



**This electronic thesis or dissertation has been  
downloaded from Explore Bristol Research,  
<http://research-information.bristol.ac.uk>**

*Author:*  
**Jiang, Jenny**

*Title:*  
**Exploring G-quadruplex topology with photoswitchable small molecules**

**General rights**

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

**Take down policy**

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact [collections-metadata@bristol.ac.uk](mailto:collections-metadata@bristol.ac.uk) and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.



# Exploring G-quadruplex topology with photoswitchable small molecules

**Y. Jennifer Jiang**

A thesis submitted to the University of Bristol as part of the requirements for  
award of the degree of Doctor of Philosophy in the Faculty of Science

School of Chemistry

Supervisors:

Professor M. Carmen Galan (School of Chemistry, University of Bristol)

Professor Adrian Mulholland (School of Chemistry, University of Bristol)

Word Count: 43,047



# Contents

Abstract .....	iii
Acknowledgments .....	iv
Author's Declaration.....	v
Abbreviations.....	vi
1 Introduction.....	1
1.1 G-quadruplex DNA .....	2
1.1.1 G-quadruplex polymorphism.....	3
1.2 G-quadruplexes as Therapeutic Targets .....	4
1.2.1 G-quadruplexes in Cancers.....	5
1.2.2 G-quadruplexes in Viruses.....	7
1.2.3 G-quadruplexes in Bacteria .....	9
1.3 Small molecule ligands for G-quadruplexes .....	10
1.4 Stimuli responsive ligands for G-quadruplexes .....	15
1.4.1 Photocaged ligands.....	15
1.4.2 Small molecule photoswitchable ligand classes as G4 binders.....	17
1.5 Project Aims .....	24
2 Design and synthesis of photoswitchable ligands .....	26
2.1 Ligand synthesis .....	30
2.1.1 Diazocine ligand synthesis .....	30
2.1.2 <i>o</i> -chlorodiazocine ligand synthesis.....	33
2.2 Ligand characterisation and photochemistry .....	38
2.3 Chapter conclusions.....	44
3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex.....	46
3.1 Biophysical Assays.....	48
3.1.1 Förster Resonance Electron Transfer thermal melting assay .....	48
3.1.2 UV-Visible absorbance titrations.....	53
3.1.3 Circular dichroism titrations .....	60
3.1.4 NMR titrations .....	67
3.2 Computational Molecular Dynamics Studies.....	75



3.2.1	Methods.....	75
3.2.2	Molecular Dynamics Results and Analysis .....	77
3.2.3	Discussion.....	91
3.3	Chapter conclusions .....	93
4	Telomeric G-quadruplex Cancer Targets .....	95
4.1	Biophysical Assays .....	96
4.1.1	FRET thermal melting assay .....	96
4.1.2	UV-Vis titrations .....	99
4.1.3	Circular Dichroism Titrations .....	105
4.2	Chapter Conclusions.....	108
5	Targeting <i>E. coli</i> G4s.....	110
5.1	Biophysical studies on microbial FRET labelled sequences .....	111
5.1.1	Ligand stabilisation.....	114
5.2	Biophysical studies with selected unlabelled sequences.....	116
5.3	Chapter Conclusions.....	122
6	Conclusions and Future work .....	124
7	Experimental.....	128
7.1	Ligand photochemistry.....	128
7.2	Biophysical studies .....	129
7.2.1	FRET melting assays .....	130
7.2.2	Titration studies .....	131
7.3	MD Computational Studies .....	133
7.4	Synthetic Procedures and compound characterisation .....	134
	References.....	156
	Appendix .....	169
A.3	Supplementary data for Chapter 3.....	169
A.4	Supplementary data for Chapter 4.....	171
A.5	Supplementary data for Chapter 5.....	173

## Abstract

G-quadruplexes (G4) are non-canonical DNA structures that can form from guanine rich sequences. They have been implicated in a variety of biological functions including genome instability, and the inhibition of transcription or translation processes. Targeting G4s by stabilising the secondary structure with small molecule ligands has been proposed as an alternative therapeutic strategy towards various diseases caused by cancers, viruses and bacteria, etc.

However, examples of stabilising G4s with small molecules are often a one-way response and thus targeting of G4s often leads to toxic side effects. In addition, as G4s are found across all cell types, ligands with poor selectivity may lead to unwanted interactions and side effects where healthy tissues are affected in off-target binding. To address the above, photoswitchable ligands can be used with an external light stimulus to control the on/off activity of G4 binding with high spatiotemporal control.

This work describes the development of a new class of G4 photoswitchable ligands known as diazocines. A small library of diazocines are synthesised and fully characterised and their potential as G4 ligands is evaluated with various techniques such as Förster resonance energy transfer (FRET) melting assays, UV-Visible spectroscopy, CD spectroscopy, NMR spectroscopy and MD simulation.

Additionally, the antibacterial activity of azobenzene ligands towards *E. coli* was investigated and a potential link with G4 stabilisation as a mode of action was explored. 17 putative quadruplex sequences that were initially identified in proteomics and bioinformatics studies, were shown to form G4s in potassium phosphate buffers. The stabilisation by the addition of ligands was also probed by spectroscopic techniques, which uncovered interesting conformational changes that could be insightful for the future study of these dynamic structures.

## Acknowledgments

Firstly, I would like to thank my supervisors Professor Carmen Galan and Professor Adrian Mulholland for their guidance and supervision over the course of this PhD project. Thank you Carmen for supporting me through my project, she was always helpful whenever I was struggling to make progress, particularly in the writing of this thesis, one of the more challenging things I have done. To Adrian, thank you for sharing your expertise in computational chemistry which helped this project become multidisciplinary.

Thanks to all members of the Galan group throughout the last four years, to the members that were here when I joined, thanks for welcoming me while I slowly got to know everyone during COVID lab rotas or through Zoom video calls. And then when the group shrunk, thanks to Simon, Eliza and Teo, who were in N309 during the quiet COVID years sharing cakes, cookies or Derbyshire cheese. Finally in later years to all the new members including, Cri, Nina, Calum, Gui who created a new atmosphere in our lab and having fun lunch sessions sat behind a computer playing geoguessr. Also thanks to those who have proofread and helped edit parts of this thesis, you know who you are.

Special gratitude goes to Mike who expertly introduced me to the field of G-quadruplexes back when I started; Javi, whose hard work and wisdom in synthesis cannot be understated; Yui, thank you for teaching me a little bit about biochemistry; in my computational work, special thanks to Sofia who has always been so knowledgeable and patient with me for whenever I needed help.

I also appreciate the help by all the technical staff in the School of Chemistry. Particularly, to Paul Lawrence and Chris Williams who have helped me with the strange NMR experiments, enduring my presence in a room with all the lights turned off.

Finally a special thanks to the CDT management team, for allowing me to do this project in the beginning and to members of my TECS 2019 cohort who have made the past few years more enjoyable.

## Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidates own work. Work done in collaboration with, or with the assistance of others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:... Y. Jennifer Jiang.. DATE: January 2024

## Abbreviations

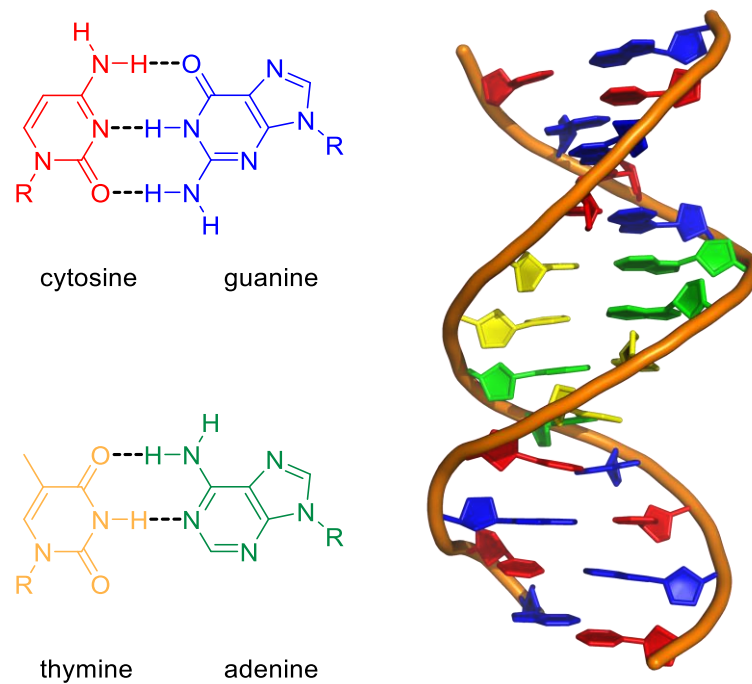
A	adenine
Abs	absorbance
Ac	acetate
AIDS	acquired immune deficiency syndrome
C	cytosine
CD	circular dichroism
DCM	dichloromethane
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
6-FAM	6-carboxyfluorescein
FRET	Förster/Fluorescence resonance energy transfer
G	guanosine
G4	G-quadruplex
ESI	electrospray ionisation
HIV	human immunodeficiency virus
HMBC	heteronuclear multiple bond correlation
HSQC	heteronuclear single quantum coherence
HRMS	high resolution mass spectrometry
hTERT	human telomerase reverse transcriptase
hTR	human telomerase RNA
IC <sub>50</sub>	half-maximal inhibitory concentration
ICD	induced circular dichroism
K <sub>a</sub>	Association constant
kb	kilobase
LRMS	low resolution mass spectrometry
LTR	long terminal repeat
<i>m</i> CPBA	meta-chloroperoxybenzoic acid

MD	molecular dynamics
NBS	N-bromosuccinimide
NMR	nuclear magnetic resonance
PDB	Protein Data Bank
PDS	pyridostatin
PQS	putative quadruplex sequence
PSS	photostationary state
RNA	ribonucleic acid
T	thymine
TAMRA	5-carboxytetramethylrhodamine
THF	tetrahydrofuran
TLC	thin layer chromatography
TRAP	telomeric repeat amplification protocol
UTR	untranslated region
UV	ultraviolet
Vis	visible
WC	Watson-Crick
$\Delta A$	change in absorbance



# 1 Introduction

The double helix structure of DNA was first discovered by Watson and Crick with X-ray crystallography data provided by Franklin in 1953.<sup>1</sup> Its double helical structure arises from a combination of the chirality of the deoxyribose-phosphate backbone and the complementary base pairing of the nitrogenous bases. There are 4 naturally occurring nitrogenous bases in DNA, two purine bases (guanine and adenine) and two pyrimidine bases (cytosine and thymine). Hydrogen bonds between guanine-cytosine and adenine-thymine Watson-Crick base pairs result in the two strands of the double helix being complementary to each other (Figure 1.1).



**Figure 1.1:** Left: Nucleic acid base pairs cytosine (C) and guanine (G) and thymine (T) and adenine (A) shown with their respective Watson-Crick hydrogen bonds. R = deoxyribose sugar in DNA and ribose sugar in RNA. Right: Helical B-form DNA, PDB ID: 1BNA. Nucleotides cytosine (C) and guanine (G) and thymine (T) and adenine (A) in red, blue, yellow and green respectively.

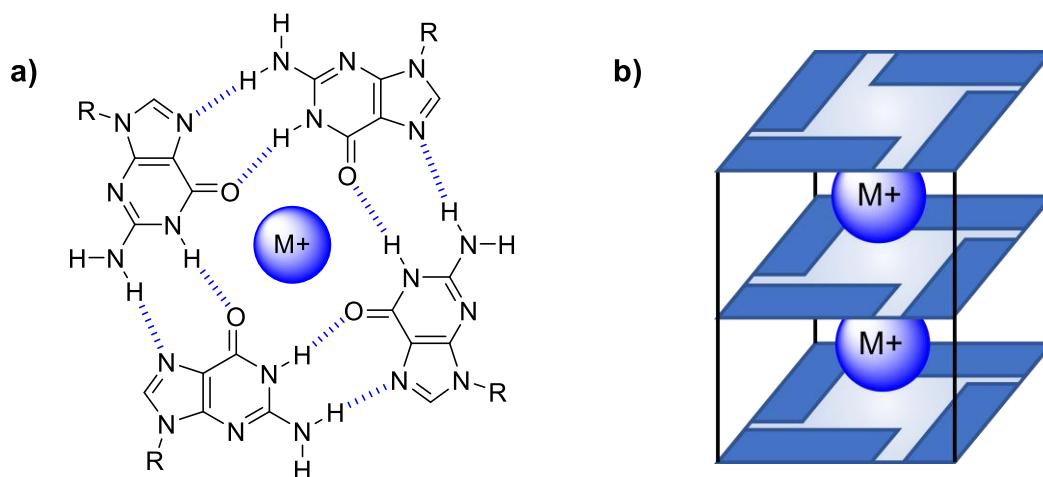
Beyond the classic double helix structure of DNA are non-canonical structures of DNA, for example A-DNA<sup>2</sup> and Z-DNA<sup>3</sup>, which are alternative double helix conformations of the classic B-DNA. Other structures include cruciform,<sup>4</sup> triplexes,<sup>5</sup> i-motifs<sup>6</sup> and quadruplexes.<sup>7</sup>



### 1.1 G-quadruplex DNA

In 1910, the German chemist Bang noticed that guanosine monophosphate solutions were able to form gels.<sup>8</sup> Roughly 50 years later, Gellert and co-workers obtained fibres from these gels and through X-ray diffraction methods determined that 4 guanylic acids could arrange themselves in a tetrameric formation through hydrogen bonding,<sup>7</sup> giving rise to particularly stable structures (**Figure 1.2a**). Later on, work by Henderson *et al.* found that guanine rich telomeric DNA was able to form these hydrogen bonded tetrameric structures.<sup>9,10</sup> This was shortly followed by work from Sen and Gilbert who noticed that guanine-rich DNA sequences could self-assemble into four-stranded structures in monovalent salt buffers at room temperature.<sup>11,12</sup>

Four guanines associate in a planar structure through eight cyclic Hoogsteen hydrogen bonds<sup>9</sup> and subsequent  $\pi$ -stacking provides the G-quadruplex (G4). This can then be further stabilised by the chelation of monovalent cations, such as  $K^+$  or  $Na^+$ , to relieve the high electron density of the carbonyl groups (**Figure 1.2**).<sup>12</sup>



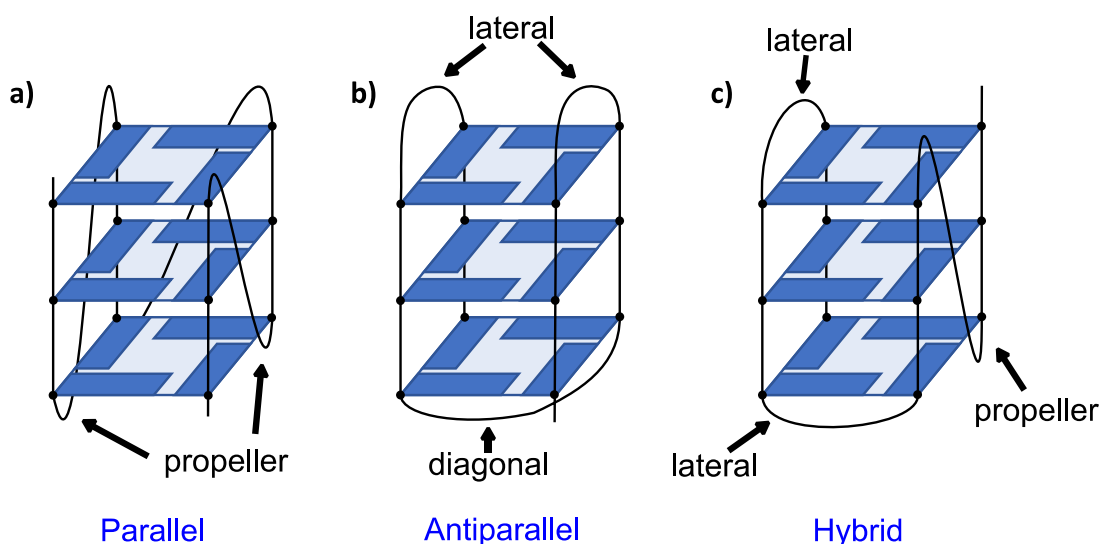
**Figure 1.2:** a) Structure of a guanine tetrad suggested by Gellert *et al.* from X-ray diffraction studies. b) Cartoon representation of the structure of a stacked G-quadruplex, made up of three tetrads and chelated by two  $M^+$  ions. Guanine residues are represented by blue rectangles.

## 1 Introduction

### 1.1.1 G-quadruplex polymorphism

Despite its simplistic core structure of  $\pi$ -stacked guanine tetrads, G-quadruplexes are an extremely diverse family of polymorphic structures which have shown to be made up of tetramolecular, bimolecular or unimolecular guanine rich sequences.

G-quadruplexes are formed from the folding of nucleic acid sequences which can occur in various different combinations. Different glycosidic bond angles, direction of guanine tracts and loop types lead to a plethora of potential variations in secondary structure. Canonical DNA G-quadruplexes have continuous guanine tracts which allow for three types of loops to connect the tracts together (outside of tetramolecular structures). These three loops are: propeller, lateral and diagonal loops which lead to parallel, antiparallel and hybrid topologies and are shown in **Figure 1.3**.<sup>13–15</sup>



**Figure 1.3:** Three polymorphic structures of monomeric G-quadruplexes. The direction of the oligomer sequence and the direction of each guanine tract is shown by black arrows. **a)** A parallel structure with propeller loops connecting the four guanine tracts. **b)** An antiparallel structure connected by two lateral loops and a diagonal loop. **c)** A hybrid structure with a propeller loop and two lateral loops.

Computational predictive algorithms have been used to identify putative G-quadruplex sequences (PQS) throughout the whole genome. Early methods used the regular expression,  $G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}$ , where a sequence is required to match four runs of 3-5 consecutive guanine nucleotides, with interspaced loops 1-7 nucleotides in length, where N corresponds to any of the four nucleotides (A,T,C,G). This algorithm predicted over 376,000 G4 forming sequences within the human genome.<sup>16–18</sup> However, many biophysical studies

## 1 Introduction

---

have shown that this restriction of loop length widely underpredicts the total quantity of G4s that could form very stable long loops by forming Watson-Crick (WC) hydrogen bonds to stabilise the secondary structure.<sup>14,19</sup> This gives rise to an even larger array of potential secondary structures for G-quadruplexes beyond the simple parallel, antiparallel and hybrid labels, to what can be considered as noncanonical G4s.

In these noncanonical G4s, features may include discontinuous guanine tracts such as V-shaped loops, bulges and snap back loops.<sup>15</sup> In some instances, the quadruplex could even contain loops or bulges that are so long that they form a stem loop that is stabilised through Watson-Crick base pairing.<sup>20</sup>

Various spectroscopic methods have been used to identify the wide array of secondary structures for G-quadruplexes. This includes CD spectroscopy, NMR spectroscopy and X-ray crystallography.<sup>18</sup> Outside of the field of targeting these unique nucleic acid secondary structures as therapeutic targets, G-quadruplexes have been utilised as hydrogels,<sup>21</sup> catalysts,<sup>22</sup> and as nanodevice sensing probes.<sup>23</sup>

### 1.2 G-quadruplexes as Therapeutic Targets

Although the majority of DNA is double stranded, G-quadruplexes can form during stages of DNA replication, transcription or DNA repair processes, when the hydrogen bonds between double stranded DNA are disrupted, providing the best opportunity for the folding of these tetrameric secondary structures.<sup>24</sup> They are also found in telomeric region of DNA due to the guanine rich sequence of the 3' overhang and their dynamic formation in live cells was recently demonstrated by Di Antonio *et al.* with the use of a fluorescent probe.<sup>25,26</sup>

G4s have been implicated in transcription, translation and genome stability.<sup>24,27</sup> Mutations of G4 helicases, enzymes used to unwind G4 secondary structure, have shown to result in increased chances of age related disorders such as, Bloom syndrome and Werner syndrome, and are characterised by premature aging and genomic instability. Beyond the human genome, G-quadruplexes have been found across all organisms including plants,<sup>28</sup> parasites,<sup>29</sup>

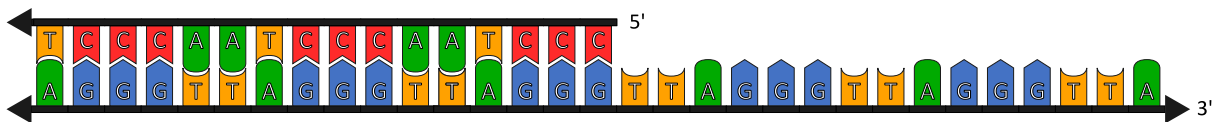
## 1 Introduction

---

bacteria<sup>30</sup> and viruses.<sup>31</sup> The following section will cover some of the roles of G4s in cancers, viruses and bacteria.

### 1.2.1 G-quadruplexes in Cancers

Telomeres are structures at the ends of DNA chromosomes composed of a repeating TTAGGG/AATCCC sequence. It is mostly double stranded (10-15 kb long) but has a guanine rich 3' single stranded section about 50-500 nucleotides long.<sup>32</sup> These single-stranded chromosome ends can be mistakenly identified as sites of DNA damage, potentially triggering cell apoptosis. To counter this, they are protected by the shelterin complex, which shields them from being wrongly identified as DNA damage sites.



**Figure 1.4:** Cartoon of telomeric DNA sequence showing a double stranded portion and a single stranded 3' overhang region. Nucleotides cytosine (C) and guanine (G) and thymine (T) and adenine (A) in red, blue, yellow and green respectively.

During cell replication in normal cells, telomeres shorten due to DNA polymerase being unable to replicate the full length of the DNA sequence. When telomeres shorten to a critical length, known as Hayflick limit, the cell can undergo senescence and apoptosis. However, the telomeres of cancerous cells are maintained and lengthened by the enzyme telomerase, allowing cancerous cells to gain immortality. Human telomerase is a ribonucleoprotein enzyme composed of hTERT (human Telomerase Reverse Transcriptase), the catalytic component, and hTR (human telomerase RNA), an RNA template used to recognise the 3' DNA telomeric overhang. While hTR levels are expressed constantly, the catalytic component hTERT is restricted to highly proliferative tissue, stem cells and tumour cells. In particular, at least 80-85% of cancerous cells are telomerase positive, significantly more than most somatic cells where telomerase is repressed.<sup>33-36</sup>

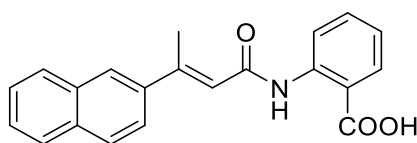
Independent work by Hahn *et al.* and Zhang *et al.* demonstrated that mutation of the hTERT catalytic site successfully disrupted telomerase activity, resulted in telomere length reduction and led to eventual cell death both *in vitro* and *in vivo*.<sup>34,35</sup> These findings suggested that

## 1 Introduction

---

telomerase inhibition could be a potential new cancer target by limiting the growth of cancer cells.

One approach is to block the template RNA, hTR, that recognises the telomeric DNA sequence by hybridising it with an antisense oligomer. Imetelstat, currently in phase II/III trials for blood cancers, is a thiophosphoramidate oligomer with a complementary sequence of d(TAGGGTTAGACAA) that hybridises with the template hTR of telomerase. It had good activity in telomerase inhibition, with an  $IC_{50}$  of 50-200 nM against pancreatic cancer cell lines.<sup>37</sup> An alternative method is to inhibit enzyme activity of the hTERT component by using small molecules like BIBR1532, which is a non-competitive telomerase binder with an  $IC_{50}$  of 93 nM *in vitro*. BIBR1532 exposed cancer cells showed steady telomere shortening, though this took a long lag period of ~120 days before observation of senescence and apoptosis, due to initial telomere shortening requiring a series of cell division cycles.<sup>38,39</sup>

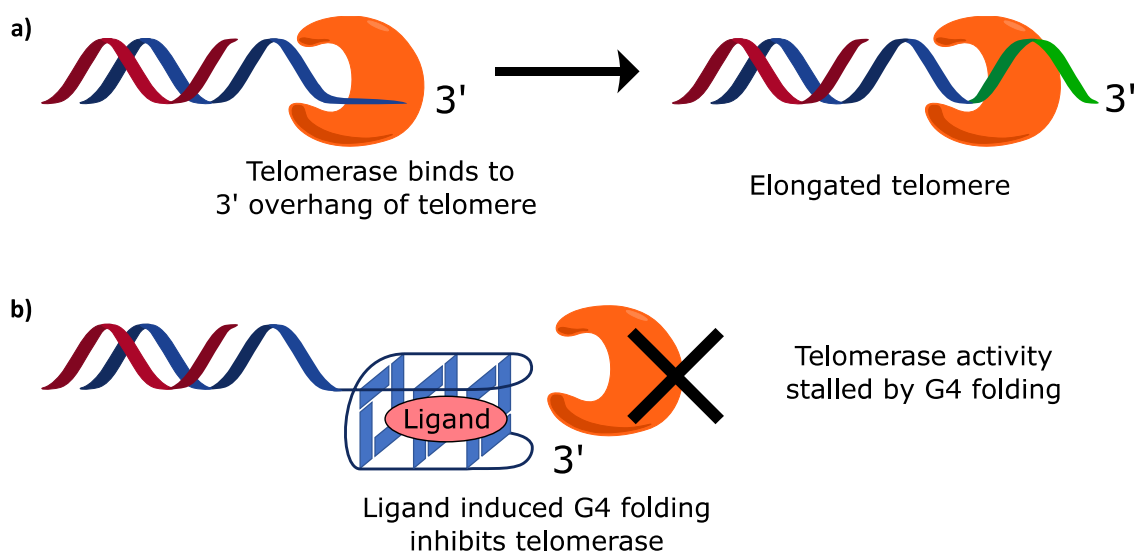


**BIBR1532**

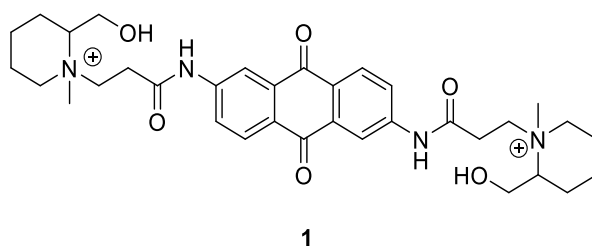
**Figure 1.5:** BIBR1532, a non-competitive small molecule inhibitor of telomerase.

The third method to indirectly inhibit the telomerase enzyme is by hindering the binding of the hTERT unit to the 3' strand by forming a quadruplex structure. Work by Zahler *et al.* in 1991 showed that the formation of  $K^+$  ion stabilised intramolecular quadruplexes in telomeres could inhibit telomerase.<sup>40</sup> This led to the proposal that small molecules could act as a quadruplex stabiliser to indirectly cause the inhibition of telomerase (**Figure 1.6**). This was successfully demonstrated by Neidle and Hurley in 1997, when they developed the first G-quadruplex stabilising ligand, anthraquinone **1**. Anthraquinone **1** selectively interacted with G-quadruplexes and was successful in the inhibition of telomerase with an  $IC_{50}$  of 23  $\mu$ M (**Figure 1.7**).<sup>41</sup> It is hypothesised that the timeframe of cellular senescence and apoptosis by ligand induced G4 stabilisation occurs faster than direct inhibition of the telomerase enzyme as small molecule ligands can also disrupt shelterin proteins, triggering DNA damage responses.

## 1 Introduction



**Figure 1.6:** Mechanism of indirect telomerase inhibition strategy by use of G4 binding ligands. a) telomerase binds to 3'-overhang of DNA leading to extension of telomere b) ligand induces G4 folding, which inhibits telomerase binding to DNA.



**Figure 1.7:** First G-quadruplex ligand anthraquinone **1** shown to inhibit telomerase function.

In addition, quadruplex forming sequences have been found in oncogene promoter regions such as c-MYC,<sup>42,43</sup> c-KIT,<sup>44</sup> BCL-2,<sup>45</sup> VEGF<sup>46</sup> and KRAS.<sup>47</sup> Small molecule ligands have shown to downregulate expression of proteins which could result in better control of cancer. Targeting these promoter region G4 structures of oncogenes may be more advantageous than targeting the expressed protein due to lower chances of developing drug resistance or due to the undrugability of the overexpressed protein.<sup>48</sup>

### 1.2.2 G-quadruplexes in Viruses

The study of G-quadruplexes in viruses has been limited (compared to cancer targets), but they have been found in a host of viruses including but not limited to: H1N1,<sup>49</sup> HSV,<sup>50</sup> Epstein-Barr virus,<sup>51</sup> coronaviruses,<sup>52</sup> and HIV.<sup>53</sup> The viruses that have caused the modern HIV/AIDS epidemic (1981-present) and COVID-19 pandemic (2019-present), rank at the top of epidemics/pandemics by death toll. Whilst there have been promising advances in mRNA

## 1 Introduction

---

vaccines that were used successfully during the COVID-19 pandemic, emergence of new variants with the potential to evade vaccine effectiveness underscores the ongoing need to explore alternative therapeutic strategies. Similar to the use of G-quadruplexes as targets for cancer therapies, studies have indicated that targeting viral G-quadruplexes could lead to effective antiviral outcomes.<sup>31,54</sup> Incorporating G4s as novel targets for antiviral treatments presents a potential alternative approach in combating these infectious diseases by directly addressing the viral genome.

### 1.2.2.1 G-quadruplexes in HIV-1 genomes

Since the widespread attention of the human immunodeficiency virus (HIV) and its associated acquired immunodeficiency syndrome (AIDS) in the 1980s, there has been an estimated 40 million deaths worldwide. Extensive research into HIV, led to the development of antiretroviral therapies which controlled HIV replication and reduced the risk of developing AIDS.<sup>55</sup> Due to this, deaths from HIV/AIDS dropped from the world's 8<sup>th</sup> leading cause of death in 2000, to the 19<sup>th</sup> leading cause of death in 2019. Even so, it still classes as an ongoing major public health issue.<sup>56</sup>

HIV is a single stranded RNA virus that is part of the family *Retroviridae*. HIV virus replication involves the use of a reverse transcriptase enzyme to produce DNA from its original RNA genome, which is then integrated into the host genome, known as the provirus.<sup>57</sup> G4 forming sequences in the HIV genome have been shown to be very conserved across primates and are primarily found in the coding region of the NEF gene, the 5'-UTR (untranslated region) of the RNA viral genome and in the long terminal repeat (LTR) of the provirus, suggesting that these sequences are potential anti-viral targets.<sup>54</sup>

Successful therapeutic applications of G4s in HIV have been shown in ligand binding to G-quadruplexes in both the viral UTR and the proviral LTR causing anti-HIV-1 activity by inhibiting transcription at the promoter level.<sup>53,58</sup> Ligands were also able to lower HIV levels of the latent virus, which is not typically susceptible to antiretroviral drugs.<sup>59</sup> Making it a novel and a promising target for the eradication of HIV from the host without the requirement of virus reactivation.

### 1.2.3 G-quadruplexes in Bacteria

As with viral G-quadruplexes, considerably less research on G4s in bacteria has been conducted when compared to studies of G4s in the human genome.<sup>30</sup> An early bioinformatic study in 2006, covering 18 prokaryotic genomes, found that there was an overrepresentation of putative quadruplex sequences (PQS) within 200 base pairs upstream of putative regulatory regions. 40 PQS motifs were conserved across multiple species, and a majority had functions pertaining to metabolism, membrane transport, transcription and translation, strongly suggesting the importance of G4s in these essential regulatory roles.<sup>60</sup>

More recently in 2019, another bioinformatics study was performed with 12 different species including two species of bacteria, *E. coli* and *R. sphaeroides*, which validated predicted PQS using the G4-seq technique.<sup>61,62</sup> Only a few of the bacterial PQS sequences formed G4s upon stabilisation by K<sup>+</sup> ions. But upon the addition of a small molecule ligand, a higher proportion of the PQS motifs were detected as observed G4 sequences (compared to nonbacterial species). This implied that many G4s in bacteria are weak but may be biologically relevant when bound to G4 binding proteins or ligands.<sup>62</sup>

Looking from a more pathogenic angle, targeting G4s in pathogenic bacterial genomes such as *Mycobacterium tuberculosis (Mtb)* may be a method to combat its increasing resistance to broad spectrum antibiotics by offering a new mechanism of action in antimicrobial activity. It was reported in a study by Mishra *et al.* that addition of small molecule ligands to *Mtb* G4 sequences found in genes associated with virulence, caused a decrease in gene transcription.<sup>63</sup> Another study by Perrone *et al.* identified *Mtb* G4 motifs located upstream of transcription start sites.<sup>64</sup> The folding and stabilisation of these G4s with ligands was demonstrated and ligands were also shown to inhibit *Mtb* growth. This potentially linked the inhibition of *Mtb* to G4 related mechanisms, suggesting that targeting G4s in bacterial genomes could be a new route of targeting tuberculosis.<sup>63,64</sup>



### 1.3 Small molecule ligands for G-quadruplexes

Seeing that G-quadruplexes are quite abundant across all organisms and have been shown to have vital biological roles. Regulation of these transient secondary structures has become of research interest to many. One of the first methods proposed (mentioned in section 1.2.1) was the use of anthraquinone **1** to indirectly inhibit telomerase. Since then, many small molecules have been designed for this purpose.

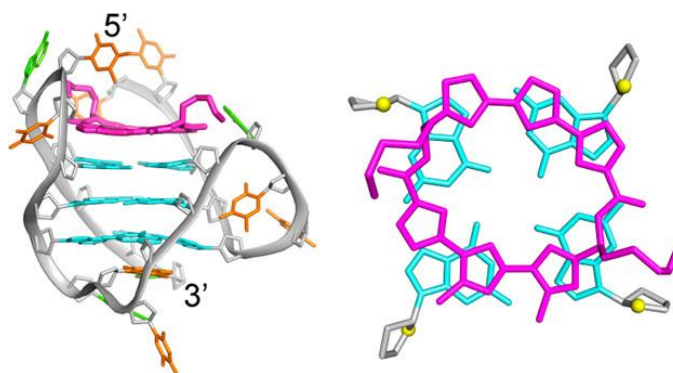
Most small molecule G-quadruplex ligands are designed to bind to G4s through a couple of common methods. Ligands generally have 1) a flat aromatic core that allows for  $\pi$ -stacking with the planar G-tetrad and 2) substituents that can be positively charged at physiological pH to provide water solubility and interact with the negatively charged sugar phosphate backbone.<sup>65,66</sup> One big challenge is that ligands must also selectively bind to G-quadruplex structures over the naturally more abundant duplex DNA. Currently over 3700 G4 ligands that have been reported can be found online at the G4 ligand database (G4LDB).<sup>67</sup> Many G4 ligands have been identified through biophysical methods, utilising melting assays such as Fluorescence Resonance Energy Transfer (FRET) to measure stabilisation that the ligand provides to the G4 sequence, while CD and NMR spectroscopy and X-ray crystallography are used to help determine binding modes.

Taking these design characteristics into account resulted in the development of the trisubstituted acridine, BRACO-19 (**2**). This compound has a cationic nitrogen that was shown to sit above the central electronegative carbonyl channel of G4s, while the 3 side chains interacted with grooves in the quadruplex structure.<sup>68</sup> BRACO-19 was shown to have  $IC_{50}$  of 0.06  $\mu$ M towards telomerase inhibition, considerably lower than when compared to anthraquinone **1**, showing the improvement in binding selectivity. It could also stabilise G-quadruplexes with a high  $\Delta T_m$  of 28 °C.<sup>69</sup> Another common G4 ligand is the cationic porphyrin TMPyP4 (**3**). This macrocycle has a large aromatic structure and cationic pyridiniums that can overlap well with an external G-tetrad. It can inhibit telomerase with an  $IC_{50}$  of 6  $\mu$ M and could stabilise telomeric G4s with  $\Delta T_m$  of 17 °C. However, it also shows a low level of duplex binding affinity, so it is often considered as having low selectivity towards G4 structures.<sup>70–72</sup> Surprisingly, it has also been reported to be a G-quadruplex destabiliser.<sup>73</sup>

## 1 Introduction

---

G4 ligands also include natural products like telomestatin (**4**) which was extracted from a GC rich bacteria *Streptomyces anulatus*. Telomestatin strongly stabilises G-quadruplexes ( $\Delta T_m = 24\text{ }^\circ\text{C}$ )<sup>74</sup> and was shown to inhibit telomerase at very low concentrations ( $IC_{50} = 5\text{ nM}$ ) without affecting other polymerase enzymes. In fact, it was shown that telomestatin was able to disrupt various protective telomeric shelterin proteins to induce apoptosis of cancer cells.<sup>75,76</sup> Its 70 fold selectivity towards quadruplex DNA is likely due to the cyclic nature of the molecule having poor interactions with duplex DNA. This can be illustrated with an NMR solution structure of a hybrid telomeric G4-telomestatin derivative complex showing a perfect overlap of the macrocyclic ring  $\pi$ -stacking with the top G-tetrad (**Figure 1.8**).<sup>77</sup> Unfortunately, telomestatin is poorly soluble in water and is difficult to obtain through natural or synthetic methods, hindering its practicality as a drug.



**Figure 1.8:** Solution structure of a telomestatin derivative (pink) and a hybrid telomeric G4 complex. **Left:** Ribbon view of a representative structure of the complex. **Right:** Top view showing the stacking of the telomestatin on the top G-tetrad. Reprinted with permission © 2013 American Chemical Society.<sup>77</sup>

Other examples of ligands include pyridostatin (**5**)<sup>78</sup> and PhenDC3 (**6**)<sup>79</sup>. These ligands are composed of small aromatic rings connected by amide bonds. Intramolecular hydrogen bonding between the nitrogen lone pair and amide NH, controls ligand conformation into a crescent-shape that can provide a good overlap with the surface area of a G-tetrad. Both pyridostatin and PhenDC3 stabilise G4s well, demonstrated by high  $\Delta T_m$  values from FRET melting assays, ( $\Delta T_m = 35\text{ }^\circ\text{C}$  and  $30\text{ }^\circ\text{C}$ ). Though pyridostatin had a relatively weak inhibition of telomerase ( $IC_{50} = 21\text{ }\mu\text{M}$ ), it was shown to disrupt the protective shelterin protein complex at even lower ligand concentrations, inducing the DNA damage response.<sup>78</sup> This suggested that DNA damage could also be induced by ligands without the stalling of telomerase. Pyridostatin has been used frequently to study the biological role of G4s, including a recent

## 1 Introduction

example in 2020 where a pyridostatin derivative was successfully used *in vivo* to demonstrate transient formation of G-quadruplexes in cells.<sup>25</sup>

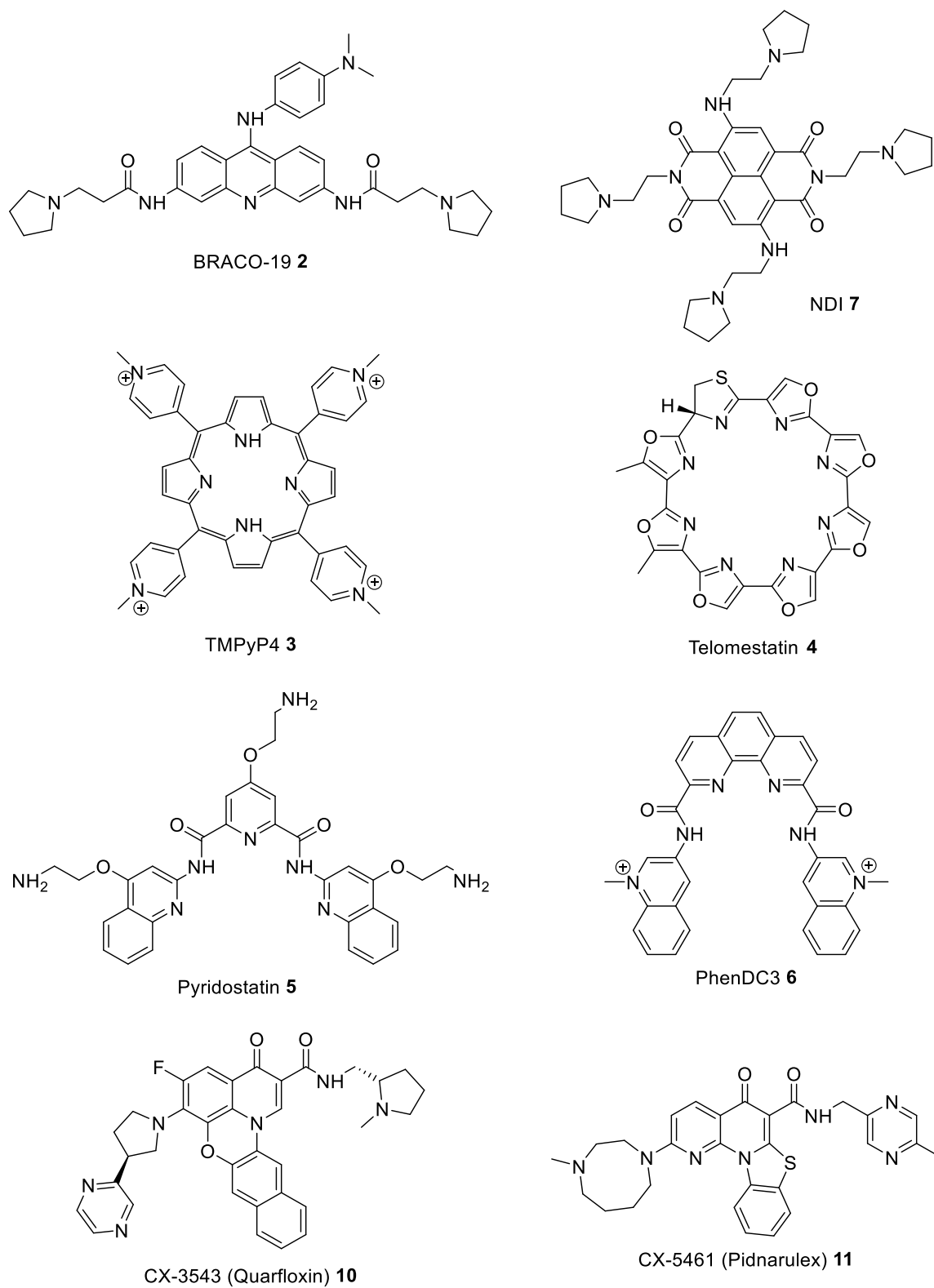
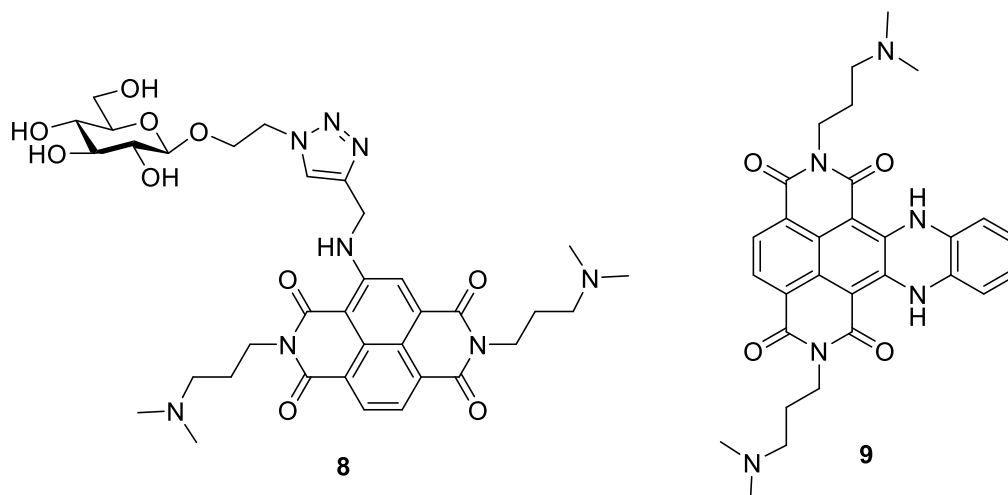


Figure 1.9: Examples of G-quadruplex ligands.

## 1 Introduction

---

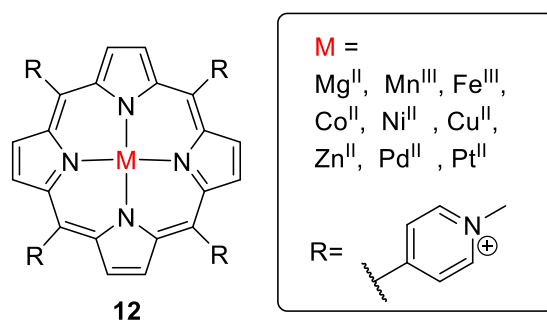
Another ligand class are naphthalene diimides (NDI). An early example, **7**, showed excellent affinity toward the c-KIT2 G4 ( $\Delta T_m = 29\text{ }^\circ\text{C}$  at  $1\text{ }\mu\text{M}$ ). NDI ligand design involves the core  $\pi$ -stacking with G-tetrads, and substituents that can branch off the central core to interact with the grooves in between the G-tracts. Interesting derivatives have included a carbohydrate NDI **8** against parasitic G4s ( $\text{IC}_{50} = 17\text{ nM}$ )<sup>29</sup> and the core-extended NDI **9**, which has been used in selective targeting of viral G-quadruplexes and bacterial G-quadruplexes ( $\text{IC}_{50} = 13\text{ nM}$ ) (Figure 1.10).<sup>80</sup>



**Figure 1.10:** NDI ligand derivatives **8** and **9** which have been shown as good binders to parasitic and viral sequences.

Of particular note are the ligands CX-3543 (**10**)<sup>48,81</sup> and CX-5461 (**11**).<sup>82</sup> These ligands, initially derived from fluoroquinolones, showed topoisomerase II and G4 interactions.<sup>83</sup> Both ligands were advanced into clinical trials for patients with cancer and were shown to selectively kill cells deficient in BRCA1/2. However, CX-3543 was withdrawn after phase II due to poor bioavailability, though CX-5461 is currently in clinical trials for breast cancer patients.<sup>82</sup>

Metal complexes have also been developed as G4 binders, which is unsurprising given the existing clinical metal complexes that target DNA (cisplatin). General design has evolved around planar complexes that can favour  $\pi$ - $\pi$  stacking and the central metal ion sitting over the electron rich channel of G-quadruplexes.

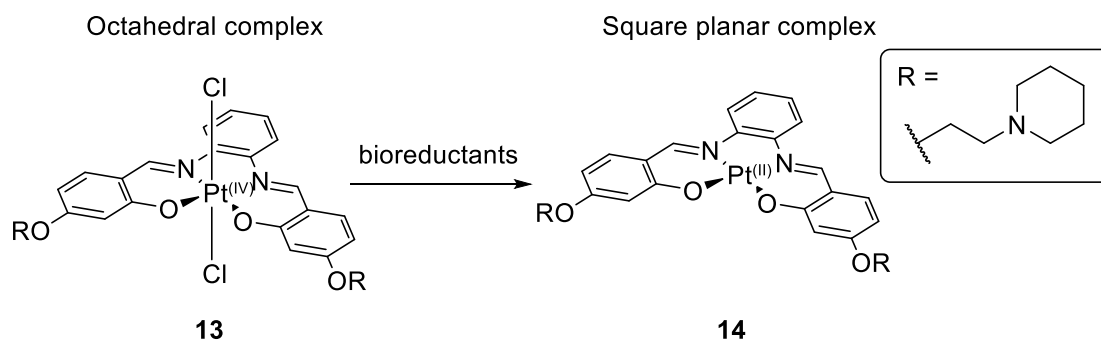


12

**Figure 1.11:** Early metal complex G4 ligands, where square planar coordinated complexes, performed best. Axial ligands = Cl.

Initial examples were derivatives of TMPyP4 porphyrins **12** with different metal cations in the center. **Figure 1.11** shows a library of metal-TMPyP4 complexes which were studied as G4 binders and telomerase inhibitors.<sup>84</sup> Among these, the Cu(II) complex with a square planar structure and the Zn(II) complex with a square-based pyramidal configuration exhibited the most potent inhibitory effect. Conversely, complexes featuring two axial ligands, such as the Mg(II) and Mn(III) complexes, demonstrated comparatively lower inhibitory activity.

In 2006, Vilar and Neidle reported the first example of a nickel metal salphen as a G4 binder.<sup>85</sup> The Schiff base salphen ligands are forced into a planar arrangement upon chelation and gave excellent selectivity towards telomeric G4s. One interesting example developed in 2017 also by the group of Vilar is the platinum salphen complex **13** (**Figure 1.12**).<sup>86</sup> When presented as the octahedral Pt(IV) complex **13**, there is no binding to G4 or duplex DNA. However, addition of bioreductants such as ascorbic acid or glutathione reduces the compound to a square planar Pt(II) species **14** that has high affinity for G4s. This approach is interesting as it requires an external stimulus to trigger G4 binding, a strategy that will be discussed in more detail in the following section.



**Figure 1.12:** Redox activated salphen metal-complex G4 ligand.

### 1.4 Stimuli responsive ligands for G-quadruplexes

G4s are found across all cell types and G4 folding is known to cause various different downstream processes such as inhibition of gene transcription, and inhibition of cell growth.<sup>24,66,87</sup> Typically, ligand binding to G-quadruplexes is a one-way response – stabilisation of an unfolded sequence towards a folded G4 structure. If a ligand has poor selectivity for its target, sometimes unwanted interactions may occur and healthy tissues are affected in off-target binding.<sup>87</sup>

Light is an attractive method to overcome some of these issues with selectivity. It is considered biorthogonal to most biological processes, is normally non-toxic and can be delivered to a selected target with high spatial and temporal control.<sup>88</sup> If it is possible to use ligands that can be activated or deactivated by light, then it may be conceivable to modulate where a ligand is actively binding to G4s and for how long.

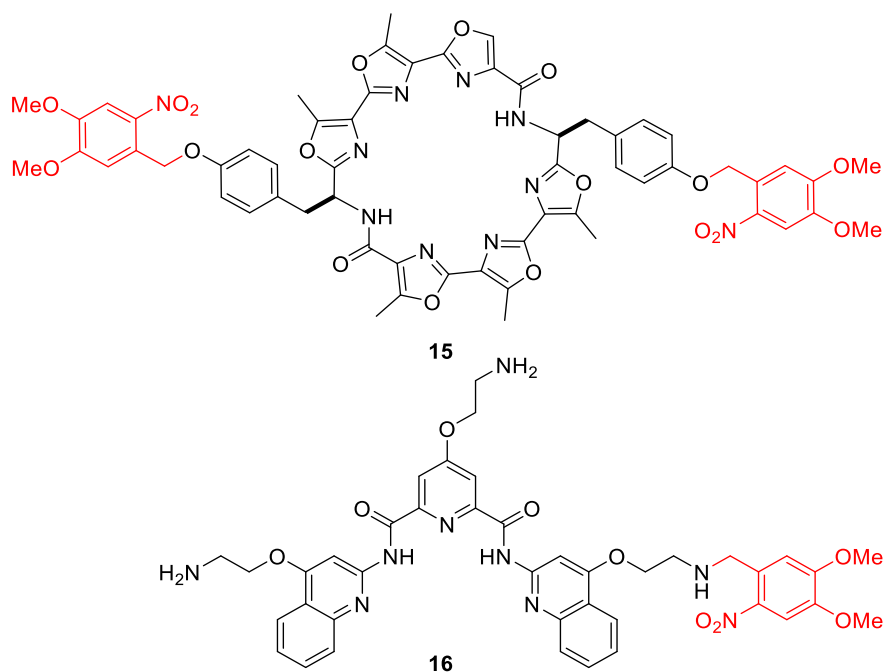
#### 1.4.1 Photocaged ligands

One strategy that has been reported is the use of photocaged ligands, where light can activate or deactivate the ligand to control selectivity. Photocages are a good example of photochromic ligands acting as prodrugs, where the activity of a ligand can be restored after a photoirradiation stimulus. For example, photocages of the classical G4 ligands telomestatin **15**<sup>89</sup> and pyridostatin **16**<sup>90</sup> (**Figure 1.13**). In both cases the caged ligands would not stabilise G4 structures. But after 365 nm UV-irradiation, the nitroveratryl groups were cleaved, restoring their quadruplex stabilising properties. These photocages allow for a one-way turn-on response of the ligand activity, which is useful as it allows for controlled activation only at desired sites for G4 binding.

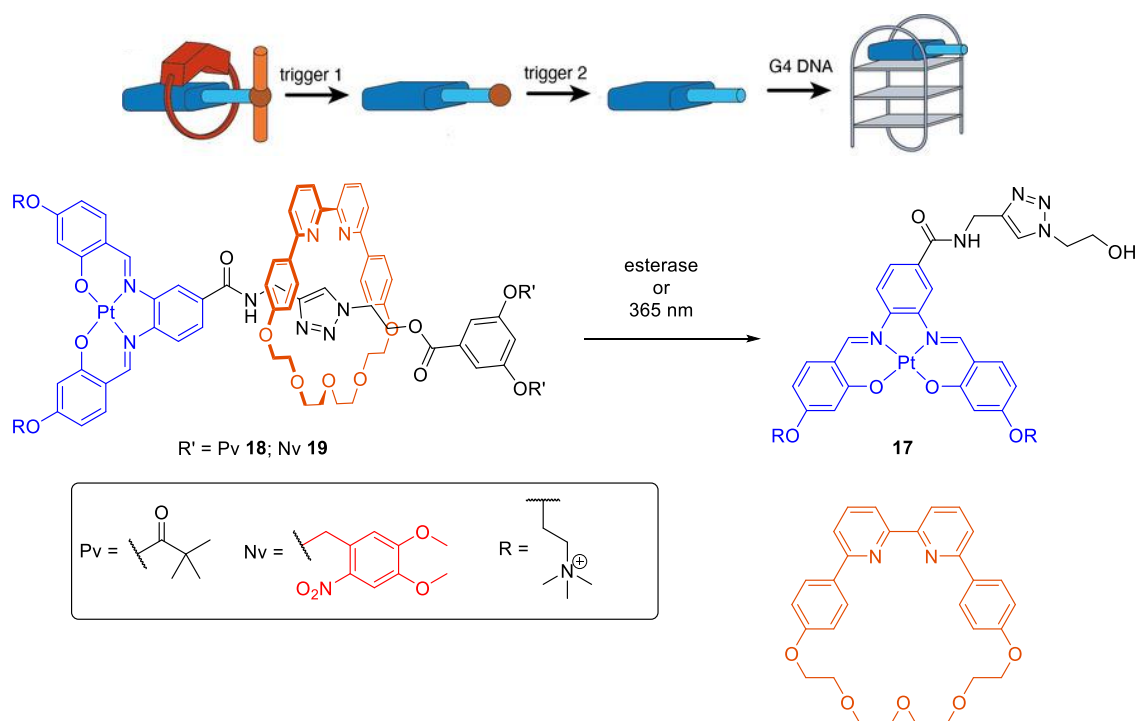
Another interesting example is a strategy based on caging the G4 ligand as a rotaxane (**Figure 1.14**). The active G4 ligand **17**, is a metal-salphen complex protected by a macrocycle that hinders binding. After triggering via enzymatic cleavage (**18**) or light irradiation (**19**), the stoppering unit is removed allowing the macrocycle to dethread, restoring G4 binding affinity. Apart from just hindering ligand activity, the rotaxane also provided increased cell

## 1 Introduction

permeability to the caged ligands **18** and **19**, allowing cellular uptake to DNA sites for binding.<sup>91</sup>



**Figure 1.13:** Photocaged G-quadruplex ligands **15** and **16**. The nitroveratryl caging group (shown in red) is removed by irradiation with 365 nm light.

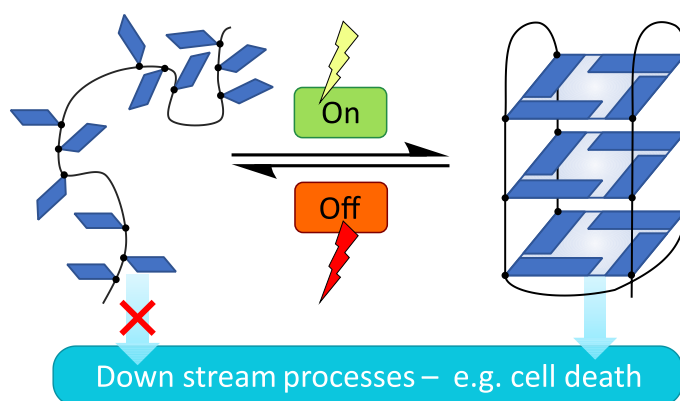


**Figure 1.14:** Rotaxane caged metal complex ligand **18** or **19**, triggered by enzyme cleavage or UV irradiation, releases macrocycle and uncages G4 ligand **17**. Figure adapted with permission from © 2021 Wiley.<sup>91</sup>

### 1.4.2 Small molecule photoswitchable ligand classes as G4 binders

While photocages are useful, a limitation of photocages is that they cannot reversibly control the on and off activity of a drug. For this, it is necessary to design a ligand with a photoswitchable moiety within the core structure. This type of strategy is employed in photopharmacology.<sup>88</sup>

Photoswitchable molecules offer control between a “toxic” active drug and a “nontoxic” inactive drug due to reversible changes in the ligand’s structure or properties upon absorption of light. Properties of a good photoswitchable ligand include: large changes in structure before and after switching and a non-biologically-active off state. Having the capability to undergo photoswitching using red or near-IR wavelength light (650-900 nm) would offer an added advantage, considering that these wavelengths can penetrate deeper into biological tissue.<sup>92</sup> Additionally there is a need to avoid UV light as a photoswitching wavelength due to its ability to damage DNA. Importantly, as the presence of the active drug is controlled by external irradiation at the desired site of activity, off-target effects at healthy tissues should be minimised. The following summary includes various classes of photoswitches, how these probes have been used (in a mostly non-covalent manner) and their effect to G-quadruplex structure and stability.<sup>93</sup>



**Figure 1.15:** Cartoon illustrating use of photoswitchable ligands to control G-quadruplex structure and the activity of downstream biological processes.

#### 1.4.2.1 Azobenzene scaffolds

Azobenzenes are a commonly used photoswitch due to their chemical and photostability, demonstrating excellent photoswitching properties without significant photofatigue. Irradiation of the stable trans isomer (*E*-) changes a long planar molecule into a bent cis isomer

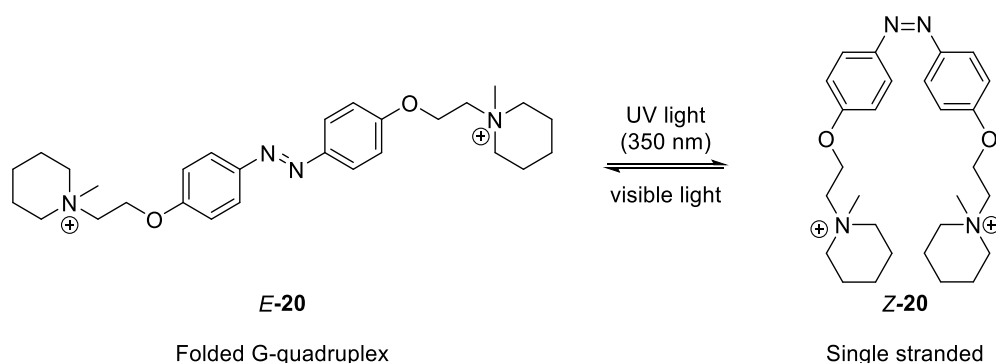


## 1 Introduction

(Z-) where the aromatic rings sit out of plane due to steric clashes. The cis isomer can then return to the trans isomer by either irradiation at another wavelength of light or thermally. There is a large structural change between the two isomers which is optimal for photopharmacology. However, one of its drawbacks is that the absorption spectra of the two isomers overlap, so photoisomerisation results in a mixture of both isomers. There is also a possibility that in cells, presence of glutathione may reduce the azobenzene to hydrazine, decomposing the compound.<sup>94,95</sup>

Azobenzene ligands have been used frequently as DNA ligands for duplex structures. However there are fewer examples where azobenzenes have been used as non-covalent small molecule ligand photoswitches for G4s. Including, but not limited to the Zhou group<sup>96-98</sup>, Matczyszyn group,<sup>99-101</sup> McCallum group<sup>102</sup> and Chen group.<sup>103</sup> Among these, the first example by Zhou and coworkers could reversibly fold and unfold a G-quadruplex sequence with azobenzene **20** (**Figure 1.15**).<sup>96,97</sup> The addition of *E*-**20** caused the folding of a single stranded G-rich telomeric sequence into a parallel G4. This was then irradiated with 350 nm light which caused the G4 structure to unfold when the azobenzene isomerised to the cis isomer. This could be repeated up to 30 cycles without appreciable photofatigue, though only without the presence of physiologically relevant potassium cations.<sup>96</sup>

Further work with the azobenzene analogues **21-23** showed that it was not possible to unfold G4 topology when potassium ions were present. (**Figure 1.16**). The same experiments were then undertaken in lithium ion containing buffers, which achieved the same effect as when using **20**. Demonstrating that trying to affect G-quadruplex topology in physiologically relevant conditions in the presence of potassium ions is not a trivial matter.<sup>97</sup>



**Scheme 1.1:** Reversible control of G-quadruplex conformation upon photoisomerisation of azobenzene **20**.

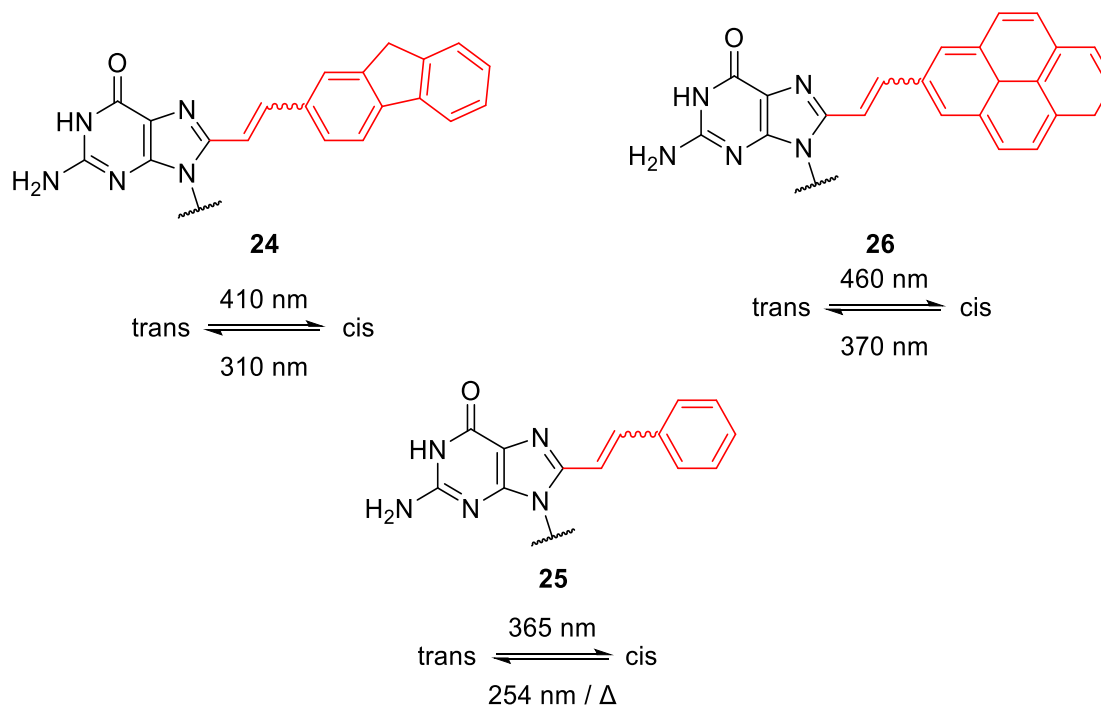


## 1 Introduction

### 1.4.2.2 Stilbene scaffolds

Similarly to azobenzenes, stilbene photoisomers structurally differ by the *cis/trans* structure of its double bond. Even in the dark, the meta stable *Z*-azobenzene is expected to eventually thermally isomerise back to the planar *E*-isomer. However, most stilbenes can only interconvert with light as there is a high energy barrier to thermal isomerisation, resulting in two kinetically stable isomers with large structural differences.<sup>105</sup>

Early examples of the use of stilbene-like photoswitches to control reversible G-quadruplex folding was demonstrated by Ogasawara and Maeda with the fluorenyl substituted guanosine **24** binding to thrombin.<sup>106</sup> This work was followed shortly by the monomer assembly and disassembly of phenyl substituted **25** monomers by Lena *et al.*<sup>107</sup> Later, Ogasawara revisited this approach to control mRNA transcription with light by using the larger pyrenyl guanosine **26** as this could be isomerised at longer wavelengths of 460 nm using less toxic UV-A light (**Figure 1.18**).<sup>108</sup> However, these approaches require the chemical modification of the guanosine bases, so is not appropriate for use in therapeutic or biological settings.



**Figure 1.18:** Guanosine derivatives **24**, **25** and **26** involved in the formation of G4 structures. Differences in photoswitching component highlighted in red.

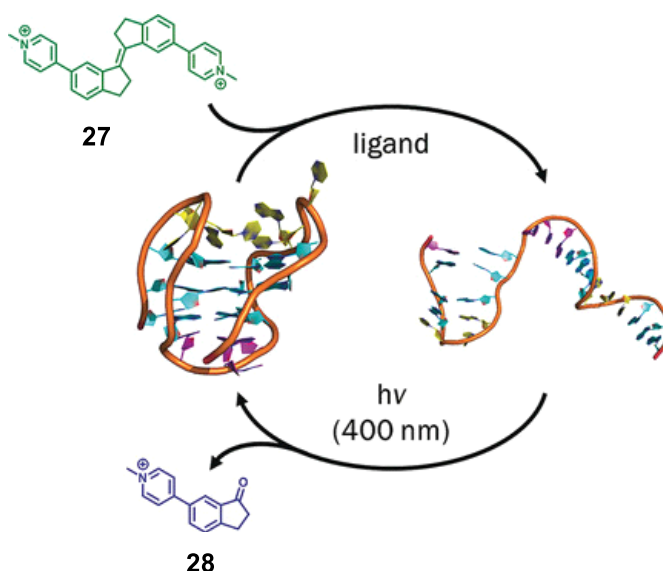
## 1 Introduction

### 1.4.2.2.1 Stiff-stilbene scaffold

Work by O'Hagan in the Galan group explored the use of a stilbene analogue, stiff-stilbenes as non-covalent photoswitchable G4 ligands. Stiff-stilbenes have additional bridging carbons to form two 5-membered rings, which increases the surface area for binding and restricts rotation about the aryl single bond.<sup>109</sup>

FRET melting studies with pyridinium stiff-stilbene **27** showed that the stiff-stilbene was selective towards G-quadruplex DNA over duplex DNA. With high thermal stabilisation values of  $\Delta T_m = 21.4$  °C in  $K^+$  and  $\Delta T_m = 11.8$  °C in  $Na^+$  at a concentration of 1  $\mu M$  of ligand. In contrast, the duplex DNA only had a  $\Delta T_m = 0.8$  °C at 1  $\mu M$  of stiff-stilbene.

CD titration experiments of *E*-stiff-stilbene **27** showed stabilisation of the telomeric quadruplex in  $K^+$  conditions, whilst this was not the case in  $Na^+$  buffer. In the latter example, the CD spectra showed large conformational changes upon increasing equivalents of ligand.  $^1H$  NMR experiments indicated that titration of the *E*-stiff-stilbene **27** into a  $Na^+$  stabilised telomeric quadruplex actually caused unfolding of the structure. Showing that despite a high thermal stabilisation, other biophysical measurements are required to fully elucidate binding modes of ligands to G4s.



