



University of  
**Reading**

# Structural Insights into Ruthenium complex- DNA Triplex Interactions

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE SCHOOL OF CHEMISTRY,  
FOOD AND PHARMACY

Ahmad Abdullrahman

Supervisor: Dr James Hall  
January 2024

## Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Ahmad Abdullrahman

## Acknowledgments

First and foremost, I wish to express my deepest gratitude to my supervisors Dr James Hall and Professor Christine Cardin for your invaluable guidance and continuous patience throughout my PhD journey. A special thanks goes to James, your role as a supervisor, line manager, and mentor has been invaluable through the highs and lows of this research experience, especially during challenging periods! I am extremely grateful for the opportunities to travel, collaborate, and explore the myriad facets of research, shaping my growth as an independent and critical thinker. This journey has been life-changing thanks to your inspiration.

My appreciation goes out to everyone I had the pleasure of collaborating with over the years. I would especially like to thank Professor Andrew Kellett for welcoming me during my secondment at Dublin City University and expanding my knowledge, and Dr Georgia Menounou for helping me get settled during my visit. I sincerely thank the patient researchers at Diamond Light Source and the Research Complex at Harwell for helping in managing the big load of experiments in short time. Thank you Dr Juan Sanchez-Weatherby, Dr James Sandy, Dr Halina Mikolajek, and Dr Christian Orr for your endless enthusiasm during tricky beamtimes. Additionally, many thanks to my fellow early career researchers in the Nature ETN project for the wonderful experiences shared. A special acknowledgment to Dr Maria Dalla Pozza for your encouragement and the memorable times we shared.

I also wish to thank emeritus Professor David Cardin for our stimulating conversations over coffee that often sparked new insights. I am also grateful to Dr Kane McQuaid for generously providing access to your precious library of compounds.

Many thanks should go to my colleagues Dr Megan Lambert and Tayler Prieto for your assistance with experiments and fruitful conversations. Thanks also to our technician Nick Spencer for keeping the X-ray diffractometer finely tuned and for your helpful suggestions.

Finally, I am grateful to my friend Adriano and my family for your unconditional support and encouragement despite the long distances. Most importantly, to my rock Mourad – thank you for your love and recent selfless patience. Your presence during my critical periods of recovery was crucial, ensuring that I remained resilient and able to move forward. Without you, this achievement wouldn't have been possible.



## Abstract

DNA triplexes, formed by the binding of a triplex-forming oligonucleotide (TFO) within the major groove of a duplex, have been shown to have potential in gene editing and DNA nanotechnology applications. Recently, metal complexes, including ruthenium polypyridyl intercalators, have been widely explored for their distinctive DNA recognition properties and ability to induce site-specific DNA cleavage. Structural information, showing how ruthenium complexes can interact with DNA triplexes, is required to aid the development of compounds capable of selectively targeting and stabilising triple helical structures. This thesis reports solution and crystal-phase characterisation of the binding of ruthenium polypyridyl complexes to DNA triplexes, including the first crystal structure of a complete triplex with intercalated Ru-dppz complexes.

UV thermal denaturation experiments were used to assess triplex stability under various conditions related to those used for crystallisation. This included pH (4.0 to 8.0), different cations ( $\text{Na}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Sr}^{2+}$ ) and spermine, all of which are known to influence triplex thermodynamic stability. The presence of  $\text{Mg}^{2+}$  increased the  $T_m$  of intermolecular triplexes by  $\sim 5^\circ\text{C}$  and intramolecular triplexes by approximately  $10^\circ\text{C}$ , compared to in the absence of magnesium ions. The observed stability profiles provided valuable guidance for the selection of systems to take forward for crystallisation and structural analysis.

The stability and binding preferences of both enantiomers of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  were then explored in solution by systematically extending the duplex component of a model triplex system. Spectroscopic analysis, including fluorescence spectroscopy and circular dichroism, revealed the  $\Lambda$ -enantiomers bind to terminal CG and TA steps of the extended duplex. While the  $\Delta$ -enantiomer exhibited fluorescence emission consistent though all the extended systems, stabilisation of the triplex (with a  $\Delta T_m$  of  $+1.2^\circ\text{C}$ ) was only observed with CG extensions, suggesting intercalation by the complex adjacent to the terminus of the TFO.

Crystallisation of a unimolecular TFO led to the first high-resolution ( $2\text{\AA}$ ) X-ray crystal structure of a complete DNA triplex with intercalated ruthenium polypyridyl complexes. Two  $\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  complexes intercalated into the minor groove of the DNA triplex, adjacent to T-A:T triplets, separated by a Watson-Crick base pair. This violates the neighbour exclusion prin-

ple due to binding in adjacent DNA steps. Two  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> complexes also intercalated into TA/TA steps within a DNA duplex cross-over region between symmetry-related triplexes.

Crystallisation screening, using the sequences studied in chapter 2, yielded additional crystal structures. A second structure, determined to near-atomic resolution (1.2 Å) revealed for the first time how Ru-dppz complexes can intercalate into the major groove of the underlying duplex, excluding the TFO from the crystal lattice. The intercalation of  $\Lambda$ -[Ru(TAP)<sub>2</sub>(11-CN-dppz)]<sup>2+</sup> in the TA/TA steps into the duplex major groove provides insight into the stacking requirements, as well as the dppz-moieties required to achieve major groove intercalation.

Finally, a crystal structure resulting from the self-assembly of a G-rich TFO was obtained. This demonstrated that the TFO could assemble into a G-quadruplex in the presence of K<sup>+</sup>. The crystal structure, determined to 1.15 Å resolution, featured G-tetrads, T-tetrads and a novel T:G octaplet motif, at the interface between two non-symmetry equivalent quadruplexes.

Overall, these findings provide insights into the intercalation of ruthenium complexes within DNA triplexes, highlighting novel structure formation while emphasizing the importance of careful TFO and DNA triplex design for future studies.

# Contents

<b>Acknowledgments</b> .....	iii
<b>Abstract</b> .....	v
<b>Abbreviations Used</b> .....	x
<b>CHAPTER 1 - Introduction</b> .....	1
1.1 Introduction .....	2
1.2 Type of Triplexes .....	5
1.2.1 Intermolecular DNA triplexes.....	5
1.2.2 Intramolecular DNA triplexes.....	7
1.2.3 G-triplexes, R-DNA and PNA.....	7
1.3 Structural analysis of triplexes .....	8
1.3.1 Similarity with B-DNA.....	10
1.4 Stability of DNA triplexes .....	13
1.4.1 Cations and anion enhancement of DNA triplex stability.....	14
1.4.2 Base modifications .....	15
1.4.3 Phosphate backbone modification.....	20
1.4.4 Sugar modification .....	22
1.4.5 DNA triplexes intercalators and groove binders .....	24
1.5 Applications based on biomolecular approaches.....	26
1.6 DNA triplex and related interactions with metal complexes.....	31
1.7 Conclusions and Project Aims .....	40
1.8 References .....	42
<b>CHAPTER 2 - Selection of DNA triplex systems for crystallisation</b> .....	55
2.1 Introduction .....	56
2.2 Material and Methods .....	60
2.2.1 Oligonucleotides and buffer solutions .....	60
2.2.2 UV thermal denaturation experiments .....	60
2.3 Results and Discussion .....	61
2.3.1 Selection of DNA triplexes.....	61
2.3.2 The effect of pH on triplex stability.....	63
2.3.3 Triplex stability supported by cations .....	66
2.3 Conclusions .....	75
2.4 References .....	78
<b>CHAPTER 3 - Intercalation of ruthenium complex in DNA triplexes with extended underlying duplex</b> .....	83

3.1 Introduction .....	84
3.2 Materials and Methods.....	86
3.2.1 Oligonucleotides and buffer solutions .....	86
3.2.2 UV absorption measurements .....	86
3.2.3 Circular dichroism measurements .....	86
3.2.4 Fluorescence measurements .....	87
3.3 Results and Discussion .....	88
3.3.1 Thermal stability of DNA triplexes by $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ .....	89
3.3.3 Determination of $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ binding in DNA triplexes .....	95
3.3.4 Triplex formation upon $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ binding using CD Spectroscopy .....	98
3.4 Conclusions.....	103
3.5 References .....	105
<b>CHAPTER 4 - Intercalation of <math>[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}</math> in DNA triplex structure .....</b>	<b>108</b>
4.1 Introduction .....	109
4.2 Materials and Methods.....	111
4.2.1 Oligonucleotide, buffer solution, and crystallisation reagents .....	111
4.2.2 Crystallisation.....	111
4.2.3 Data collection, refinement, and analysis .....	111
4.3 Results and Discussion .....	113
4.3.1 The triplex region accommodates $\Delta$ -Ru enantiomers .....	118
4.3.2 Intercalation of $\Lambda$ -Ru in Watson-Crick base pairs.....	122
4.4 Conclusion .....	126
4.5 References .....	128
<b>CHAPTER 5 - Major groove intercalation of ruthenium polypyridyl complexes.....</b>	<b>132</b>
5.1 Introduction .....	133
5.2 Materials and Methods.....	135
5.2.1 Oligonucleotide, buffer solution, and crystallisation reagents .....	135
5.2.2 Crystallisation.....	135
5.2.3 Data collection, refinement, and analysis .....	135
5.3 Results and Discussion .....	138
5.3.3 Minor groove intercalation .....	140
5.3.2 Major groove intercalation .....	144
5.4 Conclusion .....	151
5.5 References .....	153
<b>CHAPTER 6 - Crystallisation of a DNA quadruplex from a TFO .....</b>	<b>157</b>

6.1 Introduction .....	158
6.2 Materials and Methods.....	160
6.2.1 Oligonucleotide, buffer solution, and crystallisation reagents.....	160
6.2.2 Crystallisation.....	160
6.2.3 Data collection, refinement, and analysis .....	161
6.3 Results and Discussion .....	164
6.3.1 The self-assembly of triplex-forming oligonucleotide .....	165
6.3.2 Stacking of a novel octaplet .....	168
6.3.2 Metal ions coordination.....	174
6.4 Conclusion.....	175
6.5 References .....	176
<b>CHAPTER 7 - Summary and Future Work .....</b>	<b>180</b>
7.1 Summary .....	180
7.1 References .....	184
<b>Appendix .....</b>	<b>185</b>
<b>Chapter 2 .....</b>	<b>185</b>
<b>Chapter 3 .....</b>	<b>190</b>
<b>Chapter 4 .....</b>	<b>197</b>
<b>Chapter 5 .....</b>	<b>202</b>
<b>Chapter 6 .....</b>	<b>206</b>
Publications from this work.....	208

## Abbreviations Used

A - Adenine

bp - base pairs

bpy - 2,2'-bipyridine

C - Cytosine

CD - Circular dichroism

chrysi - Chrysene-5,6-quinone diammine

D3 - 1-(2-deoxy-beta-D-ribofuranosyl)-4-(3-benzamido)phenylimidazole

DNA - Deoxyribonucleic acid

dppz - dipyridophenazine

dsDNA – Double-strand DNA

G - Guanine

MAD – Multi-wavelength anomalous

MLCT - Metal-to-Ligand Charge Transfer

MPD - 2-Methyl-2,4-pentanediol

NMR – Nuclear magnetic resonance

PEG - Polyethylene glycol

phen - 1,10-phenanthroline

RNA – Ribonucleic acid

T - Thymine

TAP - 1,4,5,8-tetraazaphenanthrene

TFO - Triplex-forming oligonucleotide

$T_m$  - Melting temperature

U - Uracil

UV - Ultraviolet

WC – Watson-Crick

# CHAPTER 1- Introduction

## Contribution Statement

This chapter is a collaborative review featured in the following publication:

Dalla Pozza, M., Abdullrahman, A., Cardin, C. J., Gasser, G. & Hall, J. P. Three's a crowd - stabilisation, structure, and applications of DNA triplexes. *Chem Sci* 13, 10193–10215 (2022).

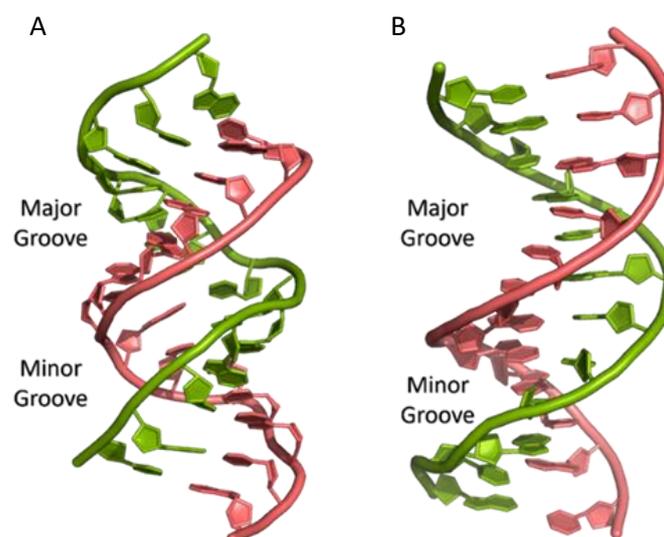
Ahmad Abdullrahman contributed to section 1 to 4, while Dr Maria Dalla Pozza contributed to section 5 and 6. All authors were involved in the editing, polishing, and publication processes.

## 1.1 Introduction

DNA is the carrier of genetic information in all cellular systems and in many viruses. As the carrier of genetic material, it directs its own replication during the cell division process and the transcription of complementary molecules of RNA. One of the defining features of DNA is its structural flexibility. DNA can adopt a wide range of higher-order structures including duplex, G-quadruplex, i-motif and Holliday Junction, all of which either have a confirmed or suspected role in gene regulation and/or transcription processes<sup>1</sup> and have been investigated in the context of ligand targeting.<sup>2</sup>

DNA triplexes are of particular interest due to its potential for exploitation in the targeting of therapeutics to specific DNA sequences. The triplex is formed when a DNA duplex is joined by a third strand, which binds in the major groove of the duplex to form a three-stranded assembly. Research efforts have increasingly focussed on TFO modification, to aid delivery *in-vivo*, reduce or prevent degradation by nucleases and increase triplex stability. However, to fully understand the structure of DNA triplexes and how modifications and bound ligands can affect their stability, it is important to first examine the structure of the DNA duplex.

The most common and best-known form of DNA is the B-DNA form, characterized by two polynucleotide strands with a right-handed helical twist about a long axis to form a double-helix, bound together by hydrogen bonds and further stabilised by  $\pi$ -stacking between adjacent bases. This winding generates two grooves: the major one is wide and deep, while the minor groove is narrow (Figure 1.1).



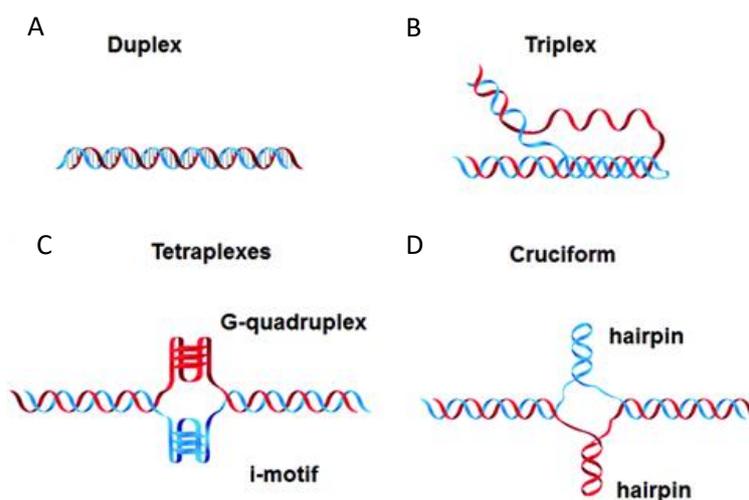
**Figure 1.1** Structure of A-DNA (A) and B-DNA (B)

This structure has been widely characterised by X-ray diffraction and occurs at high humidity and with a variety of DNA counterions including  $\text{Na}^+$ , which serves to balance the negative charge of the phosphate backbone.<sup>3,4</sup> The most significant characteristic of B-DNA is the possibility to accommodate only two types of naturally occurring base pairs (i.e., adenine-thymine A-T and cytosine-guanine C-G). In B-DNA both base pairs can be replaced by each other without altering the position of the sugar-phosphate backbone, although runs of A-T base pairs are known to have a narrower minor groove. Similarly, the double-helix is not disturbed by swapping the partners (i.e., changing a C:G with a G:C or a T:A with a A:T). However, different combinations of bases lead to the formation of non-Watson-Crick base pairs with a significant distortion of the double-helix. Since the variation of pairing causes distortions, DNA is a molecule able to adopt different non-canonical structures whilst exposed to physiological and non-physiological conditions. When the relative humidity is reduced to 75%, the B-DNA changes conformation, adopting the so-called A-DNA form, which presents a wider and flatter right-handed helix compared to the B-DNA form (Figure 1.1). In contrast to the right-handed form, Z-DNA is a left-handed analogue which has a deep minor groove and a shallow but wide major groove.<sup>5,6</sup> Z-DNA is formed as a function of DNA sequence and contains long sections of alternating purine-pyrimidine bases, most commonly as GC repeat units.

In addition to these, DNA can also form other non-canonical structures as a function of sequence, which are especially stable in the crowded intracellular environment.<sup>7</sup> These arrangements were demonstrated to play a role in different biological processes such as replication, transcription, translation and reverse translation.<sup>8</sup> Three strands of DNA can form a triplex structure, which was initially predicted to exist in 1953 by Pauling and co-workers and subsequently observed by Rich and co-workers after mixing poly U and poly A ribonucleotides in a 2:1 ratio.<sup>9,10</sup> Triplex formation has been identified both *in vitro* and *in vivo*,<sup>11</sup> as will be discussed in section 1.5. Tetraplex structures, known as G-quadruplexes, have also been observed in G-rich strands. They are formed in sequences containing multiple guanine tracts within a G-rich sequence and are bound together by Hoogsteen hydrogen bonding.<sup>12</sup> G-quadruplexes have interestingly been observed in many different locations, correlated with genomic regions that play a functional role such as replication origin sites, telomeres and promoter regions.<sup>13</sup> Another type of tetraplex structure is the intercalated motif (i-motif), formed between C-rich strands in acidic conditions. C-rich sequences are found in telomeres, and in promoter regions of many

human genes, indicating a probable role in biological processes.<sup>14</sup> Finally, the cruciform structure is formed by intra-strand base pairing of inverted repeat sequences. It can be either a four-way junction or a three-way junction depending on the number of hairpins present (Figure 1.2).<sup>15</sup>

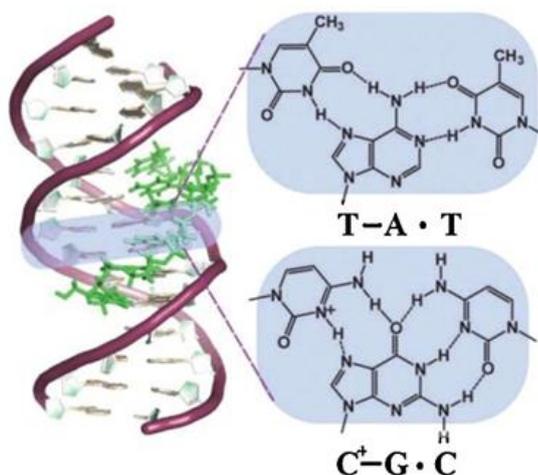
**Figure 1.2** Canonical DNA structure and non-canonical structures including (A) duplex, (B) triplex, (C) G-quadruplex and i-motif and (D) hairpin.<sup>183</sup> Reprinted from H. Tateishi-Karimata and N. Sugimoto, *Chem. Commun.*, 2020, **56**, 2379.



DNA triplexes have been investigated for decades as a very promising tool in gene editing, but development has been challenging due to the low thermal stability of the structure, the poor cellular uptake of triplex-forming oligonucleotides and degradation by nucleobases. The application and interaction of ruthenium polypyridyl complexes with DNA triplexes will also be discussed, to explore potential future therapeutic applications in areas such as photodynamic therapy (PDT). Ruthenium polypyridyl complexes possess useful properties which are particularly suitable for biological applications, as presented in section 1.6. Indeed, Ru-based compounds have been intensively studied in the last decades as antiparasitic, antimicrobial or anticancer drug candidates.<sup>16,17</sup> In particular, ruthenium polypyridyl compounds have attracted much interest.<sup>18</sup> Their ability to absorb light *via* a metal-to-ligand charge transfer (MLCT) process among other charge transfers have made them very interesting tools for photodynamic therapy (PDT).<sup>19,20</sup> Therefore, it is suggested that the intrinsic triplexes' sequence-specific binding properties combined with the phototoxicity of ruthenium derivatives can be exploited together to obtain breakthrough tools in gene editing technology.

## 1.2 Type of Triplexes

Triplex structures can be formed by DNA, RNA or hybrids of the two. DNA triplexes can be grouped based on the origin of the third strand. Intermolecular triplexes are formed between a double-stranded DNA (dsDNA) and an independent molecule termed triplex-forming oligonucleotide (TFO). If the third strand is part of a single strand which also contains the dsDNA, the triplex is referred to as an intramolecular triplex. The hydrogen bonds between the two helices of DNA are typically Watson-Crick bonds, whereas the bonds between the duplex and TFO are either Hoogsteen or reverse-Hoogsteen bonds (Figure 1.3). The directionality of the TFO can be either parallel or anti-parallel relative to the polypurine strand of the duplex.

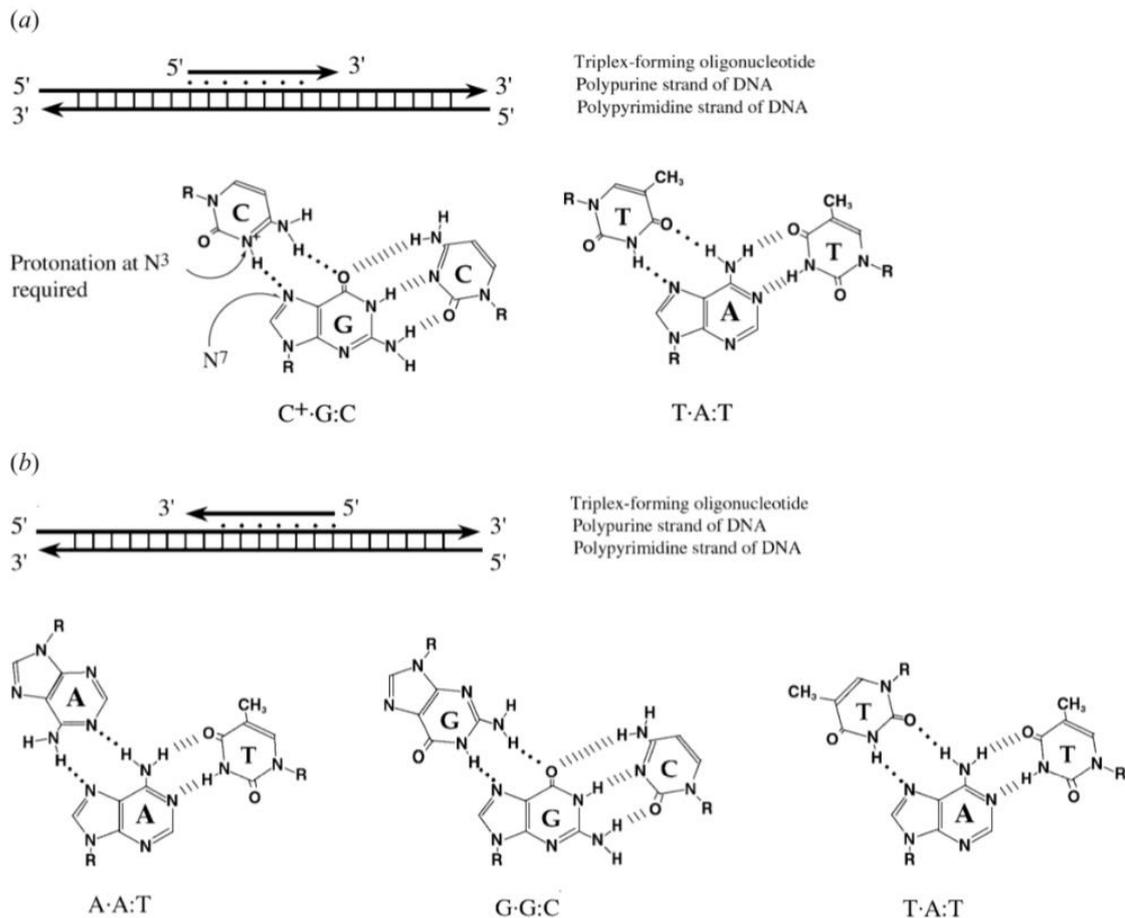


**Figure 1.3** Schematic drawing of a triplex forming oligonucleotide that specifically recognises a DNA sequence, with the TFO binding in the major groove of the DNA duplex.<sup>184</sup> Reprinted from *Coordination Chemistry Reviews*, Vol 257, Tarita Biver, *Stabilisation of non-canonical structures of nucleic acids by metal ions and small molecules*, Pages 2765-2783, Copyright (2013), with permission from Elsevier."

### 1.2.1 Intermolecular DNA triplexes

To explain the possible combination of intermolecular DNA triplexes, a close examination of the sequence of triplex-forming species is required (Figure 1.4). In a polypyrimidine TFO that consists entirely of pyrimidines, the thymine will bind to the adenosine (T-A:T) or cytosine binds to guanine (C-G:C), forming a triplex. Cytosine of the TFO, however, requires a protonation of its N3 atom to form a second Hoogsteen bond with the guanine. Therefore, these parallel triplexes require a mildly acidic environment.<sup>21</sup> However, there is a limit to protonation that, if not respected, will result in electrostatic repulsion between the adjacent protonated cytosines.<sup>22</sup> When a TFO contains only purine bases, adenine binds to adenine (A-A:T) or guanine binds to

guanine (G-G:C) with reverse-Hoogsteen hydrogen bonds. In contrast, a polypurine TFO forms a triple-helix by binding the duplex with an anti-parallel conformation.<sup>23</sup> Additionally, in the anti-parallel conformation, it is also possible to have T-A:T steps within DNA triplexes.<sup>24</sup>



**Figure 1.4** Intermolecular triplexes and canonical base triplets. (a) polypyrimidine triplexes Y-R:Y (b) polypurine triplexes R-R:Y<sup>96</sup> Reprinted from *Quarterly Reviews of Biophysics*, vol 35, K.M. Vasquez and P.M. Glazer *Triplex-forming oligonucleotides: principles and applications*, pages 89-107, copyright 2002, with permission from Cambridge University Press.

The base identity plays a key role in determining the local and overall twist angle of the DNA triplex. This residual twist is calculated based on the measurement of the angle between the two C1'-C1' carbon atoms of the adjacent Hoogsteen base pairs and the base of interest.<sup>25</sup> G-G:C triplets have the effect of increasing twist within the triplex, with an average increase of 10.6° per step, whilst T-A:T steps reduce the twist by the same value, with the overall twist angle of the helix being maintained at 30°. This is lower than the average twist for a B-DNA duplex of ca. 34° and therefore suggests that the binding of a TFO induces a slight unwinding of the duplex. This results in significant distortion after each A-T bond of the duplex within the polypurine triplexes.

By contrast, the polypyrimidine triplex has much less backbone distortion and a higher number of hydrogen bonds between the TFO and the duplex, compared with polypurine. This reduction in distortion is one possible reason why parallel triplexes are generally more stable than antiparallel helices.<sup>23</sup>

### 1.2.2 Intramolecular DNA triplexes

In addition to the intermolecular DNA triplexes, where the TFO is an external oligo, the triplex can be formed by one DNA strand which folds back on itself, to form an intramolecular assembly. These are commonly referred to as H-DNA, as their stability depends on the presence of acidic pH and negative superhelical stress. H-DNAs may be formed under supercoiled conditions with a mirror repeat polypurine-polypyrimidine sequence and the base motifs are the same as in the intermolecular triplexes with a pyrimidine third strand. Moreover, an intramolecular triplex composed with anti-parallel motifs in the DNA stretches, and a non-mirror repeat, is defined as \*H-DNA.<sup>26</sup>

### 1.2.3 G-triplexes, R-DNA and PNA

It is also possible to form a triplex from guanines – the G-triplex, which contains a strand rich in guanine bases, and can be formed as an intermediate during the formation of a DNA G-quadruplex.<sup>27,28</sup> Using fluorescence resonance energy transfer (FRET), it was determined that G-triplexes can assume both parallel and anti-parallel topologies. A parallel DNA triplex may also be formed during homologous recombination and assists the recruitment of the homologous sequences. During the formation of the recombinant DNA (R-DNA), a complex with Rec-A may be formed, leading to a triplex with an extended rise distance of 5.1 Å, compared to a standard rise distance of 3.4 Å.<sup>29</sup> Peptide nucleic acids, PNA, are modified oligonucleotides that contain a polyamide chain, instead of the sugar-phosphate backbone.<sup>30</sup> Whilst the bases retain the canonical Watson-Crick pairing scheme, the PNA backbone lacks the negative charge associated with a phosphate backbone and therefore PNA can form a highly stable triplex with one or more DNA strands with reduced electrostatic repulsion. The binding directionality respect of the ds-DNA molecule can be both parallel, or anti-parallel forming a stable D-loop, i.e., forming a triple-strand with one of the DNA strands.<sup>31</sup>

### 1.3 Structural analysis of triplexes

At the time of writing, structural characterizations of triplexes are limited. Only 32 structures, with the majority solved using NMR, have been published in the Protein Data Bank<sup>32</sup> (Table 1.1).

