured (Eperon et al., 2000).

The ability of SRSF1 to control the activation of the splice sites is also regulated. For example, the cell cycle kinase NEK2 was shown to phosphorylate SRSF1 and prevent the activation of the upstream site (Naro et al., 2014). In a somewhat similar way to hnRNP A1, SRSF1 is also regulated by the protein hnRNP I (also known as polyperimidine tract binding protein 1 (PTBP1)). This protein was shown to bind to a polypyrimidine tract within exon 2 in between the two 5'ss. Binding of PTBP1 was shown to displace SRSF1, activating the use of the upstream site and inhibiting the use of the downstream site (Bielli et al., 2014). In a similar way, the protein RBM4 was also shown to antagonise SRSF1, by binding to the same regulatory element as SRSF1 on the pre-mRNA (Wang et al., 2014).

The proteins hnRNP F and hnRNP H (F/H) are two *trans-acting* factors that have been shown to influence Bcl-x alternative splicing, and bind to and encage G-rich RNA sequences (called G-tracts), which are present just downstream of the  $X_S$  5'ss (Dominguez et al., 2010; Garneau et al., 2005). As well as providing a putative binding site for hnRNP F, G-tracts also have the ability to form structures known as G-quadruplexes (G4s), which consist of base pairing of G nucleotides in a Hoogsteen dependent manner to form a G-tetrad (see next section). Using various biophysical techniques, it has been shown that binding of hnRNP F at the Bcl-x G-tracts and G4 formation are mutually exclusive, and that protein binding is more favourable than G4 formation (Samatanga et al., 2013). This suggests that hnRNP F can indeed compete and prevent the formation of secondary structures in Bcl-x, and this has been implicated as a mechanism to influence Bcl-x alternative splicing. Increasing the amount of hnRNP F leads to an increase in the usage of the upstream site, perhaps by opening up potential *cis* regulatory

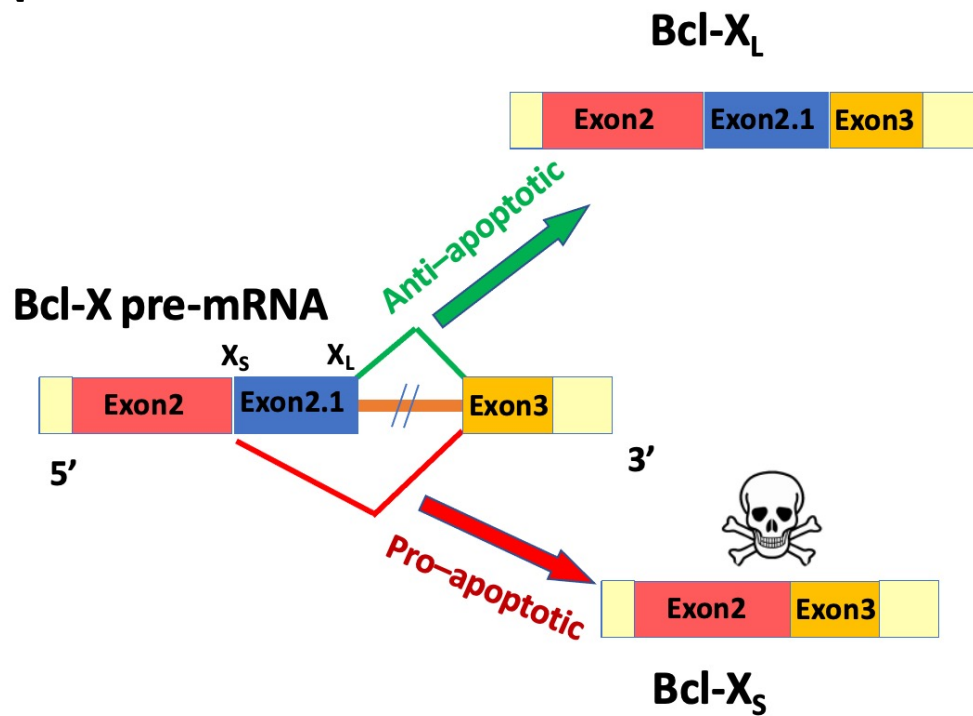
elements that would otherwise have been sequestered in a G4 structure (Dominguez et al., 2010).

A recent study has linked the SR protein SRSF10 to hnRNP F through a DNA damage response. Under normal growing conditions, SRSF10 binds to hnRNP K as well as hnRNP F, sequestering hnRNP F from binding to its *cis* acting G-tract elements. However, upon treating cells with the DNA damaging agent oxaliplatin, SRSF10 is dephosphorylated, causing hnRNP F to dissociate and bind to its regulatory element on the RNA, thereby increasing the X<sub>S</sub> isoform (Cloutier et al., 2018).

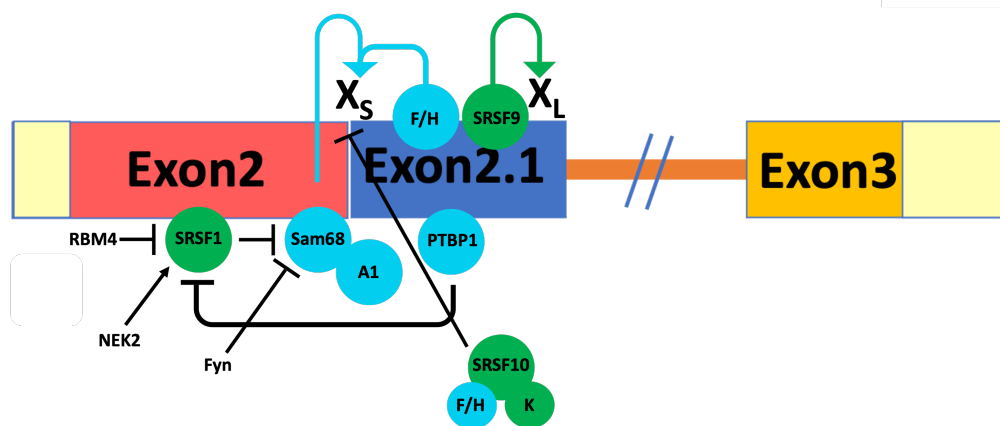
Other splicing factors such as SRSF2, SRSF9 and hnRNP K have also been implicated in Bcl-x splicing, with SRSF2 showing an increase in the X<sub>S</sub> splice site usage, whereas hnRNP K and SRSF9 showed X<sub>S</sub> repression and X<sub>L</sub> activation respectively (Merdzhanova et al., 2008; Michelle et al., 2012; Revil et al., 2009).

In conclusion, there are many different proteins that regulate the splicing pattern of Bcl-x. This paves the pathway for research into intervening at the molecular level with some of these splicing factors as a potential therapeutic strategy, some of which will be explored in the next section.

**A**



**B**



**Figure 1.25 Mechanisms of Bcl-x alternative splicing.**

(A) The two splice variants of Bcl-x. (B) Splicing factors which regulate splice site selection on the Bcl-x pre-mRNA. Green circles are those that suppress the X<sub>S</sub> and activate the X<sub>L</sub> isoform, whereas blue circles are factors that activate the X<sub>S</sub> and suppress the X<sub>L</sub> isoform (adapted from Mercatante et al., 2001; adapted from Stevens and Oltean., 2019).

### 1.2.6 Targeting the splicing machinery as a therapeutic strategy

One of the key factors in cancer cell survival is their ability to evade apoptosis (see previous section), giving them the ability to be insensitive to chemotherapeutic agents (Stevens and Oltean, 2019). One mechanism cancer cells deploy in their ability to evade apoptosis is by overexpressing the anti-apoptotic Bcl-X<sub>L</sub> isoform (Fulda, 2009). Therefore, finding treatments which alter the splicing pattern of Bcl-x to inhibit the expression of the X<sub>L</sub> isoform and activate the X<sub>S</sub> isoform would have great therapeutic potential and ultimately increase the likelihood of success with therapeutic agents.

As mentioned previously, the splicing factor SRSF1 has been shown to upregulate Bcl-X<sub>L</sub>, suggesting that downregulation of this protein would potentially increase the X<sub>S</sub>/X<sub>L</sub> ratio. The cell cycle kinase AURKA was shown to regulate the expression of SRSF1, and consequently inhibition of AURKA by the inhibitor VX-680 resulted in down regulation of SRSF1 and an increase in the X<sub>S</sub> isoform in cervical cancer cell lines (Moore et al., 2010). Furthermore, another study showed that the chemotherapeutic agent, gemcitabine, which inhibits the DNA synthesis pathway, also results in overexpression of SRSF1, meaning that its effects as an anti-cancer compound are attenuated. Silencing SRSF1 using siRNA knockdown increased the potency of gemcitabine as a chemotherapeutic agent (Adesso et al., 2013). In a similar fashion, the anti-cancer compound NSC606985 (a camptothecin derivative) was shown to be more effective in triggering apoptosis only after silencing hnRNP K (known to activate the X<sub>L</sub> isoform (see earlier)) (Gao et al., 2009). Therefore, targeting the key splicing factors known to influence splice site selection is a promising indirect approach to increase the effectiveness and potency of chemotherapy.

Another effective therapeutic strategy to alter pre-mRNA splicing is to use splice-switching oligonucleotides (SSOs). SSOs are anti-sense oligos that are designed to hybridise to the pre-mRNA at specific positions to block either RNA-RNA or RNA-protein interactions to bias splice site selection (Havens and Hastings, 2016). These SSOs require a specific 2'-O-methoxyethyl (MOE) modification to the ribose to prevent RNase H degradation once in the cell, and can be administered *in vivo* using lipid nanoparticles to increase cellular uptake efficiency (Bauman et al., 2010). Various different studies using a variety of different cancer cell lines have shown

that SSOs that target the downstream  $X_L$  5'ss increase the  $X_S$  isoform and reduce cell viability (Li et al., 2016; Mercatante et al., 2002). However, the effectiveness of SSOs depends heavily on the expression profile of Bcl- $X_L$ , in which cancer cells that have greater levels of Bcl- $X_L$  expression are more susceptible to SSO treatment (Mercatante et al., 2001).

In addition to blocking protein or RNA binding sites using SSOs, accessibility of splicing factors (and thus determining splice site selection) can also be modulated through the secondary structure of the pre-mRNA alone. For example, aberrant splicing of the tau mRNA (thought to be involved in dementia) was shown to be modulated by its secondary structure. A stable stem loop was shown to sequester the 5'ss in the tau mRNA, thereby preventing its interaction with U1 and inhibiting splicing (Grover et al., 1999). The aminoglycoside antibiotic, neomycin, was shown to bind and stabilise this stem loop and has the potential to alter the splicing pattern of Tau (Varani et al., 2000). Furthermore, our lab has shown that the putative G4 specific ligand GQC-05 has the ability to stabilise a G4 in exon 2 of the Bcl-x pre-mRNA, and has been shown to increase the  $X_S/X_L$  ratio (see section 1.7) (Weldon et al., 2018).

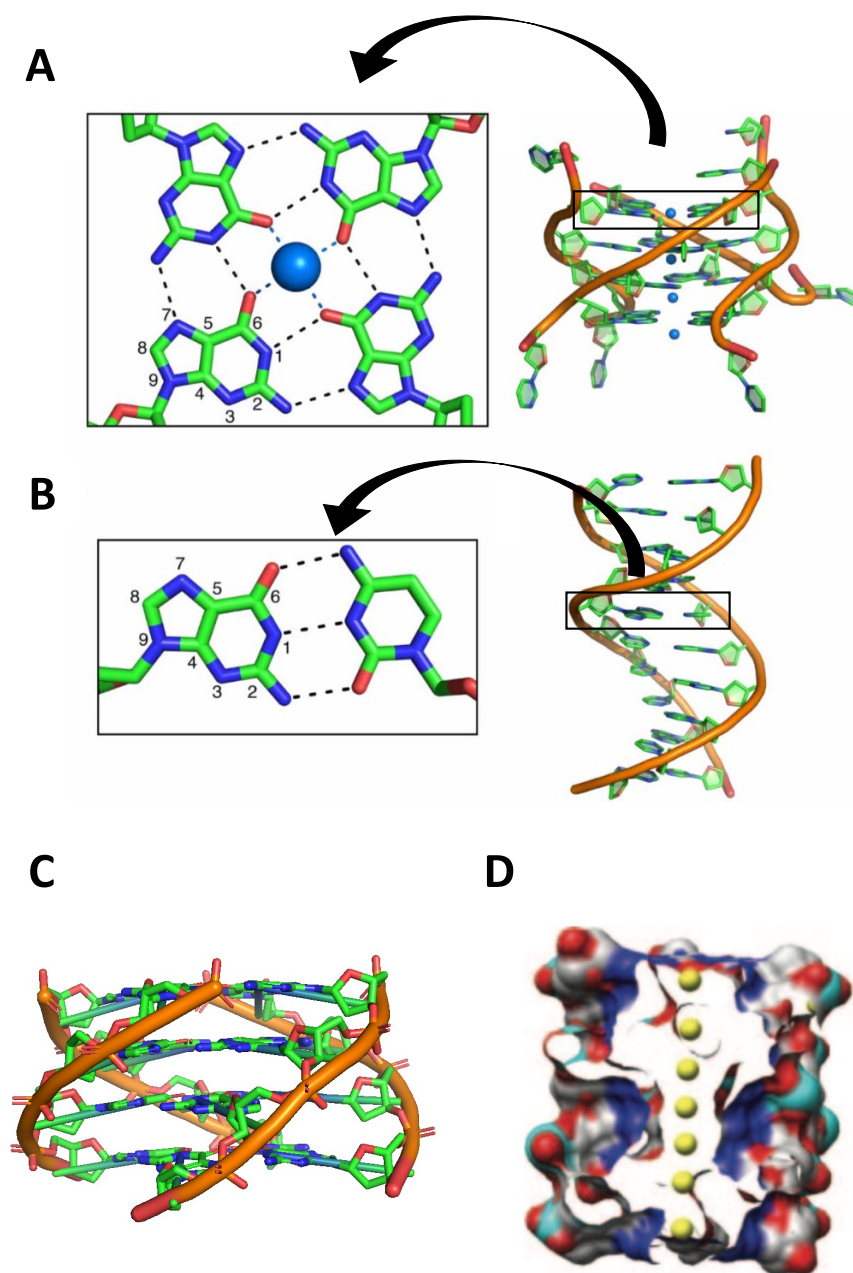
Despite these vast arrays of treatment strategies, another consideration is how we can target these therapeutic agents to tumour cells without having off-target effects. For example, compounds or SSOs which increase the  $X_S$  isoform may be advantageous for cancer therapy, but would not be beneficial if they also promote apoptosis of pancreatic  $\beta$ -cells. Therefore, targeted treatment is also very important to consider. Recent studies have shown the use of exosomes, which are small vesicular structures that can be artificially packaged with anti-cancer compounds, have the ability to specifically target tumour cells (Dai et al., 2020; Kim et al., 2016).

In conclusion, there are several strategies to target the splicing machinery as potential treatment options in cancer. Of particular interest to our study is the ability of RNA secondary structure elements to alter splice site selection in Bcl-x, particularly the relevance and effects of potential G-quadruplex structures that have been predicted to form in specific areas of the Bcl-x pre-mRNA (see later) (Weldon et al., 2018).



### 1.3 The G-quadruplex (G4)

The best known DNA structure is the B-form DNA helix, in which base pairing occurs between adenine and thymine (A-T) or Guanine and Cytosine (G-C) (Crick and Watson, 1953). However, over a century ago, the formation of another type of DNA secondary structure, known as a DNA quadruplex, was suggested. This was discovered by the formation of a viscous gel like material when guanylic acids were present at high concentrations in solution, suggesting higher order polymerisation (Bang, 1910). In 1962, the first proposed structure of a G-quadruplex was published by X-ray diffraction, where they diffracted the fibres formed from GMP rich solutions. From their data they proposed that four guanines come together with a fourfold rotational axis, with each guanine forming four hydrogen bonds with adjacent guanines, consisting of an inner and outer network of hydrogen bonds between N<sub>1</sub>-O<sub>6</sub> (inner) and N<sub>2</sub>-N<sub>7</sub> (outer) atoms, thus forming a tetramer like arrangement of the guanines (known as the G-tetrad, figure 1.31 A) (Gellert et al., 1962). Note that in the classical Watson-Crick base pairing, only the N<sub>1</sub>-N<sub>2</sub>-O<sub>6</sub> interface forms the three hydrogen bonds with the cytosine, with no hydrogen bonds emanating from N<sub>7</sub> (figure 1.31 B) (Weldon et al., 2016). Their diffraction pattern data was also consistent with the formation of tetrad aggregates, where the large planar surface of each tetrad would form strong  $\pi$ - $\pi$  stacking interactions with another tetrad, forming a four stranded cylindrical structure with a cavity in the middle, able to fit water molecules or monovalent cations (figure 1.31 C & D) (Burge et al., 2006; Gellert et al., 1962; Zimmerman et al., 1975). G4 structures are particularly stable in the presence of monovalent cations, as the electronegative inward pointing O<sub>6</sub> from each of the guanine bases are in close proximity, resulting in electrostatic repulsion and thus making the structure thermodynamically unfavourable. The presence of a monovalent cation within the cavity enables co-ordination of the O<sub>6</sub>, thereby increasing the stability of the G4 (Bochman et al., 2012). The type of ion that best stabilises a particular G4 structure can vary depending on the structure of the quadruplex (see later), however the general rule that fits many of the known G4 DNA structures is: Potassium (K<sup>+</sup>) > Rubidium (Rb<sup>+</sup>) > Sodium (Na<sup>+</sup>) > Lithium (Li<sup>+</sup>) or Caesium (Cs<sup>+</sup>) (Sen and Gilbert, 1990).

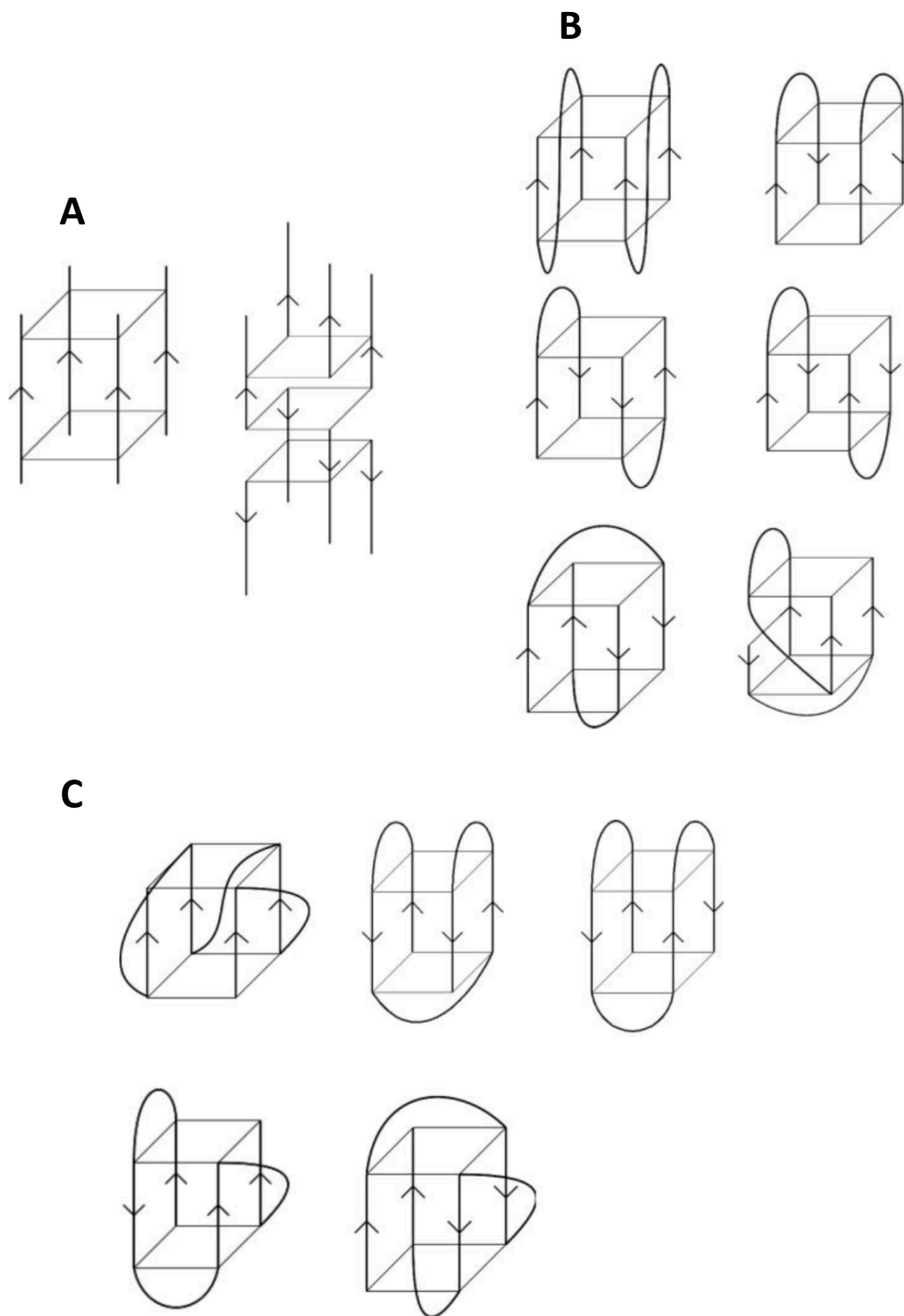


**Figure 1.31 G-quadruplex and duplex DNA structures.**

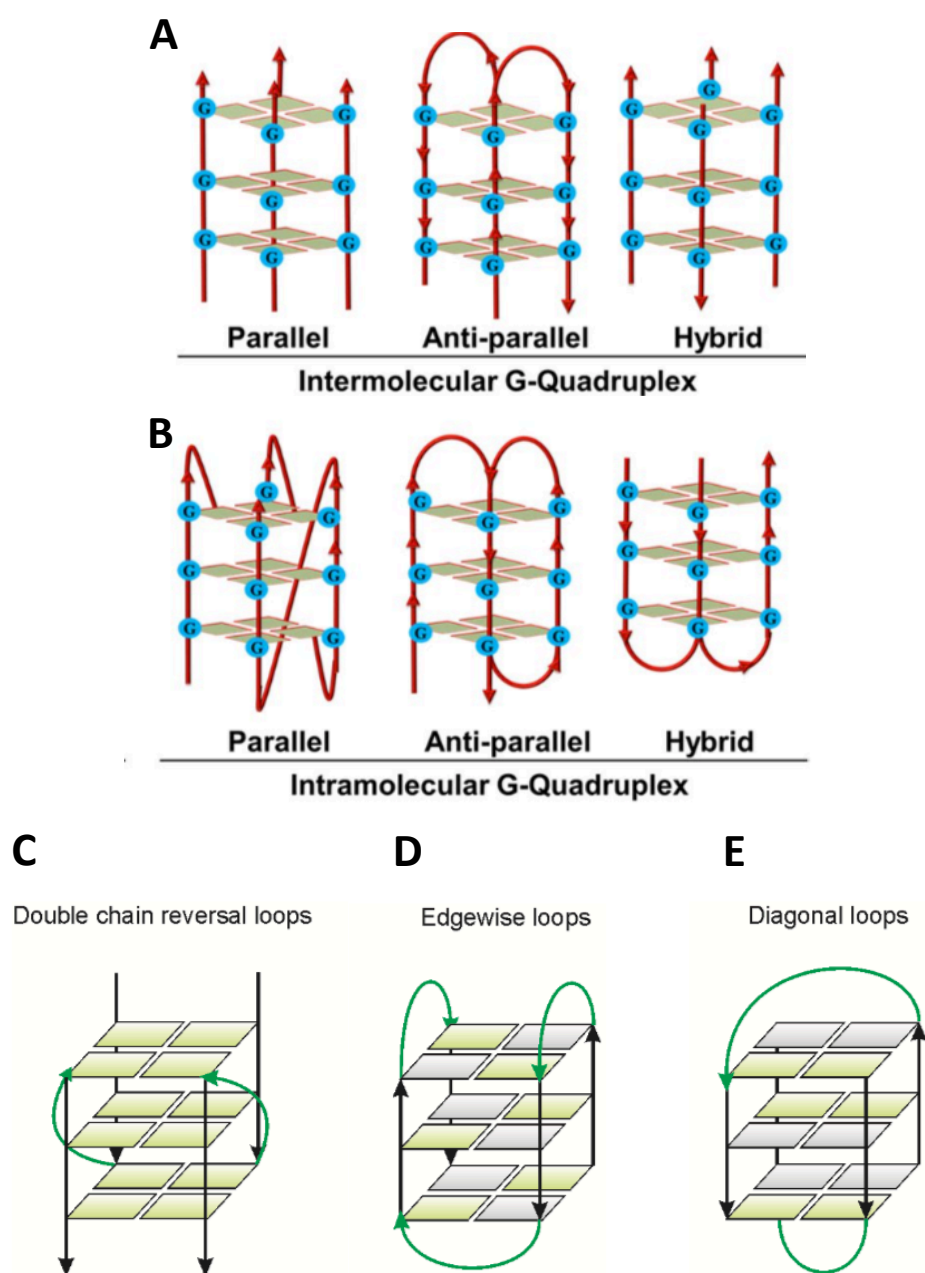
(A) Parallel G4 structure of the planar tetrad stabilized a Hoogsteen hydrogen bond network and a monovalent cation (left). Also shown is the four stranded parallel G4 structure formed by the association of the tetrads (right) (PDB: 244D). (B) Duplex structure of the Watson-Crick hydrogen bonds between GC base pairs (left). And the DNA duplex structure formed by Watson-crick base pairings (right) (PDB: 1BNA). (C) Aggregation of the tetrads shown in (A) (PDB: 244D) to form the G4 structure. (D) The central cavity of the G4, which is able to accommodate monovalent cations (Burge et al., 2006; adapted from Weldon et al., 2016; Zimmerman et al., 1975).

In addition to the stabilising ion present in solution, DNA and RNA quadruplexes can be described as having at least two stacked planar tetrads formed from either one, two or four individual strands, in which the tetrads can be interconnecting through a series of loops (in the case of two or four strands) consisting of a random sequence of nucleotides (Burge et al., 2006).

An example of a unimolecular (or intramolecular) quadruplexes is shown in figure 1.32 C, which consists of a single DNA or RNA strand, able to fold up into four strands, with the two stacked tetrads forming between the stretches of guanines within the primary sequence. Another possibility is the formation of intermolecular quadruplexes, in which the guanines in the G-tetrad arise through the contribution of either two (as is the case for bi-molecular) or four (tetra-molecular) strands of DNA or RNA (shown in figures 1.32 A & B respectively) (Bochman et al., 2012; Burge et al., 2006). Diversity of quadruplexes can also be enhanced depending on the polarity of the strands. Parallel quadruplexes have all four strands (inter or intra-molecular) pointing in the same direction (i.e. the 5' and 3' ends of each strand are on the same side), whereas anti-parallel G4s have strands in opposing directions. Hybrid conformations are also possible, for example the 3+1 hybrid in which three of the four strands point in the same direction, with one strand in the opposing direction (figures 1.33 A & B) (Mishra et al., 2019). In addition to the polarity and strand stoichiometry, further diversification of G4s have been reported, in which the length and location of the interconnecting loops can be variable. For example, parallel type G4s require a loop connection between the bottom tetrad and the top tetrad, and hence often display so-called double chain reversal (or propeller) loops (figure 1.33 C). Certain G4s can also display edge-wise (or lateral) loops that connect together two adjacent G strands (figure 1.33 D). A third possibility is a diagonal loop that links two opposite G-strands (figure 1.33 E) (Malgowska et al., 2016). Therefore, this shows that G4s can be very structurally diverse in nature and extensive research is being undertaken to understand the functional relevance of such a diverse array of quadruplex topologies, some of which will be explored later.



**Figure 1.32 Different possible strand stoichiometry of G4s.**  
 (A) Tetramolecular, (B) bimolecular and (C) unimolecular G4s (Burge et al., 2006).



**Figure 1.33 Strand polarities and loop orientations.**

(A and B) Schematic of the different strand polarities for both inter and intramolecular quadruplexes (C – E) The different loop orientations in G4s. (C) Double chain reversal (or propeller) loops. (D) Edgewise (or laterel) loops and (E) diagonal loops (Mishra et al.,2019; Malgowska et al., 2016).

## 1.4 G4s in Biology

To date, there have been around 300 structures of quadruplexes that have been deposited in the PDB, with 75% of them being DNA structures (Miskiewicz et al., 2020). Discussed below are some examples of G4 structures and their function, with more emphasis on RNA quadruplexes which only make up a quarter of all known G4 structures in the PDB.

### 1.4.1 DNA G4s

The first DNA G4 that was identified and implicated to have a biological role was the human telomeric repeat sequence G4, in which formation of G4s at the 3' overhangs prevented the action the *Oxytricha nova* telomerase *in vitro* (Zahler et al., 1991). Initial structural work using NMR showed that the telomeric overhang sequence d(G4T4G4) in *Oxytricha nova*, can fold into an antiparallel bimolecular quadruplex, and by increasing the repeat sequence to d(G4T4G4T4G4T4G4) resulted in a antiparallel intramolecular quadruplex which the latter was shown to specifically inhibit telomerase elongation (Oganesian et al., 2006; Smith and Feigon, 1992). As telomerase is required for the replicative immortality of certain cells, it is not surprising that this enzyme is upregulated in >85% of tumours, meaning that telomeric G4 sequences could be potent therapeutic targets in the treatment of certain cancers (Feldser and Greider, 2007; Moye et al., 2015; Shay and Bacchetti, 1997). As such, extensive work has been undertaken to solve the structure of the human telomere G4. Studies into the human telomeric repeat sequence, d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>], showed that this G4 displays a wide array of different topologies depending on the nature of the cationic environment, with the formation of an anti-parallel G4 in the presence of Na<sup>+</sup> and a hybrid conformation in the presence of K<sup>+</sup>, indicating the vast structural polymorphism of this sequence (Parkinson et al., 2002; Wang and Patel, 1993).

In addition to their presence at telomeres, DNA G4s have also been shown to be abundantly clustered in about 60% of human gene promotor regions, suggesting that they may have a regulatory mechanism to control gene expression patterns (Huppert and Balasubramanian., 2005; Chambers et al., 2015). G4s were identified in the promoters of various oncogenes, such as C-myc, Bcl-2 and KRAS, in which their functional relevance and structures have been determined (Agrawal et al., 2014; Ambrus et al., 2005; Dai et al., 2006; Dickerhoff et al., 2019; Kerkour et al., 2017).

### 1.4.2 RNA G4s

In contrast to DNA G4s, the structure and functions of RNA G4s have been reported to a much lesser extent. This could be due to the fact that RNA is much more prone to degradation than its DNA counterpart, both by the presence of ubiquitously found RNases and high pH values, making RNA studies more laborious than DNA. In addition, the cost of RNA oligonucleotides are considerably higher than DNA, making structural and biophysical work more expensive as large quantities of material are required. In addition, many of the DNA G4 structures solved by NMR relied on the use of site-specific labelling of DNA nucleotides in order to assign the NMR spectra for structural determination. However, due to the lack of affordable labelled RNA phosphoramidites, site-specific labelling of RNA oligonucleotides becomes rather expensive, making this technique unaffordable, resulting in a lower proportion of RNA structures on the PDB (Adrian et al., 2012). Despite these drawbacks, there have been a limited number of RNA structures that have been solved by both NMR and X-ray crystallography.

One of the most extensively studied RNA G-quadruplexes, both structurally and functionally, is the non-coding telomeric repeat-containing RNA (TERRA). It was initially thought that the telomeres of chromosomes were transcriptionally silent, however this was disproven over 10 years ago when it was demonstrated that telomeres were actually being transcribed into TERRA RNA (with heterogeneous lengths, containing repeats of r(UUAGGG)) (Azzalin et al., 2007).

Several studies have demonstrated that one of the major functions of the TERRA G4 RNA is to modulate histone modification and heterochromatin formation at the telomeres (Arnoult et al., 2012; Takahama et al., 2013). It was shown that TERRA G4 RNA acts in a negative feedback loop to suppress its own transcription from the DNA telomeric sequence (Arnoult et al., 2012). The G4 formed by both TERRA and the DNA telomere was shown to be a docking site for the telomere binding protein TLS, which binds via its Arg-Gly-Gly (RGG) motif in a G4 specific manner. Upon forming the TLS-G4 complex, histone methyltransferases (such as the Suv4-20h) are recruited, resulting in histone H4K20 and H3K9 tri-methylation which in turn leads to telomere shortening, heterochromatin formation and transcriptional repression (Takahama et al., 2013). Other functions of TERRA RNA have also been reported, such as the

suppression of innate immune genes in cancer cells and the competitive inhibition of telomerase (Hirashima and Seimiya, 2015; Redon et al., 2010).

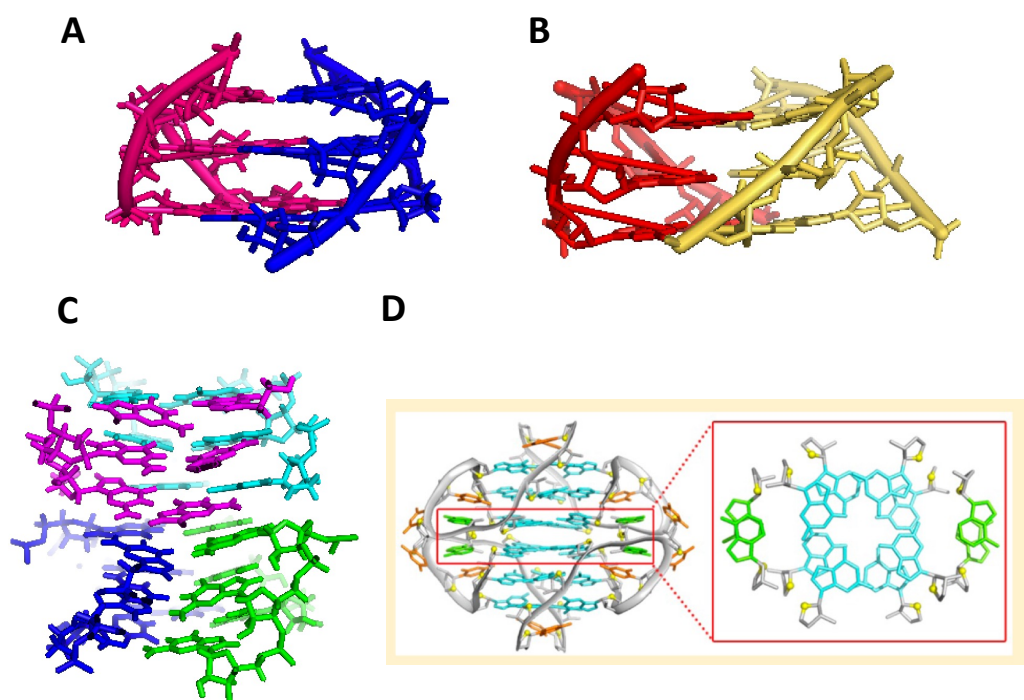
RNA G4s have considerably less topological diversity than their DNA counterparts, with most structures exhibiting the parallel conformation over the anti-parallel and hybrid forms. This occurs as a consequence of the 2'hydroxyl group, which locks the bond between the nucleobase and the sugar in the *anti*-conformation, making the parallel conformation more favourable (Song et al., 2016). This is indeed the case for the published structures of TERRA (figures 1.41 A & B). The very first structure of TERRA was solved by NMR in 2009, in which the 12 mer RNA sequence r(UAGGGUUAGGGU) resulted in the formation of a bimolecular parallel propeller type quadruplex (figure 1.41 A) (Martadinata and Phan, 2009). It was also shown by gel electrophoresis that shorter 9 or 10 mer variations of the sequence resulted in higher order polymerisation via 5' stacking of the G4 blocks (Martadinata and Phan, 2009). A year later, an almost identical structure was solved for the same sequence by X-ray crystallography (figure 1.41 B) (Collie et al., 2010a). The possibility of the formation of higher order structures through 5' stacking attracted much attention, due to the potential of deriving small molecules that may bind to the 5' interface with the ability to potentially stabilise the stacking interactions. However, very little structural information about these higher order structures were known at the time. In 2013, the first solution structure of the stacked TERRA RNA was published, which involved the 10 mer RNA sequence: r(GGGUUAGGGU) that was previously shown to form higher order structures by gel electrophoresis (Martadinata and Phan., 2009). This structure (shown in figure 1.41 C) showed the formation of a bi-molecular dimer, in which 2 bimolecular quadruplexes form 5' to 5' stacking interactions with another bi-molecular dimer, mediated by the 5' tetrads as well the adenine in the UUA loop (figure 1.41 D), forming a planar A(G.G.G.G)A hexad at the interface (Martadinata and Phan, 2013). This paved the way for future work in deriving ligands that are able to bind at the interface and stabilise its structure.

In addition to their roles in telomeric RNAs, there is growing evidence on the role of RNA G4s in regulating the alternative splicing of mRNAs, such as the TP53 mRNA and the mRNA of the human transcription factor Pax9 (Marcel et al., 2011; Ribeiro et al., 2015). Another study has shown the involvement of an RNA G4 in promoting exon inclusion in the CD44 pre-mRNA (Huang et al., 2017). More recently, our lab has postulated the involvement of RNA G4s in the



alternative splicing of the Bcl-x pre-mRNA, which will be discussed in more detail in section 1.7 (Weldon et al., 2018).

RNA G4s have also been reported to have many other functions such as: mRNA transcription and 3' end processing (Dalziel et al., 2007; Zheng et al., 2013); mRNA localisation (Subramanian et al., 2011) and mRNA translation (Bugaut and Balasubramanian, 2012). Therefore, methods that can help identify the presence of G4 structures in the human transcriptome will be crucial into understanding their functional relevance in various biological processes, and will be explored in the next section.



**Figure 1.41 Structure of the TERRA RNA G4.**

(A and B) Structure of the 12 mer TERRA RNA sequence r(UAGGGUUAGGGU) by NMR (PDB: 2KBP) (A), and X-ray crystallography (PDB: 3IBK) (B). (C) First higher order NMR structure formed by the TERRA 10 mer sequence r(GGGUUAGGGU) with each of the four chains highlighted in magenta, cyan, blue and green. Magenta and cyan chains form one bi-molecular G4, which is stacked on top of another bi-molecular G4 formed from the blue and green chains. (D) Zoom of the stacking interface of (C), with the adenine (green) contributing to the A(G.G.G.G)A hexad (structures in A-C were generated in pyMOL) (Collie et al., 2010a; Martadinata and Phan., 2009; Martadinata and Phan., 2013).

## **1.5 Characterisation of G4 structures in nucleic acids**

There are many ways to characterise putative G4 forming regions in RNA and DNA, which involve bioinformatic, biophysical, biochemical and functional approaches. However, each approach has its advantages and limitations and will be discussed below.

### **1.5.1 Bioinformatic approaches**

The very first step to characterise a putative G4 (or any nucleic acid structure) is to analyse its nucleotide sequence. In fact, bioinformatic analysis alone predicts the formation of over 1 million G4 forming sequences within the human genome (Vannutelli et al., 2020). There have been many studies which have used bioinformatic tools to predict the presence of G4s in the human genome, such as in promoters of proto-oncogenes, telomeres and at 5'UTRs, and is therefore the initial stage of structural characterisation of a given sequence (Beaudoin et al., 2014; Eddy and Maizels, 2006; Huppert et al., 2008; Huppert and Balasubramanian, 2005; Jodoin et al., 2014).

The first bioinformatic program that could predict G4 structures from a putative quadruplex sequence (PQS) was published in 2004 and 2005 (D'Antonio and Bagga, 2004; Huppert and Balasubramanian, 2005). Since then, several others have been developed which differ in the nature of the PQS that they are able to predict, and are summarised in table 1.511. These vast arrays of different bioinformatic tools can be used to predict G4s in a wide variety of different contexts, including those that may not follow the canonical G4 topology (see below) (Miskiewicz et al., 2020). For example, G4 hunter can be used to predict G4s with mismatching or bulging nucleotides which do not follow the canonical fold, whereas QGRS mapper can only predict those following a canonical topology (Bedrat et al., 2016; Kikin et al., 2006). These two programmes will be discussed in more detail below.

Tool	DNA	RNA	Multiple entry	Allow mismatches	Number of tetrads in G4	Loop length	Max PQS'
G4Catchall	✓	✓	✓	✓	$\geq 2$	1–50	99
G4Hunter	✓	✓		✓	$\geq 0$	$\geq 0$	
G4-iM Grinder	✓	✓		✓	3	$\geq 0$	
G4P Calculator	✓		✓	✓	$\geq 0$	$\geq 0$	
G4Predict	✓	✓		✓	$\geq 0$	$\geq 0$	
G4-Predictor V.2	✓	✓			2–6	0–36	45
G4PromFinder	✓				2–4	1–10	30
G4RNA screener		✓	✓	✓	$\geq 0$	$\geq 0$	
ImGQfinder	✓	✓		✓	2–10	$\leq 25$	
pqsfinder	✓	✓		✓	2–20	0–9	
QGRS Mapper	✓	✓			2–6	0–36	45
QPARSE	✓	✓	✓	✓	$\geq 2$	$\geq 0$	
Quadron	✓				$\geq 3$	1–12	
TetraplexFinder	✓	✓	✓	✓	2–5	1–50	170

**Table 1.511 Features of known G4 prediction bioinformatic tools.**

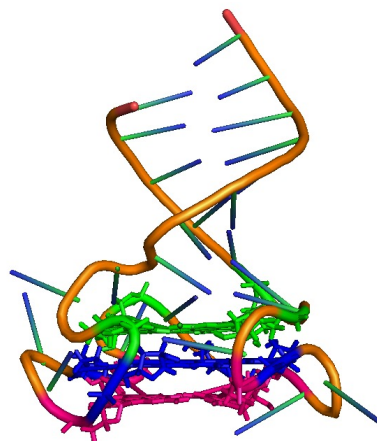
With regards to each of the columns in the table: DNA and RNA indicates if the programme accepts DNA and/or RNA sequences; multiple entry means that multiple sequences can be imputed in a single run; allow mismatches implies that the programme will accept bulges and interruptions within the G-tracts thereby providing a platform to predict non-canonical as well as canonical PQS; number of tetrads in the G4 indicates the maximum number of tetrads that a particular programme can predict; loop length indicates the maximum number of nucleotides a loop can adopt for the programme to predict a G4; max PQS is the maximum sequence length of a PQS that can be imputed into the programme for a successful prediction (review by Miskiewicz et al., 2020).

QGRS mapper can predict the formation of a unimolecular RNA or DNA G4 with the canonical G4 sequence of  $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$ , where  $x$  is the number of consecutive guanines (which must be between 2 and 6) and  $y$  is the length of the connecting loops (which can be up to 36 nucleotides) (Kikin et al., 2006). Depending on the number of tetrads and the size of the connecting loops, a given sequence is assigned a G-score, the higher the score the more likely that this will fold into a canonical G4. The greater number of tetrads gives a higher score, as these G4s are more stable due to a greater number of  $\pi$ - $\pi$  stacking interactions. In addition, shorter loops give higher scores as longer loops have been shown to reduce the stability of G4s (Guédin et al., 2010). The highest score predicted by QGRS mapper is 105, consisting of 6 tetrads with 1 nucleotide connecting loops (Kikin et al., 2006). This programme can accurately predict the structure of a DNA G4 aptamer, which consists of a four tetrad unimolecular quadruplex in which the QGRS output prediction was consistent with the NMR determined structure (Smith et al., 1995). However, a disadvantage of using QGRS mapper is that it will not correctly predict the formation of quadruplexes outside of the canonical sequence, such as the structure of the RNA G4 *sc1*, which is involved in binding the Fragile X mental

retardation protein (FMRP) (Phan et al., 2011). The solution structure showed the formation of a three-tiered quadruplex, but QGRS mapper did not predict the formation of one of the tetrads and only predicts a two stack G4, emphasising its limitation with non-canonical structures (figure 1.512). False positives have also been identified, in which the programme predicts the formation of a stable G4 when in fact the sequence does not fold into a quadruplex *in vitro* , further emphasising on the limitations of QGRS mapper to predict such structures (Beaudoin et al., 2014). Therefore, sequence analysis alone cannot give an accurate representation on the number of PQS that may be present in the human genome, as the true value may be an over or underestimate due the limitations of predicting non-canonical structures (Beaudoin and Perreault, 2010; Bedrat et al., 2016).

*Sc1*: GCUGC**GGUGUGGAAGGAGUGGUCGGG**UUGCGCAGCG

Position	Length	QGRS	G-Score
6	16	<b>GGUGUGGAAGGAGUGG</b>	20



**Figure 1.512 Structure of a non-canonical G4s not predicted by QGRS mapper.**

The nucleotide sequence of the *Sc1* RNA (top panel) with the G's involved in the tetrad formation color coded. Also shown is the output of QGRS mapper for this sequence, with the predicted G's involved in tetrad formation highlighted in blue, suggesting a 2 stacked tetrad conformation (middle panel). Shown in the lower panel is the structure of this RNA (PDB: 2LA5), showing a 3 stacked tetrad conformation, with the tetrads color coded according to the G's in the nucleotide sequence in the top panel (adapted on pyMOL) (Phan et al.,2011).

To overcome the issue of predicting non-canonical PQS, other bioinformatic tools have also been developed such as G4 hunter. This particular programme uses an algorithm that can detect PQS based on the G-richness and G-skewness of a particular sequence (Brázda et al., 2019). Within the input sequence, each G or C nucleotide is given an arbitrary score, depending on the number of G-tracts it is present in. For example, a single guanine (G) will be assigned a score of 1, (with GG = 2, GGG = 3, GGGG = 4.....). Cytosines will have a negative score in a similar fashion, depending on the number of C-rich tracts present (C = -1, CC = -2, CCC = -3, CCCC = -4.....). The overall score of a sequence will indicate its propensity to fold into a G4, meaning that guanines which may not be part of a continuous G-tract (i.e not a canonical PQS) may also be represented and evaluated as potentially participating in tetrad formation, opening up the possibility of G4 formation in a much wider context (Brázda et al., 2019).

Based on sequence analysis, our lab has identified six putative regions within the Bcl-x pre-mRNA that can potentially adopt a PQS, termed Q1-Q6 in table 1.513 (Weldon et al., 2017). Analysis of the QGRS score (which maps out canonical G4s only), indicates that the PQS within the Q5 segment has the greatest propensity to form a G4 structure, shown by the highest QGRS score of 21, in contrast to the other sequences which fall short of this score. However, when comparing the scores obtained by G4 hunter, the Q2 PQS has the greatest propensity to fold a G4, shown by a score of 1.059. This suggests that although Q5 has the potential to fold a relatively more stable canonical G4 compared to Q2 (based on the QGRS scores), Q2 has the ability to potentially fold into a more stable non-canonical structure than a canonical Q5 quadruplex. Therefore, an array of different G4 predictive tools can give insight into the nature of a quadruplex that can be formed by a particular sequence, which then paves the pathway for experimental evidence of such structures.

Name	Bcl-x PQS	QGRS score	G4 hunter score
Q1	GAGGGAGGCAGGCGACGAGUUUGAACUGCGGUA	6	0.667
Q2	CCGGGAUGGGGUAAACUGGGGUCGCAUUGUGGCC	17	1.059
Q3	UCGGCGGGGCACUGUGCGUGGAA	11	0.957
Q4	AAGGAGAUGCAGGUAAUUGGUGAGUCGGAU	18	0.621
Q5	UUGGAUCCAGGAGAACGGCGGCUGGGUA	21	0.679
Q6	CUGGUUCCUGACGGGCAUGACUGUGGCCAGGUA	15	0.364

**Table 1.513 QGRS and G4 hunter score outputs for the six PQS regions of the Bcl-x pre-mRNA (adapted from Weldon et al., 2017).**

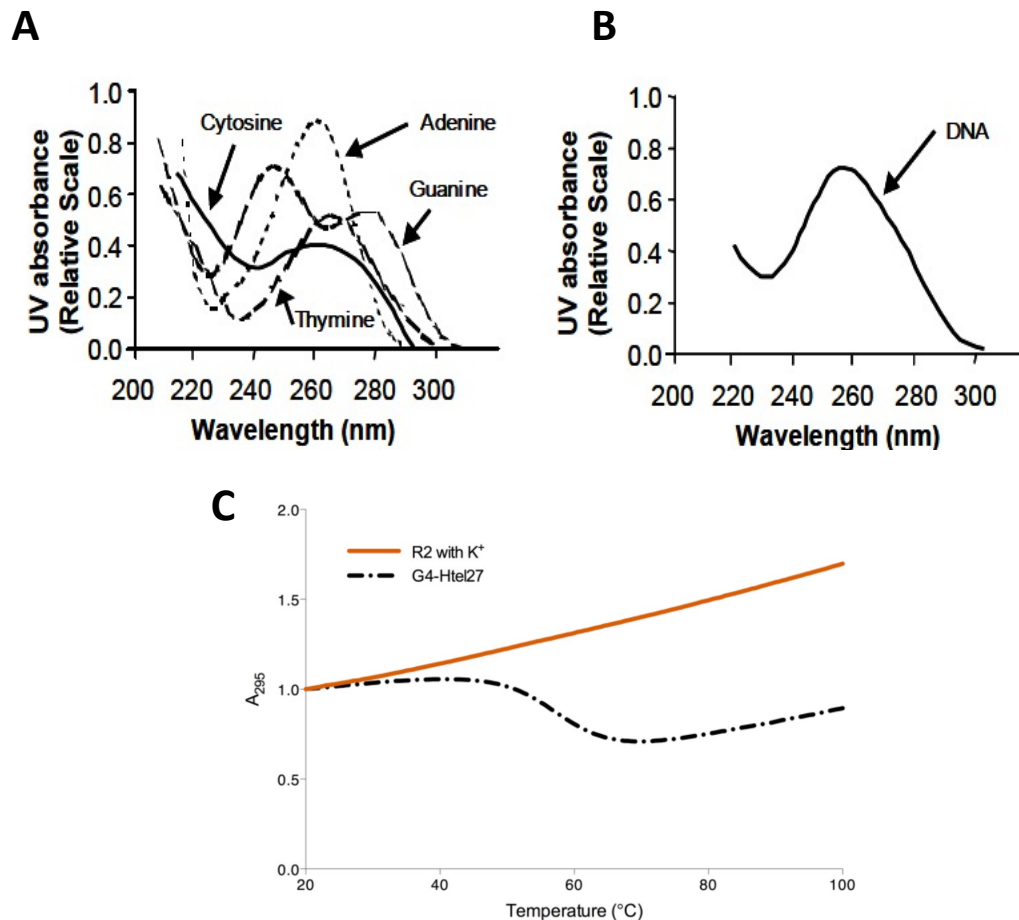
### 1.5.2 Biophysical approaches

After analysing a nucleotide sequence containing a putative G4 forming region using bioinformatic tools described above, these short identified regions can then be isolated and examined further using various biophysical techniques to confirm or disprove the presence of a G4. Discussed below are several methods that have been employed to characterise G4s in nucleotide sequences, each with their advantages and limitations.

#### 1.5.2.1 UV spectroscopy

Each of the four nucleotide bases display a unique absorption spectrum, in which the maximum is distinctive of that particular nucleotide (figure 1.514 A). Within a given sequence there would be a mixture of each of the bases, which would result in the summation of the absorption spectrum of each individual base, giving rise to an overall absorption with a maximum at 260 nm (figure 1.514 B) (Otim, 2018). During denaturation, double-stranded DNA or RNA displays hyperchromicity at 260 nm. This is due to reduced base stacking from structural unfolding, resulting in an increase in the UV absorption of the nucleotide bases (Harvey Lodish et al, 2000; Puglisi and Tinoco, 1989). The overall stability of a particular RNA or DNA structure can be measured by following the absorption at 260 nm as a function of temperature, which would provide information regarding the overall stability of a particular RNA or DNA structure. However, this would not provide much information as to the type of secondary structure that is undergoing the melting process, although it has been shown that

DNA G4s display hypochromicity at 295 nm over other structures, for example with the human telomere 27-mer G4 forming sequence (TTA GGG)<sub>4</sub> TTA (G4-Htel27) (figure 1.514 C), meaning that one can follow this wavelength (instead of at 260 nm) as more of a direct measure of the G4 melting process (Del Mundo et al., 2017; Mergny and Lacroix, 2009).

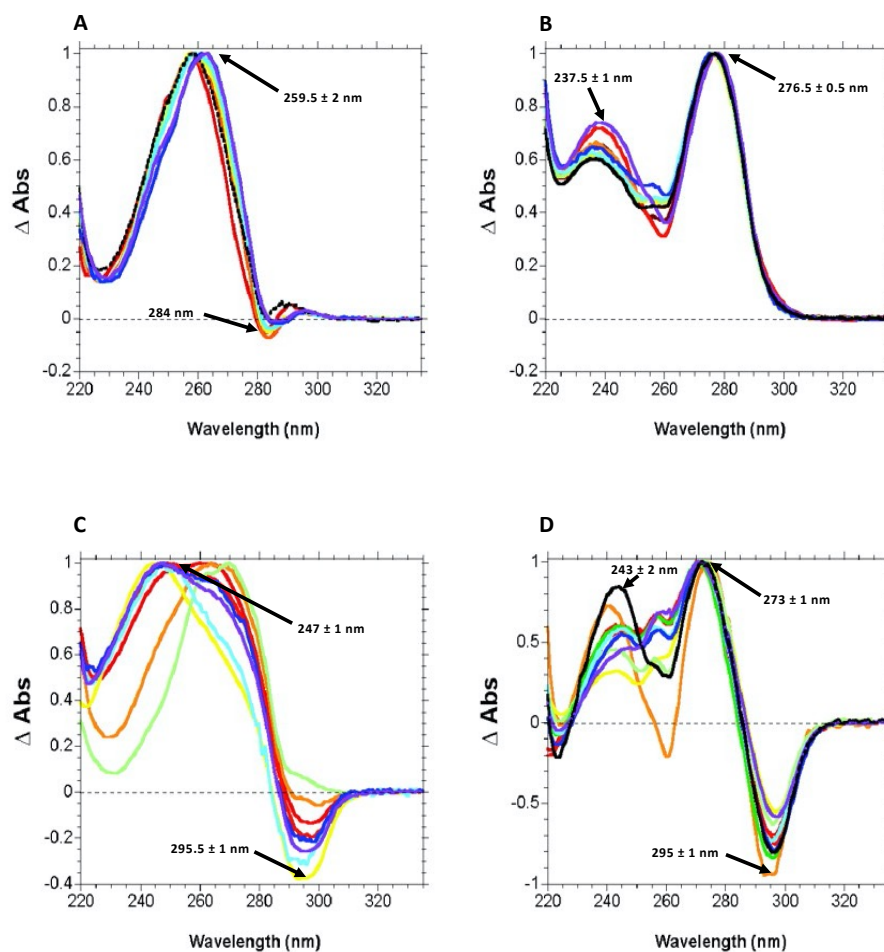


**Figure 1.514 UV characterization of DNA/RNA structures.**

(A) UV spectra showing the absorbances of each of the DNA bases. (B) the overall summed effect of each DNA base, giving an overall maxima at 260nm. (C) UV melting at 295nm of a triplex sequence (R2) and the human telomere G4 sequence which shows that the hypochromic signature at 295nm is specific to the G4 forming sequence (Del Mundo et al., 2017; Mergny and Lacroix, 2009; Otim., 2018).

Another way of characterising nucleic acid structures by UV is to record a thermal difference spectra (TDS), which is the difference between UV spectra (typically recorded between 220-320 nm) recorded at high (denaturing) temperature and low (folded) temperature. The resulting spectrum provides a signature for the type of secondary structure that is being formed from the nucleotide sequence, and has been tested on various known DNA structures (figure 1.515 A-D) (Mergny et al., 2005). However, this only provides some evidence as to the structure being formed and not a direct proof, as many of the structures display similar TDS signatures, for example that of triplexes and G4s both show hypochromicity at 295 nm (Mergny et al., 2005). Therefore, in the absence of an atomic resolution structure, further experimentation is necessary to better characterise the species in solution.





**Figure 1.515 TDS spectra of different DNA structures.**

(A) DNA self-complementary duplexes, 100% AT; (B) DNA self-complementary duplexes 100% GC; (C) Pyrimidine triplexes; (D) DNA G-quadruplexes. The different colours in each spectrum represent different DNA sequences that display that particular structure. Labelled are the major peaks involved in characterising each structure (adapted from Mergny et al., 2005).

### 1.5.2.2 CD spectroscopy

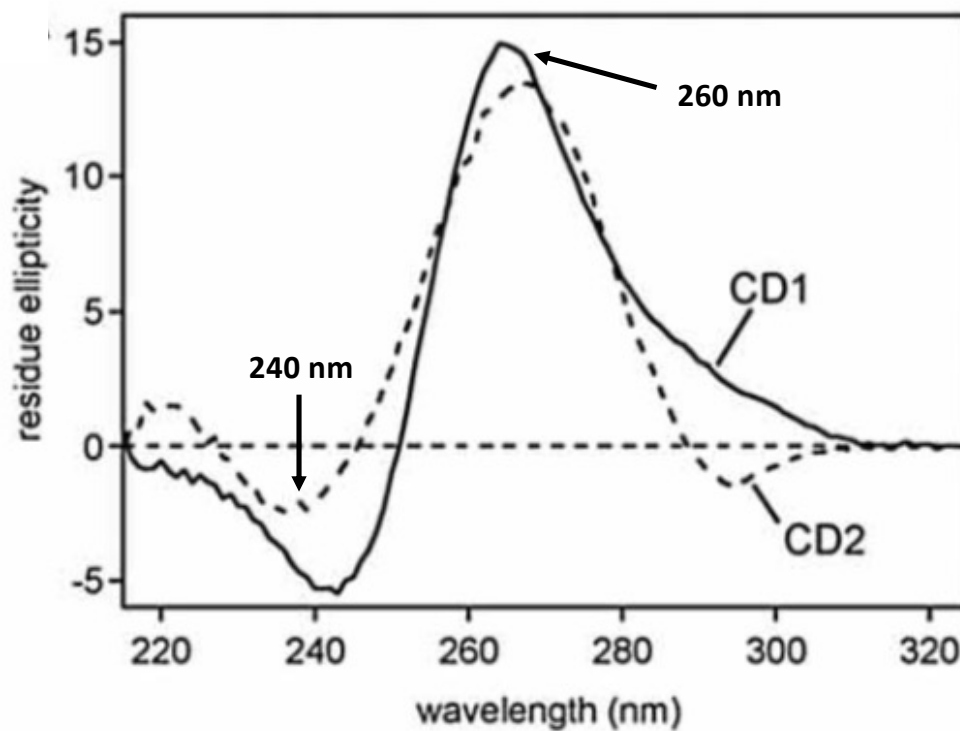
Circular dichroism (CD) is another powerful biophysical technique to characterize the secondary structure in solution of biomolecules such as proteins and nucleic acids (del Villar-Guerra et al., 2018b; Greenfield, 2009). One of the major advantages of CD and UV spectroscopy over other structural techniques is that it has incredibly high throughput, requiring relatively low sample concentrations (1-2  $\mu\text{M}$ ), giving accurate structural information in a time efficient manner (Corrêa and Ramos, 2009). CD relies on the differential absorption of left and right-handed circularly polarised light by optically active samples. Optically active samples are defined as having chirality as well as an absorbance, both of which DNA and RNA nucleotides satisfy, making them amenable for CD characterisation (Kejnovská et al., 2019). Depending on the particular conformation of a nucleotide containing structure, a characteristic CD signature is observed.

Analysis of the CD spectrum enables the user to easily distinguish between the three major types of quadruplexes: Parallel, anti-parallel and hybrid conformations (Kejnovská et al., 2019). Parallel quadruplexes display a positive peak around 260 nm and a negative peak around 240 nm (CD1 in figure 1.516), whereas anti-parallel G4s display a negative peak at 260 nm and a positive peak at 295 nm. Hybrid conformations display mixtures of both the parallel and anti-parallel type G4s, with a reduced positive peak at 295 nm and 260 nm (Vorlíčková et al., 2012). However, similarly to the TDS spectra, CD spectra of different structures can overlap, for example those of a parallel type G4 and an A-type helix (CD1 and CD2 in figure 1.516), making them very difficult to distinguish (Kralovicova et al., 2014; Vorlíčková et al., 2012; Weldon et al., 2016). To overcome this, we can exploit the fact that, unlike duplexes, G4 formation is dependent on the presence of certain monovalent cations such as  $\text{K}^+$  or  $\text{Na}^+$ , and relatively less stable in the presence of smaller cations such as  $\text{Li}^+$  (Bochman et al., 2012). Therefore, observing a change to the CD spectrum, or a greater stability from CD melting assays in the presence of  $\text{K}^+$  compared to  $\text{Li}^+$ , would indicate the presence of a G4 (Weldon et al., 2016). In addition to providing structural information, CD spectroscopy has also been employed as a technique to measure the kinetics of G4 folding, for example with the human telomere G4 after the addition of  $\text{K}^+$  (Gray et al., 2014). Furthermore, CD spectroscopy is also a very useful tool in identifying binding modes of achiral ligands to macromolecules (such as nucleic acids). As described earlier, detection of a CD signal requires the molecule under study

to be chiral. This is true for DNA and RNA molecules (Garbett et al., 2007). It is also true for many DNA or RNA binding ligands, such as ellipticine and porphyrins (see section 1.6), which are naturally achiral, but become chiral when interacting with a macromolecule and therefore can give rise to a distinct peak in the far UV range of the CD spectrum, a phenomenon known as an induced CD (ICD) signal. This phenomenon has been used in many studies to probe the interaction between small molecules and nucleic acids, and is particularly useful for determining dissociation constants when the molecule is not fluorescent (Głuszyńska et al., 2018). In addition, the shape of the induced CD signature can also be correlated with the binding mode of a ligand to a nucleic acid structure. For example, for intercalating compounds, if the molecular axis of the ligand is parallel to the pseudo dyad, a positive ICD signal is produced, whereas if it is perpendicular, a negative ICD is observed (Garbett et al., 2007). Furthermore, minor groove binders which can interact with B-DNA where the molecular axis of the ligand is approximately 45° to the base give rise to positive ICD signals (Garbett et al., 2007). This makes CD a very useful tool to determine a vast array of measurable parameters.

Previous work on each of the six putative G4 forming sequences found within the Bcl-x pre-mRNA display a CD spectrum that is typical for either a parallel G4 or a duplex structure, indicated by the presence of a positive peak at 260 nm and a negative peak at 240 nm (figure 1.517) (Weldon et al., 2017). The CD melting experiments indicate a monophasic melting profile for each of the six PQS tested. Interestingly, the Q2 sequence displayed a significantly higher structural stability than all of the other PQS regions. However, as described above, the formation of a potential G4 structure in each of these sequences will need to be more thoroughly characterised by conducting the experiments in potassium or lithium-containing buffers in order to differentiate the G4 and duplex structures with less ambiguity.

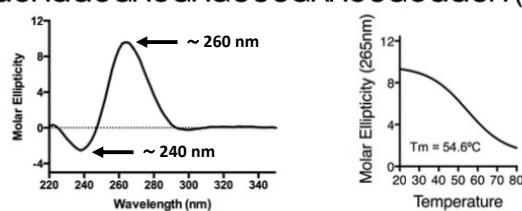
A major disadvantage of CD spectroscopy is that it only provides an overall picture of the type of structure in solution, meaning that in a heterogeneous sample you may have multiple peaks corresponding to the mixture of each of the conformers, some of which may or may not contribute to a major peak, making the data difficult to interpret. Therefore, other techniques that provide structural insight at nucleotide resolution are required, for example NMR spectroscopy.



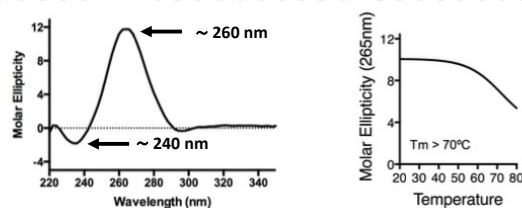
**Figure 1.516 CD spectra of a parallel G4 structure and an A-helix.**

CD spectra of a parallel G4 (CD1) and a duplex (CD2), showing that the two have almost overlapping CD spectra and hence difficult to distinguish (adapted from Kralovicova et al.,2014).

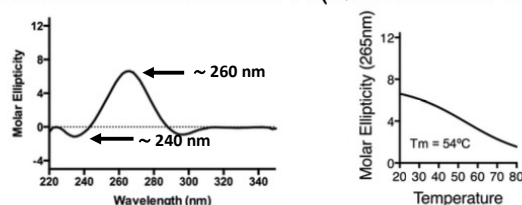
Q1: GAGGGAGGCAGGCGACGAGUUUGAACUGCGGUA (QGRS score: 6)



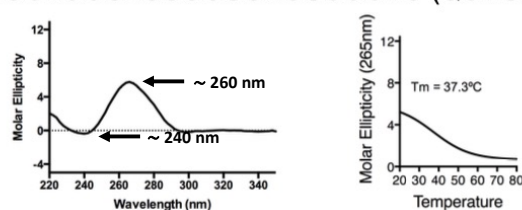
Q2: CCGGGAUGGGGUAAACUGGGUCGCAUUGUGGCC (QGRS score: 17)



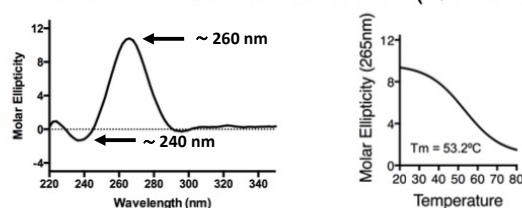
Q3: UCGGCGGGGCACUGUGCGUGGAA (QGRS score: 11)



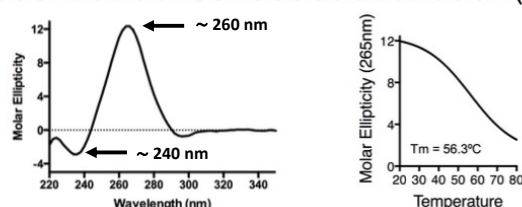
Q4: AAGGAGAUGCAGGUAAUUGGUGAGUCGGAU (QGRS score: 18)



Q5: UUGGAUCCAGGAGAACGGCGGCUGGGUA (QGRS score: 21)



Q6: CUGGUUCCUGACGGGCAUGACUGUGGCCAGGUA (QGRS score: 15)



**Figure 1.517 CD spectra and melts of each PQS sequence of the Bcl-x pre-mRNA.**

Room temperature spectrum and melting experiments of each of the six indicated PQS regions carried out in 100 mM KCl. Labelled are the major peaks at 240 and 260 nm (adapted from Weldon et al., 2017).

