Molecular tools and biophysical studies to probe the role of triplex DNA in Friedreich's Ataxia

Caitlin Mary Brown

Department of Chemistry

Imperial College London

Thesis submitted for the degree of Doctor of Philosophy

December 2019

Declaration of originality

I, Caitlin Brown, declare that the work presented herein was carried out by myself, and that any work performed by collaborators has been clearly attributed in the relevant sections. Furthermore, I declare that this thesis was written entirely by myself, and all source materials have been appropriately referenced and acknowledged.

Declaration of Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution Non-Commercial 4.0 International Licence (CC BY-NC).

Under this licence, readers are free to copy, distribute or transmit the material in any medium or format. Readers may also create and distribute modified versions of the work. This is on the condition that: they attribute it, that they do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, the user must ensure to make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, the user should indicate that the work has been changed and describe those changes.

Acknowledgements

I would first like to express my gratitude to my supervisors, Prof. Ramon Vilar, Prof. Richard Festenstein and Prof. Christian Speck who made this work possible with their continued support, guidance and kind words of encouragement throughout the project. I have been lucky to have three wonderful supervisors willing to share their expertise with me. I would also like to thank the Engineering and Physical Sciences Research Council (EPSRC) for funding this research.

Secondly, I would like to thank all the group members I have had the pleasure of working with throughout my time at Imperial. Particular thanks go to Dr Beata Klejevskava for introducing me to the world of weird DNA structures when I first began the project, Dr Peter Summers for the assistance in all things fluorescent, Dr Jackson Chan for teaching me molecular biology, Dr Indiana Magdalou for helping me solve my biological mysteries, and Dr Alice Pyne and Phil Haynes for the AFM work.

Thank you especially to the Vilar group (Vilar Village) for the fun times over the years; the office table-tennis sessions, frisbee breaks and Fireball shots will be missed! Special thanks to Tiffany and Rainbow for their moral support through the highs and lows.

And lastly, thank you to my friends and family for their continued belief in me.

I would like to dedicate this thesis to the memory of my Dad.

Thesis abstract

Friedreich's ataxia (FRDA) is a recessive inherited neurodegenerative disorder, characterised by a trinucleotide repeat expansion mutation on the *FXN* gene consisting of a $(GAA)_n$ base triplet, where n = 66-1700. This mutation results in the transcriptional silencing of the *FXN* gene. One hypothesis for how this mutation causes gene silencing is the formation of a persistent DNA, or DNA-RNA, triplex. The work presented herein focuses on the biophysical study of DNA and DNA-RNA triplexes formed from sequences containing the $(GAA)_n$ sequence of varying lengths, and subsequently the development of fluorescence lifetime probes with the aim to image triplexes *in cellulo.*

Initially, the formation and stability of triplexes made from short oligonucleotides containing the $(GAA)_n$ FRDA repeat sequence, where n = 5, 10 and 20, was studied under various pH and salt conditions. It was found that triplexes only formed under acidic conditions.

A molecular biology approach was taken to produce polynucleotides containing a $(GAA)_n$ FRDA repeat sequence of pathological length, where n = 55, 75 and 200. The formation of DNA-DNA and DNA-RNA hybrid structures under varying conditions was studied using gel electrophoresis and AFM. It was found that multiple higher order structures formed at acidic pH with DNA, and physiological pH with DNA-RNA. The DNA triplexes formed at acidic pH were visualised with AFM.

Lastly, fluorescence lifetime probes based on the structure of thiazole orange were synthesised, and their fluorescence properties when bound to different DNA topologies were investigated. It was found that all probes exhibited long fluorescence lifetimes when bound to triplex DNA, presenting a strategy for imaging triplexes. These probes were then incubated with cells containing a $(GAA)_{310}$ mutation and the fluorescence lifetime was measured using FLIM. However, no difference in average lifetime was seen between these cells containing the mutation and the wild type.

List of figures and tables

Table P1: List of figures from **Chapter 1**

Figure	*Caption*
1.1	Mature human frataxin (1EKG). Alpha helices (two) are displayed in blue, beta sheets in red (four) and loops in pink.[198]
1.2	Schematic showing intramolecular triplex formation during transcription. Once the duplex is denatured for transcription, non-template DNA strand (blue) folds back onto the duplex to form an intramolecular triplex while RNA polymerase is reading the template strand.
1.3	Schematic showing R-loop formation during transcription. Once the duplex is denatured for transcription, the RNA strand binds to the template strand of the duplex.
1.4	A solution structure of triplex DNA (1D3X). The two strands corresponding to the duplex are blue, while the TFO is shown in orange, bound in the major groove of the duplex.
1.5	Schematic showing the bonding in pyrimidine triplexes (left) and purine triplexes (right). Watson-Crick hydrogen bonding is indicated with a striped line, while Hoogsteen hydrogen bonding is indicated with a dashed line. The bases in the TFO are indicated with the coloured boxes; blue for pyrimidine and orange for purine.
1.6	The chemical structure of the backbone of DNA, PNA and LNA. The backbones of PNA and LNA lack the negatively charged phosphorus, reducing charge repulsion between strands. The LNA is conformationally restrained, enhancing base-stacking properties.
1.7	Solution structure of a G-quadruplex (2MS9, taken from the PDB).
1.8	Schematic demonstrating the principles of gel electrophoresis. An electric current is applied across the gel and the negatively charged DNA moves towards the positively charged terminal at different speeds depending on size and conformation.
1.9	Schematic demonstrating the principles of CD. Here, the right polarised light (blue) is absorbed to a larger extent than the left polarised light (green). The degree of ellipticity (θ) is given by the angle in the triangle made between the sum of the contributions (E_R+E_L) and the difference between the contributions (E_R-E_L).
1.10	Schematic demonstrating the principles of FRET. Here, when the G–quadruplex is unfolded (left), the donor fluorophore (green) emits upon excitation. However, when the G-quadruplex is folded (right), FRET occurs and the acceptor fluorophore (pink) emits.
1.11	Schematic demonstrating the use of AFM to image proteins embedded in live cell membranes. Tapping (or oscillation) mode allows imaging of fragile biological samples. The deflection of the cantilever is measured and mapped to the surface to give information about the topology.

1.12	Jablonski energy level diagram demonstrating excitation by absorption of a photon, partial relaxation by nonradiative decay and then decay back to ground state through fluorescence. Intersystem crossing is another form of nonradiative decay, in which energy is transferred from a singlet to a triplet excited state, and then quenched through phosphorescence.
1.13	The solution structure of TOTO intercalated into DNA, as determined by NMR spectroscopy (108D). TOTO is shown here in orange; the aromatic stacking interactions between the aromatic moieties of the TO monomer and the DNA bases is apparent.
1.14	Chemical structures of triplex intercalators (left), groove binders (right), and the structure of the neomycin-BQQ conjugate incorporating both intercalation and groove binding modes below.
1.15	Fig. 1.15: Netropsin (in green) bound to d(GGCCAATTGG) dsDNA in the minor groove (C and H in grey, O in red, N in blue and S in orange, 1Z8V).[199]
1.16	A) Chemical structure of triangulenium **DOATA-M2**, B) structure of **DAOTA-M2** (in pink) bound to G-quadruplex c-myc (5LIG), taken from the PDB.[151] **DOATA-M2** stacks on either end of the quadruplex structure, with the aromatic R substituents interacting with the side pockets of the quadruplex.[151]

Table P2: List of figures from **Chapter 2**

Figure	*Caption*
2.1	Schematic demonstrating the annealing of a triplex structure from the oligonucleotides through heating and then slow cooling. The two oligonucleotides comprising the duplex are coloured in blue, while the TFO is coloured in green.
2.2	Fig. 2.2: Schematic demonstrating the design of the fluorophore-labelled oligonucleotides and how the FRET occurs between the pair. The purine strand of the duplex (coloured blue) is labelled with TAMRA (pink), while the purine TFO (green) is labelled with FAM (yellow). Upon association of the strands, FRET can occur, increasing emission of TAMRA, the acceptor fluorophore.
2.3	15% PAGE of 5 repeat unit purine and pyrimidine sequences annealed in 10 mM NaCac buffer (pH 5.5) containing 100 mM NaCl and 30 mM $MgCl_2$. Ladder: O'RangeRuler 5 bp (Thermo Scientific). Lanes: 1) pyrimidine TFO, 2) duplex, 3) duplex with pyrimidine TFO, 4) 1:2 duplex with pyrimidine TFO, 5) purine TFO, 6) duplex, 7) duplex with purine TFO.
2.4	A) CD spectra of duplex and pyrimidine TFO annealed at pH 5.5 (pink) and pH 7.4 (orange). The characteristic negative band at 210 nm is present at pH 5.5. B) CD spectra of duplex and purine TFO annealed at pH 5.5 (pink) and pH 7.4 (orange). There is no characteristic triplex band present in either spectrum. C) CD melting spectra of pyrimidine triplex annealed at pH 5.5. D) CD melting spectra of purine triplex annealed at pH 7.4.
2.5	A) FRET melting profile of TAMRA emission for $(GAA)_5$ TAMRA-labelled duplex annealed with FAM-labelled purine TFO at pH 7.4. B) Fluorescence titration of increasing equivalents of $(GAA)_5$ FAM-labelled purine TFO into TAMRA-labelled duplex at pH 7.4.

2.6	12% PAGE of $(GAA)_{10}$ pyrimidine sequences annealed in 10 mM NaCac buffer (pH 5.5) containing 100 mM NaCl and 30 mM $MgCl_2$, and subsequently incubated in the same buffer at increasing pH. Ladder: O'RangeRuler 5 bp (Thermo Scientific). Lanes: 1) duplex, 2) duplex with pyrimidine TFO annealed at pH 5.5, 3) duplex with pyrimidine TFO incubated at pH 6, 4) duplex with pyrimidine TFO incubated at pH 6.5, 5) duplex with pyrimidine TFO incubated at pH 7, 6) duplex with pyrimidine TFO incubated at pH 7.4.
2.7	12% PAGE of $(GAA)_{10}$ purine sequences annealed in 10 mM NaCac buffer (pH 7.4) containing 15 mM $MgCl_2$. Ladder: O'RangeRuler 10 bp (Thermo Scientific). Lanes: 1) purine TFO, 2) duplex, 3) duplex with purine TFO, 4) 1:2 duplex with purine TFO.
2.8	6% PAGE of $(GAA)_{20}$ purine sequences annealed in 10 mM NaCac buffer (pH 7.4) containing 100 mM NaCl, 30 mM $MgCl_2$ and increasing amounts of PEG_{200}. Ladder: O'RangeRuler 20 bp (Thermo Scientific). Lanes: 1) purine TFO, 2) duplex, 3) duplex with purine TFO, 4) duplex with purine TFO and 10% PEG, 5) duplex with purine TFO and 20% PEG, 4) duplex with purine TFO and 40% PEG.
2.9	CD spectra of $(GAA)_{10}$ purine TFO annealed with duplex with increasing amounts of PEG_{200} in 10 mM NaCac buffer at pH 7.4. The introduction of 20% PEG_{200} (v/v) appeared to worsen the distortion below 220 nm.

Table P3: List of figures from **Chapter 3**

Figure	*Caption*
3.1	A) Schematic of plasmid backbone (pUC57) containing the ampicillin resistance gene (Amp), the origin of replication or ORI (rep(pMB1)), and the lacZ gene containing the multiple cloning site (MCS). B) More detailed schematic of the MCS containing M13 forward and reverse sequencing primers (M13fwd and M13rev), restriction enzyme sites (BamHI and HindiIII), promoter sites for polymerases (T7 and T3) and the $(GAA)_n$ insert of interest.
3.2	A) Schematic showing how the PCR product is annealed to form a triplex structure. The $(GAA)_n$ repeating sequence is in orange, and the flanking bases each side in grey. The single-stranded overhangs can be seen either side of the triplex structure. B) Schematic showing how the PCR product and RNA is annealed to form a triplex structure. C) Schematic showing how the PCR product is incubated with RNA to form a triplex structure.
3.3	1.0% agarose gel showing gel electrophoresis of purified samples of clones of **200r** grown in SURE2 bacterial cells and digested with BamHI and HindiIII. Digestion is incomplete as there is a prominent band corresponding to the undigested plasmid.
3.4	1.0% agarose gel showing gel electrophoresis of purified PCR products from amplification of **200r** with increasing amounts of DMSO (0-10%). Smearing is seen in each sample indicating polymerase slippage along the repeat sequence.
3.5	1.0% agarose gel showing the gel electrophoresis of purified PCR products amplified from clones of **200r** grown in SURE2 bacterial cells. From sequencing,

	the clones have $(GAA)_n$ inserts ranging in size from n = 55-75 (clones 1-3 with 75 repeats, clones 4-6 with ~55 repeats.
3.6	1.0% agarose gel showing gel electrophoresis of purified samples of A) **55r** and B) **75r** PCR products annealed at pH 5.5 and pH 7.4.
3.7	1.0% agarose gel showing gel electrophoresis of purified samples of **55r** annealed at pH 7.4 with increasing amounts of PEG_{200}. No evidence of purine triplex formation can be seen.
3.8	1.0% agarose gel showing gel electrophoresis of purified samples of purine and pyrimidine RNA TFOs annealed with **55r** PCR product at pH 5.5 and pH 7.4. Heavier bands can be seen in all samples annealed at acidic pH.
3.9	1.0% agarose gel showing gel electrophoresis of purified samples of purine and pyrimidine RNA incubated with **55r** PCR product at 37 °C for 24 hours in pH 5.5 and pH 7.4 buffer. Heavier bands can be seen in all samples with RNA TFO added (lanes 5-8).
3.10	1.0% agarose gel showing gel electrophoresis of purified samples of purine and pyrimidine RNA incubated with **55r** PCR product at 4 °C in pH 5.5 and pH 7.4 buffer. Evidence of secondary structures can be seen in samples containing purine and pyrimidine RNA TFOs at pH 5.5 and pH 7.4 (lanes 5-6).
3.11	Schematic showing the formation of a 'H-loop' during replication. The ssRNA stabilizes the ssDNA created by the formation of an intramolecular DNA-DNA triplex structure during transcription.Click or tap here to enter text.
3.12	1.0% agarose gel showing gel electrophoresis of purified samples of **55r** annealed at pH 5.5 and 7.4 before after gel purification with spin column. The samples annealed at pH 5.5 exhibit heavier bands corresponding to secondary structures (lane 1) whereas the samples annealed at pH 7.4 do not (lane 2). After purification using spin columns, the heavier bands appear to have been abolished.
3.13	AFM images of **55r** PCR product annealed at pH 5.5 and purified by ethanol precipitation. Apparent triplex structures can be seen throughout the sample, highlighted as rigid regions in the centre of the strand, most apparent in the high contrast image in D).
3.14	Length and height analysis of apparent triplex structures in AFM images of samples of **55r** annealed at pH 5.5. The apparent triplex structures are 0.5-0.9 nm taller than surround dsDNA.

Table P4: List of figures from **Chapter 4**

Figure	*Caption*
4.1	Some examples of molecules that bind to triplex DNA through intercalation, featuring the crescent-shaped structure, positively charged chromophores and substituents with hydrogen-bonding ability.
4.2	Schematic showing the standard fluorescence lifetime decay after excitation by a pulse. The lifetime (τ) is defined as the time taken for the number of excited molecules to decay to 1/*e*.
4.3	A representation of the structure of TO (black) superimposed over a DNA triplex (1D3X, taken from the PDB) with the two duplex strands in pink and blue, and

	the third triplex-forming strand in orange. The site where the aromatic substituents were added is indicated by the green sphere.
4.4	Reaction scheme for the synthesis of **3** and **4a-e** and their molecular structures. Reagents and conditions: (i) $NaHCO_3$ (aq.), methanol, r.t., 1h, (ii) reflux, 30 mins, sat. KI, (iii) aromatic aldehyde, reflux, 3h.
4.5	A) An example of an absorption titration, probe **3** with increasing equivalents of pyrimidine triplex at pH 5.5, B) an example of an emission titration, probe **3** with increasing equivalents of pyrimidine triplex at pH 5.5, C) fluorescence binding curve of probe **3** with different DNA topologies at pH 5.5, D) fluorescence binding curve of probe **3** with different DNA topologies at pH 7.4.
4.6	Graphs showing the fluorescence enhancement ($F-F_0/F_0$) of probes **3** and **4a-e**, calculated from the integrated spectrum of the probe in the absence of DNA (F_0) and the integrated spectrum of probe with 10.4 eq. of DNA (F) with different topologies at pH 5.5 (pink) and pH 7.4 (blue).
4.7	An example of the fitting of the binding curve data obtained from the emission titrations using the Aldrich-Wright method. This shows the binding of probe **3** to DNA topologies at pH 5.5. The binding curve data is indicated in colour and the fitting is indicated by the black lines.
4.8	A) Representative fluorescence lifetime decay: probe **3** (2 μM, 10 mM NaCac buffer at pH 5.5, 100 mM NaCl and 15 mM $MgCl_2$) with pyrimidine triplex (10 μM). For this plot, the data traces are shown in black and the fitting shown in red (bi-exponential, $\chi^2 = 1.32$). B) The residual trace indicating how good the fit is, 0 being the perfect fit.
4.9	Graphs showing the fluorescence lifetimes (weighted τ) of probes **3** and **4a-e** (2μM, 10 mM NaCac buffer at pH 5.5 (pink) or 7.4 (blue), 100 mM NaCl and 15 mM $MgCl_2$) when bound to different DNA topologies (10 μM, 5 eq.). The error bars represent 3 separate repeats.
4.10	1.0% agarose gel showing gel electrophoresis of purified PCR product from amplification of **55r** annealed at pH 7.4 with no probe (lane 1), probe **3** (lane 2), probe **4a** (lane 3), probe **4c** (lane 4), probe **4d** (lane 5) and probe **4e** (lane 6).
4.11	FLIM images of probe **3** (5 μM) incubated with **WT** cells (~30,000). A) Fluorescence intensity, B) fluorescence intensity within segmented nuclei, C) fluorescence lifetime within segmented nuclei, D) fluorescence intensity and lifetime within segmented nuclei merged. Pockets of longer lifetime can be seen in the nucleoli.
4.12	The average weighted fluorescence lifetime (weighted τ) of probes **3**, **4a** and **4d** incubated with HEK293 cells, **WT** and **mut**. Lifetime was calculated as a weighted average across 4 images, and all nucleic within each image, and fitted to a bi-exponential decay.
4.13	FLIM parameters extracted from an image of probe **3** (2 μM) incubated with **WT** cells (~30,000). A) Representative decay fitted with a bi-exponential, and measure of goodness of fit below, B) correlation between intensity and weighted average lifetime throughout the pixels in the image, C) histogram representing the distributions of fluorescence lifetimes across the image.

Table P4: List of tables

Table	*Title*
2.1	The name and sequence of oligonucleotides used for the triplex formation and stability assays.
2.2	Salt conditions used in annealing. Concentration of NaCl and $MgCl_2$ was varied.
4.1	K_a values for probes **3**, **4a**, **4c**, **4d** and **4e** when bound to different DNA topologies under different conditions, reported as $x10^5$ M^{-1}.

List of abbreviations

A, C, G, T, U	Adenine, cytosine, guanine, thymidine, uracil	**G4**	G-quadruplex
ACN	Acetonitrile	**G-rich**	Guanine-rich
C, N, O	Carbon, nitrogen, oxygen	**HEK**	Human embryonic kidney
AFM	Atomic force microscopy	**LNA**	Locked nucleic acid
CD	Circular dichroism	**lncRNA**	Long non-coding RNA
DAPI	4',6-diamidino-2-phenylindole	**PAGE**	Polyacrylamide gel electrophoresis
DCM	Dichloromethane	**PDB**	Protein DataBank
DMF	N,N-dimethylformamide	**PEG**	Polyethylene glycol
DMSO	N,N-dimethylsulfoxide	**PEV**	Positive effect variagation
DNA	Deoxyribonucleic acid	**PCR**	Polymerase chain reaction
dsDNA	Double-stranded DNA	**PNA**	Peptide nucleic acid
EDTA	Ethylenediaminetetraacetic acid	**ssDNA/RNA**	Single-stranded DNA/RNA
EMSA	Electrophoretic shift-mobility assay	**TAE**	Tris-acetic acid-EDTA buffer
ESI-MS	Electron spray ionisation mass spectrometry	**TBE**	Tris-borate-EDTA buffer
EtOH	Ethanol	**Tris**	Tris(hydroxymethyl)-aminomethane
FAM	6-carboxyfluorescein		
FID	Fluorescence intercalator displacement	**TFO**	Triplex-forming oligonucleotide
FLIM	Fluorescence lifetime imaging microscopy	**tsDNA/RNA**	Triple-stranded DNA/RNA
FRDA	Friedreich's ataxia	**(v/v)**	Concentration expressed as vol/vol
FRET	Förster resonance energy transfer		

Table of Contents

Acknowledgements **3**

Thesis abstract **4**

1 Introduction **15**

1.1 Friedreich's ataxia *15*

1.2 Triplexes and other non-canonical DNA structures *21*

1.2.1 DNA and RNA triplexes 22

1.2.2 Evidence of triplexes in vivo 28

1.2.3 G-quadruplexes 30

1.3 Detection of non-canonical DNA in vitro *32*

1.3.1 Electrophoretic gel shift mobility assay ..... **Error! Bookmark not defined.**

1.3.2 Circular dichroism 33

1.3.3 Förster resonance energy transfer 36

1.3.4 Atomic force microscopy 37

1.3.5 Fluorescent probes 39

1.3.6 Intercalators for triplex DNA 43

1.3.7 Groove binders for triplex DNA 46

1.3.8 Fluorescence lifetime imaging 47

1.4 Aims and objectives *49*

2 Biophysical studies of triplex formation from short synthetic oligonucleotides **51**

2.1 Introduction *51*

2.2 Results and discussion *55*

2.2.1 Effect of pH 55

2.2.2 Effect of salt concentration 61

2.2.3 Effect of crowding conditions 64

2.3 Conclusions *66*

3 Biophysical studies of secondary structure formation from polynucleotides of pathological length **67**

3.1 Introduction *67*

3.2 Results and discussion *72*

3.2.1 Bacterial transformation 72

3.2.2 PCR amplification 74

3.2.3 DNA-DNA triplex studies 78

3.2.4 DNA-RNA triplex studies 83

3.2.5 Microscopy of DNA-DNA triplexes 90

3.3 Summary and outlook *94*

4 Development of fluorescence lifetime probes for the detection of triplex DNA 96

4.1 Introduction 96

4.2 Results and discussion 101

4.2.1 Synthesis 103

4.2.2 Fluorescence properties upon DNA binding 104

4.2.3 Fluorescence lifetime studies 115

4.2.4 Fluorescence lifetime imaging microscopy (FLIM) 121

4.3 Summary and future outlook 127

5 Outlook and future work ..... Error! Bookmark not defined.

6 Experimental methods and protocols 134

6.1 General methodology 134

6.2 Nucleic acids in vitro protocols 134

6.2.1 Preparation of buffer solutions for nucleic acid experiments 134

6.2.2 Annealing of DNA for triplex studies 135

6.2.3 Gel electrophoresis – short oligonucleotides 135

6.2.4 Circular dichroism 135

6.2.5 pH stability studies 136

6.2.6 Crowding conditions 136

6.2.7 FRET melting studies 136

6.2.8 FRET emission titrations 137

6.3 Molecular biology techniques 137

6.3.1 Plasmid design 137

6.3.2 Plasmid DNA digestion with restriction endonucleases 137

6.3.3 Phenol:chloroform extraction of DNA from proteins 138

6.3.4 Ethanol precipitation of DNA and RNA 138

6.3.5 Gel electrophoresis – long sequences 139

6.3.6 Transformation of plasmids in bacteria 139

6.3.7 Preparation of bacterial glycerol stock 140

6.3.8 Growth of bacteria from glycerol stock 140

6.3.9 Sequencing of $(GAA)_n$ repeats 140

6.3.10 PCR amplification 140

6.3.11 Purification of DNA from PCR 141

6.3.12 In vitro transcription 141

6.3.13 Purification of DNA from agarose gel electrophoresis 141

6.3.14 Annealing of nucleic acids for triplex studies 142

6.3.15 AFM samples preparation 142

6.4 Synthesis 143

6.5 DNA Binding Analysis 149

6.5.1 UV-Vis absorption titrations 149

6.5.2 Fluorescence emission titrations 149

6.5.3 Time-correlated single-photon counting (TCSPC) 150

6.5.4 Fluorescence lifetime imaging microscopy 151

Bibliography ..... Error! Bookmark not defined.

Appendix **167**

1 Introduction

1.1 Friedreich's ataxia

Friedreich's ataxia (FRDA) was first described by German pathologist and neurologist Nikolaus Friedreich in 1863. Friedreich wrote five papers on a seemingly new spinal disease, detailing his observations on the thinning of the spinal cord and dorsal roots. It wasn't until 1876 that the hereditary nature of FRDA was discussed; Friedreich was studying two siblings with the disease, despite their parents not displaying any symptoms. Friedreich did not receive recognition for this work in his lifetime, though as more research was conducted into this disease the name 'Friedreich's ataxia' persisted.[1]

FRDA is an autosomal recessive genetic disorder of the cerebellum, spinal cord and nervous system. It is the most common form of inherited ataxia, with a prevalence of $2x10^{-5}$ in the Caucasian population and a carrier frequency of around 1 in 100. Symptoms typically develop between the ages of 5 and 15, beginning with walking difficulties and poor balance. These symptoms become steadily worse as the disease progresses and the extent of muscle degeneration increases, with sufferers generally becoming wheelchair bound within 15 years of diagnosis. There are many other complications associated with FRDA, including diabetes, scoliosis, heart failure and vision and hearing loss. FRDA can significantly shorten life expectancy, with the most common form of death being heart disease.[2]

FRDA is characterised by a triplet repeat expansion mutation on intron 1 of the *FXN* gene, located on chromosome 9, which encodes for the protein frataxin (crystal structure seen in Fig. 1.1).[3] Autosomal recessive refers to the requirement for both inherited alleles to contain the mutation. As such, it falls within a group of approximately 20 human disorders known as 'Repeat Expansion Diseases'. The

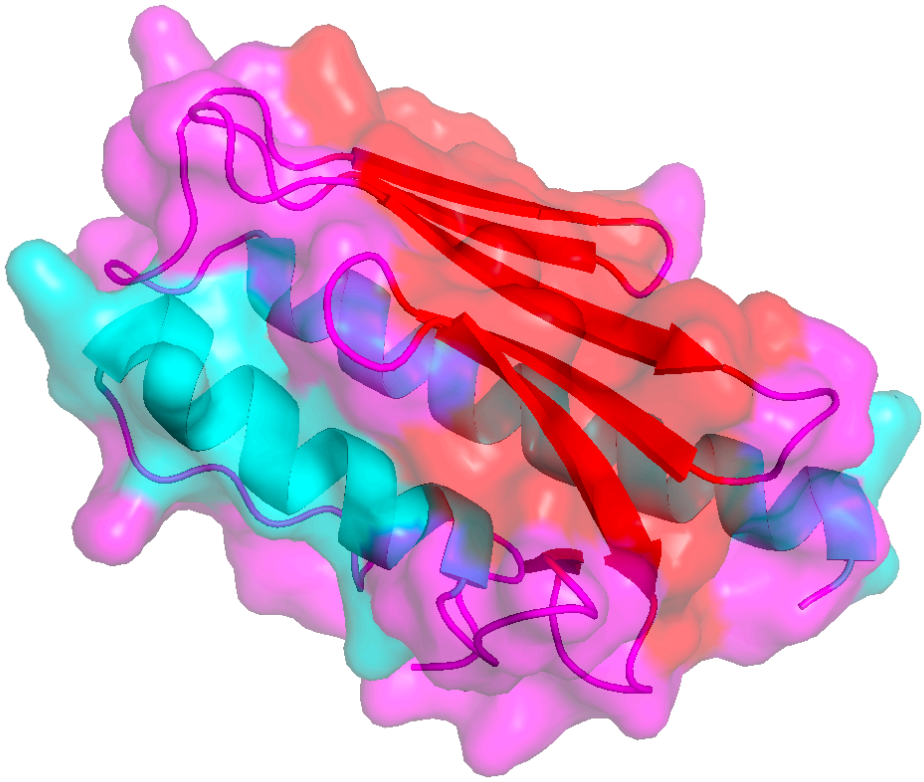


Fig. 1.1: Crystal structure of mature human frataxin (1EKG, taken from the PDB). Alpha helices (two) are displayed in blue, beta sheets in red (four) and loops in pink.[198]

repeat expansion consists of a $(GAA)_n$ base triplet. In the normal version of the gene this base triplet is repeated between 7 and 22 times, however in the defective version of *FXN* gene, this triplet can be repeated up to 1700 times. It has been reported that 96% of FRDA sufferers have $(GAA)_n$ repeat lengths of 44-1700, with 600-900 being the average. The number of repeat units is correlated with the severity of the disease, with longer repeat lengths resulting in an earlier onset of more severe symptoms owing to the increased instability of the repeat.[3] This mutation greatly disrupts the normal production of frataxin, causing progressively less protein in the body as the disease advances - although the function of frataxin is not currently known.[4] It is located in the mitochondria and has been proposed to act as an iron chaperone or storage protein involved in the assembly of iron clusters.[5,6] Frataxin is primarily expressed in tissues with a high metabolic rate, with frataxin mRNA being most abundant in the heart and spinal cord, followed by the liver, skeletal muscle and the pancreas.[7] Without normal levels of frataxin these tissues produce energy much less effectively. It has been reported that binding of iron by frataxin also reduces the amount of free iron in the cell,

and hence reduces the amount of iron available for free radical-generating reactions.[6] Without this protein, excess free iron can accumulate in cells, generating free hydroxy radicals by acting as a catalyst for the Haber-Weiss reaction with hydrogen peroxide, or through oxidation by hydrogen peroxide in the Fenton reaction, and causing oxidative stress which can eventually lead to cell damage.[6]

The disruption of frataxin production is ultimately caused by the partial transcriptional silencing of the *FXN* gene. The repeat expansion mutation allows the synthesis of some frataxin, as the coding sequence is not located within the mutation intron containing the mutation; frataxin levels in peripheral blood leukocytes were found to be between 5 and 30% of normal, depending on the length of the mutation.[8] The $(GAA)_n$ repeat expansion mutation is hypothesised to inhibit transcription by two broad mechanisms: the formation of non-canonical DNA and DNA-RNA hybrid structures and the triggering of chromatin condensation, both of which block access by transcriptional machinery.[9]

The non-canonical DNA structures suggested to form are triple-stranded, or triplex, structures.[10] These triplexes are also hypothesised to aggregate with each other to form 'sticky DNA'.[11] The DNA-RNA hybrid structures are known as 'R-loops'.[12] These triplexes are the primary focus of the work presented herein and the general characteristics of which are discussed in more detail below. A DNA triplex may form in the context of FRDA by the folding back and hybridisation of one strand of the duplex into the major groove of the duplex (see Fig. 1.2). As such, this third strand can either consist of the purine base strand (GAA) or the pyrimidine base strand (TTC), forming a purine or pyrimidine triplex respectively.[10] This can be initiated at a small denaturation bubble in the duplex, whereby the duplex can rotate slightly either side of the bubble, allowing it to fold back and form a triplex.[10] A model was proposed by Grabczyk and Usdin (2000) that suggested intramolecular triplexes can form at GAA:TTC repeats when RNA polymerase is reading the template strand, and the non-template strand folds back onto the duplex.[13] This is driven by the negative supercoiling behind the replication fork.[13] This triplex can then block subsequent RNA polymerase at the promoter proximal end, reducing production of frataxin.[13] Indeed, triplex formation at $(GAA)_n$ repeats has been reported to impede DNA polymerisation *in vitro*. Another study has reported the stalling of replication at the FRDA repeat

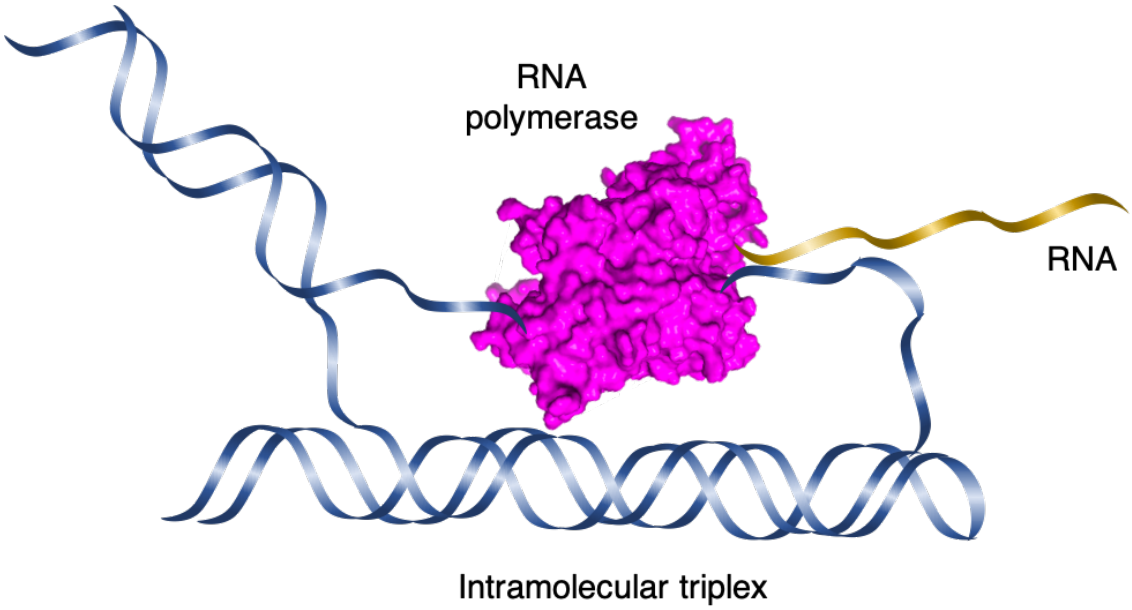


Fig. 1.2: Schematic showing intramolecular triplex formation during transcription. Once the duplex is denatured for transcription, non-template DNA strand (blue) folds back onto the duplex to form an intramolecular triplex while RNA polymerase is reading the template strand.

sequence $(GAA)_{40}$ in plasmids *in vivo*, which was not seen with a normal length of $(GAA)_{20}$ repeats.[14] The threshold repeat length for replication stalling in yeast was reported to be n = 40; similar to the FRDA pathological threshold in humans.[14] It was also reported that the stalling was primarily seen when the purine (GAA) strand was the lagging strand, leading to a hypothesis that this stalling was caused by the formation of an intramolecular purine triplex.[14] Alternatively, another study suggests that pyrimidine triplexes form at $(GAA)_n$ repeats *in vitro*.[15] This study uses peptide nucleic acids (PNAs) consisting of $(GAA)_n$ or $(TTC)_n$ units to study their effects on plasmids containing FRDA repeats under negative supercoiling. Chemical probing found triplex structure and purine ssDNA present, indicating the pyrimidine strand was involved in the triplex.[15] It was also reported that the (GAA) PNAs formed a duplex-invasion complex, abolishing any detectable triplex structure, whereas the (TTC) PNAs did not.[15] This also indicates pyrimidine triplex, as the (TTC) PNA will stabilise the purine ssDNA while the pyrimidine strand participates in the triplex.[15]

'Sticky DNA' is a DNA structure described to form at $(GAA)_n$ repeats in FRDA. It has been previously reported to form by the association of two or more purine triplexes in negatively supercoiled plasmids at neutral pH.[16] In a later study, it was

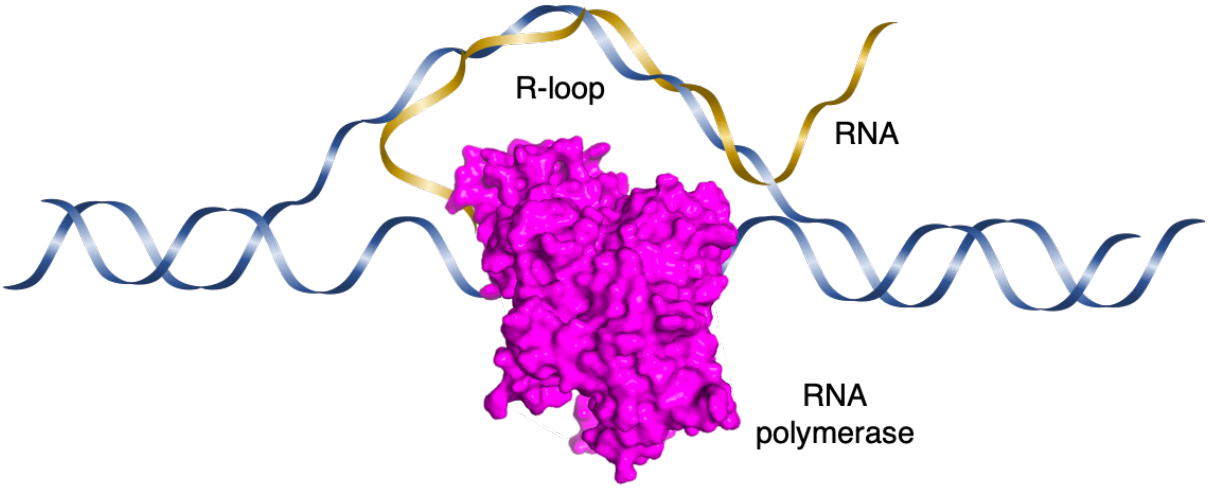


Fig. 1.3: Schematic showing R-loop formation during transcription. Once the duplex is denatured for transcription, the RNA strand binds to the template strand of the duplex.

reported that *in vitro* transcription of a supercoiled plasmid containing a 'sticky DNA' region formed at $(GAA)_n$ repeats of pathological length was greatly reduced in comparison to the linear template without the 'sticky DNA'.[11] A subsequent study used β-Alanine-linked pyrrole-imidazole polyamides to block 'sticky DNA' formation through high-affinity binding to $(GAA)_n$ tracts, resulting in the increase in *FXN* expression in cells from patients with FRDA.[17]

Several studies implicated DNA-RNA hybrid structures in FRDA such as R-loops, as mentioned above. They can form during transcription, when nascent RNA hybridises to the DNA template, leaving the non-template strand as ssDNA (see Fig. 1.3). A recent study showed that R-loops form at expanded repeats on the *FXN* gene in cells derived from FRDA patients, by using a specific antibody that recognises DNA-RNA hybrids.[12] The level of R-loops was also found to correlate with the length of the $(GAA)_n$ repeats.[12] Moreover, it was reported that these R-loops were stable and impeded transcription of the *FXN* gene by RNA polymerase.[12] Relating to heterochromatin formation (another hypothesised mechanism of transcriptional silencing in FRDA), the study also reported increased R-loop levels correlate with increased repressive chromatin markers.[12] However, this study also suggests that formation of heterochromatin is not the cause of transcriptional silencing of *FXN*, but a downstream effect of R-loop formation at expanded $(GAA)_n$ repeats.[12] A more recent study reported that knocking out the enzymes in yeast that counteract formation of

DNA-RNA hybrids resulted in increased rate of $(GAA)_n$ expansion and contraction upon transcription of the repetitive sequence.[18] This study postulates that $(GAA)_n$ expansion is induced by DNA-RNA hybrids that trigger break-induced replication, which is independent of which strand (purine or pyrimidine) is in the transcript.[18] Hence, it is suggested that rather than conventional R-loops, the structures formed could be triplexes stabilised by DNA-RNA hybrids (H-loops).[18]

There is some debate over whether FRDA is an epigenetic disorder.[19] Epigenetics deals with non-genetic mechanisms of regulating gene activity and expression, including changes to DNA that do not alter the underlying sequence; these include DNA methylation and histone modification. It is known that some other Repeat Expansion Diseases such as fragile X syndrome (FXS) are associated with heterochromatin formation,[20] however the notion that FRDA repeats are linked to unusual epigenetic modifications is relatively new. Heterochromatin is a form of tightly packed, or condensed, DNA, and is involved in the regulation of gene expression.[21] There are many varieties falling under two broad categories, constitutive and facultative heterochromatin. Constitutive heterochromatin affects the expression of genes in its vicinity through positive-effect variegation (PEV).[22] Constitutive heterochromatin is typically repetitive and plays structural roles such as formation of telomeres and centromeres. All cells of a particular species contain constitutive heterochromatin in the same regions, and as such any genes that fall within this region will always be poorly expressed.[22] Facultative heterochromatin is however not consistent between cells of the same species and hence a gene in one cell may be packaged in heterochromatin and be poorly expressed, yet in another cell may not be packaged and exhibit normal expression.[23] Facultative heterochromatin is generally regulated and is heavily associated with differentiation. It has been previously reported that $(GAA)_n$ repeats can induce heterochromatinization in a transgenic mouse model.[24] In addition to this, epigenetic modifications such as hypoacetylation of histones H3 and H4 (H3K9me2 and H3K9me3), consistent with high-order chromatin structures, were found in the vicinity of the $(GAA)_n$ repeat sequence in FRDA human and mouse tissues.[24] Acetylation of core histones is a primary factor in influencing gene expression; for example, K9 and K27 trimethylation of histone H3 promotes condensation of chromatin, favouring gene silencing.[25] Hence, inhibition of

deacetylase enzymes has been investigated as a strategy to increase *FXN* expression.[26] One study reports the re-activation of *FXN* in several FRDA models by treating with nicotinamide (a histone deacetylase inhibitor), resulting in apparent 'de-condensation' of heterochromatin.[30,31] There are many proteins that bind to heterochromatin; the heterochromatin protein 1 (HP1) family are a group of highly conserved proteins that have functional importance in the cell.[27]

DNA methylation is another epigenetic process also associated with gene silencing. The extent of DNA methylation has been reported to be much higher in the region flanking the $(GAA)_n$ repeats in FRDA patient cells in comparison to normal alleles.[28] A direct correlation between extent of methylation and length of $(GAA)_n$ repeats in patient cells has also been reported.[29]

While there is abundant evidence that unusual DNA and DNA-RNA hybrid structures and 'sticky DNA' form at the site of $(GAA)_n$ repeats *in vitro* under particular conditions, it remains to be proven whether they occur *in cellulo*.

1.2 Triplexes and other non-canonical DNA structures

DNA is typically found in the classical Watson-Crick double helical form. In this structure, two oligonucleotide strands run anti-parallel to each other, held together by hydrogen bonding and base stacking interactions between complementary base pairs.[32] There are two types of natural DNA base: purine and pyrimidine. The purine bases consist of guanine (G) and adenine (A), while the pyrimidine bases are cytosine (C), thymine (T) and the RNA base uracil (U). In the double helix, G is complementary to C and A is complementary to T or U. DNA duplexes come in at least three forms in nature: A-DNA (right-handed 11 bp/turn), B-DNA (right-handed 10 bp/turn) and Z-DNA (left-handed 12 bp/turn), with B-DNA being the most prevalent.[33]

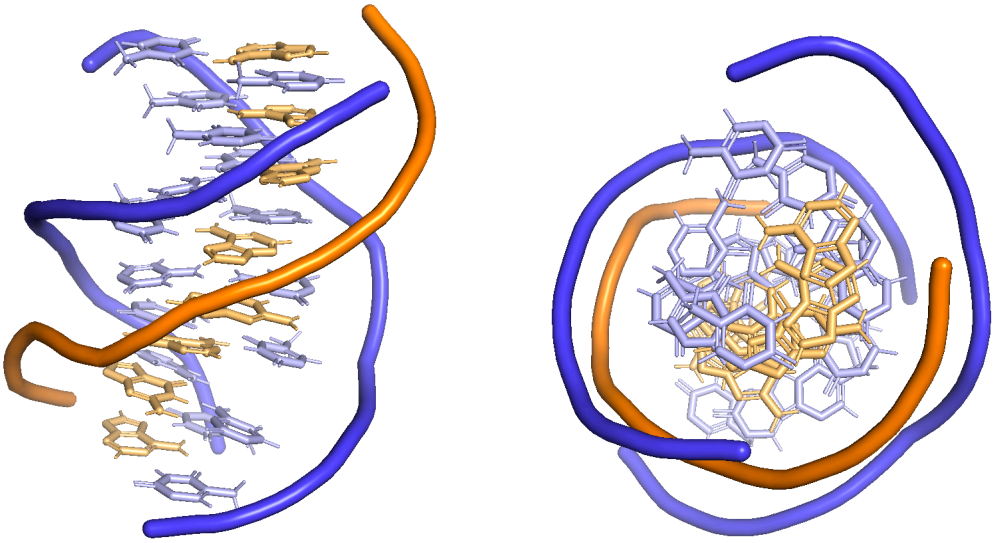


Fig. 1.4: A solution structure of triplex DNA (1D3X, taken from the PDB). The two strands corresponding to the duplex are blue, while the TFO is shown in orange, bound in the major groove of the duplex.

1.2.1 DNA and RNA triplexes

Since the elucidation of the double-helix, there have been more and more secondary DNA structures discovered. One such DNA structure is the triplex, first described in 1957.[34] As already mentioned above, DNA triplexes are triple-stranded structures consisting of a duplex with a third nucleic acid strand bound in the major groove (see Fig. 1.4). This can happen in an intermolecular or intramolecular fashion.[35] Triplexes can be 'homopurine' or 'homopyrimidine', depending on the DNA bases making up the triplex-forming oligonucleotide (TFO). Other common annotations are R*R-Y and Y*R-Y for purine and pyrimidine triplexes respectively. Homopurine triplexes contain a B-DNA duplex and a TFO consisting of only G and A bases bound in the major groove of the duplex in an antiparallel orientation. Homopyrimidine triplexes contain a B-DNA duplex with a TFO consisting of only C and T bases (or U if RNA) bound in the major groove of the duplex in a parallel fashion.[35] The third strand is held in the major groove by Hoogsteen hydrogen bonding with the homopurine strand of the duplex (Fig. 1.5). The homopurine triplex base triplets consist of G*G-C and A*A-T, whereas the homopyrimidine triplex base triplets consist

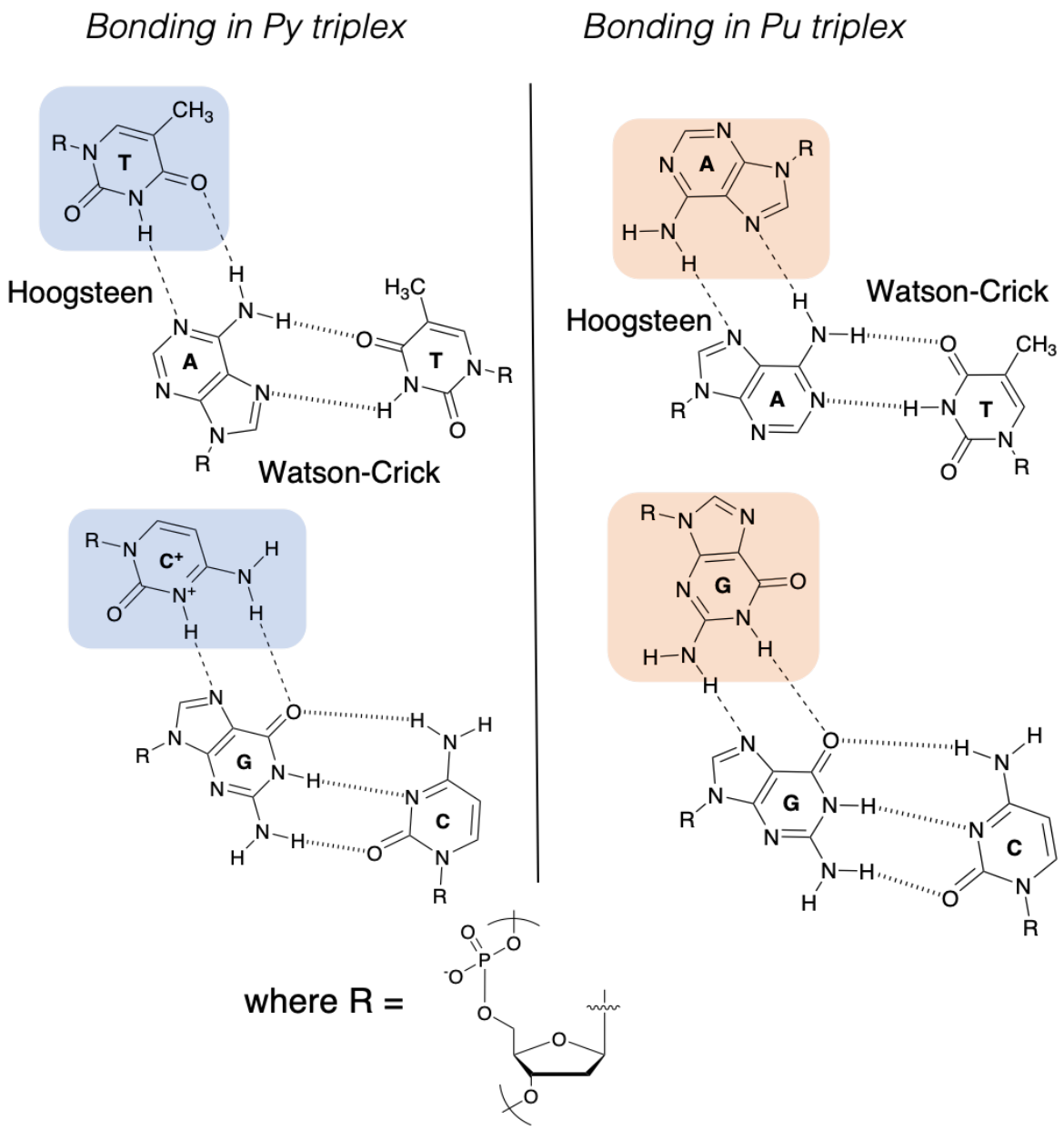


Fig. 1.5: Schematic showing the bonding in pyrimidine triplexes (left) and purine triplexes (right). Watson-Crick hydrogen bonding is indicated with a striped line, while Hoogsteen hydrogen bonding is indicated with a dashed line. The bases in the TFO are indicated with the coloured boxes; blue for pyrimidine and orange for purine.

of C$^+$*G-C and T*A-T. It can be seen that protonation, at the N7 position of the cytosine, in the homopyrimidine TFO is required for formation of one of the Hoogsteen hydrogen bonds. Hence, acidic conditions are favourable for pyrimidine triplex formation, while this is not required for purine triplex formation. Purine triplex formation is strongly favoured by the presence of divalent metal ions such as magnesium (Mg^{2+}). Generally, triplexes have a much higher charge density in comparison to dsDNA due to the three polyanionic strands in close proximity.[36] Due to this, it can be expected that the charge

repulsion effect between the negatively charged phosphate backbones of the three nucleic acid strands is more significant. The presence of cations can act to shield this repulsion in solution, allowing the strands to come together with less resistance. Stabilisation of purine triplexes by divalent metal cations has been extensively reported in the literature. One theoretical study found a large stabilisation of the G*G-C base triplet with Mg^{2+}, predominantly arising from the balance of electrostatic interactions between the salt cations and the three negatively charged DNA bases in the triplet.[37] It was also reported that the cation-induced polarisation, which was centred on the guanine of the third strand, contributed considerably to the stabilisation of the triplet.[37] Experimental studies have shown similar results.[38] It has been previously reported that divalent metal cations are instrumental in the formation of a purine triplex from a G-rich sequence as seen by mass spectrometry.[38] In particular, divalent cations have been demonstrated to coordinate to certain sites on DNA nucleotides. This has been reported for the binding of Mg^{2+} and Zn^{2+} to the N7 position of purine nucleotides G and A, enhancing the formation of Hoogsteen hydrogen bonds with these bases.[39] This is supported by another study showing the stabilisation of G-rich intramolecular triplexes with divalent cations including Mg^{2+} and Ca^{2+}.[40] Some studies have suggested that monovalent cations are actually detrimental to triplex formation, especially in G-rich sequences.[41] This could be due to competition arising with self-association structures such as dimers or tetramers, which have been previously reported for sequences containing (GA) and (GT) repeats.[42] Monovalent cations are known to stabilise alternative non-canonical DNA structures, as discussed in more detail later. The stability of triplexes is highly dependent on solvent environment, sequence and whether it is an intermolecular or intramolecular structure. Triplex stability also greatly depends on the nature of the underlying duplex, though triplexes tends to be much less stable than their counterpart duplexes.[43] In fact, there are few studies on unmodified triplexes at physiological pH.

Molecular crowding conditions also stabilise triplexes and other non-canonical DNA structures.[44] The molecular environment *in cellulo* is inherently crowded, with biomacromolecules such as proteins occupying 20-40% of the cellular environment and creating steric hinderance in the nucleus.[44] This can be emulated using large molecules like polyethylene glycol (PEG). Triplexes have exhibited increased melting

temperatures in the presence of 19% (v/v) PEG, with greater stabilisation seen with larger PEG molecules.[45,46] Another study reported the promotion of G-triplex formation in the presence of 40% (v/v) PEG.[40] This is related to the change in the thermodynamics of base pairing caused by the presence of PEG. An increased concentration of PEG in solution is accompanied by a decrease in the activity of water, hence favouring the pairing of bases in order to reduce the hydrophobic effect.[44] Triplex formation in the presence cellular concentrations of macromolecules *in vitro* provides further indication that triplexes may form *in vivo* under similar conditions.

As well as DNA triplexes, DNA-RNA hybrid and pure RNA triplexes have also been reported. In fact, RNA triplexes were the first to be observed *in vitro*.[47] DNA-RNA hybrid triplexes are subject to more sequence restraints than pure DNA triplexes; the RNA bases are required to be pyrimidine only, with uracil also being favoured over cytosine. The rU base forms a triplet with an A-T DNA base pair and protonated rC^+ base forms a triplet with a G-C base pair; these structures are among the most stable triple-helical structures.[48] The polypyrimidine RNA strand binds in an antiparallel orientation with respect to the polypyrimidine DNA strand, and hence the RNA cannot generally form a triplex with its own transcript unless it's a mirrored or repeat sequence. Despite these restraints, it has been found that triplexes where the TFO is RNA instead of DNA are generally more stable.[49] It has been previously reported that triplexes with a RNA TFO also maintain their thermal stability upon the introduction of dehydrating co-solutes such as EtOH and ACN, whereas triplexes formed with a DNA TFO do not.[50] More recently, it has been suggested that long noncoding RNAs (lncRNAs) form DNA-RNA hybrid triplex structures in order to effect biological change. Only in the last decade or so has it been discovered that although most mammalian DNA is transcribed, only a small fraction of it actually encodes for functional proteins, and the majority of RNA transcribed is 'noncoding'.[51] These lncRNAs have now been implicated in many biological processes, including epigenetic regulation.[52] Though the mechanisms by which they regulate gene expression is still relatively unknown, lncRNAs have been shown to form DNA-RNA triplexes in many cases.[53] One such example is the recruitment of DNA methyltransferases by specific binding to a DNA-RNA triplex formed from a promoter lncRNA bound to a transcription factor target

sequence.[54] This suggests that DNA-RNA triplexes could be instrumental in the epigenetic regulation of chromatin through recruitment of DNA methylation enzymes.

Modifying oligonucleotides to encourage triplexes to form under physiological conditions has been the basis of much investigation in recent years. This has been prompted by the concept of using sequence-specific targeting oligonucleotides as potential gene editing tools.[55] It has been previously reported that photoreactive psoralens conjugated with oligonucleotides can induce site-specific DNA damage and control gene expression through formation of cross-linked intermolecular triplexes, which halt the progression of RNA polymerase.[56] This was also seen to happen in plasmids in human cells, demonstrating that triplex formation can occur within cells.[57] TFOs that have been modified by incorporating unnatural bases to facilitate triplex formation are also a promising strategy for gene therapy. Peptide nucleic acids (PNAs) are synthetic DNA mimics that contain a pseudo-peptide backbone rather than the natural deoxyribose phosphate backbone. Without the negatively charged backbone, this eliminates the charge repulsion between nucleic acid strands and allows secondary structures to form more easily, although also makes them much less soluble. The backbone is also inherently more flexible, meaning that PNAs often bind to polynucleotides with higher affinity than natural DNA and RNA.[58] The first example of this showed the formation of a stable triplex between a polypyrimidine PNA hairpin and a polypurine DNA strand.[59] Since then, PNAs have been investigated for their potential in antigene applications. PNAs can form several different complexes with dsDNA, including duplex invasion complexes, triplex invasion complexes and standard triplex structures. It has been previously reported that PNAs can form triplex invasion structures within the transcribed region of a template strand, preventing elongation by RNA polymerase and inhibiting transcription.[60] PNAs can also inhibit transcription by preventing the unwinding of duplex DNA and blocking the binding of transcription factors to promoter regions of target genes.[60] A recent study employed a poly-pyrimidine triplex-forming PNA with thiazole orange (TO) as a base surrogate for fluorescence sensing of dsRNA.[61] The TO-PNA conjugate formed a thermally stable triplex at acidic pH with dsRNA, prompting a ‘light up’ response from the TO unit and allowing real-time monitoring of triplex formation.[61] As previously mentioned, PNAs

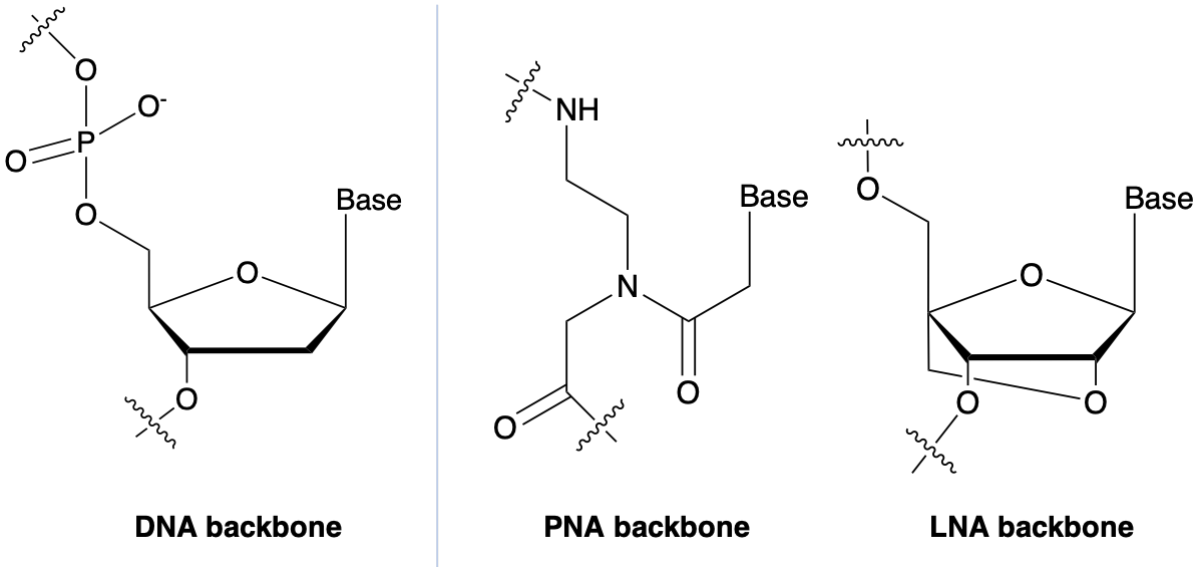


Fig. 1.6: The chemical structure of the backbone of DNA, PNA and LNA. The backbones of PNA and LNA lack the negatively charged phosphate group, reducing charge repulsion between strands. The LNA is conformationally restrained, enhancing base-stacking properties.

have also been used to investigate secondary structure formation in supercoiled plasmids containing the FRDA $(GAA)_n$ sequence.[15] It was reported that $(GAA)_4$-PNAs abolished all detectable triplex structures, yet $(TTC)_5$-PNAs did not.[15] It was shown that $(GAA)_4$-PNAs can invade the duplex at the repeat region by binding the $(TTC)_n$ DNA strand and preventing triplex formation, and $(TTC)_5$-PNAs can form triplex invasion complexes at the site of the repeats.[15] The same study also reported the use of locked nucleic acids (LNAs) to disrupt the higher order structures formed at the repeat region. Locked nucleic acids (LNAs) are modified RNA bases containing an extra bridge between the 2' oxygen and 4' carbon (see Fig. 1.6). The 'locked' conformation enhances the base-stacking properties and increases melting temperatures of structures.[62] LNAs are also more stable than natural RNA, as they cannot undergo base-catalysed hydrolysis and cleavage of the phosphodiester backbone, as LNAs do not have hydroxyl group at the 2' position. Incorporation of LNA bases into TFOs increases the stability of the resulting triplex, including promoting pyrimidine triplex formation at physiological pH.[63] A recent study showed that incorporation of LNA bases into the polypyrimidine DNA strand of dsDNA, forming a *hetero*-duplex, enhances triplex formation with TFOs by distorting the duplex structure.[64] Another recent study reports the increased thermodynamic stability of

RNA triplexes with incorporation of LNA and 2-thiouridine residues into the TFO.[65] In addition to this, these triplex structures were reported to inhibit GFP expression and regulate GFP gene silencing in *HeLa* cells as seen by quantitative analysis of GFP expression.[65] This research demonstrates how modified nucleic acids can be used to form more stable secondary structures for possible therapeutic applications.

1.2.2 Evidence of triplexes in vivo

Despite implications that triplexes play a role *in cellulo*, this evidence is based on *in vitro* studies and is primarily circumstantial. Bioinformatics tools developed to analyse the genome and locate any potential triplex-forming sequences have suggested triplex-forming motifs are prevalent across the genome.[66] These motifs have primarily been found in gene promoter regions, thus adding to the hypothesis that triplexes may play an important part in gene regulation.

More recently, proteins have been discovered that are thought to bind to triplex structures in cells. For example, the high mobility group proteins HMG1 and HMG2, associated with chromatin and distribution of heterochromatin, have the ability to bind triplexes.[67] Another example is the GAGA transcription factor in *Drosophilia*, which specifically binds d(GA:TC) sequences known for forming triple-stranded structures and other non-B conformations.[68] GAGA has been previously reported to bind with as high affinity to a pyrimidine intermolecular DNA triplex as the corresponding dsDNA, and these triplexes are seen to inhibit transcription of genes situated downstream.[68] Furthermore, the GA factor in *Drosophilia* is known to be involved in the epigenetic maintenance of heterochromatin and hence the silencing of genes.[69] Alternatively, the *STM1* gene in yeast encodes for a protein that binds to a purine triplex motif.[70] The protein Stm1p was seen to bind purine triplexes through electrophoretic mobility shift assay.[70] Interestingly, Stm1p had already been identified as binding to G-quartets, an alternative non-B DNA structure also held together by Hoogsteen-bonded guanine bases.[71] Perhaps most interestingly, it has been found that the highly conserved human initiation protein Orc4 binds preferentially to triplex DNA.[72] The protein Orc4 is essential for initiation of DNA replication in cells, and its binding site coincides with known triplex-forming sequences, suggesting possible significance of unusual DNA structures in recognition by ORC proteins.[72]

As well as cellular proteins that bind to triplexes presenting indirect evidence that they form *in cellulo*, the presence of proteins that destabilise triplex structures provides further justification. For example, different families of helicase enzymes can dissociate intramolecular DNA triplexes upon binding to the single-stranded loop region.[73,74] Increased genomic instability has been reported in the absence of these enzymes, particularly in regions with triplex-forming potential.[75] Warsaw breakage syndrome is a rare genetic disorder associated with mutations of the human *ChIR1* (*DDX11*) gene encoding for ChIR1 helicase. The disorder is characterised by defective genome maintenance. A study reported that ChIR1 selectively and efficiently unwinds intermolecular and intramolecular DNA triplex substrates and was the preferred substrate over G-quadruplexes and replication forks.[76] In contrast to this, the mutated protein from Warsaw breakage syndrome failed to unwind these triplexes.[76] Cellular assays also showed increased levels of triplex DNA in ChIR1-deficient cells, suggesting physiological relevance of triplexes in this disease.[76] In the study relating to psoralen-conjugated TFOs discussed earlier, it was found that nucleic excision repair proteins XPC-hHR23B and XPA-RPA recognise the triplex-directed psoralen DNA interstrand crosslinking and may act in a cooperative manner to repair the DNA damage.[77]

There is little direct evidence of naturally occurring triplex structures in cells. Triplex-specific antibodies have been employed in an attempt to image triplexes *in vivo*. There are early examples of this: a monoclonal antibody Jel 318 was produced by immunising mice with poly[d(TmC)]-poly[d(GA)]-poly[d(mCT), a sequence which forms a stable triplex at physiological pH.[78] This antibody was seen to bind strongly to triplex-forming sequences, but not to ssDNA or dsDNA such as calf thymus DNA.[78] Subsequently, another monoclonal antibody called Jel 466 was produced in an analogous way and found to have similar properties, however did not bind to poly(dT)-poly(dA)-poly(dT) base triplets.[79] Both Jel 318 and Jel 466 showed little structural homology, however both contained numerous positively-charged amino acids. Another study showed both antibodies inhibited transcription in nuclei by up to 20%, suggesting a role for triplex DNA in regulating transcription.[80] One study uses triplex-specific antibodies to demonstrate the presence of triple-stranded DNA sequences in the centromeric regions in *Drosophilia melanogaster* mitotic chromosomes.[81] This was

validated by staining with thiazole orange (TO), previously shown to be able to reproduce results obtained from antibody binding to triplex DNA.[82] Triplex-specific antibodies present a promising approach for imaging triplex DNA *in vivo*, however they do demonstrate differing sequence specificity between species.