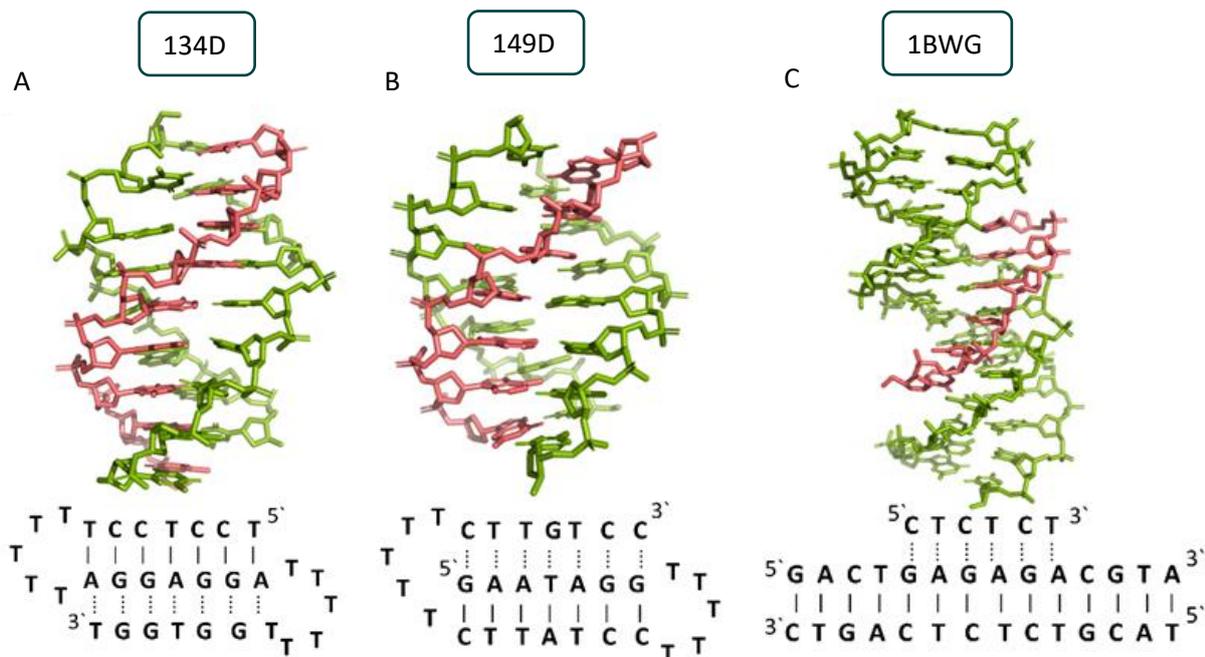
**Table 1** DNA-containing structures of triplexes deposited in the Protein Data Bank.<sup>32</sup>

Intramolecular or intermolecular	Triplex type	Nucleic acid type	Method	PDB ID	Year	Ref.
<b>DNA-only</b>						
Intramolecular	Antiparallel	DNA	NMR	134D	1993	33
Intramolecular	Antiparallel	DNA	NMR	135D	1993	33
Intramolecular	Antiparallel	DNA	NMR	136D	1993	33
Intramolecular	Antiparallel	DNA	NMR	177D	1994	34
Intermolecular	Parallel	DNA	NMR	149D	1994	33
Intermolecular	Parallel	DNA	X-ray Dif-fraction	208D	1995	35
Intramolecular	Parallel	DNA	NMR	1AT4	1997	36
Intramolecular	Parallel	DNA	NMR	1D3X	1998	37
Intramolecular	Parallel	DNA	NMR	1BCB	1998	38
Intramolecular	Parallel	DNA	NMR	1BCE	1998	38
Intermolecular	Parallel	DNA	X-ray Dif-fraction	1D3R	1999	39
Intermolecular	Parallel	DNA	NMR	1BWG	1999	40
Intramolecular	H-DNA H-Y5 isomer	DNA	NMR	1B4Y	1999	41
Intermolecular	G-triplex	DNA	NMR	2MKM	2014	42
Intermolecular	G-triplex	DNA	NMR	2MKO	2014	42
<b>Modified DNAs</b>						
Intermolecular	P-form	DNA+PNA	X-ray Dif-fraction	1PNN	1995	43
Intramolecular	Parallel	D3	NMR	1WAN	1996	44
Intramolecular	Parallel	DNA+N7-glycosylated guanine	NMR	1GN7	1997	45
Intramolecular	Parallel	DNA+1-propynyl deoxyuridine in third strand		1P3X	1998	46
Intramolecular	Parallel	DNA+LNA	NMR	1W86	2004	47
Intermolecular	P-form	PNA	X-ray Dif-fraction	1XJ9	2005	48
Intramolecular	Antiparallel	DNA+TINA intercalator	NMR	6QHI	2019	N/A

Structure determinations of DNA-containing triplexes are summarized in Table 1.1. This includes triplexes composed of hybrids of DNA-RNA, DNA-PNA and RNA-RNA triplexes, some of which

contain modified bases, sugars or intercalators, and excluding any structures which contain proteins. Only four of the structures published, containing DNA, have been determined using X-ray diffraction. DNA-containing triplex structures obtained by X-ray analysis were either formed with protein nucleic acid (PNA), intercalators, or as a result of DNA overlap with only a small number of bases forming Hoogsteen bonds and therefore do not represent a full and complete true DNA triplex, unlike several of the structures solved using NMR.

To better illustrate the structural influence of the binding of a TFO to a DNA duplex to yield a triplex, three DNA triplexes named after their PDB code, triplex **134D**, **149D** and **1BWG**, were selected. The three structures were chosen as examples of triple-helix structures which did not contain intercalators, other small molecules or chemical modifications. As the structures were solved using NMR, they are representative of triplex species in solution (Figure 1.5).

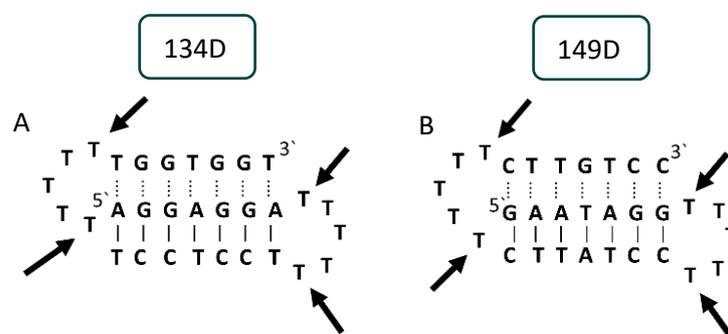


**Figure 1.5** Representation of and schematic diagram of (A) triplex **134D** (intramolecular antiparallel), (B) triplex **149D** (intramolecular parallel) (C) triplex **1BWG** (intermolecular parallel). The TFO is displayed in red and the DNA duplex is in green. In the schematic diagrams, Watson-Crick hydrogen bonding is displayed using lines with Hoogsteen bonds illustrated in dashed lines.

### 1.3.1 Similarity with B-DNA

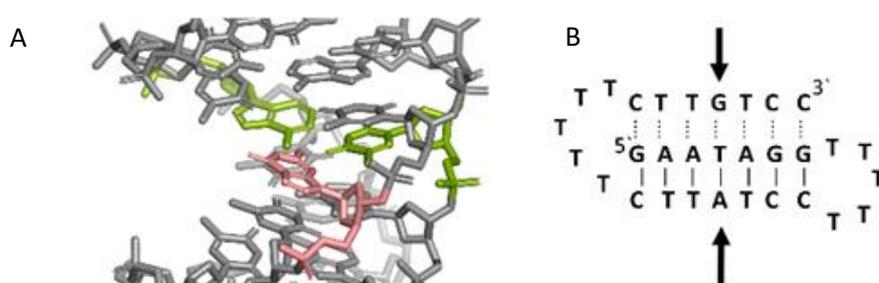
B-DNA is a right-handed form of the double-helix, with 10.5 base pairs in each turn<sup>4</sup> and a helix diameter of 20 Å.<sup>49</sup> Fibre diffraction data, obtained from X-ray studies, show that the average value of the helical twist per base pair is 36.1°, but that this can vary from 24° to 51°. The distance between bases (rise) is 3.4 Å per base pair. Whilst B-DNA is the most frequently encountered DNA conformation in physiological conditions,<sup>50</sup> others are possible, including A- and Z-forms, and are promoted by both sequence and changes in the DNA microenvironment. The B-DNA structure forms two grooves, a minor and major with a width of ~5.7 Å and ~11.7 Å, respectively. The value is obtained by subtracting 5.8 Å from the distance between the phosphate groups on opposing strands, which is the van der Waals radius of one phosphate group.<sup>51</sup>

The DNA triplex possesses significant similarity in structure to the B-form duplex. The base rise distance remains consistent at 3.3 Å and the twist value of triplex **134D** is also similar to a standard B-form duplex. at ca. 34°. Triplex A contains a poly-purine TFO, as illustrated in Figure 1.5. Triplex **134D** is an intramolecular triplex constructed from a single oligonucleotide. However, the loop positions could not be assigned due to disorder and are therefore not included in the structural coordinates. Whilst loop bases may not form hydrogen bonds with each other or with the TFO, and therefore disorder within this region is expected, the T bases indicated by arrows in Figure 1.6 adopt T-T wobble pairs, indicating two hydrogen bonds are present between the first T bases in each loop. The average base pair twist at this pair is 32° which is slightly reduced compared to the average helical twist value for B-DNA (36.1°). However, other than this there is no significant perturbation to the duplex part of the triplex structure compared to B-DNA, highlighting that the interaction of the TFO-region in the major groove does not significantly alter the structure of the template duplex.



**Figure 1.6** Schematic representation of (A) triplexes A (**134D**) and (B) triplex B (**149D**). The arrows indicate the four thymine that are reported in the analysis, but do not bind to any complementary base.

Triplex **149D** is also an intramolecular triplex but with a TFO composed of purine bases that bind the duplex strand in a parallel arrangement (Figure 1.5). Whilst the overall structure shows little difference with that of triplex **134D**, which adopts an antiparallel arrangement, local distortions can be observed in individual base triplets. The most significant of these is in the central step within the triplex **149D**, as indicated by arrows in Figure 1.7. At this step, the G base in the triplex strand is unable to form a proper binding interaction with the T-A base pair (a T base would be needed for this to occur). Whilst this mismatch of bases would be expected to reduce the overall stability of triplex assembly, individual sites of mismatched bases do not necessarily prevent triplex formation.<sup>52</sup>



**Figure 1.7** (A) 3D and (B) schematic representations of the G-T:A triplet of the triplex **149D**. Green indicate the duplex bases, guanine and adenine, while the pink base is the guanine of the TFO.

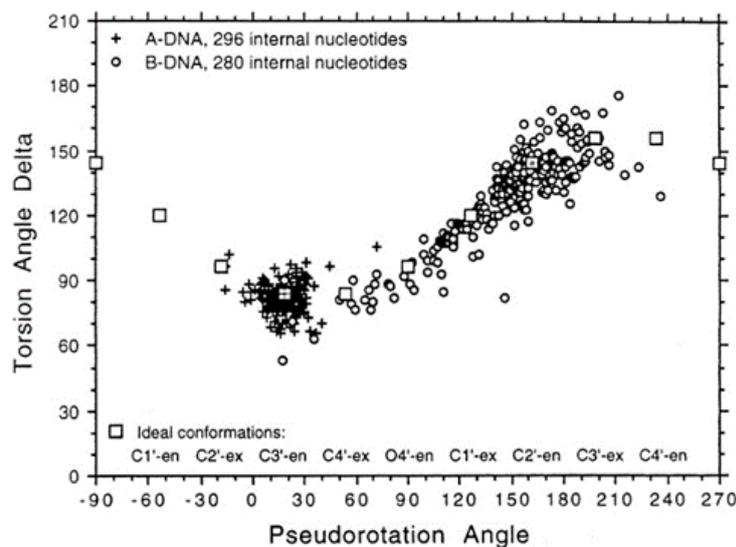
Perturbation of other derived parameters within the structure, including changes in the propeller and buckle value either side of the mismatch site, indicate that this single step of instability may result in an overall reduction of stability or rigidity across the triplex assembly<sup>6</sup> despite an overall twist value of  $30.8^\circ$ , which is reduced compared to that found for B-DNA.

In triplex **1BWG** the TFO is a purine-rich hexamer oligonucleotide that binds in the major groove of a 13 base-pair duplex, forming a parallel triplex assembly (Figure 1.5). In contrast to triplexes **134D** and **149D**, triplex **1BWG** is an intermolecular assembly and the length of the TFO is shorter than the duplex to which it is bound. As illustrated in the schematic representation of the structure in Figure 1.5, the TFO forms both T-A:T and C<sup>+</sup>-G:C triplets, with charge neutralization of the C<sup>+</sup> bases by the phosphate backbone being expected to confer greater stability on the assembly.<sup>40</sup>

Whilst the triplex section of triplex **1BWG** is structurally similar to **134D** and **149D**, this structure gives insight into the structure of the helix either side of the TFO. Whilst the overall twist angle

per step within the triplex region is maintained at ca. 33°, the remaining duplex steps display much greater variability, with twist angles ranging from 29-45° per step.

The reduced twist angle common to triplex structures raises the question of whether the duplex component is closer to B-DNA or A-DNA in conformation, the latter of which is characterized by a reduced twist of ca. 32° per base in combination with a dominant C3'-endo sugar pucker for the ribose ring in the bases. To determine whether the DNA triplex has an A or B conformation, the angle values needed are the backbone sugar torsion  $\delta$ , the glycosyl torsion  $\chi$  or the pseudorotation angle of sugar rings  $P''''P$ .<sup>53</sup> Typically, the duplex can adopt the A- or B- form, and this is dependent on the sugar pucker adopted in each nucleotide. An A-form is adopted when the dominant sugar pucker is C3'-endo, with a pseudorotation value of between -30° and 40°, while a wider range of pseudorotation values can be indicative of the B-DNA conformation. Indeed, the B-conformation is not limited to the C2'-endo pucker, where the majority of the nucleotides can be found, but can adopt several other forms including C4'-exo, O4'-endo, C1'-exo, C3'-exo and C4'-endo<sup>54</sup> (Figure 1.8). The dominant sugar pucker can be used to assign the overall conformation of the helix and is particularly important for the development of ligands designed to target specific steps, as a change in sugar pucker will change the spatial arrangement of atoms around the binding site, potentially changing the mode of interaction by the ligand. The overall conformation of the duplex component of the triplex can be assigned to a conformation using the pseudorotation value (P) for each base.<sup>55</sup>



**Figure 1.8** Representation of sugar rings of B-DNA (circle) and A-DNA (crosses) based on pseudorotation and torsion angle.<sup>185</sup> Reproduced with permission from R. E. Dickerson, *International Tables for X-ray Crystallography, Volume F: Macromolecular Crystallography*, eds M. G. Rossmann, E. Arnold (International Union of Crystallography, Chester, U.K. (2001)).

In triplex **134D**, the P value for the bases forming the duplex lies within the range of 100-160°, indicating a majority B-DNA conformation. Whilst the terminal bases in the duplex lie outside of this range, this could be because of torsional stress placed on the structure due to the folding of the loops, which have not been presented in the coordinates for the structure. The TFO strand, however, displays much less variation in the sugar pucker values. Whilst these again indicate a B-like conformation, the majority of values are either ca. 176° or are within a range of 50-70°, indicating less flexibility in the TFO strand compared to the duplex. This trend, of an overall B-DNA conformation for the duplex matched with less variation in the P for the TFO, is observed in all three structures. However, there are exceptions such as the central G:T-A triplet in Triplex B, which has a P value of 19.0°, indicating an A-like C3'-endo sugar pucker. This may be a pucker which is sequence dependent or could indicate torsional strain in the TFO which is corrected in the central step by the adoption of this unusual pucker.

Due to the relatively small number of DNA triplex structures available, it is not yet possible to identify the expected structural variation as a function of sequence. However, a better understanding of the structural variation expected for DNA triplexes may assist with the development of ligands designed to bind to specific sites within the assembly, to understand the distinctive behavior of triple-helices more structural analysis is essential.

## 1.4 Stability of DNA triplexes

DNA triplexes are less stable than their duplex counterparts in part due to the increased negative charge density from the phosphate backbones, which increases repulsion between the strands. However, multiple factors can affect the stability of a triplex assembly including the presence and concentration of monovalent or divalent cations, pH and temperature. Additionally, triplex hybridization can be promoted by the presence of molecular crowding and chromatin accessibility in the biological environment.

Efforts have been made to increase triplex stability through chemical modification of the base, sugar, or phosphate backbone of DNA.<sup>56,57</sup> Base modifications have been the focus of extensive synthetic efforts due to pH being one significant factor which can negatively affect triplex stability. Modification has also focused on changing the phosphate and sugar within the nucleotides to enhance resistance to nucleases in the cell, in order to reduce degradation, and to enhance the ability of the TFO to enter and bind in the major groove of the duplex.

Finally, the use of ligands, such as intercalators or groove binders, has been explored as one approach to increase triplex stability without chemical modification to the TFO, although this is a secondary effect of targeting the triplex assembly with such a molecule.

#### 1.4.1 Cations and anion enhancement of DNA triplex stability

The cellular microenvironment exercises direct control over triplex stability and activity at the molecular level. Considering the intense negative charge of a structure that is formed by three strands of DNA, a high concentration of multivalent cations will mitigate the repulsion.<sup>58</sup> Generally, it is agreed that the formation of intermolecular triplexes with a polypurine sequence requires divalent cations<sup>59</sup> such as  $Mg^{2+}$ , whereas for the intramolecular assembly,  $Na^+$  are sufficient. It has also been demonstrated that the inclusion of  $Mg^{2+}$  can contribute to an increase in stability of reverse Hoogsteen bonds, resulting in an increased thermal stability for intramolecular triplexes.<sup>60</sup>

Several cations can increase triplex stability. For divalent cations, the order of stabilisation is  $Mg^{2+} > Mn^{2+} > Ca^{2+} > Ba^{2+}$ , which can be attributed to the ionic radius of each ion - the smaller the radius is, the greater the alignment between nucleotides and hence the greater the stability of the triplex assembly.<sup>61</sup>

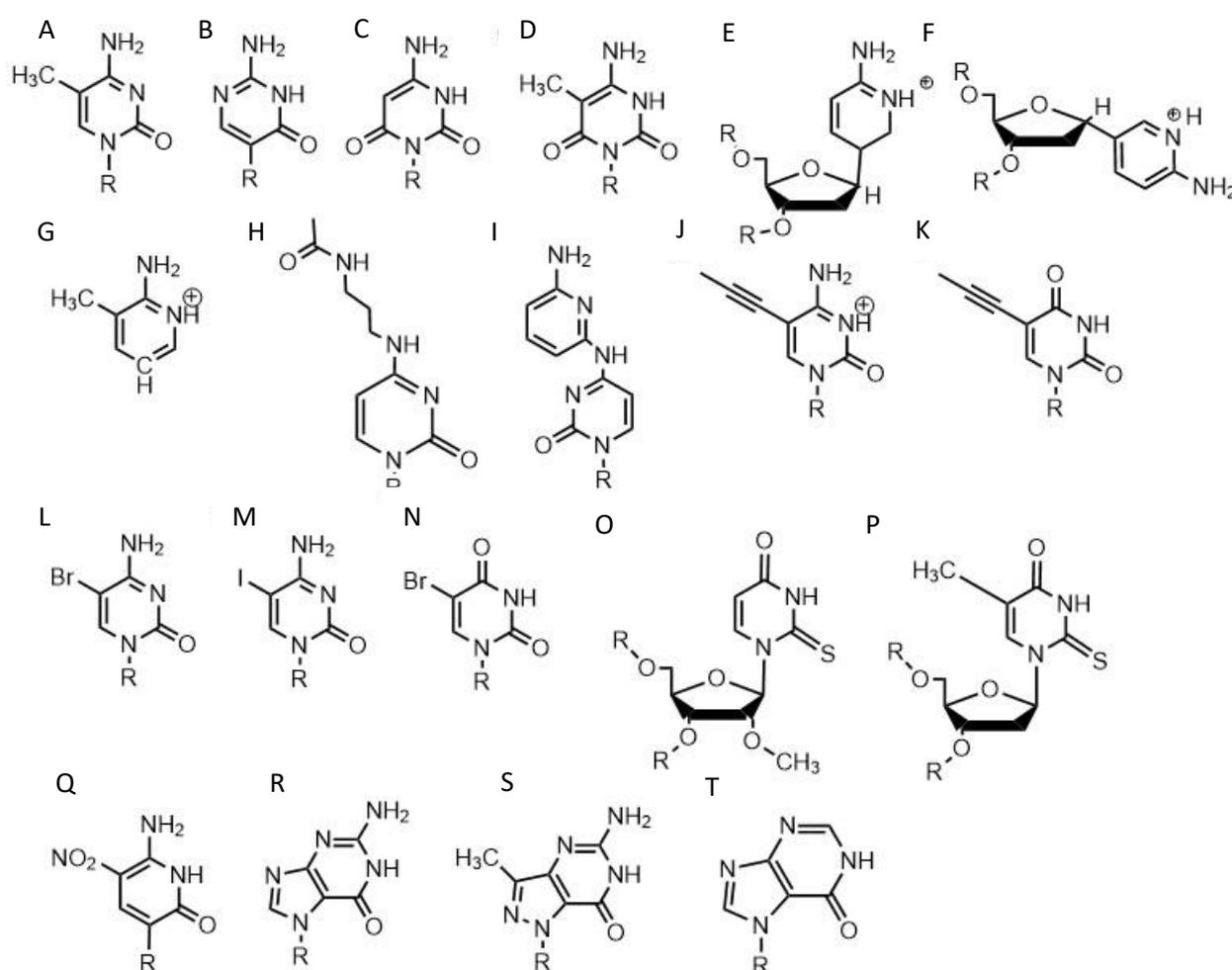
In contrast, monovalent ions, such as the physiological concentration of  $K^+$ , reduce the propensity of a G-rich strand to form a triplex. The presence of molecular crowding conditions, which are often simulated in vitro by using high concentrations of polyethylene glycol, such as PEG 200, can also affect the formation of triplexes, and with a G-rich strand in the presence of  $Ca^{2+}$ , the formation of a G-triplex is promoted with endothermic energy.<sup>62</sup> Molecular crowding conditions can also promote triplex formation and change the effect on stability of adding monovalent ions. For example, in the absence of crowding conditions, the addition of  $K^+$  has been demonstrated to increase triplex stability as a function of  $K^+$  concentration. However, in crowding conditions, the addition of  $K^+$  actually reduces the stability of the triplex assembly.<sup>63</sup>

Using a crowding agent along with ions to simulate the environment in which triplexes might be found, short triplexes tend to stack together and form a highly condensed structure.<sup>64</sup> Since this effect was also observed with duplexes, it has been argued that DNA triplexes may affect the genome structure with modification at a chromosome level.<sup>65</sup>

## $\alpha$ 1.4.2 Base modifications

An increasing number of oligonucleotide analogues have been developed to obtain TFOs with increased stability (both of the resulting triplex and increased resistance to degradation by nucleases) and enable greater selectivity of targeting towards specific structures or DNA sequences.<sup>56,66</sup>

**Base modifications in parallel triplexes.** Parallel triplex stability can be increased when the sequence contains a greater number of C<sup>+</sup>-G:G triads rather than T-A:T steps, but the observed stability is still pH-dependent, with an optimal pH below 6.2.<sup>67</sup>



**Figure 1.9** Base modifications in parallel triplexes. (A) 5-methyl-cytosine, (B) pseudoisocytidine, (C) 6-oxo-cytosine, (D) 5-methyl-6-oxo-cytosine, (E)  $\alpha$ -AP, (F)  $\beta$ -AP, (G) 2'-aminoethoxy-thymine, (H)  $N^4$ -3-acetamidopropyl-cytosine, (I)  $N^4$ -6-aminopyridinyl-cytosine, (J) 5-propynyl-cytosine, (K) 5-propynyl-uracil, (L) 5-bromo-cytosine, (M) 5-iodo-cytosine, (N) 5-bromo-uridine, (O) 2'-O-methyl-2-thio-uridine, (P) 2-thio-thymidine, (Q) 6-amino-5-nitropyridin-2-one, (R)  $N^7$ -glycosylated-guanine, (S)  $P1$ -guanine, (T) inosine.

Indeed, the protonation of the cytosine bases will provide a second hydrogen bond between the N3 of cytosine itself and the N7 of guanine, favouring a Hoogsteen bond, and consequently the triplex formation, in mildly acidic conditions.<sup>68</sup>

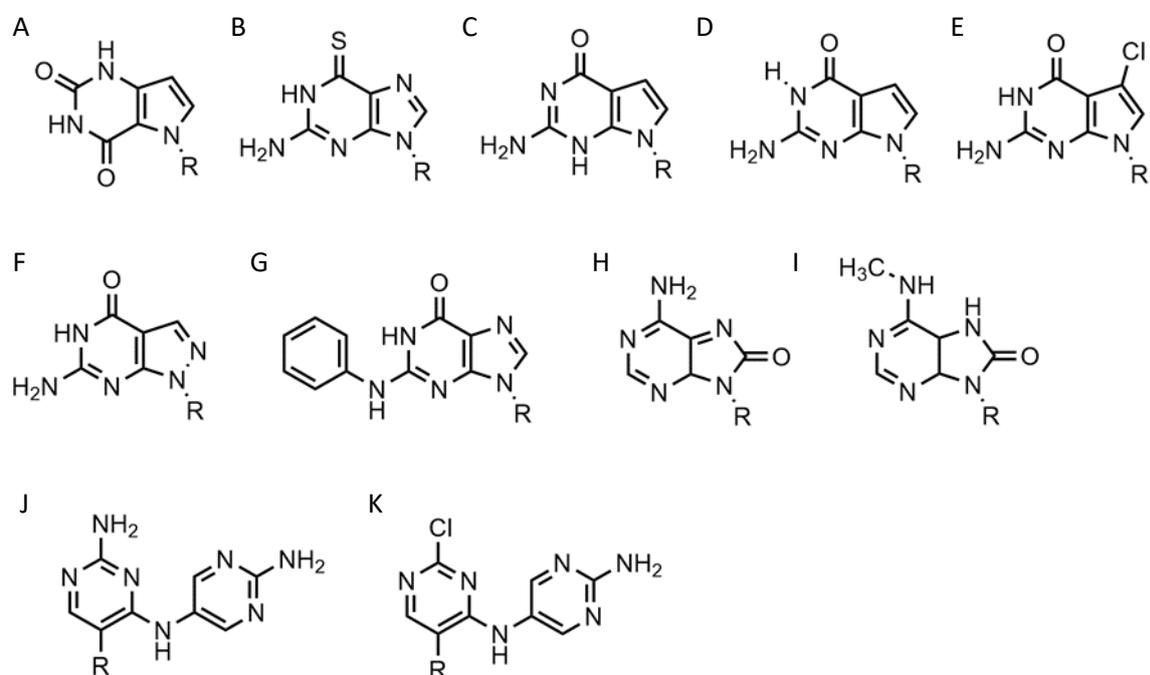
However, a series of C bases in a tract will result in lower triplex stability, due to the proximity of multiple charges from the protonated bases, which require more acidic conditions to stabilise.<sup>22</sup> This has prompted researchers to focus on cytosine analogues that support pH-independent triplex formation e.g., neutral cytosines with two hydrogen donor groups, or analogues that protonate more easily (Figure 1.9). To reduce pH-dependency, modifications to cytosine have been explored, with the aim of increasing triplex stability in a wider pH range. The methylation of cytosine in the TFO, in 5-methyl-cytosine, contributes to the base stacking, increasing stability. (Figure 1.9A).<sup>56,67</sup> In recent calculations, it was demonstrated that the methylation of C also shifts the pKa from 4.6 (for cytidine) to 4.9.<sup>69</sup> The pseudoisocytidine (Figure 1.9B) promotes the formation of triplexes in a neutral environment in a TFO that will recognize GC-tracts, exemplified using poly(GC). However, this modification is not widely used because the synthesis is highly challenging, even though it is an excellent candidate for increasing the stability of parallel triplexes. The stabilising effect of this pseudo-isocytosine is reported in intramolecular triplexes with a loop composed of only two bases.<sup>70</sup> The incorporation of 6-oxo-cytosine can increase triplex stability to above pH 7, however, when compared with triplexes containing protonated or methylated cytosines in more acidic conditions, the stability decreases (Figure 1.9C). The analogue 6-oxo-cytosine can be further modified by the addition of a methyl group in position 5, obtaining 5-methyl-6-oxo-cytosine (Figure 1.9D), which can also promote the stability of DNA triplexes. The use of glycerol linkers combined with 6-oxocytosine has been proposed as a modification, which reduces the steric interaction between the 6-carbonyl and the sugar, increasing the stability of the triplex in comparison to the stability observed with no linker present. The absence of glycerol linkers particularly reduces the stability of the triplex if it contains a G-tract.<sup>71</sup> The incorporation of 2-aminopyrimidine (AP) can promote increased triplex stability at physiological pH without protonation due to the low basicity of the modified base. AP can be incorporated in the TFO as  $\beta$  and  $\alpha$ -anomers, the first cytosine anomer has a lower pH dependency due to its pKa of 6.5, resulting in stable triplexes (Figure 1.9E-F).<sup>72,73</sup> Unsurprisingly, the addition of a 5-methyl-cytosine in the same TFO containing the  $\beta$ -AP does not form triplexes because of the unfavourable steric interaction. An alternative is a combination of the methylated version of the 2-aminopyrimidine with the 2'-aminoethoxy-thymine (Figure 1.9G) reaching a binding affinity at pH 9.0.<sup>74</sup>

In terms of base recognition, modified oligonucleotides play a crucial role in enhancing sequence-specific recognition. In the case of parallel triplexes, a TFO with N4-3-acetamidopropyl-cytosine can recognise a GC base pair, by forming a more stable triplex due to the increased chain flexibility (Figure 1.9H) and has higher stability than the equivalent TFO containing only cytosine.<sup>75</sup> A similar example reported is N4-6-aminopyridinyl-cytosine, which can recognise pyrimidine base interruptions in a polypurine sequence (Figure 1.9I).<sup>76</sup> The addition of a propynyl group can increase triplex hydrophobicity and consequently stacking interaction. Another example reported is 5-propynyl-cytosine, which replaced the cytosine, but when the propynyl group was attached to uracil, the TFO with 5-propynyl-uracil is more favourable for the stability of parallel triplex compared to the 5-propynyl-cytosine (Figure 1.9J-K).<sup>56,75</sup>

Cytosine analogues containing bromine or iodine atoms at position 5 have also been explored, obtaining 5-bromo-cytosine and 5-iodo-cytosine respectively, but the incorporation of these into a TFO, by replacement of cytosine, actually reduced triplex stability (Figure 1.9L-M). Instead, the substitution of thymine by a 5-bromo-uridine (Figure 1.9N) enabled the formation of triplex at room temperature. The inability to obtain triplex structures with 5-halocytosine derivatives is explained by their lower pKa and the requirement of protonation.<sup>77</sup> Some studies show that the use of modification on both uracil and thymine in a TFO, such as 2'-O-methyl-2-thio-uridine and 2-thio-thymidine increases the stability of a DNA parallel triplex and the reason is the stacking properties of the 2-thiocarbonyl on the 5' of the upper thiouracil base and the nitrogen atom of the 3' of the lower pyrimidine (Figure 1.9O-P). Additionally, it is emphasized that a TFO that includes thiocarbonyl moieties recognizes a base mismatch, a key feature for antibody therapies.<sup>78</sup> A recently published study proposed a TFO containing 6-amino-5-nitropyridin-2-one that overcomes the need for protonation, by acting as an uncharged mimic that can form a parallel triplex, with in vitro evidence demonstrating that this approach shows promise (Figure 1.9Q). The modified nucleobase was included in the TFO through an enzymatic process at physiological pH, relying on the thermodynamic stability of 6-amino-5-nitropyridin-2-one compared to other mismatched bases. Additionally, the modified TFO enhanced protection from DNA from nucleases.<sup>79</sup> The modification of purine bases has also been explored, although this has received less attention than the pyrimidines. A substitution of N7-glycosylated-guanine or P1-guanine with a cytosine has a remarkable impact on the triplex stability when in the presence of a G-tract (Figure 1.9R-S).<sup>45,80,81</sup> If guanine is converted into inosine by removal of the guanine-N2 amino group, then this is able

to recognise a GC base pair and form a triplex structure. Additionally, the absence of the amino function gives space to an unusual bond between the carbonyl group of the modified base and the CH of the guanine of the duplex, resulting in a higher electrostatic stability (Figure 1.9T).<sup>82</sup>

**Base modification in anti-parallel triplexes.** The principal concern when working with anti-parallel triplexes is the competitive formation of a G-quadruplex structure due to large numbers of guanine residues. It has been reported that the physiological level of  $K^+$  (over 100 mM) will stabilize the formation of quadruplexes rather than triplexes. Therefore, the aim of chemical modification is to produce analogues that will prevent quadruplex formation whilst promoting the formation of a parallel triplex (Figure 1.10).



**Figure 1.10** Base modification for anti-parallel triplexes. (A) 7-deaza-xanthine, (B) 6-thioguanine, (C) 9-deaza-guanine, (D) 7-deaza-guanine, (E) 7-chloro-7-deaza-guanine, (F) 8-aza-7-deaza-guanine, (G) PhdG, (H) 8-oxo-adenine, (I) N<sup>6</sup>-methyl-8-oxo-adenine, (J) AY-d(Y-NH<sub>2</sub>), (K) AY-d(Y-Cl).

To stabilise an antiparallel DNA triplex, it was proposed to replace thymine with a 7-deaza-xanthosine (Figure 1.10A). The introduction of this modification will reduce the likelihood of the oligo assuming a G-quadruplex structure, since the N7 needed as a hydrogen donor in the modified guanine is absent.<sup>83</sup> The essential role of potassium suggests, 6-thio-guanine should prevent the formation of quadruplexes due to the very weak electron pair donor properties of the S lone pairs to  $K^+$  ions compared to the carbonyl groups. (Figure 1.10B).<sup>84,85</sup> Other examples of analogues that prevent  $K^+$  coordination are 9-deaza-guanine, 7-deaza-guanine and 7-chloro-7deaza-guanine, and although they will form triple-helical structures, there is no sign of significantly increased triplex stability (Figure 1.10C-E).<sup>86-88</sup> Instead, a purine modification in parallel triplexes that can form triplexes in G-rich TFO at physiological  $[K^+]$  is 8-aza-7-deaza-guanine (PPG) (Figure 1.10F). Furthermore, modified TFOs containing this modification were used in cells to generate triplex-induced mutations and to cause double-strand breaks (DSBs) that will lead to cell death. These results show the modification forming a stable triplex, preventing G-quartet formation and inducing gene modification, editing and cell apoptosis.<sup>89</sup> Another more recent modification included in antiparallel triplex DNA is the product from a synthesis of a guanine derivative N2-Phenyl-2'-deoxyguanosine (PhdG), which was shown to form a stable and selective triplex with the GC base pair (Figure 1.10G). As a drawback, as more PhdG bases are introduced, there is an increased likelihood that the structure could assume a higher order.<sup>90</sup> Finally, to support triplex formation in the presence of high  $[K^+]$ , the protonation of the backbone is often used as an alternative approach,<sup>91</sup> which will be discussed in the next section.