### 1.5.2.3 NMR spectroscopy

Alongside CD and UV spectroscopy, NMR has been widely used to study both the structure and dynamics of G-quadruplexes (Ambrus et al., 2005; Črnugelj et al., 2002; Phan et al., 2005; Zhang et al., 2009). This technique relies on the phenomenon of nuclear spin, in which atoms that have a spin of  $I = \frac{1}{2}$  are NMR active. Spins of  $\frac{1}{2}$  have two distinct orientations when an external magnetic field ( $B_0$ ) is applied, with populations aligned with or against  $B_0$  ( $+\frac{1}{2}$  and  $-\frac{1}{2}$  respectively). Those that align against  $B_0$  have a greater energy than those that align with  $B_0$ , and it is this difference in energy that is detected by NMR (Roth., 1984). NMR active nuclei that are commonly used for biological and chemical applications are:  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  (Zou and Sadler, 2015). While  $^1\text{H}$  is a natural isotope,  $^{15}\text{N}$  and  $^{13}\text{C}$  are rare isotopes (the natural ones being  $^{14}\text{N}$  and  $^{12}\text{C}$ ) and NMR studies of these atoms requires isotopic labelling of the molecules. Outlined below are some of the approaches that have been used to characterise the presence of G4s in nucleic acid sequences by NMR spectroscopy.

The first and simplest way to assess if a nucleotide sequence can fold into a quadruplex is by analysing the one-dimensional  $^1\text{H}$  NMR spectrum. In this spectrum, we observe all of the exchangeable and non-exchangeable protons in a structured macromolecule. Figures 1.518 A-D highlights regions within a 1D spectrum that correspond to distinct protons in DNA or RNA. Firstly, the non-exchangeable protons of the ribose sugar moiety display chemical shifts between 4-6ppm, which include the protons  $\text{H1}' - \text{H5}'$  (see ribose numberings in figure 1.518 A). Secondly, the non-exchangeable amino ( $\text{NH}_2$ ) resonances of G, C and A bases which include protons H2, 6 and 8 and resonate between 7-8ppm (see purine and pyrimidine base numberings in figures 1.518 B and C). Finally, the exchangeable imino ( $\text{NH}$ ) protons of G, T and U bases resonate between 10-15ppm if involved in hydrogen bonds (Adrian et al., 2012; Fürtig et al., 2003; Germann, 2012). In particular, this region (10.0-15.0ppm), provides valuable information regarding base pairings. This is because these imino protons would normally be in exchange with deuterium when exposed to the solvent, resulting in the proton signal to disappear. However, these imino protons become protected when involved in hydrogen bonds via Watson-Crick or Hoogsteen base pairs, causing their signal to appear (Fürtig et al., 2003). Therefore, chemical shifts in the 10-15ppm region of the NMR spectrum provide useful information regarding the folding of the molecule and the nature of the base pairs in a given structure. For example, G-C base pairs would display signals around 12.0-

13.5ppm (with some appearing in the 8-9ppm region, see table 1.520), whereas A-U would give rise to chemical shifts between 13-15ppm (table 1.520). Therefore, characteristic Watson-Crick base pairs result in chemical shifts between 12-15ppm (CD2 in figure 1.519 A) (Patel and Tonelli, 1974). It is also possible to observe signals for non-canonical G-U wobble base pairings in the 10-12ppm region (Fürtig et al., 2003). Unlike CD spectroscopy, NMR allows to distinguish with greater ease the spectrum of a quadruplex or a duplex, as Hoogsteen base pairings give rise to imino peaks between 10.0-12.0ppm (figure 1.518 D and CD1 figure 1.519 A) (Feigon et al., 1995; Webba da Silva, 2007; Weldon et al., 2016). In addition, imino protons in the internal G-tetrads do not exchange with the solvent as readily as their Watson-Crick or external-tetrad counterparts, making these peaks appear much sharper (Smith and Feigon, 1992; Wang and Patel, 1993). Within a G4 structure, each guanine tetrad gives rise to four imino peaks (one per guanine), meaning that simply counting the number of peaks in the 10-12ppm region of a well resolved 1D spectrum gives an indication as to the number of tetrads which form the putative G4. For example, 12 sharp imino peaks would suggest 3 stacked tetrads as was seen with an optimised (mutated) c-myc promotor quadruplex sequence (figure 1.521 B, lower panel) (Ambrus et al., 2005).

Analysis of the 1D NMR spectra can also give an idea on the heterogeneity of the sample in solution, indicating whether structural determination by NMR would be feasible. For example, the two-repeat tetrahymena telomeric G4 sequence d(TGGGGTTGGGGT) was shown to exist in two different conformations via analysis of its 1D spectrum (Phan et al., 2004). It is also very common that heterogeneity gives rise to broad overlapping peaks in the 1D spectra, such as the c-myc promotor wild-type sequence (figure 1.521 B, top panel), making structural determination near impossible (Adrian et al., 2012; Ambrus et al., 2005; Le et al., 2012). A commonly used approach to reduce sample heterogeneity, and thereby improving the quality of the 1D spectrum, is by making mutations to the wild-type sequence in order to isolate a single species in solution (figure 1.521 A). For example, making mutations that remove the excess guanines and G-tracts from the c-myc wild-type sequence (which may be involved in G4 formation in minor conformational states) gave rise to a single species and a much better resolved 1D NMR spectrum, enabling structure determination (Ambrus et al., 2005; Le et al., 2012).

In summary, analysis of the  $^1\text{H}$  NMR spectra can provide information on the type of base pairs existing within the structure as well as providing insight into the feasibility of NMR for structural determination.

If the 1D NMR spectrum displays sharp, well resolved peaks, further 2D NMR characterisations can take place. One example that has been implemented widely in G4s is the nuclear overhauser effect correlation spectroscopy (NOESY), which enables identification of spatial distances between nuclei (Jeener et al., 1979). Within a G-tetrad, there are distinct NOESY peaks between the guanine H1 imino proton to the next guanine's H8 (figure 1.519 B), giving distinct NOESY patterns which characterise the G-tetrad (figure 1.519 C) (Phan et al., 2011).

#### **1.5.2.3.1 *In vivo* NMR detection of G4s**

With all the biophysical methods described earlier providing useful tools to characterise the presence or absence of a G4 *in vitro*, a major question in the nucleic acid field is to prove their existence *in vivo*, in which there have been contradictory studies which argue either case for RNA G4s (Biffi et al., 2014; Guo and Bartel, 2016). Therefore, trying to apply some of these biophysical tools to more physiologically relevant conditions would provide insight into the structures being formed within cells, for which NMR spectroscopy has been used.

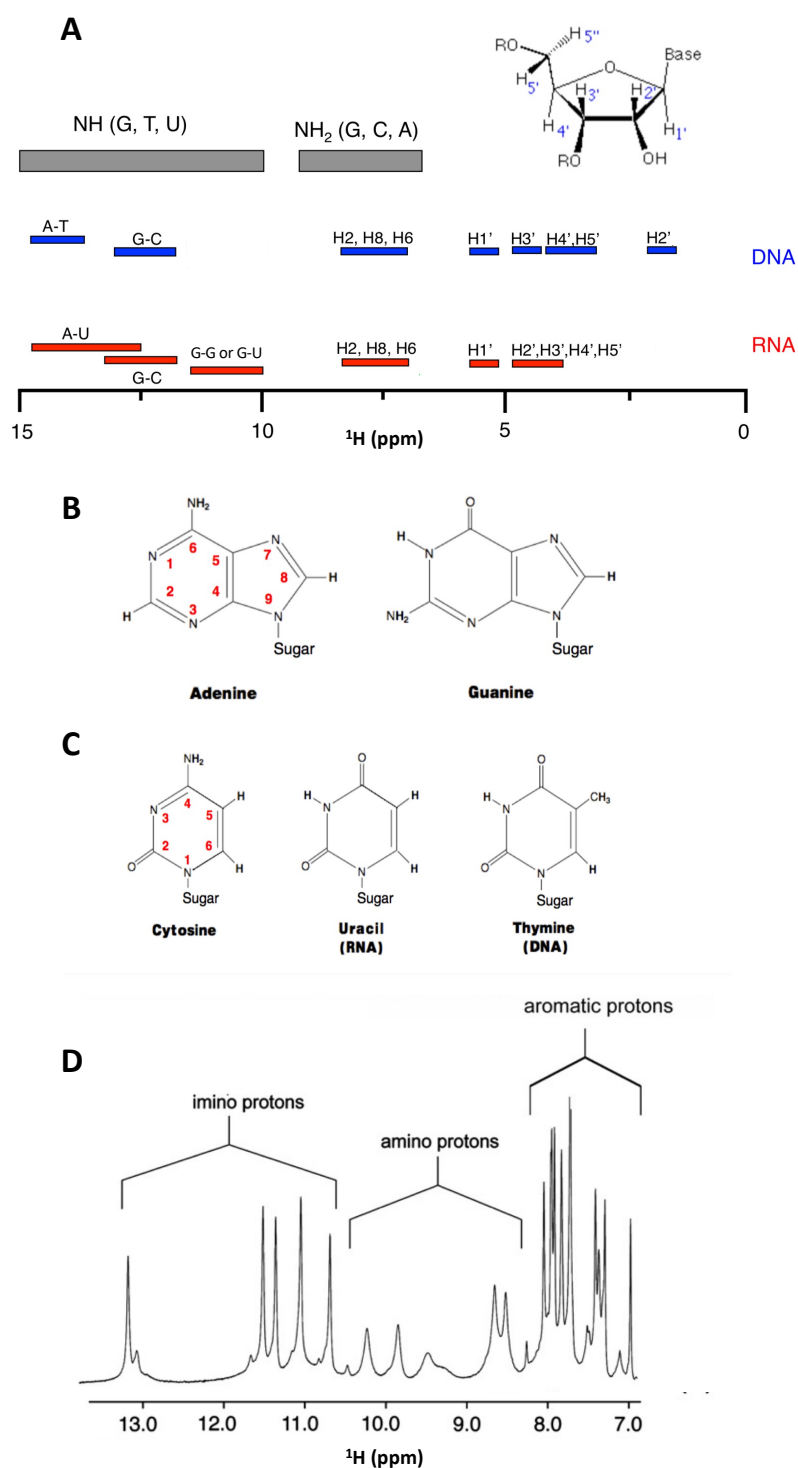
In cell NMR spectroscopy of nucleic acids has relied upon the use of *Xenopus laevis* oocytes to mimic the cellular environment. Due to their large diameters (approx 1 mm), it enables easy microinjection of exogenous nucleic acid sequences directly into the oocyte, enabling NMR analysis of the query sequence in a physiologically relevant context (Giassa et al., 2018). The first report of using in cell NMR to probe G4 structures came from the study of the human telomere G4 DNA sequence  $\text{d}(\text{G}_3(\text{TTAG}_3)_3\text{T})$ , which adopts an anti-parallel basket type structure *in vitro* (figure 1.522 A) with the presence of sharp imino peaks as shown in figure 1.522 B (Lim et al., 2009). Injection into oocytes resulted in a broadening of these imino peaks, which may arise as a result of increased viscosity around the nucleic acid sequence due to the presence of proteins and other endogenous material within the oocyte (figure 1.522 C). Incubating the sequence in the oocyte lysate resolved some of the peak broadening issues, whilst still maintaining the physiological environment originally present in the oocyte (figure 1.522 D). Furthermore, the imino spectrum in the oocyte lysate shows much greater spectral overlap than its buffer only counterpart (figures 1.522 B and D respectively), suggesting the



presence of multiple G4 structures in the oocyte lysate, indicating the structural polymorphism of G4s in the intracellular environment (Hänsel et al., 2009). This suggests that the structures observed *in vitro* may not necessarily reflect the structures *in vivo*, meaning that interpretation of biophysical data in normal buffers should be taken with caution.

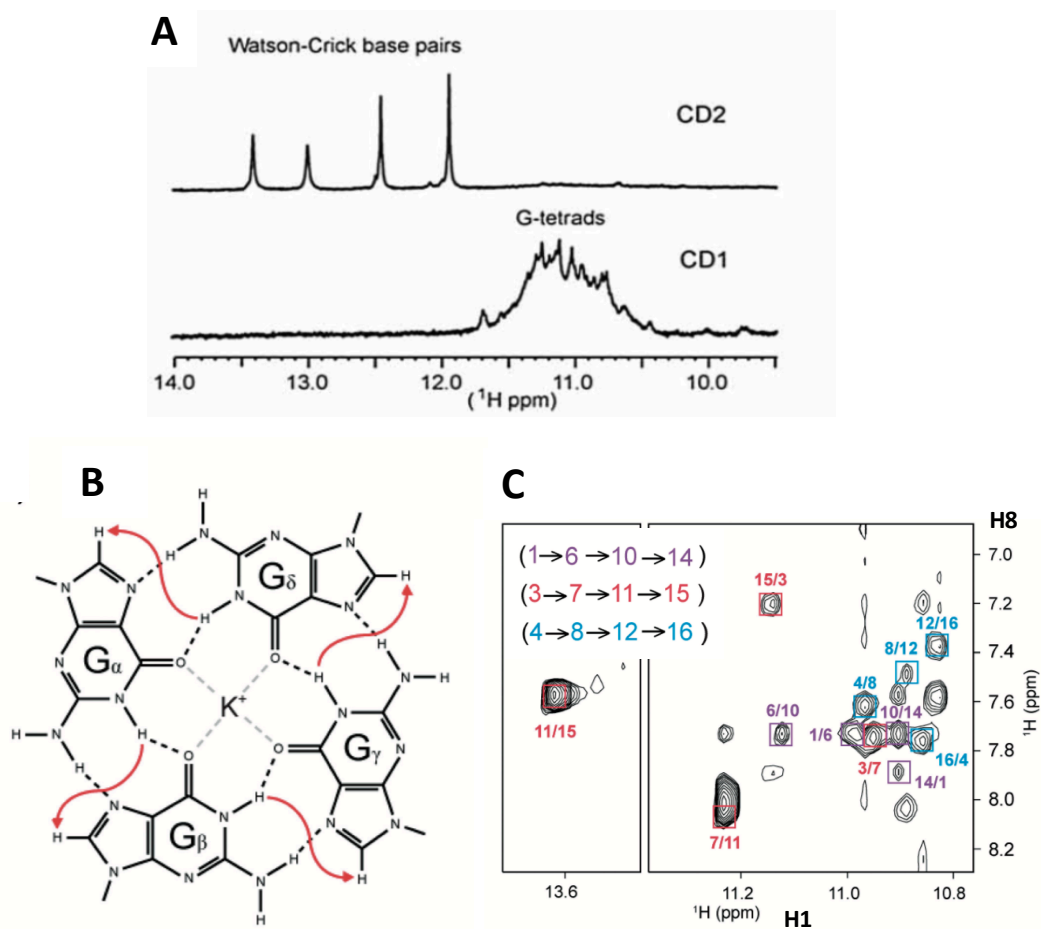
Another advantage of NMR spectroscopy for in cell analysis over other biophysical techniques is that it allows detection of labelled exogenous nucleic acid sequence with NMR active nuclei that would not be present in the cellular extract, thereby eliminating the background signal. For example, the TERRA RNA telomeric sequence: r(UAGGGUUAGGGU) was shown to adopt multimeric G4 structures by *in vitro* mass spectrometry (Collie et al., 2010b), and was revealed to adopt these same structures in *X.laevis* oocytes using <sup>19</sup>F labelled RNA, providing further evidence that RNA G4s can fold in cells (Bao et al., 2017; Biffi et al., 2014)

Only in the last few years, there has been some progress to translate the in-cell NMR from oocytes to mammalian human cells, thereby making the system even more relevant. Instead of direct microinjection, electroporation or chemical induced transfection protocols have been developed (Dzatko et al., 2018; Yamaoki et al., 2018).



**Figure 1.518** The expected chemical shift regions within the 1D spectrum for all of the exchangeable and non-exchangeable protons in nucleic acids.

(A) Chemical shift regions corresponding to each of the protons in nucleic acids. Also shown are the numbering of each proton in the ribose ring of RNA. (B and C) Base numberings within purine (B) and pyrimidine (C) bases. (D) Example of a proton NMR spectrum of the G4 DNA sequence d(GCGGTCGGT)<sub>4</sub> (adapted from Germann., 2012; Webba da Silva, 2007).



**Figure 1.519 NMR characterization of G4s**

(A) 1D NMR spectrum showing the characteristic imino peaks of a Watson-Crick duplex (CD2) and of a G4 (CD1). (B and C) NOE connectivities observed within a G-tetrad between H1 and H8 of adjacent guanines (Adrian et al., 2012; Kralovicova et al., 2014; Phan et al., 2011; Weldon et al., 2016).

Base pair	Base-pair atoms	Chemical-shift regions [ppm]	Distance [Å]
G:C	G N1–H1 ... N3 C	12–13.5	1.89
	G N2–H2 ... O2 C	8–9	2.08
	G O2 ... H4–N4 C	8–9	1.71
	G N2–H2	6.5–7	–
	C N4–H4	6.5–7	–
U:A	U N3–H3 ... N1 A	13–15	1.93
	U O4 ... H6–N6 A	7.5–8.5	1.82
	A N6–H6	6.5–7	–
G:U	G N1–H1 ... O2(O4) U	10–12	1.76
	G O2 ... H3–N3 U	11–12	1.96

**Table 1.520 Expected NMR chemical shifts.**

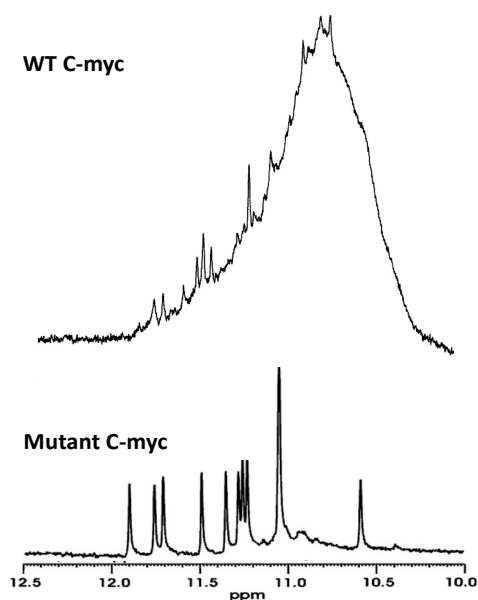
Expected chemical shifts of Watson-Crick and wobble base pairings in nucleic acids (Fürtig et al., 2003).

**A**

**WT C-myc: TGGGGAGGGTGGGGAGGGTGGGGAAGG**

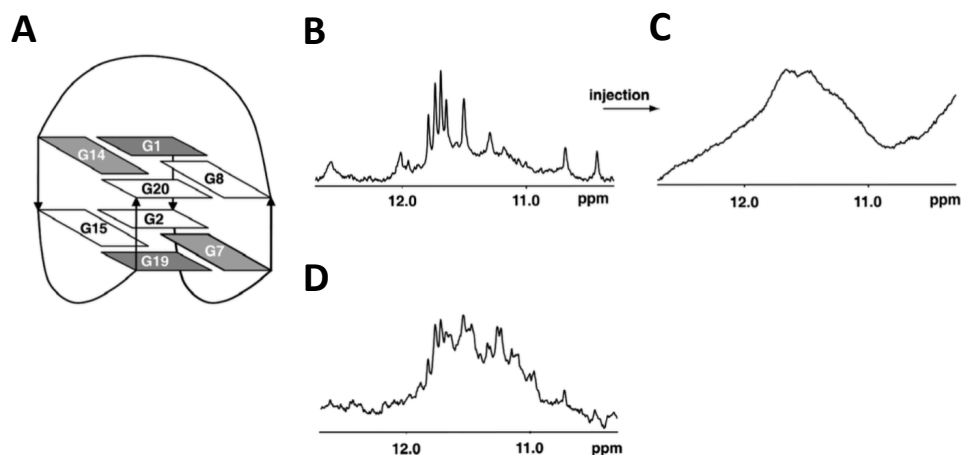
**Mutant C-myc: TGAGGGTGGGTAGGGTGGGTAA**

**B**



**Figure 1.521 Heterogeneity in the c-myc DNA sequence and how this relates to the 1D NMR spectrum.**

(A) Sequences of the c-myc wild-type (WT) sequence and the mutant c-myc sequence designed to reduce sample heterogeneity. (B) 1D NMR spectrum in the imino region of the WT and mutant c-myc sequences (Le et al., 2012).



**Figure 1.522** *In vivo* NMR studies on the human telomere G4 DNA sequence: d(G<sub>3</sub>(TTAG<sub>3</sub>)<sub>3</sub>T).

(A) The anti-parallel basket type structure of the G4. (B) 1D NMR spectrum of the imino region of the sequence in buffer: 10.5 mM NaCl, 110 mM KCl, 130 nM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 10% D<sub>2</sub>O, pH=7.5). (C and D) Imino region of the same sequence recorded in the oocyte (C) or the oocyte lysate (D) (Hänsel et al., 2009).

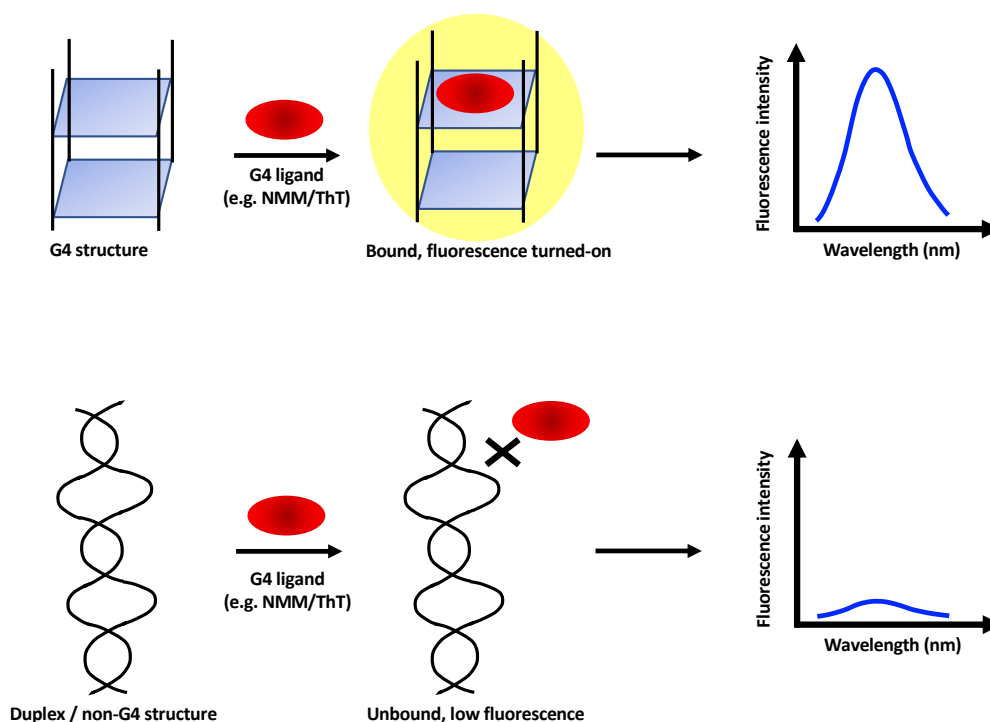
#### 1.5.2.4 Fluorescence detection of G4 structures

Fluorescence spectroscopy is a sensitive and high throughput technique that has been used extensively to study the structure and function of nucleic acid sequences. In brief, the fluorescence phenomenon arises as a result of the photo-selective excitation of certain molecules by specific wavelengths of light, which result in electronic transitions within the molecule. After absorbing a certain wavelength, electrons are excited from a ground state ( $S_0$ ), to a higher first or second electronic singlet state ( $S_1$  and  $S_2$  respectively). Some molecules in  $S_2$  can convert to  $S_1$  by a radiation-less de-excitation of the electrons to a lower energy excited state. Molecules in the excited singlet state then de-excite to  $S_0$ , resulting in the emission of a photon, giving rise to fluorescence (Lakowicz, 2006).

Fluorescence spectroscopy has played an important part in providing evidence for the existence of RNA G4s in living cells (though somewhat controversial) (Biffi et al., 2014). This was achieved by immunofluorescence microscopy, in which they used an engineered G4 specific antibody BG4 (Biffi et al., 2013). Secondary and tertiary antibodies that were conjugated with the fluorophore Alexa 594 were able to bind to the BG4-G4 complex and

resulted in the appearance of cytoplasmic foci, in which they confirmed that these were indeed caused by the presence of endogenous RNA G4s within living cells (Biffi et al., 2014). A more direct approach to detect G4 and other nucleic acid secondary structures is to use specific fluorescent ligands that can bind to and recognise a specific structure. In general terms, DNA or RNA binding molecules have various possible binding modes when recognising nucleic acid structural motifs. These include major and minor groove binders, electrostatic binding to the sugar phosphate backbone and base intercalators (Almaqashi et al., 2016). G4 specific ligands often display additional forms of interactions, such as end-stacking with the external tetrad (see below for examples). One of the best characterised DNA binders that is being used in a vast array of applications associated with nucleic acid biology is ethidium bromide, which was shown to display enhanced fluorescence when intercalating in dsDNA or RNA, and is often used as a staining method to detect duplex structures (Olmsted and Kearns, 1977). However, later studies also showed that ethidium derivatives could also bind to G-tetrads, lacking the structural specificity needed to confirm the presence of such structures (Koeppel et al., 2001). Therefore, using fluorescent ligands that are able to show structural specificity would be very helpful to characterise the presence of nucleic acid structures in solution. For example, the ligand N-methyl mesoporphyrin (NMM) was identified as a highly specific parallel G4 binder, making it useful as a probe to detect the presence of parallel G4s (del Villar-Guerra et al., 2018a; Kreig et al., 2015).

G4 specific ligands can be used as a method to confirm the presence of a putative G4 *in vitro* or *in vivo* by a method called G4 fluorescence turn on assay (summarised in figure 1.523), in which changes in the fluorescence property of that specific ligand upon binding to its target would be a marker to indicate the presence of a G4 in a given sequence (Umar et al., 2019). For example, the aforementioned porphyrin class molecule NMM (figure 1.524 A) was shown to bind with high specificity to parallel G4s over single stranded, anti-parallel, hybrid, or duplex structures, which was shown by a greater fluorescent enhancement ( $\lambda_{exc}=393\text{nm}$ ,  $\lambda_{emm}=620\text{ nm}$ ) when bound to parallel G4s over the other structures (figure 1.524 B)(Kreig et al., 2015; Nicoludis et al., 2012a; Sabharwal et al., 2014). In 2012, a crystal structure of NMM bound to the human telomere 22 mer DNA G4 sequence (Tel22) d[AGGG(TTAGGG)<sub>3</sub>] was published (PDB:4G0F), which showed end staking to the Tel22 G4 dimer (figure 1.524 C) (Nicoludis et al., 2012b).



**Figure 1.523 Principles of the fluorescence turn on assay**

An overview of how fluorescent G4 specific ligands can be used as probes to confirm the existence of G4s in a given nucleic acid oligonucleotide sequence (adapted from Umar et al., 2019).

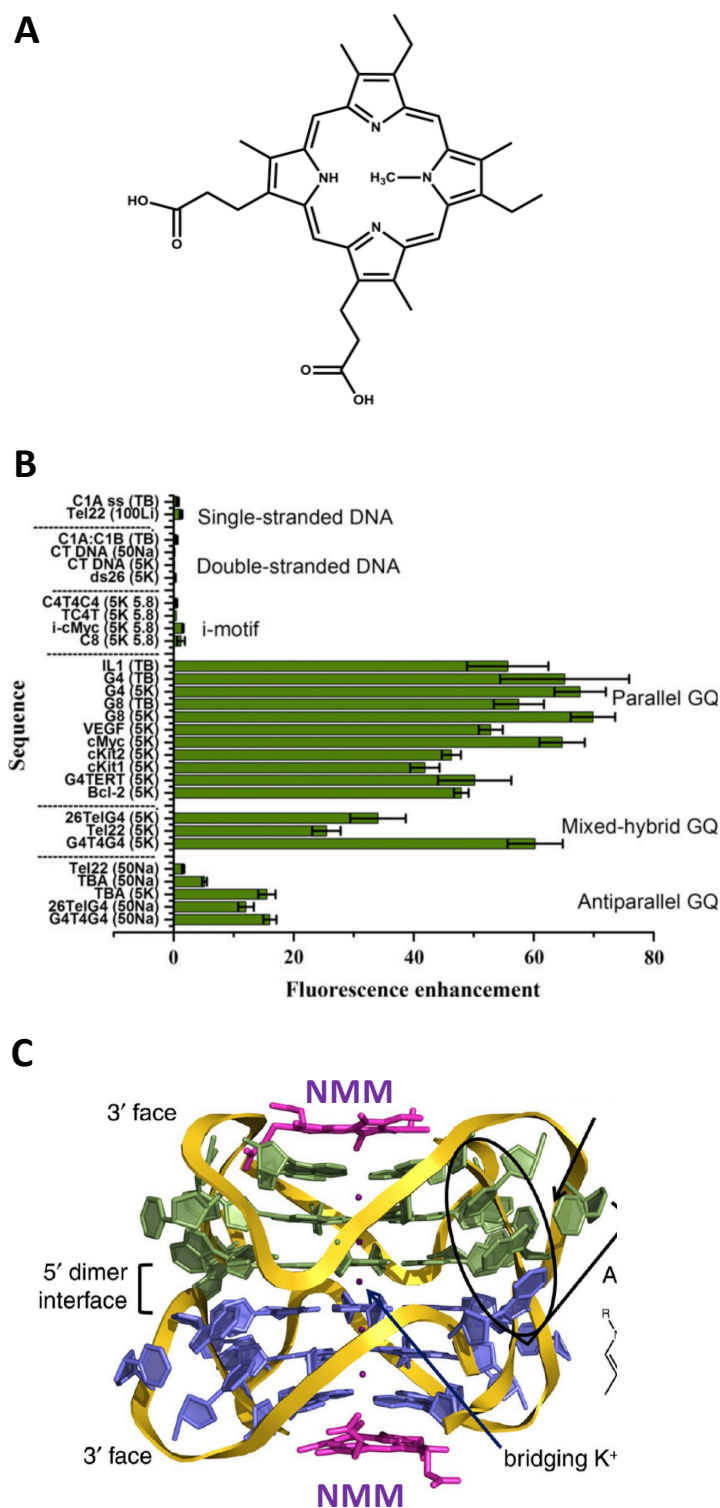
The Benzothiazole ligand, thioflavin T (ThT) (figure 1.525 A), was initially shown to bind protein fibrils and stack with the aromatic amino acids (Biancalana and Koide, 2010), but later studies demonstrated its involvement as a G4 specific ligand, in which ThT showed greater fluorescence intensity ( $\lambda_{\text{exc}} = 450 \text{ nm}$ ,  $\lambda_{\text{em}} = 485 \text{ nm}$ ) and affinity for human telomere G4 forming sequences over duplex, triplex and single stranded DNA (figure 1.525 B) (Gabelica et al., 2013). However, unlike with NMM, there is no structure of ThT in complex with a G4 structure so the exact binding mode and the nature of the G4 conformation that is recognised by ThT is yet to be determined. Still, molecular docking results show a much lower binding free energy for the end-stacking mode over the groove binding mode in RNA G4s, suggesting a possible end-stacker to the G4 tetrad in a similar manner to NMM (figure 1.525 C) (De La Faverie et al., 2014; Xu et al., 2016).

A third class of ligand that has been extensively studied as a G4 turn on sensor is the Triphenylmethane family members, particularly malachite green (MG) and crystal violet (CV).

In 2009, the specificity of both MG and CV was reported by fluorescence, with CV displaying much greater selectivity (and hence greater fluorescence) for G4 DNA sequences than MG, in which the latter also showed fluorescent enhancement to a similar extent for duplex sequences. In addition, CV also displayed greater selectivity for intramolecular rather than intermolecular G4 conformations (Kong et al., 2009b). The same group also identified further selectivity and fluorescence enhancement ( $\lambda_{\text{exc}} = 540 \text{ nm}$ ,  $\lambda_{\text{emm}} = 640 \text{ nm}$ ) of CV for anti-parallel G4 structures over parallel (which contrasts that of NMM), and the proposed hypothesis was that the diagonal and lateral loops of anti-parallel G4s provide protection from quenching of the ligand fluorescence by the solvent, whereas side loops in parallel type G4s may not provide as much protection (Kong et al., 2009a ; Kreig et al., 2015).

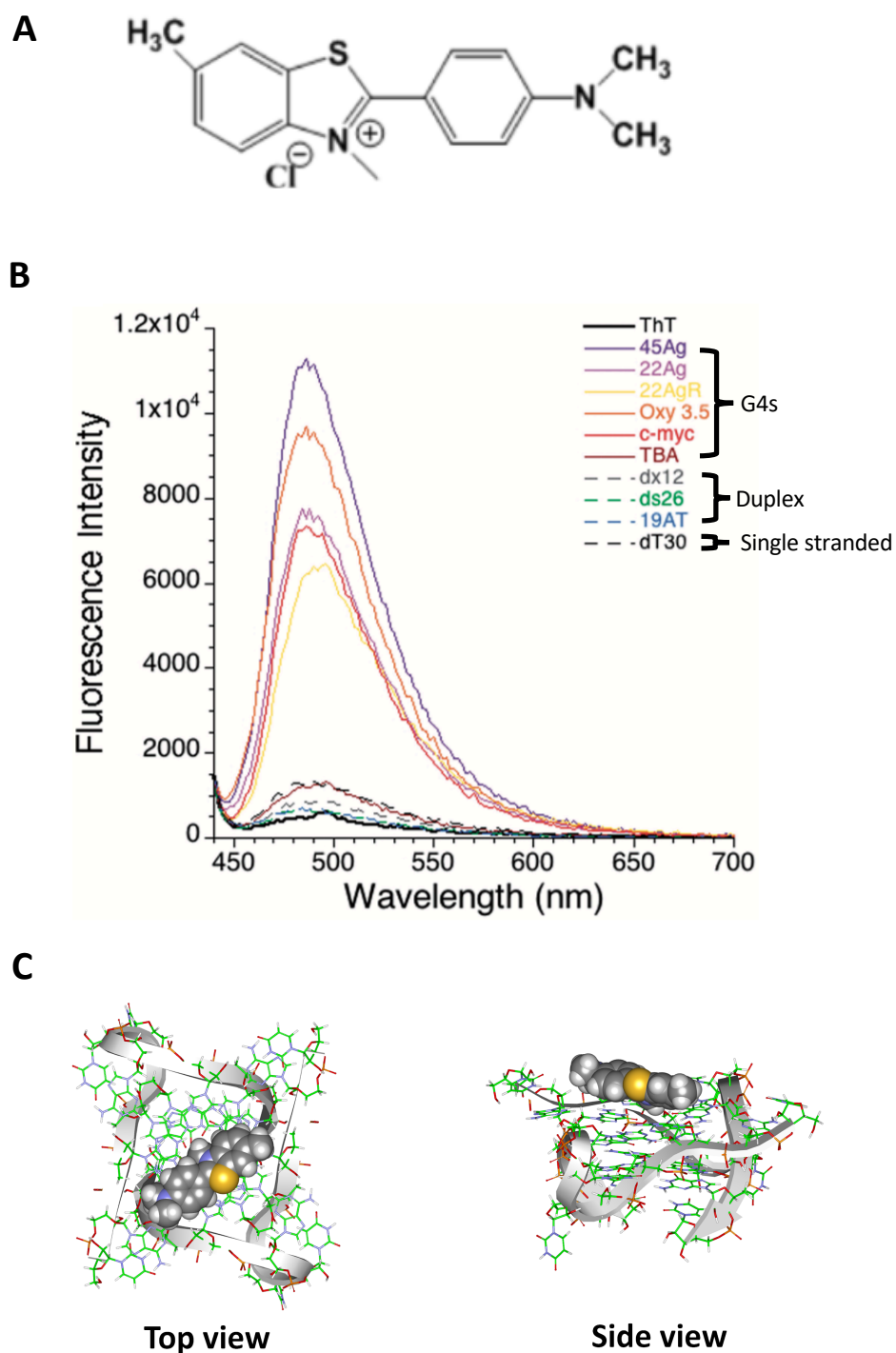
Therefore, fluorescence spectroscopy is a very powerful tool for confirming the presence of G4s as well as providing more detailed structural information such as the strand polarity. However, it is important to consider that these ligands also have the ability to alter the topology of the structure by themselves, meaning that structures that may not exist naturally could be induced in the presence of the ligand, giving false positive results (Weldon et al., 2016). Therefore, data interpretation of fluorescence measurements should always be assessed alongside data obtained with the other biophysical techniques described above.





**Figure 1.524 Structure and specificity of the porphyrin class molecule NMM**

(A) Chemical structure of NMM. (B) Fluorescence enhancement assays of NMM with different nucleic acid secondary structures. (C) Structure of NMM bound to the human telomeric G4 dimer (PDB: 4G0F), showing a 1:1 stacking stoichiometry (Nicoludis et al., 2012).



**Figure 1.525 Structure and specificity of the Benzothiazole ligand ThT.**

(A) Chemical structure of ThT. (B) Fluorescence emission spectra of ThT in the presence of different nucleic acid structures, with the maximum fluorescence occurring for G4 structures. (C) Proposed binding mode of ThT based on molecular docking simulations (De La Faverie et al., 2014; Xu et al., 2016).

### 1.5.3 Biochemical and functional approaches

As well as providing evidence for the formation of G4 structures by biophysical approaches described above, further evidence is necessary to prove their existence in a cellular context. Biophysical methods rely on the use of small RNA or DNA fragments that have the ability to form into putative G4 structures based on bioinformatic analysis. However, a limitation with these techniques is that it does not consider the structural consequences of these sequences when imbedded in larger nucleic acid molecules (in which they are normally present), where there will be competition between G4 and other secondary structures. Therefore, methods that are able to provide evidence for the existence of G4s in long functional RNA transcripts or DNA sequences and in the presence of nuclear or cell extracts, will help provide evidence for their existence in a more cellular context (Weldon et al., 2017).

Primer extension assays have been used extensively as a tool to detect the formation of both DNA and RNA G4s *in vitro* (Cogoi and Xodo, 2006; Kumari et al., 2015; Kwok and Balasubramanian, 2015). In this technique, annealed labelled primers are extended either by a DNA polymerase (to detect DNA G4s) or reverse transcriptase (to detect RNA G4s) enzymes and probing the position at which extension stalls as a consequence of a secondary structure. Indeed, formation of a stable secondary structure would terminate the extension of the cDNA, resulting in a truncated product that can be observed on a denaturing PAGE sequencing gel. By analysing the resulting cDNA fragments in the presence of either K<sup>+</sup> or Li<sup>+</sup> would provide information as to the nature of the secondary structure causing the extension stall, with greater stalling in the presence of K<sup>+</sup> indicating a G4 mediated termination (Kumari et al., 2015). This method has been used to successfully identify the presence of RNA G4s in long RNAs, such as the 5'UTR of NRAS (Kumari et al., 2008; Kwok and Balasubramanian, 2015). Similar studies have also shown the formation of DNA G4s when short stretches of G4 sequences are imbedded within plasmids (Kumari et al., 2015). However, complications in data interpretation do arise when trying to attempt these experiments in nuclear or cellular extracts, as there is also possibility of stalling as a result of protein-RNA interactions rather than a presence of a stable secondary structure (Smith et al., 2014).

Once a particular G4 has been identified by bioinformatic and biophysical analysis, the functional effect of the putative G4 can then be tested by observing the effects of G4 abolishing mutations in functional assays. Many studies have looked at the effects of DNA and RNA G4s on gene expression, both within the promoters of DNA sequences and within mRNA sequences and 5'UTRs (Beaudoin and Perreault, 2010; Kong et al., 2018; Murat et al., 2014). To determine such changes, reporter genes are placed downstream of the sequence of interest, for example the luciferase gene from the firefly *Photinus pyralis*, which when expressed after translation, gives a quantifiable fluorescence signal proportional to gene expression (Smale, 2010). This technique has been widely implemented to assess the effects of G4s and G4 abolishing mutations in a wide variety of biological functions. For example, presence of G4 forming sequences in the 5'UTRs resulted in translational repression which correlated with a lower luciferase fluorescence compared to G4 abolishing mutations which had the opposite effect (Beaudoin and Perreault, 2010). Another study used a similar technique with the GFP gene instead of luciferase by determining that G4s within the mRNA sequence prevents ribosomal elongation (Murat et al., 2014). However, to confirm that G4s are the functional consequences of such biological phenomenon, G4 abolishing mutations were necessary. This has the disadvantage of yielding false negative results, as changes to the primary sequence also has the ability to alter the binding site of proteins involved in the biological reactions. For example, the binding site of the transcription factor Sp1 overlaps that of a G4 in the promoter of the pseudorabies virus, and it was demonstrated that G4 abolishing mutations did not only abolish the formation of the G4 itself, but also the binding of Sp1 (Kong et al., 2018). Therefore, interpretation of such data should be taken with care when using G4 mutations in functional assays.

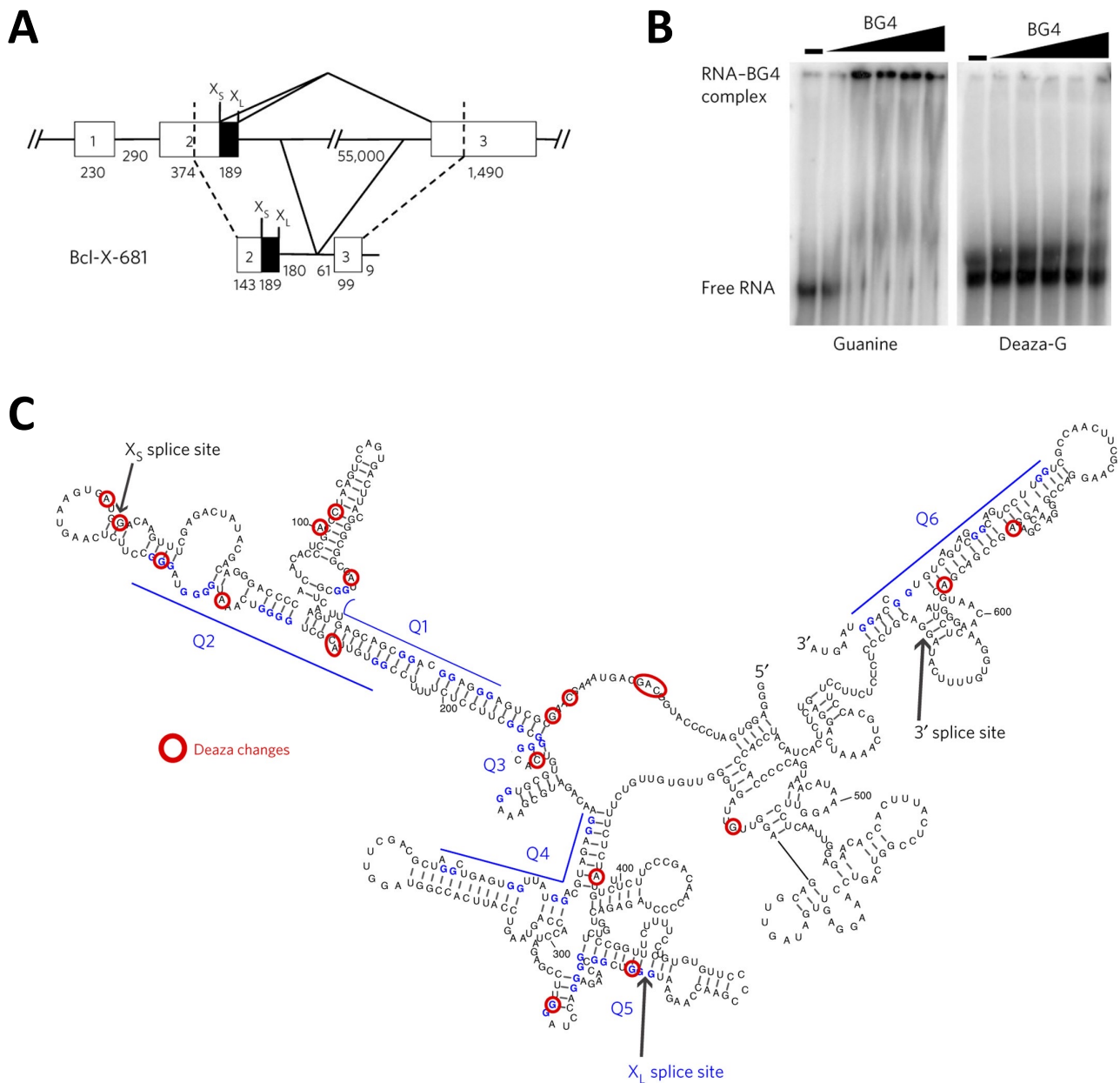
GFP and luciferase has also been helpful to study the effects of G4s in alternative splicing assays. For example, the gene TP53 (a major tumour suppressor) was shown to form different splice variants. The wild-type fully spliced variant (FS-p53) lacks intron 2, whereas the alternatively spliced non-functional splice variant (AS-p53) retains intron 2 (Courtois et al., 2002). Using biophysical and primer extension assays, regions within intron 3 were shown to fold into G4 structures, and it was shown that G4 abolishing mutations reduced the levels of FS-p53 and increased levels of AS-p53, suggesting that a G4 in intron 3 regulates splicing of intron 2 and is necessary for the tumour suppressive phenotype of p53 (Marcel et al., 2011). As intron 2 contains a stop codon, GFP was easily used as a reporter gene downstream of the

pre-mRNA, as only the FS-p53 would express GFP due to removal of the intron 2 containing the stop codon (Marcel et al., 2011). Results pointing to the effects of a G4 mediated alternative splicing event were made clear by comparing wild type and G4 abolishing mutant versions by RT-PCR and flow cytometry (Marcel et al., 2011). Similar experiments using the luciferase reporter also showed the effects of a G4 in mediating the splicing of the human transcription factor Pax9 (Ribeiro et al., 2015). However, as described previously, the limitation with these functional assays is that the RNA may also be part of a *cis*-acting regulatory sequence able to bind to *trans*-acting factors, which could influence the biological mechanisms described above (Weldon et al., 2016). Therefore, in addition to mutagenesis, an alternative strategy is to use G4 specific stabilising ligands, (for which > 800 small molecules have been identified), and investigate whether the effects of the compound are enhanced in the wild-type but not the mutant sequences (Li et al., 2013; Marcel et al., 2011; Ribeiro et al., 2015). Though it is important to note that the effect of introducing stabilising ligands also has its drawbacks, with the potential to artificially stabilise a G4 which would otherwise be non-existent, giving a false positive result (Weldon et al., 2016). Consequently, results obtained using both approaches should be taken with caution and only imply on the possibility of a G4 mediated function.

As discussed previously, the Bcl-x pre-mRNA has six PQS regions identified by bioinformatic and CD analysis (Weldon et al., 2017). However, a method that can prove their existence in more physiological conditions where each of these PQS regions would be imbedded in longer sequences, will be beneficial in understanding their functional relevance. One method that has been used to prove that a G4 can exist in the Bcl-x pre-mRNA is by using electromobility shift assays (EMSA) of wild-type and 7-deaza-substituted RNA in the presence of the previously described G4 specific antibody BG4 (Biffi et al., 2013). The 7-deaza-substituted RNA contains a modified guanine base, in which the N<sub>7</sub> position of the guanine is changed to a carbon, abolishing the G4 through disruption of the Hoogsteen base pairs whilst still retaining its Watson-Crick potential (Murchie and Lilley, 1992). Figure 1.526 shows the EMSA assays conducted on the Bcl-x -681 pre-mRNA, which is a mini gene model of the full length Bcl-x pre-mRNA, containing both alternative splice sites but a shortened intron, exon 2 and 3 with the removal of exon 1 and displays a similar splicing profile to the full length RNA (figure 1.526 A). Titrations of the BG4 antibody showed a significant shift in the wild-type but not the 7-

deaza substituted RNA, suggesting that a G4 can exist in this pre-mRNA (figure 1.526 B) (Weldon et al., 2017).

The next stage is identifying where these G4s lie within the Bcl-x sequence. To achieve this, a better understanding of the Bcl-x-681 structure in solution was required, resulting in the use of RNases T1, T2 and V1 in RNA foot-printing assays. These enzymes have the ability to cleave RNA sequences at specific positions: RNase T1 cleaves 3' to single stranded guanines; T2 cleaves 5' of single stranded bases; V1 cleaves 5' to double stranded bases (Ehresmann et al., 1987). Accessibility of these RNases will therefore be dictated by the secondary structure adopted by the given RNA sequence, giving restraints at nucleotide resolution which enabled structure determination (figure 1.526 C). The structure shown in figure 1.526 C was calculated using the multiple-fold (M-fold) software with the restraints collected from the foot-printing data, which shows that the Bcl-x pre-mRNA is highly structured with several stem loops. M-fold uses an algorithm that is able to predict the minimum free energy conformation of a particular nucleic acid sequence, with the option to input restraints guided by experimental data (such as by foot-printing or NMR experiments). However, the limitation with this programme is that it is only able to predict the formation of duplex structures (Watson-Crick base pairings), and not that of quadruplexes (Hoogsteen base pairings). Therefore, the actual structure in solution may be more complex and heterogeneous, with the potential of G4s and other structures to be in a dynamic equilibrium. To identify the location of the G4 structures within the Bcl-x pre-mRNA, a method known as FOLDeR (foot-printing of long 7-deazaguanine-substituted RNAs) was implemented (Weldon et al., 2017). This method involves comparing the cleavage pattern of RNases V1 and T2 in the wild-type and 7-deazaguanine substituted RNA, with the sections of the Bcl-x structure that display the most significant changes mapped, implying the presence of a G4 in these regions. The regions of the Bcl-x-681 pre-mRNA that displayed the most changes coincided with the G-tracts of the Q2 and Q5 regions, which lie downstream and upstream of the X<sub>S</sub> and X<sub>L</sub> 5'ss respectively (red circles in figure 1.526 C), suggesting G4 forming potential in these areas of the mRNA.

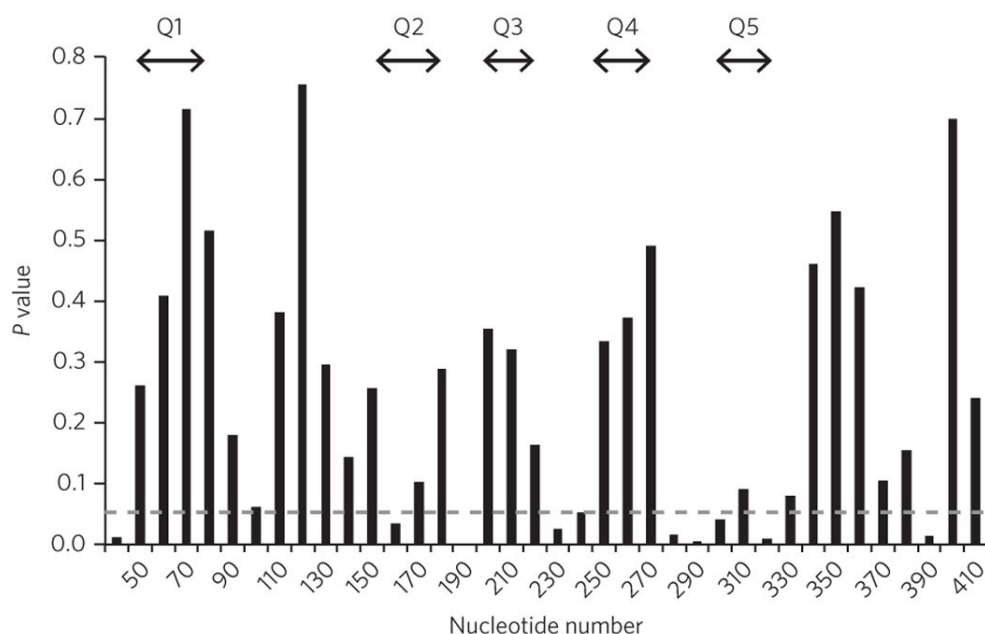


**Figure 1.526 EMSA and foot-printing assays on the Bcl-x 681 construct**

(A) Schematic showing the formation of the mini-gene model (Bcl-x 681 pre mRNA) from the full length Bcl-x pre-mRNA. (B) EMSA assays of the WT and 7-deaza Bcl-x 681 pre-mRNA transcripts with a titration of the BG4 antibody. (C) Predicted structure of the Bcl-x 681 mini gene model based on constraints from foot-printing data. Red circles indicate significant changes in RNase accessibility after 7-deaza guanine substitution (adapted from Weldon *et al.*, 2017).

To further confirm the observations made in the foot-printing assay, another strategy to probe G4s is by RNase H assays, which relies on the accessibility of short DNA oligonucleotides to hybridise to the RNA sequence. This DNA:RNA hybrid is recognised and cleaved by the enzyme RNase H. Regions that cannot bind the oligo (perhaps due to the formation of a stable secondary structure), will be protected from RNase H cleavage. Therefore, comparing the regions that show the most significant differences in the cleavage pattern between wild-type and 7-deaza substituted RNA is also an indication of G4 forming potential in these areas. The histogram represented in figure 1.527 shows the changes that occurred within the Bcl-x pre-mRNA when comparing the wild-type and 7-deaza substituted RNA, with *P* values below 0.05 indicating the most significant changes (Weldon et al., 2017). The regions which displayed these most significant differences corresponded to the Q2 and Q5 region, consistent with the foot-printing data. Together, these data show the possibility of G4 formation in these two respective regions, which will need to be validated further with more extensive biophysical analysis which will be explored in later chapters of this thesis. Knowing that a G4 has the potential to form in close proximity to both alternative 5'ss, the next stage was to examine the functional relevance of such structures in altering the splicing pattern of Bcl-x with the use of known G4 ligands that have the ability to bind and stabilise such structures (Weldon et al., 2017).





**Figure 1.527 Significant changes occurring to the Bcl-x 681 RNA in the oligo accessibility RNase H assays.**

All the changes occurring to the Bcl-x 681 RNA between 7-deazaguanine and WT RNA. *P* values below 0.05 (dashed line) indicate significant changes between the two transcripts (Weldon et al., 2017).

## 1.6 G4 binding ligands

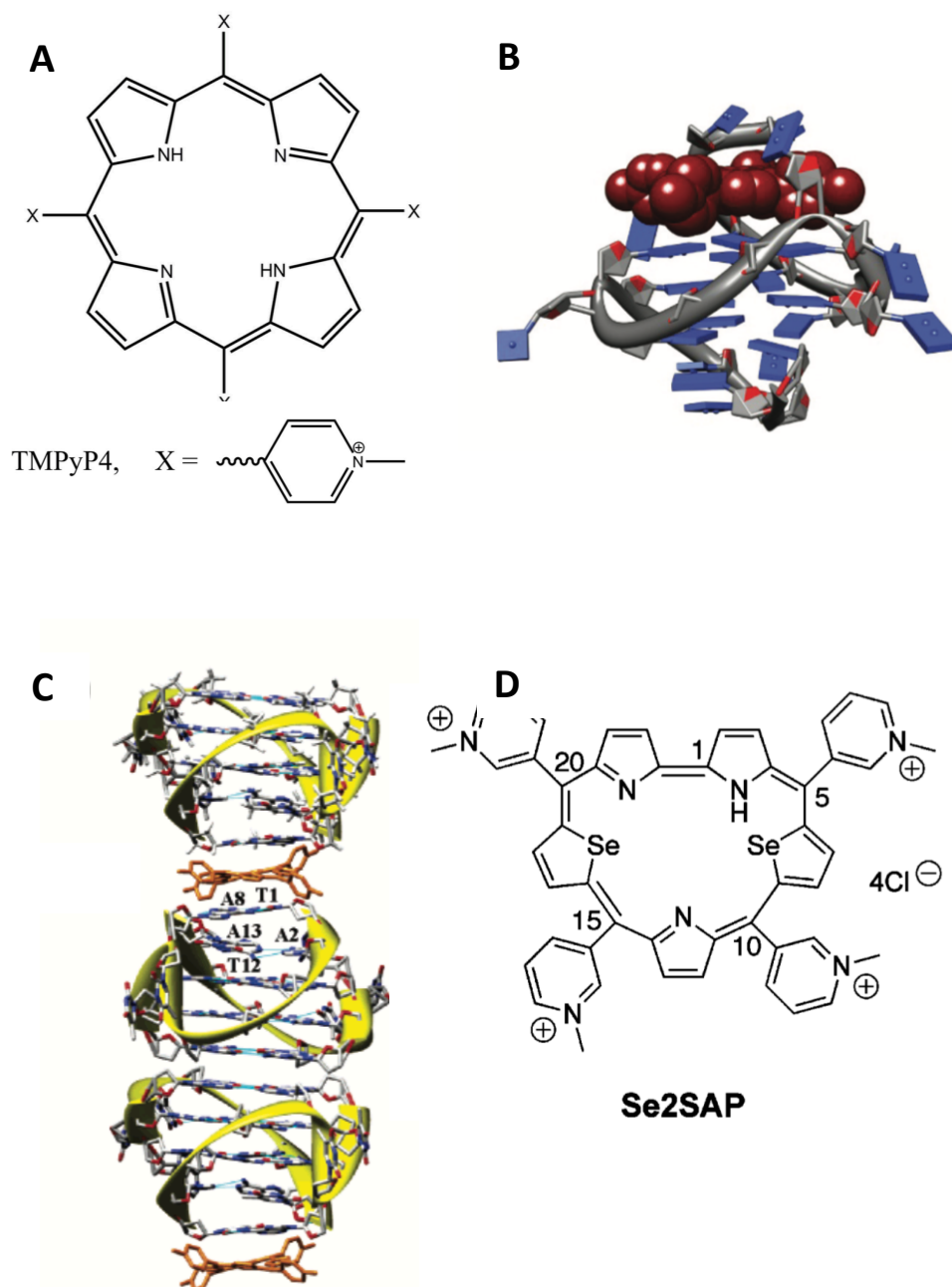
Interfering with some of the G4 dependant biological processes described above, has led to extensive research into finding putative G4 binding ligands for cancer therapy. The very first of these G4 ligands was discovered in 1997, in which molecule 2,6-diamidoanthraquinone was shown to inhibit telomerase action by stabilising the human telomere G4 (Sun et al., 1997). Subsequent research into other G4 binding ligands has resulted in the discovery of over 800 small molecules that are able to bind to G4 structures, some of which show selectivity to particular topologies of quadruplexes and vary in their binding modes, all of which have been deposited in an online database (Li et al., 2013). G4 ligands have been shown to dramatically alter the dynamics and structure of nucleic acids, not just by favouring the equilibrium towards G4s over other structures, but also to specific topologies of G4s over other G4 species (Lacroix et al., 2011; Tippana et al., 2014). There are several possible binding modes of G4 ligands, such as tetrad end stacking and intercalation, as well as binding to grooves and loops contained within the G4 structure (Sun et al., 2019). G4 binding ligands very often have certain characteristic features which aids their binding to the quadruplex structure, for example: (a) delocalised  $\pi$  bonds to enable G-quartet stacking; (b) a partial positive charge

which aids in G4 stabilisation by simulating the effects of monovalent cations like K<sup>+</sup> or Na<sup>+</sup>; (c) positively charged functional groups that can form additional interactions with the negatively charged phosphate backbone to increase affinity and hence potency (Arola and Vilar, 2008). The pyridostatin (PDS) class have been used extensively as G4 ligands, and was originally identified as a selective inhibitor of translation by stabilising a G4 in the UTR of the NRAS mRNA (Bugaut et al., 2010). PDS has also been used to trap both DNA and RNA G4s in cells (Biffi et al., 2014). A derivative of PDS, carboxyPDS, was identified to have greater specificity to RNA G4s over their DNA counterparts, and has been used to confirm the existence of RNA G4s in cells (Di Antonio et al., 2012; Biffi et al., 2014). Bisquinolinium compounds, particularly PhenDC3 and 360A, have been identified as highly specific G4 ligands and have been used widely (Halder et al., 2012). Other examples of G4 binding ligands include the porphyrin, quindoline and ellipticine class of molecules along with their different derivatives, which will be discussed below (Li et al., 2013).

### 1.6.1 Porphyrins

Porphyrins are a ubiquitous class molecules found naturally in plants (chlorophyll) and animals (haemoglobin). Their general structure consists of four pyrrole groups connected by methene linkers, forming a tetrapyrrole structure which very much resembles that of a G4 tetrad (figure 1.61 A) (Lesage et al., 1993). This similarity to the G4 tetrad has attracted much attention as putative G4 binding agents. One of the extensively studied porphyrin molecules is TMPYP4, in which four additional positively charged moieties are attached to the ends of the tetrapyrrole core (figure 1.61 A). Early studies using energy minimisation suggested that TMPYP4 binds via tetrad intercalation, and this was supported by stoichiometric analysis in which TMPYP4 was bound to a G4 structure with a stoichiometry of n-1, where n is the number of tetrads in the G4 structure (Haq et al., 1999). This finding was consistent with later studies on the wild type c-myc and Bcl2 promotor quadruplex, where two low affinity binding events occurred via intercalation ( $K_a = 4.2 \times 10^5 \text{ M}^{-1}$ ) as well as two high affinity binding sites through end stacking ( $K_a = 1.6 \times 10^7 \text{ M}^{-1}$ ) (Freyer et al., 2007; Nagesh et al., 2010). An earlier study showed a solution structure of TMPYP4 complexed with a modified c-myc quadruplex displaying only the end stacking mode (figure 1.61 B) (Phan et al., 2005). Despite all of the evidence pointing to the TMPYP4-G4 interaction, there is now growing evidence that its

interaction is non G4 specific. The first of these evidence came from a crystal structure of the bimolecular telomeric G4 in complex with TMPYP4, which showed that the ligand does not form any contact with the G-tetrads at all, but is rather sandwiched between two T-A base pairs at the 5' interface of the structure (figure 1.61 C) (Parkinson et al., 2007). The ability of TMPYP4 to stabilise G4 structures is also questionable, as numerous studies showed that addition of the ligand results in the unfolding of G4s, for example of the usually stable RNA G4 present in the 5'UTR of the matrix metalloproteinase mRNA (Morris et al., 2012). Despite all these controversies, TMPYP4 has displayed a wide variety of anti-tumour properties such as preventing transcription of oncogenes like c-myc and K-Ras as well as preventing telomere elongation (Cogoi and Xodo, 2006; Mikami-Terao et al., 2009; Siddiqui-Jain et al., 2002). Efforts into designing more specific G4 binding porphyrins are also being investigated. A previously discussed example is NMM, which was shown to bind with high specificity to parallel type quadruplexes ( $K_a = 1.7 \times 10^6 \text{ M}^{-1}$ ) (explained in section 1.5.6) (Nicoludis et al., 2012a, 2012b; Sabharwal et al., 2014). Other strategies are also being investigated, for example modifying the core porphyrin analogue with selenium to generate a selenium substituted expanded porphyrin, Se2SAP (figure 1.61 D), which showed greater selectivity to the c-myc quadruplex than TMPYP4 (Seenisamy et al., 2005).

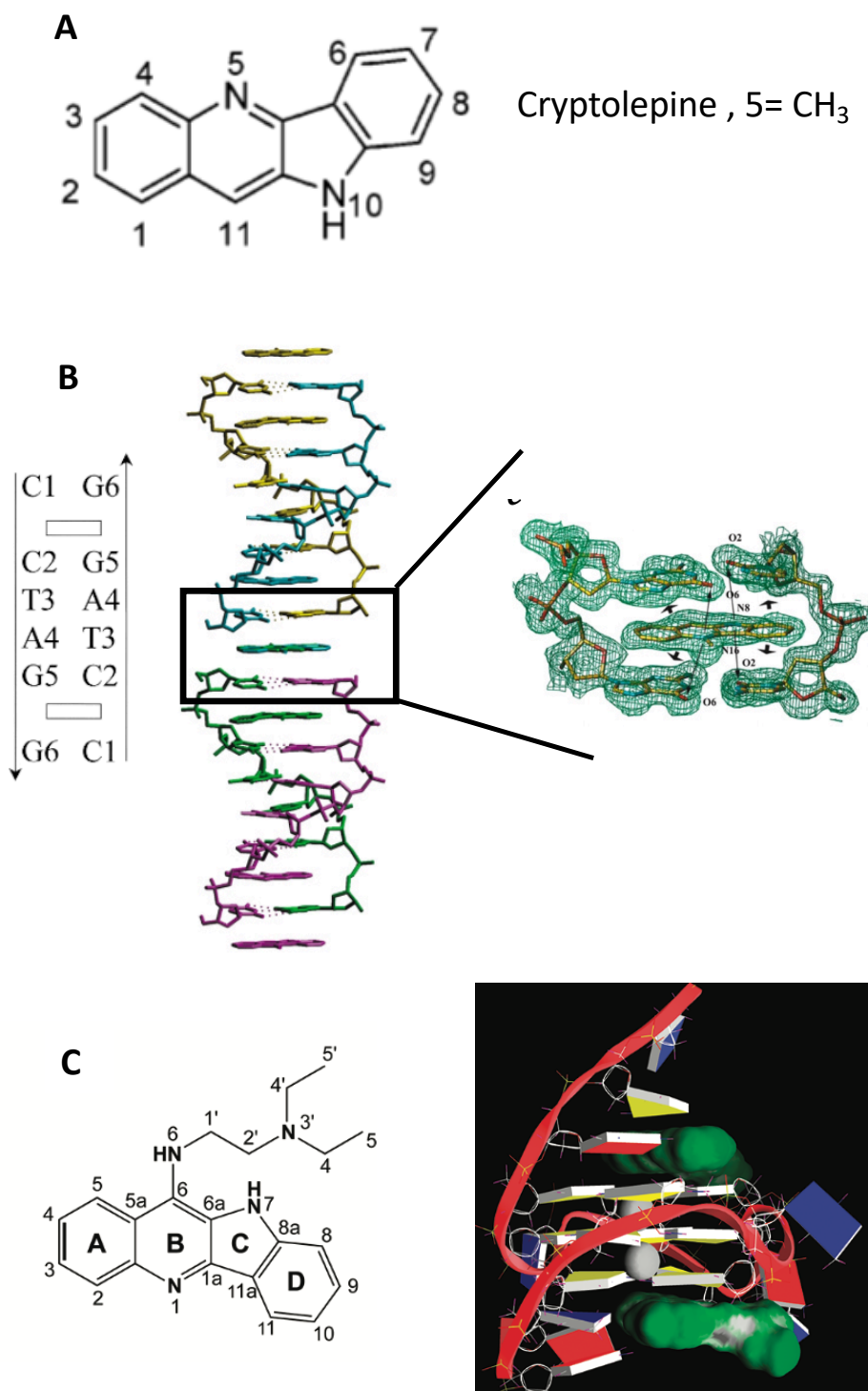


**Figure 1.61 Structure and complexes of the porphyrin class of molecules.**

(A) General structure of the porphyrin ring. (B) NMR structure of the c-myc quadruplex in complex with TMPYP4 (PDB: 2A5R). (C) Crystal structure the human telomere quadruplex, showing a series of bi-molecular quadruplex stacks with TMPYP4 (in orange) sandwiched between two T-A base pairs (PDB: 2HRI). (D) The selenium substituted derivative of the porphyrin ring (Le et al., 2012; 2013; Parkinson et al., 2007; Phan et al., 2005; Seenisamy et al., 2005).