**Figure 1.19:** Stiff-stilbene **27** binding to an antiparallel G-quadruplex causing unfolding of the sequence. After irradiation with 400 nm light, the stiff stilbene photooxidises and then cleaves allowing the G-quadruplex to refold.<sup>109</sup>

## 1 Introduction

---

The ability of the stiff stilbenes to act as a photoswitch was then tested with irradiation of the ligand. However, it was seen that irradiation of *E*-stiff-stilbene **27** caused photooxidation and fragmentation to give ketone **28** (Figure 1.19). This allowed it to act as a photoresponsive fuel, where fragmentation of the ligand **27** allowed the unfolded sequence to refold back into a quadruplex, up to 5 times before significant photofatigue was observed.<sup>110</sup> Whilst this ligand showed promise as a small molecule to control G4 folding, it was not a truly reversible system.

### 1.4.2.3 Dithienylethene scaffolds

With a different class of photoresponsive molecule, O'Hagan *et al.* reported selective targeting of telomeric G-quadruplexes with dithienylethenes. These molecules photoswitch under visible light via an electrocyclic ring closing and opening process (Figure 1.20).<sup>111</sup>

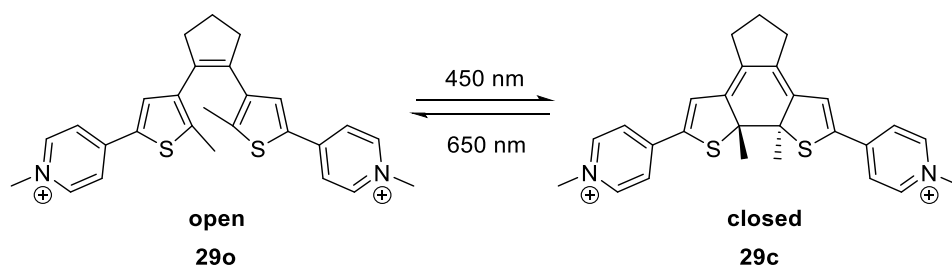


Figure 1.20: Photoswitchable dithienylethene ligands **29** that bind to G-quadruplexes.

FRET melting assays showed that both photoisomers could selectively bind to G-quadruplexes over duplex DNA. These binding modes were investigated by <sup>1</sup>H NMR studies with telomeric sequences. It was discovered that the two ligands have differing binding modes towards G-quadruplexes, and it was possible to reversibly switch between these binding modes upon alternating irradiation using visible light. Furthermore, toxicity assays showed that the two photoisomers displayed a two-fold difference in cytotoxicity towards HeLa cancer cells, demonstrating its potential as a G-quadruplex drug in photopharmacology.

### 1.4.2.4 Spiropyran scaffolds

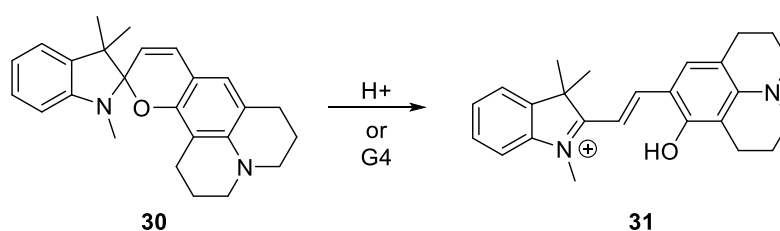
Another class of photoswitchable G-quadruplex ligands are the spiropyrans. These molecules switch between a closed sp<sup>3</sup> spiropyran and the flatter more intensely coloured merocyanine species. This provides a structural distinction between an inactive ligand and an active ligand for DNA binding. Examples of spiropyrans ligands for duplex DNA were reported by

## 1 Introduction

---

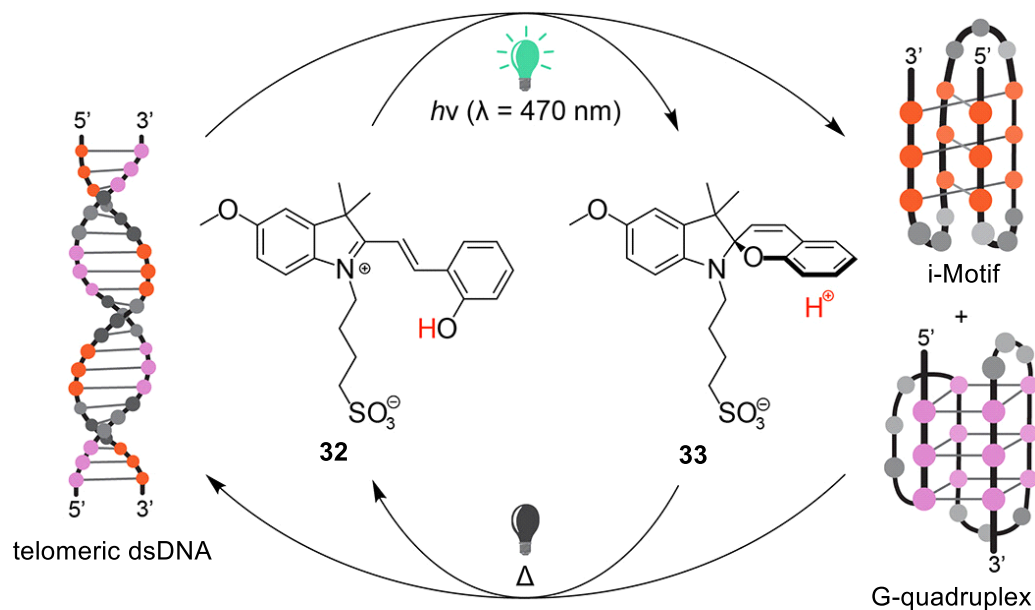
Andreasson. The closed  $sp^3$  spiropyran does not interact with the DNA, while its merocyanine photoisomer is able to intercalate in a fully reversible process.<sup>112,113</sup>

Spiroyrans have also been shown to bind to G-quadruplex DNA, such as the cMYC oncogene promoter sequence.<sup>114,115</sup> A structure-based ligand screen led to the fluorescent molecule spiropyran **30** which contains a quinolizidine moiety and has high selectivity towards G4 DNA over single stranded or double stranded DNA. Upon binding to G-quadruplexes, the emission wavelength is red-shifted from 458 nm to 610 nm and was attributed to the formation of merocyanine **31** (**Scheme 1.2**). CD titrations of spiropyran **30** demonstrated that the isomerisation towards merocyanine **31** occurs *in situ* without the need for external light excitation between the two photoisomers. Despite the lack of external control over isomerisation, spiropyran **30** was developed as a fluorescent probe instead of a photopharmacological drug, and it was successful at imaging endogenous G4s in live cells.



**Scheme 1.2:** Spiropyran **30** was able to isomerise to its merocyanine form **31** in the presence of G4 DNA or in acidic conditions.

A recent use of spiropyran was published with a peculiar method of controlling telomeric DNA structure. When merocyanine **32** was irradiated with light, it generated its corresponding spiropyran **33** and a proton which reduced the pH of the system. At a low pH, the cytosine bases on the C-rich strand were protonated and was able to transform the double stranded telomeric sequence into an i-Motif. G-quadruplex structures then formed on the complementary strand due to the  $K^+$  ion conditions stabilising the G4 secondary structure. This is a novel and indirect use of photoswitches to control G4 formation by changing the pH of the system.



**Figure 1.21:** Diagram showing control of DNA topology with photoswitch **32** by altering the pH environment. Figure adapted with permission © 2023 American Chemical Society.<sup>116</sup>

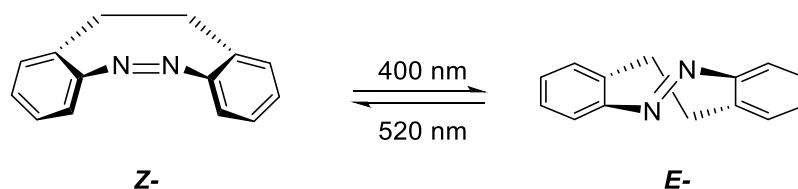
### 1.5 Project Aims

As discussed above, significant research has been undertaken to explore the prospect of G-quadruplexes as a therapeutic target. The development of the two small molecule ligands quarfloxin **10** and pidnarulex **11** as candidates for G4 mediated cancer therapy and their subsequent clinical trials is of note. Quarfloxin was unsuccessful after Phase I clinical trials, with its poor bioavailability highlighted. Due to the on/off switch activity of photoswitchable ligands, they have the potential to avoid poor bioavailability issues where the active ligand can be turned on only at the desired site, having an enhanced spatiotemporal control.

Existing photoswitchable ligands for G-quadruplexes such as azobenzenes, stilbenes, dithienylethenes and spiropyrans have been designed to selectively target G4 DNA structures. Yet, there are more classes of photoswitches that can be employed as cores for new G-quadruplex ligand design. The aim of this project is to develop a new class of photoresponsive ligands to understand, control and modulate G4 DNA topology with hope in the ability to eventually control G-quadruplex function.

## 1 Introduction

---



**Scheme 1.3:** *Z* to *E* isomerisation of diazocines.

This task will be accomplished first by the synthesis of diazocine photoswitches, an analogue of azobenzenes with different photophysical properties (**Scheme 1.3**). Then their photoswitching properties will be studied to investigate their applicability as ligands in aqueous ionic solutions. Finally, their interactions with various G-quadruplex forming sequences with different biological or pharmacological importance will be investigated using spectroscopic techniques, with a particular focus on the differences between the two photoisomers and their binding modes studied by molecular dynamics simulations.

Additionally, the antibacterial properties of different G4 ligands with photoswitchable moieties will also be studied with a selection of *E. coli* sequences that were selected through bioinformatics and proteomics studies; exploring the potential correlation of a G-quadruplex mediated mechanism and an antibacterial effect.



## 2 Design and synthesis of photoswitchable ligands

Supramolecular photoswitchable DNA ligands have been shown to bind to DNA or control topology reversibly with light.<sup>117</sup> Most of this early work was aimed towards molecular design and DNA binding, with preliminary efforts for selective control of topology and biological function being less widely reported. In contrast, a large library of covalently incorporated photoswitchable moieties have been studied in their capacity to reversibly control topology, but these cannot be used in a pharmacological setting, except as aptamers, due the necessity of synthetically modifying oligonucleotide sequences.

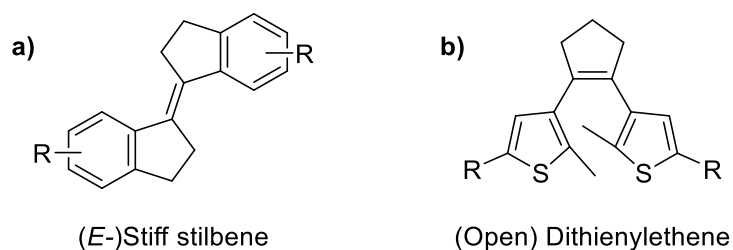


Figure 2.1: Known G-quadruplex ligands previous designed in the Galan group

Previous work in the Galan group has shown the development of two photoswitchable scaffolds that can selectively bind to G-quadruplex structures, as discussed in Chapter 1 (Figure 2.1). The stiff-stilbene, designed with a large planar surface area to  $\pi$ -stack with planar G4 tetrads, afforded remarkably high thermal stabilisation of G4s.<sup>109,118</sup> However, it could not photoisomerise between the *E* and *Z* isomers in aqueous buffer solutions, instead decomposing via an oxidative fragmentation pathway. This unfortunately required it to be treated as an exhaustive photofuel, limited to 5 cycles of ligand binding and unbinding events before fatigue of the system.

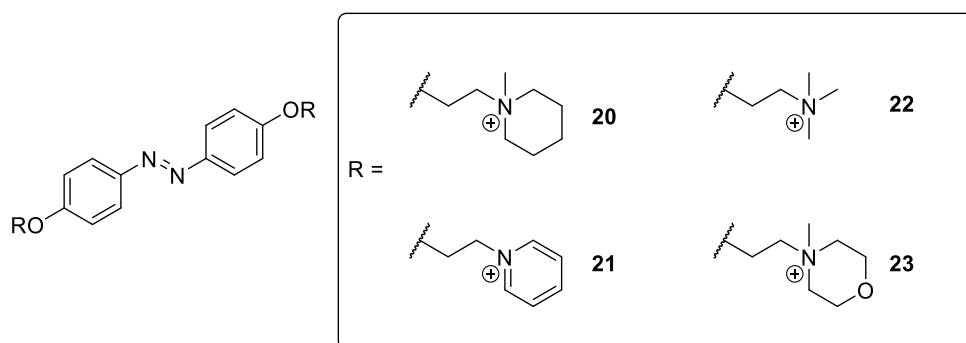
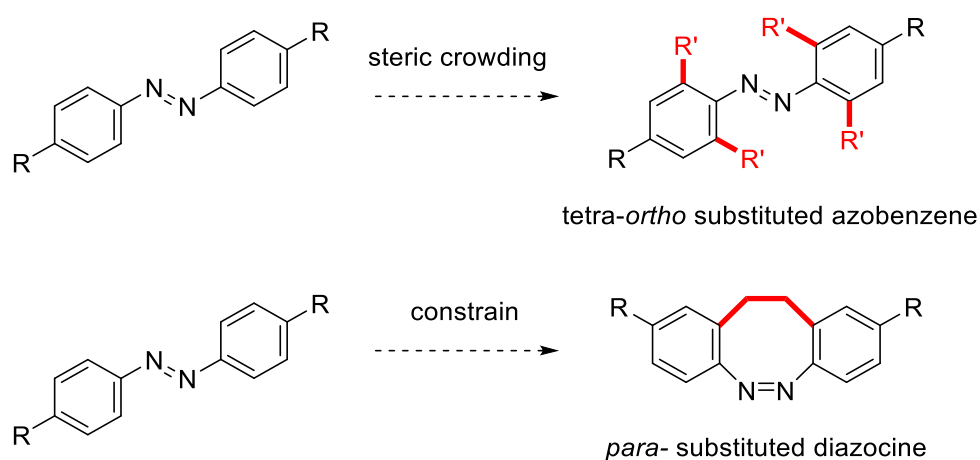


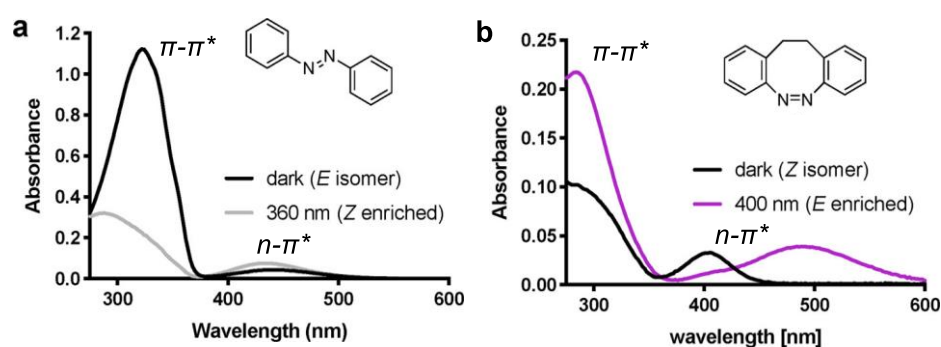
Figure 2.2: Set of azobenzene ligands prepared by Zhou and coworkers to control G4 topology.

## 2 Design and synthesis of photoswitchable ligands

With this in mind, azobenzenes are considerably more robust towards photoswitching without showing decomposition. Given the existing research conducted by Zhou and colleagues on azobenzene ligands along with its limitations (as discussed in the introduction), it was decided to focus on designing azobenzene analogues with different photophysical properties. Features that can be changed include the capacity to photoswitch without the use of UV light, a known DNA damaging agent. There are many ways to tune azobenzene photophysical properties, one option can be to change the aromatic rings to other heteroaromatic moieties, such as pyrazole. This has been shown to alter the half-life, in one case resulting in a *cis*-azopyrazole with a half-life of around 46 years.<sup>119</sup>



**Figure 2.3:** Methods to design different azobenzene analogues by affecting the steric environment of the N=N bond.



**Figure 2.4:** Comparison of UV-Vis absorbance spectra for dark and irradiated **a)** azobenzene and **b)** diazocine in DMSO. Figure adapted with permission © 2019 American Chemical Society.<sup>120</sup>

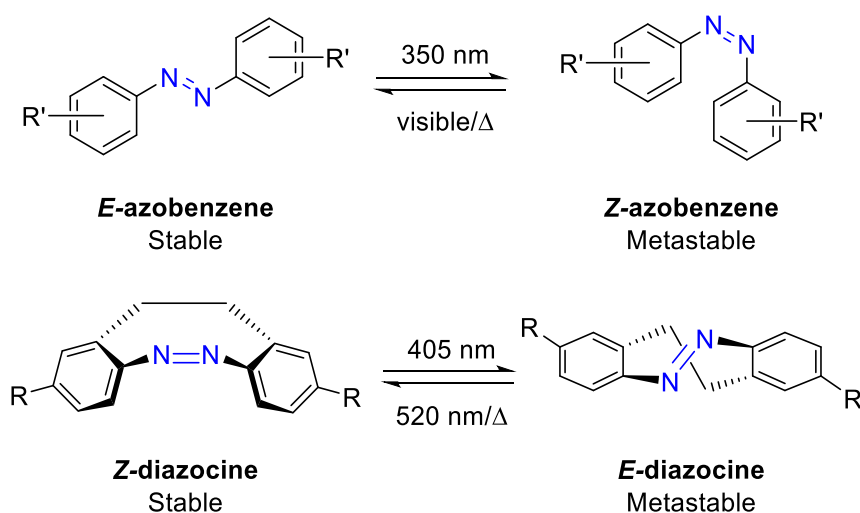
Another potential strategy to overcome previous limitations is to change the steric environment of the N=N double bond (**Figure 2.3**). Traditional azobenzenes have overlapping  $S_1$   $n-\pi^*$  absorption bands in the visible range, thus requiring UV light to excite the higher energy  $S_2$   $\pi-\pi^*$  band for *trans*  $\rightarrow$  *cis* isomerisation (**Figure 2.4a**). Substituting the *ortho*-

## 2 Design and synthesis of photoswitchable ligands

positions of the N=N double bond with -OMe, -Cl or -F results in tetra-*ortho*-substituted azobenzenes (**Figure 2.3**). This causes the  $n\text{-}\pi^*$  absorption band of trans-azobenzenes to red-shift and separate, so photoswitching between trans and cis isomers with visible light is now possible.<sup>121–123</sup>

The other method to affect the sterics around the azo bond is by introducing a conformational constraint by connecting the two aryl rings through an ethylene bridge. The added constraint increases the photoswitching quantum yield due to decreased conformational flexibility of the azobenzene. The 8-membered ring also results in a ring strain of  $17.2\text{ kcal mol}^{-1}$  for the *E*-isomer, causing the favoured stable isomer to be the *Z*-isomer (**Scheme 2.1**).<sup>124,125</sup> This change in geometry also separates the two  $n\text{-}\pi^*$  absorption bands, so diazocines can be photoswitched using non-overlapping lower-energy visible light (**Figure 2.4b**).

These properties make diazocines an attractive photoswitch class with the advantages of azobenzenes – robust photoswitching, large differences in structure. But with improvements to the red-shifted photoswitching wavelengths and quantum yields. The difference in the shapes of the stable and metastable isomers for diazocines and azobenzenes also means that alternative strategies can be used to design light-sensitive drugs with opposite selectivity. Azobenzenes will require the desired inactive state to be linear, and diazocines require the inactive state to be in the bent configuration.

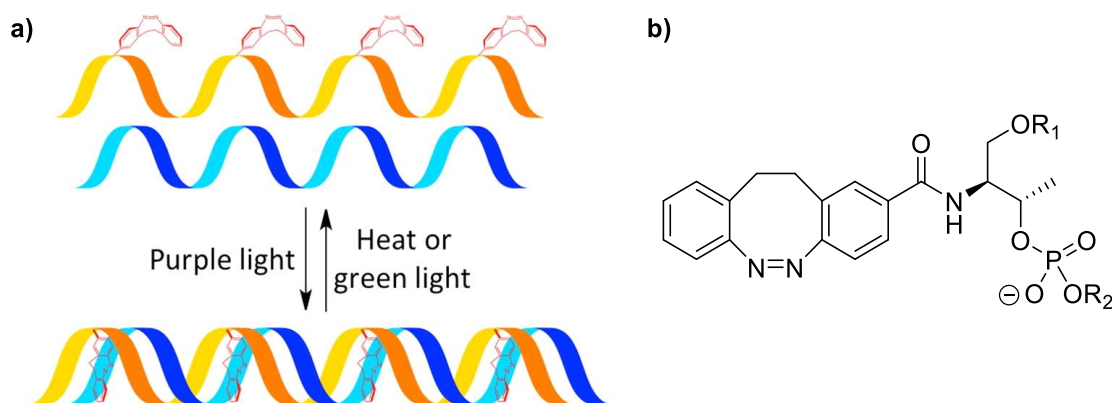


**Scheme 2.1:** Comparison of azobenzene and diazocine photoswitching behaviours.

## 2 Design and synthesis of photoswitchable ligands

Diazocines have shown use in multiple settings (photoswitchable pharmacophores/agonists,<sup>126–128</sup> coordination cages,<sup>129</sup> and peptide cross-linking<sup>130</sup>). They are novel ligands for G-quadruplexes where the only DNA related use is a covalently incorporated phosphoramidite building block that was synthesised in to a DNA oligomeric sequence as photoswitchable oligomer which could hybridise upon irradiation (**Figure 2.5**).<sup>131</sup>

From previous work, we have found that divalent ligands selectively bind to G4s over duplex forming sequences.<sup>132</sup> Ligands bearing pyridinium substituents directly linked to the central core showed excellent binding and stabilisation properties compared to alkyl piperazine substituted ligands.<sup>118</sup>



**Figure 2.5:** a) Diazocines incorporated into an oligonucleotide sequence for photoregulation of DNA hybridisation b) Diazocine phosphoramidite building block. Figure a) used with permission, © 2015 Elsevier.<sup>131</sup>

With this in mind, the final design of the initial ligand targets were set to be *para*-disubstituted diazocines with two methyl pyridinium substituents. For further structural investigations, the position of the nitrogen atom would be changed by exchanging it with *para*-, *meta*- and *ortho*-pyridinium moieties. Larger heterocycles were also considered as side-chains, to observe the potential effects in increasing the size of the heterocycle could have on the photoswitch and as a potential G4 ligand.

### 2.1 Ligand synthesis

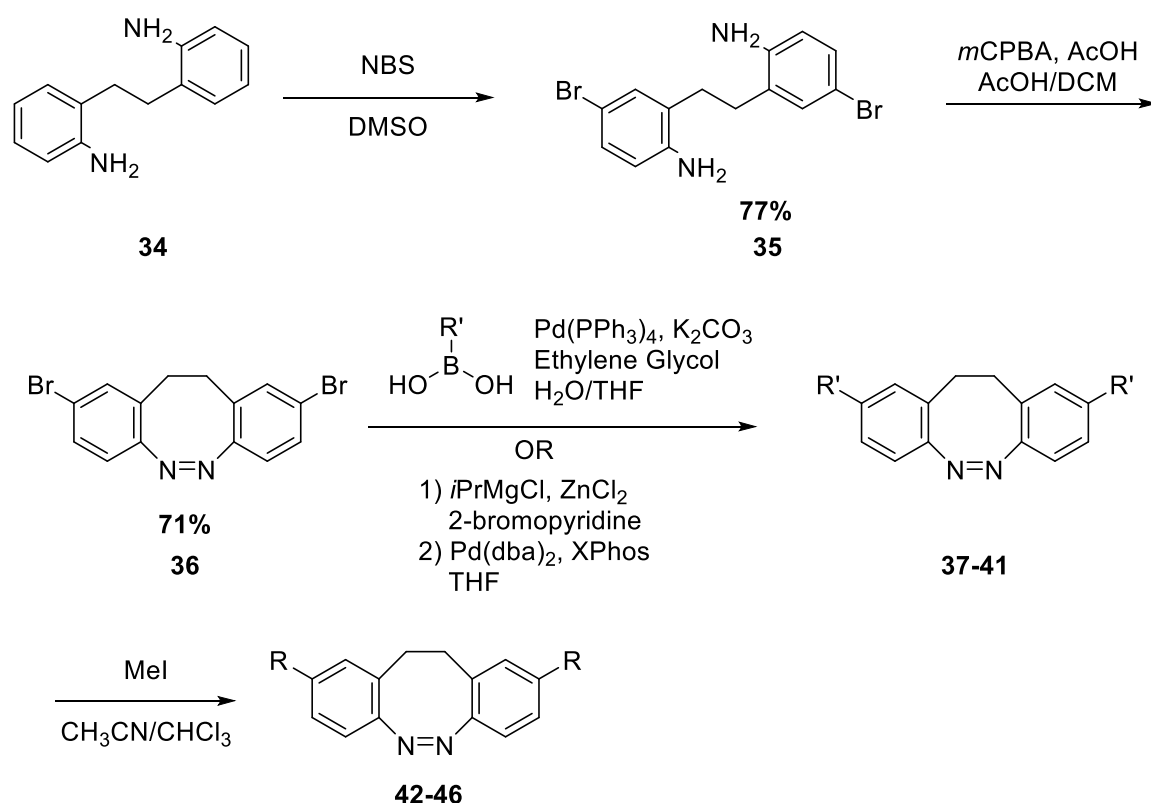
#### 2.1.1 Diazocine ligand synthesis

The synthesis of the diazocine core began from commercially available dianiline **34** which was functionalised with N-bromosuccinimide with the  $S_EAr$  mechanism favouring *para*-substitution to give dibromide **35** with a good yield of 77%. Azo bond formation was then performed using the oxidative cyclisation route to synthesise diazocines. This uses mild oxidation conditions through the slow addition of *m*CPBA over 12 hours in a mixture of acetic acid and DCM solvent, via the Baeyer-Mills reaction.<sup>120</sup> The slow addition allows the oxidation of a single aniline group to a nitroso intermediate which can then be reacted with the remaining free amine in an intramolecular condensation reaction to form the azo bond, while also preventing over-oxidation side products. The reaction is also carried out at high dilution to prevent intermolecular dimers or oligomer formation. Using these conditions the brominated diazocine core **36** can be obtained in good yields of 71%. This core can be later functionalised with aryl groups to diversify the substituents. Diazocine core **36** was also crystallised, and X-ray crystallography showed the expected bent geometry of the *Z*-isomer.



**Figure 2.6:** X-ray crystal of brominated diazocine core **36**. Showing a characteristic bent structure of the *Z*-isomer.

## 2 Design and synthesis of photoswitchable ligands



R =						
		<b>37</b>	<b>38</b>	<b>39</b>	<b>40</b>	<b>41</b>
Cross-coupling	<b>37-41</b>	<b>99%*</b>	<b>94%*</b>	<b>69%</b>	<b>59%</b>	<b>&lt;72%**</b>
N-methylation	<b>42-46</b>	<b>66%</b>	<b>98%*</b>	<b>21%</b>	<b>99%</b>	<b>70%</b>

\* Synthesis by Dr Ramos-Soriano, \*\* 30% O=PPh<sub>3</sub> by NMR

Scheme 2.2: Synthesis route to diazocine ligand library **42-46**.

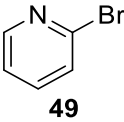
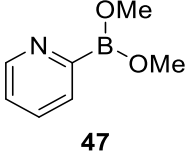
Two different cross coupling methods were then utilised to introduce a nitrogen containing aromatic heterocycle. Diazocines **37/38/40/41** were synthesised by Suzuki cross coupling conditions, with 3 equivalents of readily available stable boronic acids and provided the desired product in yields of 59-99% (**Scheme 2.2**).

Synthesis of **39** was more challenging as the 2-pyridyl boronic acid starting material is prone to protodeboronation and thus is not readily available,<sup>133</sup> necessitating an alternative route

## 2 Design and synthesis of photoswitchable ligands

that did not require this reagent. Four different conditions were tested in the synthesis of **39** (Table 2.1). The first method began with the attempted synthesis of boronate ester **47** (Entry 1). Boronate **47** was used without purification in the next cross coupling step.<sup>134</sup> However, the instability of the isolated boronate ester **47** caused the subsequent cross coupling reaction to fail and the starting material was recovered.<sup>133,135</sup>

Table 2.1: Conditions used in cross coupling attempts of 2-pyridyl reagents to the brominated diazocine core **36** to give 2-pyridyl **39**.

Entry	Starting material	2-pyridyl boronate synthesis / yield	2-pyridyl group	Cross Coupling Conditions	Yield
1		B(OMe) <sub>3</sub> , 1.6 M <i>n</i> BuLi, Et <sub>2</sub> O, -78 °C <sup>134</sup> <b>Unstable product</b>		Pd(PPh <sub>3</sub> ) <sub>4</sub> , 2M K <sub>2</sub> CO <sub>3</sub> toluene, 85 °C <sup>135</sup>	No reaction
2	<b>49</b>	1) B(O <i>i</i> Pr) <sub>3</sub> , 2.5 M <i>n</i> BuLi, THF, -78 °C 2) MIDA, DMSO, 115 °C, <b>50%</b> <sup>136</sup>	<b>48</b>	Pd(OAc) <sub>2</sub> , XPhos, 3M K <sub>3</sub> PO <sub>4</sub> , THF, 60 °C, overnight <sup>137</sup>	Incomplete
3	<b>49</b>	See Entry 2	<b>48</b>	Pd( <i>dba</i> ) <sub>2</sub> , XPhos, K <sub>2</sub> CO <sub>3</sub> , Cu(OAc) <sub>2</sub> , DMF 100 °C, 4 h <sup>138</sup>	Incomplete
4	-	-	<b>49</b>	1) 2M <i>i</i> PrMgCl, ZnCl <sub>2</sub> , RT 2) Pd( <i>dba</i> ) <sub>2</sub> , XPhos, THF, 65 °C <sup>139</sup>	<b>69%</b>

The next two conditions used methyliminodiacetic acid (MIDA) boronate **48** (Entry 2 and Entry 3).<sup>136</sup> MIDA compounds are designed to be stable to protodeboronation because the amine nitrogen in the 9 membered ring complexes with the empty boron orbital, protecting it from unwanted decomposition products (Figure 2.7). Synthesis of MIDA boronate **48** was somewhat tedious as the distillation of DMSO under vacuum was slow, but was achieved with a 50% yield. The subsequent “slow-release” cross coupling conditions with boronate **48** then allows for transiently generated boronic acid to participate in cross coupling with diazocine **36**. However, conditions 2 and 3 led to incomplete reactions with large amounts of starting material remaining.<sup>137,138</sup> It is possible that the cross coupling reaction time was too short, and the reaction was stopped before completion.

## 2 Design and synthesis of photoswitchable ligands

---

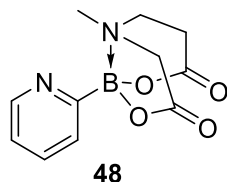


Figure 2.7: B-MIDA **48**, showing the nitrogen atom chelating boron species.

For the final method, a Negishi cross coupling condition that did not require a boron species was trialled (Entry 4). 2-bromopyridine underwent metal-halogen exchange with isopropyl magnesium chloride. This generated a new Grignard which was transformed *in situ* into an organozinc reagent using a solution of zinc chloride in a transmetalation step. Negishi cross coupling conditions were deemed suitable for the diazocine substrate **36** as there were no functional groups that could not tolerate nucleophilic organozinc or Grignard reagents.<sup>139</sup> Fortunately, the conditions including the cross coupling step provided the desired disubstituted 2-pyridyl **39** with a good yield of 69%.

Methylation of the N-heterocycles was then carried out using iodomethane, where in most cases (**37/38/40/41**) the product was easily isolated as a yellow solid precipitate without any further purification steps required, with yields ranging from 66-99%. In the case of the methylation of **41**, the presence of triphenylphosphine oxide byproducts from the palladium catalyst in the previous step was observed. Fortunately, this did not result in any complications as <sup>31</sup>P NMR could identify it as the phosphine oxide that lacked a nucleophilic lone pair which could compete with the desired N-methylation step that provided 3-quinolinium **46**. However, the methylation process encountered challenges with 2-pyridyl **39**, primarily due to steric hindrance arising from *ortho*-methylation. This compound required longer reaction times without any precipitation of product, though leaving it for too long led to degradation, likely due to over methylation at the azo-bond and much poorer yields of 21% were achieved.

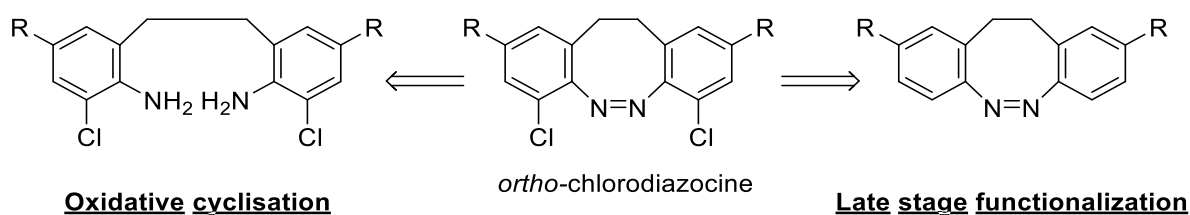
### 2.1.2 *o*-chlorodiazocine ligand synthesis

There were also attempts to synthesise another derivative of the diazocine ligand by the addition of *ortho*-chloro substituents on the diazocine core. The rationale behind this design is to 1) provide additional handles to derivatise upon and 2) mimic tetra-*ortho*-substituted



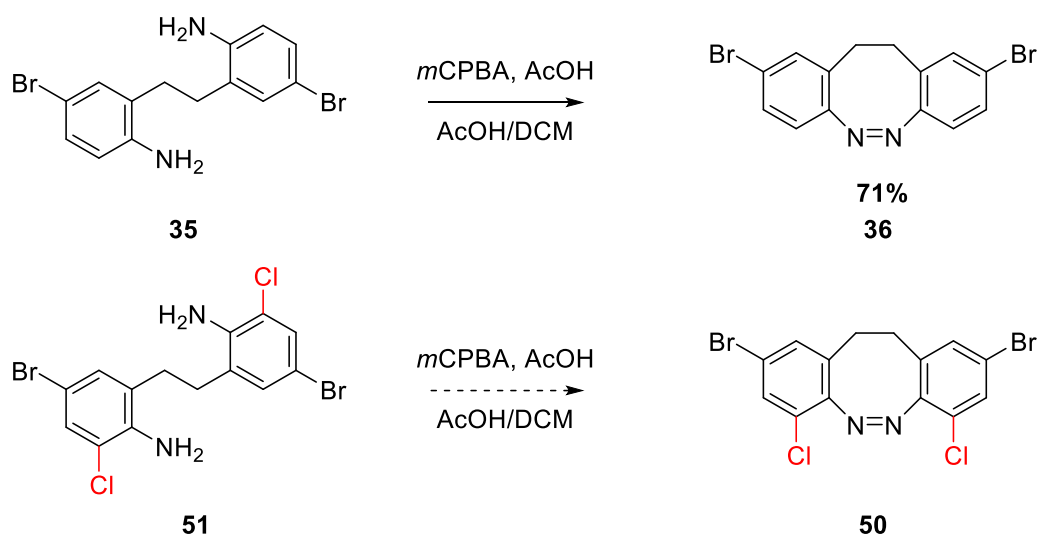
## 2 Design and synthesis of photoswitchable ligands

azobenzenes. This follows literature precedent where the steric crowding of tetrasubstituted linear azobenzenes have a red shifted photoswitching behaviour.<sup>121,122</sup>



**Scheme 2.3:** Strategies to synthesise *ortho*-chlorodiazocines (center), **Left:** oxidative cyclisation of a dichloroaniline. **Right:** late stage functionalisation of a disubstituted diazocine.

There were two different strategies to achieve the intermediate pre-cross coupling target **50**, a good target molecule that would allow for easy diversification. The initial route involved the synthesis of a chlorinated dianiline **51** which could then be treated with the same oxidative cyclisation conditions as with the formation of diazocine **36** (**Scheme 2.4**).<sup>120</sup> The other strategy would be the late stage functionalisation of a preexisting diazocine, covered in more detail in a later section.

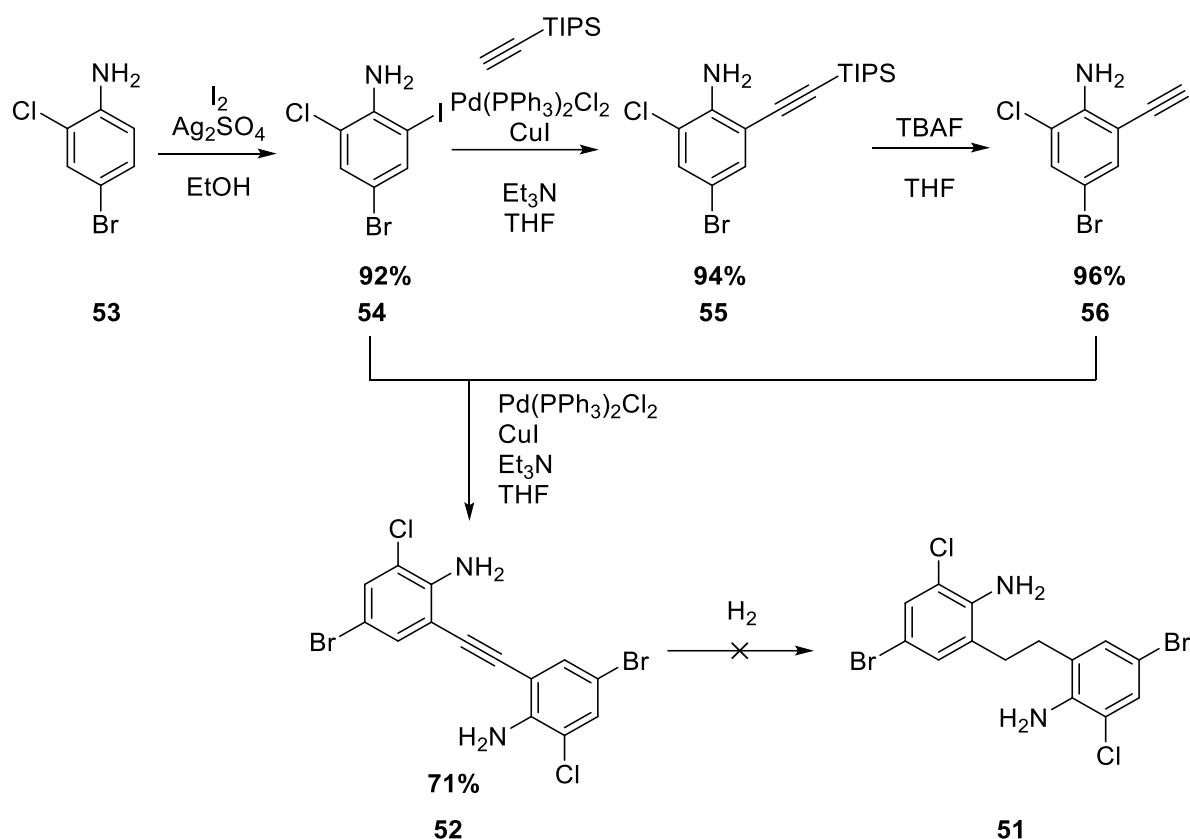


**Scheme 2.4:** Target diazocine core scaffold **50**, synthesised from *o*-chlorodianiline **51** by oxidative cyclisation method.

The synthetic route for the initial dichloroaniline target **51**, was to couple together two aniline building blocks with the correct halogen substitution. Initially, an  $sp^3$ - $sp^3$  methyl-methyl homocoupling approach was assessed and dismissed as this method did not show any substrates with aniline functionality.<sup>140</sup> Eventually it was decided to build the 2-carbon ethylene bridge via reduction of an alkyne. The synthetic route is shown below and involved

## 2 Design and synthesis of photoswitchable ligands

two Sonogashira cross coupling steps to synthesise alkyne **52**, followed by an alkyne reduction to give the desired target **51** (Scheme 2.5).<sup>120</sup>



**Scheme 2.5:** Synthetic route towards target molecule **51** via the synthesis of alkyne **52** from two Sonogashira reactions.

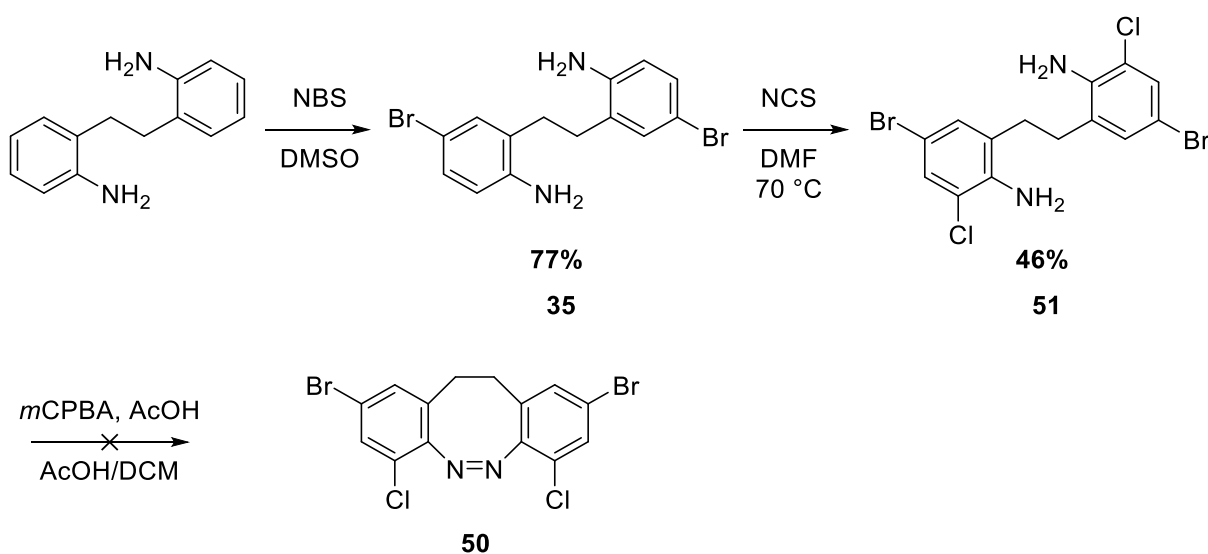
The synthetic steps towards the cross coupling partners **54** and **56** began from the same initial starting material, aniline **53**. The iodination of **53** was catalysed by the addition of silver sulfate, then a portion of **54** was then transformed into alkyne **56** through the introduction of a TIPS protected alkyne in Sonogashira and silyl deprotection steps. These were all high yielding reactions, performed at gram scales (92-96%).

However, the  $^1H$  NMR of the crude products from the second Sonogashira cross coupling step of **54** and **56** showed at least two aniline compounds. This was unsurprising as some side products such as homocoupling of the alkyne **56** or cross coupling at the bromo position (instead of iodo) could also occur. However, it was challenging to purify due to the alkyne **52** having limited solubility and the side products having very similar separation.

## 2 Design and synthesis of photoswitchable ligands

A small amount of purer alkyne **52** was taken through to the next step. However, the hydrogenation reactions of alkyne **52** using Pd/C gave complicated mixtures due to the competing reduction of the carbon-halogen bonds from oxidative insertion reactions of the palladium catalyst.<sup>141</sup> An alternative method by using a platinum catalyst was attempted, as literature examples have shown successful hydrogenation of aromatic alkynes containing halogens.<sup>142</sup> However, this still caused a complicated mixture and was attributed to the aniline poisoning the reactivity and selectivity of the hydrogenation process. It is possible that using an alternative hydrogenation method with diimides would be possible as it would avoid hydrogenolysis.<sup>143</sup>

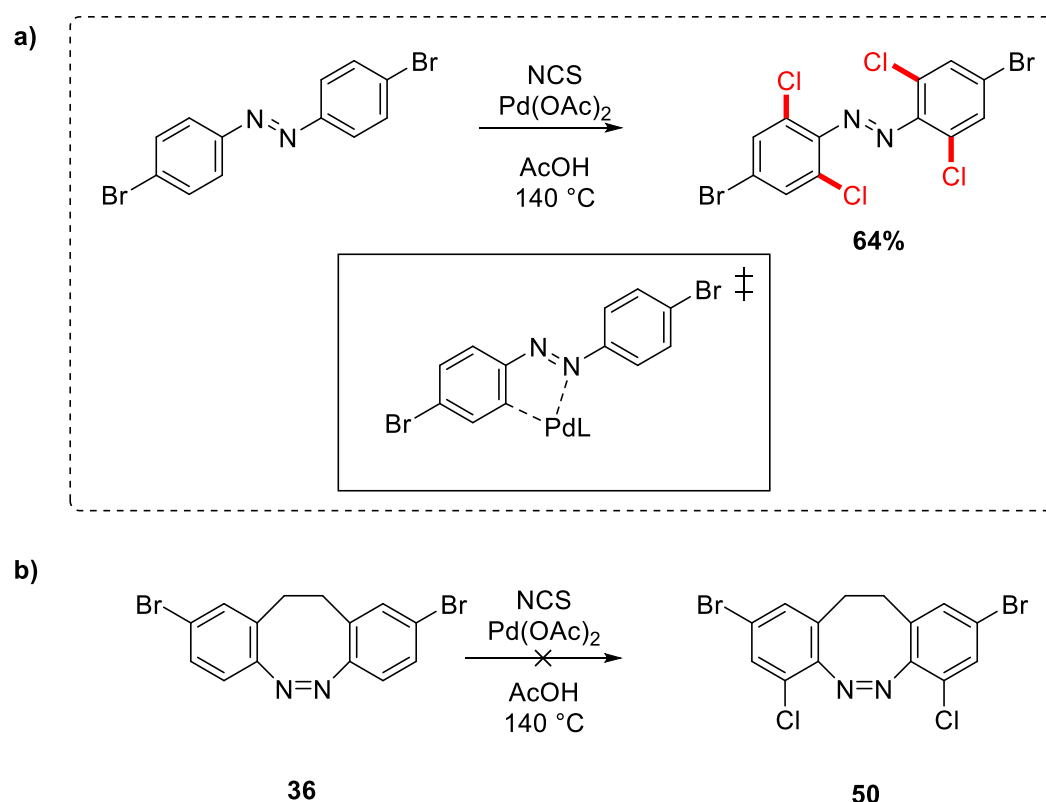
Another method to reach **51** focused on dichlorination of dibromoaniline **35**, using N-chlorosuccinimide in DMF, analogous to the bromination reaction (**Scheme 2.6**). However, this reaction was low yielding and of the 46% isolated yield, the product was only 80% pure as determined by analytical HPLC and NMR (with the likely side products being a tetra-bromo species carried over from the previous step). Attempts to purify compound **51** at a large scale were unsuccessful because of the poor solubility of anilines, and synthesis moved forwards to the target *o*-chloro diazocine **50**. However, unlike the oxidative cyclisation reaction using *m*CPBA in acetic acid for diazocine **36**, this reaction was not successful and can be attributed to an increased steric barrier towards cyclisation of the aniline groups or the presence of impurities.



**Scheme 2.6:** Successful synthetic route towards synthesis of *ortho*-chloroaniline **51** and the attempted oxidative cyclisation step to synthesise *o*-chloro diazocine scaffold **50**.

## 2 Design and synthesis of photoswitchable ligands

Examination of literature precedent shows the oxidative cyclisation approach to diazocines does not tolerate well *ortho*-substituted substrates. Only a single substrate, *ortho*-fluorodiazocine, which had medium yield of 46% was shown.<sup>120</sup> With it being apparent that the *ortho*-chloro functionality was likely sterically hindering cyclisation, the second strategy to *ortho*-chlorodiazocines by means of late-stage installation was attempted. Late-stage functionalisation *ortho*- to an azobenzene double bond has been reported by Trauner<sup>144</sup> by direct chlorination using a palladium catalyst (**Scheme 2.7a**). This method had high functional group tolerance and is advantageous as typical tetra-*ortho*-azobenzene synthesis through formation of the azo bond gives low yields (~19-26%) due to the steric hindrance from the *ortho*-chlorines.<sup>144</sup>

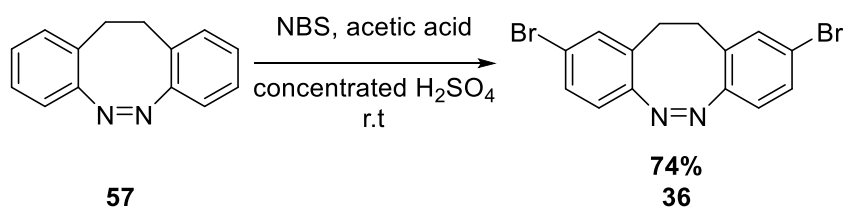


**Scheme 2.7:** a) Example of late stage *ortho*-chlorination of azobenzenes, with a palladium 5-membered-ring transition state.<sup>144</sup> b) Attempted late-stage functionalisation route towards *o*-chloro diazocine scaffold **50**.

Unfortunately, this method was also unsuccessful, with the starting material being observed in the NMR after work up of the crude reaction. It was hypothesised that these conditions only apply for linear azobenzenes (**Scheme 2.7a**) as it can form 5-membered ring intermediates with the palladium catalyst, not possible with the constrained diazocine system.<sup>144,145</sup>

## 2 Design and synthesis of photoswitchable ligands

Both tetra-*o*-chlorinated azobenzenes and diazocine **36** are already sterically constrained systems. Thus, it was no surprise that the target *o*-chloro diazocine **50** was an even more challenging compound to synthesise, since it attempts to combine both steric constraints together, resulting in an extremely high energy barrier towards the oxidative cyclisation route. In 2020, electrophilic aromatic substitution of **57** was achieved by bromination of an unfunctionalised diazocine. This raises the possibility that the sterically challenging *ortho* positions may be possible without metal catalysis, albeit in strongly acidic conditions (**Scheme 2.8**).<sup>146</sup> Despite this, further attempts to synthesise **50** were stopped due to the challenges in synthesising this compound and efforts were directed to screening the synthesised library to determine what patterns were suitable for G4 binding.



Scheme 2.8: S<sub>E</sub>Ar late-stage functionalisation of diazocine **57**.

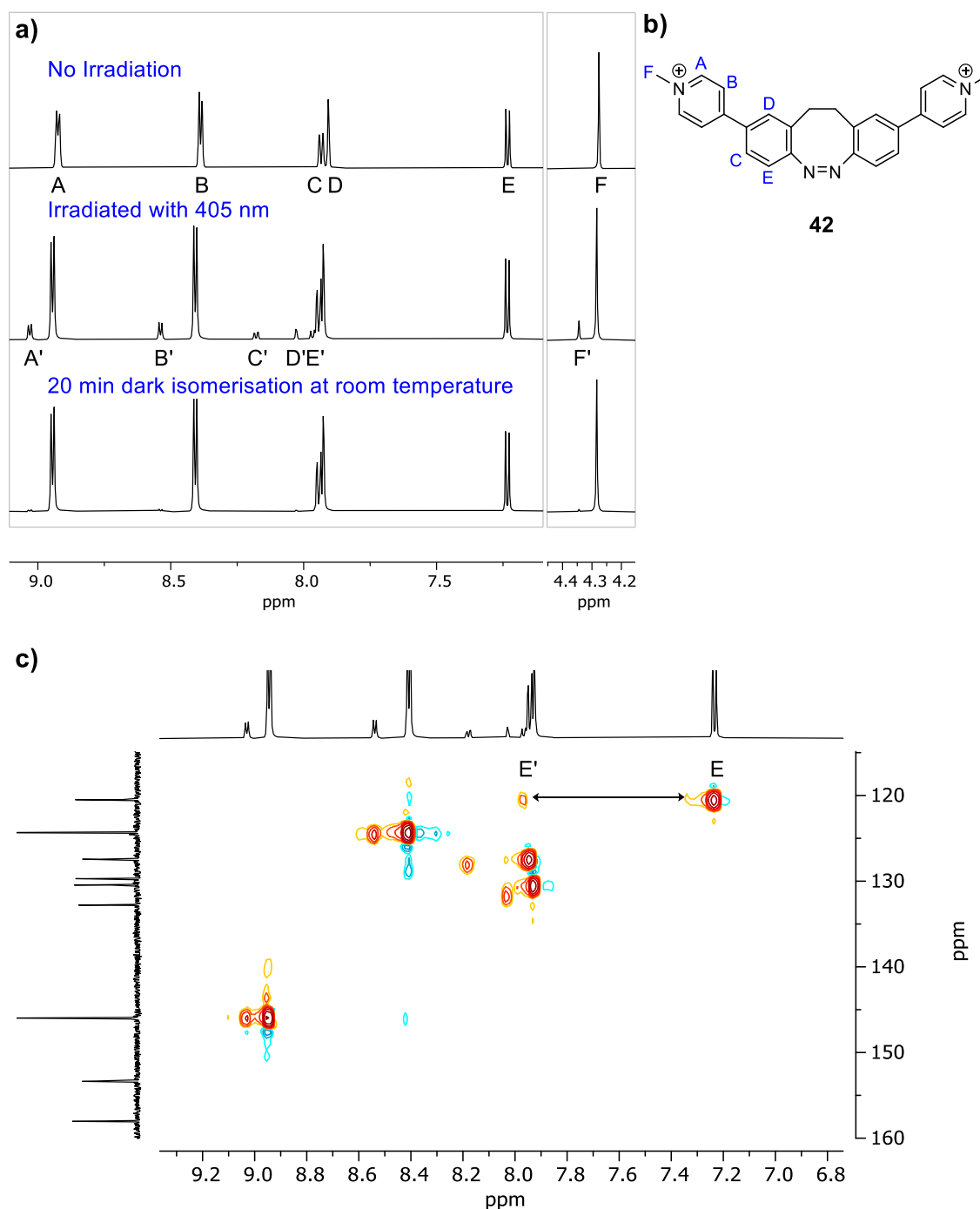
## 2.2 Ligand characterisation and photochemistry

Diazocine ligands **42-46** were isolated as their *Z*-photoisomers, with none of the <sup>1</sup>H NMR spectra showing any peaks belonging to the *E*-isomer. Characterisation of the metastable *E*-isomer requires photoisomerisation of the *Z*-species, where it will achieve a particular steady state *E:Z* ratio, formed by constant irradiation at a given wavelength, known as the photostationary state (PSS). Two types of light sources were used to photoisomerise the diazocine compounds. For most experiments, a 4.5 mW 405 nm or 520 nm Collimated Laser Diode Module (CPS405 and CPS520) were suitable to achieve desired photoswitching. These laser sources had narrow beam shapes so to more efficiently irradiate samples with larger volumes and concentrations, a 60 W 395-405 nm LED strip was used. However, this generated a significant amount of heat after long use, causing it to affect the *E/Z* thermal equilibrium so was not useable for biophysical assays which required room temperature conditions.

Structural characterisation of the *Z*-isomers by NMR spectroscopy showed that the aromatic chemical shifts in the diazocine core were always upfield (7-8 ppm) relative to the electron

## 2 Design and synthesis of photoswitchable ligands

deficient N-heterocyclic substituent (8-9 ppm). Whilst a higher number of scans typically lead to better signal to noise ratios for NMR, it was important to achieve a careful balance to characterise the metastable *E*-photoisomers, as increasing number of scans also meant increasingly lower populations of *E* to *Z* ratio.

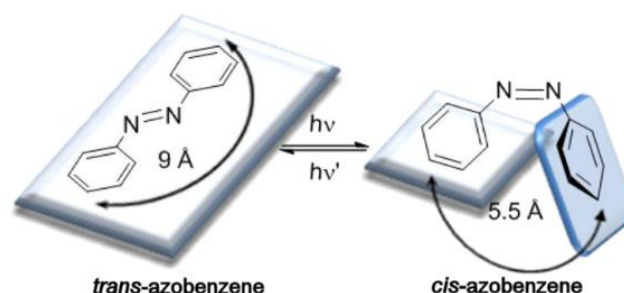


**Figure 2.8:** NMR characterisation of *Z*-**42**, peaks A-F and *E*-**42**, peaks A'-F'. **a)**  $^1\text{H}$  NMR spectra of **42** in DMSO- $d_6$ , Top: Prior to irradiation, Middle: irradiated at 405 nm, Bottom: 20 minutes dark isomerisation at room temperature **b)** Structure of **42** with key peaks labelled A-F. **c)** HSQC of irradiated **42**. Proton E, showing the largest change in chemical shift is labelled.

## 2 Design and synthesis of photoswitchable ligands

After irradiation with 405 nm light, most of the *E*-isomer peaks were deshielded compared to the stable *Z*-isomer. The shorter distance between the two aromatic rings in the bent *Z*-structure leads to increased shielding of the protons due to magnetic anisotropy (**Figure 2.9**).<sup>147</sup> The proton *ortho* to the N=N double bond, closest to the opposite aromatic ring, always had the largest downfield shift after irradiation. This effect is illustrated in **Figure 2.8a** as proton E of **42** has a 0.7 ppm downfield shift after isomerisation and proton C has the next largest downfield shift by virtue of it having the next nearest proximity to the N=N double bond. The exception to this was the chemical shifts of the ethylene protons, which were more shielded in the *E*-isomers. These protons were both broad and overlapped with each other, making identification more challenging.

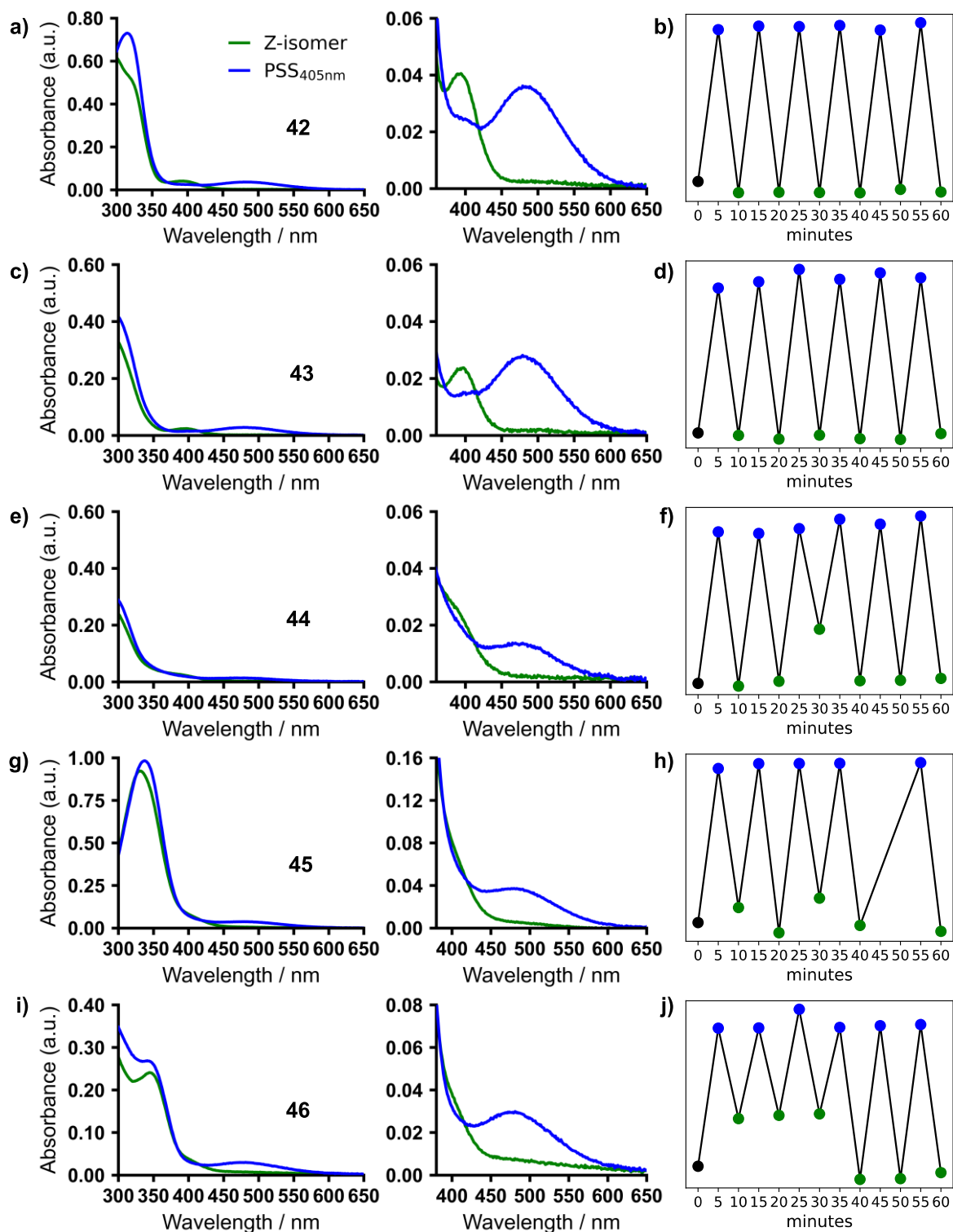
Pleasingly, HSQC was helpful in assigning the *E*-isomer NMR spectra, as it was able to acquire spectra within a relatively short period of time and also provide information of <sup>13</sup>C chemical shifts. For compounds with longer half-lives, a good <sup>13</sup>C spectra was acquired more easily, where it is clear that the signals corresponding to the *E*-isomer are also more downfield than *Z*-isomer carbon signals. For those with shorter half-lives, only non-quaternary carbons were visible through HSQC spectra.



**Figure 2.9:** Bent *Z*-azobenzene structure results in *Z*- protons experiencing aromatic ring magnetic shielding effect due to increased proximity of the two rings.<sup>147</sup>

Following this, the photoswitching capabilities of the newly synthesised diazocine ligands were investigated by UV-Visible absorption and NMR spectroscopy. The ligands were dissolved as a 10 mM stock solution in DMSO- $d_6$  and then further diluted in buffer and irradiated with alternating 405 nm and 520 nm light.

## 2 Design and synthesis of photoswitchable ligands



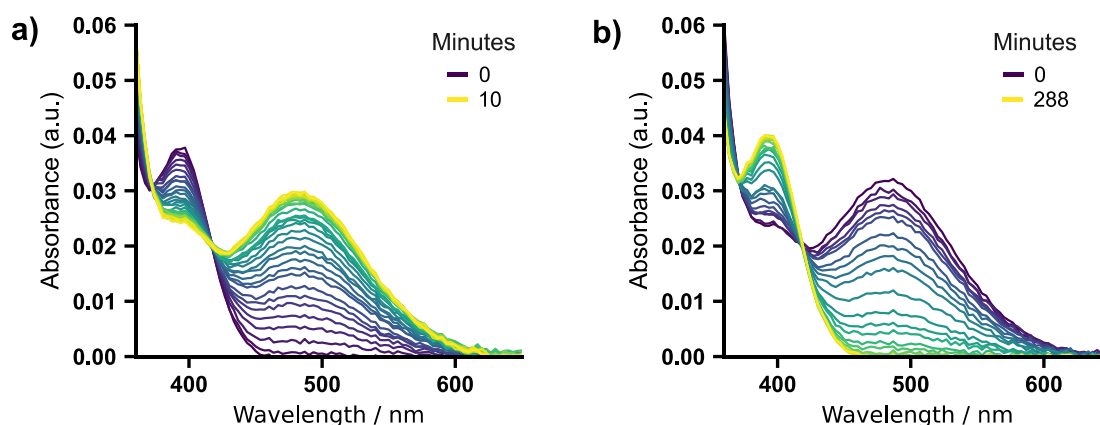
**Figure 2.10:** Photoswitching of diazocine **42-46** at 20  $\mu\text{M}$  in 100 mM potassium phosphate buffer, pH 7.4. Excitation at  $\lambda_{\text{abs}} = 405$  nm to PSS, followed by excitation at  $\lambda_{\text{abs}} = 520$  nm back to the Z-diazocine. **Left and middle:** blue – absorption spectra of PSS after 405 nm irradiation, green – absorption spectra of Z-diazocine after 520 nm irradiation. **Right:** Photoswitching cycles at 480 nm over 60 minutes. **a-b)** 4-pyridinium **42**, **c-d)** 3-pyridinium **43**, **e-f)** 2-pyridinium **44**, **g-h)** biaryl **45** and **i-j)** quinolinium **46**.



## 2 Design and synthesis of photoswitchable ligands

UV-VIS spectra of the diazocines were measured from 300-700 nm in 100 mM potassium phosphate buffer solutions with a pH of 7.4. For the pyridinium species **42-44**, there is a blue shift of the maxima of the *Z* and *E*  $\pi$ - $\pi^*$  absorption bands across the pyridinium series (<300 nm and <315 nm) (**Figure 2.10**). In contrast, changes to the substituent did not affect the *Z*  $n$ - $\pi^*$  absorption band at 390 nm which upon photoisomerisation with a 405 nm laser, disappears with a new maxima appearing at 480 nm. For biaryl **45** and quinolinium **46**, there is a general red shift of the  $\pi$ - $\pi^*$  absorption bands relative to the pyridinium species. This is expected for these species with increased  $\pi$ -conjugation. This likely results in the  $n$ - $\pi^*$  band being less pronounced in the *Z*-isomer, potentially resulting in a lower PSS of the excited *E*-diazocine due to the  $\pi$ - $\pi^*$   $n$ - $\pi^*$  band overlap, and a poorer distinction between the two extinction coefficients at 405 nm.<sup>148,149</sup> All ligands also showed robust photoswitching with no photofatigue after 6 cycles.

In addition to photoswitching capability, the thermal half-life of the *E*-diazocine ligands were also investigated, where it was shown that their stability was sensitive to thermal environmental changes. The thermal relaxation of all *E*-diazocines followed a first-order rate of decay with the results shown in **Table 2.2**. It can be seen that the diazocines with the shortest half-lives were *para*-substituted 4-pyridinium **42** and 4-biaryl **45**, rationalised by full conjugation of  $\pi$ -electrons across the whole molecule.



**Figure 2.11:** a) Excitation with 405 nm to  $PSS_{405}$  and b) thermal relaxation at 25 °C of 20  $\mu$ M **42** in 100 mM potassium phosphate buffer, pH 7.4.

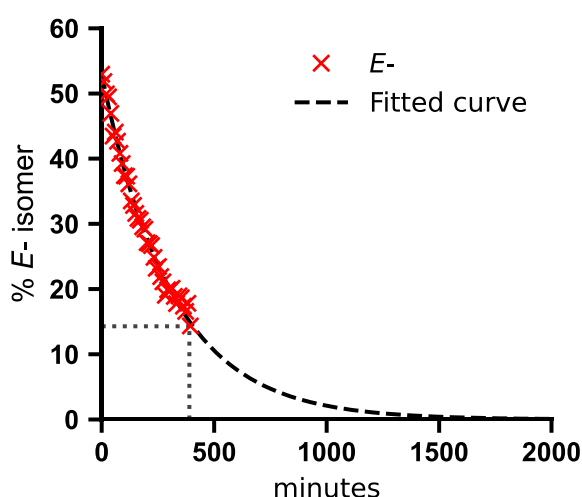
## 2 Design and synthesis of photoswitchable ligands

---

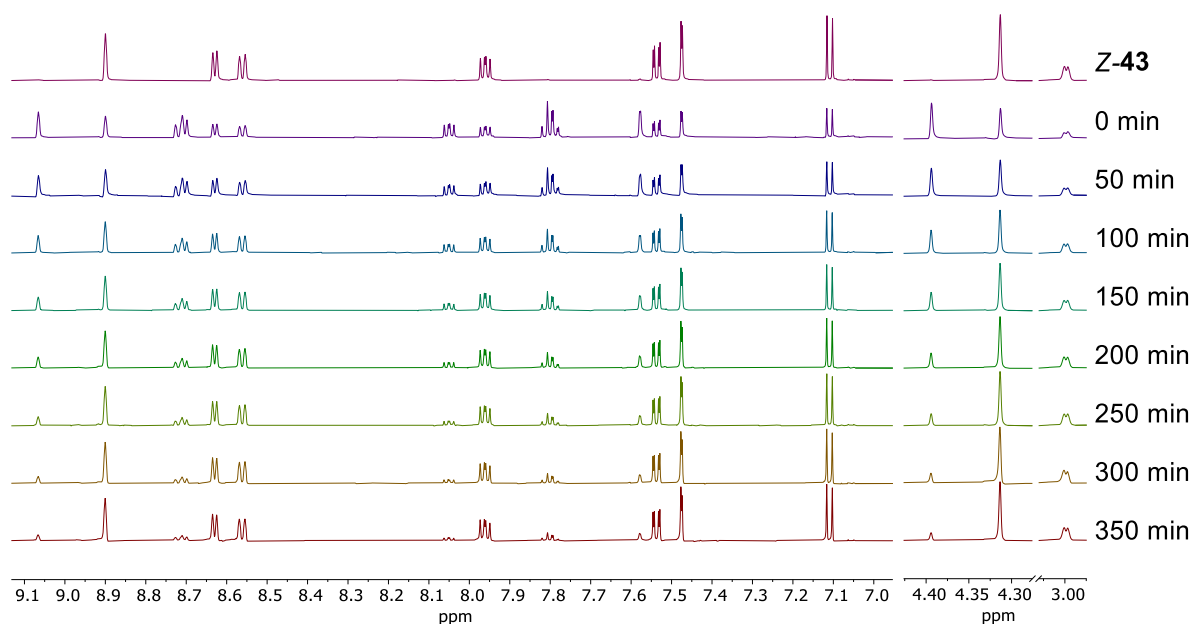
**Table 2.2:** Table of *E*-diazocine thermal half-lives at 25 °C as measured by UV-Vis spectroscopy in 100 mM potassium phosphate buffer, pH 7.4.