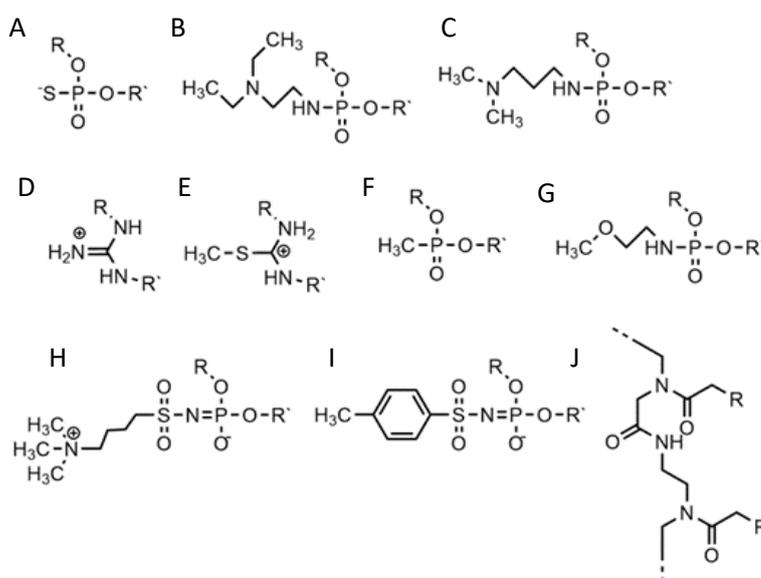
The modification of adenine has also been reported as a potential route to enhancing triplex stability. The purine analogue 8-oxo-adenine forms stable Hoogsteen bonds with a G:C Watson-Crick base pairing (Figure 1.10H). Additionally, an N6-methyl-8-oxo-adenine binds a purine sequence improving the triple-helical stability (Figure 1.11I).<sup>92</sup> The 8-NH<sub>2</sub> modification of the 8-amino-purine creates a stable interaction either with cytosine or guanine. Therefore, numerous 8-amino-purine derivatives were tested in DNA triplexes, demonstrating that, regardless of structural alterations to the chemical structure, antiparallel triplexes are found to be more stable in physiological pH conditions.<sup>93</sup>

Pyrimidine derivatives have been exploited to stabilize anti-parallel triplexes. The incorporation of a cytosine nucleoside containing an amino-pyrimidine unit AY-d(Y-NH<sub>2</sub>) or AY-d(Y-Cl), results

in stable triplexes able to recognise the inverted G:C instead of the canonical C:G, or T:A instead A:T with a duplex (Figure 1.10J-K).<sup>94</sup>

### 1.4.3 Phosphate backbone modification

An alternative strategy to promote triplex stability is to focus on modifications to the phosphate backbone of the oligonucleotide.<sup>95</sup> In general, TFOs are more likely to form a self-associated structure when the backbone is neutral or cationic, due to a decrease of the electrostatic repulsion between the three anionic strands. Modifications have been designed to promote a higher affinity between the TFO and the duplex strands whilst also increasing TFO nuclease resistance, which is important for the longevity of a TFO strand inside a cellular environment. A significant number of backbone modifications have been explored in the context of DNA triplexes (Figure 1.11).



**Figure 1.11** Phosphate backbone modifications. (A) phosphorothioates, (B) DEED, (C) DMAP, (D) Guanidino, (E) methylthiourea, (F) methyl-phosphonates, (G) PNHME, (H) azide-phosphoramidate, (I) tosyl sulfonyl phosphoramidite, (J) PNA.

One of the first to be produced, the phosphorothioate modification (S-oligos), included a substitution to one of the non-bridging oxygen atoms in the phosphate group, replacing the O with S. This modification presents a significant drawback when applied *in vivo*, as TFOs containing this modification tend to bind proteins non-specifically. So, whilst this modification does confer nuclease resistance to the TFO, increasing longevity in the cell, it still maintains the negatively charged backbone, which is thought to reduce triplex stability (Figure 1.11A).<sup>96</sup>

The formation of positively charged backbones have been explored, with the incorporation of cationic amine groups into the DNA backbone, including groups such as N,N-diethyl-ethylenediamine (DEED) or N,N-dimethyl-aminopropylamine (DMAP) (Figure 1.11B-C). These modifications increase the binding affinity of the TFO in vitro and make them increasingly nuclease-resistant. Increasingly complex changes in the backbone modification have been also proposed. Guanidino and methylthiourea are some examples of a complete substitution of the phosphate group with a cationic linked nucleoside, resulting in more stable triplex oligomers even though an increasing proportion of T and A in parallel triplexes decreases the melting temperature (Figure 1.11D-E).<sup>97,98</sup>

Options to obtain non-ionic alternatives are available as well, such as methyl-phosphonates, phosphotriesters and non-phosphate hydrazide derivatives (Figure 1.11F). However, the lack of charge makes them highly insoluble, and they are therefore less suited to in vivo applications.

Different phosphoramidate-linkage modified TFOs have been proposed to bind the dsDNA efficiently by enhancing DNA stability.<sup>99</sup> An example of a phosphoramidate-modified oligonucleotide is methoxyethylphosphoramidate (PNHME) for pyrimidine with the  $\alpha$ -anomeric configuration (Figure 1.11G).<sup>100</sup> Whilst backbone repulsion is decreased, the triplex is only formed at pH 7 or lower and therefore this process is still protonation dependent.

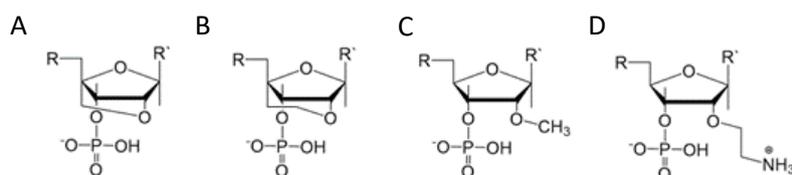
Recently, an increasing number of studies have incorporated zwitterionic modifications in oligonucleotides, yielding thermostable triplexes. A quaternary ammonium group substitutes for the negative phosphoramidate, neutralising the backbone (Figure 1.11H), and, as a consequence, the duplex formation is less dependent on ionic strength. While this change increased the hydrophobicity of the molecule compared to the unmodified DNA, stable parallel triplexes form at a pH optimum of 5, and only when the modification is at the 3' end. Furthermore, the presence of a tosylsulfonyl phosphoramidite (Ts) can be exploited as a negatively charged phosphate (Figure 1.11I). Both modifications, when introduced into the same oligonucleotide, form stable parallel DNA triplexes and show promise for in vivo applications, especially as nuclease resistance and cellular uptake were increased compared to non-modified oligonucleotides.<sup>101</sup>

A more drastic modification of the phosphodeoxyribose backbone features the use of peptide nucleic acids, PNA (Figure 1.11J). A PNA strand was conceived as a triplex-forming oligonucleotide, able to bind to a dsDNA due to its neutral charge. Subsequently, it appeared that

two PNA strands, where the phosphate backbone is substituted by units of N-(2-aminoethyl) glycine, form remarkably stable triplexes when binding the unmodified TFO. The high stability of the triple-helical structure arises primarily from the neutral charge, drastically reducing backbone repulsion. Molecular dynamics simulations confirm that PNA backbones provide additional flexibility to the triplex and in some cases can assume A-type conformations.<sup>102,103</sup> An alteration of PNA was proposed with an arginine instead of glycine, forming the G-PNA. This modification has overcome the solubility issue of PNA.<sup>104</sup> Two other modifications reported are olefinic peptide nucleic acids (OPA) and oxy-PNA. These alternatives seek to improve cellular uptake rather than the triplex stability itself.<sup>105,106</sup> Alternatively, PNA can contain ligands with coordinated metal ions instead of nucleobases. The outcome is a stable triplex in solution experiments, due to the strength of coordinative bonds compared to hydrogen bonds, but which is reduced by the steric interactions of the metal-complex and the triplexes.<sup>107</sup> The use of PNA has been explored in a number of different areas including cellular uptake<sup>108</sup>, regulation of gene expression<sup>109</sup>, interruption of the RNA polymerase and inhibition of translation and activation of DNA repair system.<sup>95</sup> It shows great potential as a future therapeutic, and work in this area is ongoing to address some of the challenges associated with its use, such as cellular delivery.

#### 1.4.4 Sugar modification

Sugar modifications focus on the sugar pucker conformations that will influence the ability of the TFO to form a stable structure (Figure 1.12). The most common approach used to increase the stability of the triplex restricts the range of sugar conformations, relying on the use of bridged nucleic acids (BNA).<sup>110</sup> The puckering characteristics of the sugar ring allows the ribofuranose structure to assume a range of conformations but, once it was realised that the C3'-endo configuration is more likely to stabilise a triplex, a range of modifications were explored, with the aim of promoting this conformation.<sup>47,111</sup>



**Figure 1.12** Sugar Backbone modifications. (A) LNA, (B) ENA, (C) 2'-OMe, (D) 2'-AE.

The first generation of BNA is locked nucleic acid (LNA), which consists of a 2'-O, 4'-C methylene bridge that restricts the sugar backbone movement (Figure 1.12A) and promotes the formation of an A-form duplex in the binding partner of the LNA strand. This reduced flexibility enhances the stability and selectivity of the TFO strand. It has been reported that including short LNA residues in pyrimidine-motif triplexes will enhance stability due to the significant puckering amplitude.<sup>59</sup> However, the modification can only be included once in every 2-3 nucleotides; a TFO composed of only LNA modifications does not form triple-helices.

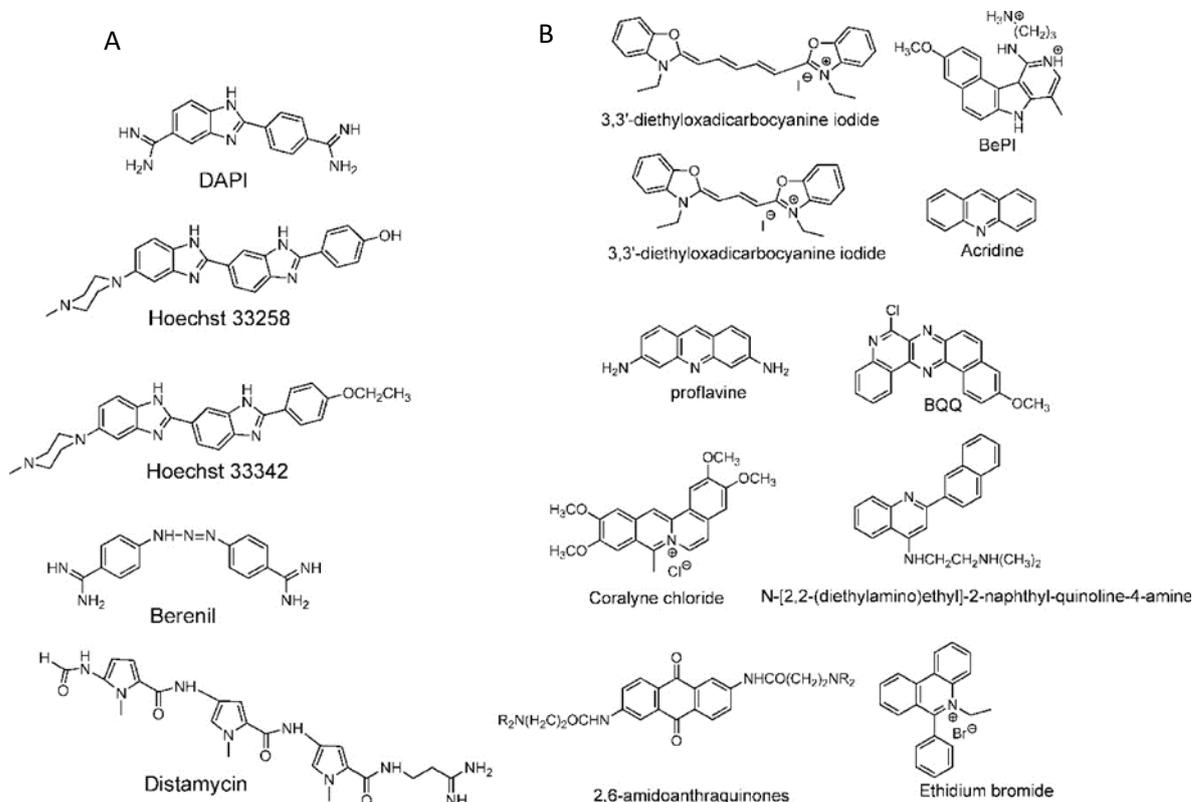
A second modification, with an ethylene link (ethylene-bridged nucleic acid, ENA, Figure 1.12B) instead of methylene, was proposed to overcome this incorporation limit and allows for the production of fully modified TFOs able to form a triplex.<sup>47,112</sup> This modification is less restrictive, and therefore allows for a greater variation in the observed LNA sugar pucker, giving more flexibility to accommodate the third strand, which can be composed fully of ENA.<sup>113</sup>

Alternative strategies to modify the sugar component of the TFO without imposing a locked conformation are the addition of an ammonium group to the sugar, 2'-O-methylribose (2'-OMe), (Figure 12C)<sup>114</sup> or a protonated aminoethyl group at C3'-endo, 2'-O-aminoethylribose (2'-AE) (Figure 12D). Both modifications bias the sugar pucker towards C3'-endo, favouring the A-form conformation, improving the stability of the TFO towards nucleases and enhancing triplex stability.<sup>115</sup>

Continued development of nucleotide analogues, modified phosphate backbone and sugar-based variants is ongoing. When evaluating a modified TFO, an ideal candidate forms DNA triplexes with a high association rate and remain thermostable, both in vitro and in vivo. Thus far, modifications have typically been investigated singularly i.e., candidate TFO strands have contained one modification, although this can be at multiple sites within a single strand. Future development should therefore focus on combining modifications to provide a successful outcome in terms of triplex stability and biological function. Indeed, for cellular applications, it must be taken into consideration that the TFO or DNA triplex must initially be delivered into the cell and therefore the hydrophobicity properties must be considered and carefully balanced. The main challenges, however, are to stabilize the triplex at physiological pH, maximise nuclease resistance and finally promote specificity in sequence targeting.

### 1.4.5 DNA triplexes intercalators and groove binders

A completely different approach to DNA triplex enhancement that does not require chemical modification, or solutes is the noncovalent intercalation of a small molecule stabiliser. The latter are molecules, widely studied over the years, which are able to specifically bind DNA triplexes, since they can provide tools to enhance triplex stability and support biological applications. As previously seen, triplex structures are less stable than duplexes. Specifically, the need for cytosine protonation in the pyrimidine third strand leads to limited triplex stability at physiological pH. For these reasons, intercalation by molecules able to selectively stabilize the triplex structure is of great interest (Figure 1.13).<sup>96</sup>



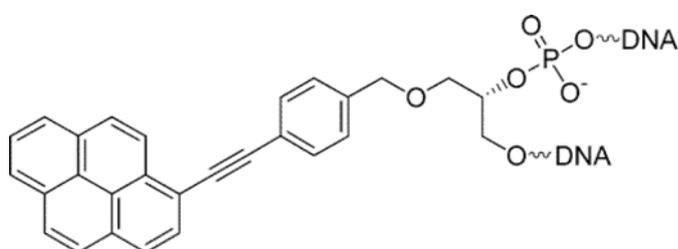
**Figure 1.13** (A) DNA triplex groove binders and (B) DNA triplex intercalators. Adapted with permission from D. P. Arya, *Acc. Chem. Res.*, 2011, **44**, 134–146. Copyright 2011 American Chemical Society.

For example, the common duplex DNA binder ethidium bromide (EtBr) can also stabilize a C-G:C structure with a triplex-specific stabilizing effect, due to the electrostatic repulsion between ethidium and cytosine. However, the stabilisation of the triplex with ligands will also depend on the concentration of the chosen ligand. It has been reported that two molecules of either EtBr or acridine orange (AO) in 10-base pair long triplex will stabilise the structure, while a third molecule

leads to destabilisation, highlighting that the effect of concentration must be carefully balanced.<sup>116</sup> Also, the increase of stability, measured as the increase in the triplex melting temperatures, depends on the DNA sequence. The melting temperatures of the 15-mer triplexes were obtained from the hyperchromicity observed at 260 nm upon thermal denaturation. A larger increase in melting temperatures for sequences having A-tract duplex structures was observed by UV spectroscopy, using a ratio of 2:1 pyrimidine to purine strand. This large thermal stabilizing effect on T-A:T triplexes is partly due to the intercalators that break up the intrinsic A-tract structure of the underlying duplex.<sup>117</sup> In fact, the intrinsically rigid and highly propeller-twisted structure of A-tract DNA disfavours triplex formation.<sup>117</sup> Propidium iodide (PI) has been reported as a potent stabiliser of the parallel triple-helix, with association constant similar to that of PI binding to duplex DNA.<sup>118</sup> PI was shown to increase the parallel triplex stability after intercalation of three molecules into the triplex, with melting temperature increasing from 21.4 up to 44.4°C in different media such as Na phosphate buffer, pH 7 and NaCl.<sup>119</sup>

Other DNA triplex binding intercalators include indolocarbazole and benzopyridoquinoxaline derivatives. These provide additional stacking interactions with the pyrimidine strand of the Watson–Crick double-helix, resulting in a very efficient and specific stabilizing effect on triple-helices and/or in inducing triple-helix formation under physiological conditions.<sup>99</sup>

Another class of intercalators able to stabilize the triple-helices are twisted intercalating nucleic acids (TINA) (Figure 1.14). These nucleic acids are bound in the TFO sequence providing a nucleobase stacking mimic and are characterized by the ability to twist around a triple bond. This twisting promotes intercalation within double-stranded DNA in order to form triplex DNA. Moreover, it has been demonstrated that these oligonucleotides can discriminate between matched and mismatched sequences of DNA.<sup>120–122</sup>



**Figure 1.14** Example of structure of a TINA intercalating unit. Reprinted with permission from I. Géci, V. V. Filichev and E. B. Pedersen, *Bioconjug. Chem.*, 2006, **17**, 950–957. Copyright 2006 American Chemical Society.

In this context, it is worth noting that intercalators usually have a stabilising effect on DNA triplexes, whereas minor groove binders will generally destabilise the assembly (Figure 1.13). Nevertheless, some aminoglycosides were tested as triplex binders, and it was shown that neomycin selectively recognizes the triplex Watson-Hoogsteen groove and stabilizes it without any effect on dsDNA. This very interesting selectivity may be related to the shape complementarity to the triplex Watson-Hoogsteen groove (the groove formed between the TFO and DNA strand which does not bind to the TFO).<sup>123</sup> Other minor groove binders that are well exploited are netropsin, spermine and cyclopolymines.<sup>124</sup> Psoralen has also been used as it can intercalate efficiently between bases and can provide a covalent linkage by forming an adduct on photoreaction with the stacked pyrimidine.<sup>125–127</sup>

Other reported groove binders are Hoechst 33258, Berenil, DAPI and Distamycin A (Figure 1.13), however, their stabilizer ability as well as the triplex stability is lower than with neomycin. In this area almost no structural characterisation of triplex-ligand systems has taken place and therefore this is an area which could be the subject of future focus to understand the DNA triplex-ligand molecular interaction.<sup>99,124</sup>

## 1.5 Applications based on biomolecular approaches

The ability to form a three-stranded complex based on base-base recognition can be exploited to develop biotechnologies suited for diagnosis, prognosis, or disease treatment. Indeed, a modified TFO included in a dsDNA is considered as a potential future for genetic medicine, exploiting sequence-specificity to target genes for manipulation. TFOs have proven to be useful tools, able to alter gene expression and cause genome modification in mammalian cells.<sup>128</sup> However, several limitations must be overcome to improve their therapeutic value. Often, these applications are restricted because of the low-affinity binding in vivo conditions, as well as TFO stability and integrity during cellular uptake. Numerous attempts have been made to modify oligonucleotides and improve these characteristics, as discussed above.<sup>96</sup>

The ability of a TFO to inhibit a transcription was demonstrated for the first time with the human c-myc protooncogene in HeLa cells. This protooncogene plays a crucial role in normal cell proliferation and programmed cell death. In particular, c-myc gene expression is present in cancer cells at an increased level compared to normal cells.<sup>129</sup> Specifically, after entering the nucleus, TFOs bind to the DNA duplex at the target sequence to form the triple-helix, which prevents the

polymerase and other transcription factors from initiating transcription. This results in the inhibition of mRNA synthesis from the c-myc promoter, demonstrating that the administration of the TFO to the cells can influence the transcription of the c-myc gene.<sup>130</sup>

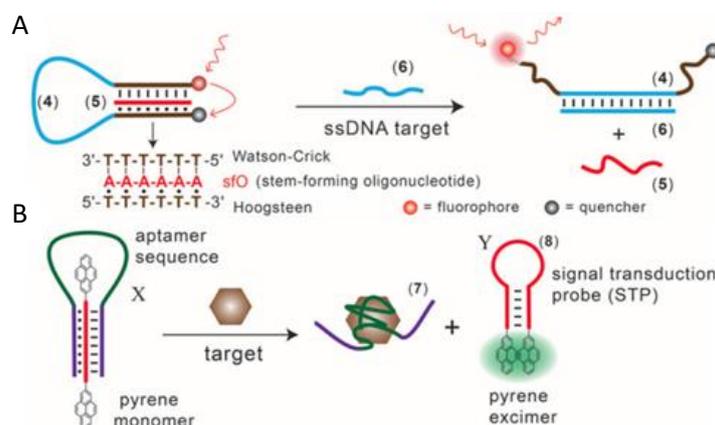
A therapeutic application that was proposed relied on the ability of the TFOs to bind a duplex structure related to the Friedreich's ataxia gene. The formation of the triplex structure stalls the RNA polymerase and decreases the frataxin protein level, which causes the disease. The GAA triplet repeat, which is responsible for the neurodegenerative disease, folds back, forming a triplex structure with the polypurine strand. In this case, disfavoring the formation of the triplex structure could be the key to restore the FXN gene transcription, and therefore regenerate the normal frataxin protein level.<sup>131,132</sup>

Since the TFO should form a DNA triplex along a gene of interest, it is useful to direct a site-specific mutation. Indeed, a psoralen-modified TFO directed to the supF reporter gene, along with UV irradiation in order to allow the cross-linking of the psoralen to the DNA, resulted in a 100-fold increase of mutations, in which 70% are TA to AT transversions. In mammalian cells, chromosomal mutations have been enhanced tenfold after targeting specific genes. Moreover, triplex formation creates a helical distortion to trigger DNA repair by different pathways, i.e. involving the nucleotide excision repair (NER) system or homologous recombination (HR).<sup>96</sup>

In addition to induced mutagenesis, another role of DNA triplexes are genome modification based on the recombination strategy. Triplex technology was used to determine whether interstrand cross-links (ICL) could be repaired through homologous recombination (HR). Indeed, a green fluorescent protein reporter forms a triplex with the psoralen-TFO and intercalates through the specific ICL sequence by confirming the HR effect.<sup>133</sup> Moreover, targeting a specific gene sequence could be used for deleting or replacing sequences on chromosomes. Therefore, a DNA break that happens during the formation of the triplex, stimulates the recombination. To support this notion, a simian virus 40 (SV40) shuttle vector was modified to present psoralen-TFO, then inoculated in human cells, resulting in DNA damage. As consequence, a mutation is induced in a NER/XPA dependent manner.<sup>134,135</sup> A result obtained with luciferase reporter assays shows that p53 was transactivated when a triplex-forming sequence, introduced via plasmid, was formed close to the p53 target sequence.<sup>136</sup>

As reported above, one of the major problems related to TFO application *in vivo* is the instability of the triplex at neutral pH, due to the requirement of cytosine protonation to form the triplex, which is not possible at physiological pH. Different strategies have been studied, such as walled nanotubes (SWNT), to stabilize C-G:C triplexes under physiological conditions. Such studies may facilitate the application of nanomaterials in the artificial control of gene expression and biosensing.<sup>137</sup> Another interesting and very recent approach proposes to modify the TFOs with the nucleobase 6-amino-5-nitropyridin-2-one (Z), which acts as uncharged replacement for the protonated cytosine. By using this method, Rusling obtained stable and selective triplex formation stable at neutral or even slightly basic pH.<sup>79</sup>

Triplex DNA structures were also used as structure-switching units to trigger a signal, following the recognition of specific targets such as proteins, antibodies, small molecules and pH.<sup>138</sup> For example, fluorophore/quencher pair molecular beacons are exploited as optical switches to detect pathogens and genetic disorders. These tools can be used with triplex structures (Figure 1.15A).



**Figure 1.15** (A) Optical sensor based on hairpin triplex structure (4) of a target gene (6) by the reconfiguration of a fluorophore/quencher-modified triplex DNA hairpin structure and the release of the stem forming oligonucleotide (5). B) A triplex DNA hairpin moiety (X) containing an aptamer sequence used as an optical aptasensor that binds the target (7) with subsequent formation of a hairpin excited structure (8). Reprinted and adapted from *Triplex DNA Nanostructures: From Basic Properties to Applications* Y. Hu, A. Cecconello, A. Idili, F. Ricci, and I. Willner, Pages 15210 –15233, Copyright (2017), *Angewandte Chemie. Chemical Society*.

Indeed, a hairpin triplex helix, functionalised with a fluorophore in one edge, and the quencher in the other edge, is reconfigured in an open structure after recognition of the target. The target recognition leads to the opening of the triplex structure and an increase in fluorescence, due to the spatial separation of the fluorophore and the quencher that were adjacent when the hairpin triplex structure was formed. This idea was applied in the design of a bimolecular triplex helix

stem for the analysis of a DNA single strand. The stem containing a T-A-T triplex incorporating a poly-T DNA and a poly-A peptide nucleic acid (PNA) strand was used to increase the stability of a molecular beacon. In this case, after recognition of the target, i.e. the single strand of DNA, the formation of the DNA duplex leads to the opening of the triplex structure with an increase in the fluorescence signal (Figure 1.15A).<sup>139</sup>

Triplex-based hairpins have also been exploited with luminescent pair to obtain a sensing platform. This system is characterised by the presence of a pyrene excimer pair attached to the two edges of a linear triplex forming oligonucleotide. Once the hairpin portion of the triplex recognises the analyte, the hairpin is opened and folds around the target molecule. The stem with the pyrenes is thus released, and able to fold into another hairpin structure, causing the contact of the luminescent pair (Figure 1.15B). This results in the emission of the pyrene excimer at 485 nm. The emission level is then proportional to the concentration of the target species. This sensing platform has been used for the detection of thrombin, ATP or L-arginamide. All these methods exploit the presence of an anti-thrombin/anti-ATP/anti-L-arginamide aptamer sequence in the triplex-based hairpin. Indeed, many sensors can be designed, but their efficacy depends on reliable opening of the hairpin triplex structure after recognition of the analyte. This could be affected by low sensitivity, so strategies to stabilise the target-recognition sequence are required.<sup>140</sup>

Triplexes can be exploited to detect a specific duplex sequence. The duplex assembly is recognised by a suitable TFO sequence folded into a hairpin loop and containing a fluorophore/quencher pair in proximity to each other. In the presence of the duplex target sequence, the fluorophore and the quencher are separated by the opening of the hairpin structure, leading to an increased fluorescence of the system. This fluorescence increase depends on the concentration of the duplex analyte. This method was applied to detect cancer cells and also non-DNA targets, like the NF- $\kappa$ B p50 transcription factor.<sup>141,142</sup>

Besides the application of triplexes as molecular beacons, triplexes have also been applied as functional units for electrochemical sensors. Electrodes have been functionalised with programmed, redox-labeled DNA structures to obtain a probe attached to the electrode surface. The concept is based on the fact that, when the analyte is present, the binding between the triplex and the target sequence leads to the formation of a duplex structure. This complex displaces the redox label from the electrode surface, suppressing the electrochemical signal

produced by the probe itself. In this way, a quantitative determination of the analyte (i.e. DNA, proteins, small molecules, metal ions) is obtained by controlling the voltametric response.<sup>143</sup> This method has been applied for the analysis of sequence-specific double-strands, adenosine, and transcription factors and to detect HIV-1 strains.<sup>144,145</sup>

Similarly, triplexes have been used also as pH probes, exploiting the ability of the oligonucleotides to change the duplex/triplex ratio depending on pH. At around pH 5.0, it was shown that cytosine bases are protonated, permitting the formation of a parallel triplex structure. This concept has been applied in the development of a construct formed by a long strand with two arms capable of bridging a fluorophore/quencher-functionalised strand via the formation of the C-G duplex. In neutral conditions, the fluorophore and the quencher are separated in the medium used. In acidic conditions, the protonation of the cytosines promotes the formation of a triplex structure, causing the proximity of the fluorophore/quencher pair and leading to the decrease of fluorescence intensity.<sup>146</sup>

Another application of the pH dependence of the duplex/triplex structure is the control of aggregation/disaggregation of nanostructures driven by the equilibrium between triplex formation and dissociation. In one example, this equilibrium was used to switch the aggregation/disaggregation of gold nanoparticles (NPs), in a reversible process. The nanoparticles were functionalised with nucleic acids that were partially self-complementary. In neutral conditions, the NPs are separated while in acidic conditions (pH 5.0) the formation of a triplex C<sup>+</sup>-G:C structure leads to NP aggregation. When the system is neutralised, the triplex structures were dissociated, and the nanoparticles disaggregated.<sup>147</sup>

In the biomedical field, the trigger release of loads is an important objective that has aroused interest. Stimuli-responsive microcapsules loaded with a substrate and stabilised by DNA shells have been used to specifically release a cargo. Elegantly, the microcapsules are released after enzymatic digestion of the DNA shells. In this context, triplexes have been attached to the microcapsules and used as pH-responsive carriers. For example, QD-loaded CaCO<sub>3</sub> microparticles, coated with poly(allylamine hydrochloride) (PAH) polyelectrolyte and functionalised with nucleic acid composites containing the caged triplex sequences, were used. The DNA-stabilised CaCO<sub>3</sub> core was dissolved by adding EDTA. At pH 5.0), the triplex structure is formed, with a subsequent separation of the microcapsules and the release of the QD loads.<sup>148</sup>

Overall, all these findings represent very intriguing and promising steps in the application of TFOs in the biomedical and DNA nanotechnology field.

## 1.6 DNA triplex and related interactions with metal complexes

Transition metal complexes have been investigated in the last decades for a large range of healthcare applications, including diagnosis and treatment of various diseases. Several characteristics are appealing for study with nucleic acids, such as the positive charge, the ability to coordinate directly to Lewis base sites on DNA, the possibility to undergo redox reactions with DNA and to generate reactive oxygen species - an attribute particularly relevant for photodynamic therapy (PDT) - make these systems exceptionally attractive for the development of new therapeutics.<sup>146</sup>

Since the serendipitous discovery of cisplatin and its ability to covalently bind the duplex DNA,<sup>149</sup> many metal complexes have been studied to obtain compounds with less side effects than cisplatin and an improved and more selective toxicity towards cancer cells. In parallel, other approaches to the use of metal complexes for targeting DNA in different ways have been developed.<sup>150–152</sup> Very interestingly, metal compounds can also be exploited with non-canonical DNA structures, to stabilise these structures and/or to functionalize them for a specific application, as presented in this section.