### 1.6.2 Quindolines

Cryptolepine (figure 1.62 A) is a naturally occurring quindoline alkaloid that has been extensively used as antimalarial drugs in Central and Western Africa (Dwuma-Badu et al., 1983). In 2002, a crystal structure of this compound bound to duplex DNA was published, which showed an intercalating binding mode in duplex DNA between two consecutive G-C base pairs in the sequence d(CCTAGG)<sub>2</sub> (figure 1.62 B) (Lisgarten et al., 2002). Later studies showed that derivatives of cryptolepine were able to stabilise DNA G4s, and the level of stability was correlated with an increase in inhibition of telomerase activity (Guyen et al., 2004). Further functionalisation of the quindoline moiety showed even greater selectivity for G4 DNA over duplex structures and enhanced the potency of telomerase inhibition, with some of the derivatives displaying over 100 fold more potency than the quindoline itself, with the ability to induce a G4 in G-rich DNA (Zhou et al., 2005). In addition to binding the human telomere G4, quindoline derivatives were also shown to bind to the c-myc promotor G4, and a solution structure was solved showing a 2:1 stoichiometry (Dai et al., 2011). The binding did not change the structure of the G4 but led to significant rearrangement of the flanking bases in the myc sequence (figure 1.62 C) (Dai et al., 2011). It was later shown that other quindoline derivatives (such as SYUIQ-5,  $K_d = 7.8 \mu\text{M}$ ) prevented the transcription of c-myc by stabilising the G4 at its promotor and preventing the binding of transcription factors (Liu et al., 2017).



**Figure 1.62 Structure and complexes of the quindoline class of molecules.**

(A) The chemical structure of the quindoline core, showing the methyl group at position 5 of the natural alkaloid Cryptolepine. (B) Crystal structure of the complex between the duplex DNA d(CCTAGG)<sub>2</sub> and Cryptolepine. (C) The binding mode of a 2:1 complex between the quindoline derivative depicted on the left and the C-myc Pu-22 sequence (PDB: 2L7V) (Dai et al., 2011; Lisgarten et al., 2002; Zhou et al., 2005).

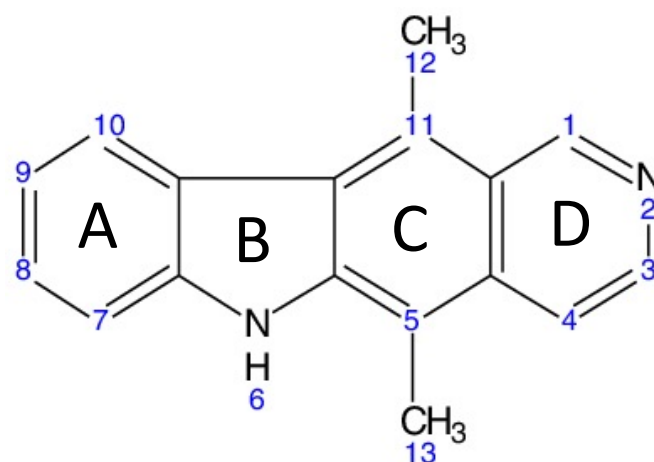
### 1.6.3 Ellipticines

Before their discovery as G4 ligands, ellipticines were originally characterised to be duplex binders. Similar to the quindoline family, the ellipticine class of compounds are also natural alkaloids and are present in the leaves of the tropical evergreen tree *Oschrosia elliptica labill* (Goodwin et al., 1959). They display a structure remarkably similar to the quindolines, in that they contain carbazole rings connected to a pyridine ring, resembling that of a purine-pyrimidine base pair and suggesting a possible role as a DNA intercalating agent (figure 1.63 A) (Auclair, 1987; Stiborová et al., 2006). In the mid 70s, the first binding assays of ellipticine was conducted on calf-thymus DNA, where they showed by spectroscopic analysis that ellipticine can indeed intercalate between base pairs in a way such that the axis of the ellipticine is parallel to the base pair (Kohn et al., 1975). Binding by intercalation of ellipticine was shown to be mainly mediated by the hydrophobic interactions between the nucleotide bases and the aromatic ring structures of ellipticine, though it was shown that in acidic conditions there is a slight increase in the binding affinity due to protonation of a nitrogen (N<sub>2</sub>) in the D ring of ellipticine, which has a pK<sub>a</sub> of 7.4 (Dodin et al., 1988). These spectroscopic studies were confirmed with the very first crystallographic evidence of ellipticine bound to a ribodinucleoside monophosphate, where they showed an intercalative binding mode with a second molecule of ellipticine stacked above or below the base pair (figure 1.63 B) (Jain et al., 1979). This was in agreement with a later crystal structure of ellipticine bound to a 6 bp oligonucleotide d(CGATCG)<sub>2</sub>, which showed the same intercalative binding mode, but also showed preference to stack between GC base pairs rather than AT, most likely due to the presence of an additional hydrogen bond leading to a higher affinity binding site (figure 1.63 C) (Canals et al., 2005). Several different mechanisms of ellipticines anti-tumour properties have been reported, which involve both RNA and protein interactions such as: their ability to inhibit and introduce DNA breaks in the presence of topoisomerase II (Douc Rasy et al., 1983; Tewey et al., 1984); inhibiting the phosphorylation of the p53 tumour suppressor (Ohashi et al., 1995) ; inhibiting the wild-type and constitutively active mutant of the c-kit protein (Vendôme et al., 2005); as well as inhibiting components in the Akt pathway (Jin et al., 2004). In addition to the protein and duplex DNA interactions, ellipticines have also been reported to bind to G4s as a mechanism of its anti-tumour properties. For example, using spectroscopic analysis, it was shown that ellipticine can bind to the human telomere G4 (K<sub>d</sub> = 0.95 μM) by

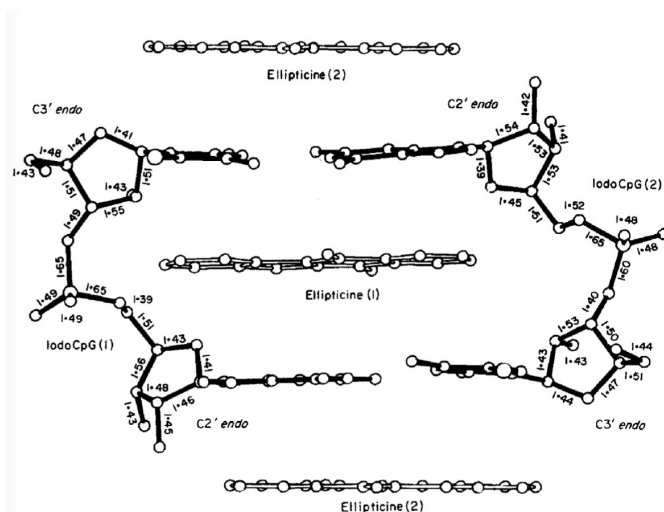
stacking on the external quartets and resulted in the inhibition of telomerase activity in breast cancer cell lines (Ghosh et al., 2013). Furthermore, it was also shown that an ellipticine derivative, GQC-05, which contains a tertiary amine at position 9 (figure 1.64 A) can selectively bind G4 structures over their duplex counterparts (figure 1.64 B), and has been shown to bind and stabilise the c-myc promotor quadruplex with a 2:1 stoichiometry, resulting in cytotoxic effects in Burkitt lymphoma cell lines (Brown et al., 2011). Molecular modelling of GQC-05 to the c-myc G4 structure suggested that the aromatic rings of GQC-05 stack into the external tetrad, with the tertiary amine at position 9 able to interact with the negatively charged phosphate group of thymine 19 in the double chain reversal loop (PDB of the c-myc structure 2L7V) (figure 1.64 C) (Kaiser et al., 2013). In addition to interacting with DNA G4s in the telomere and the myc promotor sequence, our lab has shown that GQC-05 has the ability to also bind to a putative RNA G4 in the Bcl-x pre-mRNA, and had been postulated as a mechanism to regulate its splicing pattern (see next section) (Weldon et al., 2018).



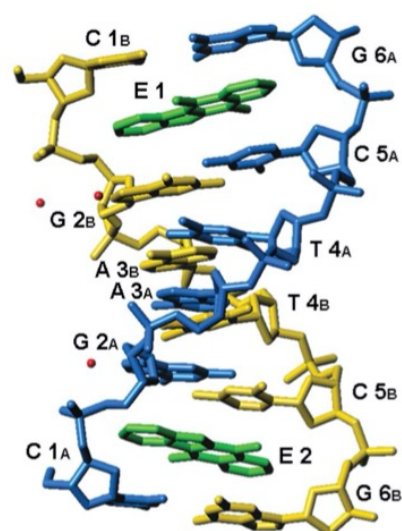
**A**



**B**

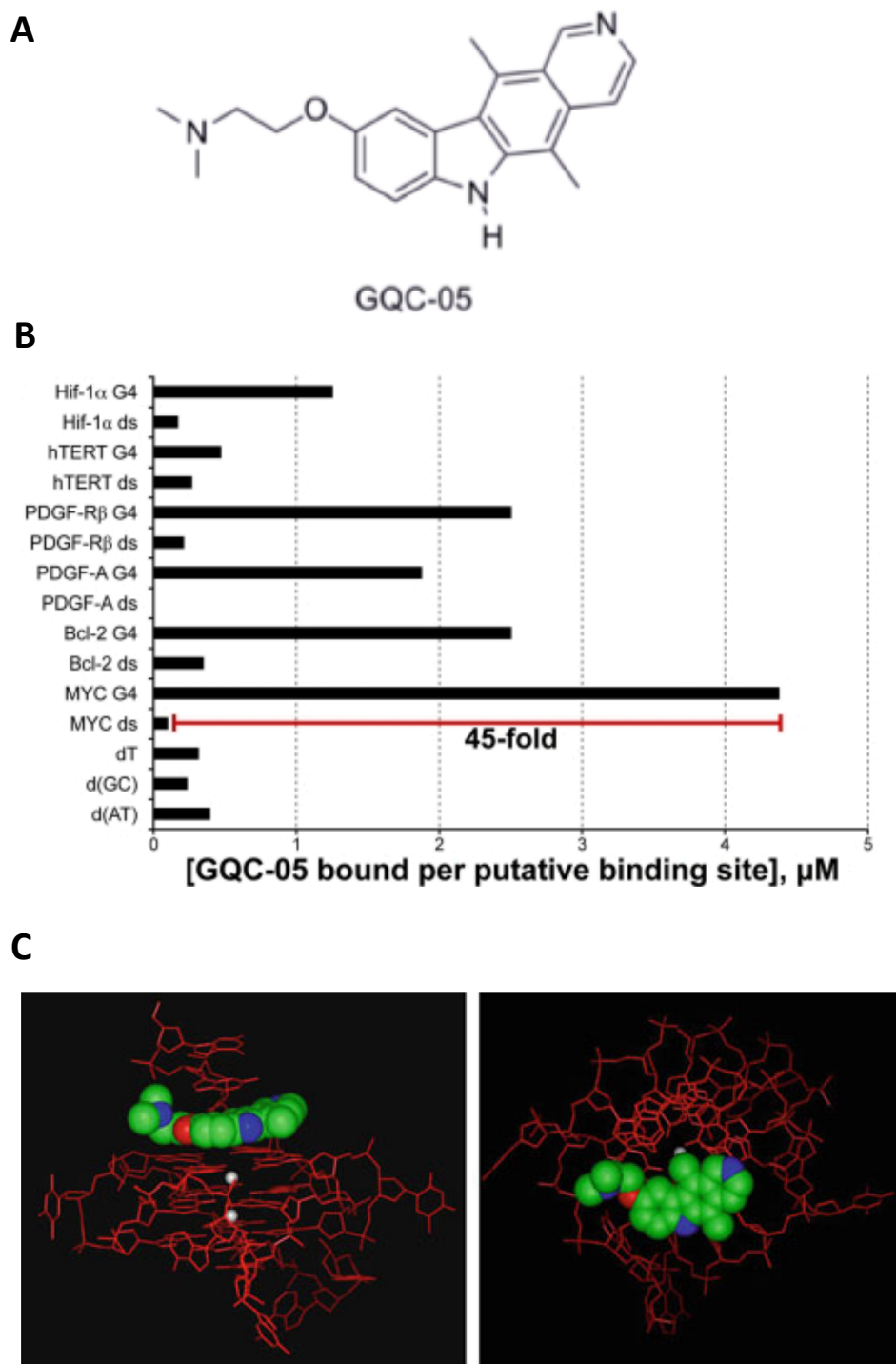


**C**



**Figure 1.63 Structure and complexes of the Ellipticine class of molecules.**

(A) The ellipticine core with the labelled rings. (B) First crystallographic evidence of ellipticine intercalating in a ribodinucleoside monophosphate. (C) Crystal structure of ellipticine bound to a 6 bp oligonucleotide d(CGATCG)<sub>2</sub>, showing a preferential intercalation between GC base pairs (Auclair., 1987; Canals et al., 2005; Jain et al., 1979; adapted from Stiborová et al., 2006).



**Figure 1.64 Binding specificity of GQC-05 and proposed binding mode.**

(A) Structure of the ellipticine derivative GQC-05. (B) Binding specificity of GQC-05 to known G4 structures and their duplex counterparts, as determined by competition dialysis experiments. (C) Binding model of GQC-05 to the c-myc DNA G4 structure (Brown et al., 2011; Kaiser et al., 2013).

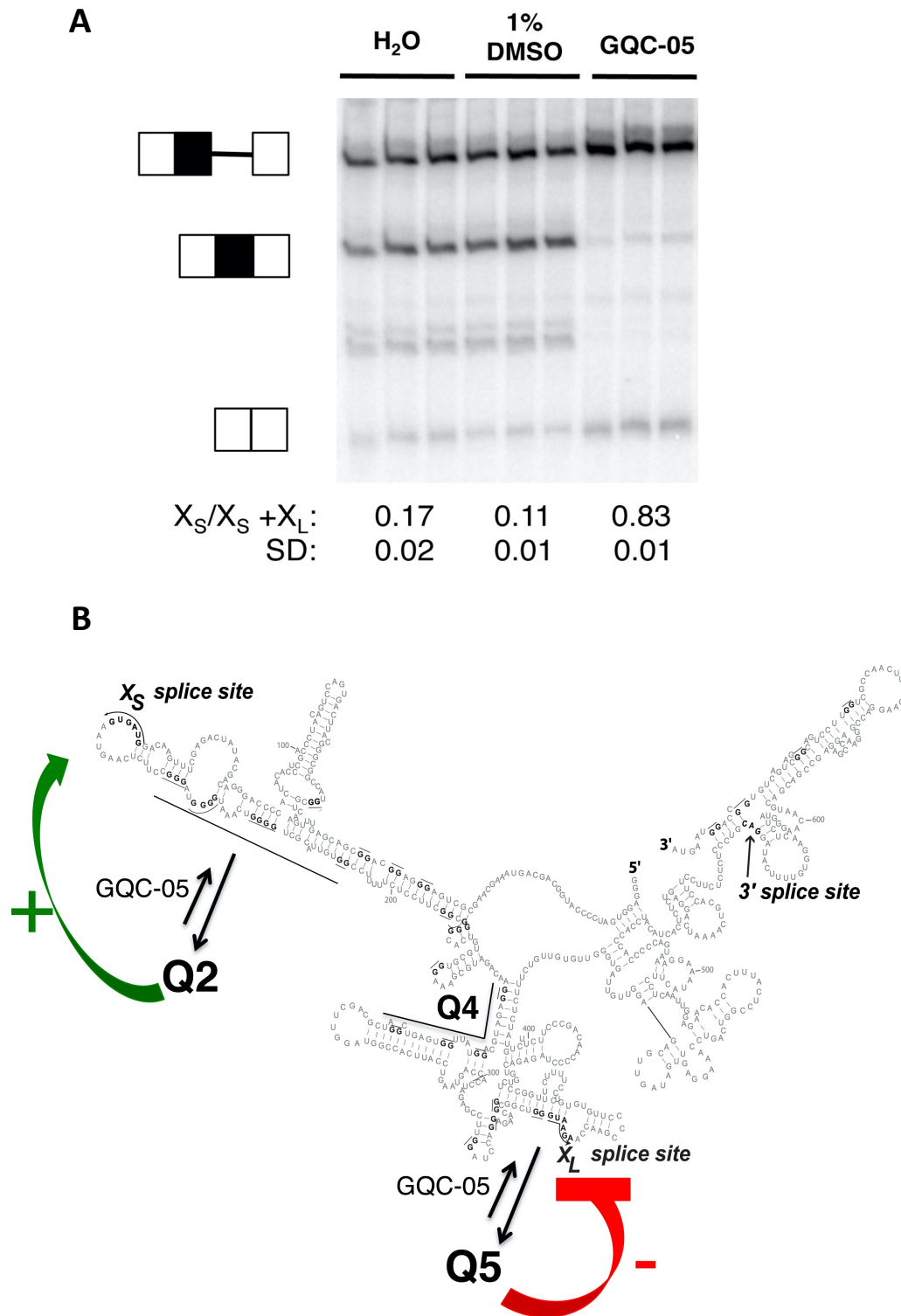
## 1.7 G4 ligands that regulate Bcl-x alternative splicing

Having shown that a G4 element exists proximal to both alternative 5'ss of Bcl-x (as explained above in section 1.5.7), the effects of known G4 ligands (and their derivatives) in shifting the splicing behaviour of the Bcl-x-681 pre-mRNA were investigated (Weldon et al., 2018).

As discussed in section 1.2.5, the Bcl-x mRNA has two antagonistic 5'ss in its exon 2. The use of the upstream site gives rise to the shorter pro-apoptotic  $X_S$  isoform, whereas the use of the downstream 5'ss gives rise to the longer anti-apoptotic  $X_L$  splice variant. The differences in the respective sizes between the two splice variants enables each of the isoforms to be easily visualised on a denaturing PAGE gel following an *in vitro* splicing reaction. Figure 1.71 A shows an *in vitro* splicing assay of the Bcl-x-681 mRNA in the absence and presence of the ellipticine derivative GQC-05 (a previously reported G4 stabilising ligand (Brown et al., 2011)). Under normal conditions, the  $X_L$  isoform predominates, giving a relatively low  $X_S/(X_S+X_L)$  ratio of around 0.1-0.2. However, upon the addition of GQC-05, there was an 8-fold increase in the  $X_S/(X_S+X_L)$  ratio, resulting from an increase in the  $X_S$  and a decrease in the  $X_L$  5'ss usage (Weldon et al., 2018). In contrast, other reported G4 stabilising ligands such as quarfloxin, PDS, carboxyPDS and TMPYP4 did not display this same increase, with some ligands like TMPYP4 completely inhibiting the splicing reaction (Weldon et al., 2018). The ligand 9-hydroxyellipticine (which contains the GQC-05 core but lacks tertiary amine), only inhibited the use of the downstream  $X_L$  site but had no effect in activating the upstream  $X_S$  site, indicating that the upregulation of the upstream site is not merely due to the downregulation of the downstream site, meaning that GQC-05 acts independently at each of the two alternative 5'ss (Weldon et al., 2018).

Having shown that a G4 has a potential to form in close proximity to the  $X_S$  and  $X_L$  5'ss (termed Q2 and Q5 respectively), G4 abolishing mutations in these regions attenuated the GQC-05 response (i.e. no  $X_S$  activation or  $X_L$  inhibition was observed in the presence of GQC-05), whilst the mutations had no effect on the normal splicing behaviour of Bcl-x in the absence of GQC-05 (Weldon et al., 2018). This suggests that GQC-05 is acting through the putative Q2 and Q5 G4s to modulate 5'ss selection, and does so in a specific manner compared to other known G4 stabilising ligands.

The exact mechanism by which GQC-05 interacts with the G4s as well as the consequences of such interactions remain to be explored. However, as many studies have highlighted the importance and relevance of RNA secondary structure elements in modulating splice site selection (Eperon et al., 1988, 1986; Grover et al., 1999), Weldon et al proposed a mechanism by which GQC-05 is able to stabilise a G4 at the upstream site, thereby making it more accessible to spliceosome components, whereas stabilisation of a G4 at the downstream would have an inhibitory effect (figure 1.71 B) (Weldon et al., 2018). A possible explanation at the X<sub>s</sub> site is that this 5'ss lies at the tip of a very large stem (based on the foot-printing structure, as in figure 1.71 B), thereby potentially making it inaccessible to the U1 snRNP. Stabilisation of the G4 within Q2 may therefore destabilise the surrounding stem structure, thereby making the site more accessible to U1 and result in splice site activation (Weldon et al., 2018). At the downstream site, the G-rich elements that have the propensity to fold a G4 (Q5) actually form part of the 5'ss sequence itself, unlike the upstream site where the Q2 G-tracts start several nucleotides downstream. Therefore, stabilisation of the Q5 G4 by GQC-05 may sequester this site in a G4 structure, thereby making it inaccessible to the U1 snRNP and hence splice site inhibition (Weldon et al., 2018). Additional questions that also need to be addressed is whether GQC-05 is binding to pre-formed G4s and stabilising them (a conformational selection mechanism), or whether GQC-05 can itself induce a G4 in expense of the surrounding duplex structure. Furthermore, it is not clear as to why GQC-05 is so unique at altering splice site selection in contrast to its derivatives and other known G4 binding ligands. Therefore, in this thesis, we aim to answer these questions by looking at the binding specificity of GQC-05 to these putative G4 structures and how such binding may alter the conformation of the RNA to achieve splice site selection. Furthermore, we will look at some derivatives of GQC-05 that do not show any changes in the splicing behaviour of Bcl-x and use biophysical tools to explain why GQC-05 is so unique at altering splice site selection.



**Figure 1.71 Bcl-x *in vitro* splicing and proposed mechanism of action of GQC-05.**

(A) *In vitro* splicing gel of Bcl-x in the absence and presence of 40  $\mu$ M GQC-05. (B) The mechanism by which GQC-05 can stabilize a G4 in Q2 and activate the  $X_S$  5'ss, whereas a GQC-05 mediated stabilization of the G4 at the  $X_L$  5'ss is having an inhibitory effect (Weldon et al., 2018).

## **CHAPTER 2: Materials and methods**

**2.1 RNA preparation**

**2.2 CD and UV spectroscopy measurements**

**2.3 NMR spectroscopy measurements**

**2.4 Fluorescence spectroscopy measurements**

**2.5 Pull down assays**

**2.6 Native gel assays**

**2.7 X-ray crystallography**

## **2.1 RNA preparation**

### **2.1.1 Short oligonucleotide preparation**

Short RNA oligonucleotides were purchased in their 2' protected form from Horizon Discovery and deprotected according to the manufacturer's protocol. In short, following centrifugation, 400  $\mu$ L of the provided deprotection buffer (consisting of 100 mM acetic acid and pH adjusted to 3.4-3.8 using TEMED) was added to the RNA until dissolved and incubated at 60°C for 30 minutes. This was then lyophilised overnight, and the pellet dissolved in milliQ water ready for use. The concentration was then determined using  $A_{260\text{nm}}$ .

### **2.1.2 RNA annealing protocol**

RNA oligonucleotides were annealed in each experimental condition separately by slow cooling. This involved incubating the RNA at 95°C for 5 minutes, which then was allowed to slowly cool to room temperature for a minimum of 12 hours.

### **2.1.3 Production of DNA templates for *in vitro* transcription**

#### **2.1.3.1 Plasmid design and stock preparation**

pEX-A128 plasmids with an ampicillin resistance gene containing the relevant inserts were purchased from Eurofins Genomics (see appendix 7.2). The inserts were designed to include a T7 RNA polymerase promotor, the desired Bcl-x fragments, a Bsa1 restriction site and a T7 terminator sequence (a detailed description of the insert can be found in Chapter 5). The plasmid pellets were dissolved in milliQ water to a concentration of 100 ng/ $\mu$ L.

#### **2.1.3.2 Transformation**

1  $\mu$ L of the plasmid stock at 100 ng/ $\mu$ L was added to 50  $\mu$ L of DH5  $\alpha$  *E.coli* competent cells (Invitrogen, 18265-017). Cells were then heat shocked for 45 seconds at 42°C and incubated on ice for 30 minutes. 500  $\mu$ L of 2TY growth media (see appendix 7.8) was added to the transformed cells and incubated for another 45 minutes at 37°C in a shaker at 200rpm. Cells were then pelleted by centrifugation (11,000 RPM for 2 minutes) and re-suspended in 200  $\mu$ L of growth media. 50  $\mu$ L of the re-suspended cells were then spread on an ampicillin containing agar plate (final concentration of 50  $\mu$ g/mL) and left to grow at 37°C for approximately 15 hours.

#### **2.1.3.3 Plasmid amplification**

A single bacterial colony was then extracted from the plate and grown for 15 hours in 750mL 2TY growth media containing 50 µg/mL ampicillin. Bacterial cells were then pelleted by centrifugation (4000 RPM with a SLC-6000 rotor for 10 minutes at 4°C).

#### **2.1.3.4 Plasmid purification**

Plasmids were then purified according to the manufacturer's protocol using a NucleoBond Xtra Maxi purification kit purchased from Macherey-Nagel (Catalogue number: 740424.50) In short, bacterial cells were re-suspended in a resuspension buffer containing RNase A and then lysed in an SDS containing lysis buffer to release the plasmid DNA from the bacterial debris. Plasmid DNA was then purified with the supplied column which consists of silica beads functionalised with methyl-amino-ethanol, a positively charged group in acidic conditions enabling it to bind to the negatively charged plasmid DNA. Following elution from the column, plasmids were precipitated using room-temperature isopropanol and dried using 70% ethanol. Plasmid pellets were then dissolved in 300 µL of milli Q water and their concentrations determined by  $A_{260nm}$ .

#### **2.1.3.5 Plasmid linearisation**

Plasmids were linearised with Bsa1 (New England Biolabs, catalogue number: R0535L). 1 unit of Bsa1 was used per µg of plasmid DNA. The digestion was done in the provided CutSmart buffer (final concentrations: 50 mM Potassium acetate, 20 mM Tris acetate, 10 mM Magnesium acetate, 100 µg/mL Bovine serum albumin (BSA), pH 7.9). After addition of each of the components, the appropriate amount of milli Q was added to adjust the final concentration of plasmid to 50 ng/µL. Samples were then incubated at 37°C, and a time course was carried out by taking samples out every hour for 4 hours and storing at -20°C. A further sample was also taken after 12 hours. Samples were then analysed on a 1% agarose gel made in Tris-Borate EDTA (TBE) buffer (see appendix 7.3 and 7.4), run for 120 minutes at 90 V and stained with ethidium bromide to check for complete digestion.



## **2.1.4 *In vitro* transcription of RNAs**

### **2.1.4.1 Transcription optimisation**

Transcription reactions were initially optimised in 20 µL final volumes by individually varying the concentrations of: MgCl<sub>2</sub> (40-100 mM), DNA (2.5-30 ng/µL), NTP mix (1.3-5.0 mM) (Sigma-Aldrich, catalogue numbers for each NTP: A7699, 51120, 94370, C1506) and T7 RNA polymerase (T7 RNAP) produced in the lab (1.7-10.2 µM) (expression and purification method of T7 RNAP described in section 2.1.9). Transcription reactions were conducted using a 20X transcription buffer stock (800 mM Tris-HCl pH 8.1, 20 mM Spermidine, 0.2% Triton X-100, 100 mM DTT), that was diluted to a final concentration of 1X in each of the transcription reactions. Samples were incubated at 37°C for 2 hours and then visualised on a 12% or 8% urea denaturing polyacrylamide gel running at 180 V for 100 minutes, stained with toluidine blue for 5 mins and de-stained with water for >30mins (see appendix 7.4 and 7.5).

### **2.1.4.2 Large scale transcriptions**

Following identification of the optimum conditions, transcription reactions were upscaled to 1 mL and left at 37°C for 4 hours. Samples were then split into 2 X 500 µL volumes followed by addition of DNase I (ThermoFisher scientific, catalogue number: EN0525) with a concentration of 2 units per µg of DNA template, incubated at 37°C for 30 minutes.

### **2.1.5 Purification of the large Xs, SD, and Q1 non-radiolabelled (cold) *in vitro* transcribed RNAs**

All purifications were done using an UltiMate 3000 Max HPLC (ThermoFisher) equipped with a column oven.

*In vitro* transcribed RNAs were centrifuged to remove the pyrophosphate pellet and filtered before column injection. Filtered transcription reactions were injected into a pre-equilibrated anion exchange preparative column (ThermoFisher DNAPac PA100 22 X 250 mm) using a 2 mL injection loop with a flow rate of 3mL/min and a temperature of 80°C. The column was equilibrated with 12.5 mM Tris HCl, pH 7.4, and the RNA was eluted using a gradient of 1M NaClO<sub>4</sub> in 12.5 mM Tris HCl, pH 7.4. Relevant fractions were collected, pooled together and then concentrated to approximately 1 mL by centrifugation at 4000 rpm for 60 minutes at 4°C using a 3 kDa cut-off amicon filter (Millipore, catalogue number: MPUFC900308). Samples were then dialysed in 2 litres of milliQ water for >12 hours using a 0.5-1 kDa Spectra-por

dialysis tubing (Cole Parmer, catalogue number: WZ-02906-21), followed by an additional 1 hour in 2 litres fresh milliQ water. Aliquots of the samples were then analysed on a 8% denaturing polyacrylamide urea gel at 180 V for 100 minutes and stained with toluidine blue to check for RNA purity. Concentrations were determined using  $A_{260\text{nm}}$ .

#### **2.1.6 Purification of the small Q2 and Q2.2 150-206 cold *in vitro* transcribed RNAs**

Filtered transcription reactions were injected into a pre-equilibrated anion exchange analytical column (DNAPac PA100 4 X 250 mm) using a 400  $\mu\text{L}$  injection loop with a flow rate of 1mL/min and a temperature of 80°C. The column was equilibrated with 12.5 mM Tris HCl, pH 7.4, and the RNA was eluted using a gradient of 1M  $\text{NaClO}_4$  in 12.5 mM Tris HCl, pH 7.4. Relevant fractions were collected, pooled together and then concentrated to approximately 1 mL by centrifugation at 4000 rpm for 60 minutes at 4°C using a 3 kDa cut-off amicon filter (Millipore, catalogue number: MPUFC900308). Samples were then dialysed in 2 litres of water for >12 hours in milliQ water using a 0.5-1 kDa Spectra-por dialysis tubing (Cole Parmer, catalogue number: WZ-02906-21), followed by an additional 1 hour in 2 litres fresh milliQ water. Samples were then run on a 12% denaturing polyacrylamide urea gel at 180 V for 100 minutes and stained with toluidine blue to check for RNA purity. Concentrations were determined using  $A_{260\text{nm}}$ .

#### **2.1.7 Single component injections of the transcription reagents**

40  $\mu\text{M}$   $\text{MgCl}_2$ , 30 ng/ $\mu\text{L}$  DNA, 3.1 mM NTPs and 1.7  $\mu\text{M}$  of T7 RNA polymerase were individually injected into the analytical column (DNAPac PA100 4 X 250 mm) using a 20  $\mu\text{L}$  injection loop with a flow rate of 1mL/min and a temperature of 80°C. The column was equilibrated with 12.5 mM Tris HCl, pH 7.4, and the components were eluted using a gradient of 1M  $\text{NaClO}_4$  in 12.5 mM Tris HCl, pH 7.4 .

#### **2.1.8 Synthesis of $^{32}\text{P}$ -RNA**

##### **2.1.8.1 *In vitro* transcription of $^{32}\text{P}$ -RNA**

$^{32}\text{P}$ -labeled RNAs were transcribed in 10  $\mu\text{L}$  volumes using the following conditions: 40 mM Tris pH 7.5, 6 mM  $\text{MgCl}_2$ , 10 mM NaCl, 2 mM spermidine HCl, 1 mM Ribo m7G cap analogue (Promega, catalogue number: P1711 ), 50 ng of the DNA template, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 0.05 mM GTP, 0.33  $\mu\text{M}$   $\alpha$   $^{32}\text{P}$ -GTP (Pelkin Elmer, catalogue number: BLU506H250UC), 4% (v/v) RNase out inhibitor cocktail (ThermoFisher, catalogue number:

10777019) and 0.069 mg/mL T7 RNA polymerase (synthesised in house, see section 2.1.9). Samples were then incubated at 37°C for 2 hours and loaded on a 8% de-naturing PAGE gel to enable gel purification.

#### **2.1.8.2 Purification of *in vitro* transcribed <sup>32</sup>P-RNAs**

RNA bands were located on the gel by exposure of an x-ray film for 10 minutes, enabling band excision. The excised bands were then soaked in elution buffer (500 mM sodium acetate pH 5.2, 0.2% SDS, 1 mM EDTA pH 8.0) for >12 hours at 4°C to allow the RNA to diffuse from the gel into the buffer. The elution buffer containing the RNA was precipitated by ethanol (see section 2.1.6.3) and the pellet re-suspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0 and RNase inhibitor cocktail) and stored at -80°C ready for use.

#### **2.1.8.3 Ethanol precipitation of <sup>32</sup>P-RNA**

The elution buffer containing the hot RNA was incubated at 30°C for 5 minutes and the empty gel piece was removed and discarded. 1 mL of 100% ethanol was added to the RNA, mixed by inversion and the RNA was pelleted by centrifugation (13000 RPM for 20 minutes). The supernatant was removed and a further 0.2 mL of ethanol was added, followed by centrifugation for 15 minutes. The supernatant was removed and the pellet dried using a vacuum desiccator.

#### **2.1.8.4 *In vitro* transcription of 7-deaza-GTP <sup>32</sup>P-RNA**

7-deaza-GTP (c<sup>7</sup>-GTP) RNAs were transcribed in 10 µL volumes using the following conditions: 40 mM Tris pH 7.5, 6 mM MgCl<sub>2</sub>, 10 mM NaCl, 2 mM spermidine HCl, 1 mM Ribo m7G cap analogue (Promega, catalogue number: P1711), 50 ng of the DNA template, 0.5 mM ATP, 0.5 mM CTP, 0.05 mM UTP, 0.5 mM c<sup>7</sup>-GTP, 0.33 µM α <sup>32</sup>P-UTP (Pelkin Elmer, catalogue number: BLU506H250UC ), 4% (v/v) RNase out inhibitor cocktail (ThermoFisher, catalogue number: 10777019) and 0.069 mg/mL T7 RNA polymerase (synthesised in house, see section 2.1.9). Samples were then incubated at 37°C for 2 hours, loaded on a 8% de-naturing PAGE gel, purified and ethanol precipitated as described in sections 2.1.6.2 and 2.1.6.3).

#### **2.1.9 Expression and purification of T7 RNA polymerase (T7 RNAP)**

The His tag T7 RNAP gene was previously cloned into the plasmid p911, which was then transformed into BL21 Rosetta competent cells. Transformed cells were grown on agar plates containing antibiotics ampicillin and chloramphenicol (at final concentrations of 50 µg/mL),

and left at 37°C for 12 hours. Resulting colonies were transferred into 5 mL of 2TY growth media, along with the antibiotics ampicillin and chloramphenicol (at final concentrations of 50 µg/mL) and incubated for 1 hour at 37°C in a shaking incubator (120rpm). The 5 mL culture was then transferred into a 50 mL starter culture with the addition of ampicillin and chloramphenicol to final concentrations of 50 µg/mL, and left to grow for 12 hours at 37°C in a shaking incubator (120rpm). The 50 mL starter culture was then inoculated into 750 mL 2TY media, along with the antibiotics ampicillin and chloramphenicol (final concentrations of 50 µg/mL), and left to grow at 37°C in a shaking incubator (120 rpm) until an OD<sub>595nm</sub> of 0.5-0.7. The 750 mL culture was then allowed to equilibrate at 20°C for 45 minutes, which was then induced for protein expression by the addition of IPTG (to final concentration of 533 µM), and left at 20°C for a minimum of 12 hours in a shaking incubator (120 rpm). Bacterial cells were then pelleted by centrifugation (4000 rpm with a SLC-6000 rotor for 20 minutes at 4°C). Pelleted cells were resuspended in 30 mL lysis buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> pH 8, 300 mM NaCl, 10 mM imidazole) and 300 µL Triton X100. Resuspended cells were then lysed by sonication (30 seconds on, 30 seconds off, 5 times) followed by centrifugation for 20 minutes at 18,000 rpm at 4°C to clarify the cell lysate. Prior to the subsequent purification steps, all buffers were chilled to 4°C to prevent protein degradation by residual proteases. The bacterial lysate was then loaded onto an equilibrated 5 mL Ni-NTA column, with all subsequent eluates collected and kept on ice. The column was then washed with 2 column volumes (C.V) of lysis buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> pH 8, 300 mM NaCl, 10 mM imidazole) and then the protein was eluted with 2 C.V of elution buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> pH 8, 300 mM NaCl, 500 mM imidazole). A further 2 C.V of 1 M imidazole was applied to ensure all the protein is eluted from the column. All eluted fractions were then analysed on an SDS PAGE gel. Fractions containing the purified T7 RNAP were then pooled together and dialysed in 2 litres of dialysis buffer (50 mM Tris pH 8, 100 mM NaCl, 0.05% beta-mecaptoethanol) for a minimum of 12 hours at 4°C. The protein concentration was then determined using A<sub>280nm</sub>, and stored in 25% glycerol at -20°C.

## **2.2 Circular dichroism (CD) and ultra violet (UV) spectroscopy measurements**

### **2.2.1 Room temperature spectrum in the absence of nuclear extract**

All CD and UV experiments were conducted on a Chirascan CD spectropolarimeter (Applied Photophysics). 2  $\mu$ M RNA samples were annealed as described in section 2.1.2 in 10 mM TBA phosphate buffer pH 7.0, with either 100 mM KCl or LiCl. Experiments were conducted in a 1 cm pathlength quartz cuvette (total volume of 1.5 mL), with a 0.5s time per point, with three scans taken and averaged per wavelength. The spectra were recorded between 200 and 400 nm wavelengths, with a 1 nm measurement interval.

### **2.2.2 Room temperature spectrum in the presence of nuclear extract**

10 $\mu$ M RNA samples were annealed in 10 mM TBA phosphate buffer pH 7.0, with 100 mM KCl or LiCl. 5% (v/v) HeLa nuclear extract (Cil Biotech, CC-01-20-10) was added post annealing. Experiments were conducted in a 1 mm pathlength quartz cuvette, with a 0.5s time per point, with three scans taken and averaged. The spectrum was recorded between 200 and 400 nm wavelengths, with a 1 nm measurement interval.

### **2.2.3 Induced CD spectrum recordings**

2  $\mu$ M c-myc DNA and RNA samples were annealed in 10 mM Tris pH 7.0 and with either 100 or 8 mM KCl. An initial spectrum was recorded, followed by another spectrum after the addition of 10 equivalents of GQC-05 and its analogues. Experiments were conducted in a 1 cm pathlength quartz cuvette, with a 0.5s time per point, with three scans taken and averaged. The spectrum was recorded between 200 and 400 nm wavelengths, with a 1 nm measurement interval.

### **2.2.4 CD melting experiments with GQC-05**

RNA samples were annealed in 10 mM TBA phosphate buffer, pH 7.0 and 10 mM KCl. Melting experiments were conducted with RNAs either without and with the addition of 10  $\mu$ M GQC-05. The data was collected using a wavelength range from 220 to 320nm from 5 to 94°C, with a heating rate of 1°C in a stepped ramp mode. Experiments were conducted in a 1 cm pathlength quartz cuvette, with a 0.5s time per point, with three scans taken and averaged at each temperature. Data at 263 nm was normalised and fitted with a Boltzmann sigmoidal curve on GraphPad prism Version 9.0.

### **2.2.5 CD spectra with NMM**

CD spectra was recorded with 1  $\mu$ M of the Q2 RNA (pre-annealed in 10 mM Tris pH 7.0) and 10  $\mu$ M NMM. Experiments were conducted in a 1 cm pathlength quartz cuvette, with a 0.5s time per point, with three scans taken and averaged. The spectrum was recorded between 200 and 400 nm wavelengths, with a 1 nm measurement interval.

## **2.3 NMR spectroscopy measurements**

### **2.3.1 1D NMR spectrum recordings of WT and mutant fragments**

All  $^1\text{H}$  NMR experiments were conducted on a 600 MHz Bruker spectrometer at 283 K using an watergate pulse sequence for suppression of the water signal. Spectral acquisition involved the collection of 256 scans with a relaxation delay of 2 seconds. A Fourier transform was applied using a cosine function. Samples of 100  $\mu$ M RNA were annealed (see 2.1.2) in 6mM  $\text{K}_2\text{HPO}_4$ , 4 mM  $\text{KH}_2\text{PO}_4$ , 100 mM KCl, pH 7.0, followed by the addition of 10%  $\text{D}_2\text{O}$  (v/v) post annealing.

### **2.3.2 KCl titrations**

100  $\mu$ M RNA sample was annealed in 10 mM Tris pH 7.0 in the absence of any salt, followed by the addition of 10%  $\text{D}_2\text{O}$  (v/v) post annealing and a 1D spectrum was record. This was followed by successive additions of small volumes of 4 M KCl, to final concentrations of 10, 20 and 100 mM, with a spectrum recorded in each instance.

### **2.3.3 GQC-05 titrations**

100  $\mu$ M of the Q2 RNA sample was annealed in 10 mM Tris pH 7.0 in the absence of any salt, followed by the addition of 10%  $\text{D}_2\text{O}$  (v/v) post annealing and a 1D spectrum was record. This was followed by increasing additions of 10 mM GQC-05 (dissolved in 100%  $\text{d}_2\text{DMSO}$ ) from 100 to 300  $\mu$ M, with a spectrum recorded in each instance.

### **2.3.4 NMR kinetics and pseudo 2D recordings**

100  $\mu$ M of the Q2 RNA sample was annealed in 10 mM Tris pH 7.0, 8 mM KCl, followed by the addition of 10%  $\text{D}_2\text{O}$  (v/v) post annealing and a 1D spectrum was record. This was followed by the addition of the Q2 complementary strand at an equimolar concentration, which defined the start of the experiment ( $t_0$ ).  $T_0$  was determined to be the void time in the NMR experimental setup with the addition of the time required to record the first 1D spectrum.

The time per spectrum was 160 seconds, which was recorded over a period of 72 hours to generate the pseudo 2D spectra. This was repeated in the presence of 4 equivalents of GQC-05, which was added to the sample post annealing of the Q2 strand, followed by addition of the complementary strand to begin the experiment. Data were analysed by peak integrals of the duplex regions using dynamic centre on Top-spin and plotted on GraphPad prism Version 9.0. Rate curves corresponding to duplex formation could be fitted using first order kinetics.

## **2.4 Fluorescence spectroscopy measurements**

All fluorescence spectra were obtained on a Fluoromax-4 fluorimeter using a 1cm pathlength quartz cuvette designed for fluorescence measurements or Hidex microplate reader on a 384 well plate (Corning).

### **2.4.1 Fluorescence emission spectra recordings with GQC-05 and c-myc DNA**

10  $\mu$ M of the c-myc DNA were annealed in 100 mM KCl, 10 mM Tris pH 7.0, followed by the addition of 20  $\mu$ M GQC-05. The emission spectrum (Fluoromax) was recorded between 350-600 nm, with an excitation wavelength of 321 nm and using a 5 nm excitation and emission slit length.

### **2.4.2 Fluorescence emission spectra with NMM and the Bcl-x RNAs**

10  $\mu$ M of the RNAs were annealed in the relevant salt concentrations in 10 mM Tris pH 7.0, followed by the addition of 1  $\mu$ M NMM. The spectrum (Fluoromax) was recorded between 450-700 nm, with a 5 nm excitation and emission slit length and an excitation wavelength of 393 nm.

### **2.4.3 Fluorescence emission spectra of RNA secondary structures**

10  $\mu$ M of the RNAs were annealed in 100 mM KCl or LiCl in Tris pH 7.0 and a spectrum recorded with an excitation wavelength of 321 nm. The spectrum (Fluoromax) was recorded between 350-600 nm, with a 5 nm excitation and emission slit length. This was then followed by the addition of 1  $\mu$ M GQC-05 to a separate RNA sample annealed in the same conditions, and the fluorescence spectrum recorded with the same parameters.

### **2.4.4 Fluorescence titrations in the absence of nuclear extract**

All fluorescence titrations were conducted on a Hidex microplate reader using 1  $\mu$ M GQC-05, with an excitation wavelength of 320 nm and an excitation and emission slit length of 5 nm.

100  $\mu$ M RNA samples was annealed in 10 mM Tris pH 7.0, 100 mM KCl, followed by the addition of 1  $\mu$ M GQC-05 post annealing. Concentrations of the RNA was varied through serial dilutions in a buffer containing 10 mM Tris pH 7.0, 100 mM KCl and 1  $\mu$ M GQC-05, in a total of 17 wells. The final fluorescence signal was obtained after subtracting the fluorescence of the RNA only sample and the GQC-05 background signal from the RNA-GQC-05 complex signal. Data were analysed using GraphPad prism Version 9.0, and fitted to a single site binding isotherm equation.

#### **2.4.5 Fluorescence titrations in the presence of nuclear extract**

Experiments were conducted in exactly the same conditions as in 2.4.4, but with the addition of 10% HeLa nuclear extract (v/v) (Cilbiotech) post annealing.

#### **2.4.6 Job plot assays**

##### **2.4.6.1 Job plot with c-myc DNA**

100  $\mu$ M c-myc DNA samples were annealed in 100 mM KCl, 10 mM Tris pH 7.0. Separate samples were then prepared with various molar ratios of [GQC-05]:[c-myc], while keeping the total molar ratio constant at 1. Each sample corresponding to a different molar ratio was loaded onto a 96 well plate and the fluorescence emission of GQC-05 was measured on the Hidex microplate reader with an excitation wavelength of 320 nm. The GQC-05 only fluorescence signal was also plotted at increasing concentrations in the same buffer. Data were analysed using GraphPad prism Version 9.0.

##### **2.4.6.2 Job plot with Q2 RNA in the absence of nuclear extract**

Experiments were conducted as in 2.4.6.1 but with 8 mM instead of 100 mM KCl.

##### **2.4.6.3 Job plot with Q2 RNA in the presence of nuclear extract**

Experiments were conducted as in 2.4.6.2, but with the addition of 10% HeLa nuclear extract (v/v) (Cilbiotech). The corrected fluorescence was calculated by subtraction of the background signal coming from GQC-05 in nuclear extract. This had to be done in this case due to the high background coming from the interaction of GQC-05 with components in the nuclear extract. Data was analysed on GraphPad prism Version 9.0.

### **2.5 Pull-down assays**

30  $\mu$ L of a 50% neutravidin slurry (ThermoFisher, catalogue number: 29200) was equilibrated to room temperature and washed with PBS. This was followed by the addition of 10  $\mu$ M



biotinylated GQC-05 (gift from Prof. Burley, University of Strathclyde) and left for 30 minutes at room temperature. 10 counts per seconds (c.p.s) of  $^{32}\text{P}$ -RNA was added followed by a further 30 minutes of incubation. Samples were washed once with PBS, and eluted by boiling the beads at 95°C for 5 mins. The unbound, wash and elution fraction were loaded onto an 8% denaturing polyacrylamide gel and ran at 180 V for 100 mins. Gels were visualised using a phosphorimager screen following exposure for >12 hours. Control experiments were conducted with the RNA and beads without the addition of the biotinylated GQC-05. Data was analysed by calculating the fraction bound (bound/total) through determination of band intensities using image J. Data was analysed on GraphPad prism Version 9.0.

## **2.6 Native gel assays**

### **2.6.1 Q2 KCl titrations**

10  $\mu\text{M}$  Q2 RNA samples were annealed in 10 mM Tris pH 7.0 with increasing KCl concentrations (0-20 mM). Samples were loaded on a 12% native PAGE gel at 180 V for 100 minutes and stained with toluidine blue for 5 minutes and de-stained in water for > 30 minutes. A mixture of poly T DNAs with sizes ranging from 20-120 nucleotides was used as a ladder.

### **2.6.2 GQC-05 titrations**

10  $\mu\text{M}$  Q2 RNA samples were annealed in 10 mM Tris pH 7.0 and 8 mM KCl. Increasing equivalents of GQC-05 was added post annealing, from 0 to 40  $\mu\text{M}$ . Samples were loaded on a 12% native PAGE gel at 180 V for 100 minutes and stained with toluidine blue for 5 minutes and de-stained in water for > 30 minutes. A mixture of poly T DNAs with sizes ranging from 20-120 nucleotides was used as a ladder.

### **2.6.3 EMSA assays with the U1 snRNA in the absence of poly T**

6  $\mu\text{M}$  of the fluorescently labelled U1 oligo was added to increasing concentrations of the X<sub>s</sub> domain RNA, with a maximum concentration of 50  $\mu\text{M}$ . Experiments were carried out in 10 mM Tris pH 7.0, 100 mM KCl, but with no annealing steps to replicate the *in vitro* splicing conditions. Samples in the presence of the ligand included 50-200  $\mu\text{M}$  GQC-05. All samples were left for 30 minutes at room temperature, and then loaded on a 16% native PAGE gel for 120 minutes at 4°C and 180 V. Samples were visualised under UV light. The fraction bound was then determined using band intensities of the bound and unbound U1 fluorescence

signals, and analysed using GraphPad prism Version 9.0 by fitting of the data to a single site binding isotherm equation.

#### **2.6.4: Measuring effects of GQC-05 on the U1 migration**

6  $\mu\text{M}$  of the fluorescently labelled U1 oligo was added to increasing concentrations of GQC-05, from 50-200  $\mu\text{M}$ . Further samples containing 50  $\mu\text{M}$  U1 with the addition of 415  $\mu\text{M}$  and 1.7 mM GQC-05 were also tested. Samples were prepared in 10 mM Tris pH 7.0, 100 mM KCl, and incubated at room temperature for 30 minutes. All samples were loaded on a 16% native PAGE gel and ran for 120 minutes at 4°C and 180 V. Samples were visualised under UV light, followed by toluidine blue staining for 5 minutes and de-stained in water for > 30 minutes.

#### **2.6.5: Poly T optimisations**

Increasing concentrations of poly T-50 (0-90  $\mu\text{M}$ ) were added to 6  $\mu\text{M}$  of the fluorescently labelled U1 oligo incubated with 200  $\mu\text{M}$  GQC-05. Samples were prepared in 10 mM Tris pH 7.0, 100 mM KCl. All samples were left for 30 minutes at room temperature, and then loaded on a 16% native PAGE gel for 120 minutes at 4°C and 180 V. Samples were visualised under UV light. Band intensities of the U1 oligo and the Poly T GQC-05 complexes were determined using image J and analysed on GraphPad prism Version 9.0

#### **2.6.6 EMSA with Poly T and GQC-05**

In all experiments, 6  $\mu\text{M}$  of the fluorescently labelled U1 oligo was added to increasing concentrations of the X<sub>s</sub> domain RNA, with a maximum concentration of 50  $\mu\text{M}$  and the addition of 70  $\mu\text{M}$  poly T-50 DNA. Experiments were carried out in 10 mM Tris pH 7.0, 100 mM KCl, with no annealing steps. Samples in the presence of ligand included 200  $\mu\text{M}$  GQC-05. All samples were left for 30 minutes at room temperature, and then loaded on a 16% native PAGE gel for 120 minutes at 4°C and 180 V. Samples were visualised under UV light. The disappearance of the U1 intensity band was determined using image J and analysed on GraphPad prism Version 9.0, by fitting of the data to a one phase decay equation

### **2.7 X-ray crystallography**

The commercially available helix screen (Molecular Dimensions), consisting of 96 different conditions was used. 2 mM of the Q2 RNA was annealed in the presence of 10 mM Tris pH

7.0. 80  $\mu$ L of each of the 96 conditions was pipetted onto a sitting drop crystallisation plate. 100 nL of the annealed Q2 RNA was pipetted onto the crystallisation plate using an Oryx8 robot (Douglas Instruments), followed by a further 100 nL addition of the crystallisation screen, giving a total volume of 200 nL per drop, per condition. The final concentration of RNA in each droplet was 1 mM. This was also repeated for the control without the RNA, where only the Tris buffer was mixed with the crystallisation screen in the same 1:1 ratio. The crystallisation plate was sealed and left at room temperature and visualised on a weekly basis.

## **CHAPTER 3: Structural and biophysical characterisation of the putative G4 region in Bcl-x (Q2)**

### **3.1 Introduction**

### **3.2 Bioinformatic analysis of the Q2 forming region**

### **3.3 Using CD and UV spectroscopy to characterise the Q2 structure**

### **3.4 Using NMR spectroscopy to characterise the Q2 structure**

### **3.5 Approaches to reduce the structural heterogeneity of Q2**

### **3.6 Characterisation of the G4 structure in the absence of salt**

### **3.7 X-ray crystallography of Q2**

### **3.8 Discussion**

### 3.1 Introduction

Previous studies from our lab have shown by *in vitro* splicing and foot-printing that a putative G-quadruplex might play an important role in the activation of the X<sub>s</sub> 5'ss, and that GQC-05 may regulate splicing by binding through a putative quadruplex, thereby enhancing the X<sub>s</sub> isoform (Weldon et al., 2018). A short stretch consisting of 30 nucleotides (termed Q2 in table 3.1), lies 14 nucleotides downstream of the X<sub>s</sub> 5'ss and has a sufficient number of guanines to follow the canonical G4 folding topology of: G<sub>x</sub>N<sub>y1</sub>G<sub>x</sub>N<sub>y2</sub>G<sub>x</sub>N<sub>y3</sub>G<sub>x</sub>, where x is the number of consecutive guanines (which must be at least 2) and y is the length of the connecting loops (Kikin et al., 2006). Our lab showed previously that disrupting the third G-tract (mutant Q2.2 in table 3.1), thereby abolishing the G4 canonical sequence, prevents GQC-05 activation of X<sub>s</sub>, suggesting that GQC-05 is acting through this quadruplex downstream of the X<sub>s</sub> 5'ss (Weldon et al., 2018). Therefore, using biophysical and structural techniques, we aim to identify if this region can fold into a quadruplex, and whether the G4 mutation is sufficient to abolish a G4 fold in order to validate the G4 model of splice site activation by GQC-05.

Sequence name	Sequence
Q2 (WT)	GGGAUGGGGUAAACUGGGGUCGCAUUGUGG
Q2.2	GGGAUGGGGUAAACUGUUUCGCAUUGUGG
ΔG4	GUGAUGUUGUAAACUGUUUCGCAUUGUGU
Q2 C	GGUAUUUGGUAAACUUUGGUCGCAUUGUGG

**Table 3.1 RNA oligonucleotide fragments.**

Showing the short RNA fragments used for biophysical and structural characterisation of the RNA and GQC-05 binding ability.

### 3.2 Bioinformatic analysis of the Q2 forming region

As discussed in section 1.5.1, following identification of a guanine rich strand, the initial stage is to validate the G4 forming potential of such sequence using bioinformatic tools. Here we have used two different tools to predict the propensity of Q2 to fold into a G4: QGRS mapper, a G4 predicting programme that has been used to predict many canonical DNA G4s with high

accuracy which shows remarkable resemblances to experimentally determined structures (Smith *et al.*, 1995); as well as RNA fold which predicts the propensity of a sequence to fold into a duplex structure. As G4s and duplexes are often competing secondary structures, using these tools in tandem will allow us to predict if a sequence is more likely to form a duplex or a quadruplex.

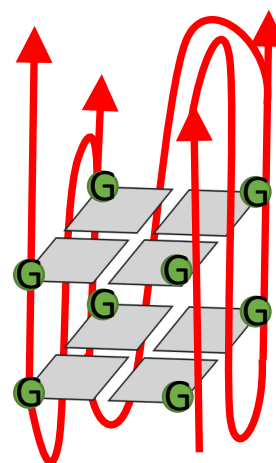
QGRS mapper has the ability to predict the propensity of a given G-rich sequence to fold into a G4 by providing a G-score. This G-score indicates the likelihood of G4 forming potential, with higher scores suggesting more stable G4 topologies, and is governed by the number of G-tetrads within the G4 as well as the length of the connecting loops (see section 1.5.1 for a more in depth discussion). A single input sequence into QGRS mapper gives rise to a cluster of G4 forming sequences, each of which is given a G score and the relevant guanines involved in tetrad formation are highlighted. This is particularly useful when looking at sample heterogeneity and isolating a single species for structural characterisation (see section 3.5). Below shows the QGRS mapper output for the Bcl-x Q2 sequence (figure 3.21 A). QGRS mapper predicts that this sequence can adopt multiple G4 topologies, indicated by a range of similar G-scores, the most stable of which has a score of 17. The wide variety of G4 topologies identified by QGRS mapper arise due to the very nature of the Q2 sequence and the fact that this programme can only predict canonical G4 structures. For canonical G4s, there must be at least four runs of two or more consecutive guanines. Based on the sequence of Q2 in table 3.1, the very last run of consecutive guanines is GG. This would mean, if Q2 is a canonical G4, the maximum number of tetrads must be two, meaning that the G4 structure predicted by QGRS mapper is a two-tiered quadruplex (figure 3.21 B). Therefore, the first three runs of Gs in Q2 which display 3,4,4 consecutive guanines respectively must only contain 2 of those guanines incorporated into the tetrad, with the remaining Gs being part of the loop. This has the ability to give rise to multiple different G4 structures with different guanines being included or excluded in the G-tetrads. For example, the difference between sequences containing G-scores of 17 and 16 arises in the third G-tract containing GGGG, the former has the last two guanines incorporated in the tetrad, whereas the one with a score of 16 has the middle two incorporated. This may result in multiple G4 species to exist in solution, each one having different guanines incorporated into the tetrad, with the potential to make structural characterisation a challenge (see section 3.5 on methods that can be used to tackle G4 heterogeneity). However, it is important to first characterise with experimental evidence if

the Q2 sequence can indeed fold into a G4 structure in the first instance, and then if such structure follows a canonical or non-canonical topology. Figure 3.22 shows the RNA fold output for Q2, which shows that a stem loop also has the potential to form (which occurs with a minimum free energy of  $-2.30 \text{ kcal mol}^{-1}$ ) and can potentially outcompete a putative G4 that may exist in Q2. Therefore, using biophysical techniques, we aim to determine if the G4 or the duplex is the major structural isoform that this sequence adopts.

**A**

Position	Length	QGRS	G-Score
1	17	GGGAU <del>GGGG</del> UAAACUGG	15
1	18	GGGAU <del>GGGG</del> UAAACUGGG	14
1	19	GGGAU <del>GGGG</del> UAAACUGGGG	13
1	19	GGGAU <del>GGGG</del> UAAACUGGGG	13
1	19	GGGAU <del>GGGG</del> UAAACUGGGG	14
1	19	GGGAUG <del>GGG</del> UAAACUGGGG	15
1	30	GGGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	2
1	30	GGGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	13
1	30	GGGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	14
1	30	GGGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	14
1	30	GGGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	14
1	30	GGGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	15
1	30	GGGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	16
1	30	GGGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	15
1	30	GGGAUG <del>GGG</del> UAAACUGGGGUCGCAUUGUGG	16
1	30	GGGAUG <del>GGG</del> UAAACUGGGGUCGCAUUGUGG	17
1	30	GGGAUG <del>GGG</del> UAAACUGGGGUCGCAUUGUGG	8
2	16	GGAU <del>GGGG</del> UAAACUGG	15
2	17	GGAU <del>GGGG</del> UAAACUGGG	14
2	18	GGAU <del>GGGG</del> UAAACUGGGG	13
2	18	GGAU <del>GGGG</del> UAAACUGGGG	13
2	18	GGAU <del>GGGG</del> UAAACUGGGG	14
2	18	GGAU <del>GGGG</del> UAAACUGGGG	15
2	29	GGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	2
2	29	GGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	12
2	29	GGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	13
2	30	GGAU <del>GGGG</del> UAAACUGGGGUCGCAUUGUGG	13

**B**



**Figure 3.21 QGRS mapper analysis of Q2 and proposed model.**

(A) Output from QGRS mapper for Q2 RNA, showing the possibilities of the different structures in solution, each with their respective G-score. (B) Schematic model of the possible two tiered Q2 RNA quadruplex.



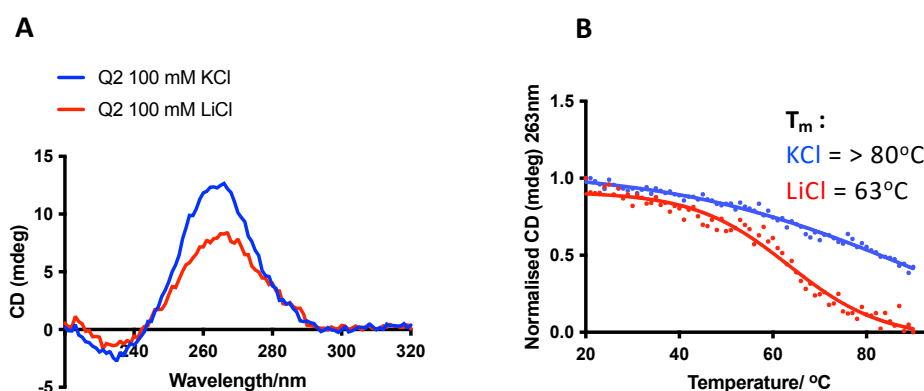
114



### 3.3 Using CD and UV spectroscopy to characterise the Q2 structure

#### 3.3.1 CD spectroscopy of Q2 and mutants

Shown below is the CD spectrum of 2  $\mu\text{M}$  of the Q2 RNA at 20 degrees, which displays a positive peak at 263 nm, a characteristic signature corresponding to either a parallel quadruplex or to a duplex (figure 3.31 A). Therefore, in order to distinguish between these structures, we exploited the fact that G4 stability is enhanced in the presence of  $\text{K}^+$  as compared to  $\text{Li}^+$ , whereas duplex structures are almost equally stable in both  $\text{K}^+$  and  $\text{Li}^+$  - containing buffers. At 20 degrees, the CD spectrum in the presence of potassium showed a greater intensity as well as a shift to 263 nm compared to lithium (figure 3.31 A), suggesting that Q2 adopts a G-quadruplex conformation and not a duplex. CD thermal melting also showed a higher stability in the presence of potassium than lithium, with a  $T_m$  in the presence of  $\text{Li}^+$  at  $63^\circ\text{C}$ , and a  $T_m$  in the presence of  $\text{K}^+$   $> 80^\circ\text{C}$ , further indicating that the Q2 sequence folds into a G-quadruplex (figure 3.31 B).



**Figure 3.31 CD spectroscopy of the Q2 fragment.**

(A) CD spectra of Q2 RNA at  $20^\circ\text{C}$  in the presence of 100 mM KCl (blue) and LiCl (red) in 10 mM TBA buffer pH 7.0. (B) CD thermal melting of Q2 RNA in the presence of 100 mM KCl (blue) and LiCl (red) in the same buffer as (A).

To confirm that these effects are G4 mediated, a mutant version of the wild-type sequence, Q2.2 (Table 3.1), was tested. This mutant contains a GG to **UU** mutation in the third G-tract, which abolishes the potential for G4 formation, shown by a relatively lower G-score from

QGRS mapper of only 2 in contrast to 17 with the wild-type sequence (figure 3.32). In addition, the RNA fold prediction shows that a duplex has a greater propensity to form in Q2.2 than that of the WT Q2 RNA, giving a minimum free energy of  $-3.80 \text{ kcal mol}^{-1}$  (figure 3.33), in contrast to Q2 which had a free energy of  $-2.30 \text{ kcal mol}^{-1}$ . Furthermore, as this mutant was previously shown to be insensitive to GQC-05 in splicing assays, we chose to use this RNA as our negative control (Weldon et al., 2018).

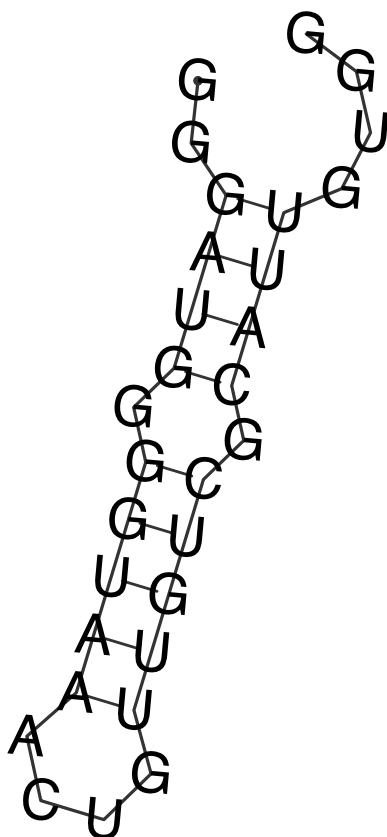
	Position	Length	QGRS	G-Score
Q2.2	1	30	GGGAUGGGGUAAACUGUUGUCGCAUUGUGG	2
	2	29	GGAUGGGUAAACUGUUGUCGCAUUGUGG	2

	Position	Length	QGRS	G-Score
Q2	1	30	GGGAUGGGGUAAACUGGGGUCGCAUUGUGG	17

**Figure 3.32 QGRS mapper comparison of Q2 and Q2.2 RNA.**

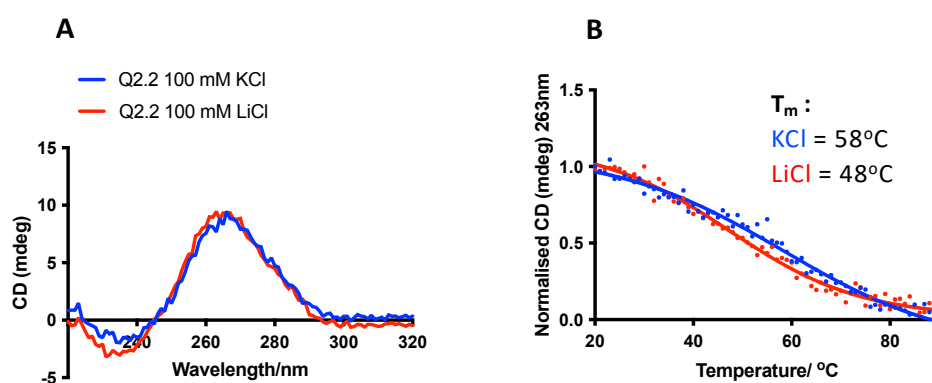
Showing the output from QGRS mapper for Q2.2 (upper table) and the best Q2 hit (lower table), in which Q2.2 shows a relatively lower G-score compared to the wild-type Q2 sequence.



**Figure 3.33: RNA-fold of Q2.2**

RNA-fold output of a putative Q2.2 duplex occurring with a minimum free energy of  $-3.80 \text{ kcal mol}^{-1}$ .