Chemical probes are another tool with which to image DNA structures *in vitro* and *in vivo.* As mentioned previously, TO has the ability to bind triplex DNA. A recent study reported that TO detects triplex DNA in the salivary gland of *Drosophilia melanogaster* in an identical location to the previously used triplex-specific antibodies.[82] However, TO also binds to quadruplex, dsDNA, ssDNA and RNA and therefore does not exhibit sufficient specificity for triplex DNA. Chemical probes for triplex DNA and other non-B conformations are discussed in more detail below.

1.2.3 G-quadruplexes

An alternative non-canonical DNA structure that is often formed at G-rich sequences is the G-quadruplex. G-quadruplexes are helical structures comprised of one, two or four strands forming guanine tetrads. Like triplexes, G-quadruplexes have been implicated in control of gene expression. Unimolecular (or intramolecular) forms in particular are found in telomeric regions and are over-represented in gene promoter regions.[83] Targeted stabilisation of G-quadruplexes has been identified as an anticancer strategy; they are prevalent in oncogene promoters and hence stabilisation of these structures may prevent transcription of these genes.[71] Also, like triplexes, most quadruplexes are held together by Hoogsteen hydrogen bonds. At the core of a G-quadruplex are two or more π-π stacked G-quartets, an arrangement of four G bases bonded together through Hoogsteen interactions. There are several possible conformations depending on the orientation of strand binding. These structures have been shown to form under physiological conditions and are notably stabilised by monovalent metal cations such as K^+ and Na^+.[84] This stability occurs through electrostatic interactions between the cations and the carbonyl moieties on the G bases, with additional stabilisation provided through interaction of the cations with the loops and grooves of the structure. It has been extensively reported that K^+ stabilises G-quadruplexes to the greatest extent. Divalent cations, especially Mg^{2+}, have been shown to destabilise G-quadruplexes under some conditions and can cause a shift in

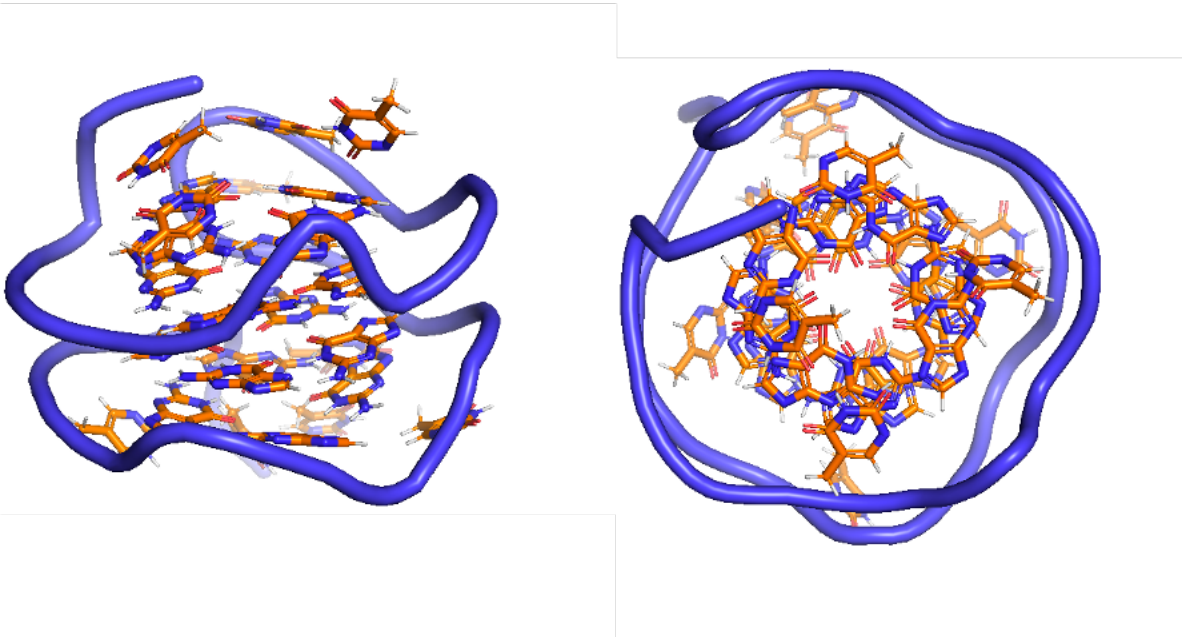


Fig. 1.7: Solution structure of a G-quadruplex (2MS9, taken from the PDB).

conformation due to the difference in ionic radii.[85] Also like triplexes, G-quadruplexes are encouraged to form by superhelical stress and crowding conditions.[86] *In vitro*, G-quadruplexes have exhibited greater stability once formed than some dsDNA structures, with slower unfolding kinetics than DNA and RNA hairpins.[85] Bioinformatics analysis of the genome shows over 300,000 potential G-quadruplex-forming sequences.[87] Like triplexes, these sequences are mostly found in functional regions of the genome. Highly enriched in gene promoters, it has been reported that 90% of human DNA replication origins contain G-quadruplex motifs.[88,89] Heavily implicated in gene regulation, G-quadruplexes are thought to form during events such as DNA replication, transcription or repair, when the duplex is temporarily denatured. In addition to this, the tips of eukaryotic telomeres have single-stranded overhanging tandem repeat sequences with the propensity to form quadruplex structures.[90] Also, transient Hoogsteen base pairs have been detected in canonical duplex DNA, especially at sites of damage and with specific transcription factors bound, further indicating the importance of non-canonical DNA structures *in vivo*.[91] This could also mean that the duplex does not necessarily need to be denatured to form a G-quadruplex, or indeed a triplex, in certain cases.

1.3 Detection of non-canonical DNA *in vitro*

There are a number of biophysical and chemical detection methods for studying and imaging non-B DNA *in vitro.* The methods discussed below are used in the following work in the analysis and characterisation of non-canonical DNA structures.

1.3.1 Gel electrophoresis

Gel electrophoresis is a useful tool for separating oligonucleotides or proteins based on size, mass or conformation. Samples are loaded onto a gel in an appropriate buffer solution and a voltage is applied across it, causing the negatively charged species to move through the gel towards to positive terminal (see Fig. 1.8). The position of species within the gel is then revealed by staining with an oligonucleotide intercalating agent such as ethidium bromide. The electrophoretic mobility of a species is dependent on its size, charge and to some extent its structural state. For example, linear oligonucleotides run more quickly than circular oligonucleotides of the same mass, and circular oligonucleotides under supercoiling runs more slowly than relaxed circular oligonucleotides.[92]

This is useful for the analysis of secondary structure formation within linear and circular nucleic acids. This technique has been used extensively for the visualisation of triplex DNA formation, both inter- and intramolecularly.[53,93–97] Intermolecular triplex formation will produce a distinctly different band within the gel than the corresponding dsDNA, as the addition of the TFO results in a heavier structure. If binding is not complete, three bands may be seen that are consistent with the intermolecular triplex, dsDNA and ssDNA representing the unbound TFO. Such an assay can be used to qualitatively assess the efficiency of TFO binding to a dsDNA sequence. On the other hand, formation of intramolecular triplexes can be investigated by observing less obvious changes in band shift. The linear unbound species will run more quickly through the gel than the more compact bound species. G-quadruplexes have also been investigated using gel electrophoresis. It has been reported that intramolecular G-quadruplexes run faster on a gel than intermolecular G-quadruplexes, which are both slower than the corresponding linear oligonucleotides. Though, as with triplexes, intramolecular quadruplexes are harder to distinguish from the corresponding linear

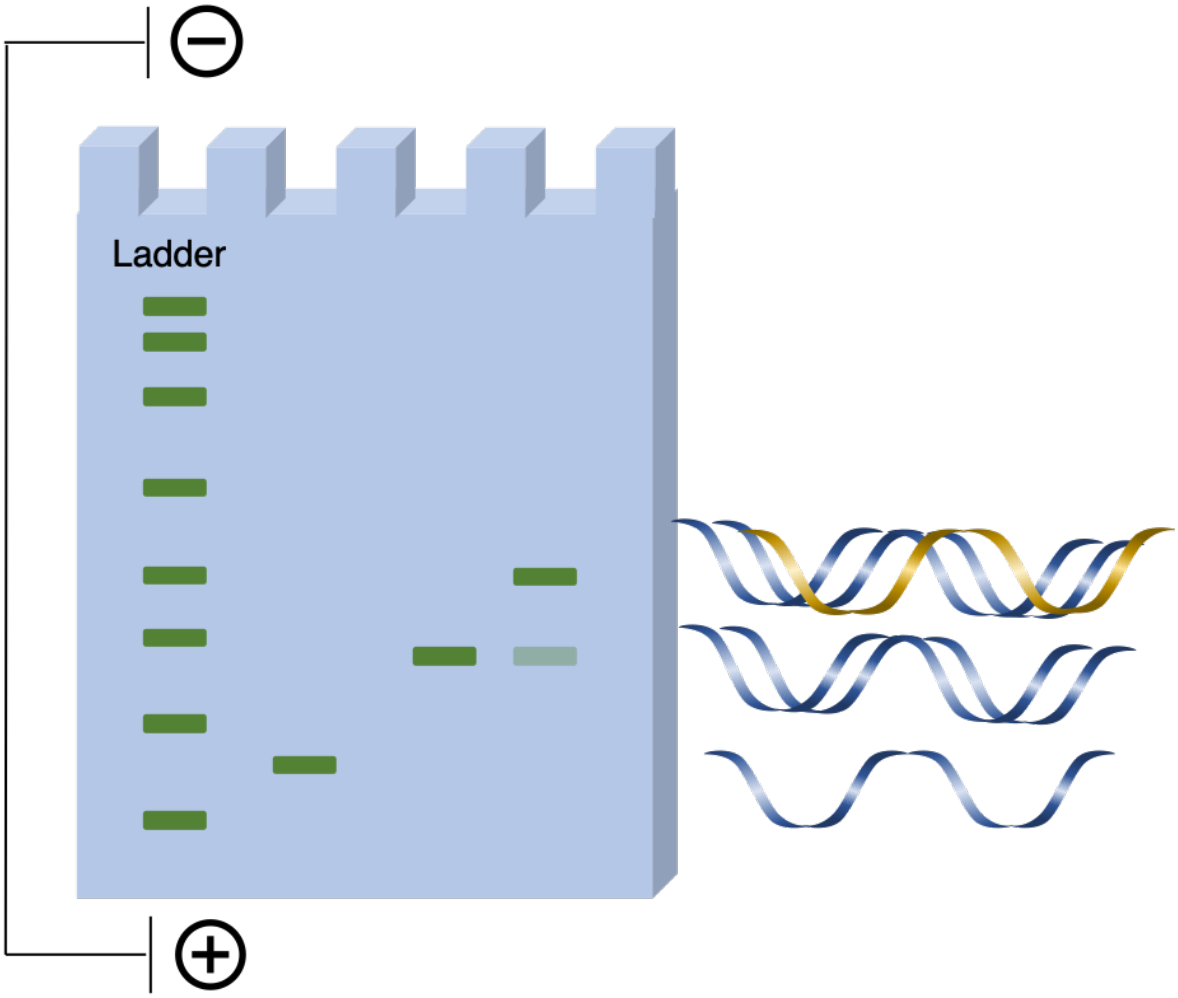


Fig. 1.8: Schematic demonstrating the principles of gel electrophoresis. An electric current is applied across the gel and the negatively charged DNA moves towards the positively charged terminal at different speeds depending on size and conformation.

DNA than intermolecular structures are.[98,99] Though gel electrophoresis does not provide any specific structural data, it is a valuable qualitative tool for determining whether secondary structures have formed and the ratio between folded and unfolded, or hybridised and unhybridized, structures.

1.3.2 Circular dichroism

Circular dichroism (CD) is a powerful technique for detecting differences in nucleic acid conformation as well as identifying particular structures through characteristic bands. CD is a form of absorption spectroscopy based on circularly polarised light used to investigate the structure of optically active chiral species. CD uses the differential absorption of left and right circularly polarised light passing

through a sample to generate a characteristic spectrum for a particular structure.[100] A chiral structure will absorb left and right polarised light by different amounts, and any absorption will change the amplitude of the incident wave. Since the amplitude of the incident wave for left- and right-handed light will be altered to different extents, the superimposed light is no longer linearly polarised, and the incident wave becomes elliptically polarised.[100] CD spectra are often reported as a degree of ellipticity (θ), which is the angle between the magnitude of the electric field vector at its maximum and at its minimum.[100]

CD is most widely used for identifying structural features of biomolecules such as DNA and proteins. For proteins, the structural elements of secondary structures such as α-helices and β-sheets all have signature CD spectra, hence the ratio of these spectral features in the CD spectrum of a protein can provide information about the relative numbers of these structural elements within a tertiary protein structure. A similar concept can be applied to DNA, though CD displays sequence dependence due to the different chemical structures of the nucleotide bases. The CD spectra of B-DNA share global characteristics such as positive bands or band at 260-280 nm and a negative band around 245 nm, though the position and amplitude of the bands is dependent on the sequence.[101] Where CD is most useful is following structural or conformational changes of polynucleotides under different conditions. For example, CD spectra can be collected as a DNA structure is denatured or annealed to follow the shifting of the bands corresponding to the formed and denatured structures. This can yield information about melting temperatures and thermal stability.

There are no characteristic CD spectra for triplex DNA as it is highly sequence dependent.[41,102,103] However, triplexes comprised of specific sequences do exhibit CD signatures. It has been previously reported that intermolecular R*R-Y triplexes formed from the FRDA sequences containing $(GAA)_5$ repeat units gave a CD spectrum with a large negative band at 210 nm.[93] Early structural studies of triplexes reported a strong negative band between 200-300 nm in CD spectra of triplexes containing GA, GT or CT repeats.[104] In addition to this, comparing CD spectra of a potential triplex structure

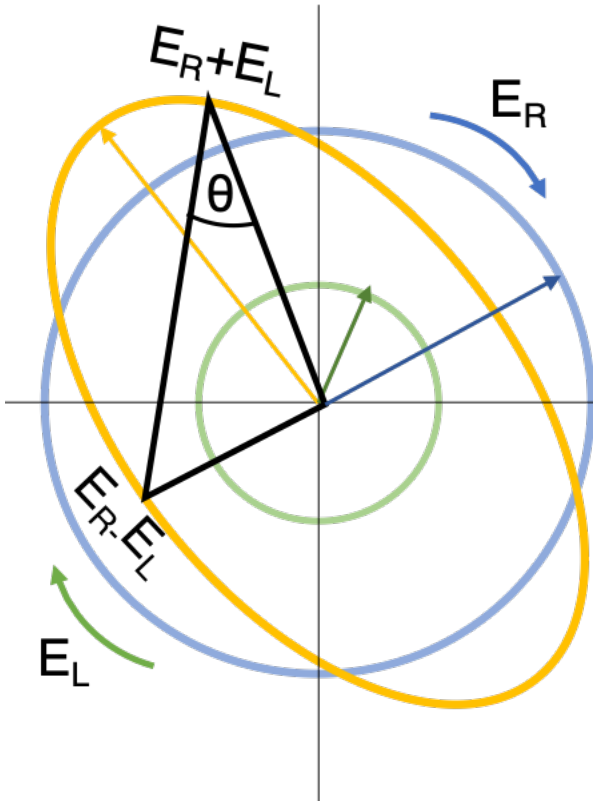


Fig. 1.9: Schematic demonstrating the principles of CD. Here, the right polarised light (blue) is absorbed to a larger extent than the left polarised light (green). The degree of ellipticity (θ) is given by the angle in the triangle made between the sum of the contributions (E_R+E_L) and the difference between the contributions (E_R-E_L).

with those of the constituent dsDNA and TFO can suggest whether a different structure is forming.

The CD spectra of G-quadruplexes, however, are not sequence dependent. Since these structures are only formed from sequences containing runs of G and the foundations of the structure are stacked G-quartets, the peripheral bases do not affect the spectrum. The two main conformations of G-quadruplex, parallel and antiparallel, do have distinct CD signatures. Parallel quadruplexes display a strong positive band at 260 nm, whereas antiparallel quadruplexes exhibit a negative band at 260 nm and a positive band at 290 nm.[105] In contrast to the FRDA triplex structure, both quadruplex conformations show a positive band at 210 nm.[105] The disparity in spectra between the two conformations is due to different folding of the DNA and hence differing glycosylic angles in the backbone.[105] Hence, CD is the most useful tool for distinguishing between these two forms and can be used to follow the transition between them induced by external stimuli.[106,107]

1.3.3 Förster resonance energy transfer

Förster resonance energy transfer (FRET) is the mechanism by which energy is transferred between a donor and an acceptor fluorophore through a nonradiative process. The efficiency of this transfer is highly dependent on the distance between the fluorophores, as shown by the equation below, and hence FRET can be used as a measure of this.[108]

$$E = \frac{1}{1 + (r/R_0)^6}$$

Where E is the FRET efficiency, r is the separation distance and R_0 is the Förster distance between the two fluorophores, or the distance at which the energy transfer efficiency is 50%.

A sample is irradiated at the excitation wavelength (λ_{ex}) of the donor fluorophore, and the donor emits at its emission wavelength (λ_{em}). However, once the donor and acceptor are in close proximity, this energy can be non-radiatively transferred between the two, and the acceptor fluorophore can then release this energy through emission at its λ_{em}, which will be distinct from the λ_{em} of the donor fluorophore. Characteristics of good donor fluorophores include high quantum yield and a high extinction coefficient, and a suitable acceptor should have an absorption spectrum that overlaps somewhat with the emission spectrum of the donor to maximise the efficiency of energy transfer.

This technique can be useful for the analysis of biomolecules because it can provide accurate measurements of distance between two points labelled with donor and acceptor fluorophores. This can be applied to analysis of protein-protein interactions, following of intramolecular nucleic acid folding and the intermolecular hybridisation of oligonucleotides, while also providing kinetic information about these processes. Increasing emission from the acceptor fluorophore indicates that the pair are sufficiently close enough for FRET to occur, which is generally within 10 nm.[108]

FRET has been used previously to follow the hybridisation of a donor-labelled TFO to an acceptor-labelled duplex to form an intermolecular triplex.[109,110] The

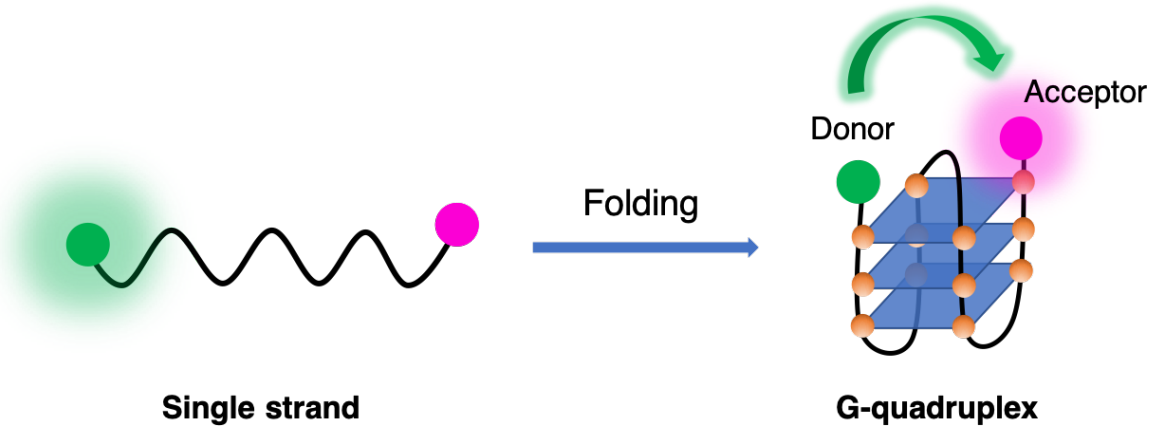


Fig. 1.10: Schematic demonstrating the principles of FRET. Here, when the G–quadruplex is unfolded (left), the donor fluorophore (green) emits upon excitation. However, when the G-quadruplex is folded (right), FRET occurs and the acceptor fluorophore (pink) emits.

increase in FRET can also be measured as a function of time, showing the kinetics of hybridisation. The folding of intramolecular triplexes can be followed in a similar way. One end of the dsDNA portion is labelled with one fluorophore and the opposite end of the ssDNA portion is labelled with the corresponding fluorophore, such that as the single-stranded sequence folds back and hybridises with the duplex, the fluorophores become close enough to each other for FRET between them. Again, the kinetics of this transition can be followed with the increasing emission of the acceptor fluorophore with time.[111]

FRET is particularly useful for following the folding of intramolecular quadruplexes (see Fig. 1.10). Since intramolecular G-quadruplexes are formed when a G-rich single-stranded region folds into the stacked G-quartet structure, the two ends of this strand go from being far apart to being close together upon folding. By labelling of each end of the region with a donor-acceptor FRET pair, this folding can be followed by observing the increasing emission of the acceptor fluorophore.[112] A previous study used FRET to follow the folding of a G-quadruplex within a fluorophore-labelled DNA mini-circle containing a G-rich single-stranded region, which was also seen with atomic force microscopy (AFM), as discussed further below.[113]

1.3.4 Atomic force microscopy

Atomic force microscopy (AFM) is a form of high-resolution scanning probe microscopy which uses a sharp tip to 'feel' a surface to give information about its

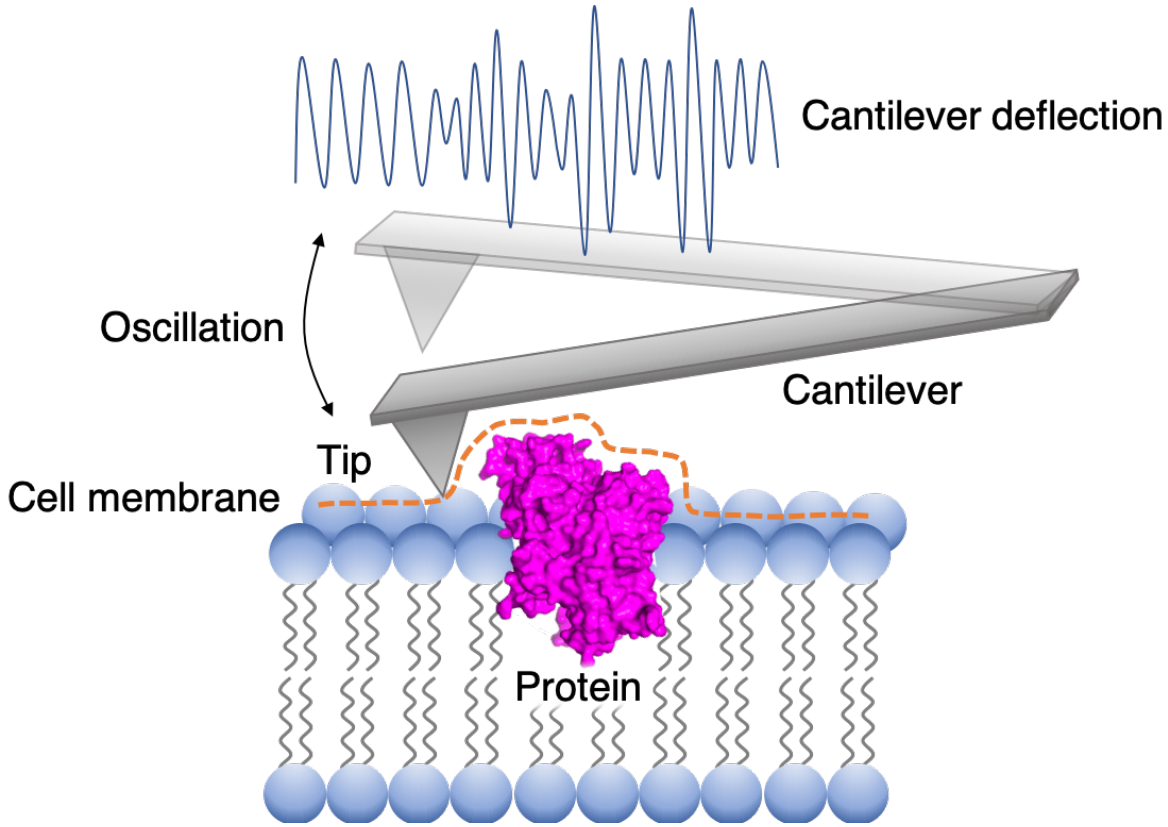


Fig. 1.11: Schematic demonstrating the use of AFM to image proteins embedded in live cell membranes. Tapping (or oscillation) mode allows imaging of fragile biological samples. The deflection of the cantilever is measured and mapped to the surface to give information about the topology.

topography. The tip is attached to a spring-like cantilever, which measures the deflection of the tip from the sample, resulting in a topographical map of the surface. AFM has a wide range of applications including surface chemistry, semiconductor technology, molecular engineering and also biology.[114]

Application of AFM to biological samples is still a growing field. AFM is useful for biological analysis as it is flexible and relatively non-invasive; it can also be used to give dynamic information about biological surfaces in an aqueous and physiologically relevant environment. Different modes of AFM imaging have been developed for different applications.[114] Contact mode, for example, is where constant force is applied through the tip and the vertical movement of the tip over the surface is tracked to map its features. This mode is generally used for solid surfaces as it requires direct interaction with the sample; the resultant frictional force can damage delicate biological samples, including proteins and live cells.[115] To combat this, tapping (or oscillating) mode was developed (see Fig 1.11). This dynamic imaging mode involves

vertical oscillation of the tip while moving over the surface. The resonant frequency of the tip changes when near the sample, largely due to Van der Waals forces or dipole interactions, and can then be used as a feedback parameter to map the surface. AFM has a vertical resolution of 0.1 nm, which is sufficient to image singular proteins on a surface, even to a sub-molecular resolution.[116]

AFM has been used to image DNA for decades, and dsDNA can now be imaged with helical resolution. Triplex DNA has been previously imaged by AFM; the first images indicate the formation of an intramolecular triplex formed in a supercoiled plasmid at acidic pH.[117] A more recent study imaged poly(dG)-poly(dG)-poly(dC) triplex DNA by AFM, showing the height of the structure was around 0.5 nm taller than the surrounding duplex.[118] This study also used AFM to follow the time-resolved enzymatic synthesis of the triplex structure by the Klenow exo- fragment DNA polymerase I, through the extension of the poly(dG) strand.[118] Secondary structures formed in supercoiled plasmids containing the $(GAA)_n$ FRDA sequence have also been imaged by AFM.[119] AFM has also been used to image G-quadruplexes in DNA mini-circles, as mentioned earlier.[113] G-quadruplexes can be seen as condensed protrusions from the mini-circle which form upon addition of certain cations.[113]

1.3.5 Fluorescent probes

Chemical probes have been developed to specifically image non-B DNA structures *in vitro* and *in vivo*. The primary detection method for these probes is fluorescence. Many fluorophores exhibit a 'switch on' fluorescence effect upon binding to nucleic acids.

Upon irradiation, fluorophores enter an excited state. To return to its electronic ground state, the excited molecule must release energy through radiative or nonradiative means (see Fig. 1.12). Radiative decay involves the emission of a photon, resulting in detectable fluorescence. Nonradiative decay involves transfer of energy through non-emissive processes such as vibration, rotation, internal conversion or intersystem crossing.[120] The fewer nonradiative decay pathways available, the higher proportion of decay occurs through radiative means, and the more emission is observed. Nonradiative decay is often governed by environmental factors, and this property can be useful for detecting binding to nucleic acids. Binding

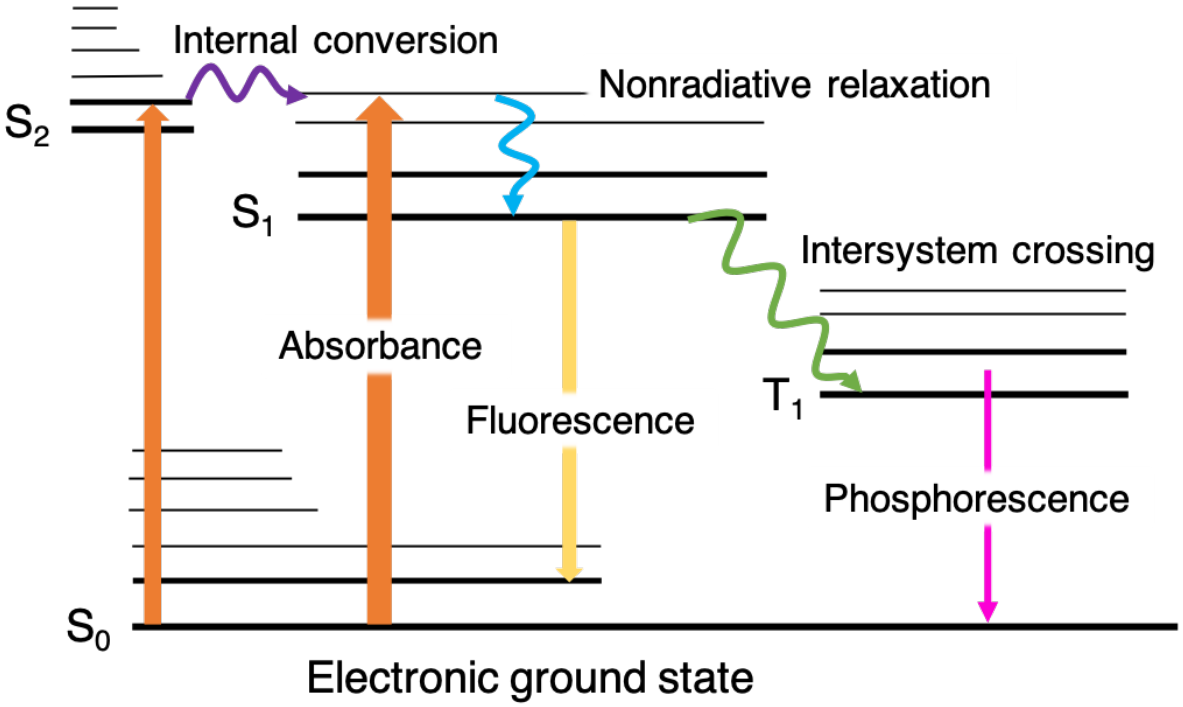


Fig. 1.12: Jablonski energy level diagram demonstrating excitation by absorption of a photon, partial relaxation by nonradiative decay and then decay back to ground state through fluorescence. Intersystem crossing is another form of nonradiative decay, in which energy is transferred from a singlet to a triplet excited state, and then quenched through phosphorescence.

to polynucleotides introduces the fluorophore to a more rigid environment, thus limiting motion and flexibility within the molecule. This will eliminate decay through nonradiative means such as rotation, and promote decay via the radiative pathway, increasing observable emission.[121] Detecting this increase in fluorescence can indicate the binding of the fluorophore to DNA/RNA. Different nucleic acid topologies can restrict the movement of bound molecules to varying extents, and hence increase the emission to a different extent also.

Such fluorescent probes are often used for imaging applications, such as staining fixed cells for fluorescence microscopy or visualising bands corresponding to nucleic acids in gel electrophoresis. One of the most common dyes for these purposes is 4′,6-diamidino-2-phenylindole (DAPI), which is a groove-binding molecule that displays a 20-fold increase in fluorescence intensity upon binding to dsDNA, particularly A-T-rich regions.[122] As described above, this fluorescence enhancement is due in part to geometric restriction within the DNA binding pocket. This prevents nonradiative relaxation through torsional movement, thus increasing radiative decay

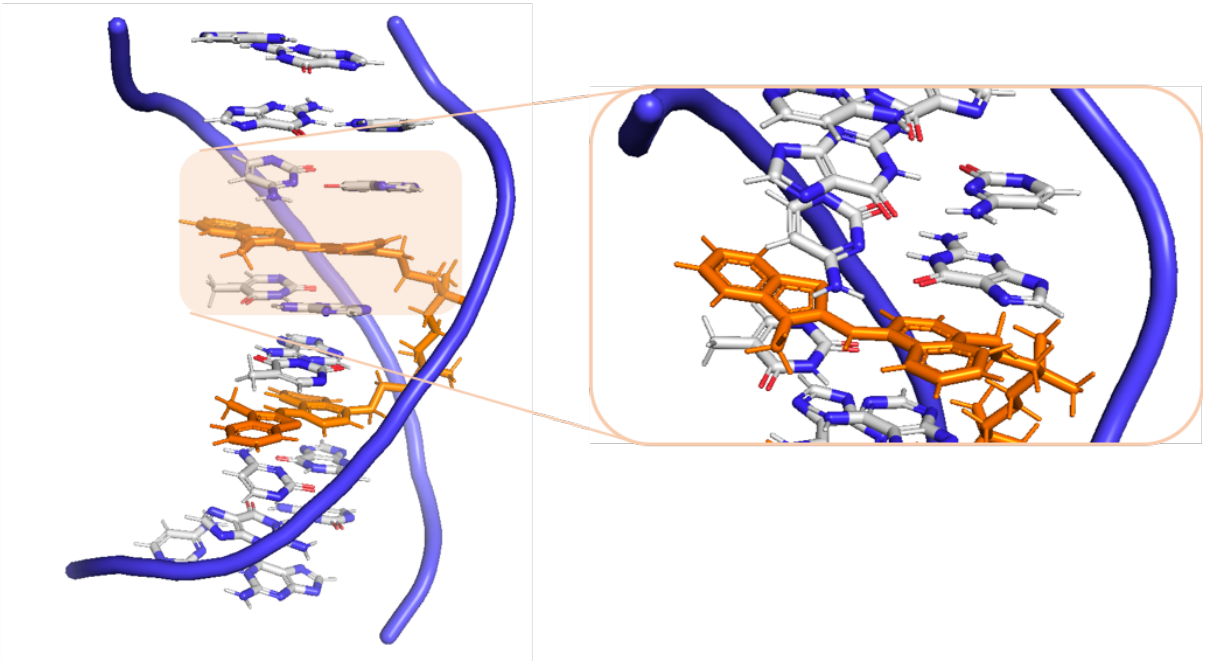


Fig. 1.13: The solution structure of TOTO intercalated into DNA, as determined by NMR spectroscopy (108D, taken from the PDB). TOTO is shown here in orange; the aromatic stacking interactions between the aromatic moieties of the TO monomer and the DNA bases is apparent.

and as a consequence, the fluorescence intensity.[123] DAPI emits in the blue region of the visible spectrum (461 nm) which makes it useful for staining applications that require more than one dye, as many fluorescent probes emit more towards the red. Another commonly used family of DNA dyes are Hoescht stains which also bind in the minor groove of dsDNA, and also to A-T-rich regions in particular.[124] These dyes are often used as substitutes for DAPI, as they are less toxic and more cell permeable, making them better suited for live cell imaging.[124] Another popular family of fluorescent DNA-binders are cyanine dyes. Pairs of cyanine fluorophores (such as Cy3 and Cy5) are often used for labelling oligonucleotides in FRET studies.[125] One cyanine-type molecule that has been investigated in particular for its non-canonical DNA binding properties is thiazole orange.

Thiazole orange (TO) has a large fluorescence response upon binding to DNA.[126] The fluorescence of free TO in solution is minimal, due to quenching of its excited state through rotation about the carbon-carbon single bond bridging the aromatic regions of the molecule. The fluorescence intensity and decay of fluorescence are dependent on the angle of twisting. TO contains two nitrogen atoms, one with a positive charge; these are in resonance with each other when TO is planar,

creating charge symmetry across the molecule. This symmetry is broken upon intramolecular twisting about the carbon-carbon single bond. Binding of TO to DNA prevents the rotation around this bond, limiting nonradiative decay and increasing fluorescence.[127] TO is a nucleic acid intercalator, meaning it stacks in-between the bases of DNA.[128] The better the structural overlap, the less rotation and the more fluorescence. Due to this, TO binding to ssDNA does not increase fluorescence by a notable amount, as the molecule is not fully intercalated into the DNA strand and hence some intramolecular twisting still occurs. Binding of TO to dsDNA, however, causes a large increase in fluorescence. TO also binds to dsDNA in a dimeric form, TOTO.[129] In this form, TOTO binds to dsDNA with such high affinity that the resulting complex is irreversible (Fig. 1.13).[129] It has been previously reported that TO also binds with high affinity to triplex and quadruplex DNA.[130] This can be attributed to better structural overlap and better intercalation (in triplex) or end-stacking (with quadruplex) over dsDNA. Despite this, no significant difference in fluorescence between TO bound to dsDNA and G-quadruplex was observed. Despite TO demonstrating enhanced binding to non-canonical structures, the lack of discrimination in fluorescence enhancement between these structures and dsDNA make using TO as a detection method for triplexes and quadruplexes challenging. However, another interesting study introduced aromatic substituents to the structure of TO, extending the conjugated system.[131] This was accompanied with a large increase in fluorescence enhancement upon binding to G-quadruplex over dsDNA.[131] Metal complexes with extended aromatic ligands have also been used as 'switch-on' probes for G-quadruplexes. These are attractive as nucleic acid binders as the ligands can be modified, allowing rational design to target certain topologies.[132] A study reported platinum(II) complexes with dipyridophenazine ligands exhibited high affinity binding to G-quadruplex over dsDNA, accompanied by a 293-fold increase in photoluminescence upon binding.[133]

Though there are many probes designed to target G-quadruplex DNA, there are few probes designed specifically to target triplex DNA. Due to the similarity in general structure between duplex and triplex, specificity is a common issue for triplex probes. The two primary modes of binding of molecules to triplex DNA are intercalation

and binding in the minor groove. Some examples of probes studied as triplex binders are described in the next two sections.

1.3.6 Intercalators for triplex DNA

Intercalators bind by slotting between nucleic acid bases and stacking with the aromatic regions. These molecules are often planar with a conjugated aromatic system, allowing intercalation between nucleotide bases within a helical structure. The extended aromatic system promotes π-π stacking interactions between the probe and the aromatic moieties of the bases.

Ethidium bromide, a popular nucleic acid stain with a conjugated aromatic structure, was one of the first compounds thought to bind to triplex DNA through intercalation.[134] Ethidium bromide is highly fluorescent upon binding to DNA, and a large increase in emission was seen upon binding to triplex DNA.[134] It was reported that ethidium bromide also stabilised the triplex, seen by increased melting temperatures, and the calculated binding constants demonstrated a stronger binding to triplex over the corresponding duplex.[134] However, ethidium bromide was found to be specifically stabilise T*A-T base triads, and destabilise C^+*G-C triads.[134] As well as this, ethidium bromide shows no selectivity for triplex DNA over other conformations.

Since then, many DNA intercalating molecules have been found to stabilise triplexes in a similar way with similar structural features. Alkaloids, such as berberine and coralyne, have shown promise as triplex stabilisers, and do not exhibit the same levels of base specificity that ethidium bromide does.[135–137] These molecules commonly have a planar extended aromatic system suitable for slotting into a helix, and a positively charged nitrogen within the ring system to increase interaction with the negatively charged phosphate nucleotide backbone. These molecules may stabilise triplexes to a larger extent than dsDNA, however they do not bind specifically to triplex.

These studies led to the rational design of one of the most selective and effective triplex ligands thus far: benzoquinoquinoxaline (BQQ) (see Fig. 1.15, left).[138] Molecular modelling studies of benzo[f]pyridoquinoxaline (BfPQ) derivatives with triplex DNA showed that they were exclusively overlapping with the Hoogsteen base pair of the triplex. It was suggested that incorporation of an additional aromatic ring

into the conjugated system would improve overlap across the whole base triplet.[138] BQQ has been reported to be cell permeable and stabilise intramolecular triplex structures formed in plasmids in living bacterial cells.[139] Subsequently, BQQ has been used as a targeted DNA cleaving agent through conjugation to ethylenediaminetetraacetic acid (EDTA) derivatives, achieving triplex-directed double-stranded cleavage of plasmid DNA.[140] This is initiated by addition and reduction of dithiothreitol (DTT) by Fe^{2+}, which produces highly reactive hydroxyl radicals that in turn cleave the dsDNA adjacent to the triplex.[140] BQQ has also been used to recognise intramolecular triplexes formed from $(GAA)_n$ FRDA sequences in supercoiled plasmids.[141] BQQ was then conjugated with a 10-phenanthroline derivate to create a triplex cleaving agent, which was shown to eliminate any secondary structures formed at the FRDA repeat mutation in the plasmid.[15,141]

Another family of intercalator molecules with triplex binding abilities are cyanine dyes, such as thiazole orange (TO), mentioned earlier (see Fig. 1.14, right). One study reported the formation of a stable triplex at physiological pH between dsDNA and a TFO conjugated with a TO moiety, which stabilised the triplex.[142] This was accompanied with a large increase in the fluorescence of TO, as is commonly seen with cyanine dyes upon binding to DNA.[142] Another study showed the preferential binding of TO to triplex and quadruplex over dsDNA, showing staining of particular regions in the nucleus that are not stained by dsDNA binder DAPI.[130] However, TO shows no particular specificity for any type of DNA.

Some anticancer compounds have a mechanism of action involving binding to DNA to prevent its transcription. A study investigating strategies for targeting the DNA sequence relating to protein HMGB1, a proposed target for cancer treatment, assessed the binding preferences of two anticancer drugs, Adriamycin and actinomycin, to dsDNA or triplex to ascertain whether targeted TFO therapy could be an option.[143] A TFO was designed to hybridise to the target dsDNA sequence encoding for HMGB1, and it was reported that both drugs bound preferentially to the triplex structure.[143] Both drugs bind to DNA through intercalation, with actinomycin also displaying minor groove-binding properties.[143]

Non-specific DNA intercalation can also be used to detect triplex formation. One study uses the concept of FRET to follow triplex formation, but rather than labelling

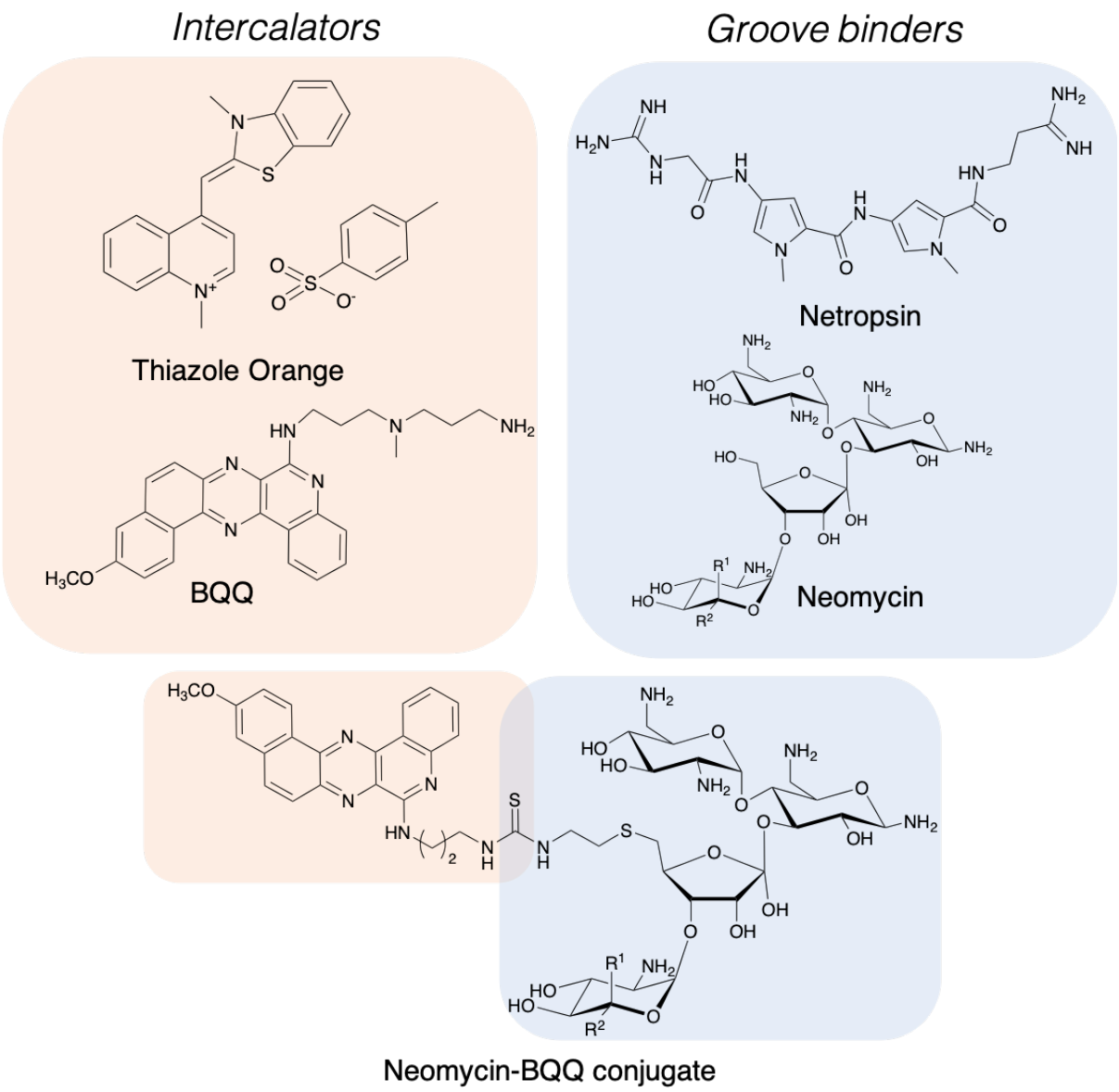


Fig. 1.14: Chemical structures of triplex intercalators (left), groove binders (right), and the structure of the neomycin-BQQ conjugate incorporating both intercalation and groove binding modes below.

both the target duplex and the TFO with fluorophores and measuring the FRET between them upon hybridisation, only the TFO was labelled with a fluorophore, and the FRET occurred between this and another fluorophore (SYBR green I) intercalated into the duplex.[144] However, this approach required 'colour compensation' to remove the fluorescence signal from the SYBR green I, and it was unclear whether the molecule was also intercalating into triplex DNA.[144]

1.3.7 Groove binders for triplex DNA

Another mode of non-covalent small molecule binding with nucleic acids is groove binding. Groove binding compounds interact with the walls of the minor groove through both van der Waals forces and directed hydrogen bonding with the peripheral functional groups of the nucleotide bases.[145] General structural characteristics of groove binding molecules include non-rigid aromatic systems with a cationic charge for interaction with the phosphate backbone, and substituents with hydrogen-bonding ability to interact with the function groups of the bases.[145] One of the most well-known groove binders is DAPI; a dsDNA binder that binds strongly to A-T sequences and is commonly used as a nuclear stain for fixed cells in fluorescence microscopy.[123] Generally, groove binders are found to destabilise triplexes. For example, netropsin (see Fig. 1.14, right), a natural DNA groove-binder, has been previously reported to destabilise triplexes while having a stabilising effect on the corresponding dsDNA (see Fig. 1.15).[146]

Aminosugars have been investigated for recognition of nucleic acids through groove-binding. Certain antibiotics, such as aminoglycosides, have been considered as triplex binders. Neomycin, an aminoglycoside, has been found to stabilise triplex DNA and interacts through minor groove binding (see Fig. 1.14, right).[147] Neomycin has been previously reported to selectively stabilise triplex DNA in the presence of salt, with no stabilising effect seen for dsDNA.[147] The downside to this molecule is that it has no inherent fluorescence, therefore eliminating a useful detection method. A subsequent study conjugated neomycin with fluorescent intercalator compounds, for which the relative binding affinities to triplex and duplex DNA could be assessed through fluorescent intercalator displacement (FID) assay.[148] The combined triplex intercalator moiety and groove-binding neomycin moiety resulted in a possible cooperative binding mode, stabilising triplex DNA to a further extent than neomycin alone.[148] The conjugate of BQQ and neomycin was found to be most effective at specifically stabilising triplex over dsDNA (see Fig. 1.14, bottom).[148] However, most intercalators used also bound to dsDNA, and hence absolute selectivity could not be achieved.

Despite fluorescence being a powerful detection tool for imaging nucleic acids, the lack of complete specificity of most probes to a particular DNA topology makes

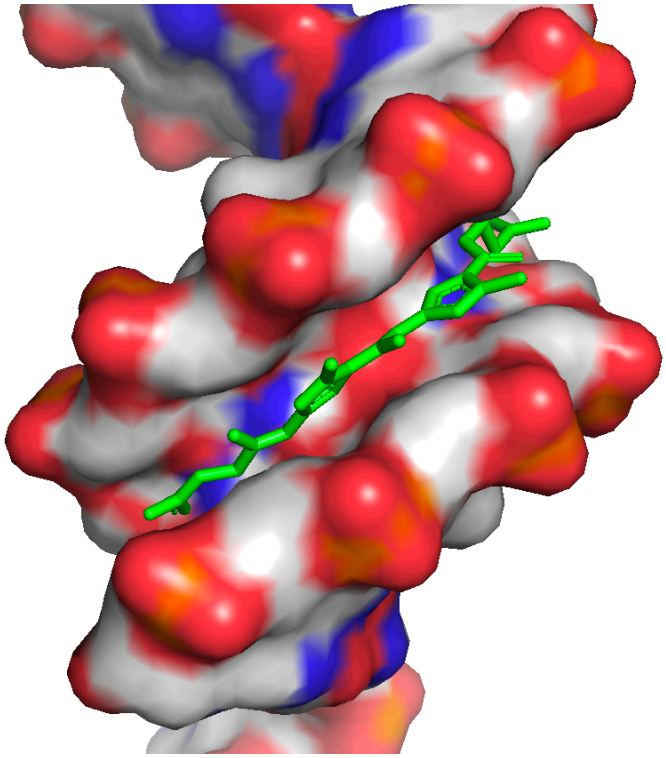


Fig. 1.15: Netropsin (in green) bound to d(GGCCAATTGG) dsDNA in the minor groove (C and H in grey, O in red, N in blue and S in orange, 1Z8V, taken from the PDB).[199]

application to live cell imaging difficult. This is especially problematic when detecting a structure that represents a small proportion of the total nucleic acid structure in the cell, such as triplexes.

1.3.8 Fluorescence lifetime imaging

One of the primary drawbacks to using intensity-based probes for cellular imaging is the inability to differentiate whether an increased fluorescence signal is due to binding to a target, or simply differences in cellular uptake. To combat this, fluorescence lifetime imaging can be used. Fluorescence lifetime imaging is a technique that measures the time-resolved decay of fluorescence following excitation and during the return to its ground state. The main advantage of this technique is that it is concentration and intensity independent. This eliminates the need for a specific probe with a high affinity for a certain DNA topology, or homogeneous cellular uptake.[120]

Once a fluorophore is in an excited state, the length of time it spends in that excited state, its fluorescence lifetime, depends on the decay pathways available. The

more decay pathways available, the faster the excited molecule will return to its ground state, and the shorter the fluorescence lifetime will be. Nonradiative decay is often governed by environmental factors. For example, for molecules that have a rotational nonradiative decay pathway, restriction of rotation by a rigid environment can impede the decay pathway and increase the fluorescence lifetime. This can be applied to the binding of probes with nucleic acids. As previously discussed, intercalation of a small organic fluorophore into a helical nucleic acid structure would restrict any possible rotation, and thus eliminate the nonradiative decay pathway associated with that rotation. This would result in an increased fluorescence lifetime in comparison to the free probe. Provided the fluorescence lifetime of a fluorophore is distinct enough when bound to different DNA topologies, these topologies can be distinguished *in vitro*, and potentially *in vivo*.

An early study using this technique reported DNA binding probes exhibited different fluorescence lifetimes in cells at different stages of their life cycle.[126] Another study used BODIPY, a fluorescent molecular rotor, linked to a nucleoside to feedback information on changes to the local environment in cells through variation in the fluorescence lifetime of BODIPY.[149] Another study uses fluorescent analogues of deoxycytidine to distinguish the transition between single- and double-stranded DNA, showing a 2 ns difference between the two topologies.[150] Trianguleniums, a family of planar aromatic DNA binders, have demonstrated significant difference in fluorescence lifetime when bound to different DNA topologies.[151] It has been previously reported that triangulenium DOATA-M2 exhibited longer lifetimes when bound to G-quadruplexes *in vitro* than single- and double-stranded nucleic acids (see Fig. 1.16).[151] In addition, fluorescence lifetime imaging microscopy (FLIM) was used to image DOATA-M2 in live cells, with evidence that suggested these probes could be used to study G-quadruplexes *in vivo*.[151]

FLIM allows the principles of fluorescence lifetime imaging to be applied to imaging in live cells. This can give information about the molecular environment *in vivo* and biological processes such as protein interactions, nucleic acid folding and condensation of chromatin. FLIM involves recording a microscopy image where each pixel has a fluorescence lifetime value associated with it. Techniques such as FLIM-FRET allow access to information not provided by steady-state FRET. It eliminates

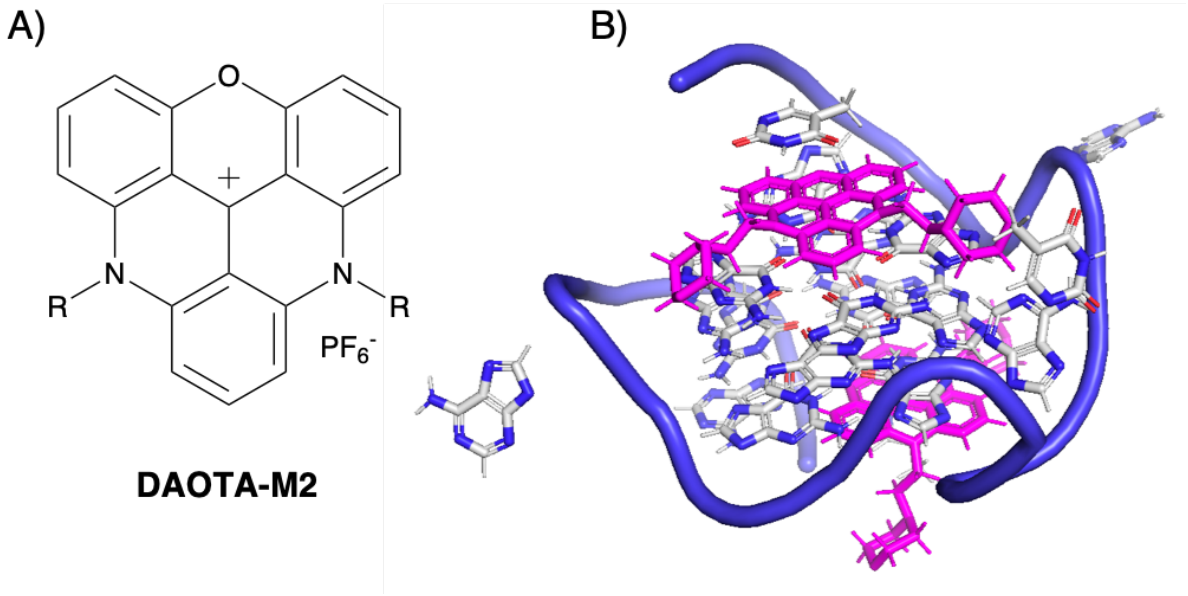


Fig. 1.16: A) Chemical structure of triangulenium **DOATA-M2**, B) structure of **DAOTA-M2** (in pink) bound to G-quadruplex c-myc (5LIG, taken from the PDB).[151] **DOATA-M2** stacks on either end of the quadruplex structure, with the aromatic R substituents interacting with the side pockets of the quadruplex.[151]

the problem of direct acceptor fluorophore excitation, increasing the range of fluorophores that can be used. FLIM-FRET can also provide information on interacting and non-interacting species, which is useful for protein binding studies.[152] FLIM has been previously used to follow changes in chromatin in living cells by measuring the change in fluorescence lifetime of DAPI. The lifetime of the probe revealed whether the chromatin was compacted or relaxed.[153] FLIM presents a powerful intensity-independent technique for probing cellular environments which can be adapted to target many different processes.

1.4 Aims and objectives

The work presented herein broadly aims to address whether triplexes play a role in the silencing of the *FXN* gene in FRDA.

Firstly, Chapter 2 aims to investigate the formation and stability of triplexes formed from synthetic oligonucleotides containing the $(GAA)_n$ repeat sequence of interest under varying conditions. Do stable triplexes form from short synthetic oligonucleotides containing the FRDA triplet repeat expansion mutation under physiological conditions, and how does the length of the repeat sequence effect the

stability of these structures? The triplex formation was assessed using gel electrophoresis, CD and FRET.

Secondly, in Chapter 3 aims to investigate the formation of triplexes from long oligonucleotides derived from plasmids, containing the $(GAA)_n$ sequence of interest at pathological and 'pre-disease' lengths. Do triplexes form from oligonucleotides of pathological length, and are any other alternative structures such as 'sticky DNA' also seen? Secondary structure formation was primarily assessed using gel electrophoresis, with structures also visualised using AFM.

Lastly, Chapter 4 aims to investigate whether triplex DNA is present in cells by employing fluorescent, DNA-binding probes based on the structure of thiazole orange. Can fluorescence lifetime imaging microscopy be used to distinguish triplex DNA or other alternative structures in cells containing the FRDA $(GAA)_n$ mutation? The optical properties of the probes when bound to various DNA topologies were investigated.

2 Biophysical studies of triplex formation from short synthetic oligonucleotides

2.1 Introduction

In this chapter, the study and biophysical characterisation of secondary structures formed from short synthetic oligonucleotides containing the FRDA repeat trinucleotide sequence of interest under varying conditions is reported. Sequences containing $(GAA)_n$ inserts, where n = 5, 10 and 20, were studied using techniques such as gel electrophoresis and circular dichroism spectroscopy (CD).

In particular, the effect of pH, salt concentration and crowding conditions on secondary structure formation was investigated. As previously discussed, pH is a crucial factor in the formation of triplex DNA *in vitro*. Formation of pyrimidine triplexes requires the protonation of a nitrogen atom on the cytosine bases of the triplex-forming oligonucleotide (TFO) to form the second Hoogsteen bond between C^+ and G for the base triplet C^+*G-C. The purine triplex can form at the physiological pH of 7.4 as no protonation is required.[154]

Triplex formation *in vitro* is highly dependent on the salt concentration present in the buffer.[97] In general, monovalent cations such as sodium and potassium facilitate the formation of secondary DNA structures.[155] This is due to the positively charged salt cations shielding the negatively charged phosphate backbone of the DNA and thus reducing charge repulsion effect between the DNA strands during annealing. It has been shown that divalent cations in particular, such as Mg^{2+}, aid in triplex formation.[37,38,41,97,156]

Molecular crowding can mimic the conditions *in cellulo* by emulating high concentrations of macromolecules such as proteins.[44] Conditions such as these are

prevalent in living cells, particularly in the nucleus where these secondary DNA structures would be found. Polyethylene glycol (PEG) is a commonly used crowding agent in *in vitro* assays.[157]

Understanding how these variable conditions affect the formation and stability of triplexes in a controlled environment at increasing $(GAA)_n$ repeat unit lengths is a critical foundation for the study of pathological repeat lengths *in vitro* and *in vivo.* The oligonucleotides used in this chapter are shown in Table 2.1 below. These sequences, though not pathological, have been previously reported to form triplex structures under controlled conditions at physiological pH.[93] By varying the length of the repeat sequence, length-dependency of stability can also be assessed.

Table 2.1: The name and sequence of oligonucleotides used for the triplex formation and stability assays.

Name	**Sequence 5' - 3'**
GAA_5	TCGC **$(GAA)_5$** CGCT
CTT_5	AGCG **$(TTC)_5$** GCGA
Purine_TFO_5	$(AAG)_5$
Pyrimidine_TFO_5	$(CTT)_5$
GAA_{10}	TCGC **$(GAA)_{10}$** CGCT
CTT_{10}	AGCG **$(TTC)_{10}$** GCGA
Purine_TFO_{10}	$(AAG)_{10}$
Pyrimidine_TFO_{10}	$(CTT)_{10}$
GAA_{20}	TCGC **$(GAA)_{20}$** CGCT
CTT_{20}	AGCG **$(TTC)_{20}$** GCGA
Purine_TFO_{20}	$(AAG)_{20}$
Pyrimidine_TFO_{20}	$(CTT)_{20}$
TAMRA-GAA_5	TCGC **$(GAA)_5$** CGCT-TAMRA
FAM-Purine_TFO_5	FAM-$(AAG)_5$

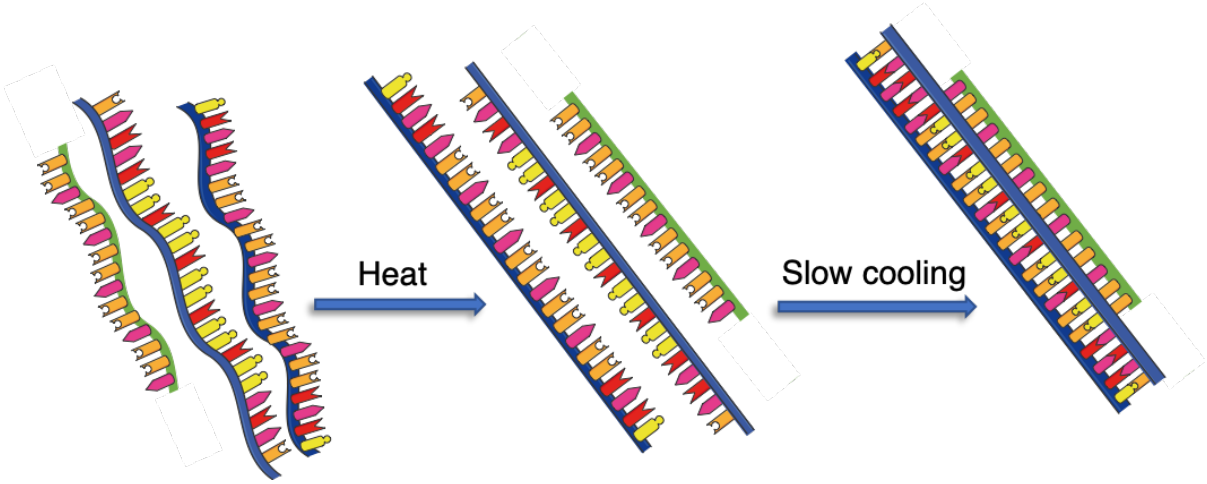


Fig. 2.1: Schematic demonstrating the annealing of a triplex structure from the oligonucleotides through heating and then slow cooling. The two oligonucleotides comprising the duplex are coloured in blue, while the TFO is coloured in green.

As described in the Methods (6.1.2), oligonucleotides were annealed by briefly heating to 95 °C and then cooling slowly. The heating serves to disrupt any unwanted secondary structures such as hairpins formed by self-hybridisation of the oligonucleotides in the stock solutions.[42] The slow rate of cooling allows for formation of the most energetically favourable product, which would be the fully hybridised structure with no loose ends. This is particularly important for these sequences that contain long stretches of repeats, as the strands have multiple complementary regions and could hybridise incorrectly with overhanging ends. The flanking bases introduced either side of the repeat insert in the oligonucleotides forming the dsDNA structure strive to resolve this by introducing non-repeat complementary regions to encourage a 'perfect' annealing (Fig. 2.1).

Following annealing, secondary structure formation within samples was investigated using a variety of techniques. Gel electrophoresis is most commonly used, as this allows direct visualisation of structures separated according to molecular weight and size. Intermolecular triplex structures have a higher molecular weight than their counterpart duplexes due to binding of the TFO and will therefore move more slowly through the gel. However, this technique relies on secondary structures being stable enough to remain intact during the assay. Circular dichroism (CD) is a powerful

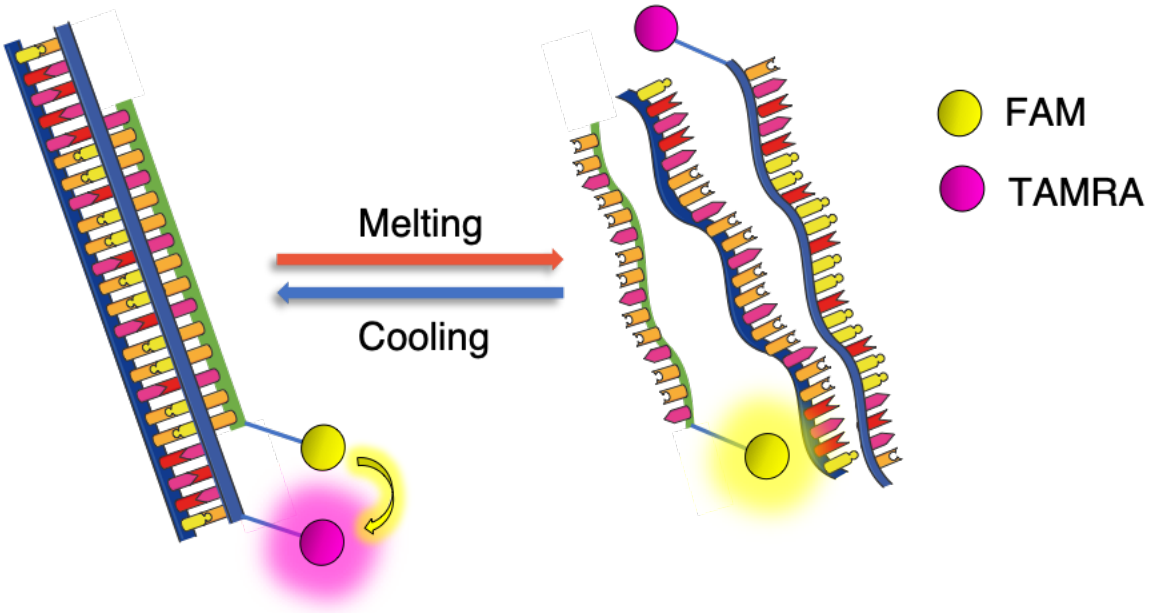


Fig. 2.2: Schematic demonstrating the design of the fluorophore-labelled oligonucleotides and how the FRET occurs between the pair. The purine strand of the duplex (coloured blue) is labelled with TAMRA (pink), while the purine TFO (green) is labelled with FAM (yellow). Upon association of the strands, FRET can occur, increasing emission of TAMRA, the acceptor fluorophore.

technique for analysing large biomolecules. CD involves the differential absorption of left- and right-handed circularly polarised light by chiral molecules, seen as absorption bands in a spectrum. Biomolecules exhibit circular dichroism due to the chirality in their higher order structures, and as such CD can be used to gain information about structure and conformation. Structures often have a 'signature' spectral feature: for example, CD is commonly used to distinguish between parallel and antiparallel G-quadruplex conformations. For triplexes, particularly ones containing GA, CT or GT repeats, this signature is a distinctive negative band at 210 nm.[158] CD melting can be used to observe the denaturation of a structure by following the disappearance/appearance of characteristic peaks. This technique is conducted entirely in solution and as such the experimental conditions are less harsh. One limitation of CD is that it is unable to distinguish multiple structures in a solution. To supplement this, Förster resonance energy transfer (FRET) between fluorophore-labelled oligonucleotides was also used. This technique relies on the quenching of fluorescence between a donor and an acceptor fluorophore. These fluorophore pairs are selected such that the excitation wavelength (λ_{ex}) of the donor is greatly blue-shifted in comparison to the acceptor. With spectral overlap, and when in close

proximity (within 10 nm), a nonradiative energy transfer from the excited state of the donor to the excited state of the acceptor can occur. In this case, the FRET pair 6-carboxyfluorescein/5-tetramethylrhodamine (FAM/TAMRA) was used. By introducing a donor fluorophore (FAM) onto the 5' end of the TFO and an acceptor fluorophore (TAMRA) onto the 3' end of the purine-based duplex strand, the FRET between these fluorophores can be measured and the extent of association between the strands can be assessed.[109] The sample was irradiated at λ_{ex} for FAM, and if the TFO is bound to the duplex and the fluorophores are spatially close, FRET can occur, causing an increase in TAMRA emission. If the TFO is not bound, only FAM emission will be seen. FRET melting experiments can be a useful tool to follow the denaturation of a secondary structure by relative increase/decrease of fluorescence of the donor/acceptor fluorophores attached (Fig. 2.2).