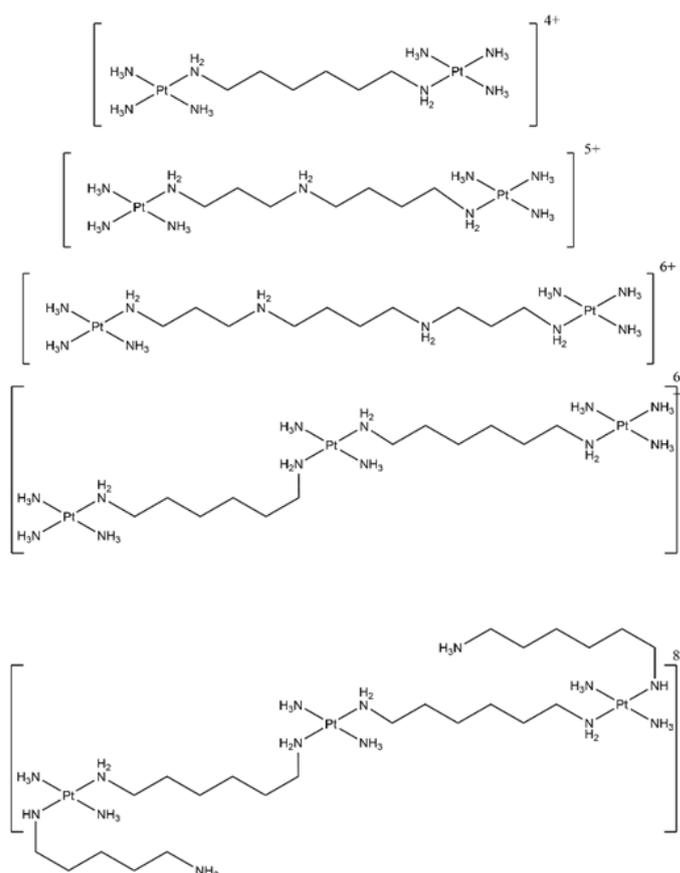
Early attempts were made to introduce Ag(I)-based complexes as artificial nucleosides to stabilize DNA triplexes through metal complexation. The incorporated Ag(I) complex significantly stabilized the DNA duplex and triplex by introducing a pair of pyridine nucleobases in the middle of the sequence. The nitrogen of the pyridyl complex coordinates with Ag(I) at the centre of the triplex, stabilizing the triplex structure.<sup>153</sup> Although it is not an independent molecule that intercalates in the DNA triplex, it is noteworthy that OsO<sub>4</sub><sup>-</sup> bipyridine stabilizes the triplexes by protecting the thymine from being disrupted. In the study, it was observed that intercalation caused a thymine base to flip out of the DNA helix. When the complex was added, the thymine was protected from this disruption.<sup>116,154</sup>

The ability to specifically recognise a non-conventional DNA structure is a very powerful tool to increase specificity in targeting biomolecular sites. For example, tetracationic supramolecular helicates such as [Fe<sub>2</sub>L<sub>3</sub>]<sup>4+</sup>, formed from Fe<sup>2+</sup> ions wrapped by three bis-pyridylimine organic strands, were used in a new approach for synthetic DNA recognition.

Intriguingly, one of the compounds ( $L=C_{25}H_{20}N_4$ ) recognised a three-way junction in duplex DNA, giving a unique hydrophobic binding site characterised by a triangular shape. The structure was determined by X-ray crystallography. This result gave information on the existence of DNA binding modes of metal-based drugs that differ from the most common ones (i.e., covalent bond, intercalation, major groove binding, minor groove binding and sugar-phosphate backbone binding).<sup>155</sup>

Bulges are sites of DNA where one or more nucleotides are not paired within the double-helix. These unpaired nucleotides arise after replication and recombination errors or after carcinogen-induced DNA damage. They are believed to play an important role in various diseases such as cancer, Alzheimer and muscular dystrophy. Thus, DNA sequences containing a bulge are an important target for developing potential therapeutic drugs. Also, small molecules able to target DNA bulges are particularly interesting for their use as potential therapies. The interaction of the above-cited compound  $[Fe_2(C_{25}H_{20}N_4)_3]^{4+}$  with bulged DNA was studied by DNA melting temperature and gel electrophoresis assays to evaluate the binding affinity of this helicate for various DNA bulges. Both enantiomers of the compound bind to bulges containing two or more unpaired nucleotides. Moreover, this compound had higher binding affinity for bulges containing unpaired pyrimidines and/or flanking pyrimidines. It is suggested that the bulge allows the triangular prismatic motif necessary to accommodate the helicate. This is an example of another uncommon DNA structure that is specifically recognised by  $[Fe_2L_3]^{4+}$  supramolecular helicates.<sup>156</sup> Brabec and co-workers described a class of dinuclear triplex-forming metallohelices able to specifically recognize and stabilise DNA bulges of different size and composition. The compounds preferably bind the DNA bulges instead of double-strand DNA. Their binding affinity was shown to be dependent on the individual metallohelices, the bulge size and the bases present in the bulge loop. In particular, pyrimidine-containing bulges are preferred compared to the purine-containing ones. These compounds were shown to have the ability to stabilise the bulge-containing sequences. In fact, increased thermal stability was obtained with DNA bulges containing three or more unpaired adenines or two unpaired thymines, indicating a stabilising effect.<sup>157</sup>

A range of antitumor substitution-inert polynuclear platinum complexes (SI-PPCs) have been studied as small molecules able to recognise, bind and stabilise the triplex structure of DNA and RNA (Figure 1.16). This class of compounds bind DNA through noncovalent interactions, in particular by “phosphate clamp”, a mode of DNA-ligand recognition different from the intercalative or minor groove binding. They had the ability to inhibit DNA synthesis by DNA polymerase when the DNA sequences used are prone to form pyrimidine- and purine-motif triplex DNAs.



**Figure 1.16** Chemical structures of the various SI-PPCs. Reproduced from Dalla Pozza, M., Abdullrahman, A., Cardin, C. J., Gasser, G. & Hall, J. P. Three's a crowd - stabilisation, structure, and applications of DNA triplexes. *Chem Sci* 13, 10193–10215 (2022).

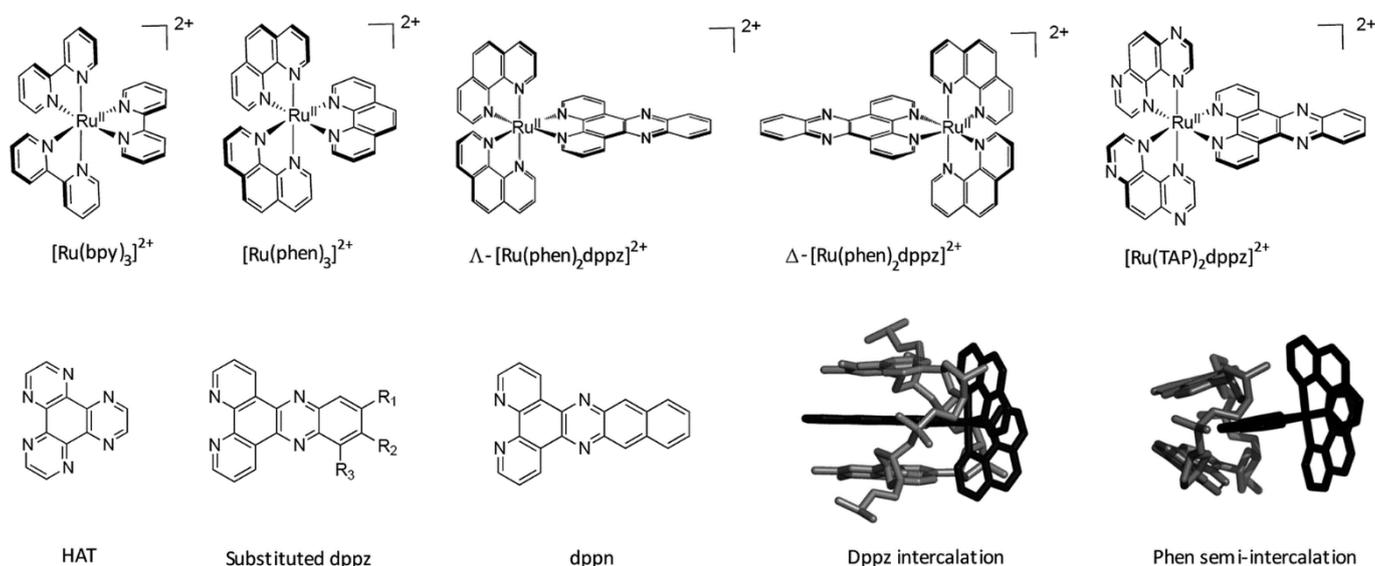
It was suggested that these compounds act as very effective stabilisers of triplex DNA and that they can play a stabilising role in triple-helical DNA. The results from a Taq DNA polymerase assay showed that the pyrimidine-rich template used for the experiment does not permit the primer extension when the SI-PPCs compounds are present. This indicates that the compounds stabilise or form a DNA topology that impedes DNA polymerisation. Interestingly, the formation of the

DNA triple-helix is not stopped in the absence of the compounds and a displacement of TO (which intercalates with high-affinity in triplex structures) takes place when the SI-PPCs are present. This indicates the ability of the Pt-derivatives to form a complex with triple-helical DNA. It was suggested that the ability to stabilise the triplex structure plays a crucial role in the cytotoxicity of this class of compounds. This is particularly important since nucleotide sequences able to form a triplex structure are present in natural DNA, preferentially near regulatory regions.<sup>158</sup>

Moreover, the ability of these class of compounds to inhibit the reverse transcription in RNA template prone to form a triplex structure was described. In particular, the ability of a class of SI-PPCs to inhibit DNA synthesis by reverse transcriptase was evaluated. A purine-rich primer and a pyrimidine-rich RNA template able (TFT) or non-able (SST) to form triplex structures were annealed together and the reverse transcriptase activity was checked by several biophysical techniques. UV melting studies were used to prove that the TFT annealed with the primer formed a triplex structure, showing a biphasic transition in the melting curve, characteristic of a triplex structure. Moreover, the primer extension was allowed in a reverse transcriptase assay without of SI-PPCs, proving that the triple-helix formation does not hamper the reverse transcriptase to extend the primer. On the contrary, in the presence of SI-PPCs, the reverse transcriptase ability to extend the primer annealed with the RNA templates was reduced. This inhibition, related to the presence of the platinum compounds, depends on the charge of the compounds and on their size. Moreover, the inhibiting activity in TFT was higher than in SST, suggesting that SI-PPCs can preferentially recognise, stabilise and inhibit the reverse transcription in RNA template prone to triplex formation rather than in SST. Overall, the ability to bind nucleic acids and inhibit protein-RNA triplex interaction is a very promising extension of the biological activity of this class of compounds.<sup>159</sup>

A wide range of octahedral ruthenium (II) complexes have been investigated for potential biomedical uses, making use of the slow rate of ligand exchange for this electron configuration, multiple and accessible oxidation states, positive charge, and ability to mimic iron in the physiological environment.<sup>160</sup> Ruthenium complexes have been associated with reduced side effects in clinical trials when compared to drugs containing other metals, such as platinum.<sup>161</sup> Ru(II) polypyridyl complexes are notable for their favourable photophysical and photochemical properties, such as visible light absorption (lower energy than 400 nm) due to metal-to-ligand charge transfer (MLCT),<sup>162</sup> and particularly important for the application of such compounds in

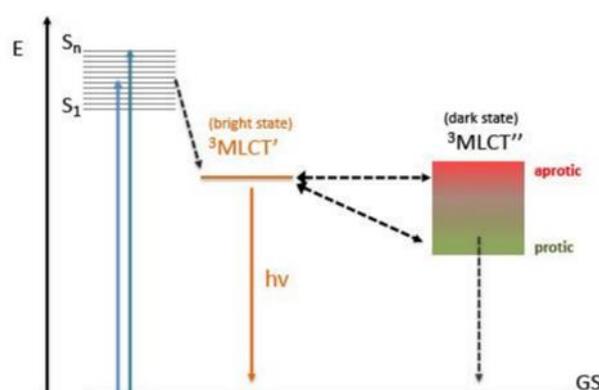
PDT. This medical technique is based on the use of an ideally non-toxic molecule, called photosensitizer (PS), which is activated by light to produce singlet oxygen with a lifetime in metabolically healthy cells of  $\sim 3 \mu\text{s}$  just at the site of irradiation, obtaining therefore a high spatial and temporal selective treatment.<sup>163,164</sup> Indeed, by varying the ligand set, Ru-based complexes can be tailored not only to obtain desired photophysical and photochemical properties in the PDT application window, but also to improve their DNA binding.<sup>165</sup> Different Ru-polypyridyl compounds have been studied for their ability to intercalate in the DNA by  $\pi$ - $\pi$  interaction between the aromatic ligands and DNA  $\pi$ -stack (Figure 1.17). In the next section, are reported some examples, from the very large range already known, of Ru complexes with interesting photophysical and photochemical properties for the application in DNA binding studies, with special attention to actual or potential triplex DNA binding.<sup>166</sup>



**Figure 1.17** Important ruthenium complexes and binding modes. Reproduced from Cardin, C. J., Kelly, J. M. & Quinn, S. J. Photochemically active DNA-intercalating ruthenium and related complexes-insights by combining crystallography and transient spectroscopy. *Chem. Sci.* **8**, 4705–4723 (2017).

A series of Ru(II) complexes with the 1,12-diazaperylene (DAP) ligand of the type  $[\text{Ru}(\text{bpy})_2(\text{DAP})]^{2+}$ ,  $[\text{Ru}(\text{bpy})(\text{DAP})_2]^{2+}$ ,  $[\text{Ru}(\text{bpy})(\text{DAP})_3]^{2+}$  (bpy= 2,2'-bipyridine) was shown to intercalate into calf thymus DNA. The DAP ligand is characterized by an extended  $\pi$ -system and a large surface area to improve the DNA intercalation. Because of the lack of water solubility of the bis- and tris-DAP species, thermal denaturation experiments were performed only with the more water-soluble compound  $[\text{Ru}(\text{bpy})_2(\text{DAP})]^{2+}$ , showing that this compound can stabilise calf thymus DNA with an efficiency comparable to that of ethidium bromide. Moreover, photocleavage of pUC18 supercoiled plasmid was observed in the presence of  $[\text{Ru}(\text{bpy})_2\text{DAP}]^{2+}$  after irradiation

with  $\lambda > 395$  nm for 30 min. The absence of photocleavage in a deoxygenated water environment demonstrated that the  $^1\text{O}_2$  species is involved in the photoreactivity with DNA.<sup>165</sup> As long ago as 1990 the “light-switch” effect of the compound  $[\text{Ru}(\text{bpy})_2(\text{dppz})]^{2+}$  (dppz=dipyrido[3,2-a:2',3'-c]phenazine) was demonstrated by Barton and co-workers, describing this compound as a highly sensitive spectroscopic reporter of double-helical DNA. They demonstrated that this compound displays luminescence only when intercalated into the duplex structure via the planar aromatic ligand dppz. It was shown that after intercalation between DNA base pairs the compound displays an intense luminescence activity, quenched in aqueous solution.<sup>167,168</sup> An accepted explanation is that  $[\text{Ru}(\text{bpy})_2(\text{dppz})]^{2+}$  has a non-emissive (dark) MLCT low-lying excited state involving the phenazine moiety of the dppz ligand, and another emissive (bright) MLCT state related to the bpy part of the dppz ligand. In aqueous solution the dark state is favoured being at lower energy compared to the bright state. On the contrary, when intercalated into DNA, the dark state gets closer in energy to the bright state, allowing thermal population and increasing the emission (Figure 1.18). The DNA duplex in which the  $[\text{Ru}(\text{bpy})_2(\text{dppz})]^{2+}$  is intercalated prevents the quenching effect of the aqueous solution, resulting in a luminescence effect. Further investigation has demonstrated that after binding the DNA via intercalation, also the  $[\text{Ru}(\text{bpy})_2(\text{dppz})]^{2+}$  compound can trigger the photocleavage of pUC18 plasmid DNA in presence of  $\text{O}_2$  ( $\lambda_{\text{irr}} > 455$  nm, 15 min).<sup>169</sup> In 1992, the light-switch effect of both  $[\text{Ru}(\text{bpy})_2(\text{dppz})]^{2+}$  and  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  was reported as a function of the nucleic acid sequence and conformation.



**Figure 1.18** Jablonski diagram indicating the electronic transition from the excited to the ground state, depending on the solvent. Reproduced from Di Pietro, M. L., La Ganga, G. La, Nastasi, F. & Puntoriero, F. *Ru(II)-Dppz Derivatives and Their Interactions with DNA: Thirty Years and Counting. Appl. Sci.* **11**, (2021) *Sci.* **8**, 4705–4723 (2017). with permission from the Royal Society of Chemistry

Indeed, the strongest luminescence effect was observed when the greatest amount of overlap between the nucleic acid structure and the complex was involved, such as when one of these complexes intercalates into triple-helices. In fact, an increased luminescence was observed when

the two compounds were bound to the triple-helical assembly, permitting the dppz ligand to be better shielded from water by the extended surface area of the triplex. Subsequently, a detailed analysis by Choi et al. using separated  $\Lambda$  and  $\Delta$  enantiomers showed that both compounds can bind to a poly(dT-dA:dT) triplex, displaying an increased luminescence compared to the duplex, assumed to be due to the larger surface area of the triplex that better protects the intercalating ligand dppz from water. This better protection and higher luminescence give a useful diagnostic of triplex formation. At the time of these solution studies, there was no clear structural evidence for any binding mode of these complexes to nucleic acids. Despite the third strand, access for intercalation is possible via the major groove, as has been proposed.<sup>170</sup> Detailed studies with separate enantiomers have elucidated by linear and circular dichroism that the Ru complexes with dppz and dppn (dppn= benzodipyrido[3,2-a:2',3'-c]phenazine) as ligands are able to intercalate between the nucleobases of a T-A:T triplex in the minor groove. These authors made a detailed study of the bound chromophore orientation, and concluded that, especially for the  $\Lambda$  complexes, the triplex binding mode had a close resemblance to that seen with duplexes. Very interestingly, the stabilisation of the third strand is related to the nature of the third phenanthroline, showing a stabilizing effect that increases in the order phen < dppn < dppz (phen=1,10-phenanthroline). Intriguingly, the stabilising effect is not related to the size of the ligand.<sup>171</sup> At the time of that publication, no structural data on duplex binding by these compounds was available. The later demonstration that the dppz chromophore intercalated exclusively from the minor groove implies that this would also be true with triplexes.<sup>172</sup> Thereafter, numerous studies have confirmed these interesting features, demonstrating the possible value of this class of compounds as photoluminescent probe for bioanalysis and application in PDT.<sup>162</sup>

Ru(II) complexes linked to triplex forming oligonucleotides could be used as photosensitizers in site-specific damaged DNA, as demonstrated by H el ene and co-workers. In fact, the complex  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  attached to the oligonucleotide and intercalated in the DNA formed a stable triplex. Different behaviours were observed between the two enantiomers of the compound, in fact the luminescence of the  $\Delta$  enantiomer linked to HIV-T oligonucleotide increased by 6-10 times, while no enhancement was observed with the  $\Lambda$  enantiomer. The  $\Delta$  enantiomer of the compound  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  linked to the 5'-phosphate group of the oligonucleotide by phenanthroline binds the DNA duplex in a sequence-specific way. The proposed mechanism is the formation of the triplex and the intercalation of the dppz ligand into the DNA molecule, leading to

the stabilisation of the structure and to an enhancement of the fluorescence. Once again, the photophysical properties of ruthenium compounds such as the ability to photocleavage, long-distance electron transfer and luminescence can be exploited for application in antigene-therapy or as photosensitizer for photodamage of the DNA by triple-helix formation.<sup>173</sup> Indeed, the  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  complex is reported to successfully bind the DNA double-helix, so that this property can be exploited to stabilise the triplex by conjugation of the complex to the 5'-end of a TFO. Importantly, the triplex formed by a TFO functionalised with  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  showed an increased stability by thermal denaturation compared to the triplex formed by the same unmodified oligonucleotide, with a  $\Delta T_m=12^\circ\text{C}$ . This indicates that the unmodified oligonucleotide forms less stable triplexes than the nucleotide decorated with the ruthenium complex.<sup>166</sup> The strong aromatic character of the dppz ligand allows for the intercalation both in duplex and triplex DNA, lying parallel to the triplex bases and intercalating into the minor groove of the triplex. Notably, the whole triplex structure is stabilised by the intercalation of the Ru-dppz complex bound to the TFO.<sup>174</sup>

Therefore, Ru polypyridyl derivatives are of great interest to obtain a stabilising effect on triplexes and to selectively cleave DNA by exploiting the high binding specificity of TFO and the photophysical properties of the ruthenium derivatives linked to the TFO. A library of Ru(II) complexes with halogenated dppz ligands was screened against several biological molecules, such as proteins, ssDNA, dsDNA, DNA triplexes and DNA G-quadruplexes to understand the main factors influencing luminescent behaviour. It was proposed that (1) intercalation in the DNA structure of these compounds mainly depends on the changes of the halogenated substituent on the dppz ligand, (2) the luminescence is increased in the presence of DNA structures but not in the presence of hydrophobic non-DNA structures such as BSA (3) the  $\pi$  stacking surface area influences the luminescence. Indeed, after studying a panel of different substituents on the dppz ligand, more luminescence effect was detected with the compound  $[\text{Ru}(\text{bpy})_2\text{dppz-11,12-Br}]^{2+}$  in the A-T:A triplex and in intrastrand G-quadruplexes compared to intercalation into the DNA duplex. The authors suggest that large Br atoms in positions 11 and 12 prevent the complex from fully intercalating in the DNA duplex, causing the phenazine N atoms to be partially exposed to water, resulting in increased luminescence quenching. The luminescence was enhanced by 89x in the presence of DNA triplexes compared to that in buffer alone. Moreover, this compound has also shown a 2.8x higher luminescence when bound to G-quadruplexes compared to DNA

triplexes, confirming that the  $\pi$  stacking surface area plays an important role in increasing the luminescence.<sup>175</sup> More structural studies are required to understand if this effect is due to the structure itself or to the DNA sequences.

DNA triplexes have also been used as part of an Enhanced Chemiluminescence (ECL) biosensor approach to detect the presence of adenosine in serum (Figure 19). The ECL based on  $[\text{Ru}(\text{bpy})_3]^{2+}$  complexes are used to detect a large number of analytes with different percentages of selectivity and sensitivity. Those characteristics change based on different elements that are part of the ruthenium complexes. Nevertheless, the advantage is to work with an approach that completely avoids radioactive labels with the limit of detection that is low and simple to use. To quantify the presence of adenosine in serum, the method is based on an aptamer, attached on the surface of a gold electrode with an ECL signal marker composed of  $[\text{Ru}(\text{bpy})_3]^{2+}$  forming the first DNA strand. The other strand used as a quenching probe binds a ferrocene carboxylic acid (FcA) at the 5' end. A complex is formed with a third strand, complementary to the quencher, and coralyne chloride as binder. This complex is stable until the concentration of the adenosine increases. At this point the first strand assumes a hairpin configuration generating an intense luminescence due to the ruthenium complex and the absence of the FcA activity. This technique based on a DNA triplex has a more sensitive adenosine detection compared to the DNA duplex-based sensor.<sup>176</sup>

## 1.7 Conclusions and Project Aims

DNA triplexes are non-canonical structures that together with other configurations, such as *i*-motif or quadruplexes, are gaining significant attention for the development of innovative therapeutic strategies. Triplexes possess large diversity in terms of stability, distortion, and environmental conditions required for the formation. In order to exploit DNA triplexes for biological uses, numerous ligands have been designed over the years to functionalise these structures and enhance their stability in physiological conditions. Despite the development of multiple metal-based compounds have been developed to interact with DNA triplexes, Ru(II) polypyridyl compounds are of significant interest due to their photophysical, electronic and biological properties.<sup>162</sup> Many interesting and promising results have been obtained. However, investigations that cover the role of ruthenium complexes in DNA triplexes are very limited.

The overall aim of this project is to gain a deeper understanding of the binding interactions of ruthenium polypyridyl complexes with DNA triplex structure. This will aid the development of rational design of metal complexes capable of selectively targeting and stabilising triple-helical DNA structures for gene editing purposes. Specifically, the main aims are as follows:

- **Identify a library of triplex structures stable in solution to enable optimal crystallisation (Chapter 2).**

DNA triplexes are known to exhibit kinetic and thermodynamic instability under certain conditions, requiring optimisation of factors such as pH and cations to maintain the triple-helical configuration, as previously discussed in this chapter. Crystallisation solution involve a variety of buffers, precipitants and salts to support the growth of nucleic acid crystals.<sup>177</sup> Thus the aim is to systematically screen for triplex stability across a variety of different aqueous solutions relevant to crystallisation. Identifying stability conditions will facilitate optimisation efforts for crystallisation of diverse triplex sequences.

- **Investigate potential metal complexes binding sites in duplex adjacent to triplex region (Chapter 3).**

Stability of the underlying duplex for overall increased triple stability has been previously explored.<sup>178</sup> Moreover, recent advances have been involved in the design of metal complex conjugated to triplex-forming oligos to promote selective recognition and

cleavage of the underlying duplex.<sup>179</sup> Since the DNA sequence is the primary factor that determines the intercalation of the complex between base pairs, the aim is to determine which triplex-adjacent site best supports binding of ruthenium complexes. The polypyridyl complex  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  was selected for this study as binds strongly compared to related ruthenium complex intercalators.<sup>180</sup> Additionally, the complex and its enantiomers display characteristic luminescence lifetimes required for interpretation of the results.<sup>181</sup> By correlating duplex extensions with ruthenium complex binding preference will aid in understand the optimal binding sites for TFO-conjugated ruthenium complexes as a function of DNA sequence.

- **Structurally define the binding modes of ruthenium polypyridyl complexes to DNA triplexes (Chapter 4, 5 & 6).**

Ruthenium polypyridyl complexes have been widely studied for various applications as DNA intercalators in nucleic acids.<sup>182</sup> However, structural details on their interaction with DNA triplexes remains limited. Thus, the aim is to provide extensive crystallographic characterisation of metal complex binding with DNA triplex structures. High resolution X-ray data will give atomic insights of the intercalation geometry of ruthenium complexes with parallel and antiparallel triplex DNA, which will enable the rational design of compounds to induce site-specific gene modification.

## 1.8 References

1. Kaushik, M. *et al.* A bouquet of DNA structures: Emerging diversity. *Biochem Biophys Rep* **5**, 388–395 (2016).
2. Tateishi-Karimata, H. & Sugimoto, N. Chemical biology of non-canonical structures of nucleic acids for therapeutic applications. *Chemical Communications* **56**, 2379–2390 (2020).
3. Alberts, B. *et al.* The Structure and Function of DNA. in *Molecular Biology of the Cell* (Garland Science, 2002).
4. Wing, R. *et al.* Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758 (1980).
5. Voet, D. & Voet, J. G. *Biochemistry*. vol. 1 (Pedersen, New York, 1995).
6. Ussery, D. W. DNA Structure: A-, B- and Z-DNA Helix Families. *Encyclopedia of life sciences* 1–7 (2002).
7. Takahashi, S. & Sugimoto, N. Stability prediction of canonical and non-canonical structures of nucleic acids in various molecular environments and cells. *Chem Soc Rev* **49**, 8439–8468 (2020).
8. Saini, N., Zhang, Y., Usdin, K. & Lobachev, K. S. When secondary comes first-The importance of non-canonical DNA structures. *Biochimie* **95**, 117–123 (2013).
9. Pauling, L. & Corey, R. B. A Proposed Structure for the nucleic acids. *Proceedings of the National Academy of Sciences* **2**, 15–60 (1953).
10. Felsenfeld, G., Davies, D. R. & Rich, A. Formation of a Three-Stranded Polynucleotide Molecule. *J Am Chem Soc* **79**, 2023–2024 (1957).
11. Frank-Kamenetskii, M. D. & Mirkin, S. M. Triplex DNA structures. *Annual Review of Biochemistry* **64**, 65–95 (1995).
12. Spiegel, J., Adhikari, S. & Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem* **2**, 123–136 (2020).
13. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* **43**, 8627–8637 (2015).
14. Brazier, J. A., Shah, A. & Brown, G. D. I-Motif formation in gene promoters: Unusually stable formation in sequences complementary to known G-quadruplexes. *Chemical Communications* **48**, 10739–10741 (2012).
15. Inagaki, H. & Kurahashi, H. Cruciform DNA. in *Brenner's Encyclopedia of Genetics* vol. 2 241–243 (2013).
16. Mandal, A. Replacement of Pt-based anticancer drugs by Ru-analogues - a mini-review. *IJRES* **8**, 17–20 (2020).

17. Allardyce, C. S. & Dyson, P. J. Ruthenium in medicine: Current clinical uses and future prospects. *Platin Met Rev* **45**, 62–69 (2001).
18. Hachey, A. C., Havrylyuk, D. & Glazer, E. C. Biological activities of polypyridyl-type ligands: implications for bioinorganic chemistry and light-activated metal complexes. *Curr Opin Chem Biol* **61**, 191–202 (2021).
19. Notaro, A. & Gasser, G. Monomeric and dimeric coordinatively saturated and substitutionally inert Ru(II) polypyridyl complexes as anticancer drug candidates. *Chem Soc Rev* **46**, 7317–7337 (2017).
20. Monro, S. *et al.* Transition Metal Complexes and Photodynamic Therapy from a Tumor-Centered Approach: Challenges, Opportunities, and Highlights from the Development of TLD1433. *Chem. Rev.* **119**, 797–828 (2019).
21. Petraccone, L. *et al.* Linkage of proton binding to the thermal dissociation of triple helix complex. *Biophys Chem* **110**, 73–81 (2004).
22. Volker, J. & Klump, H. H. Electrostatic Effects in DNA Triple Helices. *Biochemistry* **33**, 13502–13508 (1994).
23. Boehm, B. J., Whidborne, C., Button, A. L., Pukala, T. L. & Huang, D. M. DNA triplex structure, thermodynamics, and destabilisation: insight from molecular simulations. *Physical Chemistry Chemical Physics* **20**, 14013–14023 (2018).
24. Mills, M. *et al.* Chemical modification of the third strand: Differential effects on purine and pyrimidine triple helix formation. *Biochemistry* **41**, 357–366 (2002).
25. Thenmalarchelvi, R. & Yathindra, N. New insights into DNA triplexes: residual twist and radial difference as measures of base triplet non-isomorphism and their implication to sequence-dependent non-uniform DNA triplex. *Nucleic Acid Research* **33**, 43–55 (2005).
26. Kohwi, Y. & Panchenko, Y. Transcription-dependent recombination induced by triple-helix formation. *Genes Dev* **7**, 1766–1778 (1993).
27. Li, W., Hou, X.-M., Wang, P.-Y., Xi, X.-G. & Li, M. Direct Measurement of Sequential Folding Pathway and Energy Landscape of Human Telomeric G-quadruplex Structures. *J Am Chem Soc* **135**, 6423–6426 (2013).
28. Hou, X.-M. *et al.* Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res* **45**, 11401–11412 (2017).
29. Zhurkin, V. B., Raghunathan, G., Ulyanov, N. B., Camerini-Otero, R. D. & Jernigan, R. L. A parallel DNA triplex as a model for the intermediate in homologous recombination. *J Mol Biol* **239**, 181–200 (1994).
30. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science (1979)* **254**, 1497–1500 (1991).