Unlike the wild-type Q2 sequence, there was no changes in the intensity of the CD spectrum between potassium and lithium (figure 3.34 A), and there is little change in the stability of the structure from the CD melting, with  $T_m$  values in Li<sup>+</sup> and K<sup>+</sup> of 48 and 58 °C respectively (figure 3.34 B), suggesting that this mutant has a lower G4 potential. However, the 10°C increase in stability in the presence of KCl may arise due to the formation of potential intermolecular G4 structures in the presence of KCl, though this will require further experimental evidence to confirm such hypothesis. Altogether, these results indicate that the Q2 RNA likely adopts a parallel quadruplex conformation.



**Figure 3.34 CD spectroscopy of the Q2.2 fragment**

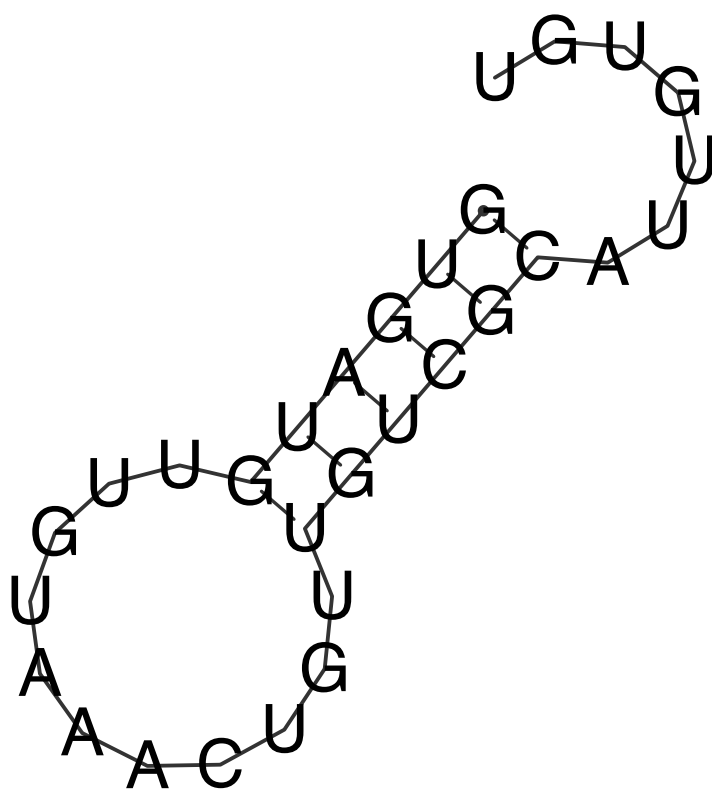
(A) CD spectra of Q2.2 RNA at 20°C in the presence of 100 mM KCl (blue) and LiCl (red) in 10 mM TBA phosphate buffer pH 7.0. (B) CD thermal melting of Q2.2 RNA in the presence of 100 mM KCl (blue) and LiCl (red) in the same buffer as (A).

### 3.3.2 CD and UV characterisation of the Q2 G4 in HeLa nuclear extract

One major drawback of biophysical techniques such as CD and UV spectroscopy, is the fact that they are conducted with the RNA in isolation in a highly purified *in vitro* system, meaning that its chemical environment would be quite different to what is present inside a cell. It is therefore difficult to prove that the presence of RNA secondary structures *in vitro* also match those present *in vivo*, where the binding of many RNA binding proteins have the ability to alter and compete with RNA secondary structure elements. For example, the splicing factor hnRNP F has been shown previously to bind to Q2 RNA and can compete with both G4 and duplex RNA structures (Dominguez et al., 2010). Because the splicing assays of Bcl-x were

conducted in the presence of commercial HeLa nuclear extract (NE), we tested whether Q2 can still adopt a G4 structure in conditions that support *in vitro* splicing.

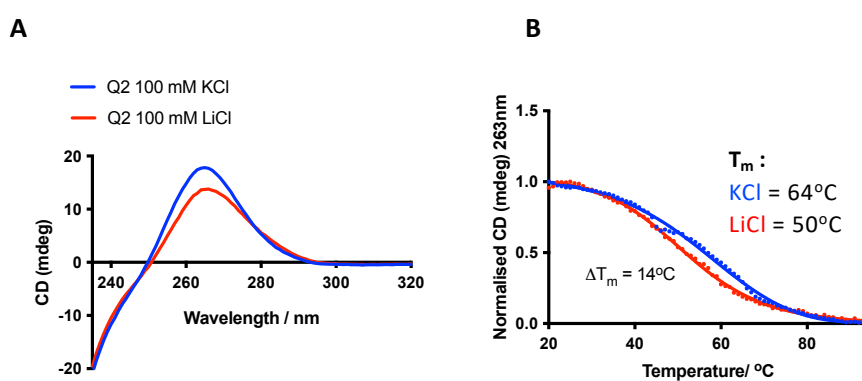
To test this, we compared the CD and UV profiles of Q2 RNA with an RNA that has an even lower G4 potential than Q2.2 (termed  $\Delta$ G4 RNA in table 3.1). As described above, the Q2.2 RNA showed a small increase in stability in the presence of KCl compared with LiCl, which may be the result of the formation of some bi-molecular G4 topologies. Therefore, the  $\Delta$ G4 RNA was generated which harbours mutations in all of its G-tracts, thereby theoretically abolishing the possibility of both inter and intramolecular G4 structures, which is reflected in the G-score of 0 from QGRS mapper. Additionally, the RNA fold output shows the minimum free energy for the  $\Delta$ G4 duplex of  $-3.50 \text{ kcal mol}^{-1}$  (figure 3.35), similar to that of Q2.2. However, as there is a lower propensity to form a competing G4 structure, a duplex is more likely to be the dominant structural isoform.



**Figure 3.35: RNA-fold of  $\Delta$ G4**

RNA-fold output of a putative  $\Delta$ G4 duplex occurring with a minimum free energy of  $-3.50 \text{ kcal mol}^{-1}$ .

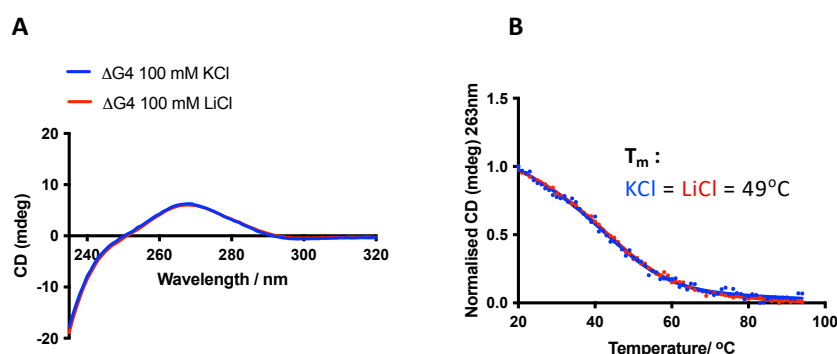
The CD profiles of Q2 and  $\Delta$ G4 RNAs in the presence of 5% HeLa nuclear extract (NE) are shown in figures 3.36 and 3.37 respectively. Analysis of the CD spectrum for Q2 at 20°C shows a greater intensity of the signal in the presence of 100 mM KCl compared with LiCl (with a difference of approximately 4 millidegrees), indicating greater  $\pi$ - $\pi$  stacking, consistent with the formation of G-tetrads as observed in the absence of NE (figure 3.36 A). A similar difference in the CD intensity between 100 mM KCl and LiCl-containing buffers was observed for the G4 sequence d(GGGA)<sub>5</sub> (Kejnovska et al., 2003). Additionally, the Q2 RNA showed a K<sup>+</sup> dependant increase in the thermal stability, with  $T_m$  values of 64 and 50°C in KCl and LiCl respectively (figure 3.36 B).



**Figure 3.36 CD data of Q2 RNA in the presence of 5% nuclear extract.**

(A) CD data at 20°C of Q2 RNA in the presence of 100 mM KCl (blue) or LiCl (red). (B) CD thermal melts in the same conditions as in (A). Experiments were conducted in 10 mM TBA phosphate buffer pH 7.0 in a 1 mm pathlength quartz cuvette.

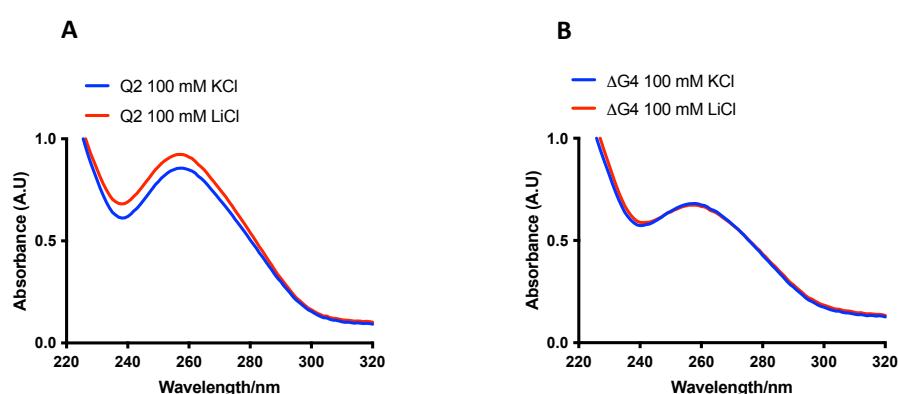
In contrast, the  $\Delta$ G4 RNA showed no K<sup>+</sup> dependence in the CD intensity, and no change in the thermal stability of the RNA, with a  $T_m$  of 49°C in both cases (figure 3.37 A and B), suggesting that the Q2 but not the  $\Delta$ G4 RNA sequence can fold a G4 in the presence of NE.



**Figure 3.37 CD data of  $\Delta G4$  RNA in the presence of 5% nuclear extract.**

(A) CD data at 20°C of  $\Delta G4$  RNA in the presence of 100 mM KCl (blue) or LiCl (red). (B) CD thermal melts in the same conditions as in (A). Experiments were conducted in 10 mM TBA phosphate buffer pH 7.0 in a 1 mm pathlength quartz cuvette.

This observation was further supported by analysing the UV profile of the two RNAs at 20 °C, in which the Q2 RNA (figure 3.38 A) showed a greater intensity at 260 nm in the presence of Li<sup>+</sup> compared to K<sup>+</sup>, consistent with a greater folding status of the RNA in the presence of K<sup>+</sup>, whereas no change was observed for the  $\Delta G4$  RNA (figure 3.38 B), indicating the formation of a G4 in the Q2 sequence.



**Figure 3.38 UV spectrum comparison of the Q2 and  $\Delta G4$  RNAs in 5% HeLa NE.**

(A and B) UV data at 20°C of Q2 (A) and  $\Delta G4$  (B) RNA in the presence of 100 mM KCl (blue) or LiCl (red). Experiments were conducted in 10mM TBA phosphate buffer pH 7.0 in a 1 mm pathlength quartz cuvette.

Interestingly, in the presence of nuclear extract, Q2 showed a reduced stability in KCl compared to Q2 without nuclear extract (figure 3.31 B). This could suggest that RNA binding proteins present in the nuclear extract may compete with the secondary structure formed by the Q2 sequence. Additionally, the melting profile of Q2 RNA observed in the presence of KCl and nuclear extract displays multiple transition points suggesting that in NE, the G4 present in Q2 may be polymorphic. This would be consistent with previous in cell NMR experiments of the human telomere G4 DNA discussed in section 1.5.2.3.1, in which those G4s were more polymorphic when introduced in oocytes and oocyte lysates (Hänsel et al., 2009). It is important to note that in each of the cases described above, the ionic strength is likely to be different when comparing samples with and without nuclear extract, as nuclear extract contains additional amounts of salts (such as 75 mM KCl and 350 mM NaCl), meaning that the ionic strength in the final mix with nuclear extract will be greater than samples without the extract.

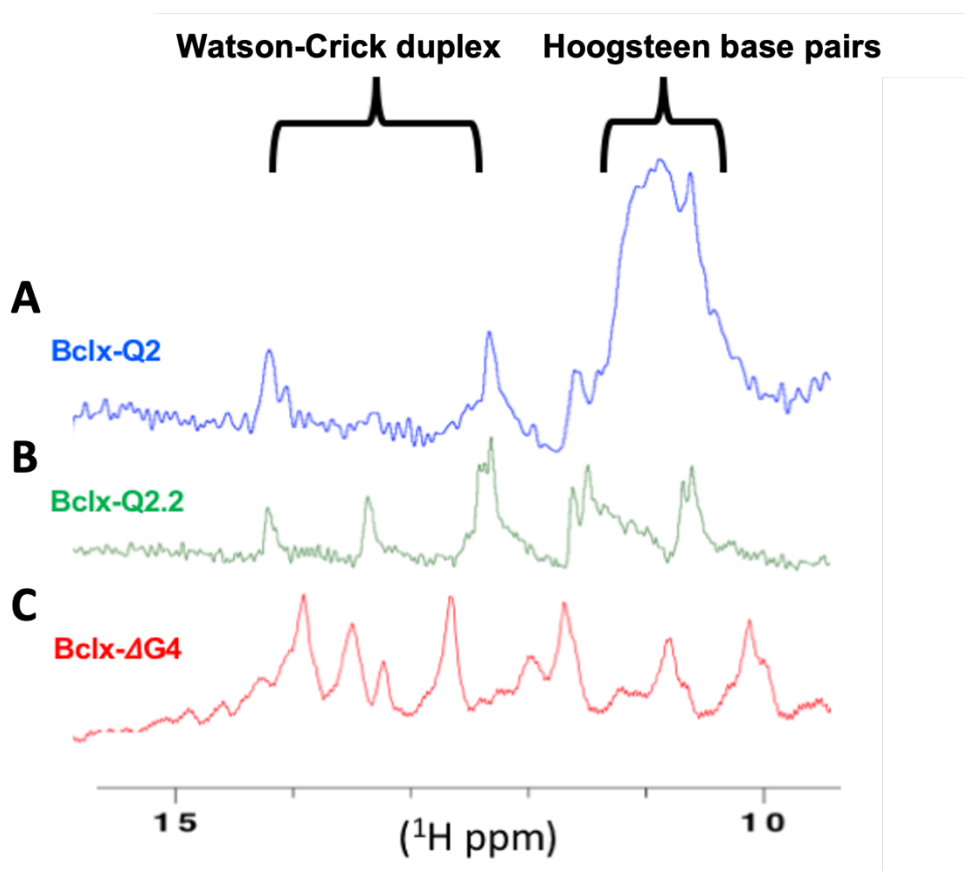
### **3.4 Using NMR spectroscopy to characterise the Q2 structure**

NMR spectroscopy can also be used to characterise G-quadruplexes. Imino protons of guanines involved in a quadruplex structure typically display chemical shifts between 10-12ppm, whereas Watson-Crick base pairings give rise to additional peaks between 13-15ppm.

#### **3.4.1 NMR spectroscopy on the Q2 and mutant sequences**

The 1D NMR spectrum obtained with Q2 RNA in 100 mM KCl is displayed in figure 3.41 A. This sequence displays broad peaks between 10-12 ppm, suggesting a population of Hoogsteen hydrogen bonding, likely coming from a G4, but also two additional peaks at 12.5 and 14ppm. Between 10-12 ppm, the individual peaks are not well defined, suggesting conformational heterogeneity of the Q2 G4 in these conditions. This could also suggest formation of higher order structures, as larger complexes consisting of dimers or tetramers may reduce the tumbling of the structure in solution, giving broader peaks. The peaks between 12.5-15 ppm could arise from Watson-Crick base pairings in the connecting loops, as was the case for the Bcl-2 promotor quadruplex (Dai et al., 2006). Alternatively, a certain population of the Q2 RNA may adopt a duplex conformation instead of a quadruplex (such as that predicted by RNA-fold in figure 3.22) . For comparison, we also measured NMR 1D spectra of the mutants Q2.2 and  $\Delta$ G4 RNAs at the same RNA concentration and in the same

buffer (figures 3.41 B and C). The NMR spectra of these RNAs display peaks between 10-15ppm, with more peaks in the 13-15ppm range and with the peaks between 10-12 ppm having lower intensities than the wild-type Q2 RNA, suggesting that these mutants favour a duplex conformation. However, as NMR spectroscopy requires nearly 100 times the concentration of RNA compared with CD or UV spectroscopy, the potential for Q2.2 to form bi or tetra-molecular G4 structures is more likely and cannot be excluded, meaning that peaks observed in the 10-12 ppm region with this RNA may form due to these intermolecular structures. However, with the  $\Delta$ G4 RNA sequence, some of these peaks remain, which could arise as a result of GU non-canonical wobble base pairings giving rise to peaks in the G4 region. Therefore, combining the results of the NMR, CD and UV experiments, it is possible that the  $\Delta$ G4 and Q2.2 RNA are folding into duplex structures (such as those predicted by RNA-fold in figures 3.33 and 3.35) and Q2 is a multi-conformational quadruplex, though the potential for the mutants (particularly Q2.2) to form G4 structures cannot be excluded entirely.



**Figure 3.41 1D NMR spectroscopy of WT and mutant fragments.**

$^1\text{H}$  NMR (600 MHz, 10°C) 6 mM  $\text{K}_2\text{HPO}_4$ , 4 mM  $\text{KH}_2\text{PO}_4$ , 15 mM KCl, pH=7.0 +100 mM KCl of Bclx-Q2 (A), Bclx-Q2.2 (B) and Bcl-x- $\Delta$ G4 (C).



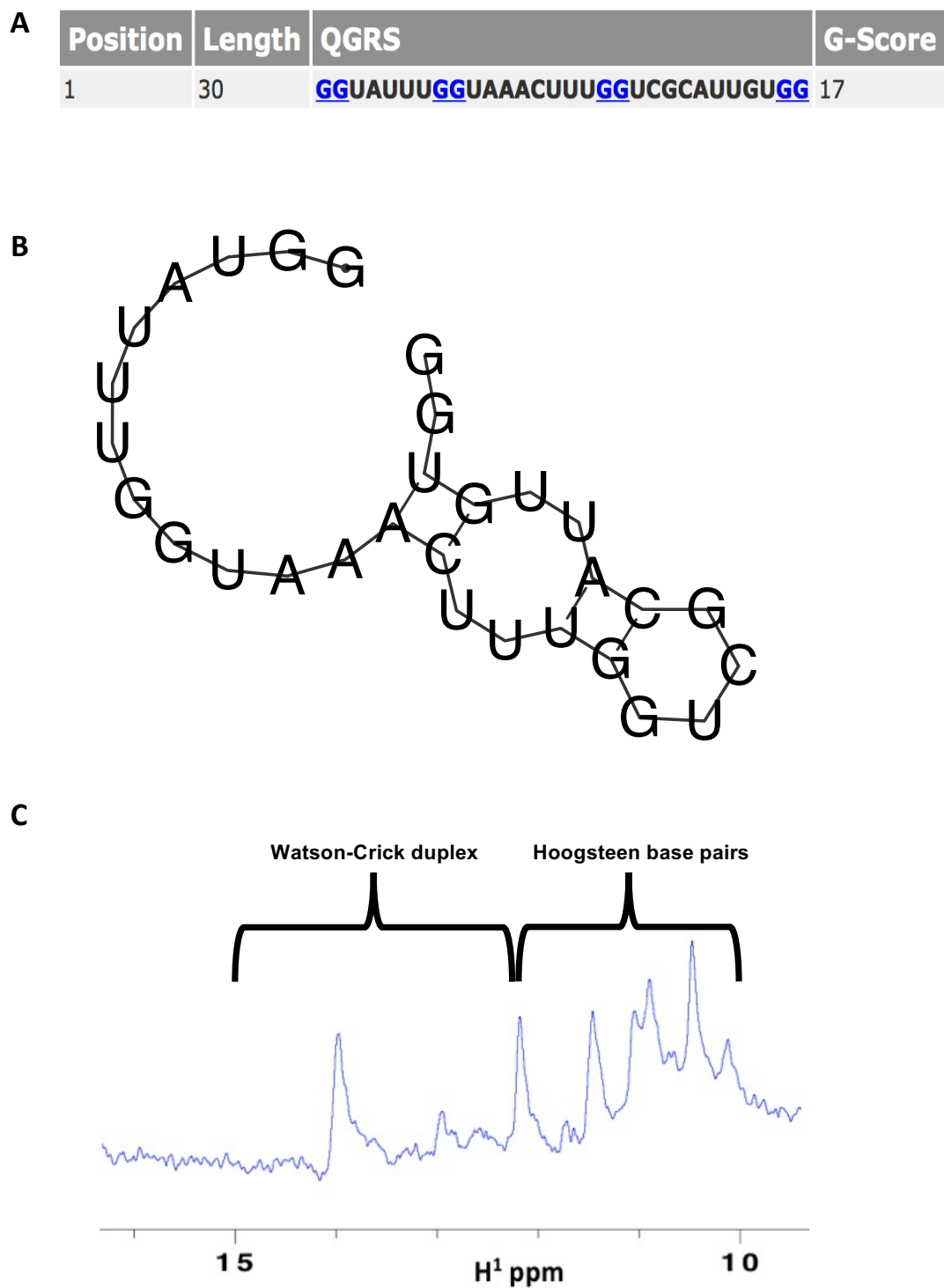
The sample heterogeneity observed with the Q2 RNA NMR makes structure determination difficult, as NMR assignments are almost impossible when there is a large spectral overlap. It also excludes the possibility of further 2D NMR characterisations. Therefore, solving the structure of this RNA by NMR requires further optimisation of the buffer and the RNA sequence in order to obtain sharp and well-defined peaks that would be easier to assign.

### **3.5 Approaches to reduce the structural heterogeneity of Q2**

#### **3.5.1 Reducing sample heterogeneity by optimisation of the Q2 sequence**

As discussed in section 3.2, the large spectral overlap in the NMR spectrum could be due to conformational heterogeneity of the Q2 sequence caused by excess guanines in the first, second and third G-tracts. Therefore, in order to isolate a single G4 species in solution, we made mutations in the first, second and third tracts to the guanines that were not expected to participate in the G4 tetrad predicted by QGRS mapper. In a similar way, this approach was also applied to the c-myc promotor G4, where deleting excess guanines resulted in the stabilisation of a predominant species in solution, enabling structure determination (figure 1.5.2.1) (Ambrus et al., 2005).

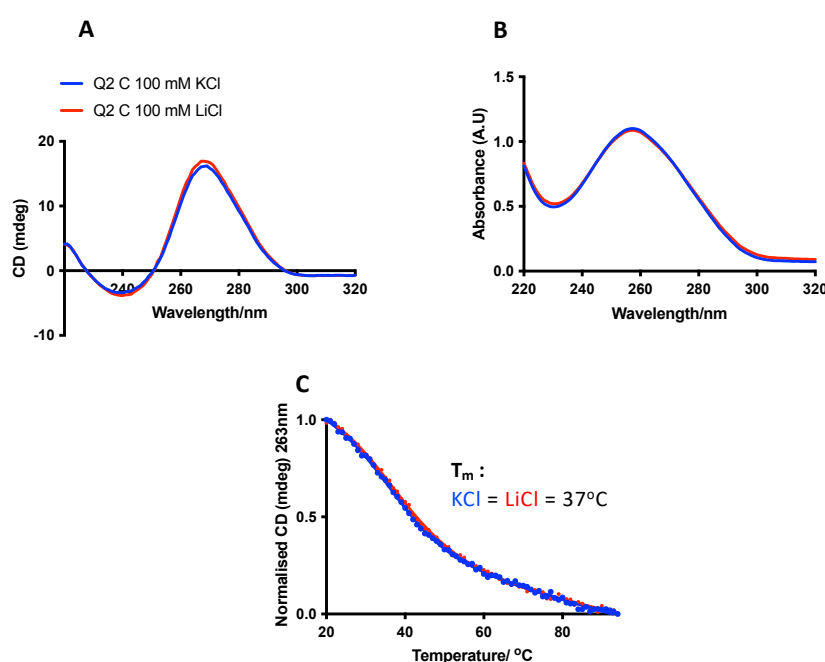
The RNA Q2 canonical (termed Q2 C in table 3.1) was synthesised by mutating out the guanines not involved in the G4 fold in the Q2 WT sequence (predicted by QGRS mapper) to uracil, which enabled the isolation of a single G4 species with the highest G-score of 17 (figure 3.51 A). Furthermore, the RNA fold prediction showed the formation of a duplex with a free energy of  $-1.40 \text{ kcal mol}^{-1}$ , meaning that this sequence has a lower propensity to fold a duplex than the WT Q2 sequence (figure 3.51 B). The NMR spectrum showed better defined peaks in the 10-12ppm region, but also displayed several peaks in the 13-15ppm region which suggests that Watson-Crick base pairing is also occurring (figure 3.51 C).



**Figure 3.51 QGRS, RNA-fold output and NMR characterisation of Q2 C.**

(A) Output from QGRS mapper for the Q2 consensus RNA, showing a single species. (B) RNA-fold output of Q2 C which has a minimum free energy of  $-1.40 \text{ kcal mol}^{-1}$ . (C)  $^1\text{H}$  NMR (600 MHz,  $10^\circ\text{C}$ ) 6 mM  $\text{K}_2\text{HPO}_4$ , 4 mM  $\text{KH}_2\text{PO}_4$ , 15 mM KCl, pH=7.0 +100 mM KCl of Bclx-Q2 consensus.

However, unlike the wild-type sequence, the CD analysis showed no change in the intensity of the CD signal at 20 degrees (figure 3.52 A), no change in the UV profile at 20 degrees (figure 3.52 B), and also no change in the melting profile (figure 3.52 C) in buffers containing either KCl or LiCl. This suggests that this sequence may not fold into a quadruplex, and hence the quadruplex seen in the wild-type sequence may not correspond to the one predicted by QGRS mapper. However, as the NMR spectrum shows that Hoogsteen hydrogen bonding is still occurring, we cannot completely rule out the possibility of a G4 structure in this sequence. However, despite not being predicted in the RNA-fold structure, peaks in the 10-12 ppm region may also arise as a result of non-canonical GU base pairings as mentioned previously, which would explain why we do not see changes in the CD profiles when comparing the K<sup>+</sup> and Li<sup>+</sup> - containing buffers. Perhaps a more stable 3 tiered (instead of 2-tiered) non-canonical fold is more likely in Q2, with the uridine in the GUGG stretch at the end of the Q2 sequence forming a bulge, thus incorporating the 5' guanine in the G4. Therefore, the G4 peak broadening we see in the wild-type NMR spectrum may not be entirely due to the excess number of guanines alone, though this may also have some contribution.



**Figure 3.52 CD characterisation of Q2 C.**

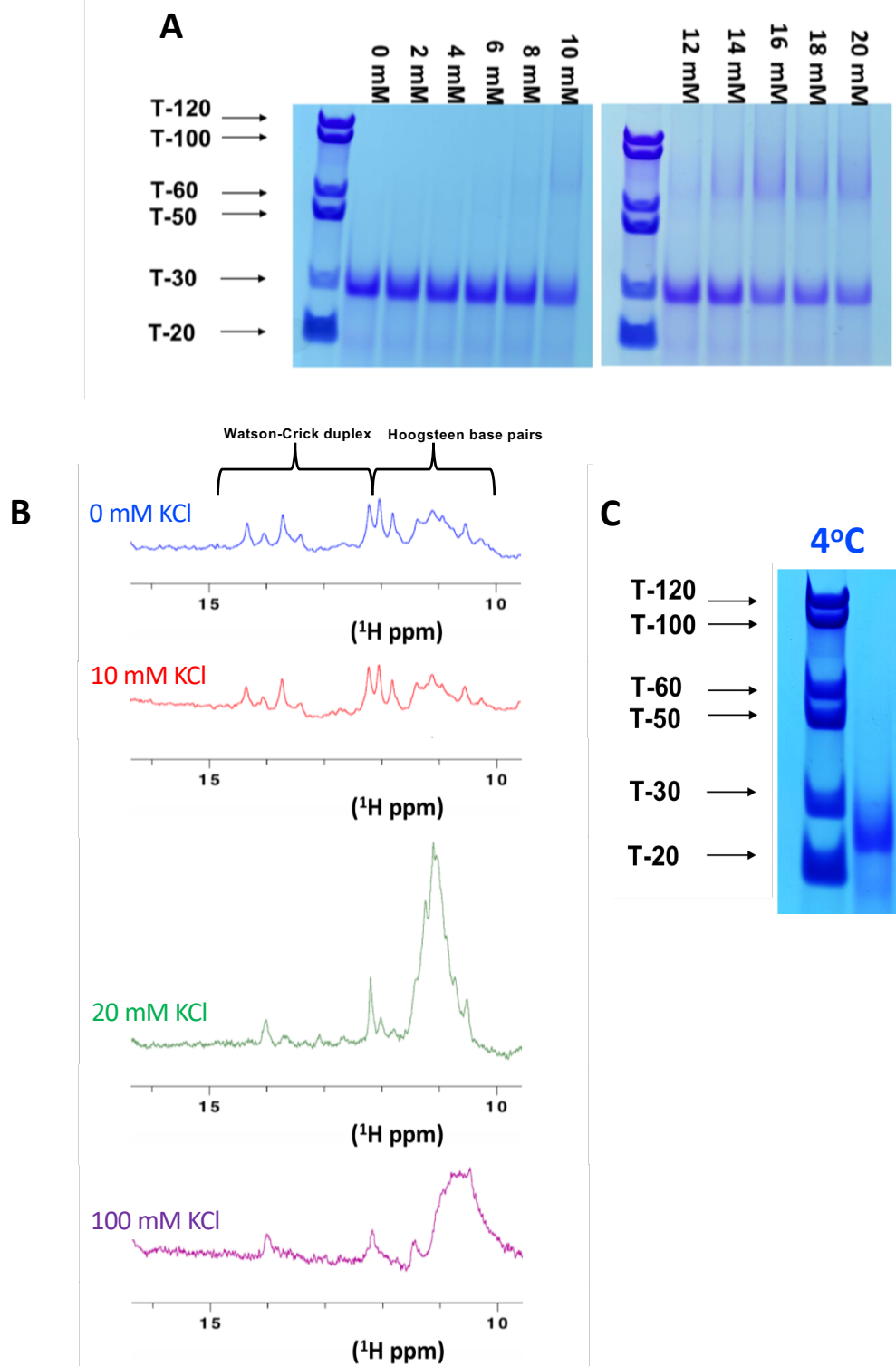
(A and B) CD (A) and UV (B) spectrum of the Q2 consensus RNA in 100 mM KCl (blue) and LiCl (red) in 10 mM TBA phosphate buffer pH 7.0. (C) CD thermal melt in the same buffer conditions as (A and B). Experiments were conducted in a 1 cm pathlength quartz cuvette.

### 3.5.2 The effect of K<sup>+</sup> concentration in the structural heterogeneity of Q2

Another explanation for the imino peak broadening of the wild-type RNA may be due to the formation of higher order structures such as G4 dimers or tetramers. In addition to it not being able to predict non-canonical folds, another limitation of QGRS mapper is that it can only predict the formation of unimolecular quadruplexes, whereas bimolecular and tetramolecular structures are highly probable at high concentrations of RNA. Particularly, G4 sequences that start with a 5' guanine have a greater propensity to form stacked dimers than those with other flanking bases. For example, the human TERRA sequence consisting of r(GGGUUAGGGU) was shown to form a bi-molecular stacked dimer in the presence of 70mM KCl (Martadinata and Phan, 2013). Salt concentration is an important parameter for dimerization since extra K<sup>+</sup> ions are required to coordinate the tetrads at the interface of the two unimolecular or bimolecular quadruplexes, and high salt concentration has been shown to induce stacked dimers in some RNA sequences (Binas et al., 2020; Kolesnikova et al., 2017). To understand the strand stoichiometry of the 30 mer Q2 RNA, native PAGE analysis was conducted in the presence of increasing concentrations of KCl (figure 3.53 A). At 0 mM KCl, there is a single predominant species in solution that migrates at a position similar to a 30-nucleotide poly-T (T-30). At 10 mM KCl, higher order structures start to be observed, which migrate around T-60 suggesting a dimeric complex. These higher order structures become more apparent between 14-20 mM KCl, with greater smearing occurring near the T-100 and T-120 mark. Therefore, the formation of these higher order structures in addition to the monomeric species may lead to large spectral overlaps and peak broadening in the NMR spectrum, as there are multiple species in solution as well as reduced tumbling of the G4 multimers. In order to test this hypothesis, KCl titrations were conducted in NMR experiments to see any potential changes to the spectrum (figure 3.53 B). In the absence of KCl (Tris buffer only), the imino peaks appear sharp and well defined. However, at 20 mM KCl, spectral broadening occurs and begin to recapitulate the effects seen with 100 mM KCl. Therefore, this suggests that the formation of multimeric species may be contributing to sample heterogeneity seen in the NMR experiments.

The next question to address is the folded species observed in the absence of salt, as imino signals are still present at 0 mM KCl, suggesting that a structure is still present under these conditions. In addition, when running Q2 on a native gel at 4°C, a band appears which

migrates lower than T-30, indicating a folded more compact structure (figure 3.53 C). Therefore, in order to determine if the folded species in the absence of external salt is a G4, we used a G4 specific fluorescent probe (NMM), which has a greater enhancement in its fluorescence emission when bound to G4 over non-G4 structures (Nicoludis et al., 2012) (see next section).



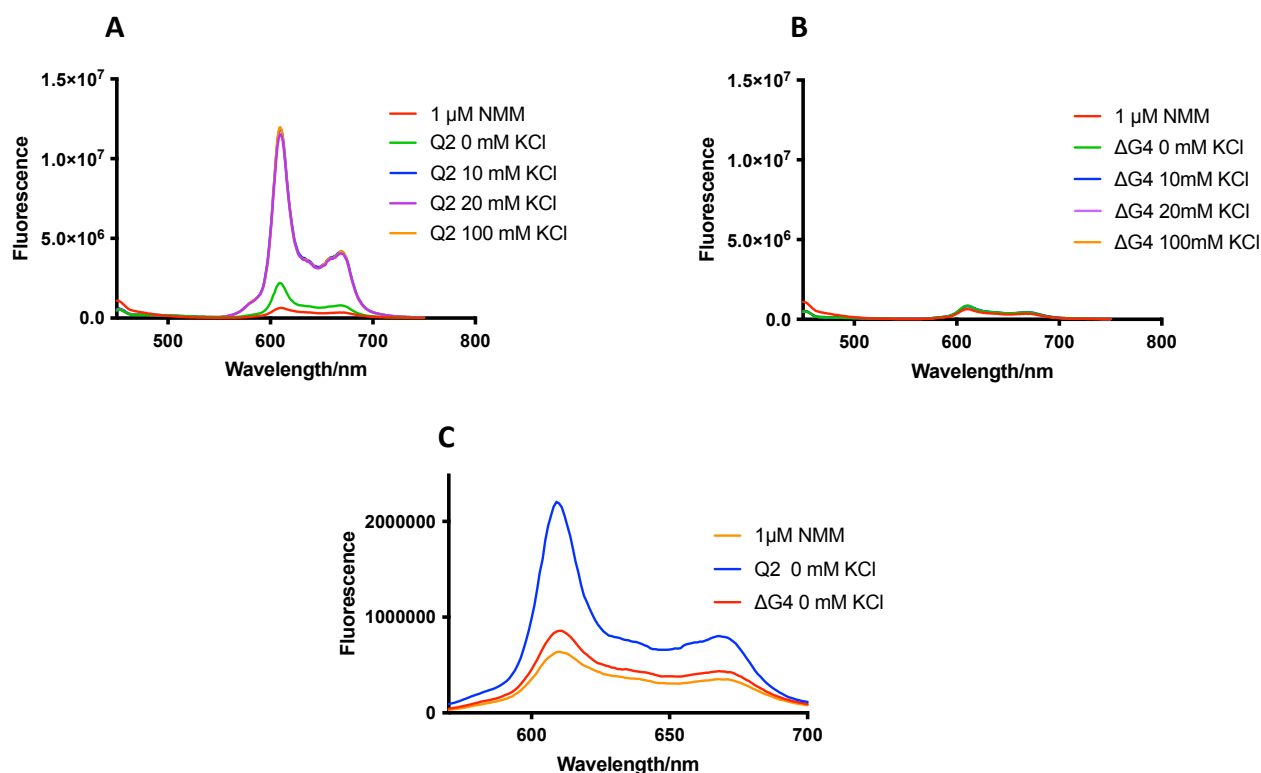
**Figure 3.53 Native gel and NMR analysis with K<sup>+</sup> titrations.**

(A) 12% Native PAGE with Q2 wild-type RNA in the presence of Tris pH 7.0, with increasing concentrations of KCl (0–20 mM) compared to Poly T standards. Gel was run at 20°C. (B)  $^1\text{H}$  NMR (600 MHz, 10°C) of the wild-type Q2 RNA in 10 mM Tris pH 7.0, with KCl titrations (0, 10, 20 and 100 mM). (C) Native gel of Q2 in Tris buffer only run at 4°C.

## 3.6 Characterisation of the G4 structure in the absence of salt

### 3.6.1 Characterisation of a G4 structure using NMM fluorescence

To probe for a putative G4 in Q2 in the absence of salt, we used the fluorescence turn on assay approach (described in section 1.5.2.4), where we utilised the fluorescence properties of a highly specific parallel G4 binder N-methyl-mesoporphyrin IX (NMM) (del Villar-Guerra et al., 2018a; Kreig et al., 2015; Sabharwal et al., 2014). NMM has a relatively low fluorescence when unbound, but when bound to parallel G4 DNA, a 50-60 fold fluorescent enhancement is observed (Nicoludis et al., 2012a). Figures 3.61 A and B shows the emission profile for NMM with the wild-type Q2 RNA and the  $\Delta$ G4 mutant, with the same salt conditions that were tested in the NMR experiment. In the absence of RNA, NMM displayed relatively low fluorescence. For the wild-type Q2 sequence, in the presence of Tris buffer, there is a small increase, and with 10, 20, 100 mM KCl, there is an approximately 4-fold increase in the fluorescence emission, suggesting that Q2 forms a G4 even at low concentrations of KCl. For the  $\Delta$ G4 mutant, there is only a small increase in the emission profile relative to NMM alone, with negligible changes in the presence of KCl, further indicating the presence of a quadruplex in the Q2 RNA and the absence of a G4 in the  $\Delta$ G4 mutant. A comparison of the emission profiles for the Q2 wild-type (WT) and  $\Delta$ G4 mutant in the presence of Tris only (figure 3.61 C), shows a 2-fold increase in the fluorescence emission for the WT as compared to the mutant, suggesting that even in the absence of KCl a quadruplex is present. However, it is important to note that these fold changes are very modest when comparing to the 50 to 60 fold enhancement reported in the literature (Nicoludis et al., 2012a). These differences may arise as a result of the present study being conducted on RNA rather than DNA G4s, where these large enhancements were observed. To identify if the reduced fluorescence enhancement in the presence of NMM is due to RNA rather than DNA G4 structures, a well characterised G4 RNA, such as TERRA, should be tested.



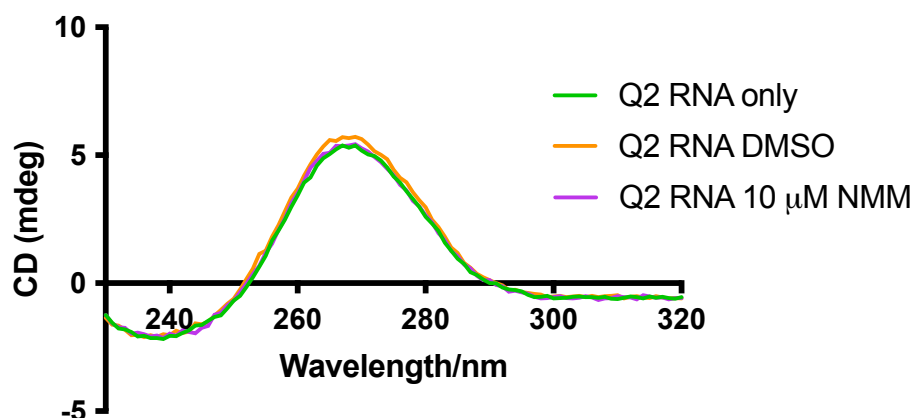
**Figure 3.61 Fluorescence analysis of Q2 RNA in the presence of NMM ( $\lambda_{exc} = 393\text{nm}$ ).**

(A and B) Fluorescence spectra in the presence of 1  $\mu\text{M}$  NMM with 10  $\mu\text{M}$  Q2 (A) and  $\Delta\text{G4}$  (B) RNA. (C) Comparison of the emission spectra for Q2 and  $\Delta\text{G4}$  RNA in the absence of KCl. Experiments were conducted in 10 mM Tris, pH 7.0.

### 3.6.2 CD spectroscopy in the presence of NMM

However, as discussed in section 1.5.2.4, a major disadvantage of using the fluorescence turn on assay approach is that NMM may artificially induce and stabilise a G4 that would otherwise not have been present. To eliminate the possibility of an induced fit over a conformational selection mechanism of NMM binding, CD spectra were recorded in the presence and absence of 10  $\mu\text{M}$  NMM (figure 3.62). In both cases, the CD spectra are overlapping, suggesting that addition of the ligand did not cause a conformational change of the RNA, meaning that the WT RNA adopts a G4 conformation in the absence of the ligand and that the ligand binds the G4 through a conformational selection mechanism. Altogether, these results show that even in the absence of KCl, the Q2 sequence can adopt a monomeric G4 conformation.





**Figure 3.62 CD spectrum of Q2 RNA in the presence of NMM.**

CD spectrum at 20°C of Q2 RNA (1  $\mu$ M) in the presence and absence of NMM (10  $\mu$ M) and DMSO.

### 3.7 X-ray crystallography of Q2

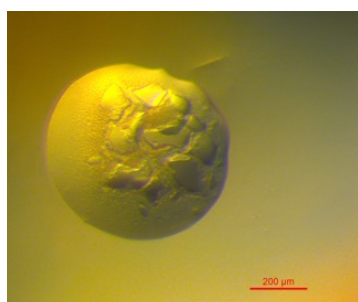
Knowing that Q2 is able to fold a stable quadruplex even in the absence of KCl, we proceeded with crystallisation trials of Q2, using a specialised screen developed for nucleic acids called HELIX. This particular screen was developed by the department of pharmacy at UCL and is based on the conditions of around 1450 nucleic acid structures deposited in the PDB, enabling the crystallisation of a wide variety of nucleic acid structures such as duplex, triplex, i-motif and quadruplexes (Viladoms and Parkinson, 2014)

Figure 3.71 below shows crystal formation in two of the conditions tested with the Q2 RNA (see section 2.7 for crystallisation conditions and sample preparation). Crystal formation is visible in the upper two wells containing the RNA, with smaller microcrystals being observed in the lower wells containing buffer only, suggesting that formation of the larger crystals is coming from the RNA. The RNA itself was annealed in the absence of salt in its monomeric state (see section 2.1.2 for the annealing protocol used and section 2.7 for the annealing conditions), though after the addition to the screen, the presence of 50 mM KCl in the screen condition A1 may lead to stacking of the G tetrads in Q2 which can aid the crystallisation process. The presence of PEG as a cryoprotectant has also been implicated in its ability to favour the parallel type G4 structure over other G4 topologies (Buscaglia et al., 2013). Furthermore, the presence polyamines, such as spermine, has been shown to interact with RNA and DNA structures, resulting in nucleic acid condensation, helping the crystallisation process (Katz et al., 2017). In condition C1, instead of potassium, the ion present is

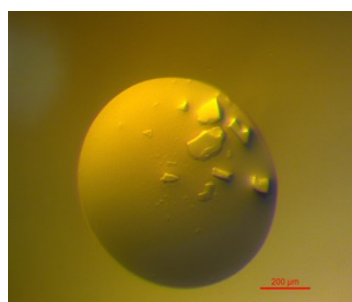
ammonium, which has also been showed to stabilise a G4 present in the human telomeric sequence (Hud et al., 1999). Therefore, as these conditions are favourable for G4 formation, it is possible that these crystals may correspond to the Q2 RNA in a G-quadruplex conformation. However, difficulties in accessing synchrotron time at the Diamond light source due to the global pandemic has put this experiment on hold.

RNA + Buffer

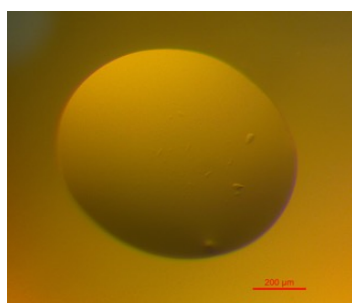
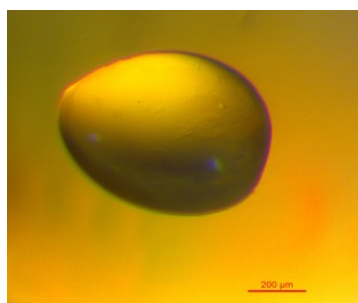
A1



C1



Buffer only



A1	C1
0.05 M KCl 0.1 M LiCl	0.1 M Ammonium sulfate
0.012 M Spermine tetrahydrochloride	0.005 M Bis-Tris pH 7.0
0.05 M MES pH 6.5	44% (v/v) PEG 200
25% (v/v) PEG 400	

**Figure 3.71 Crystallisation trials of Q2 in the Helix screen.**

Crystals obtained in the helix screen of the RNA (top two quadrants), with the negative control buffer only samples in the lower two quadrants. Also shown in the table are the two conditions of A1 and C1 within the HELIX screen in which the crystals were obtained.

### 3.8 Discussion

We have shown here by biophysical characterisations that the G-rich region just downstream of the X<sub>s</sub> splice site can fold into a G-quadruplex, even in the absence of potassium ions. Much of the sample heterogeneity observed in the NMR spectrum likely arose through the stacked multimerization of external tetrads as suggested from the native gel analysis in figure 3.53 A. Lower concentrations of salt reduced the effect of this multimerization, resulting in sharper better defined peaks in the NMR spectrum. However, it is important to note that sample heterogeneity is still observed even in the absence of salt, as both duplex (13-15ppm) and quadruplex (10-12ppm) signals are still observed, as well as some peak overlaps in the 10-12ppm region. The overlapping peaks in the G4 region may arise due to an excess number of guanines in the first three G-tracts of Q2. We showed here that simply mutating out the excess guanines to form the canonical G4 (as predicted by QGRS mapper) did not give rise to a G4 structure according with the CD data. Therefore, the potential to form a non-canonical G4 structure may be the more likely conformation adopted by Q2 RNA, especially considering that two tiered quadruplexes are not as stable as three or four stacked structures due to a reduced number of  $\pi$ - $\pi$  stacking interactions of the tetrads. Furthermore, computational analysis using other bioinformatic software that can also predict non-canonical G4 folds, such as G4 hunter, also implied a non-canonical G4 topology of Q2 (see section 1.5.1). One possibility is that the uridine in the GUGG stretch is forming a bulge, meaning that the 5'G in this sequence may be incorporated into the tetrad itself, resulting in a three stacked quartet. Therefore the following Q2 sequence with the highlighted Gs may be participating in G4 formation: **GGG**AUG**GGG**UAAACUG**GGG**UCGCAUU**GUGG**. A similar observation was made for the KRAS promotor G4, in which a guanine tract interrupted with a single nucleotide was still incorporated into the tetrad, resulting in a nucleotide bulge (Kerkour et al., 2017). Therefore, in a future study, studying a sequence which removes this bulge as well as mutating out the excess guanines in the second and third runs (which contain GGGG) such that only a three stack G4 has the potential to form, may help to isolate the single major species of the Q2 RNA.

The observation of imino peaks in the 13-15ppm region in the absence of KCl, which reduced in the presence of increasing [KCl] (figure 3.53 B), starts to open up the possibility of hairpin formation (figure 3.22), implying that in these conditions we have both quadruplex and

hairpin populations. This is likely to arise as a result of a lower stability of the G4 at low KCl concentrations, which results in a shift in the equilibrium to a different structure (such as a duplex), which will make it problematic for structural determination. Increasing the KCl concentration does solve this issue, but on the other hand, induces the dimerization or multimerization of the G4 causing spectral overlap. Therefore, to get around this two-way problem, further modifications could be made to the Q2 sequence to prevent dimerization at high KCl concentrations. Studies have shown that sequences starting with a 5'G often result in the formation of G4 5'-5' stacking interactions, due to the formation of an external tetrad at the 5' interface (Binas et al., 2020; Kogut et al., 2019; Kolesnikova et al., 2017; Martadinata and Phan, 2013 ). Though this stacking may favour the crystallisation process, it is not suitable for NMR studies as it leads to significant spectral overlap. Consequently, as well as removing the bulge and mutating out the excess guanines, placing an additional nucleotide to the 5' end of the sequence may enable the isolation of a monomeric homogenous G4 at high KCl concentrations. However, making mutations to the natural sequence in this way will need to be validated against the wild-type sequence such that you are indeed isolating the major species and not completely changing the structure of the RNA. In addition, the functional relevance of such a mutated sequence should also be tested along with the binding specificity and activity of the compound in question, GQC-05, by *in vitro* splicing assays.

## **CHAPTER 4: Binding mode and specificity of GQC-05 for the G4 region (Q2) in Bcl-x**

- 4.1 Introduction**
- 4.2 Fluorescence spectroscopy to study the structural specificity of GQC-05**
- 4.3 Native gel analysis to probe conformational changes of Q2 in the presence of GQC-05**
- 4.4 CD spectroscopy to study the binding mode of GQC-05**
- 4.5 NMR spectroscopy to determine the structural specificity of GQC-05**
- 4.6 Discussion**

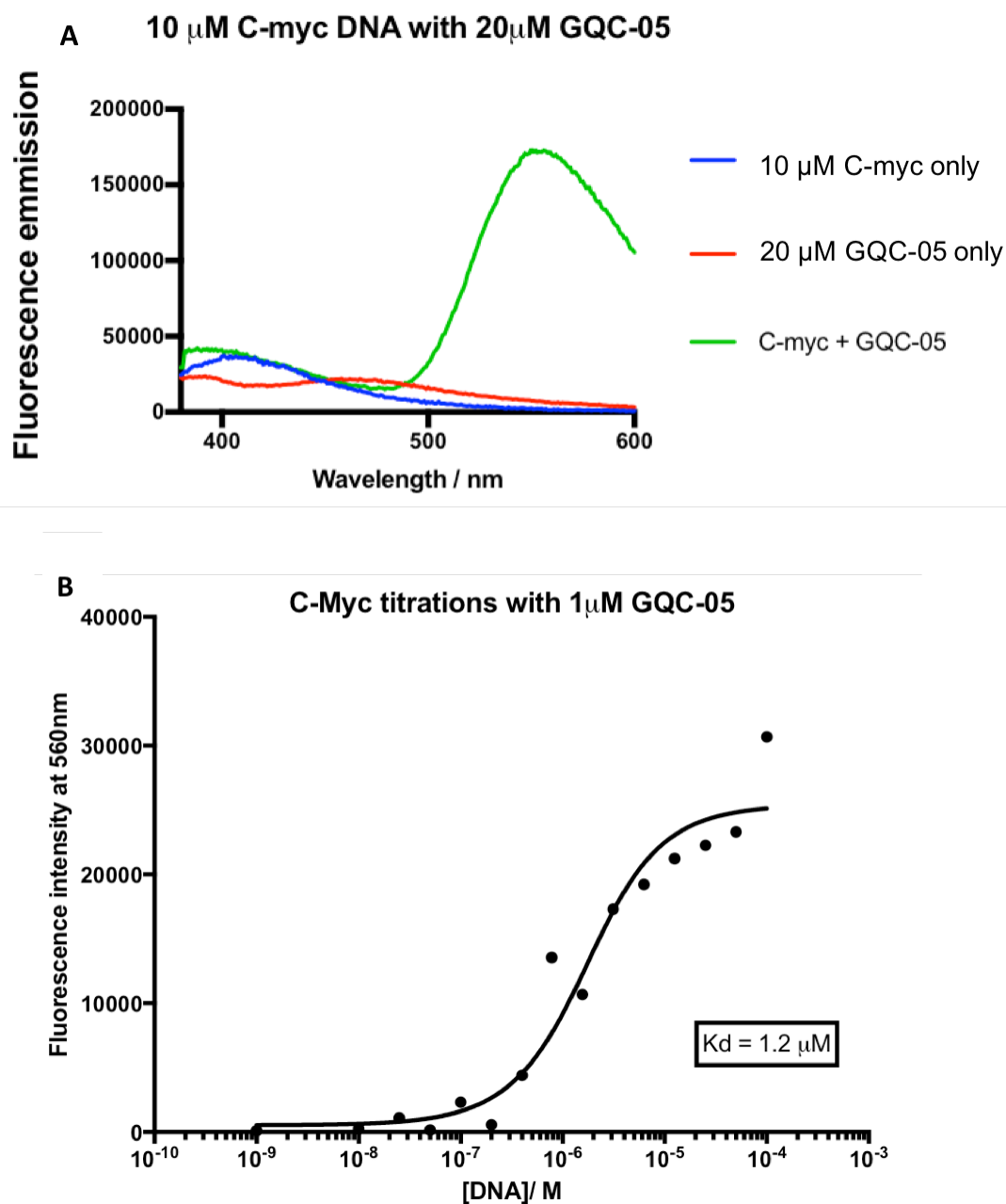
## 4.1 Introduction

As discussed in section 1.7, out of all of the G4 binding compounds tested, GQC-05 was shown to be the most effective in inhibiting the  $X_L$  and activating the  $X_S$  5'ss, thereby warranting further investigation on its mechanism of action. GQC-05 was initially identified as a DNA G4 specific ligand through observing an interaction with the myc-promotor quadruplex. In addition, it was also shown through competition dialysis experiments that GQC-05 is indeed more selective for other G4 structures over their duplex counterparts (see figure 1.64 B in section 1.6.3) (Brown et al., 2011). Therefore, following the identification of a quadruplex proximal to the  $X_S$  5'ss discussed in the previous chapter, in this chapter we will explore the interaction of GQC-05 with this G4 region, in order to understand its specificity for the G4 structure and how this may relate with the observed activation of the  $X_S$  5'ss upon addition of the compound. In this chapter, we will discuss the use of various biophysical techniques that has been applied to examine this interaction, and how this may reflect on the unique properties of GQC-05 to shift the splicing of Bcl-x.

## 4.2 Fluorescence spectroscopy to study the structural specificity of GQC-05

### 4.2.1 Determining the binding constant ( $K_d$ ) of GQC-05 to the c-myc G4 using fluorescence

Fluorescence spectroscopy is a very sensitive method used to determine binding parameters such as the dissociation constant ( $K_d$ ) and stoichiometry. GQC-05 has intrinsic fluorescence properties, enabling easy and efficient binding analysis (Weldon et al., 2018). Figure 4.21 A shows the emission profile of GQC-05 ( $\lambda_{exc} = 320$  nm) in the absence and presence of the C-myc promotor quadruplex sequence, which was previously shown to bind GQC-05 by Surface Plasmon Resonance and is a well characterised parallel G4 structure (Brown et al., 2011). GQC-05 has a relatively low fluorescence emission when unbound, with an emission peak around 490nm. Upon binding to a target sequence, the emission shows a shift to 560 nm and is greatly enhanced (Weldon et al., 2018). Therefore, following the emission at 560 nm as a function of DNA/RNA concentration would enable binding parameters to be determined. Following the emission peak of GQC-05 at 560 nm as a function of the c-myc DNA concentration, provided a  $K_d$  similar to that reported by SPR ( $K_d = 1.14 \pm 0.25$   $\mu$ M) (figure 4.21 B) (Brown et al., 2011).

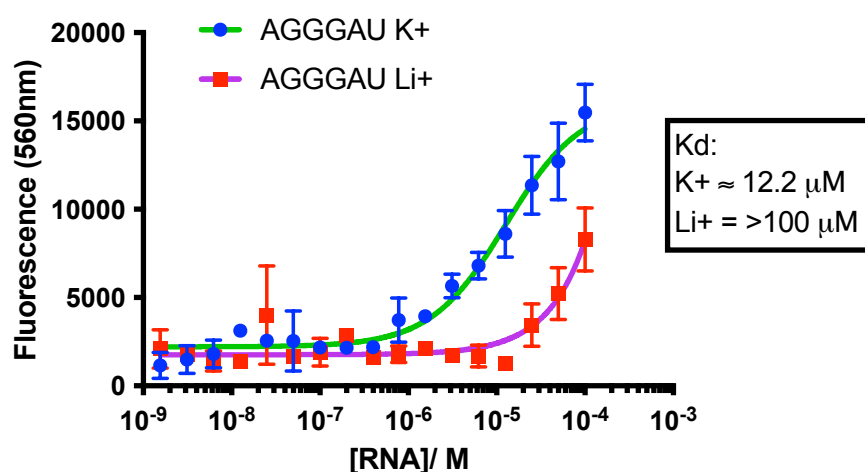


**Figure 4.21 GQC-05 and c-myc fluorescence emission spectra ( $\lambda_{\text{exc}} = 320\text{nm}$ ).**

(A) Fluorescence emission spectra of GQC-05 in the presence and absence of the C-myc promotor quadruplex (10 mM Tris pH 7.5 and 100 mM KCl). (B) Fluorescence of GQC-05 (1  $\mu\text{M}$  in 10 mM Tris pH 7.0, 100 mM KCl) measured at 560nm as a function of the c-myc quadruplex sequence.

#### 4.2.2 Structural vs sequence specificity of GQC-05

Several studies have shown that GQC-05 can interact specifically to G4 structures over their duplex counterparts (Brown et al., 2011; Kaiser et al., 2013). For example, GQC-05 displayed a 45 fold preference in the concentrations bound per binding site for the c-myc G4 over its double stranded counterpart, as determined by competition dialysis experiments (Brown et al., 2011). However a question as to whether or not GQC-05 is showing sequence rather than structural specificity remains to be determined, as the interaction may only be mediated simply by the presence of guanines rather than an actual G4. To determine this, we tested the binding of GQC-05 on the 6 mer RNA sequence AGGGAU, which has been previously shown to form a 3-tetrad tetra-molecular quadruplex in the presence of KCl and remains single stranded in the presence of LiCl (Dominguez et al., 2010). This will allow us to determine if GQC-05 is recognising the G4, or simply the Gs, by comparing the binding affinities in both the K<sup>+</sup> and Li<sup>+</sup> - containing buffers. A greater affinity in the presence of K<sup>+</sup> would indicate that GQC-05 is recognising the G4, whereas a higher affinity in Li<sup>+</sup> would indicate that it is only recognising the guanine base. RNA titration experiments showed that GQC-05 has a higher affinity for the AGGGAU RNA in the presence of K<sup>+</sup> compared to Li<sup>+</sup>, suggesting that the binding of GQC-05 is structural (G4) rather than sequence (guanine) specific (figure 4.22).



**Figure 4.22** Fluorescence emission at 560nm of GQC-05 with the RNA AGGGAU ( $\lambda_{exc} = 320\text{nm}$ ).

Fluorescence binding isotherm of the tetra-molecular G4 sequence AGGGAU in the presence of either 100 mM KCl or LiCl. Experiment was conducted in 10 mM Tris pH 7.0.



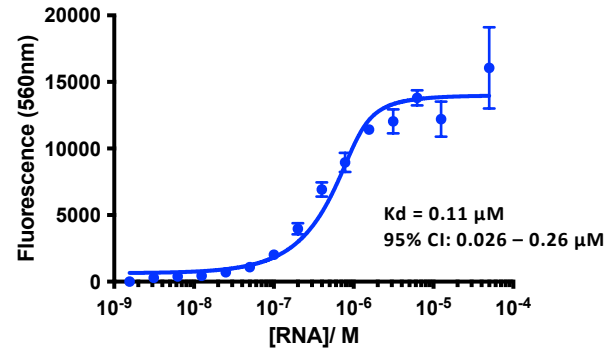
#### 4.2.3 Determining the binding constant of GQC-05 for the Bcl-x Q2 sequence

Using aforementioned fluorescence emission binding assay, we next determined the ability of GQC-05 to bind to the Q2 RNA quadruplex, along with the mutants  $\Delta$ G4 and Q2.2 (putative duplex forming sequences) as a control.

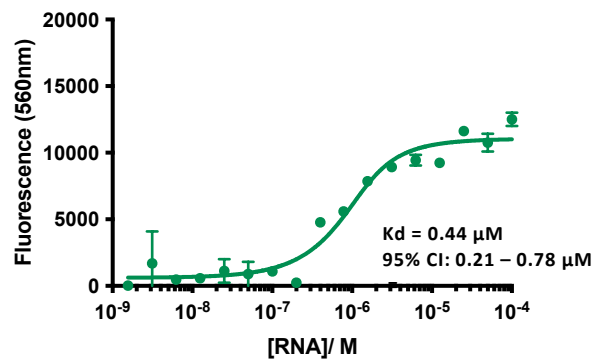
The fluorescence emission of GQC-05 at 560 nm when titrating increasing amounts of each of the RNAs to a maximum concentration of 100  $\mu$ M is shown in figure 4.23. Saturation of the fluorescence signal is obtained with an observed  $K_d$  of 0.11  $\mu$ M for the WT Q2 RNA (figure 4.23 A). Surprisingly, with the mutant  $\Delta$ G4, a similar  $K_d$  of 0.44  $\mu$ M was obtained, shown by overlapping 95% confidence intervals (CI) with the WT Q2 RNA (figure 4.23 B). Despite  $\Delta$ G4 being less prone to fold into G4 structures, the similar affinities displayed between these RNAs suggests that GQC-05 may not bind G4s specifically. This very much contradicts the competition dialysis experiments described previously (Brown et al., 2011), though it is important to note that these experiments were done on DNA rather than RNA G4s, which may lead to some differences. Our results are however consistent with previous findings, as discussed in section 1.6.3, showing that the ellipticine core is able to bind to duplex DNA (Canals et al., 2005), suggesting that the ellipticine core (in which GQC-05 is derived from) may contribute to the non-G4 interactions of GQC-05 seen with  $\Delta$ G4. Nevertheless, it is important to consider that as  $\Delta$ G4 displayed some residual Hoogsteen hydrogen bonds (figure 3.41 C), we cannot be certain that GQC-05 does not bind G4s entirely, as a G4 may exist in this RNA albeit with a shorter lifetime or a different conformation. To confirm the non-G4 mediated interactions of GQC-05, a well characterised non-G4 (or duplex) RNA will need to be tested. Interestingly, mutant Q2.2 displayed an approximately six-fold lower affinity ( $K_d$  = 0.58  $\mu$ M) than the WT Q2 sequence (figure 4.23 C), with non-overlapping 95% CIs. This may explain why Q2.2 was insensitive to GQC-05 by the *in vitro* splicing assays (Weldon et al., 2018). However, from these data, we cannot conclude that the binding of GQC-05 to the Q2 G-quadruplex is necessary for splice site activation, as a similar level of binding is also observed with the  $\Delta$ G4 mutant RNA which has less G4 forming potential.

**A**

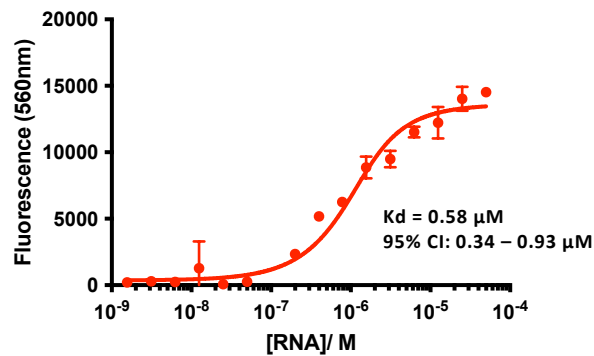
Bclx-Q2 =GGAUGGGGUAACUGGGGUCGCAUUGUGG

**B**

Bclx-ΔG4 =GUGAUGUUGUAAACUGUUGUCGCAUUGUGU

**C**

Bclx-Q2.2 =GGAUGGGGUAACUGUUGUCGCAUUGUGG



**Figure 4.23 Fluorescence emission at 560 nm of GQC-05 with different RNA fragments ( $\lambda_{\text{exc}} = 320\text{nm}$ ).** (A-C) Fluorescence binding isotherms of GQC-05 with Q2 (A),  $\Delta\text{G4}$  (B) and Q2.2 (C) RNAs in 10 mM Tris pH 7.0, 100 mM KCl. In all cases, the background fluorescence was subtracted to obtain the final fluorescence signal. Shown are the  $K_d$  values along with their respective 95% confidence intervals (CI).

#### 4.2.4 Determining the binding constant of GQC-05 in conditions that support *in vitro* splicing

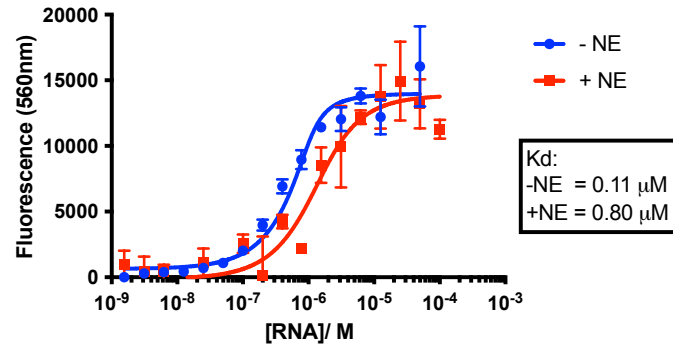
A major difference between the fluorescence assays described in section 4.2.3 and splicing assays is that fluorescence was done with naked RNA, whilst splicing assays are conducted when the RNA is covered by multiple RNA binding proteins. We therefore conducted fluorescence binding experiments in the latter condition by the addition of a nuclear extract, where the presence of RNA binding proteins (and other factors) may affect the interaction between GQC-05 and the RNA. We have already shown that the G4 structure observed in Q2 is still present in nuclear extract (see chapter 3), so any changes in binding of GQC-05 as a result of adding nuclear extract is not likely to be the result of structural remodelling of the RNA in the presence of RNA binding proteins.