2.2 Results and discussion

2.2.1 Effect of pH

To study how pH affects the formation of triplexes, two conditions were investigated; acidic (pH 5.5) and physiological (pH 7.4). Samples were annealed as described in the Methods section (6.1.2), with the appropriate stoichiometry of DNA strands and total DNA concentration. The stoichiometry was kept 1:1 for the strands forming the duplex, with the TFO added at 1 or 2 equivalents to the duplex. Triplex formation in the samples was then assessed using gel electrophoresis, CD spectroscopy and FRET melting.[201]

Triplex formation from the shortest oligonucleotides, with just 5 (GAA) repeat units, were tested initially. These sequences had been previously reported to form triplexes seen by CD and gel electrophoresis.[93] It was found that the pyrimidine triplex formed readily at pH 5.5, but not at pH 7.4. This can be seen by a heavier band present in the gel that corresponds to the weight of the dsDNA plus the TFO (lanes 3 and 4, Fig. 2.3). In addition, the CD spectrum shows a negative band at 210 nm that is characteristic of triplex structure (Fig. 2.4, A, pink). This band is absent in the sample annealed at pH 7.4 (Fig. 2.4, A, orange). The CD spectra also display characteristic

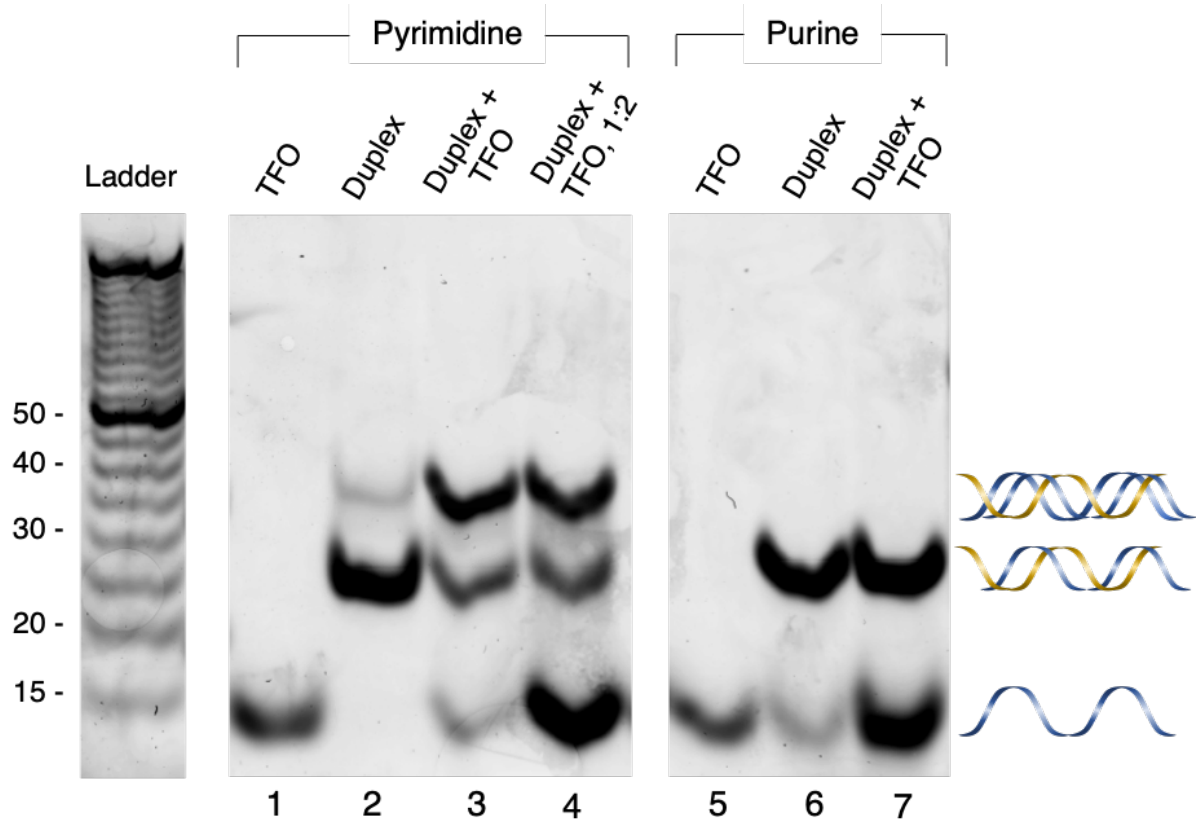


Fig. 2.3: 15% PAGE of 5 repeat unit purine and pyrimidine sequences annealed in 10 mM NaCac buffer (pH 5.5) containing 100 mM NaCl and 30 mM $MgCl_2$. Ladder: O'RangeRuler 5 bp (Thermo Scientific). Lanes: 1) pyrimidine TFO, 2) duplex, 3) duplex with pyrimidine TFO, 4) 1:2 duplex with pyrimidine TFO, 5) purine TFO, 6) duplex, 7) duplex with purine TFO.

duplex bands: a broad positive band at 279 nm and a broad negative band at 245nm. This is to be expected as the base structure of a triplex is very similar to a duplex. There is still dsDNA present in the sample as can be seen on the gel (lanes 3 and 4, Fig. 2.3), with the middle band corresponding to the double stranded structure. The unbound TFO can also be seen as the band that has travelled the furthest. This is especially prominent in the sample with a 1:2 ratio of duplex to pyrimidine TFO (lane 4, Fig. 2.3), demonstrating that the majority of the TFO added is in an unbound state. The purine triplex structure was not observed at either pH. The gel above shows that the sample containing the duplex annealed with the purine TFO displays a band corresponding to dsDNA with no trace of a heavier structure present (lane 7, Fig. 2.3). The prominence of the band relating to the purine TFO indicates that most, if not all, is unbound. The CD spectra show no negative band at 210 nm at either pH 5.5 or pH 7.4 (Fig. 2.4, B). CD melting experiments were carried out on both purine and pyrimidine triplexes annealed at pH 5.5 and pH 7.4. The pyrimidine triplex annealed

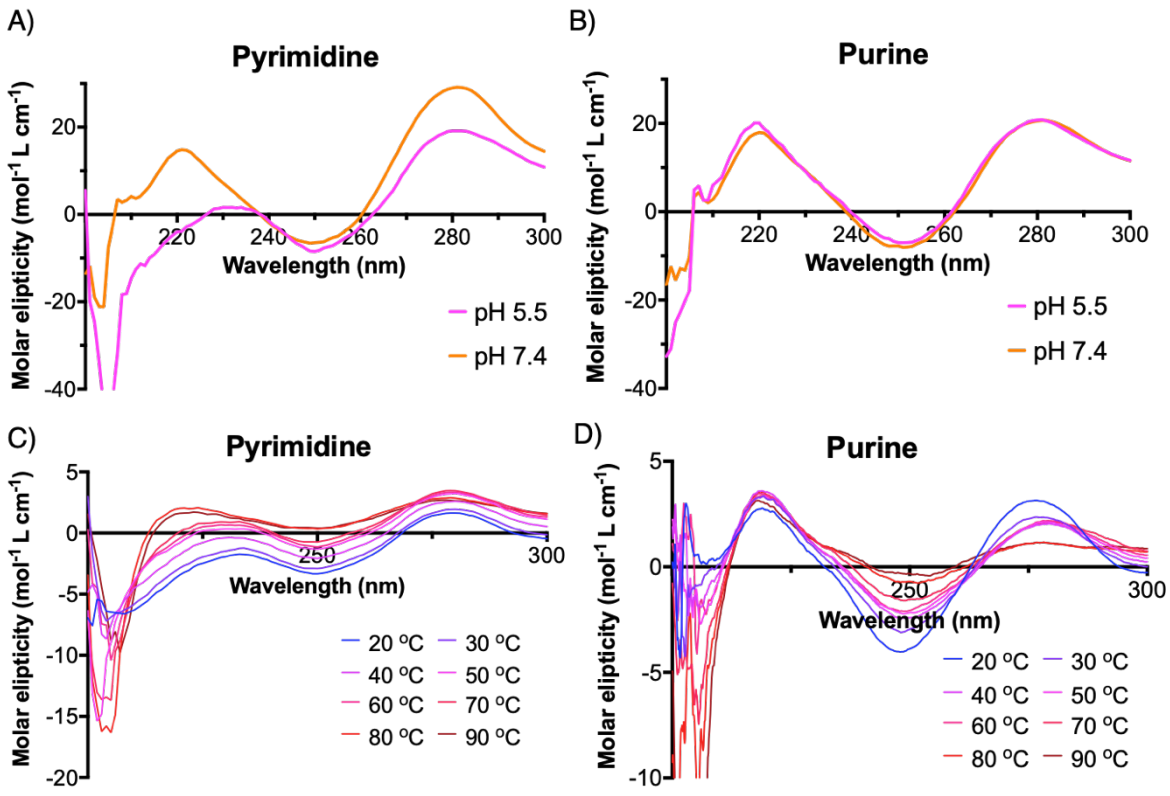


Fig. 2.4: A) CD spectra of duplex and pyrimidine TFO annealed at pH 5.5 (pink) and pH 7.4 (orange). The characteristic negative band at 210 nm is present at pH 5.5. B) CD spectra of duplex and purine TFO annealed at pH 5.5 (pink) and pH 7.4 (orange). There is no characteristic triplex band present in either spectrum. C) CD melting spectra of pyrimidine triplex annealed at pH 5.5. D) CD melting spectra of purine triplex annealed at pH 7.4.

at acidic pH shows a strong negative band at 210 nm at cooler temperatures, which disappears as the sample is heated (Fig. 2.4, C); this band was not observed at pH 7.4. However, the instrumental distortion at around 210 nm obscures this slightly as it seems to worsen with increased temperature. This distortion is also seen in the CD spectra of the purine sample; however, it is clear there is no negative band at 210 nm at any temperature, suggesting the purine triplex did not form (Fig. 2.4, D).[201]

Though the gel electrophoresis and CD suggest that there is no purine triplex forming, it is possible there is association between the strands that is not strong enough to withstand the conditions in the gel or too transient to detect with CD. Therefore, FRET melting experiments were carried out with a FAM-labelled purine TFO and a TAMRA-labelled purine duplex strand. The sample was annealed as previously described and then irradiated at the FAM excitation wavelength (λ_{ex}) of 494 nm, while the temperature was slowly increased from 25-95 ·C and then decreased. The intensity of TAMRA emission at a wavelength (λ_{em}) of 580 nm was measured. If

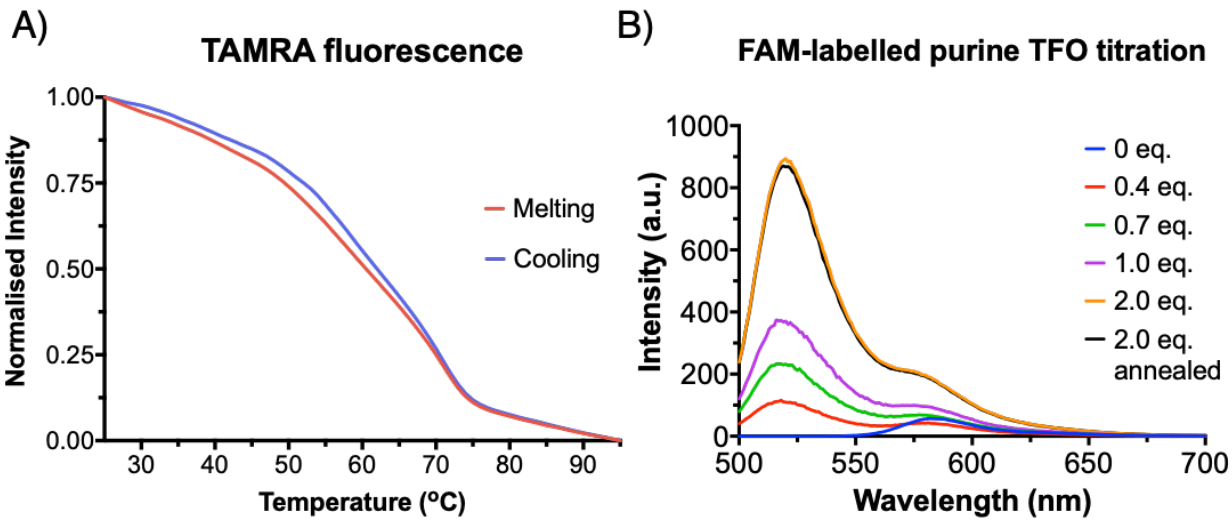


Fig. 2.5: A) FRET melting profile of TAMRA emission for $(GAA)_5$ TAMRA-labelled duplex annealed with FAM-labelled purine TFO at pH 7.4. B) Fluorescence titration of increasing equivalents of $(GAA)_5$ FAM-labelled purine TFO into TAMRA-labelled duplex at pH 7.4.

a triplex was forming, it would be expected that TAMRA emission would be increased at low temperatures and then decrease as the structure melted with the increasing temperature. Additionally, triplexes have a biphasic melting profile: the TFO should be lost from the structure at a lower temperature of around 50 °C, and subsequently the duplex should melt at around 70 °C. It can be seen from the plots (Fig. 2.5, A) that TAMRA fluorescence is at its maximum at low temperature and decreases as the temperature is increased. This could suggest the presence of a secondary structure between the duplex and the purine TFO that melts between 50 and 70 °C. Both melting and cooling profiles are in accordance with each other.[201]

In addition to this, a fluorescence titration was performed with increasing amounts of FAM-labelled TFO titrated into a solution containing the TAMRA-labelled duplex, annealed as before. Following the same principles of FRET, if a triplex is forming the spectra would show an increase in TAMRA emission at 580 nm with increasing equivalents of TFO added. Irradiating the TAMRA-labelled duplex at FAM λ_{ex} with 0 equivalents of FAM-labelled TFO should give a near-baseline spectrum; this can be seen in Fig. 2.5, B (the blue trace). There is a small amount of TAMRA emission which can be attributed to non-specific excitation. Upon the addition of 0.4 equivalents of FAM-labelled TFO, the TAMRA emission in fact decreases slightly and there is an increase in FAM emission at 520 nm. This increase continues as more equivalents of TFO are added, however the band corresponding to TAMRA emission does also

increase. Once 2 equivalents of TFO were added, the sample was annealed overnight and the emission spectrum was recorded again, giving the same result. This suggests that the purine TFO is not binding to the duplex, as the increase in FAM emission is due to the increasing concentration of FAM in solution and is not being significantly dampened by FRET with TAMRA. The increase in TAMRA emission is likely due to the presence of more and more FAM-labelled TFO in solution, meaning the pair will inevitably be in closer proximity despite not being bound.[201] This could be further validated with similar experiments with free FAM and TAMRA dyes.

Following the conclusion that the purine triplex was not formed from the oligonucleotides containing the 5 repeat unit insert, the insert length was increased to 10 repeat units. By increasing the length of the $(GAA)_n$ repeat sequence, the stability of any resulting structures should increase due to more interactions between the participating DNA strands. Again, it was found that pyrimidine triplex formed at pH 5.5,

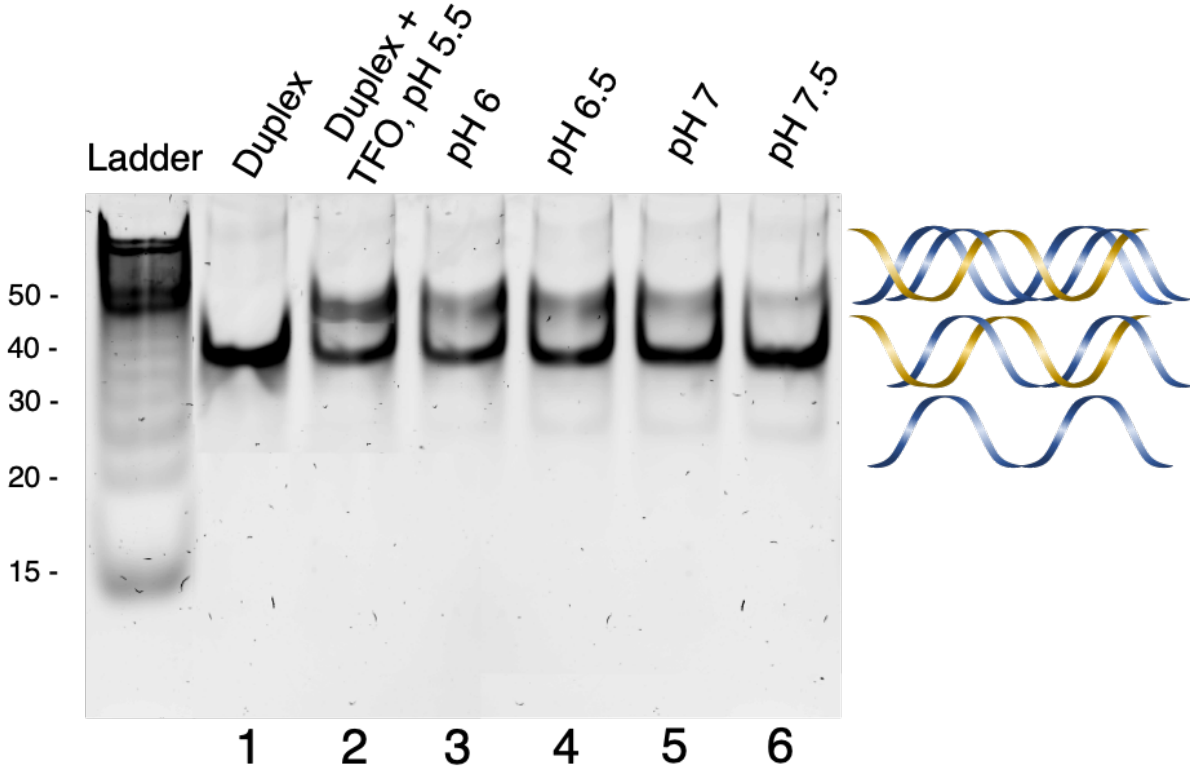


Fig. 2.6: 12% PAGE of $(GAA)_{10}$ pyrimidine sequences annealed in 10 mM NaCac buffer (pH 5.5) containing 100 mM NaCl and 30 mM $MgCl_2$, and subsequently incubated in the same buffer at increasing pH. Ladder: O'RangeRuler 5 bp (Thermo Scientific). Lanes: 1) duplex, 2) duplex with pyrimidine TFO annealed at pH 5.5, 3) duplex with pyrimidine TFO incubated at pH 6, 4) duplex with pyrimidine TFO incubated at pH 6.5, 5) duplex with pyrimidine TFO incubated at pH 7, 6) duplex with pyrimidine TFO incubated at pH 7.4.

but not at pH 7.4, and the purine triplex was not seen at either pH, as assessed by gel electrophoresis and CD spectroscopy.[201]

To increase the stability of potential secondary structures, the insert length was increased to 20 repeat units. In a normal copy of the *FXN* gene, the average amount of (GAA) repeats is between 8-22. Thus, although 20 repeats are not believed to be pathological, it is nearing 'pre-disease' length. However, the same results were observed; pyrimidine triplexes formed readily at pH 5.5 but not at pH 7.4, and purine triplexes were not observed to form at either pH.[201]

Since neither triplex was seen to form when annealed at pH 7.4, the resilience of pyrimidine triplexes formed at pH 5.5 to increase in pH was investigated. Pyrimidine triplex structures annealed at pH 5.5 were subsequently incubated in sodium cacodylate (NaCac) (10 mM) buffers at increasing pH from 6-7.5 and samples were analysed by gel electrophoresis. It was found that structures appeared to persist to some extent up to pH 7.5, though it can be seen that above pH 7 that the primary structure in the sample is duplex (lanes 5 and 6, Fig. 2.6).[201]

The CD, gel electrophoresis and FRET results presented in this section show that the pyrimidine triplexes form readily at pH 5.5. This was anticipated, as pyrimidine triplex formation requires protonation of nitrogen on the cytosine bases of the TFO in order to form the second Hoogsteen hydrogen bond in the C^{+}*G-C triplet. Alongside this, pyrimidine triplexes have been shown to have higher strand affinity than their corresponding duplexes under low pH conditions.[159] This can especially be seen in the sample corresponding to the $(GAA)_5$ duplex annealed at pH 5.5 (lane 2, Fig. 2.3), which shows some evidence of triplex formation despite the complementary strand stoichiometry being 1:1. The readiness to form is due in part to the presence of the positive charge, which reduces the charge repulsion between approaching strands caused by the negatively charged phosphate backbone of the DNA.

The results showed that purine triplexes were not seen to form at either pH. This was in contrast to what was reported in the literature, as Jain *et. al.* previously reported the formation of intermolecular purine triplexes from identical sequences with a $(GAA)_5$ repeat insert.[93] These triplexes were observed with CD and gel electrophoresis using a similar annealing protocol with radiolabelled oligonucleotides.[93] Although, Bergquist

et. al. investigated binding of a 20-mer (GAA) TFO to a target duplex at neutral pH and did not observe any purine triplex formation by gel electrophoresis, even in the presence of triplex-stabilising agent BQQ; however pyrimidine triplexes were shown to form readily.[141] Purine triplexes are known to be much less stable than their pyrimidine counterparts.[160] This is partly due to the lack of positive charge to compensate for the charge repulsion between nucleotide strands, but is also related to binding orientation. The pyrimidine TFO binds in the major groove of the duplex in a parallel orientation, whereas purine TFOs bind in an antiparallel orientation, similar to Watson-Crick pairing. The triplets formed in the pyrimidine triplex, T*A-T and C^+*G-C, are structurally isomorphic in the parallel motif.[161] This means that when structurally imposed, the positions of the C-1' atoms of the Watson-Crick pairs and the TFO are almost identical.[161] This results in minimal backbone distortion of both the duplex and the TFO, and thus only slight perturbation of the duplex structure.[161] This has also been confirmed in previously reported NMR studies, where it was shown that the major groove was only slightly widened upon binding of the TFO in a parallel orientation.[162] Lastly, the formation of an antiparallel purine triplex requires G-rich oligonucleotides. G-rich sequences have a tendency to form other non-canonical structures such as G-quadruplexes and GA-duplexes, which can in turn out-compete formation of a relatively unstable triplex structure.[42,163]

2.2.2 Effect of salt concentration

The effect of varying salt concentrations in annealing buffers was tested. As discussed previously, the presence of salt cations is essential for the formation of many secondary DNA structures. They serve to reduce the charge repulsion between the negatively charged phosphate backbones of approaching DNA strands.[155] In particular, divalent cations such as Mg^{2+} have been shown to stabilise triplexes.[40] NaCac buffers (10 mM, pH 5.5 and 7.4) with increasing or decreasing amount of NaCl or $MgCl_2$ were made, and samples were annealed as described in Methods (6.1.2). As with the pH studies described above, oligonucleotides containing $(GAA)_n$ insert lengths of n = 5, 10 and 20 repeats were investigated. Triplex formation in samples was assessed using gel electrophoresis and CD spectroscopy, with the salt conditions

used summarised in Table 2.2. Generally, 150 mM NaCl and 10 mM $MgCl_2$ are considered to be physiological conditions.

Table 2.2: Salt conditions used in annealing. Concentration of NaCl and $MgCl_2$ was varied.

Salt condition	*NaCac/mM*	*NaCl/mM*	*$MgCl_2$/mM*
1	10	100	10
2	10	150	10
3	10	150	15
4	10	0	15

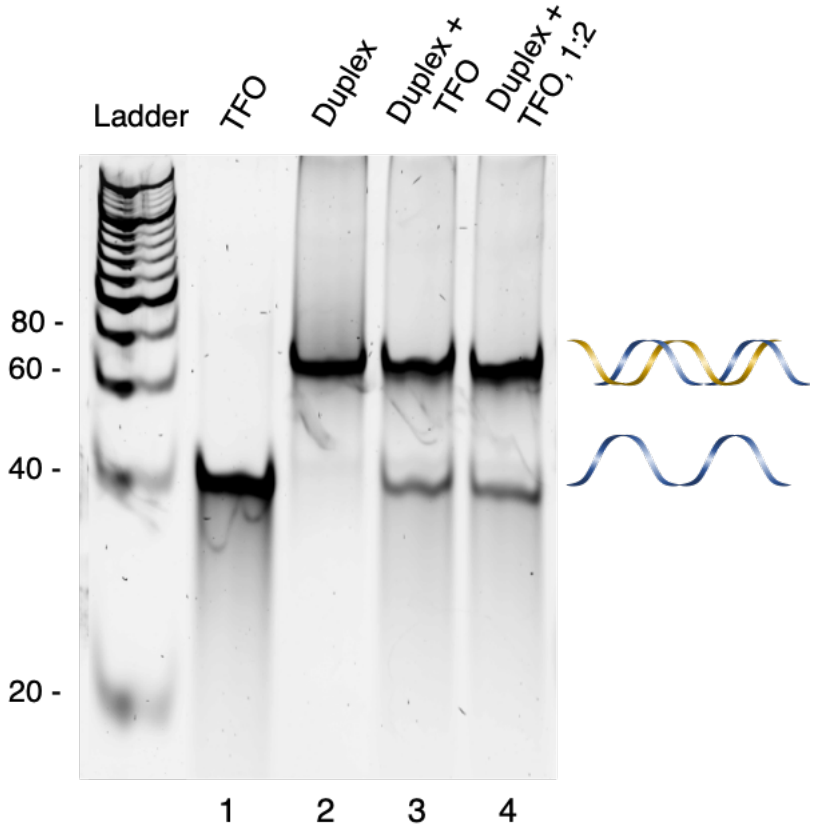


Fig. 2.7: 12% PAGE of $(GAA)_{10}$ purine sequences annealed in 10 mM NaCac buffer (pH 7.4) containing 15 mM $MgCl_2$. Ladder: O'RangeRuler 10 bp (Thermo Scientific). Lanes: 1) purine TFO, 2) duplex, 3) duplex with purine TFO, 4) 1:2 duplex with purine TFO.

Pyrimidine triplexes were seen to form when annealed at pH 5.5 under all salt conditions, from 5 repeat units up to 20 repeat units. Also, as previously observed, the pyrimidine triplex did not form at pH 7.4 under any of the salt conditions tested, from any length repeat sequence.

Purine triplexes were not seen to form under any condition. On the gel shown (Fig. 2.7), the only bands present in the lanes corresponding to the samples of duplex annealed with the purine TFO (lanes 3 and 4, Fig. 2.7) represent the dsDNA and below it the unbound TFO only.

It can be seen that the variation in salt concentration did not affect the formation of pyrimidine triplex at pH 5.5; the structure still formed readily under all salt conditions tested. The alternative salt conditions did not encourage the pyrimidine triplex to form at pH 7.4, which was to be expected as without protonation of cytosine in the TFO, the structure would be lacking the second Hoogsteen bond between the TFO and purine duplex strand required to form the C*G-C base triplet. The lack of protonation would also reduce the shielding of the charge repulsion between the DNA strands.

Despite increasing the concentration of Mg^{2+} and decreasing the concentration of Na^{+}, the purine triplex was still not seen to form at either pH. It has been shown that monovalent cations such as Na^{+} and K^{+} can inhibit triplex formation for the purine motif by promoting and stabilising competing dimers and tetramers formed from the G-rich TFO.[164] In previous studies, it was shown that divalent cations such as Mg^{2+} are essential for intermolecular triplex formation.[93,156] Such divalent cations have a higher ratio of charge to ionic radius than monovalent cations, which has been reported to be beneficial for the formation of certain types of triplexes. It has been previously reported that as well as binding to the negatively charged phosphate backbone of the DNA, divalent cations can also interact with certain sites on the bases themselves, including particular nitrogen and oxygen atoms on guanine bases.[37] The interaction of an electrophile with this guanine nitrogen atom can cause the protons bound to other guanine nitrogen atoms to become more acidic, which in turn will enhance the formation of reverse-Hoogsteen hydrogen bonds and increase the stability of the purine-rich triplex.[37] This is particularly relevant for oligonucleotides with runs of guanine bases. Adjacent adenines can repel the Mg^{2+} due to the positive electrostatic

potential created by the amino group, thus preventing the aggregation of Mg^{2+} and hence reducing its stabilising effect.[165] This could explain why the increase in Mg^{2+} does not facilitate purine triplex formation, as the purine TFO is primarily comprised of adjacent adenine bases. Other studies have suggested that Mg^{2+} only has a stabilising effect on the G*G-C triplet, and can in fact destabilise $(TC)_n(GA)_n$-$(AG)_n$ triplexes.[166]

Thus, the relationship between divalent cations and the stabilisation of triplex structures is more complicated than simple shielding of charge repulsion between strands and has been previously demonstrated to be highly sequence specific.

2.2.3 Effect of crowding conditions

Molecular crowding was introduced using PEG_{200} to mimic the crowded conditions *in cellulo*. The percentage of PEG_{200} used (v/v), calculated by molecular weight, was increased from 10% to 40%, and samples were annealed in buffer as before, keeping the salt concentration the same as in the pH experiments.

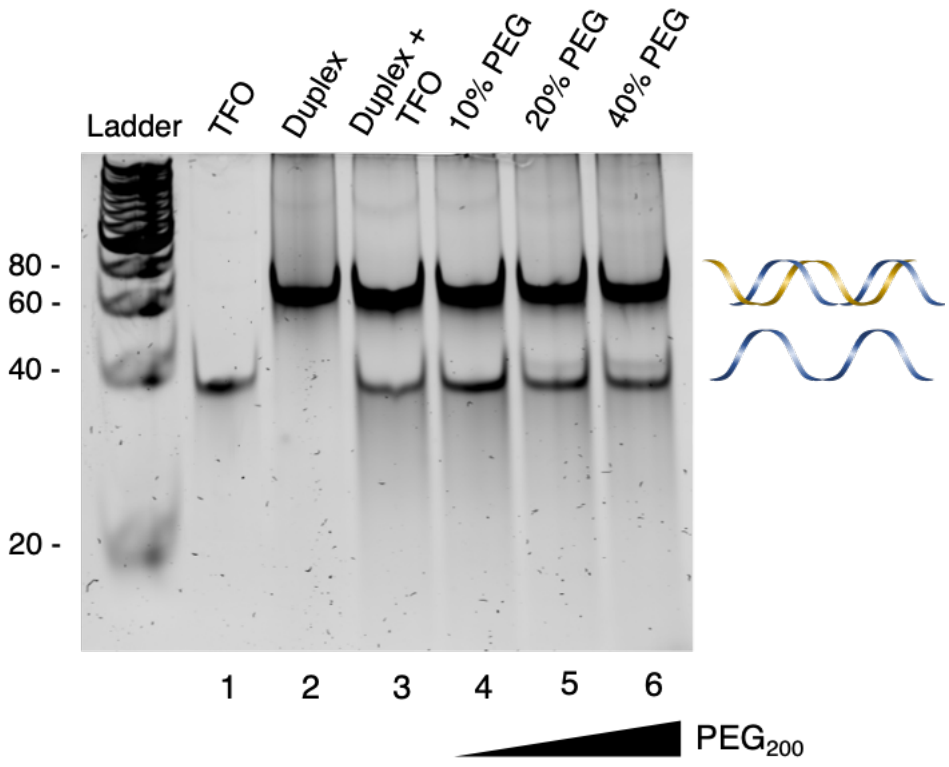


Fig. 2.8: 6% PAGE of $(GAA)_{20}$ purine sequences annealed in 10 mM NaCac buffer (pH 7.4) containing 100 mM NaCl, 30 mM $MgCl_2$ and increasing amounts of PEG_{200}. Ladder: O'RangeRuler 20 bp (Thermo Scientific). Lanes: 1) purine TFO, 2) duplex, 3) duplex with purine TFO, 4) duplex with purine TFO and 10% PEG, 5) duplex with purine TFO and 20% PEG, 4) duplex with purine TFO and 40% PEG.

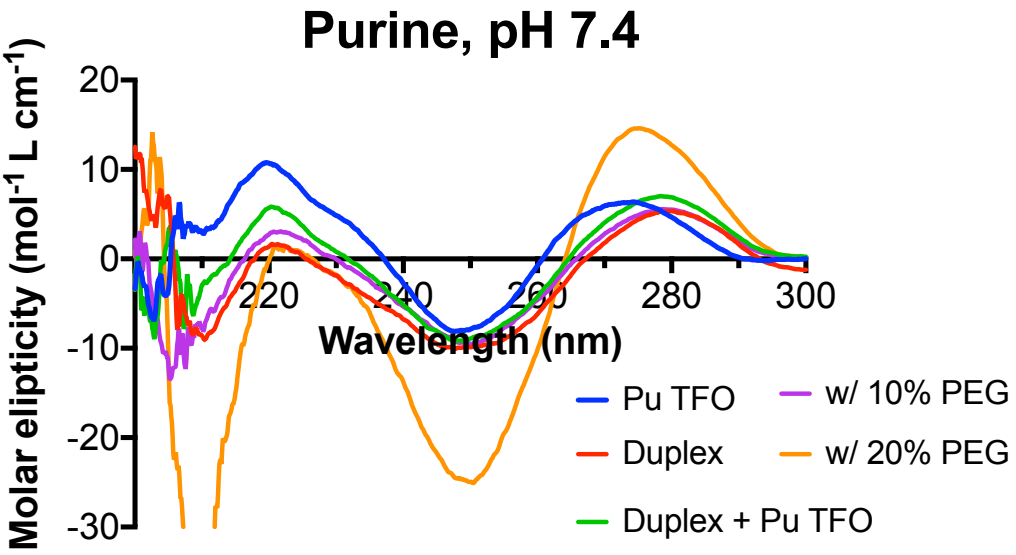


Fig. 2.9: CD spectra of $(GAA)_{10}$ purine TFO annealed with duplex with increasing amounts of PEG_{200} in 10 mM NaCac buffer at pH 7.4. The introduction of 20% PEG_{200} (v/v) appeared to worsen the distortion below 220 nm.

Oligonucleotides containing $(GAA)_n$ insert lengths of n = 5, 10 and 20 repeats were investigated. Samples were analysed with gel electrophoresis and CD spectroscopy.

Despite the introduction of molecular crowding, there was no evidence of purine triplex formation. It can be seen on the gel that only bands corresponding to the duplex and unbound purine TFO (Fig. 2.8). The CD results are less conclusive. The presence of PEG appears to worsen the instrumental distortion below 220 nm, masking the region containing the characteristic triplex band at 210 nm (Fig. 2.9). However, as seen in previous CD spectra of the formed pyrimidine triplex, the negative band begins at around 230 nm and is quite obvious; evidence of this cannot be seen in the CD spectra of the purine samples with the PEG (Fig. 2.9). The gel electrophoresis and the CD indicate that the purine triplex is not forming.

It might be expected that the molecular crowding introduced by PEG would influence the purine triplex structure to form at physiological pH by mimicking the intracellular environment in which these structures are hypothesised to form.[44] All cells have a high concentration of macromolecules, in the range of 20-40% of the total volume.[40] The increased viscosity of the buffer solution changes the kinetics and

energetics of the DNA binding.[167] The crowded environment serves to decrease the freedom of movement of the oligonucleotides and force them closer together, perhaps increasing the chance of binding events. PEG has been shown previously to stabilise triplex structures under physiological conditions in solution.[40] However, it has also been reported that the addition of cosolutes can negatively affect the thermal stability of the underlying duplex.[50] This could be beneficial for triplex formation as binding of the purine TFO in the major groove in an antiparallel orientation requires a slight conformational change of the duplex, which would be easier if this was already perturbed. However, despite the introduction of molecular crowding and the increased concentration of divalent cations, no evidence of purine triplex formation was observed at acidic or neutral pH.

2.3 Conclusions

It was found that pyrimidine triplex structures formed readily from all repeat lengths tested when annealed in buffer at pH 5.5 with a variety of concentrations of Na^+ and Mg^{2+} salts. This was to be expected as the binding of the pyrimidine TFO to the duplex requires the protonation of a nitrogen atom on cytosine, which can occur when under acidic conditions. Purine triplexes were not observed to form under any conditions. This was unexpected, as triplex formation has been previously reported in literature from these sequences with a $(GAA)_n$ insert under physiological conditions. Once formed at acidic pH, pyrimidine triplexes were seen to be somewhat resilient to increase in pH up to 7.5. Despite the addition of divalent cations and crowding conditions, purine triplexes were not observed.

Purine triplexes are known to be less stable than pyrimidine triplexes, though pyrimidine triplexes are not able to form at physiological pH as they require protonation. The pathological threshold for $(GAA)_n$ repeat number in FRDA is approximately 66, thus the sequences investigated here are far from pathological length. It is intuitive that the stability of the triplex structure is correlated with the length, hence these sequences could be too short to form stable triplexes at physiological pH. To further investigate more physiologically relevant scenarios, longer sequences of pathological length were investigated as described in the following chapter.

3 Biophysical studies of secondary structure formation from polynucleotides of pathological length

3.1 Introduction

In this chapter, the study and biophysical characterisation of secondary structures formed from long DNA and RNA strands containing the FRDA $(GAA)_n$ repeat sequence of interest under varying conditions is reported, where $n \geq 55$. Structures were studied using techniques such as gel electrophoresis and atomic force microscopy (AFM).

In order to obtain nucleic acid strands containing $(GAA)_n$ of pathological or near-pathological length, molecular biology techniques can be used. Synthesis of polynucleotides longer than 100 bp is extremely inefficient; the proportion of full-length sequences successfully synthesised decreases drastically with the increasing length. The FRDA pathological threshold for $(GAA)_n$ repeats is approximately $n = 67$, giving a total length of over 200 bp. Thus, oligonucleotide synthesis as used in the previous chapter was not a feasible strategy to produce the long sequences required.

By designing a plasmid containing the $(GAA)_n$ insert in the multiple cloning site (MCS) flanked by cloning promoters and restriction enzyme digest sites, the long $(GAA)_n$ sequence could be amplified or extracted. Initially, significant time was spent working with a plasmid cloned previously in the group, consisting of a Bluescript

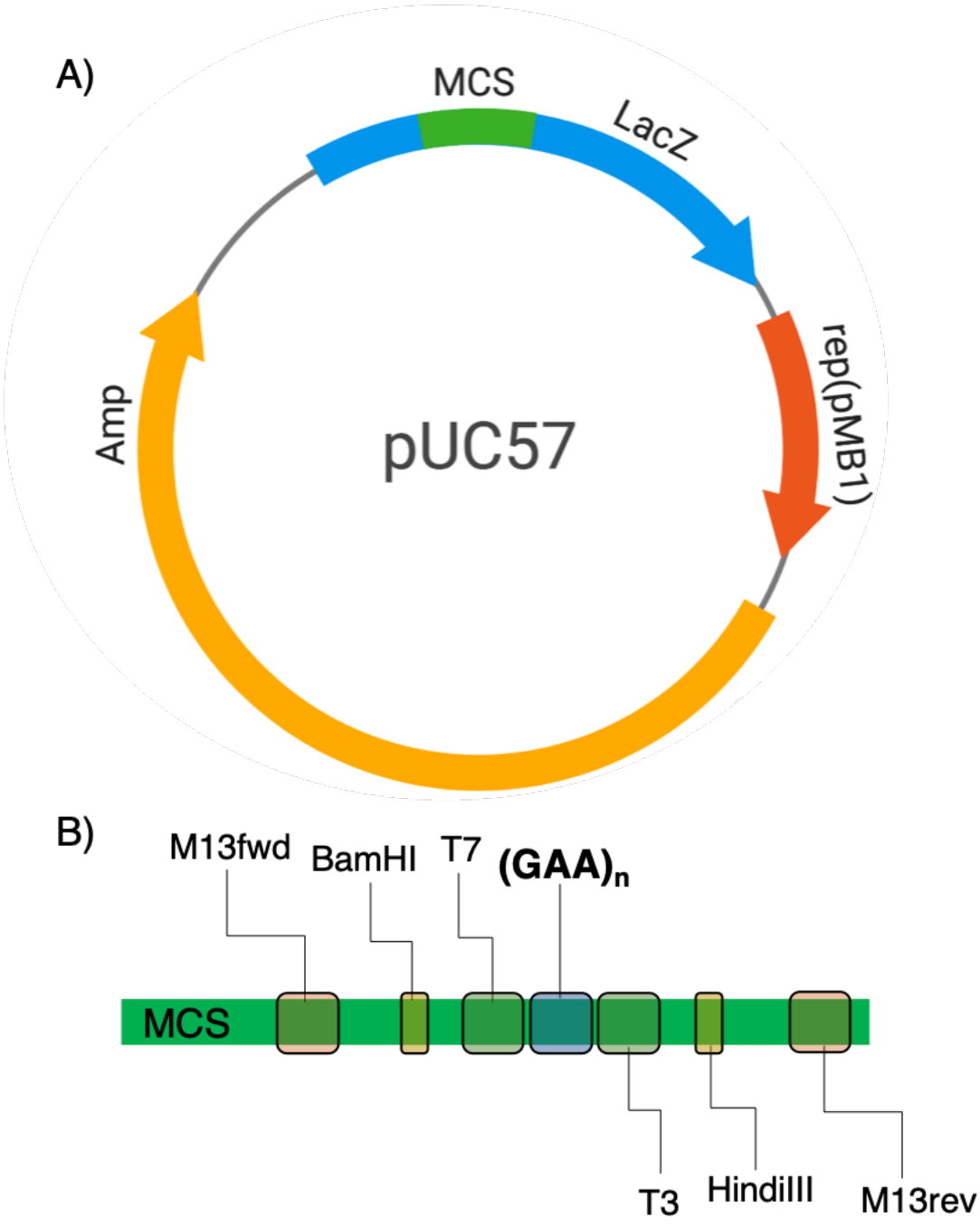


Fig. 3.1: A) Schematic of plasmid backbone (pUC57) containing the ampicillin resistance gene (Amp), the origin of replication or ORI (rep(pMB1)), and the lacZ gene containing the multiple cloning site (MCS). B) More detailed schematic of the MCS containing M13 forward and reverse sequencing primers (M13fwd and M13rev), restriction enzyme sites (BamHI and HindIII), promoter sites for polymerases (T7 and T3) and the $(GAA)_n$ insert of interest.

backbone and a $(GAA)_{200}$ insert cloned into the MCS. However, in the absence of a digital plasmid map, successful amplification of the insert was not possible due to uncertainty of promotor location, and hence this was not a viable approach.

Subsequently, two plasmids with a pUC57 backbone was designed to contain the $(GAA)_n$ insert within the MCS, where n = 55 and 200 respectively (see Fig. 3.1). The $(GAA)_n$ inserts were cloned into the MCS of the plasmid backbone and flanked by both BamHI/HindIII restriction enzyme sites and T7/T3 promoters. These plasmids were produced by GenScript® and the $(GAA)_n$ insert was checked by sequencing and restriction enzyme digest with BamHI and HindIII. The plasmids were also designed to contain an ampicillin resistance gene so they could be grown in competent bacteria in the presence of ampicillin to produce more copies of the DNA.

As an alternative to growing the plasmid and extracting the insert by digestion followed by purification, the insert could also be amplified by polymerase chain reaction (PCR), with the use of primers complementary to sections either side of the segment to be amplified. PCR is a widely used method in molecular biology to produce many copies of particular DNA segments. PCR uses the principles of enzyme-driven DNA replication and thermal cycling to produce millions of copies of the desired sequence. The first step involves denaturing the dsDNA by briefly heating to high temperature. As the temperature is then lowered, the primers added bind to their complementary DNA sequences in the target region. The two separated DNA strands act as templates for the DNA polymerase (in this case Herculase II Fusion DNA polymerase), guided by the primers, to assemble the new DNA strand from the free nucleotides added to the reaction. This newly generated strand can in turn act as a template, and thus the amount of DNA produced increases almost exponentially.

Once the long $(GAA)_n$ polynucleotides were produced, triplex formation was investigated in a similar manner as described in the previous chapter. The DNA produced from both the extraction by restriction enzyme digest and PCR amplification methods was double stranded, and hence had to first be denatured by heating and then reannealed to allow formation of higher order structures. In the case of the DNA extracted by digestion, the $(GAA)_n$ insert was flanked each side a short sequence corresponding to the T7 and T3 promoter sequences respectively. Regarding the DNA produced by PCR amplification, the insert was flanked each side by approximately 80 base pairs of additional sequence. This included the T7 and T3 promoters, the targeting sequences for BamHI and HindIII restriction enzymes, and some plasmid backbone sequence upstream of the primers used. This extra sequence can increase

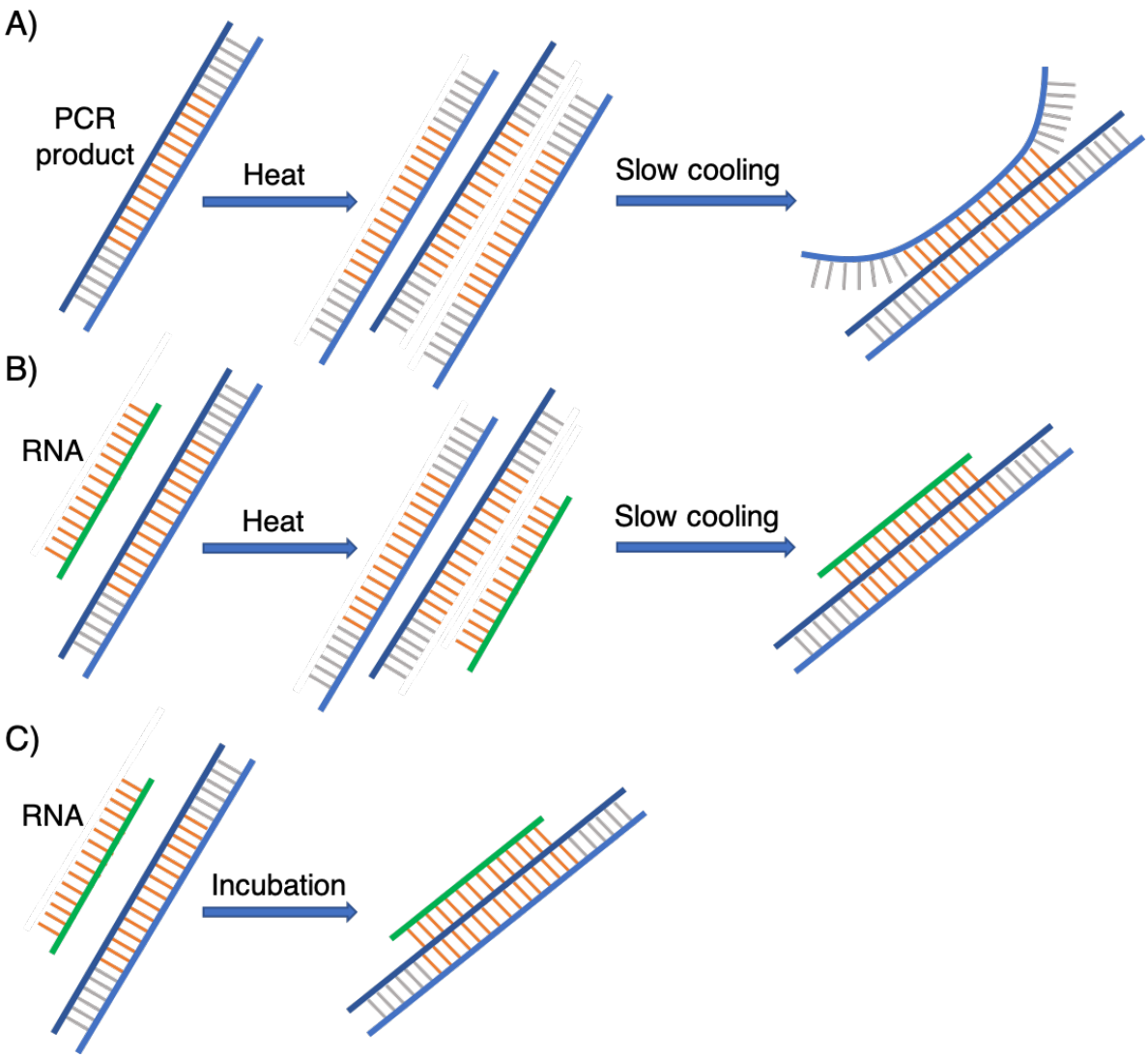


Fig. 3.2: A) Schematic showing how the PCR product is annealed to form a triplex structure. The $(GAA)_n$ repeating sequence is in orange, and the flanking bases each side in grey. The single-stranded overhangs can be seen either side of the triplex structure. B) Schematic showing how the PCR product and RNA is annealed to form a triplex structure. C) Schematic showing how the PCR product is incubated with RNA to form a triplex structure.

the stability of the duplex by providing complementary regions either side of the repeat insert. In the event of triplex formation, this extra sequence may remain as loose single-stranded overhangs (see Fig. 3.2, A).

In addition to DNA triplexes, DNA-RNA hybrid triplexes were also investigated. As discussed previously, DNA-RNA hybrid triplexes can also form at homopurine/homopyrimidine sequences.[159] Similar to the binding of a TFO, single-stranded RNA can bind in the major groove of a duplex through Hoogsteen bonding. In fact, it has been previously reported that triplexes consisting of a DNA duplex and

an RNA third strand are more stable than purely DNA-based triplexes.[168] In addition to this, DNA-RNA hybrids have been reported to play a role in FRDA.[169] One study reported that DNA-RNA hybrids (R-loops) form in the cells of FRDA patients over expanded $(GAA)_n$ repeat regions within the endogenous *FXN* gene.[12] These transcription-dependent R-loops were seen to colocalise with repressive chromatin marker H3K9me2 and impede transcription by RNA polymerase II in patient cells.[12] Repressive chromatin has also been heavily implicated in FRDA.[9] Another study has since suggested that expansion of $(GAA)_n$ repeats is caused by formation of DNA-RNA hybrids that induce break-induced replication (BIR), seen in an experimental yeast model.[18]

Following annealing, triplex formation was primarily investigated using gel electrophoresis, as for the short oligonucleotides used in the previous chapter. However, for longer polynucleotides such as these, agarose gels are normally used rather than polyacrylamide. Polyacrylamide is best used to resolve smaller fragments of DNA: gels containing a low percentage of acrylamide (3.5-5%) can be used to resolve fragments up to 1,000 base pairs in length, but gels with such low acrylamide content are extremely fragile. In addition to this, acrylamide is a potent neurotoxin. Agarose gels can be used to resolve fragments ranging from 200-30,000 base pairs in size, depending on the percentage of agarose used. As such, agarose gel electrophoresis is most commonly used when working with plasmids. The fragments produced here by PCR or from restriction enzyme digest were also assessed by gel electrophoresis with agarose gels. Secondary structures formed during annealing can be visualised directly through gel electrophoresis by staining with DNA intercalating dyes such as ethidium bromide and imaging under UV light. As before, these secondary structures will run more slowly through the gel than the corresponding duplex due to higher molecular weight.

In the previous chapter, CD was also used to detect triplex formation, seen from the characteristic negative band at 210 nm. However, CD requires relatively high DNA concentrations of 5-10 μM (at a path length of 1 cm), which was not possible to achieve using these techniques after purification. As an alternative to this, because of the increased length of the DNA, microscopy can be used to visualise the structures. In this case, atomic force microscopy (AFM) was used. AFM is a type of scanning probe

microscopy that uses a mechanical probe to 'feel' a surface. AFM has been extensively used to visualise DNA and images of DNA-minicircles containing folded G-quadruplex motifs have been previously reported with high enough resolution to visualise the helical topology.[113]

3.2 Results and discussion

3.2.1 Bacterial transformation

Plasmids containing the $(GAA)_{55}$ insert (**55r**) and the $(GAA)_{200}$ insert (**200r**) were received from GenScript® in the form of a lyophilised stock and reconstituted in ultra-pure water to give a concentration of 100 ng/μL. To produce more plasmid, **55r** and **200r** were transformed into competent bacteria and grown in large colonies before being harvested. The antibiotic resistance gene in the plasmid allows bacteria containing the plasmid to grow on a plate or in a culture containing said antibiotic (in this case ampicillin), whereas bacteria that have not successfully taken up the plasmid cannot grow since they lack the necessary resistance gene.

Initially, the DH5α strain was selected since it is the most commonly used bacterial strain for routine cloning applications. Plasmids **55r** and **200r** were transformed into the bacteria as described in Methods (6.2.6) and spread onto LB agar plates containing ampicillin (100 ng/μL) and incubated overnight at 37 •C, as per a routine transformation procedure. It was found that **55r** grew well and yielded many single colonies, however almost no colonies were seen for **200r**. It was expected that **200r** would be more difficult to clone, as the repeat number is in the FRDA pathological range and hence will naturally introduce problems with transcription. For difficult plasmids containing long stretches of repeats, the concentration of antibiotic can be decreased to allow for more transcripts of low copy number plasmids to be produced. Lowering the incubation temperature can also slow down the bacterial growth, reducing the incidence of replication error due to non-specific replication. In addition to this, DH5α cells are not specifically designed for cloning unstable sequences. Thus, SURE2 super-competent cells (Agilent) were chosen. These cells are specifically deficient in *E. Coli* genes involved in the deletion and rearrangement of DNA, and therefore can improve efficiency of cloning. Hence, **200r** was transformed into SURE2

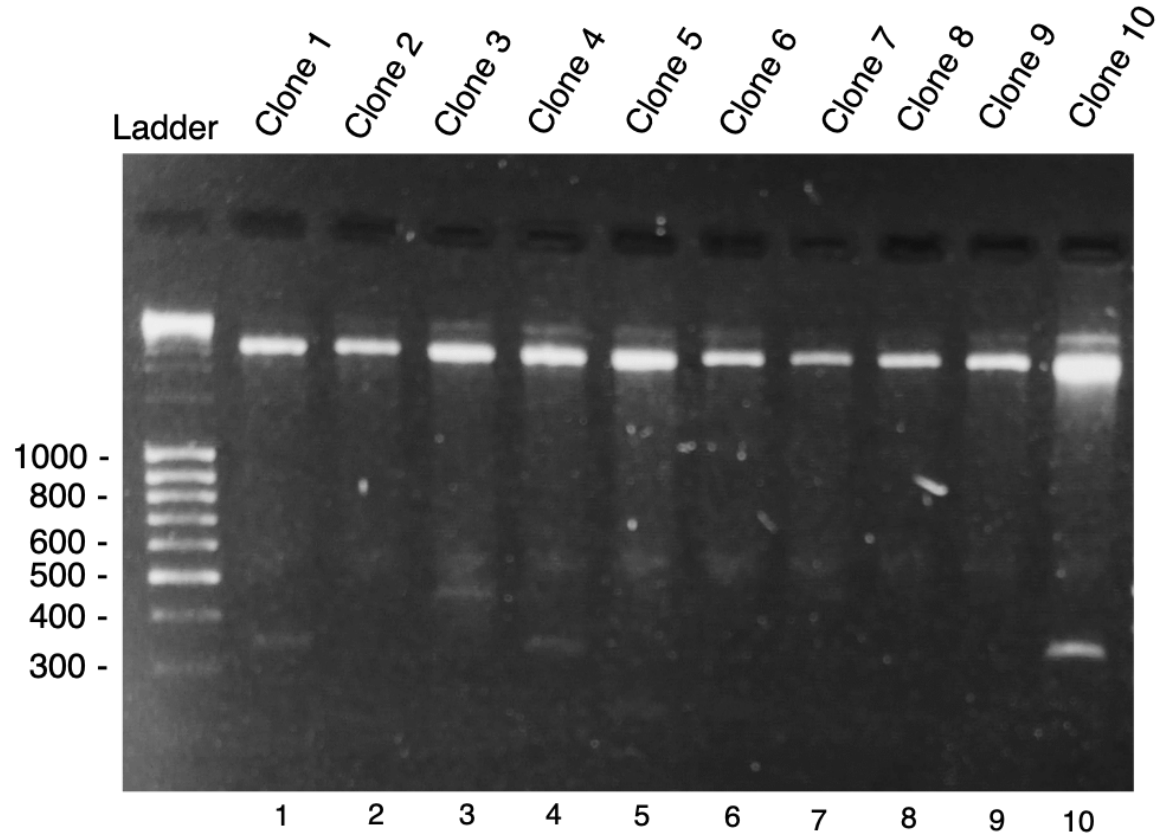


Fig. 3.3: 1.0% agarose gel showing gel electrophoresis of purified samples of clones of **200r** grown in SURE2 bacterial cells and digested with BamHI and HindIII. Digestion is incomplete as there is a prominent band corresponding to the undigested plasmid. A band corresponding to the extracted insert can be seen between 300 and 400 bp.

bacterial cells and grown on LB agar plates containing ampicillin (50 ng/μL) and incubated overnight at 30 ^{o}C. This resulted in a small number of single colonies growing successfully. From each plate containing **55r** and **200r** respectively, 10 single colonies were picked and inoculated into small starter cultures of LB broth (2-3 mL) containing ampicillin (50 ng/μL). These starter cultures were incubated at 30 $^{\cdot}$C for 6-8 hours before being expanded into larger cultures (~100 mL). These were grown at 30 $^{\cdot}$C overnight, or until the optical density (OD) indicated sufficient growth. Smaller cultures were harvested through miniprep and the plasmid DNA was extracted and purified to be checked by double digestion with BamHI and HindIII to ensure the $(GAA)_n$ insert was of the expected size, while the larger cultures were pelleted for processing later. All clones of **55r** produced bands corresponding to the expected size of around 165 bp. However, none of the clones of **200r** produced bands corresponding to the expected size of 600 bp (see Fig. 3.3). The largest insert band was consistent

with a length of 220 bp (Clones 1, 4 and 10, Fig. 3.3). The larger cultures were also processed and found to be the same. To confirm this, each clone of **200r** was sequenced using a M13 forward primer. It was found that the longest $(GAA)_n$ insert isolated from any of the **200r** clones was around 225 bp, or 75 repeat units. 50% of the sequencing runs failed and were flagged for 'homopolymeric or repetitive region'. A few of the **55r** clones were also checked by sequencing and the $(GAA)_n$ insert was found to be of the expected size.

Overall, the **55r** plasmid was grown successfully and all clones were consistent in size and purity. A repeat length of $(GAA)_{55}$ is not pathological (though it is considered 'pre-disease'), and hence transcription would not be expected to be affected.[170] However, a previous study reported transcriptional instability in uninterrupted $(GAA)_n$ repeat sequences of n = 40.[171] The **200r** plasmid, however, was much more difficult to grow, and once grown it was found that large portions of the repeat sequence had been deleted. Repeat sequences such as these are notoriously unstable and prone to recombination and deletions during cloning.[172,173] A previous study reported that long GAA repeats are unstable in bacteria and that the instability occurs at a threshold length of between 50 and 150 repeats.[174] Isolation of long fragments from the cells of FRDA patients was also reported, and it was found that although the fragments initially contained 575 repeats, after a few hours of growth in bacteria, 95% of samples had experiences deletions to bring the size down to 250 repeats, and the instability continued with further growth.[174] This supports the findings with the **55r** and **200r** plasmids shown here, suggesting that the instability threshold for these plasmids is approximately 75 repeats, as the **200r** plasmid experienced deletions to a minimum of 75 repeats.

3.2.2 PCR amplification

As growing the plasmids in bacteria was somewhat problematic and time consuming, PCR amplification was explored as an alternative to generate large amounts of DNA for triplex formation assays. A high-fidelity DNA polymerase is designed to read difficult sequences such as repeat regions, so has the potential to produce a full length $(GAA)_{200}$ insert from the **200r** plasmid.

Initially, Platinum™ *Taq* DNA polymerase (Invitrogen™) and primers complementary to the T3 and T7 promoter regions that flank the $(GAA)_n$ insert were used. The DNA produced in the PCR reactions was then purified to remove the proteins and primers and analysed on an agarose gel. The **55r** plasmid samples showed a band corresponding to 165 bp (55 repeats), but the band was smeared suggesting some polymerase slippage. The **200r** plasmid samples showed heavy smearing with no discernible bands, suggesting heavy polymerase slippage. The control plasmid showed a band corresponding to the expected size. Since Platinum™ *Taq* is not a specifically high fidelity polymerase it has not been designed for difficult sequences such as repeat regions. In addition to this, the primers are targeting a region directly adjacent to the stretch of repeats meaning that the polymerase begins reading the strand at the site of the repeat; this could increase slippage. By using primers further upstream of the $(GAA)_n$ insert, the polymerase can securely bind before reaching the repeats.

Herculase II Fusion DNA polymerase (Agilent) is a high fidelity PCR enzyme designed to amplify difficult targets such as sequences with high GC content. Sequences with high GC content (i.e. > 60% G and C bases), have a propensity to form stable secondary structures much like repeat regions. Forward and reverse primers (M13F and M13R, same as for the sequencing) were used, as this gave a ~60 bp buffer before the repeat region. The PCR was run according to the conditions suggested by the polymerase manufacturer (and in Methods section 6.2.10) and reactions were purified (as described in Methods 6.2.11) before being checked on a gel. The **55r** samples resulted in a sharp band of the correct size with very minimal smearing, indicating an efficient PCR reaction. The **200r** samples once again resulted in a smear, despite the control sample giving a clean band of the correct size. Changing the conditions, such as the concentrations of reagents or template DNA, can improve yield and purity of the product.

The amount of template DNA for both **55r** and **200r** added to the reaction was varied between 1 and 50 ng. The reaction was run and purified as before, with the concentration and purity of DNA being measured with a Nanodrop, using the absorbance at 254 nm, and samples were then loaded onto a gel to check the size. From this, it was determined that the optimum amount of template DNA for **55r** was

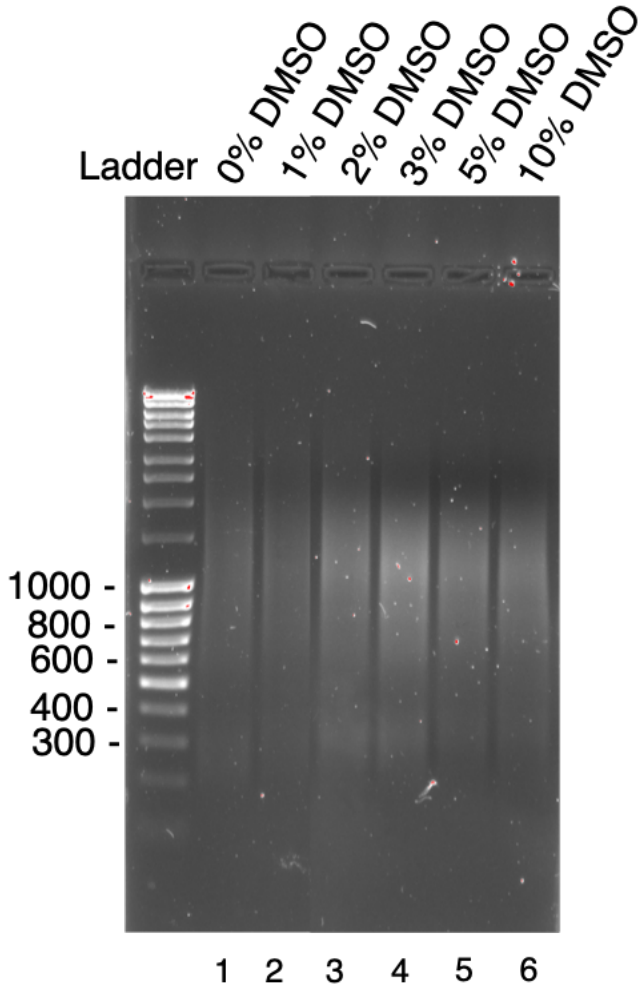


Fig. 3.4: 1.0% agarose gel showing gel electrophoresis of purified PCR products from amplification of **200r** with increasing amounts of DMSO (0-10%). Smearing is seen in each sample indicating polymerase slippage along the repeat sequence.

around 2 ng. For the **200r** plasmid, the concentration of the DNA obtained from all reactions was reasonable (approximately 100 ng/μL), however there were no discernible bands on the gel, with a large smear seen for each, indicating polymerase slippage.