31. Gupta, A., Mishra, A. & Puri, N. Peptide nucleic acids: Advanced tools for biomedical applications. *J Biotechnol* **259**, 148–159 (2017).
32. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).
33. Radhakrishnan, I. & Patel, D. J. Solution structure of a purine·purine·pyrimidine DNA triplex containing G·GC and T·AT triples. *Structure* **1**, 135–152 (1993).
34. Radhakrishnan, I. & Patel, D. J. Solution structure and hydration patterns of a Pyrimidine·Purine·Pyrimidine DNA triplex containing a novel T·CG base-triple. *J Mol Biol* **241**, 600–619 (1994).
35. Van Meervelt, L. *et al.* High-resolution structure of a DNA helix forming (C·G)\*G base triplets. *Nature* **374**, 742–744 (1995).
36. Bartley, J. P., Brown, T. & Lane, A. N. Solution Conformation of an Intramolecular DNA Triplex Containing a Nonnucleotide Linker: Comparison with the DNA Duplex †. *Biochemistry* **36**, 14502–14511 (1997).
37. Tarköy, M., Phipps, A. K., Schultze, P. & Feigon, J. Solution structure of an intramolecular DNA triplex linked by hexakis(ethylene glycol) units: d(AGAGAGAA-(EG)<sub>6</sub>-TTCTCTCT-(EG)<sub>6</sub>-TCTCTCTT). *Biochemistry* **37**, 5810–5819 (1998).
38. Asensio, J. L., Lane, A. N. & Brown, T. Comparison of the solution structures of intramolecular DNA triple helices containing adjacent and non-adjacent CG·C<sup>+</sup> triplets. *Nucleic Acids Res* **26**, 3677–3686 (1998).
39. Rhee, S., Han, Z., Liu, K., Miles, H. T. & Davies, D. R. Structure of a Triple Helical DNA with a Triplex–Duplex Junction. *Biochemistry* **38**, 16810–16815 (1999).
40. Asensio, J. L., Brown, T. & Lane, A. N. Solution conformation of a parallel DNA triple helix with 5' and 3' triplex-duplex junctions. *Structure* **7**, 1–11 (1999).
41. Van Dongen, M. J. P. *et al.* Structure and mechanism of formation of the H-y5 isomer of an intramolecular DNA triple helix. *Nat Struct Biol* **6**, 854–859 (1999).
42. Cerofolini, L. *et al.* G-triplex structure and formation propensity. *Nucleic Acids Res* **42**, 13393–13404 (2014).
43. Betts, L., Josey, J. A., Veal, J. M. & Jordan, S. R. A Nucleic Acid Triple Helix Formed by a Peptide Nucleic Acid-DNA Complex. *Science (1979)* **270**, 1838 (1995).
44. Wang, E., Koshlap, K. M., Gillespie, P., Dervan, P. B. & Feigon, J. Solution structure of a pyrimidine·purine·pyrimidine triplex containing the sequence-specific intercalating non-natural base D3. *J Mol Biol* **257**, 1052–1069 (1996).
45. Koshlap, K. M., Schultze, P., Brunar, H., Dervan, P. B. & Feigon, J. Solution structure of an intramolecular DNA triplex containing an N7- glycosylated guanine which mimics a protonated cytosine. *Biochemistry* **36**, 2659–2668 (1997).

46. Phipps, A. K., Tarköy, M., Schultze, P. & Feigon, J. Solution Structure of an Intramolecular DNA Triplex Containing 5-(1-Propynyl)-2'-deoxyuridine Residues in the Third Strand ,. *Biochemistry* **37**, 5820–5830 (1998).
47. Sørensen, J. J., Nielsen, J. T. & Petersen, M. Solution structure of a dsDNA:LNA triplex. *Nucleic Acids Res* **32**, 6078–6085 (2004).
48. Petersson, B. *et al.* Crystal Structure of a Partly Self-Complementary Peptide Nucleic Acid (PNA) Oligomer Showing a Duplex-Triplex Network †. *J. Am. Chem. Soc.* **127**, 1424–1430 (2005).
49. Wang, A. H.-J., Hakoshima, T., van der Marel, G., van Boom, J. H. & Rich, A. AT base pairs are less stable than GC base pairs in Z-DNA: The crystal structure of d(m5CGTAm5CG). *Cell* **37**, 321–331 (1984).
50. Baker, E. S. & Bowers, M. T. B-DNA Helix Stability in a Solvent-Free Environment. *J Am Soc Mass Spectrom* **18**, 1188–1195 (2007).
51. Neidle, S. *DNA Structure and Recognition*. (1994).
52. Rusling, D. A., Brown, T. & Fox, K. R. DNA triple-helix formation at target sites containing duplex mismatches. *Biophys Chem* **123**, 134–140 (2006).
53. Esguerra, M., Nilsson, L. & Villa, A. Triple helical DNA in a duplex context and base pair opening. *Nucleic Acids Res* **42**, 11329–11338 (2014).
54. Dickerson, R. E. & Ng, H.-L. DNA structure from A to B. *Proceedings of the National Academy of Sciences* **98**, 6986–6988 (2001).
55. Lu, X. J. & Olson, W. K. 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* **3**, 1213–1227 (2008).
56. Robles, J. *et al.* Nucleic Acid Triple Helices: Stability Effects of Nucleobase Modifications. *Curr Org Chem* **6**, 1333–1368 (2002).
57. Rusling, D. R., Brown, T. & Fox, K. R. DNA Recognition by Triple Helix Formation. in *Sequence-Specific DNA Binding Agents* (ed. Waring, M.) 1–28 (The Royal Society of Chemistry, 2006). doi:10.1039/9781847555304-00001.
58. Malkov, A. V., Voloshin, N. O., Soyfer, N. V. & Frank-Kamenetskii, D. M. Cation and sequence effects on stability of intermolecular pyrimidine-purine-purine triplex. *Nucleic Acids Res* **21**, 585–591 (1993).
59. Svinarchuk, F., Paoletti, J. & Malvy, C. An Unusually Stable Purine(Purine-Pyrimidine) Short Triplex. *Journal of Biological Chemistry* **270**, 14068–14071 (1995).
60. Kaushik, S. *et al.* Presence of divalent cation is not mandatory for the formation of intramolecular purine-motif triplex containing human c-jun protooncogene target. *Biochemistry* **50**, 4132–4142 (2011).
61. Sugimoto, N., Wu, P., Hara, H. & Kawamoto, Y. pH and Cation Effects on the Properties of Parallel Pyrimidine Motif DNA Triplexes. *Biochemistry* **40**, 9396–9405 (2001).

62. Jiang, H. X. *et al.* Divalent cations and molecular crowding buffers stabilize G-triplex at physiologically relevant temperatures. *Sci Rep* **5**, 1–11 (2015).
63. Teng, Y., Tateishi-Karimata, H., Ohyama, T. & Sugimoto, N. Effect of potassium concentration on triplex stability under molecular crowding conditions. *Molecules* **25**, 1–13 (2020).
64. Chen, G. & Chen, S.-J. Quantitative analysis of the ion-dependent folding stability of DNA triplexes. *Phys Biol* **8**, 066006 (2011).
65. Goobes, R., Cohen, O. & Minsky, A. Unique condensation patterns of triplex DNA: physical aspects and physiological implications. *Nucleic Acids Res* **30**, 2154–2161 (2002).
66. Ohkubo, A. *et al.* Synthesis and triplex-forming properties of oligonucleotides capable of recognizing corresponding DNA duplexes containing four base pairs. *Nucleic Acids Res* **43**, 5675–5686 (2015).
67. Carr, C. E., Ganugula, R., Shikiya, R., Soto, A. M. & Marky, L. A. Effect of dC → d(m5C) substitutions on the folding of intramolecular triplexes with mixed TAT and C+GC base triplets. *Biochimie* **146**, 156–165 (2018).
68. Ono, A., Ts'o, P. O. P. & Kan, L. sing. Triplex Formation of an Oligonucleotide containing 2'-O-Methylpseudoisocytidine with a DNA Duplex at Neutral pH. *Journal of Organic Chemistry* **57**, 3225–3230 (1992).
69. Jones, E. L., Mlotkowski, A. J., Hebert, S. P., Schlegel, H. B. & Chow, C. S. Calculations of pK<sub>a</sub> Values for a Series of Naturally Occurring Modified Nucleobases. *J Phys Chem A* **126**, 1518–1529 (2022).
70. Chin, T. M. *et al.* 'Paper-clip' type triple helix formation by 5'-d-(TC)3T(a)(CT)3C(b)(AG)3 (a and b = 0-4) as a function of loop size with and without the pseudoisocytosine base in the Hoogsteen strand. *Biochemistry* **39**, 12457–12464 (2000).
71. Fox, K. R., Brown, T. & Rusling, D. A. Chapter 1. DNA Recognition by Parallel Triplex Formation. in *DNA-targeting Molecules as Therapeutic Agents* 1–32 (2018). doi:10.1039/9781788012928-00001.
72. Bates, P. J. *et al.* Efficient triple helix formation by oligodeoxyribonucleotides containing α- or β-2-amino-5-(2-deoxy-D-ribofuranosyl) pyridine residues. *Nucleic Acids Res* **24**, 4176–4184 (1996).
73. Dong Li Chen & McLaughlin, L. W. Use of pK(a) differences to enhance the formation of base triplets involving C-G and G-C base pairs. *Journal of Organic Chemistry* **65**, 7468–7474 (2000).
74. Lou, C., Shelbourne, M., Fox, K. R. & Brown, T. 2'-aminoethoxy-2-amino-3-methylpyridine in triplex-forming oligonucleotides: High affinity, selectivity and resistance to enzymatic degradation. *Chemistry - A European Journal* **17**, 14851–14856 (2011).

75. Huang, C. Y., Cushman, C. D. & Miller, P. S. Triplex Formation by an Oligonucleotide Containing N4-(3-Acetamidopropyl)cytosine. *Journal of Organic Chemistry* **58**, 5048–5049 (1993).
76. Huang, C. Y., Bi, G. & Miller, P. S. Triplex formation by oligonucleotides containing novel deoxycytidine derivatives. *Nucleic Acids Res* **24**, 2606–2613 (1996).
77. Povsic, T. J. & Dervan, P. B. Triple helix formation by oligonucleotides on DNA extended to the physiological pH range. *J Am Chem Soc* **111**, 3059–3061 (1989).
78. Okamoto, I., Seio, K. & Sekine, M. Triplex forming ability of oligonucleotides containing 2'-O-methyl-2-thiouridine or 2-thiothymidine. *Bioorg Med Chem Lett* **16**, 3334–3336 (2006).
79. Rusling, D. A. Triplex-forming properties and enzymatic incorporation of a base-modified nucleotide capable of duplex DNA recognition at neutral pH. *Nucleic Acids Res* **49**, 7256–7266 (2021).
80. Koh, J. S. & Dervan, P. B. Design of a Nonnatural Deoxyribonucleoside for Recognition of GC Base Pairs by Oligonucleotide-Directed Triple Helix Formation. *J Am Chem Soc* **114**, 1470–1478 (1992).
81. Radhakrishnan, I., Patel, D. J., Priestley, E. S., Nash, H. M. & Dervan, P. B. NMR structural studies on a nonnatural deoxyribonucleoside which mediates recognition of GC base pairs in pyrimidine.cntdot.purine.cntdot.pyrimidine DNA triplexes. *Biochemistry* **32**, 11228–11234 (1993).
82. Marfurt, J., Parel, S. P. & Leumann, C. J. Strong, specific, monodentate G-C base pair recognition by N7-inosine derivatives in the pyrimidine·purine·pyrimidine triple-helical binding motif. *Nucleic Acids Res* **25**, 1875–1882 (1997).
83. Milligan, J. F., Krawczyk, S. H., Wadwani, S. & Matteucci, M. D. An anti-parallel triple helix motif with oligodeoxynucleotidescontaining 2'-deoxyguanosine and 7-deaza-2'-deoxy xanthosine. *Nucleic Acids Res* **21**, 327–333 (1993).
84. Olivas, W. M. & Maher, L. J. Overcoming potassium-mediated triplex inhibition. *Nucleic Acids Res* **23**, 1936–1941 (1995).
85. Marathias, V. M., Sawicki, M. J. & Bolton, P. H. 6-Thioguanine alters the structure and stability of duplex DNA and inhibits quadruplex DNA formation. *Nucleic Acids Res* **27**, 2860–2867 (1999).
86. Rao, T. S., Lewis, A. F., Durland, R. H. & Revankar, G. R. A total synthesis of 2'-deoxy-9-deazaguanosine (9-deaza-dG) and its incorporation into triple helix forming oligodeoxyribonucleotides with antiparallel motif. *Tetrahedron Lett* **34**, 6709–6712 (1993).
87. Raynaud, F., Asseline, U., Roig, V. & Thuong, N. T. Synthesis and characterization of O6-modified deoxyguanosine-containing oligodeoxyribonucleotides for triple-helix formation. *Tetrahedron* **52**, 2047–2064 (1996).

88. Aubert, Y., Perrouault, L., Hélène, C., Giovannangeli, C. & Asseline, U. Synthesis and properties of triple helix-forming oligodeoxyribonucleotides containing 7-chloro-7-deaza-2'-deoxyguanosine. *Bioorg Med Chem* **9**, 1617–1624 (2001).
89. Rogers, F. A., Lloyd, J. A. & Tiwari, M. K. Improved bioactivity of G-rich triplex-forming oligonucleotides containing modified guanine bases. *Artif DNA PNA XNA* **5**, 1–10 (2014).
90. Taniguchi, Y. *et al.* Stable and selective antiparallel type triplex DNA formation by targeting a GC base pair with the TFO containing one N2-phenyl-2-deoxyguanosine. *Chem Pharm Bull (Tokyo)* **66**, 624–631 (2018).
91. Dagle, J. M. & Weeks, D. L. Positively charged oligonucleotides overcome potassium-mediated inhibition of triplex DNA formation. *Nucleic Acids Res* **24**, 2143–2149 (1996).
92. Miller, P. S., Bhan, P., Cushman, C. D. & Trapane, T. L. Recognition of a Guanine-Cytosine Base Pair by 8-Oxoadenine. *Biochemistry* **31**, 6788–6793 (1992).
93. Aviñó, A., Cubero, E., González, C., Eritja, R. & Orozco, M. Antiparallel Triple Helices. Structural Characteristics and Stabilization by 8-Amino Derivatives. *J Am Chem Soc* **125**, 16127–16138 (2003).
94. Taniguchi, Y. *et al.* Development of novel C-nucleoside analogues for the formation of antiparallel-type triplex DNA with duplex DNA that includes TA and dUA base pairs. *Org Biomol Chem* **18**, 2845–2851 (2020).
95. Nielsen, P. E. PNA technology. *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology* **26**, 233–248 (2004).
96. Vasquez, K. M. & Glazer, P. M. Triplex-forming oligonucleotides: principles and applications. *Q Rev Biophys* **35**, 89–107 (2002).
97. Arya, D. P. & Bruice, T. C. Triple-helix formation of DNA oligomers with methylthiourea-linked nucleosides (DNmt): A kinetic and thermodynamic analysis. *Proceedings of the National Academy of Sciences* **96**, 4384–4389 (1999).
98. Blaskó, A., Minyat, E. E., Dempcy, R. O. & Bruice, T. C. Fidelity of binding of the guanidinium nucleic acid (DNG) d(Tg)<sub>4</sub>-T- azido with short strand DNA oligomers (A5G3A5, GA4G3A4G, G2A3G3A3G2, G2A2G5A2G2). A kinetic and thermodynamic study. *Biochemistry* **36**, 7821–7831 (1997).
99. Escudé, C. & Sun, J. S. DNA major Groove Binders: Triple helix-forming oligonucleotides, triple helix-specific DNA ligands and cleaving agents. *Top Curr Chem* **253**, 109–148 (2005).
100. Sun, B. W. *et al.* FTIR and UV spectroscopy studies of triplex formation between pyrimidine methoxyethylphosphoramidates  $\alpha$ -oligodeoxynucleotides and ds DNA targets. *J Biomol Struct Dyn* **19**, 1073–1081 (2002).
101. Su, Y., Bayarjargal, M., Hale, T. K. & Filichev, V. V. DNA with zwitterionic and negatively charged phosphate modifications: Formation of DNA triplexes, duplexes and cell uptake studies. *Beilstein Journal of Organic Chemistry* **17**, 749–761 (2021).

102. Brown, S. C., Thomson, S. A., Veal, J. M. & Davis, D. G. NMR Solution Structure of a Peptide Nucleic Acid Complexed with RNA. *Science (1979)* **265**, 778–780 (1994).
103. Shields, G. C., Laughton, C. A. & Orozco, M. Molecular dynamics simulation of a PNA·DNA·PNA triple helix in aqueous solution. *J Am Chem Soc* **120**, 5895–5904 (1998).
104. Zhou, P. *et al.* Novel binding and efficient cellular uptake of guanidine-based peptide nucleic acids (GPNA). *J Am Chem Soc* **125**, 6878–6879 (2003).
105. Schütz, R. *et al.* Olefinic peptide nucleic acids (OPAs): New aspects of the molecular recognition of DNA by PNA. *Angewandte Chemie - International Edition* **39**, 1250–1253 (2000).
106. Kuwahara, M., Arimitsu, M. & Sisido, M. Novel peptide nucleic acid that shows high sequence specificity and all- or-none-type hybridization with the complementary DNA [6]. *J Am Chem Soc* **121**, 256–257 (1999).
107. Jayarathna, D. R., Stout, H. D. & Achim, C. Metal Coordination to Ligand-Modified Peptide Nucleic Acid Triplexes. *Inorg Chem* **57**, 6865–6872 (2018).
108. Koppelhus, U. & Nielsen, P. E. Cellular delivery of peptide nucleic acid (PNA). *Adv Drug Deliv Rev* **55**, 267–280 (2003).
109. Pooga, M., Land, T., Bartfai, T. & Langel, Ü. PNA oligomers as tools for specific modulation of gene expression. *Biomol Eng* **17**, 183–192 (2001).
110. Soler-Bistué, A., Zorreguieta, A. & Tolmasky, M. E. Bridged nucleic acids reloaded. *Molecules* **24**, 1–17 (2019).
111. Højland, T. *et al.* LNA (locked nucleic acid) and analogs as triplex-forming oligonucleotides. *Org Biomol Chem* **5**, 2375–2379 (2007).
112. Fox, K. R., Brown, T. & Rusling, D. A. DNA Recognition by Parallel Triplex Formation. in *DNA-targeting Molecules as Therapeutic Agents* 1–32 (The Royal Society of Chemistry, 2018). doi:10.1039/9781788012928-00001.
113. Morita, K. *et al.* Synthesis and properties of 2'-O,4'-C-ethylene-bridged nucleic acids (ENA) as effective antisense oligonucleotides. *Bioorg Med Chem* **11**, 2211–2226 (2003).
114. Shimizu, M., Konishi, A., Shimada, Y., Inoue, H. & Ohtsuka, E. Oligo(2'- O - methyl)ribonucleotides Effective probes for duplex DNA. *FEBS Lett* **302**, 155–158 (1992).
115. Alam, M. R. *et al.* Extensive sugar modification improves triple helix forming oligonucleotide activity in vitro but reduces activity in vivo. *Biochemistry* **46**, 10222–10233 (2007).
116. Shcholkina, A. K. Protein-free parallel triple-stranded DNA complex formation. *Nucleic Acids Res* **29**, 986–995 (2001).
117. Sandstrom, K., Warmlander, S., Graslund, A. & Leijon, M. A-tract DNA Disfavours Triplex Formation. *J. Mol. Biol.* 737–748 (2002) doi:10.1006/jmbi.2001.5249.

118. Wilson, W. D. *et al.* DNA Triple-Helix Specific Intercalators As Antigene Enhancers: Unfused Aromatic Cations. *Biochemistry* **32**, 10614–10621 (1993).
119. Borisova, O. F. *et al.* Stabilization of parallel (Recombinant) triplex with propidium iodide. *J Biomol Struct Dyn* **13**, 15–27 (1995).
120. Filichev, V. V. & Pedersen, E. B. Stable and selective formation of Hoogsteen-type triplexes and duplexes using twisted intercalating nucleic acids (TINA) prepared via postsynthetic sonogashira solid-phase coupling reactions. *J Am Chem Soc* **127**, 14849–14858 (2005).
121. Paramasivam, M. *et al.* Purine twisted-intercalating nucleic acids: A new class of anti-gene molecules resistant to potassium-induced aggregation. *Nucleic Acids Res* **36**, 3494–3507 (2008).
122. Gécí, I., Filichev, V. V. & Pedersen, E. B. Synthesis of twisted intercalating nucleic acids possessing acridine derivatives, thermal stability studies. *Bioconjug Chem* **17**, 950–957 (2006).
123. Arya, D. P. *et al.* Neomycin binding to watson-hoogsteen (W-H) DNA triplex groove: A model. *J Am Chem Soc* **125**, 3733–2744 (2003).
124. Arya, D. P. New approaches toward recognition of nucleic acid triple helices. *Acc Chem Res* **44**, 134–146 (2011).
125. Song, J., Intody, Z., Li, M. & Wilson, J. H. Activation of gene expression by triplex-directed psoralen crosslinks. *Gene* **324**, 183–190 (2004).
126. Ping, Y. H. & Rana, T. M. Mechanism of site-specific psoralen photoadducts formation in triplex DNA directed by psoralen-conjugated oligonucleotides. *Biochemistry* **44**, 2501–2509 (2005).
127. Liu, Y., Nairn, R. S. & Vasquez, K. M. Targeted gene conversion induced by triplex-directed psoralen interstrand crosslinks in mammalian cells. *Nucleic Acids Res* **37**, 6378–6388 (2009).
128. Knauert, M. P. & Glazer, P. M. Triplex forming oligonucleotides: Sequence-specific tools for gene targeting. *Hum Mol Genet* **10**, 2243–2251 (2001).
129. Spencer, C. A. & Groudine, M. Control of c-myc regulation in normal and neoplastic cells. *Adv Cancer Res* **56**, 1–48 (1991).
130. Postel, E. H., Flint, S. J., Kesslert, D. J. & Hogant, M. E. Evidence that a triplex-forming oligodeoxyribonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-mnc mRNA levels. *Proc. Natl. Acad. Sci. USA* **88**, 8227–8231 (1991).
131. Rajeswari, M. R. DNA triplex structures in neurodegenerative disorder, Friedreich's ataxia. *J Biosci* **37**, 519–532 (2012).
132. Ruan, H. & Wang, Y. H. Friedreich's Ataxia GAA·TTC Duplex and GAA·GAA·TTC Triplex Structures Exclude Nucleosome Assembly. *J Mol Biol* **383**, 292–300 (2008).

133. Nakanishi, K., Cavallo, F., Brunet, E. & Jasin, M. Homologous recombination assay for interstrand cross-link repair. in *Methods in molecular biology (Clifton, N.J.)* vol. 745 283–291 (2011).
134. Faruqi, A. F., Datta, H. J., Carroll, D., Seidman, M. M. & Glazer, P. M. Triple-Helix Formation Induces Recombination in Mammalian Cells via a Nucleotide Excision Repair-Dependent Pathway. *Mol Cell Biol* **20**, 990–1000 (2000).
135. Thoma, B. S., Wakasugi, M., Christensen, J., Reddy, M. C. & Vasquez, K. M. Human XPC-hHR23B interacts with XPA-RPA in the recognition of triplex-directed psoralen DNA interstrand crosslinks. *Nucleic Acids Res* **33**, 2993–3001 (2005).
136. Brázdová, M. *et al.* P53 specifically binds triplex DNA in vitro and in cells P53 specifically binds triplex DNA in vitro and in cells. *PLoS One* 1–25 (2016) doi:10.1371/journal.pone.0167439.
137. Du, Y. & Zhou, X. Targeting Non-B-Form DNA in Living Cells. *The Chemical Record* **13**, 371–384 (2013).
138. Hu, Y., Ceconello, A., Idili, A., Ricci, F. & Willner, I. Triplex-DNA-Nanostrukturen: von grundlegenden Eigenschaften zu Anwendungen. *Angewandte Chemie* **129**, 15410–15434 (2017).
139. Conlon, P. *et al.* Pyrene Excimer Signaling Molecular Beacons for Probing Nucleic Acids are utilized in many types of DNA / RNA targeting applications. *J Am Chem Soc* **130**, 336–342 (2008).
140. Huang, J. *et al.* Competition-Mediated Pyrene-Switching Aptasensor : Probing Lysozyme in Human Serum with a Monomer-Excimer Fluorescence Switch. *Anal. Chem* **82**, 10158–10163 (2010).
141. Xi, D., Wang, X., Ai, S. & Zhang, S. Detection of cancer cells using triplex DNA molecular beacons based on expression of enhanced green fluorescent protein (eGFP). *Chemical Communications* **50**, 9547–9549 (2014).
142. Chen, Y., Murayama, K., Kashida, H., Kamiya, Y. & Asanuma, H. A triplex-forming linear probe for sequence-specific detection of duplex DNA with high sensitivity and affinity. *Chemical Communications* **56**, 5358–5361 (2020).
143. Patterson, A. *et al.* Using Triplex-Forming Oligonucleotide Probes for the Reagentless , Electrochemical Detection of Double-Stranded DNA. *Anal. Chem* **82**, 9109–9115 (2010).
144. Online, V. A. *et al.* A signal-on fluorescence biosensor for detection of adenosine triphosphate based on click chemistry. *Anal. Methods* **6**, 3370–3374 (2014).
145. Zhang, J. *et al.* A signal-on electrochemiluminescence aptamer biosensor for the detection of ultratrace thrombin based on junction-probe. *Biosens Bioelectron* **26**, 2645–2650 (2011).
146. Porchetta, A., Idili, A., Valle, A. & Ricci, F. General Strategy to Introduce pH-Induced Allostery in DNA-Based Receptors to Achieve Controlled Release of Ligands. *Nano Lett.* **15**, 4467–4471 (2015).

147. Xiong, C., Wu, C., Zhang, H. & Ling, L. Gold nanoparticles-based colorimetric investigation of triplex formation under weak alkalic pH environment with the aid of Ag<sup>+</sup>. *Spectrochimica Acta Part A* **79**, 956–961 (2011).
148. Liao, W. C., Riutin, M., Parak, W. J. & Willner, I. Programmed pH-Responsive Microcapsules for the Controlled Release of CdSe/ZnS Quantum Dots. *ACS Nano* **10**, 8683–8689 (2016).
149. Rosenberg, B. Cisplatin: Its history and possible mechanism of action. in *Cisplatin* 9–20 (Elsevier, 1980). doi:10.1016/b978-0-12-565050-2.50006-1.
150. Moucheron, C. From cisplatin to photoreactive Ru complexes: Targeting DNA for biomedical applications. *New Journal of Chemistry* **33**, 235–245 (2009).
151. Komor, A. C. & Barton, J. K. The path for metal complexes to a DNA target. *Chemical Communications* **49**, 3617–3630 (2013).
152. Flamme, M., Figazzolo, C., Gasser, G. & Hollenstein, M. Enzymatic construction of metal-mediated nucleic acid base pairs. *Metallomics* **13**, 1–20 (2021).
153. Tanaka, K., Yamada, Y. & Shionoya, M. Formation of silver(I)-mediated DNA duplex and triplex through an alternative base pair of pyridine nucleobases. *J Am Chem Soc* **124**, 8802–8803 (2002).
154. Paleček, E. Probing DNA structure with osmium tetroxide complexes in vitro. *Methods Enzymol* **212**, 139–155 (1992).
155. Oleksi, A. *et al.* Molecular Recognition of a Three-Way DNA Junction by a Metallosupramolecular Helicate. *Angewandte Chemie* **118**, 1249–1253 (2006).
156. Malina, J., Hannon, M. J. & Brabec, V. Recognition of DNA bulges by dinuclear iron(II) metallosupramolecular helicates. *FEBS Journal* **281**, 987–997 (2014).
157. Hrabina, O., Malina, J., Scott, P. & Brabec, V. Cationic FeII Triplex-Forming Metallohelices as DNA Bulge Binders. *Chemistry - A European Journal* **26**, 16554–16562 (2020).
158. Malina, J., Farrell, N. P. & Brabec, V. Anticancer Agents Substitution-Inert Polynuclear Platinum Complexes That Inhibit the Activity of DNA Polymerase in Triplex-Forming Templates *Angewandte. Angewandte Chemie - International Edition* **2006**, 8535–8539 (2018).
159. Malina, J., Farrell, N. P. & Brabec, V. Substitution-Inert Polynuclear Platinum Complexes Inhibit Reverse Transcription Preferentially in RNA Triplex-Forming Templates. *Inorganic Chemistry* **11**, 15135–15143 (2020).
160. Lee, S. Y., Kim, C. Y. & Nam, T.-G. Ruthenium Complexes as Anticancer Agents: A Brief History and Perspectives. *Drug Des Devel Ther* 5375–5392 (2020) doi:10.2147/DDDT.S275007.
161. Zeng, L. *et al.* The development of anticancer ruthenium(II) complexes: From single molecule compounds to nanomaterials. *Chem Soc Rev* **46**, 5771–5804 (2017).
162. Zhang, S., Ding, Y. & Wei, H. Ruthenium Polypyridine Complexes Combined with Oligonucleotides for Bioanalysis: A Review. *Molecules* **19**, 11933–11987 (2014).

163. Mari, C., Pierroz, V., Ferrari, S. & Gasser, G. Combination of Ru(II) complexes and light: New frontiers in cancer therapy. *Chem Sci* **6**, 2660–2686 (2015).
164. O'Connor, A. E., Gallagher, W. M. & Byrne, A. T. Porphyrin and nonporphyrin photosensitizers in oncology: Preclinical and clinical advances in photodynamic therapy. *Photochem Photobiol* **85**, 1053–1074 (2009).
165. Knoll, J. D. & Turro, C. Control and utilization of ruthenium and rhodium metal complex excited states for photoactivated cancer therapy. *Coord Chem Rev* **282–283**, 110–126 (2015).
166. Flamme, M., Clarke, E., Gasser, G. & Hollenstein, M. Applications of ruthenium complexes covalently linked to nucleic acid derivatives. *Molecules* **23**, 28–31 (2018).
167. Friedman, A. E., Chambron, J., Sauvage, J., Turro, N. J. & Barton, J. K. Molecular 'Light Switch' for DNA: Ru(bpy)<sub>2</sub>(dppz)<sup>2+</sup>. *J. Am. Chem. Soc.* **2**, 4960–4962 (1990).
168. Hartshorn, R. M. & Barton, J. K. Novel Dipyridophenazine Complexes of Ruthenium(II): Exploring Luminescent Reporters of DNA. *J Am Chem Soc* **114**, 5919–5925 (1992).
169. Hu, L. *et al.* [Ru(bpy)<sub>2</sub>dppz]<sup>2+</sup> electrochemiluminescence switch and its applications for DNA interaction study and label-free ATP aptasensor. *Anal Chem* **81**, 9807–9811 (2009).
170. Jenkins, Y., Friedman, A. E., Turro, N. J. & Barton, J. K. Characterization of dipyridophenazine complexes of ruthenium(II): The light switch effect as a function of nucleic acid sequence and conformation. *Biochemistry* **31**, 10809–10816 (1992).
171. Choi, S. D. *et al.* Binding mode of [ruthenium(II) (1,10-phenanthroline)<sub>2</sub>L]<sup>2+</sup> with poly(dT\*dA-dT) triplex. Ligand size effect on third-strand stabilization. *Biochemistry* **36**, 214–223 (1997).
172. Cardin, C. J., Kelly, J. M. & Quinn, S. J. Photochemically active DNA-intercalating ruthenium and related complexes-insights by combining crystallography and transient spectroscopy. *Chem Sci* **8**, 4705–4723 (2017).
173. Grimm, G. N., Boutorine, A. S., Lincoln, P., Nordén, B. & Hélène, C. Formation of DNA triple helices by an oligonucleotide conjugated to a fluorescent ruthenium complex. *ChemBioChem* **3**, 324–331 (2002).
174. Li, G., Sun, L., Ji, L. & Chao, H. Ruthenium(II) complexes with dppz: From molecular photoswitch to biological applications. *Dalton Transactions* **45**, 13261–13276 (2016).
175. Wachter, E., Moyá, D. & Glazer, E. C. Combining a Ru(II) "Building Block" and Rapid Screening Approach to Identify DNA Structure-Selective "Light Switch" Compounds. *ACS Comb Sci* **19**, 85–95 (2017).
176. Ye, S., Li, H. & Cao, W. Electrogenated chemiluminescence detection of adenosine based on triplex DNA biosensor. *Biosens Bioelectron* **26**, 2215–2220 (2011).
177. Holbrook, E. L. & Holbrook, S. R. Crystallisation of Nucleic Acids. in *Encyclopedia of Life Sciences* (Wiley, 2010). doi:10.1002/9780470015902.a0002719.pub2.