Ligand	Half-life / min
<b>42</b>	30
<b>43</b>	200
<b>44</b>	160
<b>45</b>	60
<b>46</b>	160

The thermal isomerisation of *E*-**43** was also investigated in a potassium buffer solution by <sup>1</sup>H NMR. The chemical shifts corresponding to *E*-**43** were also deshielded relative to *Z*-**43**, like the rest of *E*-diazocines in DMSO-d<sub>6</sub>. <sup>1</sup>H spectra were measured every 10 minutes, over a 6.5 hour period (**Figure 2.13**). This was plotted as an exponential decay from a 53:47 *E*:*Z* mixture at 0 min after irradiation and ends with a 15:85 *E*:*Z* ratio, 390 min after irradiation. The resulting half-life was calculated to be 215 minutes and is consistent with results from UV-Vis thermal relaxation investigations (**Figure 2.12**).



**Figure 2.12:** Change in quantity of *E*-**43** over 390 minutes by <sup>1</sup>H NMR, in 100 mM potassium phosphate buffer, pH 7.4.



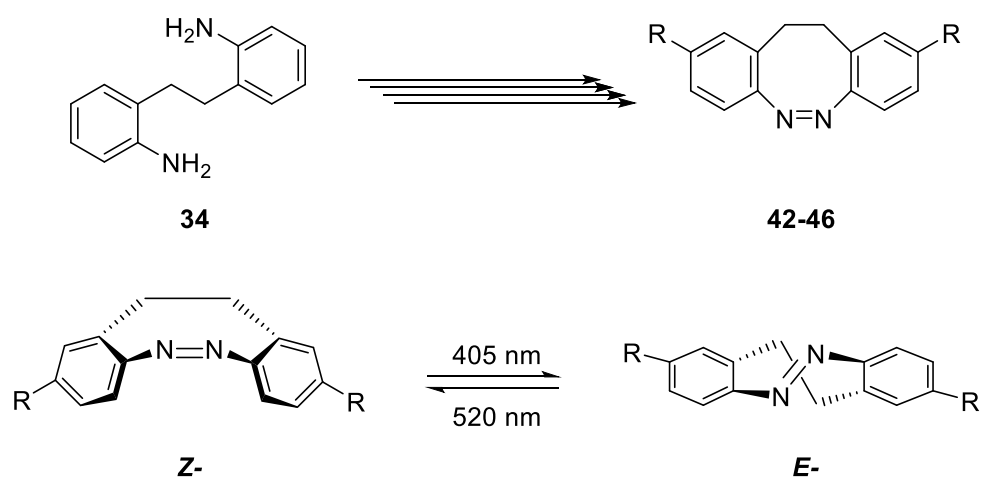
**Figure 2.13:** <sup>1</sup>H NMR showing thermal relaxation of *E*-**43** to *Z*-**43** at 25 °C, in 100 mM potassium phosphate buffer, pH 7.4. Total concentration of **43** was 446 μM.

### 2.3 Chapter conclusions

The novel disubstituted diazocines **42-46** were successfully synthesised in 4 steps with isolated yields of 59-99% for the cross coupling step and yields of 21-99% for the final methylation, and then fully characterised. Attempts at the synthesis of an *o*-dichlorodiazocine analogue was unsuccessful after various different synthetic routes.

## 2 Design and synthesis of photoswitchable ligands

---



**Scheme 2.9:** Top: successful synthesis of novel disubstituted diazocines **42-46**. Bottom: Synthesised diazocine ligands showing expected photoswitching properties, isomerising with 405 nm blue light and 520 nm green light.

These new ligands show similar photophysics with other known diazocine molecules.<sup>124,125</sup> The photoswitching kinetics and half-lives were measured by UV-Vis and NMR. All ligands demonstrated robust photoswitching after irradiating with 405 nm and 520 nm blue and green light. After irradiation to the metastable *E*-isomer, they displayed thermal half-lives in the range of 30-250 minutes, with 4-pyridinium **42** at the lowest half-life. Changing the position of the nitrogen atom within the pyridinium ring had an effect on increasing the thermal stability of the ligands, shown by their increased half-life of the *E*-isomer after irradiation with 405 nm.

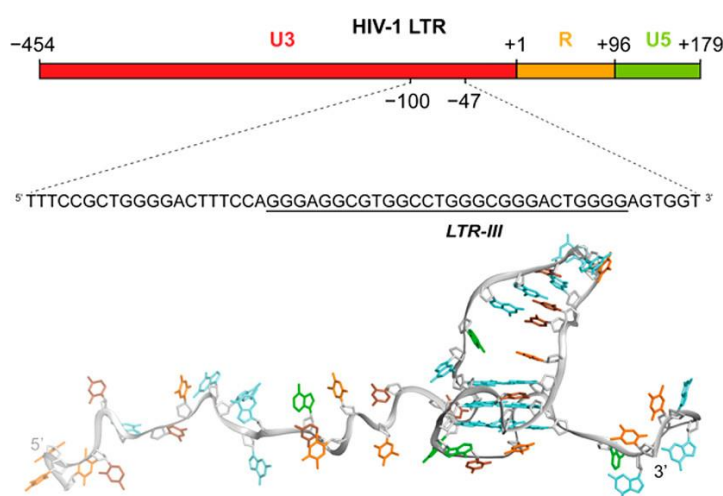
Having identified that all the ligands were successful at photoswitching, it was decided to explore the effects of this new ligand class upon G-quadruplex DNA and also the effect of these ligands on DNA structure and binding affinity after irradiation with light.

## 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

As initially described in Chapter 1, viral G4s are potential therapeutic targets with many found in the HIV RNA genome and the HIV proviral DNA genome. The Long Terminal Repeat (LTR) U3 promoter region of the proviral genome, contains guanine rich sequences which have been shown to form G-quadruplex structures. Transcription of the proviral genome is regulated in this region upstream of the transcription start site making it an interesting therapeutic target through the mediation of viral DNA transcription.<sup>53,150,151</sup>

Studies by Perrone *et. al.* showed that the addition of known G-quadruplex ligands (BRACO-19, TMPyP4) can successfully inhibit viral replication and stall transcription of the provirus with antiviral effects.<sup>53,58</sup> Other ligands that have demonstrated antiviral effects to HIV-1 included the extended core NDI, where it had good selectivity towards the HIV-1 G-quadruplexes and resulted in the inhibition of LTR promoter activity within cells.<sup>80</sup>

The two most promising sequences discovered within the LTR region are LTR-III and LTR-IV. These two sequences are overlapping and thus mutually exclusive, with LTR-III forming without the presence of external ligands, whereas LTR-IV only forms with an external ligand. Interestingly, they induce opposite effects on the LTR promoter activity. Stabilisation of LTR-III could inhibit viral transcription,<sup>53</sup> while stabilisation of LTR-IV G4 enhanced promoter activity.<sup>152</sup>



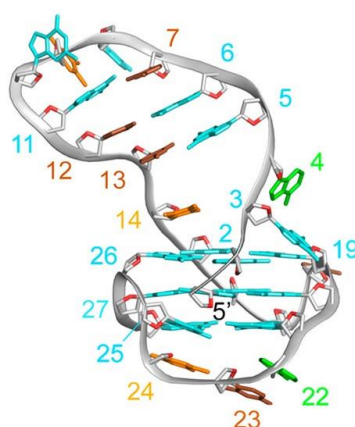
**Figure 3.1:** Cartoon illustrating location of HIV-1 guanine rich region in U3 of LTR, LTR-III sequence underlined. Cartoon of HIV-1 LTR-III structure. © 2018 American Chemical Society.<sup>57</sup>

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

Both LTR-III and LTR-IV also have detailed high-resolution NMR solution structures that can be utilised in designing highly selective ligands for these sequences. LTR-IV was shown to be a parallel G-quadruplex structure with a single nucleotide thymine bulge ( $T_m = 50.5\text{ }^\circ\text{C}$ ).<sup>152</sup> On the other hand, LTR-III ( $T_m = 65.5\text{ }^\circ\text{C}$ ) is a G-quadruplex with loop lengths of 12, 1 and 3 nucleotides. In particular, the relatively long 12 nucleotide loop is able to form a diagonal stem loop which is stabilised by 3 Watson-Crick base pairs across the face of the G-quadruplex giving a unique quadruplex-duplex hybrid structure (**Figure 3.2**).<sup>57</sup> Other high-resolution structures of naturally occurring quadruplex-duplex hybrid structures have not been reported despite bioinformatic studies suggesting that over 80,000 sequences within the human genome have the potential to form these hybrid structures. This means LTR-III with its unique stem loop structure provides an interesting structural target for ligand binding with a large, open junction for ligand binding.

On the basis of the LTR-III distinctive quadruplex-duplex topology, we set to explore ligands that could target this G4 sequence. This chapter focuses on the evaluation of the synthesised diazocine ligands **42-46** as ligands for the LTR-III sequence using a range of biophysical assays. In addition, molecular dynamic simulations of the LTR-III sequence with *E* and *Z* ligands were also used to provide supporting structural information of binding interactions.



**Figure 3.2:** Cartoon representation of the 28-nucleotide HIV-1 LTR-III sequence. Watson-Crick pairs G5-C13, G6-C12 and C7-G11 stabilise diagonal the stem loop. © 2018 American Chemical Society.<sup>57</sup>

## 3.1 Biophysical Assays

### 3.1.1 Förster Resonance Electron Transfer thermal melting assay

#### 3.1.1.1 Introduction to FRET

A quick method to analyse the stability of the secondary structure of DNA structures is through a Förster Resonance Electron Transfer (FRET) melting assay. This is a high throughput method used to determine structure stability and the effect of ligands on inducing further stability.

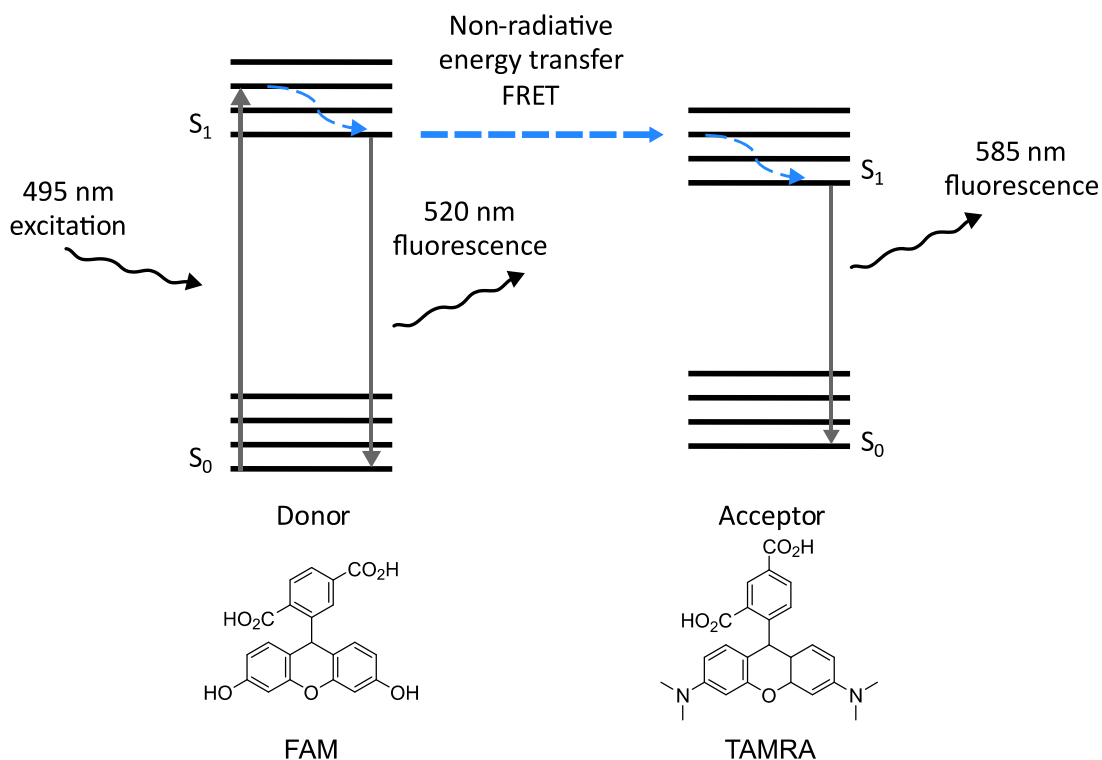
In a FRET melting assay, G4 sequences are labelled with a donor and an acceptor fluorophore. The donor fluorophore is 6-carboxyfluorescein (FAM) and the acceptor is 5-carboxytetramethylrhodamine (TAMRA). When these two fluorophores are close together in space, excitation of the donor ( $\lambda_{\text{ex}} = 495 \text{ nm}$ ) allows for a non-radiative energy transfer from the  $S_1$  energy levels of the donor and acceptor, which effectively quenches the fluorescence of the donor. If the donor and acceptor fluorophores are far apart in space, then no energy transfer mechanism takes place, and the donor fluorescence is observed ( $\lambda_{\text{em}} = 520 \text{ nm}$ ) (**Figure 3.3** and **Figure 3.4**).

The main principle behind FRET melting assays is that it allows for a high throughput method to test which ligands can stabilise a G-quadruplex. When a ligand binds and stabilises a G-quadruplex, it is expected to result in an increased melting temperature ( $T_m$ ) during the thermal denaturation process. This method can quickly screen many compounds and DNA sequences as it is done using a 96 well plate. The  $T_m$  values for each ligand-DNA pair are then compared against negative control G-quadruplex sequences that had no ligand added to calculate a  $\Delta T_m$  value, allowing for a quick comparison of ligand-G4 binding effects.

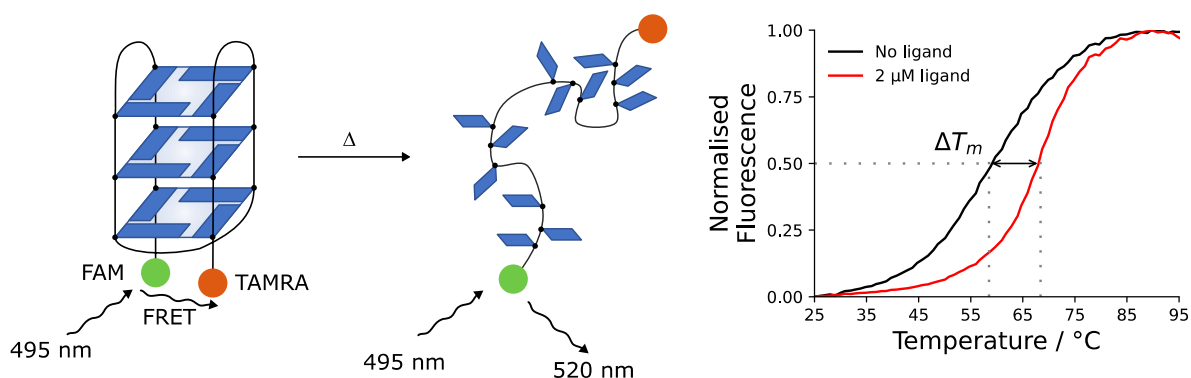
Despite this, there are some disadvantages to FRET melting assays, and results need to be taken with caution. For example, the presence of the fluorophore tags do not always properly represent the unlabelled sequence, as the ligands could interact with the fluorophores hindering binding interactions with the DNA sequence. Additionally, artefacts can appear if the ligand molecule being studied is highly fluorescent at the emission wavelength of the

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

donor molecule. Requiring non-labelled oligonucleotide studies to validate the results from the melting assay.<sup>153</sup>



**Figure 3.3:** Jablonski diagram showing Förster Resonance Energy Transfer mechanism for the FAM/TAMRA pair.



**Figure 3.4:** Overview of the FRET melting assay. **a)** Cartoon representation of FRET melting assay. Initially, fluorescence of FAM is quenched by the presence of TAMRA through non-radiative FRET. Thermal denaturation of the G4 structure causes the fluorophores FAM and TAMRA to move apart in space, leading to increased fluorescence from the donor fluorophore as temperature increases. **b)** An example melting curve showing  $\Delta T_m$  after ligand stabilisation of secondary structure.

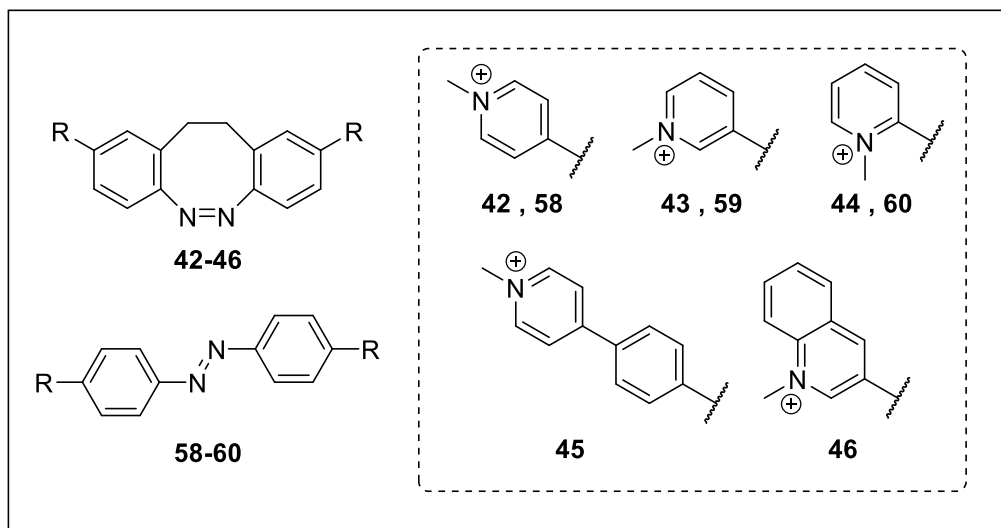
#### 3.1.1.2 FRET melting assay results

The 28 nucleotide oligomer sequence LTR-III (FLTR-IIIT) and a duplex DNA hairpin oligomer sequence (F10T) were used in the FRET melting assay, each labelled with a FAM and TAMRA



### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

fluorophore on the 5' and 3' ends respectively. It was also recognised that *E*-diazocines could not be used in the FRET melting assay as these metastable photoisomers would thermally revert back to the *Z*-diazocine during the course of the experiment. Thus, linear *E*-azobenzene mimics **58-60** were employed in the assay as a thermally stable *E*-“diazocine”. (**Figure 3.5**).



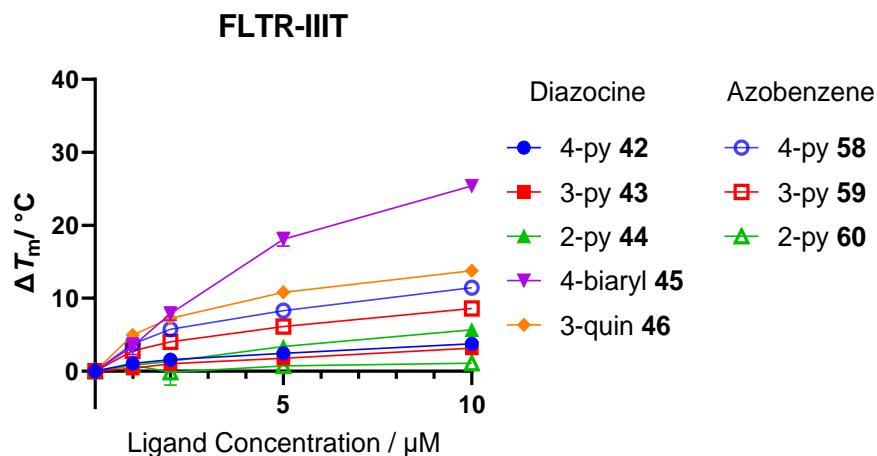
**Figure 3.5:** Ligands used in FRET melting assay. Acyclic azobenzene mimics **58-60** were required as a substitute for the *E*-diazocines. Azobenzene **58** was synthesised by Dr Ramos Soriano, azobenzene **59-60** was synthesized by M. Holbrow-Wilshaw.

It was expected that the  $\Delta T_m$  for *E*-azobenzenes **58-60** were to be higher than the bent cyclic *Z*-diazocines **42-46**, due to the hypothesis that more planar aromatic ligands have better binding interactions with the planar G-tetrads. The  $\Delta T_m$  results are summarised graphically in **Figure 3.6** where it can be seen that there is a large range of  $\Delta T_m$  values for FLTR-IIIIT after the addition of 10  $\mu\text{M}$  ligand (**Table 3.1**). Representative melting curves are shown in Appendix A3.

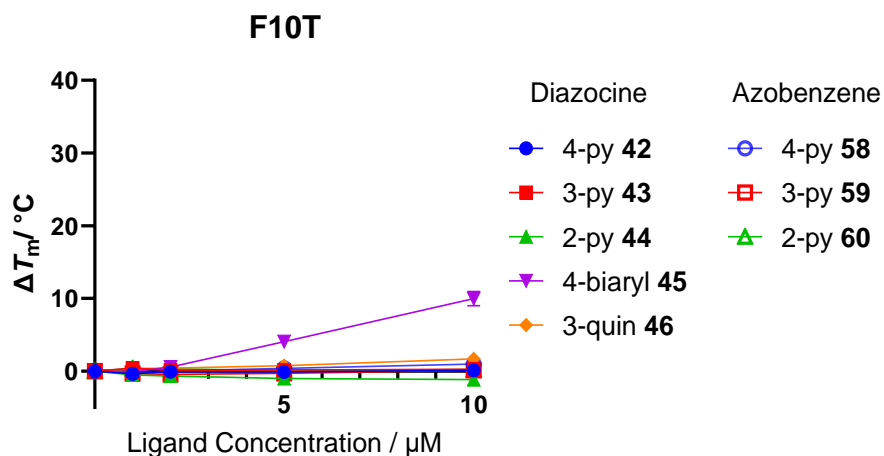
Despite poor expectations of the bent cyclic diazocines to provide any significant stabilisation of FLTR-IIIIT, 4-biaryl **45** and 3-quinolinium **46** exhibited high stabilisation of the FLTR-IIIIT sequence ( $\Delta T_m = 25$  and  $14$  °C at  $10$   $\mu\text{M}$ ). This is in contrast to the pyridinium diazocines **42-44** which had much lower stabilisation of the sequence ( $\Delta T_m = 3-6$  °C at  $10$   $\mu\text{M}$ ). Comparison of the azobenzene thermal stabilisation results against their diazocine analogues did not fully follow the expected trend. The 4- and 3-pyridinium azobenzenes **58** and **59** did show higher stabilisation ( $\Delta T_m = 11$  and  $9$  °C at  $10$   $\mu\text{M}$ ) than their diazocine counterparts ( $\Delta T_m = 4$  and  $3$  °C at  $10$   $\mu\text{M}$ ). But the 2-pyridinium ligands displayed opposite selectivity with the bent diazocine

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

**44** stabilisation FLTR-IIIIT more than the linear azobenzene **60**, albeit quite low in both cases, and whilst the explanation behind this is unclear, we should be cautious to overinterpret given the low  $T_m$  values and potential inaccuracy of the FRET assays as stated above.



**Figure 3.6:** Thermal stabilisation values ( $\Delta T_m$ ) for the FAM and TAMRA labelled FLTR-IIIIT oligonucleotide sequence (200 nM), with ligands **42-46** and **58-60** in 10 mM KCl, 90 mM LiCl and 10 mM Li cacodylate buffer, pH 7.2.



**Figure 3.7:** Thermal stabilisation values ( $\Delta T_m$ ) for the FAM and TAMRA labelled F10T oligonucleotide sequence (200 nM), with ligands **42-46** and **58-60** in 10 mM KCl, 90 mM LiCl and 10 mM Li cacodylate buffer, pH 7.2.

Though all pyridinium ligands (Entry 1-3, 6-8) had lower stabilisation temperatures than non-pyridinium ligands (Entry 4 and 5), it may be more beneficial that the substituent is not the key influencing factor towards G4 binding but rather the photoswitchable core. None of the pyridinium ligands showed higher stabilisation of the G-quadruplexes than 4-biaryl **45** and 3-quinolinium **46**, suggesting that the nature of the side-chain substituent was a significant

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

factor to the stabilisation of G-quadruplexes as Perez-Soto *et al.* have previously shown in other instances.<sup>154</sup>

**Table 3.1:** FRET thermal stabilisation values ( $\Delta T_m$ ) of FAM and TAMRA labelled G4 (FLTR-IIIIT) and duplex (F10T) sequences. DNA concentration = 200 nM, diazocine and azobenzene ligand concentration = 10  $\mu$ M.

		$\Delta T_m / ^\circ\text{C}$	
		FLTR-IIIIT	F10T
Entry	Ligand	10 $\mu$ M	10 $\mu$ M
1	Z-42	4 $\pm$ 0.1	0 $\pm$ 0.8
2	Z-43	3 $\pm$ 0.6	0 $\pm$ 0.6
3	Z-44	6 $\pm$ 0.2	-1 $\pm$ 0.4
4	Z-45	25 $\pm$ 0.4	10 $\pm$ 1.0
5	Z-46	14 $\pm$ 0.4	2 $\pm$ 0.4
6	E-58	11 $\pm$ 0.2	1 $\pm$ 0.2
7	E-59	9 $\pm$ 0.2	0 $\pm$ 0.1
8	E-60	1 $\pm$ 0.4	0 $\pm$ 0.5

We can see this clearly from the results of the melting assay with the duplex stem loop structure F10T (**Figure 3.7**). The outcomes from the melting assay of the negative control F10T sequence indicate that only Z-45 provides stabilisation (Entry 4,  $\Delta T_m = 10$   $^\circ\text{C}$ ) at ligand concentrations of 10  $\mu$ M. This behaviour is not seen in the remaining 7 ligands ( $\Delta T_m = -1$ -2  $^\circ\text{C}$ ). Where it becomes clear that the 4-biaryl pyridinium substituent is likely a general good binder to DNA structures, resulting in poor selectivity between G-quadruplexes and duplex DNA.

High stabilisation of F10T is not observed with the quinolinium substituent of Z-46, which only provided a 2  $^\circ\text{C}$  stabilisation of the duplex structure (Entry 5). This is promising as many prominent G4 ligands such as pyridostatin and PhenDC3 contain a quinoline group, further demonstrating that quinolinium G4 ligands have good selectivity between duplex and quadruplex structures.

The overall results from the FRET melting assays were encouraging. We could see that the biaryl-pyridinium substituent in 45 seemed to stabilise any DNA structure, regardless of

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

topology. There were also differences in the *E* and *Z* species for the 4- and 3-pyridinium ligands, where the more planar azobenzene provided increased stabilisation. Owing to the necessary use of the linear *E*-azobenzene mimics for the thermal melting experiments and the use of labelled oligonucleotides, there is a level of caution necessary in the interpretation of the results. It is unclear if *E*-azobenzenes accurately portray how well *E*-diazocines can bind to G4s. It is also possible that the *E*-isomers of **45** and **46** could have even higher binding affinities upon isomerisation despite the high stabilisation already brought about by their larger substituents in comparison to the pyridinium ligands. Additional biophysical experiments such as UV-Visible and CD spectroscopy titrations can be run with the *E*- and *Z*-isomers of diazocines as they are performed at room temperature and may provide additional data to support the results from the FRET melting assays.

#### 3.1.2 UV-Visible absorbance titrations

##### 3.1.2.1 Introduction to UV-Vis titration

UV-Vis titrations allow for an understanding of the ligand binding association constants and also provides information of potential binding modes of the ligands to DNA. The titration of ligand (*L*) with DNA (*D*), follows the equilibrium:



This can be rearranged to define the association constant ( $K_a$ ):

$$K_a = \frac{[D.L]}{[D] \cdot [L]} \quad (3.2)$$

Additionally, assuming low concentrations where the Beer-Lambert law is valid, the absorption spectra can be defined as the sum of absorption of the DNA, Ligand and DNA-Ligand complex:

$$A = \varepsilon_D[D] + \varepsilon_L[L] + \varepsilon_{\Delta D.L}[D.L] \quad (3.3)$$

DNA absorption has a maxima at 260 nm which decreases to zero at 320 nm. Therefore measurements beyond 320 nm can only be attributed to the ligand (*L*) or the DNA-Ligand complex (*D.L*), allowing for the simplification of eqn 3.3 to:

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

$$A = \varepsilon_L[L] + \varepsilon_{\Delta D.L}[D.L] \quad (3.4)$$

We can therefore solve the change in absorbance ( $\Delta A$ ) with the equation 3.5 with  $[D.L]$  derived from the rearrangement of equation 3.2. This allows us to calculate the association constant  $K_a$  in terms of total DNA and total ligand concentrations ( $[D]_0$  and  $[L]_0$ ).<sup>155</sup>

$$\Delta A = \varepsilon_{\Delta D.L} \cdot [D.L] \quad (3.5)$$

$$[D.L] = \frac{\left(N \cdot [D]_0 + [L]_0 + \frac{1}{K_a}\right) - \sqrt{\left(N \cdot [D]_0 + [L]_0 + \frac{1}{K_a}\right)^2 - 4 \cdot N \cdot [D]_0 [L]_0}}{2} \quad (3.6)$$

In particular, the equation uses a multiple independent binding sites (MIS) model that allows for the assumption that the binding of multiple ligands are independent and do not act cooperatively.<sup>155,156</sup>

As mentioned previously, aside from calculating association constants, UV-Vis titrations allow for the understanding of ligand binding modes due to changes in the absorption spectrum of the ligand and DNA-ligand complex. Bathochromic (red) shifts typically occur due to the decrease in the  $\pi$ - $\pi^*$  electronic transition of the ligand. This energy gap is lowered due to the  $\pi$  bonding orbitals of the DNA bases stacking with the empty  $\pi^*$  antibonding orbital of the ligand.<sup>157</sup> Hypochromic and hyperchromic shifts may also occur because of changes to the electronics of the ligand.

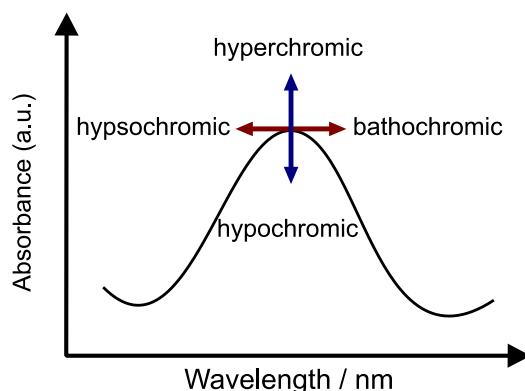
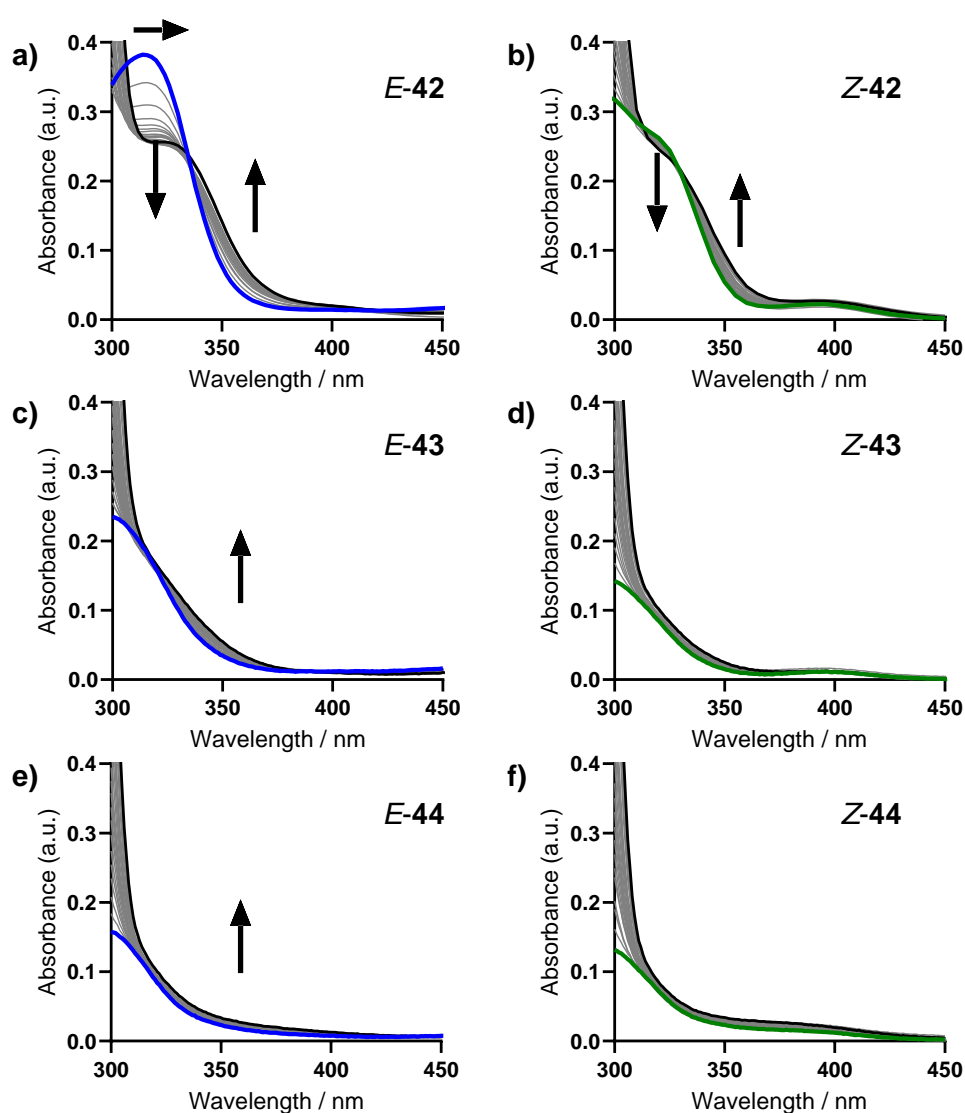


Figure 3.8: Possible spectral shifts of an absorption band.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

#### 3.1.2.2 UV-Vis titrations with LTR-III

In the titrations of **42-46** with the LTR-III sequence, ligand concentration is held constant at 10  $\mu\text{M}$  with increasing additions of DNA sequences. Titrations with *E*-diazocines were carried out by continuous irradiation at 405 nm, to be able to measure an *E*-rich titration. Only *E*-**42** showed hypochromic and bathochromic shifts upon ligand binding, suggesting end-stacking binding modes<sup>158</sup> (Figure 3.9a). Unfortunately, this was not visible for the other two pyridinium ligands as their blue shifted  $\pi$ - $\pi^*$  bands overlap significantly with the absorption of the DNA (Figure 3.9c-f).



**Figure 3.9:** UV-Vis titration studies of 10  $\mu\text{M}$  ligands **42-44** with increasing concentrations (0-31  $\mu\text{M}$ ) of LTR-III in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. **Left:** *E*-**42-44** and **Right:** *Z*-**42-44**. Blue or green line represents 10  $\mu\text{M}$  ligand with no DNA added, black line represents addition of 3 equivalents of DNA.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

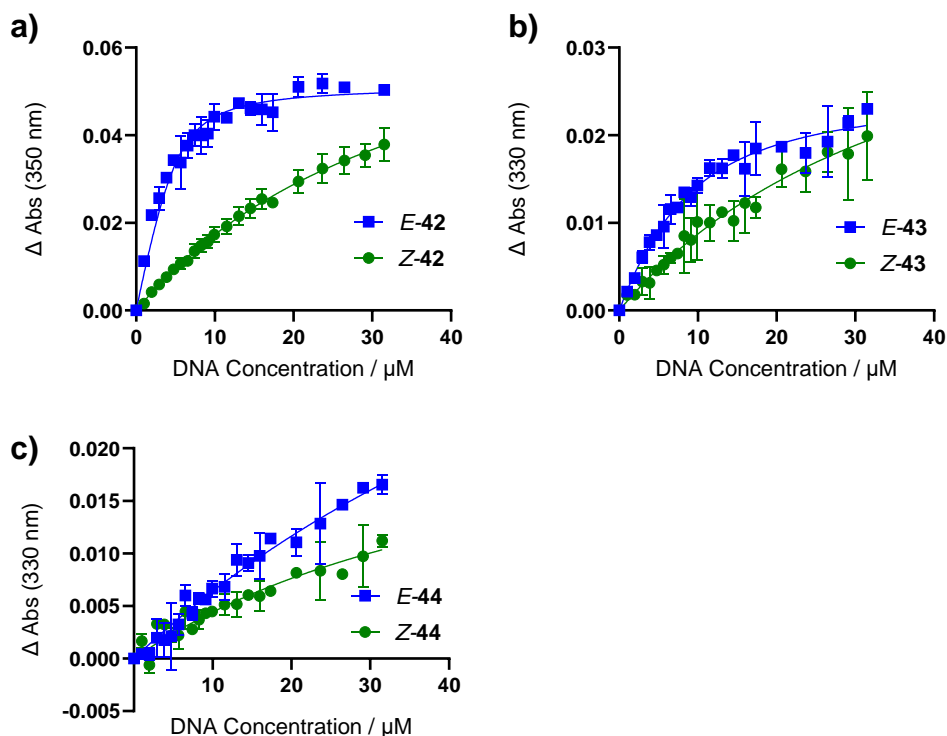


Figure 3.10: Binding isotherms for **42-44** with LTR-III from UV-Vis titration studies.

Table 3.2: Association constants and *E*-/*Z*- ratios for pyridinium diazocines **42-44** and pyridinium azobenzenes **58** and **59** calculated from the binding isotherms with LTR-III, *N*=2.

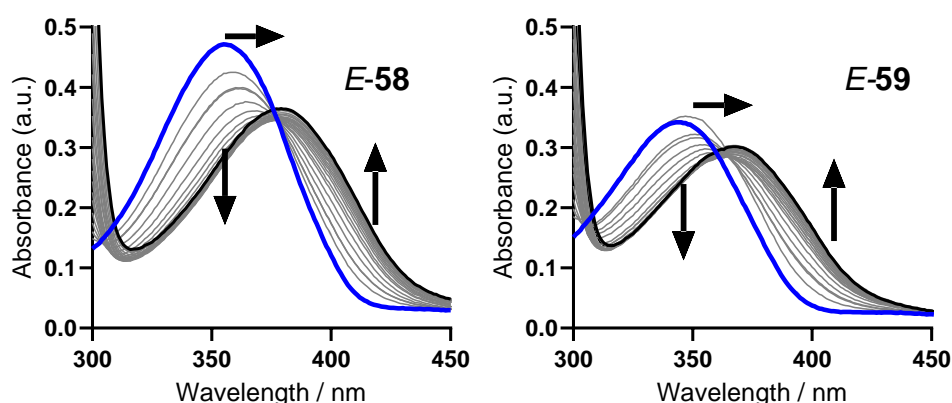
Ligand	$K_a / 10^6 \text{ M}^{-1}$			$\lambda / \text{nm}$
	<i>E</i> -	<i>Z</i> -	<i>E</i> -/ <i>Z</i> -	
<b>42</b>	$0.54 \pm 0.10$	$0.016 \pm 0.002$	34	350
<b>43</b>	$0.09 \pm 0.01$	$0.015 \pm 0.005$	5.8	330
<b>44</b>	$0.006 \pm 0.003$	$0.011 \pm 0.005$	0.55	330
<b>58</b>	$0.69 \pm 0.10$	-	-	400
<b>59</b>	$0.82 \pm 0.10$	-	-	390

Binding isotherms were selected at wavelengths with the largest changes in absorption for both isomers (>320 nm). By setting *N*=2, we can see that the association constant decreases across the pyridinium series (Table 3.2). Additionally, the results show that association constants follow the same trend as the results from the FRET melting assay with FLTR-IIIIT. The *E*/*Z* ratio of **42** has an impressive 34 fold difference and **43** still has a good 5.8 fold difference between its two photoisomers. With the more linear *E*-diazocine binding stronger to LTR-III.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

The opposite is true for **44**, where Z-**44** has a 2-fold larger  $K_a$  than E-**44**. However, the association constants for **44** are accompanied by large errors, indicating generally poor binding of **44** to LTR-III.

Differences in the  $K_a$  of E- and Z-diazocines across the pyridinium series is not necessarily strictly due to higher binding affinities of the E-ligand. There is a possibility that different PSS for each E-isomer results in different effective concentrations of the “active” species. E-diazocines with higher PSS can be expected to have a higher apparent  $K_a$ . We can illustrate this with the linear azobenzenes **58** and **59**, which unlike their cyclic E-diazocines analogues, 100% of **58** and **59** will be as the E-isomer. Binding isotherms of the two azobenzenes gives comparable association constants **58** ( $K_a = 0.69 \times 10^6 \text{ M}^{-1}$ ) and **59** ( $K_a = 0.82 \times 10^6 \text{ M}^{-1}$ ), as opposed to the 6-fold difference between E-diazocines **42** and **43** (Table 3.2). This could suggest that the apparent  $K_a$  may be influenced by the actual population of E-diazocine in solution.



**Figure 3.11:** UV-Vis titrations of 10  $\mu\text{M}$  E-azobenzenes **58** and **59** with increasing concentrations (0-31  $\mu\text{M}$ ) of LTR-III in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. Blue line represents 10  $\mu\text{M}$  ligand with no DNA added, black line represents addition of 3 equivalents of DNA.

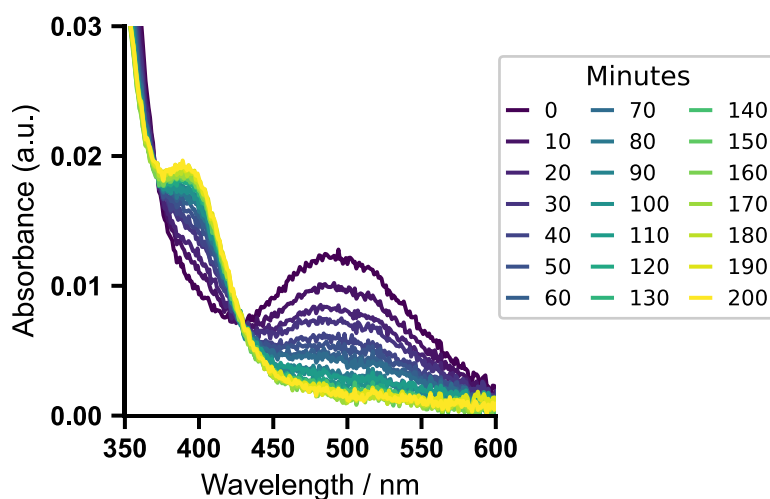
The  $\pi$ - $\pi^*$  band of azobenzenes is also more red shifted than the E-diazocines, free from overlapping with DNA absorption. This allows for the observation of the hypochromic and bathochromic shifts of the  $\pi$ - $\pi^*$  bands. **58**, has a 23 nm bathochromic shift with a 27% hypochromic shift followed by a slight hyperchromic shift. **59** shows a comparable 25 nm bathochromic shift but a smaller hypochromic shift of 17% also followed by a slight hyperchromic shift. These suggest that at least in the case of E-azobenzenes, the ligand



### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

binding modes could also be end-stacking to guanine tetrads, though more studies may be required to confirm this.

Investigations into the kinetics of *E*-diazocine ligands have not been studied in detail, but there seems to be evidence that in these bound states, the half-life of the *E*-isomers is shortened. **Figure 3.12** shows an example of the thermal isomerisation of *E*-**43** after titration. Without the presence of G-quadruplex DNA, *E*-**43** has  $t_{1/2}$  of 200 min which decreases to 60 minutes when bound to DNA at a ratio of 1:3, (Ligand:DNA). This could be due to molecular crowding or the isomerisation be catalysed by the G4 structure. This effect is not surprising, but would require a more systematic investigation.<sup>128</sup>

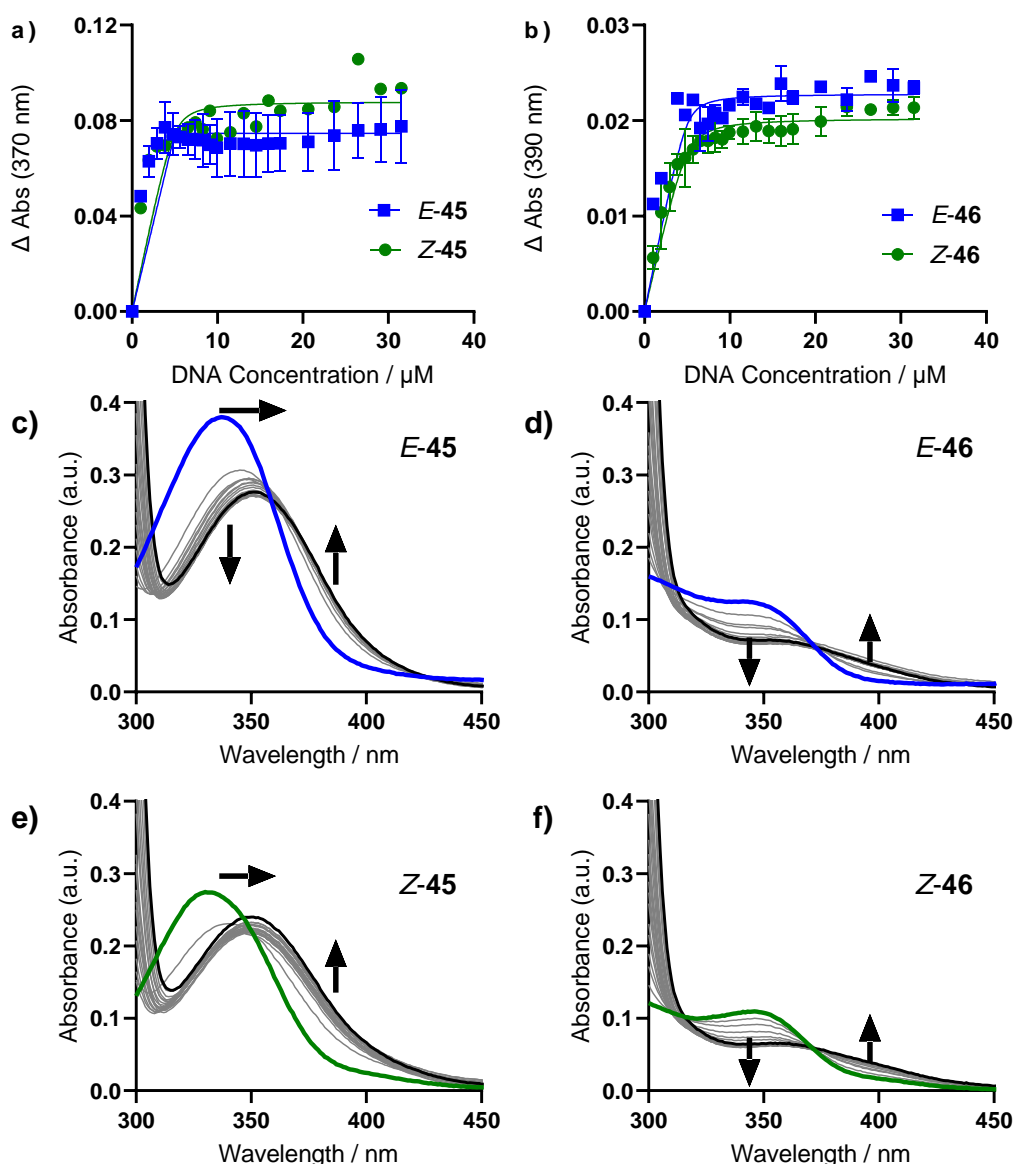


**Figure 3.12:** Thermal isomerisation of *E*-**43** after the addition of 3 equivalents of LTR-III at 20 °C. On the other hand, the non-pyridinium ligands **45** and **46** provided high thermal stabilisation of the DNA sequences by the *Z*-species in the FRET melting assay. As there were no linear *E*-azobenzene analogues, it was of interest to see if isomerising the core diazocine structure would provide any differences in binding affinities afforded by the change in structure beyond the highlighted properties of the larger substituents.

Titration of 4-biaryl **45** and 3-quinolinium **46** with LTR-III (**Figure 3.13**) both showed strong binding isotherms with quick saturation of binding sites, which was expected given the FRET melting assay results. The similarities between the *E*- and *Z*- binding isotherms of **45** (**Figure 3.13a**) provided evidence that the 4-biaryl pyridinium substituent provides more significant binding interactions than any differences in the *E*/*Z*- structural isomerisation of the diazocine core. The isotherms poorly fit the curve because the initial concentration of the ligand is too

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

high relative to the association constant, in what is considered the titration regime.<sup>159</sup> The addition of any DNA is immediately depleted to bind with the ligand until there is no more ligand remaining. Overall, this gives an underestimation of the association constant which cannot be calculated. For a more accurate estimation of the association constant, would require repeating the titrations at a lower initial ligand concentration.



**Figure 3.13:** UV-Vis titration binding isotherms of **45** and **46** with LTR-III in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. **a-b)** Extracted binding isotherms of **a)** **45** and **b)** **46**. **c-f)** Titrations of 10  $\mu$ M **c),e)** **45** and **d),f)** **46** with LTR-III G4. Blue or green line represents 10  $\mu$ M ligand with no DNA added, black line represents addition of 3 equivalents of DNA.

On the other hand, quinolinium **46** shows better fitted isotherms and a 2-fold difference between the estimated association constants from the *E* and *Z* titrations ( $K_a = 4.2 \times 10^6 \text{ M}^{-1}$  and

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

$2.3 \times 10^6 \text{ M}^{-1}$ ) (Figure 3.13b). These are larger association constants than the pyridinium ligands, but result in smaller  $E/Z$  ratios compared to the pyridinium ligands **42** and **43**.

#### 3.1.3 Circular dichroism titrations

Circular dichroism (CD) is a spectroscopic technique that can be employed to study the secondary structures of biomolecules such as proteins and nucleic acids.<sup>72,160–162</sup> CD can be employed to explore how ligand binding may affect G4 secondary structure.

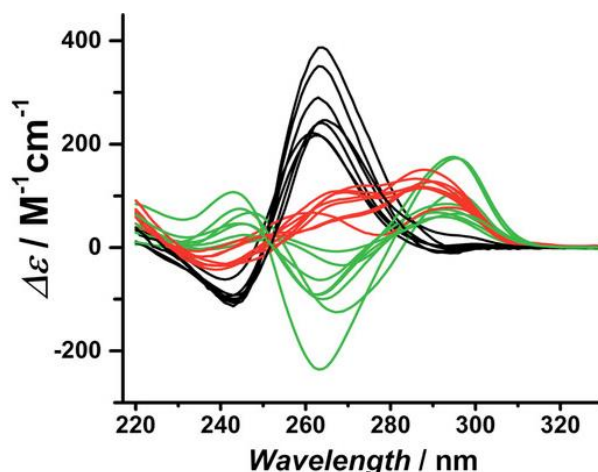
Each topology has a characteristic structure due to the chiral environment of stacked guanine bases causing exciton coupling between transition dipole moments. Upon addition of ligand, any changes to the G4 topology caused by ligand binding will be reflected in the CD spectrum. As all the ligands are achiral, the appearance of absorption bands beyond 320 nm suggests induced circular dichroism (ICD) of the ligands, caused by the interaction of achiral ligands to the chiral DNA, either by structurally induced circular dichroism of the ligand itself or by excitonic coupling with the DNA bases.<sup>163,164</sup> If ICD bands become available, it can also be useful to monitor if ligands can photoswitch in the presence of the DNA whilst also causing changes in topology.

Circularly polarised light occurs when 2 perpendicular planes of light are 90° out of phase. This results in either left circularly polarised light or right circularly polarised light. Achiral molecules absorb both left and right circularly polarised light equally, so the resulting CD spectrum is zero. However, chiral molecules absorb polarised light differently, resulting in a positive or negative CD spectrum.

$$\Delta A = A_L - A_R \quad (3.7)$$

The recorded signal is measured as the observed ellipticity, ( $\theta$ ). Molar ellipticity,  $[\theta]$ , can then be calculated by converting observed ellipticity with equation 3.8, where  $c$  is the concentration of the DNA and  $l$  is the pathlength of the sample:

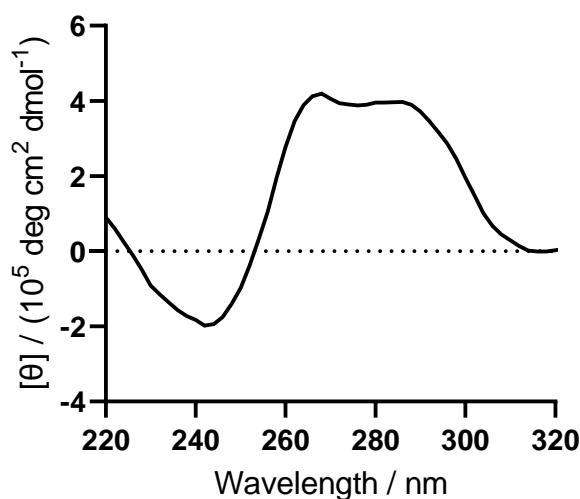
$$[\theta] = \frac{100 \cdot \theta}{c \cdot l} \quad (3.8)$$



**Figure 3.14:** Characteristic G4 CD spectra, parallel G4 (black), hybrid (red) and antiparallel (green). Figure used with permission, © 2018 Wiley.<sup>160</sup>

**Figure 3.14** shows the characteristic CD spectra of different G4 structures. Parallel G4s (black) have a positive band at ~260 nm and a negative band ~240 nm. Antiparallel G4s (green) have a positive band at 295 nm and a negative band at ~260 nm. Finally hybrid G4 structures (red) show positive bands at 295 and ~260 nm with a negative band at 240 nm.<sup>72,160</sup>

#### 3.1.3.1 LTR-III CD titrations



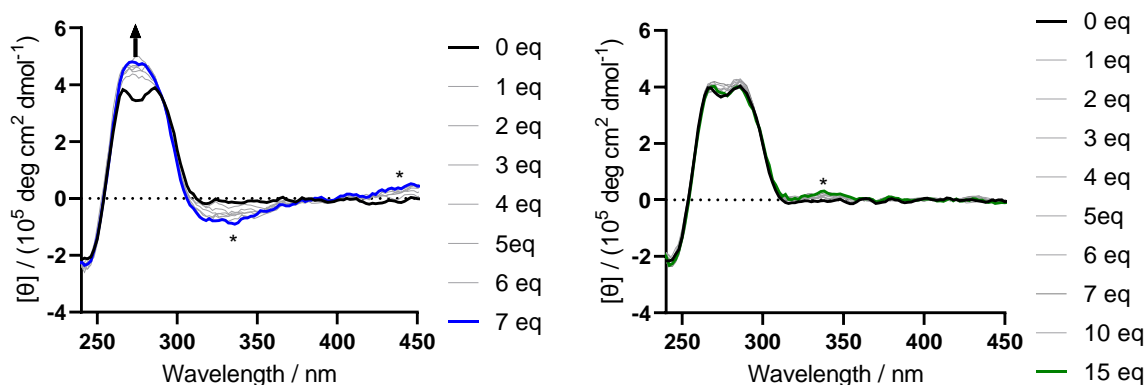
**Figure 3.15:** CD spectrum of a 5  $\mu\text{M}$  HIV-1 LTR-III sequence at 20  $^{\circ}\text{C}$  in a 20 mM K-phosphate and 70 mM KCl buffer, pH 7.

The CD spectrum of LTR-III shows a large negative band at 245 nm, a positive band at 265 nm with a shoulder 285 nm, demonstrating a nonparallel G4 topology, which is concordant with the eventual elucidated structure. Like most other DNA structures, there is no absorption of light beyond 320 nm. Apart from 2-pyridinium diazocine **44** that demonstrated poor binding

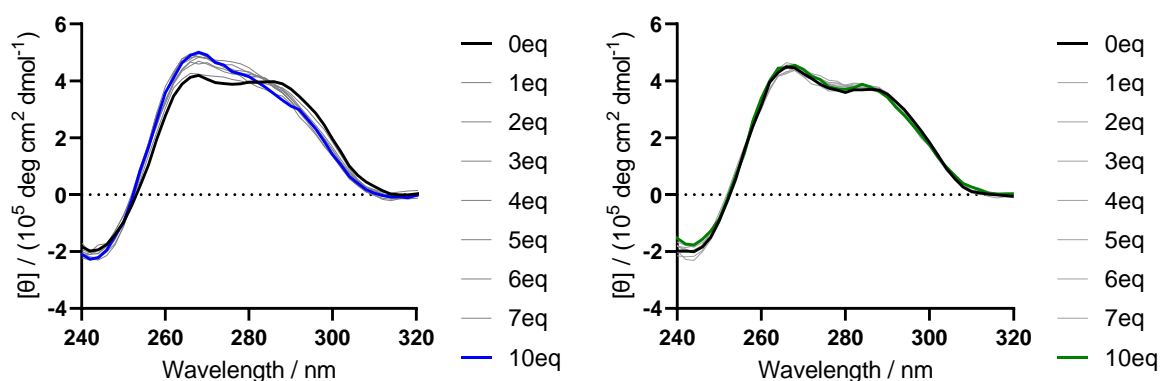
### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

through the FRET melting assay, all remaining diazocine ligands were employed in CD titration experiments with LTR-III at 5  $\mu\text{M}$  concentration.

Upon the addition of **42** to LTR-III, *E*-**42** causes a hyperchromic shift at 275 nm between the two positive bands in LTR-III CD spectrum, typically indicative of stabilisation of the DNA structure (**Figure 3.16**). There are also two induced circular dichroism bands, a negative band at 340 nm and a wide positive band between 400-600 nm. Both of these correspond to the maxima of the absorption bands that can be seen in UV-Vis titration experiments and confirms that the ligand is interacting with the chiral DNA environment. There is less of an effect of *Z*-**42** to LTR-III topology, with only a small induced CD band at 340 nm which also matches the UV-Vis absorption spectrum for *Z*-**42**.



**Figure 3.16:** CD titration of **42** with LTR-III at 20 °C in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. **Left:** *E*-, **Right:** *Z*-. Induced CD marked with \*.

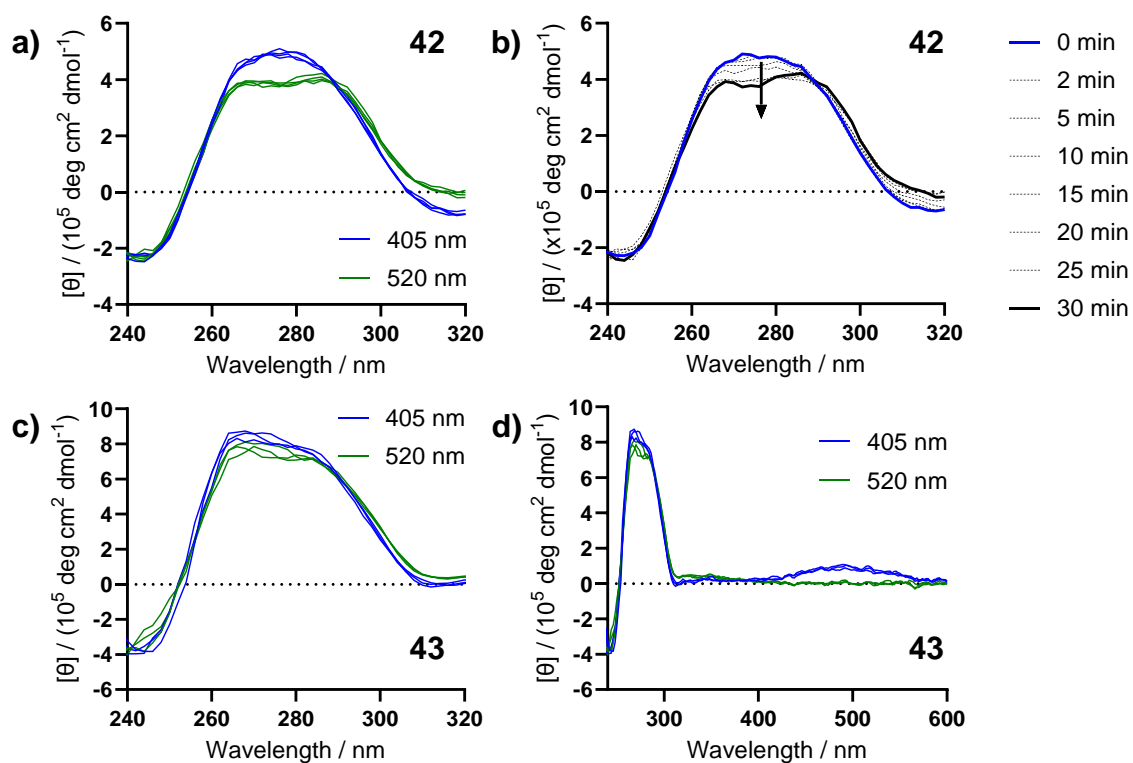


**Figure 3.17:** CD titration of **43** with LTR-III at 20 °C in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. **Left:** *E*-, **Right:** *Z*-.

A similar effect is seen with the addition of **43** to LTR-III (**Figure 3.17**). The addition of *E*-**43** perturbs the spectra of the LTR-III sequence towards a more parallel topology, seen by the

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

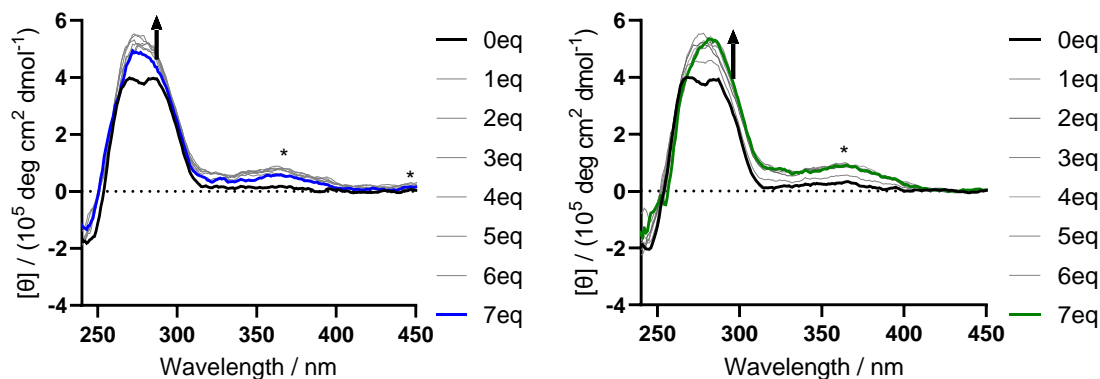
increase in the positive band at 270 nm with a slight decrease in the band at 295 nm. Whereas no change is observed with **Z-43**. Overall, the CD titrations of the two pyridinium ligands compliment previous FRET melting assay and UV-Vis titration studies. There are larger changes to the CD spectrum upon the addition of *E*-ligands with this effect maximised with 4-pyridinium **42** which had the highest association constant as calculated from UV-Vis.



**Figure 3.18:** a) Photoswitching **42** in the presence of LTR-III at 20 °C in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. b) Thermal relaxation of **42** in the presence of LTR-III c-d) Photoswitching **43** in the presence of LTR-III at a higher DNA concentration, c) DNA region between 240-320 nm. d) Full range between 240 nm and 600 nm.

Irradiation of both 4-pyridinium **42** and 3-pyridinium **43** diazocines in the presence of DNA show that the ligands can be reversibly change LTR-III topology just by the irradiation of blue and green light. In fact, as *E-42* is the ligand with the shortest thermal half-life, it is possible to monitor the thermal relaxation of the photoswitch over a 30 minute period. Photoswitching **43** at slightly higher concentrations of DNA (**Figure 3.18c-d**) achieves the same effect. Ligand isomerisation is fully reversible as shown by the ICD regions at the further red shifted wavelengths.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

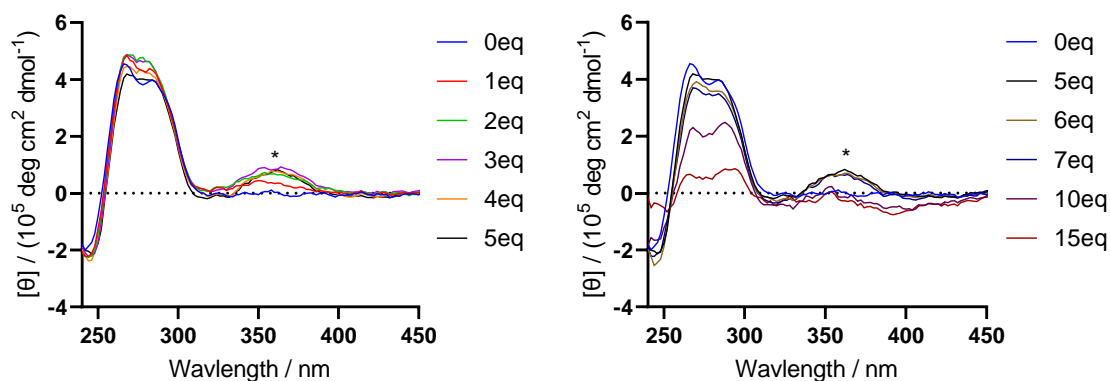


**Figure 3.19:** CD titration of **46** with LTR-III at 20 °C in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. **Left:** *E*-, **Right:** *Z*-. Induced CD marked with \*.

Titration of the quinolinium **46** ligands show similar results to the pyridinium ligands (**Figure 3.19**). The higher affinity of **46** to LTR-III is demonstrated in that both photoisomers cause a hyperchromic shift the DNA region of the CD spectra. It is also possible to see induced CD bands of the ligands, both isomers show a positive ICD band between 320-400 nm and in the titration of *E*-**46** has another positive ICD band between 400-600 nm. Slight differences between the two titrations can be observed after the addition of 1 equivalent of ligand. *E*-**46** already reaches saturation after 1 equivalent of ligand, whereas this is delayed for *Z*-**46**, demonstrating *E/Z* selectivity to LTR-III that was shown in UV-Vis titrations.