DMSO is a common additive to PCR reactions as it serves to interfere with the self-complementarity of the DNA template and primers and thus disrupt any competing structures such as hairpins or, in this case, triplexes or 'sticky DNA'.[175] Magnesium chloride ($MgCl_2$) is crucial for the activity of polymerases, as it acts as a cofactor and catalyst for PCR. If the concentration is too low, the enzyme will be inactive and there will be no product, however if the concentration is too high, the activity will become less specific and the PCR product may contain mismatches. A titration with each of

these reagents was performed to determine the optimum concentration. DMSO concentration was varied between 1-10%, while $MgCl_2$ concentration was varied between 2-10 mM (the Herculase II reaction buffer contains $MgCl_2$ to a final reaction concentration of 2 mM).

For **55r**, varying the concentration of DMSO and $MgCl_2$ in the reaction made little difference, as the PCR reaction already produced the correct product with good yield and purity. For **200r**, significant smearing was still seen on the gel (see Fig. 3.4). This indicates that the issue is not related to formation of secondary structures in the primers, or inactivity of the enzyme, but the occurrence of polymerase slippage along the repeat or transcriptional blockage during replication.[176]

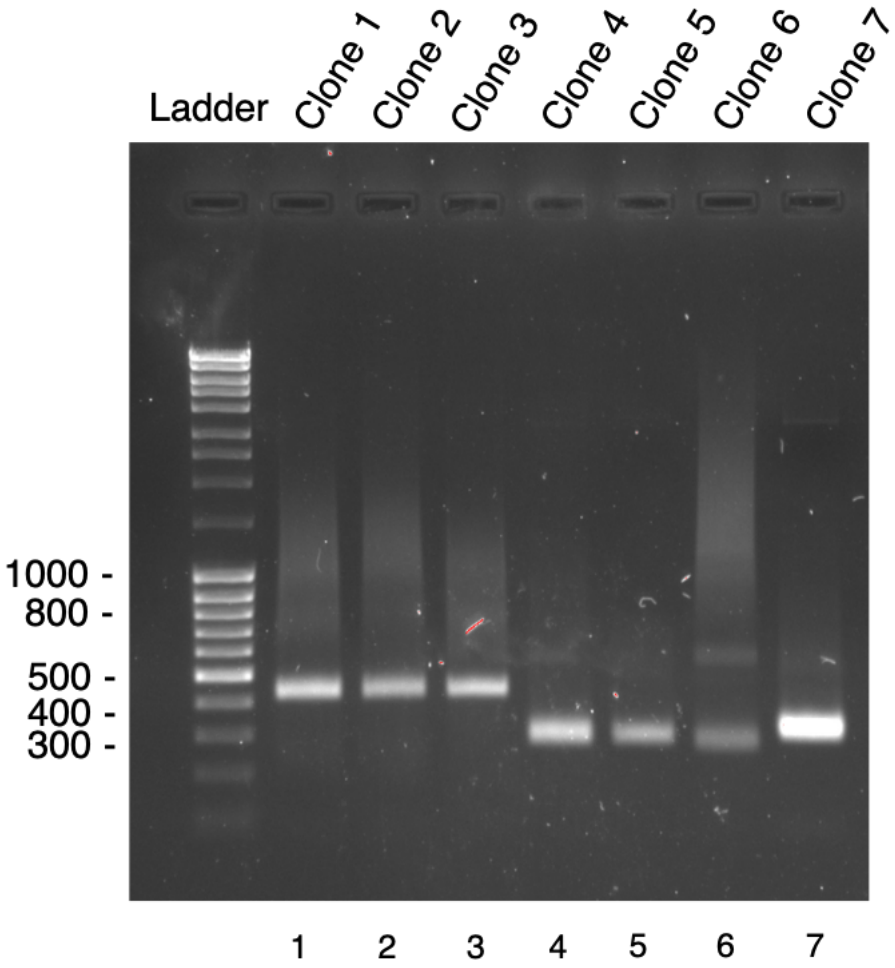


Fig. 3.5: 1.0% agarose gel showing the gel electrophoresis of purified PCR products amplified from clones of **200r** grown in SURE2 bacterial cells. From sequencing, the clones have $(GAA)_n$ inserts ranging in size from n = 55-75 (clones 1-3 with 75 repeats, clones 4-6 with ~55 repeats.

The clones containing 75 repeats (**75r**) that were grown in bacteria as described in the previous section were also amplified by PCR. The 7 clones that were successfully sequenced (with a $(GAA)_n$ repeat number ranging from n = 55-75) were amplified using Herculase II Fusion polymerase as per conditions suggested by the manufacturer (Methods 6.2.10). The PCR reactions were then purified, and the DNA was checked on a gel. Concentrations of DNA produced were reasonable (approximately 300 ng/μL) as determined by Nanodrop. It can be seen on the gel that the larger repeat lengths of 75 bp (lanes 1-3, Fig. 3.5) give a clear band of the correct size yet exhibit some smearing, which could be due to polymerase slippage resulting in PCR products of varying lengths. The clones with around 55 repeats (lanes 4-7, Fig. 3.5) give a clear band of the expected size; the clone in lane 6 also shows a second heavier band and some smearing, indicating non-specific amplification. This heavier band is close to 600 bp in length, suggesting that clone 6 could contain a small proportion of longer $(GAA)_n$ insert which has been transcribed during PCR. The significant smearing centred around 1000 bp is consistent with the previous attempts to amplify **200r** (Fig. 3.4).

PCR of repetitive sequences can be difficult. A previous study investigating the stability of $(GAA)_n$ repeats reported that a replication block prevented transcription by polymerases in an *in vitro* replication assay performed on longer repeat lengths of n = 250.[174] However, efficient replication of shorter repeat lengths of 62 was also reported, suggesting that repeat regions of pathological length caused blocking of transcription.[174] Another study into the phenomena of replication slippage at repeat sequences reported that polymerases pause at repeat sequences and can dissociate from the strand, causing premature termination of transcription or misalignment of the sequence if the polymerase re-associates at the wrong site when replication resumes.[176]

3.2.3 DNA-DNA triplex studies

Using the DNA produced by PCR, the formation of secondary structures under various conditions was investigated as for the short oligonucleotides described in the

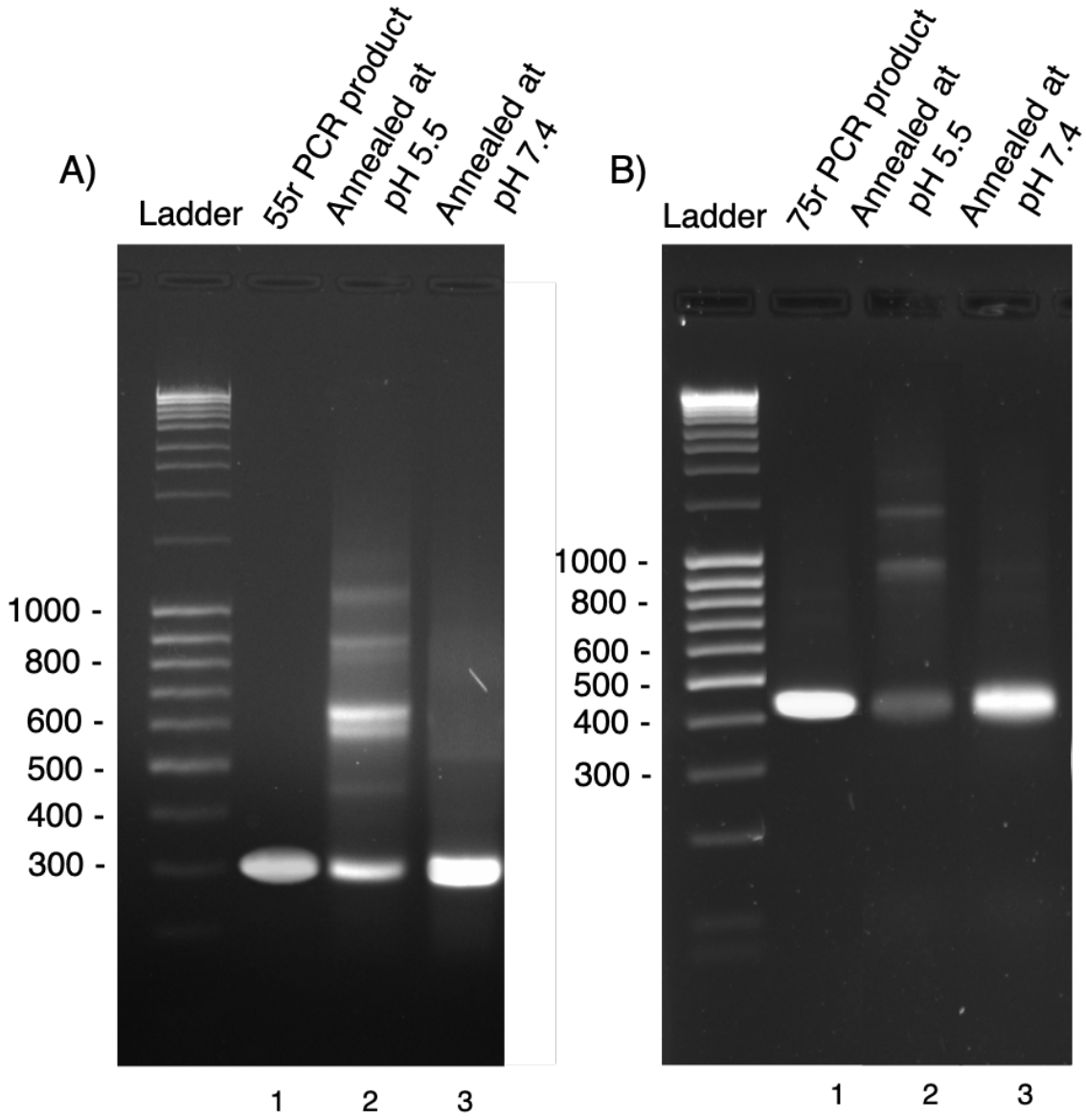


Fig. 3.6: 1.0% agarose gel showing gel electrophoresis of purified samples of A) **55r** and B) **75r** PCR products annealed at pH 5.5 and pH 7.4.

previous chapter. The effect of pH, salt concentration and crowding conditions was explored.

To study the effect of pH on the formation of secondary structures, two conditions were investigated, as before: acidic (pH 5.5) and physiological (pH 7.4). The DNA used was purified PCR product from **55r** and **75r** clones, previously checked for purity on a gel. Samples were annealed as described in Methods section (6.2.14). This differed from the annealing in the previous chapter for the short oligonucleotides as no third strand was added as a TFO; the double-stranded DNA was simply denatured by heating and re-annealed by cooling slowly to room temperature

overnight. Initially, the annealed sample was then loaded directly onto the gel to assess triplex formation using gel electrophoresis. However, the salt in the annealing buffers consistently caused distortion of the bands to the extent that it was difficult to judge the size against the ladder. Hence the samples were purified by ethanol precipitation after annealing, and the pellet was resuspended in buffer before being loaded onto the gel.

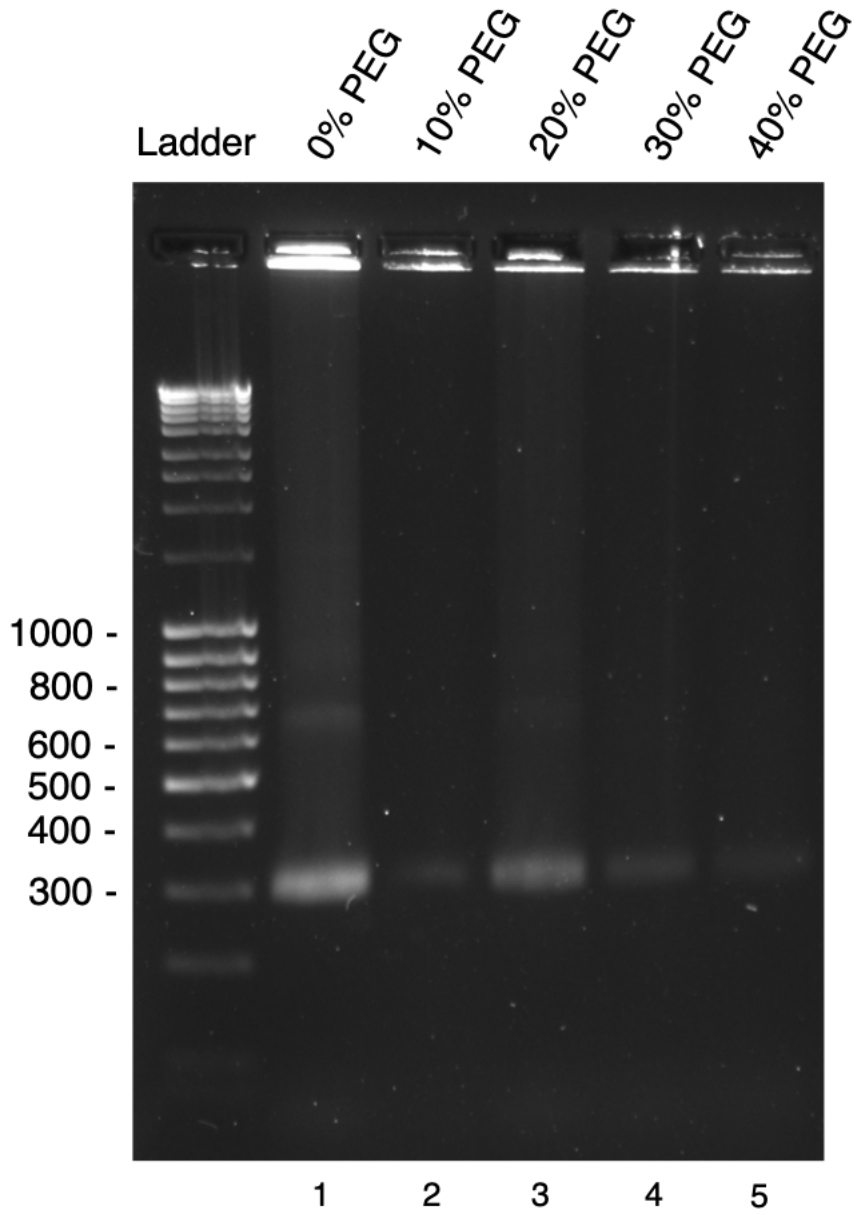


Fig. 3.7: 1.0% agarose gel showing gel electrophoresis of purified samples of **55r** annealed at pH 7.4 with increasing amounts of PEG_{200}.

The **55r** DNA showed evidence multiple secondary structures when annealed at pH 5.5, which can be seen on the gel as heavier bands above the band corresponding to the dsDNA (lane 2, Fig. 3.6, A). When annealed at pH 7.4, only bands corresponding to dsDNA were seen (lane 3, Fig. 3.6, A). A similar result is seen for the **75r** DNA annealed at each pH (Fig. 3.6, B).

The effect of salt concentration in annealing buffers was also investigated, with the concentration of NaCl and $MgCl_2$ being varied as described in the previous chapter. The concentrations used can be seen in Table 2.1 in the previous chapter. As previously described, salt cations have a significant stabilising effect on higher order nucleic acid structures. The positive charge acts to shield the negatively charged phosphate backbone of the DNA or RNA, minimising charge repulsion and allowing strands to come together. Despite altering the salts in the buffer, no evidence of triplex formation was seen at pH 7.4 from either **55r** or **75r** samples.

To mimic conditions in the cell and encourage secondary structures to form, crowding conditions were introduced using PEG_{200}. Samples of **55r** were annealed at pH 7.4 in buffer containing PEG, with the percentage volume (v/v) of PEG being varied from 0-40%. After annealing, samples were purified by ethanol precipitation, resuspended in buffer and immediately loaded onto an agarose gel. However, there was no evidence of triplex formation seen by gel electrophoresis (see Fig. 3.7), with the only visible band corresponding to the weight of the dsDNA (300 bp).

The results of the gel electrophoresis studies of **55r** and **75r** presented in this section indicate that triplexes and other heavier secondary structures form from these sequences at pH 5.5. This was consistent with findings in the Chapter 2, as pyrimidine triplexes were seen to form readily at pH 5.5 from the short oligonucleotides investigated. A study by Bergquist *et. al.* investigating intramolecular triplex formation in plasmids containing $(GAA)_n$ repeats suggested that a pyrimidine triplex structure formed under the influence of negative supercoiling at physiological pH.[141] An increase in the extent of intramolecular triplex formation was reported upon the addition of a $(TTC)_n$ single-stranded oligonucleotide, as this oligonucleotide would bind to the loose $(GAA)_n$ single-stranded region produced by the formation of a pyrimidine triplex and thus stabilise the structure.[141] The same increase was not reported upon the addition

of a $(GAA)_n$ oligonucleotide.[141] Here, it is not possible to determine whether the third strand involved in the higher order structure is the purine or the pyrimidine strand. There are much heavier bands (>1000 bp) present in the samples annealed at pH 5.5, indicating aggregation of several nucleotide strands, or the bands are significantly retarded due to the formation of a compact higher order structure. $(GAA)_n$ tracts have been previously reported to adopt a structure called 'sticky DNA'. A study by Sakamoto *et. al.* suggests that intramolecular purine triplexes formed in plasmids containing long $(GAA)_n$ tracts under the influence of negative supercoiling can also form stable dimers ('sticky DNA'), seen by reduced gel mobility and electron microscopy.[16] This is proposed to happen through a strand exchange mechanism. In the study, 'sticky DNA' was reported to form at repeat lengths of 59 and above.[16] A later study also showed that low pH destabilised the 'sticky DNA' conformation.[11] The same study proposes that 'sticky DNA' can as well take the form of a single long triplex, formed from the alignment of two $(GAA)_n$ tracts in living *E. Coli* cells.[11] The retarded bands seen in the gels of **55r** and **75r** could correspond to an association of two or more triplex structures. However, as gel electrophoresis cannot provide structural information, it is not possible to predict the conformation from these analyses alone.

The results of the gel electrophoresis studies of **55r** and **75r** annealed at pH 7.4 showed no evidence of secondary structures. This is contradictory to the study discussed previously in which intramolecular triplexes formed at pH 7.5.[15] However, it appears that supercoiling plays an essential role in promoting the formation of these structures in plasmids. In the same study, the plasmid had to be linearised to eliminate the intramolecular triplex structure in order to study the interaction between the duplex and a PNA, suggesting that triplex was not present in linear DNA.[15] Perhaps there is an intramolecular triplex present in the **55r** and **75r** plasmids induced by negative supercoiling, which is then absent when the plasmid is digested or amplified with PCR as the strand becomes linear. There are no studies previously reporting the investigation of intermolecular triplexes containing $(GAA)_n$ tracts of pathological length in this linear form; however, following on from the investigation of intermolecular triplex formation from short oligonucleotides in the previous chapter, this result is consistent.

Varying the salt concentration of the buffers did not encourage secondary structure formation at pH 7.4. This is also consistent with the short oligonucleotide

study in the previous chapter, however it was suggested that oligonucleotides of pathological length might have formed a triplex at neutral pH, as was hypothesised to happen in FRDA. As discussed previously, 'sticky DNA' has been reported to form at neutral pH from the association of two purine triplexes within negatively supercoiled plasmids.[11] It was also reported that divalent metal cations (such as Mg^{2+}) served to stabilise this structure, and the 'sticky DNA' was destabilised by the addition of EDTA which would chelate the cations, therefore nullifying the stabilising effect.[11] This was presented as further evidence that the triplexes involved in the structure were purine triplexes, as divalent cations are essential for purine triplex formation.[166] Again, the lack of secondary structures seen at pH 7.4 could be due to the lack of supercoiling in this system. Although, in plasmids the supercoiling acts to encourage the duplex to disrupt itself and re-form in a H-DNA structure, whereas in this linear model the duplex is similarly disrupted by heating before being slowly annealed. Supercoiling is known to be present during DNA replication *in vivo*; as the two strands of the duplex are separated, the replication fork is created, and movement of the replication machinery through the immobilised DNA template strand causes massive positive supercoiling ahead of the fork.[177]

Introducing crowding conditions through use of PEG_{200} also did not encourage secondary structures to form at neutral pH. As discussed before, crowding conditions serve to emulate the high concentrations of macromolecules found inside the nucleus. By simulating the environment found *in vivo*, it might be expected that this would trigger secondary structures to form as hypothesised in cells of FRDA sufferers.

Several higher order structures were seen to form when **55r** and **75r** were annealed at under acidic conditions. It is not possible to decipher the conformations of these structures from gel electrophoresis alone, hence microscopy can be employed to visualise them as described below.

3.2.4 DNA-RNA triplex studies

DNA-RNA hybrid structures have also been implicated in FRDA, with R-loops being reported to form in FRDA patient cells.[12] Therefore, it was of interest to synthesise RNA sequences with the potential to assemble into triplex structures with dsDNA. Thus, RNA was synthesised directly from the **55r** PCR product using the T3

and T7 promoters either side of the $(GAA)_{55}$ repeat region. The T7 polymerase reads the sense $(GAA)_n$ strand, whereas the T3 polymerase reads the complementary antisense $(TTC)_n$ strand. Hence, T7 polymerase produces pyrimidine (CUU) RNA and T3 polymerase produces purine (AAG) RNA. The size and purity of the RNA was assessed using a Bioanalyser. The **55r** double stranded PCR product was incubated with either the purine or the pyrimidine RNA strand in three different ways: annealed by heating to 95 °C and slowly cooling, incubating at 37 °C overnight and incubating at 4 °C for 48 hours. The effect of pH was also investigated, with acidic (pH 5.5) and neutral (pH 7.4) conditions being tested. Formation of triplex structures was assessed

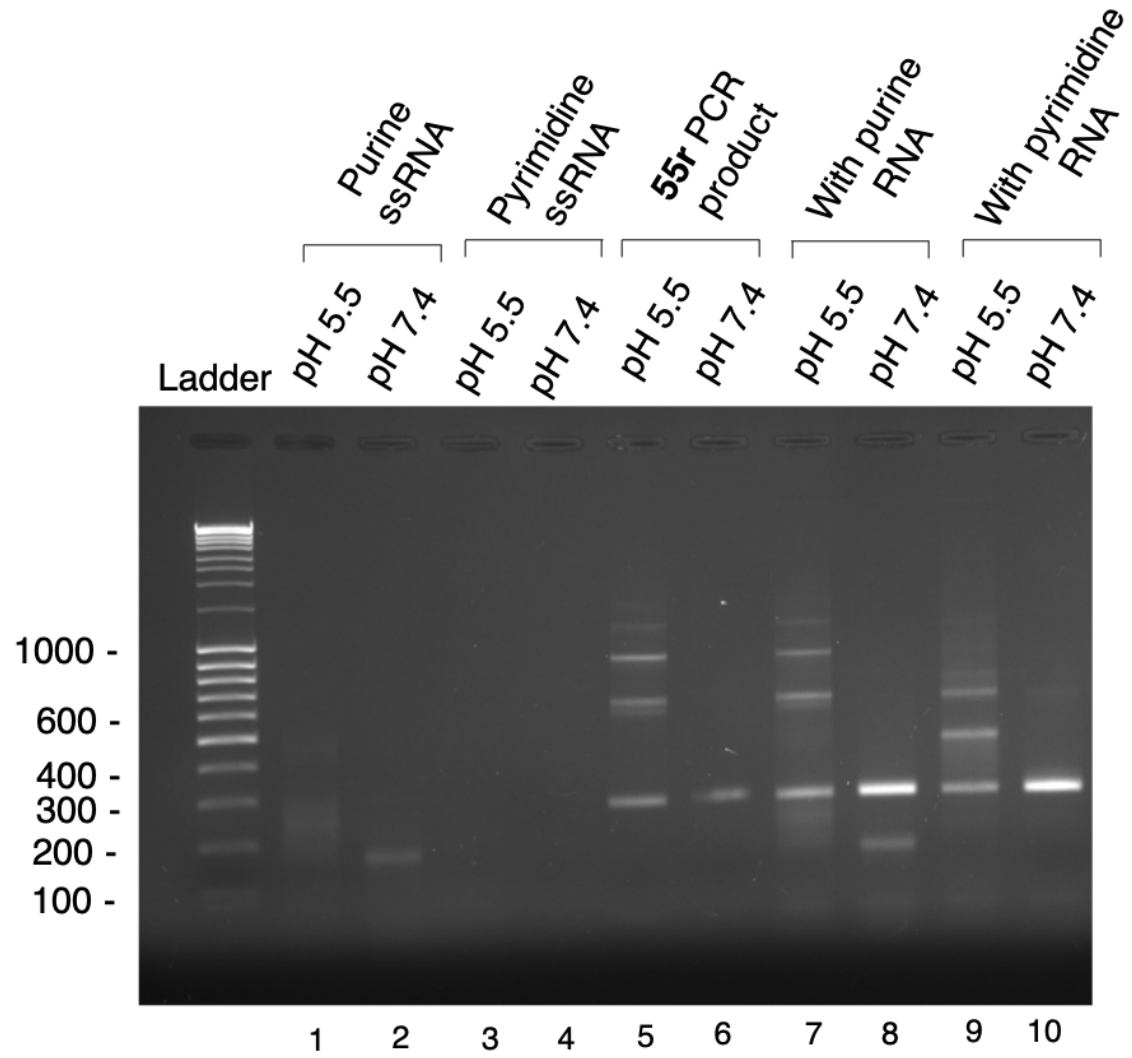


Fig. 3.8: 1.0% agarose gel showing gel electrophoresis of purified samples of purine and pyrimidine RNA TFOs annealed with **55r** PCR product at pH 5.5 and pH 7.4.

using gel electrophoresis. After incubation in annealing buffer, the samples were purified by ethanol precipitation and resuspended in buffer, before being loaded onto a gel.

First, samples were annealed as done previously for the purely DNA triplexes (Methods 6.2.14) (Fig. 3.2, B). Purine or pyrimidine RNA TFOs were added to **55r** PCR product in a large excess (approximately 10-fold) in annealing buffer at either pH 5.5 or pH 7.4 and heated briefly to 95 °C before being cooled slowly to room temperature overnight, purified, resuspended and loaded onto a gel. It can be seen from the gel that there are secondary structures present in the samples that were annealed at pH 5.5, for both purine and pyrimidine RNA TFOs, as well as just **55r** dsDNA (lanes 5, 7 and 9, Fig. 3.8). Annealing the samples by heating and subsequent cooling also results in denaturation of the duplex during heating, hence these bands could be the result of DNA-DNA interactions as discussed in the previous section. The heavier bands observed in the sample containing pyrimidine RNA TFO annealed at pH 5.5 (lane 9) appear to have a different mobility to the heavier bands seen for the sample containing purine RNA TFO annealed at pH 5.5 (lane 7), and for just the **55r** duplex annealed at pH 5.5 (lane 5). This disparity could indicate the formation of an alternative structure that involves the pyrimidine RNA. To investigate this, RNAse H was added to the samples and they were ran on a gel to determine whether the position of the bands had changed; RNAse H digests only RNA that is involved in a DNA-RNA hybrid structure. However, the concentration of enzyme was too high and also digested the DNA in the sample. This experiment could not be repeated due to time constraints.

So as not to denature the duplex during incubation, purine and pyrimidine RNA was incubated with **55r** PCR product at 37 °C for 24 hours (Fig. 3.2, C). By keeping the duplex intact, the likelihood of DNA-DNA secondary structures forming decreases as the duplex remains relatively stable at this temperature. Thus, any heavier bands observed on the gel corresponding to higher order structures can likely be attributed to the interaction of DNA and RNA. Samples were incubated in buffer at pH 5.5 or pH 7.4, purified by ethanol precipitation, resuspended and loaded onto the gel. It can be seen from the gel that there are similar sized bands above the band corresponding to **55r** dsDNA in every sample with added RNA TFO incubated at 37 °C, from purine and

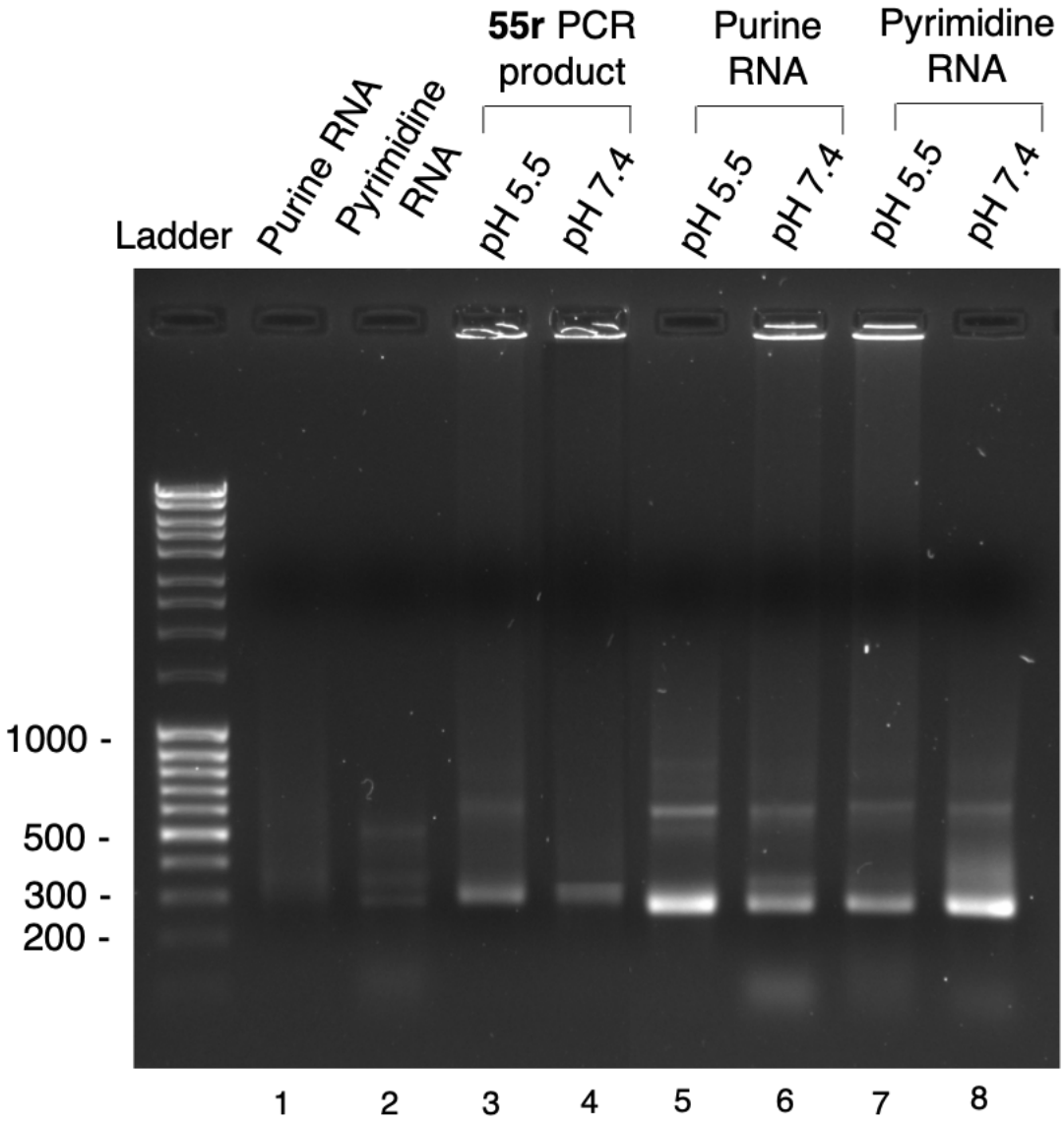


Fig. 3.9: 1.0% agarose gel showing gel electrophoresis of purified samples of purine and pyrimidine RNA incubated with **55r** PCR product at 37 °C for 24 hours in pH 5.5 and pH 7.4 buffer. Heavier bands can be seen in all samples with RNA TFO added (lanes 5-8).

pyrimidine RNA at pH 5.5 and pH 7.4 (Fig. 3.9, lanes 5, 6, 7 and 8). This is promising evidence that the RNA is associating with the DNA at acidic and neutral pH. In the samples annealed with the DNA and the purely DNA triplex samples annealed at pH 5.5, there were much heavier bands (> 1000 bp) which could be due to 'sticky DNA'; these bands are absent in the samples incubated at 37 •C. This is further indication that the heavier bands resulting from the samples incubated with RNA are due to DNA-RNA hybrids and not pure DNA structures, as the heavier 'sticky DNA' bands are absent since the duplex remains intact during the incubation.

To further prevent any denaturation of the duplex and prevent secondary DNA structures forming, samples were incubated at 4 •C (Fig. 3.2, C). Purine and pyrimidine RNA TFOs were incubated with **55r** PCR product at pH 5.5 and pH 7.4 at 4 °C for 48 hours, before being purified, resuspended and ran on a gel. A higher concentration of sample was added in order to visualise the fainter bands. It can be seen from the gel that multiple structures were seen in all samples containing RNA TFOs (see Fig. 3.10, lanes 5, 6, 7 and 8). For the purine RNA at both pH 5.5 and pH 7.4 a single heavier band can be seen above the band corresponding to the **55r** duplex, which is equivalent to approximately 700 bp (lanes 5 and 6, Fig. 3.10). This is approximately the same

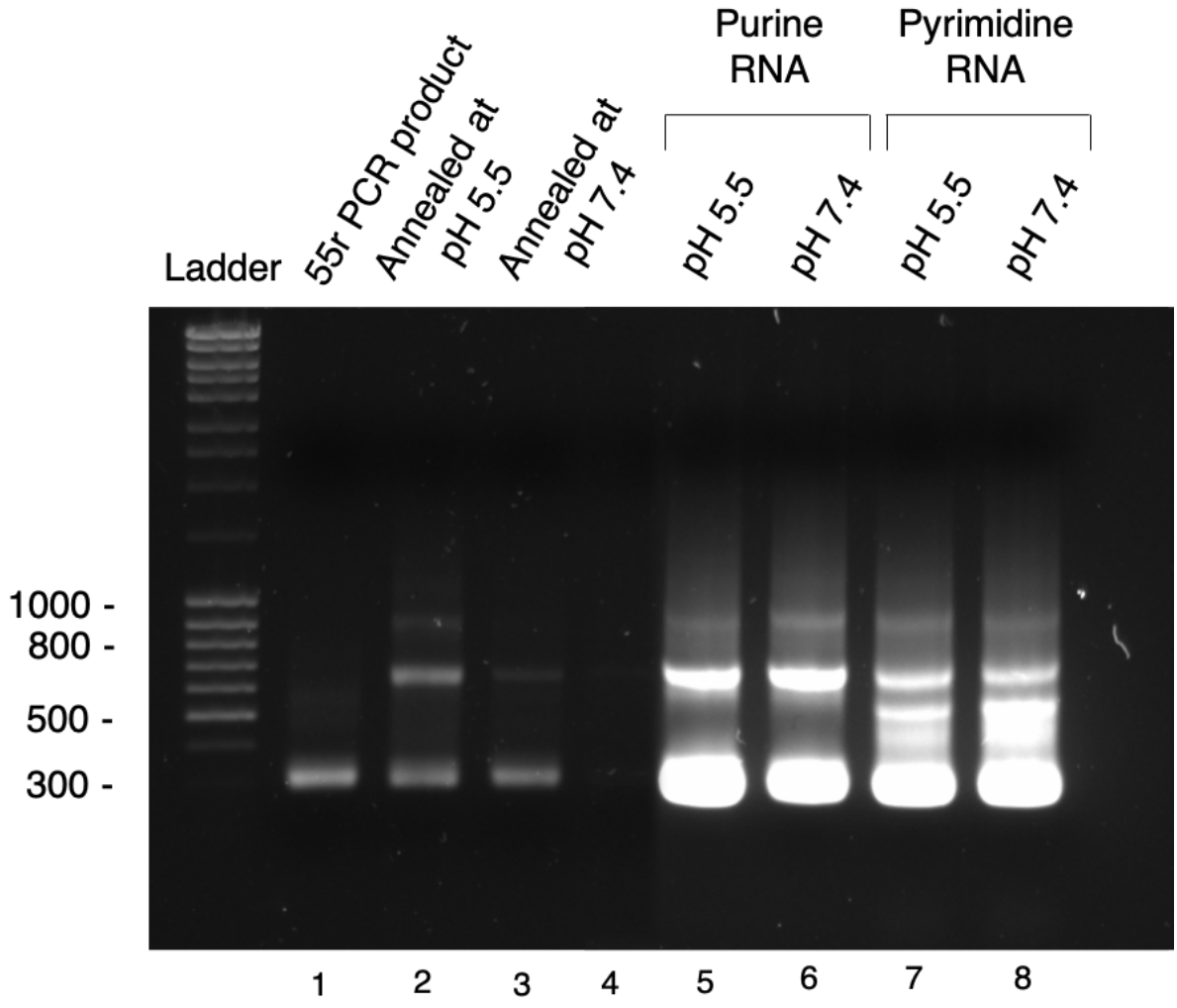


Fig. 3.10: 1.0% agarose gel showing gel electrophoresis of purified samples of purine and pyrimidine RNA incubated with **55r** PCR product at 4 °C in pH 5.5 and pH 7.4 buffer. Evidence of secondary structures can be seen in samples containing purine and pyrimidine RNA TFOs at pH 5.5 and pH 7.4 (lanes 5-6).

weight as the DNA secondary structure formed when annealed at pH 5.5. The single-stranded RNA produced is around 165-mer, as the T3 and T7 promoters are directly adjacent to the $(GAA)_{55}$ tract in the sequence, hence it would be expected that the triplex structure formed with the RNA would be around 130 bp lighter than the pure DNA triplex, which is formed from three strands of around 300-mer each. For the pyrimidine RNA at both pH 5.5 and pH 7.4, there is a band corresponding to a structure equivalent to 700 bp, as for the purine RNA samples, but there are also multiple structures between this 700 bp band and the band corresponding to the **55r** duplex (lanes 7 and 8, Fig. 3.10). This has not been observed in any of the previous DNA-RNA or DNA-DNA samples. The samples are over-loaded, so the individual bands are difficult to separate, however the effect seems to be even more prominent at pH 7.4. In previous cases, a band corresponding to the unbound ssRNA TFO can also be seen to some extent on the gel, whereas here none can be seen, hence suggesting that a large portion of the RNA TFO is bound.

The results of the gel electrophoresis studies of **55r** DNA incubated with purine and pyrimidine RNA TFOs at pH 5.5 and pH 7.4 present some interesting evidence of secondary structure formation even at neutral pH. Initially, samples were annealed in the same manner as for the DNA triplex studies, but with purine or pyrimidine RNA TFO added as a third strand. As this method involves heating the sample to 95 •C, the duplex is denatured in the process. It follows that during cooling, DNA-DNA structures can form as seen in the DNA triplex studies in the previous section, especially under acidic conditions. This is reflected in the gel electrophoresis results, as a heavier band is seen in the DNA-RNA samples annealed at pH 5.5 that is the same weight as for the pure DNA samples annealed at pH 5.5 (see Fig. 3.8). This band is not seen in samples annealed at pH 7.4. It could be suggested that these DNA structures block the association of RNA, leaving the RNA unbound in solution. This can also be seen on the gel, as there appears to be a lighter species of around 160-mer that could correspond to unbound ssRNA (lane 8, Fig. 3.8). This band is also seen in the control lanes with just the purine and pyrimidine ssRNA respectively (lanes 1 and 2, Fig. 3.8). The only difference that can be seen in the samples is that the pyrimidine RNA TFO and **55r** annealed at pH 5.5 seem to give bands with different mobility to the purine

RNA TFO and **55r** annealed at pH 5.5, and just **55r** annealed at pH 5.5 without RNA. This suggests that perhaps the pyrimidine RNA is affecting the secondary structures formed. In fact, previous studies investigating incorporation of RNA into DNA triplexes reported that pyrimidine RNA can form very stable intermolecular triplexes with DNA, whereas purine RNA cannot form intermolecular triplexes at all.[159,178] Samples were then incubated at 37 •C rather than annealed by denaturation. This resulted in heavier bands being seen at both acidic and neutral pH, with purine and pyrimidine RNA (see Fig. 3.9). It could be suggested that since the **55r** duplex was not denatured, that no additional DNA secondary structures would have formed, and hence the heavier bands can be attributed to RNA associating with DNA. The retarded bands are very similar in weight to those seen in the DNA samples annealed at pH 5.5 (lane 5, Fig 3.8). Although, the RNA is around 130-mer shorter than the DNA and thus the resulting associated hybrid structure would be lighter than the DNA triplex, which is not represented on the gel. This is difficult to confirm, as the separation between the bands on the ladder is small. When the samples were incubated at 4 •C for 48 hours, evidence for a significant amount of secondary structure formation was seen across all DNA-RNA samples at both acidic and neutral pH (see Fig. 3.10). Samples with pyrimidine RNA TFOs appeared to show multiple different structures, with the effect being more pronounced at neutral pH. The increased incubation time could have allowed the system to reach thermal equilibrium and allow for more stable structures to form.

RNA-DNA hybrids have been previously investigated in relation to triplet repeat diseases and causes of transcriptional blockage.[179–181] R-loops have been implicated in FRDA as a mechanism for the blocking of transcription of the *FXN* gene.[12,182] R-loops involve a disrupted duplex, with complementary RNA bound to one of the DNA strands from the duplex, leaving the other DNA strand unbound. This has primarily been demonstrated in plasmids, and as with the intramolecular triplexes discussed in the previous section, requires supercoiling to form.[183] This has been shown to depend on which DNA strand, homopurine or homopyrimidine, is the template strand. It was reported that R-loop formation occurs when the RNA contains the $(GAA)_n$ tract, and this is bound to the $(TTC)_n$ strand of the duplex, indicating that the purine DNA strand is involved in the intramolecular DNA triplex. A recent study reports that inactivation of

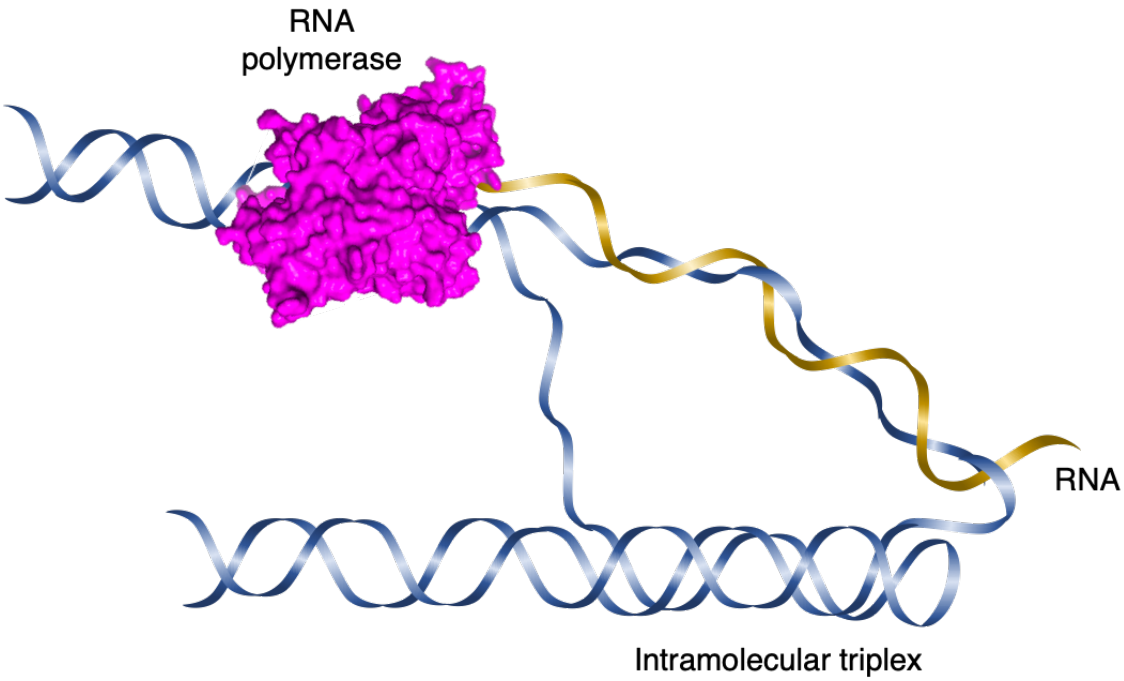


Fig. 3.11: Schematic showing the formation of a 'H-loop' during replication. The ssRNA stabilizes the ssDNA created by the formation of an intramolecular DNA-DNA triplex structure during transcription.[18]

RNAse H stimulates massive $(GAA)_n$ repeat instability in yeast, implying that DNA-RNA hybrids play a role in the expansion process.[18] This study also hypothesises the formation of a more complex DNA-RNA hybrid structure that is independent of a canonical R-loop.[18] One proposed species for this structure is a 'H-loop': an intramolecular triplex as described previously and a loose, unbound DNA strand, which is then stabilised by an RNA strand in an R-loop-type structure (see Fig. 3.11).[18]

More RNA digestion experiments using RNAse A+H would be beneficial to help confirm whether the RNA is participating in the secondary structure. RNAse H would digest any RNA that is participating in a DNA-RNA hybrid structure, while RNAse A would digest any unbound single-stranded RNA in the sample. Resulting change in band mobility in gel electrophoresis could be used to infer whether the heavier bands correspond to DNA-RNA or DNA-DNA secondary structures.

3.2.5 Microscopy of DNA-DNA triplexes

All AFM studies were performed by Philip Haynes, with supervision from Dr Alice Pyne (University College London).

AFM is a powerful technique for the visualisation of biomacromolecules and has been previously used for the imaging of higher order nucleic acid structures, including triplex and G-quadruplex.[113,184] AFM can provide structural information that

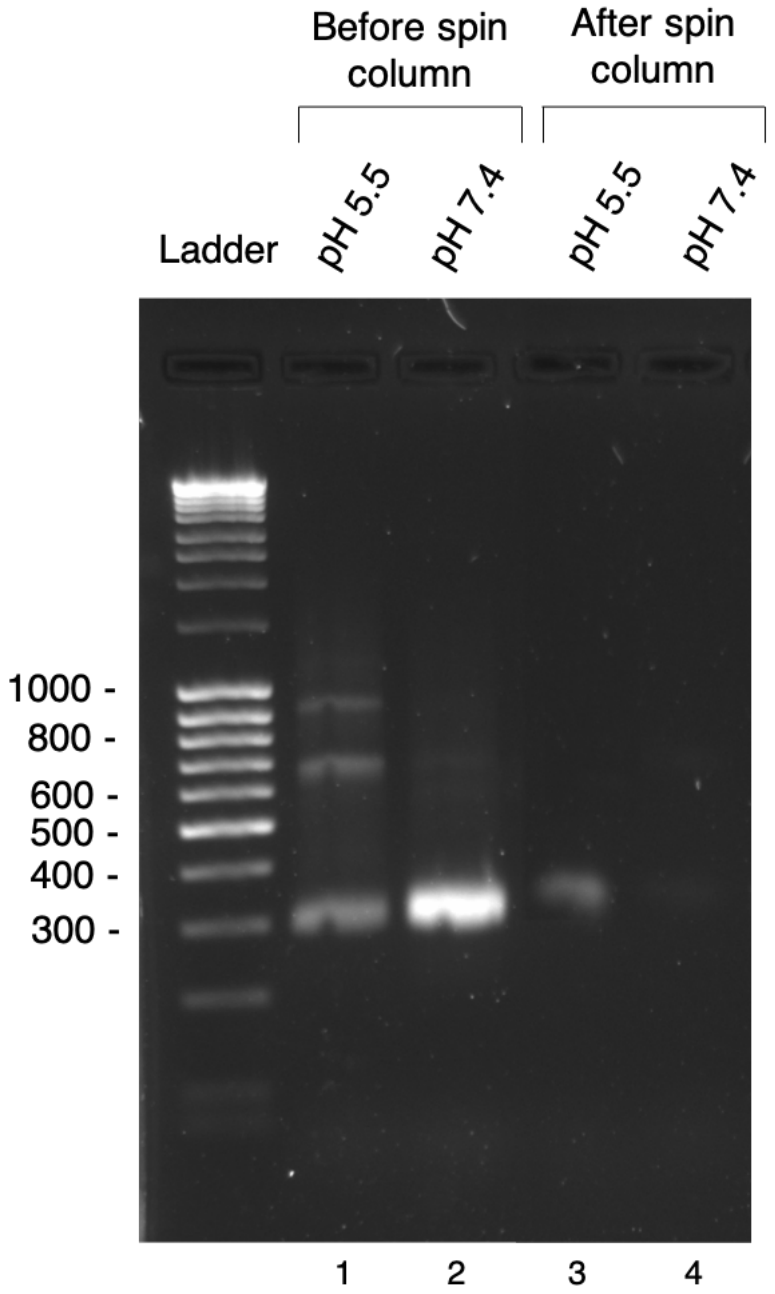


Fig. 3.12: 1.0% agarose gel showing gel electrophoresis of purified samples of **55r** annealed at pH 5.5 and 7.4 before after gel purification with spin column. The samples annealed at pH 5.5 exhibit heavier bands corresponding to secondary structures (lane 1) whereas the samples annealed at pH 7.4 do not (lane 2). After purification using spin columns, the heavier bands appear to have been abolished.

other techniques used cannot, and thus can be used to shed light on the nature of secondary structures indicated by gel electrophoresis.

Samples were annealed at pH 5.5 and pH 7.4 as before, with the double stranded **55r** PCR product being heated briefly to 95 •C and cooled slowly to room temperature overnight. This was then purified twice by ethanol precipitation to remove the salts from the annealing buffer, resulting in a DNA pellet. This was resuspended in buffer and checked on a gel. It was seen from the gel that there are multiple structures present in the sample annealed at pH 5.5, but only a band corresponding to the **55r** duplex present in the sample annealed at pH 7.4, consistent with the previous experiments. The bands in the pH 5.5 sample range in size from 300 bp to the equivalent of 1000 bp, indicating the presence of the duplex, a species that could be a triplex and a species that is heavier still, which could be 'sticky DNA'. Ideally, the band corresponding to each structure should be excised from the gel and purified, giving a pure sample of each structure to be compared using AFM. However, it was found that the gel purification step, more specifically the spin columns used to bind the DNA, resulted in fragmentation and disassembly of secondary structures. This was first seen with AFM and subsequently confirmed by running samples on a gel before and after spin column purification (see Fig. 3.12).

Following this, the gel purification step was bypassed to avoid disruption of secondary structures formed during annealing. This meant that each sample contained multiple structures and could be compared directly within each image. The samples were ultimately prepared and imaged according to Methods (6.2.15). From the images of **55r** annealed at pH 5.5, there is an indication of another structure present alongside dsDNA (see Fig. 3.13, particularly C and D). This structure seems to be more rigid than the duplex and occurs in the central section of the DNA strand, where the $(GAA)_{55}$ repeat region is located. Based on the DNA annealed in the sample, the triplex structure should theoretically consist of a duplex of approximately 300 bp in length with a third strand bound to the duplex in the triplex-forming region only; this being the central region containing the repeats of 165 bp in length, with loose, single-stranded ends either side. The potential triplex structures seen in the AFM images obtained from the **55r** samples annealed at pH 5.5 mostly support this theory, though the single-stranded unbound ends cannot clearly be seen. This structure is not

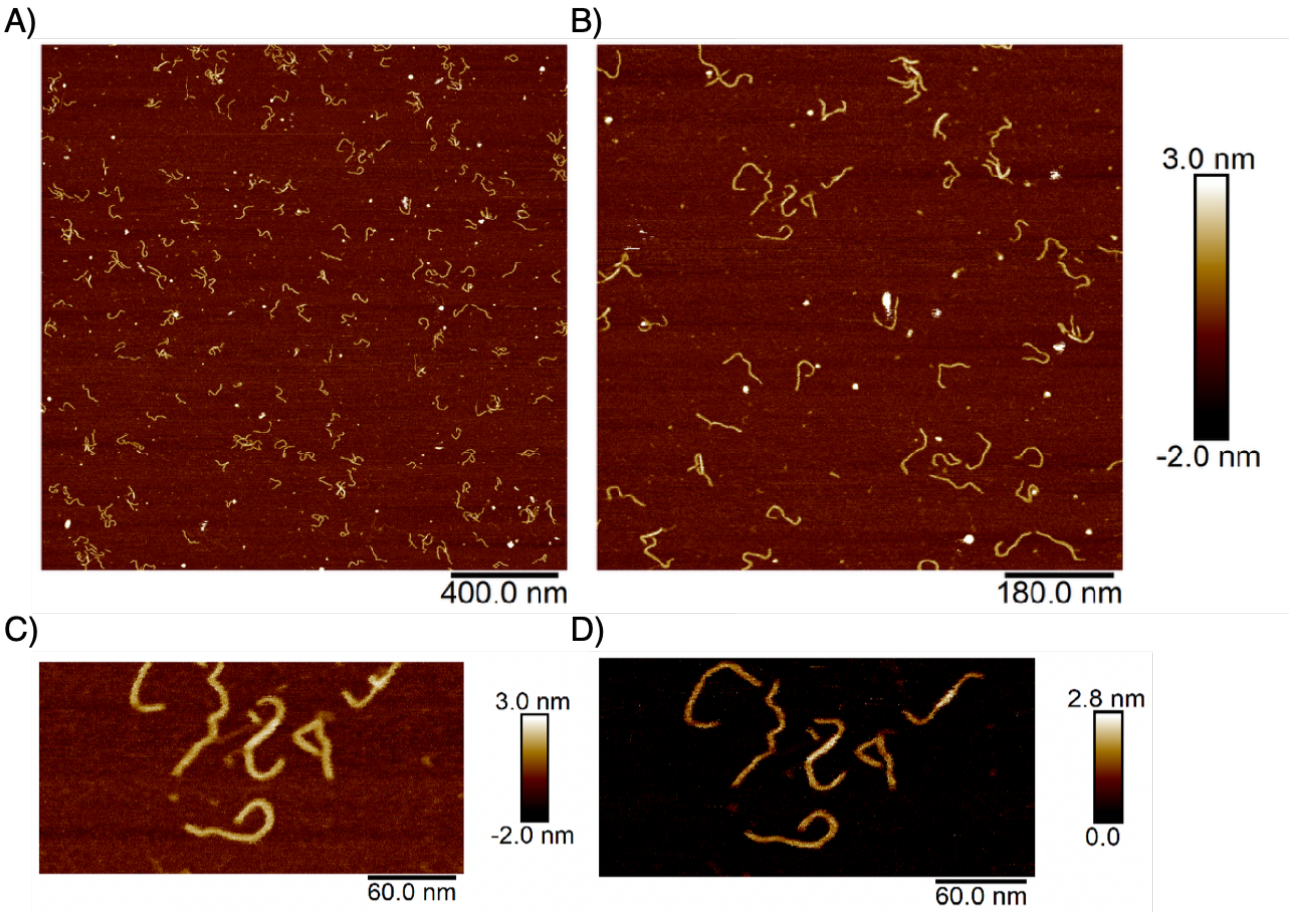


Fig. 3.13: AFM images of **55r** PCR product annealed at pH 5.5 and purified by ethanol precipitation. Apparent triplex structures can be seen throughout the sample, highlighted as rigid regions in the centre of the strand, most apparent in the high contrast image in D).

seen in the AFM images obtained from **55r** annealed at pH 7.4, where only dsDNA is seen, which is consistent with the gel electrophoresis.

Some length and height analyses performed on the AFM images of potential triplex structures also support the presence of something other than duplex. The rigid section within the DNA that could correspond to a triplex is 0.5-0.9 nm taller than the surrounding dsDNA, which is consistently 2 nm (see Fig. 3.14, B). This is observed across all areas imaged. Height analysis performed on samples of **55r** annealed at pH 7.4 shows a height of 2 nm throughout the images, suggesting just the presence of duplex in the sample. This is consistent with what is observed in the images as well as the gel electrophoresis results reported previously.

Triplex structures have been studied by AFM previously, predominantly in supercoiled plasmids.[117,118] A previous study reported the formation of intramolecular

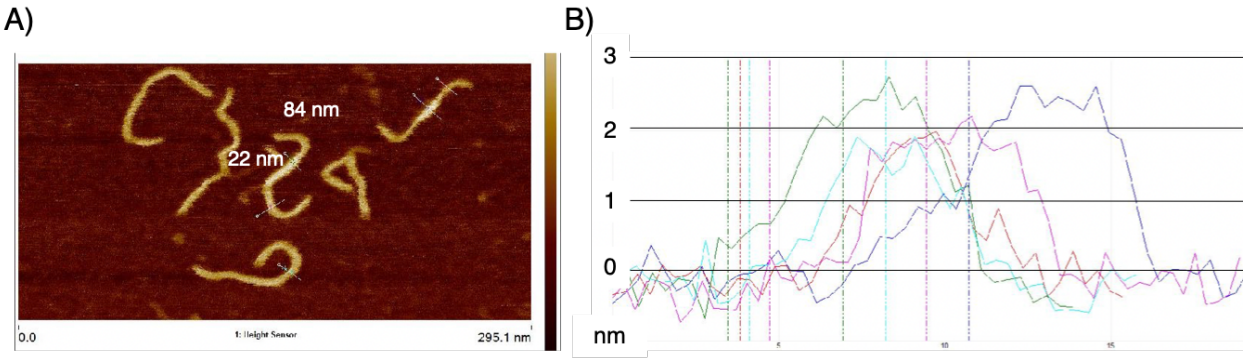


Fig. 3.14: Length and height analysis of apparent triplex structures in AFM images of samples of **55r** annealed at pH 5.5. The apparent triplex structures are 0.5-0.9 nm taller than surround dsDNA.

triplex structures which were seen as protrusions from a supercoiled plasmid.[117] Another study reported AFM images of linear triplexes formed from poly(dG)-poly(dG)-poly(dC) synthesised by Klenow exo, a fragment of DNA polymerase.[118] This study reports in depth height analyses of the images showing triplex portions to be up to 1 nm taller than the duplex counterpart.[118] This supports the analyses performed on the $(GAA)_{55}$ DNA triplexes shown here.

3.3 Summary and outlook

It was found that plasmids containing 55 repeats (**55r**) could be produced in bacteria and amplified by PCR, resulting in the expected size of bands as seen on an agarose gel and with acceptable purity. However, plasmids containing 200 repeats (**200r**) experienced massive deletions when grown in bacteria, with the largest insert achieved being 75 repeats. PCR amplification of **200r** was also unsuccessful, with extensive polymerase slippage and non-specific replication being observed with gel electrophoresis. Clones containing 75 repeats (**75r**) that were initially grown in bacteria from the **200r** plasmid were successfully amplified with PCR to give a band of the expected size as seen by gel electrophoresis.

Multiple secondary structures were observed when purified PCR products of **55r** and **75r** were annealed at acidic pH. These structures were not seen when the DNA was annealed at neutral pH. This is consistent with the findings from the short oligonucleotide studies reported in the previous chapter, however it cannot be

confirmed which strand is acting as the TFO and hence whether the structures observed here are pyrimidine or purine triplexes. The much heavier band also seen by gel electrophoresis could be attributed to the association of two or more triplex structures, such as 'sticky DNA', although the formation of 'sticky DNA' has been previously reported to be encouraged by negative supercoiling in a non-linear system.[16]

Potential DNA-RNA hybrid structures were also observed by gel electrophoresis, especially when the samples were incubated for 4 •C for 48 hours. Since the incubation at low temperature does not involve disruption of the duplex by heating, it implies that the multiple bands observed could correspond to the interaction of DNA and RNA. These bands were observed in samples containing purine and pyrimidine RNA TFOs at both acidic and neutral pH, though were more prominent with pyrimidine RNA at neutral pH. AFM imaging of the DNA-RNA hybrid samples without gel purification (that caused fragmentation and dissociation) would perhaps shed light on the physical nature of the structures seen on the gel.

AFM images of **55r** annealed at pH 5.5 and pH 7.4 were collected. The images of samples annealed at pH 5.5 show apparent triplex structures, seen as rigid regions in the centre of the DNA strand where the $(GAA)_{55}$ repeat insert is situated. Height and length analysis showed that these apparent triplex regions were 0.5-0.9 nm taller than surrounding dsDNA, consistent with previously reported AFM triplex studies.[118] Samples annealed at pH 7.4 showed no evidence of triplex structures, which was in accordance with the gel electrophoresis.

Along with biological and biophysical studies, these structures can be chemically probed to elucidate more about their composition. The way chemical probes interact with nucleic acids and how their properties change upon binding has been employed in the following chapter to determine a method to image triplex DNA in novel manner.

4 Development of fluorescence lifetime probes for the detection of triplex DNA

4.1 Introduction

In this chapter, the use of small molecule fluorescent probes based on the structure of thiazole orange to study triplex DNA through fluorescence lifetime imaging is reported. Fluorescence enhancement and fluorescence lifetimes of probes bound to single stranded, duplex, triplex and quadruplex DNA at pH 5.5 and pH 7.4 *in vitro* were measured, and probes were incubated with both wild-type and $(GAA)_{310}$-containing HEK293 cells and fluorescence lifetimes were collected using fluorescence lifetime imaging microscopy.

In order to investigate whether triplexes play a role in blocking the transcription of the *FXN* gene *in vivo*, small molecule probes with the ability to image triplexes selectively would be of great benefit. Small molecules that fluoresce when bound to DNA are useful tools, as they are often cell permeable and properties can be easily tuned with chemical modifications.[185] However, the difficulty comes with designing a probe that binds specifically to a certain DNA topology. The basic structure of a triplex is similar to that of duplex, being a double helix structure with a third strand bound in the major groove. It follows a molecule that binds strongly to triplex DNA will almost certainly also bind to duplex, making the design of a triplex-specific small molecule very challenging. To overcome this, an alternative detection method can be used. Fluorescence lifetime is an intensity-independent measure of the time a molecule

spends in its excited state and can vary depending on the environment of the fluorophore. Hence, a probe that has a different fluorescence lifetime when bound to triplex DNA than any other topology would be a novel approach to detecting triplexes in solution and *in cellulo.*

There are three primary methods of small molecule binding to DNA: intercalation, groove binding and covalent binding. DNA intercalating agents bind to DNA by non-covalently inserting between adjacent DNA bases and are predominantly planar aromatic systems that can bind through π-π stacking with aromatic moieties on the DNA bases. This binding is predominantly reversible and can also act to stabilise the underlying DNA structure, making it the most suitable binding mode for many DNA probing and targeting molecules. Many intercalating agents have 'switch-on' optical properties, such as fluorescence, upon binding to DNA, due to elimination of decay through nonradiative means such as intramolecular rotation. This makes them useful probes for detection. Groove-binding agents can bind in the major or minor groove of the DNA structure, though tend to show preference for the minor groove due to the favourable positioning of the nucleotide functional groups.[202] These molecules are generally crescent-shaped to match the curvature of the minor groove, and their binding is strongly driven by van der Waals, hydrogen bonding, and electrostatic interactions with the nucleotide functional groups and the negatively charged backbone.[202] Covalently binding small molecules bind strongly and irreversibly to DNA, often causing damage. Such compounds are frequently used as chemotherapy agents in the treatment of cancer. *Cis*-platin, for example, acts by covalently binding to two separate DNA strands and forming an inter-strand cross-link, thus preventing the DNA from being replication, thus preventing cell growth. It follows that the best modes of binding of small molecules for live cell imaging are intercalation and groove-binding. Groove-binding molecules are often sequence-specific, and triplex groove-binding molecules have been reported to bind primarily in T*A-T triplets.[186] This would not be suitable for imaging of purine triplexes containing G*G-C and A*A-T triplets. Hence, a DNA intercalating agent would be most suitable for the imaging of triplexes in live cells. In general, features of a good triplex intercalator are an extended, crescent-shaped polyaromatic surface area which is large enough to provide π-π overlap with the 3 bases in each triplet. A cationic charge on the chromophore

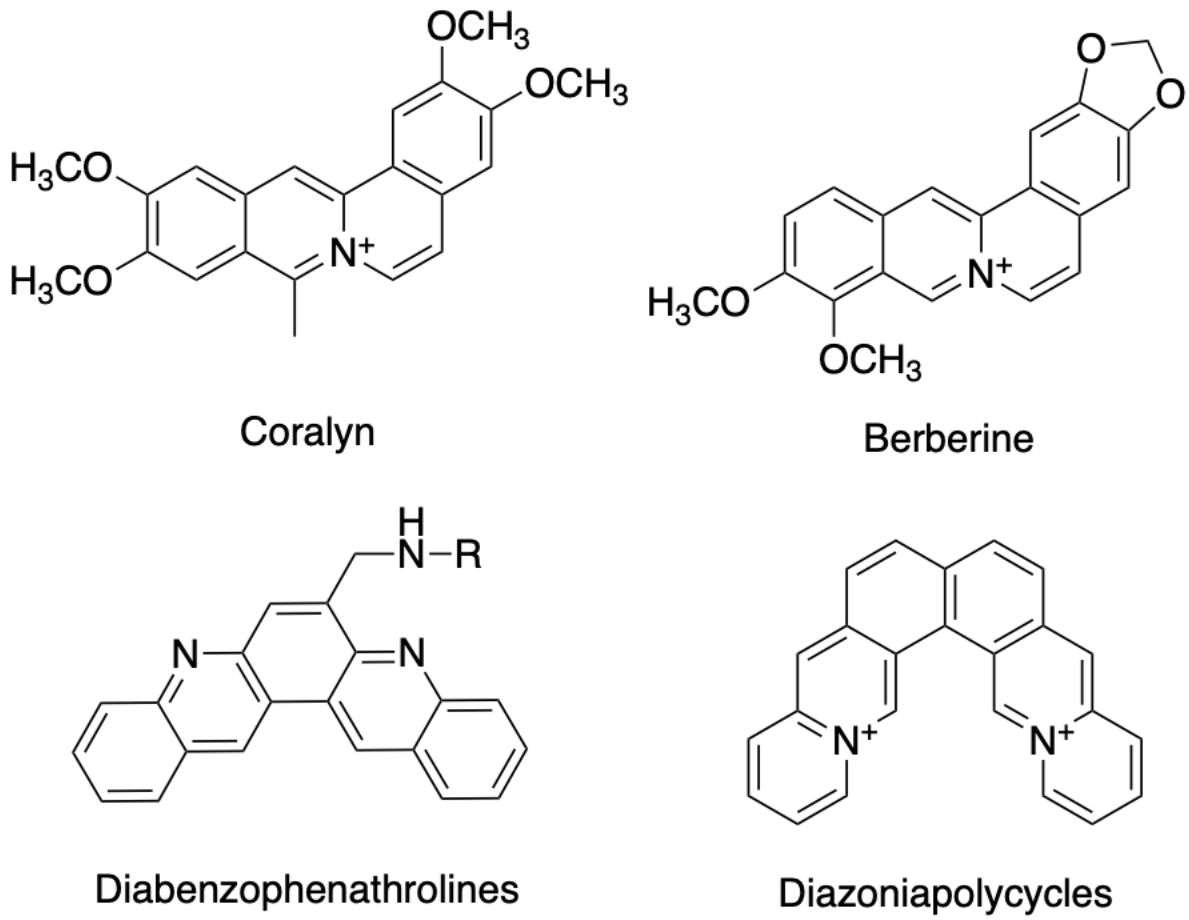


Fig. 4.1: Some examples of molecules that bind to triplex DNA through intercalation, featuring the crescent-shaped structure, positively charged chromophores and substituents with hydrogen-bonding ability.

encourages binding to the negatively charged phosphate backbone of the DNA, and a substituent with groove-binding properties such as a basic amine residue could also be beneficial (see Fig. 4.1). A substituent that can be chemically altered in order to tune the fluorescence properties would be useful in order to ensure the fluorescence lifetime is different enough when bound to triplex over duplex. Thiazole orange, a well-known DNA intercalator, was chosen for the base structure of the probe.