178. Rusling, D. A., Rachwal, P. A., Brown, T. & Fox, K. R. The stability of triplex DNA is affected by the stability of the underlying duplex. *Biophys Chem* **145**, 105–110 (2009).
179. Zuin Fantoni, N. *et al.* Development of Gene-Targeted Polypyridyl Triplex-Forming Oligonucleotide Hybrids. *ChemBioChem* **21**, 3563–3574 (2020).
180. Ossipov, D., Pradeepkumar, P. I., Holmer, M. & Chattopadhyaya, J. Synthesis of [Ru(phen)2dppz]2+-tethered oligo-DNA and studies on the metallointercalation mode into the DNA duplex. *J Am Chem Soc* **123**, 3551–3562 (2001).
181. Ossipov, D., Pradeepkumar, P. I., Holmer, M. & Chattopadhyaya, J. Synthesis of [Ru(phen)2dppz]2+ -Tethered Oligo-DNA and Studies on the Metallointercalation Mode into the DNA Duplex. *J Am Chem Soc* **123**, 3551–3562 (2001).
182. Gill, M. R. & Thomas, J. A. Ruthenium(ii) polypyridyl complexes and DNA—from structural probes to cellular imaging and therapeutics. *Chem Soc Rev* **41**, 3179 (2012).
183. Tateishi-Karimata, H. & Sugimoto, N. Chemical biology of non-canonical structures of nucleic acids for therapeutic applications. *Chemical Communications* **56**, 2379–2390 (2020).
184. Biver, T. Stabilisation of non-canonical structures of nucleic acids by metal ions and small molecules. *Coord Chem Rev* **257**, 2765–2783 (2013).
185. Dickerson, R.E. International Tables for Crystallography Volume B: Reciprocal Space. in *International Tables for Crystallography, Volume F* (eds. Arnold, E. & Himmel, M. D.) 610 (Wiley, 2012).

## CHAPTER 2- Selection of DNA triplex systems for crystallisation

## 2.1 Introduction

The structural characterisation of DNA triplexes and TFO-ligand interaction is both useful and important in order to design gene editing molecules. Currently, structural characterisation of DNA triplexes reported in the literature is mostly limited to solution structures solved by NMR.<sup>1</sup> Although NMR models provide information about these molecules in solution, the integration with crystallographic data can result in a more comprehensive representation of triple-helical structures. DNA crystallography specifically, can provide atomic structural details including bond lengths and angles, as well as atomic details about interactions with complexes.

The crystallisation process has recently undergone a transformation through an intensive effort from researchers to shift towards a systematic approach.<sup>2</sup> This crystallisation process has radically improved through the implementation of automated methods, which facilitate the systematic screening of a wide range of conditions. Although there is an element of randomness in some experimental outcomes, the use of a systematic screening aims to minimise the chance and optimise the crystal growth by varying the screen parameters to increase the likelihood of obtaining high-quality crystals.<sup>3</sup> A method of DNA crystallisation used as screening is the vapor diffusion method. In this method, the solution of purified DNA in its buffer and cofactors is mixed with the crystallisation agent and allowed to equilibrate against a reservoir of the same reagent but in a higher concentration. A typical crystallisation screen may include buffers, salts, additives, and precipitants. The crystallisation plates are subsequently placed in incubators set to temperatures ranging from 4 °C to 20 °C, ensuring optimal conditions for crystal growth.

Buffers, cofactors, and ionic strength can contribute to the overall stability of DNA triplexes.<sup>4</sup> Identifying and optimizing these parameters before starting crystallisation trials can often be the key to successful crystallisation of the complex. Indeed, the more a complex is thermodynamically stable, forming in solution a specific configuration, the more the probability to have that construct as crystal increases. A number of conditions have been tested to determine the most stable melting condition for a short library of DNA triplexes.

The buffer solution plays a pivotal role in a crystallisation screen to maintain the stability of macromolecules, particularly proteins as these are sensitive to pH variations.<sup>5</sup> Regarding DNA triplexes, the pH of the environment has a significant influence on their stability, particularly those with a parallel configuration. In this configuration, the N3 atom of the cytosine in the third strand

within a C-G:C triad needs to be protonated, forming a positively charged cytosine. The protonated cytosine can then form an additional Hoogsteen hydrogen bond with N7 of the guanine contributing to the overall stabilisation of the structure.<sup>6-8</sup> all cytosine will form Hoogsteen bonds, only half of the cytosine-forming triplets need to be protonated. Proton exchange studies, as well as calorimetric measurements, confirm that the Hoogsteen bond stability is similar to the Watson-Crick pairs in a protonated environment.<sup>9,10</sup> Moreover, similar experiments show the pH-independence of antiparallel triplexes as here the protonation is not necessary for the formation of Hoogsteen bonds.<sup>11</sup> The peculiar pH sensitivity of DNA triplexes was explored to design nanoswitch platforms in the field of DNA nanotechnology.<sup>12</sup> An example is the *in vivo* application of a triplex DNA-based nanoswitch responsive to pH changes to target the delivery of multiple drugs.<sup>13</sup>

To counteract the electrostatic repulsion between the negatively charged DNA strands, besides the protonated environment, cations help by neutralising the negative charges of the phosphate groups in the DNA backbone.<sup>14,15</sup> This is confirmed by studies based on the tightly bound ion model (TBI) that examines the interaction of ion and nucleic acids and by calorimetric studies that investigate the increased ionic strength when triple-helical structures are formed in a cation solution.<sup>16-18</sup>

In the crystallisation process, various salts are introduced significantly effecting the crystallisation of macromolecules.<sup>19</sup> Commonly studied monovalent ions such as Na<sup>+</sup> or K<sup>+</sup> influence the triplex stability.<sup>20</sup> Studies show monovalent ions such as sodium ions stabilise the triplexes by incorporating counterions between strands.<sup>21</sup> In an acidic environment, a high concentration of Na<sup>+</sup> can have the opposite effect because of the large uptake of protons.<sup>22</sup> Conversely, potassium ions destabilise the triplex formation and instead support the formation of the quadruplex structure due to their atomic radius that is large enough to fit in the central cavity of a quadruplex structure.<sup>23</sup> The presence of NaCl is also responsible for the compacting of DNA molecules, particularly in the presence of crowding agents such as polyethylene glycol (PEG).<sup>24</sup> This compaction is facilitated by the water distribution around the phosphate and bases of the DNA, attracting the binding of positively charged ions around the phosphate resulting in the overall compactness of DNA molecules<sup>25,26</sup>, potentially aiding in the crystallisation process.

In triplexes, divalent ions interact with the nucleobase adenine (atom N7) and guanine (atoms N7 and O6) influencing the triplet's stability.<sup>27</sup> Extensively studies showed the divalent  $Mg^{2+}$  ion interacting with a triplex structure<sup>28,29</sup> and other cations influence triplex stability with different intensities, following the order ( $Mg^{2+} > Mn^{2+} > Ca^{2+} > Sr^{2+}$ ).<sup>30</sup> A particular aspect to take into consideration is the role that these divalent ions can have during the crystallisation process.  $Mg^{2+}$  for instance, condenses the triplex molecules through ion binding in two ways: in the grooves connecting the bases and externally at the phosphate level.<sup>31</sup> Furthermore,  $Ca^{2+}$  and  $Sr^{2+}$  are commonly used as crystallisation reagents, therefore their effect on triplex stability in solution was investigated here.

Aside from the cations, polyamines such as spermine can also provide counterions. Spermine molecules are commonly present in crystallisation screens as they support the condensation of DNA structure to form crystals.<sup>32-34</sup> Polyamines, such as putrescine, spermidine, or spermine also have an effect as DNA triplex stabilisers. Spermine binds in the minor groove of the DNA, as the major groove is occupied by the TFO, reducing the negative charge repulsion due to their amino group. Crystallographic studies of spermine-binding DNA molecules demonstrate the formation of hydrogen bonding with nitrogen atoms of the DNA and water, thus compacting the DNA molecules.<sup>35,36</sup> Interestingly, DNA segments are sometimes used as additives to promote the crystal growth.<sup>37</sup>

In a crystallisation screen, precipitants are employed to promote the crystal lattice formation by altering the macromolecule-solvent contact and facilitate its precipitation out of the solution. The widely used precipitant 2-Methyl-2,4-pentanediol (MPD) also serves as a dehydrating agent. While reducing the water activity slows the evaporation and promotes high quality crystal formation, the used of dehydrating agents could cause perturbation of the water molecules around nucleic acids leading oligonucleotides to unfold. Dehydrating agents such as MeOH, EtOH, ethylene glycol and DMSO could provoke the separation of the TFO.<sup>38</sup> Nonetheless, the effect of a high concentration of MPD included in crystallisation screen, has no direct effect on the structure of DNA molecules.<sup>39</sup> MPD was found in the crystallisation drop of previously crystallised non-canonical triplexes.<sup>40,41</sup> Additionally, Polyethylene glycol (PEG) a polymer with varying molecular weights (PEG 200, PEG 400, PEG 1,000, PEG 4,000, PEG 8,000) often used as a precipitant, excludes a comparable number of water molecules from both DNA duplex and triplex. Alt-

though the crowding environment composed of PEG and solutes influences the formation of triple-helical structures, maintaining them more stable than duplex,<sup>42</sup> this can be attributed to the reduced number of water molecules around the triplex molecules. Indeed, compared to the duplex DNA, the presence of a TFO in the major groove, along with cations prevents water molecules from binding to nucleobases.<sup>43</sup> PEG, used as a crowding agent that mimics the cellular environment has been shown to stabilise the triplex molecules and condensate into liquid-crystalline aggregates.<sup>14</sup>

The objective of this study is to identify triplex systems that are stable in solution and may therefore yield suitable crystals for X-ray crystallography analysis. To determine the stability of the triple-helix complexes in solution the UV-Vis spectrometry technique was employed. Normalised first derivatives of the thermal denaturation curves are used to calculate the melting temperature ( $T_m$ ). The lower temperature transition, denoted as the first transition corresponds to the triplex denaturation (TFO dissociation), while the second transition at higher temperatures indicates the denaturation of the duplex. UV-Vis spectroscopy analysis shows that the triplex melting stability in solution differs between systems depending on the strand composition, length and conformation of the triplex. Here, the  $T_m$  serves as the thermodynamic parameter used to establish the temperature at which the oligonucleotides form a triple-helical structure in solution.

## 2.2 Material and Methods

### 2.2.1 Oligonucleotides and buffer solutions

RP-HPLC purified oligonucleotides were purchased from Merck or Eurogentec and were used with no further purification. Other materials, including buffers and salts, were purchased from Sigma-Aldrich.

### 2.2.2 UV thermal denaturation experiments

UV melting experiments were carried out using on Cary 100 UV-VIS spectrophotometer. The temperature was controlled with a Cary temperature controller. Normalised melting curves (between 0 to 1A) were obtained at 260 nm. Sodium cacodylate at pH ranges 4.0-8.0 was used as a buffer along with 100 mM NaCl and 10 mM MgCl<sub>2</sub> unless stated otherwise. Each sample had a total volume of 600 µL and an absorbance below 1 A. Each stand of intermolecular triplexes AA-t1, AA-t2, AA-3 had a concentration of 2 µM and intramolecular DNA strands AA-t4, AA-t5 and AA-t6 had a final concentration of 1 µM. Samples were annealed by heating them to 90 °C for 5 minutes and then allowing them to cool slowly to room temperature. During the acquisition of the melting profile, the samples were heated from 20 °C to 90 °C at a rate of 1 °C/min. All experiments were run in triplicate. First derivatives curves are shown in Figures A2.1-A2.5

## 2.3 Results and Discussion

### 2.3.1 Selection of DNA triplexes

Based on literature analysis, an initial set of DNA triplexes were selected for investigation as model systems (Table 2.1). These triplexes were selected to represent both inter- and intramolecular configurations. The intermolecular triplexes, composed of two strands ( $d_1$  and  $d_2$ ) and a TFO ( $d_{TFO}$ ), have various characteristics such as oligo length, directionality, and the third strand composition, polypurine or polypyrimidine. Additionally, three intramolecular triplexes with a parallel orientation were included.

**Table 2.1** Intermolecular and intramolecular DNA triplexes used in this study

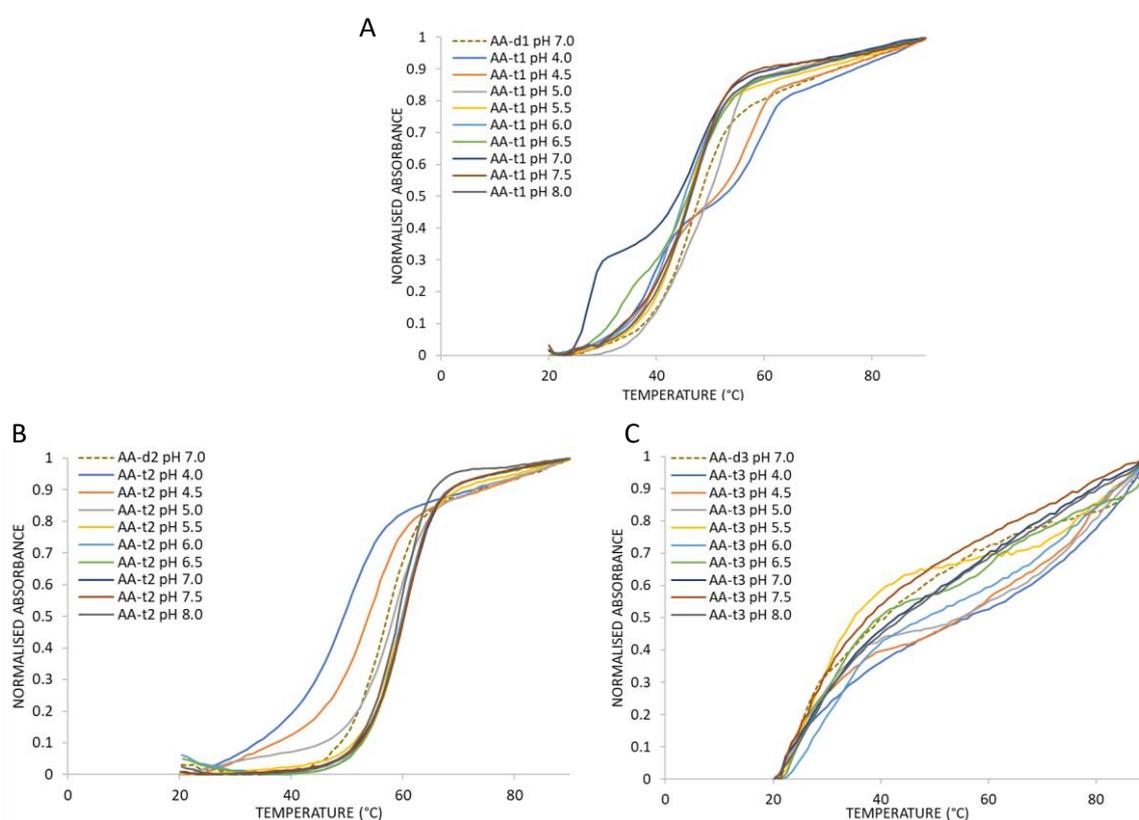
DNA system	DNA triplexes 5'→3'		
	Intermolecular Triplexes		
	$d_1$	$d_2$	$d_{TFO}$
AA-t1	AAGAAAGAAGAGA	TCTCTTCTTTCTT	TTCTTTCTTCTCT
AA-t2	GACTGAGAGACGTA	TACGTCTCTCAGTC	CTCTCT
AA-t3	AGGAGGA	TCCTCCT	TGGTGGT
Intramolecular Triplexes			
AA-t4	AAGAAAGAAGAGATTTTTCTTCTTTCTTTTTTTCTTTCTTCTCT		
AA-t5	TCTTCTCTTTCTTTTTTCTTTCTTCTTTTTAGAAGAGAAAGA		
AA-t6	AGATAGAACCCCTTCTATCTTATATCTTCTT		

AA-t1 is a parallel triplex with 13-mer oligonucleotides forming a triplex with the same length TFO. This triplex has been previously characterised with NMR with the reduced number of 8 nucleotides for each oligonucleotide linked by two ethylene glycol linker to form an intramolecular configuration (PDB ID: **1D3X**).<sup>44</sup> The sequence was also used as part of a DNA Tensegrity Triangle complex, where the TFO was modified with psoralen useful for its role in cross-linking.<sup>45</sup> The distinctive feature of AA-t2 is the short (hexamer) TFO compared to the double-strand DNA of 14-mer oligonucleotide (PDB ID: **1GWR**).<sup>46</sup> Successfully crystallising a structure with a short triplex region may provide structural insights into both the triplex and duplex regions of the molecule. Finally, AA-t3 is the shortest DNA triplex in the library composed of 7-mer bases long oligonucleotides that form an anti-parallel configuration. The NMR structure of AA-t3 was deposited in the PDB database at the time of selection but removed in July 2023 after the completion of this study. The selection of a triplex with short sequence motifs was due to their ability to assemble in crystalline phases.<sup>47</sup>

In addition to the intermolecular DNA triplexes, where the TFO is an external oligo, sequences that by unbinding and folding back to form an intramolecular triplex have been considered. AA-t4 was designed by introducing two loops of four thymine nucleotides that connect the duplex strands  $d_1$  with  $d_2$  and  $d_2$  with the TFO of AA-t1 resulting in the formation of an intramolecular conformation. If the crystallisation process is successful, will enable a comparison between both structures at an atomic level. The size of the loop can vary between 4- to 8-bases with the optimal number being up to 5-bases.<sup>48,49</sup> The loops can be composed of a series of purines, pyrimidines or a combination of both. Loops composed of thymine were chosen to prevent the need for protonation, although it is shown that cytosine-based loops can help the triplex formation by reducing the repulsion upon protonation.<sup>50</sup> Alternatively, loops can be composed of ethylene glycol oligomers, showing an even higher stability effect, although it is expected to be the opposite given the high flexibility of the linker.<sup>51</sup> For the purpose of obtaining the crystal structure of DNA triplexes with natural loops these modifications have not been considered. AA-t5 was previously used to understand better the role of water molecules in crowding conditions.<sup>52</sup> AA-t6 represents a pyrimidine-purine-pyrimidine triplex and its NMR structure has been solved (PDB ID: 1WAN). Moreover, AA-t6 has been characterised containing a non-canonical base.<sup>53</sup> While the modification enhances triplex stability, it is not necessary to the screening scope of this study and consequently was excluded from the triplex design. Further details about AA-t6 are discussed in Chapter 4.

### 2.3.2 The effect of pH on triplex stability

The melting stability for the selected systems was first evaluated with sodium cacodylate buffer at pH 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, and 8.0, keeping constant the concentration of NaCl of 100 mM and 10 mM MgCl<sub>2</sub>. Figures 2.1 illustrate the melting curve collected at 260 nm for the intermolecular triplexes.



**Figure 2.1** Normalised melting profile of AA-t1 (A), AA-t2 (B), AA-t3 (C) and their corresponding duplexes (AA-d1, AA-d2, AA-d3). Samples were prepared with 20 mM sodium cacodylate buffer at ranges of pH from 4.0 to 8.0 with a 0.5 pH unit increment, 100 mM NaCl, and 10 mM MgCl<sub>2</sub>. The absorbance was recorded at  $\lambda = 260$  nm

The characteristic two-transition melting profile is visible exclusively in AA-t1 at pH 4.0, 4.5, 6.5 and 7.0. The first derivative analysis (Figure A2.1) reveals that  $T_m$  at pH 5.0 and 5.5 of the triplex transitions align with those of the duplex transition (Table 2.2). The gradual  $T_m$  decrease suggests the possibility that a triplex transition occurs below 20 °C. Although AA-t2 did not present clear triplex melting curves, and the first derivative did not show any significant values, compared to the duplex melting transition a different melting behaviour was observed. Indeed, by examining the melting curves the end of the triplex transition is evident below pH 5.0, corresponding with the termination of the triplex' hyperchromic effect. Hence, it is plausible to conclude that the

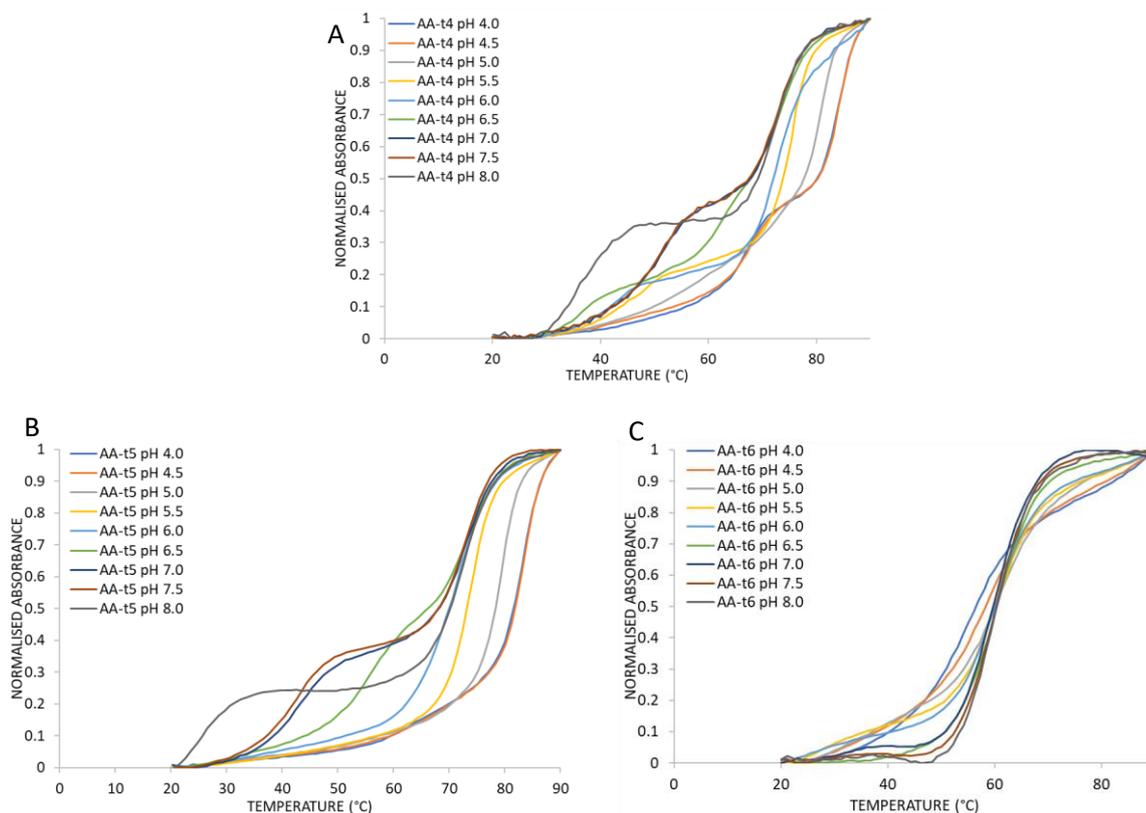
triplex formation could be maintained below 20°C. Similarly, to AA-t2, AAt-3 does not exhibit a clear  $T_m$  for the triplex transition, whereas the duplex transition has a  $T_m$  of ~23.2 °C. However, the shape of the duplex transition, AA-d3, indicates that a triplex transition could be formed below 20 °C. It's important to highlight that there is no variation in pH, as these systems are unaffected by protonation.

**Table 2.2** Melting temperature (°C) of DNA triplexes and their duplexes transition at pH ranges 4.0-8.0. "---" indicate the absence of a triplexes melting transition.

	AA-t1		AA-t2		AA-t3		AA-t4		AA-t5		AA-t6	
	Triplex	Duplex										
pH 4.0	38.9	57.8	< 20	48.5	< 20	21.5	65.5	83.5	61.5	82.5	< 20	53.5
pH 4.5	40	56	< 20	53.6	< 20	21.6	66.7	83.6	63.6	82.6	< 20	57.7
pH 5.0	44	52.1	< 20	57.8	< 20	21.7	71.7	79.7	69.7	77.7	< 20	59.8
pH 5.5	44.9	44.9	---	58.9	< 20	23.8	61.1	74.8	72.9	78.8	24.8	58.9
pH 6.0	42	42	---	59	---	26.9	55	71	68	72.9	---	59
pH 6.5	32.1	46.1	---	59.1	---	24.1	61.1	72.1	54.1	70	---	59.2
pH 7.0	26.5	46.5	---	58.8	---	20.9	50.5	71.5	40.8	71.9	---	59.5
pH 7.5	< 20	44.6	---	59.9	---	21.9	49.6	71.6	40.9	72	---	58.6
pH 8.0	< 20	44.7	---	58	---	26.1	33.8	70.8	24.1	71.1	---	58.7

Different conclusions can be drawn within the intramolecular DNA triplex melting profile (Figure 2.2). The intramolecular version of AA-t1, AA-t4, shows a hyperchromic effect relative to the formation of triple-helices across of all the pH ranges with the highest at pH 5.0 ( $T_m$  71.7 °C) and the lowest at pH 8.0 ( $T_m$  of 33.8 °C); subsequently decreasing as the solution pH becomes basic. (Table 2.2) The analysis of the first derivative of the melting absorptions of AA-t5 that the most stable triplex observed at pH 5.5 with a  $T_m$  of 72.9 °C.

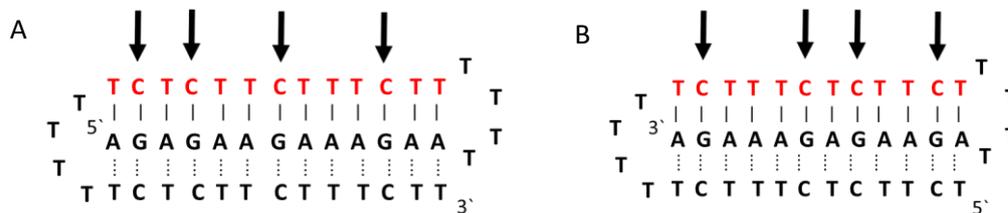
Finally, the triplex formation in solution of AA-t6 can be observed from the melting profile. Upon analysis of the first derivative, it becomes evident that a triplex transition happens at pH 5.5 with a  $T_m$  of 24.8 °C, thus suggesting that triplex formation may occur below this pH threshold.



**Figure 2.2** Normalised melting profile of AA-t4 (A), AA-t5 (B) and AA-t6 (C). Samples were prepared with 20 mM sodium cacodylate buffer at pH ranges from 4.0 to 8.0 with a 5 percent pH increment, 100 mM NaCl and 10 mM MgCl<sub>2</sub>. The absorbance was recorded at  $\lambda = 260$  nm.

The acidic environment is necessary for the triplex formation due to the need of cytosine protonation. Triplexes AA-t1 and AA-t2 exhibit triplex transition at acidic environment. However, the same cannot be observed for AA-t3. In this case the short sequence might have undergone a self-assembly due to its length. A comparable melting profile reported in other studies suggests that the strands might self-assemble in supramolecular conformations.<sup>54</sup> UV melts of the intramolecular triplexes show the biphasic transitions at a wide range of pH. Some systems are formed even at pH 8.0, particularly AA-t4 showing triplex formation ranging with  $T_m$  values from 65.5 °C and  $T_m$  33.8 °C, across the pH range of 4.0 to 8.0. Compared to its intermolecular triplex counterpart, this structure shows enhanced stability, arguing that the thymine loops impose steric constraint to facilitate the three strands to remain in physical proximity to form hydrogen bond. In terms of free energy, the loop reduces the  $\pi$ - $\pi$  base stacking strength, thus decreasing the enthalpy resulting in more stable triplexes.<sup>55</sup> Slightly different melting behaviour can be observed between the two parallel intramolecular triplexes AA-t5 and AA-t4, with the highest triplex formation at pH 5.5 and 5.0, respectively (Table 2.2). Triplex AA-t4 has one more triplet compared

to AA-t5, the but the same number of four C-G:C triplets (Figure 2.3). The additional triplet would have influenced the stability of the structure, if it was a C-G:C instead of a T-A:T.<sup>56,57</sup> On other hand, the formation at lower pH could be attributed to the position of the protonated cytosine. In AA-t4 the C-G:C triplet is two triplets away from the loop and previous studies report that this may lead to a less stable triplex structure.<sup>58</sup>

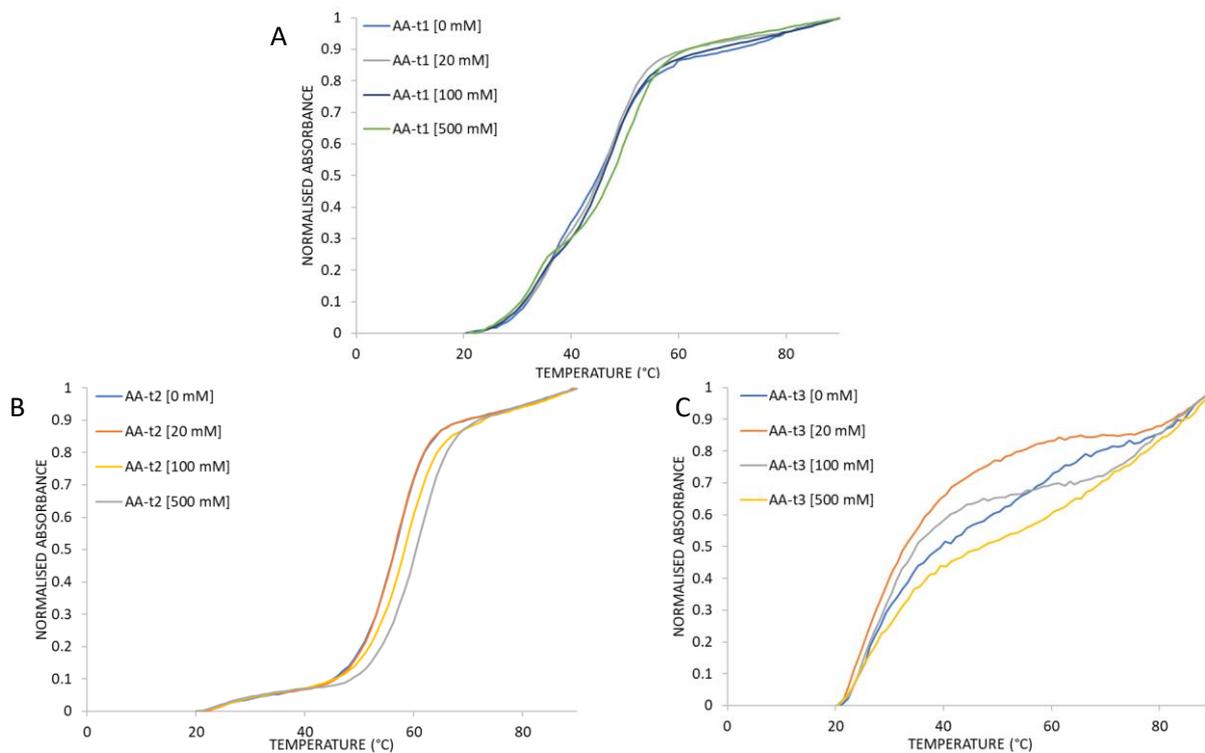


**Figure 2.3** Schematic representation of AA-t4 (A) and AA-t5(B). Arrow point to the C-G:C triplets. TFO shown in red.

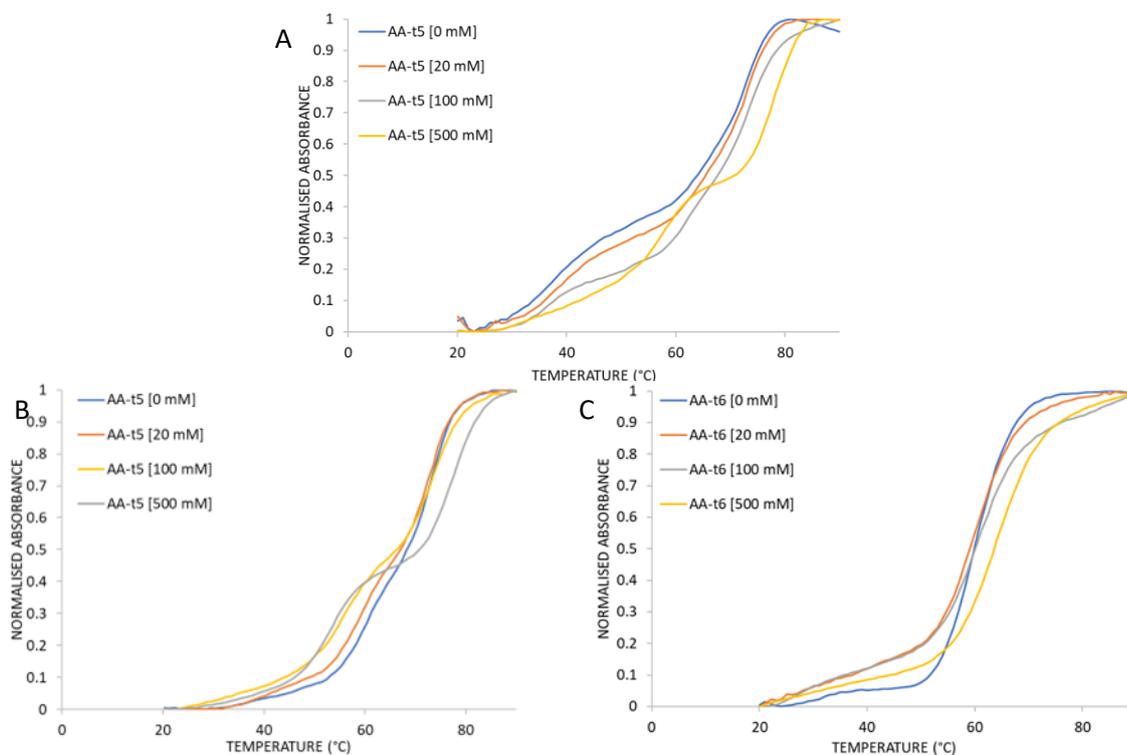
### 2.3.3 Triplex stability supported by cations

Triplex structures necessitate cations in solution to be stable.<sup>59</sup> The aim is to investigate the role that monovalent, divalent, and trivalent ions on the melting behaviour of the selected triplex systems. This is particularly relevant the common use of salts in crystallisation screens.