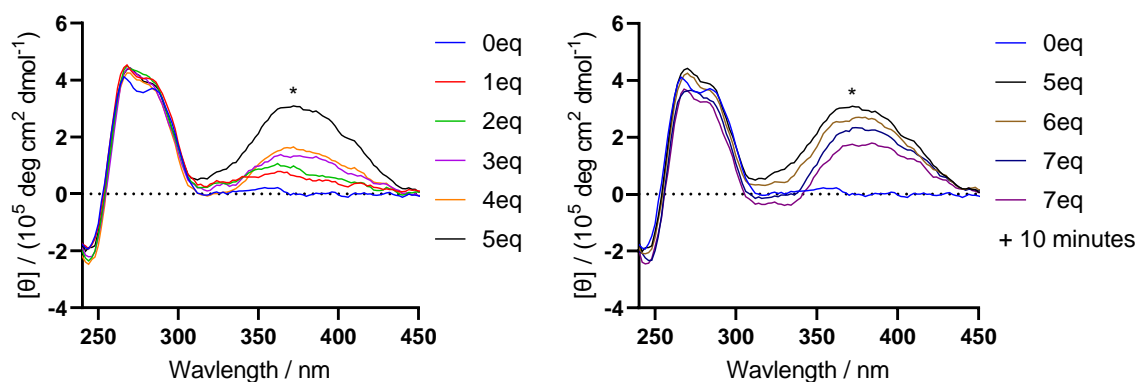
In the titration of biaryl *E*-**45**, we initially see an increase to the CD signal of the DNA region along with an induced CD band showing ligand interactions (**Figure 3.20**). After the addition of 7 equivalents of ligand, the CD spectrum of the DNA deteriorates rapidly. Observation of the sample shows the formation of a pink precipitate, which after an hour under visible light turns yellow, suggesting that it has returned to *Z*-**45**. This precipitate is the most likely reason for the large drop in CD signal, as the DNA is removed from solution this results in an effective drop in actual DNA concentration in solution at high equivalents of ligand. The lack of an induced CD band between 400-600 nm suggests that unlike the other *E*-diazocines, the planar diazocine chromophore is not interacting with the DNA. These results might imply that it has a different mode of interaction with LTR-III than the other *E*-diazocines.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex



**Figure 3.20:** CD titration of *E-45* with LTR-III at 20 °C in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. Deterioration of the DNA CD band occurs after the addition of 7 equivalents of ligand through precipitation of DNA.

As for the *Z*-photoisomer, at lower ligand equivalents of *Z-45*, there is the same hyperchromic shift in the DNA region of LTR-III, like the titration of *E-45* (Figure 3.21). However, behaviour in the induced CD region is particularly interesting and not observed for other *Z*-diazocine ligands. After the addition of 2 equivalents of ligand, there is a small positive ICD band from 315-450 nm. This quickly shifted to larger peak after the addition of 5 equivalents of ligand, suggesting the formation of a new binding mode for multiple ligands. The particularly large ICD band is strongly indicative for groove binding of the ligand, where there could be a combination of ligand-DNA exciton coupling and ligand-ligand exciton coupling.<sup>162–164</sup> Beyond 5 equivalents of ligand, addition of more ligand caused the CD signal to decrease. This decrease is particularly significant 10 minutes after the addition 7 ligand equivalents, as it potentially signifies the destabilisation of the G-quadruplex structure upon high equivalents of ligand binding after the aggregation of ligand.

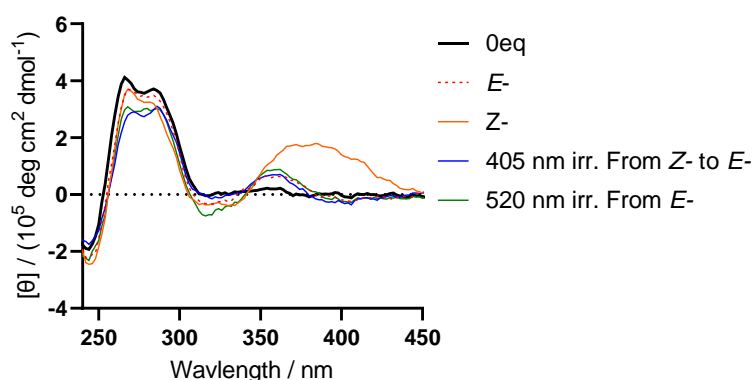


**Figure 3.21:** CD titration of *Z-45* with LTR-III at 20 °C in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. Deterioration of the DNA CD band occurs after 5 equivalents of ligand. Large ICD band of ligand also occurs at high ligand equivalents.



### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

Photoswitching of 4-biaryl **45** after titration into LTR-III was also attempted and gave interesting results (Figure 3.22). After irradiation with 405 nm light, it was possible to photoswitch the ICD band of the Z- titration (orange->blue) to match the ICD band of the equivalent E-titration (red). However, the reverse photoswitching from E was not possible (green). Afterwards, the DNA-ligand complex precipitated again suggesting that the irreversibility of the E to Z isomerisation could be due to the strong affinity of the E-ligand that disfavours the dissociation of the ligand.



**Figure 3.22:** Photoswitching studies of **45** with LTR-III at 20 °C in a 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7. LTR-III + 7 equivalents of Z-**45** (orange), irradiation of orange with 405 nm light (blue), irradiation of blue 520 nm light (green). Black, LTR-III without ligand. Red, reference spectra of LTR-III with E-**45**.

Overall, CD spectroscopy provided useful information that the two pyridinium ligands **42** and **43** do show different binding modes between their E- and Z-isomers. While induced CD signals can serve to supply useful information about ligand binding modes, this method is less well understood for G-quadruplexes than for duplex DNA. Large positive ICD bands often suggest ligand groove binding modes. On the other hand, negative or weak positive ICD bands can suggest stacking interactions.<sup>163</sup> Thus, it is unclear whether the binding modes of the pyridinium E-**42** and E-**43** show groove or stacking binding modes just based on CD data alone. Often any ICD bands can be a combination of multiple factors including the ligand's own CD signal due to the chiral DNA environment and from excitonic coupling between DNA-Ligand or Ligand-Ligand interactions.<sup>164</sup> Quantum mechanical calculations can be used to understand the final resolved ICD bands and provide binding mode information, but this not an efficient approach. So alternative methods are required to further understand where the ligands bind to the DNA structure.

#### 3.1.4 NMR titrations

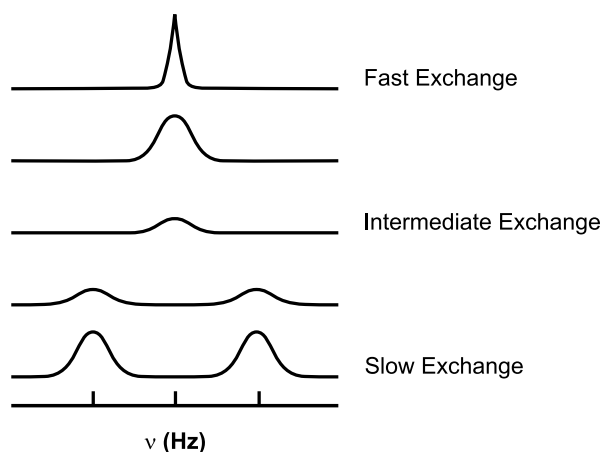
As LTR-III has a well-defined NMR solution structure, NMR spectroscopy is a helpful tool to define the structural interactions of the ligand with the G4, allowing determination of the binding modes of the ligand at an atomic level. The NMR solution structure of LTR-III has been solved by the groups of Richter and Phan, where it showed a relatively unique feature quadruplex-duplex junction.<sup>57</sup>

The <sup>1</sup>H NMR of Hoogsteen hydrogen bonded imino protons (**Figure 3.24a**) of G-tetrads occur quite downfield between the regions of 10-12 ppm and the resonances of Watson-Crick hydrogen bonded imino protons occur even further downfield between 12-14 ppm.<sup>165,166</sup> This is helpful as this region can be clearly identified without overlapping with peaks from the ligand or other aromatic resonances. Upon ligand binding, perturbations of specific guanine residues may signify the association of the ligand at that site, providing useful structural information in the determination of ligand binding modes.

Perturbations into lineshapes during titrations can show slow, intermediate or fast exchange binding regimes due to the differences between the rate of exchange ( $k_{ex}$ ) and the change in chemical shift ( $\Delta\nu$ ) (**Figure 3.23**).<sup>167-170</sup> In slow exchange binding regimes, ligand binding is strong with protons experiencing slow rates of exchange ( $k_{ex} \ll |\Delta\nu|$ ). At sub-stoichiometric quantities of ligand, two sets of peaks may appear for the same proton, one bound and one unbound to ligand. For intermediate exchange regimes ( $k_{ex} \sim |\Delta\nu|$ ), the averaging of the bound and unbound states will cause broadening of a signal where it may not be visible. Finally, in fast exchange regimes, it is expected that ligand binding is weak with the proton able to switch quickly between bound and unbound states ( $k_{ex} \gg |\Delta\nu|$ ). The time averaged chemical shift will appear as a well-resolved sharp peak, where signals shift towards the chemical shift of the bound state upon further ligand addition.

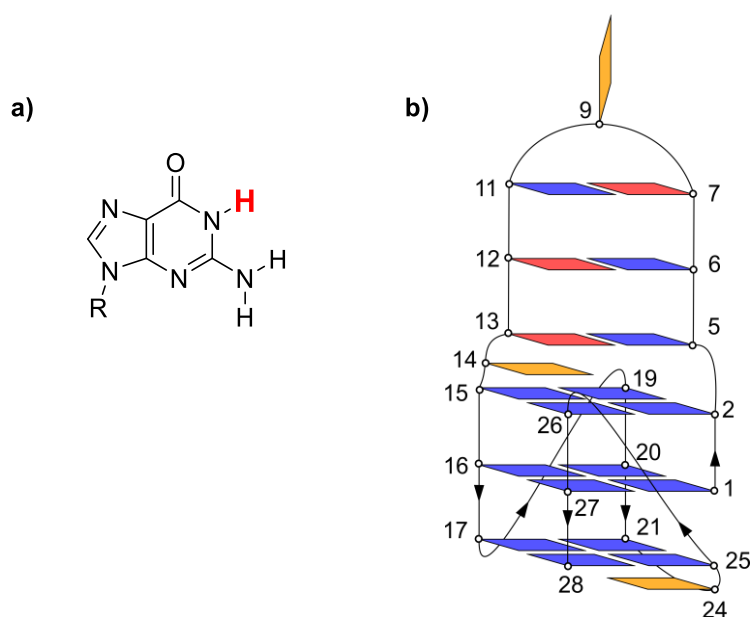
We decided to investigate how the binding of the two pyridinium diazocines **42** and **43** affected the LTR-III sequence by NMR spectroscopy. As both of these two ligands seemed to have the larger differences between *E/Z* unlike quinolinium **46** and biaryl pyridinium **45**.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex



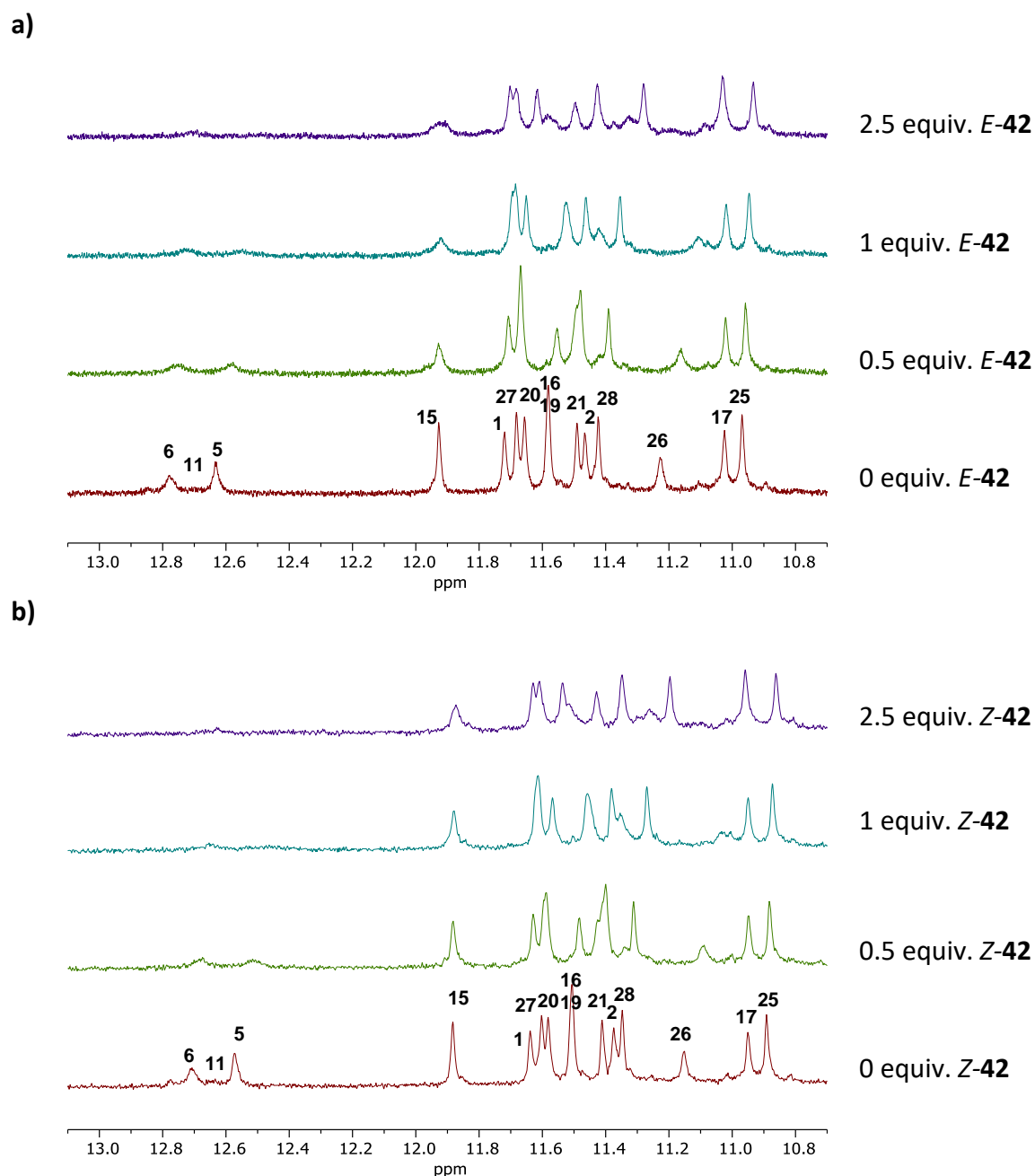
**Figure 3.23:** Illustration of fast, intermediate and slow exchange regimes for a proton in bound and unbound states.

In the titration of LTR-III with **42**, there are some differences between the two isomers (**Figure 3.25**). The endpoint of both titrations show that most of the G4 imino residues remain intact. This suggests that the overall G4 structure is retained, as is also demonstrated in the CD spectroscopy experiments. However, the biggest issue with the titration study of **42** was that this ligand has a relatively fast thermal relaxation. So a significant amount of *E*-**42** isomerises back to *Z*-**42** during measurement at room temperature, resulting in the wrong observations for the *E* titration, where the differences in the two spectra are smaller than expected. One method of dealing with this issue is by lowering the temperature of the system to 5 °C. This slows the rate of thermal isomerisation allowing for a better comparison of *E* and *Z* isomers.



**Figure 3.24:** **a)** Imino proton (red) of guanine. **b)** Cartoon of LTR-III structure. Guanine (blue), Cytosine (red), Thymine (yellow).

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex



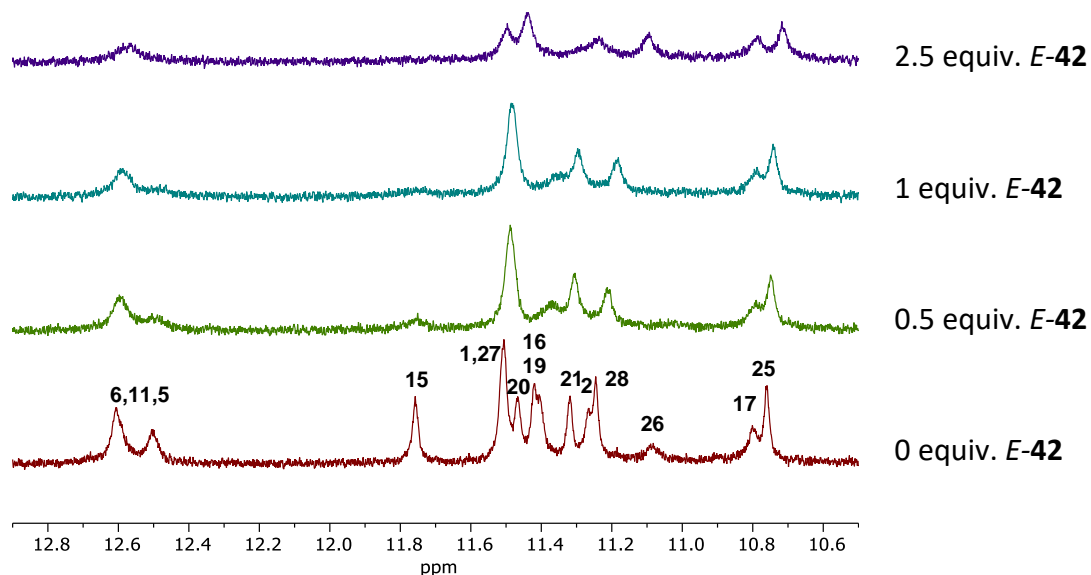
**Figure 3.25:** Imino  $^1\text{H}$  NMR titration studies of LTR-III G4 at 25 °C in a 20 mM potassium phosphate buffer (pH 7) containing 70 mM potassium chloride and 10%  $\text{D}_2\text{O}$  with **a)** *E*-42 and **b)** *Z*-42. Initial DNA concentration = 186  $\mu\text{M}$ . Imino protons of each guanine base is labelled. Titrations of 42 with LTR-III were done by Dr J Ramos Soriano.

Upon cooling to 5 °C, the initial chemical shifts for the imino residues shift upfield and slightly broaden, leading to a slight overlapping of chemical shifts for some of the residues between 11.2-11.6 ppm. However, as the general trend of chemical shifts is retained, the assignment of peaks is still straightforward (**Figure 3.26**). At the lower temperature, differences between *E*- and *Z*- titrations are much clearer. Peaks in the titration of *E*-42 are broader than in the titration of *Z*-42. This is the most clear with residue G15, a guanine in the top tetrad at the

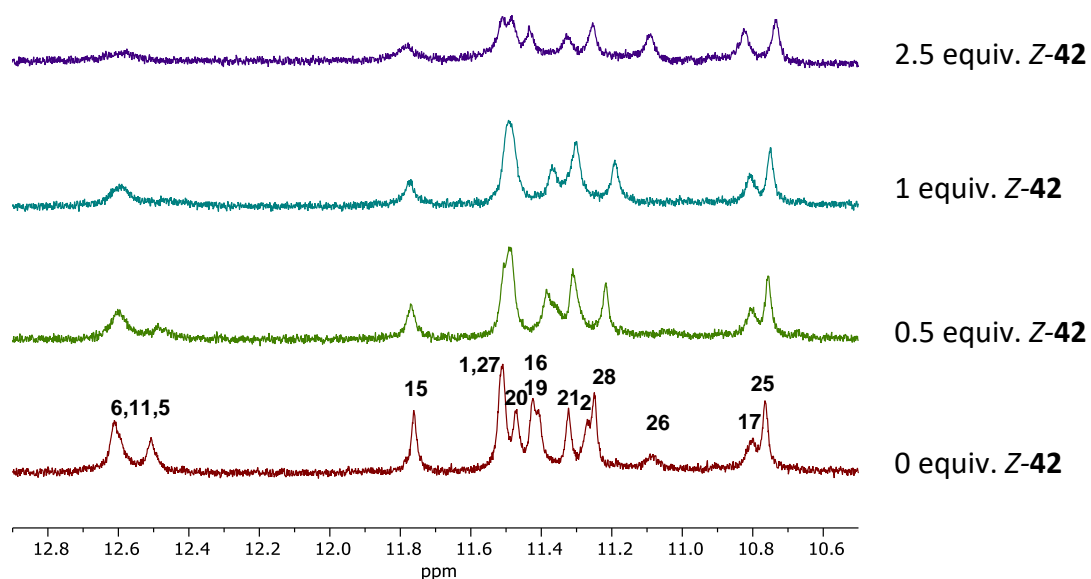
### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

quadruplex-duplex junction (**Figure 3.24**). G15 disappears after the addition of 2.5 equivalents of *E-42*, but remains as a broad peak after 2.5 equivalents of *Z-42* (**Figure 3.26**).

a)



b)



**Figure 3.26:** Imino <sup>1</sup>H NMR titration studies of LTR-III G4 at 5 °C in a 20 mM potassium phosphate buffer (pH 7) containing 70 mM potassium chloride and 10% D<sub>2</sub>O with **a)** *E-42* and **b)** *Z-42*. Initial DNA concentration = 186 μM. Imino protons of each guanine base is labelled. Titrations of **42** with LTR-III were done by Dr J Ramos Soriano.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

The reason for this disappearance could be attributed to the intermediate exchange regime, where the rate of ligand exchange between the bound and unbound conformations occur at NMR timescales resulting in a broadened peak that can disappear into the baseline. However, the eventual recovery of the signal at higher ligand concentrations is expected for intermediate exchange regimes, due to the eventual saturation of the binding sites with excess quantities of ligand.<sup>169</sup> However, the reappearance of G15 is not observed, suggesting other reasons are effecting the disappearance of this imino peak. An alternative explanation is that the imino proton of G15 is no longer protected from solvent exchange if the hydrogen bond between G15-G26 is disrupted after ligand binding.

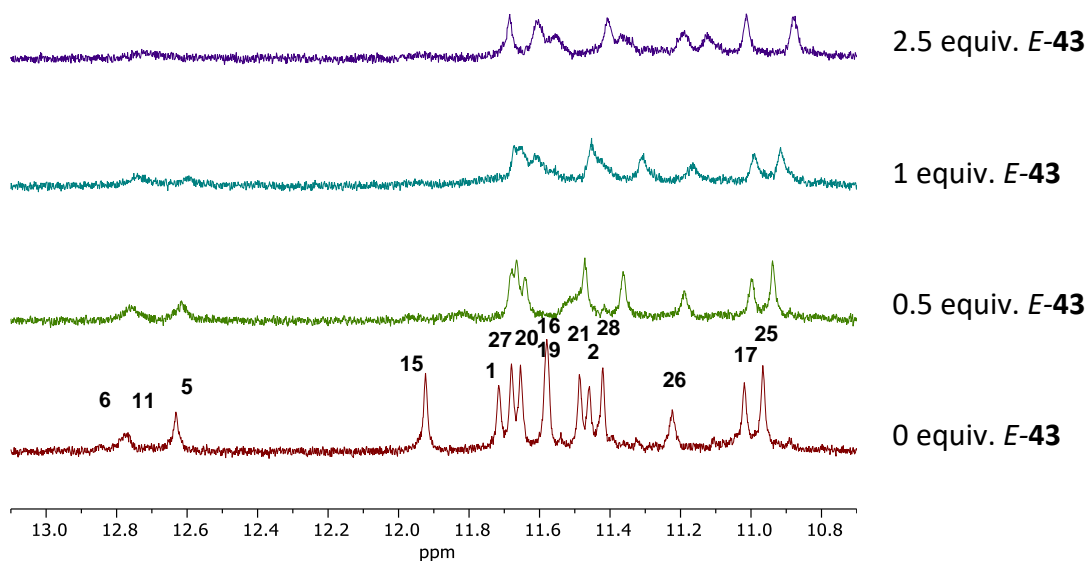
Other trends in the titration of **42** show an upfield shift of the bottom tetrad residues G17-G21-G25-G28. This suggests possible end-stacking binding modes of the ligands at the opposite face of the G4 tetrad are also occurring, as  $\pi$ -stacking interactions of the ligand is expected to cause the protons above or below to experience shielding due to ring current effects.<sup>169,171</sup>

**Figure 3.27** shows the titration of 3-py diazocine **43** with LTR-III G4. This can be carried out at room temperature as the half-life of **43** is longer, and the population of *E-43* should not be significantly reduced over the course of the NMR titrations. Once again, the most significant interaction is the disappearance of the imino signal for G15 upon the addition of 0.5 equivalents of *E-43* (**Figure 3.27a**). Whereas in the *Z-43* titration, this signal persists at a weaker intensity even after 2.5 equivalents of ligand added (**Figure 3.27b**). This gave a reasonable justification that the locations at which both *E-42* and *E-43* are interacting is at the quadruplex-duplex junction.

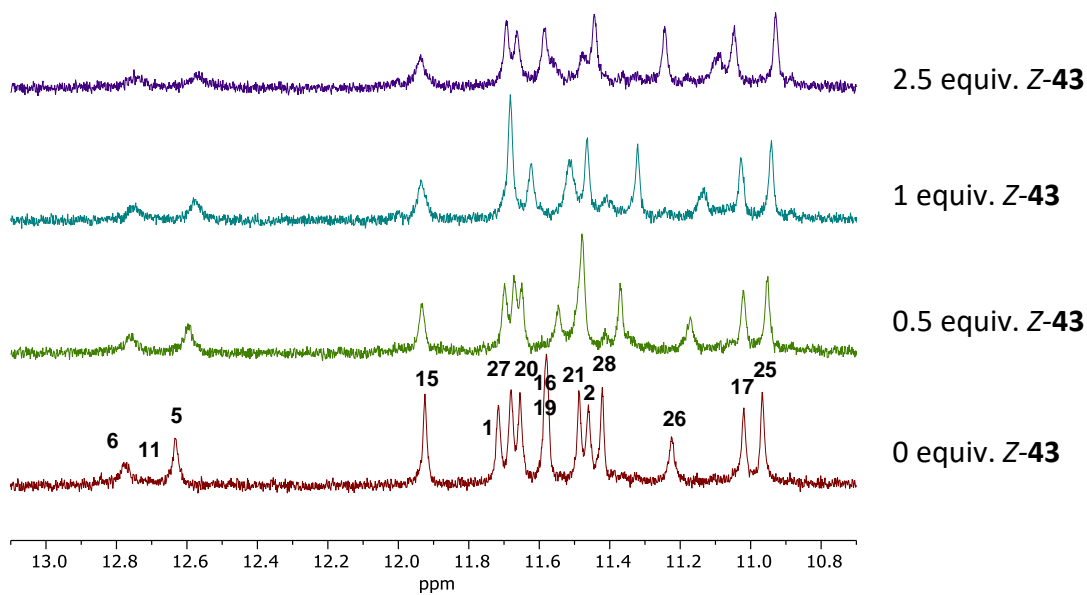
For both titrations, most of the guanines in the bottom tetrad G17-G21-G25-G28 also show an upfield shift, again suggesting that it is a secondary binding site of ligand stacking interactions.<sup>169,171</sup> The duplex region shows similar results, where it is clear that the effect of ligand binding from the *E*-isomer is more significant.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

a)

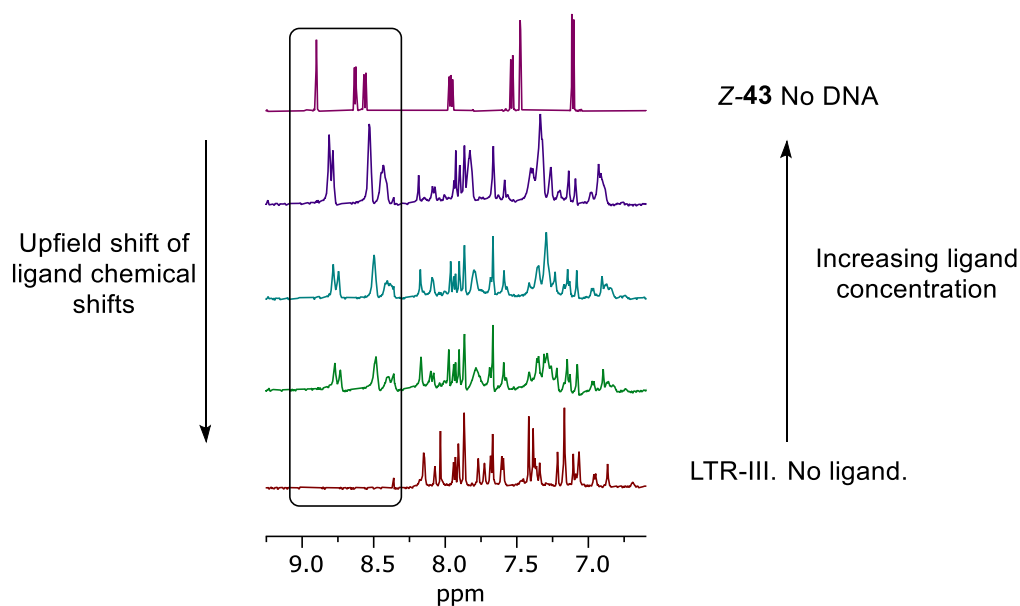


b)

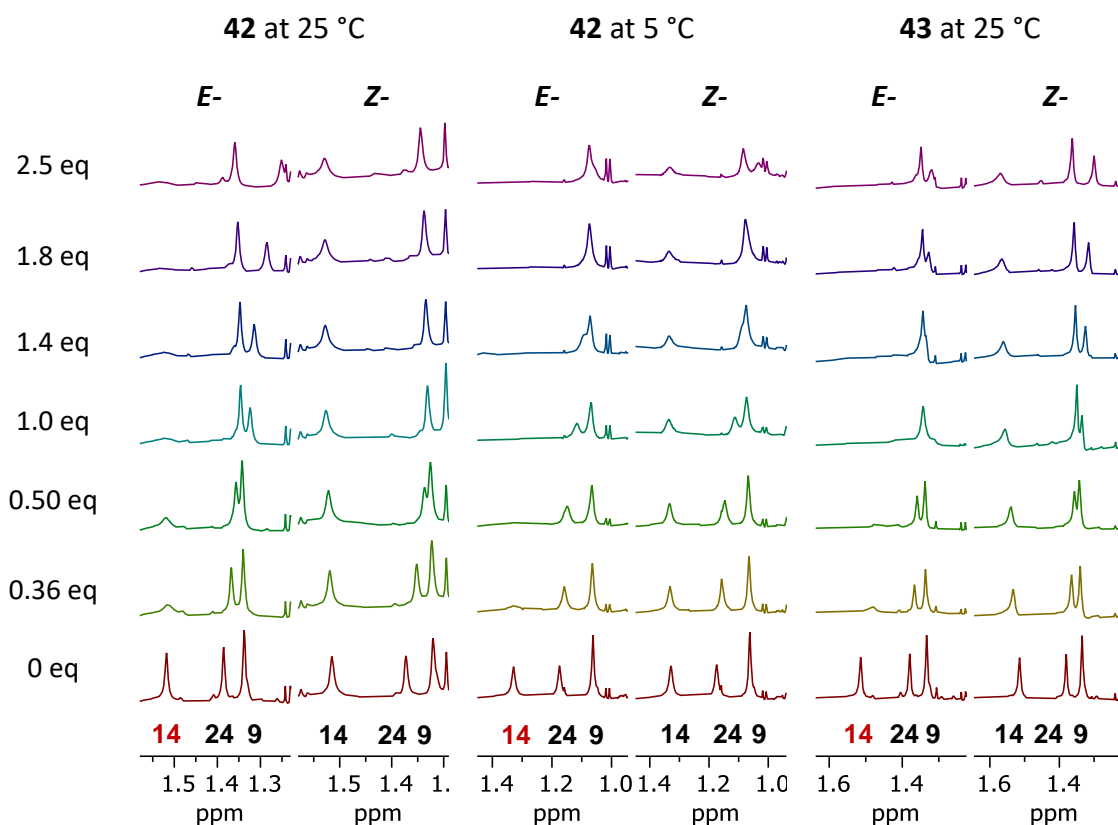


**Figure 3.27:** Imino <sup>1</sup>H NMR titration studies of LTR-III G4 titrated at 25 °C in a 20 mM potassium phosphate buffer (pH 7) containing 70 mM potassium chloride and 10% D<sub>2</sub>O with **a)** E-43 and **b)** Z-43. Initial DNA concentration = 186 μM. Imino protons of each guanine base is labelled.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex



**Figure 3.28:**  $^1\text{H}$  NMR showing aromatic region of Z-43 ligand titration. Ligand experiences an upfield shift upon a relative increase in LTR-III concentration, suggesting  $\pi$ -stacking interactions.



**Figure 3.29:**  $^1\text{H}$  NMR of LTR-III thymine methyl region in NMR titration with diazocines. Affected residue T14 labelled in red. **Left:** 42 at 25 °C ; **Middle:** 42 at 5 °C ; **Right:** 43 at 25 °C .

Changes in the chemical shifts outside of the imino region may also provide ligand binding structural information. A few peaks of Z-43 around 8.5-9.0 ppm do not overlap with any DNA



### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

chemical shifts. By comparing a DNA free spectrum of ligand in buffer, it can be seen that the ligand also experiences an upfield shift and peak broadening towards higher concentrations of DNA, suggesting it experiences its own  $\pi$ -stacking interactions (**Figure 3.28**). There may also be changes in chemical shifts in the aromatic or sugar regions, however these can be challenging to assign due to overlap with other DNA peaks and the chemical shifts of the ligand.

Fortunately, at the opposite end of the  $^1\text{H}$  NMR spectra are thymine methyl peaks which reside between 1-2 ppm. In LTR-III, 3 singlet peaks corresponding to T14, T9 and T24 can be easily identified (**Figure 3.24**). T9 and T24 are loop residues and remain fairly sharp in all titrations. The T14 residue is located at the quadruplex-duplex junction, and the methyl group of T14 quickly broadens out and disappears in titrations with *E*-diazocines. In comparison, *Z*-diazocines cause the T14 methyl signal to broaden and is accompanied with a downfield shift (**Figure 3.29**). These changes supports the claim that the ligand is interacting at the quadruplex-duplex junction for *E*-**42** and *E*-**43**. Additionally, the upfield shift of T24 (a loop residue situated at the bottom tetrad) across all titrations supports evidence for a potential secondary binding site at the bottom tetrad due to  $\pi$ -stacking interactions.

Other studies that investigated ligand binding with NMR titrations to the LTR-III structure showed similar results.<sup>172,173</sup> Where they also observed larger changes to the chemical shifts of residues around the quadruplex-duplex junction region along with a secondary binding site at the bottom tetrad. Indeed, literature studies of synthetically designed sequences with quadruplex-duplex junctions concluded that there was a preferential selectivity of ligands binding at the junction G-tetrad over the external bottom G-tetrad.<sup>174</sup> To properly elucidate binding modes, 2D-NMR techniques such as NOESY or TOCSY could be utilised to identify any changes in the spectrum for the aromatic region. Alternatively, running molecular dynamics (MD) simulations by an *in silico* approach allows the opportunity to gain information at atomistic levels on how the two ligands bind with the quadruplex structure.

## 3.2 Computational Molecular Dynamics Studies

MD simulations is a computational tool that can also provide information for the study of the binding modes for the diazocine ligands with LTR-III. This is a useful method to approach understanding biomolecules such as DNA, as it can look at structural behaviours of DNA and the ligand at an atomistic level.

Molecular dynamics have been used in various ways to further understanding of G-quadruplexes. For example, it can provide structural information to support experimental results by providing data that cannot be measured experimentally and furthering understanding of different ligand binding poses. Or it can be used to test hypotheses in G-quadruplex folding mechanisms by exploring conformations that are unlikely to form in real experiments or by the modification of parameters to test the effect of different ions on stability.<sup>175–177</sup> Here, we are interested in seeing if utilising a computational approach can provide more insight into the experimental results that have been seen from the biophysical studies.

### 3.2.1 Methods

In the approach to study the LTR-III sequence by utilising *in silico* approaches, it was decided to examine the differences in binding between the LTR-III sequence and the two photoisomers of ligand **42** with molecular dynamics simulations.

The behaviour of atomic interactions in biomolecules is described by a combination of molecular mechanics energy terms known as a force field. There are many existing force fields for nucleic acid molecular dynamics studies.<sup>178–180</sup> The force fields utilised to describe the nucleic acids are composed of a pair additive combination of Lennard-Jones van der Waals and Coulomb electrostatic non-bonded interactions and bonded terms defined by bond lengths, angles and dihedrals.<sup>180</sup> The majority of nucleic acid MD simulations have been carried out with AMBER force fields. Many benchmark studies have compared their stability and validity at modelling nucleic acid structures,<sup>178–180</sup> with many recent MD simulations for DNA carried out with either OL15<sup>181</sup> or bsc1<sup>182</sup> force fields. Refinements from the original parm99 force field to current ones have been made by the reparameterisation of the dihedral

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

energy terms of the DNA backbone and the sugar glycosidic torsion angles. These changes have allowed for longer stable simulations that do not unfold, allowing for meaningful information to be extracted.<sup>175</sup>

The nucleic acid force fields were developed for the canonical duplex B-DNA structures, so there are certain limitations of these force fields in MD simulations of G-quadruplexes. One example is that the aforementioned non-bonded interactions are modelled as point charges which cannot be polarised, this results in significant repulsive forces between the positively charged cations within the central G-quadruplex channel.<sup>175</sup> This potentially causes the expulsion of ions from the central channel, leading to unstable structures.<sup>183</sup> Alternative force fields have attempted to take into account polarisable groups in the developments of the AMOEBA<sup>184</sup> and Drude polarizable force fields.<sup>180,185–187</sup>

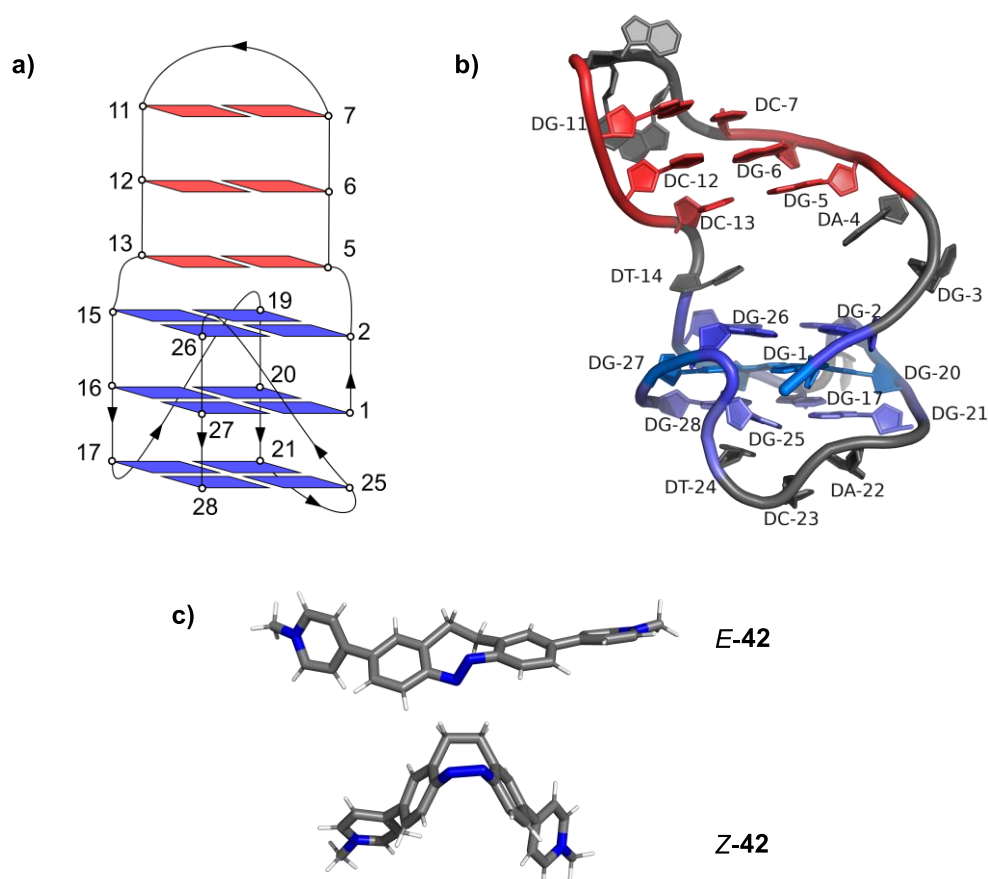
Another issue is the dynamic nature of G-quadruplexes. While most simulations have shown that residues contributing to the G4 tetrads are very stable, the flexible loop residues can adopt many different conformations that may require up to microsecond timescales to explore. In particular, propellor loops have been reported to be poorly modelled in MD simulations with the existing force fields. This could be due to a combination of either the force field not describing the behaviour of loop residues well and the complicated energy landscape of G4s.<sup>188–191</sup>

The choice of the bsc1 force field for this work was in large due to a recent study by Castelli *et al.*, who examined the dynamics and stability of the same LTR-III sequence.<sup>192</sup> Their study of LTR-III by molecular dynamics simulations was due to its more complex structural nature, having a duplex stem loop component in addition to a G-quadruplex. They carried out simulations with the bsc1 force field in 4 different solvent environments: K-TIP3P, KCl-TIP3P, K-TIP4P, and KCl-TIP4P. These involved two different water models (TIP3P and TIP4P) and two different ion concentrations (“K”: K<sup>+</sup> ions required to neutralise the system and “KCl”: additional salt to reach 100 mM ionic strength). The overall observations were that the quadruplex component was very stable and there were no observed disruptions to the tetrad. In contrast, the duplex component was more dynamic and very flexible.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

In our work, we used the 10 different NMR conformations as separate starting structures to run the simulations (PDB: 6H1K). The sequence was simulated using the explicit water model, TIP3P with potassium chloride ions at 100 mM concentration with Joung-Cheatham ion parameters.<sup>193–195</sup> The ligands *E-42* and *Z-42* were parameterised with the general AMBER force field, GAFF.<sup>196</sup> Based on the NMR titration data, we were confident that the *E*-ligand was interacting strongly at the quadruplex-duplex junction and thus manually docked both photoisomers into the junction of LTR-III, running each model for 255 ns at 300 K.

#### 3.2.2 Molecular Dynamics Results and Analysis

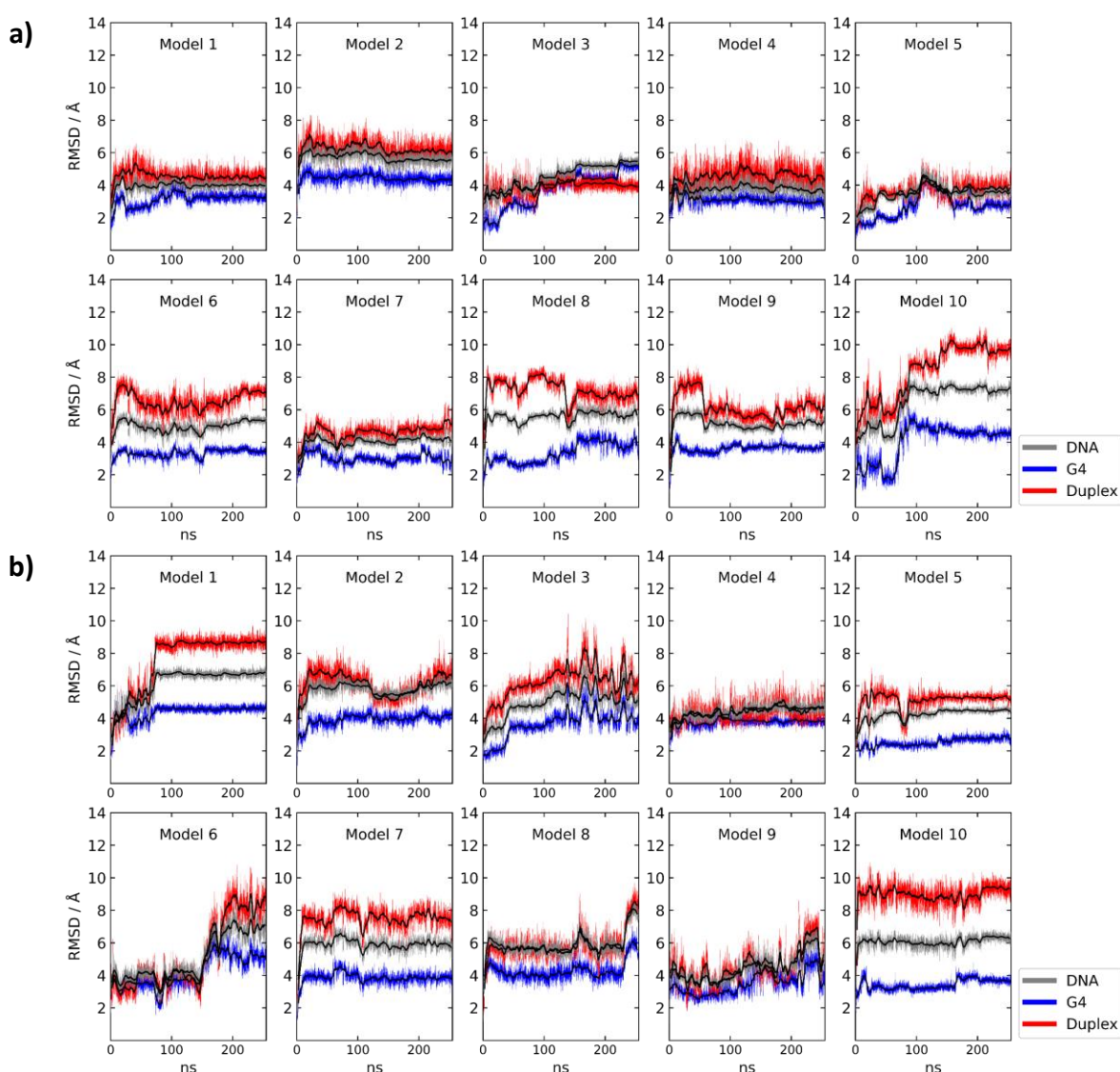


**Figure 3.30:** **a)** Simplified cartoon of LTR-III, the guanine bases in the quadruplex are shown as blue rectangles and cytosine and guanine base pairs in the duplex region are shown as red rectangles. **b)** Cartoon representation of LTR-III (PDB: 6H1K), guanine-quadruplex bases shown in shades of blue and guanine-cytosine duplex base pairs shown in red. Other loop residues in grey. **c)** Stick representation of ligands *E-42* and *Z-42*.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

**Table 3.3:** Average ( $\pm$  standard deviation) and median RMSD of DNA residues for 10 replicate MD simulations with *E-42* and *Z-42*.

	RMSD <sub>E</sub>		RMSD <sub>Z</sub>	
	Mean (Å)	Median (Å)	Mean (Å)	Median (Å)
All DNA residues	4.76 $\pm$ 1.1	4.72	5.28 $\pm$ 1.1	5.45
G4 residues	3.44 $\pm$ 0.8	3.36	3.74 $\pm$ 0.8	3.79
Duplex residues (5-13)	5.58 $\pm$ 1.7	5.31	6.16 $\pm$ 1.8	5.84
Duplex G-C pairs	4.92 $\pm$ 1.8	4.74	5.26 $\pm$ 2.2	4.98



**Figure 3.31:** RMSD of the atomic positions for DNA (grey), G4 (blue), and duplex stem loop (red) during MD simulations with **a)** *E-42* or **b)** *Z-42*. Duplex stem loop positions are represented by residues 5-13, duplex base pairs are G5-C13, G6-C12, C7-G11. Black line shows moving averages over 500 neighbouring points.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

The first analysis carried out was the root mean squared deviation (RMSD) of the DNA during the MD simulations. The RMSD measures the average distance that atoms have travelled from a reference position. In this case, the reference position is the starting position at 0 ns of the simulation. **Table 3.3** shows the average and median root mean squared deviation (RMSD) values for the two sets of MD simulations. As expected, the atoms within the G4 residues have deviated the least from the initial structures indicating that the G4 residues are the most stable component ( $E$ : 3.44 Å and  $Z$ : 3.74 Å). This is clearly visible in the RMSD time plots (**Figure 3.31**) that show the G4 residues (blue) typically between 2-5 Å. The duplex residues 5-13 (red) (3 guanine-cytosine pairs and 3 loop residues 8-10), deviate the most showing that this component is more flexible ( $E$ : 5.58 Å and  $Z$ : 6.16 Å).

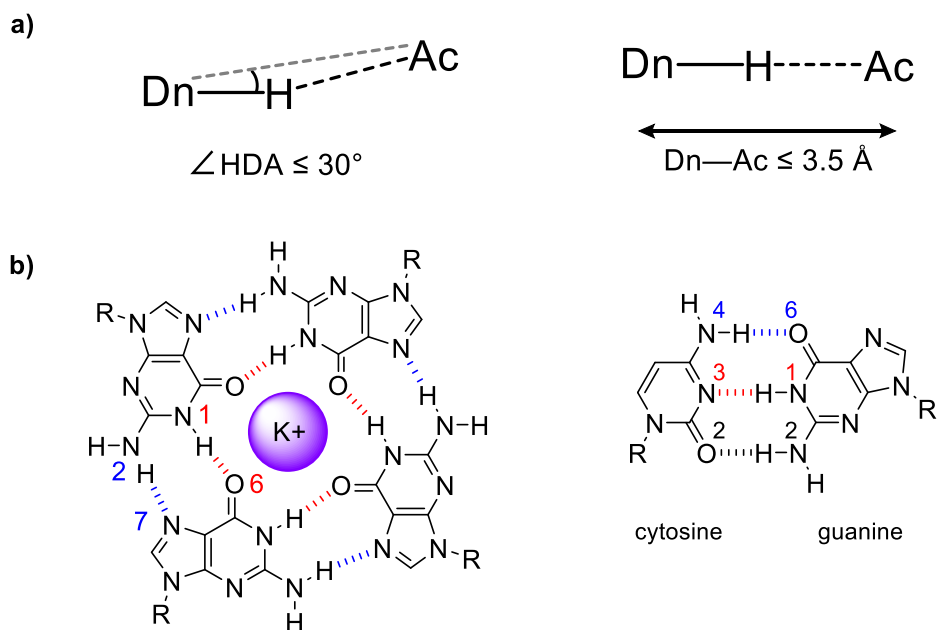
Within the 255 ns timeframe of the MD simulations, some models have reached a local conformational equilibrium. Whereas other models have either unfolded or have adopted multiple conformations. This can be seen in the large fluctuations of the RMSD and illustrates the dynamic nature of these short oligomers (**Figure 3.31**). Additionally, these RMSD values were relatively high compared to the RMSD values of LTR-III without a ligand.<sup>192</sup> Though it is unsurprising that the system deviates far from structures that began without a ligand present, and thus requires significantly more time to equilibrate. Finally, comparing the two sets of simulations with different ligands, we see that  $\text{RMSD}_E < \text{RMSD}_Z$ , showing that the DNA atoms move further in the presence of the bulkier  $Z$ -diazocine.

However, looking at the RMSD for the DNA sequence does not give a full picture of what is occurring, as it only demonstrates changes in conformations relative to a reference and cannot identify if a structure is misfolded. Additionally, different conformations may also have the same RMSD relative to the reference. A good method for judging the stability of the simulations is to analyse the hydrogen bonds that build up the secondary structures of the quadruplex and duplex motifs (**Figure 3.32b**).

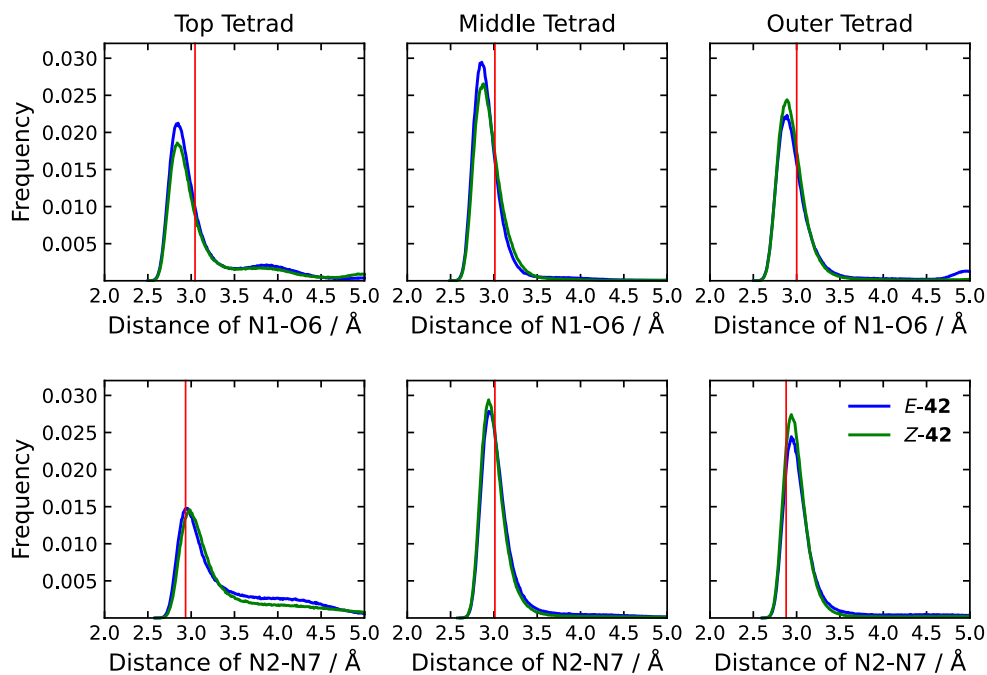
Hydrogen bonds are typically defined as having a maximum distance of 3.5 Å between the donor and acceptor atoms, and having a hydrogen-donor-acceptor angle of less than 30° (**Figure 3.32a**).<sup>180,197</sup> Though in the study by Castelli *et al.*, they defined a cutoff distance of

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

3.0 Å and donor-hydrogen-acceptor angle of 135° to calculate the hydrogen bond persistence.<sup>192</sup>



**Figure 3.32:** a) Definition of a hydrogen bond. HDA angle  $\leq 30^\circ$  and Donor-Acceptor distance  $\leq 3.5 \text{ \AA}$ . b) **Left:** Hydrogen bonds between guanine tetrads. N1-O6 (red), N2-N7 (blue). Chelation of O6 atoms to  $\text{K}^+$  stabilises the tetrad. **Right:** Cytosine-Guanine base pair with Watson-Crick hydrogen bonds. N1-N3 (red), N2-O2 (black), N4-O6 (blue).



**Figure 3.33:** Histogram of hydrogen bond distances per tetrad. Blue: MD simulations with *E-42*. Green: MD simulations with *Z-42*. Red line: Average hydrogen bond distance reported in NMR structure. Populations above 5.0 Å not shown.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

**Table 3.4:** Summary of hydrogen bond persistence for all models in simulations with the ligand *E-42* and *Z-42*. **Top:** values for G4 tetrad hydrogen bonds. **Bottom:** values for duplex hydrogen bonds. Cut off values for hydrogen bond persistence was 3.0 Å and 135°. Top refers to bases G2←G26 ←G15←G19. Mid refers to bases G1→G27→G16→G20. Out refers to G25→G28→G17→G21. (Arrows indicate the direction of hydrogen bond donor to hydrogen bond acceptor).

Hydrogen Bond Persistence (%)			
G4 Tetrads – All Models			
<i>E-42</i>	top	mid	out
N1-O6	52	71	52
N2-N7	25	47	42
<i>Z-42</i>	top	mid	out
N1-O6	45	64	56
N2-N7	22	51	48
Duplex Bases – All Models			
	G5-C13	G6-C12	C7-G11
<i>E-42</i>	69	69	67
<i>Z-42</i>	64	77	63

For comparison, we have reported hydrogen bond persistence with the same criteria (**Table 3.4**). **Figure 3.33** shows histograms of G4 tetrad hydrogen bond distances, which includes the average hydrogen bond distances of the NMR models (ranging around 2.8-3.0 Å) shown as a red line. It is clear that the middle tetrad of LTR-III has the highest hydrogen bond persistence which could be because these hydrogen bonds are protected from the solvent by the tetrads above and below it. We also see that the top tetrad has the lowest hydrogen bond persistence, as the histograms show significant populations with distances above 3.5 Å. This outcome is unsurprising as there was a lower probability to find stabilising potassium ions chelating to the top tetrad in the central channel, leading to higher degrees of motion and further distances for those guanines.

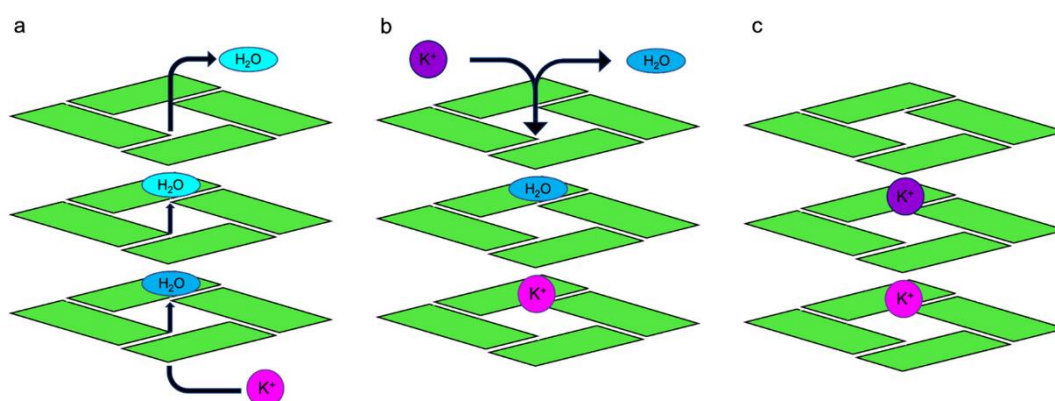
The stability of the tetrads and thus the quadruplex in these simulations were highly dependent on the presence of potassium ions. The presence and chelation of an ion restricts the motions of the guanine residues and improves hydrogen bonding interactions as a consequence, resulting in stable G-tetrads for the rest of the simulation. Potassium ions within



### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

the central channel are positioned between two tetrads and chelate to 8 guanine residues. There are two potassium sites available in the LTR-III G4 structure, however 18 out of 20 simulations only had one ion enter the central channel over the 255 ns timeframe.

Castelli *et al.* noted in their work, that the entry of both potassium ions into the LTR-III channel occurred from opposite ends of the tetrads (**Figure 3.34**).<sup>192</sup> In our 18 simulations (with a single ion site occupied), there were a few cases where the potassium ion moved between the two sites. However, none resulted in the addition of a second potassium ion within the simulation timeframe. Most ions entered through the bottom tetrad and occupied the site between the bottom and middle tetrads. It is unclear how significant the ligand placement at the quadruplex-duplex junction above the top tetrad hindered the access of potassium ions to the top tetrad, thus effecting the stability of the top tetrad.



**Figure 3.34:** Simplified scheme showing initial potassium ion entry into the central G-channel of LTR-III displaces a water molecule. Then, a second potassium ion enters from opposite face. © 2022 American Chemical Society.<sup>192</sup>

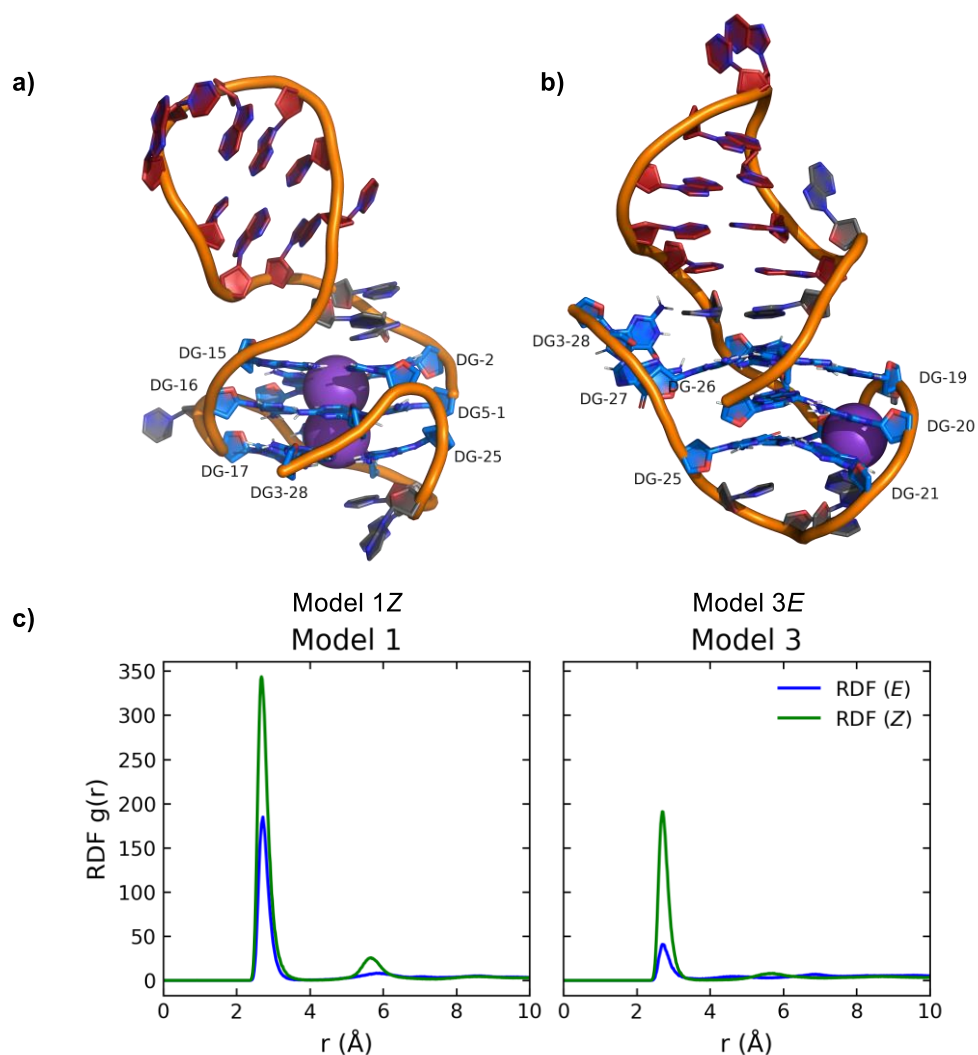
Outside of the 18 simulations with single potassium ion occupation, there were two notable outliers. Model 1Z was successful at incorporating two ions at the expected sites, with the first K<sup>+</sup> ion entering from the top face at the junction around 12 ns and the second K<sup>+</sup> ion entering from the external bottom face by 50 ns (**Figure 3.35a**). We can see in the radial distribution functions that Model 1Z has double the probability than Model 1E to find a potassium ion near the chelating G4-O6 atoms (**Figure 3.35c**). The stabilisation brought by the chelation to both potassium ions is likely why the RMSD of Model 1Z is stable after the rearrangement of the chelating guanines ~75 ns (**Figure 3.31b**). It is also the only simulation where the hydrogen bond persistence of the G4 tetrads are above 40% (**Table 3.5**).

Tetrad Hydrogen Bond Persistence (%)															
	Model 1			Model 2			Model 3			Model 4			Model 5		
<i>E-42</i>	top	mid	out	top	mid	out	top	mid	out	top	mid	out	top	mid	out
N1-O6	55	77	69	55	77	73	58	27	22	52	82	74	26	70	58
N2-N7	22	53	57	14	56	54	21	17	2	21	42	53	17	54	55
<i>Z-42</i>	top	mid	out	top	mid	out	top	mid	out	top	mid	out	top	mid	out
N1-O6	64	41	63	16	67	71	37	70	72	71	67	11	38	68	63
N2-N7	52	67	47	12	55	53	13	60	58	39	36	2	12	60	58
	Model 6			Model 7			Model 8			Model 9			Model 10		
<i>E-42</i>	top	mid	out	top	mid	out	top	mid	out	top	mid	out	top	mid	out
N1-O6	54	78	73	65	68	10	37	71	59	57	75	35	64	83	49
N2-N7	20	52	56	51	46	3	21	62	63	25	47	17	37	43	63
<i>Z-42</i>	top	mid	out	top	mid	out	top	mid	out	top	mid	out	top	mid	out
N1-O6	50	74	72	68	79	53	16	29	30	45	72	63	40	70	64
N2-N7	17	55	54	27	39	60	6	30	29	19	53	58	20	59	60

**Table 3.5:** Percentage hydrogen bond persistence of each G4 tetrad in LTR-III by the two types of hydrogen bonds, N1-O6 and N2-N7 for all 10 models of MD simulation data over the full 255 ns of simulation time. Cut off values for hydrogen bond persistence was 3.0 Å and 135°. Top refers to bases G2←G26←G15←G19. Mid refers to bases G1→G27→G16→G20. Out refers to G25→G28→G17→G21. (Arrows indicate the direction of hydrogen bond donor to hydrogen bond acceptor).

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

In a contrasting scenario, Model 3E was not able to incorporate any potassium ions within its central channel. Instead, a potassium ion was located in a pocket coordinated to the phosphate atoms of C18 and the G20-N7 and G21-O6 atoms, respectively (**Figure 3.35b**). G20 and G21 are crucial to maintaining the tetrad hydrogen bond network of LTR-III. As these residues preferentially coordinate to the ion, this distorts the G4 structure of Model 3E and leads to an unfolded G4. Compared to Model 1Z, the tetrads in Model 3E are not planar and the 3'-end of the sequence begins to unfold. A rare example where the RMSD of the guanine residues deviated further than the duplex stem loop (**Figure 3.31a**).



**Figure 3.35:** a-b) Representative cartoons of LTR-III demonstrating the effect that potassium ions (purple) provide to the stability of the G-tetrads. Ligand not included for clarity. a) Model 1Z with two potassium ions within the central channel shows a well folded G4 structure. b) Model 3E with no potassium ions within the central channel shows unfolding of the G4 structure at the 3'-end of the sequence. c) Radial distribution function between O6-G4 atoms and potassium ions for Models 1 and 3. Blue: MD simulations with E-42. Green: MD simulations with Z-42.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

**Table 3.6:** Percentage hydrogen bond persistence of each G-C pair in the duplex stem loop of LTR-III for all 10 models of MD simulation data over the full 255 ns of simulation time. Cut off values for hydrogen bond persistence was 3.0 Å and 135°.