Thiazole orange (TO) is one of the oldest synthetic cyanine dyes and has been widely used as a nucleic acid stain for decades. Today, TO is commonly used in reticulocyte analysis, and is known to permeate live cell membranes and stain residual RNA in reticulocytes.[130] The fluorescence of TO is extremely weak in aqueous solution. However, upon binding to DNA or RNA, the fluorescence yield increases dramatically, accompanied by a shift in the absorption spectrum and an induced CD signal. It has been demonstrated that TO has a binding stoichiometry to dsDNA of 1

dye molecule per 2 base pairs.[130] Despite the considerable increase in fluorescence yield, binding of TO to ssDNA and dsDNA is relatively weak; complexes of TO and DNA readily dissociate during gel electrophoresis and chromatography. In contrast to this, TO has been reported to bind strongly to quadruplex and triplex DNA.[130] One study reported the strong binding of TO to a long poly(dG-dG)-poly(dC) triplex and G-quadruplex DNA, resulting in massive increase in fluorescence yield in comparison to binding to dsDNA, and formation of a TO-DNA complex that persisted through gel electrophoresis.[130] This study also investigated the binding mode of TO to triplex DNA. Intercalation of organic molecules between base pairs has been previously reported to result in lengthening of DNA, whereas groove-binding molecules did not.[130] It was reported that the binding of TO to the poly(dG-dG)-poly(dC) triplex caused lengthening of the DNA as observed by AFM, which suggests that TO binds to triplex DNA through intercalation.[130] Another study reports the molecular engineering of TO to create a G-quadruplex-selective fluorescence probe by introducing styrene-like substituents at the ortho position.[131] With the addition of these substituents, the fluorescence quantum yield of the probes upon binding to G4-DNA was significantly increased in comparison to binding with ssDNA and dsDNA, and exhibited stabilisation of G4-DNA as seen by the increase in melting temperature (ΔT_m). Although triplexes were not investigated in this study, the structures reported are good candidates for triplex fluorescent lifetime probes due to their extended aromatic systems and modifiable styrene-like substituents. These substituents provide another planar aromatic surface with which to stack into DNA, and further resemble the more crescent-shaped triplex binders such as berberine.

Fluorescence lifetime spectroscopy presents an intensity-independent approach to identifying different DNA topologies *in vitro* and *in cellulo*. The fluorescence lifetime of a fluorophore is a measure of the time that fluorophore spends in an excited state before emitting a photon and returning to the ground state; this is characteristic of each molecule. The lifetime is defined as the time taken for the number of excited molecules to decay by 1/*e* (Fig. 4.2). This is a function of competition between radiative and nonradiative decay pathways. Fluorescence lifetimes are generally on the order of nanoseconds, though can be as short as picoseconds in some cases. The benefit of lifetime measurements is that they are concentration and intensity independent. This

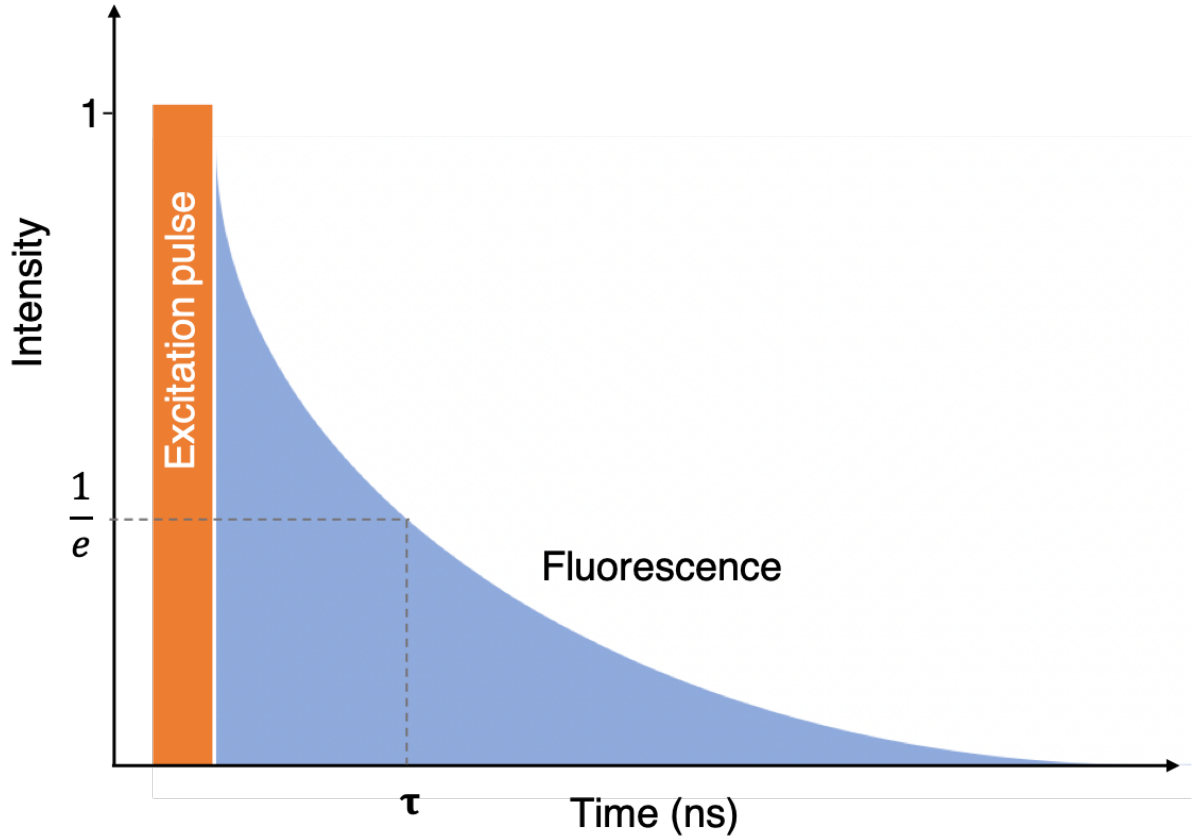


Fig. 4.2: Schematic showing the standard fluorescence lifetime decay after excitation by a pulse. The lifetime (τ) is defined as the time taken for the number of excited molecules to decay to 1/*e*.

is particularly beneficial for live cell imaging, where fluorophores can accumulate in certain organelles which can skew intensity-based measurements. Primarily, this eliminates the need for a probe with high affinity and selectivity for triplex DNA. If triplex DNA does indeed play a role in gene silencing in FRDA, this structure would make up a small percentage of the total DNA present in the cell, which would be predominantly double stranded. Hence, any intensity-based probe would need 100% selectivity for triplex DNA in order to eliminate the background noise caused by binding to duplex DNA. This concept can be applied to cell imaging also. Fluorescence lifetime imaging microscopy (FLIM) involves collecting an image where the fluorescence lifetime decay of each pixel is also measured. This technique allows the pinpointing of areas of long/short lifetime within cells, as well as an average fluorescence lifetime across the whole cell. Theoretically, with a probe showing a significantly different lifetime when bound to triplex compared to other DNA topologies, FLIM could be used to shed light on whether triplexes are present in the cell and where they might be located.

4.2 Results and discussion

4.2.1 Probe design

As stated above, thiazole orange (TO) was chosen as the base structure for the probe design. TO has been reported previously to bind strongly to triplex DNA and other non-canonical DNA structures, such as G-quadruplex, and exhibits a large increase in fluorescence emission upon binding to DNA in general.[130] Introduction into a rigid environment, such as DNA, prevents intramolecular rotation about the carbon-carbon single bond connecting the two aromatic moieties in TO, thus preventing the quenching of fluorescence through this nonradiative decay pathway.

General structural characteristics of triplex intercalator molecules include a crescent-shaped aromatic surface, extended conjugation and positive charges within the aromatic ring system; binding can also benefit from a sidechain with groove-binding properties to provide additional stabilisation. The crescent shape allows for most complete overlap with the aromatic surface of the nucleotides on each of the three strands in the triplex. TO consists of two aromatic moieties connected by a carbon-carbon single bond which allows free rotation. These planar regions provide a good basis for stacking with nucleotides. TO also has a positively charged nitrogen atom within the ring system, which is also in resonance with the second nitrogen atom when the molecule is in a planar conformation, which can interact electrostatically with the negatively charged phosphate backbone of the nucleic acid structure. TO binds strongly to dsDNA and becomes fully embedded in the structure upon intercalation.[131] It follows that TO would be fully embedded upon intercalation to triplex DNA also, as this would present a larger aromatic surface across the three strands, introducing further surface area with which a small molecule can stack (see Fig. 4.3). Extending the conjugated system of TO by adding substituents that can also stack with the nucleotides of the third strand of the triplex would increase stability of the resulting complex and also increase the effect of binding on the fluorescence properties of the probe.

TO-based molecules with aromatic substituents have been previously reported to bind strongly to G-quadruplex, as discussed in Chapter 1.[131] These molecules were

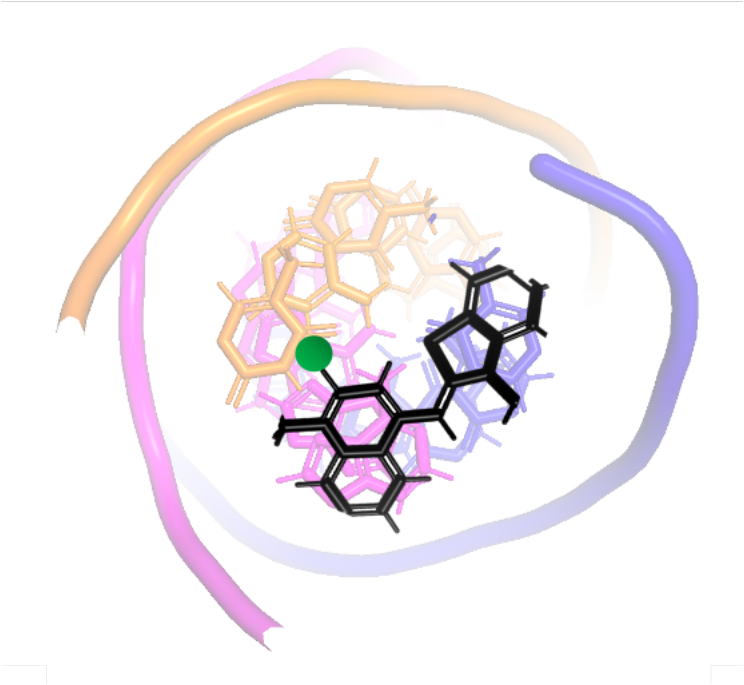


Fig. 4.3: A representation of the structure of TO (black) superimposed over a DNA triplex (1D3X, taken from the PDB) with the two duplex strands in pink and blue, and the third triplex-forming strand in orange. The site where the aromatic substituents were added is indicated by the green sphere.

chosen as suitable candidates for triplex-binding fluorescent probes due to their fluorescence properties, preferential binding to alternative DNA structures and ease of functionalisation. In Fig. 4.3, the structure of TO can be seen superimposed over a surface view of a triplex. The two aromatic regions in TO stack well with the nucleotides of the two duplex strands, however there is no overlap with the triplex-forming strand. Triplex binding would be improved through introduction of a further aromatic region that can overlap with the triplex-forming nucleotide.

The structures of these molecules are shown below (4.2.2). Probes **3** and **4a-d** were reported previously as G-quadruplex binders, and probe **4e** was also investigated to assess the effect of having functional groups on the additional aryl moiety, though **4e** has not been reported in the context of DNA binding before.[131,197]

4.2.2 Synthesis

The TO derivative (**3**) was accomplished via a 3-step synthesis involving the methylation of 4-chloro-2-methylquinoline (**1**) and 2-methylbenzothiazole (**2**), followed by the coupling of the methylated species.[131] The coupling step was low yielding to begin with, but the introduction of a reflux step improved the yield and purity significantly. Introduction of the styrene-like substituents proceeded by reaction with the desired aromatic aldehyde in the presence of a base under reflux (Fig. 4.4).[131] The products were purified by column chromatography using a solvent gradient of methanol (0-10%) in DCM. The purity of **3** and **4a-e** was assessed by ^{1}H NMR spectroscopy, mass spectrometry and elemental analysis.

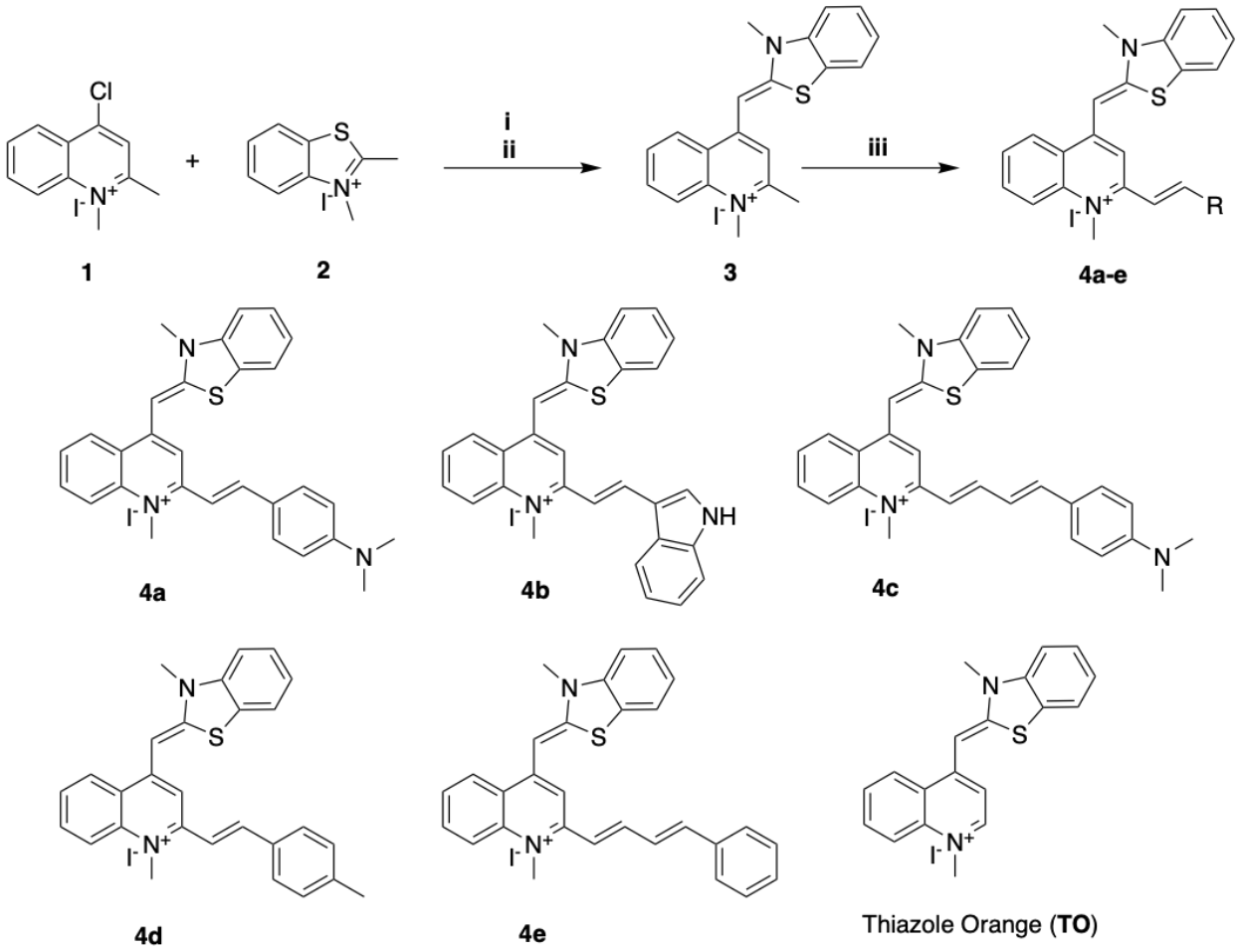


Fig. 4.4: Reaction scheme for the synthesis of **3** and **4a-e** and their molecular structures. Reagents and conditions: (i) $NaHCO_3$ (aq.), methanol, r.t., 1h, (ii) reflux, 30 mins, sat. KI, (iii) aromatic aldehyde, 4-methylpiperidine, reflux, 3h.

4.2.3 Fluorescence properties upon DNA binding

To assess the binding of probes **3** and **4a-e** to different DNA topologies, UV-Vis absorbance and fluorescence emission titrations were performed. For the fluorescence lifetime measurements, it is important that there is no free probe in solution to skew the average measurement. The fluorescence lifetime of the free probe in solution is likely to be significantly different to when it is bound to DNA, and hence would change the average lifetime measurement of the sample. In order to ensure the equivalence of DNA to probe is sufficient for complete binding, absorbance and emission titrations with increasing equivalents of DNA titrated into a solution of the probe at a fixed concentration were used. Once a plateau was reached in the increase of absorbance/emission of the probe upon binding to increasing equivalents of the DNA, it was assumed that all probe was bound. These titrations also provide information on the affinity of binding. The absorbance titrations are also important to evaluate the maximum absorption wavelength throughout the experiment, which can shift upon binding to DNA. This ensures that the excitation wavelength used for the fluorescence emission titrations remains suitable throughout.

The DNA models considered include single stranded DNA (Pu TFO containing a $(GAA)_{10}$ repeat and Py TFO, containing a $(TTC)_{10}$ repeat), double stranded DNA (FRDA duplex containing a $(GAA:TTC)_{10}$ repeat, and ds26), triple stranded DNA (purine and pyrimidine FRDA triplex, 10 repeat units) and G-quadruplex DNA (c-myc and h-telo). For the ssDNA, dsDNA and tsDNA models, conditions of pH 5.5 and 7.4 were investigated for each. The pyrimidine triplex has been shown in Chapter 2 to only form at pH 5.5, and the purine triplex was not formed at either pH, though would be most expected to form under neutral conditions. The quadruplex models were only investigated at pH 7.4, to ensure that the structures remained intact throughout the measurements.

The fluorescence intensity measurements cannot be directly compared between different probes and different DNA topologies, as the slit width was varied between experiments. This was due to the varying intensity of emission between probes. Probe **3**, for example, was by far the brightest probe, and hence the smallest slit width was used to collect the emission spectra to avoid saturation of detector. The

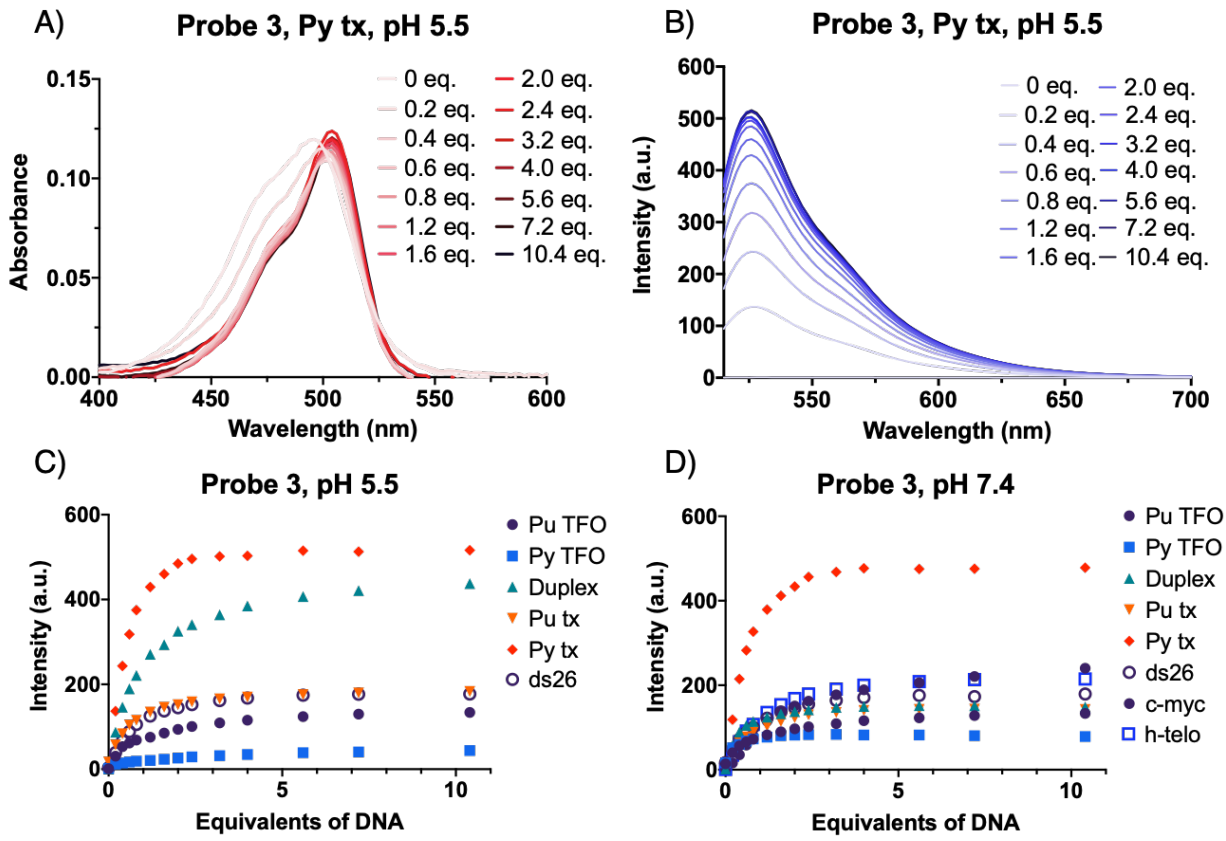


Fig. 4.5: A) An example of an absorption titration, probe **3** with increasing equivalents of pyrimidine triplex at pH 5.5, B) an example of an emission titration, probe **3** with increasing equivalents of pyrimidine triplex at pH 5.5, C) fluorescence binding curve of probe **3** with different DNA topologies at pH 5.5, D) fluorescence binding curve of probe **3** with different DNA topologies at pH 7.4.

modified probes, **4a-e**, had much lower fluorescence intensity and thus a wider slit width was used in order to collect enough light. However, fluorescence enhancement upon binding can be compared between probes when bound to each topology.

Example absorption and emission titration spectra, along with binding curves for different DNA topologies at each pH condition, can be seen in Fig. 4.5. Fluorescence enhancement values $((F-F_0)/F_0)$ of probes **3** and **4a-e** upon binding to DNA are summarised in Fig. 4.6. Probe **4b** was found to be quite insoluble in the buffer used, and thus the fluorescence intensity was so low that reliable spectra could not be recorded. The lack of solubility also meant that no increase in fluorescence was seen upon addition of DNA in solution, indicating that no binding was taking place. Due to this issue, **4b** was eliminated from further investigation.

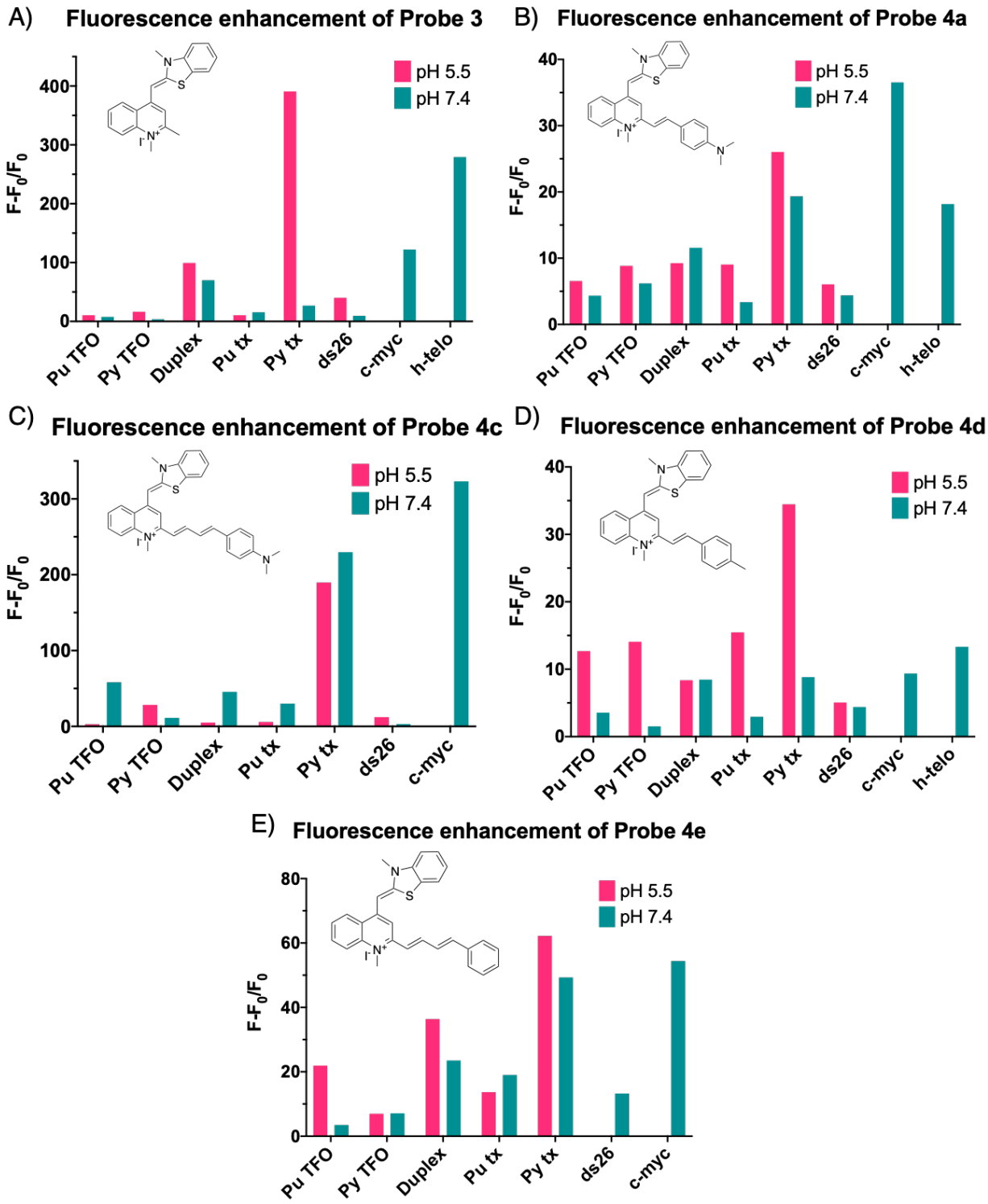


Fig. 4.6: Graphs showing the fluorescence enhancement ($(F-F_0)/F_0$) of probes **3** and **4a-e**, calculated from the integrated spectrum of the probe in the absence of DNA (F_0) and the integrated spectrum of probe with 10.4 eq. of DNA (F) with different topologies at pH 5.5 (pink) and pH 7.4 (blue).

It can be seen from the graphs in Fig. 4.6 that all probes exhibit significant fluorescence enhancement ($(F-F_0)/F_0$) upon binding to pyrimidine triplex annealed at

pH 5.5. Probes **4a** and **4c** display the greatest fluorescence enhancement upon binding to c-myc, with **4e** also displaying large enhancement upon binding to c-myc, though slightly larger enhancement with pyrimidine triplex at pH 5.5. Generally, the fluorescence enhancement appears to be greater at pH 5.5 than pH 7.4. Under acidic conditions, there may be protonation of the nitrogen in the dimethylamine group at the end of the substituents of **4a** and **4c**. This protonation could increase the interaction with this group and the negatively charged phosphate backbone of the DNA. Fluorescence enhancement upon binding in the purine triplex samples is generally low and in a similar range to that seen for duplex/ds26. This was expected, as the purine triplex was not observed to form by gel electrophoresis or CD as described in the previous chapters, and as such it can be expected that the purine triplex samples contain dsDNA and unbound purine TFO.

The binding curves obtained by taking the fluorescence intensity at the maximum emission wavelength (λ_{em}) for all probes at increasing equivalents of DNA show that saturation occurs at 4-5 equivalents of DNA to probe (examples in Fig. 4.5 C and D). This is true for all DNA topologies under both pH conditions tested. Hence, 5 equivalents of DNA to probe was deemed sufficient to ensure all probe in solution was bound to DNA for fluorescence lifetime measurements.

From these binding curves, preliminary values for the binding affinity constants (K_a) were calculated for each probe with each DNA topology at both pH 5.5 and pH 7.4. As these titrations could not be performed in triplicate due to time constraints, these calculated values can only be used to give an initial approximation of the binding affinities of the probes for each DNA topology. The data was fit via the least-squares fitting method for non-competitive binding developed by the Aldrich-Wright group as per the following equation:[203]

$$I = 0.5 * I_{max} * [\frac{1}{K_a} + L + n[DNA] - \left\{\left(\frac{1}{K_a} + +n[DNA]\right)^2 - 4 * L * n * [DNA]\right\}^{0.5}]$$

Where I is the fluorescence intensity, I_{max} is the maximum fluorescence intensity at the plateau, K_a is the affinity binding constant derived using a Levenberg-Marquardt nonlinear least-squares iteration algorithm, L is the concentration of ligand added (2

μM), *[DNA]* is the concentration of DNA added (in M), and *n* is the number of binding sites (set to a constant of 2).

However, this model could not be used to reliably fit all the data. The resulting fitting for probe **4d** in particular was unreliable, and it was hypothesised that this was due to a significant proportion of unspecific and incomplete binding, which is further validated with the fluorescence lifetime values reported later in this chapter. An alternative method was used to fit this data: the Swillens method, which takes any non-specific binding into account.[204] Data was fit according to the following equation:

$$I = \underbrace{\frac{(I_{max} * [DNA])}{(K_d + [DNA])}}_{\text{specific}} + \underbrace{(\alpha * [DNA])}_{\text{non-specific}} + I_0$$

Where *I* is fluorescence intensity, I_{max} is the maximum fluorescence intensity at the plateau, [DNA] is the concentration of DNA added (in M), K_d is the dissociation constant (in M) derived using a Levenberg-Marquardt nonlinear least-squares iteration algorithm, α is the gradient of the initial slope (in M^{-1}) and I_0 is the background fluorescence intensity of the compound in the absence of DNA. The affinity constant (K_a) was calculated as the reciprocal of the dissociation constant (K_d).

The values for K_a fit using either the Aldrich-Wright method, or the Swillens method (shown in **bold**), are reported in Table 4.1. The R^2 value for the fitting is also reported in brackets to indicate the goodness of fit, with a value closer to 1.0 indicating a better fit. Mathematically, $R^2 = 1.0 - SS_{reg}/SS_{tot}$, where SS_{reg} is the sum-of-squares of the points from the regression curve, SS_{tot} is the sum-of-squares of the distance of points from a horizontal line where *y* equals the mean of all the data points. An example of the fitting can be seen in Fig. 4.7.

Table 4.1 - K_a values for probes **3**, **4a**, **4c**, **4d** and **4e** when bound to different DNA topologies under different conditions, reported as $x10^5$ M^{-1}.

		3	4a	4c	4d	4e
pH 5.5	**Pu TFO**	4.2 (0.97)	4.6 (0.95)	-	**2.0 (0.93)**	4.6 (0.99)
	Py TFO	2.5 (0.94)	3.0 (0.99)	1.4 (0.99)	**23 (0.97)**	0.6 (0.99)
	Duplex	4.1 (0.99)	4.7 (0.95)	3.4 (0.90)	**6.8 (0.99)**	4.0 (0.99)
	Pu tx	9.7 (0.98)	1.2 (0.99)	2.9 (0.86)	**18 (0.99)**	4.9 (0.97)
	Py tx	12.9 (0.99)	0.7 (0.98)	3.5 (0.96)	**16 (0.99)**	7.9 (0.99)
	ds26	6.1 (0.99)	1.8 (0.99)	5.1 (0.95)	**20 (0.98)**	-
pH 7.4	**Pu TFO**	5.9 (0.95)	2.3 (0.94)	1.5 (0.97)	**13.1 (0.99)**	5.1 (0.99)
	Py TFO	12 (0.81)	4.0 (0.96)	1.3 (0.97)	**8.7 (0.97)**	2.2 (0.80)
	Duplex	22 (0.98)	3.0 (0.96)	4.7 (0.97)	**18 (0.99)**	6.1 (0.98)
	Pu tx	8.0 (0.99)	1.3 (0.95)	2.7 (0.99)	**12 (0.99)**	3.6 (0.96)
	Py tx	10 (0.99)	2.6 (0.99)	34 (0.97)	**12 (0.99)**	28 (0.98)
	ds26	6.1 (0.98)	1.3 (0.92)	15 (0.98)	**17 (0.99)**	10 (0.99)
	c-myc	1.6 (0.99)	1.8 (0.97)	2.8 (0.98)	10 (0.98)	1.2 (0.97)
	h-telo	3.9 (0.99)	2.4 (0.99)	-	16 (0.98)	1.2 (0.97)

The fluorescence intensity of the free probes in solution is so low in all cases that a spectrum above the baseline is barely discernible, with the exception being probe **3** which is far brighter than the modified probes **4a-e**. This was also reported in the investigation of the fluorescence properties of these probes when bound to quadruplex DNA.[131] This is a well-known property of **TO**, and precisely what makes it such a useful nucleic acid stain.[128] The extremely low quantum yield demonstrated by the free dye in solution can be attributed to the nonradiative relaxation allowed by the twisting around the bridge between the benzothiazole and quinoline moieties.[187] The two aromatic regions in the structure of **TO** are linked by a carbon-carbon single bond with free rotation, and resonance can only occur when these regions are coplanar to each other, increasing energetic overlap between the excited state and the ground state.[187] This isomerisation is hindered upon DNA intercalation, blocking the nonradiative decay pathway and triggering relaxation by radiative means, thus greatly increasing the fluorescence emission.[188,189] This is the mechanism that regulates the proportional relaxation between radiative and nonradiative pathways, and in turn also governs the lifetime of the excited state. It follows that intercalation into ssDNA would

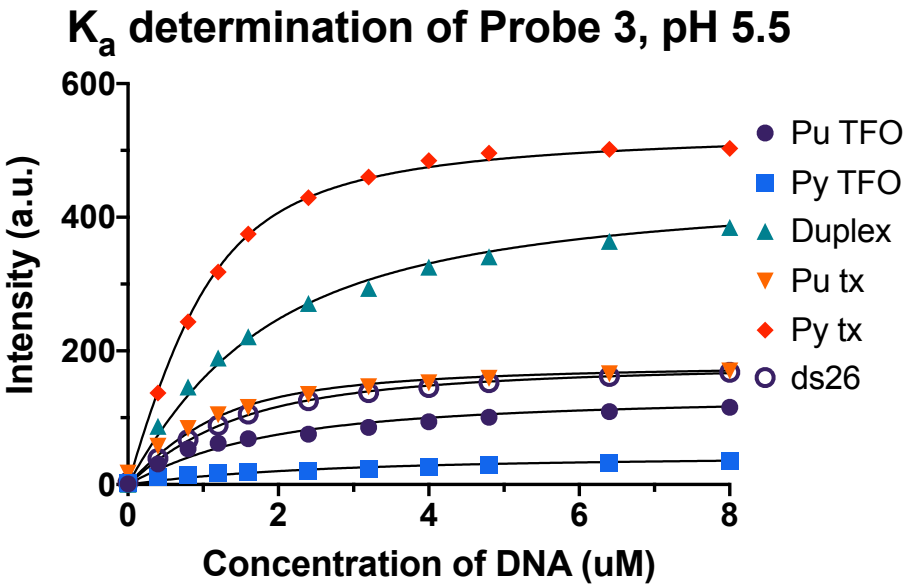


Fig. 4.7: An example of the fitting of the binding curve data obtained from the emission titrations using the Aldrich-Wright method. This shows the binding of probe **3** to DNA topologies at pH 5.5. The binding curve data is indicated in colour and the fitting is indicated by the black lines.

not result in a large increase in fluorescence, as **TO** (and TO-based structures) would not be fully embedded into the single-stranded structure, and thus this may allow some rotation along the single bond connecting the two planar aromatic moieties of **TO**. This movement will induce nonradiative relaxation and consequently the radiative decay pathway will be less prominent. This is indeed what is observed here; the fluorescence enhancement ($(F-F_0)/F_0$) of all probes investigated is predominantly the lowest when bound to ssDNA (Pu TFO and Py TFO). From the emission titrations and the binding curves for the ssDNA models, it can be seen that the fluorescence intensity is extremely low in comparison to the other models. In contrast to this, more significant fluorescence enhancement was seen upon binding to dsDNA (FRDA duplex and ds26). This is to be expected, as **TO** has been historically used for staining dsDNA. Previous studies have reported the intercalation of **TO** into the 5′-CpT-3′ binding site of a duplex, stacking in a planar configuration as demonstrated by computational modelling.[131] This is also supported by NMR studies of homodimeric **TO** (or **TOTO**) similarly stacking at the same site in duplex DNA.[190] This intercalation is much more effective than with ssDNA, as the **TO** molecule is fully embedded in the binding site. Consequently, movement around the carbon-carbon bridge is restricted, thus causing an increase in radiative decay and as such increasing the observed emission. The

modified probes **4a-e** generally have a lower fluorescence intensity. Computational docking studies performed previously on probes **4a-d** show that they intercalate at the same site as **TO**, but the added styrene-like substituents at the ortho position protrude from the binding pocket.[131] It can be assumed that **4e** behaves similarly. It was hypothesised that this substituent could interact with solvent molecules and have a negative effect on binding to dsDNA, resulting in decreased fluorescence enhancement.[131] The substituents could also cause a decrease in emission due to more movement within the molecule. The styrene-like substituents are free to move about the single carbon-carbon bond connecting them to the **TO** framework, and since the whole system is conjugated, this may increase the incidence of nonradiative relaxation and thus decrease the emission in general.

The modified probes **4a-d** were previously reported as fluorescence intensity-based probes to image G4-quadruplexes in cells.[131] The binding of **TO** to quadruplexes and triplexes has been investigated before, and it was found that **TO** binds with higher affinity to these non-B DNA conformations over dsDNA.[130] The resulting probe-DNA complexes were also more stable, and not seen to dissociate during gel electrophoresis.[130] This particular study also reported that the fluorescence enhancement of **TO** upon binding to intramolecular poly(dG-dG)-poly(dC) triplexes and G-quadruplexes was much greater than upon binding to dsDNA.[130] This is contradicted, however, in the study that reported the modified **TO** probes with the styrene-like substituents. This study reports that **TO** exhibits equal fluorescence enhancement in the presence of dsDNA and G-quadruplex, and thus can't be used as a method to differentiate between them.[131] Following this, the styrene-like substituents were incorporated onto the **TO** derivative (**3**) to increase the affinity for G-quadruplex. It was reported that with the modified probes, the fluorescence enhancement upon binding to G-quadruplex was greatly increased in comparison to dsDNA.[131] This is mostly in accordance with the measurements recorded here with probes **4a-e** and the G4-quadruplex models (c-myc and h-telo), though increased fluorescence enhancement upon G4 binding was seen for **3** also. From the docking studies reported previously, it can be seen that **TO** stacks on top of the G-quadruplex with a binding stoichiometry of 1:1.[131] Similar studies for the modified probes showed improved stacking on the quadruplex, with the substituent situated close to the side of the G-

quartet pocket.[131] This additional stabilisation is the primary factor causing the increase in binding affinity and subsequent fluorescence enhancement.

The binding of probes **3** and **4a-e** to purine and pyrimidine triplexes containing the FRDA insert at pH 5.5 and pH 7.4 was investigated. Generally, a large increase in fluorescence was seen upon binding to the pyrimidine triplex at pH 5.5. This was also observed for the pyrimidine triplex at pH 7.4 to a lesser extent. This level of fluorescence enhancement was not seen in the purine triplex samples at either pH, which was similar to the fluorescence enhancement observed in the presence of dsDNA. This is to be expected, as the purine triplex was not seen to form at either pH as demonstrated in the previous chapters, hence it can be assumed that this is purely a mixture of dsDNA and ssDNA in solution. Probe **3** exhibited the greatest fluorescence enhancement upon triplex binding, followed by **4c** and then **4a, d** and **e**. As previously discussed, the emission observed is a function of the competition between radiative and nonradiative decay pathways. The greater fluorescence enhancement upon triplex binding compared to dsDNA binding is caused by a reduction in relaxation by nonradiative means as a result of further restriction of rotation and movement within the **TO** molecule. The introduction of the third strand into the major groove of the dsDNA to form the triplex further limits any movement of molecules intercalated into the structure. The previously reported docking studies of **TO** bound to dsDNA show that the molecule is fully embedded in the helix, however complexes of **TO** and dsDNA easily dissociate, and the binding affinity is relatively low.[130,131] Without confirmation from docking studies, it can be hypothesised that the third strand in the triplex structure could act to limit dissociation of the probe once intercalated into the structure. The stronger interactions could aid binding and increase affinity, further enhancing fluorescence. It has been previously reported that **TO** binds to triplex through intercalation with a stoichiometry of 1 dye molecule to 2 triads.[130]

As previously reported in docking studies of **4a-d** to dsDNA, the dyes intercalate into the same site as **TO**.[131] The styrene-like substituents were seen to protrude from the structure and may have acted to hinder binding.[131] The third strand in the triplex conformation could provide the aromatic surface with which the substituent can stack, hence increasing the binding affinity. The probes exhibit different extents of fluorescence enhancement, which can be attributed to the variation in substituents.

Probe **4c** showed the greatest enhancement, which could be due to the longer conjugated chain in the substituent. The rotation about this chain arguably provides more opportunity for nonradiative relaxation, and hence when its movement is limited by intercalation into triplex DNA, this results in the largest increase in emission. The dimethylamine group situated at the end of the substituent could provide additional stabilisation of the probe-DNA complex.[191] This could occur through interaction with the minor groove or hydrogen bonding with the DNA bases themselves. This could explain why **4c** shows greatest fluorescence enhancement when bound to G-quadruplex, as this substituent has been shown in docking studies to be positioned at the G-quartet pocket such that the amine could accept hydrogen bonds here.[131] Probe **4e** also has a longer conjugated sidechain, however, does not have the terminal amine. It exhibits lower fluorescence enhancement and less discrimination between DNA topologies. This could be due to the absence of the dimethylamine group making binding less specific and with lower affinity. Probe **4d** appears to exhibit good fluorescence enhancement specificity towards triplex DNA. Previous investigation of the binding of these probes towards G-quadruplex suggests that the methyl group at the end of the substituent in **4d** is not beneficial for binding to quadruplex DNA when compared to the dimethylamine group on **4a** and **4c**.[131] Whilst this could be detrimental for quadruplex binding, it appears to have no negative effect on triplex binding, as the planar aromatic surface is still able to intercalate into the base triads. Probe **4a** is structurally similar to **4c** except for the shorter chain length of the substituent. This was seen previously to be beneficial for quadruplex binding, as the extended substituent of **4c** resulted in the dimethylamine group being positioned outside of the G-quartet binding pocket, whereas this was not the case for **4a**.[131] Due to this, **4a** shows a greater extent of fluorescence enhancement for quadruplex than other topologies.

The K_a values calculated are generally highest when bound to Py triplex and G-quadruplex. The K_a values when bound to the ssDNA TFOs are unexpectedly high; it could be suggested that this is mostly due to skewed data from the very low fluorescence enhancement, and hence there is not a discernible plateau, making data fitting unreliable. The fluorescence binding curves for all probes when bound to ssDNA demonstrate a sharp increase in emission upon addition of DNA and then a quick plateau at a very low intensity value. This is most likely due to partial intercalation of

the probes into the ssDNA, which does result in some fluorescence enhancement, but the effect is not significant as the probes are not fully intercalated into the DNA, allowing for some quenching of fluorescence through rotation. The sharp initial increase in emission upon DNA addition could be construed as strong specific binding. The K_a literature values for probes **4a**, **4c** and **4d** when bound to telo21, another G-quadruplex structure, were reported to be 9.65, 5.38 and 2.33 x10^5 M^{-1} respectively.[131] The values calculated here for probes **4a**, **4c** and **4d** when bound to c-myc were 1.8, 2.8 and 10 x10^5 M^{-1} respectively, and for probes **4a** and **4d** bound to h-telo were 2.4 and 1.2 x10^5 M^{-1} respectively. The values obtained here differ from literature, as do the experimental conditions; however, the values are of the same order of magnitude. Since these experiments were performed only once, this data can be taken as a preliminary approximation of the binding affinities of the probes for each DNA topology.

Despite there being a clear increase in fluorescence enhancement upon binding of all the probes to triplex and quadruplex DNA over dsDNA, there is still significant fluorescence when bound to dsDNA. Thus, fluorescence enhancement is not a sufficient measure of binding to triplex and cannot be used *in cellulo,* as the background fluorescence from binding to dsDNA would be too great. Once the fluorescence properties of the probes when bound to different DNA topologies including triplex were established, a fluorescence-based intensity-independent method for triplex detection was developed. Difference in fluorescence lifetime between binding to different topologies would eliminate background noise and also the need for selectivity for a given topology.

4.2.4 Fluorescence lifetime studies

The fluorescence lifetimes of probes **3** and **4a-e** when bound to different DNA topologies *in vitro* were measured. This was done using time correlated single photon counting (TCSPC). It was determined from the absorption/emission titrations described in the previous section that 5 eq. of DNA to 1 eq. of probe was sufficient to ensure that there was no free probe in solution that would skew the average fluorescence lifetime. The DNA models used were the same as for the titrations in the previous section i.e. single stranded DNA (Pu TFO and Py TFO), double stranded DNA (GAA:TTC_{10} and ds26), triple stranded DNA (purine and pyrimidine triplex, 10 repeat units) and G-quadruplex DNA (c-myc and h-telo). For the ssDNA, dsDNA and triplex models, conditions of pH 5.5 and 7.4 were investigated for each. The

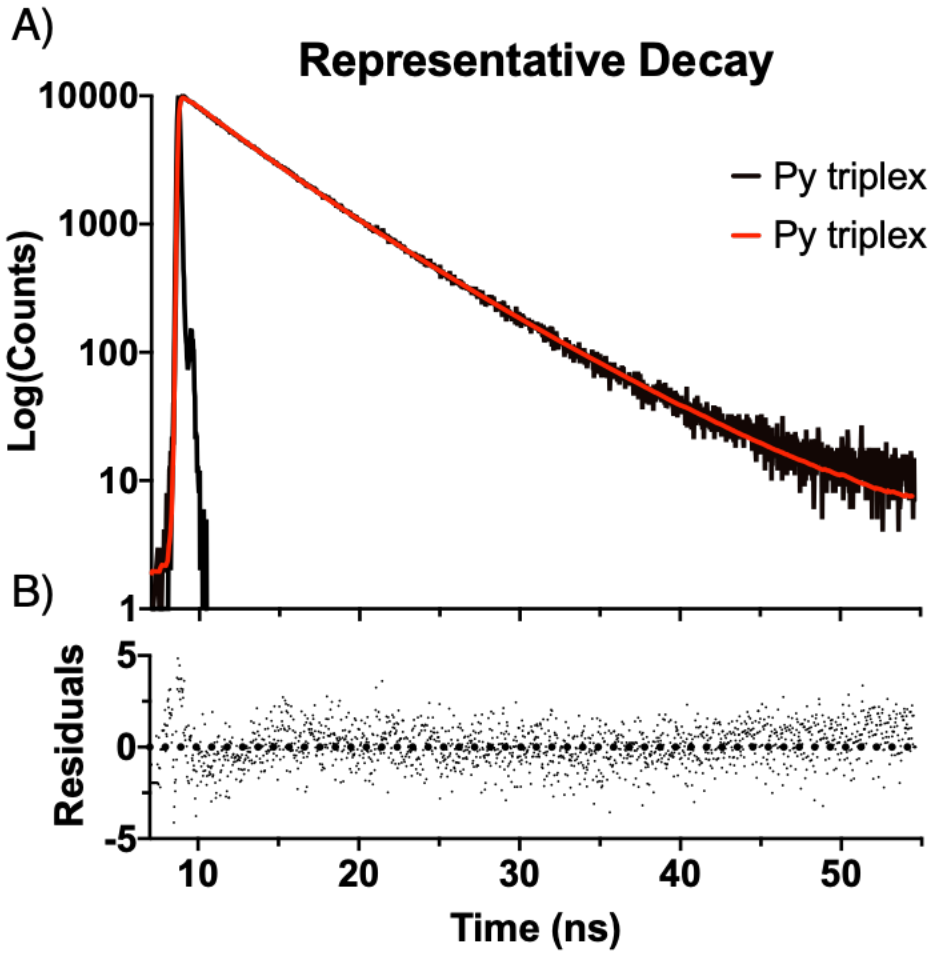


Fig. 4.8: A) Representative fluorescence lifetime decay: probe **3** (2μM, 10 mM NaCac buffer at pH 5.5, 100 mM NaCl and 15 mM $MgCl_2$) with pyrimidine triplex (10 μM). For this plot, the data traces are shown in black and the fitting shown in red (bi-exponential, χ^2 = 1.32). B) The residual trace indicating how good the fit is, 0 being the perfect fit.

quadruplex model was only investigated at pH 7.4, to ensure that the structure remained intact throughout the measurements. TCSPC was used to determine the lifetimes for each probe with each DNA topology. The data was analysed using the DAS6 fluorescence decay analysis software (HORIBA Scientific) and fit to an appropriate exponential function (Fig. 4.8). The individual components of each decay (τ_1, τ_2, etc.) and the contributions of each component (α_1, α_2, etc.) were extracted and a weighted lifetime (τ_w) was calculated using the following equation (for a bi-exponential decay):

$$\tau_w = \frac{\alpha_1 {\tau_1}^2 + \alpha_2 {\tau_2}^2}{\alpha_1 \tau_1 + \alpha_2 \tau_2}$$

Where τ is the lifetime of the component and α is the contribution of that lifetime value to the weighted lifetime (τ_w). The goodness of fit can be assessed using the χ^2 parameter; a value between 1 and 2 indicates an acceptable fit, with closer to 1 being a better fit.

It can be seen from the graphs in Fig. 4.9 that the fluorescence lifetimes of probes **3** and **4a-e** are seen to be generally longer when bound to pyrimidine triplex in comparison to dsDNA. Interestingly, the fluorescence lifetime is similar for the pyrimidine triplex at both pH 5.5 and pH 7.4. This could indicate that the probes are stabilising the pyrimidine triplex structure and encouraging it to form at neutral pH. The fluorescence lifetimes when bound to purine triplex are similar to dsDNA, which would be expected as previously discussed. There is no evidence of purine triplex formation at either pH as described in the previous chapters, hence it can be assumed that this sample consists of dsDNA and unbound purine TFO. Long fluorescence lifetimes were seen for probes when bound to G-quadruplex DNA also, particularly c-myc. A tri-exponential function was used to fit most of the data, which gave χ^2 values between 1 and 2, however this did not fit the data collected when bound to the triplex structure. For this data, a bi-exponential function was required to yield a good fit. This was seen for the purine and pyrimidine triplex samples at pH 5.5 and 7.4. Only two repeats have been shown for probe **4d** as the third repeat showed a large disparity between the first

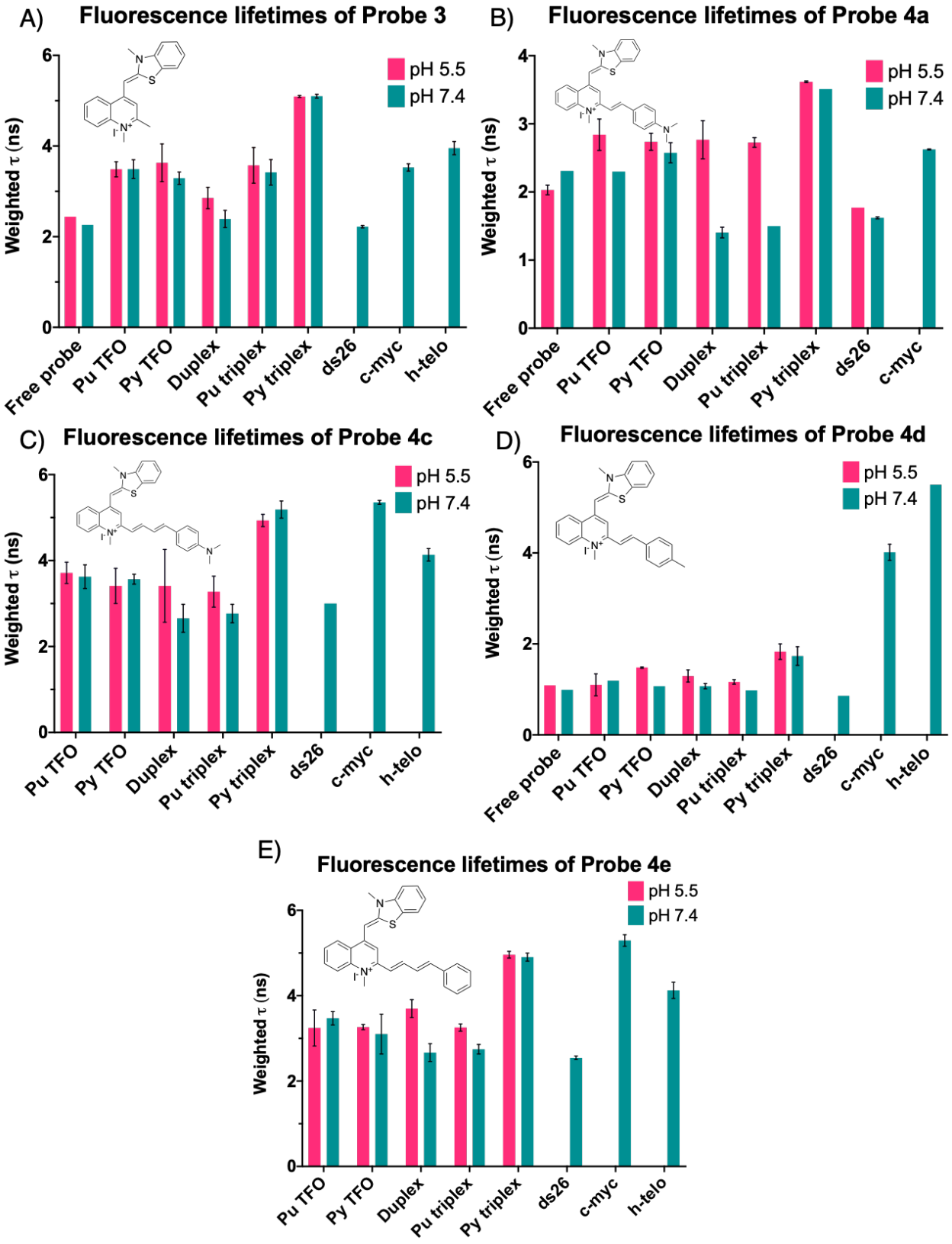


Fig. 4.9: Graphs showing the fluorescence lifetimes (weighted τ) of probes **3** and **4a-e** (2 μM, 10 mM NaCac buffer at pH 5.5 (pink) or 7.4 (blue), 100 mM NaCl and 15 mM $MgCl_2$) when bound to different DNA topologies (10 μM, 5 eq.). The error bars represent 3 separate repeats, with the exception of **4d** which consists of 2 repeats.

two repeats. The first set of measurements resembled that seen for **4c**, with a longer

lifetime seen for pyrimidine triplex at acidic and neutral pH, as was expected from these probes. However, the two subsequent sets of repeats (shown on the graph) show much shorter lifetimes for ssDNA, dsDNA and triplex samples in the order of 1-2 ns. This is more similar to the lifetime of the free probe in solution (τ_w = 0.99-1.09 ns) and suggests the probe was not fully bound for these measurements.

In order to assess whether the probes were encouraging the formation of the pyrimidine triplex at pH 7.4, gel electrophoresis was performed on samples of **55r** (purified dsDNA PCR product obtained by amplification of plasmid **55r** containing $(GAA)_{55}$ repeat insert) were annealed at pH 7.4 with probes **3** and **4a-e**, purified and

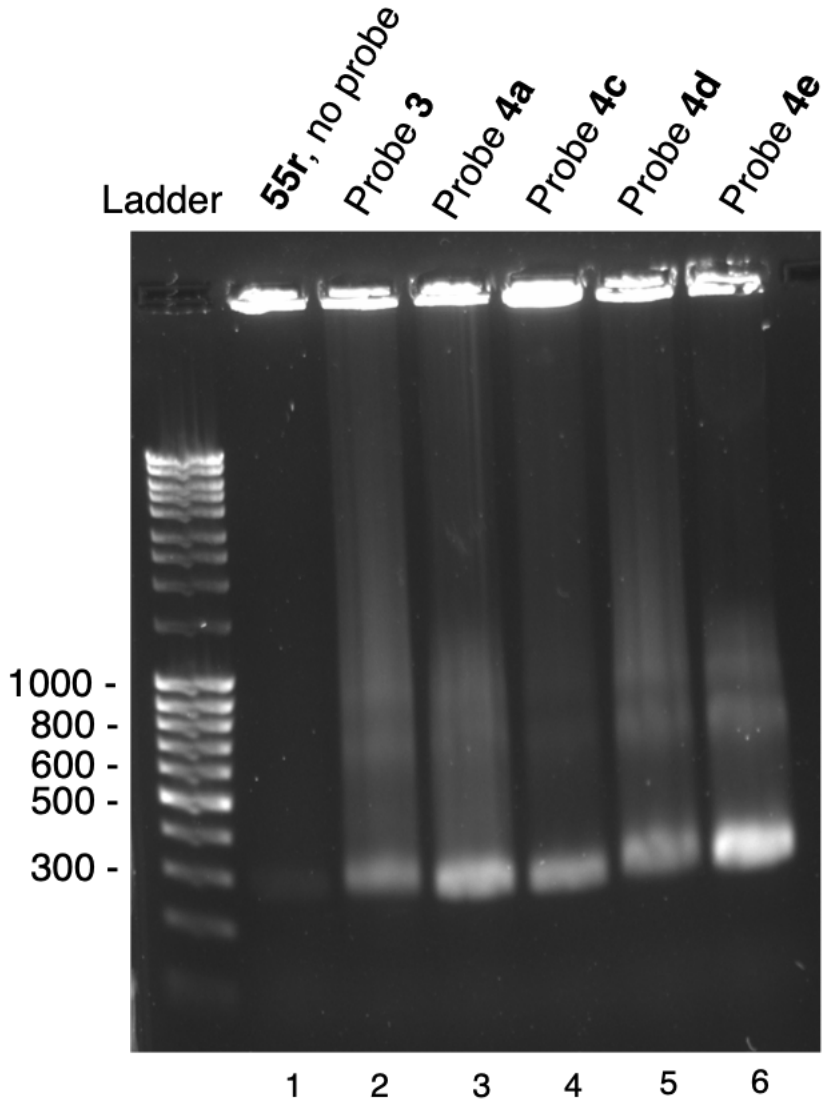


Fig. 4.10: 1.0% agarose gel showing gel electrophoresis of purified PCR product from amplification of **55r** annealed at pH 7.4 with no probe (lane 1), probe **3** (lane 2), probe **4a** (lane 3), probe **4c** (lane 4), probe **4d** (lane 5) and probe **4e** (lane 6).

ran on an agarose gel (Fig. 4.10). If the pyrimidine triplex as forming at pH 7.4 in the presence of the TO-based probes, it might be expected that a heavier band corresponding to the weight of the triplex might be seen on the gel, especially as complexes of TO-triplex DNA have been previously reported to withstand electrophoresis.[130] The result was a clear band at 300 bp as expected from the **55r** dsDNA, seen in the control (lane 1, Fig. 4.10) containing just **55r** annealed at pH 7.4, and also the samples containing the probe with the **55r** dsDNA (lanes 2-6, Fig. 4.10). The samples with the probes also show significant smearing above the band at 300 bp, with faint heavier bands apparent at 700-900 bp that could indicate a secondary structure, though this was inconclusive.

The fluorescence lifetimes of all probes were notably longer when bound to non-B DNA conformations (triplex and G-quadruplex) than ssDNA and dsDNA. This is expected in much the same way that the fluorescence intensity increases dramatically upon binding to non-B DNA in particular. By being completely embedded in a structure by intercalation or stacking, the movement and rotation of the molecule is fully limited. This eliminates the possibility of relaxation through nonradiative pathways and means the compound spends more time in an excited state, increasing the fluorescence lifetime. Similar to fluorescence intensity, the extent to which the fluorescence lifetime of **TO** is increased is correlated with the extent to which rotation within the molecule is limited. As discussed previously, the fluorescence enhancement for probe **3** was largest when bound to pyrimidine triplex followed by G-quadruplex due to the probe being fully embedded or stacked onto the DNA structure, immobilising it and preventing isomerisation. This is in conjunction with the lifetime measurements.

The use of bi- and tri-exponential functions to describe the data indicate several components to the lifetime (τ_1, τ_2, τ_3) with differing contributions (α_1, α_2, α_3), all of which were considered to calculate the weighted lifetime. Each lifetime component suggests a different species; for example, a very short component (τ_1 = 1-2 ns) may correspond to the free probe in solution, and a long component (τ_3 > 5 ns) may correspond to the lifetime when bound to triplex DNA. Taking the measurements for probe **3** when bound to DNA at pH 5.5 as an example, the lifetime for the probe bound to pyrimidine triplex is described with a bi-exponential function and has two lifetime components. The first

(τ_1) is a relatively short lifetime of 2.79 ns and has an α_1 value of 0.34, meaning it contributes 34% to the total lifetime. The second component (τ_2) is a longer lifetime of 5.69 ns with an α_2 value of 0.66, contributing 66% to the total lifetime. This indicates the presence of two species in the sample, one with a shorter lifetime than the other. This could be due to some dsDNA in the sample, or perhaps probe **3** not being fully embedded in the triplex structure and allowing some torsional movement, thus having some extent of nonradiative relaxation. Interestingly, all samples containing possible triplex structure required a bi-exponential function to fit whereas all other conformations required fitting with a tri-exponential function. Attempting to fit a triplex lifetime decay with a tri-exponential resulted in an extremely long (μs) τ_3 component with a small α_3 component of 0.0004, which was sufficient to skew the weighted lifetime to the microsecond timescale.

As with the fluorescence enhancement measurements in the previous section, there is a similarity between the pyrimidine triplex at pH 5.5 and pH 7.4, despite there being no evidence of pyrimidine triplex formation at neutral pH as described in the previous chapters. Previous studies incorporating **TO** onto a TFO and hybridising it to dsDNA reports that the subsequent polypyrimidine triplex formed is stable at pH 7 and above.[142] This effect was shown to be cumulative, with the most stabilisation being observed for triplexes formed from TFOs featuring three **TO** units.[142] This was accompanied by a 20-fold increase in fluorescence of **TO**, indicating that **TO** is intercalating into the triplex structure.[142] This study suggests that **TO** could stabilise pyrimidine triplexes at neutral pH, which would be consistent with the increased fluorescence enhancement and longer fluorescence lifetimes recorded. In the future, this could be further investigated with ^{1}H NMR spectroscopy, by observing potential shifts in the peaks corresponding to the protons on the cytosines in the TFO involved in Hoogsteen hydrogen bonding if the triplex is formed.