Fluorescence binding assays between GQC-05 and each of the RNAs were now conducted in the presence of 10% HeLa nuclear extract, with the same buffer conditions described previously (figure 4.24), though as nuclear extract contains additional amounts of salt (such as 75 mM KCl and 350 mM NaCl), in 10% extract we would expect an additional 7.5 mM KCl and 35 mM NaCl, which would be present in the fluorescence assay. For the wild-type Q2 RNA, binding was observed both in the absence and presence of nuclear extract, giving an approximate 8-fold lower affinity in the presence of the nuclear extract (figure 4.24 A). However, for the  $\Delta$ G4 mutant RNA, addition of nuclear extract strongly reduced the binding of GQC-05 ( $K_d > 90 \mu\text{M}$ ) (figure 4.24 B). To determine whether the reduction in binding to the  $\Delta$ G4 RNA is related to the elimination of the G-tracts, the mutant Q2.2 RNA was also tested which only lacks the central G-tract (figure 4.24 C). Upon addition of nuclear extract, this sequence showed reduced binding compared with Q2, but displayed a greater level of binding than the  $\Delta$ G4 RNA. An explanation of this could be that GQC-05 can still interact with (or induce) bi or tetra-molecular structures formed by Q2.2, as disrupting the central G-tract would only abolish the formation of a unimolecular quadruplex species (in theory), whereas bi or tetramolecular G4 structures remain possible (see section 3.3). This is consistent with the fact that GQC-05 only displayed binding at higher concentrations of RNA, suggesting interactions with higher order quadruplex structures. These results suggest that nuclear extract confers specificity of GQC-05 for the Q2 G4 structure, whereas in the absence of nuclear extract, non-G4 mediated interactions may also be taking place. Collectively, these results suggest that GQC-05 is able to recognise and bind to the Q2 RNA G4 and potentially

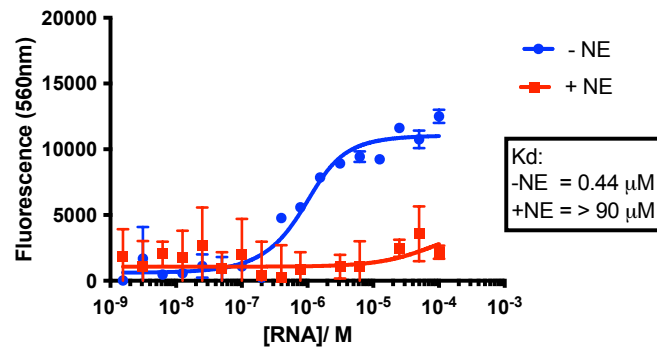
has increased specificity for this G4 over the secondary structures displayed by Q2.2 and  $\Delta$ G4 in the presence of nuclear extract. However, we cannot be certain that nuclear extract results in only G4 specific interactions of GQC-05 or that it abrogates non-G4 interactions entirely, as this would require binding assays to be conducted on well characterised G4 structures and known duplex sequences. However, these data are consistent with the previous splicing results, as the binding of GQC-05 is reduced to a greater extent on the Q2.2 mutant RNA compared to Q2 (and more in the presence than in the absence of nuclear extract), which could explain why GQC-05 was ineffective in the splicing assay of the mutant Q2.2. A possible mechanism could be that a G4 on the Q2 RNA is able to displace potential RNA binding proteins, thereby enabling it to recruit molecules of GQC-05 which will interact with the G4, whereas with the G4 abolishing mutant RNAs, the RNA binding proteins remain attached (as there is no stable secondary structure to displace them) and compete with GQC-05 for RNA binding, reducing its overall affinity. This hypothesis will require further investigation.

**A**

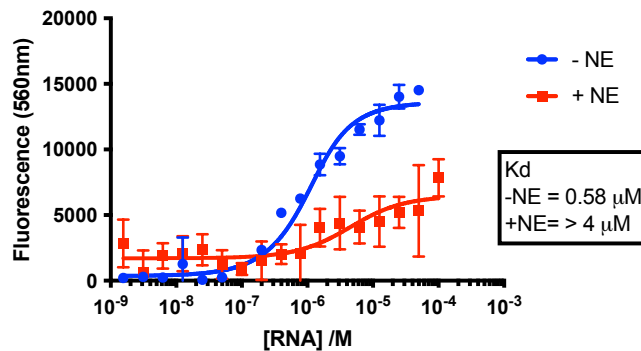
Bclx-Q2 =GGGAUGGGGUAACUGGGGUCGCAUUGUGG

**B**

Bclx-ΔG4 =GUGAUGUUGUAAACUGUUGUCGCAUUGUGU

**C**

Bclx-Q2.2 =GGGAUGGGGUAACUGUUGUCGCAUUGUGG



**Figure 4.24** Fluorescence emission at 560nm of GQC-05 with different RNA fragments in the presence of 10% HeLa nuclear extract ( $\lambda_{exc} = 320\text{nm}$ ).

(A-C) Fluorescence binding isotherms of GQC-05 in the absence (blue) and presence (red) of 10% nuclear extract (NE) with Q2 (A),  $\Delta$ G4 (B) and Q2.2 (C) RNAs in 10 mM Tris pH 7.0, 100 mM KCl. In all cases, the background fluorescence was subtracted to obtain the final fluorescence signal.

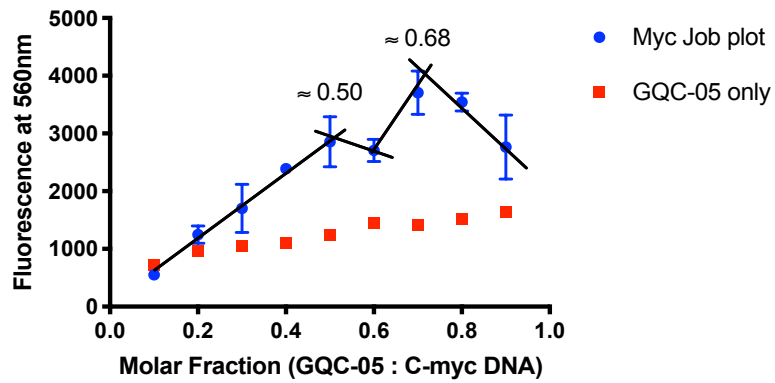
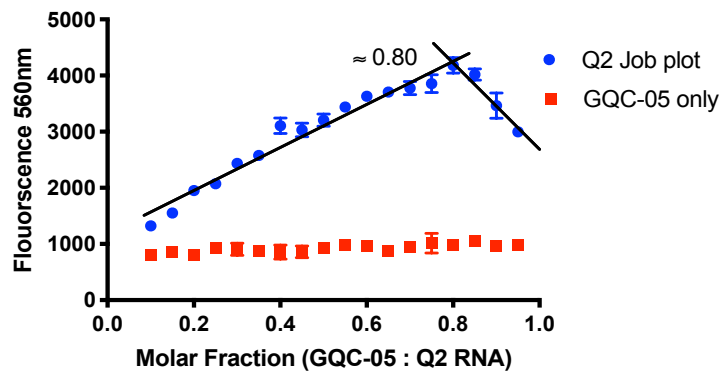
#### 4.2.5 Binding stoichiometry of GQC-05 to the c-myc DNA and Q2 RNA G4 sequences

As well as measuring the binding affinity, fluorescence spectroscopy can also be used to characterise/estimate the stoichiometry of the interaction between two molecules, using the method of continuous variation (MCV) or Job plot, which was originally described by Paul Job in 1928 (Job., 1928). In this method, the total molar concentration of both molecules are kept constant, but the molar ratios of each are changed (Renny et al., 2013). Any signal which corresponds to the bound form (such as the bound fluorescence emission signal of GQC-05) can be followed as a function of the molar ratio. This technique has previously been successfully used to predict the binding stoichiometry of various compounds to different nucleic acid secondary structures (including G4s) (Loontjens et al., 1990; Verma et al., 2018). The stoichiometry between GQC-05 and c-myc was previously shown to be 2:1 by SPR analysis (Brown et al., 2011). As a control, we therefore measured a Job plot based on the fluorescence of GQC-05 bound to the c-myc G4 (figure 4.25 A). In this assay, the molar concentration of both components was kept constant at 1  $\mu$ M, but the molar ratio of GQC-05:DNA was altered. Points of inflexion in the plot mark the binding events taking place between the small molecule and the macromolecule. Results indicate two points of inflexion, which suggest that two different binding events are taking place between GQC-05 and the c-myc G4. These correspond to molar ratios of 0.5 and 0.68, indicating binding events at 1:1 and 2:1 ratios of GQC-05:DNA respectively, consistent with the published data (figure 4.25 A) (Brown et al., 2011). To confirm that no further points of inflexion are occurring, more data points would need to be collected in future experiments.

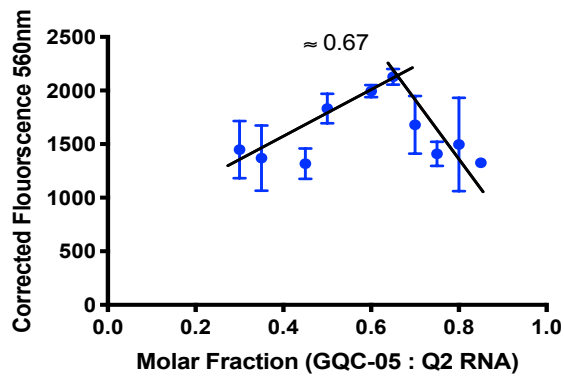
We next expanded the analysis to the binding of GQC-05 to Q2 RNA in 8 mM KCl (predominantly in monomeric conditions). The Job plot displays a major inflexion point at a molar ratio of 0.8 (determined by inspection), indicating a 4:1 stoichiometry between GQC-05 and the Q2 RNA (figure 4.25 B). As these points of inflexion were determined by inspection, more accurate methods may identify other binding events taking place.

Because in buffer GQC-05 is able to bind G4 but also non-G4 structures (figure 4.23 A-C above), the observed 4:1 stoichiometry might reflect the two modes of binding since these Job plots were conducted in 8 mM KCl, and we have shown that there is a mixture of both duplex and quadruplex structures under these conditions (see figure 3.53 in the previous chapter). We therefore measured a Job plot in the presence of nuclear extract, as we have

observed that addition of nuclear extract potentially abolishes non-G4 interactions whilst maintaining G4 contacts (figure 4.24 above). The Job plot analysis of GQC-05-Q2 binding in the presence of 10% HeLa nuclear extract shows a major inflexion point occurring at a molar ratio of 0.67 (figure 4.25 C), which clearly indicates a 2:1 binding stoichiometry between GQC-05 and Q2. This suggests that proteins in the nuclear extract are able to displace 50% of GQC-05 molecules from the RNA. A possible explanation could be that the nuclear extract is displacing the non-G4 weak interactions (such as duplex interactions) whilst retaining the stronger interactions with the G4 itself. The binding stoichiometry observed with GQC-05 in nuclear extract is very much consistent with many other specific G4 ligands, which have been shown to bind G4s through an end stacking binding mode (5' and/or 3' stacking on the external tetrads) (see chapter 1, section 1.6), which often give rise to stoichiometries at 1:1 or 2:1 (see chapter 1, section 1.6), supporting the idea that the 2:1 stoichiometry observed in nuclear extract with GQC-05 and Q2 may be that of a G4 specific interaction. However, this hypothesis would need clarification with further experiments such as a structure of the bound Q2-GQC-05 complex in the presence of a nuclear extract.

**A****B****C**

$$\text{FL (corrected)} = (\text{GQC-05} + \text{NE} + \text{RNA}) - (\text{GQC-05} + \text{NE})$$



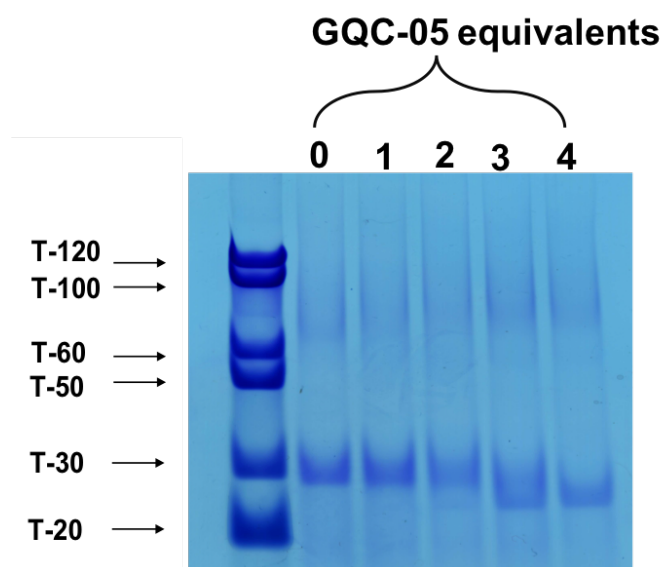
**Figure 4.25 Job plots on the G4 forming DNA/RNA.**

(A and B) Job plots of GQC-05 with C-myc DNA (A) and Q2 RNA (B), in 10 mM Tris pH 7.0, in 100 mM KCl with (A) and 8 mM KCl with (B). (C) Job plot of Q2 as in (B) but with the addition of 10% HeLa nuclear extract.



### 4.3 Native gel analysis to probe conformational changes of Q2 in the presence of GQC-05

To understand the conformational changes that may be occurring to Q2 as a result of GQC-05 addition, in addition to the binding events that may be taking place, we conducted native PAGE experiments in order to compare the migration of Q2 before and after incubation with increasing equivalents of GQC-05. This would allow us to see possible conformational changes occurring to the RNA as a result of GQC-05 addition, and enable us to identify the molar ratio of GQC-05 needed to sufficiently remodel the RNA. The addition of 4 equivalents of GQC-05 resulted in the formation of a band migrating lower than T-30, indicating a folded more compact conformation (figure 4.31). This is consistent with the results of the Job plot (in the absence of nuclear extract) which showed a binding stoichiometry of 4:1, as 4 equivalents of GQC-05 was needed bind and hence completely remodel the RNA into a putative G4 conformer. Native gel analysis in the presence of nuclear extract was not attempted due to the possibility of protein induced band shifts, making the interpretation very ambiguous.



**Figure 4.31 Native PAGE of Q2 with increasing equivalents of GQC-05.**

12 % native PAGE gel with 10  $\mu$ M Q2 RNA in 8 mM KCl with increasing equivalents of GQC-05.

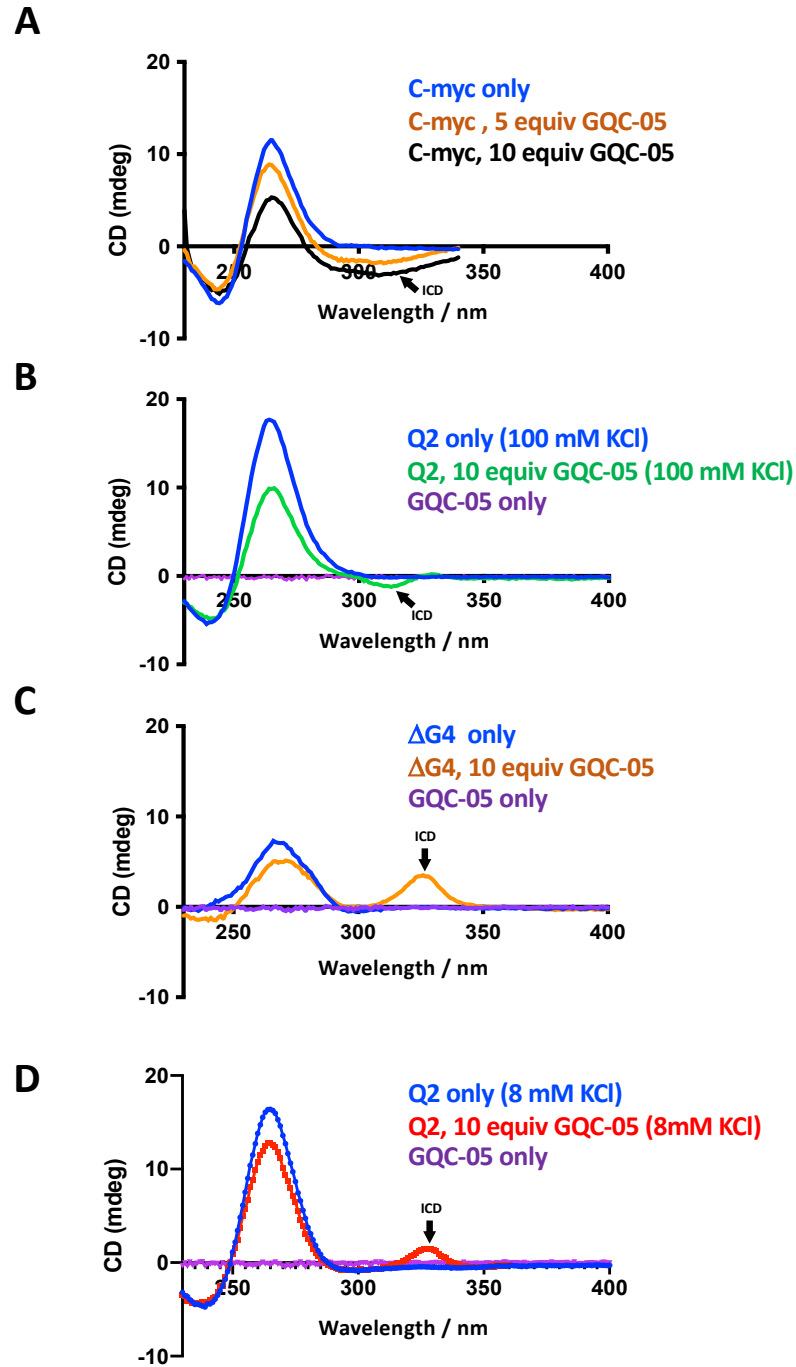
## 4.4 CD spectroscopy to study the binding mode of GQC-05

CD spectroscopy can also be used to study the binding mode of GQC-05 to nucleic acid sequences by analysing shape of the induced CD (ICD) spectrum (see section 1.5.2.2), and is explored below for the c-myc, Q2 and mutant sequences.

### 4.4.1 Induced CD analysis of GQC-05

As described in section 1.5.2.2, the shape of the ICD signature will provide information on the binding mode of GQC-05 to the RNA. For example, if we observe a positive ICD peak at 320 nm (the absorbance wavelength of GQC-05), this would indicate that GQC-05 is intercalating with its molecular axis parallel to the pseudo dyad. A positive ICD would also be produced if GQC-05 is binding in the minor groove of the RNA structure. The formation of a negative ICD at 320 nm would indicate that GQC-05 is intercalating perpendicularly to the pseudo dyad.

The ICD signature of GQC-05 when interacting with the c-myc G4 displays a negative ICD peak (figure 4.41 A). This suggests that binding to the G4 grooves is unlikely, and the interaction is rather through a possible intercalation or base stacking mode, which is consistent with the structural model of the GQC-05 – c-myc complex (figure 1.64 C, chapter 1). In a similar way, in the presence of 100 mM KCl, the ICD signature also is negative for the interaction between GQC-05 and Q2, which also eliminates the possibility of groove binding in these conditions (figure 4.41 B). In contrast, the  $\Delta$ G4 RNA showed a strong positive ICD at 320 nm, suggesting an intercalative binding mode where the molecular axis is parallel to the dyad, or minor groove binding (figure 4.41 C). Therefore, the observation that GQC-05 bound to Q2 displays a similar ICD to the c-myc G4 in the presence of 100 mM KCl is a very rough indication of a potential Q2 G4 interaction. Interestingly, when conducting the experiment with Q2 in 8 mM KCl, the ICD signal now becomes positive (figure 4.41 D). A possible explanation is that GQC-05 may also be interacting with duplex structures due to sample heterogeneity in these conditions, whereas at 100 mM KCl, there is a greater population of G4 structures and thus the G4 specific binding mode is predominant. This will need to be confirmed by measuring the ICD of GQC-05 when bound to well characterised duplex RNAs. Alternatively, the structure of the G4 may be different as a result of the dimerization observed at high KCl concentrations, altering the accessibility of the minor groove of the G4, which would otherwise be more accessible at the lower KCl concentrations.



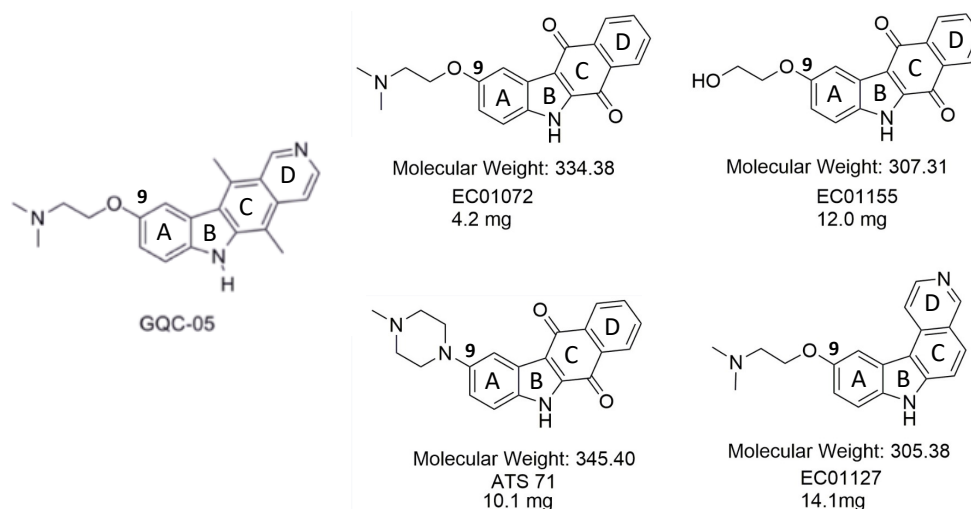
**Figure 4.41 Induced CD spectrums of GQC-05 with different DNA/RNA sequences at 20  $\mu$ M.**

(A) The c-myc DNA sequence. (B) Q2 in 100 mM KCl, (C)  $\Delta$ G4 in 100 mM KCl and (D) Q2 in 8 mM KCl. All experiments were conducted in 10 mM Tris pH 7.0.

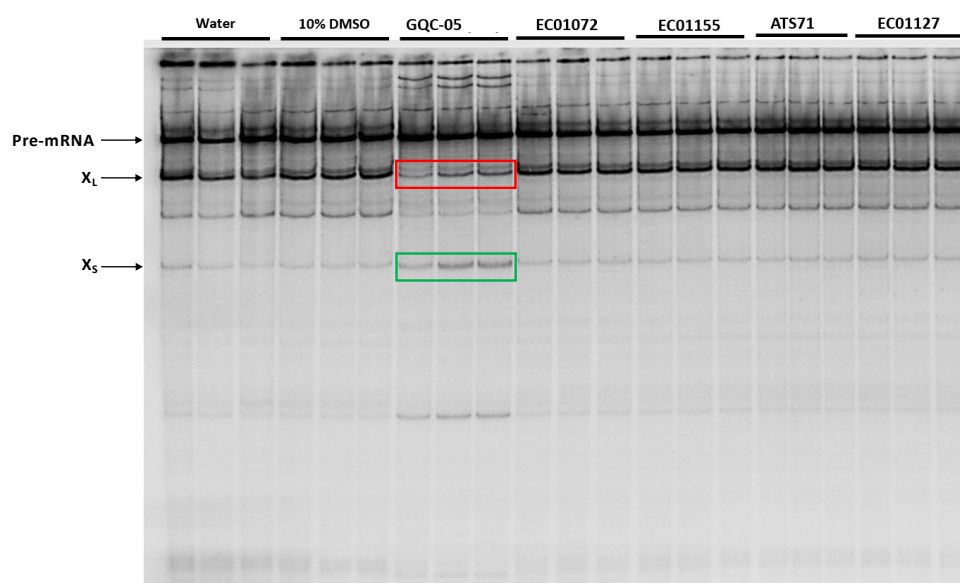
#### 4.4.2 Structure-activity profiling of GQC-05 analogues by ICD

As well as looking at the binding mode, the ICD signal can also be used to investigate the binding affinity of the complex in a similar way as with fluorescence measurements described above. A major advantage of following the ICD signal is that it does not rely on the ligand to have fluorescent properties, enabling the screening of a much wider variety of compounds. Therefore, we have taken this approach for structure activity relationship (SAR) profiling of GQC-05, to try and identify key functional groups within GQC-05 necessary for binding to the Q2 G4 and if this correlates with splice site selection. The laboratory of Professor Glenn Burley (University of Strathclyde) synthesised four GQC-05 derivatives (shown in figure 4.42 A) that were tested using *in-vitro* splicing by Dr Sudipta Ghosh (University of Leicester) (figure 4.42 B). Compound EC01072 has a similar structure to GQC-05, but displays ketone oxygens on the C-ring and lacks the nitrogen on the D-ring. EC01155 and ATS 71 are similar to EC01072, but EC01155 has an ethylene glycol chain instead of the tertiary amine at position 9, and compound ATS 71 has a piperazine side chain at position 9. Molecule EC01127 has a different ring structure to the other compounds, but contains a tertiary amine at position 9 and a nitrogen on the D ring alike GQC-05, but lacks the methyl groups on the C-ring. From the *in vitro* splicing assay (figure 4.42 B), addition of GQC-05 led to an inhibition of the X<sub>L</sub> 5' splice site usage (red box in figure 4.42 B), whilst an increase in X<sub>S</sub> 5' splice site usage was observed as expected (green box in figure 4.42 B). As can be seen in the figure, all of the other compound derivatives fail to switch Bcl-x splicing from X<sub>L</sub> to X<sub>S</sub>. This data suggests that the methyl groups on the C-ring and the nitrogen on the D-ring are necessary to achieve Bcl-x splice switch, and the correct ring system which resembles that of the ellipticine class of molecules is also required.

**A**



**B**



**Figure 4.42 SAR profiling of GQC-05 analogues.**

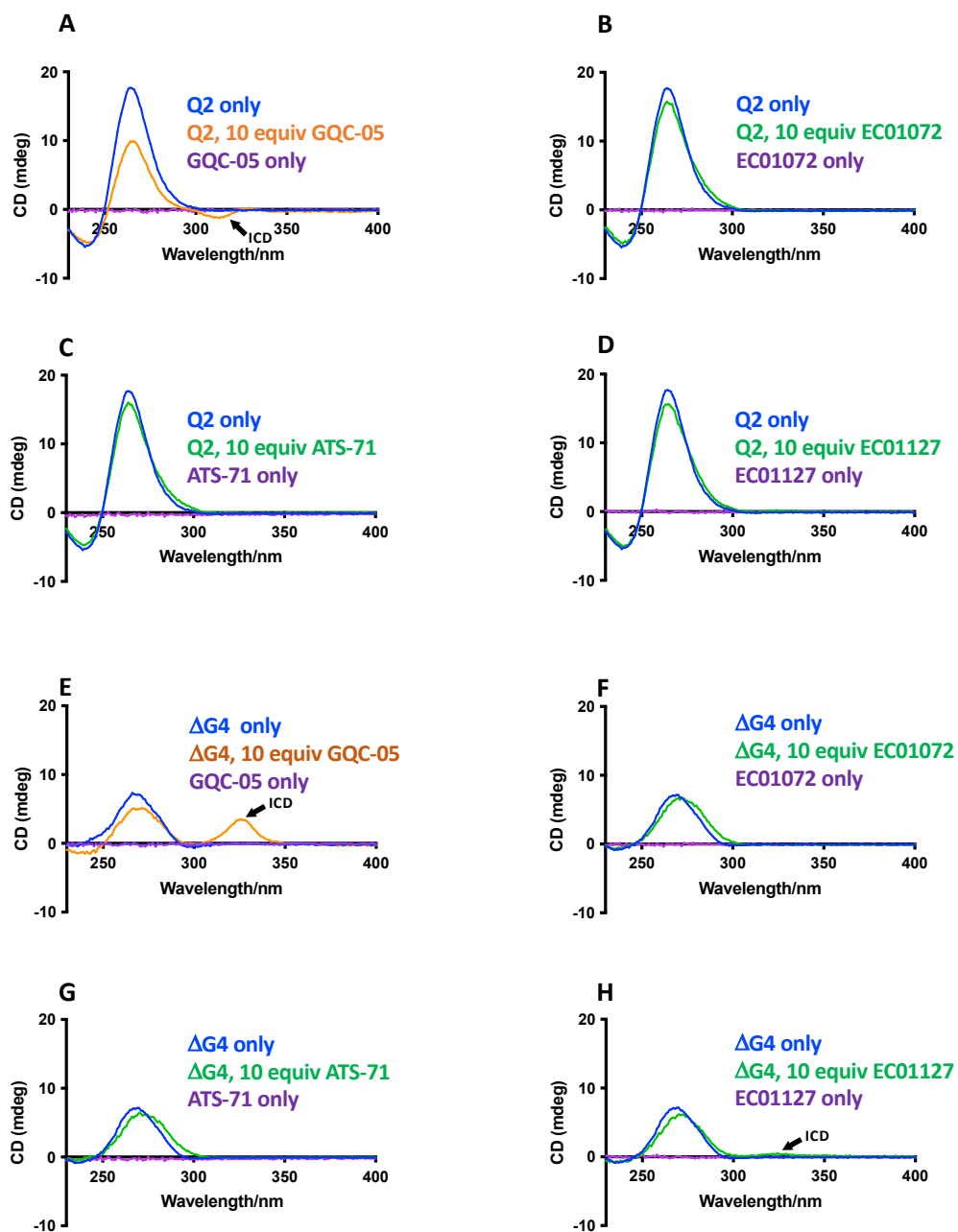
(A) GQC-05 and its derivatives generated by the lab of Professor Glenn Burley at the University of Strathclyde. (B) *In vitro* splicing assays of GQC-05 and its derivatives conducted by Dr Sudipta Ghosh.

Knowing that these compounds cannot shift the splicing pattern of Bcl-x, we next wanted to examine if the lack of X<sub>5</sub> 5'ss activation was due to the inability of these compounds to bind the Q2 G4. As all of these GQC-05 analogues did not display fluorescent properties unlike GQC-05, we used the ICD signal to determine binding. However, due to the solubility limit of some of these compounds, conducting a full titration binding assay was not possible. Furthermore, we did not conduct any additional experiments with EC01155, due to compound insolubility at 20  $\mu$ M (concentration used in the CD experiments). We therefore did a single point measurement of the ICD signal to determine the relative affinities of each of the compounds and to determine structural changes occurring to the RNA after addition of the compounds. I chose to use a condition of 100 mM KCl in Tris buffer to monitor the interaction of the compounds with the Q2 G4. Although this condition is favourable for dimerization of the G4, it also reduces the population of duplex structures which may also exist at the lower salt concentrations which could contribute to the shape of the ICD signal of the ligand. Additionally, as some of these compounds exhibit a positive charge due to the presence of a tertiary amine tail, conducting the experiments in a high ionic strength buffer would be necessary to eliminate any non-specific electrostatic interactions the compounds make with the RNA. Therefore, to monitor only the G4 interactions, I used this condition.

Shown below are the induced CD spectrums obtained after 10 equivalents of each of the compounds to both Q2 (figure 4.43 A-D) and the  $\Delta$ G4 (figure 4.43 E-H) RNAs. As described earlier, with GQC-05, we see changes to the structure of Q2 (shown by a decrease in the intensity of the CD spectrum in the 220-290nm range) as well as a negative ICD at 320 nm as expected in these conditions (figure 4.43 A). However, compounds EC01072, ATS-71, and EC01127 did not display any ICD signals upon addition to Q2, and the CD spectrum of the RNA is almost overlapping, suggesting that these compounds may not be able to interact with Q2 and do not elicit a conformational change to the Q2 RNA (figures 4.43 B-D). With the  $\Delta$ G4 RNA, GQC-05 resulted in substantial remodelling of the RNA and showed a positive ICD as described earlier (figure 4.43 E). This is consistent with the fluorescence binding assays which showed binding to both the Q2 and  $\Delta$ G4 RNA in the absence of nuclear extract. With the compound derivatives, the shape of the CD spectrum corresponding to the RNA display less overlaps unlike the Q2 RNA after addition of the compounds, suggesting a greater level of

structural remodelling of the  $\Delta$ G4 mutant RNA (figure 4.43 F-H). A positive ICD can also be seen with compound EC01127, suggesting a non-G4 interaction, but this is much smaller than GQC-05. However, as the ICD is rather qualitative, and the fact that binding titrations by ICD could not be conducted due to limitations in compound solubility, other more quantitative techniques will be required to determine detailed biophysical parameters (such as affinity constants).

In summary, the absence of an ICD with the GQC-05 derivatives as well as an inability to remodel the structure of Q2 suggests that the interaction with the Q2 G4 is necessary to achieve activation of the X<sub>5</sub> 5'ss, in which only GQC-05 succeeded. The presence of the nitrogen on the D-ring as well as the methyl groups on the C-ring of GQC-05 may provide additional contacts with the Q2 G4, enabling it to bind with a high enough affinity to remodel the RNA and achieve splice site activation.



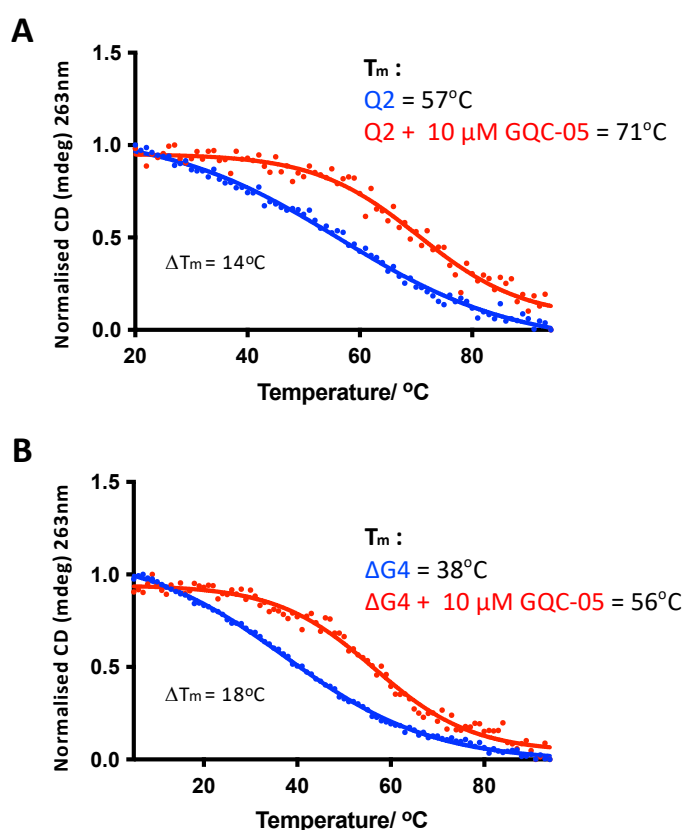
**Figure 4.43 Induced CD spectrums of GQC-05 and its analogues.**

ICD signatures of each compound on Q2 RNA (A-D) and the  $\Delta$ G4 RNA (E-H) All experiments were conducted in 20  $\mu$ M RNA, 10 mM Tris pH 7.0, 100 mM KCl.



#### 4.4.3 Thermal melting of Bcl-x Q2 and $\Delta$ G4 RNA in the presence of GQC-05

To examine the effect of GQC-05 on the overall stability of the Q2 and the  $\Delta$ G4 RNAs, we conducted CD melting experiments in the presence of 10 mM KCl (figures 4.44 A and B). For the WT Q2 RNA we observed a 14°C increase in the melting temperature (from 57 to 71°C). We also noticed a similar level of increase (18 °C) in the  $\Delta$ G4 RNA (from 38 to 56°C). This confirms that both RNAs are structured, which is consistent with our NMR data (figure 3.41), and that GQC-05 is able to interact with both quadruplex and duplex forming sequences, which is consistent with our fluorescence binding data (figure 4.23), and suggests that GQC-05 stabilises both structures. However, as Q2 in 10 mM KCl most probably adopts a mixture of G4 and duplex structures, it is difficult to conclude if the increase in melting temperature is caused by duplex or quadruplex stabilisation, as both duplex and parallel G4s display similar CD signatures. Therefore, to distinguish the effects of GQC-05 on quadruplex or duplex structures, we used NMR spectroscopy.



**Figure 4.44 CD thermal melts in the presence of GQC-05.**

(A and B) CD melt in the presence of 10 mM TBA phosphate pH 7.0 and 10 mM KCl in the presence (red) and absence (blue) of GQC-05 with the wild type Q2 (A) and the  $\Delta$ G4 mutant (B) RNAs.

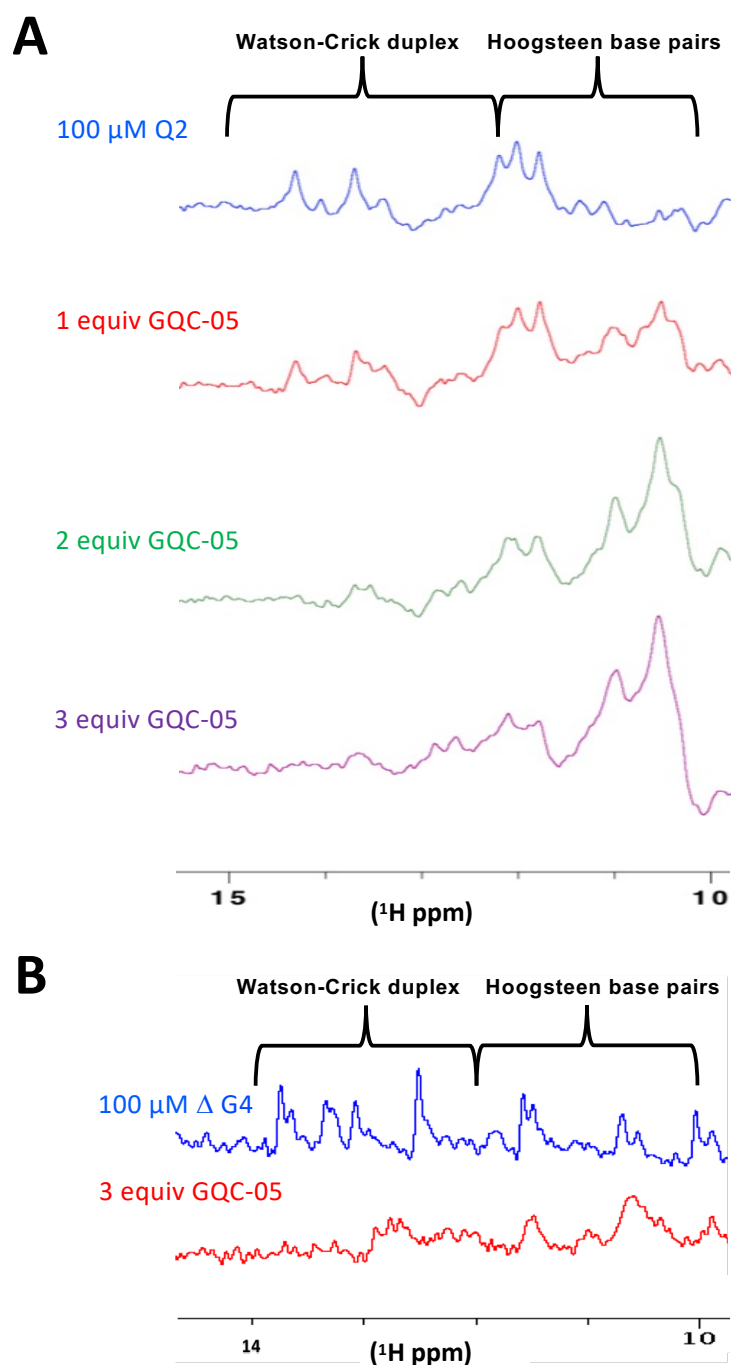
## 4.5 NMR spectroscopy to determine the structural specificity of GQC-05

Unlike CD spectroscopy, NMR can be used to distinguish between duplex and quadruplex structures as they give rise to imino peaks in distinct regions in the NMR spectrum (see chapter 3). Therefore, we aim to use this technique to identify the structural specificity of GQC-05 recognition of the Q2 region.

### 4.5.1 1D NMR titrations with GQC-05 and RNAs

The NMR spectrum of Q2 in the absence of salt suggests that the RNA adopts a mixture of both G4 and duplex structures, implying the presence of two populations in these conditions (figure 3.53 B). We proceeded with NMR titrations of Q2 with increasing amounts of GQC-05 to determine its effect on each of the structures (figure 4.51 A). Addition of GQC-05 resulted in a decrease of the imino peaks corresponding to a duplex (13-15ppm), but showed an increase in the peaks corresponding to the G4 (10-12ppm) in a dose dependant manner. This suggests that GQC-05 can elicit a conformational change of Q2 shifting the equilibrium from the duplex population towards the stabilization of a G4 structure. This also suggests that the band observed in the native gel at 4 equivalents of GQC-05 (figure 4.31), in which we concluded a folded more compact formation of the RNA, could be because of a G4 species being induced as a result of GQC-05 addition. We also conducted a similar experiment with the  $\Delta$ G4 RNA (figure 4.51 B), where we noticed a reduction of the signals at 3 equivalents of GQC-05, particularly those between 12-14 ppm in the Watson-Crick region. In the Hoogsteen region, there were small changes, though this was not as distinct as with the Q2 RNA. Nonetheless, we cannot exclude the possibility of an induced G4 in the  $\Delta$ G4 RNA in the presence of GQC-05. The reduction seen the duplex region could suggest that the binding of GQC-05 is resulting in unfolding of the duplex RNA, or that the binding event is altering solvent exchange. The second possibility seems more plausible because we previously observed an increase in the stability of the  $\Delta$ G4 RNA in the presence of GQC-05 (figure 4.44 B), indicating that the unfolding mechanism is unlikely. This observation however leads to another hypothesis observed with the Q2 RNA, in that instead of shifting the equilibrium from a duplex towards a G4, GQC-05 may actually be binding (and stabilising) both structures (duplex and quadruplex), and that binding to the duplex population results in peak disappearance whereas binding to the G4 results in chemical shift changes and peak intensity increase.

Therefore, to answer the question of the preferred structural selectivity of GQC-05 (duplex or quadruplex), we used NMR kinetics to investigate the rate of duplex and quadruplex formation in the presence and absence of GQC-05.



**Figure 4.51 NMR titrations with GQC-05 on Q2 and  $\Delta$ G4 RNA.**

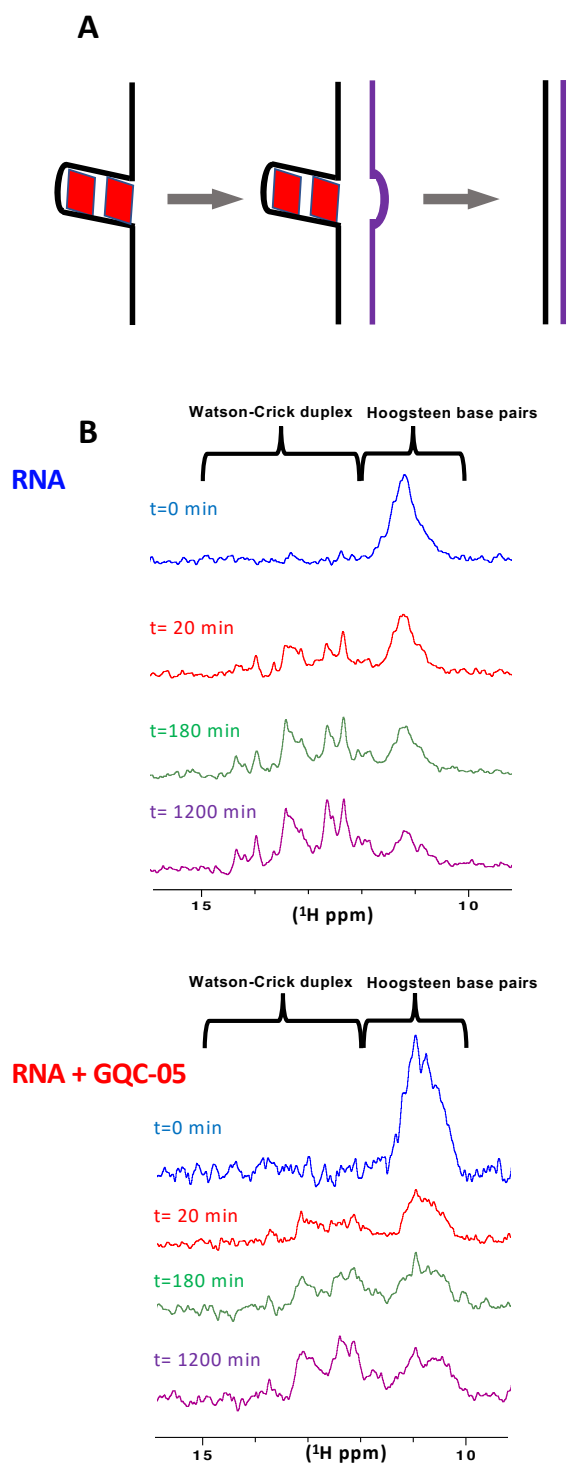
(A) Titration up to 3 equivalents of GQC-05 on Q2 RNA. (B) With 3 equivalents of GQC-05 on the  $\Delta$ G4 RNA.  $^1\text{H}$  NMR (600 MHz, 10°C) in 10 mM Tris pH 7.0 for (A), and 100 mM KCl addition in (B).

#### 4.5.2 Complement trapping to determine structural selectivity of GQC-05

To determine if GQC-05 can preferentially stabilise duplex or G4 structures, we performed a previously described complement trapping NMR kinetic experiments, shown by the scheme depicted in figure 4.52 A (Gray et al., 2019). Here, we started with the Q2 sequence annealed in 8 mM KCl (see section 2.1.2 on the annealing method), which we showed to preferentially fold into a monomeric G4 by giving rise to imino peaks in the 10-12ppm region (figure 3.53 and figure 4.52 B). This was followed by addition of the Q2 complementary strand sequence (at equimolar concentration to Q2), which is able to anneal and form Watson-Crick base pairings with Q2, outcompeting the Hoogsteen base pairs that were initially present in the Q2 G4. Over time, the G4 would be out-competed by the duplex, which can be observed in 1D NMR spectra over time by the disappearance of the G4 signals (10-12 ppm) and the appearance of Watson-Crick imino peaks (12-15 ppm). By comparing the rate at which the duplex forms (or the G4 disappears) in the presence and absence of GQC-05, we can determine if GQC-05 has preferential stability for a particular structure

The experiment was performed on Q2 RNA either in the absence or in the presence of 4 equivalents of GQC-05. At  $t=0$ , we see an envelope of peaks in the 10-12 ppm, despite our previous NMR and native PAGE analysis indicating predominant monomeric conformations (figures 4.52 B and 3.53). Firstly, one possibility is that the NMR KCl titrations conducted in figure 3.53 B were done on an RNA already pre-annealed in the absence of KCl, in which KCl was then added directly on top of that pre-annealed sample before recording the spectrum. In this case however, we have annealed the RNA in 8 mM KCl before recording the spectrum at  $t=0$ , which may favour the formation of higher order G4 structures. Secondly, the differences observed between the NMR spectrum and the native PAGE (in figure 3.53 A) is likely to arise as a result of the RNA concentration, as a 10-fold higher concentration of RNA is used in NMR than on a native PAGE (100  $\mu$ M as compared to 10  $\mu$ M on the native PAGE), which will tend to favour multimerization of the RNA. Additionally, NMR is more sensitive in detecting multimeric conformations than a native PAGE stained with toluidine blue. However, we proceeded in these conditions as we are interested in monitoring the structural specificity of GQC-05 for G4 over duplex structures, and at 8 mM KCl, there is likely a predominant G4 conformation based on our previous analysis. After addition of the complementary strand, both in the absence and presence of GQC-05, we observe in a time-dependent manner the

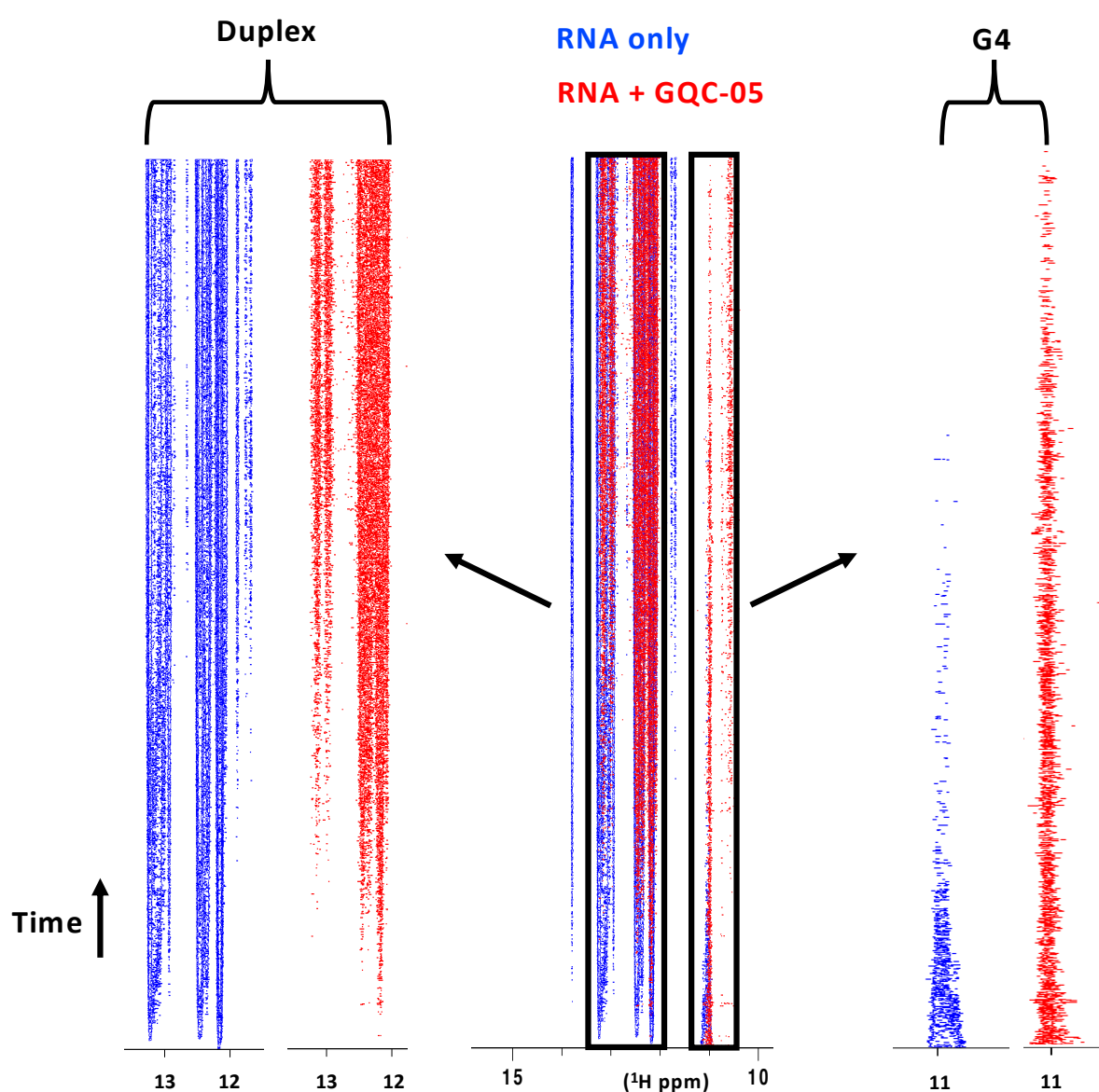
expected appearance of the duplex and the disappearance of the G4 imino proton signals, indicating the formation of Watson-Crick base-pairings to the complementary strand over time. Because we measured spectra every 160 seconds over a period of 72 hours, we can compare the rate of quadruplex-duplex exchange in the absence or presence of GQC-05.



**Figure 4.52 Complement trapping theory and 1D NMR kinetics.**

(A) Scheme showing the experimental design of the NMR experiment, with the Q2 G4 strand in black and its complementary strand in purple. (B) <sup>1</sup>H NMR spectrum at certain time points after the addition of the complementary strand in the absence and presence of 4 equivalents of GQC-05. Experiments were carried out using 100 μM RNAs at 283 K in 8 mM KCl, 10 mM Tris pH 7.0.

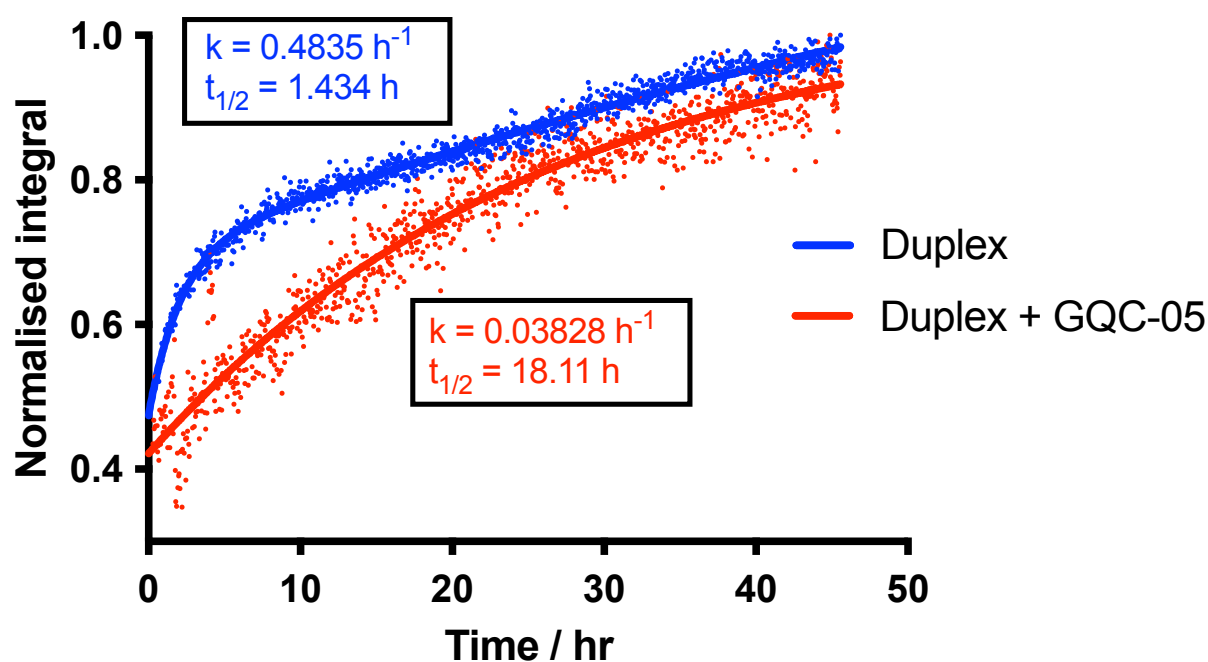
The differences in rate of formation and disappearance of the duplex and G4 peaks can be clearly visualised using a pseudo 2D trace (shown in figure 4.53) that displays the peak intensities across the imino regions (10-15 ppm) as a function of time (in the y direction). Analysis of the G4 trace shows that the peaks disappear more rapidly in the absence of GQC-05 (blue) than in the presence of GQC-05 (red). The reverse is observed for the duplex region, where the appearance of Watson-Crick pairing appears almost instantly in the absence of GQC-05 (blue), but much later in the presence of GQC-05 (red).



**Figure 4.53 Pseudo 2D traces in the absence and presence of GQC-05.**

Pseudo 2D traces showing the peak intensities across the imino region as a function of time in the y direction in the absence (blue) and presence (red) of GQC-05. Zoomed in on the right and left hand panels are the peaks corresponding to the G4 and duplex regions respectively. Experiments were carried out using 100  $\mu$ M RNAs at 283 K in 8 mM KCl, 10 mM Tris pH 7.0.

We next plotted the integrals of the peaks in the 13-15 ppm region as a function of time to obtain the rates of duplex formation using first order kinetics (figure 4.54). The estimated half-time ( $t_{1/2}$ ) of duplex formation in the absence of GQC-05 is 1.4 hours, whereas it is >18 hours in the presence of GQC-05. These data clearly indicate that GQC-05 can bind and selectively stabilise the Q2 G-quadruplex structure over the duplex.



**Figure 4.54 Rate curves showing duplex formation in the absence and presence of GQC-05.**

Rate curves fitted to a first order reaction, showing the rate of duplex formation (based on peak integrals in the duplex region) as a function of time (data was analysed using dynamic centre on Topspin).



## 4.6 Discussion

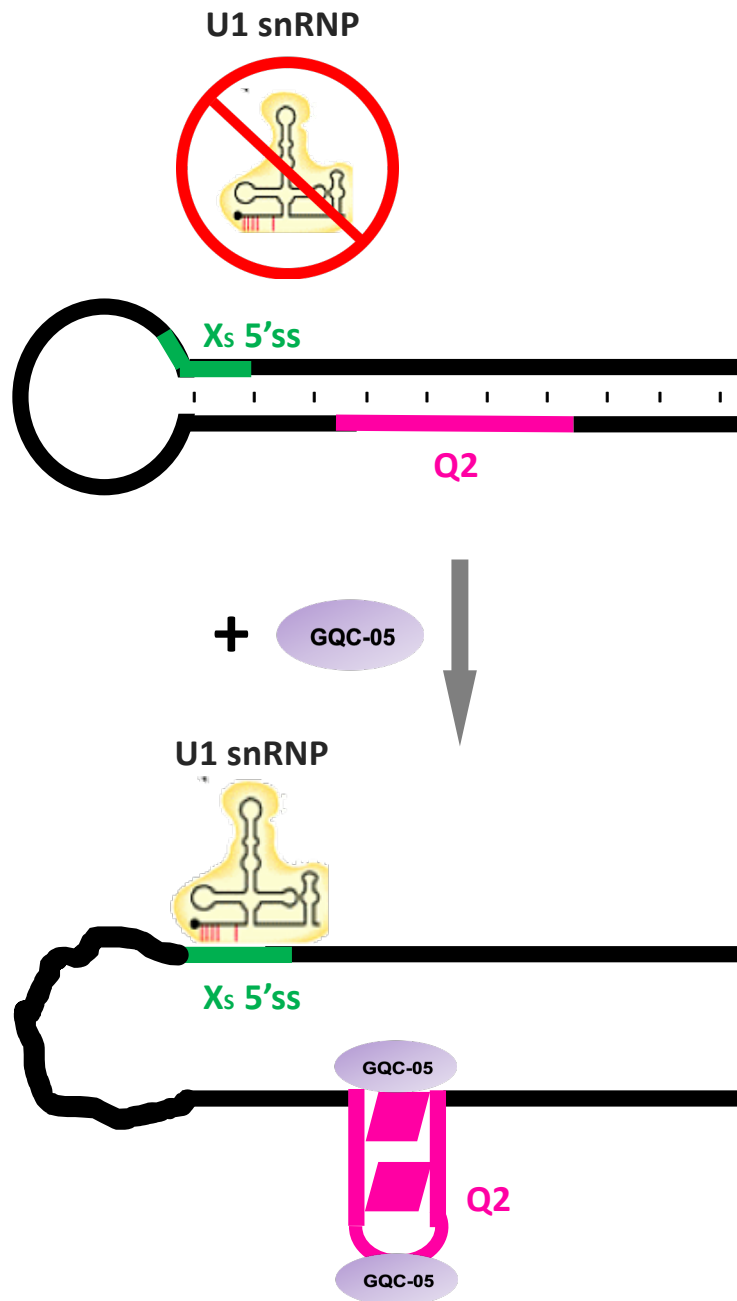
In this chapter we have combined CD, fluorescence and NMR to demonstrate that GQC-05 can bind to both the Q2 and  $\Delta$ G4 RNA and with  $\mu$ M affinities, suggesting that it may lack the specificity for G4 structures, very much contrary to what was described previously in the literature (Brown et al., 2011). However, as  $\Delta$ G4 still displayed Hoogsteen hydrogen bonds, we cannot exclude the possibility that GQC-05 may be binding to a residual G4 in this sequence, or that it may be inducing the formation of one. Binding assays with better characterised duplex structures will need to be tested to confirm if GQC-05 lacks G4 specificity. Very interestingly, in the presence of a nuclear extract which contains multiple RNA binding proteins, the binding of GQC-05 to Q2 was retained, whilst the binding decreased with the mutant Q2.2 and was abolished with the  $\Delta$ G4 RNA (figure 4.24 A-C). Earlier experimentation suggested that this is indeed a structural rather than sequence specific interaction of GQC-05 with G4 structures, since the G4 in Q2 is present even in the presence of nuclear extract (figures 3.36 – 3.38). This observation is now consistent with our splicing assays, in that altering the G4 in Q2 has the ability to disrupt the GQC-05 interaction with the RNA, which may prevent the formation of a G4 induced by GQC-05 and hence reducing its ability to activate the X<sub>s</sub> 5'ss. The exact mechanism by which the nuclear extract is synergising with GQC-05 to achieve its structural specificity remains to be explored. An avenue worth exploring is the cooperativity that may exist between GQC-05 and hnRNP F/H. Similar to GQC-05, hnRNP F/H was shown to interact with Q2 and maintaining it in a single stranded conformation (Dominguez et al., 2010). The possibility of cooperativity that may exist between GQC-05 and hnRNP F/H cannot be excluded, especially considering that the mutant RNAs, particularly  $\Delta$ G4, also lacks the hnRNP F/H binding site.

Further evidence suggesting that the GQC-05/G4 interaction in Q2 is necessary for X<sub>s</sub> splice site activation came from our SAR analysis of the GQC-05 derivatives. These derivatives did not show any effect in the alternative splicing of Bcl-x, unlike GQC-05. Our induced CD results demonstrate that these compounds do not interact with the Q2 RNA in contrast to GQC-05 and show substantially less re-modelling of the RNA. This suggests that the inactivity of these compounds in Bcl-x splicing assays are due to their inability to bind the Q2 region.

Knowing that GQC-05 can remodel the RNA in such a way that may influence splice site selection, we wanted to see if the binding can indeed bias a G4 structure in Q2 over a duplex.

We showed here by NMR spectroscopy that GQC-05 can kinetically slow down the rate of duplex formation in Q2 and cause G4 structures to persist for much longer periods of time than in the absence of the ligand. This observation is particularly important when we consider the proposed mechanism of action of GQC-05 in activating the upstream site (figure 4.61). Based on previous findings in our lab, we proposed a mechanism by which the poor usage of the X<sub>s</sub> 5'ss in the absence of a ligand is due to the formation of a stem-loop that prevents it to bind the U1 snRNP spliceosome component (figure 4.61) (Weldon et al., 2018). However, the hypothesis was that upon addition of the G4 stabilising ligand GQC-05, the G4 in the Q2 region will be stabilised, inducing a destabilisation of the stem-loop and therefore increasing the accessibility of the X<sub>s</sub> splice site to U1 snRNP. Therefore, this equilibrium that may exist between the duplex and the quadruplex is something we wanted to explore by NMR, and the NMR experiment described here supports the hypothesis of a G4 stabilisation in expense of the duplex.

In conclusion, by collectively taking all of the experiments described in this chapter, we have obtained growing evidence to show that the presence of a Q2 G4 induced by GQC-05 is necessary to achieve splice site activation. However, it is important to note that under all of the conditions described in chapters 3 and 4, we have characterised the structure of the Q2 G4 and assayed for GQC-05 binding with the Q2 RNA in isolation, and not when in context of longer more relevant RNAs for which it is imbedded into. This is particularly essential as the Q2 G4 will exhibit much more competition with other secondary structures in the longer RNAs than shorter ones, where more stable duplex structures may out compete the G4, even in the presence of GQC-05, meaning that the *in vitro* characterisation conducted with the RNA in isolation may be biologically irrelevant. Therefore, in the next chapter, we aim to address if the Q2 G4 can still form when imbedded in the longer X<sub>s</sub> domain and how GQC-05 may engage and remodel this longer RNA to achieve splice site selection.



**Figure 4.61 Schematic showing proposed mechanism of GQC-05 at the X<sub>5</sub> 5'ss.**

The proposed structural changes near the Q2 region before (top panel) and after (bottom panel) the addition of GQC-05. Before GQC-05, the X<sub>5</sub> 5'ss (green) is inaccessible to the U1 snRNP. Following GQC-05 addition, a G4 is stabilised in Q2 (magenta), destabilising the stem (black), making the X<sub>5</sub> 5'ss accessible to the U1 snRNP.

## **CHAPTER 5: Structural characterisation and binding specificity of GQC-05 for longer RNAs surrounding the Q2 region**

- 5.1 Introduction**
- 5.2 Rationale behind the design of each of the larger Bcl-x constructs**
- 5.3 Generating template DNA for transcription of the longer fragments of Bcl-x**
- 5.4 Transcription and purification of the large RNAs**
- 5.5 G4 characterisation on the larger Bcl-x RNAs**
- 5.6 Important structural elements for the specificity of the GQC-05 interaction**
- 5.7 Effects of GQC-05 in the structural remodelling of the X<sub>S</sub> domain**
- 5.8 The effect of GQC-05 in altering the U1 accessibility at the X<sub>S</sub> 5'ss**
- 5.9 Discussion**

## 5.1 Introduction

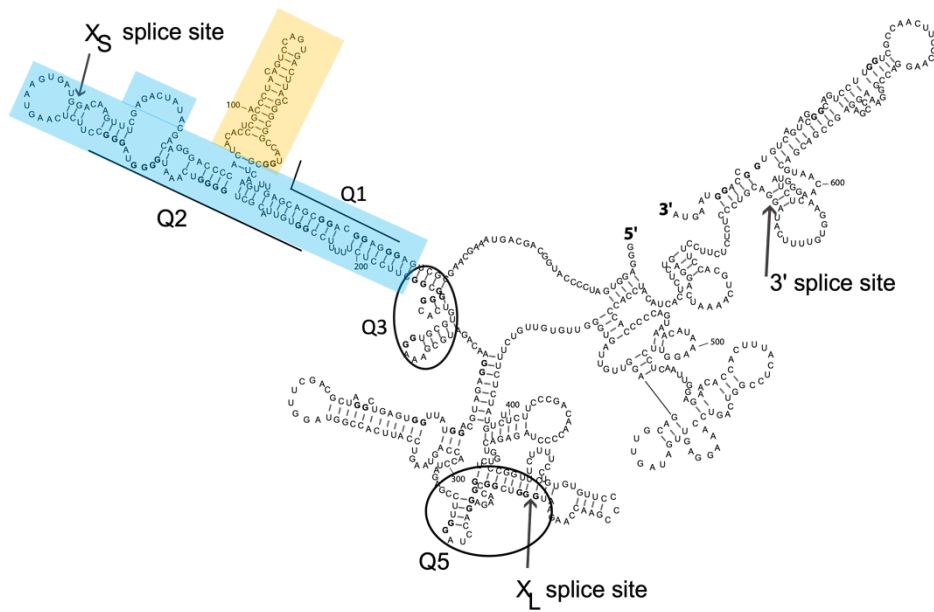
In the previous two chapters, we have shown that the G-rich region termed Q2 when studied in isolation is able to fold into a quadruplex and that GQC-05 can bind this structure in the presence of a nuclear extract. We also showed that binding of GQC-05 can shift the equilibrium from a duplex towards a G4 structure, giving an insight into its putative mechanism of action. A major advantage of studying shorter regions of RNA in isolation is that it makes biophysical characterisation considerably easier, as techniques such as NMR are limited in the size of the macromolecule that is under investigation. Furthermore, longer RNAs are significantly more heterogeneous, giving rise to multiple species in solution which often lead to difficult and ambiguous interpretation of structural and biophysical data. However, the major disadvantage of studying short RNA regions is the physiological relevance of the biophysical observations, as longer RNAs have much more structural diversity, meaning that a structure formed in isolation may not be as favourable when imbedded in its physiological sequence, where a different structure(s) may predominate. Therefore, in this chapter we aim to address this issue, in order to understand the physiological relevance of the observations made in chapters 3 and 4 when contextualising Q2 in longer more physiologically relevant RNAs.