Duplex Guanine-Cytosine Hydrogen Bond Persistence (%)						
Model	E-42			Z-42		
	G5-C13	G6-C12	C7-G11	G5-C13	G6-C12	C7-G11
1	79	78	78	16	79	70
2	73	73	66	75	77	18
3	78	77	82	79	78	74
4	80	75	80	82	78	78
5	77	79	77	74	73	81
6	75	80	70	76	79	73
7	75	75	78	81	79	10
8	76	76	62	72	77	78
9	78	76	75	77	77	67
10	0	4	5	5	77	81

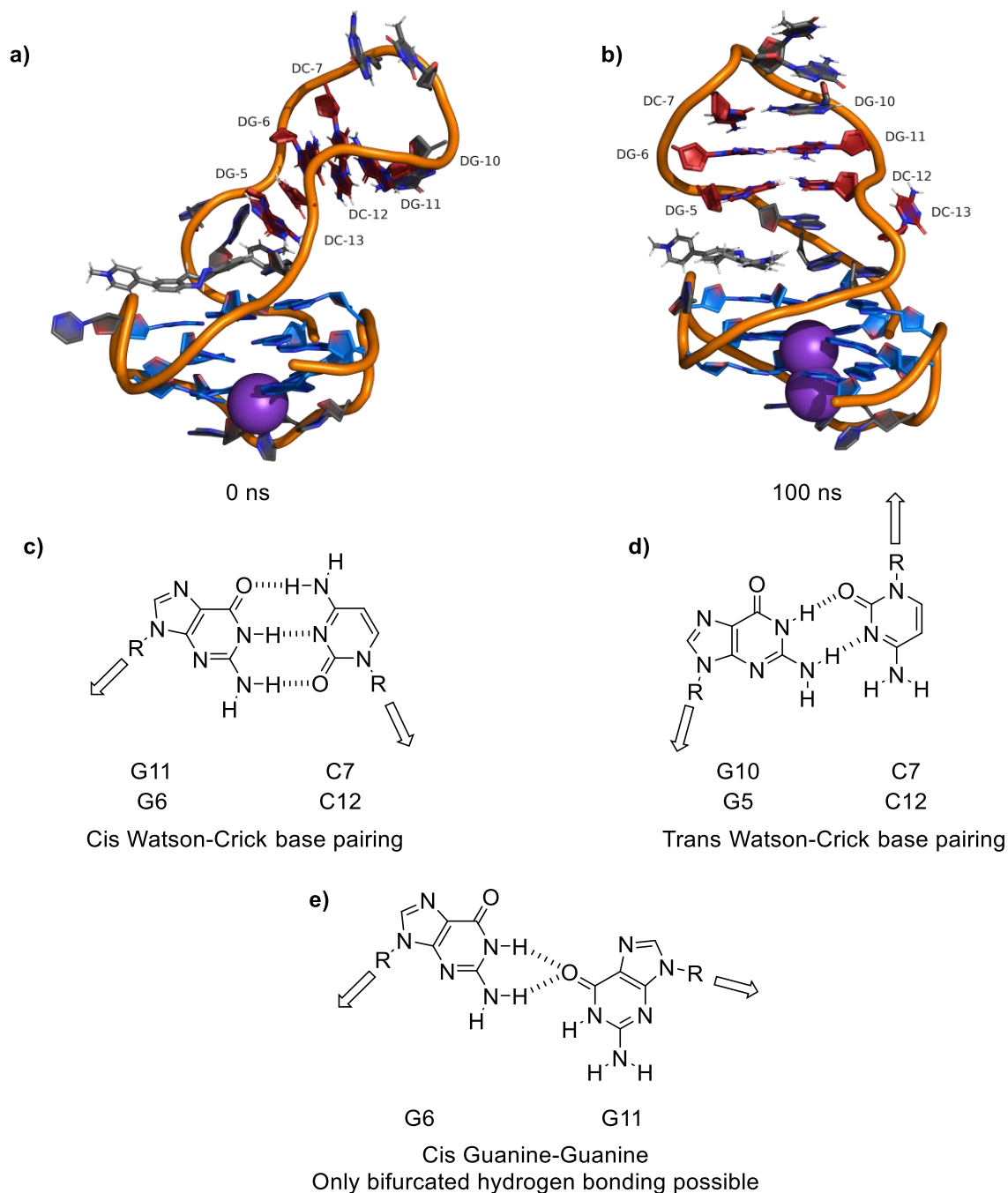
While the stability of hydrogen bonds in the G4 moiety was highly dependent on potassium ion site occupancy, this was not as significant for the duplex stem loop. The RMSD of the duplex stem loop suggested that it is more flexible, but the higher values of 60% hydrogen bond persistence for the duplex stem loop show that the base pairs form stable hydrogen bonds though the stem loop as a whole is dynamic (**Table 3.4**).

Typically the lowest hydrogen bond persistence occurred with the top most pair, C7-G11, where the bases have a higher probability of fraying apart as it is more exposed towards the solvent. This could match the experimental evidence from <sup>1</sup>H NMR spectra where the imino peak for G11 is the broadest. Other weaker hydrogen bonds in the stem loop are the G5-C13 pair in some of the simulations with Z-42, as the bulky nature of the ligand could disrupt the hydrogen bonds for the base pair nearest the quadruplex-duplex junction.

**Table 3.6** breaks down the hydrogen bond persistence by model for the duplex stem base pairs. The only outlier, Model 10E, has full disruption between all 3 G-C base pairs. Unfortunately, this is not because of the ligand, but likely due to the starting positions of the bases where G5 begins below C13. This leads to a cascade of motions where there is strand slippage and new hydrogen bonds form between G5-C12 and C7-G10 in the misfolded stem loop (**Figure 3.36**). However, the new base pairs cannot form typical Watson-Crick pairs as the orientation of the sugar phosphate backbone are trans to each other. Thus, it can only form

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

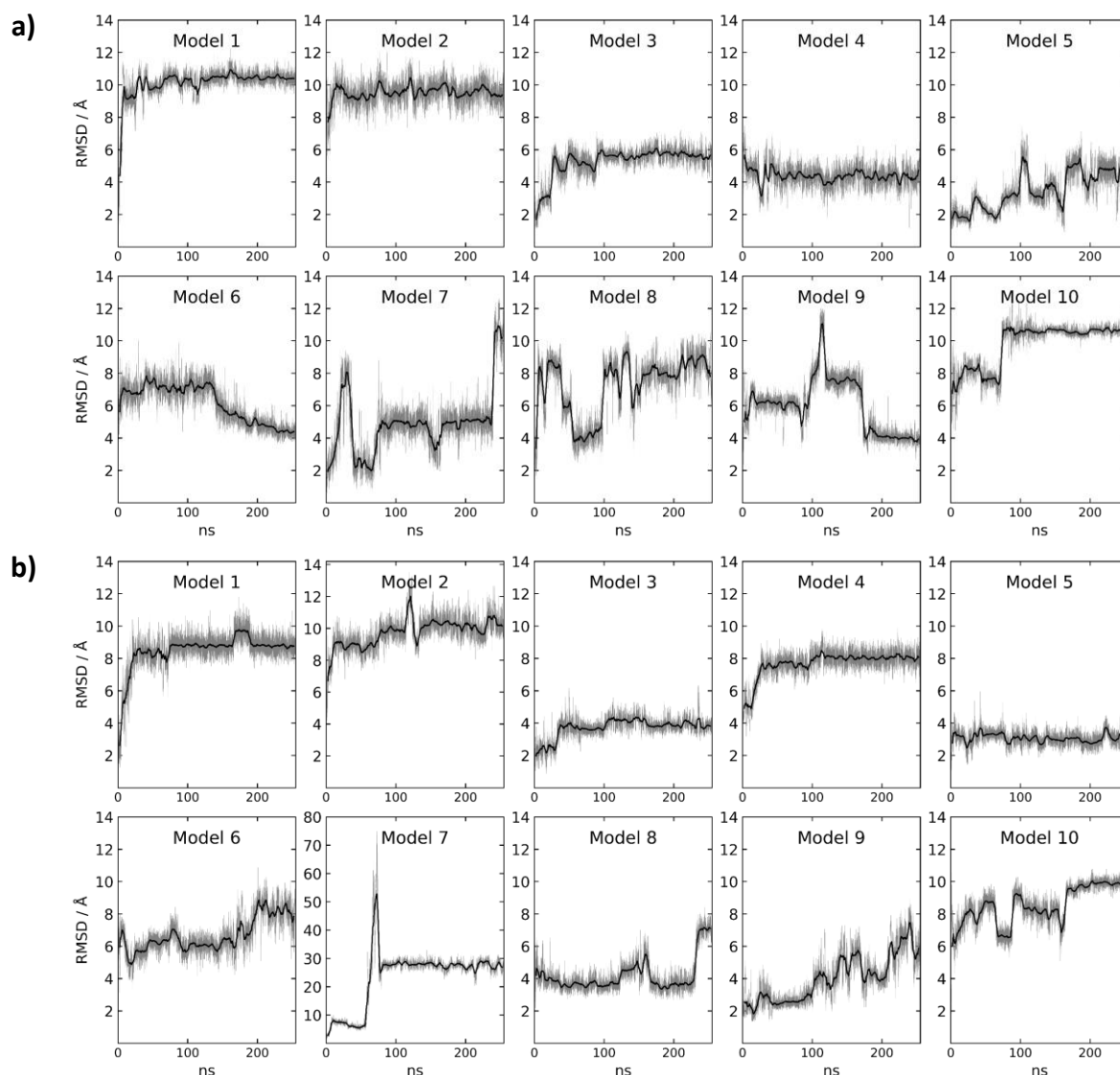
trans/reverse Watson-Crick pairs, another type of non-canonical hydrogen bonding (**Figure 3.36c-d**).<sup>198</sup> The other nucleotides, G6 and G11 can only form a bifurcated hydrogen bond where the imino N1 and amino N2 protons coordinate to the O6 atom as the sugars are orientated on the same side (**Figure 3.36e**).



**Figure 3.36:** Cartoons showing the misalignment of the duplex base pairs in Model 10E. **a)** At 0 ns, only G5-C13 is disrupted. However this quickly results in the whole duplex disruption. **b)** which eventually stabilises as a misfolded stem loop from 100 ns onwards. **c-d)** Types of guanine-cytosine base pairing in Model 10E stem loop. **c)** Canonical cis WC base pair where both sugars are on the same side. **d)** Non-canonical trans WC base pair where sugars on the opposite side. **e)** Bifurcated cis G6-G11 hydrogen bonding in Model 10E at 100 ns.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

Having established the importance of potassium ions in maintaining a stable G4 structure, we can now look to see how the two ligands *E-42* and *Z-42* have behaved in the simulations. The binding site at the junction is rather large and open, unlike many enzymes with small binding pockets. So on average, the ligand moves significantly during the simulation time (more than the DNA does), adopting multiple conformations. However, due to the different starting positions of the *E*- and *Z*- ligands, it is difficult to properly compare two separate models.



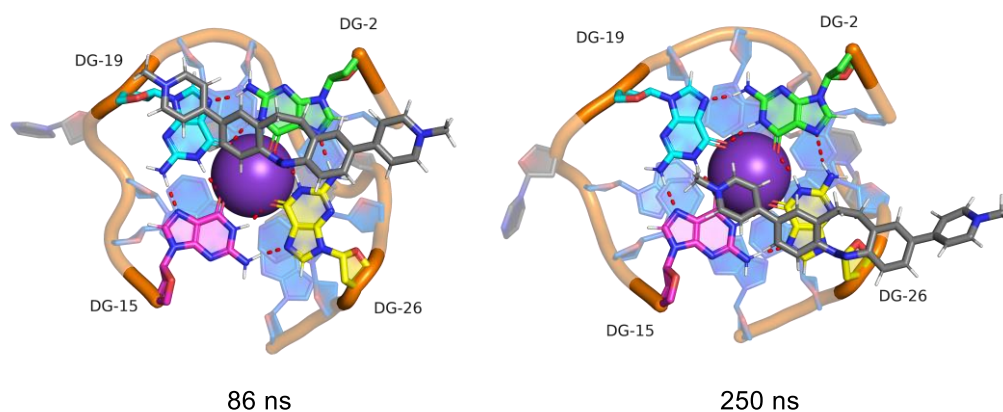
**Figure 3.37:** RMSD of the atomic positions for the ligand during simulation. Black line shows moving averages over 500 neighbouring points. **a)** *E-42*, **b)** *Z-42*.

Most of the interactions the ligand can make are  $\pi$ -stacking interactions and there appears to be hydrogen bond interactions. Thus the effect that the ligand has towards G4 residues while potassium ions occupy the upper ion site in the G4 channel the ligand is relatively miniscule, with little effect on G4 topology, something that is observed in CD titration experiments.

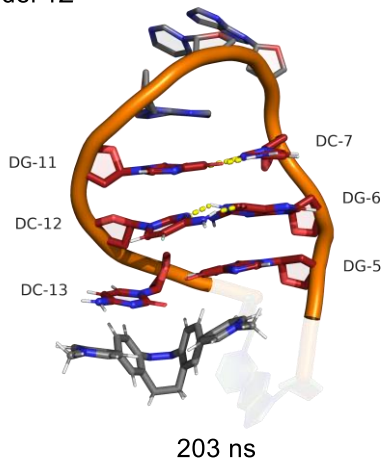
### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

Models 7E, 1Z and 4Z were the few examples where the top tetrad was stable due to significant potassium ion occupancy. Ligand *E-42* in Model 7E is able to adopt multiple binding poses by shifting across the tetrad with large changes in the RMSD (**Figure 3.38**). Whereas in the two models with the bulkier Z-diazocine, there is less movement of the ligand and the pyridinium rings rotate out of plane to maximise  $\pi$ -stacking interactions with residues in the duplex or the top tetrad.

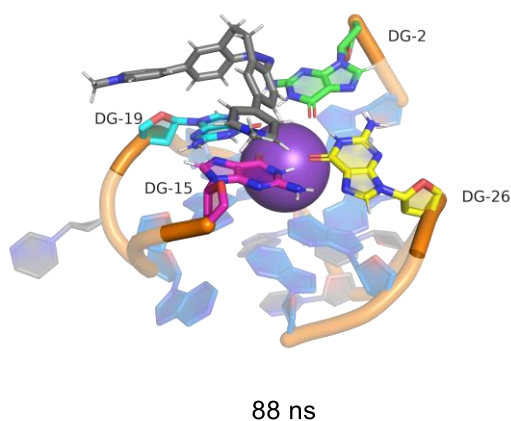
a) Model 7E



b) Model 1Z

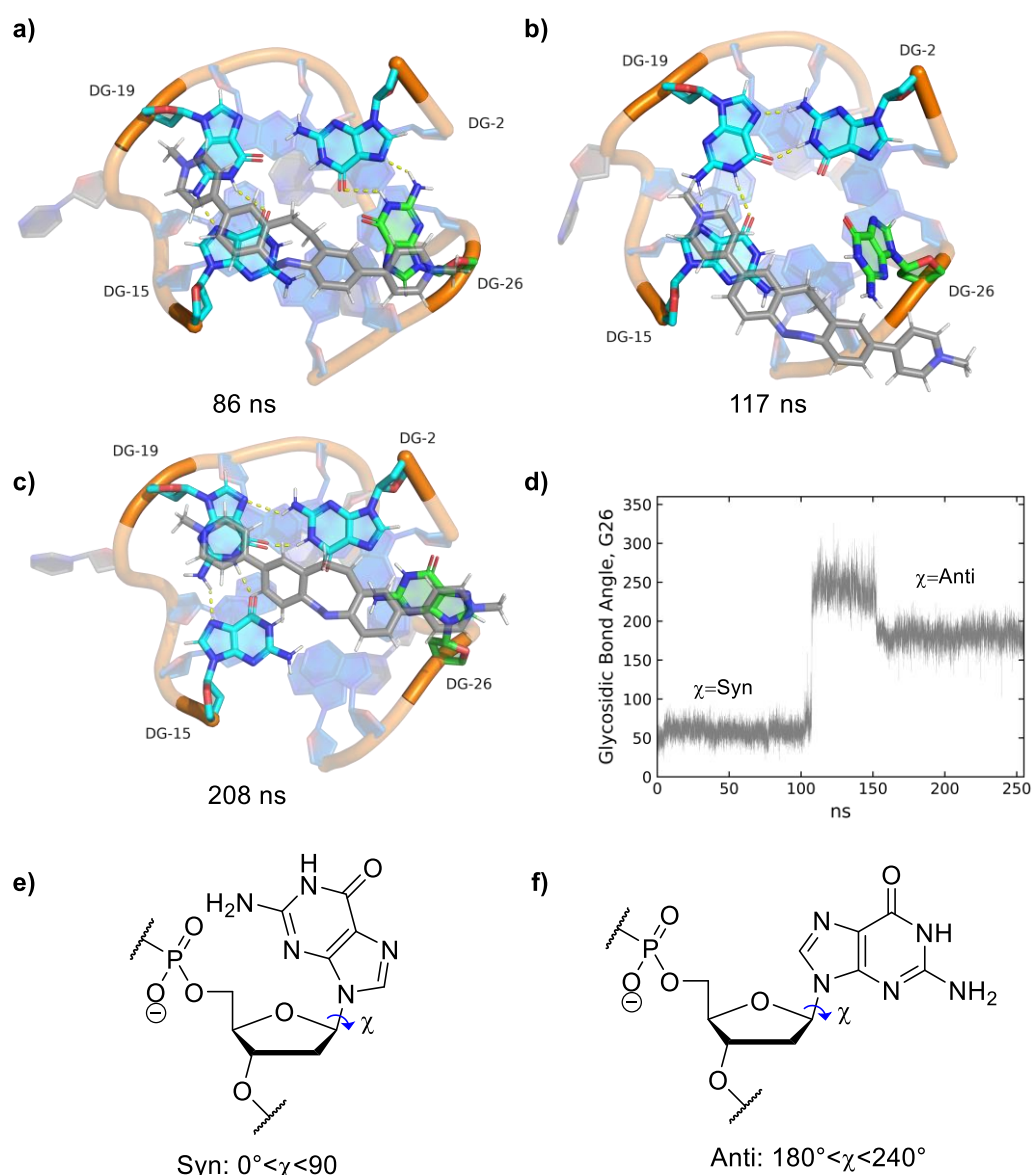


c) Model 4Z



**Figure 3.38:** a) Two different conformations of *E-42* in Model 7E after the top ion site is occupied with a potassium ion after 75 ns. **Left:** Ligand  $\pi$ -stacking above residue G19 and G2 between 80-240 ns, (86 ns shown). **Right:** New conformation adopted by ligand near end of simulation, where the ligand transitions to  $\pi$ -stack over G26, (250 ns shown). **b-c)** Conformations of Z-42 showing smaller  $\pi$ -stacking interactions of just the pyridinium ring **b)** to the duplex stem loop residues G5 and C13. **c)** to residue G15 of the top tetrad of LTR-III.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex



**Figure 3.39:** a-c) Top tetrad of LTR-III in Model 5E. a) 86 ns, a distorted top tetrad with E-42  $\pi$ -stacking with G19, G15, G26, b) 117 ns, flipping of G26 during interactions with E-42, c) 208 ns, flipped G26 that cannot participate in Hoogsteen hydrogen bonding with its neighbours. d) glycosidic bond angles of G26 in Model 5E over time. e-f) Illustrations of *syn* and *anti* glycosidic bond angles for guanines.

Yet these were far from the only binding modes that were explored over the full set of simulations. Simulations without a potassium ion stabilising the top tetrad regularly showed that residue G26 had the longest and weakest hydrogen bonds. This particular nucleotide is part of a V-shaped loop (G25→G26) that connects the corners of the adjacent first and last columns (Figure 3.30a). In Model 5E, the ligand further affects G26 by flipping the base, and changes the glycosidic bond angle from a *syn* to an *anti* conformation (Figure 3.39). This completely disrupts hydrogen bonding, as G26 now faces the opposite direction. Whilst the

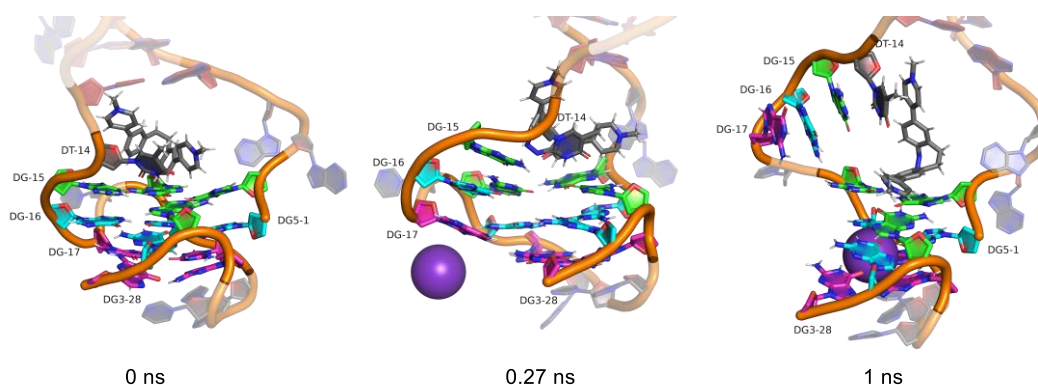


### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

*syn* to *anti* conformational change suggests an unsuccessful simulation, there are a couple interesting implications as a result.

G26 is one of the more dynamic residues in LTR-III when there is no potassium stabilising the top tetrad. It is unclear why this is the case, but one possibility is that V-shaped loops are poorly described by the bsc1 force field, leading to regular distortions at this position that can only be noticed when potassium atoms are absent. The other possibility is that G26 is naturally a dynamic residue. If it is the latter, then it raises the question if its flexible behaviour is visible in the NMR titrations as the disappearance of the G15-N1 imino proton peak, tentatively postulated to be due to a disruption of the hydrogen bond between G15-G26.

The final two examples show contrasting behaviours of LTR-III in the presence of Z-42. The simulation of Model 8Z had very low hydrogen bond persistence, which was because of a very quickly unfolding G4 structure within the first nanosecond of simulation time (**Figure 3.40**). This is unlikely to be a valid or accurate simulation, particularly since significant unfolding is not observed in CD or NMR titration experiments. It is most likely due to poor starting atomic positions favouring an energetic pathway to an unfolded conformations. If this sort of unfolding does occur, it could only have very low probability, with most outcomes preferring to retain a folded state.



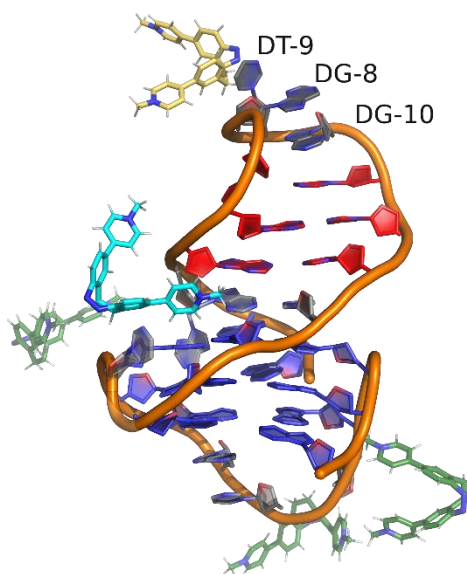
**Figure 3.40:** Unstable G4 structure in Model 8Z, which unfolds within 1 ns. Potentially due to favourable interactions between Z-42 and residues T14. Residues of each tetrad are coloured: top tetrad (green), middle tetrad (cyan), bottom tetrad (magenta).

Finally, there is Model 7Z where after ~60 ns of simulation time, the bulky Z-diazocine was expelled from binding at the junction after the formation of an A4-T14 WC hydrogen bond. The ligand eventually found the top of the stem loop ~80 ns and remained there for the rest

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

of the simulation (**Figure 3.41**). This presents a unique case as none of the other repeats had the ligand leave the junction. The expulsion of the bulky **Z-42** does correlate with experimental results as this ligand consistently showed poorer binding affinity towards LTR-III than **E-42** in all biophysical experiments and NMR did not suggest strong interactions of **Z-42** at the quadruplex-duplex junction. Another potential binding site for **Z-42** may in fact be at the outer tetrad, also an accessible site for the bulky ligand. We do see that sort of binding in the NMR titrations of both isomers, and it could be possible that this may be due to populations of **Z**-diazocines in both titrations.



**Figure 3.41:** **Z-42** is expelled from the quadruplex-duplex junction in Model 7Z. Eventually binding to the residues at the top of the duplex stem loop. Cyan: Ligand position before expulsion. Green: Ligand in various regions of solvent. Beige: Ligand  $\pi$ -stacking with loop residues, G8, T9 and G10.

#### 3.2.3 Discussion

The results from the computational MD simulations of LTR-III with diazocine **42** remain inconclusive. However, several observations could still be made which included a combination of the importance of ions in stabilising the secondary structure, the impact of what the starting structure positions were and the length of the simulations. Additional controls would be essential to clarify some of these points for these MD simulations, such as running simulations that began with both  $K^+$  ions sites occupied within the central channel without any ligand and simulations with no  $K^+$  ions or ligand.

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

Potassium was important throughout these simulations at retaining stable structures. It could stabilise the G4 structure through chelation, causing the restriction of motion. There was a larger impact from the presence of potassium ions within the G-channel towards G-quadruplex stability than from any of the simulated ligands. While ions were not reported in the NMR structures of LTR-III, we know that 100 mM K<sup>+</sup> conditions lead to a 27.5 °C thermal stabilisation of LTR-III.<sup>53,57</sup> In our simulations, comparisons of tetrad stability with or without chelation to potassium strongly suggested that the G4 channel in LTR-III must also have potassium ions as reported throughout literature.

Even so, the starting structures of simulations had a strong impact on the pathway that the simulation took. Based on <sup>1</sup>H NMR experiments, the ligands were manually docked at the quadruplex-duplex junction of the 10 starting structures from the NMR models. These 10 conformations had wide or narrow junction openings and attempts to place the bulkier **Z-42** at that position were challenging, so comparisons between all of the simulations were difficult. Biased manual docking may have contributed to the expulsion of **Z-42** from the junction in Model 7Z, or the rapid unfolding of the G4 within 1 ns of the simulation in Model 8Z. Use of docking programs such as AutoDock<sup>111,199,200</sup> could provide better starting ligand placement and remove manual docking errors.<sup>201</sup>

Throughout the simulations, another major issue was the relatively short timescales of the simulations compared to the timescales of experimental ligand binding events. The RMSD plots of many of the repeats showed that more run time was required to reach a stable conformation. This is highlighted by reports of long G4 simulations (1-10 μs) that G4 simulations under 300 ns are still part of the equilibrium phase.<sup>188</sup> Aside from running controls which would allow the identification of which factors affect the stability of G4s, future work could consider further extending the simulations for Models with optimal ion placement such as Models 1Z and 7E (**Figure 3.38**). Principle component analysis could also be used to differentiate the conformations that could be forming over the course of the simulation.

## 3.3 Chapter conclusions

The binding affinities and effects on LTR-III G4 topology were studied with the newly synthesised diazocine ligands **42-46** by using various different biophysical techniques. Comparisons between the pyridinium ligands **42-44** and non-pyridinium ligands **45-46** showed that the non-pyridinium ligands had stronger affinities for nucleic acid structures. However, this higher affinity was due to the ligand substituent instead of the *E*-/*Z*-photoswitchable moiety, weakening their applicability as photoswitchable ligands. Two examples for comparison are 4-pyridinium **42** and 3-quinolinium **46**. The former has an impressive 34 fold *E*/*Z* ratio of association constants compared to a measly 2 fold difference for the latter.

UV-Vis and CD titrations could confirm that the ligands were able to isomerise by photoirradiation or thermally in the presence of DNA. And investigations into the ligands effect on LTR-III topology and structure allows us to tentatively suggest that the planar *E*-pyridinium diazocines interacted with the DNA in  $\pi$ -stacking interactions. As for the biaryl ligand **45**, the large induced CD effect, potentially excitonic coupling between the DNA and ligand suggested that there was groove binding.<sup>164</sup> The strong binding affinity of biaryl **45** was validated as the DNA-Ligand complex precipitated out of solution during CD titration experiments.

Following the investigations in binding affinities and the effect on DNA topology, we inspected how the <sup>1</sup>H NMR of LTR-III would be affected by the addition of the pyridinium ligands. It was important to consider the thermal isomerisation kinetics of *E*-**42** making it difficult to see any differences between *E* and *Z* at room temperature as *E*-**42** kept on reverting back to *Z*-**42**. So additional titrations with **42** and the more thermally stable **43** were performed at 5 °C and 25 °C respectively, giving strong suggestions that the planar ligand interacted most strongly at the quadruplex-duplex junction. Prompting computational analysis to build on this observation by running MD simulations with ligands *E*-**42** and *Z*-**42**.

These simulations were set up with both photoisomers of the ligand and manually docked at the junction. Unfortunately, the results from these simulations were not conclusive. Large fluctuations of the RMSD showed that the 255 ns simulation timescales was not long enough

### 3 Photoresponsive ligands targeting LTR-III HIV G-quadruplex

---

to explore all the conformations of dynamic structures like G-quadruplexes. For the few repeats that were the most stable, we could see that the ligands were interacting with the DNA with mostly  $\pi$ -stacking interactions.

Additionally, we were also interested to see if the changes that the G4 residues experience in the NMR titrations could be noticed by molecular dynamics simulations. This was hard to determine due to multiple factors. For example, certain residues were more prone to distortion than others. This was often solved by chelation to potassium ions in the central channel of the G4. However, this begs the question if the reported underestimation of hydrogen bond strength and overestimation of base stacking interactions of the force field whilst no potassium ions were present led to these destabilised structures in the first place.<sup>183</sup> Finally, some of the changes in the 10-12 ppm imino region of the DNA overlapped with each other. This made it harder to notice changes that occurred to residues that were overlapping, unlike other more well resolved signals. Thus any changes it experienced during the MD simulations could not be correlated with experimental results.

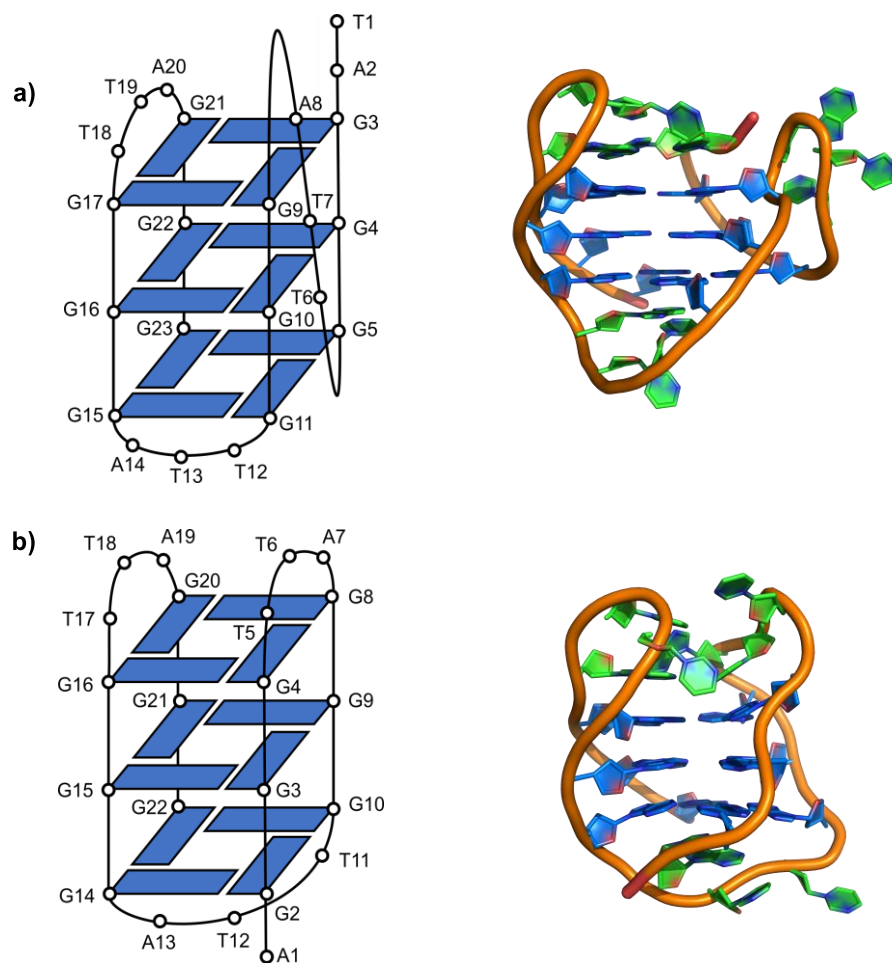
Yet some of the simulation repeats gave interesting results, with potential new insights into how the ligands can interact with the DNA. One of these was Model 7Z, where the bulky ligand was expelled from the junction. Matching experimental results as we do not expect Z-42 to remain at the quadruplex-duplex junction due to its lower binding affinity for LTR-III than the more linear E-42 and the lack of interactions inferred from NMR titrations.

### 4 Telomeric G-quadruplex Cancer Targets

While many forms of cancer therapies currently exist, such as chemotherapy, many of the current therapeutic strategies cause unpleasant side effects due to their modes of action and inability to just hit the desired therapeutic target. As mentioned in the introduction, the 3' overhang region of telomeric DNA is a single stranded guanine rich sequence,  $d(\text{GGGTTA})_n$ . This single stranded region is protected by a shelterin complex and reduces in length as the cell replicates, until it reaches the Hayflick limit resulting in cell senescence. However, 80-85% of cancer cells have upregulated telomerase activity, which maintains telomere length, and allows the cell continuous division cycles. Since the demonstration that stabilising G-quadruplex structures with small molecules leads to the indirect inhibition of telomerase activity (and eventual cell senescence and apoptosis),<sup>41</sup> there has been significant research into investigating new G4 ligands and their interactions with telomeric G4 structures as a novel anticancer strategy.<sup>36,66</sup>

The telomeric sequences are polymorphic and can form multiple different G4 conformations.<sup>202,203</sup> This polymorphic characteristic can be attributed to the exact nature of the sequence and the specific buffer conditions. Different flanking sidechains can result in different conformations of telomeric G-quadruplexes.<sup>204</sup> As for the buffer conditions, the concentration and identity of the central chelating cations are as important to G4 stabilisation as the sequence. Different cations are known to have different influences on the stability of G4s and will favour different G4 conformations.<sup>205</sup> The physiologically relevant monovalent cations are potassium and sodium and follow the trend  $\text{K}^+ > \text{Na}^+ > \text{Li}^+$ , where potassium ions offer the highest stabilisation of G4s. In particular, G4 conformations in the presence of  $\text{K}^+$  ions are considered biologically relevant due to a higher intracellular potassium concentration.

NMR studies have shown that in potassium rich conditions, the major conformation (~70%) of a 23-mer telomeric DNA sequence,  $d[\text{TAGGG}(\text{TTAGGG})_3]$ , (telo23) is the hybrid-1 conformation.<sup>206</sup> Whereas in sodium rich conditions, a 22-mer sequence,  $d[\text{AGGG}(\text{TTAGGG})_3]$ , (telo22) can fold into a basket antiparallel topology.<sup>207</sup> (**Figure 4.1**). Other variations in sequence length such as a 26-mer sequence in  $\text{K}^+$  favours the hybrid-2 conformation,<sup>204</sup> which is not discussed in this work.



**Figure 4.1:** Cartoon and NMR solution structures of **a) telo23** (PDB: 2JSM) and **b) telo22** (PDB: 143D). **a) telo23** folds into a hybrid-1 conformation in potassium buffer and **b) telo22** folds into an antiparallel basket conformation in sodium buffer. PDB structures show guanine residues in blue; thymine and adenine residues in green.

The two different conformations of telo23 and telo22 have unique loops and grooves which may offer different selectivity between different ligands.<sup>208,209</sup> Thus, we became interested to evaluate whether the diazocine ligands showed any differences in binding affinity and selectivity towards these sequences depending on the G4 conformation by using biophysical assays e.g. UV-Vis and CD spectroscopy as previously described.

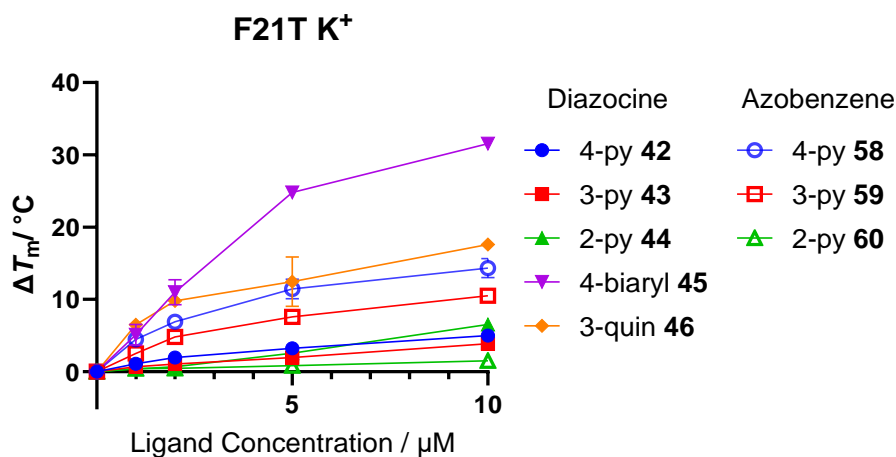
## 4.1 Biophysical Assays

### 4.1.1 FRET thermal melting assay

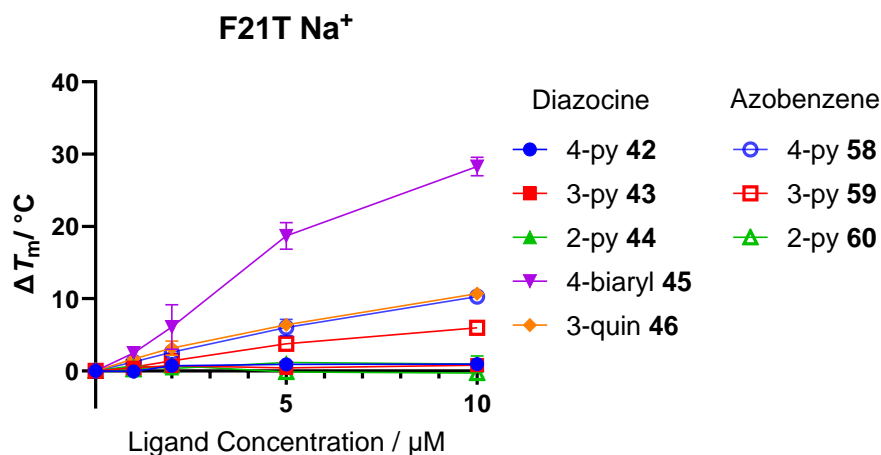
Knowing that telomeric G4s are polymorphic, it was interesting to see how the ligands could affect the stabilisation of the same G4 sequence in different buffers. Here, we use the

## 4 Telomeric G-quadruplex Cancer Targets

polymorphic sequence F21T, a 21 long nucleotide with a fluorescein tag and a rhodamine tag at the 5' and 3' ends.<sup>210</sup> The same diazocine and azobenzene ligands were employed in this melting assay with the cyclic *Z*-diazocine ligands **42-46** and the linear *E*-azobenzene ligands **58-60**.



**Figure 4.2:** Thermal stabilisation values ( $\Delta T_m$ ) for the FAM and TAMRA labelled F21T oligonucleotide sequence (200 nM) in potassium buffer (10 mM KCl, 90 mM LiCl, 10 mM Li cacodylate buffer, pH 7.2) with ligands **42-46** and **58-60**.



**Figure 4.3:** Thermal stabilisation values ( $\Delta T_m$ ) for the FAM and TAMRA labelled F21T oligonucleotide sequence (200 nM) in sodium buffer (100 mM NaCl, 10 mM Li cacodylate buffer, pH 7.2) with ligands **42-46** and **58-60**.

The results from the FRET melting assay can be seen in **Table 4.1** (Representative melting curves are shown in Appendix A4). We achieve very similar results for the stabilisation of the F21T-K<sup>+</sup> sequence when compared to the FLTR-IIIT sequence in the previous chapter. However, F21T-Na<sup>+</sup> has lower  $\Delta T_m$  values than F21T-K<sup>+</sup>, with a decrease of 3-7 °C across all the ligands.



## 4 Telomeric G-quadruplex Cancer Targets

This is particularly evident for the pyridinium diazocines which drop to a 1 °C stabilisation of the sequence at 10  $\mu\text{M}$  ligand concentrations, suggesting a lack of interaction (Entry 1-3).

**Table 4.1:** FRET thermal stabilisation values ( $\Delta T_m$ ) of FAM and TAMRA labelled G4 F21T sequences in potassium and sodium rich buffers. DNA concentration = 200 nM, diazocine and azobenzene ligands = 10  $\mu\text{M}$ .

		$\Delta T_m / ^\circ\text{C}$	
		F21T K <sup>+</sup>	F21T Na <sup>+</sup>
Entry	Ligand	10 $\mu\text{M}$	10 $\mu\text{M}$
<b>1</b>	<b>Z-42</b>	5 $\pm$ 0.2	1 $\pm$ 0.6
<b>2</b>	<b>Z-43</b>	4 $\pm$ 0.2	1 $\pm$ 0.2
<b>3</b>	<b>Z-44</b>	7 $\pm$ 0.7	1 $\pm$ 1.1
<b>4</b>	<b>Z-45</b>	32 $\pm$ 0.8	28 $\pm$ 1.3
<b>5</b>	<b>Z-46</b>	18 $\pm$ 0.6	11 $\pm$ 0.3
<b>6</b>	<b>E-58</b>	14 $\pm$ 1.3	10 $\pm$ 1.0
<b>7</b>	<b>E-59</b>	11 $\pm$ 0.4	6 $\pm$ 0.4
<b>8</b>	<b>E-60</b>	2 $\pm$ 0.3	0 $\pm$ 1.0

The ligands **45** and **46** with the larger substituents showed higher  $\Delta T_m$  values than the pyridinium ligands (Entry 4 and 5). However it is important to remember that the selectivity of biaryl **45** is poor as it was also shown to result in the stabilisation of the duplex stem loop F10T (**Table 3.1**). On the other hand, 3-quinolinium **46** only stabilises G4 structures and is among the strongest binders as it stabilises F21T-K<sup>+</sup> by 18 °C and F21T-Na<sup>+</sup> by 11 °C at a 10  $\mu\text{M}$  ligand concentration.

Within the pyridinium ligands, the *E*-azobenzenes stabilised the two telomeric structures more than the *Z*-diazocines following a trend of 4-py > 3-py > 2-py, with the 2-pyridinium **44** and **60** ligands once again being the exception. This continues to support our hypothesis that the more planar *E*-azobenzenes will have stronger interactions with G4 structures than the bent cyclic *Z*-diazocines.

## 4 Telomeric G-quadruplex Cancer Targets

---

These results still need to be treated with caution as the structural differences between *E*-azobenzenes and *E*-diazocines could still provide differing results. This can be confirmed further using other biophysical techniques such as UV-Vis and CD spectroscopy which are performed at room temperature with more regular irradiation of the ligand samples to maintain high *E*-diazocine concentrations in its binding to nucleic acid structures.

### 4.1.2 UV-Vis titrations

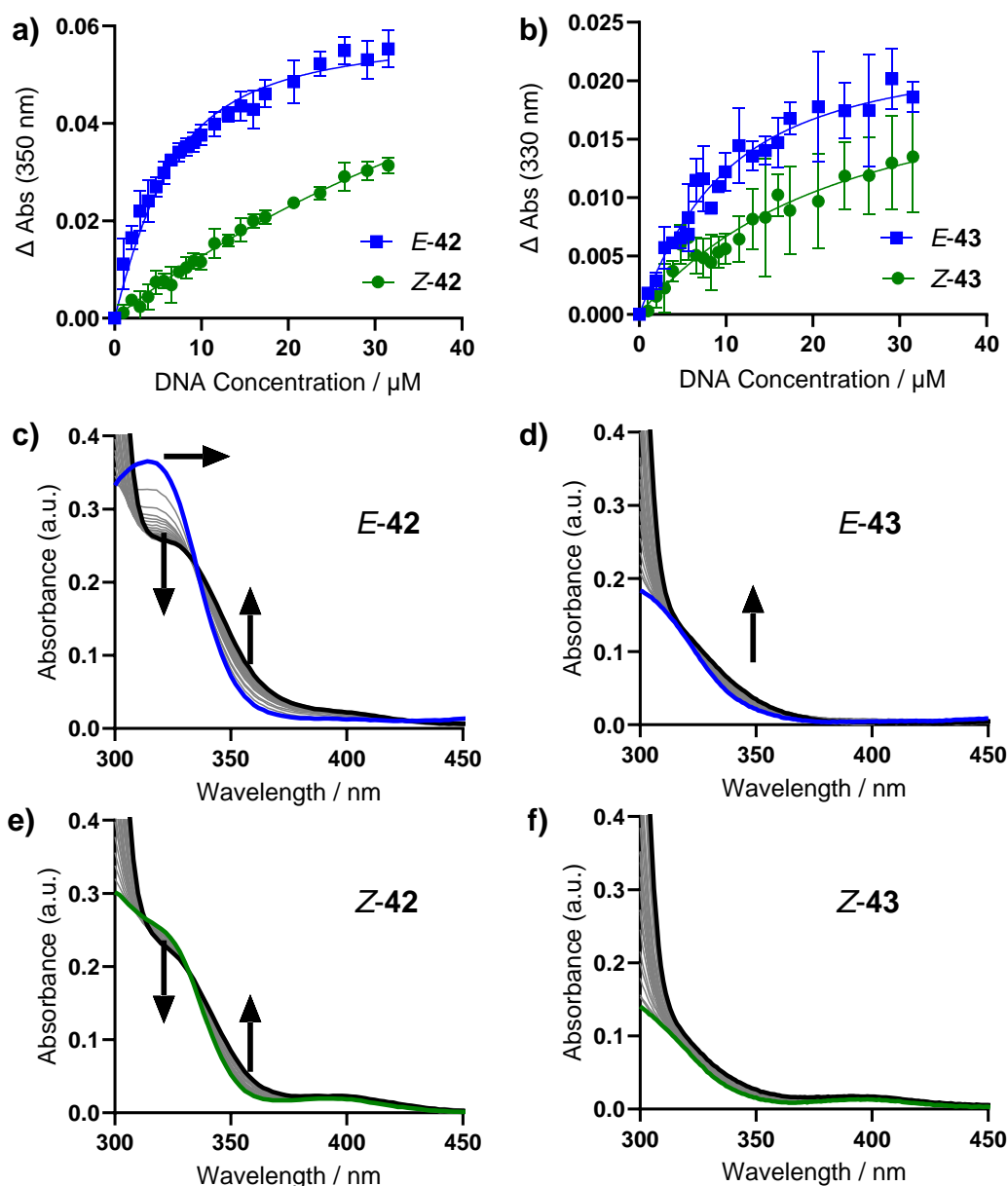
Numerically comparing the binding affinity of *E*- and *Z*-diazocines through UV-Vis titrations allowed us to see whether the ligands will also bind to the more typical G4 structures as well as it did with LTR-III. In these studies, instead of using the 21-mer sequence, we used the 23-mer (telo23) and 22-mer (telo22) sequences since they have been shown in literature to fold into two different conformations just by altering the buffer conditions.

#### 4.1.2.1 UV-Vis titrations with telo23

We can first look at the titrations with the telo23 sequence in potassium buffer which has a hybrid conformation. The titrations with the 2-pyridinium ligands are not included as they did not result in large enough changes in the ligand absorption spectra, resulting in no meaningful association constants that could be extracted.

**Table 4.2:** Association constants,  $K_a$ , and *E*-/*Z*- ratios for pyridinium diazocines **42** and **43** and pyridinium azobenzenes **58** and **59** calculated from the binding isotherms with telo23 in 100 mM potassium buffer conditions (pH 7.4), N=2.

Ligand	$K_a / 10^6 \text{ M}^{-1}$			$\lambda / \text{nm}$
	<i>E</i> -	<i>Z</i> -	<i>E</i> -/ <i>Z</i> -	
<b>42</b>	$0.15 \pm 0.02$	$0.0070 \pm 0.002$	21	350
<b>43</b>	$0.074 \pm 0.02$	$0.026 \pm 0.01$	2.9	330
<b>58</b>	$0.51 \pm 0.04$	-	-	400
<b>59</b>	$0.25 \pm 0.01$	-	-	390



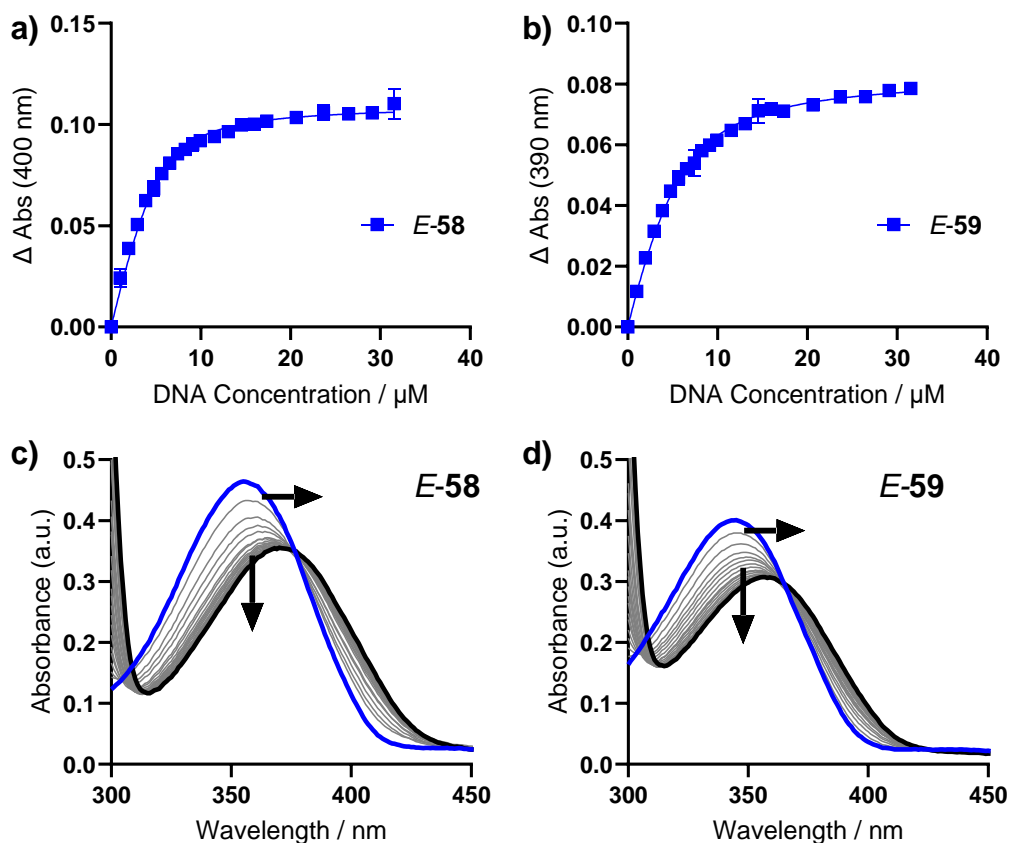
**Figure 4.4:** UV-Vis titration of 10  $\mu\text{M}$  diazocines **42** and **43** with increasing concentrations (0-31  $\mu\text{M}$ ) of telo23 in 100 mM potassium phosphate buffer (pH 7.4). **a-b)** Extracted binding isotherms of **a)** **42** and **b)** **43**. **c-f)** Titrations of 10  $\mu\text{M}$  **c),e)** **42** and **d),f)** **43** with telo23 G4. Blue or green line represents 10  $\mu\text{M}$  ligand with no DNA added, black line represents addition of 3 equivalents of DNA.

**Table 4.2** shows the summary of the association constants for the titration of the pyridinium ligands with the binding isotherms shown in **Figure 4.4** and **Figure 4.5**. It was clear that the  $E$ -diazocine ligands had higher association constants than their  $Z$ -isomers, demonstrating that the structural differences impacted upon binding affinity. The linear  $E$ -azobenzenes (which were assumed to have a 100:0  $E$ : $Z$  ratio) demonstrated even higher binding affinities to telo23

## 4 Telomeric G-quadruplex Cancer Targets

than the *E*-diazocines. We also observed hypochromic and bathochromic shifts for the  $\pi$ - $\pi^*$  bands of the *E*-azobenzenes, suggesting that the ligands had  $\pi$ -stacking interactions with the hybrid G4 (**Figure 4.5**). Unfortunately, the overall  $\Delta A$  in the titration of 3-pyridinium **43** was smaller and had relatively larger errors, signifying weaker interactions that *E*-**43** makes with telo23 (**Figure 4.4b**).

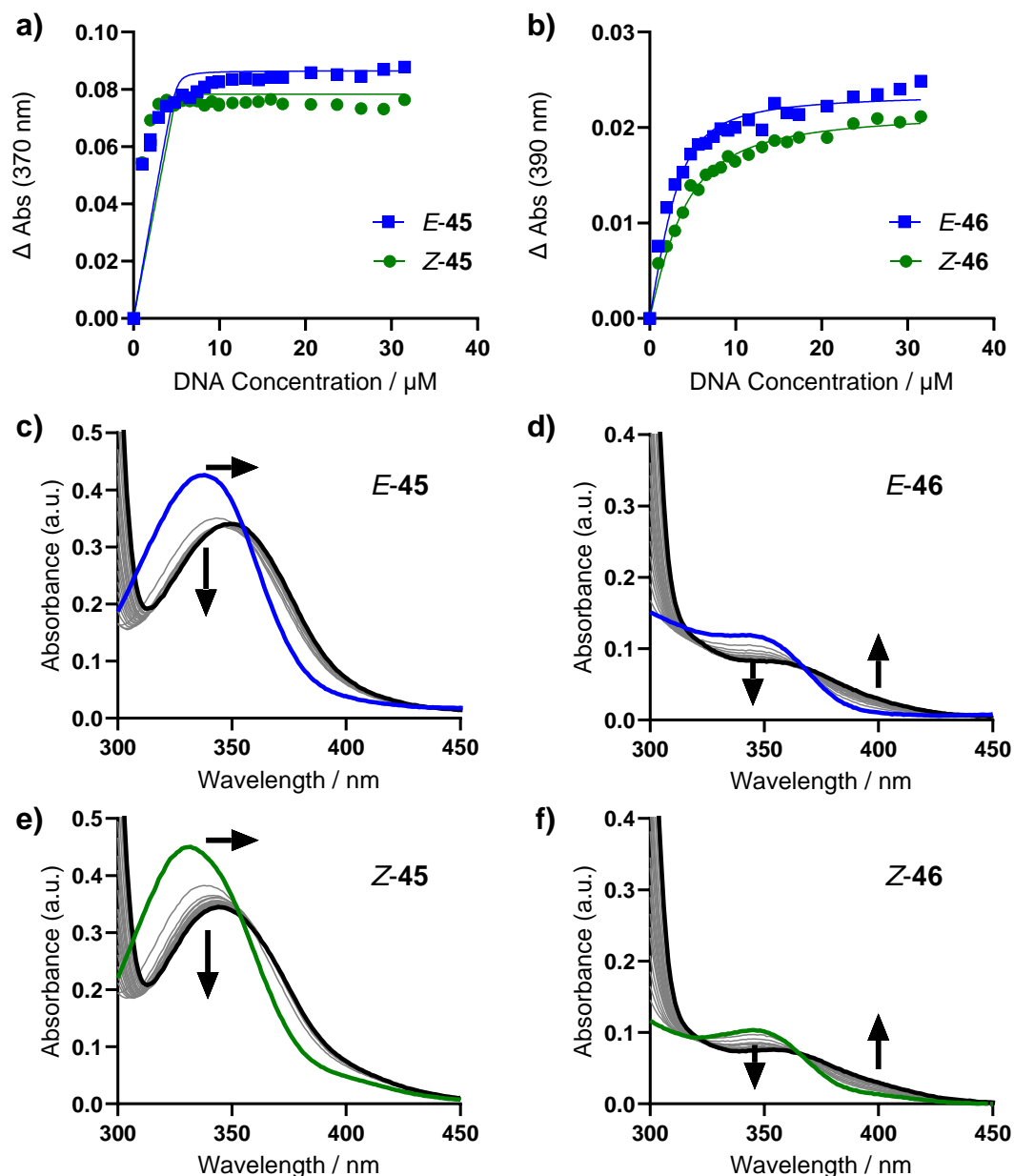
We also observed that the association constants ( $K_a$ ) for all ligands were lower than the values achieved with LTR-III (**Table 3.2**). Additionally, the *E/Z* ratios of the  $K_a$  values for the two diazocine ligands are smaller than in the titrations with LTR-III. These two results may suggest that the unique quadruplex-duplex conformation of LTR-III potentially creates more ligand binding sites that favour planar *E*-ligands than the hybrid telo23 G4.



**Figure 4.5:** UV-Vis titration of *E*-azobenzene **58** and **59** with telo23 in 100 mM potassium phosphate buffer (pH 7.4). **a-b**) Extracted binding isotherms of **a)** **58** and **b)** **59**. **c-d**) Titrations of 10  $\mu$ M **c)** **42** and **d)** **43** with telo23 G4. Blue line represents 10  $\mu$ M ligand with no DNA added, black line represents addition of 3 equivalents of DNA. Clear hypochromic and bathochromic shifts are observed.

## 4 Telomeric G-quadruplex Cancer Targets

As with the titrations with LTR-III, the ligands with larger substituents showed stronger binding affinities which can be seen in the quick saturation of the ligand upon the addition of telo23 DNA (Figure 4.6). Once again, this suggested that the initial concentration at which the titration with biaryl 45 was performed at was too high, leading to poorly fitted binding isotherms where no meaningful data can be extracted.<sup>159</sup>



**Figure 4.6:** UV-Vis titration of diazocines 45 and 46 with telo23 in 100 mM potassium phosphate buffer (pH 7.4). a-b) Extracted binding isotherms of a) 45 and b) 46. c-f) Titrations of 10  $\mu$ M c), e) 45 and d), f) 46 with LTR-III G4. Blue or green line represents 10  $\mu$ M ligand with no DNA added, black line represents addition of 3 equivalents of DNA.

## 4 Telomeric G-quadruplex Cancer Targets

Finally, the 3-quinolinium **46** ligand was a weaker binder than biaryl **45**, so association constants could be extracted from the *E*- and *Z*- titrations ( $K_a = 1.0 \times 10^6 \text{ M}^{-1}$  and  $0.40 \times 10^6 \text{ M}^{-1}$ ), showing a 2.5 fold difference between the two isomers. Following the same trend as the pyridinium ligands, it had a smaller  $K_a$  when compared with the binding to LTR-III.

### 4.1.2.2 UV-Vis titrations with telo22

To see how a different G4 conformation would impact ligand binding, we carried out titrations with telo22 only with the 4- and 3-pyridinium ligands. The binding isotherms are shown in **Figure 4.7** with results in **Table 4.3**. The trends from these titrations with telo22 are the same as telo23. The two diazocines continued to show that the linear *E*-ligands demonstrated higher binding affinity to the G4 than *Z*-ligands. Though the binding affinities of the ligands with telo22 were lower than with telo23, with 2 fold difference for *E*-**42**, a 3 fold difference for *E*-**58** and a 4 fold difference with *E*-**59**.

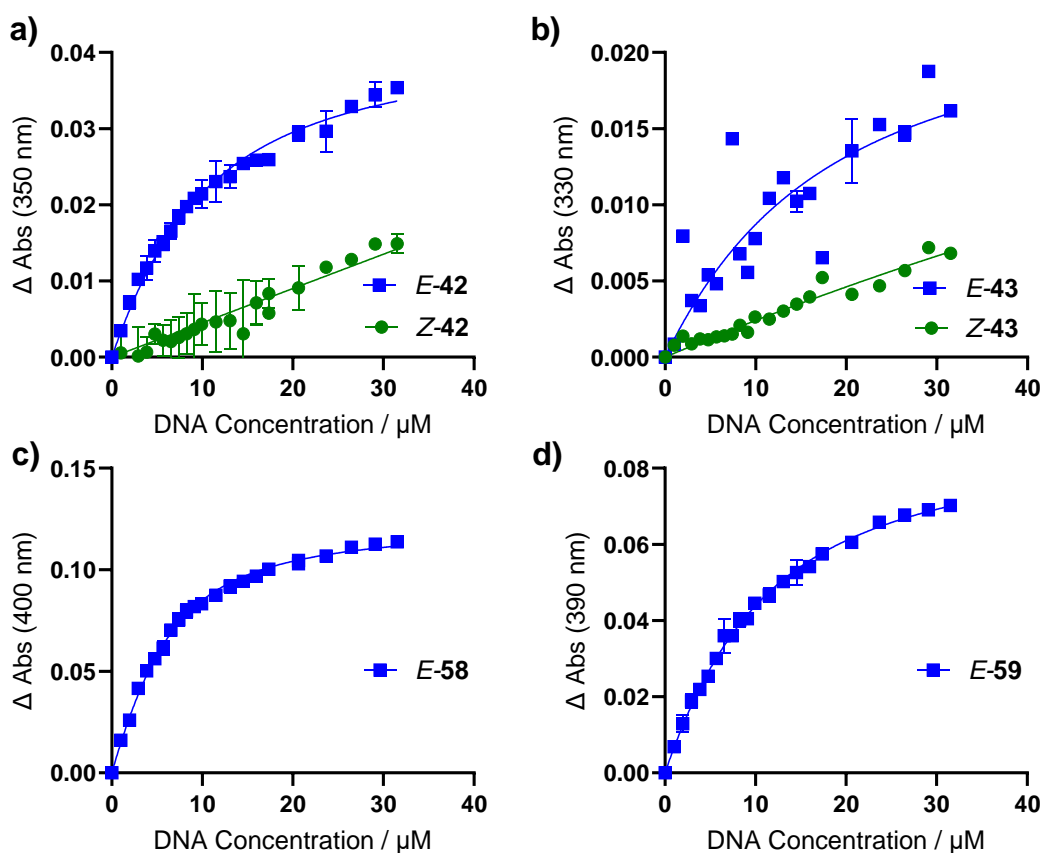
**Table 4.3:** Association constants,  $K_a$ , and *E*-/*Z*- ratios for pyridinium diazocines **42** and **43** and pyridinium azobenzenes **58** and **59** calculated from the binding isotherms with telo22 in 100 mM sodium phosphate buffer conditions (pH 7.4), N=2.

Ligand	$K_a / 10^6 \text{ M}^{-1}$			$\lambda / \text{nm}$
	<i>E</i> -	<i>Z</i> -	<i>E</i> -/ <i>Z</i> -	
<b>42</b>	$0.069 \pm 0.006$	$0.0029 \pm 0.002$	23	350
<b>43</b>	$0.034 \pm 0.02$	$0.0024 \pm 0.003$	14	330
<b>58</b>	$0.16 \pm 0.008$	-	-	400
<b>59</b>	$0.062 \pm 0.003$	-	-	390

Large errors in the titration with 3-pyridinium **43** were noted again, as previously seen in the titration with telo23 (**Figure 4.4b** and **Figure 4.7b**). The large variations in measurement are most likely driven by the poorer interactions of the 3-pyridinium (and 2-pyridinium) substituent with G4s. Changing the position of the methylated nitrogen in the pyridine ring resulted in a decrease of the binding affinity of the ligand to the G4. This could be explained as the pyridinium ring rotating out of plane to accommodate the methyl group as it changes from *para*- to *meta*- to *ortho*-. The decreased planarity of the ligand reduced the optimised stacking interactions with the G-tetrads and decreased the binding affinity as a result. This

## 4 Telomeric G-quadruplex Cancer Targets

effect can be seen with a series of porphyrin ligands, TMPyP4, TMPyP3 and TMPyP2 with a bcl2 G4 sequence.<sup>211</sup>



**Figure 4.7:** UV-Vis binding isotherms titration of 10  $\mu$ M **a-b)** diazocines **42** and **43** and **c-d)** azobenzenes **58** and **59** with telo22 in 100 mM sodium phosphate buffer (pH 7.4).

In both azobenzene and diazocine titrations with telo23 and telo22, we see hypochromic and bathochromic shifts. These can be calculated for the azobenzene ligands as their  $\pi$ - $\pi^*$  absorption band does not overlap with the DNA absorption (<320 nm). We achieve very similar changes for both ligands with both G4 structures, hypochromic shifts of 18-24% and bathochromic shifts of 10-15 nm (**Table 4.4**). The similar shifts of  $\pi$ - $\pi^*$  absorption band, suggests that the two ligands are likely interacting with telomeric G4s in the same manner (either both  $\pi$ -stacking or groove binding), thus affecting the electronics of each ligand similarly. On the other hand, the different binding affinities suggest there should be differences in the interactions of the ligands and the telomeric G4s. This could be the different loop residues of the hybrid and antiparallel structures forming additional binding interactions to improve the affinity of the azobenzene ligand with the G4. There is a possibility that the changes in the absorption bands are due to the binding just to the central azobenzene core.

## 4 Telomeric G-quadruplex Cancer Targets

And changes in binding affinity are due to the changes in side-chain substituent which have a lower contribution to the absorption spectra.

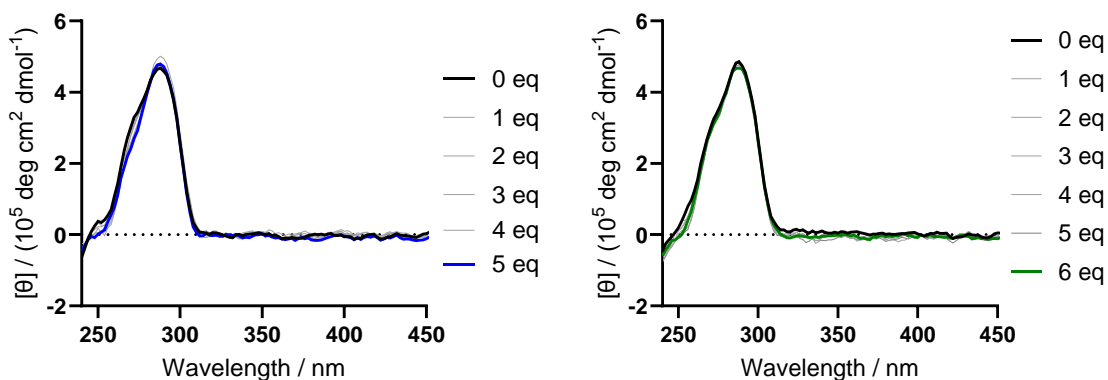
**Table 4.4:** Spectral shifts of *E*-azobenzene ligands **58** and **59** (10  $\mu$ M) with telo23 and telo22 sequences after 3 equivalents of DNA.

Ligand	G4	Bathochromicity / nm	Hypochromicity / %
<i>E</i> - <b>58</b>	telo23	15	24
	telo22	14	23
<i>E</i> - <b>59</b>	telo23	12	24
	telo22	10	18

Overall, the binding affinities of all the diazocines is strongly dependent on the side-chain substituent. The larger substituents can make more aromatic  $\pi$ -stacking interactions, at the cost of selectivity towards G4s and ignores the impact that the diazocine *E/Z* structural differences may interact with G4s. The 4-pyridinium ligands represent a happy “Goldilocks zone” which show both selectivity for G4s and have significant *E/Z* differences. This however, does not explain why there is a decrease in the binding affinity between the hybrid telo23 structure and antiparallel telo22 structure, and other methods are required to gain more information about these systems.

### 4.1.3 Circular Dichroism Titrations

As it had shown good binding affinity for telomeric G4s and also the largest *E/Z*- ratio, we focused on CD titration studies of the 4-pyridinium ligands diazocine **42** and azobenzene **58**.

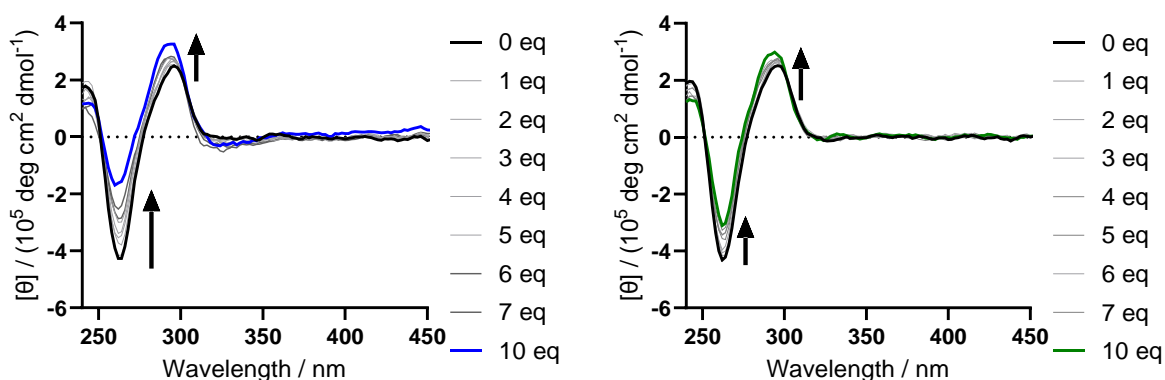


**Figure 4.8:** CD titration of **42** with telo23 at 20 °C in a 100 mM potassium phosphate buffer (pH 7.4). Left: *E*-**42**, Right: *Z*-**42**.