From the data, **3** and **4a** would be the most suitable lifetime probes for triplex, as they have longer lifetimes when bound to triplex than all other topologies tested. Probe **4d** also shows promise, as the initial measurements showed a longer lifetime when bound to triplex DNA. Probes **4c** and **4e** both exhibit very similar lifetimes when bound to triplex and G-quadruplex, which would make assigning signals to specifically triplex *in cellulo* difficult.

This data has shown that fluorescence lifetime is a potentially viable strategy for imaging triplex DNA in solution, and that **TO** and **TO-based** probes are a promising basis for this strategy. These probes have already been shown to be cell permeable and have generally good cell tolerability; **4a**, for example, has an IC_{50} = 0.598 μM after 24h.[131] Although this is still toxic to an extent, the timeframe of imaging is sufficiently short to ensure the survival of the cells after incubation with the probes. Hence, the next step is to assess whether these probes can distinguish triplex structures in cells containing the FRDA mutation versus wild-type cells though fluorescence lifetime imaging microscopy.

4.2.5 Fluorescence lifetime imaging microscopy (FLIM)

All FLIM measurements were performed by Dr Peter Summers, Imperial College London.

A method to image triplex structures in cells would be greatly beneficial in order to confirm whether the formation of triplexes play a part in the silencing of the *FXN* gene in FRDA. The promising *in vitro* results reported in the previous section indicate that **TO-based** probes could be suitable for this purpose through use of fluorescence lifetime imaging. For this, time correlated single photon counting fluorescence lifetime imaging microscopy (TCSPC-FLIM or FLIM) was used. FLIM records an image, with each pixel of the image being a measure of a fluorescence lifetime decay. These individual lifetimes can be fitted and produce a distribution of lifetimes across the image.

The cells used were HEK293; both wild-type (**WT**) and containing the FRDA repeat expansion mutation with $(GAA)_{310}$ repeats (**mut**). The cells were incubated with probes **3** (2 and 5 μM), **4a** (5 μM) and **4d** (5 μM) for 1h before imaging. This is sufficient time for the probes to permeate the cell. Probe **4a** has been previously reported for live cell imaging, and the IC_{50} found to be 0.6 μM in PC3 cells (after 24 hours at a concentration of 5 μM). The cells were inspected after incubation and found to be sufficiently healthy. Cells were imaged under the same conditions as incubation (37 •C, same CO_2 level). For each well, three standard images and one high resolution image were taken (as described in Methods 6.5.2). Images were analysed using

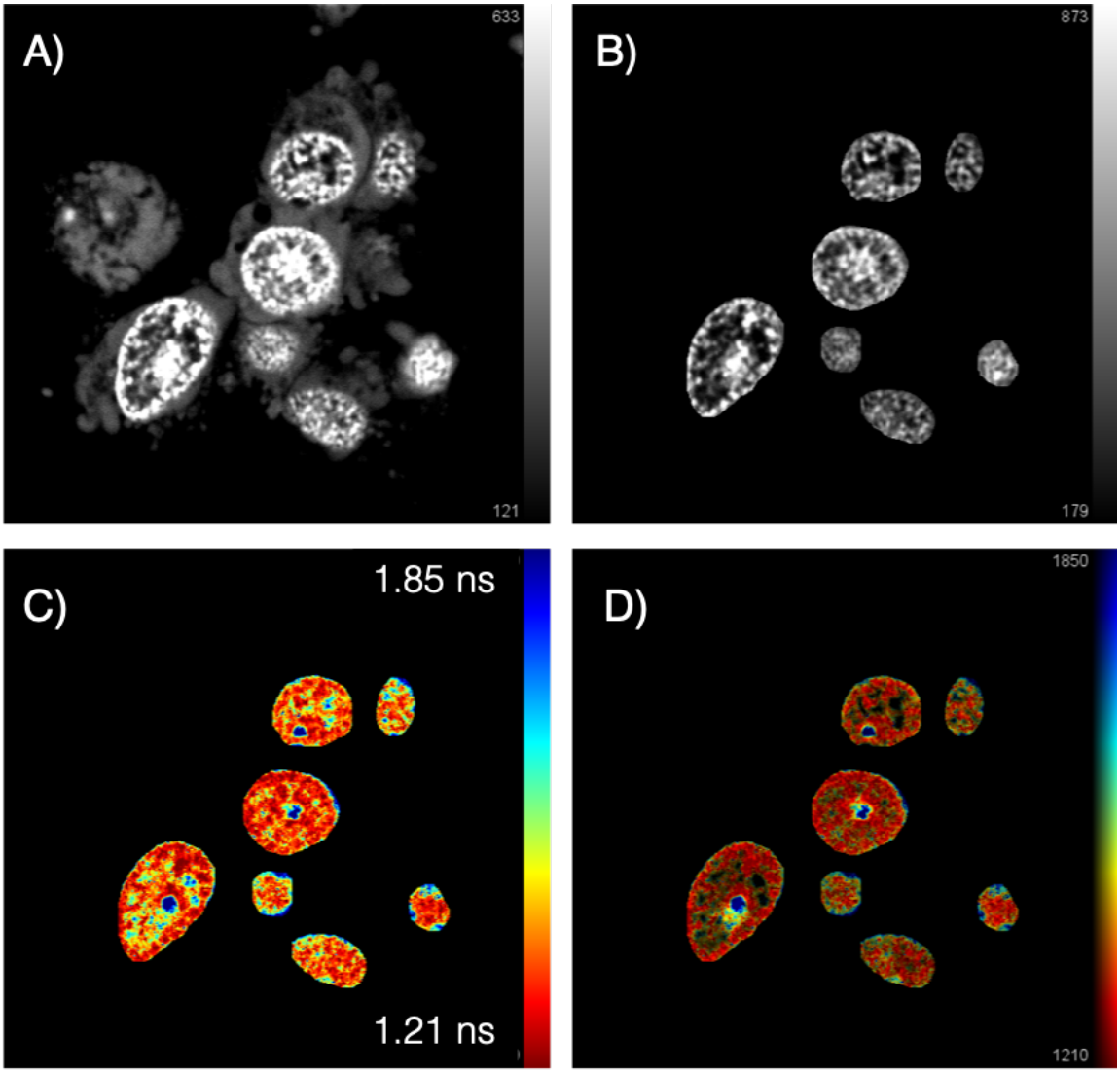


Fig. 4.11: Example FLIM images of probe **3** (5 µM) incubated with **WT** cells (~30,000). A) Fluorescence intensity, B) fluorescence intensity within segmented nuclei, C) fluorescence lifetime within segmented nuclei, D) fluorescence intensity and lifetime within segmented nuclei merged. Pockets of longer lifetime can be seen in the nucleoli.

FLIMfit (5.1.1), from which information about the fluorescence intensity, lifetime and goodness of fit were extracted. For each image, the cell nuclei were separated using the segmentation tool (Fig. 4.11B), to eliminate readings from the cytoplasm. Data from the nuclei was collected and lifetime parameters were all fit to a bi-exponential function across all images. For each image a correlation between the fluorescence intensity and the fluorescence lifetime was also plotted to assess whether there was any intensity-dependent effect on the lifetime (Fig. 4.11, D). The average weighted lifetime of each nuclei was calculated, and from that an overall average weighted lifetime for each image, eliminating the points with very low fluorescence intensity that

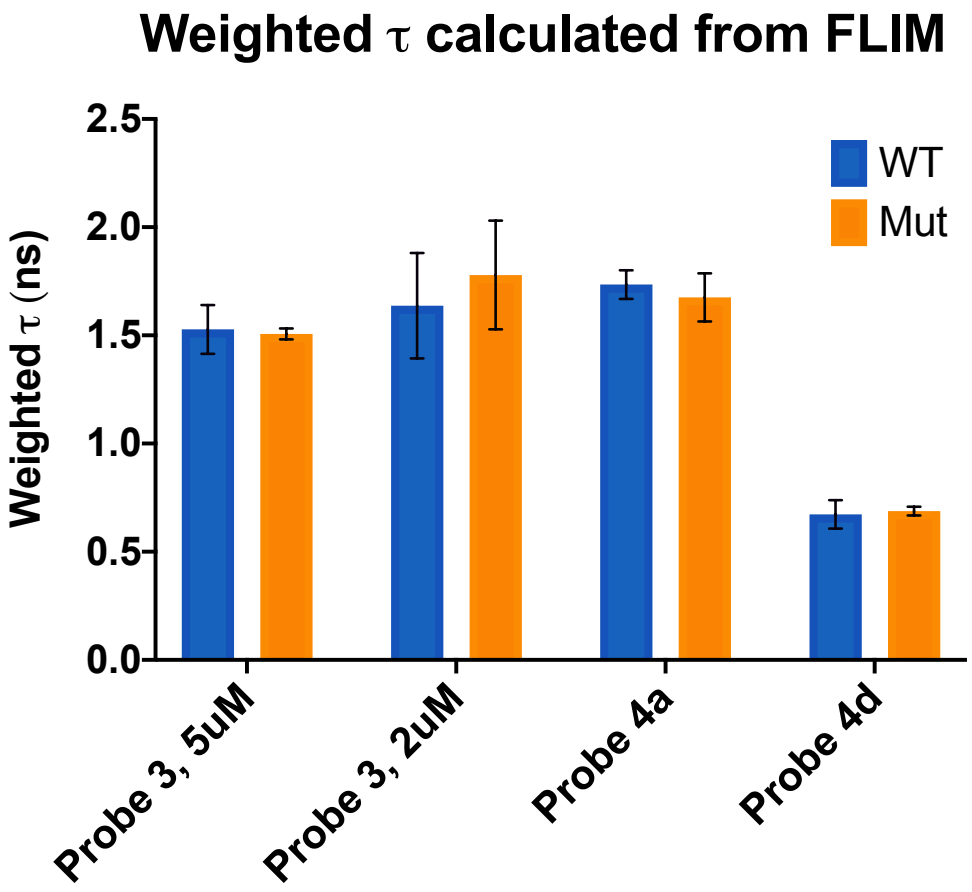


Fig. 4.12: The average weighted fluorescence lifetime (weighted τ) of probes **3**, **4a** and **4d** incubated with HEK293 cells, **WT** and **mut**. Lifetime was calculated as a weighted average across 4 images, and all nucleic within each image, and fitted to a bi-exponential decay.

exhibited skewed lifetimes. These values were then compared between **WT** and **mut** for each probe.

It can be seen from the final data that there is no significant difference in overall average weighted lifetime between the **WT** and the **mut** cells for any of the probes tested (Fig. 4.12). The data is consistent for each probe across both cell types. Probe **3** (at 5 μM) exhibits an average weighted lifetime in **WT** cells of $\tau_w = 1.53 \pm 0.14$ ns, and an average weighted lifetime in **mut** cells of $\tau_w = 1.51 \pm 0.03$ ns. Probe **3** (at 2 μM) exhibits an average weighted lifetime in **WT** cells of $\tau_w = 1.64 \pm 0.25$ ns, and an average weighted lifetime in **mut** cells of 1.78±0.28 ns. The lifetime is generally longer at the lower concentration; this is consistent with observations of the correlation between fluorescence intensity and lifetime (Fig. 4.13, B). Often, at low intensities, there is a

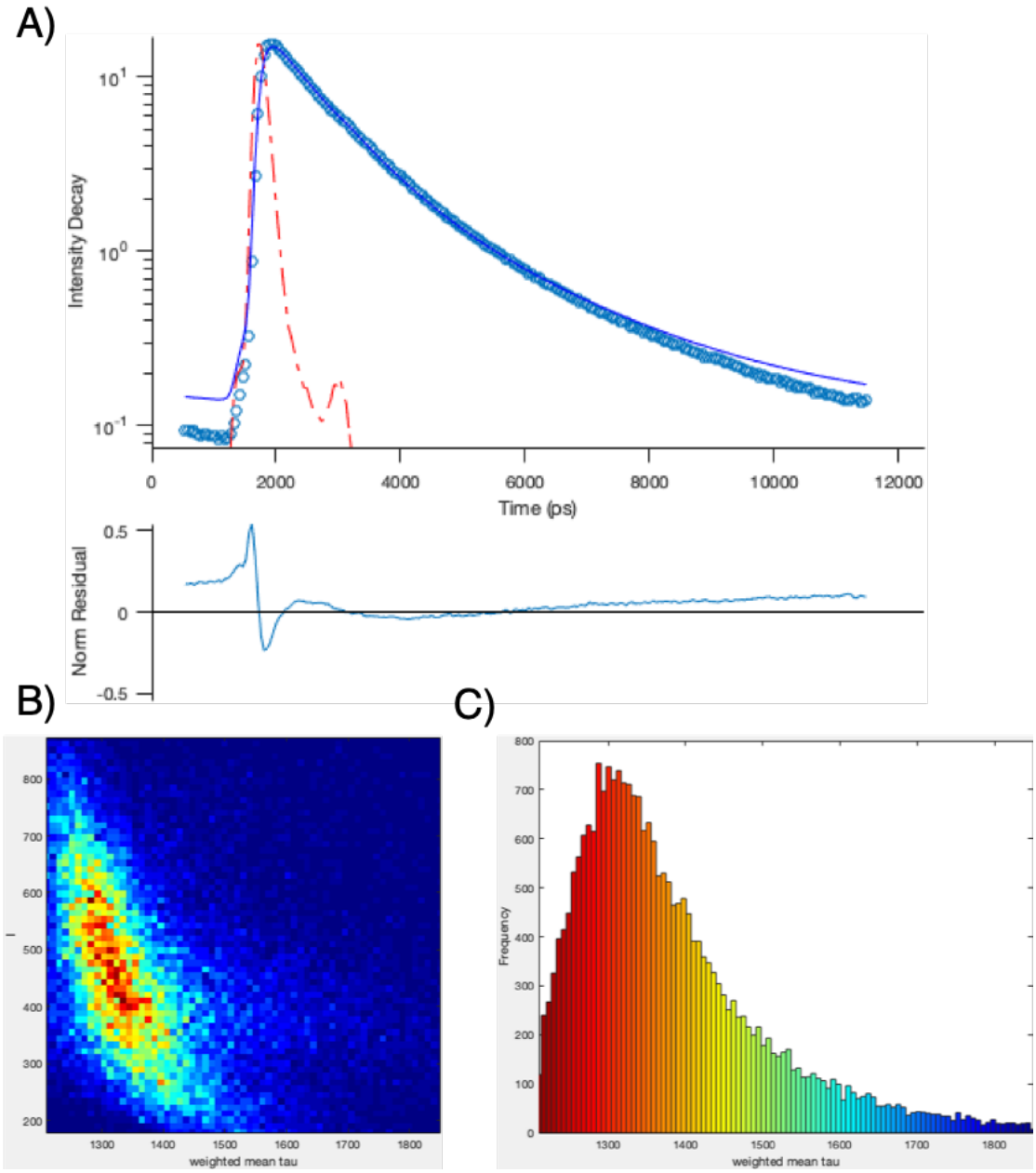


Fig. 4.13: FLIM parameters extracted from an image of probe **3** (2 µM) incubated with **WT** cells (~30,000). A) Representative decay fitted with a bi-exponential, and measure of goodness of fit below, B) correlation between intensity and weighted average lifetime throughout the pixels in the image, C) histogram representing the distributions of fluorescence lifetimes across the image.

skew of the data towards longer lifetimes. At lower photon counts, the offset, which accounts for the background noise, is lower and this can lead to an overestimation of lifetime. The lower concentration used for the sample of probe **3** at 2 µM results in lower fluorescence intensity, which may result in slightly longer lifetimes (Fig. 4.13, B). Probe **4a** (at 5 µM) exhibits an average weighted lifetime in **WT** cells of $\tau_w = 1.74 \pm 0.06$

ns, and an average weighted lifetime in **mut** cells of τ_w = 1.68±0.11 ns. This is slightly longer than probe **3** at the same concentration, though the fluorescence intensity of **4a** is considerably lower than **3** both *in vitro* and *in cellulo.* Probe **4d** (at 5 μM) exhibits an average weighted lifetime in **WT** cells of τ_w = 0.67±0.06 ns, and an average weighted lifetime in **mut** cells of τ_w = 0.69±0.02 ns. This is much lower than probes **3** and **4a** and is close to the lifetime of the free probe in solution (τ_w = 1.04±0.05 ns) recorded with TCSPC as reported in the previous section.

Probe **3** is structurally similar to **TO**, and this is represented by the strong fluorescence intensity both *in vitro* and *in cellulo*. The average weighted lifetimes are very similar across both cell lines, and consistent between images. It would be expected that the lifetimes of the probes in the nuclei would be close to that measured by TCSPC when bound to dsDNA *in vitro*. The lifetime of probe **3** when bound to ds26 at pH 7.4 in solution was τ_w = 2.22±0.02 ns; the disparity in the *in vitro* and *in cellulo* data could be due to non-specific binding to other biomolecules in the nucleus such as ssRNA, or the presence of free unbound probe, which could account for the shorter lifetime. At a lower concentration (2 μM), **3** displays a slightly longer lifetime in comparison to the sample with increased concentration (5 μM). This is likely due to the lower photon count as described before, lowering the offset and overestimating the lifetime (Fig. 4.13, B).

Probe **4a** is similar in structure to **4d**, but with a dimethylamine group rather than a methyl group. This probe appears to bind well to DNA, triggering a large increase in fluorescence intensity and discernible change in fluorescence lifetime between topologies. The *in vitro* experiments indicate when bound to ds26 at pH 7.4, **4a** has an average fluorescence lifetime of τ_w = 1.62±0.01 ns. This is consistent with the *in cellulo* data (τ_w = 1.74±0.06 ns and 1.68±0.11 ns for **WT** and **mut** respectively), suggesting that **4a** is intercalated into dsDNA in the nucleus.

Probe **4d** measured a significantly lower weighted average lifetime in both the **WT** and **mut** cell lines. This is similar to the *in vitro* experiments described in the previous section, as for two of three repeats it appeared that the probe was not bound to DNA at all. The *in vitro* studies showed that the average lifetime of the free probe at pH 7.4 was τ_w = 0.99 ns, and the presence of DNA did not increase this by a large

amount, indicating some interaction but most likely non-specific. This was also possibly an overestimation, as the emission of the free probe was so low that the baseline was skewed by increased contribution of the background noise. The low average lifetime observed in the nuclei indicates **4d** is not specifically bound. The general fluorescence intensity recorded is also around 10% of the intensity for **3** and **4a**; another indication that **4d** is not intercalated into the DNA. The correlation between the intensity and the average weighted lifetime was visibly skewed towards a longer lifetime at lower intensities. Although the lowest intensity points with the worst skewing effect were removed during analysis, the effect was most pronounced for **4d** and may have led to an overestimation of the average lifetime *in cellulo*.

It might have been expected that the overall fluorescence lifetime of the probes within the **mut** cells would be longer than that of the **WT** cells, due to the presence of the $(GAA)_{310}$ repeat expansion mutation. It was hypothesised that this expansion mutation could cause the formation of either a triplex or R-loop structure, or the condensation of chromatin: the gene-silencing mechanisms discussed previously.[9] From the *in vitro* fluorescence lifetime data presented in the previous section, it could be suggested that this might increase the fluorescence lifetime of the probes tested. This would have been observed as a slight increase in the overall average lifetime of the probe within the **mut** cell line in comparison to the **WT** cell line. This may have provided evidence that a triplex-like secondary structure was forming in the **mut** cells. However, an increase in the average fluorescence lifetime was not observed. This could be due to the comparatively small proportion of cellular DNA involved in the mutation, at just 930 base pairs.

Across both cell lines, there are defined circular pockets of significantly longer lifetimes (indicated in blue, Fig. 4.11, C), probably corresponding to the nucleoli. The nucleolus is the site of ribosomal RNA synthesis and ribosome assembly and has also been implicated in chromatin organisation.[192] Similar probes based on **TO** conjugated with styryl moieties were previously developed and found to show better staining of nucleolus RNA than existing commercial stains.[193] Also, in the study that developed these probes as selective G-quadruplex imaging agents, it was found that they strongly stained the nucleoli in human pancreatic cancer cells, exhibiting increased fluorescence intensity as seen upon binding to quadruplexes *in vitro*.[131] The nucleolus

is the site of transcription of rDNA (ribosomal DNA) which contains G-rich sequences that have been previously reported to potentially form G-quadruplexes.[194] This could suggest that probes **3** and **4a** are binding to G-quadruplexes formed from rDNA and rRNA in the nucleolus, causing increased fluorescence intensity and longer fluorescence lifetime, similar to that observed *in vitro* for quadruplexes c-myc and h-telo. As well as the nucleolus, there are 'speckles' of longer lifetimes within the main body of the nucleus also, evident in the higher resolution images (Fig. 4.11, C. This area is the location of the DNA packaged around histones with chromatin, and where transcription occurs. It is possible that the pockets of longer lifetime indicate secondary structures formed during transcription of particular genes, or regions of condensed chromatin.

Although no difference in the average weighted fluorescence lifetime was seen between the two cell lines tested, **WT** and **mut**, fluorescence lifetime imaging and FLIM presents an interesting tool for the study of higher order DNA structures *in vitro* and *in cellulo*. This technique eliminates the need for a specific, high-affinity probe, allowing imaging of distinct environments within the cell using a non-specific stain. From this data, no conclusions can be drawn as to whether there are triplex structures or similar playing a part in the transcriptional silencing of the *FXN* gene in FRDA.

4.3 Summary and future outlook

Probes based on the structure of **TO** were successfully synthesised and purified using procedures taken and improved from literature. These probes contained the scaffold of **TO** with styrene-like substituents incorporated at the ortho position to the carbon-carbon single bond connecting the two aromatic moieties. It was rationalised that these substituents would provide another planar aromatic surface with which to stack between base triads in the triplex structure, some with a terminal dimethylamine group to introduce hydrogen bonding ability. The binding of these probes to several DNA topologies was tested, including ssDNA, dsDNA, purine and pyrimidine triplex, and G-quadruplex. Absorption and emission titrations were performed in order to determine the fluorescence enhancement of each probe upon DNA binding, and also a suitable stoichiometry for the lifetime measurements that would ensure no free probe

was in solution. Probe **4b** was found to have some solubility issues and hence at measurement concentrations there was not enough probe in solution to produce a fluorescence signal or bind to the DNA, so this probe was eliminated from further investigation. The remaining probes **3** and **4a-e** exhibited large fluorescence enhancement upon binding to triplex and G-quadruplex, especially pyrimidine triplex at pH 5.5 where the triplex structure is very stable. Large fluorescence enhancement was also seen for the pyrimidine triplex annealed at pH 7.4, which was interesting as the pyrimidine triplex structure was not seen to form under these conditions as reported in the previous chapters. This indicated that the probes were potentially stabilising the formation of the triplex. The fluorescence lifetimes of the probes when bound to different DNA topologies was also measured. It was found that probe **3** and **4a** had longer lifetimes when bound to pyrimidine triplex at acidic and neutral pH than any other topology, including G-quadruplex. Probes **4c** and **4e**, also displayed promising fluorescence properties, though their lifetimes when bound to G-quadruplex and triplex were very similar which would make them difficult to distinguish *in cellulo.* Probe **4d** exhibited great inconsistency between repeats, with one set of measurements indicating long lifetime when bound to triplex, but the following two sets of measurements suggesting the probe was not bound at all, with a very short lifetime consistent with that of the free probe in solution. Again, the lifetime measurements indicated a pyrimidine triplex structure was forming at pH 7.4. To investigate whether the probes were encouraging triplex formation at neutral pH, the probes were annealed with **55r** (purified from PCR amplification) at pH 7.4 from the previous chapter and ran on an agarose gel. There was large amount of smearing in the lanes containing the DNA with the probes, with some faint heavier bands which could indicate secondary structure. An ^{1}H NMR experiment would be useful to confirm this by assessing whether the TFO is interacting with the dsDNA and whether the probes are involved in this interaction. ^{1}H NMR shifts corresponding to protons involved in hydrogen bonding between bases will be shifted in comparison to the same protons when not involved in hydrogen bonding.

The FLIM of the **WT** and **mut** HEK293 cell lines with probes **3**, **4a** and **4d** did not provide any evidence as to whether there were triplex structures present in the cells containing the $(GAA)_{310}$ mutation, or whether they played a role in the transcriptional

silencing of *FXN*. The probes, however, were cell permeable and non-cytotoxic within the imaging timeframe at imaging concentrations, and afforded homogenous cell staining with varying fluorescence lifetime between compartments of the nucleus. Hence, this technique is still a promising strategy for elucidating location of higher order nucleic acid structures within cells. Larger sample sizes and more high-resolution images, coupled with transcriptional inhibition assays, could shed light on the nature of gene silencing in FRDA and other disorders.

5 Outlook and future work

The work presented in this thesis aimed to study and biophysically characterise intermolecular triplexes formed from oligonucleotides containing the FRDA repeat expansion mutation of varying lengths and present a strategy for the detection of triplexes *in cellulo* by using fluorescence lifetime probes.

Firstly, in Chapter 2, short synthetic oligonucleotides containing the $(GAA)_n$ sequence (where n = 5, 10 and 20) were synthesised, and the formation of triplex structures was investigated under varying conditions of pH and salt concentration. There were two possible triplex conformations: purine and pyrimidine. It was found that the pyrimidine triplexes formed readily at pH 5.5 from all three lengths of oligonucleotides investigated, as seen by gel electrophoresis and CD spectroscopy. Purine triplexes, however, were not detected under any conditions by gel electrophoresis or CD, even under molecular crowding conditions introduced by PEG_{200}. FRET melting and fluorescence titrations with a FAM-labelled TFO and a TAMRA-labelled duplex indicated an association between the TFO and the duplex, with the intensity of TAMRA fluorescence increasing with added FAM-labelled TFO. However, this could be due to non-specific FRET and hence not clear evidence of purine triplex formation. It was thought that the oligonucleotides containing the $(GAA)_{5-20}$ mutation were too short to form higher order nucleic acid structures under physiological conditions, as the pathological length of $(GAA)_n$ repeats is approximately n = 66. Thus, a molecular biology approach was taken to synthesise long polynucleotides containing repeat mutations of pathological and near pathological length.

In Chapter 3, plasmids were designed containing the $(GAA)_n$ mutation at n = 55 and 200, flanked by restriction enzyme sites and polymerase promoters for extraction and amplification of the insert respectively. A mutation length of 55 repeats (**55r**) was 'pre-mutation'; much longer than normal but not quite at the pathological limit. A length of 200 repeats (**200r**) was fully pathological. When growing the plasmids in bacteria, it was found that **200r** experienced large deletions of the repeat sequence, as seen by sequencing and gel electrophoresis following a restriction digest. Attempting to amplify the 200 repeat insert using PCR with high fidelity enzymes resulted in no full-length product, with a large smear being seen by gel electrophoresis. Several PCR conditions were investigated, including addition of DMSO which abolishes secondary structures formed in primers and templates, however the **200r** insert could not be amplified. It was suggested that this pathological length of repeat was so unstable that that polymerase was slipping along it, resulting in the smear of different sized products observed. The bacterial clone with the largest repeat insert was n = 75 (**75r**), which was interestingly just above the pathological limit. This clone was successfully amplified using PCR. The smaller plasmid, **55r**, was grown in bacteria and amplified by PCR without issue. Secondary structure formation under various conditions was investigated. Multiple bands were observed by gel electrophoresis when the DNA was annealed in buffer at acidic pH, indicating the presence of at least two higher order structures. It was suggested that these could be triplexes and 'sticky DNA'; a novel structure formed by the association of two or more triplexes containing (GAA) repeats. Some very faint heavy bands were also observed upon annealing at physiological pH. The formation of DNA-RNA hybrid structures was also investigated. RNA was synthesised from **55r** and incubated with the dsDNA from the **55r** PCR product at acidic and physiological pH. Under both pH conditions, multiple heavy bands were observed by gel electrophoresis, suggesting higher order structures. AFM was used to elucidate the structures of the species responsible for the bands seen on the gel from the DNA-DNA samples. Images taken of **55r** DNA annealed at acidic pH showed possible triplex structures, apparent from the thicker, more rigid regions that were 0.5-0.9 nm taller than the surrounding dsDNA. These regions were not observed in AFM images of **55r** DNA annealed at physiological pH, and thus this supported the previous findings.

In order to elucidate whether triplex DNA plays a role in gene silencing in FRDA, and whether they are present in cells containing the $(GAA)_n$ repeat mutation, a strategy for detecting them *in cellulo* would be beneficial. Fluorescent DNA-binding small-molecule probes were chosen for this due to their tuneable properties, ease of detection and cell permeability. In Chapter 4, six probes based on the structure of thiazole orange (TO) were synthesised, each with different substituents. Effects of binding to different topologies of DNA (ssDNA, dsDNA, triplex, G-quadruplex) at acidic and physiological pH were investigated by UV-Vis absorption and fluorescence emission titrations. It was found that the fluorescence intensity increased dramatically upon binding to dsDNA, but even more for triplex and G-quadruplex. However, these probes are not specific for triplex DNA, and thus standard fluorescence could not be used as a specific detection method. Fluorescence lifetime imaging was proposed as an alternative, as it eliminates the need for a specific probe as long as the probe exhibits a distinct fluorescence lifetime when bound to different DNA topologies. It was found that the probes did indeed exhibit a longer fluorescence lifetime when bound to triplex DNA in comparison to dsDNA. Generally, a longer lifetime was also observed for G-quadruplex, though with probes **3** and **4a** this was not as long as when bound to triplex. Consequently, probes **3**, **4a** and **4d** (for comparison) were tested *in cellulo.* The probes were incubated HEK293 cells, wild-type (**WT**) and with a $(GAA)_{310}$ repeat mutation (**mut**), and FLIM was used to measure the fluorescence lifetime of the probes within the cell nuclei. It was hypothesised that the average lifetime of the probes in the **mut** cells might be slightly longer than the **WT** cells if there was higher order structure formation from the (GAA) repeats, or significant condensation of chromatin. However, the average lifetimes were the same across both cell lines for each probe, showing no discernible difference. Despite this, the data was consistent and demonstrated the potential of this technique for *in cellulo* imaging of biomacromolecules. Pockets of long lifetimes coincided with the nucleoli, which have been previously reported to containing G-rich rDNA with the propensity to form G-quadruplexes. Hence, these probes could be investigated in the future for quadruplex detection in cells.

The biophysical studies presented in Chapter 2 indicate that a triplex structure is forming only under acidic, but not physiological, conditions. It was hypothesised that perhaps a weak triplex was forming under neutral conditions and it was too transient

and unstable to be detected by gel electrophoresis. This could be investigated further with ^{1}H NMR. Protons involved in hydrogen bonding will be shifted when compared to the unbound version, and hence this could be used to assess whether the DNA bases of the TFO is interacting with the duplex through hydrogen bonding. The pathological length of $(GAA)_n$ repeats for FRDA is n = ~66, and thus it would be unexpected for unusual DNA structures to form under physiological conditions at repeat lengths shorter than this such as the sequences investigated in this chapter.

The biophysical studies of the pathological sequences described in Chapter 3 present some interesting findings. For these longer sequences, bands corresponding to heavier structures were seen when samples containing a DNA duplex with an RNA TFO were annealed or incubated at physiological pH, as well as acidic pH. This is strong evidence that sequences containing the $(GAA)_n$ repeat can form DNA-RNA hybrid structures under physiological conditions. This supports the hypothesis of R-loop/H-loop formation in FRDA, which has been previously reported to block transcription in plasmids.[182,183] These structures could be further investigated with AFM, as was reported here for the DNA-DNA structures.

The *in vitro* fluorescence lifetime data presented in Chapter 4 suggested that perhaps the probes were stabilising pyrimidine triplex formation at neutral pH, as well as acidic pH, since the lifetime values under both conditions were similar. This was also investigated with gel electrophoresis, to elucidate whether a band corresponding to a triplex structure could be seen at pH 7.4 after incubation with the probes. This was inconclusive, however. The phenomenon could further be investigated using CD to determine whether a triplex is forming at neutral pH in the presence of the probes.

The data presented in this thesis sheds further light on the nature of higher order structures formed at repeat regions, which have been previously hypothesised to a play a large part in gene expression across the genome. There is still no cure for FRDA but understanding a little more about the basis of this rare disease on a molecular level may aid in eventually developing a treatment. Fluorescence lifetime imaging has been presented as a versatile and useful tool to overcome specificity issues in DNA imaging, with applications *in vitro* and *in cellulo*.

6 Experimental methods and protocols

6.1 General methodology

All oligonucleotides were purchased from Eurogentec as a lyophilised stock, or synthesised using an ABI Oligo Synthesiser, Model 39, using a solid CpG support and phosphoamidites purchased from Sigma. Synthesised oligos were purified using illustra NAP-10 columns.

All commercial reagents and solvents were purchased from Sigma Aldrich, VWR, Fluorochem and Fisher Scientific, unless otherwise stated.

NMR spectra were obtained on a Bruker Advance Ultrashield 400 MHz NMR spectrometer at 295 K, unless otherwise stated.

Electrospray (ESI) mass spectra were obtained on a Waters LCT Premier (ES-Tof)/Acquity i-Class spectrometer by Dr. Lisa Haigh, Imperial College London.

C,H,N elemental microanalyses were run by Mr. Alan Dickerson, University of Cambridge.

6.2 Nucleic acids *in vitro* protocols

6.2.1 Preparation of buffer solutions for nucleic acid experiments

Milli-Q water (18 Ω) was used to prepare all buffer solutions for nucleic acid experiments. Each buffer was prepared from a double concentrated (40 mM) stock of sodium or lithium cacodylate solution at pH 5.5 or pH 7.4 and dilution of a concentrated stock (1 M) of the appropriate metal cation chloride in milli-Q water. Buffers were filtered with a 0.2 μm filter and stored at 4 •C until use.

6.2.2 *Annealing of DNA for triplex studies*

Oligonucleotides were reconstituted in milli-Q water to the desired concentration. These were then diluted in sodium cacodylate buffer (10 mM, pH 5.5 or pH 7.4) with an appropriate concentration of metal cation chloride salt (NaCl and $MgCl_2$) in milli-Q water to the necessary concentration for the experiment. The oligonucleotide solution was heated to 95 •C for 5 minutes and then cooled slowly to room temperature overnight. Solutions were then stored at 4 •C until analysis.

6.2.3 *Gel electrophoresis – short oligonucleotides*

Polyacrylamide gels (12-15% acrylamide/bis-acrylamide, 40%) were used to analyse short oligonucleotides. These gels were cast containing TBE buffer (for pH 7.4) and TAE buffer (for pH 5.5) with the appropriate concentration of metal cation chloride salt in milli-Q water at a plate width of 1 mm. Gels were pre-run for 30 minutes in a mini-Protean (Bio-Rad) tank in 1x TBE buffer (for pH 7.4) or 1x TAE buffer (for pH 5.5). Samples were loaded onto the gel with 30% glycerol and alongside an appropriate ladder. The gel was run for 120 minutes at 4 •C and stained with SybrGold before being imaged using an Ettan DIGE Imager scanner (GE Healthcare, UK).

6.2.4 *Circular dichroism*

DNA was annealed as described in (6.2.2) to a total concentration of 5 μM in sodium cacodylate buffer (10 mM) at pH 5.5 or pH 7.4 containing the appropriate concentration of metal cation chloride salt in milli-Q water, with a total volume of 500 μL. CD spectra were collected on a JASCO-715 spectrometer at 20 •C with a slit width of 2 nm and a step-size of 0.1 nm, between 200 and 600 nm. A baseline spectrum of just the buffer was first recorded, and this was subtracted from the final spectra. CD melting studies were performed in a similar manner; however, the temperature was increased/decreased between 20 and 95 •C at a rate of 0.5 •C/min, using a Peltier Controller.

6.2.5 *pH stability studies*

Oligonucleotides were annealed (6.2.2) at a total concentration of 10 μM in sodium cacodylate buffer (10 mM, pH 5.5) containing the appropriate concentration of metal cation chloride salt in milli-Q water, with a total volume of 500 μL. Additional sodium cacodylate buffers of varying pH were made by titrating NaOH (1 M, aq.) into sodium cacodylate stock solution (40 mM) until the desired pH value was reached (a range of pH 6-8), and the final sodium cacodylate concentration was adjusted to 20 mM by addition of milli-Q water. To the annealed samples, 500 μL of each variable pH buffer was added, and the final pH was measured. The samples were then incubated at room temperature overnight, before being analysed by gel electrophoresis (6.2.3) under neutral pH conditions, with 1x TBE running buffer.

6.2.6 *Crowding conditions*

Oligonucleotides were annealed as described above (6.2.2) at a total concentration of 5 μM in sodium cacodylate buffer (10 mM, pH 7.4) containing the appropriate concentration of metal cation chloride salt in milli-Q water and varying concentrations of PEG_{200}, with a total volume of 500 μL. Total PEG_{200} concentration ranged from 10-40% (v/v), and the annealed samples were analysed by gel electrophoresis (6.2.3) and CD spectroscopy (6.2.4) as described above.

6.2.7 *FRET melting studies*

For FRET melting experiments, oligonucleotides containing 6-carboxyfluorescein (FAM) and 5-carboxytetramethylrhodamine (TAMRA) labels were purchased from Eurogentec and received as lyophilised stock. Stocks were reconstituted in milli-Q water to a concentration of 100 μM, as determined by UV-Vis spectroscopy (absorption at 260 nm) and stored at -20 •C until use. The duplex strand containing $(GAA)_5$ was TAMRA-labelled, and the purine $(GAA)_5$ TFO was FAM-labelled. The sequences used can be seen in Table 2.1, Chapter 2. Oligonucleotides were diluted in sodium cacodylate buffer (10 mM, pH 7.4) to a total concentration of 0.4 μM and annealed as described above (6.2.2). After annealing, samples were added into 0.2 mL PCR-tubes with a total volume of 40 μL. For FRET melting experiments, an RT-qPCR instrument (Agilent Mx3005P) was used. The sample was

irradiated at FAM λ_{ex} of 494 nm and TAMRA emission was followed at 580 nm. Temperature was increased/decreased from 20-95 •C with melting/cooling being carried out at a 1 °C/min rate, and data collected at 1•C intervals. Data was normalised (0 to 1).

6.2.8 FRET emission titrations

DsDNA was annealed in sodium cacodylate buffer (10 mM, pH 7.4) as described above (6.2.2), with the $(GAA)_5$-containing strand TAMRA-labelled, with a final total DNA concentration of 5 μM and a final volume of 500 μL. A solution of purine $(GAA)_5$ TFO labelled with FAM in sodium cacodylate buffer (10 mM, pH 7.4) was diluted to a concentration of 50 μM, and titrated into the dsDNA sample with increasing equivalents from 0 to 2. The sample was irradiated at the FAM λ_{ex} of 494 nm and the emission spectra recorded between 500-700 nm. The final sample containing 2 eq. of the purine TFO to the dsDNA was annealed as described above (6.2.2) and the emission spectrum was recorded again the following day.

6.3 Molecular biology techniques

6.3.1 Plasmid design

Plasmids were designed and ordered from Genscript. Plasmids contained a Puc57 backbone, with the repeat sequence of interest (GAA_{55} or GAA_{200}) cloned into the multiple cloning site (MCS). This sequence was flanked by transcription promoter sites for T7 and T3 *in vitro* transcription polymerases, PCR/seqeuncing primer sites M13F (5'-GTAAAACGACGGCCAGT-3') and M13R (5'-GTCATAGCTGTTTCCTG-3') and restriction enzyme digest sites for BamHI and HindiIII. Plasmids also contained an Ampicillin resistance gene, and were grown by Genescript in *E.coli* TOP10 bacterial cells. Plasmids were received as a lyophilised stock and were reconstituted in milli-Q water to a final concentration of 40 ng/μL.

6.3.2 Plasmid DNA digestion with restriction endonucleases

Plasmid DNA was digested using restriction enzymes alongside the manufacturers' recommended 10x restriction digest buffers. Between 500 ng and 1 μg

of plasmid DNA was digested overnight at the appropriate temperature, with 1x suitable restriction buffer and 1x BSA. Generally, 2 units of enzyme were used for each 1μg of DNA, with the volume of enzyme not exceeding the recommended 10% of the total reaction volume. Double digests using restriction enzymes were performed in a mutually compatible buffer following the same protocol. Digested DNA was then purified by phenol:chloroform extraction as described below (6.3.3).

6.3.3 Phenol:chloroform extraction of DNA from proteins

Proteins, such as restriction enzymes, were removed from aqueous solutions of DNA by extraction with equal volumes of phenol and chloroform, containing a small amount of glycogen to assist in visualisation of the pellet. The solution was centrifuged at 12,000 g for 15 minutes at room temperature to separate the phases. The proteins were left at the interface between the aqueous and organic phases, and the upper aqueous phase containing the DNA was retrieved. The solution was then extracted with equal volumes of chloroform to remove traces of phenol, and the DNA was recovered by ethanol precipitation as described below (6.3.4).

6.3.4 Ethanol precipitation of DNA and RNA

DNA and RNA in low concentration in salt solutions following PCR amplification, *in vitro* transcription or secondary structure annealing were purified/concentrated by ethanol precipitation. Solutions were mixed with one tenth volume of sodium acetate (NaOAc, 3 M, pH 5.0) followed by three volumes of 100% ethanol. This solution was then chilled at -20 •C for 30 minutes, and subsequently centrifuged at 15,000 g for 30 minutes at 4 •C. The supernatant was removed, the pellet was washed once with cold 70% ethanol (500 μL), the solution was centrifuged at 12,000 g for 20 minutes and the pellet was air dried after the supernatant had been removed. This process was repeated for DNA and RNA with downstream processes other than gel analysis. The DNA pellet was resuspended in an appropriate volume of EB buffer (QIAGEN) and the concentration was determined using a Nanodrop One (Thermo Scientific).

6.3.5 *Gel electrophoresis – long sequences*

Agarose gels (0.8-1%) were used for longer DNA sequences isolated from plasmid DNA. These gels were cast containing 1x TBE buffer with 1x SybrSafe or ethidium bromide (EtBr) for imaging. Samples were loaded containing 1x Loading Dye (with 30% glycerol) and ran against a MassRuler ladder (Thermoscript). Gels were run for 75 minutes in 1x TBE buffer before being imaged using a Bio-Rad Molecular Imager® Gel Dock™ XR+ Imaging System.

6.3.6 *Transformation of plasmids in bacteria*

The *E. coli* strain DH5α was initially used, followed by SURE2 cells. For growth, 50 μg/mL ampicillin was added to agar plates and LB broth to select for the desired plasmids containing the ampicillin resistant gene. Transformation was performed under sterile conditions. For bacterial transformation, 1 μL of plasmid (10 ng/μL) was mixed with 35 μL of competent cells (such as SURE2) and kept on ice for 30 minutes. The mixture was then 'heat shocked' at 42 •C for 30 seconds and placed back on ice for 2 minutes. To the cells, 500 μL of SOC media was added and the solution was incubated at 30 •C with shaking for 1 hour.

Cell mixture was spread onto agar plates (200 μL per plate) and incubated at 30 •C overnight. From each plate, 3 single cultures were picked and inoculated in starter cultures of LB broth (3 mL) and incubated at 30 •C with shaking for 7 hours, before half of the starter culture (1.5 mL) was expanded into LB broth (150 mL, 50 ng/μL Amp) and incubated at 30 •C with shaking overnight. Starter cultures were topped up with 1.5 mL of LB broth and also incubated at 30 •C with shaking overnight.

Once the cells had suitably propagated (as seen by measurement of OD), 500 μL of each culture was taken to prepare a glycerol stock as described below (6.3.7). From each of the larger cultures, 3 mL was initially taken, and the plasmid DNA was extracted and purified by mini prep (QIAGEN®) and checked by restriction digest with BamHI and HindiII as described above. All other cultures were pelleted for processing.

6.3.7 *Preparation of bacterial glycerol stock*

A 50% glycerol solution was prepared by diluting pure glycerol in distilled water. To the 50% glycerol solution (500 μL), 500 μL of bacterial culture was added in 2 mL screw top cryovial. The solution was gently mixed and stored at -80 •C until use.

6.3.8 *Growth of bacteria from glycerol stock*

The protocol was performed under sterile conditions. Glycerol stocks were removed from storage at -80 •C, though not allowed to thaw, and a pipette tip was used to scratch the surface and pick up some material. This was then streaked across an agar plate (50 ng/μL Amp) and the plate was incubated overnight at 30 •C and single colonies were picked for growth in LB media as described above (6.3.6).

6.3.9 *Sequencing of $(GAA)_n$ repeats*

The plasmid DNA purified from the bacterial cultures was sequenced by Sanger Sequencing (GENEWIZ®). Template and primer were pre-mixed in a 1.5 mL Eppendorf tube according to sample preparation guidelines provided (10 μL template at 50 ng/μL in water, 5 μL primer at 5 μM in water). Sequencing files were analysed using SnapGene.

6.3.10 *PCR amplification*

The insert of interest was amplified from the Genscript plasmids by PCR with Herculase II Fusion DNA Polymerase (Agilent). The reaction was carried out in 50 μL and consisted of Herculase II Fusion DNA polymerase (Agilent), 1x Herculase II buffer (with a final concentration of 2 mM Mg^{2+}), dNTPs, sterile water, 2 ng of plasmid DNA and 0.2 μM final concentration of each primer (M13F and M13R). A control plasmid with corresponding primers was used to validate the reagents. The reaction began with a heated lid step (105 •C), followed by an initial denaturation step (98 •C, 4 mins), denaturation step (94 •C, 30 s), annealing step (55 •C, 30 s), extension step (72 •C, 15 s for **55r**, 36 s for **200r**) and final extension step (72 •C, 5 mins) over 35 cycles for **55r** and 30 cycles for **200r**. For the optimisation of PCR conditions, a gradient of temperatures ranging from 50 to 60 •C for the annealing step was set up across the plate, and the other variables were kept consistent. For the DMSO titration, a (v/v)

concentration of 1-10% DMSO was added to the PCR reaction, reducing the amount of sterile water to keep the total reaction volume consistent. The same was done for the $MgCl_2$ titration, with the concentration being varied from 2-10 mM (the 1x Herculase II buffer had an inherent Mg^{2+} concentration of 2 mM) and the added volume being compensated for my reducing the amount of sterile water added.

6.3.11 Purification of DNA from PCR

After PCR amplification, DNA was purified using the QIAquick® PCR Purification Kit. Five volumes of binding Buffer PB was added to the PCR reaction, and the pH was adjusted with sodium acetate (NaOAc, 3 M, pH 5.0) as per the pH indicator solution. The solution was applied to the QIAquick spin column and centrifuged for 60 seconds at 12,000 g, and the flow-through discarded. Wash Buffer PE was applied to the column and centrifuged for 60 seconds at 12,000 g, the flow-through discarded, and the column centrifuged again to ensure complete removal of buffer. The DNA was then eluted with Buffer EB and the concentration and purity was determined with a Nanodrop One (Thermo Scientific), and checked by gel electrophoresis.

6.3.12 In vitro transcription

Using the DNA purified from the PCR amplification of the plasmid, RNA was synthesised using the MAXIscipt™ T7/T3 Transcription Kit (Invitrogen™) and the T3 and T7 promoter sites flanking the $(GAA)_n$ insert region. The reaction was carried out in 20 μL and consisted of T7 or T3 enzyme mix, 1x Transcription Buffer, 0.2–1 μg of DNA template, ATP, CTP, GTP and UTP free nucleotides, and nuclease-free water. Samples were incubated at 37 •C overnight, before being purified by ethanol precipitation as described above (6.3.4) and resuspended in EB buffer. The concentration of RNA was measured using a Nanodrop One (Thermo Scientific) and checked by gel electrophoresis.

6.3.13 Purification of DNA from agarose gel electrophoresis

DNA was purified from the gel using the QIAEX® Gel Extraction Kit. Samples were loaded onto an agarose gel (0.8%) and ran in 1x TBE for approximately 40

minutes at 90 V. The gel was then post-stained in 1x EtBr solution for 15 minutes, before being visualised using a Bio-Rad Molecular Imager® Gel Dock™ XR+ Imaging System next to a ruler. Desired bands were cut out of the gel using a sterile scalpel. The gel slice was weighed and solubilised in 3 volumes of Buffer QG to 1 volume of gel slice by vortexing briefly. The solution was heated to 50 •C for 10 minutes to bind the DNA and the pH was adjusted with NaOAc (10 µL, 3 M, pH 5.0. To this, 1 volume of isopropanol was added, and the solution was applied to a QIAquick spin column and centrifuged (12,000 g, 30 seconds) to bind the DNA to the membrane. The DNA was washed with Buffer PE and eluted with Buffer EB, before the concentration and purity was measured using a Nanodrop One (Thermo Scientific) and checked by gel electrophoresis.

6.3.14 Annealing of nucleic acids for triplex studies

Samples were annealed in a similar fashion to the shorter oligonucleotides. The double-stranded PCR insert was diluted in sodium cacodylate buffer (10 mM, pH 5.5 or 7.4) with an appropriate concentration of metal cation chloride salt in milli-Q water to a concentration of 100 ng/µL. The oligonucleotide solution was heated to 95 •C for 5 minutes and then cooled slowly to room temperature over approximately 3 hours, or overnight. Alternatively, the dsDNA was taken directly from the PCR amplification after purification, and the RNA TFO was added to solution and incubated at 37 •C for 24 hours or left at 4 •C for 48 hours. Following annealing/incubation, samples were purified by ethanol precipitation (6.3.4) and secondary structure formation was assessed by gel electrophoresis (6.3.5).

6.3.15 AFM samples preparation

Samples for AFM were annealed as described above (6.3.14) in sodium cacodylate buffer (10 mM, pH 5.5 or pH 7.4). After annealing, DNA was purified by ethanol precipitation (6.3.4) and resuspended in EB buffer to create a highly concentrated solution. The purity was measured using a Nanodrop One (Thermo Scientific) and DNA was checked on an agarose gel (1%, TBE buffer, EtBr), run at 100V for 60 minutes, before being visualised under UV light.

AFM sample preparation, imaging and data processing was performed by Philip Haynes, UCL, as described below:

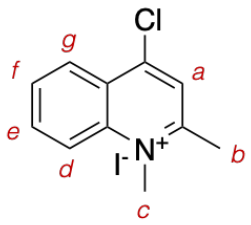
DNA pellets were resuspended in buffer; samples annealed at pH 5.5 and pH 7.4 were dissolved in 50mM NaOAc (pH 5.5) and 20mM HEPES (pH 7.4) respectively. For DNA adsorption at room temperature, 10 µL of poly-L-lysine (PLL, 0.1%, 150,000-300,000 MW, Sigma-Aldrich) was added to a freshly cleaved mica disk (diameter 6 mm, Agar Scientific, UK) and left to incubate for 1 minute. The disk was then held at an angle and put under a stream of double-deionised water (Milli-Q, Merck Millipore). The disk was then blotted and 20 µL DNA sample (0.25 ng/µL) was added. After incubating for 10 minutes, the sample was washed 5 times with imaging buffer (same as resuspension buffer).

All images were recorded by force–distance curve-based imaging (PeakForce Tapping) on a FastScan Bio system (Bruker, Santa Barbara, USA) at a frame rate of 0.2-0.8 /min. We used commercial silicon nitride FastScan-D cantilevers (Bruker) that have a resonance frequency of ~110 kHz and nominal spring constant ~0.25 Nm^{-1}. The force–distance curves were recorded at 8 kHz with a tip-sample separation (PeakForce amplitude) between 10 and 20 nm. Recorded images were either 512 or 1024 pixels per line. All AFM measurements were performed in liquid at room temperature.

The images were processed using Nanoscope Analysis software (Bruker). They were first flattened by a zeroth-order line-by-line flattening where the mean value of each row is subtracted to centre the data. If there was tilt in the image, a first order plane fit was applied.

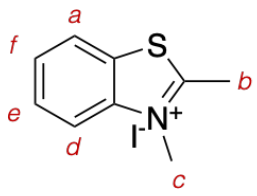
6.4 Synthesis

*6.4.1 Synthesis of 4-chloro-1,2-dimethylquinolin-1-ium iodide (**1**)*



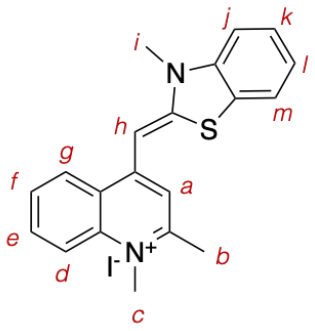
This compound was prepared following a previously reported procedure.[131] A solution of 4-chloro-2-methylquinoline (0.2 g, 1.12 mmol) in sulfolane (10 mL) and iodomethane (0.5 mL, 8.03 mmol) was stirred at 50 •C for 18 hours. Once the solution was cooled to room temperature, anhydrous diethyl ether (~300 mL) was added to the reaction mixture and the resulting solids were collected by filtration under reduced pressure. The solid was dried under vacuum to give a brown powder (1) (0.25 g, 0.77 mmol, 79%). ^{1}H NMR (400 MHz, DMSO-d$_6$): δ 8.67 (d, *J* = 9.0 Hz, 1H$_d$), 8.54 (d, *J* = 8.4 Hz, 1H$_g$), 8.52 (s, 1H$_a$), 8.32 (t, *J* = 8.0 Hz, 1H$_e$), 8.11 (dd, *J* = 8.0 and 7.4, 1H$_f$), 4.42 (s, 3H$_b$), 3.05 (s, 3H$_c$) ppm; ESI-MS: *m/z* 192.1 [M^{35}Cl – I]$^+$.

6.4.2 *Synthesis of 1,2-dimethylbenzothiazol-1-ium iodide (**2**)*



This compound was prepared following a previously reported procedure.[195] A solution of 2-methylbenzothiazole (0.25 g, 1.68 mmol) and iodomethane (0.7 mL, 11.2 mmol) in DMF (10 mL) was stirred at 60 •C for 20 hours. Once the solution had cooled to room temperature, ethyl acetate (~300 mL) was added and the light pink precipitate was collected by filtration under reduced pressure. The solid was resuspended in ethyl acetate (20 mL) and heated to reflux for 20 minutes. The solution was cooled to room temperature, and the solid was collected by filtration under reduced pressure and dried under vacuum to afford a pale pink powder (2) (0.42 g, 1.46 mmol, 87%). ^{1}H NMR (400 MHz, DMSO-d$_6$): δ 8.44 (d, *J* = 8.2, 1H$_a$), 8.20 (d, *J* = 8.5, 1H$_d$), 7.91 (t, *J* = 7.9, 1H$_e$), 7.81 (t, *J* = 7.7, 1H$_f$), 4.20 (s, 3H$_c$), 3.17 (s, 3H$_b$) ppm; ESI-MS: *m/z* 164.1 [M – I]$^+$.

6.4.3 *Synthesis of (Z)-1,2-dimethyl-4-((3-methylbenzo[d]thiazol2(3H)-ylidene)methyl) quinolin-1-ium iodide (**3**)*

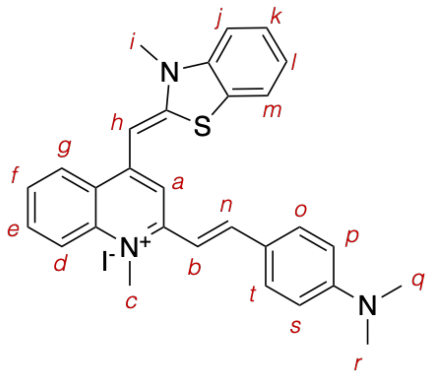


This compound was prepared following a previously reported procedure.[196] A mixture of (**1**) (0.5 g, 1.60 mmol), (**2**) (0.5 g, 1.72 mmol) and aqueous sodium bicarbonate solution (0.5 mol/L, 2 mL) in methanol (10 mL) was stirred at RT for 1 hour, and then heated to reflux for 30 minutes. Saturated potassium iodide solution (4 mL) was then added, and the reaction mixture was stirred for 15 minutes. The resulting solid was collected by filtration under reduced pressure, washed with water (5 mL) and acetone (5 mL) and dried under vacuum to give compound (3) as a bright red powder (47 mg, 0.11 mmol, 66%). ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.77 (br. dd, J = 8.7 and 1.0 Hz, 1H$_d$), 8.19 (br. dd, J = 8.9 and 0.6 Hz, 1H$_g$), 8.04-7.97 (m, 2H$_{f,e}$),7.79-7.71 (m, 2H$_{j,m}$), 7.60 (ddd, J = 8.4, 7.4 and 1.2 Hz, 1H$_k$), 7.43-7.37 (m, 1H$_l$), 7.36 (s, 1H$_a$), 6.86 (s, 1H$_h$), 4.08 (s, 3H$_i$), 3.99 (s, 3H$_c$), 2.88 (s, 3H$_b$) ppm; ESI-MS: *m/z* 320.1 [M – I]$^+$. Elemental analysis: Calc. C 53.82, H 4.29, N 6.28%, Found C 53.49, H 4.31, N 6.17%.

*6.4.4 Synthesis of **4a-e**, general procedure*

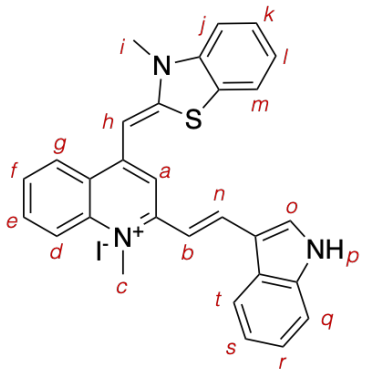
These compounds were prepared following previously reported procedures.[131,197] A mixture of **3** (0.072 g, 0.16 mmol), 4-methylpiperidine (0.5 mL) and n-butanol (10 mL) was stirred at RT for 1 hour. The corresponding aldehyde (0.32 mmol) was added and the solution was heated to 135 •C for 3 hours. This was cooled to room temperature and the resulting solid was collected by suction filtration and washed with n-butanol (10 mL). The solid was then purified by column chromatography with a solvent system of DCM:MeOH (gradient, MeOH 1-10%). Solid was collected by and dried under reduced pressure and lyophilised to remove residual water. Purity was analysed by ^{1}H NMR, mass spec and elemental analysis.

6.4.5 *N-methyl-2-((E)-4-(dimethylamino)styryl)-4-((Z)-(3-methylbenzo[d]thiazol-2(3H)-ylidene)methyl)quinolin-1-ium iodide (**4a**)*



Reported previously.[131] Purple pearlescent solid (54 mg, 0.09 mmol, 58%). ^{1}H NMR (400 MHz, DMSO-d$_6$): δ 8.73 (d, *J* = 8.9 Hz, 1H$_d$), 8.18 (d, *J* = 9.0 Hz, 1H$_g$), 8.04 (d, *J* = 7.7 Hz, 1H$_b$), 7.97 (t, *J* = 8.1 Hz, 1H$_f$), 7.81 (d, *J* = 8.8, 2H$_{j,m}$), 7.75-7.64 (m, 4H$_{a,e,o,t}$), 7.58 (t, *J* = 7.8 Hz, 1H$_l$), 7.48 (d, *J* = 15.6 Hz, 1H$_n$), 7.37 (t, *J* = 7.2 Hz, 1H$_k$), 6.86-6.80 (m, 3H$_{h,p,s}$), 4.17 (s, 3H$_i$), 3.97 (s, 3H$_c$), 3.05 (s, 6H$_{q,r}$) ppm; ESI-MS: *m/z* 450.2 [M – I]$^+$. Elemental analysis: Calc. C 60.31, H 4.89, N, 7.28%, Found C 59.63, H 4.90, N 7.08%.

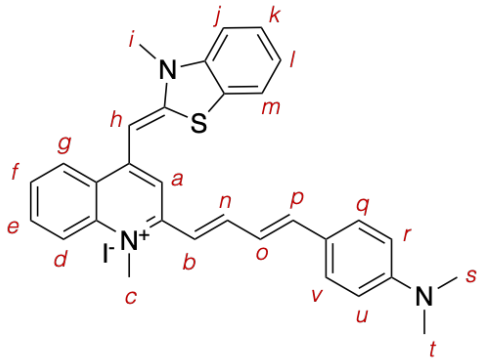
6.4.6 *2-((E)-2-(1H-indol-3-yl)vinyl)-1-methyl-4-((Z)-(3-methylbenzo[d]thiazol-2(3H)-ylidene)methyl)quinolin-1-ium iodide (**4b**)*



Reported previously.[131] Dark purple solid (43 mg, 0.07 mmol, 47%). ^{1}H NMR (400 MHz; DMSO-d$_6$): δ 11.85 (s, 1H$_p$), 8.76 (d, *J* = 8.6 Hz, 1H$_d$), 8.19 (d, *J* = 8.9 Hz, 1H$_g$), 8.06 (dd, *J* = 7.8 and 0.8 Hz, 1H$_b$), 8.00 (ddd, *J* = 8.5, 7.0 and 1.2, 1H$_f$), 7.81-7.71 (m,

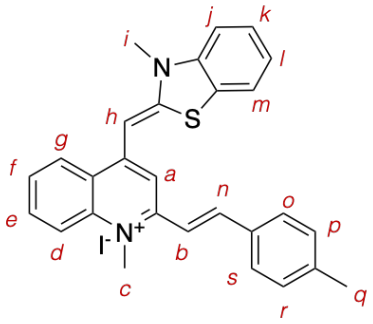
5H$_{a,e,j,m,o}$), 7.68-7.63 (m, 1H$_k$), 7.60 (ddd, *J* = 8.4, 7.4 and 1.2 Hz, 1H$_l$), 7.48 (dd, *J* = 8.2 and 0.8 Hz, 1H$_t$), 7.43-7.39 (ddd, *J* = 8.1, 7.1 and 1.1 Hz, 1H$_q$), 7.26 (ddd, *J* = 8.1, 7.1 and 1.1 Hz, 1H$_s$), 7.12 (s, 1H$_n$), 7.08 (ddd, *J* = 7.9, 7.1 and 1.1 Hz, 1H$_r$), 6.91 (s, 1H$_h$), 4.20 (s, 3H$_i$), 4.01 (s, 3H$_c$) ppm; ESI-MS: *m/z* 446.1 [M – I]$^+$. Elemental analysis: Calc. C 70.74, H 4.22, N 7.33%, Found C 68.92, H 4.35, N 7.22%.

6.4.7 *2-((1E,3E)-4-(4-(dimethylamino)phenyl)buta-1,3-dien-1-yl)-1-methyl-4-((Z)-(3-methylbenzo[d]thiazol-2(3H)-ylidene)-methyl)quinolin-1-ium iodide (**4c**)*



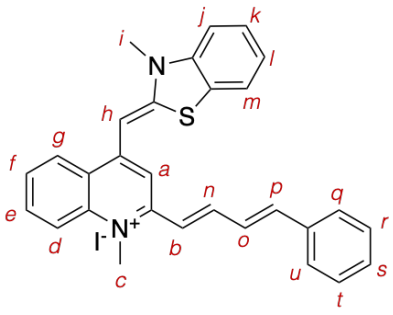
Reported previously.[131] Purple/gold pearlescent solid (47 mg, 0.08 mmol, 49%). ^{1}H NMR (400 MHz; DMSO-d$_6$): δ 8.72 (d, *J* = 8.2 Hz, 1H$_d$), 8.14 (d, *J* = 8.7 Hz, 1H$_g$), 8.07 (d, *J* = 7.8 Hz, 1H$_b$), 7.99-7.93 (m, 1H$_f$), 7.75-7.69 (m, 2H$_{j,m}$), 7.64 (s, 1H$_a$), 7.63-7.55 (m, 2H$_{e,n}$), 7.49 (d, *J* = 8.8 Hz, 2H$_{r,u}$), 7.41-7.37 (m, 1H$_k$), 7.27-7.13 (m, 3H$_{l,q,v}$), 6.85 (s, 1H$_h$), 6.79 (s, 1H$_p$), 6.76 (s, 1H$_o$), 4.11 (s, 3H$_c$), 3.97 (s, 3H$_i$), 3.00 (s, 6H$_{s,t}$) ppm; ESI-MS: *m/z* 476.2 [M – I]$^+$. Elemental analysis: Calc. C 61.69, H 5.01, N 6.96%, Found C 60.25, H 4.99, N 6.49%.

6.4.8 *1-methyl-4-((Z)-(3-methylbenzo[d]thiazol-2(3H)- ylidene)methyl)-2-((E)-4-methylstyryl)quinolin-1-ium iodide (**4d**)*



Reported previously.[131] Light red pearlescent solid (46 mg, 0.08 mmol, 52%). ^{1}H NMR (400 MHz; DMSO-d_6): δ 8.75 (d, *J* = 8.4 Hz, 1H$_d$), 8.17 (d, *J* = 8.6 Hz, 1H$_g$), 8.05 (d, *J* = 7.8 Hz, 1H$_b$), 7.99 (t, *J* = 8.1 Hz, 1H$_f$), 7.83 (d, *J* = 8.0 Hz, 2H$_{j,m}$), 7.77-7.72 (m, 3H$_{a,e,n}$), 7.67-7.57 (m, 3H$_{k,o,s}$), 7.39 (t, *J* = 7.7 Hz, 1H$_l$), 7.34 (d, *J* = 7.9 Hz, 2H$_{p,r}$), 6.90 (s, 1H$_h$), 4.15 (s, 3H$_c$), 3.99 (s, 3H$_i$), 2.38 (s, 3H$_q$) ppm; ESI-MS: *m/z* 421.1 [M – I]$^+$. Elemental analysis: Calc. C 61.32, H 4.59, N 5.11%, Found C 54.93, H 4.23, N 4.24%.