For each system, samples were prepared with a sodium cacodylate buffer at the pH that previously showed a clear two melting transitions. High acidic buffer (pH 4.0 and 4.5) was avoided to remain close to the physiological conditions. Samples were prepared with a constant concentration of 10 mM MgCl<sub>2</sub> and screened the systems at concentrations of 20 mM, 100 mM, 500 mM of NaCl and without Na<sup>+</sup> ions. The triplex stability is independent of the concentration of NaCl tested, forming a triplex transition also in total absence of NaCl (Figures 2.4-2.5). However, it should be taken into consideration that sodium cacodylate used as a buffer in our group contributes to the total number of Na<sup>+</sup> ions in the solution.



**Figure 2.4** Normalised melting profile of AA-t1 (A), AA-t2 (B) and AA-t3 (C). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t1), pH 5.0 (AA-t2) and pH 5.5 (AA-t3); 10 mM MgCl<sub>2</sub> and NaCl at a concentration of 0 mM, 20 mM, 100 mM and 500 mM. The absorbance was recorded at  $\lambda = 260$  nm



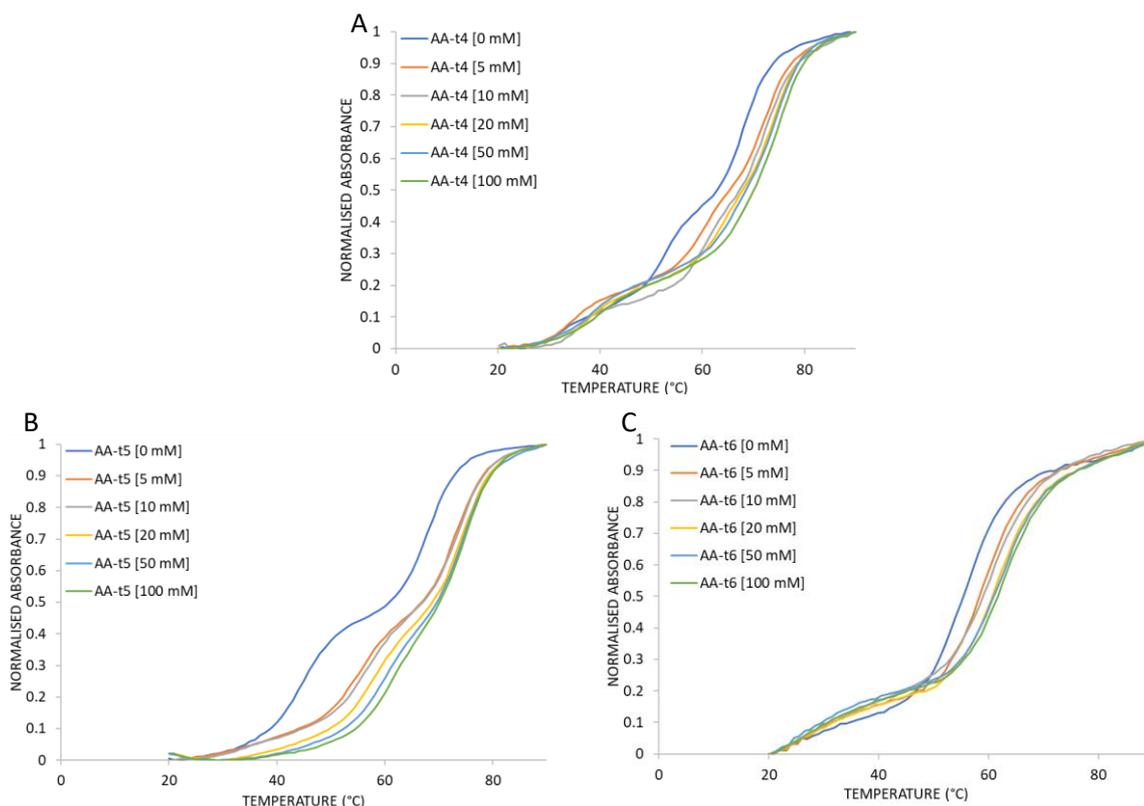
**Figure 2.5** Normalised melting profile of AA-t4 (A), AA-t5 (B) and AA-t6 (C). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t4), and pH 5.5 (AA-t5, AA-t6); 10 mM MgCl<sub>2</sub> and NaCl at a concentration of 0 mM, 20 mM, 100 mM and 500 mM. The absorbance was recorded at  $\lambda = 260$  nm.

Studies shows that monovalent ions, such as Na<sup>+</sup> ions, contribute to stabilising triplex structures by incorporating counterions between strands. However, in an acidic environment, a high concentration of Na<sup>+</sup> can have the opposite effect due to the large uptake of protons. Linearly with the increase of the pH, protons decrease the overall negative charge density, therefore a counterion uptake is not necessary.<sup>22</sup> This explains the  $T_m$  decrease proportionally with the NaCl concentration in the systems analysed here (Table 2.3). A drastic melting temperature decrease is observed in the AA-t4 and AA-t5, which were prepared at lower pH. Nevertheless, unless divalent ions are absent, a solution containing a concentration above 200 mM NaCl suggests triplex formation.<sup>60</sup>

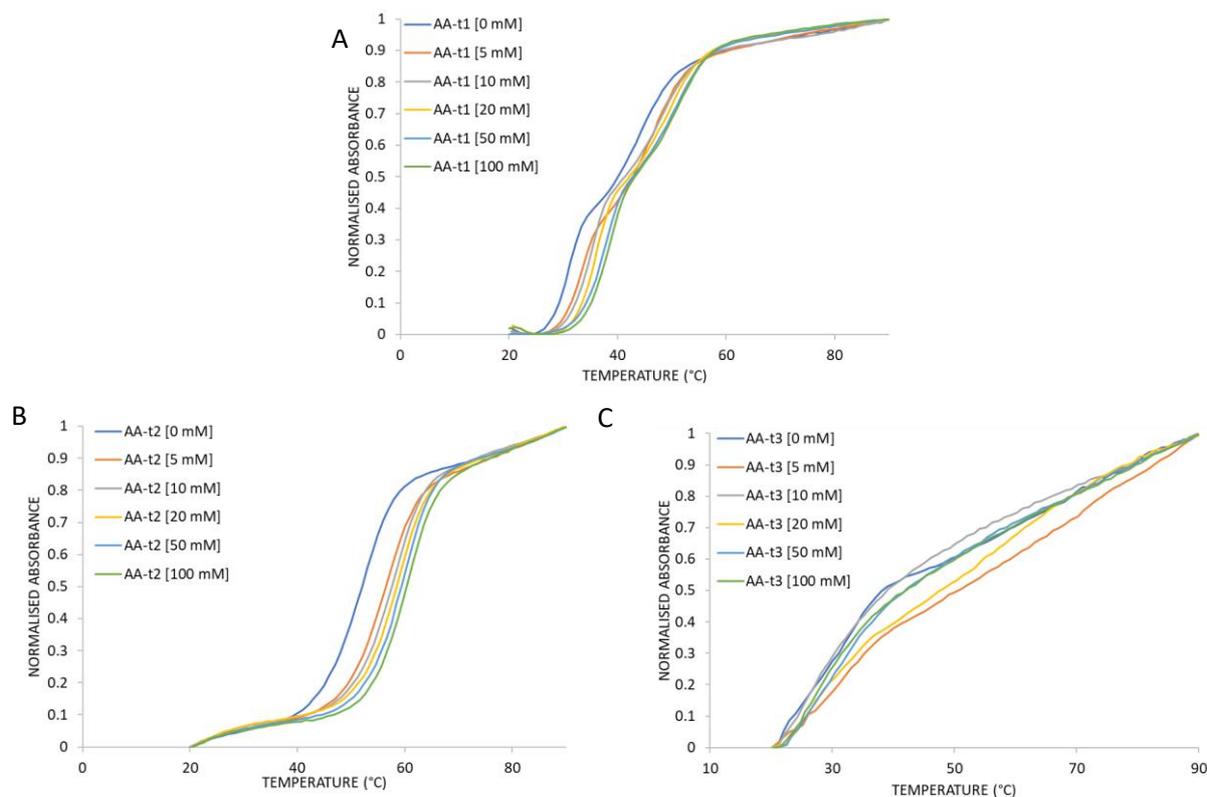
**Table 2.3** Melting temperature (°C) of the triplex and duplex transition at concentration of NaCl of 0 mM, 20 mM, 100 mM and 500 mM. Additional ions are Na<sup>+</sup> from 20 mM sodium cacodylate buffer at pH 6.5 (AA-t1,AA-t4), pH 5.0 (AA-t2) and pH 5.5 (AA-t3, AA-t5, AA-t6) and 10 mM MgCl<sub>2</sub>.

	<b>AA-t1</b>		<b>AA-t2</b>		<b>AA-t3</b>		<b>AA-t4</b>		<b>AA-t5</b>		<b>AA-t6</b>	
	Triplex	Duplex										
<b>0 mM</b>	35.5	45.5	< 20	55.5	< 20	25.1	64.5	71.5	59.9	71.8	< 20	57.5
<b>20 mM</b>	34.7	46.7	< 20	55.6	< 20	22.9	63.7	71.6	59	71.9	< 20	58.6
<b>100 mM</b>	33.1	46.1	< 20	57.8	< 20	24.8	61.1	72.1	54.1	73.1	< 20	58.9
<b>500 mM</b>	32.1	49.1	< 20	59.7	< 20	24	56.7	76.7	52.1	76.1	< 20	62.8

The most commonly divalent magnesium cation used is Mg<sup>2+</sup>, due to its small ionic size this enhances the stability of DNA triplexes by reducing the electrostatic repulsion of the three negatively charged strands.<sup>30,61</sup> Figures 2.6-2.7 shows the melting profile of both inter- and intra-molecular triplexes at concentrations of Mg<sup>2+</sup> ions (5 mM, 10 mM, 20 mM 50 mM, 100 mM) and in the absence of MgCl<sub>2</sub> in solution. A consistent pattern emerges, showing an increase in the melting temperature as the concentration of MgCl<sub>2</sub> increases.



**Figure 2.6** Normalised melting profile of AA-t1 (A), AA-t2 (B) and AA-t3 (C). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t1), pH 5 (AA-t2), and pH 5.5 (AA-t3); 100 mM NaCl and MgCl<sub>2</sub> at a concentration of 0 mM, 5 mM, 10 mM, 20 mM, 50 mM and 100 mM. The absorbance was recorded at  $\lambda = 260$  nm.



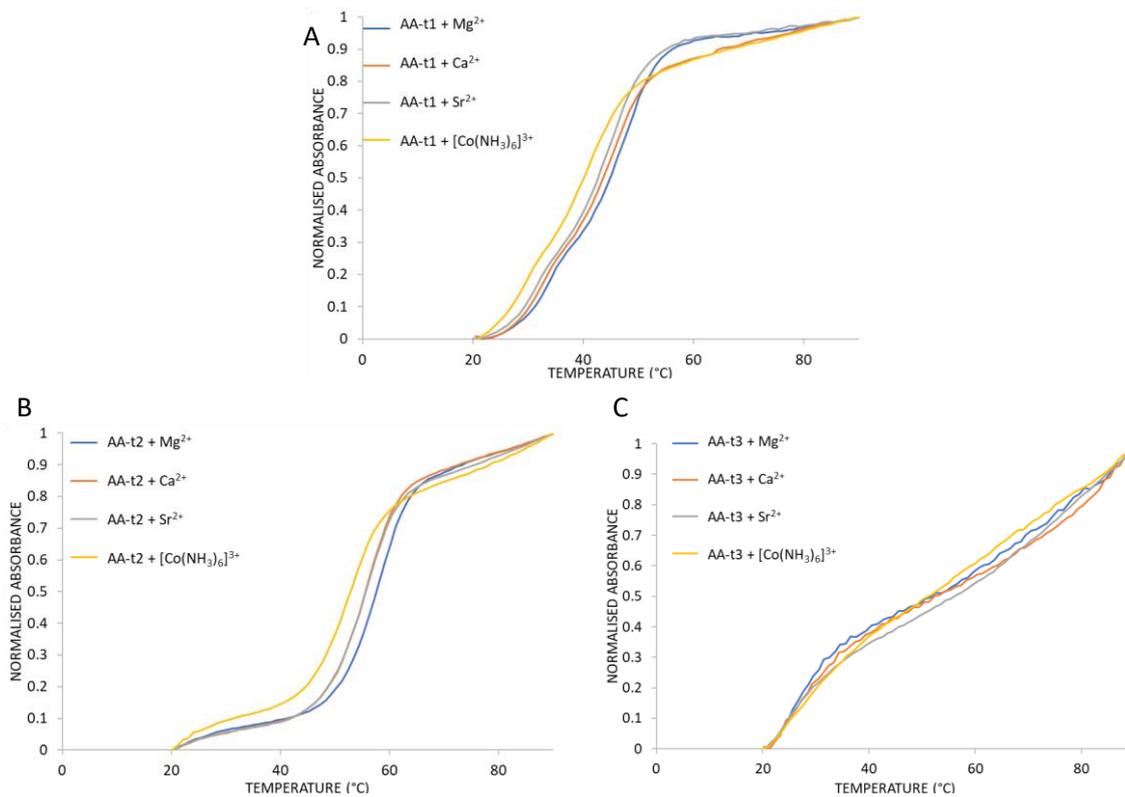
**Figure 2.7** Normalised melting profile of AA-t4 (A), AA-t5 (B) and AA-t6 (C). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t4), and pH 5.5 (AA-t5, AA-t6); 100 mM NaCl and a concentration of 0 mM, 5 mM, 10 mM, 20 mM, 50 mM and 100 mM MgCl<sub>2</sub>. The absorbance was recorded at  $\lambda = 260$  nm.

DNA triplexes show the lowest stability when the triplex formation is not supported by divalent cations (Table 2.4). To note that the mitigation of electrostatic repulsion between the negatively charged phosphate groups is further supported by the presence of Na<sup>+</sup> ions given that the samples were prepared with NaCl 100 mM and 20 mM sodium cacodylate.<sup>62</sup>

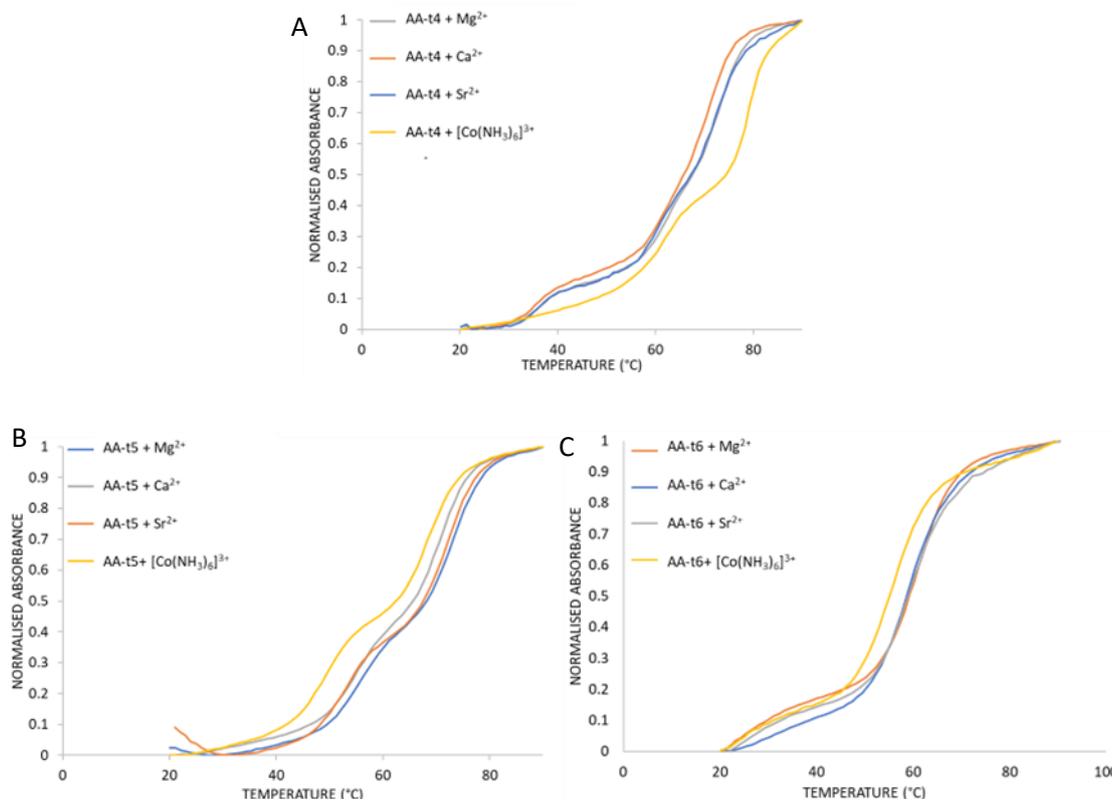
**Table 2.4** Melting temperature (°C) of the triplex and duplex transition at concentration of MgCl<sub>2</sub> of 0 mM, 5 mM, 10 mM, 20 mM, 50 mM and 100 mM.

	<b>AA-t1</b>		<b>AA-t2</b>		<b>AA-t3</b>		<b>AA-t4</b>		<b>AA-t5</b>		<b>AA-t6</b>	
	Triplex	Duplex										
<b>0 mM</b>	27.5	38.5	< 20	51.5	< 20	32.5	51.5	67.5	43.5	67.5	< 20	55.5
<b>5 mM</b>	30.6	44.6	< 20	55.6	< 20	30.6	59.6	71.6	54.6	71.6	< 20	57.6
<b>10 mM</b>	32.7	46.8	< 20	57.8	< 20	24.7	58.8	69.8	54.7	72.7	< 20	59.8
<b>20 mM</b>	33.8	47.9	< 20	57.8	< 20	24.8	62.8	71.8	56.8	71.9	< 20	59.9
<b>50 mM</b>	35.9	48.9	< 20	59	< 20	---	66	73	58.5	72.5	< 20	62
<b>100 mM</b>	37.1	50.1	< 20	60	< 20	---	69	73	60.6	72.6	< 20	62.1

Other divalent ions, such Ca<sup>2+</sup> and Sr<sup>2+</sup>, as well as trivalent ions [Co(NH<sub>3</sub>)<sub>6</sub>]<sup>3+</sup> are known to stabilise DNA triplexes and are commonly employed as crystallisation reagents.<sup>30</sup> UV spectroscopy was used to compare the melting profile of individual triplexes with a concentration of ions of 10 mM of Ca<sup>2+</sup>, Sr<sup>2+</sup> or 0.1 mM of [Co(NH<sub>3</sub>)<sub>6</sub>]<sup>3+</sup> cations to those annealed in the presence of Mg<sup>2+</sup>. Figures 2.8-2.9 illustrate the stabilising effects of the various cations with the sigmoidal curves being shifted more towards lower temperatures proportionate to their atomic radius.



**Figure 2.8** Normalised melting profile of AA-t1 (A), AA-t2 (B) and AA-t3 (C). Samples were prepared with sodium cacodylate buffer at pH 6.5 (AA-t1), pH 5.0 (AA-t2) and pH 5.5 (AA-t3); 100 mM NaCl and 10 mM MgCl<sub>2</sub>, 10 mM CaCl<sub>2</sub>, 10mM SrCl<sub>2</sub> or 0.1 mM of [Co(NH<sub>3</sub>)<sub>6</sub>]Cl<sub>3</sub>. The absorbance was recorded at  $\lambda = 260$  nm.



**Figure 2.9** Normalised melting profile of AA-t4 (A), AA-t5 (B) and AA-t6 (C). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t4), and pH 5.5 (AA-t5, AA-t6); 100 mM NaCl and 10 mM MgCl<sub>2</sub>, 10 mM CaCl<sub>2</sub> 10mM SrCl<sub>2</sub> or 0.1 mM of [Co(NH<sub>3</sub>)<sub>6</sub>]Cl<sub>3</sub>. The absorbance was recorded at  $\lambda = 260$  nm.

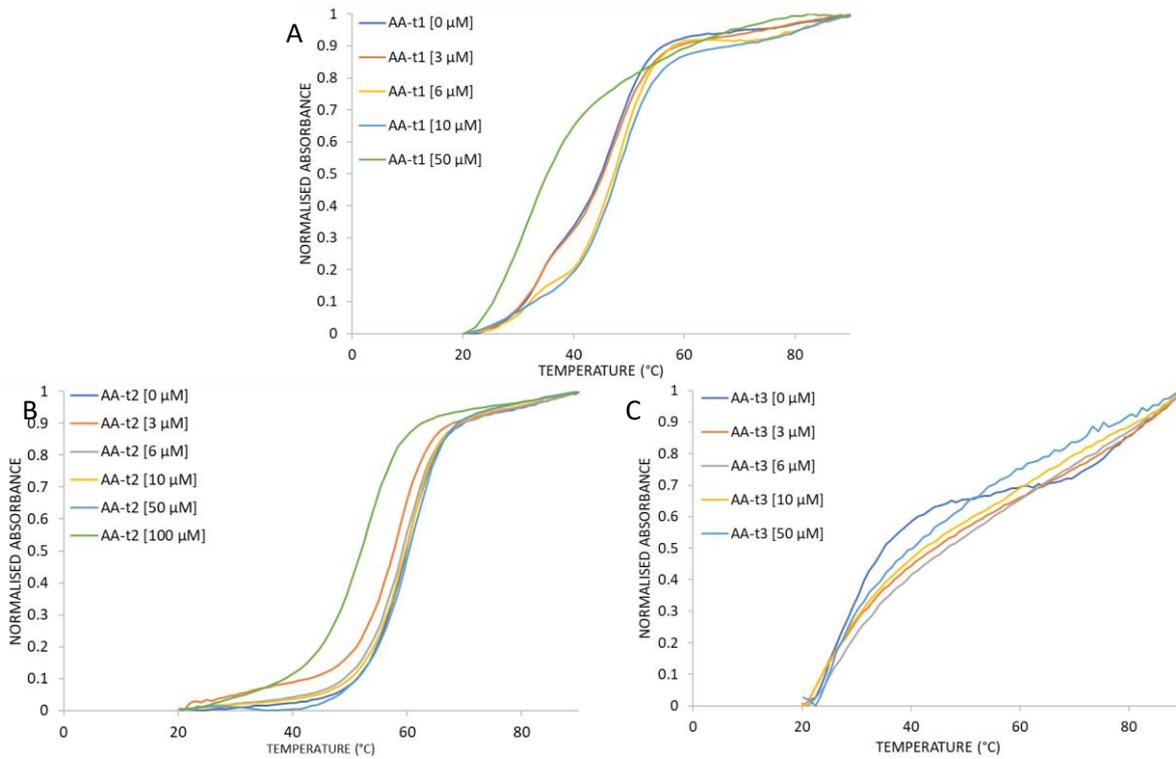
A linear decrease of the triplex melting temperature is observed for the divalent ions in the order  $Mg^{2+} > Ca^{2+} > Sr^{2+}$  in both intermolecular and intramolecular triplexes, except for the systems where the triplex transition is not visible (Table 2.5).

**Table 2.5** Melting temperature ( $^{\circ}C$ ) of the triplex and duplex transition annealed in the presence of 10 mM  $MgCl_2$ , 10 mM  $CaCl_2$ , 10 mM  $SrCl_2$  or 0.1 mM of  $[Co(NH_3)_6]Cl$ .

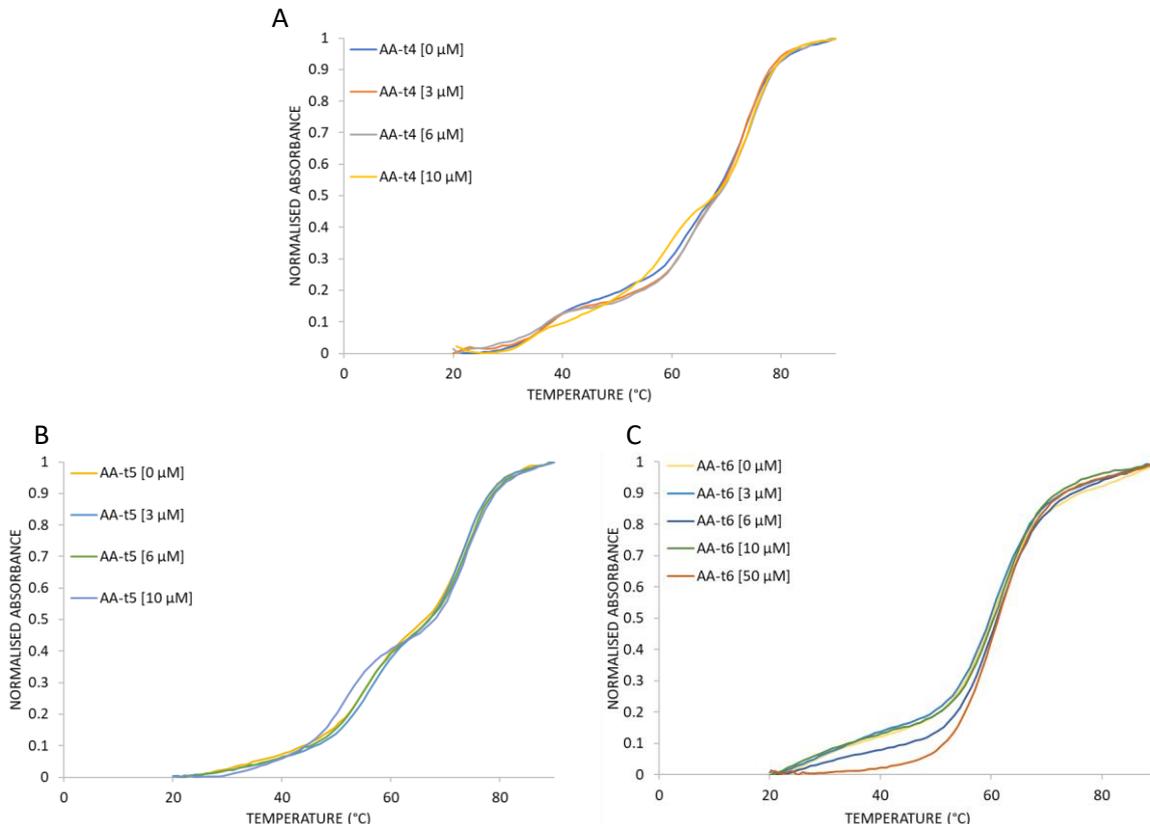
	<b>AA-t1</b>		<b>AA-t2</b>		<b>AA-t3</b>		<b>AA-t4</b>		<b>AA-t5</b>		<b>AA-t6</b>	
	Triplex	Duplex										
<b><math>MgCl_2</math></b>	32.7	46.7	< 20	56.8	< 20	25.1	62	71.1	54.7	71.8	< 20	59.7
<b><math>CaCl_2</math></b>	32	44	< 20	54.9	< 20	21.1	60.8	70	53.9	71.5	< 20	57
<b><math>SrCl_2</math></b>	30.8	43.8	< 20	55	< 20	23	59.8	69.8	53	69	< 20	58.8
<b><math>[Co(NH_3)_6]Cl_3</math></b>	28.5	39.5	< 20	50.6	< 20	< 20	61.7	78.7	47.8	66.9	< 20	55.5

Hexamine cobalt (III) chloride consisting of cobalt with six ammonia ligand is widely used as cryoprotectant in crystallisation screens and been used in triplex studies as triplex stabiliser.<sup>63</sup> Here  $[Co(NH_3)_6]Cl_3$  shows lower stabilisation at a concentration of 0.1 mM compared to the divalent ions. Precipitation occurred at concentrations of 5 mM and 10 mM (data not shown), concluding that to obtain similar stable complexes compared to other cations, a reduced quantity of the trivalent ions is sufficient. Higher concentration of  $[Co(NH_3)_6]Cl^{3+}$  seems to bind other ions in the solution which can lead to the precipitation of the triplexes in the solution.<sup>64</sup>

Finally, the effect of the positive charged polyamine was examined. Figures 2.10-2.11 shows the melting absorbance of DNA triplexes at different concentrations of spermine (ranging from 3  $\mu M$  to 500  $\mu M$ ) compared with the melting profile of the triplexes without spermine.



**Figure 2.10** Normalised melting profile of AA-t1 (A), AA-t2 (B) and AA-t3 (C). Samples were prepared with sodium cacodylate buffer at pH 6.5 (AA-t1), pH 5.0 (AA-t2) and pH 5.5 (AA-t3); 100 mM NaCl, 10 mM  $\text{MgCl}_2$  and different concentration of spermine (3  $\mu\text{M}$ , 6  $\mu\text{M}$ , 10  $\mu\text{M}$ , 50  $\mu\text{M}$ , 100  $\mu\text{M}$ ).



**Figure 2.11** Normalised melting profile of AA-t4 (A), AA-t5 (B) and AA-t6 (C). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t4, AA-t5) and pH 5.5 (AA-t6); 100 mM NaCl, 10 mM  $\text{MgCl}_2$  and different concentration of spermine (3  $\mu\text{M}$ , 6  $\mu\text{M}$ , 10  $\mu\text{M}$ , 50  $\mu\text{M}$ ).