## 5.2 Rationale behind the design of each of the larger Bcl-x constructs

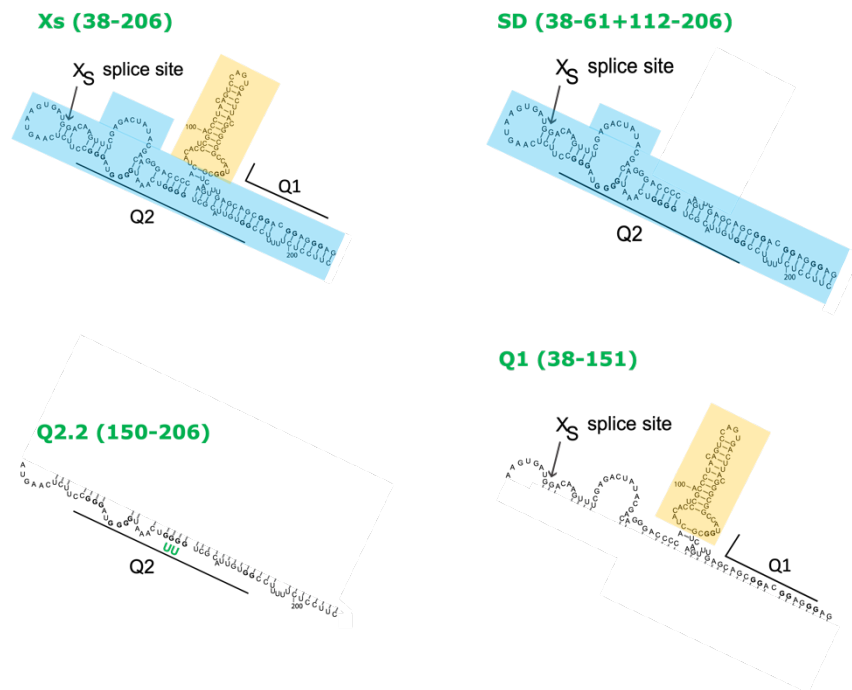
In order to characterise the structure and binding specificity of GQC-05 in larger RNAs, we designed various fragments of Bclx-681 which surround the Q2 region (figure 5.21 A and appendix figure 7.2). The first of the fragments is the X<sub>S</sub> domain which covers nucleotide positions 38-206 of Bclx-681 (highlighted in blue and yellow in figures 5.21 A and B) and is the longest RNA used in this study. Previous foot-printing analysis suggested that this fragment consists of a large stem, with the X<sub>S</sub> 5' ss shown at the apex of this structure. A smaller stem covering positions 62-111 (highlighted in yellow in figures 5.21 A and B) was shown to branch off this larger stem structure. Additionally, this particular domain also consists of both the Q1 and Q2 G-tracts, both of which contain the sequence requirements to fold into a G4 as described previously. The second construct, SD 38-61 + 112-208 in figure 5.21 B (known termed as SD), is similar to the X<sub>S</sub> domain, but contains a deletion of the smaller stem loop at positions

62-111 (highlighted in yellow in figures 5.21 A and B), enabling us to examine the role of this structure in the GQC-05 interaction. The third and fourth constructs (Q2 and 2.2 150-206) consists of the nucleotides immediately downstream of the X<sub>s</sub> 5'ss (which stops just before the start of Q3) and contains the G-tracts belonging to Q2 but not Q1. We also made a mutant version of this construct (Q2.2 150-206), which has the third run of G-tracts within Q2 disrupted by changing a GG to **UU**, thereby abolishing the canonical sequence of a G4 (though a G4 may have the potential to form as seen in chapter 4). The final construct (Q1 38-151) contains the nucleotides upstream of the X<sub>s</sub> 5'ss, consisting of the smaller stem and the G-tracts of Q1 but not Q2. It is important to note that the respective structures of each of these domains will not necessarily be as depicted in figure 5.21 B, since these are the structures that were observed when part of Bcl-x 681 (Weldon et al., 2017). Rather, these are to illustrate which fragments, and hence sequence (shown in appendix table 7.2), of the full length RNA we are isolating for our biophysical analysis. Collectively, these constructs will help us identify the regions within the X<sub>s</sub> domain that are necessary for G4 formation as well as the structural and sequence specific requirements for the GQC-05 interaction. As these RNAs are considerably longer than those studied in chapters 3 and 4, purchasing them commercially by chemical synthesis can be very expensive, particularly considering the quantities needed for biophysical studies. Therefore, we proceeded to synthesise these RNAs by T7 polymerase *in vitro* transcription, which requires the use of a DNA template and an RNA polymerase.

**A**



**B**



**Figure 5.21 Larger Bcl-x constructs around the Q2 region generated for G4 characterization and GQC-05 binding.**

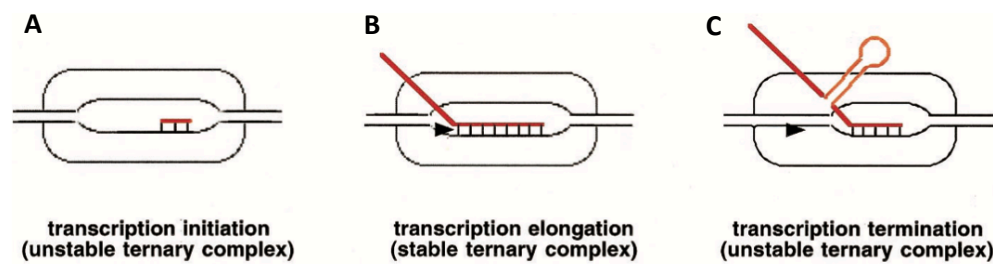
(A) Bclx-681 predicted RNA structure, which the X<sub>s</sub> domain region highlighted in blue and yellow. (B) Structures of the domains shown here are taken from the predicted structure of the larger Bclx-681 RNA (shown in A). The name of each construct is shown above each domain, with the numberings according to Bcl-x-681.

### 5.3 Generating template DNA for transcription of the longer fragments of Bcl-x

To generate large quantities of the RNA constructs described above, we exploited the highly processive properties of the bacteriophage T7 RNA polymerase (T7 RNAP), which was originally isolated in the late 60s following viral infection in *E.coli* cells (Chamberlin et al., 1970). This enzyme is approximately 100 kDa and only consists of a single polypeptide chain with four domains. Similar to the eukaryotic system, transcription by T7 RNAP occurs through 3 stages: initiation, elongation and termination. During the initiation phase, the promotor binding domain of T7 RNAP recognises and binds to the major groove of the promotor in the region -17 to -6 in a highly specific manner, resulting in the formation of the 'closed' complex (Li et al., 1996). The 'open' complex then occurs as a result of conformational changes occurring to the DNA after T7 binding, resulting in the template strand to enter the enzyme active site which is coordinated by a magnesium ion. This results in a transcription bubble occurring between positions -6 to +6, hence the term 'opened state' and is often referred to as promotor melting, which is where the initiation of transcription begins (Li et al., 1996). Before entering the elongation phase, the T7 RNAP synthesises short stretches of RNA before falling off and restarting the initiation phase, resulting in the formation of small RNA abortive products. After the polymerisation of approximately 9-12 RNA nucleotides, the elongating RNA interacts with the N-terminal domain of the T7 RNAP, resulting in a conformational change to the T7 enzyme causing it to enter the elongation phase (Ikeda and Richardson, 1987; Muller et al., 1988). The T7 RNAP then moves along the template strand in a 5' to 3' direction, synthesising the full-length RNA. Termination of transcription often occurs due to the formation of an RNA stem loop on the elongating RNA, thereby reducing the affinity of the T7 RNAP for the double stranded DNA, causing it to dissociate (figure 5.31) (Severinov, 2001). Alternatively, transcription may be terminated through cleavage of the DNA template, which is often known as 'run off' transcription (Severinov, 2001). One major advantage of using the T7 RNAP system is that it has a highly efficient elongation rate of around 230 nt/s, in contrast to bacterial polymerases, such as *E.coli*, which has an elongation rate of just over 10 nt/s (Kochetkov et al., 1998; Wang et al., 1998). Secondly, as the enzyme only consists of a single polypeptide chain, it is significantly easier to express and purify in *E.coli* than other



RNAPs which contain multiple polypeptides, making it an attractive tool for the efficient synthesis of large quantities of RNAs.



**Figure 5.31 Stages in T7 RNAP transcription.**

(A) The initiation complex in the opened conformation, showing production of small abortive RNA products (red). (B) Showing transcription elongation, in which the elongating RNA contacts the T7 RNAP, forming a stable ternary complex. (C) Transcription termination caused by the formation of an RNA hairpin in the elongating transcript (Severinov, 2001).

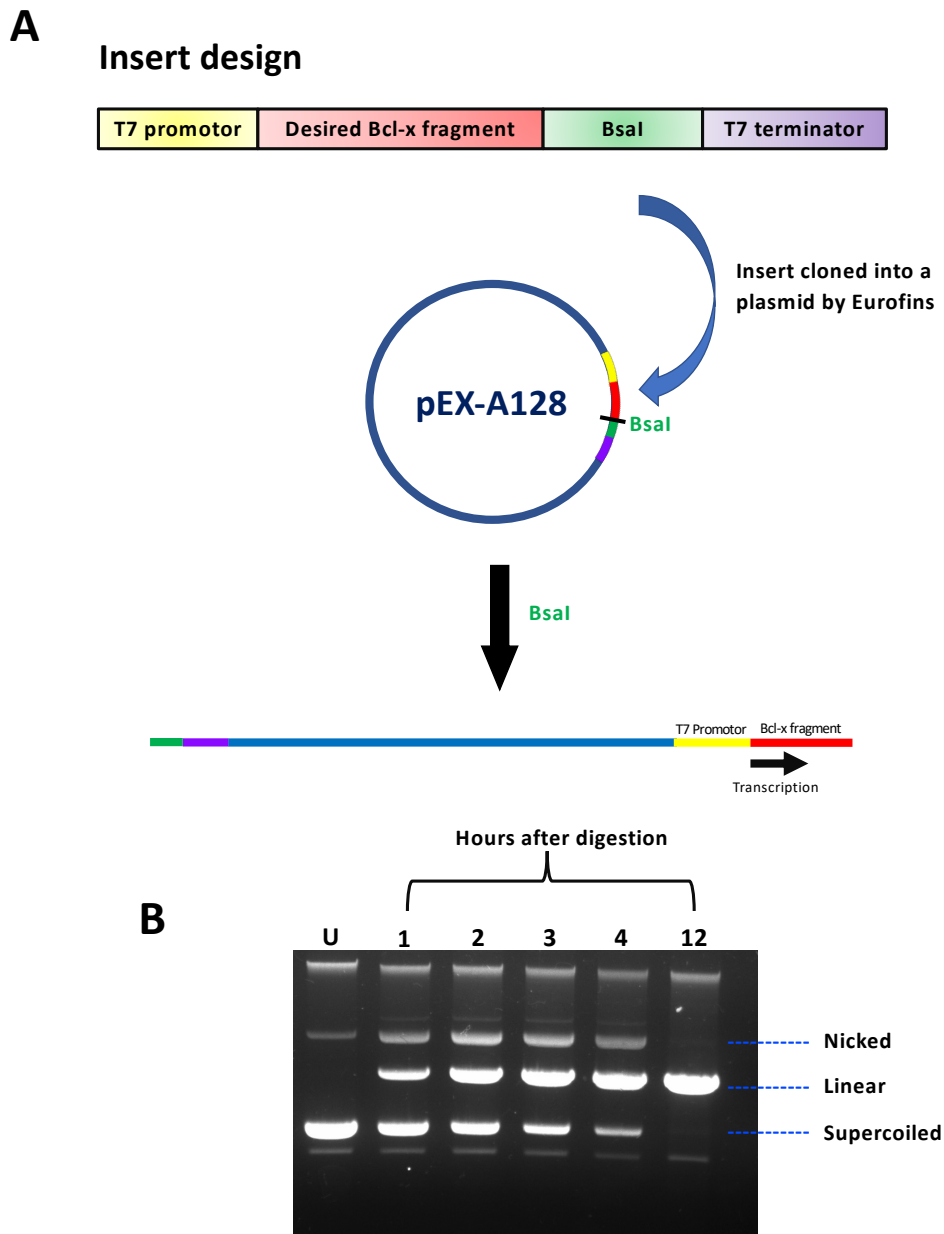
To synthesise the RNA fragments described above, we designed the constructs in a way that would enable the efficient initiation, elongation and termination of the transcription in order to give us high yields of RNA (figure 5.32).

The first stage was to generate large quantities of each of the template DNAs by inserting the template DNA into a plasmid vector to enable the exponential increase in plasmid DNA yields, allowing large supply of the DNA template for large scale transcriptions. The construct we inserted contained the following: a consensus T7 RNAP promotor sequence spanning positions -17 to -1; the desired Bcl-x fragments described in section 5.1; a Bsa1 restriction enzyme site for plasmid linearisation and a T7 terminator sequence (figure 5.32 A). The rationale behind each of these components is discussed below.

As described previously, the T7 RNAP promotor sequence is necessary for the interaction of the enzyme to the DNA template. However, despite the promotor sequence between -17 and -1 not being part of the final RNA transcript, positions +1 (the transcription start site) to +6 (which are incorporated in the final RNA transcript) have been shown to have specific sequence requirements to enable efficient RNA production by T7, as these nucleotides are involved in the initiation of transcription (through formation of the transcription bubble described earlier). Using different DNA starting sequences, a study conducted in the late 80s showed that the best sequence between positions +1 and +6 for generating the highest yield of RNA is GGGATC, with the initial GGG being the most critical component for high yields (Milligan et al., 1987). However, the G-rich nature of this sequence poses a big problem when trying to characterise G4s in these larger transcripts, as the introduction of G-rich artificial

nucleotides at positions +1 to +6 have the potential to generate false positive G4s that would otherwise not be present in the naturally occurring RNA sequence. Therefore, we opted for a different RNA starting sequence containing the nucleotides GAGACT, which results in approximately 50% reduction in RNA yields compared to the optimum sequence, but eliminates the G-tracts thereby minimising the risk of false positive G4 formations. At position +7, the desired Bcl-x sequence is placed. Downstream of the Bcl-x sequence is a Bsa1 restriction enzyme site, allowing linearization of the plasmid to enable run off transcription for termination. The advantage of using Bsa1 compared to other restriction enzymes is that it is able to cleave 5 nt upstream of its recognition sequence, meaning that the nucleotides involved in the restriction site itself are no longer part of the DNA template once cleaved with Bsa1, reducing further incorporation of artificial nucleotides into the final RNA transcript. Downstream of the Bsa1 restriction site we placed a T7 terminator sequence (which forms a stable stem loop upon transcription, terminating elongation), to ensure that transcription comes to a halt even if any undigested plasmid DNA remains, thereby preventing the transcription of the entire 3 kb plasmid DNA. This construct with all the described elements was then inserted into the plasmid pEX-A28 by Eurofins (see appendix 7.1), which we amplified in *E-coli* bacterial cells and proceeded with the digestion and linearisation by Bsa1. To ensure complete digestion of the plasmid DNA by Bsa1, we used agarose gel electrophoresis to look at the observed banding pattern of the plasmid before and after digestion. Figure 5.32 B shows a Bsa1 time-course digestion of the cloned plasmid. The plasmid pEX-A128 does not contain an endogenous Bsa1 restriction site, therefore the cloned plasmid will only have a single Bsa1 site coming from the insert, resulting in a single point of cleavage, generating a distinct linear fragment. Undigested plasmid normally runs with 3 banding patterns which correspond to the nicked, linear and supercoiled conformations (from the lowest to the highest migration speeds respectively). Following digestion with Bsa1, the plasmid loses its supercoil and becomes linear, with a small portion initially being nicked before becoming linear. Between 1-2 hrs after digestion, the proportion of supercoiled plasmid decreases, with a marked increase in the linear and nicked fragments. Between 2-4 hrs, there is a further reduction in the supercoiled plasmid and a further increase in the linear fragments, but also a reduction in the proportion of the nicked conformation, suggesting that Bsa1 is converting the nicked into the linear conformer at these more extensive time periods. After 12 hrs of digestion, almost 100% of the plasmid DNA is in the linear conformation,

indicating near complete digestion, which can now be used in transcription. The long time period of BsaI digestion may reflect in its unique ability in cleaving 5 nt upstream of its recognition sequence, which could reduce its overall efficiency.



**Figure 5.32 Scheme in generating the larger Bcl-x DNA templates.**

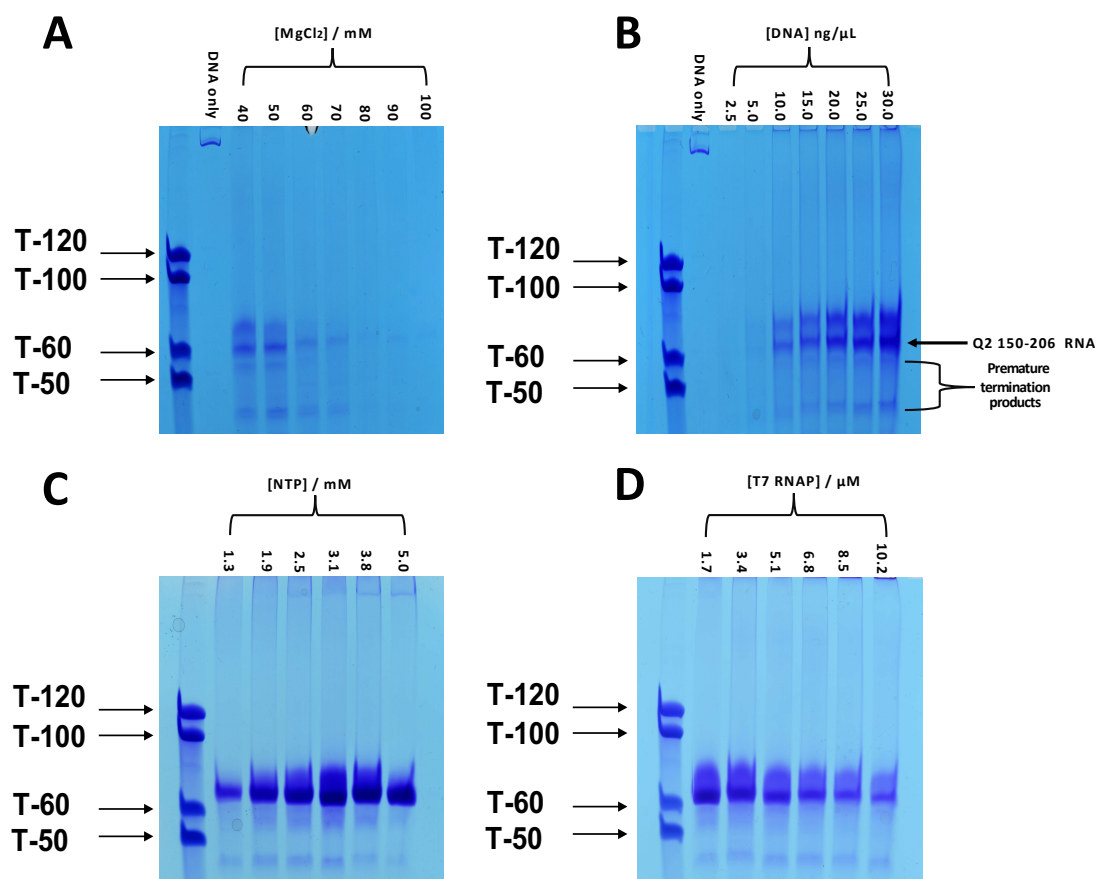
(A) Schematic showing the steps involved in generating the Bcl-x constructs from fragment design to BsaI linearization of the plasmid. (B) Time course of BsaI digestions carried out on the plasmid with the undigested (U) sample used as a control. Plasmids were run on a 1% agarose gel, 90 V and stained with EtBr (1:1000).

## 5.4 Transcription and purification of the large RNAs

### 5.4.1 *In vitro* transcriptions of the large Bcl-x RNAs

Following the successful digestion with Bsa1, we proceeded with transcription optimisation, by conducting small scale (20  $\mu$ L) reactions, in which we individually and sequentially varied the concentrations of each of the following components in the transcription reaction:  $\text{MgCl}_2$ , DNA template, NTPs and T7 RNAP. Following incubation for 2 hours, we determined the best concentration for each component through denaturing PAGE analysis. After identifying the best concentration for one single component, we kept it constant and optimised the next component, until all reagents were optimised. Figure and tables 5.41 and 5.42 show the transcription optimisation gels and conditions for the Q2 150-206 and the X<sub>s</sub> 38-206 constructs, both of which are at the extremes in terms of their relative sizes used in this study, with Q2 150-206 being the smallest and X<sub>s</sub> 38-206 being the largest construct. Due to the size differences between these constructs, I optimised both of these separately, as the optimum concentrations of each of the components, such as the NTPs, may vary (larger fragments may require a greater concentration of NTPs). Interestingly, despite the 112 nt difference between the two constructs, the condition optimisation follow a remarkably similar pattern. In both cases, with increasing  $\text{MgCl}_2$  concentrations, there was a dose dependant reduction in the RNA yield, with the lowest concentration of 40 mM being optimum in this test (though lower concentrations may be more optimum but were not assayed). Despite magnesium being an essential cofactor for T7 RNAP, one possible explanation for the reduction in RNA yield is that higher concentrations of  $\text{MgCl}_2$  is stabilising secondary structures in the DNA template or the elongating RNA, thereby reducing the ability of T7 RNAP to bind and elongate the DNA. Another explanation could be due to magnesium dependant hydrolysis of the RNA which can occur at the higher concentrations of magnesium. This has been reported to be the cause of the reduce yield in a recent paper, where they observed degradation of the RNA at concentrations of  $\text{MgCl}_2$  above 45 mM (Kanwal et al., 2018). Additionally, increasing the ionic strength may reduce protein – DNA interactions, resulting in reduced transcription by T7 RNAP. For both fragments, the opposite pattern was observed with the DNA template concentration, where increasing the concentration from 2.5 - 30 ng/ $\mu$ L resulted in a dose dependant increase in the RNA yield. With the NTPs, a plateau was reached at 3.1 mM, in which no further increase was seen in the RNA yield above this concentration, and

interestingly, despite the differences in the sizes of these two constructs, the optimum concentration of the NTPs remained the same. The T7 RNAP concentration did not have much effect in altering the RNA yields, with the lowest concentration of 1.7  $\mu\text{M}$  being the optimum in both cases. However, with construct Q2 150-206, there does seem to be a slight reduction of the RNA yield at the higher T7 RNAP concentrations. In conclusion, despite the 112 nt discrepancy in the size of the smallest and largest fragment, the optimal conditions for the transcription reaction remained similar. As the remaining two constructs (SD and Q1) have intermediate sizes, I decided to use the same conditions obtained with the Xs and Q2 150-206 fragments. After having identified the optimum condition for both fragments, we then upscaled the reaction to 1 mL and proceeded with HPLC purification optimisations.



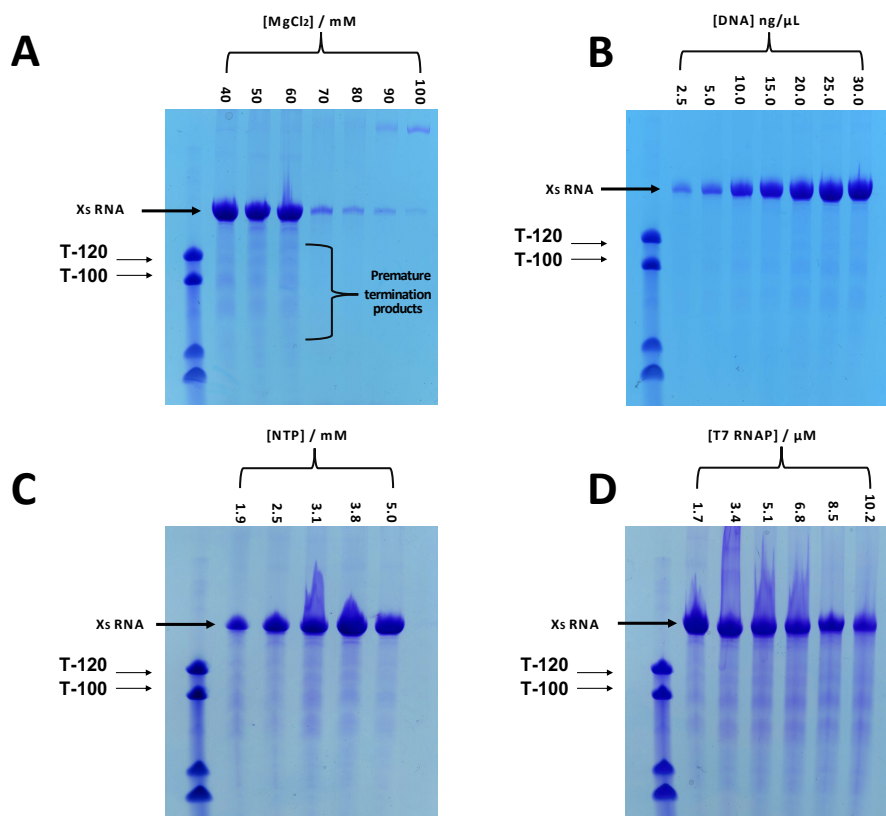
**Figure 5.41 Transcription optimization gels for the Q2 150-206 fragment.**

Optimised for (A)  $\text{MgCl}_2$ , (B) DNA, (C) NTPs and (D) T7 RNAP concentrations. Samples were run on an 8% denaturing PAGE gel at 180 V and stained with toluidine blue.

Reagent	Gel A	Gel B	Gel C	Gel D	Optimised condition
$[\text{MgCl}_2] / \mu\text{M}$	Varied	40	40	40	40
$[\text{DNA}] \text{ ng}/\mu\text{L}$	20	Varied	30	30	30
$[\text{NTP}] / \text{mM}$	2.5	2.5	Varied	3.1	3.1
$[\text{T7 RNAP}] / \mu\text{M}$	8.5	8.5	8.5	Varied	1.7

**Table 5.41**

Concentrations of each reagent used in the four different gels and the final optimised condition.



**Figure 5.42** Transcription optimization gels for the Xs 38-206 fragment.

Optimised for (A)  $\text{MgCl}_2$ , (B) DNA, (C) NTPs and (D) T7 RNAP concentrations. Samples were run on an 8% denaturing PAGE gel at 180 V and stained with toluidine blue.

Reagent	Gel A	Gel B	Gel C	Gel D	Optimised condition
$[\text{MgCl}_2] / \mu\text{M}$	Varied	40	40	40	40
$[\text{DNA}] \text{ ng}/\mu\text{L}$	30	Varied	30	30	30
$[\text{NTP}] / \text{mM}$	3.1	3.1	Varied	3.1	3.1
$[\text{T7 RNAP}] / \mu\text{M}$	1.7	1.7	1.7	Varied	1.7

**Table 5.42**

Concentrations of each reagent used in the four different gels and the final optimised condition.

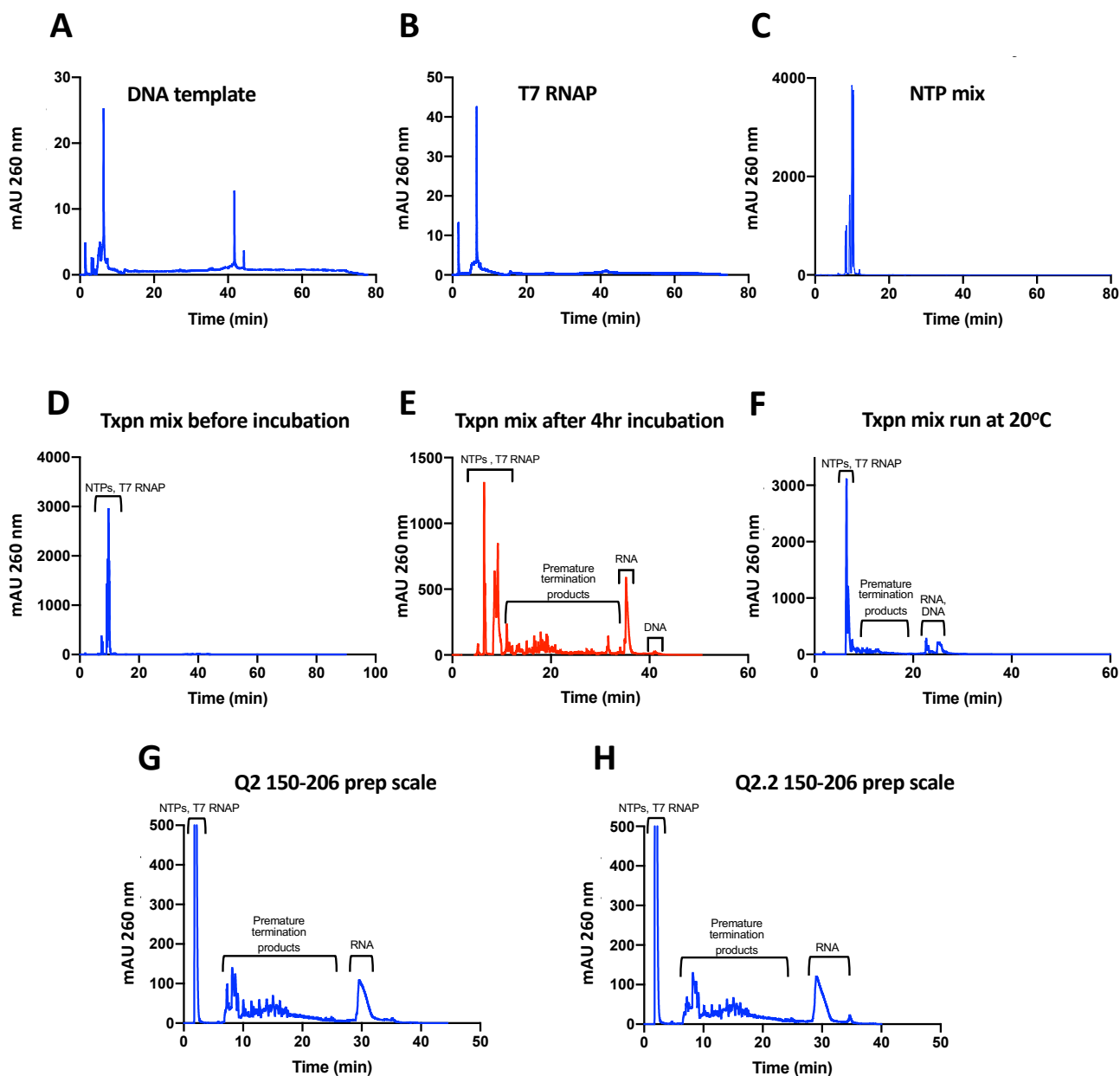
#### 5.4.2 HPLC purifications of the 150-206 Bcl-x RNAs

The large-scale transcription reactions were purified by HPLC using denaturing anion exchange chromatography. The column we used consists of a resin made from diethylamino ethanol (DEAE) cellulose, enabling negatively charged species to be retained on the column and be eluted out using an ionic strength gradient. This has been used previously as an efficient technique to separate out components of the transcription reaction, with the smaller NTPs and uncharged species having a relatively lower retention time than the prematurely terminated RNA products and the full length RNA (Zlobina et al., 2016). By using a gradient of chloride ions (in the form of  $\text{NaClO}_4$ ), we can separate out the different components of the transcription mixture as a function of their charge (which is proportional to their size in the case of nucleic acids), enabling us to purify the RNA of interest from the polymerase and the premature termination product contaminants. Therefore, in order to successfully identify every component in the transcription mix chromatogram, we performed small 20  $\mu\text{L}$  single component injections of each reagent used in the transcription reaction on the analytical anion exchange column (figure 5.43).

Figure 5.43 A shows the chromatogram for the digested plasmid DNA template (at 30  $\text{ng}/\mu\text{L}$ ) containing the Q2 150-206 fragment run at 1  $\text{mL}/\text{min}$ . At approximately 5 mins (which corresponds to 1 column volume) a large peak is observed, followed by another smaller peak at around 40 minutes. The larger peak eluted at 5 minutes is likely coming from the Bsa1 restriction enzyme following plasmid digestion, and the latter smaller peak is the linearised plasmid DNA. With the T7 RNAP injection (figure 5.43 B), a large peak is observed at 5 minutes which is consistent with the T7 RNAP protein being uncharged relative to nucleic acids. Similarly, the NTP mix also gave rise to a large peak at 1 column volume (figure 5.43 C). Despite being negatively charged, they are small enough to behave as an uncharged species (due to fewer phosphate groups compared with longer nucleic acid sequences) and therefore have a comparably low retention time. Next, we wanted to identify the retention time of the RNA, and we tested this by running the HPLC of the transcription mix before and after incubation at 37°C (figures 5.43 D and E). Before incubation, the transcription mix gave rise to a large peak at 1-2 column volumes (at approximately 5-10 minutes), which corresponds to each of the transcription components (DNA template, T7 RNAP and the NTP mix) (figure 5.43 D). Following an incubation period of 4 hours, we observed a more than 50% reduction in the



peak at 1-2 column volumes, and the appearance of peaks at later retention times, with a major peak around 37 minutes (figure 5.43 E). Based on the transcription gels shown in figures 5.41 and 5.42, we could unambiguously assign this major peak as the fully-transcribed RNA. This observation is also consistent with the reduction of the peak seen at 1-2 column volumes after the 4 hour incubation, as during the transcription reaction, free NTP concentration reduces. However, with the gradients used in this protocol, the DNA template and the RNA have very similar retention times, which runs the risk of potential co-purification when upscaling to larger volumes where the resolution may be lower. Therefore, in all subsequent purifications, the samples were pre-treated with DNase-I for 30 minutes before loading onto the HPLC, such that the plasmid DNA is hydrolysed into its dNTP constituents. A major advantage of using an anion exchange over a size exclusion chromatography (SEC) column is that we can perform the experiments at high temperatures, which allows denaturing conditions without the use of urea. Figure 5.43 F shows a chromatogram of a transcription mixture run at 20°C as opposed to 80°C in figure 5.43 E. From this chromatogram, there is not a distinct peak corresponding to the RNA transcript (as in figure 5.43 E), making it difficult to identify the major RNA product. This is consistent with the fact that at these lower temperatures, the RNA (as well as the premature termination products) can adopt multiple different conformations, giving rise to an array of different retention times depending on the shape of the RNA, making it difficult to distinguish between the RNA product and the smaller premature termination product fragments. Therefore, denaturing conditions are essential for the successful purification of the major RNA transcript from the transcription mixture. Having successfully identified the RNA product for fragments Q2 150-206, we proceeded with a large-scale purification (0.5 mL injections) of both the Q2 and Q2.2 150-206 fragments, with their respective chromatograms (figures 5.43 G and H). Interestingly, the retention time of the RNA is approximately 5 minutes earlier as compared to the smaller scale purification. The lower retention time is likely to be the result of a reduction of the number of binding sites for the RNA to the column resin, as there is now more material which will compete with the RNA for binding, hence molecules are displaced sooner, but their relative positions on the chromatogram remain constant.

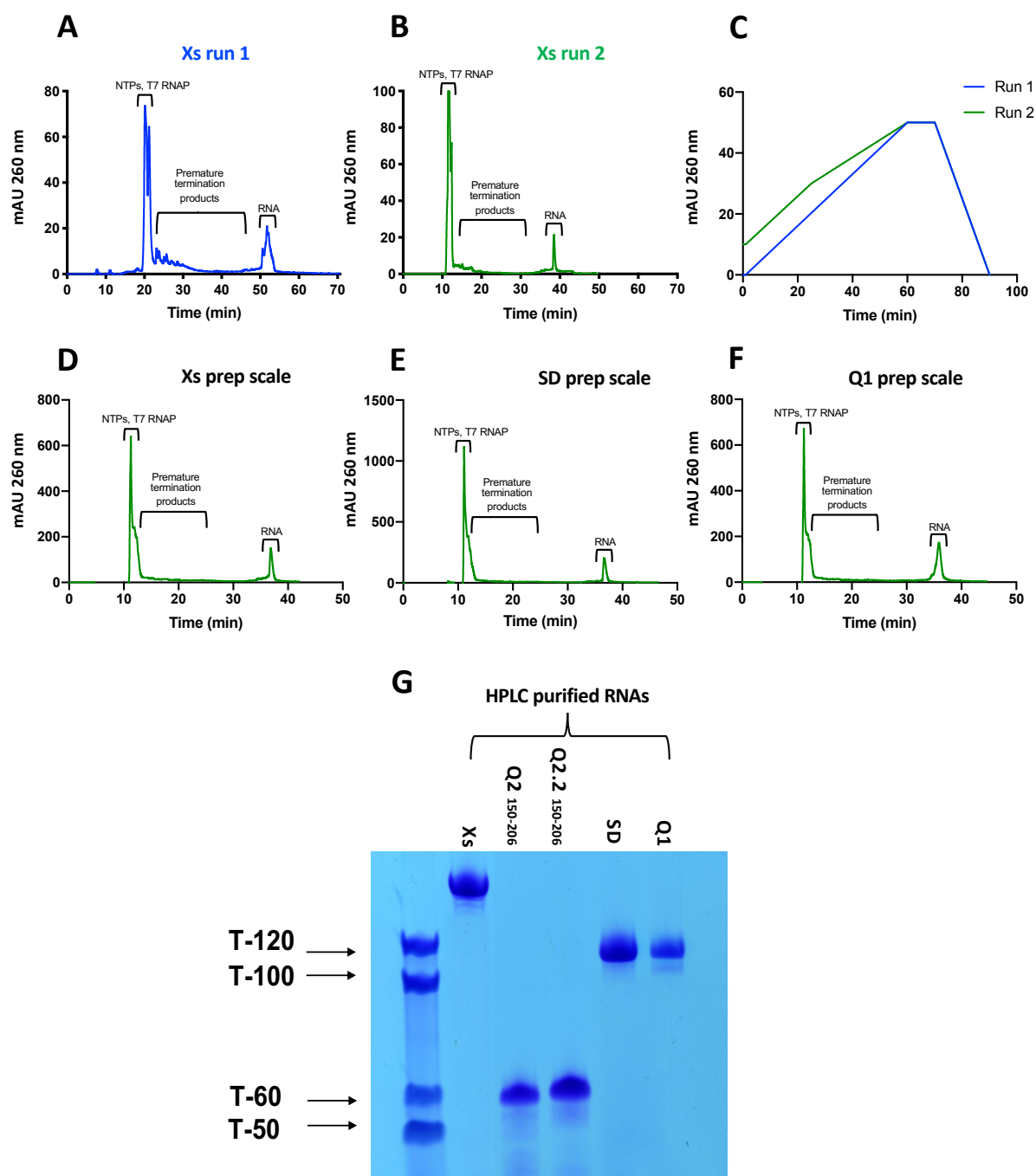


**Figure 5.43 Analytical anion exchange chromatography purification optimisation on Q2 150-206.**

(A-C) Single component injections of each transcription reagent, (A) DNA template, (B) T7 RNA polymerase (T7 RNAP) and (C) NTP mix. (D-F) Injection of the transcription mix before incubation (D), after 4hrs incubation (E), and after 4hrs with the column at 20°C (F). (G-H) Preparative scale purification of: Q2 150-206 (G) and Q2.2 150-206 (H) RNAs. All experiments were carried out in 12.5 mM Tris HCl pH 7.4 with a gradient of 1 M NaClO<sub>4</sub> up to 50%, flow rate of 1 mL/min with a pre-heated analytical DNA pac PA-100 column at 80°C (except for figure F in which the column was at 20°C). Scales on panels A-H are different due to the vast differences in concentrations between each of the components, but the relevant peaks are labelled for cross reference.

### 5.4.3 HPLC purifications of X<sub>s</sub> Q1 and SD Bcl-x RNAs

In a similar approach to the transcription optimisation gels described above, we next investigated how the largest construct (X<sub>s</sub> 38-206) would behave on ion exchange chromatography, as this is significantly larger and heterogeneous than the shorter 150-206 constructs. Figure 5.44 A shows the chromatogram of the X<sub>s</sub> domain with the same gradient used for the purification of the 150-206 constructs. The run-off RNA peak is clearly distinguishable from the prematurely terminated RNA products. However, as these RNAs are larger in size, and with the consideration that we are using a preparative flow rate with these constructs at 3mL/min, waiting approximately 50 minutes for the RNA to elute would use up nearly 150 mL of mobile phase per injection. I therefore modified the gradient used from the previous protocol (run 1 in figure 5.44 C) by starting at a higher percentage of NaClO<sub>4</sub> (10%) at time zero as compared to 0 % with the previous protocol. This will result in the elution of the NTPs, T7 RNAP, DNA template and premature termination products much earlier and together, causing the entire chromatogram to shift to the left, thereby reducing the retention time of the RNA and hence the run time of the experiment. The chromatogram obtained for the X<sub>s</sub> fragment with this new gradient shows an approximately 10 minute reduction in the retention time of the RNA whilst still maintaining sufficient resolution to distinguish the major RNA product (figure 5.44 B). I therefore chose this gradient to do the subsequent large-scale purifications on the remaining constructs (X<sub>s</sub>, Q1 and SD) (figures 5.44 D-F). Despite the size differences between these three constructs, their retention times are almost identical. To confirm that the X<sub>s</sub>, Q1, SD and the 150-206 constructs are of the expected sizes, we eluted, dialysed and loaded the major RNA peaks of each chromatogram on an 8% de-naturing PAGE gel (figure 5.44 G). X<sub>s</sub> being the largest construct of 175 nt, displayed a band above the 120 nt marker. SD and Q1 are very similar in size of 125 and 120 nt respectively, making them run almost identically to the T-120 marker. Finally, the 58 mer 150-206 fragments also migrated at the expected size of around T-60. Collectively, this confirms that we have correctly transcribed and purified each of the Bcl-x RNA fragments using an HPLC anion exchange column, enabling us to confidently conduct biophysical experiments.



**Figure 5.44 Preparative anion exchange chromatography purifications of X<sub>s</sub>, SD and Q1.**

(A and B) Showing X<sub>s</sub> chromatograms using two different gradient patterns shown in (C). (C) Gradients of NaClO<sub>4</sub> for runs 1 and 2: Run 1 starts at 0%, then a linear gradient from 0-50% in 60 minutes. Run 2 starts at 10%, then linear from 10-30% for 20 minutes then 30-50% in 40 minutes. (D-F) Preparative scale purification of: X<sub>s</sub> (D) and SD (E) and Q1 (F) RNAs using gradient run 2. (G) 8% de-naturing PAGE of each construct after HPLC purification and dialysis in H<sub>2</sub>O. All HPLC experiments were carried out in 12.5 mM Tris HCl pH 7.4 with a gradient of 1M NaClO<sub>4</sub> up to 50% using a flow rate of 3mL/min with a pre-heated preparative DNA pac PA-100 column at 80°C. Scales on panels A-F are different due to the vast differences in concentrations between each of the components, but the relevant peaks are labelled for cross reference.

## 5.5 G4 characterisation on the larger Bcl-x RNAs

Having previously identified the presence of a G4 in the Q2 region, the next step is to examine if a G4 can still form within Q2 when imbedded in the X<sub>S</sub> domain. As discussed previously, foot-printing data on the full length Bcl-x 681 RNA shows that there are large changes occurring in the Q2 region when comparing wild-type and 7-deaza guanine substituted RNA, implying that a G4 still has the potential to form in this region. In this section, we will use biophysical data to either support or disprove this observation.

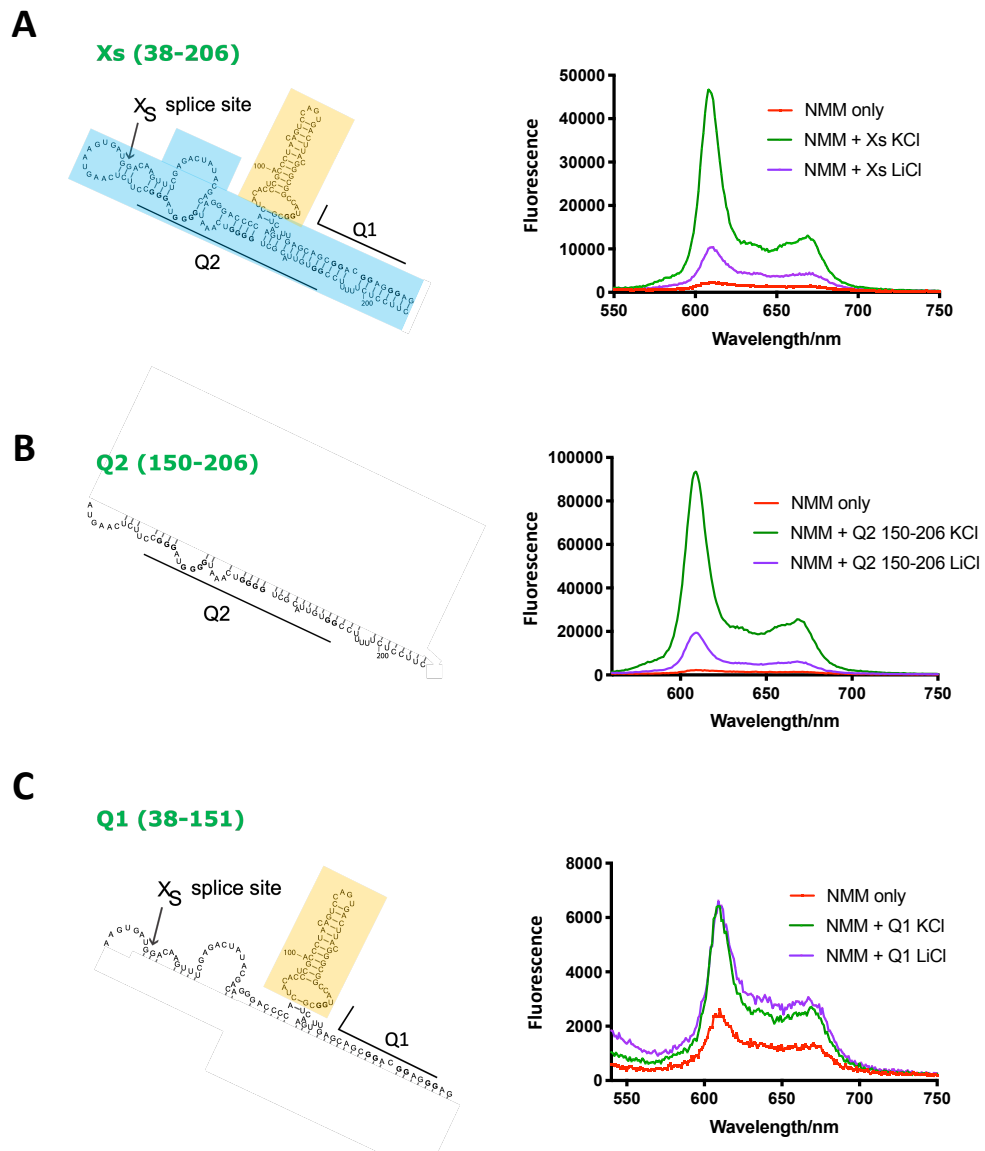
### 5.5.1 Fluorescence spectroscopy to characterise a G4 in the longer RNAs

Unlike NMR, steady state fluorescence spectroscopy does not have a limit on the size of the macromolecule that can be studied. We therefore chose this initial approach to evaluate the presence of a G4 in the X<sub>S</sub> 38-206 domain. As described previously, NMM was shown to be a highly specific binder of parallel type quadruplexes and was successfully used to identify the presence of a G4 in the Bcl-x Q2 RNA. We therefore chose this approach to identifying a putative G4 in the X<sub>S</sub> domain (figures 5.51 A – C).

Figure 5.51 A shows the NMM emission spectra in the absence or presence of the X<sub>S</sub> domain either in KCl (green) or LiCl (purple). NMM unbound (red) has a relatively low fluorescence emission as expected, however there is an approximately 5-fold increase in the NMM emission in the presence of KCl compared with LiCl, suggesting that a G4 can exist in the X<sub>S</sub> domain. This domain contains two possible regions that have the ability to fold into a G4: the G-tracts of Q1 which lie upstream of the 5'ss, and the G-tracts of Q2 which lies downstream of the 5'ss. Therefore, to try and narrow down the region(s) within the X<sub>S</sub> domain that are responsible for the G4 formation, we tested the two halves of the X<sub>S</sub> domain through the constructs Q2 150-206, which lies downstream of the 5'ss and contains the G-tracts of Q2 only, and Q1 38-151, which is the sequence upstream of the X<sub>S</sub> 5'ss containing the G-tracts of Q1 only shown in figures 5.51 B and C respectively. For the Q2 150-206 fragment, we also see an approximately 5-fold enhancement of the NMM emission in the presence of KCl compared with LiCl, suggesting that a G4 still exists between positions 150-206. However, in contrast, the Q1 fragment did not show any fluorescence enhancement of NMM in the presence of KCl relative to LiCl, suggesting that this region does not fold into a G4. Therefore, from this data, we can conclude that a G4 can form in the X<sub>S</sub> domain, and this is more than

likely coming from the G-tracts of Q2 rather than Q1, supporting the foot-printing data (Weldon et al., 2017). It is important to note that the Q2 G4 found within the X<sub>5</sub> domain may not exhibit the same conformation as those characterised on the shorter fragments discussed in chapter 4, as the presence of additional nucleotide sequences (and competing secondary structures) may result in a different minimum free energy of the G4 structure. This is also true when comparing the structure of the G4 in the X<sub>5</sub> domain and the same stretch of sequence imbedded in Bcl-x 681, highlighting that biophysical characterisation of structures with fragmented RNAs should always be taken with caution.

With regards to the fluorescence emission, we did not observe the expected 50-60 fold fluorescent enhancement of NMM as was previously described when bound to DNA G4 structures (Nicoludis et al., 2012a). This may be due to changes in the fluorescent properties of NMM when bound to RNA G4s, though this will require further investigation using well characterised RNA G4s, such as TERRA.



**Figure 5.51 NMM emission spectra on and around the X<sub>s</sub> domain.**

(A-C) NMM emission spectra on X<sub>s</sub> (A), Q2 150-206 (B), and Q1 (C) in the presence of 100 mM KCl or LiCl, with their respective structures depicted on the left. As we are comparing relative changes in the fluorescence intensity between KCl and LiCl in graphs A-C, the Y axis scaling is different in each case. Experiments were carried out with 1  $\mu$ M NMM in 10 mM Tris pH 7.0, 100 mM salt ( $\lambda_{\text{exc}} = 393\text{nm}$ ).

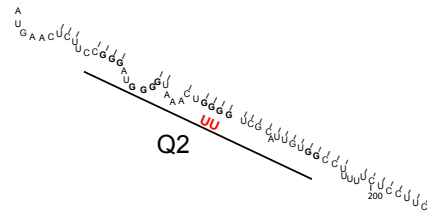
### 5.5.2 CD spectroscopy to characterise a G4 in the longer RNAs

We next investigated whether the G4 could exist even in the absence of NMM on Q2 150-206 and the Xs 38-206 constructs, as NMM may be artificially inducing and stabilising a G4 which would otherwise not be present. For this we used CD spectroscopy in the presence of either KCl or LiCl (figure 5.52).

Figures 5.52 A and B shows the CD spectra recorded at 20°C (figure 5.52 A) and the thermal melt for the Q2 150-206 construct (figure 5.52 B) in the presence of KCl (blue) or LiCl (red). There are significant conformational differences in the presence of potassium and lithium, as indicated by non-overlapping CD spectra. Furthermore, in a similar way to Q2, there is a K<sup>+</sup> dependence in the thermal stability of this RNA, which shows a 21°C increase in the melting temperature in the presence of KCl compared with LiCl. In contrast, with the G4 mutation Q2.2 150-206, the CD spectra at 20°C are overlapping in both buffers (figure 5.52 C), and there are no differences in the thermal stability of the mutant RNA (figure 5.52 D). Collectively, this suggests that a G4 has the ability to form in KCl-containing buffers for the fragment Q2 150-206.



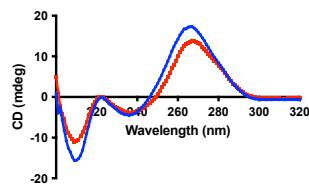
**Q2 (150-206)**  
**Q2.2 (150-206)**



— 100 mM KCl  
— 100 mM LiCl

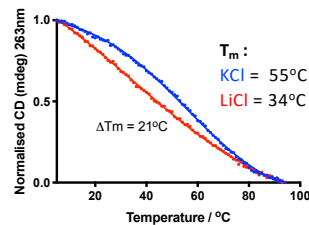
**B**

**Q2 150-206**



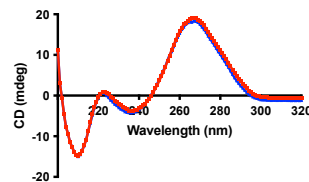
**C**

**Q2 150-206**



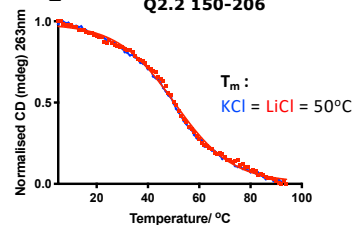
**D**

**Q2.2 150-206**



**E**

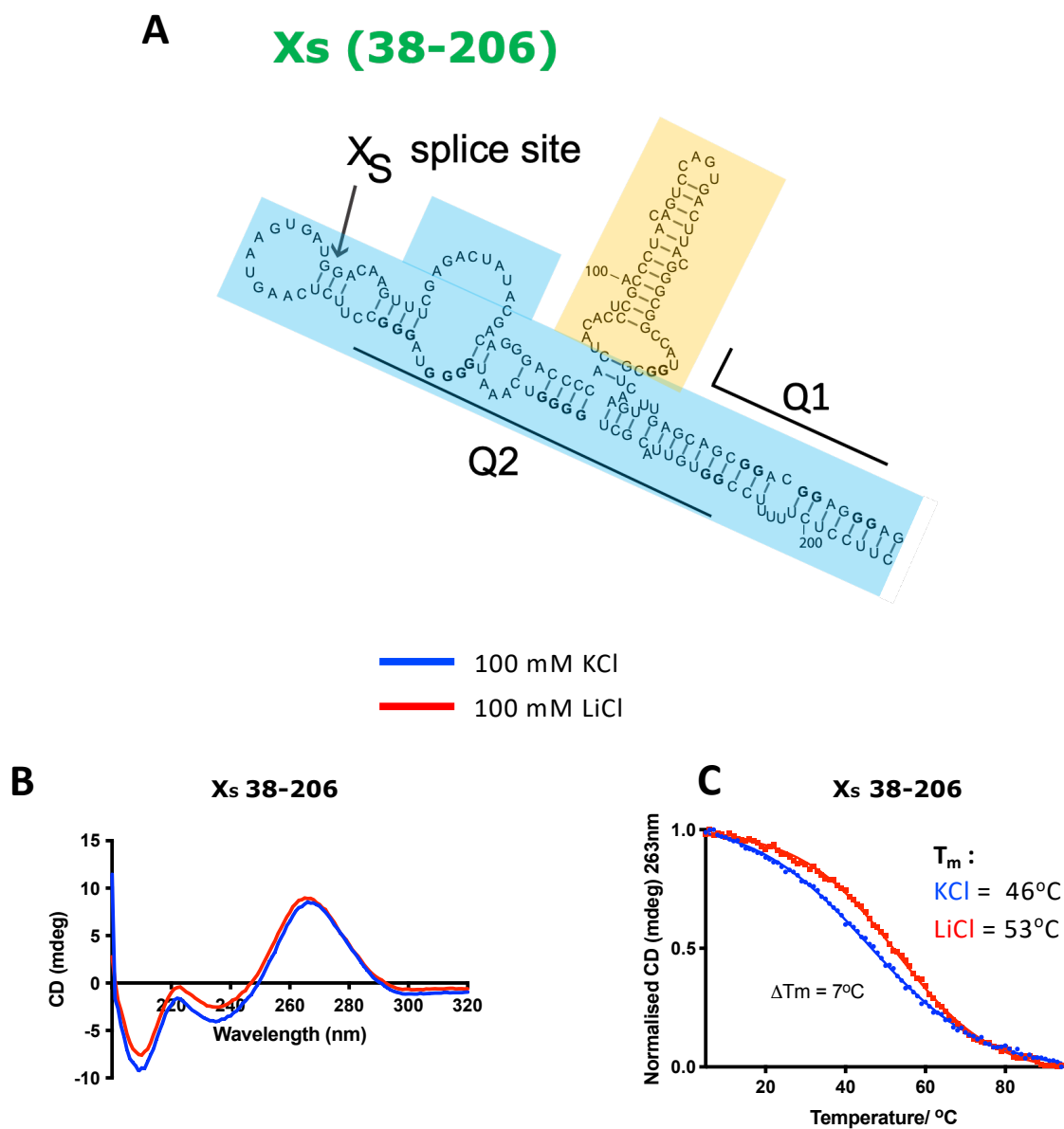
**Q2.2 150-206**



**Figure 5.52 CD spectra and thermal melts of the 150-206 fragments.**

(A) Structure of the Q2 150-206 construct, and highlighted in red the Q2.2 150-206 change. CD and thermal melt in the presence of KCl (blue) and LiCl (red) of Q2 150-206 (B-C) and Q2.2 150-206 (D-E). All experiments were carried out in 10 mM TBA phosphate buffer pH 7.0 with 100 mM salt in a 1 cm pathlength quartz cuvette.

Similarly, we tested by CD spectroscopy the presence of a G4 in the X<sub>S</sub> domain. At 20°C, we do not see overlapping CD signatures in the presence of KCl or LiCl, with bigger changes occurring at wavelengths below 250 nm, suggesting that there are conformational differences of this RNA in KCl and LiCl-containing buffers (figure 5.53 A). The negative peak at around 210 nm is unique to an RNA duplex, and this region is being affected in the presence of the different buffers, whilst the signal at 260 nm (corresponding to duplex and G4) is almost not affected by the buffer. As X<sub>S</sub> is considerably larger than the 150-206 fragments, sample heterogeneity is likely to be much more prominent, as we predict that the X<sub>S</sub> RNA to have two competitive conformations (stem-loop versus G4). The CD spectrum shows an average of all of the structure's adopted by the RNA in solution, and cannot differentiate between stem-loop and parallel G4s. This could mean that changes in secondary structure induced by the presence of different salts may only give rise to subtle differences in the CD spectrum, which may explain why the CD signal at 260 nm is not being effected by the buffer. However, the larger changes that are observed at 210 nm suggests that the RNA duplex is adopting different structures in KCl and LiCl. Perhaps a G4 induced by the presence of KCl is re-modelling and changing the structure of the duplex surrounding this newly formed G4. Interestingly, with the thermal melting experiments (figure 5.53 B), the stability of the X<sub>S</sub> domain is lower in the presence of KCl compared with LiCl (with a difference of 7°C), contrasting with what we have observed on the shorter G4 fragments. This observation is consistent with the idea that in the longer RNA, there are more Watson-Crick hydrogen bonds than those in the shorter fragments (consistent with the foot-printing data, showing that X<sub>S</sub> adopts a large stem-loop structure), and that stabilising a G4 through Hoogsteen base pairings would result in a much larger enthalpic cost to the surrounding duplex structure, as more Watson-Crick hydrogen bonds will need to be broken to form the G4 structure in the presence of KCl. Therefore, stabilising a G4 in this RNA has the effect of destabilising the rest of the surrounding structure, hence a reduction in the T<sub>m</sub> in the presence of KCl. In summary, these data suggest that this RNA sequence can adopt a G4 (likely forming with the Q2 but not the Q1 G-tracts), but that it is in competition with a more stable duplex structure. This is fully consistent with the model previously described (Weldon et al.,2018).



**Figure 5.53 CD spectra and thermal melts of the X<sub>S</sub> 38-206 domain.**

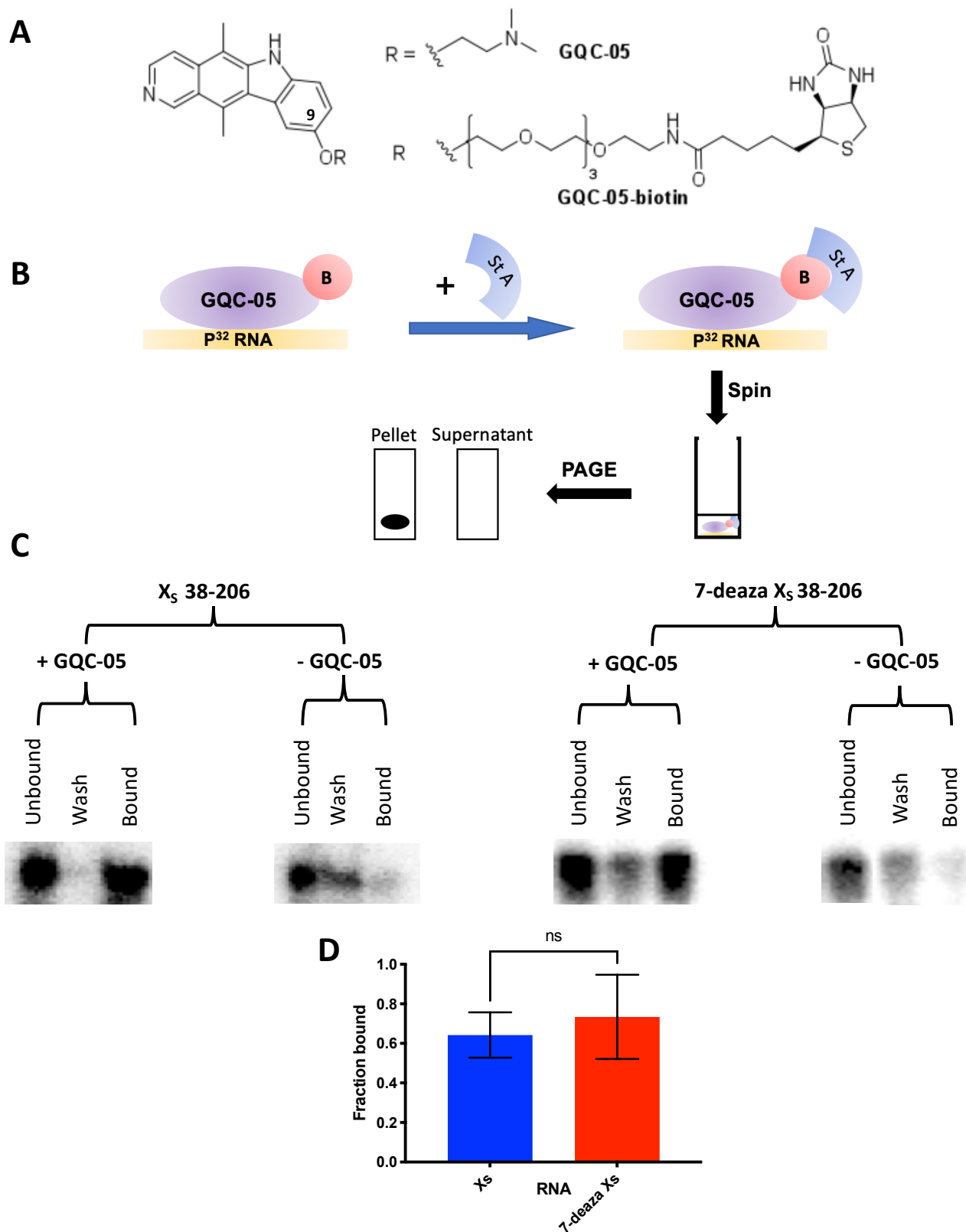
(A) Structure of the X<sub>S</sub> domain. CD (B) and thermal melt (C) in the presence of KCl (blue) and LiCl (red) of the X<sub>S</sub> domain. All experiments were carried out in 10 mM TBA phosphate buffer pH 7.0 with 100 mM salt in a 1cm pathlength quartz cuvette.

## 5.6 Important structural elements for the specificity of the GQC-05 interaction

Knowing that a G4 could still form in the X<sub>s</sub> domain, we next tested the binding of GQC-05 to this larger fragment, in order to determine the critical structural elements necessary for its interaction.

### 5.6.1 Biotin pull-down assays to probe the structural specificity of GQC-05

To determine if the G4 in the X<sub>s</sub> domain is necessary for the GQC-05 interaction, we performed pull down experiments with a biotinylated derivative of GQC-05 (bio-GQC-05) synthesised by the lab of Professor Glenn Burley at the University of Strathclyde. Bio-GQC-05 differs from the normal compound through an addition of a biotin tag at position 9 via a PEG linker, instead of the tertiary amine which is present on the normal GQC-05 (figure 5.61 A). We incubated bio-GQC-05 with the <sup>32</sup>P-labeled X<sub>s</sub> RNA and pulled down the bio-GQC-05-RNA complex using streptavidin beads. After several washes, the beads were loaded onto a denaturing PAGE gel to detect any RNA bound to the biotinylated compound (figure 5.61 B). In order to determine if the G4 in the X<sub>s</sub> domain is important for the GQC-05 interaction, we also synthesised a 7-deazaguanine-substituted RNA, in which the N<sub>7</sub> position of the guanine is changed to a carbon, abolishing G4 formation through disruption of the Hoogsteen base pairs whilst still retaining its Watson-Crick potential. For the wild-type sequence, there was a large proportion of GQC-05 bound, yielding a fraction bound ratio of about 0.7. Interestingly, the 7-deaza RNAs showed a remarkably similar level of pull down, with a fraction bound ratio of about 0.75, showing no significant difference to the WT sequence (figure 5.61 C and D). This suggests that the G4 in the X<sub>s</sub> RNA is not a key structural element in the GQC-05 interaction.