6.4.9 *1-methyl-4-((Z)-(3-methylbenzo[d]thiazol-2(3H)-ylidene)methyl)-2-((1E,3E)-4-phenylbuta-1,3-dien-1-yl)quinolin-1-ium iodide (**4e**)*



Reported previously.[131] Purple/green pearlescent solid (34 mg, 0.06 mmol, 38%). ^{1}H NMR (400 MHz; DMSO-d_6): δ 8.78-8.75 (m, 1H$_d$), 8.17 (br. dd, *J* = 8.7 and 0.7 Hz, 1H$_g$), 8.08 (br. dd, *J* = 7.9 and 0.8 Hz, 1H$_b$), 7.99 (ddd, *J* = 8.3, 6.8 and 1.0 Hz, 1H$_f$), 7.79-7.72 (m, 2H$_{e,n}$), 7.67-7.56 (m, 5H$_{a,j,k,l,m}$), 7.49-7.28 (m, 7H$_{o,p,q,r,s,t,u}$), 6.91 (s, 1H$_h$), 4.13 (s, 3H$_c$), 4.01 (s, 3H$_i$) ppm; ESI-MS: *m/z* 433.1 [M – I]$^+$. Elemental analysis: Calc. C 60.68, H 4.34, N 5.24%, Found C 58.92, H 4.51, N 5.09%.

6.5 DNA Binding Analysis

6.5.1 UV-Vis absorption titrations

Stock solutions of compounds **3** and **4a-e** were made in DMSO to a concentration of 2 mM. Stock solutions of DNA were annealed at a total concentration of 200 μM in cacodylate buffer (10 mM, pH 5.5 or 7.4) containing the appropriate concentration of metal cation chloride salt in milli-Q water. Measurements were taken on a Cary 60 UV-Vis instrument (Agilent). A baseline was measured with only the buffer. The solution of the compound being tested was diluted in sodium cacodylate buffer (10 mM, pH 5.5 or 7.4) to a concentration of 2 *μ*M in a 500 *μ*L cuvette, and an absoprtion spectrum was taken. The stock DNA solution (200 μM) was titrated into the cuvette at equivalents of 0.2, 0.4, 0.6, 0.8, 1.2, 1.6, 2.0, 2.4, 3.2, 4.0, 5.6, 7.2 and 10.4, incubated for 3 minutes, and a spectrum were recorded for each point.

6.5.2 Fluorescence emission titrations

Stock solutions of compounds **3** and **4a-e** were made in DMSO to a concentration of 2 mM. Stock solutions of DNA were annealed at a total concentration of 200 μM in cacodylate buffer (10 mM, pH 5.5. or 7.4) containing the appropriate concentration of metal cation chloride salt in milli-Q water. Measurements were taken on a Cary Eclipse Fluorescence Spectrophotometer (Agilent). A baseline was measured with only the buffer. A solution of the compound being tested was diluted in sodium cacodylate buffer (10 mM, pH 5.5 or 7.4) to a concentration of 2 *μ*M in a 500 *μ*L cuvette, and a fluorescence spectrum was taken. The stock DNA solution was titrated into the cuvette at equivalents of 0.2, 0.4, 0.6, 0.8, 1.2, 1.6, 2.0, 2.4, 3.2, 4.0, 5.6, 7.2 and 10.4, incubated for 3 minutes, and spectra were recorded for each point.

Binding curves were plotted showing the plateau at the maximum emission wavelength (λ_{em}) of each probe with each DNA topology. Affinity constants (K_a) were extracted from these curves using the Aldrich-Wright method, according the equation below:[203]

$$I = 0.5 * I_{max} * [\frac{1}{K_a} + L + n[DNA] - \left\{\left(\frac{1}{K_a} + +n[DNA]\right)^2 - 4 * L * n * [DNA]\right\}^{0.5}]$$

Where *I* is the fluorescence intensity, I_{max} is the maximum fluorescence intensity at the plateau, K_a is the affinity binding constant derived using a Levenberg-Marquardt nonlinear least-squares iteration algorithm, *L* is the concentration of ligand added (2 μM), *[DNA]* is the concentration of DNA added (in M), and *n* is the number of binding sites (set to a constant of 2).

Alternatively, data that could not be fit using this model (as it did not reach a plateau) was fit using to the Swillens method according to the equation below:[204]

$$I = \frac{(I_{max} * [DNA])}{(K_d + [DNA])} + (\alpha * [DNA]) + I_0$$

Where *I* is fluorescence intensity, I_{max} is the maximum fluorescence intensity at the plateau, [DNA] is the concentration of DNA added (in M), K_d is the dissociation constant (in M) derived using a Levenberg-Marquardt nonlinear least-squares iteration algorithm, α is the gradient of the initial slope (in M^{-1}) and I_0 is the background fluorescence intensity of the compound in the absence of DNA. The affinity constant (K_a) was calculated as the reciprocal of the dissociation constant (K_d).

6.5.3 *Time-correlated single-photon counting (TCSPC)*

Stock solutions of compounds **3** and **4a-e** were made in DMSO to a concentration of 2 mM. Stock solutions of DNA were annealed at a total concentration of 200 μM in cacodylate buffer (10 mM, pH 5.5 or 7.4) containing the appropriate concentration of metal cation chloride salt in milli-Q water. Measurements were taken using a Jobin Yvon IBH data station (5000F, HORIBA Scientific) TCSPC set up from Horiba Scientific. The excitation sources used were 467 or 535 nm NanoLED diode lasers, depending on the λ_{ex} of the probe being studied. An IRF was recorded using diluted LUDOX at the maximum wavelength of the excitation source, and a neutral density filter was used. During measurements, short pass filters were used (< 467 and 535 nm) to eliminate interference from external or scattered light. Compounds were diluted from the stock solution with sodium cacodylate buffer (10 mM, pH 5.5 or 7.4) in a 500 μL cuvette to a concentration of 2 μM, and a measurement was taken in the

absence of DNA (if compound was emissive enough), with the decay being recorded at the maximum emission wavelength (λ_{em}) of the corresponding compound. The annealed DNA was then added into the cuvette to a final concentration of 10 μM (5 eq.), and measurements were recorded at the maximum emission wavelength (λ_{em}) of the compound bound to that DNA topology, until the 10,000 photon maxima was reached.

Data was plotted and parameters were extracted using DAS-6 fluorescence decay analysis software (HORIBA Scientific). All decay curves were fit to a bi- and tri-exponential model, with the goodness of fit being assessed using the Chi parameter (the closer the value is to 1, the better the fit). The weighted lifetime (τ_w) of each decay was calculated, taking into account the lifetime components (τ) and their fractional contributions to the overall average lifetime (α), using the following equation:

$$\tau_w = \frac{\alpha_1 {\tau_1}^2 + \alpha_2 {\tau_2}^2}{\alpha_1 \tau_1 + \alpha_2 \tau_2}$$

Where τ is the lifetime of the component and α is the contribution of that lifetime value to the weighted lifetime (τ_w). The goodness of fit was assessed using the χ^2 parameter; a value between 1 and 2 indicates an acceptable fit, with closer to 1 being a better fit.

6.5.4 Fluorescence lifetime imaging microscopy

HEK293 cells, both the wild-type (**WT**) and containing the $(GAA)_{310}$ repeat expansion mutation (**mut**) were grown at 37 •C with 5% CO_2 in DMEM with 10% FBS and 1% Pen strep and L-glutamine. Cells were counted using a Guava® easyCyte flow cytometer and diluted to a concentration of 300,000 cells in 2.5 mL of media. In each well of the 8-well FLIM sample chamber slide, 30,000 cells were seeded in 250 μL of media. In the 4 left-hand wells, the **WT** cells were seeded, and in the 4 right-hand cells the **mut** cells were seeded. These cells were incubated overnight in the chamber slide at 37 °C with 5% CO_2. The probes, **3**, **4a** and **4d**, were added to the appropriate wells to a final concentration of 2 μM and 5 μM as stated and incubated for an hour before imaging.

FLIM was performed by Dr Peter Summers, Imperial College London, as follows:

Fluorescence lifetime imaging microscopy was performed through time-correlated single-photon counting (TCSPC), using a confocal laser scanning microscope (Leica, SP5 II). A mode-locked femtosecond Ti:Sapphire laser (Coherent, Chameleon Vision II) tunable over 680-1080 nm (140 fs pulse duration, 80 MHz) was used as the excitation source, operated at the excitation wavelength corresponding to the probe being imaged. A PMC-100-1 photomultiplier tube (Hamamatsu) and a SPC-830 single-photon counting card (Becker-Hickl) were used for data acquisition. Fluorescence was collected between 500 and 700 nm; short-pass filters of > 470 nm were used. Samples on chamber slides were mounted in the microscope chamber heated by a thermostat (Lauda GmbH, E200) to 37 •C under an atmosphere of 5% CO_2 in air. SHG signal from urea crystals on a glass cover slide was used to obtain the instrument response function (IRF). Images were collected using a 100x objective with immersion oil, 3 per well, at 155 µm across, 255x256 pixel resolution. For each well, a high-resolution picture with 512x512 pixel resolution was taken.

Data analysis was performed using FLIMfit (5.1.1), where images were individually uploaded, and the nuclei were selected using a segmentation tool. Each pixel in each nucleus was fitted with a bi-exponential function, giving individual decays for each pixel and an averaged decay across the whole image. Each nuclei segment has a calculated weighted average lifetime (τ_w), which gave an average τ_w across each image. Points corresponding to low fluorescence intensity that resulted in a fluorescence lifetime value skewed towards a longer value were eliminated by assessing the severity of the skew by plotting the intensity against the weighted lifetime.

References

(1) Friedreich, N. Über Degenerative Atrophie Der Spinalen Hinterstrânge. *Virchows Arch. Pathol. Anat.* **1863**, *26*, 391.

(2) Pandolfo, M. Friedreich Ataxia: The Clinical Picture. *Journal of Neurology* **2009**, *256* (1), 3–8.

(3) Filla, a; de Michele, G.; Cavalcanti, F.; Pianese, L.; Monticelli, a; Campanella, G.; Cocozza, S. The Relationship between Trinucleotide (GAA) Repeat Length and Clinical Features in Friedreich Ataxia. *American Journal of Human Genetics* **1996**, *59* (3), 554–560.

(4) Pastore, A.; Puccio, H. Frataxin: A Protein in Search for a Function. *Journal of Neurochemistry* **2013**, *126 Suppl*, 43–52.

(5) Babcock, M.; de Silva, D.; Oaks, R.; Davis-Kaplan, S.; Jiralerspong, S.; Montermini, L.; Pandolfo, M.; Kaplan, J. Regulation of Mitochondrial Iron Accumulation by Yfh1p, a Putative Homolog of Frataxin. *Science* **1997**, *276*, 1709–1712.

(6) Pandolfo, M. Iron Metabolism and Mitochondrial Abnormalities in Friedreich Ataxia. *Blood Cells, Molecules, and Diseases* **2002**, *29* (3), 536–547.

(7) Campuzano, V.; Montermini, L.; Lutz, Y.; Cova, L.; Hindelang, C.; Jiralerspong, S.; Trottier, Y.; Kish, S. J.; Faucheux, B.; Trouillas, P.; et al. Frataxin Is Reduced in Friedreich Ataxia Patients and Is Associated with Mitochondrial Membranes. *Human Molecular Genetics* **1997**, *6* (11), 1771-1780.

(8) Gellera, C.; Castellotti, B.; Mariotti, C.; Mineri, R.; Seveso, V.; Didonato, S.; Taroni, F. Frataxin Gene Point Mutations in Italian Friedreich Ataxia Patients. *Neurogenetics* **2007**, *8* (4), 289–299.

(9) Festenstein, R. Breaking the Silence in Friedreich's Ataxia. *Nature Chemical Biology* **2006**, *2* (10), 512–513.

(10) Rajeswari, M. R. DNA Triplex Structures in Neurodegenerative Disorder, Friedreich's Ataxia. *Journal of Biosciences* **2012**, *37* (3), 519–532.

(11) Sakamoto, N.; Ohshima, K.; Montermini, L.; Pandolfo, M.; Wells, R. D. Sticky DNA, a Self-Associated Complex Formed at Long GAA*TTC Repeats in Intron 1 of the Frataxin Gene, Inhibits Transcription. *The Journal of Biological Chemistry* **2001**, *276* (29), 27171–27177.

(12) Groh, M.; Lufino, M. M. P.; Wade-Martins, R.; Gromak, N. R-Loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLoS Genetics* **2014**, *10* (5).

(13) Grabczyk, E.; Usdin, K. The GAA*TTC Triplet Repeat Expanded in Friedreich's Ataxia Impedes Transcription Elongation by T7 RNA Polymerase in a Length and Supercoil Dependent Manner. *Nucleic Acids Research* **2000**, *28* (14), 2815-2822.

(14) Krasilnikova, M. M.; Mirkin, S. M. Replication Stalling at Friedreich's Ataxia (GAA)n Repeats in Vivo. *Molecular and Cellular Biology* **2004**, *24* (6), 2286–2295.

(15) Bergquist, H.; Rocha, C. S. J.; Álvarez-Asencio, R.; Nguyen, C.-H.; Rutland, Mark. W.; Smith, C. I. E.; Good, L.; Nielsen, P. E.; Zain, R. Disruption of Higher

Order DNA Structures in Friedreich's Ataxia (GAA)n Repeats by PNA or LNA Targeting. *PLOS ONE* **2016**, *11* (11).
(16) Sakamoto, N.; Chastain, P. D.; Parniewski, P.; Ohshima, K.; Pandolfo, M.; Griffith, J. D.; Wells, R. D. Sticky DNA: Self-Association Properties of Long GAA · TTC Repeats in R · R · Y Triplex Structures from Friedreich's Ataxia. *Molecular Cell* **1999**, *3* (4), 465–475.
(17) Burnett, R.; Melander, C.; Puckett, J. W.; Son, L. S.; Wells, R. D.; Dervan, P. B.; Gottesfeld, J. M. DNA Sequence-Specific Polyamides Alleviate Transcription Inhibition Associated with Long GAA·TTC Repeats in Friedreich's Ataxia. *Proceedings of the National Academy of Sciences* **2006**, *103* (31), 11497-11502.
(18) Neil, A. J.; Liang, M. U.; Khristich, A. N.; Shah, K. A.; Mirkin, S. M. RNA–DNA Hybrids Promote the Expansion of Friedreich's Ataxia (GAA)n Repeats via Break-Induced Replication. *Nucleic Acids Research* **2018**, *46* (7), 3487-3497.
(19) Kumari, D.; Usdin, K. Is Friedreich Ataxia an Epigenetic Disorder? *Clinical Epigenetics* **2012**, *4* (1), 2.
(20) Oberle, I.; Rousseau, F.; Heitz, D.; Kretz, C.; Devys, D.; Hanauer, A.; Boue, J.; Bertheas, M. F.; Mandel, J. L. Instability of a 550-Base Pair DNA Segment and Abnormal Methylation in Fragile X Syndrome. *Science* **1991**, *252* (5009), 1097–1102.
(21) Grewal, S. I. S.; Moazed, D. Heterochromatin and Epigenetic Control of Gene Expression. *Science* **2003**, *301* (5634), 798–802.
(22) Dillon, N.; Festenstein, R. Unravelling Heterochromatin: Competition between Positive and Negative Factors Regulates Accessibility. *Trends in Genetics* **2002**, *18* (5), 252–258.
(23) Trojer, P.; Reinberg, D. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell* **2007**, *28* (1), 1–13.
(24) Chan, P. K.; Torres, R.; Yandim, C.; Law, P. P.; Khadayate, S.; Mauri, M.; Grosan, C.; Chapman-Rothe, N.; Giunti, P.; Pook, M.; et al. Heterochromatinization Induced by GAA-Repeat Hyperexpansion in Friedreich's Ataxia Can Be Reduced upon HDAC Inhibition by Vitamin B3. *Human MolecularGgenetics* **2013**, *22* (13), 2662–2675.
(25) Schotta, G.; Lachner, M.; Sarma, K.; Ebert, A.; Sengupta, R.; Reuter, G.; Reinberg, D.; Jenuwein, T. A Silencing Pathway to Induce H3-K9 and H4-K20 Trimethylation at Constitutive Heterochromatin. *Genes & Development* **2004**, *18* (11), 1251–1262.
(26) Gottesfeld, J. M.; Rusche, J. R.; Pandolfo, M. Increasing Frataxin Gene Expression with Histone Deacetylase Inhibitors as a Therapeutic Approach for Friedreich's Ataxia. *Journal of Neurochemistry* **2013**, *126 Suppl* (1), 147–154.
(27) Herman, D.; Jenssen, K.; Burnett, R.; Soragni, E.; Perlman, S. L.; Gottesfeld, J. M. Histone Deacetylase Inhibitors Reverse Gene Silencing in Friedreich's Ataxia. *Nature Chemical Biology* **2006**, *2* (10), 551–558.
(28) Chan, P. K.; Torres, R.; Yandim, C.; Law, P. P.; Khadayate, S.; Mauri, M.; Grosan, C.; Chapman-Rothe, N.; Giunti, P.; Pook, M.; et al. Heterochromatinization Induced by GAA-Repeat Hyperexpansion in Friedreich's Ataxia Can Be Reduced upon HDAC Inhibition by Vitamin B3. *Human Molecular Genetics* **2013**, *22* (13), 2662–2675.

(29) Saveliev, A.; Everett, C.; Sharpe, T.; Webster, Z.; Festenstein, R. DNA Triplet Repeats Mediate Heterochromatin-Protein-1-Sensitive Variegated Gene Silencing. *Nature* **2003**, *422*, 909–913.
(30) Al-Mahdawi, S.; Pinto, R. M.; Ismail, O.; Varshney, D.; Lymperi, S.; Sandi, C.; Trabzuni, D.; Pook, M. The Friedreich Ataxia GAA Repeat Expansion Mutation Induces Comparable Epigenetic Changes in Human and Transgenic Mouse Brain and Heart Tissues. *Human Molecular Genetics* **2007**, *17* (5), 735–746.
(31) Castaldo, I.; Pinelli, M.; Monticelli, A.; Acquaviva, F.; Giacchetti, M.; Filla, A.; Sacchetti, S.; Keller, S.; Avvedimento, V. E.; Chiariotti, L.; et al. DNA Methylation in Intron 1 of the Frataxin Gene Is Related to GAA Repeat Length and Age of Onset in Friedreich Ataxia Patients. *Journal of Medical Genetics* **2008**, *45* (12), 808–812.
(32) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.
(33) Dickerson, R. E. DNA Structure from A to Z. *Methods in Enzymology* **1992**, *211*, 67–111.
(34) Felsenfeld, G.; Davies, D. R.; Rich, A. Formation of a Three-Stranded Polynucleotide Molecule. *Journal of the American Chemical Society* **1957**, *79* (8), 2023–2024.
(35) Frank-Kamenetskii, M. D.; Mirkin, S. M. Triplex DNA Structures. *Annual Review of Biochemistry* **1995**, *64*, 65–95.
(36) Lipanov, A. A.; Chuprina, V. P. The Structure of Poly(DA):Poly(DT) in a Condensed State and in Solution. *Nucleic Acids Research* **1987**, *15* (14), 5833–5844.
(37) Muñoz, J.; Gelpí, J. L.; Soler-López, M.; Subirana, J. A.; Orozco, M.; Luque, F. J. Can Divalent Metal Cations Stabilize the Triplex Motif? Theoretical Study of the Interaction of the Hydrated Mg2+ Cation with the G–G·C Triplet. *The Journal of Physical Chemistry B* **2002**, *106* (34), 8849–8857.
(38) Wan, C.; Cui, M.; Song, F.; Liu, Z.; Liu, S. Evaluation of Effects of Bivalent Cations on the Formation of Purine-Rich Triple-Helix DNA by ESI-FT-MS. *Journal of the American Society for Mass Spectrometry* **2009**, *20* (7), 1281–1286.
(39) Šponer, J.; Sabat, M.; Gorb, L.; Leszczynski, J.; Lippert, B.; Hobza, P. The Effect of Metal Binding to the N7 Site of Purine Nucleotides on Their Structure, Energy, and Involvement in Base Pairing. *The Journal of Physical Chemistry B* **2000**, *104* (31), 7535–7544.
(40) Jiang, H.-X.; Cui, Y.; Zhao, T.; Fu, H.-W.; Koirala, D.; Punnoose, J. A.; Kong, D.-M.; Mao, H. Divalent Cations and Molecular Crowding Buffers Stabilize G-Triplex at Physiologically Relevant Temperatures. *Scientific Reports* **2015**, *5* (1), 9255.
(41) Floris, R.; Scaggiante, B.; Manzini, G.; Quadrifoglio, F.; Xodo, L. E. Effect of Cations on Purine ´ Purine ´ Pyrimidine Triple Helix Formation in Mixed-Valence Salt Solutions. *European Journal of Biochemistry* **1999**, *809*, 801–809.
(42) Noonberg, S. B.; François, J.-C.; Garestier, T.; Héiène, C. Effect of Competing Self-Structure on Triplex Formation with Purine-Rich Oligodeoxynucleotides Containing GA Repeats. *Nucleic Acids Research* **1995**, *23* (11), 1956–1963.

(43) Rusling, D. A.; Rachwal, P. A.; Brown, T.; Fox, K. R. The Stability of Triplex DNA Is Affected by the Stability of the Underlying Duplex. *Biophysical Chemistry* **2009**. https://doi.org/10.1016/j.bpc.2009.09.007.
(44) Nakano, S.; Miyoshi, D.; Sugimoto, N. Effects of Molecular Crowding on the Structures, Interactions, and Functions of Nucleic Acids. *Chemical Reviews* **2014**, *114* (5), 2733–2758.
(45) Spink, C. H.; Chaires, J. B. Effects of Hydration, Ion Release, and Excluded Volume on the Melting of Triplex and Duplex DNA. *Biochemistry* **1999**, *38* (1), 496–508.
(46) Spink, C. H.; Chaires, J. B. Selective Stabilization of Triplex DNA by Poly(Ethy1ene Glycols). *J. Am. Chem. Soc.* **1995**, *4* (6), 12887–12888.
(47) Roberts, R. W.; Crothers, D. M. Stability and Properties of Double and Triple Helices: Dramatic Effects of RNA or DNA Backbone Composition. *Science* **1992**, *258* (5087), 1463–1466.
(48) Cheng, Y.-K.; Pettitt, B. M. Stabilities of Double- and Triple-Strand Helical Nucleic Acids. *Progress in Biophysics and Molecular Biology* **1992**, *58* (3), 225–257.
(49) Escude, C.; Garestier, T.; Sun, J. S. Drug Interaction with Triple-Helical Nucleic Acids. *Methods in Enzymology* **2001**, *340*, 340–357.
(50) Aviñó, A.; Mazzini, S.; Gargallo, R.; Eritja, R. The Effect of Small Cosolutes That Mimic Molecular Crowding Conditions on the Stability of Triplexes Involving Duplex DNA. *International Journal of Molecular Sciences* **2016**, *17* (2), 211.
(51) Wilusz, J. E.; Sunwoo, H.; Spector, D. L. Long Noncoding RNAs: Functional Surprises from the RNA World. *Genes & Development* **2009**, *23* (13), 1494–1504.
(52) Rinn, J. L.; Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry* **2012**, *81*, 145–166.
(53) Li, Y.; Syed, J.; Sugiyama, H. RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chemical Biology* **2016**, *23* (11), 1325–1333.
(54) Schmitz, K.-M.; Mayer, C.; Postepska, A.; Grummt, I. Interaction of Noncoding RNA with the RDNA Promoter Mediates Recruitment of DNMT3b and Silencing of RRNA Genes. *Genes & development* **2010**, *24* (20), 2264–2269.
(55) le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.-L.; Thuong, N. T.; Lhomme, J.; Héène, C. Sequence-Specific Recognition, Photocrosslinking and Cleavage of the DNA Double Helix by an Oligo-(α]-Thymidylate Covalently Linked to an Azidoproflavine Derivative. *Nucleic Acids Research* **1987**, *15* (19), 7749–7760.
(56) Ping, Y.-H.; Rana, T. M. Mechanism of Site-Specific Psoralen Photoadducts Formation in Triplex DNA Directed by Psoralen-Conjugated Oligonucleotides. *Biochemistry* **2005**, *44* (7), 2501–2509.
(57) Liu, Y.; Nairn, R. S.; Vasquez, K. M. Targeted Gene Conversion Induced by Triplex-Directed Psoralen Interstrand Crosslinks in Mammalian Cells. *Nucleic Acids Research* **2009**, *37* (19), 6378–6388.
(58) Nielsen, P. E. Peptide Nucleic Acids (PNA) in Chemical Biology and Drug Discovery. *Chemistry & Biodiversity* **2010**, *7* (4), 786–804.
(59) Betts, L.; Josey, J. A.; Veal, J. M.; Jordan, S. R. A Nucleic Acid Triple Helix Formed by a Peptide Nucleic Acid-DNA Complex. *Science* **1995**, *270* (5243), 1838–1841.

(60) Hu, J.; Corey, D. R. Inhibiting Gene Expression with Peptide Nucleic Acid (PNA)--Peptide Conjugates That Target Chromosomal DNA. *Biochemistry* **2007**, *46* (25), 7581–7589.
(61) Sato, T.; Sato, Y.; Nishizawa, S. Triplex-Forming Peptide Nucleic Acid Probe Having Thiazole Orange as a Base Surrogate for Fluorescence Sensing of Double-Stranded RNA. *Journal of the American Chemical Society* **2016**, *138* (30), 9397–9400.
(62) Owczarzy, R.; You, Y.; Groth, C. L.; Tataurov, A. v. Stability and Mismatch Discrimination of Locked Nucleic Acid-DNA Duplexes. *Biochemistry* **2011**, *50* (43), 9352–9367.
(63) Torigoe, H.; Hari, Y.; Sekiguchi, M.; Obika, S.; Imanishi, T. 2'-O,4'-C-Methylene Bridged Nucleic Acid Modification Promotes Pyrimidine Motif Triplex DNA Formation at Physiological PH: Thermodynamic and Kinetic Studies. *The Journal of Biological Chemistry* **2001**, *276* (4), 2354–2360.
(64) Xu, Y.; Gissberg, O.; Pabon-Martinez, Y. V.; Wengel, J.; Lundin, K. E.; Smith, C. I. E.; Zain, R.; Nilsson, L.; Villa, A. The Ability of Locked Nucleic Acid Oligonucleotides to Pre-Structure the Double Helix: A Molecular Simulation and Binding Study. *PLOS ONE* **2019**, *14* (2).
(65) Szabat, M.; Kierzek, E.; Kierzek, R. Modified RNA Triplexes: Thermodynamics, Structure and Biological Potential. *Scientific Reports* **2018**, *8* (1).
(66) Buske, F. A.; Bauer, D. C.; Mattick, J. S.; Bailey, T. L. Triplexator: Detecting Nucleic Acid Triple Helices in Genomic and Transcriptomic Data. *Genome research* **2012**, *22* (7), 1372–1381.
(67) Suda, T.; Mishima, Y.; Takayanagi, K.; Asakura, H.; Odani, S.; Kominami, R. A Novel Activity of HMG Domains: Promotion of the Triple-Stranded Complex Formation between DNA Containing (GGA/TCC)11 and d(GGA)11 Oligonucleotides. *Nucleic Acids Research* **1996**, *24* (23), 4733–4740.
(68) Jimenez-Garcia, E.; Vaquero, A.; Espinas, M. L.; Soliva, R.; Orozco, M.; Bernues, J.; Azorin, F. The GAGA Factor of Drosophila Binds Triple-Stranded DNA. *The Journal of Biological Chemistry* **1998**, *273* (38), 24640–24648.
(69) Lehmann, M. Anything Else but GAGA: A Nonhistone Protein Complex Reshapes Chromatin Structure. *Trends in Genetics: TIG* **2004**, *20* (1), 15–22.
(70) Nelson, L. D.; Musso, M.; van Dyke, M. W. The Yeast STM1 Gene Encodes a Purine Motif Triple Helical DNA-Binding Protein. *The Journal of biological chemistry* **2000**, *275* (8), 5573–5581. https://doi.org/10.1074/jbc.275.8.5573.
(71) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nature Reviews: Drug Discovery* **2011**, *10* (4), 261–275.
(72) Kusic, J.; Tomic, B.; Divac, A.; Kojic, S. Human Initiation Protein Orc4 Prefers Triple Stranded DNA. *Molecular Biology Reports* **2010**, *37* (5), 2317–2322.
(73) Sommers, J. A.; Rawtani, N.; Gupta, R.; Bugreev, D. v; Mazin, A. v; Cantor, S. B.; Brosh Jr, R. M. FANCJ Uses Its Motor ATPase to Destabilize Protein-DNA Complexes, Unwind Triplexes, and Inhibit RAD51 Strand Exchange. *The Journal of Biological Chemistry* **2009**, *284* (12), 7505–7517.
(74) Brosh, R. M. J.; Majumdar, A.; Desai, S.; Hickson, I. D.; Bohr, V. A.; Seidman, M. M. Unwinding of a DNA Triple Helix by the Werner and Bloom Syndrome Helicases. *The Journal of Biological Chemistry* **2001**, *276* (5), 3024–3030.

(75) Chu, W. K.; Hickson, I. D. RecQ Helicases: Multifunctional Genome Caretakers. *Nature Reviews: Cancer* **2009**, *9* (9), 644–654.
(76) Guo, M.; Hundseth, K.; Ding, H.; Vidhyasagar, V.; Inoue, A.; Nguyen, C.-H.; Zain, R.; Lee, J. S.; Wu, Y. A Distinct Triplex DNA Unwinding Activity of ChlR1 Helicase. *The Journal of Biological Chemistry* **2015**, *290* (8), 5174–5189.
(77) Thoma, B. S.; Wakasugi, M.; Christensen, J.; Reddy, M. C.; Vasquez, K. M. Human XPC-HHR23B Interacts with XPA-RPA in the Recognition of Triplex-Directed Psoralen DNA Interstrand Crosslinks. *Nucleic Acids Research* **2005**, *33* (9), 2993–3001.
(78) Lee, J. S.; Burkholder, G. D.; Latimer, L. J.; Haug, B. L.; Braun, R. P. A Monoclonal Antibody to Triplex DNA Binds to Eucaryotic Chromosomes. *Nucleic Acids Research* **1987**, *15* (3), 1047–1061.
(79) Agazie, Y. M.; Lee, J. S.; Burkholder, G. D. Characterization of a New Monoclonal Antibody to Triplex DNA and Immunofluorescent Staining of Mammalian Chromosomes. *The Journal of Biological Chemistry* **1994**, *269* (9), 7019–7023.
(80) Agazie, Y. M.; Burkholder, G. D.; Lee, J. S. Triplex DNA in the Nucleus: Direct Binding of Triplex-Specific Antibodies and Their Effect on Transcription, Replication and Cell Growth. *The Biochemical Journal* **1996**, *316* (2), 461–466.
(81) Gorab, E. Triple-Helical DNA in Drosophila Heterochromatin. *Cells* **2018**, *7* (12), 227.
(82) Gorab, E.; Pearson, P. L. Thiazole Orange as an Alternative to Antibody Binding for Detecting Triple-Helical DNA in Heterochromatin of Drosophila and Rhynchosciara. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society* **2018**, *66* (3), 143–154.
(83) Rhodes, D.; Lipps, H. J. G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Research* **2015**, *43* (18), 8627–8637.
(84) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Research* **2006**, *34* (19), 5402–5415.
(85) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. Stability and Kinetics of G-Quadruplex Structures. *Nucleic Acids Research* **2008**, *36* (17), 5482–5515.
(86) Miyoshi, D.; Karimata, H.; Sugimoto, N. Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions. *Journal of the American Chemical Society* **2006**, *128* (24), 7957–7963.
(87) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Research* **2005**, *33* (9), 2908–2916.
(88) Besnard, E.; Babled, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J.-M.; Lemaitre, J.-M. Unraveling Cell Type-Specific and Reprogrammable Human Replication Origin Signatures Associated with G-Quadruplex Consensus Motifs. *Nature Structural & Molecular Biology* **2012**, *19* (8), 837–844.
(89) Cayrou, C.; Gregoire, D.; Coulombe, P.; Danis, E.; Mechali, M. Genome-Scale Identification of Active DNA Replication Origins. *Methods* **2012**, *57* (2), 158–164.
(90) 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. Telomeric G-Quadruplexes Are a

Substrate and Site of Localization for Human Telomerase. *Nature Communications* **2015**, *6* (1).
(91) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen Base Pairs in Canonical Duplex DNA. *Nature* **2011**, *470*, 498–502.
(92) Oppenheim, A. Separation of Closed Circular DNA from Linear DNA by Electrophoresis in Two Dimensions in Agarose Gels. *Nucleic Acids Research* **1981**, *9* (24), 6805–6812.
(93) Jain, A.; Rajeswari, M. R.; Ahmed, F. Formation and Thermodynamic Stability of Intermolecular (R*R*Y) DNA Triplex in GAA/TTC Repeats Associated with Freidreich's Ataxia. *Journal of Biomolecular Structure & Dynamics* **2002**, *19* (4), 691–699.
(94) Fox, K. R.; Brown, T. Formation of Stable DNA Triplexes. *Biochemical Society Transactions* **2011**.
(95) Carmona, P.; Molina, M. Binding of Oligonucleotides to a Viral Hairpin Forming RNA Triplexes with Parallel G*G*C Triplets. *Nucleic Acids Research* **2002**.
(96) Tirogoe, H.; Nakagawa, O.; Imanishi, T. Chemical Modification of Triplex-Forming Oligonucleotide to Promote Pyrimidine Motif Triplex Formation at Physiological PH. *Biochimie* **2012**, *94* (4), 1032–1040.
(97) Beck, A.; Vijayanathan, V.; Thomas, T.; Thomas, T. J. Ionic Microenvironmental Effects on Triplex DNA Stabilization: Cationic Counterion Effects on Poly(DT)·poly(DA)·poly(DT). *Biochimie* **2013**, *95* (6).
(98) Dexheimer, T. S.; Sun, D.; Hurley, L. H. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the Bcl-2 P1 Promoter. *Journal of the American Chemical Society* **2006**, *128* (16), 5404–5415.
(99) Sun, D.; Hurley, L. H. Biochemical Techniques for the Characterization of G-Quadruplex Structures: EMSA, DMS Footprinting, and DNA Polymerase Stop Assay. *Methods in Molecular Biology* **2010**, *608*, 65–79.
(100) Purdie, N.; Swallows, K. A. Circular Dichroism — I. Theory and Practice. *Trends in Analytical Chemistry* **1990**, *9* (3), 94–97.
(101) Kypr, J.; Kejnovská, I.; Renciuk, D.; Vorlícková, M. Circular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Research* **2009**, *37* (6), 1713–1725.
(102) Khomyakova, E. B.; Gousset, H.; Liquier, J.; Huynh-Dinh, T.; Gouyette, C.; Takahashi, M.; Florentiev, V. L.; Taillandier, E. Parallel Intramolecular DNA Triple Helix with G and T Bases in the Third Strand Stabilized by $Zn^{(2+)}$ Ions. *Nucleic Acids Research* **2000**, *28* (18), 3511–3516.
(103) Gray, D. M.; Hung, S. H.; Johnson, K. H. Absorption and Circular Dichroism Spectroscopy of Nucleic Acid Duplexes and Triplexes. *Methods in Enzymology* **1995**, *246*, 19–34.
(104) He, Y.; Scaria, P. v; Shafer, R. H. Studies on Formation and Stability of the d[G(AG)5]*d[G(AG)5]. d[C(TC)5] and d[G(TG)5]*d[G(AG)5].d[C(TC)5] Triple Helices. *Biopolymers* **1997**, *41* (4), 431–441.
(105) Gray, D. M.; Wen, J.-D.; Gray, C. W.; Repges, R.; Repges, C.; Raabe, G.; Fleischhauer, J. Measured and Calculated CD Spectra of G-Quartets Stacked with the Same or Opposite Polarities. *Chirality* **2008**, *20* (3–4), 431–440.

(106) Miyoshi, D.; Nakao, A.; Sugimoto, N. Structural Transition from Antiparallel to Parallel G-Quadruplex of d(G4T4G4) Induced by Ca2+. *Nucleic Acids Research* **2003**, *31* (4), 1156–1163.
(107) Zhou, B.; Geng, Y.; Liu, C.; Miao, H.; Ren, Y.; Xu, N.; Shi, X.; You, Y.; Lee, T.; Zhu, G. Characterizations of Distinct Parallel and Antiparallel G-Quadruplexes Formed by Two-Repeat ALS and FTD Related GGGGCC Sequence. *Scientific Reports* **2018**, *8* (1).
(108) Eilert, T.; Kallis, E.; Nagy, J.; Rocker, C.; Michaelis, J. Complete Kinetic Theory of FRET. *The Journal of Physical Chemistry. B* **2018**, *122* (49), 11677–11694.
(109) Reither, S.; Jeltsch, A. Specificity of DNA Triple Helix Formation Analyzed by a FRET Assay. *BMC Biochemistry* **2002**, *3* (1), 27.
(110) Schneider, U. v; Severinsen, J. K.; Geci, I.; Okkels, L. M.; Johnk, N.; Mikkelsen, N. D.; Klinge, T.; Pedersen, E. B.; Westh, H.; Lisby, G. A Novel FRET Pair for Detection of Parallel DNA Triplexes by the LightCycler. *BMC Biotechnology* **2010**, *10*, 4.
(111) Lee, I.-B.; Hong, S.-C.; Lee, N.-K.; Johner, A. Kinetics of the Triplex-Duplex Transition in DNA. *Biophysical Journal* **2012**, *103* (12), 2492–2501.
(112) Shirude, P. S.; Balasubramanian, S. Single Molecule Conformational Analysis of DNA G-Quadruplexes. *Biochimie* **2008**, *90* (8), 1197–1206.
(113) Klejevskaja, B.; Pyne, A. L. B.; Reynolds, M.; Shivalingam, A.; Thorogate, R.; Hoogenboom, B. W.; Ying, L.; Vilar, R. Studies of G-Quadruplexes Formed within Self-Assembled DNA Mini-Circles. *Chemical Communications* **2016**, *52* (84), 12454–12457.
(114) Dufrêne, Y. F.; Ando, T.; Garcia, R.; Alsteens, D.; Martinez-Martin, D.; Engel, A.; Gerber, C.; Müller, D. J. Imaging Modes of Atomic Force Microscopy for Application in Molecular and Cell Biology. *Nature Nanotechnology* **2017**, *12* (4), 295–307.
(115) Müller, D. J.; Dufrêne, Y. F. Atomic Force Microscopy: A Nanoscopic Window on the Cell Surface. *Trends in Cell Biology* **2011**, *21* (8), 461–469.
(116) Kada, G.; Kienberger, F.; Hinterdorfer, P. Atomic Force Microscopy in Bionanotechnology. *Nano Today* **2008**, *3* (1), 12–19.
(117) Tiner, W. J.; Potaman, V. N.; Sinden, R. R.; Lyubchenko, Y. L. The Structure of Intramolecular Triplex DNA: Atomic Force Microscopy Study. *Journal of Molecular Biology* **2001**, *314* (3), 353–357.
(118) Klinov, D.; Dwir, B.; Kapon, E.; Borovok, N.; Molotsky, T.; Kotlyar, A. High-Resolution Atomic Force Microscopy of Duplex and Triplex DNA Molecules. *Nanotechnology* **2007**, *18* (22).
(119) Potaman, V. N.; Oussatcheva, E. A.; Lyubchenko, Y. L.; Shlyakhtenko, L. S.; Bidichandani, S. I.; Ashizawa, T.; Sinden, R. R. Length-Dependent Structure Formation in Friedreich Ataxia (GAA)N*(TTC)n Repeats at Neutral PH. *Nucleic Acids Research* **2004**, *32* (3), 1224–1231.
(120) Berezin, M. Y.; Achilefu, S. Fluorescence Lifetime Measurements and Biological Imaging. *Chemical Reviews* **2010**, *110* (5), 2641–2684.
(121) Goh, W. L.; Lee, M. Y.; Lim, T. X.; Chua, J. S.; Brenner, S.; Ghadessy, F. J.; Teo, Y. N. A Novel Molecular Rotor Facilitates Detection of P53-DNA Interactions Using the Fluorescent Intercalator Displacement Assay. *Scientific Reports* **2018**, *8* (1).

(122) Kapuscinski, J. DAPI: A DNA-Specific Fluorescent Probe. *Biotechnic & Histochemistry: Official Publication of the Biological Stain Commission* **1995**, *70* (5), 220–233.
(123) Biancardi, A.; Biver, T.; Secco, F.; Mennucci, B. An Investigation of the Photophysical Properties of Minor Groove Bound and Intercalated DAPI through Quantum-Mechanical and Spectroscopic Tools. *Physical Chemistry Chemical Physics: PCCP* **2013**, *15* (13), 4596–4603.
(124) Hard, T.; Fan, P.; Kearns, D. R. A Fluorescence Study of the Binding of Hoechst 33258 and DAPI to Halogenated DNAs. *Photochemistry and Photobiology* **1990**, *51* (1), 77–86.
(125) on der Haar, M.; Heuer, C.; Pähler, M.; von der Haar, K.; Lindner, P.; Scheper, T.; Stahl, F. Optimization of Cyanine Dye Stability and Analysis of FRET Interaction on DNA Microarrays. *Biology* **2016**, *5* (4), 47.
(126) Sailer, B. L.; Steinkamp, J. A.; Crissman, H. A. Flow Cytometric Fluorescence Lifetime Analysis of DNA-Binding Probes. *European Journal of Histochemistry* **1998**, *42*, 19–27.
(127) Biancardi, A.; Biver, T.; Marini, A.; Mennucci, B.; Secco, F. Thiazole Orange (TO) as a Light-Switch Probe: A Combined Quantum-Mechanical and Spectroscopic Study. *Phys. Chem. Chem. Phys.* **2011**, *13* (27), 12595–12602. https://doi.org/10.1039/C1CP20812H.
(128) Nygren, J.; Svanvik, N.; Kubista, M. The Interactions between the Fluorescent Dye Thiazole Orange and DNA. *Biopolymers* **1998**, *46* (1), 39–51.
(129) Spielmann, H. P.; Wemmer, D. E.; Jacobsen, J. P. Solution Structure of a DNA Complex with the Fluorescent Bis-Intercalator TOTO Determined by NMR Spectroscopy. *Biochemistry* **1995**, *34* (27), 8542–8553.
(130) Lubitz, I.; Zikich, D.; Kotlyar, A. Specific High-Affinity Binding of Thiazole Orange to Triplex and G-Quadruplex DNA. *Biochemistry* **2010**, *49* (17), 3567–3574.
(131) Lu, Y.-J.; Deng, Q.; Hou, J.-Q.; Hu, D.-P.; Wang, Z.-Y.; Zhang, K.; Luyt, L. G.; Wong, W.-L.; Chow, C.-F. Molecular Engineering of Thiazole Orange Dye: Change of Fluorescent Signalling from Universal to Specific upon Binding with Nucleic Acids in Bioassay. *ACS Chemical Biology* **2016**, *11* (4), 1019–1029.
(132) Umar, M. I.; Ji, D.; Chan, C. Y.; Kwok, C. K. G-Quadruplex-Based Fluorescent Turn-on Ligands and Aptamers: From Development to Applications. *Molecules* **2019**, *24* (13).
(133) Ma, D.-L.; Che, C.-M.; Yan, S.-C. Platinum(II) Complexes with Dipyridophenazine Ligands as Human Telomerase Inhibitors and Luminescent Probes for G-Quadruplex DNA. *Journal of the American Chemical Society* **2009**, *131* (5), 1835–1846.
(134) Scaria, P. v; Shafer, R. H. Binding of Ethidium Bromide to a DNA Triple Helix. Evidence for Intercalation. *The Journal of Biological Chemistry* **1991**, *266* (9), 5417–5423.
(135) Sinha, R.; Saha, I.; Kumar, G. S. Protoberberine Alkaloids Berberine, Palmatine, and Coralyne Binding to Poly(DT)·(Poly(DA)·Poly(DT)) Triplex: Comparative Structural Aspects and Energetics Profiles of the Interaction. *Chemistry & Biodiversity* **2011**, *8* (8), 1512–1528.
(136) Basu, A.; Jaisankar, P.; Kumar, G. S. Binding of Novel 9-O-N-Aryl/Arylalkyl Amino Carbonyl Methyl Berberine Analogs to Poly(U)-Poly(A).Poly(U) Triplex

and Comparison to the Duplex Poly(A)-Poly(U). *Molecular Biology Reports* **2014**, *41* (8), 5473–5483.
(137) Lee, J. S.; Latimer, L. J.; Hampel, K. J. Coralyne Binds Tightly to Both T.A.T- and C.G.C(+)-Containing DNA Triplexes. *Biochemistry* **1993**, *32* (21), 5591–5597.
(138) Escude, C.; Nguyen, C. H.; Kukreti, S.; Janin, Y.; Sun, J. S.; Bisagni, E.; Garestier, T.; Helene, C. Rational Design of a Triple Helix-Specific Intercalating Ligand. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95* (7), 3591–3596.
(139) Amiri, H.; Nekhotiaeva, N.; Sun, J.-S.; Nguyen, C.-H.; Grierson, D. S.; Good, L.; Zain, R. Benzoquinoquinoxaline Derivatives Stabilize and Cleave H-DNA and Repress Transcription Downstream of a Triplex-Forming Sequence. *Journal of Molecular Biology* **2005**, *351* (4), 776—783.
(140) Zain, R.; Polverari, D.; Nguyen, C.-H.; Blouquit, Y.; Bisagni, E.; Garestier, T.; Grierson, D. S.; Sun, J.-S. Optimization of Triple-Helix-Directed DNA Cleavage by Benzoquinoquinoxaline-Ethylenediaminetetraacetic Acid Conjugates. *ChemBiochem* **2003**, *4* (9), 856–862.
(141) Bergquist, H.; Nikravesh, A.; Fernández, R. D.; Larsson, V.; Nguyen, C. H.; Good, L.; Zain, R. Structure-Specific Recognition of Friedreich's Ataxia (GAA)n Repeats by Benzoquinoquinoxaline Derivatives. *ChemBioChem* **2009**.
(142) Walsh, S.; El-Sagheer, A. H.; Brown, T. Fluorogenic Thiazole Orange TOTFO Probes Stabilise Parallel DNA Triplexes at PH 7 and Above. *Chemical Science* **2018**, *9* (39), 7681–7687.
(143) Lohani, N.; Rajeswari, M. R. Preferential Binding of Anticancer Drugs to Triplex DNA Compared to Duplex DNA: A Spectroscopic and Calorimetric Study. *RSC Advances* **2016**, *6* (46), 39903–39917.
(144) Chiou, C.-C.; Chen, S.-W.; Luo, J.-D.; Chien, Y.-T. Monitoring Triplex DNA Formation with Fluorescence Resonance Energy Transfer between a Fluorophore-Labeled Probe and Intercalating Dyes. *Analytical Biochemistry* **2011**, *416* (1), 1–7.
(145) Lauria, A.; Montalbano, A.; Barraja, P.; Dattolo, G.; Almerico, A. M. DNA Minor Groove Binders: An Overview on Molecular Modeling and QSAR Approaches. *Current Medicinal Chemistry* **2007**, *14* (20), 2136–2160.
(146) Durand, M.; Nguyen Thanh Thuong; Maurizot, J. C. Binding of Netropsin to a DNA Triple Helix. *Journal of Biological Chemistry* **1992**, *267* (34), 24394–24399.
(147) Arya, D. P.; Micovic, L.; Charles, I.; Coffee, R. L. J.; Willis, B.; Xue, L. Neomycin Binding to Watson-Hoogsteen (W-H) DNA Triplex Groove: A Model. *Journal of the American Chemical Society* **2003**, *125* (13), 3733–3744.
(148) Xue, L.; Xi, H.; Kumar, S.; Gray, D.; Davis, E.; Hamilton, P.; Skriba, M.; Arya, D. P. Probing the Recognition Surface of a DNA Triplex: Binding Studies with Intercalator-Neomycin Conjugates. *Biochemistry* **2010**, *49* (26), 5540–5552.
(149) Dziuba, D.; Jurkiewicz, P.; Cebecauer, M.; Hof, M.; Hocek, M. A Rotational BODIPY Nucleotide: An Environment-Sensitive Fluorescence-Lifetime Probe for DNA Interactions and Applications in Live-Cell Microscopy. *Angewandte Chemie International Edition* **2016**, *55* (1), 174–178.
(150) Bhavsar, Y. P.; Reilly, S. M.; Wadkins, R. M. Evaluation of Fluorescent Analogs of Deoxycytidine for Monitoring DNA Transitions from Duplex to Functional Structures. *Journal of Nucleic Acids* **2011**, *2011*.

(151) Shivalingam, A.; Izquierdo, M. A.; Marois, A. le; Vyšniauskas, A.; Suhling, K.; Kuimova, M. K.; Vilar, R. The Interactions between a Small Molecule and G-Quadruplexes Are Visualized by Fluorescence Lifetime Imaging Microscopy. *Nature Communications* **2015**, *6*, 8178.
(152) Margineanu, A.; Chan, J. J.; Kelly, D. J.; Warren, S. C.; Flatters, D.; Kumar, S.; Katan, M.; Dunsby, C. W.; French, P. M. W. Screening for Protein-Protein Interactions Using Förster Resonance Energy Transfer (FRET) and Fluorescence Lifetime Imaging Microscopy (FLIM). *Scientific Reports* **2016**, *6* (1), 28186.
(153) Estandarte, A. K.; Botchway, S.; Lynch, C.; Yusuf, M.; Robinson, I. The Use of DAPI Fluorescence Lifetime Imaging for Investigating Chromatin Condensation in Human Chromosomes. *Scientific Reports* **2016**, *6* (1).
(154) Chan, P. P.; Glazer, P. M. Triplex DNA: Fundamentals , Advances , and Potential Applications for Gene Therapy. *J. Mol. Med.* **1997**, *75* (4), 267–282.
(155) Stellwagen, E.; Muse, J. M.; Stellwagen, N. C. Monovalent Cation Size and DNA Conformational Stability. *Biochemistry* **2011**, *50* (15), 3084–3094.
(156) Šponer, J.; Sabat, M.; Burda, J. v; Doody, A. M.; Leszczynski, J.; Hobza, P. Stabilization of the Purine•Purine•Pyrimidine DNA Base Triplets by Divalent Metal Cations. *Journal of Biomolecular Structure and Dynamics* **1998**, *16* (1), 139–143.
(157) Spink, C. H.; Chaires, J. B. Selective Stabilization of Triplex DNA by Poly(Ethylene Glycols). *Journal of the American Chemical Society* **1995**, *117* (51), 12887–12888.
(158) Gondeau, C.; Maurizot, J. C.; Durand, M. Circular Dichroism and UV Melting Studies on Formation of an Intramolecular Triplex Containing Parallel T*A:T and G*G:C Triplets: Netropsin Complexation with the Triplex. *Nucleic Acids Research* **1998**, *26* (21), 4996–5003.
(159) Hoyne, P. R.; Gacy, A. M.; McMurray, C. T.; Maher, L. J. 3rd. Stabilities of Intrastrand Pyrimidine Motif DNA and RNA Triple Helices. *Nucleic Acids Research* **2000**, *28* (3), 770–775.
(160) Chandler, S. P.; Fox, K. R. Specificity of Antiparallel DNA Triple Helix Formation. *Biochemistry* **1996**, *35* (47), 15038–15048.
(161) Keppler, M. D.; Neidle, S.; Fox, K. R. Stabilisation of TG-and AG-Containing Antiparallel DNA Triplexes by Triplex-Binding Ligands. *Nucleic Acids Research* **2001**, *29* (9), 1935–1942.
(162) Koshlap, K. M.; Schultze, P.; Brunar, H.; Dervan, P. B.; Feigon, J. Solution Structure of an Intramolecular DNA Triplex Containing an N7-Glycosylated Guanine Which Mimics a Protonated Cytosine. *Biochemistry* **1997**, *36* (9), 2659–2668.
(163) Olivas, W. M.; Maher, L. J. Competitive Triplex/Quadruplex Equilibria Involving Guanine-Rich Oligonucleotides. *Biochemistry* **1995**, *34* (1), 278–284.
(164) Cheng, A. J.; van Dyke, M. W. Monovalent Cation Effects on Intermolecular Purine-Purine-Pyrimidine Triple-Helix Formation. *Nucleic Acids Research* **1993**, *21* (24), 5630–5635.
(165) Li, W.; Nordenskiold, L.; Mu, Y. Sequence-Specific Mg2+-DNA Interactions: A Molecular Dynamics Simulation Study. *The Journal of Physical Chemistry. B* **2011**, *115* (49), 14713–14720.

(166) Malkov, V. A.; Voloshin, O. N.; Soyfer, V. N.; Frank-Kamenetskii, M. D. Cation and Sequence Effects on Stability of Intermolecular Pyrimidine-Purine-Purine Triplex. *Nucleic Acids Research* **1993**, *21* (3), 585–591.
(167) Miyoshi, D.; Nakamura, K.; Muhuli, S.; Karimata, H. T.; Sugimoto, N. Thermodynamics of DNA Structures under Molecular Crowding Conditions with Neutral and Positive Charged Cosolutes. *Nucleic Acids Symposium Series (2004)* **2008**, *52*, 413–414.
(168) Kunkler, C. N.; Hulewicz, J. P.; Hickman, S. C.; Wang, M. C.; McCown, P. J.; Brown, J. A. Stability of an RNA•DNA–DNA Triple Helix Depends on Base Triplet Composition and Length of the RNA Third Strand. *Nucleic Acids Research* **2019**, *47* (14), 7213–7222.
(169) Grabczyk, E.; Mancuso, M.; Sammarco, M. C. A Persistent RNA.DNA Hybrid Formed by Transcription of the Friedreich Ataxia Triplet Repeat in Live Bacteria, and by T7 RNAP in Vitro. *Nucleic Acids Research* **2007**, *35* (16), 5351–5359.
(170) Soragni, E.; Herman, D.; Dent, S. Y. R.; Gottesfeld, J. M.; Wells, R. D.; Napierala, M. Long Intronic GAA*TTC Repeats Induce Epigenetic Changes and Reporter Gene Silencing in a Molecular Model of Friedreich Ataxia. *Nucleic Acids Research* **2008**, *36* (19), 6056–6065.
(171) Sharma, R.; Bhatti, S.; Gomez, M.; Clark, R. M.; Murray, C.; Ashizawa, T.; Bidichandani, S. I. The GAA Triplet-Repeat Sequence in Friedreich Ataxia Shows a High Level of Somatic Instability in Vivo, with a Significant Predilection for Large Contractions. *Human Molecular Genetics* **2002**, *11* (18), 2175–2187.
(172) de Bustos, A.; Cuadrado, A.; Jouve, N. Sequencing of Long Stretches of Repetitive DNA. *Scientific Reports* **2016**, *6* (1).
(173) Bzymek, M.; Lovett, S. T. Instability of Repetitive DNA Sequences: The Role of Replication in Multiple Mechanisms. *Proceedings of the National Academy of Sciences* **2001**, *98* (15), 8319–8325.
(174) Pollard, L. M.; Sharma, R.; Gómez, M.; Shah, S.; Delatycki, M. B.; Pianese, L.; Monticelli, A.; Keats, B. J. B.; Bidichandani, S. I. Replication-Mediated Instability of the GAA Triplet Repeat Mutation in Friedreich Ataxia. *Nucleic Acids Research* **2004**, *32* (19), 5962–5971.
(175) Bhagya, C. H. W. M. R. C.; Wijesundera Sulochana, W. S.; Hemamali, N. P. Polymerase Chain Reaction Optimization for Amplification of Guanine-Cytosine Rich Templates Using Buccal Cell DNA. *Indian Journal of Human Genetics* **2013**, *19* (1), 78–83.
(176) Viguera, E.; Canceill, D.; Ehrlich, S. D. Replication Slippage Involves DNA Polymerase Pausing and Dissociation. *The EMBO Journal* **2001**, *20* (10), 2587–2595.
(177) Witz, G.; Stasiak, A. DNA Supercoiling and Its Role in DNA Decatenation and Unknotting. *Nucleic Acids Research* **2010**, *38* (7), 2119–2133.
(178) Bacolla, A.; Wang, G.; Vasquez, K. M. New Perspectives on DNA and RNA Triplexes as Effectors of Biological Activity. *PLoS Genetics* **2015**, *11* (12).
(179) Belotserkovskii, B. P.; Soo Shin, J. H.; Hanawalt, P. C. Strong Transcription Blockage Mediated by R-Loop Formation within a G-Rich Homopurine-Homopyrimidine Sequence Localized in the Vicinity of the Promoter. *Nucleic Acids Research* **2017**, *45* (11), 6589–6599.
(180) Belotserkovskii, B. P.; Liu, R.; Tornaletti, S.; Krasilnikova, M. M.; Mirkin, S. M.; Hanawalt, P. C. Mechanisms and Implications of Transcription Blockage by

Guanine-Rich DNA Sequences. *Proceedings of the National Academy of Sciences* **2010**, *107* (29), 12816–12821.
(181) Andrade-Lima, L. C.; Veloso, A.; Ljungman, M. Transcription Blockage Leads to New Beginnings. *Biomolecules* **2015**, *5* (3), 1600–1617.
(182) Reddy, K.; Tam, M.; Bowater, R. P.; Barber, M.; Tomlinson, M.; Nichol Edamura, K.; Wang, Y. H.; Pearson, C. E. Determinants of R-Loop Formation at Convergent Bidirectionally Transcribed Trinucleotide Repeats. *Nucleic Acids Research* **2011**.
(183) Gowrishankar, J.; Leela, J. K.; Anupama, K. R-Loops in Bacterial Transcription: Their Causes and Consequences. *Transcription* **2013**, *4* (4), 153–157.
(184) Abdollahi, E.; Taucher-Scholz, G.; Jakob, B. Application of Fluorescence Lifetime Imaging Microscopy of DNA Binding Dyes to Assess Radiation-Induced Chromatin Compaction Changes. *International Journal of Molecular Sciences* **2018**, *19* (8).
(185) Boutorine, A. S.; Novopashina, D. S.; Krasheninina, O. A.; Nozeret, K.; Venyaminova, A. G. Fluorescent Probes for Nucleic Acid Visualization in Fixed and Live Cells. *Molecules* **2013**, *18* (12), 15357–15397.
(186) Arya, D. P. New Approaches toward Recognition of Nucleic Acid Triple Helices. *Accounts of Chemical Research* **2011**, *44* (2), 134–146.
(187) Karunakaran, V.; Perez Lustres, J. L.; Zhao, L.; Ernsting, N. P.; Seitz, O. Large Dynamic Stokes Shift of DNA Intercalation Dye Thiazole Orange Has Contribution from a High-Frequency Mode. *Journal of the American Chemical Society* **2006**, *128* (9), 2954–2962.
(188) Silva, G. L.; Ediz, V.; Yaron, D.; Armitage, B. A. Experimental and Computational Investigation of Unsymmetrical Cyanine Dyes: Understanding Torsionally Responsive Fluorogenic Dyes. *Journal of the American Chemical Society* **2007**, *129* (17), 5710–5718.
(189) Rye, H. S.; Yue, S.; Wemmer, D. E.; Quesada, M. A.; Haugland, R. P.; Mathies, R. A.; Glazer, A. N. Stable Fluorescent Complexes of Double-Stranded DNA with Bis-Intercalating Asymmetric Cyanine Dyes: Properties and Applications. *Nucleic Acids Research* **1992**, *20* (11), 2803–2812.
(190) Petersen, M.; Hamed, A. A.; Pedersen, E. B.; Jacobsen, J. P. Bis-Intercalation of Homodimeric Thiazole Orange Dye Derivatives in DNA. *Bioconjugate Chemistry* **1999**, *10* (1), 66–74.
(191) Shen, R.; Wang, W.; Yang, G. DNA Binding Property and Antitumor Evaluation of Xanthone with Dimethylamine Side Chain. *Journal of Fluorescence* **2014**, *24* (3), 959–966.
(192) Chubb, J. R.; Boyle, S.; Perry, P.; Bickmore, W. A. Chromatin Motion Is Constrained by Association with Nuclear Compartments in Human Cells. *Current Biology: CB* **2002**, *12* (6), 439–445.
(193) Lu, Y.-J.; Deng, Q.; Hu, D.-P.; Wang, Z.-Y.; Huang, B.-H.; Du, Z.-Y.; Fang, Y.-X.; Wong, W.-L.; Zhang, K.; Chow, C.-F. A Molecular Fluorescent Dye for Specific Staining and Imaging of RNA in Live Cells: A Novel Ligand Integration from Classical Thiazole Orange and Styryl Compounds. *Chemical Communications* **2015**, *51* (83), 15241–15244.
(194) Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; et al. Anticancer Activity of CX-

3543: A Direct Inhibitor of RRNA Biogenesis. *Cancer Research* **2009**, *69* (19), 7653–7661.
(195) Shi, Q.-Q.; Sun, R.; Ge, J.-F.; Xu, Q.-F.; Li, N.-J.; Lu, J.-M. A Comparative Study of Symmetrical and Unsymmetrical Trimethine Cyanine Dyes Bearing Benzoxazolyl and Benzothiazolyl Groups. *Dyes and Pigments* **2012**, *93* (1), 1506–1511.
(196) Hu, M.-H.; Guo, R.-J.; Chen, S.-B.; Huang, Z.-S.; Tan, J.-H. Development of an Engineered Carbazole/Thiazole Orange Conjugating Probe for G-Quadruplexes. *Dyes and Pigments* **2017**, *137*, 191–199.
(197) Sun, N.; Du, R.-L.; Zheng, Y.-Y.; Guo, Q.; Cai, S.-Y.; Liu, Z.-H.; Fang, Z.-Y.; Yuan, W.-C.; Liu, T.; Li, X.-M.; et al. Antibacterial Activity of 3-Methylbenzo[d]Thiazol-Methylquinolinium Derivatives and Study of Their Action Mechanism. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2018**, *33* (1), 879–889.
(198) Dhe-Paganon, S.; Shigeta, R.; Chi, Y. I.; Ristow, M.; Shoelson, S. E. Crystal Structure of Human Frataxin. *The Journal of Biological Chemistry* **2000**, *275* (40), 30753–30756.
(199) van Hecke, K.; Nam, P. C.; Nguyen, M. T.; van Meervelt, L. Netropsin Interactions in the Minor Groove of d(GGCCAATTGG) Studied by a Combination of Resolution Enhancement and Ab Initio Calculations. *The FEBS Journal* **2005**, *272* (14), 3531–3541.
(200) Brown, C; Development of novel synthetic DNA-minicircles to study *in cellulo* the molecular basis of Friedreich's ataxia. *MRes Report* **2015**.
(201) Brown, C; Studying the role of triplex DNA in Friedreich's ataxia. *Early Stage Assessment* **2016**.
(202) Nanjunda, R.; Wilson, W. D. Binding to the DNA Minor Groove by Heterocyclic Dications: From AT-Specific Monomers to GC Recognition with Dimers. *Current Protocols in Nucleic Acid Chemistry* **2012**, *Chapter 8*.
(203) Stootman, F. H.; Fisher, D. M.; Rodger, A.; Aldrich-Wright, J. R. Improved Curve Fitting Procedures to Determine Equilibrium Binding Constants. *Analyst* **2006**, *131* (10), 1145–1151.
(204) Swillens, S. Interpretation of Binding Curves Obtained with High Receptor Concentrations: Practical Aid for Computer Analysis. *Molecular Pharmacology* **1995**, *47* (6), 1197–1203.

Appendix

Figure	*Caption*
A.1	^{1}H NMR spectrum of **1** in DMSO-d_6.
A.2	ESI-MS of **1**
A.3	^{1}H NMR spectrum of **2** in DMSO-d_6
A.4	ESI-MS of **2**
A.5	^{1}H NMR spectrum of **3** in DMSO-d_6
A.6	ESI-MS of **3**
A.7	^{1}H NMR spectrum of **4a** in DMSO-d_6
A.8	ESI-MS of **4a**
A.9	^{1}H NMR spectrum of **4b** in DMSO-d_6
A.10	ESI-MS of **4b**
A.11	^{1}H NMR spectrum of **4c** in DMSO-d_6
A.12	ESI-MS of **4c**
A.13	^{1}H NMR spectrum of **4d** in DMSO-d_6
A.14	ESI-MS of **4d**
A.15	^{1}H NMR spectrum of **4e** in DMSO-d_6
A.16	ESI-MS of **4e**
A.17	A) Example emission spectrum of **4a**, with pyrimidine triplex at pH 5.5, B) example absorption spectrum of **4a**, with pyrimidine triplex at pH 5.5, C) binding curve of **4a** with different DNA topologies at pH 5.5, D) binding curve of **4a** with different DNA topologies at pH 7.4.
A.18	A) Example emission spectrum of **4c**, with pyrimidine triplex at pH 5.5, B) example absorption spectrum of **4c**, with pyrimidine triplex at pH 5.5, C) binding curve of **4c** with different DNA topologies at pH 5.5, D) binding curve of **4c** with different DNA topologies at pH 7.4.
A.19	Example emission spectrum of **4d**, with pyrimidine triplex at pH 5.5, B) example absorption spectrum of **4d**, with pyrimidine triplex at pH 5.5, C) binding curve of **4d** with different DNA topologies at pH 5.5, D) binding curve of **4d** with different DNA topologies at pH 7.4.
A.20	Example emission spectrum of **4e**, with pyrimidine triplex at pH 5.5, B) example absorption spectrum of **4e**, with pyrimidine triplex at pH 5.5, C) binding curve of **4e** with different DNA topologies at pH 5.5, D) binding curve of **4e** with different DNA topologies at pH 7.4.