## 4 Telomeric G-quadruplex Cancer Targets

The telo23 CD spectrum shows a positive peak at 295 nm and a shoulder around 265 nm, characteristic of the hybrid conformation of telo23 (**Figure 4.8**).<sup>160,200,212</sup> The addition of 4-pyridinium **42** to the hybrid-1 G4 in potassium buffer does not lead to any significant changes to the CD spectrum for the DNA region. Beyond 320 nm, there is also no observed induced CD bands. We do see a slight difference between the two isomers, with the addition of *E*-**42** leading to a loss of the shoulder peak at ~265 nm. This shoulder peak has often been reported to be a minor conformation of the telo23 sequence in potassium buffers.<sup>213</sup> As this change is not observed in the addition of *Z*-**42**, this confirms that the *Z*-ligand has the lower binding affinity between the two and suggests that the hybrid-1 conformation in potassium is very stable, where the minor conformation is driven to the hybrid-1 topology.

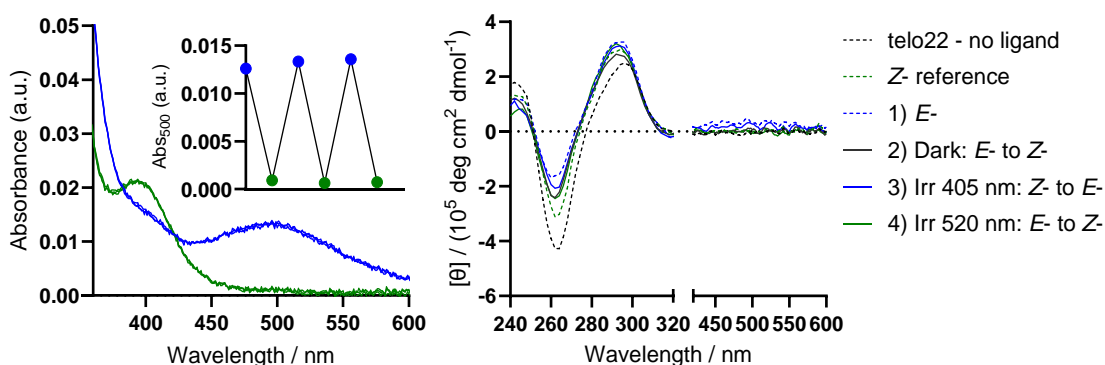


**Figure 4.9:** CD titration of **42** with telo22 at 20 °C in a 100 mM sodium phosphate buffer (pH 7.4). Left: *E*-**42**, Right: *Z*-**42**.

In contrast, **Figure 4.9** shows the titration of diazocine **42** with telo22 in sodium buffer. The initial antiparallel structure is characterised by a negative band at 260 nm and positive bands at 295 and 245 nm.<sup>200</sup> We see that both ligands are affecting the G4 topology and the *E*-**42** titration incurs larger spectra shifts and induced CD bands at 330 and 450 nm, supporting our expectation that *E*-diazocines are interacting more strongly. Both ligands cause a hypochromic shift of the negative band but also cause a hyperchromic shift at the positive band. The combination of these changes suggests a shift towards more parallel structures including the hybrid conformations. Despite the weaker binding affinity, there are a few possibilities for why the changes in the DNA region of the CD spectrum is larger for telo22. One possibility is that the ligand can bind at a site that perturbs the structure quite significantly which does not exist in telo23. Another possibility relates to the known effect that K<sup>+</sup> ions can stabilise G4s more

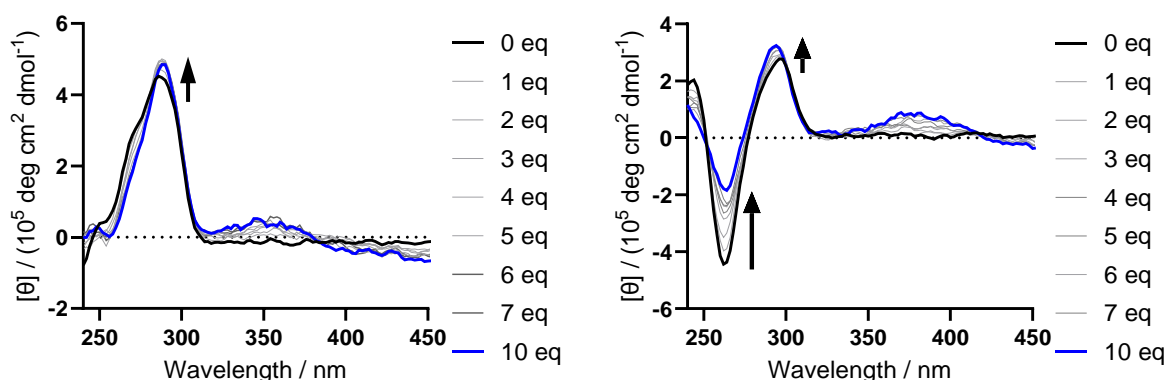
## 4 Telomeric G-quadruplex Cancer Targets

than  $\text{Na}^+$  ions. So the effect the ligand has upon the sodium stabilised structure seems to be stronger, as the whole structure is held together more weakly.



**Figure 4.10:** Photoswitching of **42** and telo22 in 100 mM sodium phosphate buffer. **Left:** UV-Vis absorption of 10 μM **42** in the presence of 3 equiv of telo22 with alternating 405 nm and 520 nm irradiation. Insert shows change in absorption at 500 nm. **Right:** CD spectrum of telo22 in the presence of 10 equiv of **42**. Photoswitching of ICD of *E*-**42** ~500 nm.

Photoswitching of the diazocine ligands in the presence of DNA was also possible and was observed by monitoring the changes of the  $n-\pi^*$  band around 500 nm for both UV-Vis and CD spectroscopy with presence of *E*-**42**. Additionally, the differences between *E*- and *Z*- titrations effect on the DNA structure are shown most clearly at 265 nm in the CD spectrum (**Figure 4.10 Right**). Ligand free conditions show the strongest absorption (black dashed). Irradiation of the DNA-Ligand sample with 405 nm light causes a hypochromic shift of the negative band (blue), and thermal relaxation (black solid) or irradiation with 520 nm light (green solid) results in a relative hyperchromic shift of the negative band.



**Figure 4.11:** CD titration at 20 °C of **Left:** telo23 in 100 mM potassium phosphate buffer (pH 7.4) with *E*-58. **Right:** telo22 in 100 mM sodium buffer (pH 7.4) with *E*-58.

**Figure 4.11** shows both azobenzene **58** titrations with telo23 and telo22. Given the larger binding affinity of the azobenzene ligands in general, we see similar but larger changes to the CD spectra of the two G4 structures, such as the loss of the shoulder peak in telo23 and the hypochromic shift of the negative band in telo22. Induced CD bands are also observed for both hybrid and antiparallel G4s, indicating that there is ligand interaction occurring.

Previous work by Xing *et al.*, which used azobenzenes with more flexible alkyl substituents (**Figure 1.16**) as ligands for G4s provided a helpful benchmark to compare the affinity of the pyridinium azobenzenes.<sup>97,102</sup> For example, they used a 24-mer telomeric sequence d(TTAGGG)<sub>4</sub> which adopts a hybrid structure in potassium buffer conditions.<sup>3</sup> The addition of azobenzene ligands resulted in the loss of the shoulder peak ~265 nm, and an appearance of a negative peak around 260 nm. Seemingly indicating that the ligand shifted the conformational equilibrium towards an antiparallel structure.<sup>16</sup> This behaviour has been observed with other ligands added to hybrid telomeric G4s, with larger shifts towards antiparallel conformations observed at lower K<sup>+</sup> concentrations.<sup>214</sup> Subsequent photoisomerisation of the ligand caused the sequence to readopt the hybrid structure, implying that the *E*-azobenzene ligands induced antiparallel topologies.

These conformational changes may also be occurring in the titrations of *E*-**58** with telo23. We see a loss of the shoulder at 270 nm and the appearance of what is similar to the negative band in antiparallel G4s around 260 nm after 10 equiv of ligand addition. It is also possible that the antiparallel structure of telo22 is shifted towards a more hybrid structure, as we observe a loss of the characteristic negative band and an increase of the positive band at 295 nm. Both sequences may be tending towards the same topology which may require long equilibration times to observe any changes by circular dichroism.

## 4.2 Chapter Conclusions

To conclude, we demonstrated that diazocines will also bind to more classical telomeric G4s. The binding affinity to the hybrid telo23 sequence in potassium and antiparallel telo22 sequence in sodium were weaker than the binding affinity to the LTR-III structure but the general trends observed in the biophysical assays with LTR-III remained the same with the telomeric cancer G4 sequences. For example, larger non-pyridinium substituents stabilised

#### 4 Telomeric G-quadruplex Cancer Targets

---

telomeric G4 structures better than ligands with pyridinium substituents. And the position of the methyl pyridinium moiety effected the binding affinity of the ligands in the order 4-py > 3-py > 2-py. We also demonstrated that the ligand could isomerise in the presence of G4 DNA by observing changes in the  $n-\pi^*$  bands for UV-VIS and CD.

However, the higher  $K_a$  values for the hybrid conformation versus the antiparallel conformations was not fully understood. Especially since CD did not show much interaction of *E*-**42** with the hybrid telo23 sequence. Additionally, while 4-pyridinium **42** had a 21 and 23 fold difference in *E*-/*Z*-association constants for telo23 and telo22. Higher *E*-ligand binding affinity did not always correlate with larger changes in CD spectra. This showed that ligands may happily bind to a structure without effecting its conformation and that the larger spectral shifts in the telo22 titrations are likely driven by the weaker stabilisation provided by the sodium ions.

Finally, the CD titration of telomeric G4s with the 4-pyridinium linear azobenzene analogue **58** led to interesting observations abouts its effect on the DNA topology as it seemed as if the two structures were converging upon a more similar conformation, which was supported by literature findings.<sup>97,214</sup>

## 5 Targeting *E. coli* G4s

Relatively little work has been conducted on bacterial G4 research. Early work in 2006 initially identified putative quadruplex forming sequences (PQS) upstream of regulatory regions.<sup>60</sup> More recent work in 2019 showed that many predicted bacteria G4s do not form. However, it also noted that a high proportion of the predicted G4s that did not form in the presence of  $K^+$ , were stabilised after the addition of the G4 ligand pyridostatin (**Figure 5.1**). Suggesting that targeting these bacterial G4s with G4 ligands could be a potential new mode of antibacterial therapy.<sup>62</sup>

Work conducted by Dr Jenny Samphire (PhD student in the Galan group) previously screened various G4 ligands to test for antibacterial properties against various drug resistant bacterial strains.<sup>215</sup> The ligand *E-58* (**Figure 5.1**) was selective for various *E. coli* strains and exhibited strong antibacterial activity, with low minimum inhibitory concentrations (MIC) of 2-4  $\mu\text{g/mL}$ . It was proposed that the antimicrobial activity observed might be attributed to the ability of this compound to target microbial G4 sequences and it was thus decided to further investigate this hypothesis. For comparison, the analogous stiff-stilbene *E-61*, a well-known G4 ligand,<sup>109,110,118</sup> which had higher MICs of 64-128  $\mu\text{g/mL}$  against the same *E. coli* strains, was selected due to structural similarities between the two molecules. And finally, the known G4 ligand pyridostatin was also selected to use as a positive control.

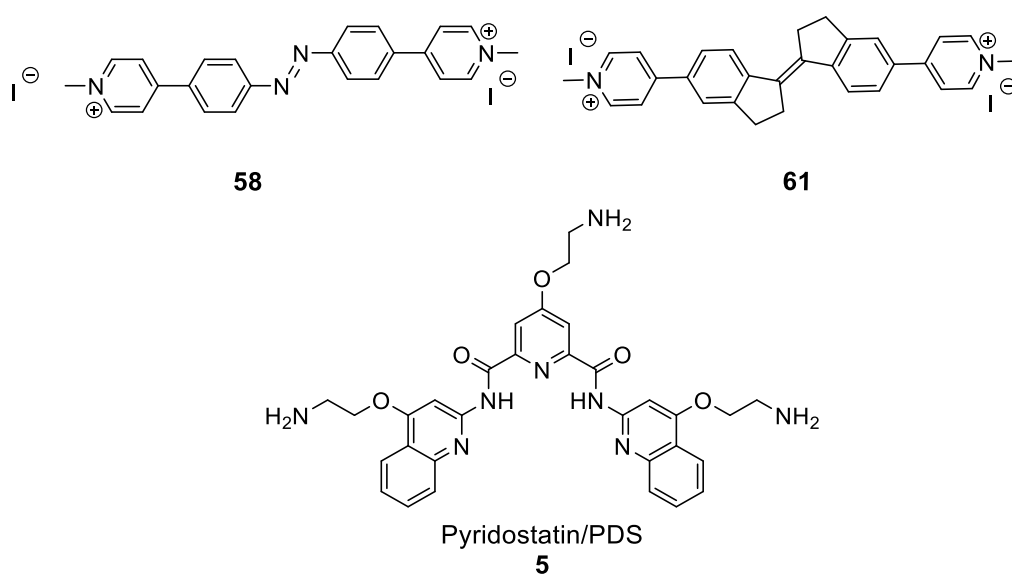


Figure 5.1: G4 ligands used in *E. coli* G4 studies.

## 5 Targeting *E. coli* G4s

---

To identify what G4 sequences were being targeted with these ligands, the proteomes of *E. coli* cells treated with **58**, **61** or pyridostatin were analysed by TMT-labelled LC-MS/MS. Proteomics analysis (carried out by Dr Yuiko Takebayashi) showed changes in the upregulation and downregulation of multiple essential proteins such as those involved in transcription and translation. Based on prior literature from Marsico *et al.*<sup>62</sup>, G4 sequences which were associated with the up- and down-regulated proteins were selected, and their potential to form G4 structures *in vitro* were scored using G4Hunter<sup>19</sup>, a predictive algorithm tool. This resulted in seven sequences (clpB, dppA, glnD, hemL, pdxA, thrA and yjbD) being selected for further investigation.

In addition, up- and downregulated gene sequences were further queried by the bioinformatics tool, G4-iM-Grinder, which is an algorithm that scores G4 formation propensity by using machine learning and accounts for long loops or bulges that may otherwise interrupt a G-tract in a stable G4.<sup>18,216</sup> This latest search highlighted a further ten PQS (asnS, yhaJ, ybiO, aroL, cobS, mtgA, maeA, yfiC, creA and mukE). A total of 17 sequences were identified *in silico*, however biophysical assays were required to determine whether these sequences could fold into G4 structures and if so, whether the three key ligands, azobenzene **58**, stiff-stilbene **61** and pyridostatin could interact with those G4 oligonucleotide sequences.

### 5.1 Biophysical studies on microbial FRET labelled sequences

To validate whether or not the 17 new sequences would fold into G4s, FRET labelled sequences were dissolved in increasing concentrations of KCl buffer solutions (0, 1, 10, 50,100 mM). Pleasingly, a majority of the structures formed characteristic G4 topologies identifiable by CD spectroscopy (**Figure 5.2** and Appendix A5.1-A5.3). Certain sequences retained the same spectra, regardless of the potassium ion concentration. Other sequences had topologies that shifted upon increasing potassium ion concentrations. Examples included hyperchromic shifts of the absorption bands, or a switch in topology (eg. dppA, mukE, yfiC, yjbD). The main conformation of the quadruplex could be quantified using equation 5.1 below, which calculated a ratio of the two absorption bands at 265 and 290 nm. Values (*r*) above 0.5 resembled parallel G4s, values between 0 and 0.5 were hybrid and values below 0 were antiparallel and are listed in **Table 5.1** at each different buffer condition.

$$r = \frac{CD_{265}}{|CD_{265}| + CD_{290}} \quad (5.1)$$

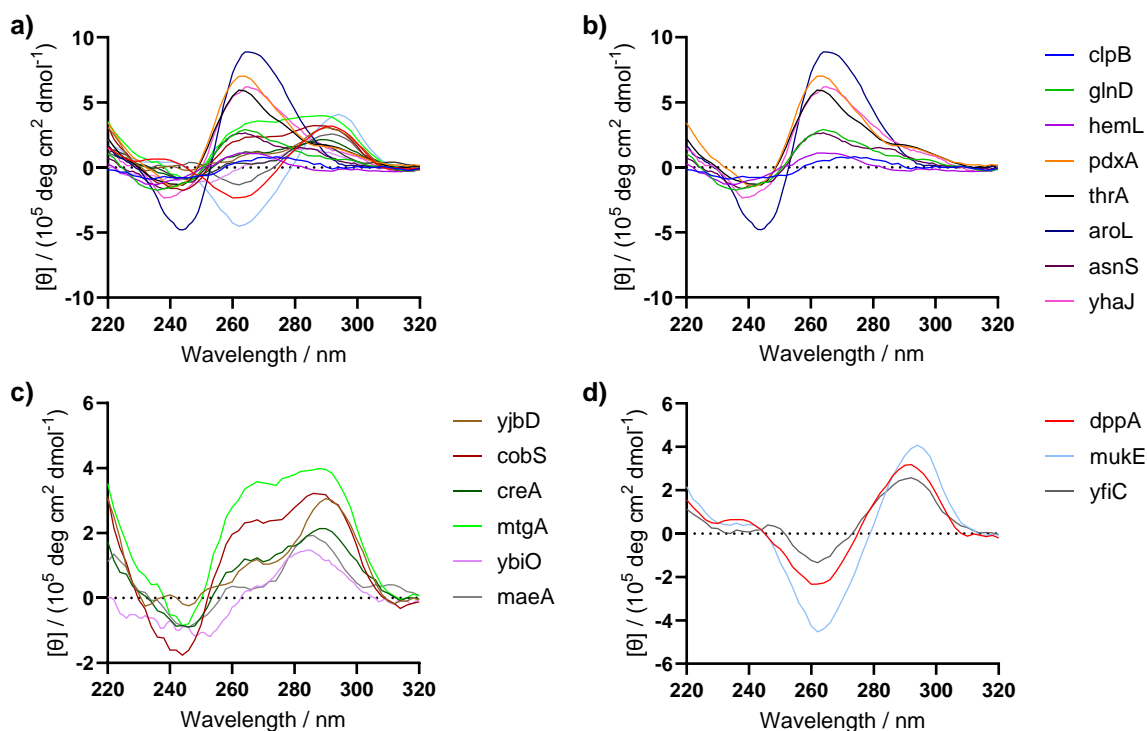
**Table 5.1:** Metric to determine G4 conformation at different KCl concentrations.  $r \geq 0.5$ ,  $0 \leq r < 0.5$ , and  $r < 0$  correspond to predominantly parallel, hybrid, and antiparallel topologies, respectively.  $T_m$  of FAM and TAMRA labelled sequences in 100 mM KCl, 10 mM Li cacodylate buffer (pH 7.2).

Sequence	G4 Conformation Ratio, $r$				Melting Temperature
	K <sup>+</sup> ion concentration				$T_m / ^\circ\text{C}$
	0 mM	10 mM	50 mM	100 mM	(100 mM KCl)
1) <b>clpB</b>	2.63	1.85	2.55	0.94	64 ± 0.1
2) <b>dppA</b>	0.59	0.45	-0.37	-0.41	51 ± 5.6
3) <b>glnD</b>	1.00	0.93	0.89	0.80	57 ± 0.1
4) <b>hemL</b>	2.40	1.15	1.30	1.30	59 ± 0.1
5) <b>pdxA</b>	0.97	0.83	0.82	0.81	56 ± 3.0
6) <b>thrA</b>	0.85	0.82	0.77	0.77	56 ± 0.1
7) <b>yjbD</b>	0.97	0.39	0.26	0.25	53 ± 0.3
8) <b>asnS</b>	0.80	0.75	0.73	0.75	54 ± 0.7
9) <b>yhaJ</b>	0.77	0.83	0.86	0.80	59 ± 0.4
10) <b>ybiO</b>	0.26	0.32	0.14	0.15	63 ± 1.4
11) <b>aroL</b>	0.92	0.91	0.93	0.90	40 ± 0.2
12) <b>cobS</b>	0.36	0.45	0.40	0.42	58 ± 0.2
13) <b>mtgA</b>	0.32	0.37	0.43	0.46	54 ± 0.5
14) <b>maeA</b>	-0.15	-0.39	-0.10	0.15	49 ± 0.7
15) <b>yfiC</b>	0.94	0.74	-0.16	-0.31	45 ± 0.4
16) <b>creA</b>	0.69	0.64	0.40	0.35	59 ± 1.7
17) <b>mukE</b>	0.77	-0.39	-0.52	-0.54	53 ± 0.2

**Figure 5.2** shows the CD spectra of the sequences at 100 mM KCl. A majority adopted parallel G4s with differing intensities (**Figure 5.2b**). These had a dominant positive band at 265 nm, and a negative band at 245 nm. Many also had a shoulder at 295 nm, suggesting other minor

## 5 Targeting E. coli G4s

conformations were present. Three sequences adopted antiparallel G4s in high K<sup>+</sup> concentrations, having switched from hybrid (dppA), parallel (yfiC) or non-G4 (mukE) structures at 0 mM KCl (**Figure 5.2d**, Appendix A5).



**Figure 5.2:** CD spectra of FAM and TAMRA labelled DNA sequences at 2.5  $\mu$ M concentration in 100 mM KCl and 10 mM lithium cacodylate buffer, pH 7.2. **a)** CD of all 17 sequences. **b)** CD of parallel sequences **c)** CD of hybrid or mixed conformations. **d)** CD of antiparallel sequences.

Sequences that were categorised as hybrid G4s featured more complex spectra as they could either be the result of a single intramolecular 3+1 hybrid G4 or mixtures of parallel, antiparallel and hybrid G4s at different conformational populations (**Figure 5.2c**). For example, cobS and mtgA ( $r=0.42$  and  $0.46$ ) seemed to have equal populations of hybrid and parallel structures in solution. On the other hand, maeA was potentially a mix of parallel and antiparallel G4 structures. It lacked the characteristic positive and negative bands at 265 nm, which would be cancelling each other out. Finally, the yjbD sequence is also noteworthy as it too initially begins as a non-G4 structure at lower KCl concentrations, before ultimately folding into a hybrid G4.



## 5 Targeting *E. coli* G4s

---

We then used a FRET melting assay to assess the thermal stability of these sequences. The melting temperatures in different concentrations of KCl buffer was measured with the  $T_m$  at 100 mM KCl ranging from 40-64 °C.

Most sequences did not show any changes in the melting temperature with increasing  $K^+$  concentration, the  $T_m$  in the absence of KCl was the same as the  $T_m$  in 100 mM KCl. The mukE and yjbD sequences were exceptions and showed increased  $T_m$  at higher potassium concentrations since these sequences did not initially form G4s in the absence of KCl, relying on  $K^+$  ions to fold in to a G4 structure (mukE:  $\Delta T_m = 5$  °C and yjbD:  $\Delta T_m = 8$  °C). The opposite effect was observed for some other sequences (thrA and pdxA), which had a decrease in  $T_m$  with increasing KCl concentration, but this could not be easily explained.

### 5.1.1 Ligand stabilisation

Next, we utilised a FRET melting assay to assess the effect that the 2 ligands (**58** and **61**, identified from the proteomics studies to have antimicrobial activity) would have on G4 stabilisation. Unlike the previous melting assays, these were performed at a single concentration of 5  $\mu$ M, and at four different buffer conditions to see the effect of the ligand at stabilising G4s of different topologies, since in the initial CD studies showed these G4s were indeed polymorphic at different  $K^+$  concentrations.

Previous work by O'Hagan *et. al.*, demonstrated the ability of the pyridinium stiff-stilbene family as good G4 ligands.<sup>109,110,118</sup> The ligand **61** was highly selective towards G4s and provided exceptional stabilisation of G4 structures when measured by a FRET melting assay (F21T:  $\Delta T_m = 33$  °C at 5  $\mu$ M, F10T:  $\Delta T_m = 3$  °C at 5  $\mu$ M).<sup>118</sup> These values are considerably higher than the effect that **58** has towards the same sequences, (F21T:  $\Delta T_m = 11$  °C at 5  $\mu$ M, F10T:  $\Delta T_m = 0$  °C at 5  $\mu$ M).

The results shown in **Table 5.2** show the  $\Delta T_m$  values for the ligands **58**, **61** and pyridostatin at 5  $\mu$ M in a 100 mM KCl buffer. Both **61** and pyridostatin strongly stabilised all 17 sequences ( $\Delta T_m = 9 - 39$  °C). This was lower for **58**, which selectively stabilised 7 sequences, with  $\Delta T_m$  values ranging from 9 – 17 °C (glnD, pdxA, thrA, yjbD, aroL, cobS, mtgA). In order to evaluate further the effect of the ligand on the G4 sequences and to elucidate a possible link between

## 5 Targeting *E. coli* G4s

the antibacterial activity of the ligands and the identified G4s, UV-Vis and CD titrations were carried out next.

**Table 5.2:** FRET thermal stabilisation values ( $\Delta T_m$ ) of FAM and TAMRA labelled *E. coli* DNA sequences (200 nM), with ligands **58**, **61** and positive control pyridostatin at 5  $\mu$ M each, in 100 mM KCl, 10 mM Li cacodylate buffer (pH 7.2).

Sequence	$\Delta T_m / ^\circ\text{C}$		
	Ligand		
	<b>58</b>	<b>61</b>	pyridostatin
1) <b>clpB</b>	2 $\pm$ 0.5	17 $\pm$ 2.4	19 $\pm$ 0.3
2) <b>dppA</b>	7 $\pm$ 2.2	28 $\pm$ 0.8	28 $\pm$ 1.4
3) <b>glnD</b>	11 $\pm$ 0.7	26 $\pm$ 1.5	32 $\pm$ 0.0
4) <b>hemL</b>	2 $\pm$ 0.0	21 $\pm$ 0.2	20 $\pm$ 0.6
5) <b>pdxA</b>	9 $\pm$ 0.3	16 $\pm$ 5.7	28 $\pm$ 0.6
6) <b>thrA</b>	12 $\pm$ 0.3	25 $\pm$ 1.9	30 $\pm$ 0.0
7) <b>yjbD</b>	15 $\pm$ 0.5	35 $\pm$ 0.5	36 $\pm$ 0.0
8) <b>asnS</b>	4 $\pm$ 0.8	31 $\pm$ 1.2	26 $\pm$ 1.3
9) <b>yhaJ</b>	5 $\pm$ 2.0	19 $\pm$ 8.2	17 $\pm$ 2.5
10) <b>ybiO</b>	5 $\pm$ 0.5	26 $\pm$ 0.2	20 $\pm$ 2.6
11) <b>aroL</b>	13 $\pm$ 0.4	33 $\pm$ 4.1	23 $\pm$ 3.0
12) <b>cobS</b>	17 $\pm$ 0.3	31 $\pm$ 0.1	31 $\pm$ 0.1
13) <b>mtgA</b>	14 $\pm$ 0.3	39 $\pm$ 0.5	39 $\pm$ 1.4
14) <b>maeA</b>	7 $\pm$ 0.2	25 $\pm$ 1.3	22 $\pm$ 1.6
15) <b>yfiC</b>	7 $\pm$ 0.2	17 $\pm$ 3.6	21 $\pm$ 1.4
16) <b>creA</b>	8 $\pm$ 0.3	30 $\pm$ 1.8	24 $\pm$ 1.2
17) <b>mukE</b>	7 $\pm$ 0.4	21 $\pm$ 1.0	9 $\pm$ 2.0

## 5.2 Biophysical studies with selected unlabelled sequences

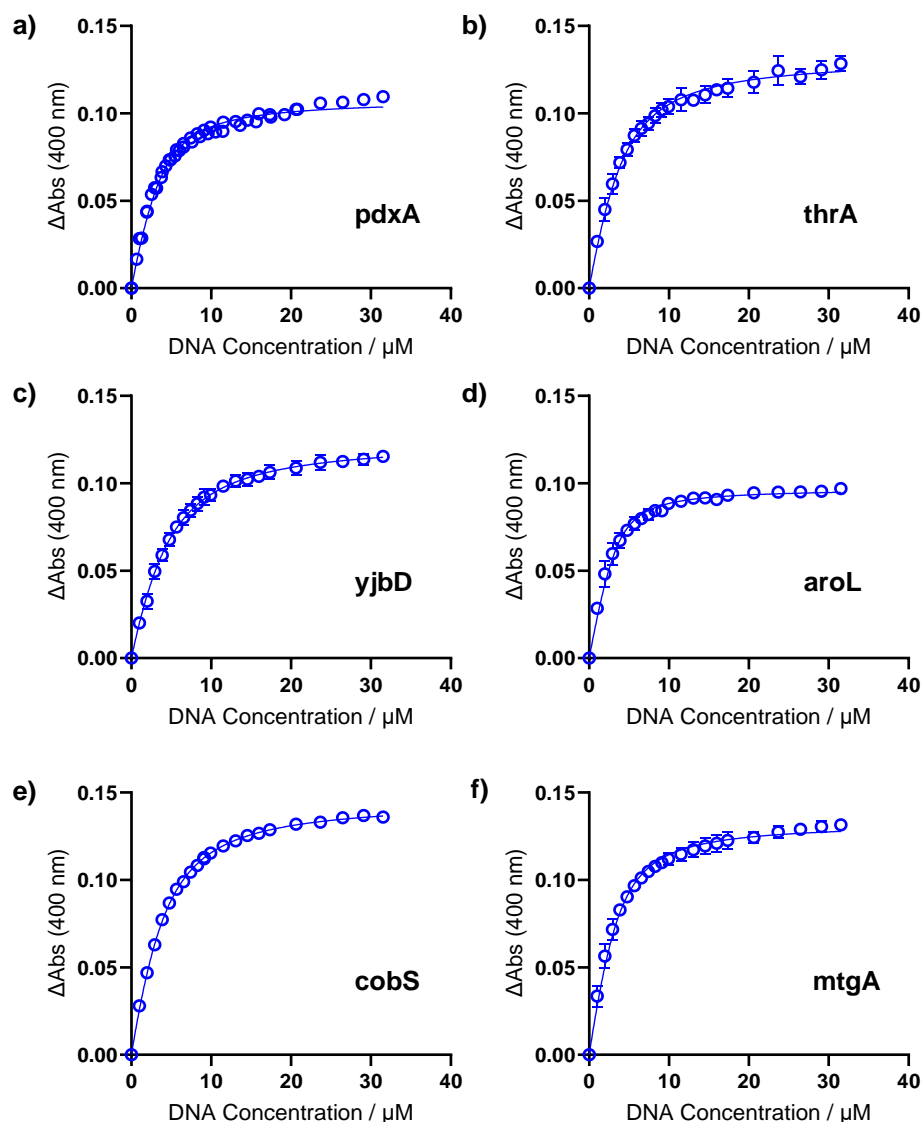
Six sequences were chosen for further biophysical studies with the ligands. These were sequences that showed good stabilisation in the FRET melting assay by the azobenzene ligand **58** (pdxA, thrA, yjbD, aroL, cobS and mtgA).

In order to assess the binding affinity of each ligand and the selected G4 sequences, UV-Vis titrations were undertaken as previously described in this thesis. The UV-Vis titrations could only be carried out with ligands that had chromophores that absorbed above 320 nm where it does not overlap with DNA absorption. This meant that only conjugated ligands like the azobenzene and stiff-stilbene could absorb light at those wavelengths. Whereas pyridostatin, which is only connected by amide bonds and did not absorb any light above 300 nm could not be assayed.

**Table 5.3:** Association constants,  $K_a$ , for the ligands **58** and **61**. Calculated from the binding isotherms with each sequence in potassium buffer conditions, N=3.

Sequence	$K_a / 10^6 \text{ M}^{-1}$	
	ligand	
	<b>58</b>	<b>61</b>
<b>pdxA</b>	$0.27 \pm 0.02$	$1.07 \pm 0.2$
<b>thrA</b>	$0.18 \pm 0.01$	$0.53 \pm 0.09$
<b>yjbD</b>	$0.13 \pm 0.007$	$0.55 \pm 0.08$
<b>aroL</b>	$0.49 \pm 0.05$	$0.60 \pm 0.2$
<b>cobS</b>	$0.17 \pm 0.003$	$0.36 \pm 0.09$
<b>mtgA</b>	$0.29 \pm 0.02$	$0.71 \pm 0.2$

**Figure 5.3** shows the binding isotherms of the 6 selected sequences with azobenzene **58**. The  $K_a$  values ranged from 0.13-0.49 x  $10^6 \text{ M}^{-1}$  (**Table 5.3**). Representative UV-Vis titrations of **58** with DNA (ie. cobS) can be seen in **Figure 5.5a**. As before, we see that upon the addition of DNA, there is a hypochromic (22-31%) and bathochromic shift (14-21 nm) of the  $n-\pi^*$  band, indicating that there is a  $\pi$ -stacking interaction of the ligand with the DNA.



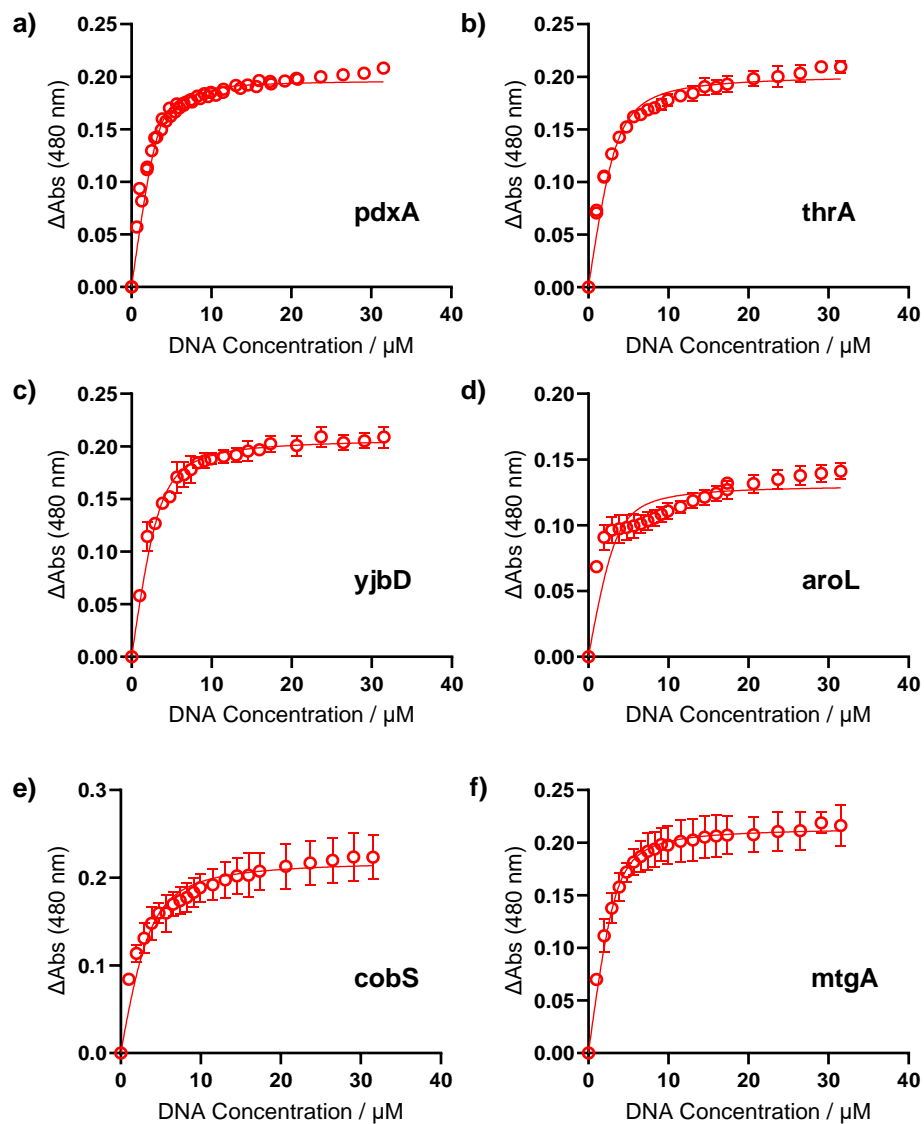
**Figure 5.3:** UV-Vis titration binding isotherms of 10  $\mu\text{M}$  **58** at 400 nm with **a)** pdxA **b)** thrA **c)** yjbD **d)** aroL **e)** cobS **f)** mtgA sequences in 100 mM potassium phosphate buffer (pH 7.4).

Similar results were achieved with stiff-stilbene **61**, showing similar trends of hypochromic (21-28%) and bathochromic (27-37 nm) shifts of the main absorption band (**Figure 5.4**). The  $K_a$  values of  $0.36\text{-}1.07 \times 10^6 \text{ M}^{-1}$  are around two-fold higher than with the azobenzene ligand, supporting prior results that suggested the stiff-stilbene to be a better general G4 ligand.

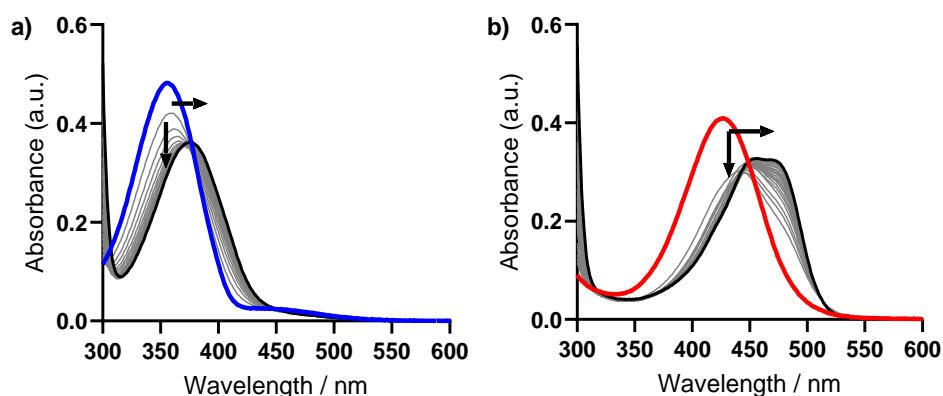
However, there are key differences between the two ligands. An additional hyperchromic (3-20%) shift and a broadened absorption band for the stiff-stilbene **61**, not observed in the titration with **58**. The change from hypo- to hyperchromic shift of the absorption band may

## 5 Targeting *E. coli* G4s

suggest that there are at least two key binding modes for **61**. An initial stacking interaction (hypochromic) followed by electrostatic groove binding (hyperchromic).<sup>158,217,218</sup>



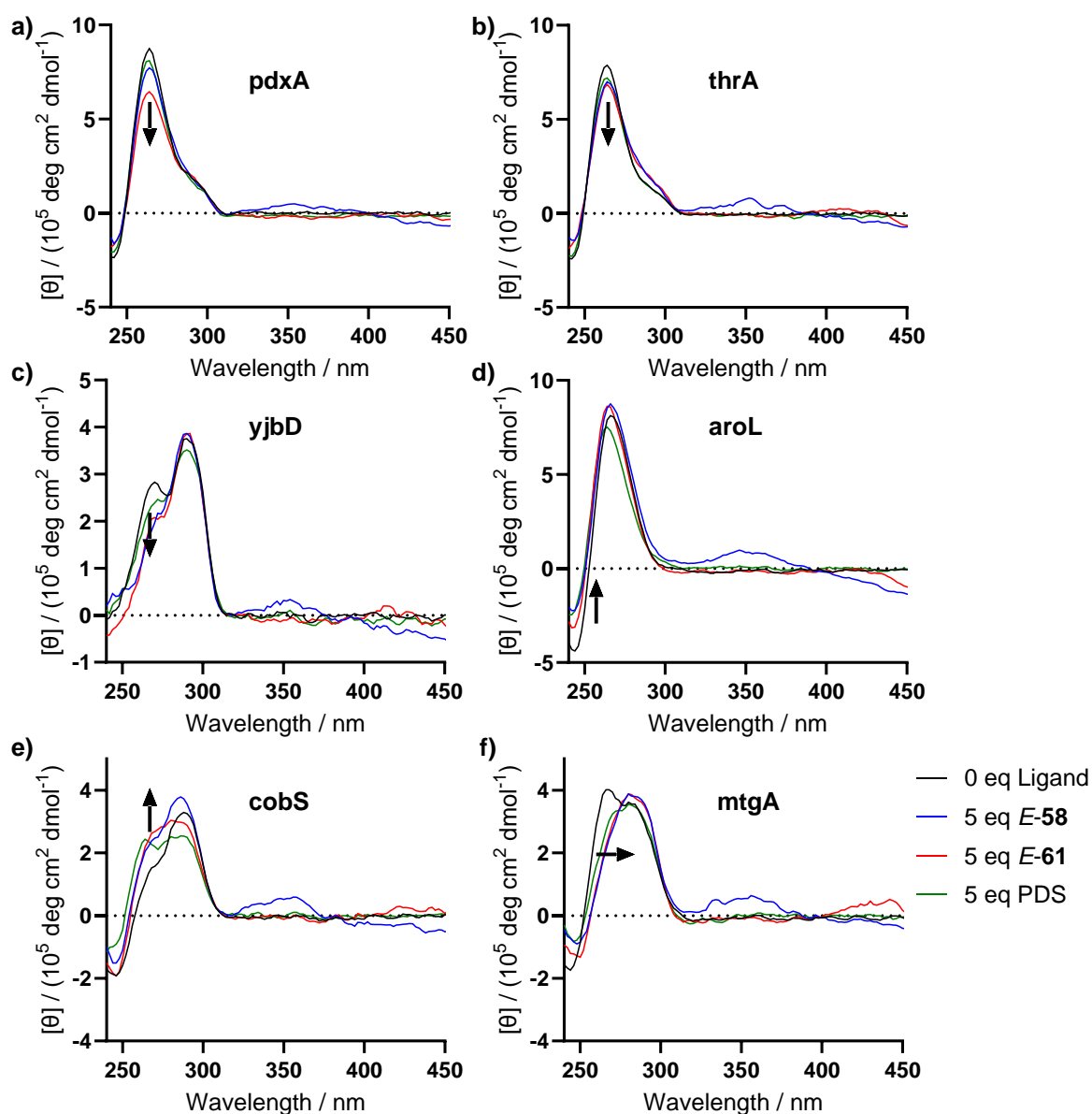
**Figure 5.4:** UV-Vis titration binding isotherms of 10  $\mu$ M **61** at 480 nm with a) pdxA b) thrA c) yjbD d) aroL e) cobS f) mtgA sequences in 100 mM potassium phosphate buffer (pH 7.4).



**Figure 5.5:** Representative UV-Vis titrations of 10  $\mu\text{M}$  ligands a) **58**. b) **61** in  $\text{K}^+$  buffer in UV-Vis titrations (with cobS) in 100 mM potassium phosphate buffer (pH 7.4). Clear hypochromic and bathochromic shifts are observed for both ligands. Additional hyperchromic shifts occur in titrations with stiff-stilbene **61**.

We then carried out titrations between the DNA and ligands with CD spectroscopy to evaluate the effects on G4 topology upon ligand addition. As this primarily measures the chiral conformation of the DNA structure, all ligands could be used in this study.

The CD of the selected unlabelled 6 sequences in 100 mM  $\text{K}^+$  phosphate buffer can be seen in **Figure 5.6** and are shown in black. These had similar spectra as the FRET labelled sequences shown above, where the sequences adopted parallel (aroL, pdxA, thrA), hybrid (yjbD, cobS) or a mix of the two conformations (mtgA). Perturbations to the CD spectra after the addition of 5 equivalents of each ligand, is also shown (**Figure 5.6** and Appendix A5.4-A5.6). Ligand induced changes to the CD spectra for each DNA sequence were similar for all ligands, indicating that any shifts were likely due to the intrinsic nature of the sequence favouring certain topologies. Four sequences showed a large decrease of the positive parallel band  $\sim 265$  nm, suggesting that these conformations favoured more hybrid structures. Ligand addition to cobS caused an increase in the parallel conformation, while the aroL topology remained parallel with a slight decrease in the absorption of the negative band  $\sim 245$ . Additionally, only ligands that had chromophores in the visible range had induced CD bands (ICD), so the lack of ICD does not categorically imply a lack of interaction by pyridostatin.

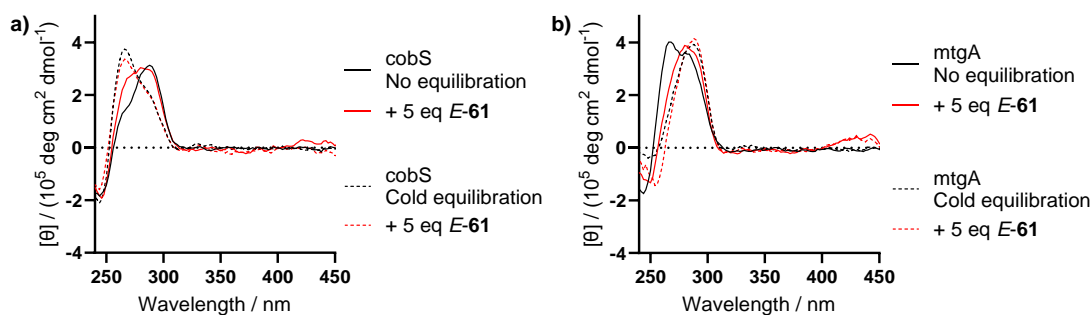


**Figure 5.6:** CD spectra at 20 °C showing titrations of selected *E. coli* sequences (black) in 100 mM potassium phosphate buffer with 0-5 eq of ligands 58 (blue), 61 (red) and PDS/pyridostatin (green). For clarity, only 0 and 5 eq of ligand added shown.

One important factor to consider is that G4s may have multiple conformations due to its dynamic nature. As newly discovered G4 forming sequences, the conditions required for optimal folding of the G4 would not be known. Variations in how the DNA was prepared had implications on the folded conformation of the sequence. We can illustrate this by comparing the *cobS* and *mtgA* sequences that were either used immediately after annealing, versus samples that were left in solution and stored at 4 °C for a few weeks (**Figure 5.7**). The annealing process involves heating DNA solutions to 90 °C for 2 minutes, which denatures the secondary

## 5 Targeting E. coli G4s

structure. It is then immediately cooled in an ice bath for 10 minutes, trapping the G4 sequence into a particular major conformation. This means that the folding process will follow a potential energy landscape that may favour the less thermodynamically stable structure.<sup>219</sup>



**Figure 5.7:** CD of **a)** cobS and **b)** mtgA showing a change in DNA conformation equilibration during long term storage in solution (black). Addition of ligand and its associated conformational change (red). Solid line shows sample used immediately after annealing. Dashed line shows sample stored at 4 °C for 1 month.

**Figure 5.7** shows two sets of titrations with cobS and mtgA sequences with stiff-stilbene **61**. The solid line represents DNA samples that were used immediately after annealing and the dashed line represents DNA samples that were left in solution for a month at 4 °C. After annealing, cobS initially shows a hybrid structure which shifts to parallel conformations upon the addition of **61**. On the other hand, mtgA has an opposite behaviour, initially showing a higher proportion of parallel G4s which shift towards a hybrid structure upon ligand addition.

However, leaving the DNA solutions at 4 °C for 1 month led to a change in the major conformation of both sequences, cobS was more parallel and mtgA was more hybrid. When ligands were added to the equilibrated structures there was no further perturbation to the CD spectra, suggesting that the equilibrated structure was more stable than the unequilibrated structure.

This highlights the importance of considering equilibration time when working with G4s. The time required to achieve an equilibrated structure could take anywhere from seconds to months at ambient temperatures.<sup>175,220,221</sup> While the CD spectra implies that the addition of ligand to the non-equilibrated sequence trends to the same topology as the equilibrated structures, the underlying mechanism is unclear. The ligand could be assisting the sequence



in folding towards a thermodynamically favoured structure by preferentially binding with the non-equilibrated topology or by the stabilisation of any unfolded conformations. Both of these may appear as a conformation shift between varying populations in CD spectroscopy. Techniques such as NMR could be used to elucidate any structural differences. Or single-molecule FRET techniques could be used to determine if there are any differences between the conformational populations of the equilibrated structures and after ligand addition.

219,222,223

It may be important to assess the influence of any kinetically stable versus thermodynamically stable conformations that could be forming in solution. Particularly since the population distribution is likely to be different at physiologically relevant temperatures. If the structure of the G4 changes considerably after equilibration, then ligand induced changes to DNA will be different given potentially different binding modes to two different conformations, making comparing the effectiveness of ligands to modulate topology more challenging.

### 5.3 Chapter Conclusions

In screening of various G4 ligands, azobenzene **58** was discovered to be a strong antibacterial candidate for *E. coli*, thus it was of interest to evaluate potential links between G4 stabilisation and G4 sequences associated to any up- or downregulated microbial genes.

17 *E. coli* PQS sequences were discovered through proteomics and bioinformatics methods. CD spectroscopy showed that a majority of these sequences folded into clear G4 structures, adopting parallel, antiparallel and hybrid structures in increasing KCl concentration.

The stiff-stilbene **61** and control G4 ligand pyridostatin both strongly stabilised all 17 sequences ( $\Delta T_m = 9 - 39$  °C), whereas **58** was only selective for a smaller subset of these sequences. We selected 6 sequences that showed good temperature stabilisation by the addition of **58**, and ran UV-Vis titrations and CD titrations with these three ligands to evaluate binding affinity and topology effects.

Whilst the results of the biophysical assays did not unequivocally show a direct link between the antibacterial activity of **58** and G4 binding, it provides confirmation that the ligands do

## 5 Targeting E. coli G4s

---

interact with G4 sequences. From UV-Vis we further identified that **58** and **61** both had similar association constants but the stiff-stilbene exhibits  $K_a$  values around 2-fold higher, signified by the quick saturation of the ligand shown in the binding isotherms. Finally, CD spectroscopy allowed us to ascertain that addition of ligand led to topology changes in the DNA, although caution needs to be taken when evaluating those experiments since equilibration of the DNA topology to the most stable one can vary between sequences. Thus any observed ligand effects could in some instances be due to a combination of (un)equilibrated DNA and ligand binding acting cooperatively, since a longer DNA equilibration time led will also lead to topology shifts.

## 6 Conclusions and Future work

Controlling G4 topology with photoswitchable ligands has previously been explored with azobenzenes, stiff-stilbenes, dithienylethenes and spiropyran photoswitchable scaffolds. In this work, we have expanded this to a newer photoswitchable class, diazocines. These are close analogues of azobenzenes and feature a connecting ethylene bridge between the two aromatic rings to form an 8 membered ring. Diazocines share many advantageous properties with azobenzenes as photoswitchable molecules. They are robust to photoswitching without showing signs of photofatigue and display large structural differences between its two photoisomers. Unlike azobenzenes however, diazocines have superior photophysical properties such as improved photoswitching quantum yields and can be isomerised using visible light which is beneficial for applications of targeting DNA. Additionally, due to ring strain the thermodynamically stable isomer is the bent *Z*-diazocine, opposite in behaviour to azobenzenes.

A small library of diazocines (Chapter 2) with various N-methylated aromatic heterocycles side chains was synthesised. These compounds were shown to behave similarly as other diazocines and could photoisomerise with 405 nm and 520 nm light. They also showed first order kinetics of thermal isomerisation. Where the 4-pyridinium *E*-**42** had the shortest  $t_{1/2}$  of 30 minutes at room temperature in a potassium phosphate buffer solution. On the other hand, attempts to synthesise an *ortho*-substituted analogue was challenging and not achieved likely due to steric constraints towards the final cyclisation step.

With these photoswitchable ligands, we investigated their binding properties with various G4s implicated in viruses and cancer (Chapter 3, Chapter 4). Ligands with larger side chains (**45**, **46**) showed stronger thermal stabilisation to G4s. However, the biaryl ligand *Z*-**45** had poor selectivity, it could strongly stabilise G4 sequences ( $\Delta T_m$ : 25-32 °C) but also had significant thermal stabilisation of the duplex stem loop sequence, F10T ( $\Delta T_m = 10$  °C), even as the bulkier *Z*-isomer. This indicated that it was important to find a balance between binding affinity and G4 selectivity. Diazocines with quinolinium (**46**) and pyridinium side-chains (**42-44**) stabilised G4 structures more weakly, and did not show any stabilisation of the stem loop structure F10T. Additionally, changing the position of the N-methyl group on the pyridinium ring also

## 6 Conclusions and Future work

---

impacted thermal stabilisation following a trend of 4-py > 3-py > 2-py. FRET melting assays also suggested that there might be different binding affinity between *E/Z*-diazocine isomers as the use of linear azobenzene pyridinium ligands (**58-60**) offered higher stabilisation of than the *Z*-diazocines to G4s.

Investigations of the binding affinity of the ligands towards G4s were then conducted which supported the results from the melting assay, as titrations with *E*-diazocines showed higher binding affinity than the corresponding *Z*-diazocines. In particular, the 4-pyridinium diazocine **42** had a 34, 21 and 23 fold *E/Z* ratio in  $K_a$  towards the sequences LTR-III, telo23 and telo22, respectively.

Given the large differences in *E/Z* binding affinities shown by **42** towards LTR-III,  $^1\text{H}$  NMR was used to identify where the ligands were interacting within the G4 structure. Titrations with **42** and **43** showed larger changes in chemical shifts for the *E*-isomer and indicated that it was associating at the quadruplex-duplex junction.

To probe further, we carried out MD simulations of LTR-III and **42** to achieve further insights into how the two structurally different ligand behaved at an atomistic level. We manually docked both *E*-**42** and *Z*-**42** at the quadruplex-duplex junction of LTR-III into 10 different NMR conformations. Then ran 255 ns long MD simulations which gave a total of 5.1  $\mu\text{s}$  simulation data. Unfortunately, due to a combination of poor starting positions and limited simulation time, some of the Models showed unfolding behaviour which was not observed from biophysical experiments. This was likely due to missing  $\text{K}^+$  ions in the central G4 channel that typically provide structural stability by coordinating and restricting motion of G4 residues. As a result, the impact of the two ligands upon G4 dynamics was not clear. Of the few Models that had stable G4 structures, the ligand was found to be bound with the G4 via  $\pi$ -stacking interactions, which is what we expected based on empirical data. To be able to further understand the influence that diazocine **42** had upon LTR-III dynamics, longer simulations and additional control simulations with ions sites occupied within the channel without any ligand present may allow for better conclusions to understand if any changes that have occurred are due to the presence of the ligand or limitations of the method.

## 6 Conclusions and Future work

---

Overall, diazocines seem like a promising photoswitchable ligand class for targeting G4 structures. The 4-pyridinium diazocine **42** demonstrated desirable *E/Z* differences where *E-42* preferentially bound to G4 targets and the binding affinity of *Z-42* was significantly weaker. Further structural modifications such as changing the substitution pattern from *para*-disubstituted diazocines to *meta*-substituted diazocines could extend the thermal half-life of the metastable *E*-diazocines, easing future study of these molecules. Alternatively, the conformational flexibility of diazocines can be further constrained by the addition of two 5-membered rings between the bridging carbons and the aromatic ring of diazocines as reported in work by Moormann *et al.*<sup>224</sup> These diindane diazocines have larger surface areas which may be beneficial, as it could increase potential binding interactions between the ligand and the equally large surface area of a G-tetrad.

Finally, to explore the link between the antibacterial activity of azobenzene **58** against *E. coli* and G4s, we also tested the propensity of 17 putative G4 sequences from the *E. coli* genome, (discovered through a combination of proteomics and bioinformatics methods) to fold into G4s *in vitro*. Using CD spectroscopy, we could see that most of the sequences folded into characteristic G4 spectra without the requirement of additional K<sup>+</sup> ions. The sequences that did not initially fold into any secondary structure at 0 mM KCl (*mukE*, *yjbD*) showed characteristic G4 spectra upon increasing K<sup>+</sup> ion concentration. This stabilisation from folding could be measured in the melting assays where a small increase in thermal stability was observed at higher potassium concentrations for both *mukE* ( $\Delta T_m = 5$  °C) and *yjbD* ( $\Delta T_m = 8$  °C).

FRET melting assays with the 17 sequences showed that the stiff-stilbene **61** and pyridostatin (**5**) ligands led to strong stabilisation of all sequences ( $\Delta T_m = 9$ – $39$  °C). Whereas azobenzene **58** only stabilised 7 sequences to a lesser degree ( $\Delta T_m = 9$ – $17$  °C). Thus, the sequences that were selectively stabilised by **58** were investigated further in titrations using UV-Vis and CD spectroscopy.

Ligand addition to the selected 6 sequences (*pdxA*, *thrA*, *yjbD*, *aroL*, *cobS* and *mtgA*) showed expected induced CD bands and perturbed the spectra of the DNA. However, two sequences, *cobS* and *mtgA*, which were left at 4 °C for weeks after annealing also showed changes in the

## 6 Conclusions and Future work

---

CD spectra as a consequence of long-term equilibration which warrants significant further study about the dynamic nature of these newly discovered bacterial G4s.

In conclusion, a new ligand class to target G4s were developed and demonstrated good photoswitchable properties in its interactions with the G4s LTR-III, telo23 and telo22. The 4-methylpyridinium substituent was identified to be a suitable side chain that will provide satisfactory binding affinities, without outperforming the innate binding differences between *E*- and *Z*- diazocines from light induced photoisomerisation. Alternative structural modifications can be made to further explore its potential to target G4s.

# 7 Experimental

## 7.1 Ligand photochemistry

The light sources used in experiments were: a ThorLabs 4.5 mW 405 nm or 520 nm Collimated Laser Diode Module (CPS405 and CPS520) or a 60 W 395-405 nm 3528 SMD UV LED strip.

Stock ligand solutions were prepared at 10 mM concentration in DMSO- $d_6$  and stored at  $-20\text{ }^{\circ}\text{C}$ . Compounds were diluted in buffer immediately prior to use. Ligand photoswitching was recorded using by UV-Visible or NMR spectroscopy at 10  $\mu\text{M}$ , 20  $\mu\text{M}$  or 100  $\mu\text{M}$  in UV-Vis experiments or 446  $\mu\text{M}$  in NMR spectroscopy.

UV-Visible spectra were collected recorded using an Agilent Cary 60 UV-Vis Spectrophotometer at ambient temperature using a room-light immune fiber optic probe with 10 mm path length. For thermal half-life measurements, the scanning kinetics mode was utilised, with the baseline corrected for with buffer. The ligand was irradiated with a ThorLabs 405 nm 4.5 mW collimated laser diode module till PSS was reached, then left in darkness where spectra were collected at timed intervals. For photoswitching measurements, the ligand was irradiated alternatingly with a 405 nm or 520 nm laser for 5 minute periods.

$^1\text{H}$  NMR and  $^{13}\text{C}$  NMR spectra were recorded at 298 K on a 600 MHz Bruker spectrometer. For thermal half-life measurements, the ligand was irradiated with a ThorLabs 405 nm 4.5 mW collimated laser diode module for at least 5 minutes. Then  $^1\text{H}$  NMR spectra was recorded using presat water suppression to record rate of thermal reversion from the metastable isomer to the stable isomer. Integrals were calculated relative to the residual DMSO peak. For *E*-isomer characterisation, the ligand was irradiated with a 60 W 395-405 nm 3528 SMD UV LED strip. All spectra were processed with MestReNova software.

Thermal half-life was processed using Scipy curvefit or linear regression functions.

## 7 Experimental

---

### 7.2 Biophysical studies

All oligonucleotides used were purchased from Eurogentec (Belgium), purified by HPLC and delivered dry. Oligonucleotide concentrations were determined by UV-absorbance using a NanoDrop 2000 Spectrophotometer from Thermo Scientific. Oligonucleotides were annealed by heating to 90 °C for 2 minutes and then cooled on ice for at least 10 minutes. The oligonucleotides are shown in the following tables.

Fluorescently labelled sequences:

DNA model	Sequence
F21T	5'-FAM-GGGTTAGGGTTAGGGTTAGGG-TAMRA-3'
F10T	5'-FAM-TATAGCTATA-HEG-TATAGCTATA-TAMRA-3'
FLTR-IIIT	5'-FAM-GGGAGGCGTGGCCTGGGCGGGACTGGGG-TAMRA-3'
hemL	5'-FAM-GGTCCGGTCTATCAGGCGGGT-TAMRA-3'
dppA	5'-FAM-GGGCTGGACTGGCGATAACGGGG-TAMRA-3'
clpB	5'-FAM-GGCGCGTTGGACGGGG-TAMRA-3'
yjbD	5'-FAM-GGGAAAAGGGTTAGGGTGAGGG-TAMRA-3'
thrA	5'-FAM-GGGCGATGGGGGTAATGGTGCGGGGG-TAMRA-3'
pdxA	5'-FAM-GGGGAGTTGGGGGAATAAGGGCGGAGGG-TAMRA-3'
glnD	5'-FAM-GCGGTTGACCGGCAGGGTGGG-TAMRA-3'
asnS	5'-FAM-GGGTAAAGTCGTGGCGTCGCCGGGCCAGGGG-TAMRA-3'
yhaJ	5'-FAM-GGGGGCGTGGGAACGGCTGGAGCAGGGG-TAMRA-3'
ybiO	5'-FAM-GGGTACGCGTGCGGGCGCTGGGTAGCGGG-TAMRA-3'
aroL	5'-FAM-GGAGATCGTCGAAAGGGAAGAGTGGGCGGG-TAMRA-3'
cobS	5'-FAM-GGGCGGGCAAACGGGCGATACGCTGGG-TAMRA-3'
mtgA	5'-FAM-GGGATGGGCGTAGCTGGGTTCGAAAAGGG-TAMRA-3'
maeA	5'-FAM-GGGGCTTGGTGACCAGGGCATCGGCGGG-TAMRA-3'
yfiC	5'-FAM-GGGCGCATGGGCACCGGTGGCTGGGG-TAMRA-3'
creA	5'-FAM-GGTGGTATTAAGGGGGATTGGGTCTGGCG-TAMRA-3'
mukE	5'-FAM-GGAACGGCTGGCGAATGAGGGG-TAMRA-3'

FAM = 6-carboxyfluorescein;

TAMRA = 6-carboxy-tetramethylrhodamine;

HEG = [(-CH<sub>2</sub>CH<sub>2</sub>O-)<sub>6</sub>]



## 7 Experimental

---

Non-fluorescently labelled sequences:

DNA	Sequence
LTR-III	5'-GGGAGGCGTGGCCTGGGCGGGACTGGGG-3'
telo23	5'-TAGGGTTAGGGTTAGGGTTAGGG-3'
telo22	5'-AGGGTTAGGGTTAGGGTTAGGG-3'
yjbD	5'-GGGAAAAGGGTTAGGGTGAGGG-3'
thrA	5'-GGGCGATGGGGGTAATGGTGCGGGGG-3'
pdxA	5'-GGGGAGTTGGGGGAATAAGGGCGGAGGG-3'
aroL	5'-GGAGATCGTCGAAAGGGAAGAGTGGGCGGG-3'
cobS	5'-GGGCGGGCAAACGGGCGATACGCTGGG-3'
mtgA	5'-GGGATGGGCGTAGCTGGGTTGAAAAGGG-3'

### 7.2.1 FRET melting assays

FRET melting assays were performed according to the procedure reported by De Cian and co-workers<sup>153</sup> on a Stratagene Mx3005P or AriaMx qPCR instrument. The assay was performed in a 96-well plate format. The final volume of solution in each well was 25  $\mu$ L and contained the appropriate oligonucleotide (200  $\mu$ M).

Buffers used were as follows: (F21T-K<sup>+</sup>, F10T and FLTR-IIIT) 10 mM KCl, 90 mM LiCl and 10 mM Li cacodylate, pH 7.2; (F21T-Na<sup>+</sup>) 100 mM NaCl and 10 mM Li cacodylate, pH 7.2. For *E. coli* FRET sequences, buffers used were variations of KCl:LiCl salt concentrations: 0 mM KCl, 100 mM LiCl and 10 mM Li cacodylate, pH 7.2; 1 mM KCl, 99 mM LiCl and 10 mM Li cacodylate, pH 7.2; 10 mM KCl, 90 mM LiCl and 10 mM Li cacodylate, pH 7.2; 50 mM KCl, 50 mM LiCl and 10 mM Li cacodylate, pH 7.2; 100 mM KCl, 0 mM LiCl and 10 mM Li cacodylate, pH 7.2.

Ligand concentrations were either 1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M or 10  $\mu$ M and prepared from an intermediate stock of 50  $\mu$ M in HPLC-grade H<sub>2</sub>O (0.5% DMSO). Each sample was tested in duplicated on the same plate and each plate was repeated in at least duplicate. Appropriate control experiments were also carried out for each sample set, DNA in the absence of ligand as a negative control; 1  $\mu$ M TMPyP4 as a positive control. Data processing was carried out using Python, with  $\Delta T_{1/2}$  used to represent  $\Delta T_m$ .

## 7 Experimental

---

### 7.2.2 Titration studies

The oligonucleotide sequences used in titration studies were LTR-III, telo23, telo22, aroL, cobS, mtgA, pdxA, thrA and yjbD.

The buffers for these sequences were: (LTR-III) 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7; (telo23, aroL, cobS, mtgA, pdxA, thr and yjbD) 100 mM potassium phosphate at pH 7.4; (telo22) 100 mM sodium phosphate at pH 7.4.

For experiments involving the addition of *E*-diazocines, both the titrant and titrate were either constantly irradiated (UV-Vis titration) or the titrate was irradiated for at least 2 minutes prior to spectroscopic measurement of data (CD and NMR) with a ThorLabs 405 nm 4.5 mW collimated laser diode module and the titration was carried out with the lights out.

#### 7.2.2.1 Circular dichroism titrations

Circular dichroism (CD) titrations were recorded using a Jasco J-815 spectrometer fitted with a Peltier temperature controller. Measurements were taken in a quartz cuvette with a path length of 5mm, at 20 °C. The CD spectra were recorded between 600 and 240 nm, and baseline corrected for the buffer used. The oligonucleotides were at an initial concentration of 5 μM. The ligand was added by aliquot from a 1 mM stock solution in the appropriate buffer (10% DMSO). The reported spectrum for each sample represents the average of 3 scans. Observed ellipticities (in mdeg) were converted to molar ellipticity  $[\theta] = \text{deg cm}^2 \text{dmol}^{-1}$ .

#### 7.2.2.2 UV-Visible titration studies

UV-Visible spectra were recorded using an Agilent Cary 60 UV-Vis Spectrophotometer at ambient temperature using a room-light immune fiber optic probe with 10 mm path length. UV-visible spectra were recorded between 700 and 300 nm and baseline corrected for the buffer used.

In the titrations, the concentration of the ligand was fixed at 10 μM in an initial volume of 500 μL. Aliquots of oligonucleotide were added using a microsyringe from a 100 μM stock solution in appropriate buffer also containing 10 μM ligand to maintain constant ligand

## 7 Experimental

---

concentration. Data were fitted to an appropriate binding model using non-linear regression in Prism 9 software to provide estimates of the association constants.

### 7.2.2.3 NMR titration studies

$^1\text{H}$  NMR spectra were recorded at 298 K or 278 K on a 600 MHz Bruker spectrometer. The oligonucleotide used was LTR-III. The DNA concentration was 185  $\mu\text{M}$  in 20 mM potassium phosphate and 70 mM potassium chloride buffer, pH 7.0, containing 10%  $\text{D}_2\text{O}$ . All experiments employed sculpted excitation water suppression. During the titrations, aliquots of ligand were added from a 10 mM stock solution in  $\text{DMSO-d}_6$  to give the required titration points. After addition of ligand aliquots, the sample was mixed thoroughly and the NMR spectrum acquired immediately. The imino resonances of LTR-III in the absence of ligand were assigned by comparison with previously published data.<sup>57</sup> All spectra were processed with MestReNova software.