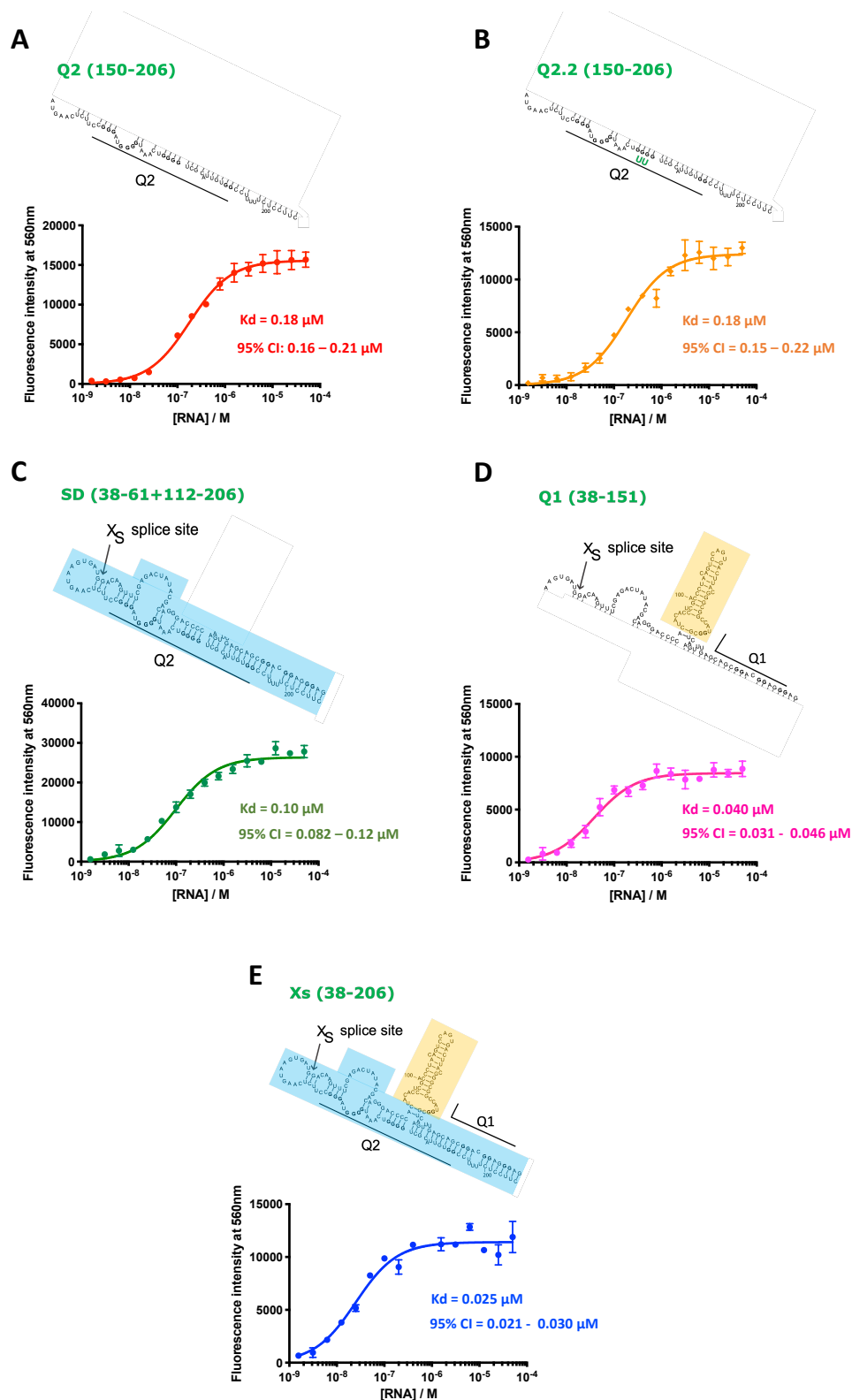
**Figure 5.61 Biotin GQC-05 pull downs of the Xs domain.**

(A) Core structure of ellipticine with the showing the position in which the biotinylated tag is attached. (B) Schematic showing the bio-GQC-05 pull down assay. (C) GQC-05 biotin pull downs on  $^{32}\text{P}$ -GTP labelled X<sub>s</sub> or 7-deaza X<sub>s</sub> RNAs on a 8% de-naturing PAGE gel. (D) Histogram comparing the fraction bound of the X<sub>s</sub> and 7-deaza X<sub>s</sub> RNA. Error bars indicate the standard error of the mean (SEM) and were done in duplicate.

### 5.6.2 Fluorescence titrations to identify key RNA structural elements in the GQC-05 interaction

The disadvantage of pull down assays is that very subtle changes in binding affinity which may be significant may not be detected. We therefore proceeded with fluorescence binding titrations in a similar way to those described in chapter 4, using GQC-05 and each of the fragments generated to determine the binding affinity and how this may correlate to the structural features present.

Figure 5.62 shows the fluorescence titrations for each of the indicated constructs using the bound emission of GQC-05 at 560 nm. For constructs Q2 and Q2.2 150-206, an almost analogous binding affinity is observed of 0.18  $\mu\text{M}$  (figure 5.62 A and B), suggesting that GQC-05 cannot discriminate between the structures of Q2 and Q2.2 150-206 constructs. This observation is consistent with the fluorescence binding assays described in chapter 4, which showed that the G4 is not relevant for the GQC-05 interaction as binding is also observed in the shorter Q2.2 and the  $\Delta\text{G4}$  mutated fragments. A similar binding affinity ( $K_d$  of 0.1  $\mu\text{M}$ ) is also observed for the SD fragment (figure 5.62 C). However, with the Q1 and  $X_s$  fragments, an approximately 10-fold higher affinity is observed, with measured  $K_d$ s of 0.04 and 0.025  $\mu\text{M}$  respectively (figure 5.62 D and E). The higher affinity observed with the Q1 fragment as opposed to the Q2 fragment also highlights that GQC-05 is not a G4 specific compound, as we described earlier that Q1 cannot fold into a G4. Interestingly, both the Q1 and  $X_s$  RNAs, which display the highest affinity interactions, contain the additional stem at position 62-111 (highlighted yellow in figure 5.21), whereas the 150-206 and the SD fragments do not have this stem and display a 10-fold lower affinity. This suggests that the nucleotides between 62-111 may be important in the GQC-05-RNA interaction. Further mutational assays and structural characterisations would need to be performed on this short stretch to understand its relevance in the interaction with GQC-05.



**Figure 5.62 : Fluorescence emission titrations at 560nm of each of the indicated Bcl-x constructs with GQC-05.**

(A) Q2 150-206, (B) Q2.2 150-206, (C) SD, (D) Q1 and (E) Xs domains. Experiments were conducted in 10 mM Tris pH 7.0, 100 mM KCl, 1  $\mu\text{M}$  GQC-05 ( $\lambda_{\text{Exc}} = 320 \text{ nm}$ ).

### 5.6.3 Binding of GQC-05 to the larger RNAs in the presence of a nuclear extract

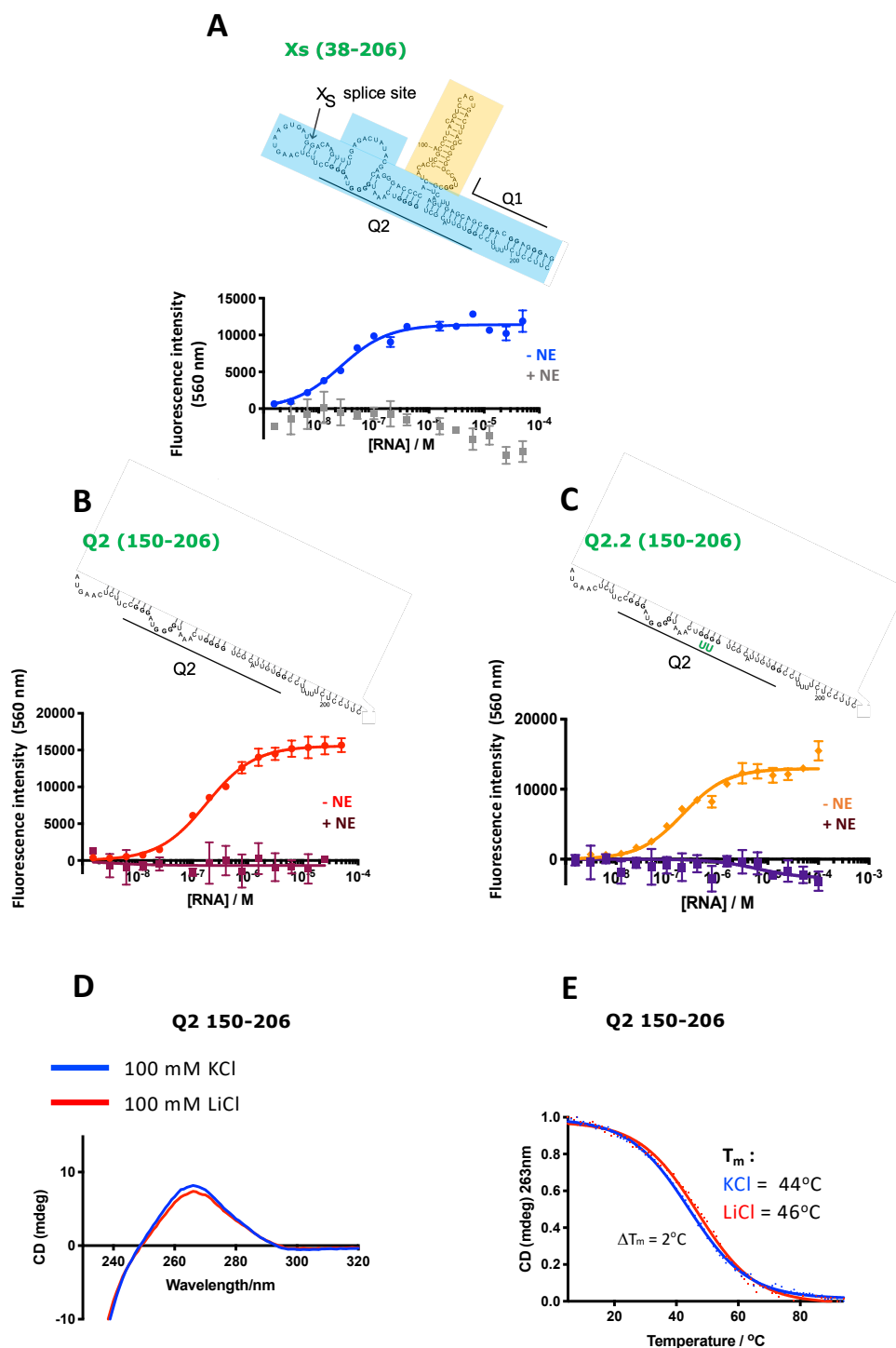
Using the shorter RNA fragments used in chapter 4, we have already shown that GQC-05 binds to RNAs that have both high and low G4 forming potential, but in the presence of a nuclear extract, it only binds to the WT Q2 G4-forming RNA (which had the highest G4 forming potential). Therefore, we next tested the effect of nuclear extract on the binding specificity of GQC-05 to these larger fragments.

Figure 5.63 A-C shows the GQC-05 fluorescence binding titrations of the X<sub>s</sub> and the 150-206 fragments in the presence and absence of nuclear extract. Interestingly, with the X<sub>s</sub> and Q2 150-206 construct, in which a G4 was shown to form in Tris buffer without the addition of nuclear extract, addition of nuclear extract completely abolished the binding GQC-05. Similarly, the mutant Q2.2 also showed the same effect. This is very much in contradiction to the results described in chapter 4, in which the shorter Q2 fragment retained binding of GQC-05 in the presence of nuclear extract. Therefore, we checked whether a G4 can still exist in the Q2 150-206 RNA in the presence of nuclear extract, using CD spectroscopy (figure 5.63 D and E). Unlike in the absence of nuclear extract, these results show that for Q2 150-206, there is very little potassium dependence on the shape of the CD spectra at 20°C or on the thermal stability, suggesting that a G4 may not form in this sequence in presence of a nuclear extract, which could explain why we do not see a binding isotherm of the GQC-05 RNA interaction in these conditions. A possible explanation could be that G4 formation is in competition with RNA binding proteins, such as RNA helicases and hnRNPs F/H, which have been shown to unfold RNA secondary structures (Dominguez et al., 2010; Gao et al., 2019). However further biophysical characterisation, such as using NMR spectroscopy, will need to be conducted to confirm the G4 disruption by nuclear extract.

The question then arises as to what the biological significance of the G4/GQC-05 interaction is in the alternative splicing of Bcl-x if the complex cannot form in functional conditions as the fluorescence binding assay suggests. The *in vitro* splicing assay of Bcl-x shown in figures 1.71 A and 4.42 B shows that 40 µM of GQC-05 is required to obtain maximum activation of the X<sub>s</sub> 5'ss. The RNA concentration used in these splicing assays is approximately 50 nM, which means that there is an 800-fold excess of GQC-05 compared to the RNA for splice site activation. In the fluorescence assay described above, we were using 1 µM of GQC-05 with the RNA being titrated from the nM range to 100 µM, meaning that only at the lowest



concentrations of the RNA we recapitulated the 800-fold excess of GQC-05. Possibly due to the detection limit of fluorescence at these lower concentrations, we were not able to see the bound signal. Unfortunately, due to the limitations of fluorescence spectroscopy, simply increasing the GQC-05 concentration to achieve the same molar ratio as in the splicing assay is impossible due to the inner filter effects as well as limitations of GQC-05 solubility. Therefore, we used a different approach to explore the effect of GQC-05 on the re-modelling of the X<sub>5</sub> RNA. However, due to the limitations described above, all subsequent experiments were conducted in the absence of nuclear extract.



**Figure 5.63** The effect of nuclear extract on the binding of GQC-05 and the formation of a G4 in Xs and Q2 150-206.

(A-C) Fluorescence titrations of Xs (A), Q2 150-206 (B), and Q2.2 150-206 (C) with 1  $\mu\text{M}$  GQC-05 in 10 mM Tris pH 7.0, 100 mM KCl with and without 10% HeLa nuclear extract. (D-E) KCl and LiCl CD spectra at 20°C (D) and thermal melt (E) in the presence of 5% HeLa nuclear extract and 1  $\mu\text{M}$  RNA, conducted in 100 mM salt, 10 mM TBA phosphate buffer pH 7.0, 1 mm pathlength quartz cuvette.

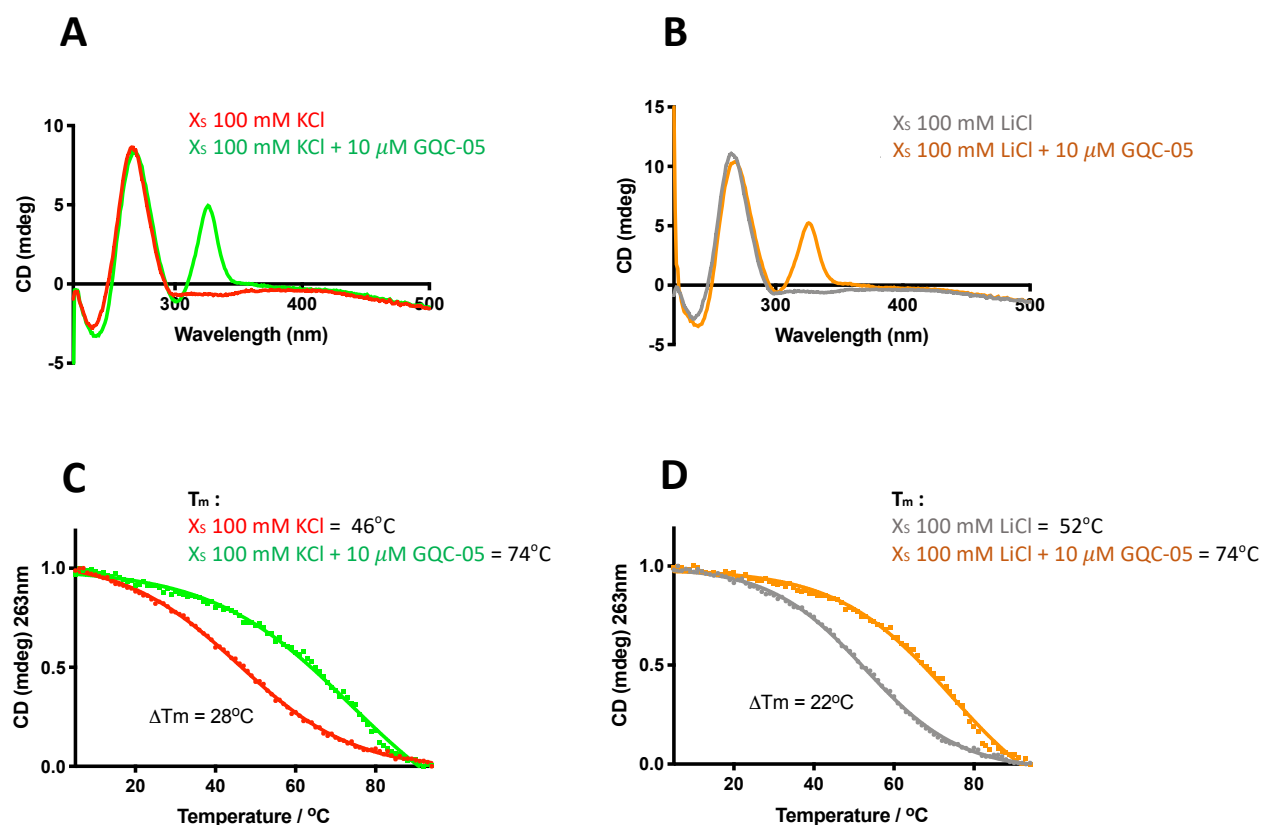
## 5.7 Effects of GQC-05 in the structural remodelling of the X<sub>S</sub> domain

### 5.7.1 CD spectroscopy to monitor the conformational changes occurring with GQC-05

In the previous section, we showed that GQC-05 can bind to multiple different RNA sequences, which include those that may or may not have G4 forming potential (though the exact binding specificity needs to be further investigated). This means that simply looking at the binding affinity and specificity for different structures may not be so conclusive in deciphering its mechanism of action. For example, even though GQC-05 was shown to bind to the Q1 construct, the accessibility of binding sites within this region may alter when in context of the X<sub>S</sub> domain due to structural remodelling of the RNA in the presence of additional nucleotide sequences. Consequently, interpreting binding data by simply looking at individual domains should be taken with caution. We therefore turned our attention to observing the changes that occur to the RNA as a result of GQC-05 binding, and how such binding it may modulate the secondary structure of the X<sub>S</sub> domain to achieve splice site selection.

We measured a CD spectrum at 20°C as well as conducting thermal melts of the X<sub>S</sub> domain in the absence or presence of 10 µM GQC-05, with either KCl (figure 5.71 A and C) or LiCl (figure 5.71 B and D) containing buffers. From the 20°C spectrum for both KCl and LiCl (figure 5.71 A and B), there is a subtle difference in the CD signature in the presence of GQC-05 below 230 nm, similar to the changes seen when comparing X<sub>S</sub> in KCl and LiCl buffers (figure 5.53 above), suggesting that there are some structural changes occurring to the RNA. Furthermore, in both KCl and LiCl buffers, we observe a sharp positive ICD peak corresponding to the increased chirality of GQC-05, suggesting potential binding to the minor groove (see section 4.4). However, as these experiments were conducted in the absence of a nuclear extract, it is likely that GQC-05 is displaying many different modes of binding due to its ability to bind to different regions of the RNA, meaning that the ICD is an average of all the binding modes present in solution (specific and non-specific). Therefore, no further conclusions can be made about the binding mode from the shape of the ICD. The thermal melting experiments in the presence of GQC-05 show an increase in stability of the X<sub>S</sub> domain in both KCl and LiCl-containing buffers, with a change in the T<sub>m</sub> of 28 and 22°C respectively (figure 5.71 C and D).

Collectively these data suggest that there may be changes occurring to the X<sub>5</sub> RNA in the presence of GQC-05, though as the CD spectrum for duplex and G4s are so similar, it is hard to determine whether the changes are specific for either structure. Therefore, another approach is required to clarify the role of GQC-05 in X<sub>5</sub> remodelling.



**Figure 5.71 CD spectra and thermal melts of X<sub>5</sub> in the presence of GQC-05.**

(A and B) CD spectra at 20°C of X<sub>5</sub> in 100 mM KCl (A) or LiCl (B) with and without 10 μM GQC-05. (C-D) Thermal melts with 100 mM KCl (C), or LiCl (D) in the presence and absence of 10 μM GQC-05. All experiments were conducted in TBA phosphate pH 7.0, 100 mM salt using a 1cm pathlength quartz cuvette.

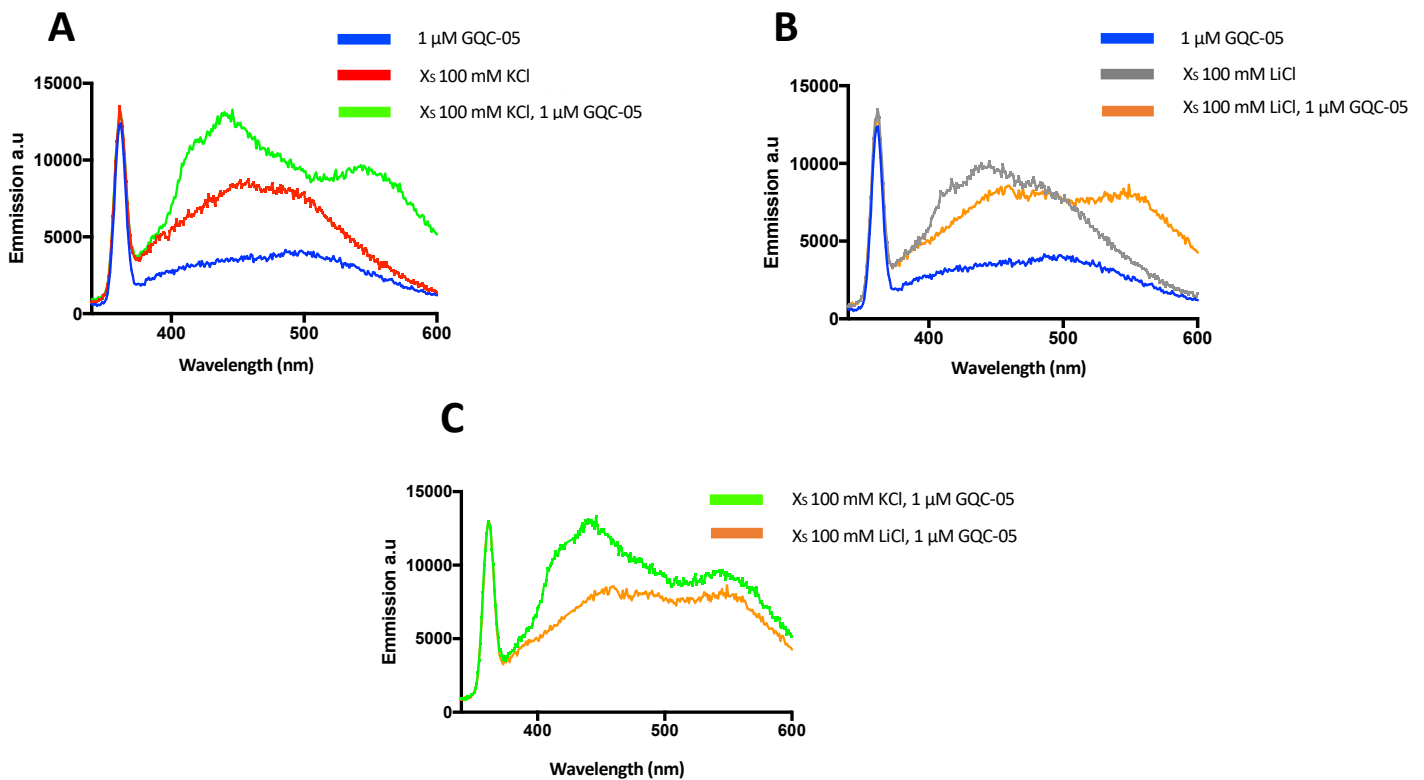
NMR would be a very informative technique to understand the remodelling of the X<sub>5</sub> RNA, as duplex and G4 species are much more discernible. However, as the X<sub>5</sub> domain has a large molecular weight of approximately 60 kDa, NMR would be difficult due to potential for extensive peak broadening. Therefore, we decided to use another method by exploiting the intrinsic fluorescence properties of G4s.

### **5.7.2 Harnessing the intrinsic fluorescence properties of G4s to probe conformational changes upon GQC-05 addition**

Previous studies have shown that nucleotides adopting a G4 structure display a much higher quantum yield ( $\phi$ ) than their duplex and single stranded counterparts (Zuffo et al., 2020). For example, the 26 mer G4 sequence d(TAAGGG)<sub>4</sub>TT displayed a  $\phi = 1.295 \times 10^{-3}$ , which was almost 5-fold more than its double-stranded counterpart (ds26  $\phi = 2.8 \times 10^{-4}$ ) and 3-fold more than a single stranded sequence of similar length (ss 24 mer  $\phi = 4.44 \times 10^{-4}$ ) (Zuffo et al., 2020). In the same study, they used the human telomere G4 sequence, which was shown to adopt a G4 structure in the presence of KCl but remained single stranded in LiCl. Analysis of the fluorescence emission spectrum showed that in the presence of KCl, there is a large emission peak between 350-450 nm, which is greatly attenuated in the presence of LiCl, indicating that this increased fluorescence can be attributed to the G4 structure (Zuffo et al., 2020). Similarly, duplex structures were also shown to exhibit reduced fluorescence emission in this region. Therefore, by comparing the fluorescence emission intensity in the 350-450 nm region in either KCl or LiCl-containing buffers both in the absence and presence of GQC-05, we can understand if GQC-05 is able to remodel the RNA towards a G4 structure, or simply maintain the RNA duplex.

The emission spectrum of X<sub>5</sub> RNA in the absence and presence of GQC-05 in KCl and LiCl-containing buffers is shown in figures 5.72 A-C. As explained previously, GQC-05 has a relatively low fluorescence when unbound. Figure 5.72 A shows the emission profile of the X<sub>5</sub> RNA in the presence of KCl before and after the addition of GQC-05. For the RNA alone, an emission is observed in the 350-450nm range. Upon addition of 1  $\mu$ M GQC-05, an almost 2-fold increase in the peak intensity is observed in this range, with a new shoulder appearing at 400nm. However, in the presence of LiCl, the addition of GQC-05 did not produce this same fluorescence enhancement, suggesting that in the presence of KCl, GQC-05 can re-model the X<sub>5</sub> domain by inducing a G4 structure (figures 5.72 B and C). This observation is consistent

with the data obtained earlier, when analysing the Q2 and its complementary strand using the NMR kinetic experiments (see section 4.5), where we showed that GQC-05 can preferentially stabilise a G4 structure over its duplex counterpart. Collectively, these results suggest that although GQC-05 may be binding to the  $X_s$  domain in different regions, the overall effect of its action is that it is able to induce a G4 structure in  $X_s$ , which is likely coming from the G-tracts of Q2. Following this, next question we wanted to address is if the G4 induced by GQC-05 leads to changes in  $X_s$  5'ss accessibility, and this will be explored in the next section.



**Figure 5.72 The intrinsic fluorescence of G4s and how this is used to monitor the conformational changes observed with GQC-05 on  $X_s$  RNA.**

(A and B) Fluorescence emission spectra of GQC-05 with  $X_s$  RNA in the presence of either 100 mM KCl (C), or LiCl (D). (E) Comparison of the emission spectra with  $X_s$  and GQC-05 in KCl and LiCl. Experiments were conducted in 10 mM Tris pH 7.0, 100 mM salt, 1 cm pathlength quartz cuvette ( $\lambda_{\text{Exc}} = 321$  nm).

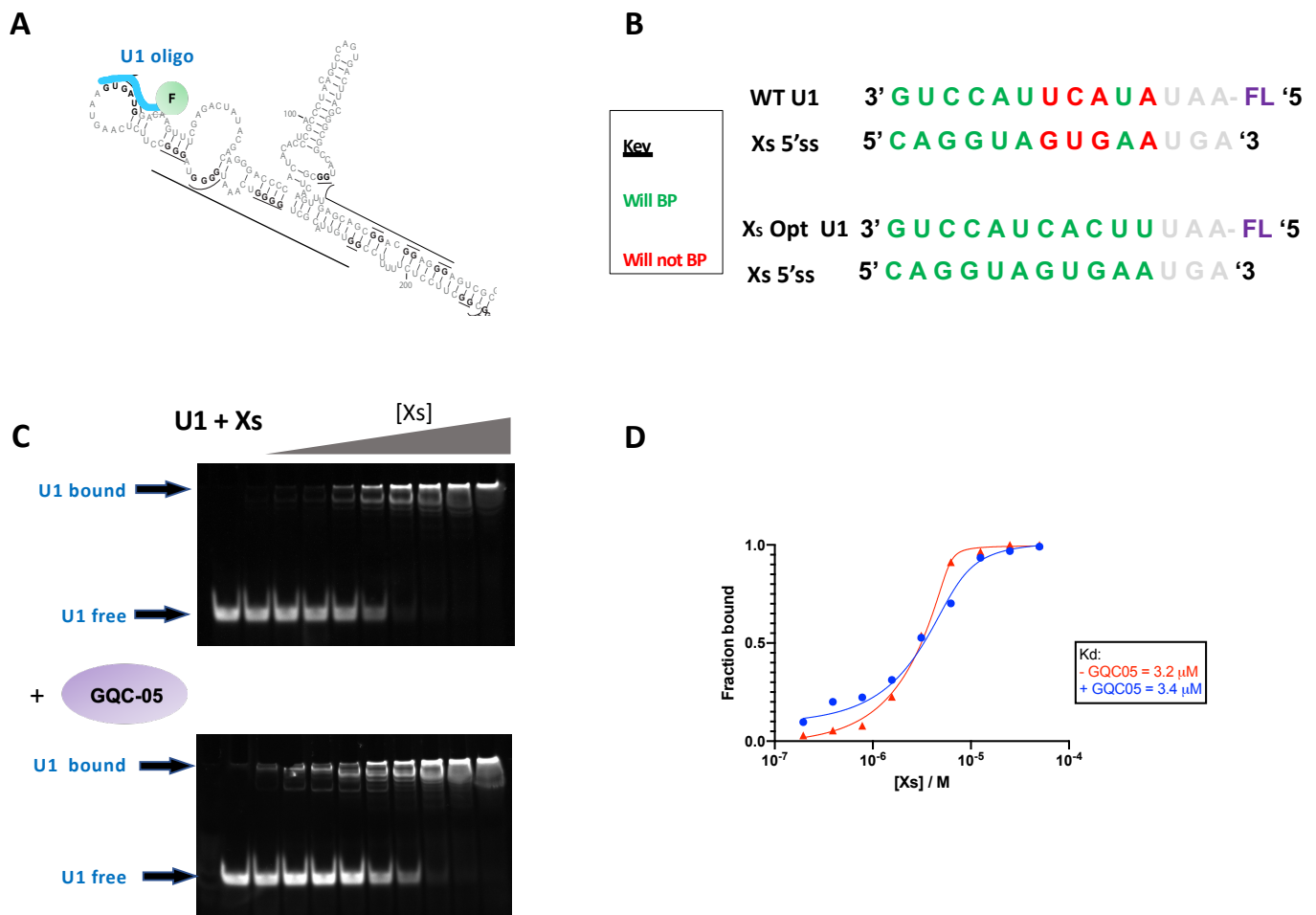
## 5.8 The effect of GQC-05 in altering the U1 accessibility at the X<sub>5</sub> 5'ss

As described previously in section 4.6, figure 4.61, we hypothesise that stabilisation of the Q2 G4 by GQC-05 results in a destabilisation of the large stem loop structure within the X<sub>5</sub> domain, making the X<sub>5</sub> 5'ss more accessible to the U1 snRNP and hence a splice site activation. Having shown that GQC-05 can re-model the RNA in favour of a G4, the next stage is to examine the effect of such structural changes on the accessibility of the U1 snRNP.

### 5.8.1 Using EMSA to monitor the U1-X<sub>5</sub> interaction

To test this hypothesis, we took the approach shown in figure 5.81 A, where we tested the accessibility of a fluorescently labelled U1 oligo which resembles that of the 5'snRNA component of the U1 snRNP. However, as we are not using the actual snRNP itself which contains proteins to help stabilise the interaction at the 5'ss (see section 1.2), we decided to use an X<sub>5</sub> optimised oligo which is 100% complementary to the X<sub>5</sub> 5'ss, as the wild-type U1 snRNA has 4 non-complementary nucleotides due to the fact that the X<sub>5</sub> 5'ss is not a consensus splice-site sequence (figure 5.81 A and B). This will help increase the affinity of the oligo at this site, meaning that the binding affinity to the X<sub>5</sub> RNA can be more readily measured as well as reducing the probability of non-specific binding to other sites of the RNA.

In order to test the binding of the U1 snRNA oligo to the X<sub>5</sub> 5'ss, we used electromobility shift assays, where we titrated increasing amounts of the X<sub>5</sub> RNA to a fixed concentration of the fluorescein labelled U1 oligo. Because fluorescein has an overlapping excitation spectra to ethidium bromide, we were able to visualise the fluorescein labelled U1 under ultraviolet (UV) light. I chose to use 6  $\mu$ M of the U1 oligo, as this was minimum concentration in order to detect the band on the gel under UV light. We then compared the gel shifts in the presence and absence of GQC-05 at different concentrations. Figure 5.81 C shows an example of the EMSA experiment, where we titrated increasing amounts of the X<sub>5</sub> RNA up to 50  $\mu$ M to the U1 oligo in the absence (top gel) and presence (lower gel) of 50  $\mu$ M GQC-05 (i.e. up to a 1:1 ratio of GQC-05 to RNA). From these data we can plot the fraction of U1 bound as a function of the X<sub>5</sub> concentration (figure 5.81 D). These results indicate that the presence of GQC-05 does not affect the binding affinity of the U1 oligo, giving rise to K<sub>d</sub> values of 3.2 and 3.4  $\mu$ M in the absence and presence of GQC-05 respectively. Following this, we next investigated if increasing the GQC-05 concentration is necessary to observe a change in the binding affinity of the U1 oligo.



**Figure 5.81 Using EMSA to examine the U1 accessibility to the X<sub>s</sub> 5'ss.**

(A) Schematic showing the expected interaction of the oligo and the X<sub>s</sub> 5'ss. (B) Comparison showing the base complementarity of the WT and X<sub>s</sub> optimised U1 oligos. (C) EMSA assay with (top gel) and without (lower gel) 50  $\mu$ M GQC-05. Experiments were conducted using 6  $\mu$ M fluorescein labelled U1 with increasing concentrations of the X<sub>s</sub> domain. Samples were run on a 16% native PAGE and visualized under a UV lamp. (D) Binding isotherm of (C), showing the fraction bound as a function of the X<sub>s</sub> RNA concentration in the absence (red) and presence (blue) of GQC-05.

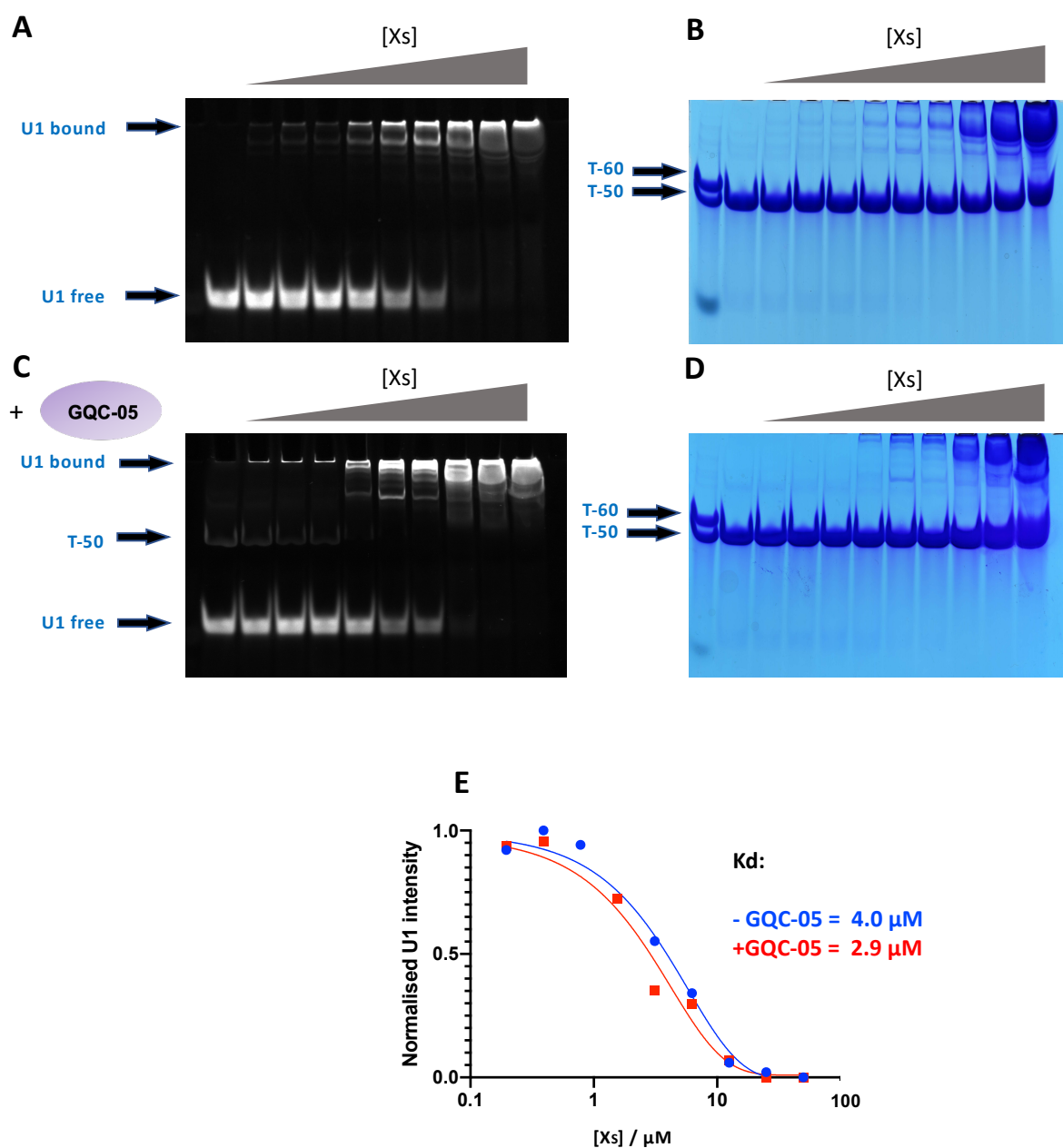


### 5.8.2 EMSA assays in the presence of higher concentrations of GQC-05

In chapter 4, we showed that GQC-05 binds to the Q2 G-tracts with a stoichiometry of 4:1. We therefore used this stoichiometric ratio in our EMSA assays, where we used 200  $\mu$ M GQC-05, giving rise to a 4:1 ratio of GQC-05 to the maximal  $X_5$  concentration of 50  $\mu$ M. However, preliminary work at this GQC-05 concentration showed that the migration of the U1 snRNA was being reduced, resulting in this RNA to be retained in the wells. This may be the result of GQC-05 binding the U1 snRNA and altering its electrophoretic mobility, due to it harbouring a positive charge in its tertiary amine. We therefore chose to use an excess of the poly-T-50 oligonucleotide, in which GQC-05 will expect to bind, preventing the interactions to the U1 snRNA, enabling U1 to migrate at the expected size (14 nucleotides). As poly T-50 is single stranded, the affinity of GQC-05 will likely be low enough to dissociate from T-50 and bind the  $X_5$  domain during the course of the titration.

Under the aforementioned conditions, we proceeded with the EMSA titrations in the absence and presence of 200  $\mu$ M GQC-05 with the addition of 70  $\mu$ M T-50 in all reactions (figures 5.82 A-D). In the absence of GQC-05, the EMSA was very similar as the one presented in figure 5.81 C, showing that poly T-50 has little to no effect on the binding of the U1 oligo to the  $X_5$  domain (figure 5.82 A). We also stained the gel with toluidine blue as proof of the presence of T-50 as well as showing that the T-50 band is not binding to the  $X_5$  RNA during the titration, both in the presence and absence of GQC-05 (figures 5.82 B and D respectively). In the presence of GQC-05, a band appears between U1 free and U1 bound (figure 5.82 C), which is likely coming from the fluorescence of GQC-05 binding to the T-50 fragment. The fluorescence of the T-50 band reduces during the course of the titration, supporting our prediction that it is dissociating from T-50 and binding the  $X_5$  domain. Another issue which arose at this GQC-05 concentration (figure 5.82 C) is the large background occurring at the bound signal, which was a result of GQC-05 binding to the  $X_5$  domain. Preliminary work showed that the proportion of the total fluorescence signal coming from just the GQC-05- $X_5$  interaction is around 90% of the total fluorescence of the U1- $X_5$ -GQC-05 ternary complex at the highest  $X_5$  concentration (50  $\mu$ M). This will make it very difficult to calculate the proportion of U1 bound, due to the large interference of the GQC-05 fluorescence. Therefore, to estimate the  $K_d$  of the U1/ $X_5$  interaction, we only monitored the disappearance of the U1 signal as a function of the  $X_5$  concentration, rather than calculating the fraction bound (figure

5.82 E). From figure 5.82 E, we noticed very little differences in the binding of the U1 oligo in the presence and absence of GQC-05, despite being at a concentration of 200  $\mu$ M. This suggests that the re-modelling of the RNA discussed in section 5.7 may not be directly affecting the accessibility of the U1 snRNA at the 5'ss to achieve splice site selection, but rather a different mechanism may be at play. However, as mentioned previously, the *in vitro* splicing assay showed that to achieve maximum activation of the X<sub>5</sub> 5'ss, an 800-fold excess of GQC-05 relative to the RNA concentration is required. In this experiment, we have used a maximum of 4 equivalents of GQC-05, which is considerably less than the molar ratio of GQC-05:RNA used in the splicing assay. Therefore, to see a difference in the U1 uptake, a higher molar ratio which more closely resembles that used in the splicing assay may need to be used. Unfortunately, with the above assay, this will require around 40 mM GQC-05, which is beyond its solubility limit even in 100% DMSO. Therefore, the EMSA approach will need to be adapted in order to recapitulate the GQC-05:RNA molar ratios necessary to activate the upstream site. One option would be to use a radio-labelled X<sub>5</sub> mRNA, meaning that only pM - nM concentrations of RNA would be required, enabling GQC-05 to be applied in excess without solubility concerns. The EMSA could then be conducted through titrations of a purified U1 snRNP to a fixed concentration of the radio-labelled RNA.



**Figure 5.82 EMSA of U1 to the X<sub>s</sub> RNA with 200 μM GQC-05 and 70 μM T-50.**

(A-D) EMSA titrations of the X<sub>s</sub> domain in the absence (A and B) and presence (C and D) of 200 μM GQC-05, and visualised through UV light (A and C) or toluidine blue staining (B and D). (E) Showing the normalised U1 intensity as a function of the X<sub>s</sub> concentration.

## 5.9 Discussion

In this chapter, we have explored the relevance of the observations made in the previous two chapters to longer more physiologically relevant RNAs, in order to validate the experimental data carried out on the Q2 G-tract. We have shown by both CD and fluorescence measurements that a G4 can still form in the 200nt long X<sub>S</sub> domain, where we provide strong evidence that this G4 is emanating from the Q2 and not the Q1 G-tracts.

Data interpretation regarding binding of GQC-05 for each of the generated constructs proved to be difficult in yielding any conclusions, as in the absence of a nuclear extract, there is high affinity binding for each of the constructs generated, regardless if the sequence has greater G4 forming potential or not, which may indicate that GQC-05 is specific for multiple different structures (or an unknown RNA secondary structure element). This is consistent with our previous findings, where the  $\Delta$ G4 RNA (which has greater duplex and less G4 forming potential) also displayed a similar binding affinity to the WT Q2 sequence (which has a higher G4 forming potential). One possible argument for such interactions is the very nature of GQC-05 by harbouring a positive charge in its tertiary amine at position 9. However, in our pull-down experiments, we used a modified analogue of GQC-05, by which the tertiary amine at position 9 is removed and replaced with a PEG linker attached to biotin. These results showed similar levels of streptavidin pull down on both the WT and 7-deaza substituted X<sub>S</sub> RNA, suggesting that simply the positively charged amine is not the only cause for the non-structural specificity of GQC-05.

However, as pull-down assays are not suitable for detecting small changes in affinity, we conducted fluorescence binding assays on each of the constructs generated, to determine the necessary domains of the RNA needed for a tighter binding interaction. Results indicated that the small stem at positions 62-111 which was predicted to form in the full length X<sub>S</sub> domain may play an important role in ensuring a tighter binding affinity, although, the exact function of this stem in the GQC-05 interaction remains to be explored. Furthermore, the formation of this stem may also be compromised when making truncations and isolating specific domains of the RNA, as the RNA in these constructs will have a different minimum free energy from the X<sub>S</sub> domain, meaning that an entirely different structure may be folding than the ones predicted from the foot-printing data. Therefore, the increased binding affinity we observe may be due to sequence rather than structural specificity of this stem. The exact structure of

the RNAs, particularly the presence of the stem at positions 62-111 in these shorter constructs could be validated with methods such as selective 2'hydroxyl acylation and primer extension (SHAPE) chemistry (Weeks and Mauger., 2011).

G4s in longer RNAs are likely to be significantly more transient than their isolated shorter domain counterparts, adding a nuclear extract, and thereby increasing the transient nature of G4s further, did not yield any potassium dependant changes in the CD spectrum, suggesting that a G4 cannot form. A possible explanation could be that G4 formation is in competition with RNA binding proteins, such as RNA helicases and hnRNPs F/H which have been shown to unfold RNA secondary structures (Dominguez et al., 2010; Gao et al., 2019). In particular, many studies have shown that the RNA helicase DHX36 can bind specifically to parallel type G4s through its N-terminal domain and unfold them in an ATP dependant manner (Chen et al., 2015; Creacy et al., 2008; Giri et al., 2011; Lattmann et al., 2010; Tippana et al., 2019; Vaughn et al., 2005). Furthermore, the GQC-05 interaction was also abolished in the X<sub>s</sub> domain in the presence of a nuclear extract, despite still harbouring the Q2 sequence which was shown to bind to GQC-05 in the presence of a nuclear extract when studied in isolation. This suggests that biologically, the G4 within the X<sub>s</sub> domain may not form readily under normal circumstances, and only after the addition of a large excess of GQC-05, a G4 structure may be stabilised and have the potential to form. This is consistent with the conditions of the splicing assay, in which an 800-fold excess of GQC-05 was required in order to achieve maximum activation of the X<sub>s</sub> 5'ss, meaning that a G4 may only form under these conditions when competing proteins are present. The lack of binding between GQC-05 and the X<sub>s</sub> domain in the presence of nuclear extract may therefore arise from the fact we were not recapitulating the same molar ratio in the fluorescence titrations as in the splicing assay. However, due to limitations with biophysical techniques and compound solubility, recapitulating the same molar ratio is probably impossible.

With all these limitations in determining the binding affinity in the presence of a nuclear extract, we next turned our attention to examining the effect of GQC-05 in modulating the overall structure of the X<sub>s</sub> domain in the absence of nuclear extract. By harnessing the higher quantum yield of G4 structures over their single stranded and duplex counterparts, preliminary data using fluorescence spectroscopy suggested that GQC-05 may have the ability to re-model the RNA and induce the formation of a G4 structure in the presence of potassium chloride, however further investigation will need to be conducted on well characterised G4

and duplex structures to validate such observations. This suggests that despite GQC-05 binding to multiple places on the RNA, the overall effect of GQC-05 in modulating the RNA secondary structure may be the critical factor in influencing splice site selection.

We next tested the effect of this re-modelling on the accessibility of the U1 snRNA through EMSA experiments. Results indicated no noticeable change in the binding affinity of the U1 snRNA after the addition of 4 equivalents of GQC-05. However, a similar issue arises as we are not fully recapitulating the same molar ratios as in the splicing reaction. Perhaps using higher GQC-05 concentrations would be necessary to see a noticeable change in the binding affinity of the U1 snRNA. Alternatively, instead of affecting the U1 snRNP accessibility, the induced G4 in the presence of GQC-05 may make another *cis* acting sequence accessible to a different *trans* acting factor as a means to regulate splice site selection. The mechanism by which this structural re-modelling can bias the X<sub>5</sub> 5'ss remains to be explored.

There are limitations with the EMSA assay described above which may have impacted on the observed results. For example, by annealing an oligonucleotide to probe the accessibility has the potential to completely change the 3D structure of the RNA, meaning that structural and biophysical characterisations of the RNA without the oligo may no longer hold true. In a future experiment, using an oligo that is not 100% complementary (such as the WT U1 5'snRNA sequence) may prevent such drastic changes to the RNA structure (as the artificial duplex would not be as strong), and potentially result in observable changes with GQC-05.

In conclusion, we have showed that GQC-05 has the ability to bind to multiple different RNA sequences, but has the ability to re-model the RNA into a G4 in regions that have G4 forming potential, meaning that simply looking at the binding affinity of this compound for different structures is not very informative on deciphering its mechanism of action.

## **CHAPTER 6: Conclusions and future perspectives**

### **6.1 Introduction**

### **6.2 The Q2 G-quadruplex**

### **6.3 Effects of GQC-05 on the Bcl-x mRNA**

### **6.4 Effect of GQC-05 on the accessibility of the X<sub>s</sub> 5'ss**

### **6.5 Final conclusions**

## 6.1 Introduction

Bcl-x is a member of the BCL-2 family of pro and anti-apoptotic proteins and has two alternative 5' splice sites (5'ss) in exon 2, giving two protein splice variants with antagonistic functions. Bcl-X<sub>S</sub> functions as a pro-apoptotic mediator by promoting the release of cytochrome c from the mitochondria, in contrast to Bcl-X<sub>L</sub> which inhibits pro-apoptotic mediators (Czabotar et al., 2014). The ratio of Bcl-X<sub>S</sub> to Bcl-X<sub>L</sub> therefore determines the fate of a cell, enabling us to exploit the splicing mechanism of Bcl-x for cancer treatment by trying to promote the pro-apoptotic X<sub>S</sub> isoform.

In our lab, previous work suggested a structural model for a Bcl-x-681 mini-gene (which has identical splicing patterns to the wild-type Bcl-x pre-mRNA) obtained using M-fold with constraints collected from foot printing data (Weldon et al., 2017). After substituting the guanine for de-aza guanine (which prevents formation of G4 structures), the largest differences in the foot-printing cleavage pattern was restricted to two distinct regions of mRNA, termed Q2 and Q5. Bioinformatic analysis of these areas suggested that certain stretches within these regions have the requirements to fold into canonical G-quadruplexes (G4s). Interestingly, the Q2 and Q5 G4 forming regions lie close to both the 5' splice sites (downstream of X<sub>S</sub> and upstream of X<sub>L</sub> respectively), which suggests their role as a regulatory mechanism to alter the splice switch of Bcl-x.

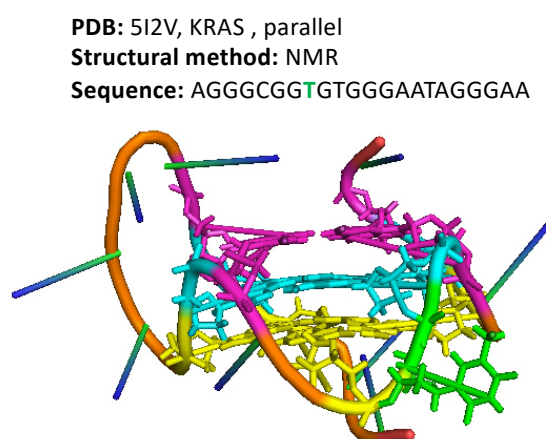
With the potential of G4s regulating the splicing behaviour of Bcl-x, G4 stabilising ligands were added into the splicing reaction and the ellipticine derivative GQC-05 showed an increase in X<sub>S</sub> and a decrease in the X<sub>L</sub> splice site usage, resulting in an 8-fold increase in the X<sub>S</sub>/X<sub>L</sub> ratio. Other classes of known G4 stabilising ligands were also tested by *in vitro* splicing such as TMPyP4 and SYUIQ-5 (porphyrin and quindoline derivatives), but none showed the same splicing switch as GQC-05.

Therefore, the overall aim of this thesis is to characterise the presence of a G4 in the Q2 region using a variety of biophysical techniques. This was followed by investigating the binding specificity of GQC-05 and its effects in the structural re-modelling of the RNA, and how this may achieve splice site selection.



## 6.2 The Q2 G-quadruplex

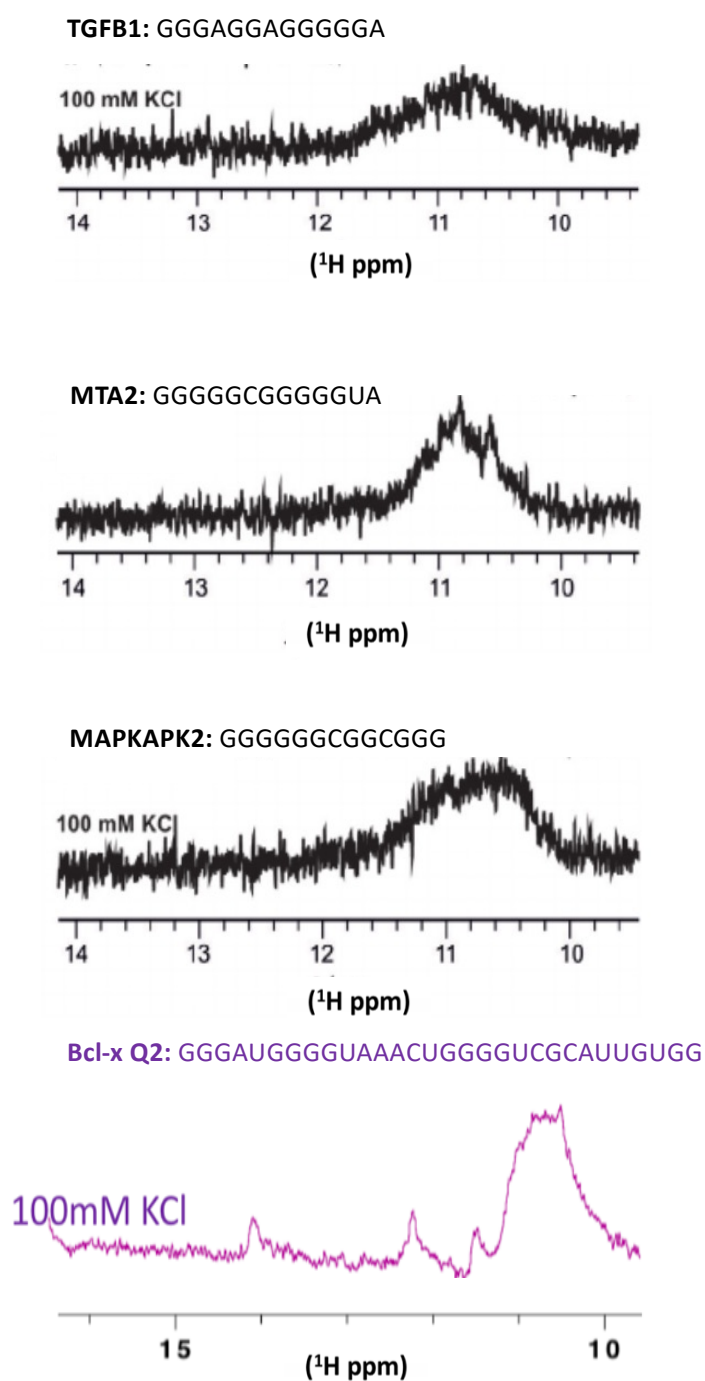
In chapter 3, we have shown that the G-tracts of Q2, which lies 13 nucleotides downstream of the X<sub>S</sub> 5'ss, can fold into a multi-conformational quadruplex as demonstrated by our NMR experiments (figure 3.41 A). We also showed that simply mutating out the excess guanines not involved in the G-tetrad, thereby to try and isolate the major canonical 2-tiered G4 conformation, prevented the G4 fold altogether (figure 3.52), suggesting that the G4 in the Q2 region may follow a non-canonical topology (Kikin et al., 2006). This observation is also consistent with the bioinformatic analysis described in section 1.5.1, where we showed that the Q2 sequence displays a relatively higher G-score in programmes that can also predict non-canonical G4 structures (G4 hunter), over those that can only predict canonical G4 topologies (QGRS mapper) (Brazda et al., 2019). One possibility of a non-canonical conformer may arise through the incorporation of the 5' G in the G4 tetrad of the GUGG stretch at the end of Q2 sequence, resulting in a single uridine bulge and a 3-tiered G4 conformer (as opposed to a canonical 2-tiered structure). A very similar observation was made with the NMR structure of the DNA KRAS promotor quadruplex, where a discontinuous run of guanines in the central G-tract (GGTG), still incorporated all of the guanines in the G4, resulting in a 3-tiered G4 with a thymine bulge (highlighted in green in figure 6.21) (Kerkour et al., 2017). Therefore, the possibility of a non-canonical G4 may be the likely conformation adopted by the Q2 sequence.



**Figure 6.21 Example of a non-canonical G4 topology of the KRAS promotor G4.**

Highlighted in magenta, cyan and yellow are the G-tetrads. Highlighted in green is the thymine bulge which interrupts the second G-tract (Kerkour et al., 2017).

As well as adopting a non-canonical fold, we also showed that the Q2 sequence can form complex higher order structures (such as dimers and multimers), which also contribute to the peak broadening observed in the NMR experiments. Higher order G4 structures have been shown to form in G4 sequences that have an unrestrained guanine at the 5' end (which Q2 contains), resulting in a highly energetically favourable 5'-5' stacking mode of the G4s (Kogut et al., 2019). This phenomenon has also been observed in many G4 sequences that have a free 5' guanine, such as the RNA G4 sequences of TGFB1, MTA2 and MAPKAPK2, which showed a remarkably similar NMR spectra to the Q2 sequence in 100 mM KCl, where broad peaks were observed in the 10-12 ppm region indicating the formation of higher order structures (figure 6.22) (Binas et al., 2020). However, at lower KCl concentrations, we showed that Q2 can adopt a monomeric G4, with a significantly better resolved NMR spectrum (figure 3.53 B). Unfortunately, despite showing an improvement in the NMR spectrum at the lower KCl concentrations, there were also more duplex structures in solution, making structural studies difficult regardless of the salt concentration used. Therefore, to isolate the major species and aid in structural determination, the next stage would be to make further mutations to the Q2 sequence, whereby we could introduce an additional nucleotide (or an additional phosphate group) at the 5' end of the sequence, removing the thymine bulge, as well as removing the excess guanines in the GGGG runs of the second and third G-tracts, thereby isolating the possible 3-stacked major G4 conformation. This new sequence would need to be verified as the biologically active conformer by *in vitro* splicing assays.



**Figure 6.22 1D NMR spectra in the imino region of G4 RNA sequences displaying higher order structures.**

Comparison of 1D NMR spectra in the imino region of the TGFB1, MTA2, MAPKAPK2 and Bcl-x Q2 RNA sequences (adapted from Binas et al., 2020).

### 6.3 Effects of GQC-05 on the Bcl-x mRNA

To date, there has not been an experimentally determined structure of a GQC-05/G4 interaction. However, using molecular modelling of the c-myc G4 and docking, it has been proposed that the most likely binding mode was stacking of the planar ligand to the external tetrad of the parallel myc G4 (Kaiser et al., 2013) (see section 1.6.3). Furthermore, using competition dialysis, GQC-05 was shown to display much more selectivity and specificity for G4 structures over their duplex counterparts (Brown et al., 2011). However, in this thesis, we found experimental evidence contradicting conclusions in Brown et al. Indeed, we showed that GQC-05 exhibits similar affinities to structures that have a high G4 as well as low G4 (high duplex) potential, suggesting that it may not only be recognising G4s. This observation is not surprising when we consider that the ellipticine core (from which GQC-05 is derived), was initially described as a duplex binder, that can intercalate between G-C Watson-Crick base pairs (Canals et al., 2005; Kohn et al., 1975). It is therefore possible that the ellipticine core of GQC-05 is responsible for its non-G4 interactions. Interestingly, to our knowledge, we are the first to report that the G4 specificity of GQC-05 is revealed only in functional conditions when in the presence of a nuclear extract, suggesting that the presence of surrounding proteins governs its specificity and perhaps its function. Furthermore, we showed that in the presence of nuclear extract, the binding stoichiometry of GQC-05 is shifted from a 4:1 to a 2:1 molar ratio, which is very consistent with many well characterised G4 binding ligands such as ThT and BMVC, all of which showed 2:1 binding stoichiometries to the c-myc promotor G4 (Liu et al., 2019; Verma et al., 2018). In light of these observations, biophysical and structural results conducted without the environment that support *in vitro* splicing should be taken with caution.

Despite showing interactions with both duplex and quadruplex structures in the absence of a nuclear extract, we showed that GQC-05 has the ability to re-model the structure of the RNA in favour of a G4 in RNAs that have G4 forming potential (such as in the Q2 and X<sub>5</sub> domains). This very much fits with the initial model that GQC-05 can shift the equilibrium from a duplex to a quadruplex structure, thereby achieving splice site selection (Weldon et al., 2018). This is also consistent with the foot-printing results described previously, where the addition of GQC-05 led to significant structural changes within the X<sub>5</sub> domain (see section 1.7, figure 1.75)

(Weldon et al., 2017). A future avenue worth exploring would be to assess the sequence requirements necessary for GQC-05 to induce a G4 structure in a given RNA sequence, and if only sequences displaying a certain topology of quadruplex (such as parallel) is induced by the presence of GQC-05. This may help identify other effects of GQC-05 transcriptome wide.

## 6.4 Effect of GQC-05 on the accessibility of the X<sub>s</sub> 5'ss

Our original hypothesis for the activation of the X<sub>s</sub> 5'ss was that in the presence of GQC-05, a G4 is induced within the Q2 region, resulting in destabilisation of the duplex to which the 5'ss may be part of, hence increasing its accessibility to the U1 snRNP (see section 4.6, figure 4.61). Despite having shown that GQC-05 can shift the equilibrium from a duplex to a G4 structure, our data showed that there is no change in the U1 accessibility upon addition of GQC-05, which does not support the hypothesis proposed by Weldon et al., 2018. This could mean that the accessibility of the U1 snRNP to the X<sub>s</sub> splice site may not be the entire reason behind the increase in splice site usage when adding GQC-05. Alternatively, instead of affecting the U1 snRNP accessibility directly, the induced G4 in the presence of GQC-05 may increase the accessibility of another *cis* acting sequence to a *trans* acting activator. For example, immediately downstream of the Q2 G-tracts lies a polypyrimidine rich sequence: CCUUUUUCUCCUUC, which was shown to be the binding site of the protein PTBP1 (Bielli et al., 2014). As discussed in section 1.2.5.1, this protein displaces SRSF1, resulting in the activation of the X<sub>s</sub> 5'ss. Because the binding site of PTBP1 is in close proximity to the Q2 G-tracts, it is possible that this site becomes single stranded (and hence more accessible) as a result of the GQC-05 induced G4. This will need to be explored in the future. As well increasing the accessibility of the surrounding structure, the induced G4 in the Q2 region may also play a more direct role by acting as a potential protein recruitment site for key splicing factors. Recently, Eperon et al. performed a protein-RNA pulldown assay and showed that GQC-05 induces a strong binding of the Bcl-x pre-mRNA to the non-POU domain-containing octamer-binding protein (NONO) (unpublished data), which has been recently identified in having G4 specific binding properties (Simko et al., 2020). The pulldown also showed that GQC-05 induces the binding of another protein SFPQ (splicing factor proline and glutamine-rich) and earlier studies have demonstrated that NONO and SFPQ can form heterodimers (Dong et al., 1993; Passon et al., 2012). SFPQ has also been shown to bind independently to PTBP1, and

during the apoptotic response, SFPQ becomes hyperphosphorylated and dissociates from PTBP1, acquiring new binding partners such as the U1-70K protein (Patton et al., 1993; Shav-Tal et al., 2001). Therefore, a possible mechanism by which GQC-05 activates the use of the upstream site is that following G4 induction, the protein NONO is recruited to the Q2 G4, which in-turn forms heterodimers with the protein SFPQ which then recruits molecules of the U1 snRNP through interaction with the U1 70K protein, thereby bringing the U1 snRNP to the X<sub>5</sub> 5'ss and enhancing its usage. This potential mechanism would imply that the accessibility of the U1 snRNP to the X<sub>5</sub> 5'ss is mediated through a complex network of both protein-protein and well as protein-RNA interactions, and not just simply through the RNA-RNA interactions described in this thesis. Therefore, despite not showing any changes in splice site accessibility with regards to the U1 oligo base pairing at the upstream site, we cannot rule out the possibility that GQC-05 may be affecting accessibility of the U1 snRNP by influencing protein recruitment rather than just the RNA secondary structure. Further experiments to test this model will need to be conducted.

## 6.5 Final conclusions

In conclusion, we have shown that GQC-05 has the ability to bind to multiple different RNA sequences (containing both high and low G4 forming potential) with  $\mu$ M affinities, meaning that it may have specificity for an unknown RNA structural element, contradicting the published data (Brown et al., 2011). Therefore, trying to correlate the binding affinity of this compound for different structures does not give any insight into its mechanism of action, but rather that the overall consequence of such binding may give GQC-05 its unique properties in altering splice site selection. Thus, the overall effect of this compound seen in splicing may come from its ability to re-model RNA rather than its selectivity to bind to different structures directly. In light of these observations, solving the 3D structure of the Q2 quadruplex bound to GQC-05 will provide greater insight into its mechanism of action, an aim which was originally part of this project but due to heterogeneity and time constraints could not be achieved. However, we have shown that using the wild-type Q2 sequence will make structural determination incredibly difficult, if not impossible, due to the heterogeneity of the RNA structure described above. Therefore, the next stage of this project would be to make

mutations in the Q2 RNA sequence suggested in section 6.2, in order to begin the process of structural determination with the hope to obtain a ligand bound complex.