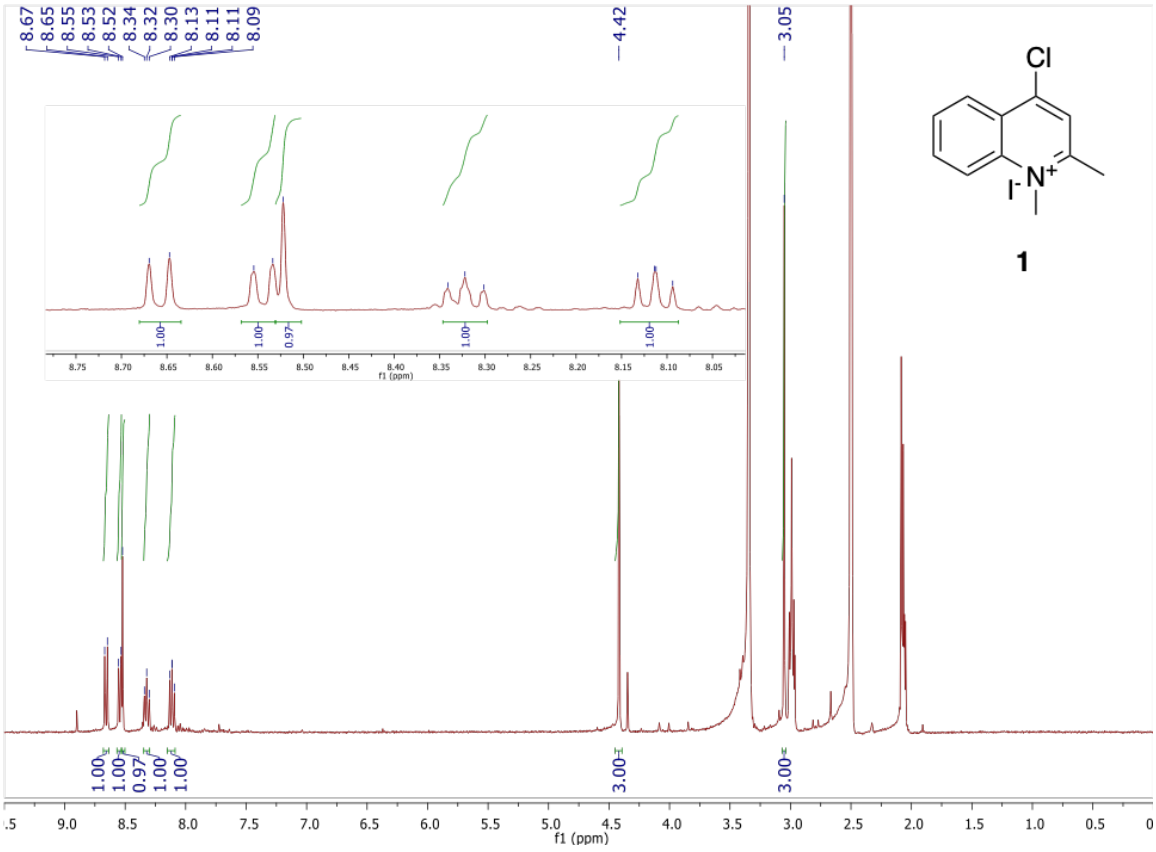


Fig. A.1: ^{1}H NMR spectrum of **1** in DMSO-d_6.

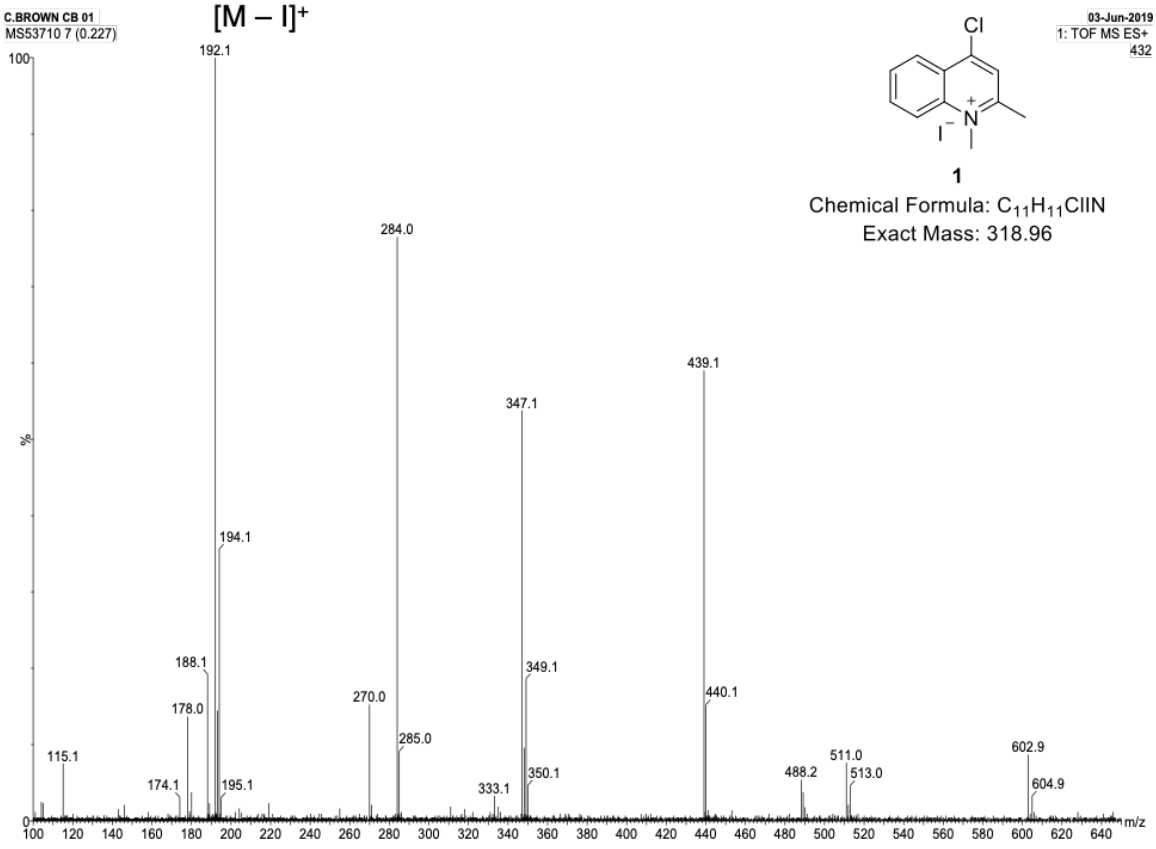


Fig. A.2: ESI-MS of **1**.

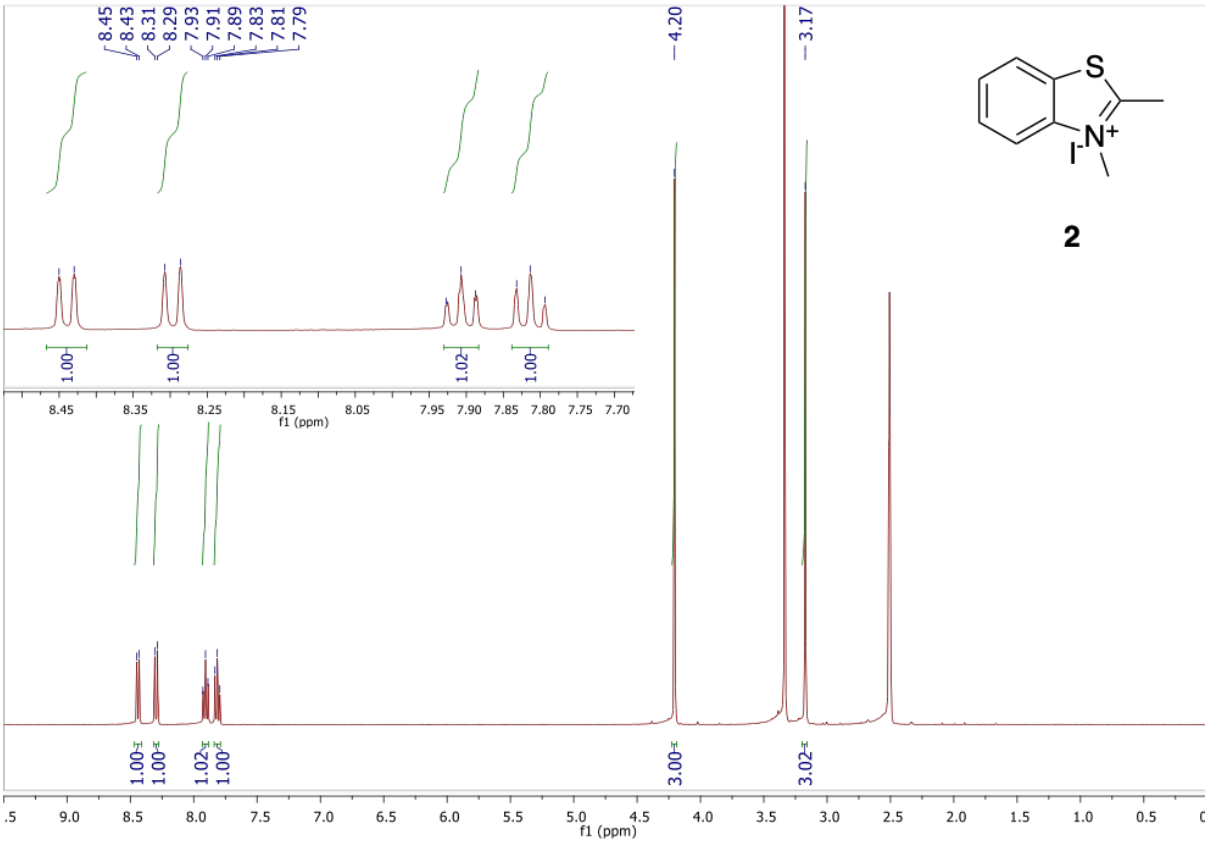


Fig. A.3: ^{1}H NMR spectrum of **2** in DMSO-d$_6$.

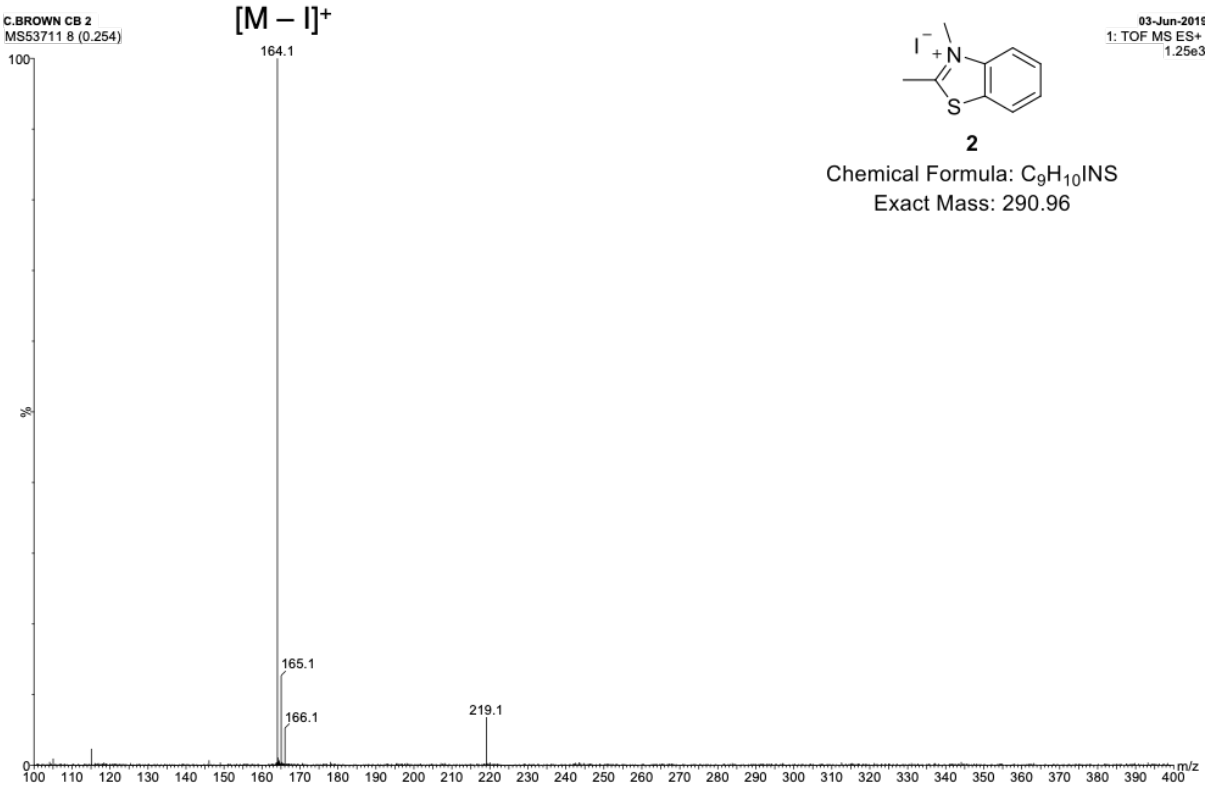


Fig. A.4: ESI-MS of **2**.

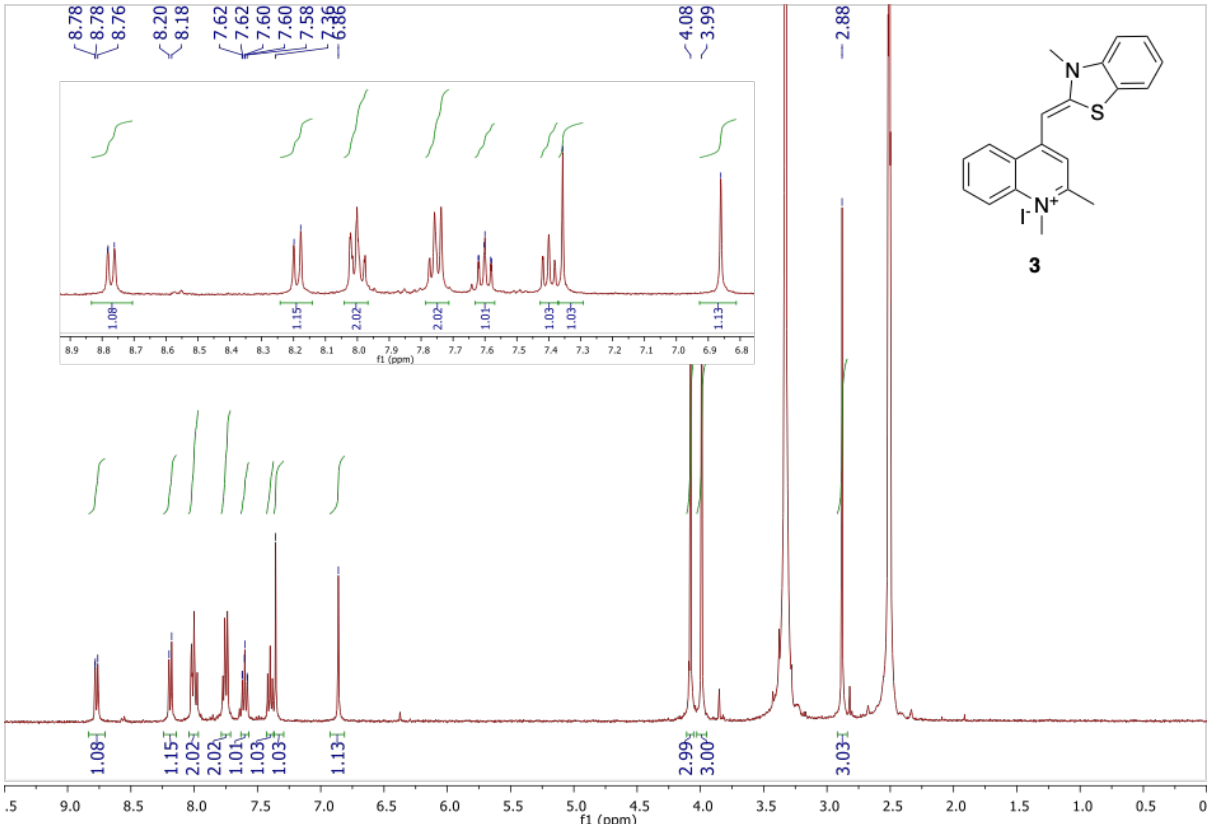


Fig. A.5: ^{1}H NMR spectrum of **3** in DMSO-d_6.

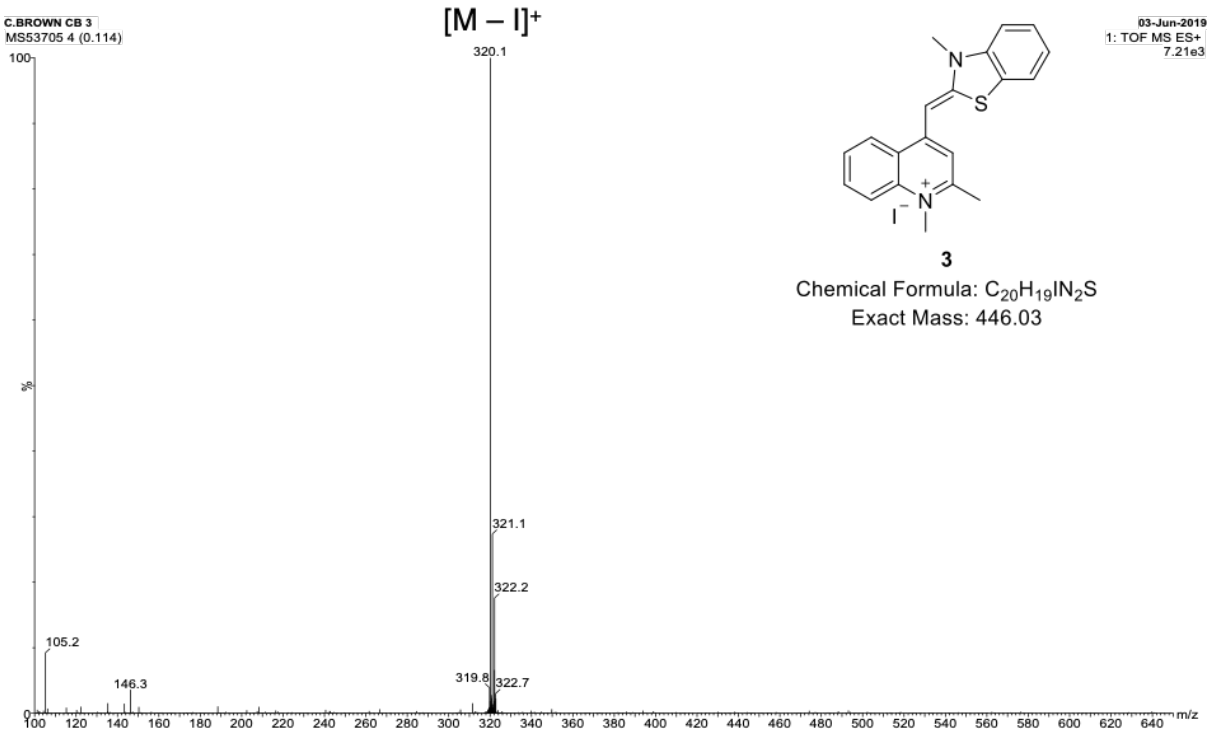


Fig. A.6: ESI-MS of **3**.

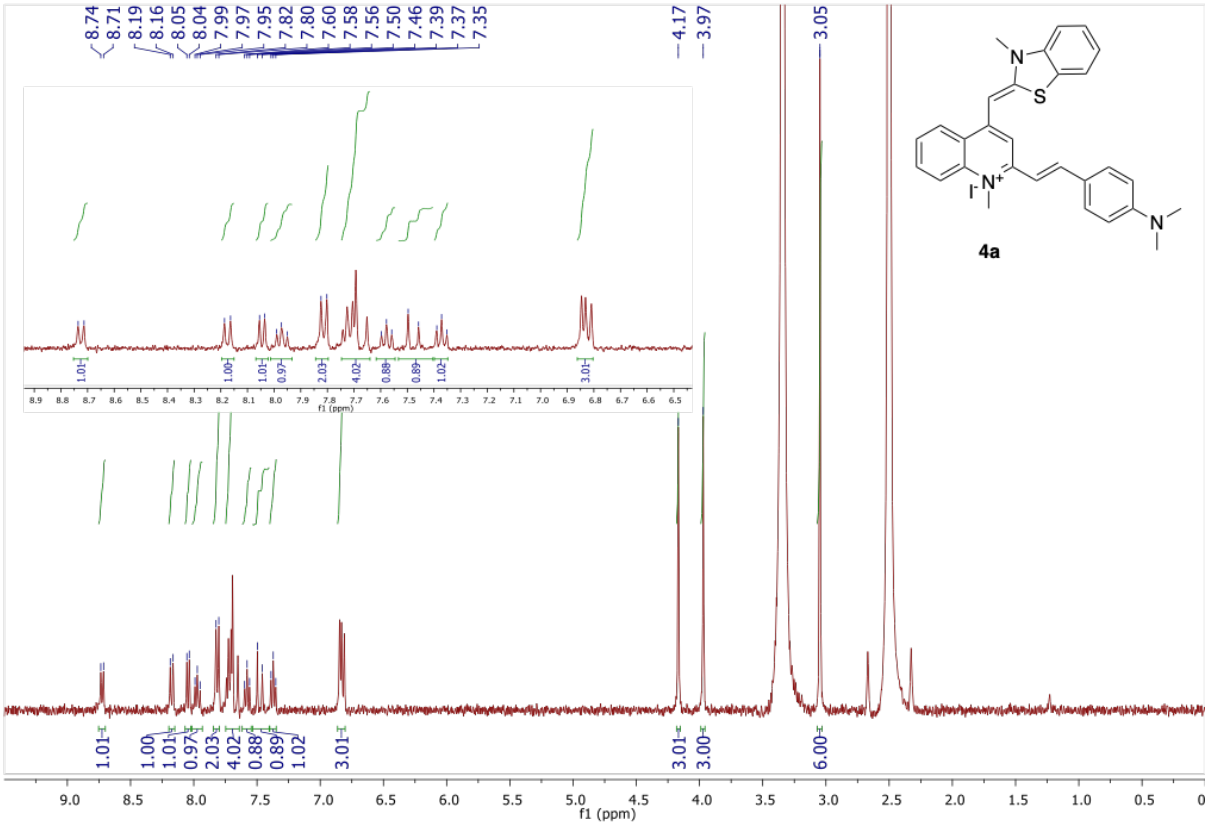


Fig. A.7: ^{1}H NMR spectrum of **4a** in DMSO-d_6.

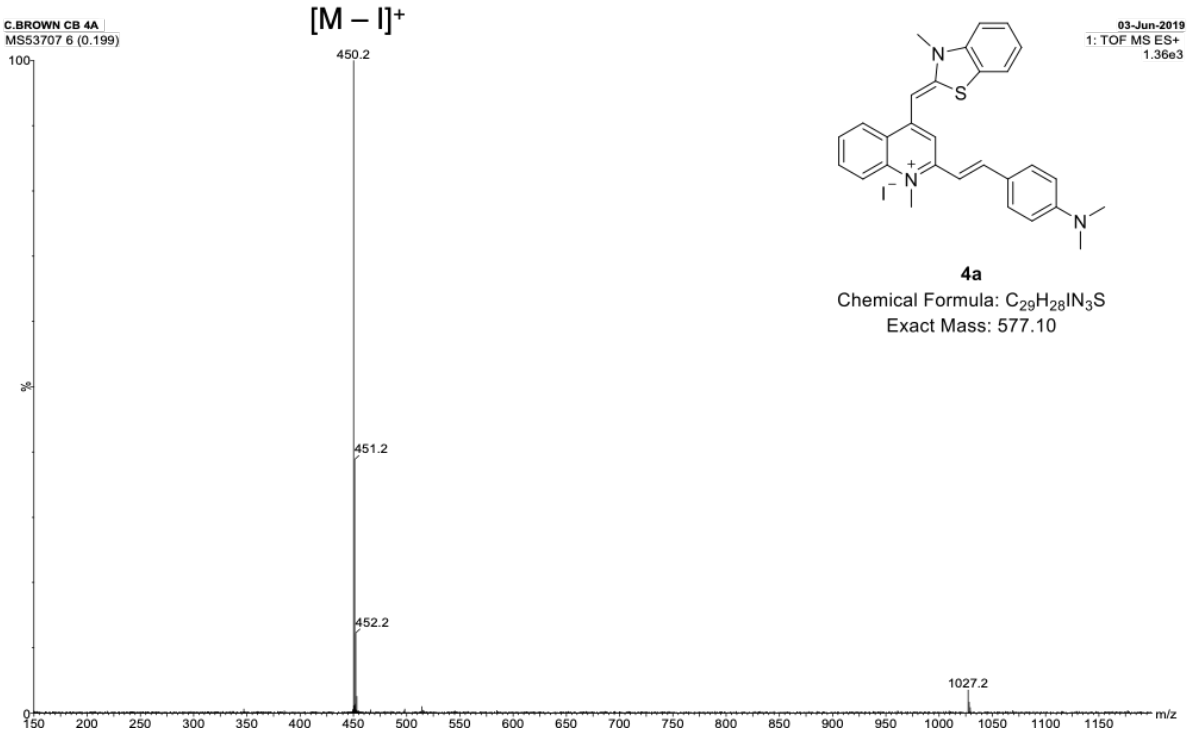


Fig. A.8: ESI-MS of **4a**.

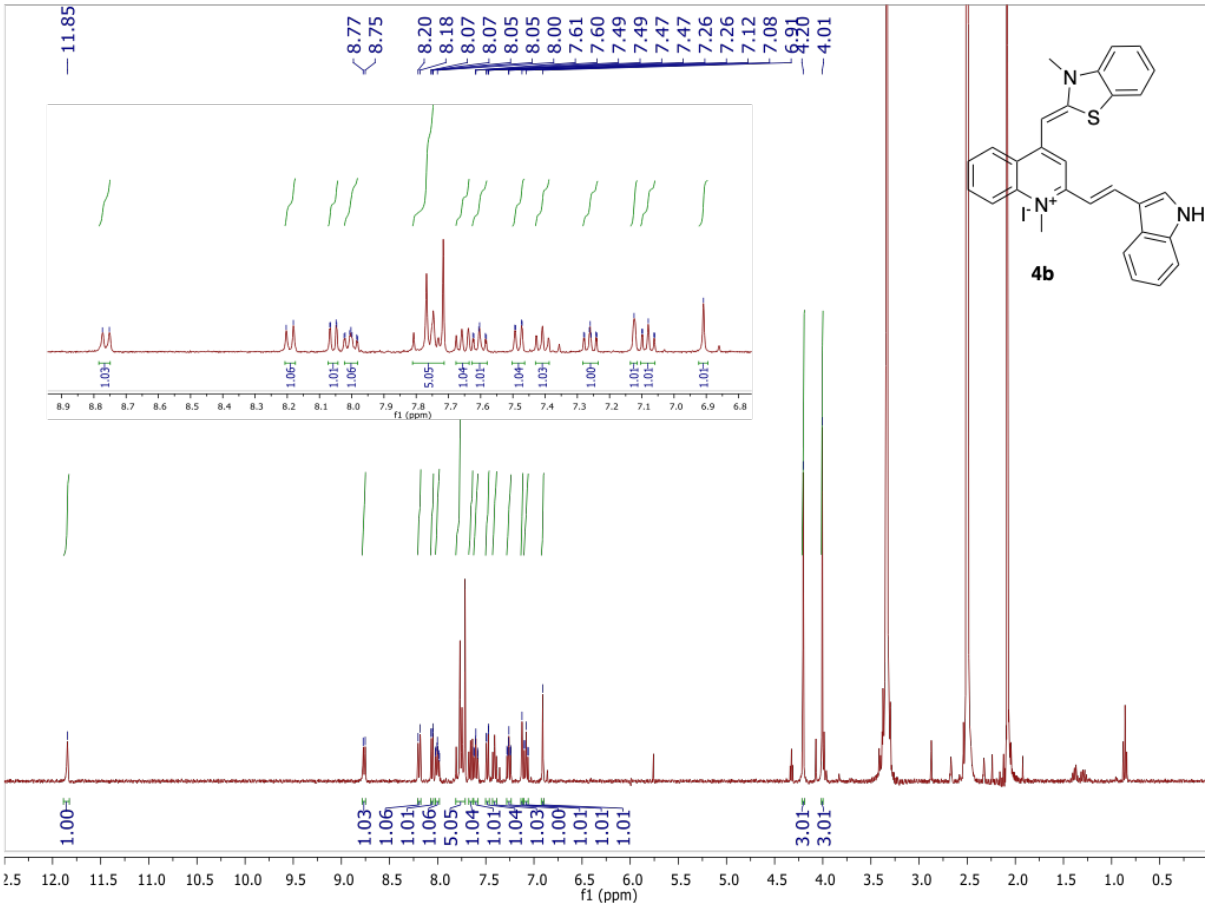


Fig. A.9: ^{1}H NMR spectrum of **4b** in DMSO-d_6.

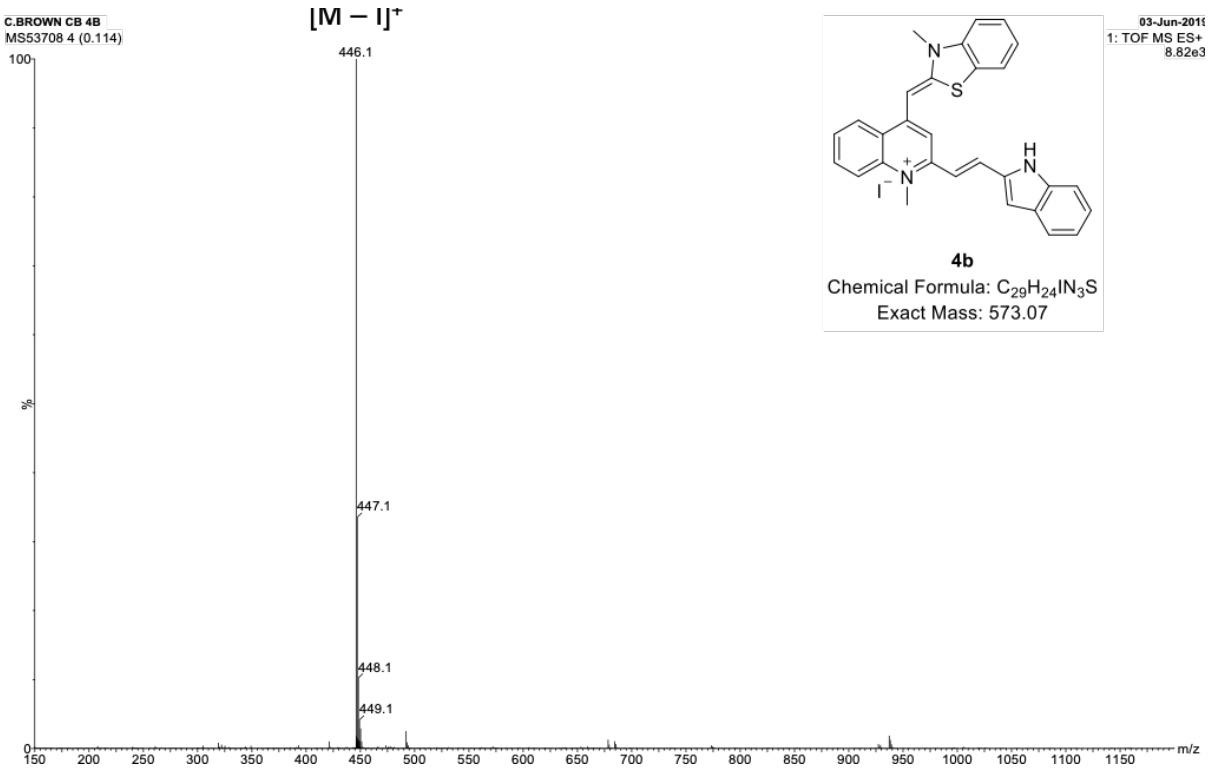


Fig. A.10: ESI-MS of **4b**.

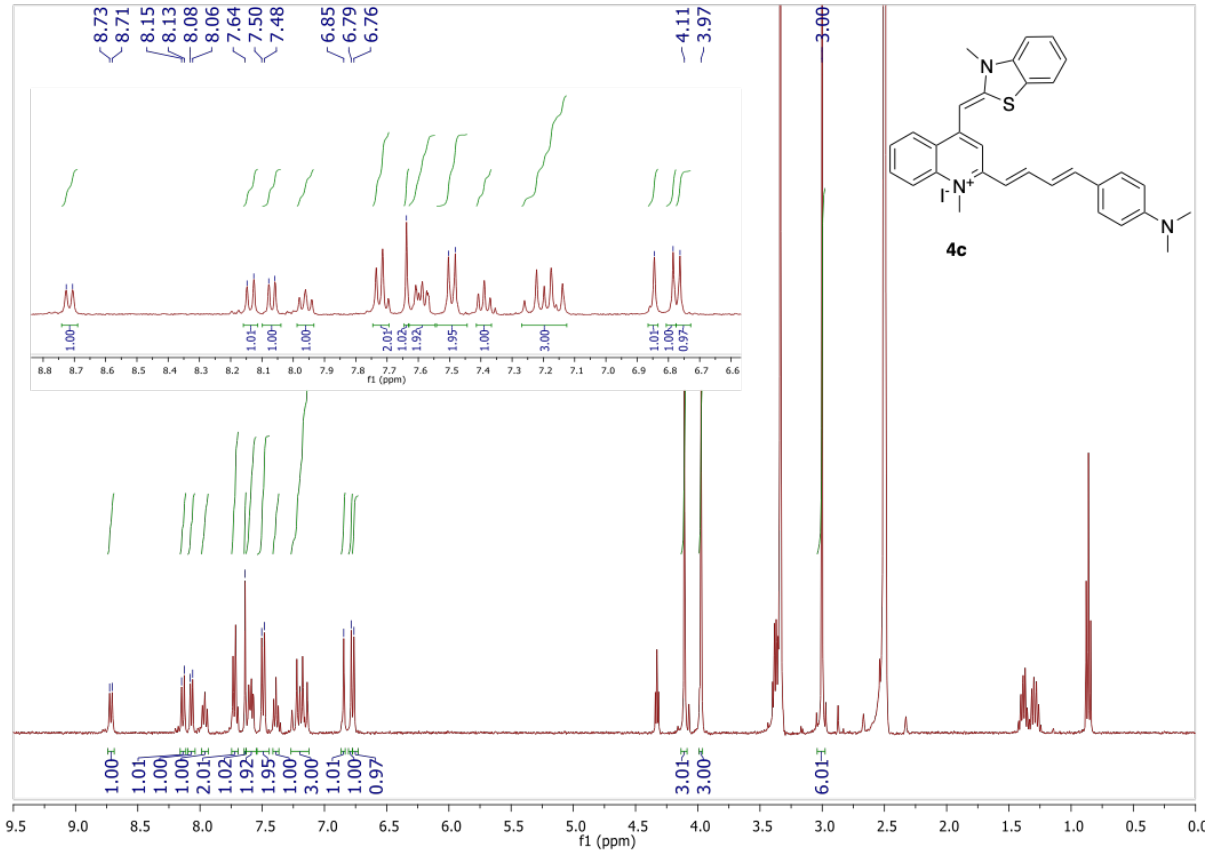


Fig. A.11: ^{1}H NMR spectrum of **4c** in DMSO-d_6.

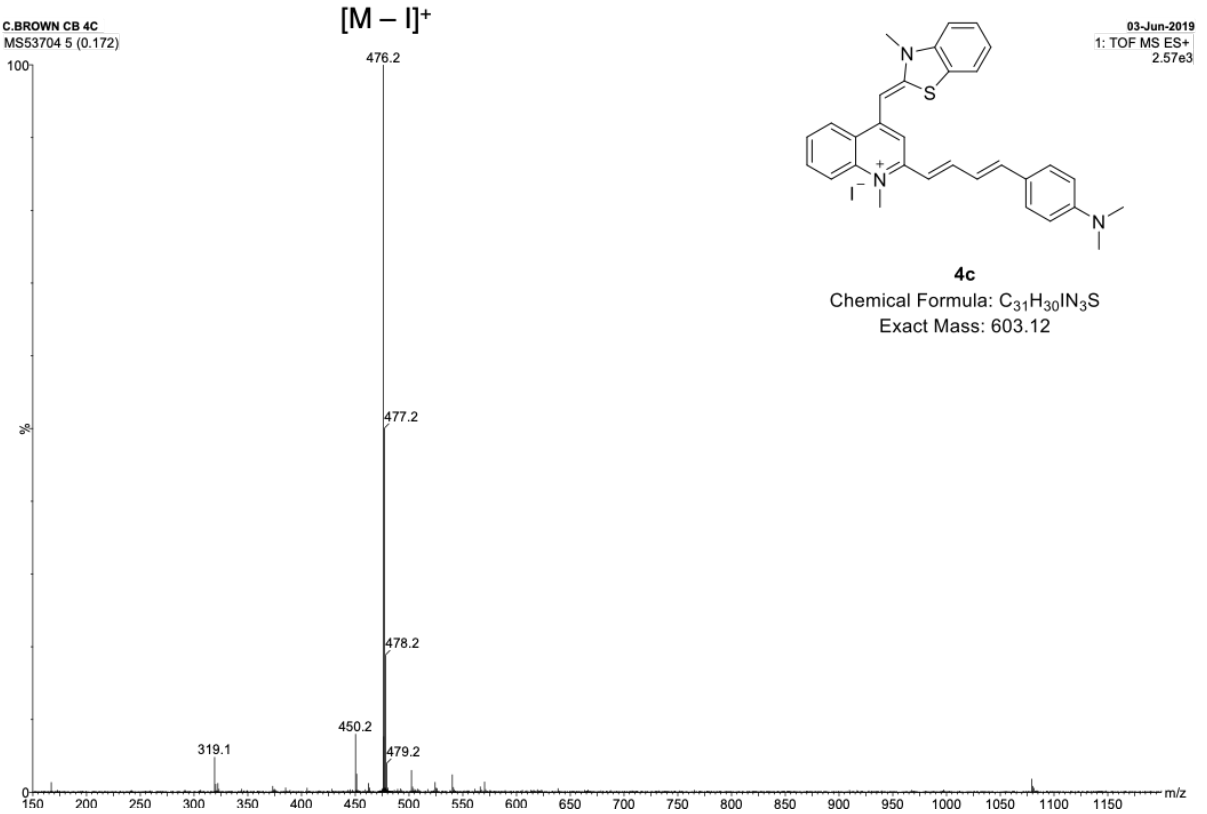


Fig. A.12: ESI-MS of **4c**.

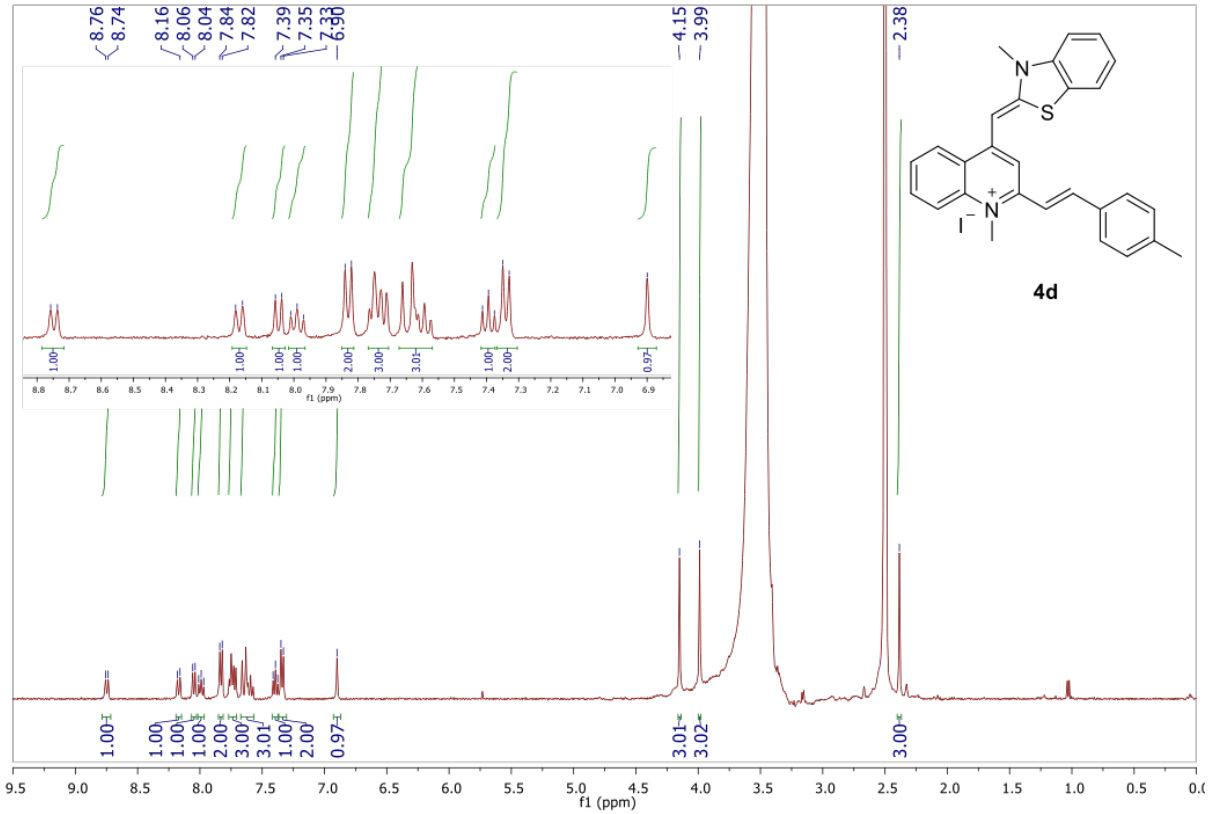


Fig. A.13: ^{1}H NMR spectrum of **4d** in DMSO-d_6.

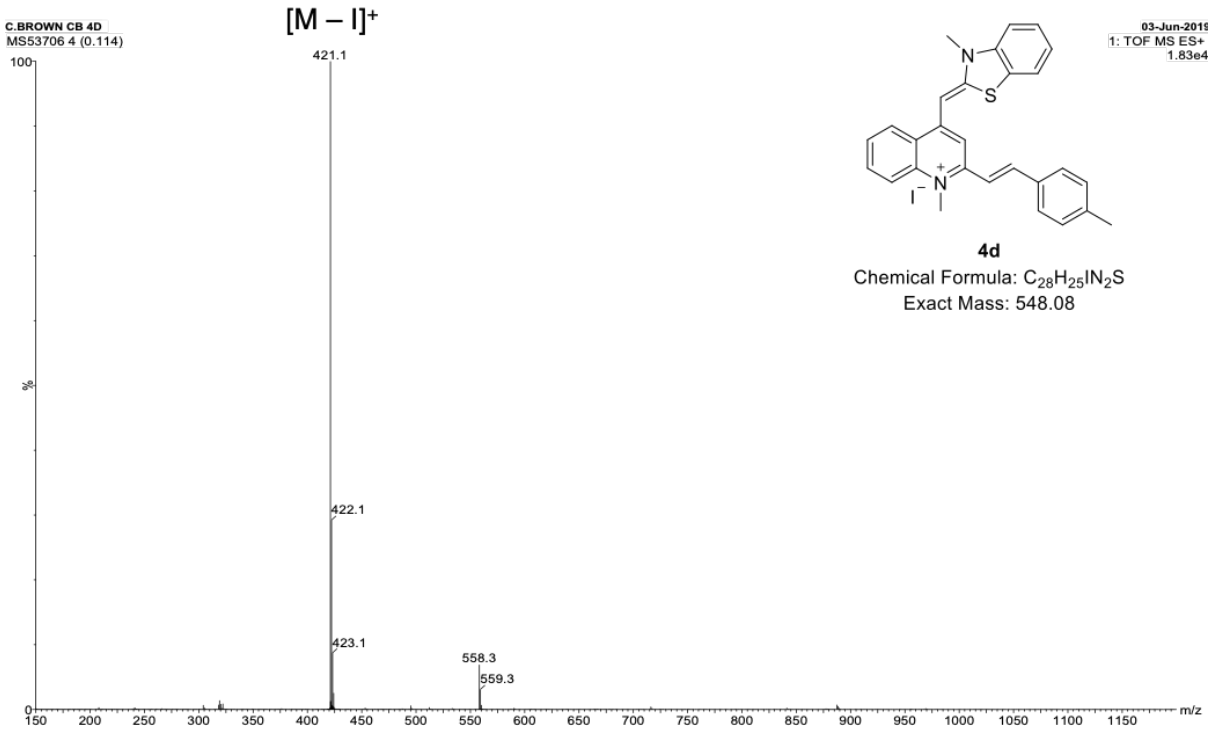


Fig. A.14: ESI-MS of **4d**.

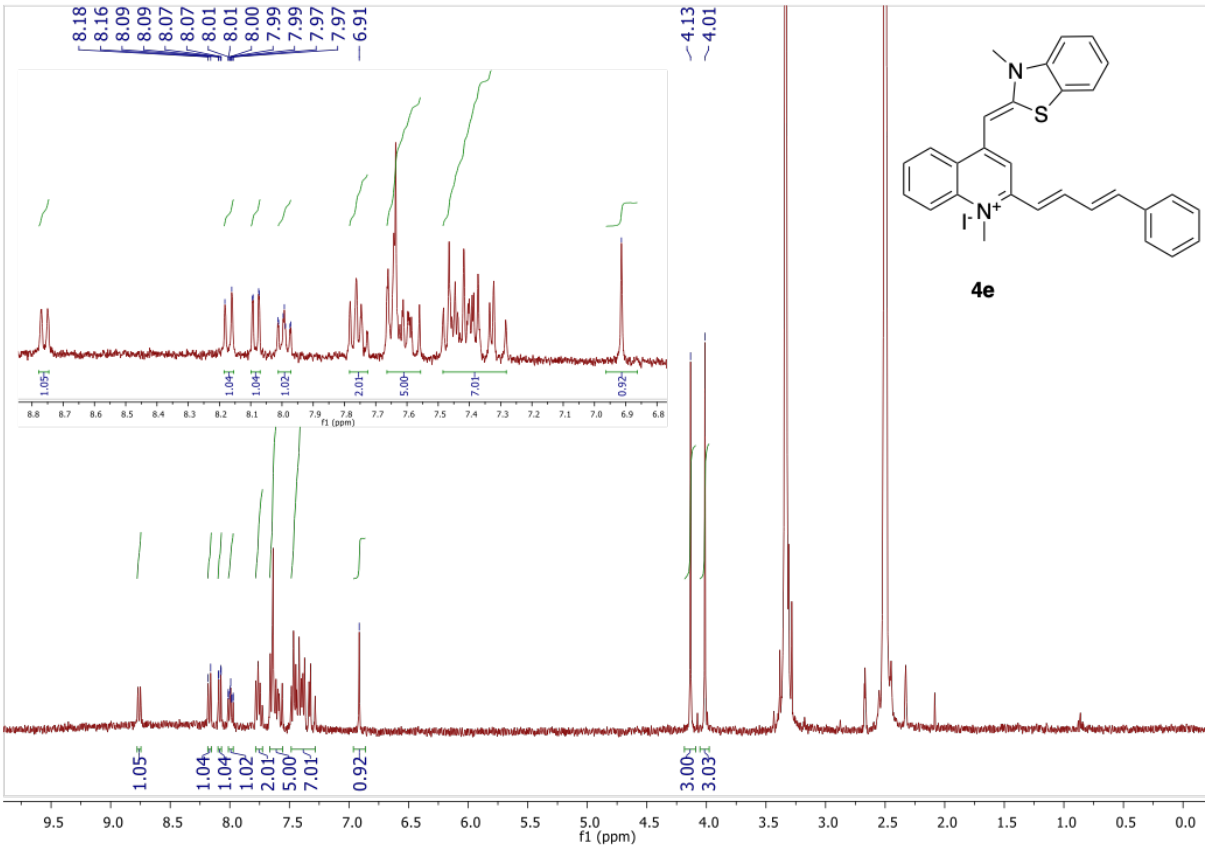


Fig. A.15: ^{1}H NMR spectrum of **4e** in DMSO-d_6.

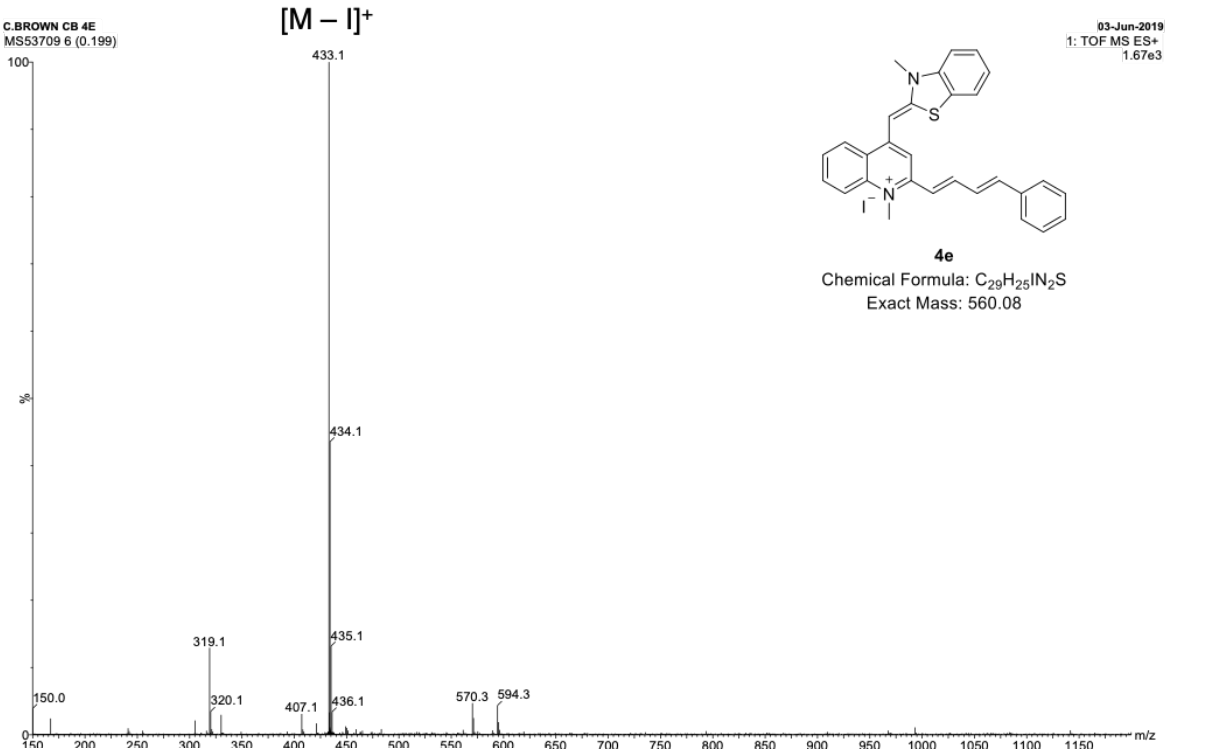


Fig. A.16: ESI-MS of **4e**.

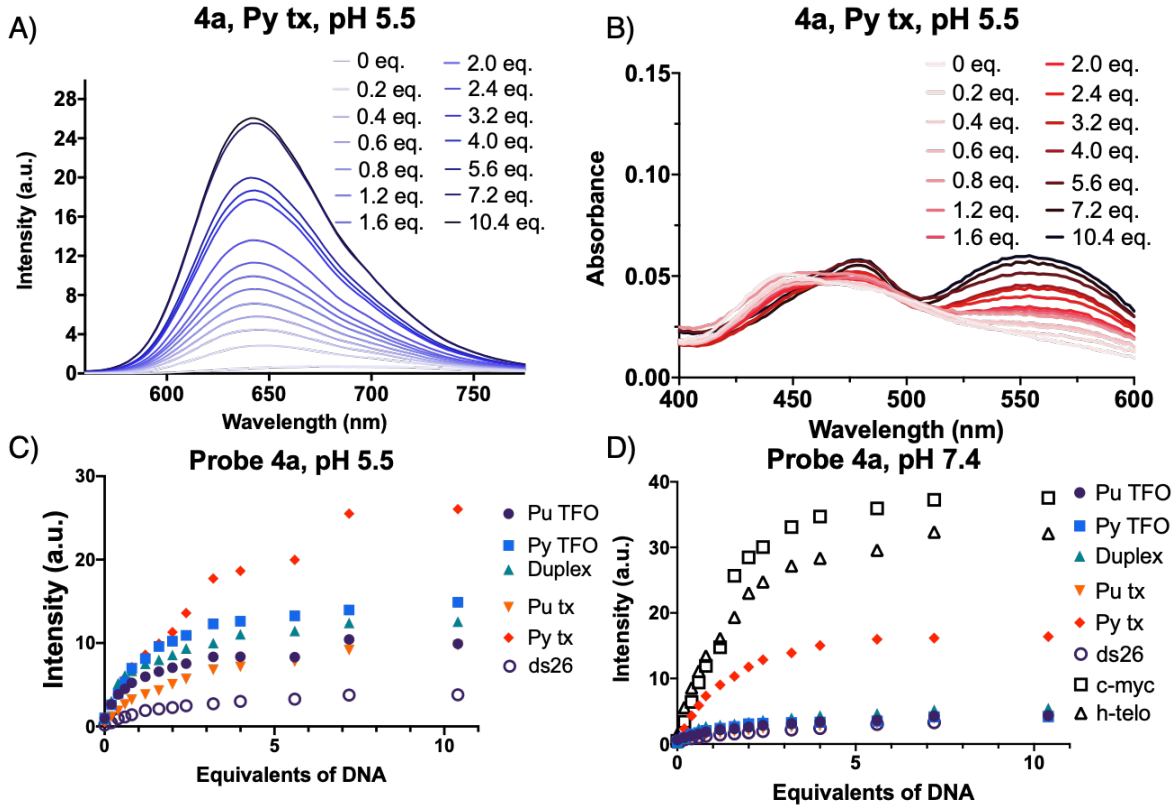


Fig. A.17: A) Example emission spectrum of **4a**, with pyrimidine triplex at pH 5.5, B) example absorption spectrum of **4a**, with pyrimidine triplex at pH 5.5, C) binding curve of **4a** with different DNA topologies at pH 5.5, D) binding curve of **4a** with different DNA topologies at pH 7.4.

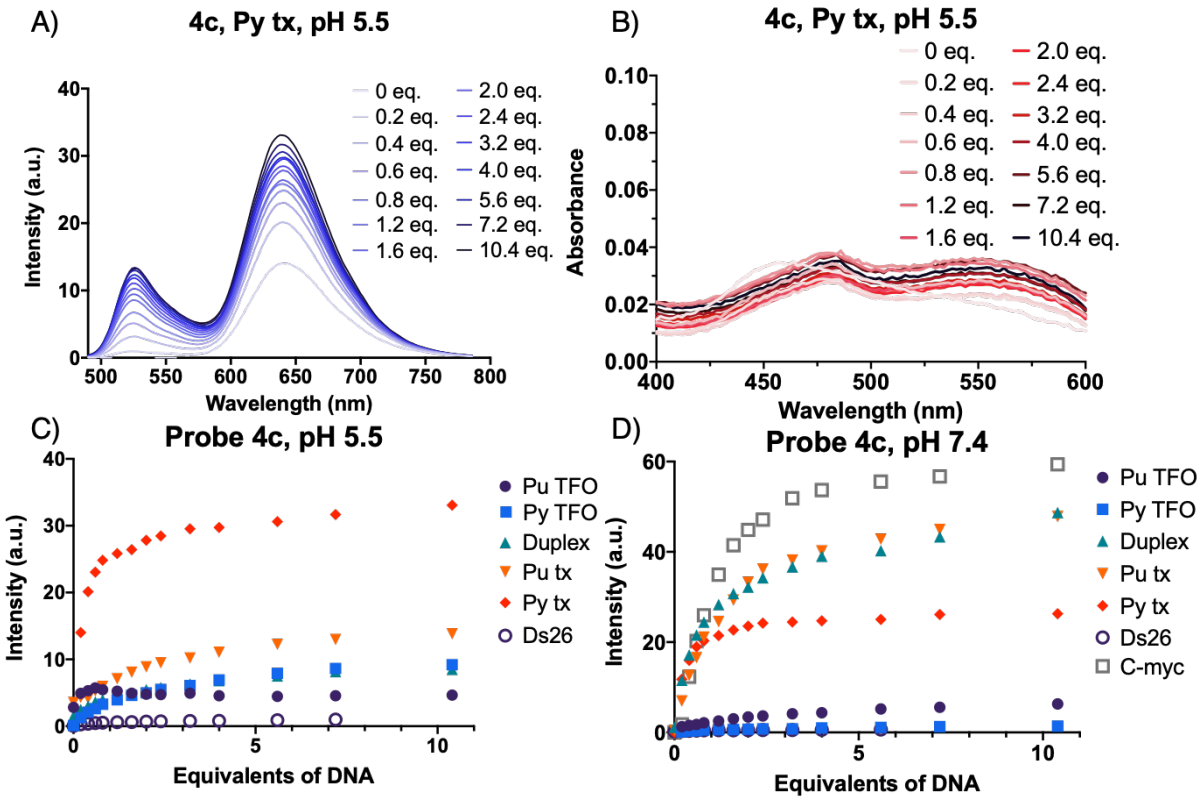


Fig. A.18: A) Example emission spectrum of **4c**, with pyrimidine triplex at pH 5.5, B) example absorption spectrum of **4c**, with pyrimidine triplex at pH 5.5, C) binding curve of **4c** with different DNA topologies at pH 5.5, D) binding curve of **4c** with different DNA topologies at pH 7.4.

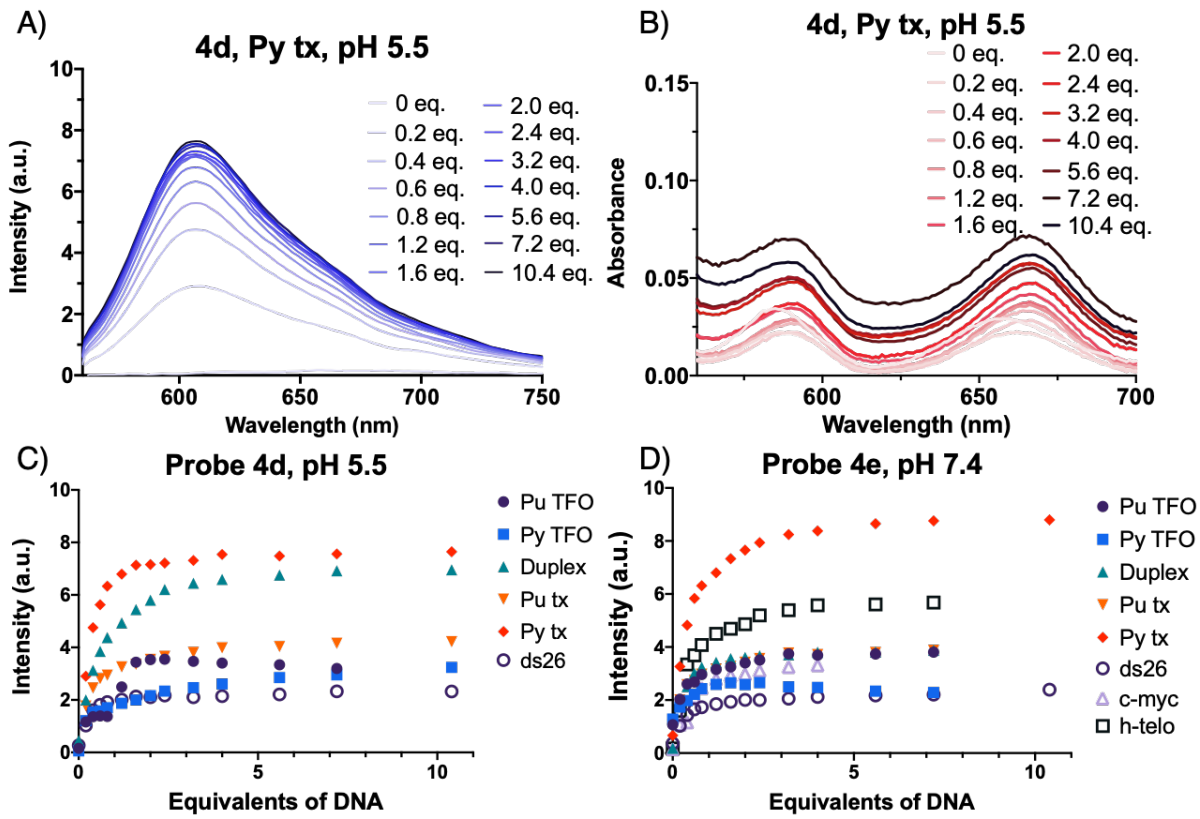


Fig. A.19: A) Example emission spectrum of **4d**, with pyrimidine triplex at pH 5.5, B) example absorption spectrum of **4d**, with pyrimidine triplex at pH 5.5, C) binding curve of **4d** with different DNA topologies at pH 5.5, D) binding curve of **4d** with different DNA topologies at pH 7.4.

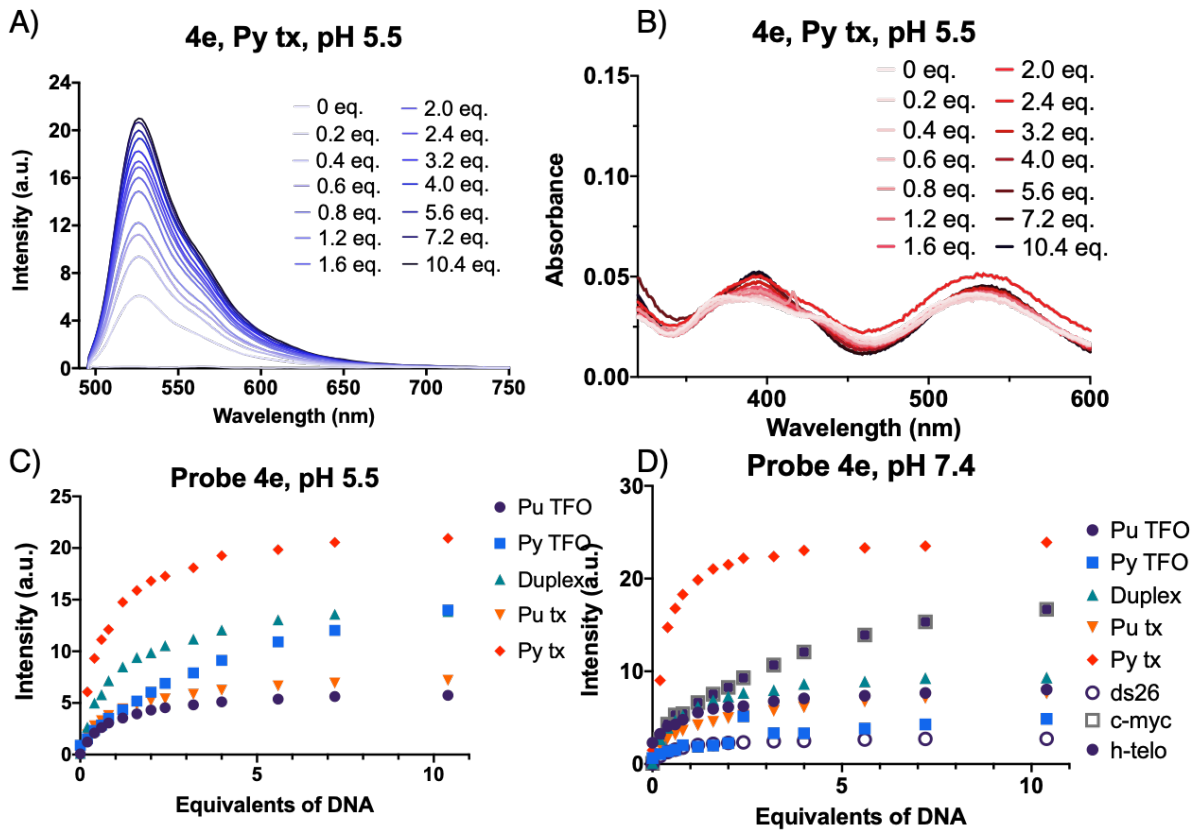


Fig. A.20: A) Example emission spectrum of **4e**, with pyrimidine triplex at pH 5.5, B) example absorption spectrum of **4e**, with pyrimidine triplex at pH 5.5, C) binding curve of **4e** with different DNA topologies at pH 5.5, D) binding curve of **4e** with different DNA topologies at pH 7.4.