### 7.3 MD Computational Studies

Ligands were manually docked at the quadruplex-duplex junction in 10 NMR structures of LTR-III. All simulations were performed using GROMACS (v 2019.5)<sup>225</sup>, using the bsc1 force field.<sup>182</sup> The ligands *E-42* and *Z-42* were parameterised with the general AMBER force field, GAFF.<sup>196</sup> The DNA-Ligand complex was placed in a cubic box with at least 2.0 nm separation from the box edges, and solvated with explicit TIP3P water<sup>226</sup> and neutralised with potassium chloride ions at an 100 mM concentration with Joung-Cheatham ion parameters.<sup>193–195</sup> The DNA-Ligand complexes were minimised prior to the equilibration and production run as follows: The minimisation of the solute hydrogen atoms on the DNA and the ligand with decreasing force constants applied to heavy atoms. After full relaxation, the system was slowly heated to room temperature (300 K) using a velocity-rescaling thermostat with a coupling constant of 0.5 ps employing an NVT (constant-temperature, constant-volume) ensemble.<sup>227</sup> As the system reached the temperature of interest, the equilibration simulation was performed for 300 ps using an NPT ensemble with a Berendsen barostat.<sup>228</sup> Finally, a production run of 255 ns using Nose-Hoover thermostat<sup>229</sup> and Parrinello-Rahman barostat<sup>230</sup> with a 0.5 ps coupling constant for both temperature and pressure. All the simulations were carried out under the periodic boundary conditions (PBC). The particle-mesh Ewald (PME) method was used to calculate the electrostatic interactions within a cut-off of 10 Å.<sup>231</sup> The same cut-off was used for Lennard Jones (LJ) interactions. All simulations were performed with a 1.0 fs time step. For the equilibrium simulations, analysis was performed using GROMACS tools (v 2019.5).<sup>225</sup> Hydrogen bonds were defined based on a combined distance ( $d_{D-A} \leq 3.0$  Å) and angle ( $\angle(D-H-A) 135^\circ \leq 180^\circ$ ) criteria.

### 7.4 Synthetic Procedures and compound characterisation

#### General experimental details

Chemicals were purchased and used without further purification. Dry solvents were obtained by distillation using standard procedures, or by passage through a column of anhydrous alumina using equipment from Anhydrous Engineering (University of Bristol) based on the Grubbs' design.<sup>232</sup> Reactions requiring anhydrous conditions were performed under N<sub>2</sub>; glassware and needles were either flame dried immediately prior to use, or placed in an oven (150 °C) for at least 2 h and allowed to cool in a desiccator or under reduced pressure. Liquid reagents, solutions or solvents were added via syringe through rubber septa; solid reagents were added via Schlenk type adapters. Reactions were monitored by TLC on Kieselgel 60F<sub>254</sub> (Merck), with UV light (254 nm) detection and by staining with basic potassium permanganate solution. Flash column chromatography was performed according to the method reported Still and co-workers using silica gel [Merck, 230–400 mesh (40–63 μm)]<sup>233</sup>. Solvents for flash column chromatography (FCC) and thin layer chromatography (TLC) are listed in volume:volume percentages. Extracts were concentrated in vacuo using both a Heidolph HeiVAP Advantage rotary evaporator (bath temperatures up to 50 °C) at a pressure of 15 mmHg (diaphragm pump) or 0.1 mmHg (oil pump), as appropriate, and a high vacuum line at room temperature.

Preparative HPLC was performed on a Grace Discovery Sciences Reveleris Prep System with a Phenomenex Luna 5 μm C18(2) 100 Å AXIA packed (250 × 21.2 mm) column. For purification, the instrument was set to monitor the ELSD signal as well as 220 nm, 254 nm and 280 nm on the UV detector. Flow rates were 14 mL/min. Water soluble compounds were freeze dried on a Lyotrap Plus (LTE Scientific LTD).

<sup>1</sup>H NMR and <sup>13</sup>C NMR spectra were measured at 25 °C in the solvent specified with Varian or Bruker spectrometers operating at field strengths listed. Chemical shifts are quoted in parts per million with spectra referenced to the residual solvent peaks. Multiplicities are abbreviated as: br (broad), s (singlet), d (doublet), t (triplet), q (quartet), p (pentet), m (multiplet) and app. (apparent) or combinations thereof. Assignments of <sup>1</sup>H NMR and <sup>13</sup>C NMR

## 7 Experimental

---

signals were made where possible, using COSY, HSQC and HMBC experiments. Mass spectra were obtained by the University of Bristol mass spectrometry service by electrospray ionisation (ESI), electron ionisation (EI) atmospheric pressure chemical ionisation (APCI), or nanospray modes. Infra-red spectra were recorded in the range 4000-600  $\text{cm}^{-1}$  on a Perkin Elmer Spectrum as solids compressed onto a diamond window.

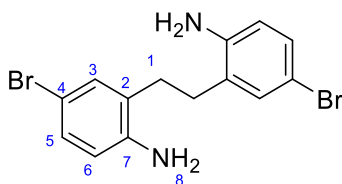
### General procedure for Suzuki cross coupling of boronic acids

A solution of 2.5 M  $\text{K}_2\text{CO}_3$  (10 equiv.) in water (2 mL) was degassed by bubbling nitrogen for 10 min. This was then added to a yellow suspension of diazocine **36** (1 equiv.), a boronic acid (3 equiv.),  $\text{Pd}(\text{PPh}_3)_4$  (0.1 equiv.) and 1 drop of ethylene glycol in anhydrous THF which was also degassed by bubbling nitrogen for 10 min. The reaction was heated overnight under nitrogen at 60 °C and then was extracted with DCM and EtOAc and washed with brine. The organic phase was dried with  $\text{MgSO}_4$  and the solvent was removed *in vacuo*. The crude product was purified by multiple column chromatography steps to give the product as a yellow solid.

## 7 Experimental

---

### 2,2'-(ethane-1,2-diyl)bis(4-bromoaniline) (**35**)



Prepared according to a literature procedure.<sup>120</sup>

To 2,2'-(ethane-1,2-diyl)dianiline (3.00 g, 14.1 mmol) dissolved in DMSO (45 mL) was added N-bromosuccinimide (5.14 g, 28.9 mmol, 2.05 equiv.) slowly over 15 min and then the reaction was stirred overnight at room temperature. The reaction was poured into a stirred solution of water (180 mL) and DCM (30 mL). This was stirred for 1 h before filtering the precipitate giving the product as a beige solid (1.92 g, 5.19 mmol, 37%).

The filtrate was collected, dried and purified by column chromatography (DCM:EtOAc, 98:2 to 90:10) to give additional product (2.12 g, 5.73 mmol, 41%).

**<sup>1</sup>H NMR** (400 MHz, DMSO-*d*<sub>6</sub>) δ 7.18 (d, *J* = 2.3 Hz, 2H, 3-CH), 7.03 (dd, *J* = 8.5, 2.4 Hz, 2H, 5-CH), 6.59 (d, *J* = 8.5 Hz, 2H, 6-CH), 5.11 (s, br, 4H, NH<sub>2</sub>), 2.62 (s, 4H, 1-CH<sub>2</sub>).

**<sup>13</sup>C NMR** (101 MHz, DMSO-*d*<sub>6</sub>) δ 145.7 (7-C), 131.1 (3-CH), 129.0 (5-CH), 127.5 (ArC), 116.4 (6-CH), 106.9 (ArC), 29.1 (1-CH<sub>2</sub>).

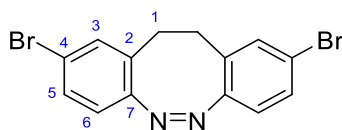
**ESI-LRMS** C<sub>14</sub>H<sub>15</sub>Br<sub>2</sub>N<sub>2</sub><sup>+</sup> [M+H]<sup>+</sup> calc: 371.0, found: 371.0.

Proton and carbon NMR were consistent with literature data.<sup>234</sup>

## 7 Experimental

---

### (Z)-2,9-dibromo-11,12-dihydrodibenzo[c,g][1,2]diazocine (**36**)



Prepared according to a literature procedure.<sup>120</sup>

To a stirred solution of dianiline **35** (1.00 g, 2.70 mmol) in DCM (50 mL) and AcOH (17 mL), a solution of 77% *m*CPBA (1.21 g, 5.40 mmol, 2 equiv.) in AcOH (9 mL) was added dropwise via a syringe pump over 12 hours under rapid stirring. After complete addition, the solution was stirred for another 3 h at room temperature. The solvent was concentrated *in vacuo* and the residue was dissolved in EtOAc (100 mL) and washed with sat. NaHCO<sub>3</sub> (2 x 15 mL) and then with brine (2 x 25 mL). The organic phase was dried with MgSO<sub>4</sub> and then the solvent was removed *in vacuo* to give the crude product as a brown solid which was purified by column chromatography (DCM:Hexane, 1:1 then 2:1) gave the product as a yellow solid (700 mg, 1.92 mmol, 71%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.28 (dd, *J* = 8.4, 2.1 Hz, 2H, 5-CH), 7.16 (d, *J* = 2.0 Hz, 2H, 3-CH), 6.72 (d, *J* = 8.4 Hz, 2H, 6-CH), 2.99 – 2.66 (m, 4H, 1-CH<sub>2</sub>).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 154.2 (7-C), 132.6 (3-CH), 130.2 (5-CH), 130.0 (4-C), 120.8 (2-C), 120.8 (6-CH), 31.4 (1-CH<sub>2</sub>).

$\nu_{\max}$  /cm<sup>-1</sup> (compressed solid) 3451 (w), 3369 (w), 1748 (m), 1614 (m), 1465 (s), 1286 (m), 893 (m), 878 (m), 861 (m), 805 (s), 677 (m).

EI-LRMS C<sub>14</sub>H<sub>10</sub>Br<sub>2</sub>N<sub>2</sub><sup>+</sup> [M]<sup>+</sup> calc: 363.9, found: 363.9.

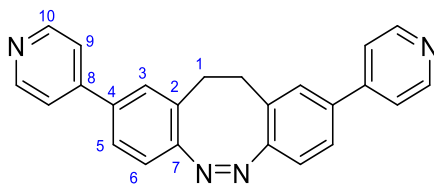
Proton and carbon NMR were consistent with literature data.<sup>120</sup>



## 7 Experimental

---

### (Z)-2,9-di(pyridin-4-yl)-11,12-dihydrodibenzo[*c,g*][1,2]diazocine (**37**)



Synthesis by Dr J. Ramos Soriano.

Prepared according to General Procedure for Suzuki cross coupling. Using diazocine **36** (400 mg, 1.099 mmol), 4-pyridinylboronic acid (450 mg, 3.30 mmol), Pd(PPh<sub>3</sub>)<sub>4</sub> (128 mg, 0.110 mmol) and 1 drop of ethylene glycol in THF (15 mL). And a 2.5 M K<sub>2</sub>CO<sub>3</sub> solution (2 mL), degassed by bubbling nitrogen. The crude product was extracted with DCM and EtOAc, washed with brine and dried with MgSO<sub>4</sub>. Then the crude product was purified by column chromatography (DCM:MeOH, 40:1) to give the product as a yellow solid (395 mg, 1.099 mmol, 99%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.66 – 8.56 (m, 4H, 10-CH), 7.45 (dd, *J* = 8.1, 1.9 Hz, 2H, 5-CH), 7.43 – 7.37 (m, 4H, 9-CH), 7.29 (d, *J* = 2.0 Hz, 2H, 3-CH), 7.01 (d, *J* = 8.1 Hz, 2H, 6-CH), 3.15 – 2.82 (m, 4H, 1-CH<sub>2</sub>).

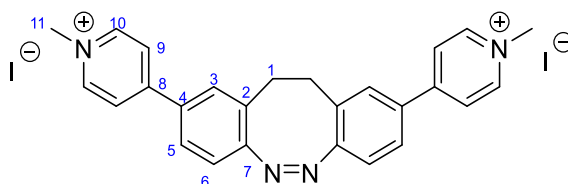
<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 156.0 (7-C), 150.4 (10-CH), 147.1 (8-C), 137.1 (4-C), 128.9 (2-C), 128.5 (3-CH), 125.7 (5-CH), 121.4 (9-CH), 120.0 (6-CH), 31.9 (1-CH<sub>2</sub>).

ESI-HRMS C<sub>24</sub>H<sub>19</sub>N<sub>4</sub><sup>+</sup> [M+H]<sup>+</sup> calc: 363.1604, found: 363.1609.

## 7 Experimental

---

(*Z*)-4,4'-(11,12-dihydrodibenzo[*c,g*][1,2]diazocine-2,9-diyl)bis(1-methylpyridin-1-ium) iodide (**42**)



To a sealed microwave vial was added disubstituted diazocine **37** (100 mg, 0.276 mmol), and dissolved in CH<sub>3</sub>CN (8 mL). To this was added CH<sub>3</sub>I (138 μL, 2.209 mmol) which was heated at 70 °C overnight. The precipitate was collected by vacuum filtration and washing with Et<sub>2</sub>O to give the product as a yellow solid (117 mg, 0.181 mmol, 66%).

**<sup>1</sup>H NMR** (600 MHz, DMSO-*d*<sub>6</sub>) δ 8.94 – 8.90 (m, 4H, 10-CH), 8.41 – 8.37 (m, 4H, 9-CH), 7.93 (dd, *J* = 8.3, 2.0 Hz, 2H, 5-CH), 7.91 (d, *J* = 1.9 Hz, 2H, 3-CH), 7.23 (d, *J* = 8.3 Hz, 2H, 6-CH), 4.28 (s, 6H, 11-CH<sub>3</sub>), 3.09 – 3.00 (m, 4H, 1-CH<sub>2</sub>).

**<sup>13</sup>C NMR** (101 MHz, DMSO-*d*<sub>6</sub>) δ 157.6 (7-C), 152.9 (8-C), 145.5 (10-CH), 132.3 (4-C), 130.0 (5-CH), 129.2 (2-C), 127.0 (3-CH), 123.9 (9-CH), 120.0 (6-CH), 47.1 (11-CH<sub>3</sub>), 30.7 (1-CH<sub>2</sub>).

**ESI-HRMS** C<sub>26</sub>H<sub>24</sub>N<sub>4</sub>I<sup>+</sup> [M]<sup>+</sup> calc: 519.1040, found: 519.1043.

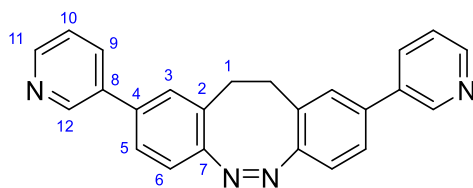
(*E*)-4,4'-(11,12-dihydrodibenzo[*c,g*][1,2]diazocine-2,9-diyl)bis(1-methylpyridin-1-ium) iodide (*E*-**42**)

**<sup>1</sup>H NMR** (600 MHz, DMSO-*d*<sub>6</sub>) δ 9.05 – 9.01 (m, 4H, 10-CH), 8.56 – 8.52 (m, 4H, 9-CH), 8.18 (dd, *J* = 8.2, 2.0 Hz, 2H, 5-CH), 8.03 (d, *J* = 2.0 Hz, 2H, 3-CH), 7.97 (d, *J* = 8.2 Hz, 2H, 6-CH), 4.34 (s, 6H, 11-CH<sub>3</sub>), 3.11 – 3.00 (m, 4H, 1-CH<sub>2</sub>).

## 7 Experimental

---

### (Z)-2,9-di(pyridin-3-yl)-11,12-dihydrodibenzo[*c,g*][1,2]diazocine (**38**)



Synthesis by Dr J. Ramos Soriano

Prepared according to General Procedure for Suzuki cross coupling. Using diazocine **36** (200 mg, 0.550 mmol), 3-pyridinylboronic acid (213 mg, 1.65 mmol), Pd(PPh<sub>3</sub>)<sub>4</sub> (64 mg, 0.055 mmol) and 1 drop of ethylene glycol in THF (8 mL). And a 2.5 M K<sub>2</sub>CO<sub>3</sub> solution (2 mL), degassed by bubbling nitrogen. The crude product was extracted with DCM and EtOAc, washed with brine and dried with MgSO<sub>4</sub>. Then the crude product was purified by column chromatography (CHCl<sub>3</sub>:MeOH, 40:1) to give the product as a yellow solid (187 mg, 0.517 mmol, 94%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.75 (dd, *J* = 2.4, 0.9 Hz, 2H, 12-CH), 8.55 (dd, *J* = 4.8, 1.6 Hz, 2H, 11-CH), 7.78 (ddd, *J* = 7.9, 2.4, 1.7 Hz, 2H, 9-CH), 7.39 (dd, *J* = 8.1, 1.9 Hz, 2H, 5-CH), 7.30 (ddd, *J* = 7.9, 4.8, 0.9 Hz, 2H, 10-CH), 7.24 (d, *J* = 2.0 Hz, 2H, 3-CH), 7.01 (d, *J* = 8.2 Hz, 2H, 6-CH), 3.15 – 2.84 (m, 4H, 1-CH<sub>2</sub>).

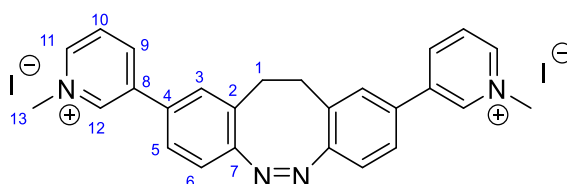
<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 155.2 (7-C), 148.7 (11-CH), 148.1 (12-CH), 136.7 (4-C), 135.4 (8-C), 134.1 (9-CH), 128.8 (2-C), 128.4 (3-CH), 125.6 (5-CH), 123.5 (10-CH), 120.0 (6-CH), 31.9 (1-CH<sub>2</sub>).

ESI-HRMS C<sub>24</sub>H<sub>19</sub>N<sub>4</sub><sup>+</sup> [M+H]<sup>+</sup> calc: 363.1604, found: 363.1603.

## 7 Experimental

---

(*Z*)-3,3'-(11,12-dihydrodibenzo[*c,g*][1,2]diazocine-2,9-diyl)bis(1-methylpyridin-1-ium) iodide (**43**)



Synthesis by Dr J. Ramos Soriano

To a sealed microwave vial was added disubstituted diazocine **38** (100 mg, 0.276 mmol), and dissolved in CH<sub>3</sub>CN (6 mL). To this was added CH<sub>3</sub>I (138 μL, 2.209 mmol) which was heated at 70 °C overnight. The precipitate was collected by vacuum filtration and washing with Et<sub>2</sub>O to give the product as a yellow solid (174 mg, 0.269 mmol, 98%). <sup>1</sup>H and <sup>13</sup>C NMR both show acetonitrile peaks.

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 9.34 (s, 2H, 12-CH), 8.91 (d, *J* = 6.2 Hz, 2H, 11-CH), 8.82 (d, *J* = 8.4 Hz, 2H, 9-CH), 8.15 (dd, *J* = 8.2, 6.0 Hz, 2H, 10-CH), 7.77 (dd, *J* = 8.2, 2.0 Hz, 2H, 5-CH), 7.74 (d, *J* = 2.0 Hz, 2H, 3-CH), 7.21 (d, *J* = 8.2 Hz, 2H, 6-CH), 4.36 (s, 6H, 13-CH<sub>3</sub>), 3.08 – 2.98 (m, 4H, 1-CH<sub>2</sub>).

<sup>13</sup>C NMR (101 MHz, DMSO-*d*<sub>6</sub>) δ 156.1 (7-C), 143.8 (11-CH), 143.5 (12-CH), 141.8 (9-CH), 137.7 (8-C), 131.9 (4-C), 129.1 (3-CH, 2-C), 127.51 (10-CH), 126.0 (5-CH), 119.8 (6-CH), 48.0 (13-CH<sub>3</sub>), 30.8 (1-CH<sub>2</sub>).

ESI-HRMS C<sub>26</sub>H<sub>24</sub>N<sub>4</sub>I<sup>+</sup> [M]<sup>+</sup> calc: 519.1046, found: 519.1044.

(*E*)-3,3'-(11,12-dihydrodibenzo[*c,g*][1,2]diazocine-2,9-diyl)bis(1-methylpyridin-1-ium) iodide (*E*-**43**)

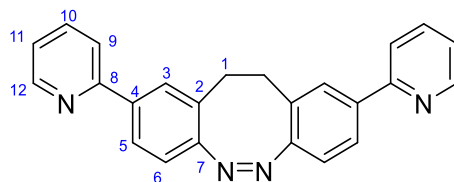
<sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ 9.47 (t, *J* = 1.6 Hz, 2H, 12-CH), 9.01 – 8.96 (m, 2H, 11-CH), 8.94 (dt, *J* = 8.4, 1.5 Hz, 2H, 9-CH), 8.23 (dd, *J* = 8.2, 6.0 Hz, 2H, 10-CH), 8.00 (dd, *J* = 8.1, 1.9 Hz, 2H, 5-CH), 7.94 (d, *J* = 8.1 Hz, 2H, 6-CH), 7.84 (d, *J* = 1.9 Hz, 2H, 3-CH), 4.44 (s, 6H, 13-CH<sub>3</sub>), 3.00 – 2.82 (m, 4H, 1-CH<sub>2</sub>).

<sup>13</sup>C NMR (151 MHz, DMSO-*d*<sub>6</sub>) δ 157.6 (7-C), 143.9 (11-CH), 143.8 (12-CH), 142.2 (9-CH), 138.5 (8-C), 135.5 (4-C), 132.6 (2-C), 130.5 (3-CH), 127.7 (10-CH), 126.8 (5-CH), 119.9 (6-CH), 48.2 (13-CH<sub>3</sub>), 31.7 (1-CH<sub>2</sub>).

## 7 Experimental

---

### (Z)-2,9-di(pyridin-2-yl)-11,12-dihydrodibenzo[*c,g*][1,2]diazocine (**39**)



Prepared according to a modified literature procedure.<sup>139</sup>

To a dry 3-neck-flask under nitrogen was added 2M isopropylmagnesium chloride in THF (0.9 mL, 1.8 mmol). Then 2-bromopyridine (0.16 mL, 1.68 mmol) was added dropwise, turning the solution yellow. This was stirred at room temperature for 3 h, upon which the solution turned brown. To another dry flask was added ZnCl<sub>2</sub> (300 mg, 2.2 mmol) dissolved in anhydrous THF (4 mL). This solution was added dropwise to the Grignard solution, where the colour turned purple. To a third flask, was added Pd(dba)<sub>2</sub> (25 mg, 0.0437 mmol), dicyclohexyl[2',4',6'-tris(propan-2-yl)[1,1'-biphenyl]-2-yl]phosphane (42 mg, 0.088 mmol) and anhydrous THF (2.7 mL). This was heated to 65 °C, then the diazocine **36** (200 mg, 0.546 mmol) was added, turning the dark red-purple solution green. The was stirred for 15 min, then the organozinc mixture was added dropwise, turning the solution brown. This was stirred overnight at 65 °C.

The reaction was quenched with saturated NaHCO<sub>3</sub> (3 mL) and H<sub>2</sub>O (3 mL). Then was extracted with EtOAc and DCM. The combined organic layers were dried with MgSO<sub>4</sub> and then dried *in vacuo*. Purification by column chromatography (CHCl<sub>3</sub>:MeOH, 98:2) gave the product as a yellow solid (136.7 mg, 0.377 mmol, 69%).

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.60 (ddd, *J* = 4.9, 1.8, 1.0 Hz, 2H, 12-CH), 7.72 (dd, *J* = 8.2, 1.9 Hz, 2H, 5-CH), 7.69 (d, *J* = 1.8 Hz, 2H, 3-CH), 7.67 (ddd, *J* = 8.0, 7.4, 1.9 Hz, 2H, 10-CH), 7.60 (dt, *J* = 8.0, 1.1 Hz, 2H, 9-CH), 7.16 (ddd, *J* = 7.4, 4.8, 1.2 Hz, 2H, 11-CH), 6.97 (d, *J* = 8.2 Hz, 2H, 6-CH), 3.11 – 2.92 (m, 4H, 1-CH<sub>2</sub>).

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 156.4 (Ar-C), 156.2 (Ar-C), 149.7 (12-CH), 138.2 (4-C), 136.8 (10-CH), 128.7 (2-C), 128.5 (3-CH), 125.4 (5-CH), 122.4 (11-CH), 120.5 (9-CH), 119.5 (6-CH), 32.0 (1-CH<sub>2</sub>).

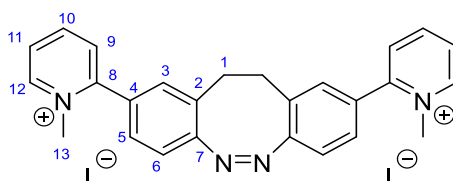
$\nu_{\max}$  /cm<sup>-1</sup> (compressed solid) 2981 (w), 1597 (m), 1460 (s), 1435 (m), 1156 (w), 786 (m), 773 (s), 742 (m).

ESI-HRMS C<sub>24</sub>H<sub>19</sub>N<sub>4</sub><sup>+</sup> [M+H]<sup>+</sup> calc: 363.1604, found: 363.1610.

ESI-HRMS C<sub>24</sub>H<sub>18</sub>N<sub>4</sub>Na<sup>+</sup> [M+Na]<sup>+</sup> calc: 385.1424, found: 385.1417.

## 7 Experimental

(*Z*)-2,2'-(11,12-dihydrodibenzo[*c,g*][1,2]diazocine-2,9-diyl)bis(1-methylpyridin-1-ium) iodide (**44**)



To a sealed microwave vial was added disubstituted diazocine **39** (74 mg, 0.204 mmol) dissolved in  $\text{CHCl}_3$  (6 mL). To this was added  $\text{CH}_3\text{I}$  (59  $\mu\text{L}$ , 0.818 mmol) which was heated at 50 °C overnight. The following day, another 7.8 equivalents of  $\text{CH}_3\text{I}$  (100  $\mu\text{L}$ , 1.61 mmol) was added and was heated overnight at 70 °C, this caused a precipitate to form. The following day, another 7.8 equiv of  $\text{CH}_3\text{I}$  (100  $\mu\text{L}$ , 1.61 mmol) was added and the reaction continued to stir at 70 °C overnight. The following day, the precipitate was collected by vacuum filtration to give an orange solid and a brown solid. The orange solid was the desired product (27.4 mg, 0.042 mmol, 21%).

$^1\text{H NMR}$  (400 MHz,  $\text{DMSO-}d_6$ )  $\delta$  9.10 (dd,  $J = 6.3, 1.4$  Hz, 2H, 12-CH), 8.60 (td,  $J = 7.8, 1.5$  Hz, 2H, 10-CH), 8.14 (ddd,  $J = 7.8, 6.1, 1.6$  Hz, 2H, 11-CH), 7.97 (dd,  $J = 8.0, 1.6$  Hz, 2H, 9-CH), 7.56 (dd,  $J = 8.1, 1.9$  Hz, 2H, 5-CH), 7.47 (d,  $J = 1.9$  Hz, 2H, 3-CH), 7.23 (d,  $J = 8.1$  Hz, 2H, 6-CH), 3.99 (s, 6H, 13- $\text{CH}_3$ ), 3.12 – 2.92 (m, 4H, 1- $\text{CH}_2$ ). Includes chloroform.

$^{13}\text{C NMR}$  (151 MHz,  $\text{DMSO-}d_6$ )  $\delta$  156.8 (7-C), 154.1 (8-C), 146.9 (12-CH), 145.5 (10-CH), 131.1 (3-CH), 130.7 (4-C), 129.9 (9-CH), 129.0 (2-C), 128.3 (5-CH), 126.9 (11-CH), 119.3 (6-CH), 47.0 (13- $\text{CH}_3$ ), 30.7 (1- $\text{CH}_2$ ).

**Nanospray-HRMS**  $\text{C}_{26}\text{H}_{24}\text{N}_4^{2+}$   $[\text{M}]^{2+}$  calc: 196.1001, found: 196.1004.

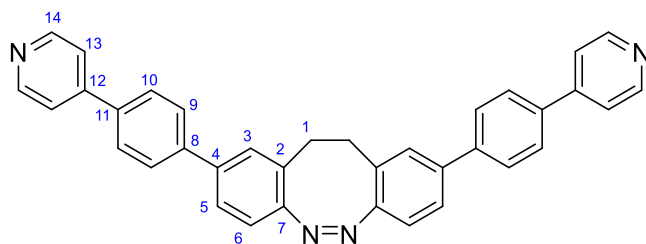
(*E*)-2,2'-(11,12-dihydrodibenzo[*c,g*][1,2]diazocine-2,9-diyl)bis(1-methylpyridin-1-ium) iodide (*E*-**44**)

$^1\text{H NMR}$  (600 MHz,  $\text{DMSO-}d_6$ )  $\delta$  9.13 (dd,  $J = 6.3, 1.4$  Hz, 2H, 12-CH), 8.65 (td,  $J = 7.9, 1.4$  Hz, 2H, 10-CH), 8.16 (ddd,  $J = 7.8, 6.1, 1.5$  Hz, 2H, 11-CH), 8.14 – 8.12 (m, 2H, 9-CH), 7.97 (d,  $J = 8.2$  Hz, 2H, 6-CH), 7.78 (dd,  $J = 8.0, 1.8$  Hz, 2H, 5-CH), 7.58 (d,  $J = 1.8$  Hz, 2H, 3-CH), 4.19 (s, 6H, 13- $\text{CH}_3$ ), 2.95 – 2.81 (m, 4H, 1- $\text{CH}_2$ ).

## 7 Experimental

---

### (Z)-2,9-bis(4-(pyridin-4-yl)phenyl)-11,12-dihydrodibenzo[c,g][1,2]diazocine (**40**)



Prepared according to General Procedure for Suzuki cross coupling. Using diazocine **36** (200 mg, 0.546 mmol), (4-(pyridin-4-yl)phenyl)boronic acid (326 mg, 1.64 mmol), Pd(PPh<sub>3</sub>)<sub>4</sub> (66 mg, 0.057 mmol) and 1 drop of ethylene glycol in THF (8 mL). And a 2.5 M K<sub>2</sub>CO<sub>3</sub> solution (2 mL), degassed by bubbling nitrogen. The crude product was extracted with DCM and EtOAc, washed with brine and dried with MgSO<sub>4</sub>. Then the crude product was purified by column chromatography (CHCl<sub>3</sub>:MeOH, 98:2) to give the product as a yellow solid (168 mg, 0.346 mmol, 59%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.69 – 8.63 (m, 4H, 14-CH), 7.69 – 7.65 (m, 4H, 10-CH), 7.65 – 7.60 (m, 4H, 9-CH), 7.54 – 7.49 (m, 4H, 13-CH), 7.46 (dd, *J* = 7.9, 1.9 Hz, 2H, 5-CH), 7.30 (d, *J* = 1.9 Hz, 2H, 3-CH), 7.02 (d, *J* = 8.1 Hz, 2H, 6-CH), 3.25 – 2.78 (m, 4H, 1-CH<sub>2</sub>).

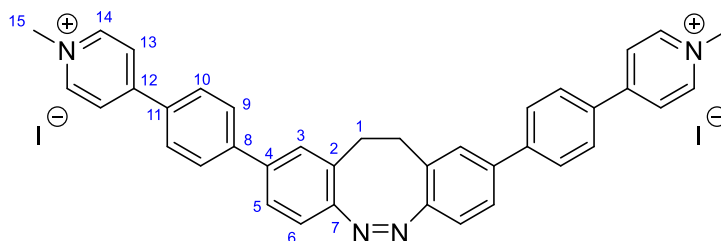
<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 155.1 (7-C), 150.4 (14-CH), 147.7 (12-C), 140.8 (8-C), 139.1 (4-C), 137.3 (11-C), 128.7 (2-C), 128.4 (3-CH), 127.7 (10-CH), 127.5 (9-CH), 125.6 (5-CH), 121.5 (13-CH), 120.0 (6-CH), 32.1 (1-CH<sub>2</sub>).

$\nu_{\max}$  /cm<sup>-1</sup> (compressed solid) 3451 (w), 3369 (w), 3035 (w) 1918 (w), 1592(s), 1475 (m), 1405 (m), 1217 (w), 853 (w), 804 (s), 752 (s).

ESI-HRMS C<sub>36</sub>H<sub>27</sub>N<sub>4</sub><sup>+</sup> [M+H]<sup>+</sup> calc: 515.2230, found: 515.2236.

## 7 Experimental

(Z)-4,4'-((11,12-dihydrodibenzo[c,g][1,2]diazocine-2,9-diyl)bis(4,1-phenylene))bis(1-methylpyridin-1-ium) iodide (**45**)



To a sealed microwave vial was added disubstituted diazocine **40** (70.6 mg, 0.137 mmol), suspended in  $\text{CHCl}_3$  (4 mL) and  $\text{CH}_3\text{CN}$  (2 mL). To the suspension was added  $\text{CH}_3\text{I}$  (35  $\mu\text{L}$ , 0.549 mmol) which was heated at 50 °C overnight. The following day, another 4 equivalents of  $\text{CH}_3\text{I}$  (35  $\mu\text{L}$ , 0.549 mmol) was added and was heated overnight at 50 °C. The precipitate was collected by vacuum filtration and washing with  $\text{Et}_2\text{O}$  to give the product as a yellow solid (109.3, 0.137 mmol, 99%).  $^1\text{H}$  and  $^{13}\text{C}$  NMR both show chloroform peaks.

$^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ )  $\delta$  8.98 (d,  $J$  = 6.9 Hz, 4H, 14-CH), 8.53 (d,  $J$  = 7.0 Hz, 4H, 13-CH), 8.14 (d,  $J$  = 8.6 Hz, 4H, 10-CH), 7.91 (d,  $J$  = 8.6 Hz, 4H, 9-CH), 7.69 (dd,  $J$  = 8.2, 1.9 Hz, 2H, 5-CH), 7.64 (d,  $J$  = 1.9 Hz, 2H, 3-CH), 7.10 (d,  $J$  = 8.2 Hz, 2H, 6-CH), 4.32 (s, 6H, 15- $\text{CH}_3$ ), 3.07 – 2.97 (m, 4H, 1- $\text{CH}_2$ ).

$^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO}-d_6$ )  $\delta$  155.1 (7-C), 153.4 (12-C), 145.5 (14-CH), 142.2 (8-C), 137.2 (4-C), 132.4 (11-C), 128.7 (2-C), 128.6 (10-CH), 128.4 (3-CH), 127.6 (9-CH), 125.5 (5-CH), 123.8 (13-CH), 119.8 (6-CH), 47.0 (15- $\text{CH}_3$ ), 31.1 (1- $\text{CH}_2$ ).

$\nu_{\text{max}}/\text{cm}^{-1}$  (compressed solid) 3481 (w), 3420 (w), 3027 (w), 1639 (s), 1605 (s), 1480, (m), 1469 (m), 1296 (m), 1190 (m), 818 (s), 800 (s), 751 (m).

Nanospray-HRMS  $\text{C}_{38}\text{H}_{32}\text{N}_4^{2+}$   $[\text{M}]^{2+}$  calc: 272.1313, found: 272.1306.

(E)-4,4'-((11,12-dihydrodibenzo[c,g][1,2]diazocine-2,9-diyl)bis(4,1-phenylene))bis(1-methylpyridin-1-ium) iodide (**E-45**)

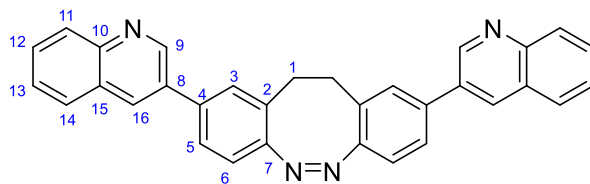
$^1\text{H}$  NMR (600 MHz,  $\text{DMSO}-d_6$ )  $\delta$  9.01 – 8.97 (m, 4H, 14-CH), 8.58 – 8.53 (m, 4H, 13-CH), 8.23 – 8.18 (m, 4H, 10-CH), 8.04 – 7.99 (m, 4H, 9-CH), 7.89 – 7.87 (m, 2H, 5-CH), 7.81 (d,  $J$  = 8.1 Hz, 2H, 6-CH), 7.70 (d,  $J$  = 1.9 Hz, 2H, 3-CH), 4.34 (s, 6H, 15- $\text{CH}_3$ ), 2.95 – 2.80 (m, 4H, 1- $\text{CH}_2$ ).

$^{13}\text{C}$  NMR (151 MHz,  $\text{DMSO}-d_6$ )  $\delta$  157.0 (7-C), 153.8 (12-C), 145.7 (14-CH), 142.9 (8-C), 138.0 (4-C), 135.6 (C), 132.6 (C), 130.0 (3-CH), 128.9 (10-CH), 128.0 (9-CH), 126.2 (5-CH), 124.0 (13-CH), 119.9 (6-CH), 47.2 (15- $\text{CH}_3$ ), 32.0 (1- $\text{CH}_2$ ).



## 7 Experimental

### (Z)-2,9-di(quinolin-3-yl)-11,12-dihydrodibenzo[c,g][1,2]diazocine (**41**)



Prepared according to General Procedure for Suzuki cross coupling. Using diazocine **36** (200 mg, 0.546 mmol), quinolin-3-ylboronic acid (285 mg, 1.65 mmol), Pd(PPh<sub>3</sub>)<sub>4</sub> (66 mg, 0.057 mmol) and 1 drop of ethylene glycol in THF (8 mL). And a 2.5 M K<sub>2</sub>CO<sub>3</sub> solution (2 mL), degassed by bubbling nitrogen. The crude product was extracted with DCM, washed with brine and dried with MgSO<sub>4</sub>. Then the crude product was purified by column chromatography three times (DCM:MeOH, 98:2), (DCM:MeOH, 98:2), (Et<sub>2</sub>O:MeOH, 98:2 to DCM:MeOH, 98:2) to remove a fluorescent blue impurity. The final product contained ~30% triphenylphosphine oxide impurity (by <sup>1</sup>H NMR) and the desired product as a bright yellow solid (182.7 mg, <0.39 mmol, <72%).

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 9.05 (d, *J* = 2.3 Hz, 2H, 9-CH), 8.18 (d, *J* = 2.3 Hz, 2H, 16-CH), 8.08 (d, *J* = 8.4 Hz, 2H, 11-CH), 7.79 (dd, *J* = 8.2, 1.4 Hz, 2H, 14-CH), 7.71 – 7.62 (m, 2H, 12-CH), 7.57 – 7.48 (m, 4H, 5-CH, 13-CH), 7.38 (d, *J* = 1.8 Hz, 2H, 3-CH), 7.07 (d, *J* = 8.1 Hz, 2H, 6-CH), 3.17 – 2.95 (m, 4H, 1-CH<sub>2</sub>).

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 155.3 (7-C), 149.5 (9-CH), 147.4 (10-C), 136.9 (4-C), 133.3 (16-CH), 132.7 (8-C), 129.7 (12-CH), 129.2 (11-CH), 129.0 (2-CH), 128.8 (3-CH), 128.0 (14-CH), 127.9 (15-CH), 127.2 (13-CH), 126.0 (5-CH), 120.2 (6-CH), 32.0 (1-CH<sub>2</sub>).

#### Triphenylphosphine oxide impurity<sup>235</sup>

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.71 – 7.62 (m, 6H), 7.57 – 7.48 (m, 3H), 7.45 (ddd, *J* = 8.4, 6.9, 3.0 Hz, 6H).

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 132.6 (d, *J* = 104.1 Hz), 132.2 (d, *J* = 9.9 Hz), 132.0 (d, *J* = 2.8 Hz), 128.6 (d, *J* = 12.1 Hz)

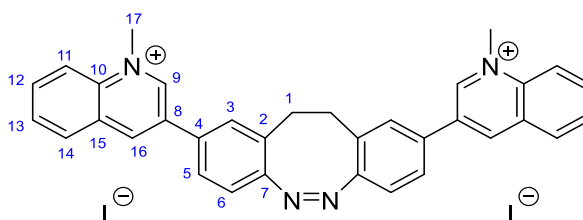
<sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 31.1.

$\nu_{\text{max}}$  /cm<sup>-1</sup> (compressed solid) 2981 (w), 1490 (w), 1437 (w), 1190 (m), 1121 (m), 955 (m), 806 (m), 750 (s), 721 (m).

ESI-HRMS C<sub>32</sub>H<sub>23</sub>N<sub>4</sub><sup>+</sup> [M+H]<sup>+</sup> calc: 463.1917, found: 463.1936.

## 7 Experimental

(*Z*)-3,3'-(11,12-dihydrodibenzo[*c,g*][1,2]diazocine-2,9-diyl)bis(1-methylquinolin-1-ium) iodide (**46**)



To a sealed microwave vial was added disubstituted diazocine **41** (63 mg, 0.136 mmol), suspended in  $\text{CHCl}_3$  (4 mL) and  $\text{CH}_3\text{CN}$  (2 mL). To the suspension was added  $\text{CH}_3\text{I}$  (35  $\mu\text{L}$ , 0.549 mmol) which was heated at 70 °C overnight. The following day, another 4 equivalents of  $\text{CH}_3\text{I}$  (35  $\mu\text{L}$ , 0.549 mmol) was added and was heated overnight at 70 °C. The precipitate was collected by vacuum filtration and washing with  $\text{CH}_3\text{CN}$ ,  $\text{CHCl}_3$ , and  $\text{Et}_2\text{O}$  to give the product as an orange solid (72 mg, 0.096, 71%).  $^1\text{H}$  NMR shows chloroform and acetonitrile peaks.  $^{31}\text{P}$  NMR shows no phosphorous species. HPLC shows 92% purity, with 8% monomethylated product.

$^1\text{H}$  NMR (500 MHz,  $\text{DMSO}-d_6$ )  $\delta$  9.83 (d,  $J = 2.0$  Hz, 2H, 9-CH), 9.50 (d,  $J = 2.1$  Hz, 2H, 16-CH), 8.46 (dd,  $J = 9.0, 0.9$  Hz, 2H, 11-CH), 8.40 (dd,  $J = 8.3, 1.4$  Hz, 2H, 14-CH), 8.23 (ddd,  $J = 8.8, 7.0, 1.5$  Hz, 2H, 12-CH), 8.04 (ddd,  $J = 8.1, 7.0, 0.9$  Hz, 2H, 13-CH), 7.89 (dd,  $J = 8.2, 2.0$  Hz, 2H, 5-CH), 7.85 (d,  $J = 2.0$  Hz, 2H, 3-CH), 7.26 (d,  $J = 8.2$  Hz, 2H, 6-CH), 4.63 (s, 6H, 17- $\text{CH}_3$ ), 3.08 (d,  $J = 1.6$  Hz, 4H, 1- $\text{CH}_2$ ).

$^{13}\text{C}$  NMR (126 MHz,  $\text{DMSO}-d_6$ )  $\delta$  156.1 (7-C), 149.3 (9-CH), 142.9 (16-CH), 137.3 (10-C), 135.5 (12-CH), 132.6 (4-C), 132.4 (8-C), 130.6 (14-CH), 130.5 (13-CH), 129.4 (2-C), 129.3 (15-C), 129.1 (3-CH), 126.3 (5-CH), 120.3 (6-CH), 119.1 (11-CH), 45.6 (17- $\text{CH}_3$ ), 31.3 (1- $\text{CH}_2$ ).

$\nu_{\text{max}}$  / $\text{cm}^{-1}$  (compressed solid) 2971 (w), 1529 (m), 1350 (m), 1130 (w), 812 (m), 770 (s), 742 (s), 679 (m).

Nanospray-HRMS  $\text{C}_{34}\text{H}_{28}\text{N}_4^{2+}$  [M] $^{2+}$  calc: 246.1157, found: 246.1146.

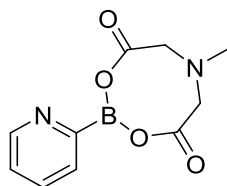
(*E*)-3,3'-(11,12-dihydrodibenzo[*c,g*][1,2]diazocine-2,9-diyl)bis(1-methylquinolin-1-ium) iodide (*E*-**46**)

$^1\text{H}$  NMR (600 MHz,  $\text{DMSO}-d_6$ )  $\delta$  9.99 (d,  $J = 2.0$  Hz, 2H, 9-CH), 9.65 (d,  $J = 2.0$  Hz, 2H, 16-CH), 8.53 (d,  $J = 8.9$  Hz, 2H, 11-CH), 8.50 (dd,  $J = 8.4, 1.4$  Hz, 2H, 14-CH), 8.29 (ddd,  $J = 8.8, 6.9, 1.5$  Hz, 2H, 12-CH), 8.15 – 8.08 (m, 4H, 5-CH and 13-CH), 7.99 (d,  $J = 8.0$  Hz, 2H, 6-CH), 7.94 (d,  $J = 1.9$  Hz, 2H, 3-CH), 4.72 (s, 6H, 17- $\text{CH}_3$ ), 2.99 (d,  $J = 25.9$  Hz, 4H, 1- $\text{CH}_2$ ).

## 7 Experimental

---

### 6-methyl-2-(pyridin-2-yl)-1,3,6,2-dioxazaborocane-4,8-dione (**48**)



A solution of 2-bromopyridine (0.86 mL, 9 mmol) and isopropylborate (2.3 mL, 9.97 mmol) in anhydrous THF (17 mL) was cooled to  $-78\text{ }^{\circ}\text{C}$  under a nitrogen atmosphere. To this solution 2.5 M *n*-butyllithium (3.8 mL, 9.5 mmol) was added dropwise and then the yellow solution was stirred at  $-78\text{ }^{\circ}\text{C}$  for 1 h. Then the reaction was warmed to room temperature and stirred for another 3 h.

To a separate 3-neck flask, equipped with a pressure equalising addition funnel, a short-path distillation apparatus and a thermometer, *N*-methyliminodiacetic acid (2.5 g, 16.99 mmol) was dissolved in anhydrous DMSO (17 mL) at  $115\text{ }^{\circ}\text{C}$ . The boronate ester solution was added to the addition funnel and washed with additional THF (9 mL), this solution was added dropwise to the hot DMSO solution at a rate to maintain the internal temperature at  $110\text{--}120\text{ }^{\circ}\text{C}$ . Upon addition, THF was immediately distilled. Following addition, the reaction was cooled to  $70\text{ }^{\circ}\text{C}$  and the bulk of the DMSO was vacuum distilled. The remaining residue was left under vacuum overnight, to remove additional DMSO. Then acetonitrile was added to the residue to create a suspension which was adsorbed onto celite and dried overnight under vacuum to remove further DMSO. The adsorbed compound was then purified by column chromatography ( $\text{Et}_2\text{O}:\text{CH}_3\text{CN}$ , 90:10 to 5:95) to achieve the product as a white solid (1.045 g, 4.46 mmol, 50%).

$^1\text{H NMR}$  (400 MHz, Acetonitrile- $d_3$ )  $\delta$  8.68 (ddd,  $J = 4.8, 1.7, 1.1\text{ Hz}$ , 1H), 7.71 (td,  $J = 7.6, 1.8\text{ Hz}$ , 1H), 7.63 (dt,  $J = 7.5, 1.3\text{ Hz}$ , 1H), 7.29 (ddd,  $J = 7.6, 4.8, 1.4\text{ Hz}$ , 1H), 4.10 (d,  $J = 16.7\text{ Hz}$ , 2H), 3.99 (d,  $J = 16.8\text{ Hz}$ , 2H), 2.55 (s, 3H).

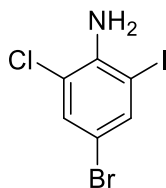
$\nu_{\text{max}}/\text{cm}^{-1}$  (compressed solid) 3451 (w), 3369 (w), 1746 (s), 1614 (m), 1467 (m), 1454 (m), 1303 (s), 1282 (s), 1053 (s) 998 (s), 867 (s), 774 (s), 681 (m).

Proton NMR was consistent with literature data.<sup>136</sup>

## 7 Experimental

---

### 4-bromo-2-chloro-6-iodoaniline (**54**)



4-bromo-2-chloroaniline (1.026 g, 4.97 mmol), iodine (1.264 g, 4.98 mmol) and  $\text{Ag}_2\text{SO}_4$  (1.555 g, 4.99 mmol) were dissolved in EtOH (16 mL) sequentially to give a dark brown solution. The reaction was stirred for 1.5 h at room temperature, until TLC indicated disappearance of the starting material. The reaction was filtered to remove the solid precipitate by washing with EtOH and then the solvent was concentrated *in vacuo* to give a brown solid. This was redissolved in DCM (40 mL) and then washed with 1M NaOH (25 mL) and water (25 mL). The organic phase was dried with  $\text{MgSO}_4$  and then concentrated *in vacuo*. The resulting brown crude product was purified by column chromatography (hexane) to give the product as a white fluffy solid (1.531 g, 4.61 mmol, 92%).

$^1\text{H NMR}$  (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.66 (d,  $J = 2.2$  Hz, 1H), 7.38 (d,  $J = 2.1$  Hz, 1H), 4.54 (br, 2H,  $\text{NH}_2$ ).

$^{13}\text{C NMR}$  (101 MHz,  $\text{CDCl}_3$ )  $\delta$  142.8, 139.3, 132.0, 118.1, 109.2, 83.3.

$\nu_{\text{max}}/\text{cm}^{-1}$  (compressed solid) 3405 (w), 3307 (br), 1727 (w), 1606 (m), 1541 (m), 1447 (s), 1381 (m), 1289 (w), 1274 (w), 1065 (m), 1051 (m), 859 (s), 844 (m), 745 (m), 708 (m).

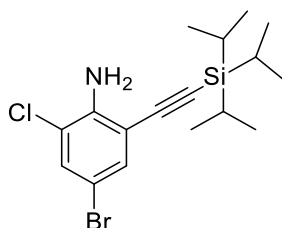
**APCI LRMS**  $\text{C}_6\text{H}_5\text{BrClIN}^+$   $[\text{M}+\text{H}]^+$  calc: 331.8, found: 331.8.

Proton and carbon NMR was consistent with literature data.<sup>236</sup>

## 7 Experimental

---

### 4-bromo-2-chloro-6-((triisopropylsilyl)ethynyl)aniline (**55**)



To an oven dried flask, iodoaniline **54** (1.504 g, 4.53 mmol) CuI (42.7 mg, 0.224 mmol), Pd(PPh<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub> (158.2 mg, 0.225 mmol) was added by flushing N<sub>2</sub> and suspended in anhydrous THF (20 mL). Then distilled Et<sub>3</sub>N (1.9 mL, 13.6 mmol) and (triisopropylsilyl)acetylene (1.3 mL, 5.80 mmol) were added and stirred at room temperature until TLC showed disappearance of the starting material. The brown solution was then diluted with Et<sub>2</sub>O and then filtered through celite. The filtrate was concentrated *in vacuo* and the crude product was purified by column chromatography (hexane) to give the product as a colourless oil (1.643 g, 94%, 4.25 mmol).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.34 (s, 2H, CH and CH), 4.65 (br, 2H, NH<sub>2</sub>) 1.14 – 1.12 (m, 21H, -CH(CH<sub>3</sub>)<sub>2</sub>).

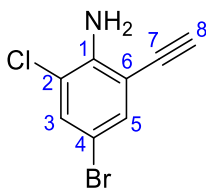
<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 144.2, 133.1, 132.1, 119.4, 110.8, 107.9, 101.47, 98.9, 18.8, 11.3.

ESI HRMS C<sub>17</sub>H<sub>25</sub>BrClINaSi<sup>+</sup> [M+Na]<sup>+</sup> calc: 408.0526 found: 408.3079.

## 7 Experimental

---

### 4-bromo-2-chloro-6-ethynylaniline (**56**)



TIPS-aniline **55** (1.643 g, 4.25 mmol) was dissolved in THF (9 mL) and then 1 M TBAF (4.7 mL, 4.7 mmol) was added. After 50 min, TLC showed loss of starting material and the solvent was concentrated *in vacuo* to give a yellow oil. This was diluted in DCM (50 mL) and washed with water (2 x 25 mL). The aqueous phase was extracted with DCM (15 mL) and the combined organic phases were dried with MgSO<sub>4</sub> and concentrated *in vacuo* to give the crude product as an off-white solid. This was purified by column chromatography (hexane) to give the product as a white solid (934 mg, 4.06 mmol, 96%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.37 (d, *J* = 2.2 Hz, 1H, 3-CH), 7.36 (d, *J* = 2.2 Hz, 1H, 5-CH), 4.65 (br, 2H, NH<sub>2</sub>), 3.46 (s, 1H, 8-CH).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 144.5 (1-C), 133.4 (5-CH), 132.6 (3-CH), 119.5 (2-C), 109.1 (6-C), 107.8 (4-C), 84.4 (8-CH), 78.7 (7-C).

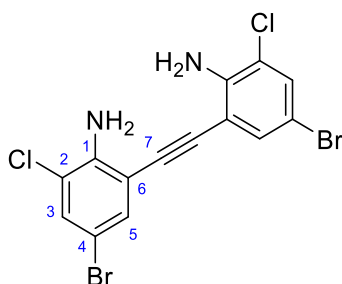
$\nu_{\max}$  /cm<sup>-1</sup> (compressed solid) 3432 (w), 3280 (m), 3054 (w), 2922 (w), 2102 (w), 1747 (w), 1613 (m), 1456 (m), 1071 (m), 864 (s), 667 (s), 614 (s).

APCI HRMS C<sub>8</sub>H<sub>6</sub>BrClN<sup>+</sup> [M+H]<sup>+</sup> calc: 229.9367, found: 229.9361.

## 7 Experimental

---

### 6,6'-(ethyne-1,2-diyl)bis(4-bromo-2-chloroaniline) (**52**)



To iodoaniline **54** (1.132 g, 3.41 mmol), alkyne **56** (861.7 mg, 3.74 mmol), Pd(PPh<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub> (120.3 mg, 0.171 mmol) and CuI (30 mg, 0.158 mmol) was added anhydrous THF (6 mL) and distilled Et<sub>3</sub>N (1.4 mL, 10.0 mmol). The yellow suspension was stirred at room temperature under N<sub>2</sub> and turned dark brown. After 3 h of stirring, the reaction was diluted by EtOAc and filtered through celite.

A pale pink solid was isolated above the celite pad which has shown to be the desired product by NMR. The filtrate was dried to give a greenish solid which was dissolved in acetone and precipitated with water. The yellow precipitate was filtered, dried and gave a lower purity product. Combined mass of pale pink solid and yellow solid (1.063 g, 2.44 mmol, 71%).

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 9.06 (s, Et<sub>3</sub>NH), 7.76 (d, *J* = 2.3 Hz, 2H, 5-CH), 7.47 (d, *J* = 2.4 Hz, 2H, 3-CH), 5.91 (s, 4H, NH<sub>2</sub>), 3.09 (q, Et<sub>3</sub>NH), 1.17 (t, Et<sub>3</sub>NH).

<sup>13</sup>C NMR (101 MHz, DMSO-*d*<sub>6</sub>) δ 144.6 (1-C), 133.4 (5-CH), 131.6 (3-CH), 118.2 (2-C), 108.9 (6-C), 105.2 (4-C), 91.0 (7-C), 45.7 (Et<sub>3</sub>NH), 8.6 (Et<sub>3</sub>NH).

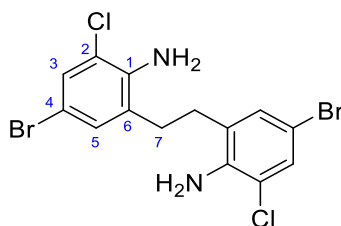
$\nu_{\text{max}}$  /cm<sup>-1</sup> (compressed solid) 3471 (w), 3376 (m), 1609 (s), 1580 (m), 1547 (w), 1463 (s), 1419 (m), 1283 (m), 1226 (m), 1072 (m), 853 (s), 843 (m), 721 (m).

APCI HRMS C<sub>14</sub>H<sub>9</sub>Br<sub>2</sub>Cl<sub>2</sub>N<sub>2</sub><sup>+</sup> [M+H]<sup>+</sup> calc: 432.8504, found: 432.8495.

## 7 Experimental

---

### 6,6'-(ethane-1,2-diyl)bis(4-bromo-2-chloroaniline) (**51**)



To a solution of dianiline **35** (2.5 g, 6.76 mmol) in DMF (10 mL) was added N-chlorosuccinimide (1.92 g, 14.17 mmol) in small portions. This solution was heated to 70 °C for 2 h then it was poured into brine and extracted with DCM to give a dark brown solid. The precipitate was filtered and purified by column chromatography (DCM:Hexane, 50:50 to 75:25 to 100:0) to give the product with about 80% purity as a brown solid (1.115 g, 2.54 mmol, 46%)

**<sup>1</sup>H NMR** (400 MHz, DMSO-*d*<sub>6</sub>) δ 7.30 (d, *J* = 2.3 Hz, 2H, 3-CH), 7.24 (d, *J* = 2.3 Hz, 2H, 5-CH), 5.34 (br, 4H, NH<sub>2</sub>), 2.74 (d, *J* = 2.0 Hz, 4H, 7-CH<sub>2</sub>).

**<sup>13</sup>C NMR** (101 MHz, DMSO-*D*<sub>6</sub>) δ 141.7 (1-C), 130.1 (5-C), 128.6 (C), 128.5 (3-C), 118.2 (C), 106.1 (C), 29.2 (7-CH<sub>2</sub>).

**v<sub>max</sub> /cm<sup>-1</sup>** (compressed solid) 3452 (w), 3369 (w), 2869 (w), 1614 (s), 1468 (s), 1455(s), 1407 (m), 1289 (m), 1197 (m), 1069 (m), 856 (s), 833 (m), 737 (m).

**APCI HRMS** C<sub>14</sub>H<sub>13</sub>Br<sub>2</sub>Cl<sub>2</sub>N<sub>2</sub><sup>+</sup> [M+H]<sup>+</sup> calc: 436.8817, found: 436.8819.

Main impurity: **<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.48 (d, *J* = 2.2 Hz, 2H), 7.11 (d, *J* = 2.2 Hz, 2H), 4.15 (br, 4H), 2.76 (s, 4H).



### Single-crystal X-Ray analysis of compound **36**



Single-crystals suitable for X-ray crystallography were grown by evaporation of a DCM solution of **36** at room temperature.

Chemical formula	C <sub>14</sub> H <sub>10</sub> Br <sub>2</sub> N <sub>2</sub>
<i>M<sub>r</sub></i> /g mol <sup>-1</sup>	366.06
Crystal system	orthorhombic
crystal size	0.379 x 0.202 x 0.08
<i>T</i> /K	100.0
radiation	MoK $\alpha$
diffractometer	Bruker D8 Venture dual source kappa CPAD
space group	P b c a
<i>a</i> /Å	15.7251
<i>b</i> /Å	9.5742
<i>c</i> /Å	16.9039
$\alpha$ /°	90
$\beta$ /°	90
$\gamma$ /°	90
<i>Z</i>	8
<i>V</i> /Å <sup>3</sup>	2544.97(11)
calc. density/g cm <sup>-3</sup>	1.911
$\mu$ /mm <sup>-1</sup>	6.352
absorption correction	multi-scan
transmission factor range	0.4413-0.7458
refls. measured	44391

## 7 Experimental

---

$R_{\text{int}}$	0.0415
mean $\sigma(I)/I$	0.0201
$\theta$ range	2.410-28.724
observed refls.	2828
$x, y$ (weighting scheme)	0.016400, 2.076000
hydrogen refinement	constr
refls in refinement	3293
parameters	163
restraints	0
$R(F_{\text{obs}})$	0.0203
$R_w(F^2)$	0.0437
$S$	1.028
shift/error <sub>max</sub>	0.001
max electron density/e $\text{\AA}^{-3}$	0.393
min electron density/e $\text{\AA}^{-3}$	-0.331

## References

- 1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
- 2 R. E. Franklin and R. G. Gosling, *Acta Crystallogr.*, 1953, **6**, 673–677.
- 3 Y. Mitsui, R. Langridge, B. E. Shortle, C. R. Cantor, R. C. Grant, M. Kodama and R. D. Wells, *Nature*, 1970, **228**, 1166–1169.
- 4 A. I. H. Murchie and D. M. J. Lilley, in *Methods in Enzymology*, 1992, pp. 158–180.
- 5 G. Felsenfeld, D. R. Davies and A. Rich, *J. Am. Chem. Soc.*, 1957, **79**, 2023–2024.
- 6 K. Gehring, J.-L. Leroy and M. Guéron, *Nature*, 1993, **363**, 561–565.
- 7 M. Gellert, M. N. Lipsett and D. R. Davies, *Proc. Natl. Acad. Sci.*, 1962, **48**, 2013–2018.
- 8 I. Bang, *Biochem. Z.*, 1910, **26**, 293.
- 9 E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco and E. H. Blackburn, *Cell*, 1987, **51**, 899–908.
- 10 W. I. Sundquist and A. Klug, *Nature*, 1989, **342**, 825–829.
- 11 D. Sen and W. Gilbert, *Nature*, 1988, **334**, 364–366.
- 12 D. Sen and W. Gilbert, *Nature*, 1990, **344**, 410–414.
- 13 J. T. Davis, *Angew. Chem. Int. Ed.*, 2004, **43**, 668–698.
- 14 A. Guédin, J. Gros, P. Alberti and J.-L. Mergny, *Nucleic Acids Res.*, 2010, **38**, 7858–7868.
- 15 F. R. Winnerdy and A. T. Phan, in *Annual Reports in Medicinal Chemistry*, 2020, pp. 45–73.
- 16 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2005, **33**, 2908–2916.
- 17 A. K. Todd, M. Johnston and S. Neidle, *Nucleic Acids Res.*, 2005, **33**, 2901–2907.
- 18 E. Puig Lombardi and A. Londoño-Vallejo, *Nucleic Acids Res.*, 2020, **48**, 1–15.
- 19 A. Bedrat, L. Lacroix and J.-L. Mergny, *Nucleic Acids Res.*, 2016, **44**, 1746–1759.
- 20 T. Q. Ngoc Nguyen, K. W. Lim and A. T. Phan, *Nucleic Acids Res.*, 2020, **48**, 10567–10575.
- 21 J. Fang, L. Zheng, Y. Liu, Y. Peng, Q. Yang, Y. Huang, J. Zhang, L. Luo, D. Shen, Y. Tan, X. Lu and G. Feng, *Int. J. Biol. Macromol.*, 2023, **247**, 125614.
- 22 P. M. Punt, M. D. Langenberg, O. Altan and G. H. Clever, *J. Am. Chem. Soc.*, 2021, **143**, 3555–3561.
- 23 B. Ruttkay-Nedecky, J. Kudr, L. Nejdil, D. Maskova, R. Kizek and V. Adam, *Molecules*, 2013, **18**, 14760–14779.
- 24 D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627–8637.

## References

---

- 25 M. Di Antonio, A. Ponjavic, A. Radzevičius, R. T. Ranasinghe, M. Catalano, X. Zhang, J. Shen, L. M. Needham, S. F. Lee, D. Klenerman and S. Balasubramanian, *Nat. Chem.*, 2020, **12**, 832–837.
- 26 G. Wang and K. M. Vasquez, *Nat. Rev. Genet.*, 2023, **24**, 211–234.
- 27 D. Varshney, J. Spiegel, K. Zyner, D. Tannahill and S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.*, 2020, **21**, 459–474.
- 28 B. D. Griffin and H. W. Bass, *Plant Sci.*, 2018, **269**, 143–147.
- 29 E. Belmonte-Reche, M. Martínez-García, A. Guédin, M. Zuffo, M. Arévalo-Ruiz, F. Doria, J. Campos-Salinas, M. Maynadier, J. J. López-Rubio, M. Freccero, J.-L. Mergny, J. M. Pérez-Victoria and J. C. Morales, *J. Med. Chem.*, 2018, **61**, 1231–1240.
- 30 R. R. Cueny, S. D. McMillan and J. L. Keck, *Crit. Rev. Biochem. Mol. Biol.*, 2022, **57**, 539–561.
- 31 E. Ruggiero and S. N. Richter, *Nucleic Acids Res.*, 2018, **46**, 3270–3283.
- 32 E. M. Smith, D. F. Pendlebury and J. Nandakumar, *Cell. Mol. Life Sci.*, 2020, **77**, 61–79.
- 33 N. Kim, M. Piatyszek, K. Prowse, C. Harley, M. West, P. Ho, G. Coviello, W. Wright, S. Weinrich and J. Shay, *Science*, 1994, **266**, 2011–2015.
- 34 W. C. Hahn, S. A. Stewart, M. W. Brooks, S. G. York, E. Eaton, A. Kurachi, R. L. Beijersbergen, J. H. M. Knoll, M. Meyerson and R. A. Weinberg, *Nat. Med.*, 1999, **5**, 1164–1170.
- 35 X. Zhang, V. Mar, W. Zhou, L. Harrington and M. O. Robinson, *Genes Dev.*, 1999, **13**, 2388–2399.
- 36 S. Neidle and G. Parkinson, *Nat. Rev. Drug Discov.*, 2002, **1**, 383–393.
- 37 K. M. Burchett, Y. Yan and M. M. Ouellette, *PLoS One*, 2014, **9**, e85155.
- 38 K. Damm, U. Hemmann, P. Garin-Chesa, N. Hael, I. Kauffmann, H. Priepke, C. Niestroj, C. Daiber, B. Enenkel, B. Guilliard, I. Lauritsch, E. Müller, E. Pascolo, G. Sauter, M. Pantic, U. M. Martens, C. Wenz, J. Lingner, N. Kraut, W. J. Rettig and A. Schnapp, *EMBO J.*, 2001, **20**, 6958–6968.
- 39 C. Bryan, C. Rice, H. Hoffman, M. Harkisheimer, M. Sweeney and E. Skordalakes, *Structure*, 2015, **23**, 1934–1942.
- 40 A. M. Zahler, J. R. Williamson, T. R. Cech and D. M. Prescott, *Nature*, 1991, **350**, 718–720.
- 41 D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S.