## **CHAPTER 7: Appendix**

**7.1 Plasmid vector used for cloning the larger Bcl-x fragments**

**7.2 Sequences of the large Bcl-x constructs**

**7.3 10 X Tris – Borate EDTA (TBE)**

**7.4 Gel recipes**

**7.5 Toluidine blue stain**

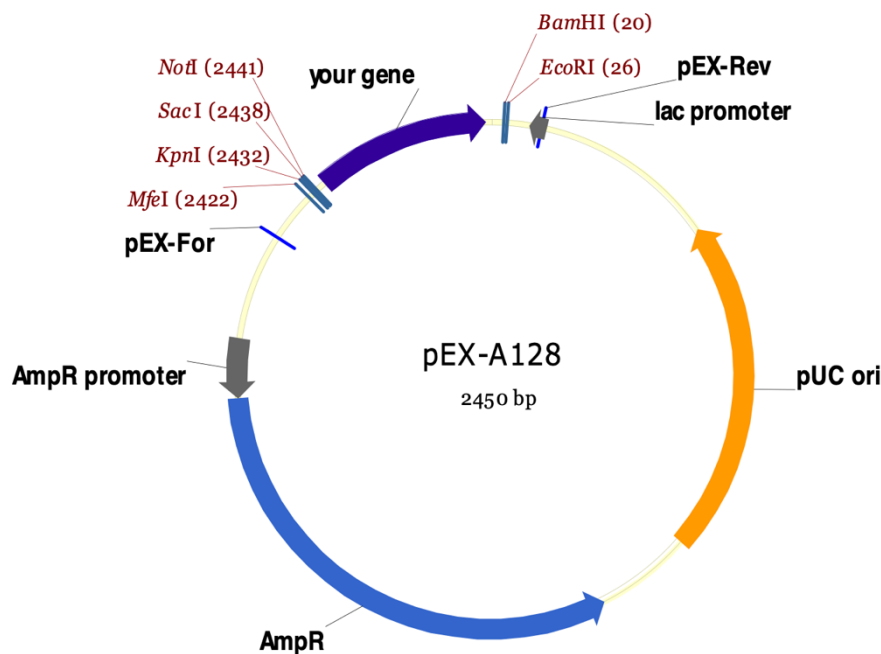
**7.6 Gel loading dyes**

**7.7 Transcription buffer**

**7.8 2TY growth media**



## 7.1 Plasmid vector used for cloning the larger Bcl-x fragments



## 7.2 Sequences of the large Bcl-x constructs

Bcl-x construct	Sequence
X <sub>5</sub> (38-206)	TAATACGACTCACTATA GAGACTGAGGGAGGCAGGCGACGAGTTTGAAGTGCGGTACCGGCGGGCATTCACTGACCTGACATCCCAGCTCCACATCACCCCAGG GACAGCATATCAGAGCTTTGAACAGGTAGTGAATGAAGTCTTCCGGGATGGGGTAACTGGGGTCGCATTGTGGCCTTTTCTCCTTC TGAGACCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
SD (38-61 + 112-206)	TAATACGACTCACTATA GAGACTGAGGGAGGCAGGCGACGAGTTTGACCCAGGGACAGCATATCAGAGCTTTGAACAGGTAGTGAATGAAGTCTTCCGGGATG GGGTAACTGGGGTCGCATTGTGGCCTTTTCTCCTTC TGAGACCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
Q1 (38-151)	TAATACGACTCACTATA GAGACTGAGGGAGGCAGGCGACGAGTTTGAAGTGCGGTACCGGCGGGCATTCACTGACCTGACATCCCAGCTCCACATCACCCCAGG GACAGCATATCAGAGCTTTGAACAGGTAGTGAA TGAGACCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
Q2 (150-206)	TAATACGACTCACTATA GAGACTAATGAAGTCTTCCGGGATGGGGTAACTGGGGTCGCATTGTGGCCTTTTCTCCTTC TGAGACCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
Q2.2 (150-206)	TAATACGACTCACTATA GAGACTAATGAAGTCTTCCGGGATGGGGTAACTGUUGTCGCATTGTGGCCTTTTCTCCTTC TGAGACCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG

### Key

T7 promotor

Bsal restriction site

T7 terminator

### 7.3 10 X Tris-Borate EDTA (TBE)

Reagents	Amount added
Tris base	108 g
Na <sub>2</sub> EDTA.2H <sub>2</sub> O	9.3 g
Boric acid	55 g
MilliQ water	Up to 1 litre

### 7.4 Gel recipes

#### 7.4.1 Agarose gel 1%

Component	Amount added
Agarose	1 g
1 X TBE	100 mL
Ethidium bromide (10 mg/mL)	10 µL

#### 7.4.2 Acrylamide gels

Component	8%	12%	16%
Accugel 40% (w/v) 29:1 Acrylamide: Bis acrylamide	2 mL	3 mL	4 mL
*9 M urea in 1 X TBE	8 mL	7 mL	6 mL
10% APS	100 µL	100 µL	100 µL
TEMED (Tetramethylethylenediamine)	10 µL	10 µL	10 µL

\*Native gels are made in the same way but with 1 X TBE only (without 9 M urea).

### 7.5 Toluidine blue gel stain

Component	Amount added	[Final]
Toluidine blue (MW = 305.83)	1 g	0.1 %
Acetic acid	100 mL	10 %
MilliQ water	900 mL	/

## 7.6 Gel loading dyes

### 7.6.1 Denaturing loading dye (2 X)

Component	[Final]	[Stock]	Amount added
Formamide	95 %	100 %	9.5 mL
Na <sub>2</sub> EDTA pH 8.0	18 mM	500 mM	360 µL
SDS	0.025 %	10 %	25 µL
Bromophenol blue	0.025 %	/	2.5 mg
Xylene cyanol	0.025 %	/	2.5 mg
MilliQ water	/	/	115 µL

### 7.6.2 Native loading dye (5 X)

Component	[Final]	[Stock]	Amount added
TBE	5 X	10 X	5 mL
Glycerol	20 %	100 %	2 mL
Bromophenol blue	0.025 %	/	2.5 mg
MilliQ water	/	/	3 mL

## 7.7 Transcription buffer (TB 20X)

Component	[Final]	[Stock]	Amount added
Tris-HCl pH 8.1	800 mM	1 M	8 mL
Spermidine	20 mM	1 M	200 µL
Triton-X	0.2 %	100 %	20 µL
DTT	100 mM	1 M	1 mL
MilliQ water	/	/	780 µL

### 7.8 2TY growth media (1 litre)

Component	Amount added
Tryptone	16 g
Yeast extract	10 g
NaCl	5 g

## Bibliography

- Adesso, L., Calabretta, S., Barbagallo, F., Capurso, G., Piloizzi, E., Geremia, R., Delle Fave, G., Sette, C., 2013. Gemcitabine triggers a pro-survival response in pancreatic cancer cells through activation of the MNK2/eIF4E pathway. *Oncogene* 32, 2848–2857.
- Adrian, M., Heddi, B., Phan, A.T., 2012. NMR spectroscopy of G-quadruplexes. *Methods* 57, 11–24.
- Agrawal, P., Lin, C., Mathad, R.I., Carver, M., Yang, D., 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. *J. Am. Chem. Soc.* 136, 1750–1753.
- Almaqwashi, A.A., Paramanathan, T., Rouzina, I., Williams, M.C., 2016. Mechanisms of small molecule-DNA interactions probed by single-molecule force spectroscopy. *Nucleic Acids Res.* 44, 3971–3988.
- Ambrus, A., Chen, D., Dai, J., Jones, R.A., Yang, D., 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.
- Anokhina, M., Bessonov, S., Miao, Z., Westhof, E., Hartmuth, K., Lührmann, R., 2013. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* 32, 2804–2818.
- Aritomi, M., Kunishima, N., Inohara, N., Ishibashi, Y., Ohta, S., Morikawa, K., 1997. Crystal Structure of Rat Bcl-x<sub>L</sub>. *J. Biol. Chem.* 272, 27886–27892.
- Arnoult, N., Van Beneden, A., Decottignies, A., 2012. Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1 $\alpha$ . *Nat. Struct. Mol. Biol.*
- Arola, A., Vilar, R., 2008. Stabilisation of G-Quadruplex DNA by metal complexes. *Curr. Top. Med. Chem.* 1405–1415.
- Ashkenazi, A., Dixit, V.M., 1998. Death Receptors : Signaling and Modulation 281, 1305–1309.
- Auclair, C., 1987. Multimodal action of antitumor agents on DNA: The ellipticine series. *Arch. Biochem. Biophys.* 259, 1–14.
- Azzalin, C.M., Reichenbach, P., Khoraiuli, L., Giulotto, E., Lingner, J., 2007. and RNA Surveillance Factors at. *Science* 318, 798–801.
- Bang, I., 1910. Untersuchungen über die Guanylsäure. *Biochem. Z* 26, 293–311.
- Bao, H.L., Ishizuka, T., Sakamoto, T., Fujimoto, K., Uechi, T., Kenmochi, N., Xu, Y., 2017. Characterization of human telomere RNA G-quadruplex structures in vitro and in living cells using 19F NMR spectroscopy. *Nucleic Acids Res.* 45, 5501–5511.
- Barbour, S.E., Nguyen, P.T., Park, M., Emani, B., Lei, X., Kambalapalli, M., Shultz, J.C., Wijesinghe, D., Chalfant, C.E., Ramanadham, S., 2015. Group VIA phospholipase  $\alpha 2$  (iPLA2 $\beta$ ) modulates Bcl-x 5'-splice site selection and suppresses anti-apoptotic Bcl-x(L) in  $\beta$ -cells. *J. Biol. Chem.* 290, 11021–11031.
- Bauman, J.A., Li, S.D., Yang, A., Huang, L., Kole, R., 2010. Anti-tumor activity of splice-switching oligonucleotides. *Nucleic Acids Res.* 38, 8348–8356.
- Beaudoin, J.D., Jodoin, R., Perreault, J.P., 2014. New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res.* 42, 1209–1223.
- Beaudoin, J.D., Perreault, J.P., 2010. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* 38, 7022–7036.
- Bedrat, A., Lacroix, L., Mergny, J.L., 2016. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746–1759.
- Biancalana, M., Koide, S., 2010. Molecular mechanism of Thioflavin-T binding to amyloid

- fibrils. *Biochim. Biophys. Acta - Proteins Proteomics*.
- Bielli, P., Bordi, M., Di Biasio, V., Sette, C., 2014. Regulation of BCL-X splicing reveals a role for the polypyrimidine tract binding protein (PTBP1/hnRNP I) in alternative 5' splice site selection. *Nucleic Acids Res.* 42, 12070–12081.
- Biffi, G., Di Antonio, M., Tannahill, D., Balasubramanian, S., 2014. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* 6, 75–80.
- Biffi, G., Tannahill, D., McCafferty, J., Balasubramanian, S., 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*
- Binas, O., Bessi, I., Schwalbe, H., 2020. Structure Validation of G-Rich RNAs in Noncoding Regions of the Human Genome. *ChemBioChem* 21, 1656–1663.
- Bochman, M.L., Paeschke, K., Zakian, V.A., 2012. DNA secondary structures: Stability and function of G-quadruplex structures. *Nat. Rev. Genet.* 13, 770–780.
- Boise, L.H., González-García, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nuñez, G., Thompson, C.B., 1993. Bcl-X, a Bcl-2-Related Gene That Functions As a Dominant Regulator of Apoptotic Cell Death. *Cell* 74, 597–608.
- Boucher, M.J., Morisset, J., Vachon, P.H., Reed, J.C., Lainé, J., Rivard, N., 2000. MEK/ERK signaling pathway regulates the expression of Bcl-2, Bcl-X(L), and Mcl-1 and promotes survival of human pancreatic cancer cells. *J. Cell. Biochem.* 79, 355–369.
- Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Šťastný, J., Mergny, J.L., 2019. G4Hunter web application: A web server for G-quadruplex prediction. *Bioinformatics* 35, 3493–3495.
- Brody, E., Abelson, J., 1985. The 'spliceosome': Yeast pre-messenger RNA associates with a 40s complex in a splicing-dependent reaction. *Science* (80-. ). 228, 963–967.
- Brown, R. V., Danford, F.L., Gokhale, V., Hurley, L.H., Brooks, T.A., 2011. Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *J. Biol. Chem.* 286, 41018–41027.
- Bugaut, A., Balasubramanian, S., 2012. 5'-UTR RNA G-quadruplexes: Translation regulation and targeting. *Nucleic Acids Res.* 40, 4727–4741.
- Bugaut, A., Rodriguez, R., Kumari, S., Hsu, S.T.D., Balasubramanian, S., 2010. Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Org. Biomol. Chem.* 8, 2771–2776
- Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., Neidle, S., 2006. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* 34, 5402–5415.
- Buscaglia, R., Miller, M.C., Dean, W.L., Gray, R.D., Lane, A.N., Trent, J.O., Chaires, J.B., 2013. Polyethylene glycol binding alters human telomere G-quadruplex structure by conformational selection. *Nucleic Acids Res.* 41, 7934–7946.
- Cain, K., Bratton, S.B., Cohen, G.M., 2002. The Apaf-1 apoptosome: A large caspase-activating complex. *Biochimie* 84, 203–214.
- Canals, A., Purciolas, M., Aymamí, J., Coll, M., 2005. The anticancer agent ellipticine unwinds DNA by intercalative binding in an orientation parallel to base pairs. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 61, 1009–1012.
- Chamberlin, M., Mcgrath, J., Waskell, L., 1970. New RNA polymerase from escherichia coli infected with bacteriophage T7. *Nature* 228, 227–231.
- Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P., Balasubramanian, S.,

2015. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33, 877–881.
- Chen, M.C., Murat, P., Abecassis, K., Ferré-D'Amaré, A.R., Balasubramanian, S., 2015. Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res.* 43, 2223–2231.
- Chen, L., Weinmeister, R., Kralovicova, J., Eperon, L.P., Vorechovsky, I., Hudson, A.J., Eperon, I.C., 2017. Stoichiometries of U2AF35, U2AF65 and U2 snRNP reveal new early spliceosome assembly pathways. *Nucleic Acids Res.* 45, 2051–2067.
- Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.J., 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell.*
- Cloutier, A., Shkreta, L., Toutant, J., Durand, M., Thibault, P., Chabot, B., 2018. HnRNP A1/A2 and Sam68 collaborate with SRSF10 to control the alternative splicing response to oxaliplatin-mediated DNA damage. *Sci. Rep.* 8, 1–14.
- Cloutier, P., Toutant, J., Shkreta, L., Goekjian, S., Revil, T., Chabot, B., 2008. Antagonistic effects of the SRp30c protein and cryptic 5' splice sites on the alternative splicing of the apoptotic regulator Bcl-x. *J. Biol. Chem.* 283, 21315–21324.
- Cogoi, S., Xodo, L.E., 2006. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* 34, 2536–2549.
- Collie, G.W., Haider, S.M., Neidle, S., Parkinson, G.N., 2010a. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* 38, 5569–5580.
- Collie, G.W., Parkinson, G.N., Neidle, S., Rosu, F., Pauw, E. De, Gabelica, V., 2010b. Electrospray mass spectrometry of telomeric RNA (TERRA) Reveals the formation of stable multimeric G-quadruplex structures. *J. Am. Chem. Soc.* 132, 9328–9334.
- Corrêa, D., Ramos, C., 2009. The use of circular dichroism spectroscopy to study protein folding, form and function. *African J Biochem Res* 3, 164–173.
- Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcárcel, J., Sattler, M., 2007. U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat. Struct. Mol. Biol.* 14, 620–629.
- Courtois, S., Verhaegh, G., North, S., Luciani, M.G., Lassus, P., Hibner, U., Oren, M., Hainaut, P., 2002.  $\Delta$ N-p53, a natural isoform of p53 lacking the first transactivation domain, counteracts growth suppression by wild-type p53. *Oncogene* 21, 6722–6728.
- Creacy, S.D., Routh, E.D., Iwamoto, F., Nagamine, Y., Akman, S.A., Vaughn, J.P., 2008. G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* 283, 34626–34634.
- Crick, F., Watson, J., 1953. © 1953 Nature Publishing Group.
- Črnogelj, M., Hud, N. V., Plavec, J., 2002. The solution structure of d(G4T4G3)2: A bimolecular G-quadruplex with a novel fold. *J. Mol. Biol.*
- Czabotar, P.E., Lessene, G., Strasser, A., Adams, J.M., 2014. Control of apoptosis by the BCL-2 protein family: Implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* 15, 49–63.
- D'Antonio, L., Bagga, P., 2004. Computational methods for predicting intramolecular G-quadruplexes in nucleotide sequences. *Proc. - 2004 IEEE Comput. Syst. Bioinforma. Conf.*

CSB 2004.

- Dai, J., Carver, M., Hurley, L.H., Yang, D., 2011. Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: Insights into G-quadruplex-interactive small molecule drug design. *J. Am. Chem. Soc.* 133, 17673–17680.
- Dai, J., Chen, D., Jones, R.A., Hurley, L.H., Yang, D., 2006. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* 34, 5133–5144.
- Dai, J., Su, Y., Zhong, S., Cong, L., Liu, B., Yang, J., Tao, Y., He, Z., Chen, C., Jiang, Y., 2020. Exosomes: key players in cancer and potential therapeutic strategy. *Signal Transduct. Target. Ther.* 5.
- Dalziel, M., Nunes, N.M., Furger, A., 2007. Two G-Rich Regulatory Elements Located Adjacent to and 440 Nucleotides Downstream of the Core Poly(A) Site of the Intronless Melanocortin Receptor 1 Gene Are Critical for Efficient 3' End Processing. *Mol. Cell. Biol.* 27, 1568–1580.
- De La Faverie, A.R., Guédin, A., Bedrat, A., Yatsunyk, L.A., Mergny, J.L., 2014. Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res.* 42.
- Del Mundo, I.M.A., Zewail-Foote, M., Kerwin, S.M., Vasquez, K.M., 2017. Alternative DNA structure formation in the mutagenic human c-MYC promoter. *Nucleic Acids Res.* 45, 4929–4943.
- del Villar-Guerra, R., Gray, R.D., Trent, J.O., Chaires, J.B., 2018a. A rapid fluorescent indicator displacement assay and principal component/cluster data analysis for determination of ligand–nucleic acid structural selectivity. *Nucleic Acids Res.* 46, 1–10.
- del Villar-Guerra, R., Trent, J.O., Chaires, J.B., 2018b. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chemie - Int. Ed.* 57, 7171–7175.
- Di Antonio, M., Biffi, G., Mariani, A., Raiber, E.-A., Rodriguez, R., Balasubramanian, S., 2012. Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. *Angew. Chem. Int. Ed. Engl.* 51, 11073–11078.
- Dickerhoff, J., Onel, B., Chen, L., Chen, Y., Yang, D., 2019. Solution Structure of a MYC Promoter G-Quadruplex with 1:6:1 Loop Length. *ACS Omega* 4, 2533–2539.
- Dodin, G., Schwaller, M., Aubard, J., Paoletti, C., 1988. Binding of ellipticine base and ellipticinium cation to calf-thymus DNA: A thermodynamic and kinetic study. *Eur. J. Biochem.* 176, 371–376.
- Dominguez, C., Fiset, J.F., Chabot, B., Allain, F.H.T., 2010. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.* 17, 853–861.
- Dong, B., Horowitz, D.S., Kobayashi, R., Krainer, A.R., 1993. Purification and cDNA cloning of HeLa cell p54nrb, a nuclear protein with two RNA recognition motifs and extensive homology to human splicing factor PSF and Drosophila Nona/BJ6. *Nucleic Acids Res.* 21, 4085–4092.
- Douc Rasy, S., Multon, E., Kayser, A., Riou, G., 1983. Inhibition by ellipticine derivatives of the reactions catalysed by the topoisomerases: preferential inhibition of topoisomerase II. *C. R. Seances Acad. Sci.* III.
- Dwuma-Badu, D., Ayim, J.S.K., Dabra, T.T., El-Azizi, M.M., Schiff, P.L., Slatkin, D.J., Knapp, J.E., 1983. constituents of west african medicinal plants XXV isolation of oblongine from tiliacora dinklagei and the synthesis of oblongine and related benzylisoquinoline alkaloids. *J. Nat. Prod.* 46, 342–349.
- Dzatko, S., Krafcikova, M., Hänsel-Hertsch, R., Fessler, T., Fiala, R., Loja, T., Krafcik, D., Mergny,



- J.L., Foldynova-Trantirkova, S., Trantirek, L., 2018. Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chemie - Int. Ed.* 57, 2165–2169.
- Eddy, J., Maizels, N., 2006. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* 34, 3887–3896.
- Ehresmann, C., Baudin, F., Mougél, M., Romby, P., Ebel, J.-P., Ehresmann, B., 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15, 9109–9128.
- Elmore, S., 2007. Apoptosis: A Review of Programmed Cell Death. *Toxicol. Pathol.* 35, 495–516.
- Eperon, I.C., Makarova, O. V., Mayeda, A., Munroe, S.H., Cáceres, J.F., Hayward, D.G., Krainer, A.R., 2000. Selection of Alternative 5' Splice Sites: Role of U1 snRNP and Models for the Antagonistic Effects of SF2/ASF and hnRNP A1. *Mol. Cell. Biol.* 20, 8303–8318.
- Eperon, L.P., Estibeiro, J.P., Eperon, I.C., 1986. The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. *Nature* 324, 280–282.
- Eperon, L.P., Graham, I.R., Griffiths, A.D., Eperon, I.C., 1988. Effects of RNA secondary structure on alternative splicing of Pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase? *Cell* 54, 393–401.
- España, L., Fernández, Y., Rubio, N., Torregrosa, A., Blanco, J., Sierra, A., 2004. Overexpression of Bcl-x L in human breast cancer cells enhances organ-selective lymph node metastasis. *Breast Cancer Res. Treat.*
- Feigon, J., Koshlap, K.M., Smith, F.W., 1995. <sup>1</sup>H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.* 261, 225–255.
- Feldser, D.M., Greider, C.W., 2007. Short Telomeres Limit Tumor Progression In Vivo by Inducing Senescence. *Cancer Cell* 11, 461–469.
- Fischer, M., 2017. Census and evaluation of p53 target genes. *Oncogene* 36, 3943–3956.
- Freedman, D.A., Wu, L., Levine, A.J., 1999. Functions of the MDM2 oncoprotein. *Cell. Mol. Life Sci.* 55, 96–107.
- Freyer, M.W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L.H., Lewis, E.A., 2007. Biophysical studies of the c-MYC NHE III1 promoter: Model quadruplex interactions with a cationic porphyrin. *Biophys. J.* 92, 2007–2015.
- Frisoni, P., Pradella, D., Di Matteo, A., Belloni, E., Ghigna, C., Paronetto, M.P., 2015. SAM68: Signal transduction and RNA metabolism in human cancer. *Biomed Res. Int.* 2015.
- Fu, X.D., Ares, M., 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701.
- Fulda, S., 2009. Tumor resistance to apoptosis. *Int. J. Cancer* 124, 511–515.
- Fürtig, B., Richter, C., Wöhnert, J., Schwalbe, H., 2003. NMR spectroscopy of RNA. *ChemBioChem* 4, 936–962.
- Gabelica, V., Maeda, R., Fujimoto, T., Yaku, H., Murashima, T., Sugimoto, N., Miyoshi, D., 2013. Multiple and cooperative binding of fluorescence light-up probe thioflavin t with human telomere DNA G-quadruplex. *Biochemistry* 52, 5620–5628.
- Gao, F.H., Wu, Y.L., Zhao, M., Liu, C.X., Wang, L.S., Chen, G.Q., 2009. Protein Kinase C-δ mediates down-regulation of heterogeneous nuclear ribonucleoprotein K protein: involvement in apoptosis induction. *Exp. Cell Res.*
- Gao, J., Byrd, A.K., Zybailov, B.L., Marecki, J.C., Guderyon, M.J., Edwards, A.D., Chib, S., West, K.L., Waldrip, Z.J., Mackintosh, S.G., Gao, Z., Putnam, A.A., Jankowsky, E., Raney, K.D., 2019. DEAD-box RNA helicases Dbp2, Ded1 and Mss116 bind to G-quadruplex nucleic acids and destabilize G-quadruplex RNA. *Chem. Commun.* 55, 4467–4470.

- Garbett, N.C., Ragazzon, P.A., Chaires, J.O.B., 2007. Circular dichroism to determine binding mode and affinity of ligand-dna interactions. *Nat. Protoc.* 2, 3166–3172.
- Garneau, D., Revil, T., Fiset, J.F., Chabot, B., 2005. Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J. Biol. Chem.* 280, 22641–22650.
- Gellert, M., Lipsett, M.N., Davies, D.R., 1962. Helix Formation By Guanylic Acid. *Proc. Natl. Acad. Sci.* 48, 2013–2018.
- Germann, M.W., 2012. *Nucleic Acids NMR Spectroscopy*.
- Ghosh, S., Kar, A., Chowdhury, S., Dasgupta, D., 2013. Ellipticine binds to a human telomere sequence: An additional mode of action as a putative anticancer agent? *Biochemistry* 52, 4127–4137.
- Giassa, I.C., Rynes, J., Fessl, T., Foldynova-Trantirkova, S., Trantirek, L., 2018. Advances in the cellular structural biology of nucleic acids. *FEBS Lett.* 592, 1997–2011.
- Giri, B., Smaldino, P.J., Thys, R.G., Creacy, S.D., Routh, E.D., Hantgan, R.R., Lattmann, S., Nagamine, Y., Akman, S.A., Vaughn, J.P., 2011. G4 resolvase 1 tightly binds and unwinds unimolecular G4-DNA. *Nucleic Acids Res.* 39, 7161–7178.
- Głuszyńska, A., Juskowiak, B., Rubiś, B., 2018. Binding Study of the Fluorescent Carbazole Derivative with Human Telomeric G-Quadruplexes. *Molecules* 23, 3154.
- Goodwin, S., Smith, A.F., Horning, E.C., 1959. Alkaloids of *Ochrosia elliptica* Labill. *J. Am. Chem. Soc.* 81, 1903–1908.
- Gray, R.D., Trent, J.O., Arumugam, S., Chaires, J.B., 2019. Folding Landscape of a Parallel G-Quadruplex. *J. Phys. Chem. Lett.* 10, 1146–1151.
- Gray, R.D., Trent, J.O., Chaires, J.B., 2014. Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.* 426, 1629–1650.
- Greenberg, D.S., Soreq, H., 2013. Alternative Splicing. In: *Brenner's Encyclopedia of Genetics: Second Edition*. Elsevier Inc., pp. 97–98.
- Greenfield, N.J., 2009. NIH Public Access structure. *ProQuest Diss. Theses* 1, 2876–2890.
- Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K., Hutton, M., 1999. 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J. Biol. Chem.* 274, 15134–15143.
- Guédin, A., Gros, J., Alberti, P., Mergny, J.L., 2010. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* 38, 7858–7868.
- Guo, J.U., Bartel, D.P., 2016. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* (80- ). 353.
- Guyen, B., Schultes, C.M., Hazel, P., Mann, J., Neidle, S., 2004. Synthesis and evaluation of analogues of 10H-indolo[3,2-b]-quinoline as G-quadruplex stabilising ligands and potential inhibitors of the enzyme telomerase. *Org. Biomol. Chem.* 2, 981–988.
- Halder, R., Riou, J.F., Teulade-Fichou, M.P., Frickey, T., Hartig, J.S., 2012. Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. *BMC Res. Notes* 5.
- Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: The next generation. *Cell* 144, 646–674.

- Hann, C.L., Daniel, V.C., Sugar, E.A., Dobromilskaya, I., Murphy, S.C., Cope, L., Lin, X., Hierman, J.S., Wilburn, D.L., Watkins, D.N., Rudin, C.M., 2008. Therapeutic Efficacy of ABT-737, a Selective Inhibitor of BCL-2, in Small Cell Lung Cancer. *Cancer Res.* 68, 2321–2328.
- Hänsel, R., Foldynová-Trantírková, S., Löhr, F., Buck, J., Bongartz, E., Bamberg, E., Schwalbe, H., Dötsch, V., Trantírek, L., 2009. Evaluation of parameters critical for observing nucleic acids inside living *Xenopus laevis* oocytes by in-cell NMR spectroscopy. *J. Am. Chem. Soc.* 131, 15761–15768.
- Haq, I., Trent, J.O., Chowdhry, B.Z., Jenkins, T.C., 1999. Intercalative G-Tetraplex Stabilization of Telomeric DNA by a. *Biochemistry* 1768–1779.
- Harvey Lodish, Arnold Berk, Lawrence Zipursky, Paul Matsudaira, David Baltimore, and J.D., 2000. *Molecular Cell Biology* - NCBI Bookshelf. 4th Ed. New York W.H.Freeman Co Ltd.
- Havens, M.A., Hastings, M.L., 2016. Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res.* 44, 6549–6563.
- Hernández-Martínez, R., Covarrubias, L., 2011. Interdigital cell death function and regulation: New insights on an old programmed cell death model. *Dev. Growth Differ.* 53, 245–258.
- Hirashima, K., Seimiya, H., 2015. Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic Acids Res.* 43, 2022–2032.
- Huang, H., Zhang, J., Harvey, S.E., Hu, X., Cheng, C., 2017. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev.* 31, 2296–2309.
- Hud, N. V., Schultze, P., Sklenář, V., Feigon, J., 1999. Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex. *J. Mol. Biol.* 285, 233–243.
- Huppert, J.L., Balasubramanian, S., 2005. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 33, 2908–2916.
- Huppert, J.L., Bugaut, A., Kumari, S., Balasubramanian, S., 2008. G-quadruplexes: The beginning and end of UTRs. *Nucleic Acids Res.* 36, 6260–6268.
- Ikeda, R.A., Richardson, C.C., 1987. Interactions of a proteolytically nicked RNA polymerase of bacteriophage T7 with its promoter. *J. Biol. Chem.* 262, 3800–3808.
- Inoue, A., Yamamoto, N., Kimura, M., Nishio, K., Yamane, H., Nakajima, K., 2014. RBM10 regulates alternative splicing. *FEBS Lett.* 588, 942–947.
- Jain, S.C., Bhandary, K.K., Sobell, H.M., 1979. Visualization of drug-nucleic acid interactions at atomic resolution. *J. Mol. Biol.* 135, 813–840.
- Jeener, J., Meier, B.H., Bachmann, P., Ernst, R.R., 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.*
- Jin, X., Gossett, D.R., Wang, S., Yang, D., Cao, Y., Chen, J., Guo, R., Reynolds, R.K., Lin, J., 2004. Inhibition of AKT survival pathway by a small molecule inhibitor in human endometrial cancer cells. *Br. J. Cancer* 91, 1808–1812.
- Jin, Z., El-Deiry, W.S., 2005. Overview of cell death signaling pathways. *Cancer Biol. Ther.* 4, 139–163.
- Job, P., 1928, Formation and Stability of Inorganic Complexes in Solution. *Ann. Chim.*, 9, 113–203.
- Jodoin, R., Bauer, L., Garant, J.M., Laaref, A.M., Phaneuf, F., Perreault, J.P., 2014. The folding of 5'-UTR human G-quadruplexes possessing a long central loop. *Rna* 20, 1129–1141.

- Kaiser, C.E., Gokhale, V., Yang, D., Hurley, L.H., 2013. Gaining insights into the small molecule targeting of the G-quadruplex in the c-MYC promoter using NMR and an allele-specific transcriptional assay. *Top. Curr. Chem.* 330, 1–21.
- Kantari, C., Walczak, H., 2011. Caspase-8 and Bid: Caught in the act between death receptors and mitochondria. *Biochim. Biophys. Acta - Mol. Cell Res.* 1813, 558–563.
- Kanwal, F., Chen, T., Zhang, Y., Simair, A., Rujie, C., Sadaf Zaidi, N.U.S., Guo, X., Wei, X., Siegel, G., Lu, C., 2018. Large-Scale in Vitro Transcription, RNA Purification and Chemical Probing Analysis. *Cell. Physiol. Biochem.* 48, 1915–1927.
- Katz, A.M., Tolokh, I.S., Pabit, S.A., Baker, N., Onufriev, A. V., Pollack, L., 2017. Spermine Condenses DNA, but Not RNA Duplexes. *Biophys. J.* 112, 22–30.
- Kejnovská, I., Kypr, J., Vorlíčková, M., 2003. Circular dichroism spectroscopy of conformers of (guanine + adenine) repeat strands of DNA. *Chirality* 15, 584–592.
- Kejnovská, I., Renčiuk, D., Palacký, J., Vorlíčková, M., 2019. CD Study of the G-Quadruplex Conformation. *Methods Mol. Biol.*
- Kelley, S.K., Harris, L.A., Xie, D., DeForge, L., Totpal, K., Bussiere, J., Fox, J.A., 2001. Preclinical Studies to Predict the Disposition of Apo2L/Tumor Necrosis Factor-Related Apoptosis-Inducing Ligand in Humans: Characterization of in Vivo Efficacy, Pharmacokinetics, and Safety. *J. Pharmacol. Exp. Ther.* 299, 31 LP – 38.
- Keren, H., Lev-Maor, G., Ast, G., 2010. Alternative splicing and evolution: Diversification, exon definition and function. *Nat. Rev. Genet.* 11, 345–355.
- Kerkour, A., Marquevielle, J., Ivashchenko, S., Yatsunyk, L.A., Mergny, J.L., Salgado, G.F., 2017. High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation. *J. Biol. Chem.* 292, 8082–8091.
- Kerr, A.H.W.A.A.R.Currie., 1972. Apoptosis: a Basic Biological Phenomenon With Wide-Ranging Implications in Tissue Kinetics. *J. Intern. Med.* 258, 479–517.
- Kikin, O., D’Antonio, L., Bagga, P.S., 2006. QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 34, 676–682.
- Kim, E., Goren, A., Ast, G., 2008. Alternative splicing: Current perspectives. *BioEssays* 30, 38–47.
- Kim, M.S., Haney, M.J., Zhao, Y., Mahajan, V., Deygen, I., Klyachko, N.L., Inskoe, E., Piroyan, A., Sokolsky, M., Okolie, O., Hingtgen, S.D., Kabanov, A. V., Batrakova, E. V., 2016. Development of exosome-encapsulated paclitaxel to overcome MDR in cancer cells. *Nanomedicine Nanotechnology, Biol. Med.*
- Kistler, A.L., Guthrie, C., 2001. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase. *Genes Dev.* 15, 42–49.
- Kline, M.P., Rajkumar, S. V., Timm, M.M., Kimlinger, T.K., Haug, J.L., Lust, J.A., Greipp, P.R., Kumar, S., 2007. ABT-737, an inhibitor of Bcl-2 family proteins, is a potent inducer of apoptosis in multiple myeloma cells. *Leukemia*.
- Knudson, A.G., 1971. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci.* 68, 820–823.
- Kochetkov, S.N., Rusakova, E.E., Tunitskaya, V.L., 1998. Recent studies of T7 RNA polymerase mechanism. *FEBS Lett.* 440, 264–267.
- Koeppel, F., Riou, J.F., Laoui, A., Mailliet, P., Arimondo, P.B., Labit, D., Petitgenet, O., Hélène,

- C., Mergny, J.L., 2001. Ethidium derivatives bind to G-quartets, inhibit telomerase and act as fluorescent probes for quadruplexes. *Nucleic Acids Res.* 29, 1087–1096.
- Kogut, M., Kleist, C., Czub, J., 2019. Why do G-quadruplexes dimerize through the 5'-ends? Driving forces for G4 DNA dimerization examined in atomic detail. *PLoS Comput. Biol.* 15, 1–22.
- Kohn, K.W., Waring, M.J., Glaubiger, D., Friedman, C.A., 1975. Intercalative Binding of Ellipticine to DNA. *Cancer Res.* 35, 71 LP – 76.
- Kong, D.M., Ma, Y.E., Guo, J.H., Yang, W., Shen, H.X., 2009a. Fluorescent sensor for monitoring structural changes of G-Quadruplexes and detection of potassium ion. *Anal. Chem.* 81, 2678–2684.
- Kong, D.M., Ma, Y.E., Wu, J., Shen, H.X., 2009b. Discrimination of G-quadruplexes from duplex and single-stranded DNAs with fluorescence and energy-transfer fluorescence spectra of crystal violet. *Chem. - A Eur. J.* 15, 901–909.
- kong, J.N., Zhang, C., Zhu, Y.C., Zhong, K., Wang, J., Chu, B.B., Yang, G.Y., 2018. Identification and characterization of G-quadruplex formation within the EP0 promoter of pseudorabies virus. *Sci. Rep.* 8, 1–13.
- Kolesnikova, S., Hubálek, M., Bednářová, L., Cvačka, J., Curtis, E.A., 2017. Multimerization rules for G-quadruplexes. *Nucleic Acids Res.* 45, 8684–8696.
- Koren, E., Lev-Maor, G., Ast, G., 2007. The emergence of alternative 3' and 5' splice site exons from constitutive exons. *PLoS Comput. Biol.* 3, 0895–0908.
- Kralovicova, J., Lages, A., Patel, A., Dhir, A., Buratti, E., Searle, M., Vorechovsky, I., 2014. Optimal antisense target reducing INS intron 1 retention is adjacent to a parallel G quadruplex. *Nucleic Acids Res.* 42, 8161–8173.
- Kreig, A., Calvert, J., Sanoica, J., Cullum, E., Tipanna, R., Myong, S., 2015. G-quadruplex formation in double strand DNA probed by NMM and CV fluorescence. *Nucleic Acids Res.* 43, 7961–7970.
- Kumari, R., Nambiar, M., Shanbagh, S., Raghavan, S.C., 2015. Detection of G-quadruplex DNA using primer extension as a tool. *PLoS One* 10, 7–10.
- Kumari, S., Bugaut, A., Huppert, J.L., Balasubramanian, S., 2008. N-Ras Gquad 3, 218–221.
- Kwok, C.K., Balasubramanian, S., 2015. Targeted detection of g-quadruplexes in cellular RNAs. *Angew. Chemie - Int. Ed.* 54, 6751–6754.
- Lacroix, L., Séosse, A., Mergny, J.L., 2011. Fluorescence-based duplex-quadruplex competition test to screen for telomerase RNA quadruplex ligands. *Nucleic Acids Res.* 39.
- Lakowicz, J.R., 2006. Principles of fluorescence spectroscopy, 3rd Principles of fluorescence spectroscopy, Springer, New York, USA, 3rd edn, 2006., Principles of fluorescence spectroscopy, Springer, New York, USA, 3rd edn, 2006.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., Levine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, Christina, Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A.,

- Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissole, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Hong, M.L., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., De La Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglu, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M. V., Kaul, R., Raymond, Christopher, Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., 2001. Erratum: Initial sequencing and analysis of the human genome: International Human Genome Sequencing Consortium (Nature (2001) 409 (860-921)). Nature 412, 565–566.
- Lattmann, S., Giri, B., Vaughn, J.P., Akman, S.A., Nagamine, Y., 2010. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. Nucleic Acids Res. 38, 6219–6233.
- Le, H.T., Miller, M.C., Buscaglia, R., Dean, W.L., Holt, P.A., Chaires, J.B., Trent, J.O., 2012. Not all G-quadruplexes are created equally: An investigation of the structural polymorphism of the c-Myc G-quadruplex-forming sequence and its interaction with the porphyrin TMPyP4. Org. Biomol. Chem. 10, 9393–9404.
- Le, V.H., Nagesh, N., Lewis, E.A., 2013. Bcl-2 Promoter Sequence G-Quadruplex Interactions with Three Planar and Non-Planar Cationic Porphyrins: TMPyP4, TMPyP3, and TMPyP2. PLoS One 8, 1–9.
- Lee, E.F., Czabotar, P.E., Smith, B.J., Deshayes, K., Zobel, K., Colman, P.M., Fairlie, W.D., 2007. Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of

- antagonists of the Bcl-2 family. *Cell Death Differ.* 14, 1711.
- Lesage, S., Hao Xu, Durham, L., 1993. The occurrence and roles of porphyrins in the environment: possible implications for bioremediation. *Hydrol. Sci. Journal/Journal des Sci. Hydrol.* 38, 343–354.
- Li, T., Ho, H.H., Maslak, M., Schick, C., Martin, C.T., 1996. Major groove recognition elements in the middle of the T7 RNA polymerase promoter. *Biochemistry* 35, 3722–3727.
- Li, Q., Xiang, J.-F., Yang, Q.-F., Sun, H.-X., Guan, A.-J., Tang, Y.-L., 2013. G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* 41, D1115–D1123.
- Li, Z., Li, Q., Han, L., Tian, N., Liang, Q., Li, Y., Zhao, X., Du, C., Tian, Y., 2016. Pro-apoptotic effects of splice-switching oligonucleotides targeting Bcl-x pre-mRNA in human glioma cell lines. *Oncol. Rep.* 35, 1013–1019.
- Lim, K.W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D.J., Luu, K.N., Phan, A.T., 2009. Structure of the human telomere in K<sup>+</sup> solution: A stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.*
- Lisgarten, J.N., Coll, M., Portugal, J., Wright, C.W., Aymami, J., 2002. The antimalarial and cytotoxic drug cryptolepine intercalates into DNA at cytosine-cytosine sites. *Nat. Struct. Biol.* 9, 57–60.
- Liu, H.Y., Chen, A.C., Yin, Q.K., Li, Z., Huang, S.M., Du, G., He, J.H., Zan, L.P., Wang, S.K., Xu, Y.H., Tan, J.H., Ou, T.M., Li, D., Gu, L.Q., Huang, Z.S., 2017. New Disubstituted Quindoline Derivatives Inhibiting Burkitt's Lymphoma Cell Proliferation by Impeding c-MYC Transcription. *J. Med. Chem.* 60, 5438–5454.
- Liu, W., Lin, C., Wu, G., Dai, J., Chang, T.C., Yang, D., 2019. Structures of 1:1 and 2:1 complexes of BMVC and MYC promoter G-quadruplex reveal a mechanism of ligand conformation adjustment for G4-recognition. *Nucleic Acids Res.* 47, 11931–11942.
- Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngrin-Molango, S., Sprangers, R., Zanier, K., Krämer, A., Sattler, M., 2001. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* (80-. ). 294, 1098–1102.
- Loontjens, F.G., Dumortier, L., Regenfuss, P., Zechel, A., Clegg, R.M., 1990. Binding Characteristics of Hoechst 33258 with Calf Thymus DNA, Poly[d(A-T)], and d(CCGGAATTCCGG): Multiple Stoichiometries and Determination of Tight Binding with a Wide Spectrum of Site Affinities. *Biochemistry* 29, 9029–9039.
- Malgowska, M., Czajczynska, K., Gudanis, D., Tworak, A., Gdaniec, Z., 2016. Overview of the RNA G-quadruplex structures. *Acta Biochim. Pol.* 63, 609–621.
- Marcel, V., Tran, P.L.T., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.P., Hall, J., Mergny, J.L., Hainaut, P., 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.
- Martadinata, H., Phan, A.T., 2009. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K<sup>+</sup> solution. *J. Am. Chem. Soc.* 131, 2570–2579.
- Martadinata, H., Phan, A.T., 2013. Structure of human telomeric RNA (TERRA): Stacking of two G-quadruplex blocks in K<sup>+</sup> solution. *Biochemistry* 52, 2176–2183.
- Massiello, A., Roesser, J.R., Chalfant, C.E., Massiello, A., Roesser, J.R., Chalfant, C.E., 2006. SAP155 Binds to ceramide-responsive RNA cis-element 1 and regulates the alternative 5' splice site selection of Bcl-x pre-mRNA. *FASEB J.* 20, 1680–1682.

- Matlin, A.J., Clark, F., Smith, C.W.J., 2005. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6, 386–398.
- Mercatante, D.R., Bortner, C.D., Cidlowski, J.A., Kole, R., 2001. Modification of alternative splicing of Bcl-x Pre-mRNA in prostate and breast cancer cells: Analysis of apoptosis and cell death. *J. Biol. Chem.* 276, 16411–16417.
- Mercatante, D.R., Mohler, J.L., Kole, R., 2002. Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents. *J. Biol. Chem.* 277, 49374–49382.
- Merdzhanova, G., Edmond, V., De Seranno, S., Van den Broeck, A., Corcos, L., Brambilla, C., Brambilla, E., Gazzeri, S., Eymin, B., 2008. E2F1 controls alternative splicing pattern of genes involved in apoptosis through upregulation of the splicing factor SC35. *Cell Death Differ.* 15, 1815–1823.
- Mergny, J.L., Lacroix, L., 2009. UV melting of G-quadruplexes. *Curr. Protoc. Nucleic Acid Chem.* 1–15.
- Mergny, J.L., Li, J., Lacroix, L., Amrane, S., Chaires, J.B., 2005. Thermal difference spectra: A specific signature for nucleic acid structures. *Nucleic Acids Res.* 33, 1–6.
- Michelle, L., Cloutier, A., Toutant, J., Shkreta, L., Thibault, P., Durand, M., Garneau, D., Gendron, D., Lapointe, E., Couture, S., Le Hir, H., Klinck, R., Elela, S.A., Prinos, P., Chabot, B., 2012. Proteins Associated with the Exon Junction Complex Also Control the Alternative Splicing of Apoptotic Regulators. *Mol. Cell. Biol.* 32, 954–967.
- Mikami-Terao, Y., Akiyama, M., Yuza, Y., Yanagisawa, T., Yamada, O., Kawano, T., Agawa, M., Ida, H., Yamada, H., 2009. Antitumor activity of TMPyP4 interacting G-quadruplex in retinoblastoma cell lines. *Exp. Eye Res.* 89, 200–208.
- Milligan, J.F., Groebe, D.R., Witherell, G.W., Uhlenbeck, O.C., 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* 15, 8783–8798.
- Mishra, S.K., Jain, N., Shankar, U., Tawani, A., Sharma, T.K., Kumar, A., 2019. Characterization of highly conserved G-quadruplex motifs as potential drug targets in *Streptococcus pneumoniae*. *Sci. Rep.* 9, 1–13.
- Miskiewicz, J., Sarzynska, J., Szachniuk, M., 2020. How bioinformatics resources work with G4 RNAs. *Brief. Bioinform.* 00, 1–14.
- Modrek, B., Lee, C., 2002. A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19.
- Momand, J., Jung, D., Wilczynski, S., Niland, J., 1998. The MDM2 gene amplification database. *Nucleic Acids Res.* 26, 3453–3459.
- Monni, O., Joensuu, H., Franssila, K., Klefstrom, J., Alitalo, K., Knuutila, S., 1997. BCL2 overexpression associated with chromosomal amplification in diffuse large B-cell lymphoma. *Blood* 90, 1168–74.
- Moore, M.J., Query, C.C., Sharp, P. a, 1993. Splicing of precursors to messenger RNAs by the spliceosome, *The RNA world*.
- Moore, M.J., Wang, Q., Kennedy, C.J., Silver, P.A., 2010. An alternative splicing network links cell-cycle control to apoptosis. *Cell* 142, 625–636.
- Morris, M.J., Wingate, K.L., Silwal, J., Leeper, T.C., 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 Res.* 40, 4137–4145.



- Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L., Bryan, T.M., 2015. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* 6.
- Mozaffari-Jovin, S., Santos, K.F., Hsiao, H.H., Will, C.L., Urlaub, H., Wahl, M.C., Lührmann, R., 2012. The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes Dev.* 26, 2422–2434.
- Muchmore, S.W., Sattlert, M., Liangt, H., Meadows, R.P., Harlant, J.E., Sup Yoont, H., Nettesheimt, D., Chang, B.S., Thompson, C.B., Wongll, S.-L., Ngll, S.-C., Fesikt, S.W., 1996. X-ray and NMR structure of human Bcl-xu an inhibitor of programmed cell death *LETTERS TO NATURE. C. Nucleic Acids Res* 381, 1747–1754.
- Muller, D.K., Martin, C.T., Coleman, J.E., 1988. Processivity of proteolytically modified forms of T7 RNA polymerase. *Biochemistry* 27, 5763–5771.
- Murat, P., Zhong, J., Lekieffre, L., Cowieson, N.P., Clancy, J.L., Preiss, T., Balasubramanian, S., Khanna, R., Tellam, J., 2014. G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* 10, 358–364.
- Murchie, A.I.H., Lilley, D.M.J., 1992. Retinoblastoma susceptibility genes contain 5' sequences with a high propensity to form guanine-tetrad structures. *Nucleic Acids Res.* 20, 49–53.
- Nagata, S., Golstein, P., 1995. The Fas Death Factor. *Science* (80-. ). 267, 1449–1456.
- Nagesh, N., Buscaglia, R., Dettler, J.M., Lewis, E.A., 2010. Studies on the site and mode of TMPyP4 interactions with Bcl-2 promoter sequence G-quadruplexes. *Biophys. J.* 98, 2628–2633.
- Naro, C., Barbagallo, F., Chieffi, P., Bourgeois, C.F., Paronetto, M.P., Sette, C., 2014. The centrosomal kinase NEK2 is a novel splicing factor kinase involved in cell survival. *Nucleic Acids Res.* 42, 3218–3227.
- Nelissen, R.L.H., Will1, C.L., Van Venrooij, W.J., Luhrmann1, R., 1994. The association of the UI-specific 70K and C proteins with UI snRNPs is mediated in part by common U snRNP proteins. *EMBO J.* 13, 4113–4125.
- Nicoludis, J.M., Barrett, S.P., Mergny, J.L., Yatsunyk, L.A., 2012a. Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.* 40, 5432–5447.
- Nicoludis, J.M., Miller, S.T., Jeffrey, P.D., Barrett, S.P., Rablen, P.R., Lawton, T.J., Yatsunyk, L.A., 2012b. Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.* 134, 20446–20456.
- Oganesian, L., Moon, I.K., Bryan, T.M., Jarstfer, M.B., 2006. Extension of G-quadruplex DNA by ciliate telomerase. *EMBO J.* 25, 1148–1159.
- Ohashi, M., Sugikawa, E., Nakanishi, N., 1995. Inhibition of p53 protein phosphorylation by 9-hydroxyellipticine: a possible anticancer mechanism. *Jpn. J. Cancer Res.* 86, 819–827.
- Olmsted, J., Kearns, D.R., 1977. Mechanism of Ethidium Bromide Fluorescence Enhancement on Binding to Nucleic Acids. *Biochemistry*.
- Otim, B., 2018. An investigation into the effectiveness of sunlight in disinfecting water from different sources.
- Parker, R., Siliciano, P.G., Guthrie, C., 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49, 229–239.
- Parkinson, G.N., Ghosh, R., Neidle, S., 2007. Structural basis for binding of porphyrin to human telomeres. *Biochemistry* 46, 2390–2397.
- Parkinson, G.N., Lee, M.P.H., Neidle, S., 2002. Crystal structure of parallel quadruplexes from

- human telomeric DNA. *Nature* 417, 876–880.
- Paronetto, M.P., Achsel, T., Massiello, A., Chalfant, C.E., Sette, C., 2007. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J. Cell Biol.* 176, 929–939.
- Passon, D.M., Lee, M., Rackham, O., Stanley, W.A., Sadowska, A., Filipovska, A., Fox, A.H., Bond, C.S., 2012. Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proc. Natl. Acad. Sci. U. S. A.* 109, 4846–4850.
- Patel, Dinshaw J and Tonelli, A.E., 1974. Proton Nuclear Magnetic Spectroscopy 71, 1945–1948.
- Patton, J.G., Porro, E.B., Galceran, J., Tempst, P., Nadal-Ginard, B., 1993. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.* 7, 393–406.
- Pedrotti, S., Busà, R., Compagnucci, C., Sette, C., 2012. The RNA recognition motif protein RBM11 is a novel tissue-specific splicing regulator. *Nucleic Acids Res.* 40, 1021–1032.
- Petros, A.M., Nettesheim, D.G., Wang, Y., Olejniczak, E.T., Meadows, R.P., Mack, J., Swift, K., Matayoshi, E.D., Zhang, H., Fesik, S.W., Thompson, C.B., 2000. Rationale for Bcl-X L /Bad peptide complex formation from structure, mutagenesis, and biophysical studies . *Protein Sci.* 9, 2528–2534.
- Phan, A.T., Kuryavyi, V., Darnell, J.C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R.B., Patel, D.J., 2011. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* 18, 796–804.
- Phan, A.T., Kuryavyi, V., Gaw, H.Y., Patel, D.J., 2005. Small-molecule interaction with a five-guanine-tract g-quadruplex structure from the human myc promoter. *Nat. Chem. Biol.* 1, 167–173.
- Phan, A.T., Modi, Y.S., Patel, D.J., 2004. Two-repeat Tetrahymena telomeric d(TGGGGTTGGGGT) sequence interconverts between asymmetric dimeric G-quadruplexes in solution. *J. Mol. Biol.* 338, 93–102.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K.W., Li, J., Nagai, K., 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 resolution. *Nature* 458, 475–480.
- Puglisi, J.D., Tinoco, I., 1989. Absorbance melting curves of RNA. *Methods Enzymol.* 180, 304–325.
- Redon, S., Reichenbach, P., Lingner, J., 2010. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* 38, 5797–5806.
- Rehman, A.G., Booth, C., Potten, C.S., 2001. What is apoptosis, and why is it important? *BMJ* 322, 1536–1538.
- Renny, J.S., Tomasevich, L.L., Tallmadge, E.H., Collum, D.B., 2013. Method of continuous variations: Applications of job plots to the study of molecular associations in organometallic chemistry. *Angew. Chemie - Int. Ed.*
- Revil, T., Pelletier, J., Toutant, J., Cloutier, A., Chabot, B., 2009. Heterogeneous nuclear ribonucleoprotein K represses the production of pro-apoptotic Bcl-xS splice isoform. *J. Biol. Chem.* 284, 21458–21467.
- Reyes, J.L., Kois, P., Konforti, B.B., Konarska, M.M., 1996. The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. *Rna* 2, 213–225.
- Ribeiro, M.M., Teixeira, G.S., Martins, L., Marques, M.R., de Souza, A.P., Line, S.R.P., 2015. G-

- quadruplex formation enhances splicing efficiency of PAX9 intron 1. *Hum. Genet.* 134, 37–44.
- Roca, X., Krainer, A.R., Eperon, I.C., 2013. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev.* 27, 129–144.
- Rogers, J., Wall, R., 1980. A mechanism for RNA splicing. *Proc. Natl. Acad. Sci. U. S. A.*
- Roth K., 1984, Basic Principles of NMR Spectroscopy. In: *NMR-Tomography and -Spectroscopy in Medicine*. Springer, Berlin, Heidelberg.
- Sabharwal, N.C., Savikhin, V., Turek-Herman, J.R., Nicoludis, J.M., Szalai, V.A., Yatsunyk, L.A., 2014. N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J.* 281, 1726–1737.
- Sakabe, N.J., de Souza, S.J., 2007. Sequence features responsible for intron retention in human. *BMC Genomics* 8, 1–14.
- Samatanga, B., Dominguez, C., Jelesarov, I., Allain, F.H.T., 2013. The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic Acids Res.* 41, 2505–2516.
- Seenisamy, J., Bashyam, S., Gokhale, V., Vankayalapati, H., Sun, D., Siddiqui-Jain, A., Streiner, N., Shin-ya, K., White, E., Wilson, W.D., Hurley, L.H., 2005. Design and synthesis of an expanded porphyrin that has selectivity for the c-MYC G-quadruplex structure. *J. Am. Chem. Soc.* 127, 2944–2959.
- Sen, D., Gilbert, W., 1990. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* 344, 410–414.
- Severinov, K., 2001. T7 RNA polymerase transcription complex: What you see is not what you get. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5–7.
- Shangary, S., Wang, S., 2009. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: A novel approach for cancer therapy. *Annu. Rev. Pharmacol. Toxicol.*
- Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H., Chen, G., Sun, H., Zhang, Y., Denise, A., Zhang, D.E., Fu, X.D., 2014. Mechanisms for U2AF to define 5' splice sites and regulate alternative splicing in the human genome. *Nat. Struct. Mol. Biol.* 21, 997–1005.
- Sharma, S., Wongpalee, S.P., Vashisht, A., Wohlschlegel, J.A., Black, D.L., 2014. Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev.* 28, 2518–2531.
- Shav-Tal, Y., Cohen, M., Lapter, S., Dye, B., Patton, J.G., Vandekerckhove, J., Zipori, D., 2001. Nuclear relocalization of the pre-mRNA splicing factor PSF during apoptosis involves hyperphosphorylation, masking of antigenic epitopes, and changes in protein interactions. *Mol. Biol. Cell* 12, 2328–2340.
- Shay, J.W., Bacchetti, S., 1997. A survey of telomerase activity in human cancer. *Eur. J. Cancer Part A* 33, 787–791.
- Shieh, J.J., Liu, K.T., Huang, S.W., Chen, Y.J., Hsieh, T.Y., 2009. Modification of alternative splicing of Mcl-1 pre-mRNA using antisense morpholino oligonucleotides induces apoptosis in basal cell carcinoma cells. *J. Invest. Dermatol.* 129, 2497–2506.
- Shkreta, L., Toutant, J., Durand, M., Manley, J.L., Chabot, B., 2016. SRSF10 Connects DNA

- Damage to the Alternative Splicing of Transcripts Encoding Apoptosis, Cell-Cycle Control, and DNA Repair Factors. *Cell Rep.* 17, 1990–2003.
- Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., 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. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11593–11598.
- Simko, E.A.J., Liu, H., Zhang, T., Velasquez, A., Teli, S., Haeusler, A.R., Wang, J., 2020. G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural lncRNA. *Nucleic Acids Res.* 48, 7421–7438.
- Slee, E.A., Adrain, C., Martin, S.J., 2001. Executioner Caspase-3, -6, and -7 Perform Distinct, Non-redundant Roles during the Demolition Phase of Apoptosis. *J. Biol. Chem.* 276, 7320–7326.
- Smale, S.T., 2010. Luciferase assay. *Cold Spring Harb. Protoc.* 2010, pdb.prot5421.
- Smith, F.W., Feigon, J., 1992. *jvlt* 356, 164–168.
- Smith, F.W., Schultze, P., Feigon, J., 1995. Solution structures of unimolecular quadruplexes formed by oligonucleotides containing Oxytricha telomere repeats. *Structure* 3, 997–1008.
- Smith, L.D., Dickinson, R.L., Lucas, C.M., Cousins, A., Malygin, A.A., Weldon, C., Perrett, A.J., Bottrill, A.R., Searle, M.S., Burley, G.A., Eperon, I.C., 2014. A targeted oligonucleotide enhancer of SMN2 Exon 7 splicing forms competing quadruplex and protein complexes in functional conditions. *Cell Rep.* 9, 193–205.
- Song, J., Perreault, J.-P., Topisirovic, I., Richard, S., 2016. RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* 4, e1244031.
- Staley, J.P., Guthrie, C., 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* 92, 315–326.
- Stark, H., Dube, P., Luührmann, R., Kastner, B., 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 409, 539–542.
- Stevens, M., Oltean, S., 2019. Modulation of the Apoptosis Gene Bcl-x Function Through Alternative Splicing. *Front. Genet.* 10, 1–9.
- Stiborová, M., Bořek-Dohalská, L., Aimová, D., Kotrbová, V., Kukačková, K., Janouchová, K., Rupertová, M., Ryšlavá, H., Hudeček, J., Frei, E., 2006. Oxidation pattern of the anticancer drug ellipticine by hepatic microsomes - Similarity between human and rat systems. *Gen. Physiol. Biophys.* 25, 245–261.
- Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J.L., Moine, H., 2011. G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.* 12, 697–704.
- Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S., Hurley, L.H., 1997. Inhibition of human telomerase by a G-Quadruplex-Interactive compound. *J. Med. Chem.* 40, 2113–2116.
- Sun, Z.Y., Wang, X.N., Cheng, S.Q., Su, X.X., Ou, T.M., 2019. Developing novel G-quadruplex ligands: From interaction with nucleic acids to interfering with nucleic acid–protein interaction. *Molecules* 24.
- Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., Oyoshi, T., 2013. Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem. Biol.* 20, 341–350.
- Tas, S.W., Quartier, P., Botto, M., Fossati-Jimack, L., 2006. Macrophages from patients with SLE and rheumatoid arthritis have defective adhesion in vitro, while only SLE

- macrophages have impaired uptake of apoptotic cells. *Ann. Rheum. Dis.* 65, 216–221.
- Tatton, W.G., Olanow, C.W., 1999. Apoptosis in neurodegenerative diseases: The role of mitochondria. *Biochim. Biophys. Acta - Bioenerg.* 1410, 195–213.
- Tewey, K.M., Chen, G.L., Nelson, E.M., Liu, L.F., 1984. Intercalative antitumor drugs interfere with the breakage-reunion reaction of mammalian DNA topoisomerase II. *J. Biol. Chem.* 259, 9182–9187.
- Tippana, R., Chen, M.C., Demeshkina, N.A., Ferré-D'Amaré, A.R., Myong, S., 2019. RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36. *Nat. Commun.* 10, 1–10.
- Tippana, R., Xiao, W., Myong, S., 2014. G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* 42, 8106–8114.
- Umar, M.I., Ji, D., Chan, C.Y., Kwok, C.K., 2019. G-quadruplex-based fluorescent turn-on ligands and aptamers: From development to applications, *Molecules*.
- Vannutelli, A., Belhamiti, S., Garant, J.-M., Ouangraoua, A., Perreault, J.-P., 2020. Where are G-quadruplexes located in the human transcriptome? *NAR Genomics Bioinforma.* 2, 1–14.
- Varani, L., Spillanti, M.G., Goedert, M., Varani, G., 2000. Structural basis for recognition of the RNA major groove in the tau exon 10 splicing regulatory element by aminoglycoside antibiotics. *Nucleic Acids Res.* 28, 710–719.
- Vaughn, J.P., Creacy, S.D., Routh, E.D., Joyner-Butt, C., Jenkins, G.S., Pauli, S., Nagamine, Y., Akman, S.A., 2005. The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* 280, 38117–38120.
- Vendôme, J., Letard, S., Martin, F., Svinarchuk, F., Dubreuil, P., Auclair, C., Le Bret, M., 2005. Molecular modeling of wild-type and D816V c-kit inhibition based on ATP-competitive binding of ellipticine derivatives to tyrosine kinases. *J. Med. Chem.* 48, 6194–6201.
- Verma, S., Ghuge, S., Velayutham, R., Ranjan, N., 2018. Spectroscopic studies of Thioflavin-T binding to c-Myc G-quadruplex DNA. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 212.
- Viladoms, J., Parkinson, G.N., 2014. HELIX: A new modular nucleic acid crystallization screen. *J. Appl. Crystallogr.* 47, 948–955.
- Von Pawel, J., Harvey, J.H., Spigel, D.R., Dediu, M., Reck, M., Cebotaru, C.L., Humphreys, R.C., Gribbin, M.J., Fox, N.L., Camidge, D.R., 2014. Phase II trial of mapatumumab, a fully human agonist monoclonal antibody to tumor necrosis factor-related apoptosis-inducing ligand receptor 1 (TRAIL-R1), in combination with paclitaxel and carboplatin in patients with advanced non-small-cell lung cancer. *Clin. Lung Cancer* 15, 188-196.e2.
- Vorlíčková, M., Kejnovská, I., Sagi, J., Renčíuk, D., Bednářová, K., Motlová, J., Kypr, J., 2012. Circular dichroism and guanine quadruplexes. *Methods*.
- Wahl, M.C., Will, C.L., Lührmann, R., 2009. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* 136, 701–718.
- Wang, Y., Chen, D., Qian, H., Tsai, Y.S., Shao, S., Liu, Q., Dominguez, D., Wang, Z., 2014. The Splicing Factor RBM4 Controls Apoptosis, Proliferation, and Migration to Suppress Tumor Progression. *Cancer Cell* 26, 374–389.
- Wang, Y., Patel, D.J., 1993. Solution structure of the human telomeric repeat d[AG3(T2AG3)3]

- G-tetraplex. *Structure* 1, 263–282.
- Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J., Block, S.M., 1998. Force and velocity measured for single molecules of RNA polymerase. *Science* (80-. ). 282, 902–907.
- Webba da Silva, M., 2007. NMR methods for studying quadruplex nucleic acids. *Methods* 43, 264–277.
- Weeks, K.M., Mauger, D.M., 2011. Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.*
- Weldon, C., Behm-Ansmant, I., Hurley, L.H., Burley, G.A., Branlant, C., Eperon, I.C., Dominguez, C., 2017. Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA. *Nat. Chem. Biol.* 13, 18–20.
- Weldon, C., Dacanay, J.G., Gokhale, V., Boddupally, P.V.L., Behm-Ansmant, I., Burley, G.A., Branlant, C., Hurley, L.H., Dominguez, C., Eperon, I.C., 2018. Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. *Nucleic Acids Res.* 46, 886–896.
- Weldon, C., Eperon, I.C., Dominguez, C., 2016. Do we know whether potential G-quadruplexes actually form in long functional RNA molecules? *Biochem. Soc. Trans.* 44, 1761–1768.
- Wiley, S.R., Schooley, K., Smolak, P.J., Din, W.S., Huang, C.P., Nicholl, J.K., Sutherland, G.R., Smith, T.D., Rauch, C., Smith, C.A., Goodwin, R.G., 1995. Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3, 673–682.
- Will, C.L., Lührmann, R., 1997. Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 9, 320–328.
- Wu, J., Manley, J.L., 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev.* 3, 1553–1561.
- Xu, S., Li, Q., Xiang, J., Yang, Q., Sun, H., Guan, A., 2016. Thioflavin T as an efficient fluorescence sensor for selective recognition of RNA G-quadruplexes. *Nat. Publ. Gr.* 1–9.
- Yamaoki, Y., Kiyoshi, A., Miyake, M., Kano, F., Murata, M., Nagata, T., Katahira, M., 2018. The first successful observation of in-cell NMR signals of DNA and RNA in living human cells. *Phys. Chem. Chem. Phys.* 20, 2982–2985.
- Yan, C., Wan, R., Bai, R., Huang, G., Shi, Y., 2017. Structure of a yeast step II catalytically activated spliceosome. *Science* (80-. ). 355, 149–155.
- Yoshida, H., Park, S.Y., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., Tsuda, K., Kuwasako, K., Unzai, S., Muto, Y., Urano, T., Obayashi, E., 2015. A novel 3' splice site recognition by the two zinc fingers in the U2AF small subunit. *Genes Dev.* 29, 1649–1660.
- Zahler, A.M., Williamson, J.R., Cech, T.R., Prescott, D.M., 1991. Inhibition of telomerase by G-quartet DNA structures. *Nature.*
- Zhang, C., Hastings, M.L., Krainer, A.R., Zhang, M.Q., 2007. Dual-specificity splice sites function alternatively as 5' and 3' splice sites. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15028–15033.
- Zhang, Z., Dai, J., Veliath, E., Jones, R.A., Yang, D., 2009. Structure of a two-G-tetrad intramolecular G-quadruplex formed by variant human telomeric sequence in K<sup>+</sup> solution: Insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.* 38, 1009–1021.
- Zheng, K.W., Xiao, S., Liu, J.Q., Zhang, J.Y., Hao, Y.H., Tan, Z., 2013. Co-transcriptional formation of DNA: RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.* 41, 5533–5541.
- Zhou, A., Ou, A.C., Cho, A., Benz, E.J., Huang, S.-C., 2008. Novel Splicing Factor RBM25 Modulates Bcl-x Pre-mRNA 5' Splice Site Selection. *Mol. Cell. Biol.* 28, 5924–5936.

- Zhou, J.L., Lu, Y.J., Ou, T.M., Zhou, J.M., Huang, Z.S., Zhu, X.F., Du, C.J., Bu, X.Z., Ma, L., Gu, L.Q., Li, Y.M., Chan, A.S.C., 2005. Synthesis and evaluation of quindoline derivatives as G-quadruplex inducing and stabilizing ligands and potential inhibitors of telomerase. *J. Med. Chem.* 48, 7315–7321.
- Zilfou, J.T., Lowe, S.W., 2009. Tumor suppressive functions of p53. *Cold Spring Harb. Perspect. Biol.* 1, 1–12.
- Zimmerman, S.B., Cohen, G.H., Davies, D.R., 1975. X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J. Mol. Biol.* 92, 181–192.
- Zlobina, M., Sedo, O., Chou, M.Y., Slepankova, L., Lukavsky, P.J., 2016. Efficient large-scale preparation and purification of short single-stranded RNA oligonucleotides. *Biotechniques* 60, 75–83.
- Zou, T., Sadler, P.J., 2015. Speciation of precious metal anti-cancer complexes by NMR spectroscopy. *Drug Discov. Today Technol.* 16, 7–15.
- Zuffo, M., Gandolfini, A., Heddi, B., Granzhan, A., 2020. Harnessing intrinsic fluorescence for typing of secondary structures of DNA. *Nucleic Acids Res.* 48, e61–e61.