## References

---

- Neidle and L. H. Hurley, *J. Med. Chem.*, 1997, **40**, 2113–2116.
- 42 A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci.*, 2002, **99**, 11593–11598.
- 43 J. Seenisamy, E. M. Rezler, T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain and L. H. Hurley, *J. Am. Chem. Soc.*, 2004, **126**, 8702–8709.
- 44 S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2005, **127**, 10584–10589.
- 45 T. S. Dexheimer, D. Sun and L. H. Hurley, *J. Am. Chem. Soc.*, 2006, **128**, 5404–5415.
- 46 D. Sun, K. Guo, J. J. Rusche and L. H. Hurley, *Nucleic Acids Res.*, 2005, **33**, 6070–6080.
- 47 J. Lavrado, H. Brito, P. M. Borralho, S. A. Ohnmacht, N.-S. Kim, C. Leitão, S. Pisco, M. Gunaratnam, C. M. P. Rodrigues, R. Moreira, S. Neidle and A. Paulo, *Sci. Rep.*, 2015, **5**, 9696.
- 48 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discov.*, 2011, **10**, 261–275.
- 49 V. Brázda, O. Porubiaková, A. Cantara, N. Bohálová, J. Coufal, M. Bartas, M. Fojta and J.-L. Mergny, *BMC Genomics*, 2021, **22**, 77.
- 50 I. Frasson, P. Soldà, M. Nadai, S. Lago and S. N. Richter, *Commun. Biol.*, 2021, **4**, 510.
- 51 P. Murat, J. Zhong, L. Lekieffre, N. P. Cowieson, J. L. Clancy, T. Preiss, S. Balasubramanian, R. Khanna and J. Tellam, *Nat. Chem. Biol.*, 2014, **10**, 358–364.
- 52 G. Qin, C. Zhao, Y. Liu, C. Zhang, G. Yang, J. Yang, Z. Wang, C. Wang, C. Tu, Z. Guo, J. Ren and X. Qu, *Cell Discov.*, 2022, **8**, 86.
- 53 R. Perrone, M. Nadai, I. Frasson, J. A. Poe, E. Butovskaya, T. E. Smithgall, M. Palumbo, G. Palù and S. N. Richter, *J. Med. Chem.*, 2013, **56**, 6521–6530.
- 54 E. Ruggiero and S. N. Richter, in *Annual Reports in Medicinal Chemistry*, 2020, pp. 101–131.
- 55 M. S. Saag, R. T. Gandhi, J. F. Hoy, R. J. Landovitz, M. A. Thompson, P. E. Sax, D. M. Smith, C. A. Benson, S. P. Buchbinder, C. del Rio, J. J. Eron, G. Fätkenheuer, H. F. Günthard, J.-M. Molina, D. M. Jacobsen and P. A. Volberding, *JAMA*, 2020, **324**, 1651.
- 56 World Health Organization, <https://www.who.int/>, (accessed 13 October 2023).
- 57 E. Butovskaya, B. Heddi, B. Bakalar, S. N. Richter and A. T. Phan, *J. Am. Chem. Soc.*, 2018, **140**, 13654–13662.
- 58 R. Perrone, E. Butovskaya, D. Daelemans, G. Palu, C. Pannecouque and S. N. Richter, *J.*

## References

---

- Antimicrob. Chemother.*, 2014, **69**, 3248–3258.
- 59 D. Piekna-Przybylska, G. Sharma, S. B. Maggirwar and R. A. Bambara, *Cell Cycle*, 2017, **16**, 968–978.
- 60 P. Rawal, V. B. R. Kummarasetti, J. Ravindran, N. Kumar, K. Halder, R. Sharma, M. Mukerji, S. K. Das and S. Chowdhury, *Genome Res.*, 2006, **16**, 644–655.
- 61 V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith and S. Balasubramanian, *Nat. Biotechnol.*, 2015, **33**, 877–881.
- 62 G. Marsico, V. S. Chambers, A. B. Sahakyan, P. McCauley, J. M. Boutell, M. Di Antonio and S. Balasubramanian, *Nucleic Acids Res.*, 2019, **47**, 3862–3874.
- 63 S. K. Mishra, U. Shankar, N. Jain, K. Sikri, J. S. Tyagi, T. K. Sharma, J.-L. Mergny and A. Kumar, *Mol. Ther. - Nucleic Acids*, 2019, **16**, 698–706.
- 64 R. Perrone, E. Lavezzo, E. Riello, R. Manganelli, G. Palù, S. Toppo, R. Provvedi and S. N. Richter, *Sci. Rep.*, 2017, **7**, 5743.
- 65 D. Monchaud and M.-P. Teulade-Fichou, *Org. Biomol. Chem.*, 2008, **6**, 627–636.
- 66 S. Neidle, *J. Med. Chem.*, 2016, **59**, 5987–6011.
- 67 Y.-H. Wang, Q.-F. Yang, X. Lin, D. Chen, Z.-Y. Wang, B. Chen, H.-Y. Han, H.-D. Chen, K.-C. Cai, Q. Li, S. Yang, Y.-L. Tang and F. Li, *Nucleic Acids Res.*, 2022, **50**, D150–D160.
- 68 M. Read, R. J. Harrison, B. Romagnoli, F. A. Tanious, S. H. Gowan, A. P. Reszka, W. D. Wilson, L. R. Kelland and S. Neidle, *Proc. Natl. Acad. Sci.*, 2001, **98**, 4844–4849.
- 69 C. Martins, M. Gunaratnam, J. Stuart, V. Makwana, O. Greciano, A. P. Reszka, L. R. Kelland and S. Neidle, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 2293–2298.
- 70 L. Martino, B. Pagano, I. Fotticchia, S. Neidle and C. Giancola, *J. Phys. Chem. B*, 2009, **113**, 14779–14786.
- 71 S. Asamitsu, T. Bando and H. Sugiyama, *Chem. Eur. J.*, 2019, **25**, 417–430.
- 72 T. Santos, G. F. Salgado, E. J. Cabrita and C. Cruz, *Pharmaceuticals*, 2021, **14**, 769.
- 73 P. Lejault, J. Mitteau, F. R. Sperti and D. Monchaud, *Cell Chem. Biol.*, 2021, **28**, 436–455.
- 74 K. Shin-ya, K. Wierzba, K. Matsuo, T. Ohtani, Y. Yamada, K. Furihata, Y. Hayakawa and H. Seto, *J. Am. Chem. Soc.*, 2001, **123**, 1262–1263.
- 75 D. Gomez, T. Wenner, B. Brassart, C. Douarre, M.-F. O’Donohue, V. El Khoury, K. Shin-ya, H. Morjani, C. Trentesaux and J.-F. Riou, *J. Biol. Chem.*, 2006, **281**, 38721–38729.
- 76 H. Tahara, K. Shin-ya, H. Seimiya, H. Yamada, T. Tsuruo and T. Ide, *Oncogene*, 2006, **25**,

## References

---

- 1955–1966.
- 77 W. J. Chung, B. Heddi, M. Tera, K. Iida, K. Nagasawa and A. T. Phan, *J. Am. Chem. Soc.*, 2013, **135**, 13495–13501.
- 78 R. Rodriguez, S. Müller, J. A. Yeoman, C. Trentesaux, J.-F. Riou and S. Balasubramanian, *J. Am. Chem. Soc.*, 2008, **130**, 15758–15759.
- 79 A. De Cian, E. DeLemos, J.-L. Mergny, M.-P. Teulade-Fichou and D. Monchaud, *J. Am. Chem. Soc.*, 2007, **129**, 1856–1857.
- 80 R. Perrone, F. Doria, E. Butovskaya, I. Frasson, S. Botti, M. Scalabrin, S. Lago, V. Grande, M. Nadai, M. Freccero and S. N. Richter, *J. Med. Chem.*, 2015, **58**, 9639–9652.
- 81 W. Duan, A. Rangan, H. Vankayalapati, M.-Y. Kim, Q. Zeng, D. Sun, H. Han, O. Y. Fedoroff, D. Nishioka and S. Y. Rha, *Mol. Cancer Ther.*, 2001, **1**, 103–120.
- 82 H. Xu, M. Di Antonio, S. McKinney, V. Mathew, B. Ho, N. J. O’Neil, N. Dos Santos, J. Silvester, V. Wei, J. Garcia, F. Kabeer, D. Lai, P. Soriano, J. Banáth, D. S. Chiu, D. Yap, D. D. Le, F. B. Ye, A. Zhang, K. Thu, J. Soong, S. Lin, A. H. C. Tsai, T. Osako, T. Algara, D. N. Saunders, J. Wong, J. Xian, M. B. Bally, J. D. Brenton, G. W. Brown, S. P. Shah, D. Cescon, T. W. Mak, C. Caldas, P. C. Stirling, P. Hieter, S. Balasubramanian and S. Aparicio, *Nat. Commun.*, 2017, **8**, 14432.
- 83 H. Xu and L. H. Hurley, *Bioorg. Med. Chem. Lett.*, 2022, **77**, 129016.
- 84 D.-F. Shi, R. T. Wheelhouse, D. Sun and L. H. Hurley, *J. Med. Chem.*, 2001, **44**, 4509–4523.
- 85 J. E. Reed, A. A. Arnal, S. Neidle and R. Vilar, *J. Am. Chem. Soc.*, 2006, **128**, 5992–5993.
- 86 S. Bandeira, J. Gonzalez-Garcia, E. Pensa, T. Albrecht and R. Vilar, *Angew. Chem. Int. Ed.*, 2018, **57**, 310–313.
- 87 J. Spiegel, S. Adhikari and S. Balasubramanian, *Trends Chem.*, 2020, **2**, 123–136.
- 88 W. A. Velema, W. Szymanski and B. L. Feringa, *J. Am. Chem. Soc.*, 2014, **136**, 2178–2191.
- 89 T. Nakamura, K. Iida, M. Tera, K. Shin-ya, H. Seimiya and K. Nagasawa, *ChemBioChem*, 2012, **13**, 774–777.
- 90 P. Murat, M. V. Gormally, D. Sanders, M. Di Antonio and S. Balasubramanian, *Chem. Commun.*, 2013, **49**, 8453.
- 91 T. Kench, P. A. Summers, M. K. Kuimova, J. E. M. Lewis and R. Vilar, *Angew. Chem. Int. Ed.*, 2021, **60**, 10928–10934.
- 92 E. Hemmer, A. Benayas, F. Légaré and F. Vetrone, *Nanoscale Horizons*, 2016, **1**, 168–

## References

---

- 184.
- 93 I. M. Welleman, M. W. H. Hoorens, B. L. Feringa, H. H. Boersma and W. Szymański, *Chem. Sci.*, 2020, **11**, 11672–11691.
- 94 I. C. D. Merritt, D. Jacquemin and M. Vacher, *Phys. Chem. Chem. Phys.*, 2021, **23**, 19155–19165.
- 95 C. Boulègue, M. Löweneck, C. Renner and L. Moroder, *ChemBioChem*, 2007, **8**, 591–594.
- 96 X. Wang, J. Huang, Y. Zhou, S. Yan, X. Weng, X. Wu, M. Deng and X. Zhou, *Angew. Chem. Int. Ed.*, 2010, **49**, 5305–5309.
- 97 X. Xing, X. Wang, L. Xu, Y. Tai, L. Dai, X. Zheng, W. Mao, X. Xu and X. Zhou, *Org. Biomol. Chem.*, 2011, **9**, 6639–6645.
- 98 T. Tian, Y. Song, L. Wei, J. Wang, B. Fu, Z. He, X.-R. Yang, F. Wu, G. Xu, S.-M. Liu, C. Li, S. Wang and X. Zhou, *Nucleic Acids Res.*, 2017, **45**, 2283–2293.
- 99 M. Dudek, M. Deiana, Z. Pokladek, P. Mlynarz, M. Samoc and K. Matczyszyn, *Nanoscale*, 2018, **10**, 11302–11306.
- 100 M. Dudek, M. Deiana, Z. Pokladek, K. Pawlik and K. Matczyszyn, *Chem. – A Eur. J.*, 2018, **24**, 18963–18970.
- 101 M. Dudek, M. Deiana, K. Szkaradek, M. J. Janicki, Z. Pokladek, R. W. Góra and K. Matczyszyn, *J. Phys. Chem. Lett.*, 2021, **12**, 9436–9441.
- 102 J. E. B. McCallum, C. W. Coyle, R. R. Elson and B. A. Titterington, *Nucleosides, Nucleotides and Nucleic Acids*, 2018, **37**, 169–178.
- 103 H. Deng, H. Xu, Y. Wang, R. Jia, X. Ma, Y. Feng and H. Chen, *Nucleic Acids Res.*, 2023, **51**, 4064–4077.
- 104 H. Deng, H. Xu, Y. Wang, R. Jia, X. Ma, Y. Feng and H. Chen, *Nucleic Acids Res.*, 2023, **51**, 4064–4077.
- 105 I. M. Welleman, M. W. H. Hoorens, B. L. Feringa, H. H. Boersma and W. Szymański, *Chem. Sci.*, 2020, **11**, 11672–11691.
- 106 S. Ogasawara and M. Maeda, *Angew. Chem. Int. Ed.*, 2009, **48**, 6671–6674.
- 107 S. Lena, P. Neviani, S. Masiero, S. Pieraccini and G. P. Spada, *Angew. Chem. Int. Ed.*, 2010, **49**, 3657–3660.
- 108 S. Ogasawara, *ACS Synth. Biol.*, 2018, **7**, 2507–2513.
- 109 M. P. O’Hagan, S. Haldar, M. Duchi, T. A. A. Oliver, A. J. Mulholland, J. C. Morales and M. C. Galan, *Angew. Chem. Int. Ed.*, 2019, **58**, 4334–4338.



## References

---

- 110 M. P. O'Hagan, S. Haldar, J. C. Morales, A. J. Mulholland and M. C. Galan, *Chem. Sci.*, 2021, **12**, 1415–1426.
- 111 M. P. O'Hagan, J. Ramos-Soriano, S. Haldar, S. Sheikh, J. C. Morales, A. J. Mulholland and M. C. Galan, *Chem. Commun.*, 2020, **56**, 5186–5189.
- 112 J. Andersson, S. Li, P. Lincoln and J. Andréasson, *J. Am. Chem. Soc.*, 2008, **130**, 11836–11837.
- 113 M. Hammarson, J. Andersson, S. Li, P. Lincoln and J. Andréasson, *Chem. Commun.*, 2010, **46**, 7130–7132.
- 114 J. Li, X. Yin, B. Li, X. Li, Y. Pan, J. Li and Y. Guo, *Anal. Chem.*, 2019, **91**, 5354–5361.
- 115 D. Avagliano, P. A. Sánchez-Murcia and L. González, *Chem. Eur. J.*, 2020, **26**, 13039–13045.
- 116 L. Wimberger, F. J. Rizzuto and J. E. Beves, *J. Am. Chem. Soc.*, 2023, **145**, 2088–2092.
- 117 A. S. Lubbe, W. Szymanski and B. L. Feringa, *Chem. Soc. Rev.*, 2017, **46**, 1052–1079.
- 118 M. P. O'Hagan, P. Peñalver, R. S. L. Gibson, J. C. Morales and M. C. Galan, *Chem. Eur. J.*, 2020, **26**, 6224–6233.
- 119 J. Calbo, A. R. Thawani, R. S. L. Gibson, A. J. P. White and M. J. Fuchter, *Beilstein J. Org. Chem.*, 2019, **15**, 2753–2764.
- 120 M. S. Maier, K. Hüll, M. Reynders, B. S. Matsuura, P. Leippe, T. Ko, L. Schäffer and D. Trauner, *J. Am. Chem. Soc.*, 2019, **141**, 17295–17304.
- 121 D. Bléger and S. Hecht, *Angew. Chem. Int. Ed.*, 2015, **54**, 11338–11349.
- 122 L. N. Lameijer, S. Budzak, N. A. Simeth, M. J. Hansen, B. L. Feringa, D. Jacquemin and W. Szymanski, *Angew. Chem. Int. Ed.*, 2020, **59**, 21663–21670.
- 123 A. A. Beharry, O. Sadovski and G. A. Woolley, *J. Am. Chem. Soc.*, 2011, **133**, 19684–19687.
- 124 P. Lentès, E. Stadler, F. Röhricht, A. Brahms, J. Gröbner, F. D. Sönnichsen, G. Gescheidt and R. Herges, *J. Am. Chem. Soc.*, 2019, **141**, 13592–13600.
- 125 R. Siewertsen, H. Neumann, B. Buchheim-Stehn, R. Herges, C. Näther, F. Renth and F. Temps, *J. Am. Chem. Soc.*, 2009, **131**, 15594–15595.
- 126 E. R. Thapaliya, J. Zhao and G. C. R. Ellis-Davies, *ACS Chem. Neurosci.*, 2019, **10**, 2481–2488.
- 127 F. Wages, P. Lentès, T. Griebenow, R. Herges, C. Peifer and E. Maser, *Chem. Biol. Interact.*, 2022, **354**, 109822.

## References

---

- 128 J. Ewert, L. Heintze, M. Jordà-Redondo, J.-S. von Glasenapp, S. Nonell, G. Bucher, C. Peifer and R. Herges, *J. Am. Chem. Soc.*, 2022, **144**, 15059–15071.
- 129 H. Lee, J. Tessarolo, D. Langbehn, A. Baksi, R. Herges and G. H. Clever, *J. Am. Chem. Soc.*, 2022, **144**, 3099–3105.
- 130 S. Samanta, C. Qin, A. J. Lough and G. A. Woolley, *Angew. Chemie*, 2012, **124**, 6558–6561.
- 131 F. Eljabu, J. Dhruval and H. Yan, *Bioorg. Med. Chem. Lett.*, 2015, **25**, 5594–5596.
- 132 S. T. G. Street, D. N. Chin, G. J. Hollingworth, M. Berry, J. C. Morales and M. C. Galan, *Chem. Eur. J.*, 2017, **23**, 6953–6958.
- 133 A. J. J. Lennox and G. C. Lloyd-Jones, *Chem. Soc. Rev.*, 2014, **43**, 412–443.
- 134 M. D. Sindkhedkar, H. R. Mulla, M. A. Wurth and A. Cammers-Goodwin, *Tetrahedron*, 2001, **57**, 2991–2996.
- 135 M. W. Ha, M.-H. Park, J. Y. Hwang, J. Kim, D.-H. Kim, T.-W. Lee and Y.-H. Kim, *Dye. Pigment.*, 2021, **185**, 108880.
- 136 G. R. Dick, D. M. Knapp, E. P. Gillis and M. D. Burke, *Org. Lett.*, 2010, **12**, 2314–2317.
- 137 Y. M. A. Mohamed, E. J. Solum and A. F. Eweas, *Monatshefte für Chemie - Chem. Mon.*, 2018, **149**, 1857–1864.
- 138 D. M. Knapp, E. P. Gillis and M. D. Burke, *J. Am. Chem. Soc.*, 2009, **131**, 6961–6963.
- 139 M. R. Luzung, J. S. Patel and J. Yin, *J. Org. Chem.*, 2010, **75**, 8330–8332.
- 140 P. Kumar, T. Guntreddi, R. Singh and K. N. Singh, *Org. Chem. Front.*, 2017, **4**, 147–150.
- 141 I. J. McManus, H. Daly, H. G. Manyar, S. F. R. Taylor, J. M. Thompson and C. Hardacre, *Faraday Discuss.*, 2016, **188**, 451–466.
- 142 E. Galán, M. L. Perrin, M. Lutz, H. S. J. van der Zant, F. C. Grozema and R. Eelkema, *Org. Biomol. Chem.*, 2016, **14**, 2439–2443.
- 143 C. E. Miller, *J. Chem. Educ.*, 1965, **42**, 254.
- 144 D. B. Konrad, J. A. Frank and D. Trauner, *Chem. Eur. J.*, 2016, **22**, 4364–4368.
- 145 T. H. L. Nguyen, N. Gigant and D. Joseph, *ACS Catal.*, 2018, **8**, 1546–1579.
- 146 J. Deng, X. Wu, G. Guo, X. Zhao and Z. Yu, *Org. Biomol. Chem.*, 2020, **18**, 5602–5607.
- 147 E. Merino and M. Ribagorda, *Beilstein J. Org. Chem.*, 2012, **8**, 1071–1090.
- 148 H. Sell, C. Näther and R. Herges, *Beilstein J. Org. Chem.*, 2013, **9**, 1–7.
- 149 R. Herges and P. Lentès, in *Molecular Photoswitches*, ed. Z. L. Pianowski, Wiley, Weinheim, Germany, 2022, pp. 65–81.

## References

---

- 150 D. Piekna-Przybylska, M. A. Sullivan, G. Sharma and R. A. Bambara, *Biochemistry*, 2014, **53**, 2581–2593.
- 151 S. Amrane, A. Kerkour, A. Bedrat, B. Vialet, M.-L. Andreola and J.-L. Mergny, *J. Am. Chem. Soc.*, 2014, **136**, 5249–5252.
- 152 B. De Nicola, C. J. Lech, B. Heddi, S. Regmi, I. Frasson, R. Perrone, S. N. Richter and A. T. Phan, *Nucleic Acids Res.*, 2016, **44**, 6442–6451.
- 153 A. De Cian, L. Guittat, M. Kaiser, B. Sacca, S. Amrane, A. Dourdoncle, P. Alberti, M. Teulade-Fichou, L. Lacroix and J.-L. Mergny, *Methods*, 2007, **42**, 183–195.
- 154 M. Pérez-Soto, P. Peñalver, S. T. G. Street, D. Weenink, M. P. O'Hagan, J. Ramos-Soriano, Y. J. Jiang, G. J. Hollingworth, M. C. Galan and J. C. Morales, *Bioorg. Med. Chem.*, 2022, **71**, 116946.
- 155 P. Thordarson, *Chem. Soc. Rev.*, 2011, **40**, 1305–1323.
- 156 L. Hahn, N. J. Buurma and L. H. Gade, *Chem. – A Eur. J.*, 2016, **22**, 6314–6322.
- 157 C. Ramos, S. Almeida, L. Lourenço, P. Pereira, R. Fernandes, M. Faustino, J. Tomé, J. Carvalho, C. Cruz and M. Neves, *Molecules*, 2019, **24**, 733.
- 158 A. Głuszyńska, B. Juskowiak, M. Kuta-Siejkowska, M. Hoffmann and S. Haider, *Molecules*, 2018, **23**, 1134.
- 159 I. Jarmoskaite, I. AlSadhan, P. P. Vaidyanathan and D. Herschlag, *Elife*, 2020, **9**, e57264.
- 160 R. del Villar-Guerra, J. O. Trent and J. B. Chaires, *Angew. Chem. Int. Ed.*, 2018, **57**, 7171–7175.
- 161 A. Randazzo, G. P. Spada and M. W. da Silva, in *Quadruplex Nucleic Acids*, eds. J. B. Chaires and D. Graves, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 67–86.
- 162 T. Šmidlehner, I. Piantanida and G. Pescitelli, *Beilstein J. Org. Chem.*, 2018, **14**, 84–105.
- 163 N. C. Garbett, P. A. Ragazzon and J. B. Chaires, *Nat. Protoc.*, 2007, **2**, 3166–3172.
- 164 N. Holmgaard List, J. Knoop, J. Rubio-Magnieto, J. Idé, D. Beljonne, P. Norman, M. Surin and M. Linares, *J. Am. Chem. Soc.*, 2017, **139**, 14947–14953.
- 165 M. A. Keniry, *Biopolymers*, 2000, **56**, 123–146.
- 166 M. Webba da Silva, *Methods*, 2007, **43**, 264–277.
- 167 I. R. Kleckner and M. P. Foster, *Biochim. Biophys. Acta - Proteins Proteomics*, 2011, **1814**, 942–968.
- 168 I. Bessi, J. Wirmer-Bartoschek, J. Dash and H. Schwalbe, in *Modern Magnetic Resonance*, Springer International Publishing, Cham, 2017, pp. 1–22.

## References

---

- 169 C. Lin, J. Dickerhoff and D. Yang, in *Methods in Molecular Biology: G-Quadruplex Nucleic Acids Methods and Protocols*, Humana, New York, 2019, pp. 157–176.
- 170 C. A. Waudby and I. Alfonso, *J. Magn. Reson. Open*, 2023, **16–17**, 100102.
- 171 J. Dash, P. S. Shirude, S.-T. D. Hsu and S. Balasubramanian, *J. Am. Chem. Soc.*, 2008, **130**, 15950–15956.
- 172 L. Díaz-Casado, I. Serrano-Chacón, L. Montalvillo-Jiménez, F. Corzana, A. Bastida, A. G. Santana, C. González and J. L. Asensio, *Chem. – A Eur. J.*, 2021, **27**, 6204–6212.
- 173 D. Y. Ryazantsev, M. Y. Myshkin, V. A. Alferova, V. B. Tsvetkov, E. Y. Shustova, P. N. Kamzeeva, P. V. Kovalets, E. R. Zaitseva, N. S. Baleeva, T. S. Zatsepin, Z. O. Shenkarev, M. S. Baranov, L. I. Kozlovskaya and A. V. Aralov, *Biomolecules*, 2021, **11**, 1409.
- 174 Y. M. Vianney and K. Weisz, *Nucleic Acids Res.*, 2022, **50**, 11948–11964.
- 175 J. Šponer, B. Islam, P. Stadlbauer and S. Haider, in *Annual Reports in Medicinal Chemistry*, 2020, pp. 197–241.
- 176 J. Šponer, G. Bussi, P. Stadlbauer, P. Kührová, P. Banáš, B. Islam, S. Haider, S. Neidle and M. Otyepka, *Biochim. Biophys. Acta - Gen. Subj.*, 2017, **1861**, 1246–1263.
- 177 M. Havrila, P. Stadlbauer, B. Islam, M. Otyepka and J. Šponer, *J. Chem. Theory Comput.*, 2017, **13**, 3911–3926.
- 178 R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2016, **12**, 4114–4127.
- 179 O. Love, R. Galindo-Murillo, M. Zgarbová, J. Šponer, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2023, **19**, 4299–4307.
- 180 N. Li, Y. Gao, F. Qiu and T. Zhu, *Molecules*, 2021, **26**, 1–11.
- 181 M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *J. Chem. Theory Comput.*, 2015, **11**, 5723–5736.
- 182 I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case and M. Orozco, *Nat. Methods*, 2016, **13**, 55–58.
- 183 H. Kim and Y. Pak, *J. Phys. Chem. B*, 2022, **126**, 6199–6209.
- 184 C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder and P. Ren, *J. Chem. Theory Comput.*, 2018, **14**, 2084–2108.
- 185 A. Savelyev and A. D. MacKerell, *J. Comput. Chem.*, 2014, **35**, 1219–1239.
- 186 J. A. Lemkul and A. D. MacKerell, *J. Chem. Theory Comput.*, 2017, **13**, 2053–2071.

## References

---

- 187 J. A. Lemkul and A. D. MacKerell, *J. Chem. Theory Comput.*, 2017, **13**, 2072–2085.
- 188 B. Islam, M. Sgobba, C. Laughton, M. Orozco, J. Sponer, S. Neidle and S. Haider, *Nucleic Acids Res.*, 2013, **41**, 2723–2735.
- 189 B. Islam, P. Stadlbauer, M. Krepl, J. Koca, S. Neidle, S. Haider and J. Sponer, *Nucleic Acids Res.*, 2015, **43**, 8673–8693.
- 190 B. Islam, P. Stadlbauer, A. Gil-Ley, G. Pérez-Hernández, S. Haider, S. Neidle, G. Bussi, P. Banas, M. Otyepka and J. Sponer, *J. Chem. Theory Comput.*, 2017, **13**, 2458–2480.
- 191 B. Islam, P. Stadlbauer, M. Krepl, M. Havrila, S. Haider and J. Sponer, *J. Chem. Theory Comput.*, 2018, **14**, 5011–5026.
- 192 M. Castelli, F. Doria, M. Freccero, G. Colombo and E. Moroni, *J. Chem. Theory Comput.*, 2022, **18**, 4515–4528.
- 193 I. S. Joung and T. E. Cheatham, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.
- 194 M. Rebič, A. Laaksonen, J. Šponer, J. Uličný and F. Mocci, *J. Phys. Chem. B*, 2016, **120**, 7380–7391.
- 195 P. Auffinger, T. E. Cheatham and A. C. Vaiana, *J. Chem. Theory Comput.*, 2007, **3**, 1851–1859.
- 196 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 197 X. Zhang, J. Barrow, T. van Mourik and M. Bühl, *Molecules*, 2023, **28**, 3447.
- 198 N. B. Leontis and E. Westhof, *RNA*, 2001, **7**, 499–512.
- 199 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
- 200 M. P. O’Hagan, J. C. Morales and M. C. Galan, *European J. Org. Chem.*, 2019, **2019**, 4995–5017.
- 201 D. Jiang, H. Zhao, H. Du, Y. Deng, Z. Wu, J. Wang, Y. Zeng, H. Zhang, X. Wang, J. Wu, C.-Y. Hsieh and T. Hou, *J. Chem. Theory Comput.*, 2023, **19**, 5633–5647.
- 202 D. Renčik, I. Kejnovská, P. Školáková, K. Bednářová, J. Motlová and M. Vorlíčková, *Nucleic Acids Res.*, 2009, **37**, 6625–6634.
- 203 K. Weisz, *ChemTexts*, 2021, **7**, 25.
- 204 J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927–4940.
- 205 D. Bhattacharyya, G. Mirihana Arachchilage and S. Basu, *Front. Chem.*, 2016, **4**, 1–14.
- 206 K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2006, **128**,

## References

---

- 9963–9970.
- 207 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263–282.
- 208 Y. Ma, K. Iida and K. Nagasawa, *Biochem. Biophys. Res. Commun.*, 2020, **531**, 3–17.
- 209 T. Biver, *Molecules*, 2022, **27**, 4165.
- 210 A. Marchand and V. Gabelica, *Nucleic Acids Res.*, 2016, **44**, 10999–11012.
- 211 V. H. Le, N. Nagesh and E. A. Lewis, *PLoS One*, 2013, **8**, e72462.
- 212 K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu and A. T. Phan, *J. Am. Chem. Soc.*, 2009, **131**, 4301–4309.
- 213 M. Vorlíčková, I. Kejnovská, J. Sagi, D. Renčiuk, K. Bednářová, J. Motlová and J. Kypr, *Methods*, 2012, **57**, 64–75.
- 214 A. Marchand, A. Granzhan, K. Iida, Y. Tsushima, Y. Ma, K. Nagasawa, M.-P. Teulade-Fichou and V. Gabelica, *J. Am. Chem. Soc.*, 2015, **137**, 750–756.
- 215 J. L. Samphire, PhD Thesis, University of Bristol, 2021.
- 216 E. Belmonte-Reche and J. C. Morales, *NAR Genomics Bioinforma.*, 2020, **2**, lqz005.
- 217 Y. Sun, S. Bi, D. Song, C. Qiao, D. Mu and H. Zhang, *Sensors Actuators B Chem.*, 2008, **129**, 799–810.
- 218 M. Sirajuddin, S. Ali and A. Badshah, *J. Photochem. Photobiol. B Biol.*, 2013, **124**, 1–19.
- 219 P. Maleki, J. B. Budhathoki, W. A. Roy and H. Balci, *Mol. Genet. Genomics*, 2017, **292**, 483–498.
- 220 M. Aznauryan, S. Søndergaard, S. L. Noer, B. Schiøtt and V. Birkedal, *Nucleic Acids Res.*, 2016, **44**, 11024–11032.
- 221 J. T. Grün and H. Schwalbe, *Biopolymers*, 2022, **113**, e23477.
- 222 X. Long and M. D. Stone, *PLoS One*, 2013, **8**, e83420.
- 223 M. Aznauryan, S. L. Noer, C. W. Pedersen, J. Mergny, M. Teulade-Fichou and V. Birkedal, *ChemBioChem*, 2021, **22**, 1811–1817.
- 224 W. Moormann, T. Tellkamp, E. Stadler, F. Röhricht, C. Näther, R. Puttreddy, K. Rissanen, G. Gescheidt and R. Herges, *Angew. Chem. Int. Ed.*, 2020, **59**, 15081–15086.
- 225 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
- 226 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 227 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

## References

---

- 228 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
- 229 D. J. Evans and B. L. Holian, *J. Chem. Phys.*, 1985, **83**, 4069–4074.
- 230 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 231 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 232 A. B. Pangborn, M. A. Giardello, R. H. Grubbs, R. K. Rosen and F. J. Timmers, *Organometallics*, 1996, **15**, 1518–1520.
- 233 W. C. Still, M. Kahn and A. Mitra, *J. Org. Chem.*, 1978, **43**, 2923–2925.
- 234 M. Reynders, A. Chaikuad, B. Berger, K. Bauer, P. Koch, S. Laufer, S. Knapp and D. Trauner, *Angew. Chem. Int. Ed.*, 2021, **60**, 20178–20183.
- 235 D. I. Bugaenko and A. V. Karchava, *Adv. Synth. Catal.*, 2023, **365**, 1893–1900.
- 236 J. Ezquerra, C. Pedregal, C. Lamas, J. Barluenga, M. Pérez, M. A. García-Martín and J. M. González, *J. Org. Chem.*, 1996, **61**, 5804–5812.

# Appendix

## A.3 Supplementary data for Chapter 3

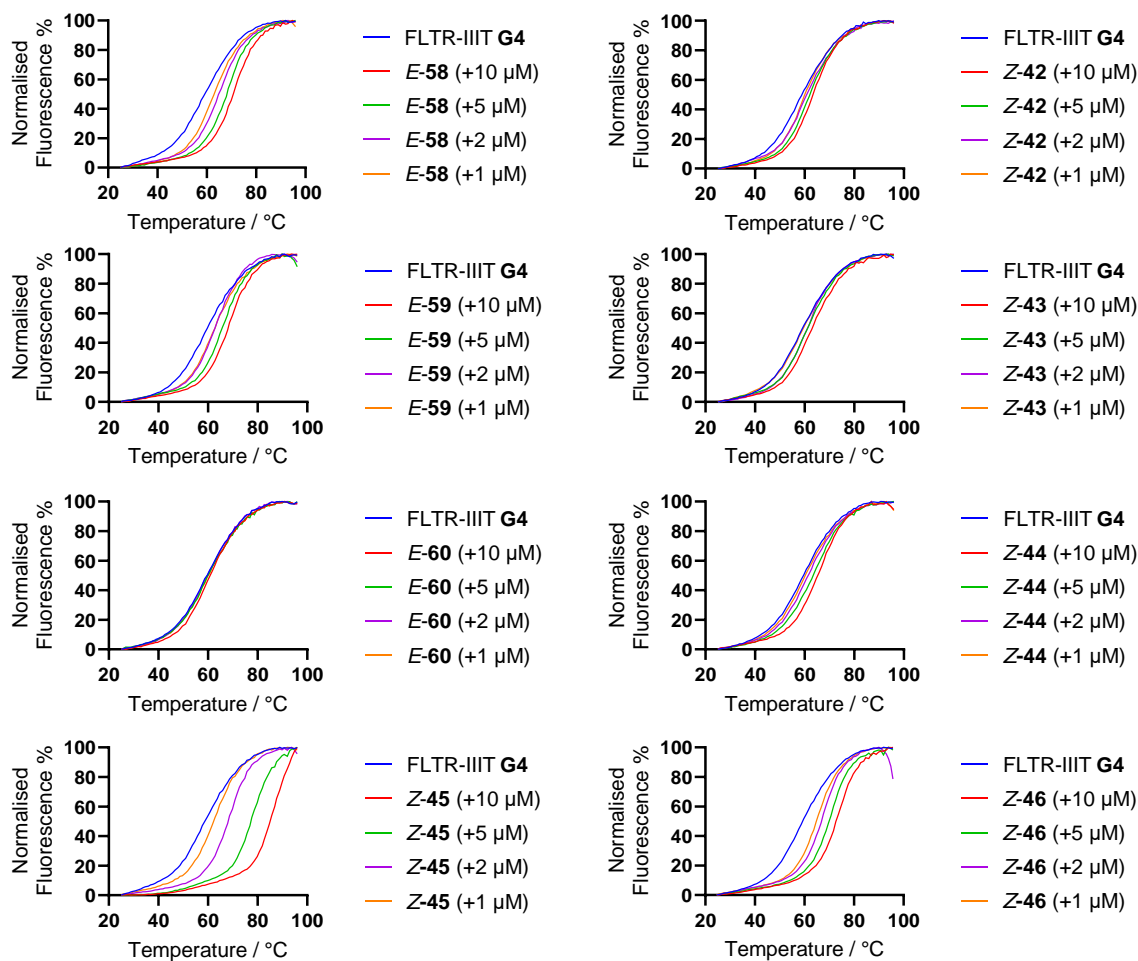


Figure A3.1: Representative FRET melting curves of FLTR-IIIIT with ligands.



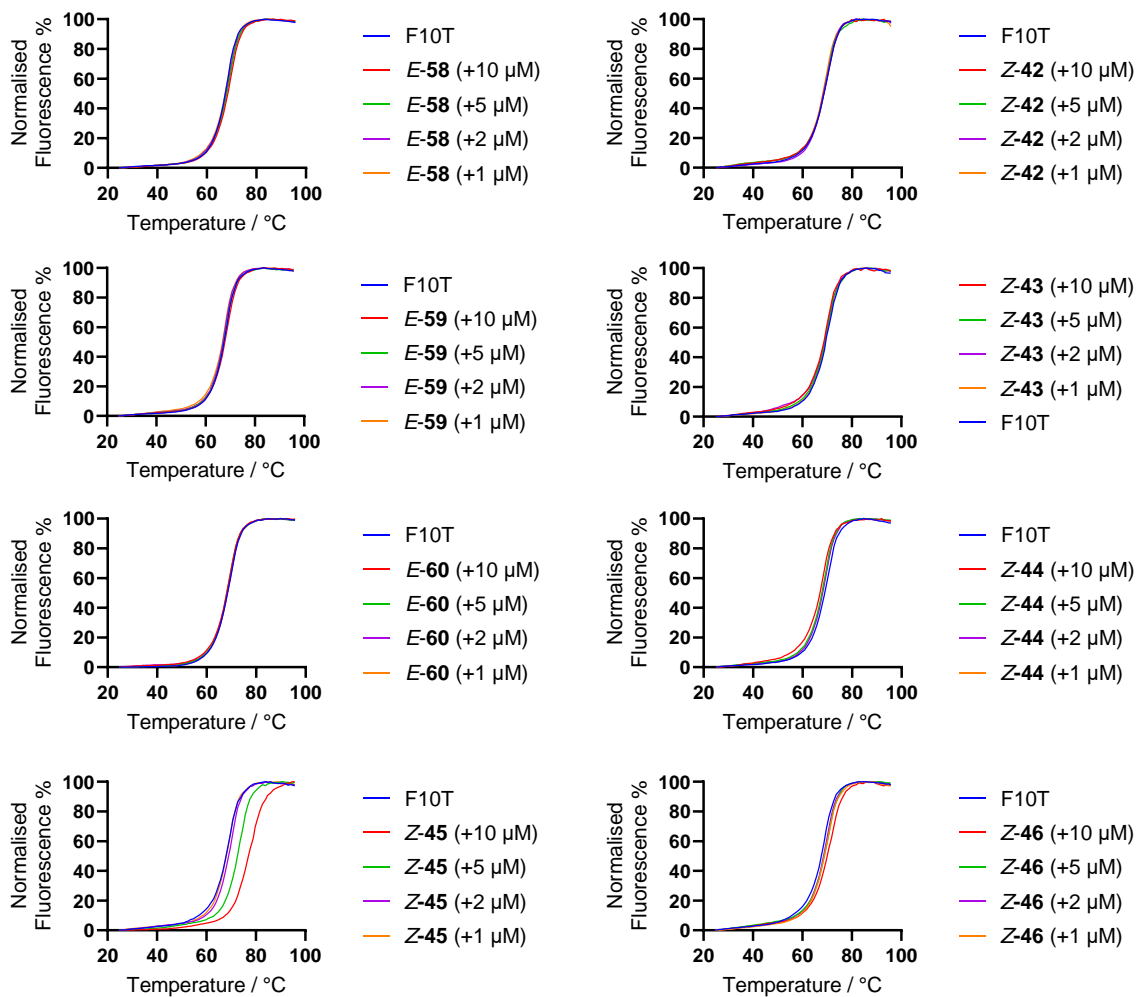
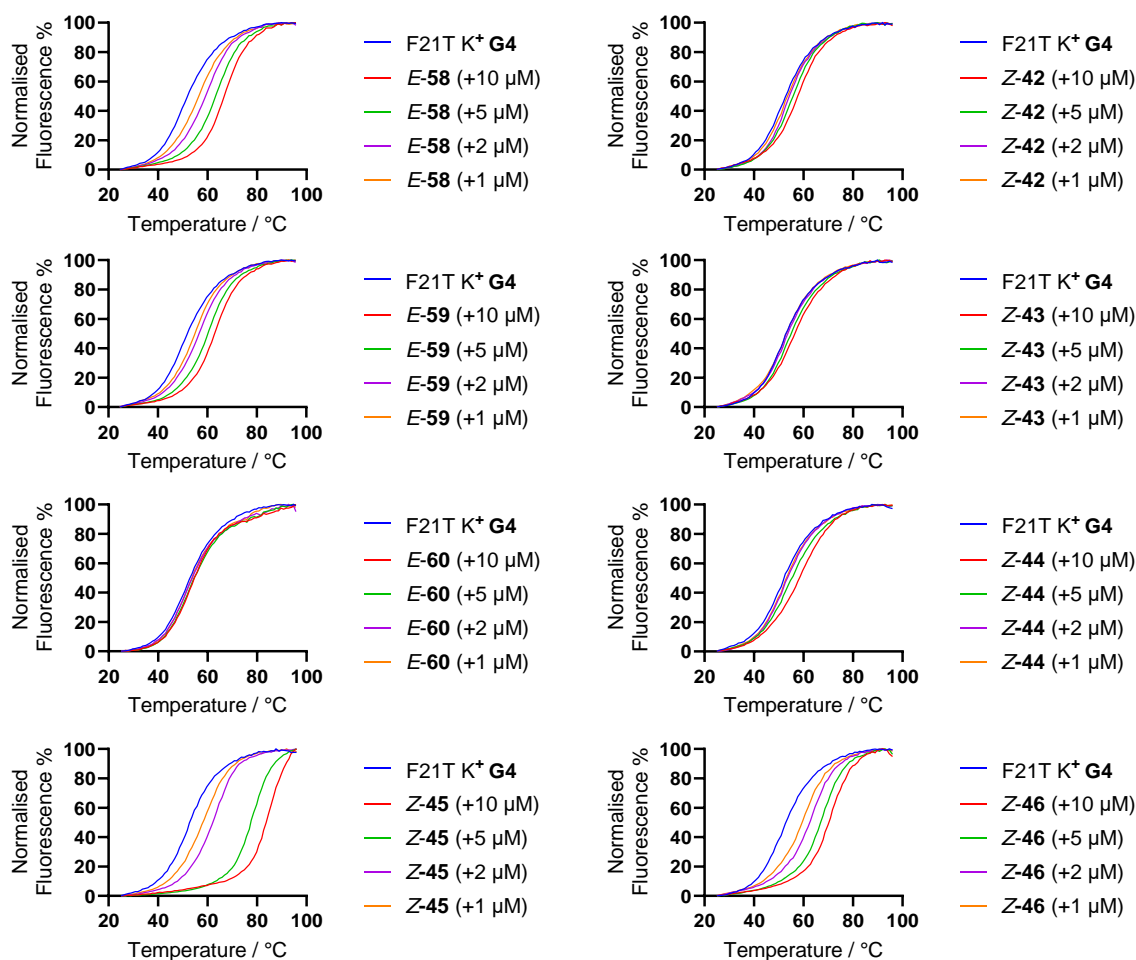


Figure A3.2: Representative FRET melting curves of F10T with ligands.

## A.4 Supplementary data for Chapter 4

Figure A4.1: Representative FRET melting curves of F21T-K<sup>+</sup> with ligands.

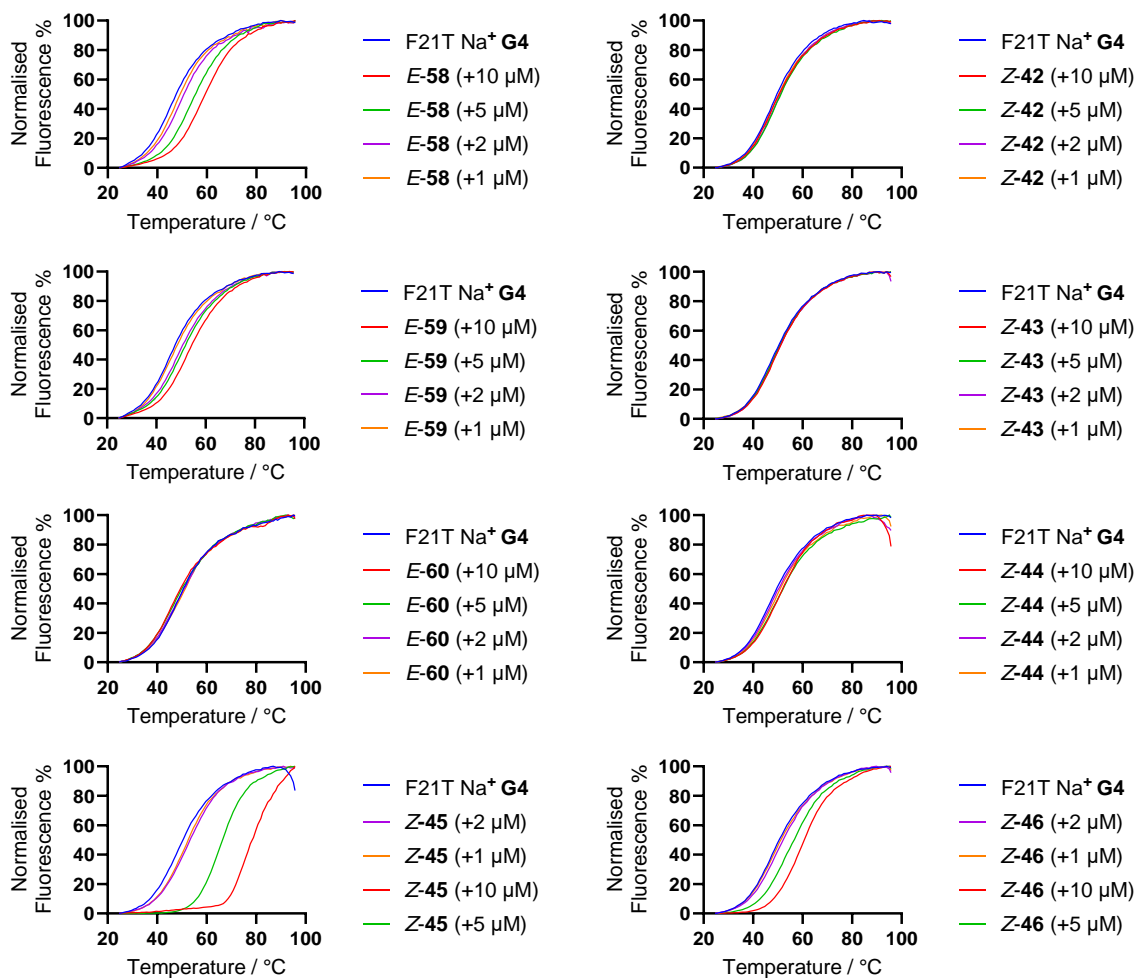


Figure A4.2: Representative FRET melting curves of F21T-Na<sup>+</sup> with ligands.

## A.5 Supplementary data for Chapter 5

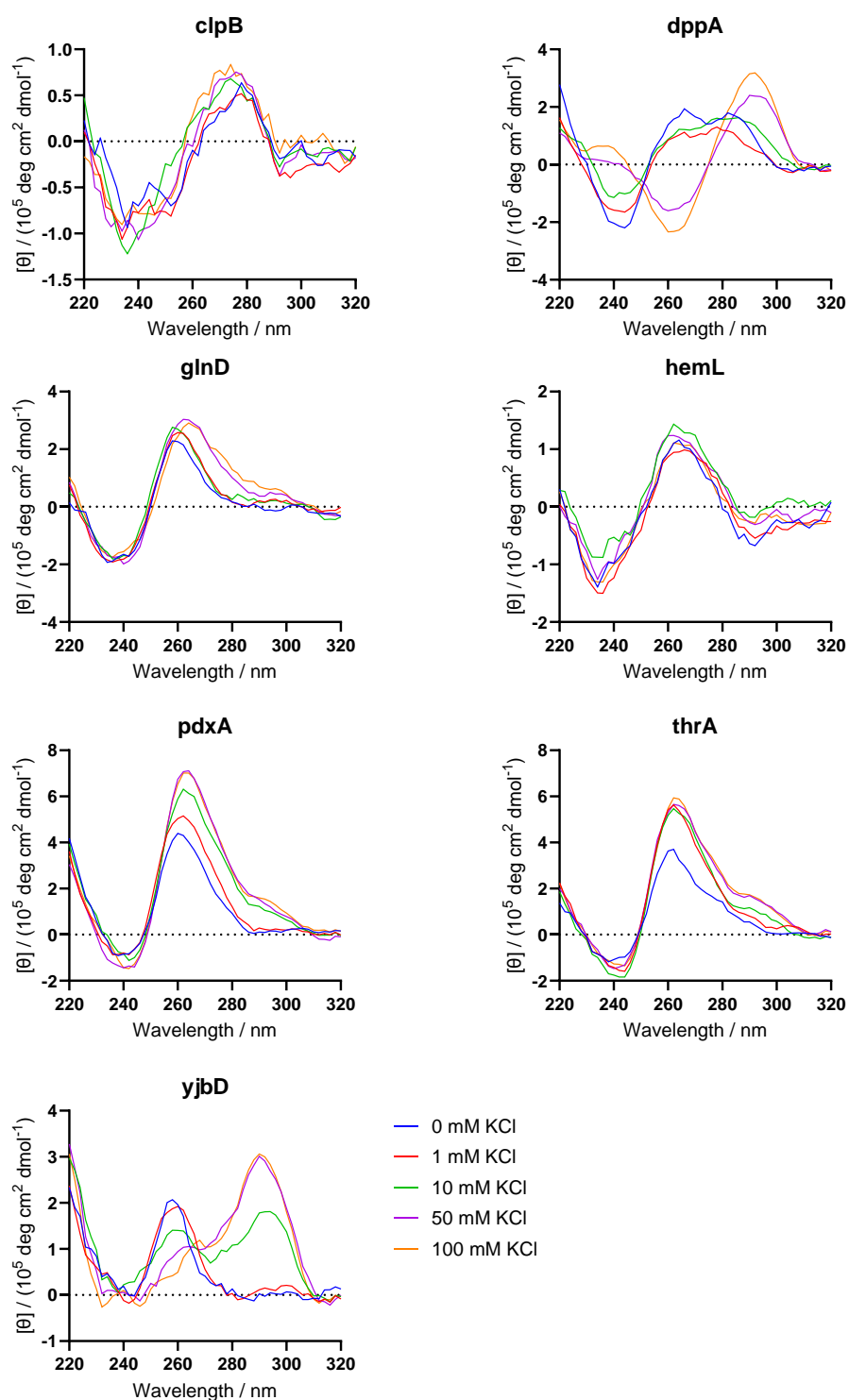


Figure A5.1: CD spectra of fluorescently labelled *E. coli* G4 sequences in different KCl concentrations.

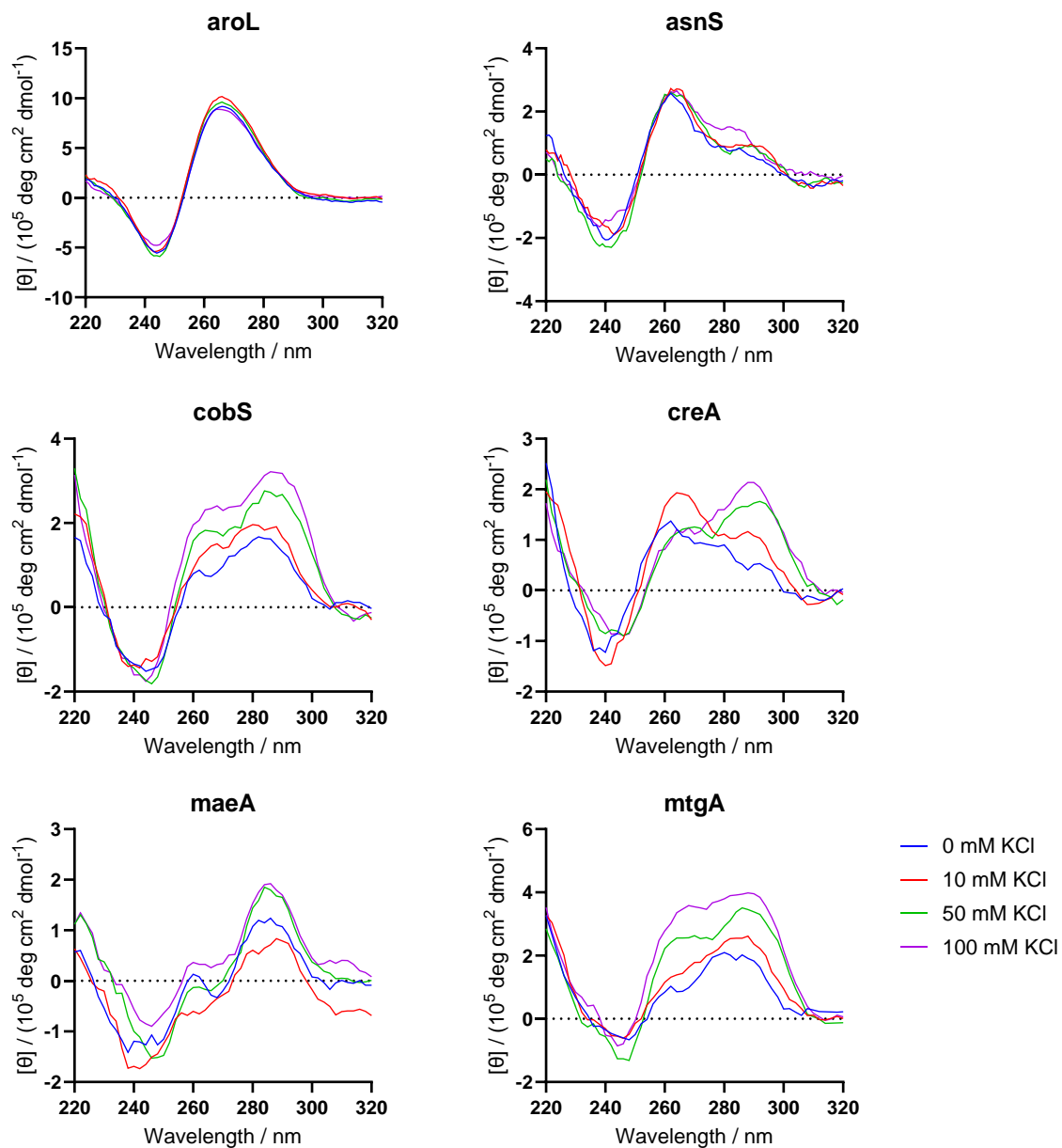


Figure A5.2: CD spectra of fluorescently labelled *E. coli* G4 sequences in different KCl concentrations.

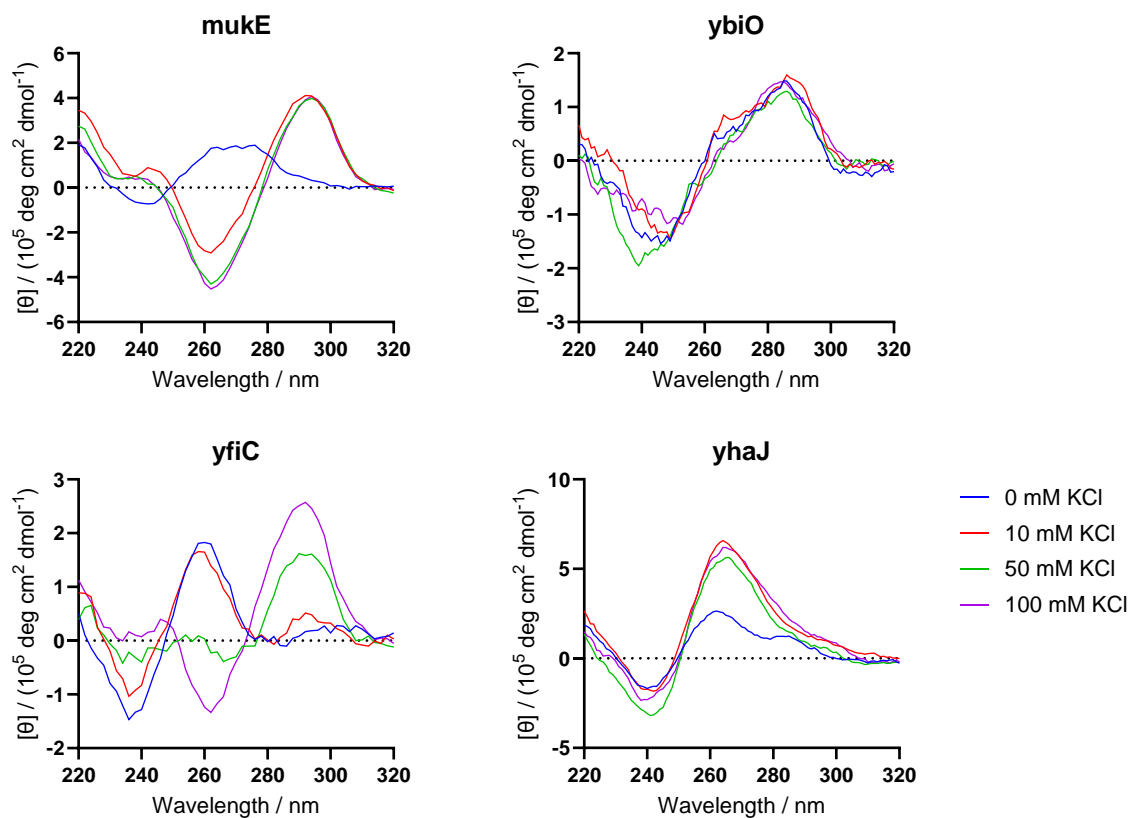


Figure A5.3: CD spectra of fluorescently labelled *E. coli* G4 sequences in different KCl concentrations.

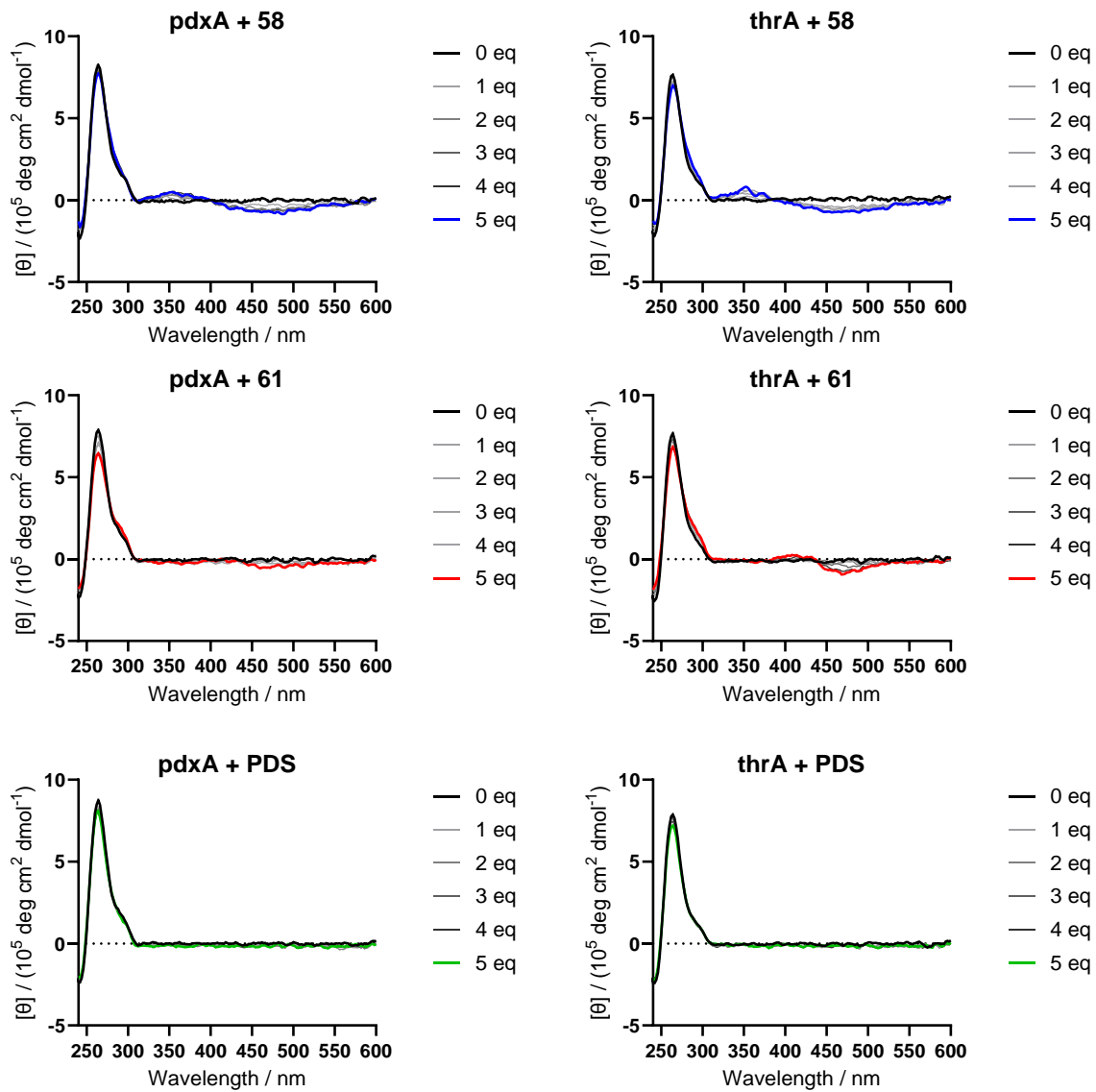


Figure A5.4: CD titration spectra of selected *E. coli* sequences and the addition of 5 equivalents of 58, 61 and PDS.

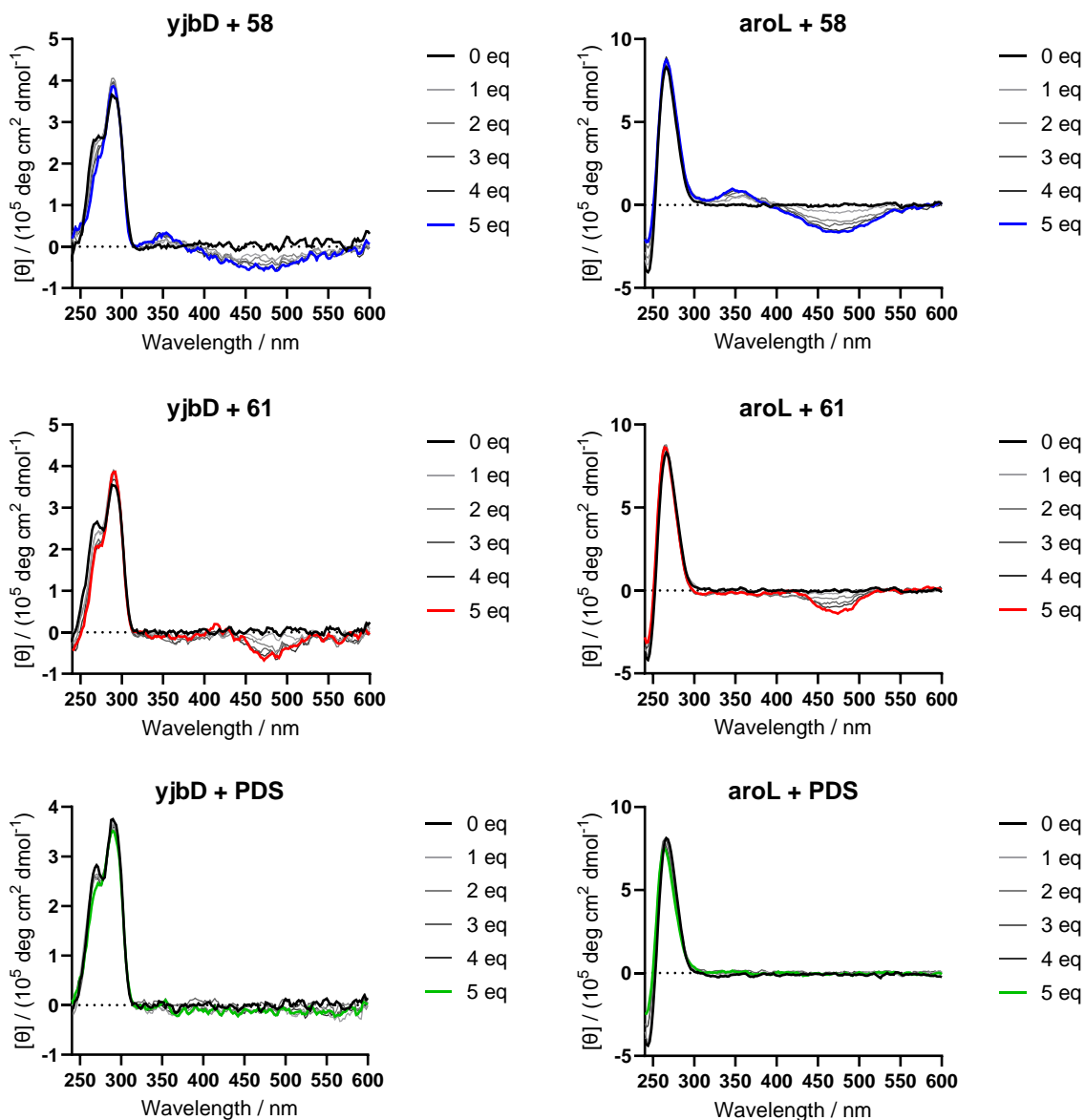


Figure A5.5: CD titration spectra of selected *E. coli* sequences and the addition of 5 equivalents of 58, 61 and PDS.



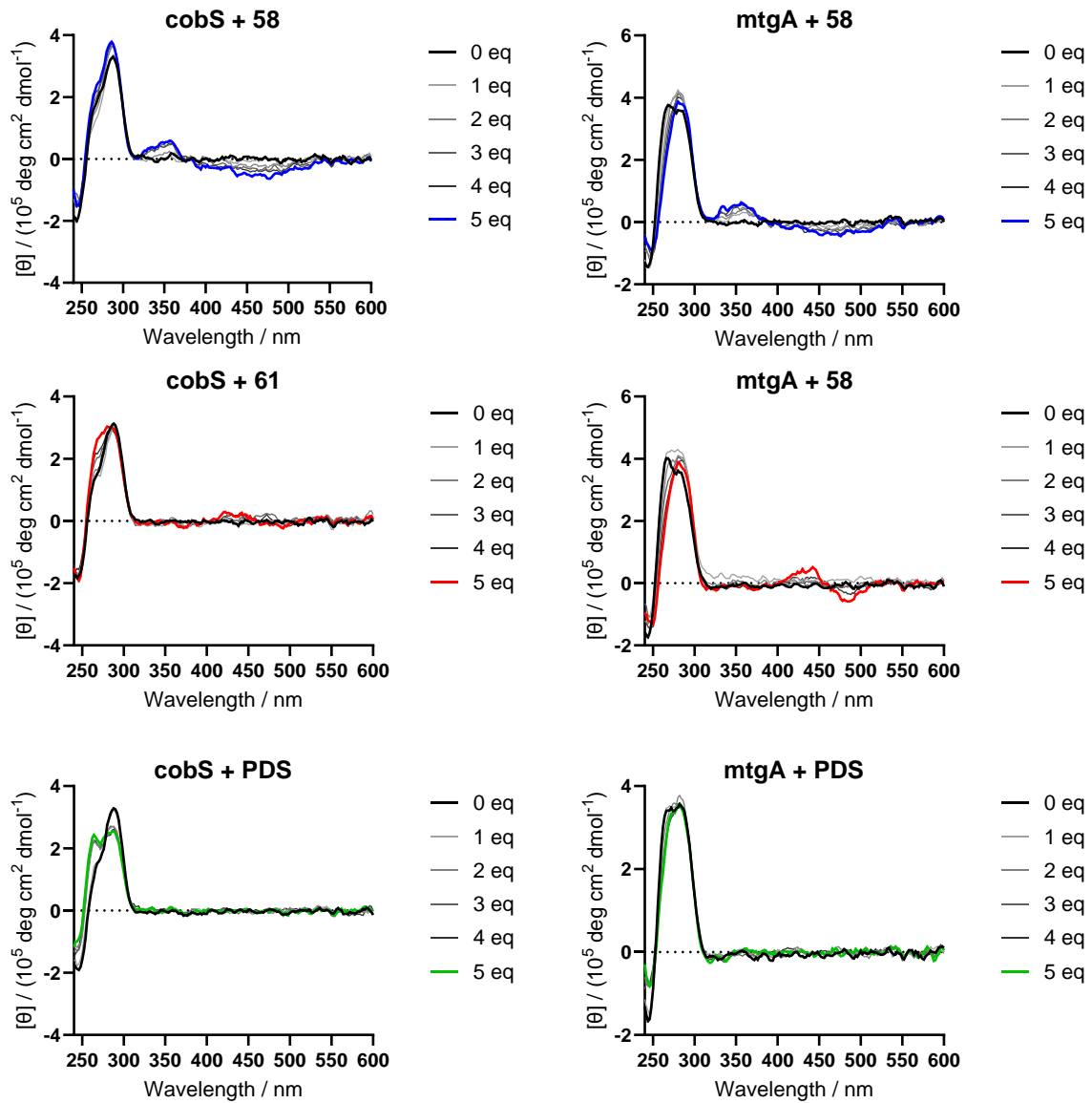


Figure A5.6: CD titration spectra of selected *E. coli* sequences and the addition of 5 equivalents of 58, 61 and PDS.