The analysis of the triplex library revealed a modest stability increase at 3  $\mu\text{M}$  spermine, with stability starting to decline with 6  $\mu\text{M}$ . Triplexes annealed with 10  $\mu\text{M}$  spermine and above shows a drastic destabilisation of the melting DNA (Table 2.6). While spermine is known to stabilise DNA triplex formation by interacting with the negatively charged phosphate groups increasing the melting temperature, at higher concentration polyamines impacts the equilibrium between different DNA structures conformation, ultimately leading to the precipitation of the molecules.<sup>65</sup>

**Table 2.6** Melting temperature ( $^{\circ}\text{C}$ ) of the triplex and duplex transition annealed in the presence of spermine concentration of 3  $\mu\text{M}$ , 6  $\mu\text{M}$ , 10  $\mu\text{M}$ , 50  $\mu\text{M}$ , 100  $\mu\text{M}$  and 500  $\mu\text{M}$ . “---” indicate that the melting temperature cannot be calculated, due to precipitation or complex formation below 20  $^{\circ}\text{C}$ . Spermine concentrations of 100  $\mu\text{M}$ , 500  $\mu\text{M}$  and 50  $\mu\text{M}$  for AA-t4 and AA-t5, thus the  $T_m$  was not determined (ND).

	AA-t1		AA-t2		AA-t3		AA-t4		AA-t5		AA-t6	
	Triplex	Duplex										
<b>0 <math>\mu\text{M}</math></b>	32.7	46.8	< 20	58.9	< 20	24.7	61.1	73.1	54.1	73.1	< 20	58.9
<b>3 <math>\mu\text{M}</math></b>	32.5	46.5	< 20	57.5	< 20	22.5	62.6	71.5	55.5	72.5	< 20	59.5
<b>6 <math>\mu\text{M}</math></b>	30.6	47.7	---	58.6	< 20	3.6	61.7	72.6	52.6	71.6	< 20	59.8
<b>10 <math>\mu\text{M}</math></b>	28.8	47.7	---	58.7	< 20	21.7	58.1	73.1	50.7	71.7	< 20	60.8
<b>50 <math>\mu\text{M}</math></b>	---	31	---	58.8	< 20	23.9	ND	ND	ND	ND	< 20	59.9
<b>100 <math>\mu\text{M}</math></b>	ND	ND	< 20	48.6								
<b>500 <math>\mu\text{M}</math></b>	ND	ND										

## 2.3 Conclusions

While much remains unknown about the crystallisation process, the final conditions of the DNA samples could impact the crystallisation outcome. Structures solved by crystallography often reveal conformations adopted by molecules in solution.<sup>66</sup>

In this chapter, UV spectroscopy was employed to determine the formation of triple-helical structures in solution. Other techniques, such as circular dichroism (CD), fluorescence resonance energy transfer (FRET), NMR or non-denaturing PAGE in conjunction with SYBR labelled DNA are also commonly employed to determine the formation of secondary structures. Although, UV-spectroscopy provides limited structural information, for the purposes of this research, UV-vis spectroscopy was selected due to its availability and cost-effectiveness, making it suitable for wide screening.

Several crystallisation screens are commercially available for protein and DNA crystallisation, providing pre-prepared solutions covering a wide range of precipitants, salts, and additives. The crystallisation screen commonly used in our lab and selected for initial screening is the Natrix HT Screen by Hampton Research. The conditions used in crystal screens, designed to support the crystal growth, can also influence the final structures. The Natrix HT Screen comprises 192 crystallisation conditions with a buffer ranging from pH 5.6 to pH 8.5. However, considering that the optimal pH shown for the triplex formation is below pH 7.0, a customised screen version with a more acidic pH was requested. The existing solutions at pH 5.5 and 5.6 were maintained, and the neutral and basic solutions were adjusted with the highest pH corresponding to 6.5, thus aligning with the requirements for triplex formation. Table 2.7 compares the several conditions analysed in this Chapter as an attempt to identify the ideal condition for an initial triplex crystallisation screening. The stable triplexes used as baseline conditions contain 10 mM MgCl<sub>2</sub>, 100 NaCl and 20 mM sodium cacodylate at pH 6.5, 5.0, and 5.5 for the intermolecular triplex AA-t1 and the intramolecular triplexes AA-t4 and AA-t5, respectively.

**Table 2.7** Comparison of the melting temperatures with the baseline condition for the triplexes AA-t1, AA-t4 and AA-t5. The baseline condition contains 10 mM MgCl<sub>2</sub>, 100 mM NaCl and 20 mM sodium cacodylate. Each component was optimised individually while keeping the constant.

		<b>AA-t1</b>	<b>AA-t4</b>	<b>AA-t5</b>
<b>Baseline</b>		32.7	62	54.7
<b>MgCl<sub>2</sub></b>	0 mM	27.5	51.5	43.5
<b>MgCl<sub>2</sub></b>	5 mM	30.6	59.6	54.6
<b>MgCl<sub>2</sub></b>	20 mM	33.8	62.8	56.8
<b>MgCl<sub>2</sub></b>	50 mM	35.9	66	58.5
<b>MgCl<sub>2</sub></b>	100 mM	37.1	69	60.6
<b>CaCl<sub>2</sub></b>	10 mM	32	61	53.9
<b>SrCl<sub>2</sub></b>	10 mM	30.8	58.8	53
<b>[Co(NH<sub>3</sub>)<sub>6</sub>]Cl<sub>3</sub></b>	0.1 mM	28.5	59.8	47.8
<b>NaCl<sub>2</sub></b>	0 mM	35.5	64.5	59.9
<b>NaCl<sub>2</sub></b>	20 mM	34.7	63.7	59
<b>NaCl<sub>2</sub></b>	500 mM	32.1	56.7	52.1
<b>Spermine</b>	3 μM	32.5	62.6	55.5
<b>Spermine</b>	6 μM	30.6	61.7	52.6
<b>Spermine</b>	10 μM	28.8	58.1	50.7

The oligonucleotide strands are more attracted to each other when the concentration of divalent ions is high.<sup>17</sup> It has also been shown that to condense more triplexes a higher concentration of MgCl<sub>2</sub> is required.<sup>64</sup> However, the concentration was maintained at 10 mM for two reasons (1) to mimic the physiological salt composed of 100-140 mM NaCl and 10 mM MgCl<sub>2</sub> as the structural information is needed for gene editing applications and (2) to prevent the crystallisation of inorganic crystals they would most likely interfere with the crystallisation of DNA triplexes. Compared to the other concentration tested, 10 mM of MgCl<sub>2</sub> is sufficient to stabilise DNA triplexes in solution. Other divalent and trivalent ions tested here show a lack of enhanced melting stability. The triplexes annealed without divalent ions, show triplex formation in terms of temperature, although less stable. It is reported that intramolecular triplexes are not dependent on divalent ions for their formation.<sup>67</sup> Although here a complete comparison with intermolecular triplexes is difficult to obtain due to their formation below 20 °C, these will still need a high concentration of salt to ensure stable triplex formation.<sup>22</sup>

The optimal outcome in terms of melting stability is achieved through a combination of NaCl and MgCl<sub>2</sub>, conditions generally accepted in research experiments with triple-helical DNA structures. For an initial crystallisation screening, additional Na<sup>+</sup> beyond that provided by the buffer is not included, thus reflecting the highest *T<sub>m</sub>* obtained in the NaCl screening (Figure).

Given that spermine has been demonstrated to relatively stabilise triplex formation, a screening of the triplex system containing spermine will be conducted. However, the concentration will be maintained at low levels. Both spermine and  $\text{Na}^+$  play a role in increasing the interspace between strands. Consequently, higher concentrations of these polyamines may disrupt the stability of triplex crystal lattice formation.<sup>32</sup>

In general, the intramolecular triplexes within the library show a more stable behaviour in solution compared to the intermolecular triplexes, as evidenced by UV-Vis showing melting curves in two sequential transitions. Despite the extensive testing under various conditions to achieve stability above 20 °C for triplexes AA-t2 and AA-t6, the efforts were unsuccessful, thus, requiring setting up crystallisation plates in a low-temperature environment.

Conversely, AA-t3 is reported to be the least stable, due to the short sequence length. Despite this instability in solution, attempts were made to crystallise this short sequence-forming DNA triplex. Previous crystallisation outcomes have demonstrated that short oligonucleotides, even if not highly unstable in solution, can still crystallise successfully at the elevated concentrations required for crystallisation experiments.<sup>68,69</sup> Additional details regarding the crystallisation of AA-t3 will be discussed in Chapter 6.

Archiving successful crystallisation involves optimisation of different parameters including the pH, temperature, and concentrations of precipitants and additives. The findings presented in this chapter will provide a guide for identifying the optimal and stable conditions of triplexes during the optimisation process.

## 2.4 References

1. Dalla Pozza, M., Abdullrahman, A., Cardin, C. J., Gasser, G. & Hall, J. P. Three's a crowd - stabilisation, structure, and applications of DNA triplexes. *Chem Sci* **13**, 10193–10215 (2022).
2. Chayen, N. E. Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol* **14**, 577–583 (2004).
3. McPherson, A. & Cudney, B. Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr F Struct Biol Commun* **70**, 1445–1467 (2014).
4. Roberts, R. W. & Crothers, D. M. Prediction of the stability of DNA triplexes. *Proceedings of the National Academy of Sciences* **93**, 4320–4325 (1996).
5. Collins, B., Stevens, R. C. & Page, R. Crystallization optimum solubility screening: Using crystallization results to identify the optimal buffer for protein crystal formation. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 1035–1038 (2005).
6. Wu, P., Kawamoto, Y., Hara, H. & Sugimoto, N. Effect of divalent cations and cytosine protonation on thermodynamic properties of intermolecular DNA double and triple helices. *J Inorg Biochem* **91**, 277–285 (2002).
7. Pilch, D. S., Brousseau, R. & Shafer, R. H. Thermodynamics of triple helix formation: spectrophotometric studies on the d(A) 10-2d(T) 10 and d(C + C + GGCC triple helices. *Nucleic Acids Res* **18**, 5743–5750 (1990).
8. Wu, P., Hara, H., Kawamoto, Y. & Sugimoto, N. Effect of cytosine protonation and cation on thermodynamic properties of parallel DNA triplex family. *Nucleic Acids Research Supplement* **1**, 39–40 (2001).
9. Powell, S. W., Jiang, L. & Russu, I. M. Proton exchange and base pair opening in a DNA triple helix. *Biochemistry* **40**, 11065–11072 (2001).
10. Sugimoto, N., Wu, P., Hara, H. & Kawamoto, Y. pH and Cation Effects on the Properties of Parallel Pyrimidine Motif DNA Triplexes. *Biochemistry* **40**, 9396–9405 (2001).
11. Wärmländer, S., Sandström, K., Leijon, M. & Gräslund, A. Base-Pair Dynamics in an Antiparallel DNA Triplex Measured by Catalyzed Imino Proton Exchange Monitored via <sup>1</sup>H NMR Spectroscopy. *Biochemistry* **42**, 12589–12595 (2003).
12. Hu, Y., Ceconello, A., Idili, A., Ricci, F. & Willner, I. Triplex DNA Nanostructures: From Basic Properties to Applications. *Angewandte Chemie - International Edition* vol. 56 15210–15233 Preprint at <https://doi.org/10.1002/anie.201701868> (2017).
13. Chen, X. *et al.* Triplex DNA Nanoswitch for pH-Sensitive Release of Multiple Cancer Drugs. *ACS Nano* **13**, 7333–7344 (2019).
14. Goobes, R., Cohen, O. & Minsky, A. Unique condensation patterns of triplex DNA: physical aspects and physiological implications. *Nucleic Acids Res* **30**, 2154–2161 (2002).
15. Wu, P., Kawamoto, Y., Hara, H. & Sugimoto, N. Effect of divalent cations and cytosine protonation on thermodynamic properties of intermolecular DNA double and triple helices. *J Inorg Biochem* **91**, 277–285 (2002).

16. Völker, J., Klump, H. H., Manning, G. S. & Breslauer, K. J. Counterion association with native and denatured nucleic acids: An experimental approach. *J Mol Biol* **310**, 1011–1025 (2001).
17. Tan, Z. J. & Chen, S. J. Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length. *Biophys J* **90**, 1175–1190 (2006).
18. Soto, A. M. & Marky, L. A. Thermodynamic Contributions for the Incorporation of GTA Triplets within Canonical TAT/TAT and C+GC/C+GC Base-Triplet Stacks of DNA Triplexes. *Biochemistry* **41**, 12475–12482 (2002).
19. Sauter, A. *et al.* Nonclassical pathways of protein crystallization in the presence of multivalent metal ions. *Cryst Growth Des* **14**, 6357–6366 (2014).
20. Floris, R., Scaggiante, B., Manzini, G., Quadrifoglio, F. & Xodo, L. E. Effect of cations on purine-purine-pyrimidine triple helix formation in mixed-valence salt solutions. *Eur J Biochem* **260**, 801–809 (1999).
21. Dadarlat, V. M. & Saxena, V. K. Stability of triple-helical poly(dT)-poly(dA)-poly(dT) DNA with counterions. *Biophys J* **75**, 70–91 (1998).
22. Lee, H. T., Khutsishvili, I. & Marky, L. A. DNA complexes containing joined triplex and duplex motifs: Melting behavior of intramolecular and bimolecular complexes with similar sequences. *Journal of Physical Chemistry B* **114**, 541–548 (2010).
23. Cheng, A.-J. & Van Dyke, M. W. Monovalent cation effects on intermolecular purine-purine-pyrimidine triple-helix formation. *Nucleic Acids Res* **21**, 5630–5635 (1993).
24. Zinchenko, A. A. & Yoshikawa, K. Na<sup>+</sup> shows a markedly higher potential than K<sup>+</sup> in DNA compaction in a crowded environment. *Biophys J* **88**, 4118–4123 (2005).
25. Schneider, B., Patel, K. & Berman, H. M. Hydration of the phosphate group in double-helical DNA. *Biophys J* **75**, 2422–2434 (1998).
26. Collins, K. D. Ion hydration: Implications for cellular function, polyelectrolytes, and protein crystallization. *Biophys Chem* **119**, 271–281 (2006).
27. Potaman, V. N. & Soyfer, V. N. Divalent Metal Cations upon Coordination to the N7 of Purines Differentially Stabilize the PyPuPu DNA Triplex due to Unequal Hoogsteen-type Hydrogen Bond Enhancement. *J Biomol Struct Dyn* **11**, 1035–1040 (1994).
28. Kankia, B. I. Mg<sup>2+</sup>-induced triplex formation of an equimolar mixture of poly(rA) and poly(rU). *Nucleic Acids Res* **31**, 5101–5107 (2003).
29. Burda, J. V., Šponer, J. S., Leszczynski, J., Hobza, P. & Heyrovský, J. Interaction of DNA Base Pairs with Various Metal Cations (Mg<sup>2+</sup>, Ca<sup>2+</sup>, Sr<sup>2+</sup>, Ba<sup>2+</sup>, Cu<sup>+</sup>, Ag<sup>+</sup>, Au<sup>+</sup>, Zn<sup>2+</sup>, Cd<sup>2+</sup>, and Hg<sup>2+</sup>): Nonempirical ab Initio Calculations on Structures, Energies, and Nonadditivity of the Interaction. *J Phys Chem B* **101**, 9670–9677 (1997).
30. Beck, A., Vijayanathan, V., Thomas, T. & Thomas, T. J. Ionic microenvironmental effects on triplex DNA stabilization: Cationic counterion effects on poly(dT)·poly(dA)·poly(dT). *Biochimie* **95**, 1310–1318 (2013).
31. Zhang, Z. L., Wu, Y. Y., Xi, K., Sang, J. P. & Tan, Z. J. Divalent Ion-Mediated DNA-DNA Interactions: A Comparative Study of Triplex and Duplex. *Biophys J* **113**, 517–528 (2017).

32. Raspaud, E., Durand, D. & Livolant, F. Interhelical spacing in liquid crystalline spermine and spermidine-DNA precipitates. *Biophys J* **88**, 392–403 (2005).
33. Saminathan, M., Thomas, T., Shirahata, A., Pillai, C. K. S. & Thomas, T. J. Polyamine structural effects on the induction and stabilization of liquid crystalline DNA: potential applications to DNA packaging, gene therapy and polyamine therapeutics. *Nucleic Acids Res* **30**, 3722–3731 (2002).
34. Tari, L. W. & Secco, A. S. Base-pair opening and spermine binding-B-DNA features displayed in the crystal structure of a gal operon fragment: implications for protein-DNA recognition. *Nucleic Acids Res* **23**, 2065–2073 (1995).
35. Thomas, T. & Thomas, T. J. Selectivity of polyamines in Triplex DNA Stabilization. *Biochemistry* **32**, 14068–14074 (1993).
36. Egli, M., Dean Williams, L., Gao, Q. & Rich Alexander. Structure of the Pure-Spermine Form of Z-DNA (Magnesium Free) at 1-Å Resolution. *Biochemistry* **30**, 11357–11484 (1991).
37. Zhang, B. *et al.* Protein crystallization with DNA templates. *Particuology* **86**, 262–268 (2024).
38. Lee, C.-H., Mizusawa, H. & Kakefuda, T. Unwinding of double-stranded DNA helix by dehydration. *Proceedings of the National Academy of Sciences* **78**, 2838–2842 (1981).
39. Dickerson, R. E., Goodsell, D. & Kopka, M. L. MPD and DNA Bending in Crystals and in Solution. *J. Mol. Biol* **256**, 108–125 (1996).
40. Vlieghe, D. *et al.* Parallel and Antiparallel (G GC)<sub>2</sub> Triple Helix Fragments in a Crystal Structure. *Science* (1979) **273**, 1702–1705 (1996).
41. Rhee, S., Han, Z., Liu, K., Miles, H. T. & Davies, D. R. Structure of a Triple Helical DNA with a Triplex–Duplex Junction. *Biochemistry* **38**, 16810–16815 (1999).
42. Spink, C. H. & Chaires, J. B. Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA. *Biochemistry* **38**, 496–508 (1999).
43. Spink, C. H. & Chaires, J. B. Selective Stabilization of Triplex DNA by Poly(ethylene glycols). *Proc. Natl. Acad. Sci. U.S.A* **117**, 12887–12888 (1995).
44. Tarköy, M., Phipps, A. K., Schultze, P. & Feigon, J. Solution structure of an intramolecular DNA triplex linked by hexakis(ethylene glycol) units: d(AGAGAGAA-(EG)<sub>6</sub>-TTCTCTCT-(EG)<sub>6</sub>-TCTCTCTT). *Biochemistry* **37**, 5810–5819 (1998).
45. Abdallah, H. O. *et al.* Stabilisation of self-assembled DNA crystals by triplex-directed photo-cross-linking. *Chemical Communications* 8014–8017 (2016) doi:10.1039/c6cc03695c.
46. Asensio, J. L., Brown, T. & Lane, A. N. Solution conformation of a parallel DNA triple helix with 5' and 3' triplex-duplex junctions. *Structure* **7**, 1–11 (1999).
47. Goobes, R. Unique condensation patterns of triplex DNA: physical aspects and physiological implications. *Nucleic Acids Res* **30**, 2154–2161 (2002).
48. Htun, H. & Dahlberg, J. E. Topology and Formation of Triple-Stranded H-DNA. *Science* (1979) **243**, 1571–1576 (1989).
49. Shimizu, M., Konishi, A., Shimada, Y., Inoue, H. & Ohtsuka, E. Oligo(2'-O-methyl)ribonucleotides Effective probes for duplex DNA. *FEBS Lett* **302**, 155–158 (1992).

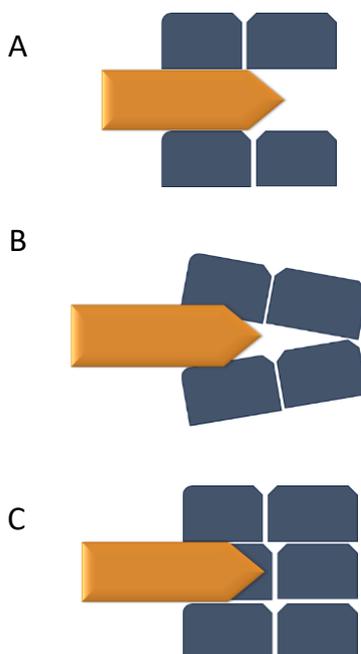
50. Soto, A. M., Loo, J. & Marky, L. A. Energetic Contributions for the Formation of TAT/TAT,TAT/CGC+, and CGC+/CGC+Base Triplet Stacks. *Biopolymers* **124**, 14355–14363 (2002).
51. Rumney, S. & Kool, E. T. Structural Optimization of Non-Nucleotide Loop Replacements for Duplex and Triplex DNAs. *J Am Chem Soc* **117**, 5635–5646 (1995).
52. Miyoshi, D., Nakamura, K., Tateishi-Karimata, H., Ohmichi, T. & Sugimoto, N. Hydration of Watson-Crick base pairs and dehydration of Hoogsteen base pairs inducing structural polymorphism under molecular crowding conditions. *J Am Chem Soc* **131**, 3522–3531 (2009).
53. Wang, E., Koshlap, K. M., Gillespie, P., Dervan, P. B. & Feigon, J. Solution structure of a pyrimidine-purine-pyrimidine triplex containing the sequence-specific intercalating non-natural base D3. *J Mol Biol* **257**, 1052–1069 (1996).
54. Shankaraswamy, J., Tyagi, S., Singh, A., Miyoshi, D. & Saxena, S. Metal sensitive and DNA concentration dependent structural rearrangement of short oligonucleotide into large suprastructures. *J Biomol Struct Dyn* **37**, 2211–2218 (2019).
55. Lee, H., Carr, C. E., Khutsishvili, I. & Marky, L. A. Effect of Loop Length and Sequence on the Stability of DNA Pyrimidine Triplexes with TAT Base Triplets. *J Phys Chem* **121**, 9081–9320 (2017).
56. Keppler, M. D. & Fox, K. R. Relative stability of triplexes containing different numbers of T·AT and C + ·GC triplets. *Nucleic Acids Res* **25**, 4644–4649 (1997).
57. Volker, J. & Klump, H. H. Electrostatic Effects in DNA Triple Helices. *Biochemistry* **33**, 13502–13508 (1994).
58. Leitner, D., Schröder, W. & Weisz, K. Influence of Sequence-Dependent Cytosine Protonation and Methylation on DNA Triplex Stability. *Biochemistry* **39**, 5886–5892 (2000).
59. Sugimoto, N., Wu, P., Hara, H. & Kawamoto, Y. pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes. *Biochemistry* **40**, 9396–9405 (2001).
60. Plum, G. E., Park, Y.-W., Singleton, S. F., Dervant, P. B. & Breslauer, K. J. Thermodynamic characterization of the stability and the melting behavior of a DNA triplex: A spectroscopic and calorimetric study. *Proceedings of the National Academy of Sciences* **87**, 9436–9440 (1990).
61. Muñoz, J. *et al.* Can divalent metal cations stabilize the triplex motif? Theoretical study of the interaction of the hydrated Mg<sup>2+</sup> cation with the G-G·C triplet. *Journal of Physical Chemistry B* **106**, 8849–8857 (2002).
62. Debin, A. *et al.* Stability of G,A triple helices. *Nucleic Acids Res* **27**, 2699–2707 (1999).
63. Gyo Kim, M. & Daniel Camerini-Otero, R. An Alteration in the Structure of the Minor Groove of Duplex DNA Induced by the Formation of an Intermolecular d(GA)<sub>n</sub>:d(GA)<sub>n</sub>·d(TC)<sub>n</sub> Triplex. *Mol Cells* **7**, 641–647 (1997).
64. Qiu, X., Parsegian, V. A., Rau, D. C. & Baldwin, R. Divalent counterion-induced condensation of triple-strand DNA. *Biophysics and computational biology* **107**, 21481–21486 (2010).
65. Saminathan, M. *et al.* Ionic and structural specificity effects of natural and synthetic polyamines on the aggregation and resolubilization of single-, double-, and triple-stranded DNA. *Biochemistry* **38**, 3821–3830 (1999).

66. Vorlíčková, M. *et al.* Comparison of the solution and crystal conformations of (G + C)-rich fragments of DNA. *Biophys J* **71**, 1530–1538 (1996).
67. Kaushik, S. *et al.* Presence of divalent cation is not mandatory for the formation of intramolecular purine-motif triplex containing human c-jun protooncogene target. *Biochemistry* **50**, 4132–4142 (2011).
68. Hall, J. P. *et al.* X-ray crystal structure of rac-[Ru(phen)<sub>2</sub>dppz]<sub>2</sub><sup>+</sup> with d(ATGCAT)<sub>2</sub> shows enantiomer orientations and water ordering. *J Am Chem Soc* **135**, 12652–12659 (2013).
69. Hall, J. P. *et al.* Guanine Can Direct Binding Specificity of Ru–dipyridophenazine (dppz) Complexes to DNA through Steric Effects. *Chemistry - A European Journal* **23**, 4981–4985 (2017).

## CHAPTER 3 - Intercalation of ruthenium complex in DNA triplexes with extended underlying duplex

### 3.1 Introduction

The DNA binding properties of Ruthenium (II) polypyridyl complexes have been extensively studied.<sup>1,2</sup> The octahedral complex  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  [phen = 1,10-phenanthroline, dppz = dipyrrophenazine] acts as a “light switch” molecule, exhibiting little luminescence in an aqueous environment and an enhanced photoluminescent effect when intercalated in the DNA. The DNA stacking bases protect the phenazine nitrogen from binding with the water molecules, thus preventing the quenching effect.<sup>3,4</sup> The intercalation of planar heterocyclic aromatic rings as a non-covalent stacking between nucleotides was first proposed in 1961 by Lerman.<sup>5</sup> In the case of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ , the dppz component is a planar and aromatic ligand that intercalates between base pairs increasing their distances and overall distortion of the DNA structure. This complete intercalation of the Ru(II) complex is described as classical intercalation. Other intercalation modes are semi-intercalation, where the molecule does not cause a complete separation of the base pairs; and the quasi-intercalation by which the complex is aligned with the stacking of the base pairs (Figure 3.1).<sup>6</sup> The phenanthroline component of the metal complexes has been found to exhibit both semi-intercalation and quasi-intercalation modes. The intercalation by ruthenium polypyridyl complexes has been extensively studied by our group.<sup>7</sup>



**Figure 3.1** Diagram illustrating the different binding modes intercalation (A), semi-intercalation (B), and quasi-intercalation (C).

While intercalation of ruthenium polypyridyl complexes in dsDNA is extensively studied, there are limited reports on their binding to DNA triplexes. The challenge with these systems lies in adjusting specific factors required for their stability, including triplex sequence, pH conditions, and ionic environment, as discussed in Chapter 2. Spectroscopic data shows that the dppz component of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  intercalates between the T-A:T bases of a DNA triplex, in a similar fashion to the duplex DNA.<sup>8</sup>

This chapter aims to investigate the intercalation effect of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  in DNA triplexes. The system selected to carry this study is the intermolecular parallel triplex AA-t1 from the library in Chapter 2 which demonstrated thermal stability across a range of different conditions. It was hypothesised that the length of the underlying duplex DNA may impact the intercalation of the metal complex, so the duplex length of the triplex system was systematically increased, keeping the third strand length constant.

The  $\Delta$ - and  $\Lambda$ -enantiomers ( $\Delta$ -Ru and  $\Lambda$ -Ru) were studied alongside the racemic form of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  (rac-Ru), as the enantiomers can have distinct binding interactions with the DNA.<sup>9</sup> Studying separated enantiomers can provide greater insight compared to the racemic mixtures. Since the racemate represents equal proportions of  $\Delta$ -Ru and  $\Lambda$ -Ru, this allows to determine if one enantiomer's binding predominates over the other.

Different spectroscopic techniques were employed to investigate the interaction of the ruthenium complex with DNA triplexes. The thermal denaturation profile and induced structural changes of the  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  intercalation in the molecules will be explored using UV melting and circular dichroism. Finally, the "light switch" effect that arises when this complex is intercalated within the base pairs of the DNA<sup>10</sup> will be used to determine the intercalation of the complex through luminescence techniques. Comparing the results across the multiple triplex sequences will elucidate the relationship between triple-helical structures and ruthenium polypyridyl complexes.

## 3.2 Materials and Methods

### 3.2.1 Oligonucleotides and buffer solutions

All chemicals and oligonucleotides were purchased from Sigma-Aldrich. The concentrations of all oligonucleotides were calculated from the absorbance value at 260 nm using their extinction coefficients of the nearest neighbour model obtained from an online calculator (<https://atdbio.com/tools/oligo-calculator>). All solvents were obtained at HPLC grade and used without further purification.  $\Delta$ - and  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> were synthesized and resolved by our research team<sup>11</sup>

### 3.2.2 UV absorption measurements

UV melting experiments were carried out using Agilent Cary 100 with a temperature-controlled six-cell changer. Samples contained three single strands for intermolecular triplex DNA triplex and one single strand for intramolecular DNA triplex with a final calculated absorbance at 1 AU. The buffer was prepared with 20 mM sodium cacodylate at a pH of 6.5 (unless otherwise stated), 100 mM NaCl and 10 mM MgCl<sub>2</sub>. Samples have been annealed at 90° for 5 minutes and left to cool down at room temperature overnight. Samples containing [Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> were prepared with the same conditions, with the metal complex added after the annealing and left intercalating for 10 minutes. The concentration of the Ru(II) complexes was 1:1 DNA/Ruthenium (II). Absorption was recorded at 260 at 1 °C intervals between 20– 90 °C, with a temperature change rate of 0.5 °C/min in a 1 cm pathlength quartz cuvette. Normalised melting curves were generated from this data. First derivative curves are shown in Figure A3.1-A3.3.

### 3.2.3 Circular dichroism measurements

The oligonucleotides were dissolved in a buffer of 20 mM sodium cacodylate at a pH of 6.5, 100 mM NaCl and 10 mM MgCl<sub>2</sub>. The final concentration of each strand of the triplex was 70 μM. The solutions were annealed by reaching a temperature of 90° for 5 minutes and slowly allowed to cool down at room temperature. Samples containing [Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> were prepared with the same conditions, with the metal complex added after the annealing and left intercalating for 10 minutes. [DNA duplex] or [DNA triplex] ratio with [Ru] of 1:1. CD spectra were recorded at beamline B23 at Diamond Light Source at temperatures from 10 °C to 70 °C with a data collection every 5°C, between 180 and 350 nm wavelengths in a 0.02 cm pathlength cuvette. Normalised CD spectra were generated from this data, with PMT cut-off below 600 HT voltage.

### 3.2.4 Fluorescence measurements

Fluorescence studies were carried out with Cary Eclipse Fluorescence Spectrophotometer (Agilent). The pathlength of the cuvettes used was 1 mm. Fluorescence spectra were obtained using a 1:1 and 2:1 ratio of oligonucleotide DNA to  $\Lambda$ -,  $\Delta$ -, *rac*-[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup>. Data collected at 20 °C with excitation at 475nm (isosbestic point for [Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup>) and emission at 550-800nm using an emission filter 550-1100 nm. The solutions were prepared with the same buffer condition as in other solution studies. Ruthenium complexes were left intercalating for one hour before measurements.

### 3.3 Results and Discussion

AA-t1 is an intermolecular DNA triplex composed of polypyrimidine-polypurine-polypyrimidine sequences. The double-strand DNA, AA-d1, is composed of 13-mer oligonucleotides,  $d_1(5'-AAGAAAGAAGAGA-3')$  and  $d_2(5'-TCTCTTCTTTCTT-3')$  that forms a parallel triple-helical structure with a 13-mer TFO,  $d_{TFO}(5'-TTCTTTCTTCTCT-3')$ . The duplex component was extended at the 3'-end for  $d_1$  and 5'-end for  $d_2$  using non-modified deoxyoligonucleotides while keeping the  $d_{TFO}$  length and base composition unchanged. (Table 3.1)

The extension of up to four base pairs was designed to systematically study the effect of a more stable duplex on the overall triplex stability. Furthermore, by extending the duplex length, the aim is to create a model to examine the binding of  $[Ru(phen)_2(dppz)]^{2+}$  into the duplex region adjacent to a tethered triplex-forming oligonucleotide. Finally, examining the binding behaviour of  $[Ru(phen)_2(dppz)]^{2+}$  would also reveal the minimum duplex length as well as the binding site size required adjacent to the triplex for stable ruthenium complex binding.

**Table 3.1** Schematic representation of AA-d1, AA-t1 and the duplex-extended triplexes. The TFO is shown in red and the base extensions in bold.

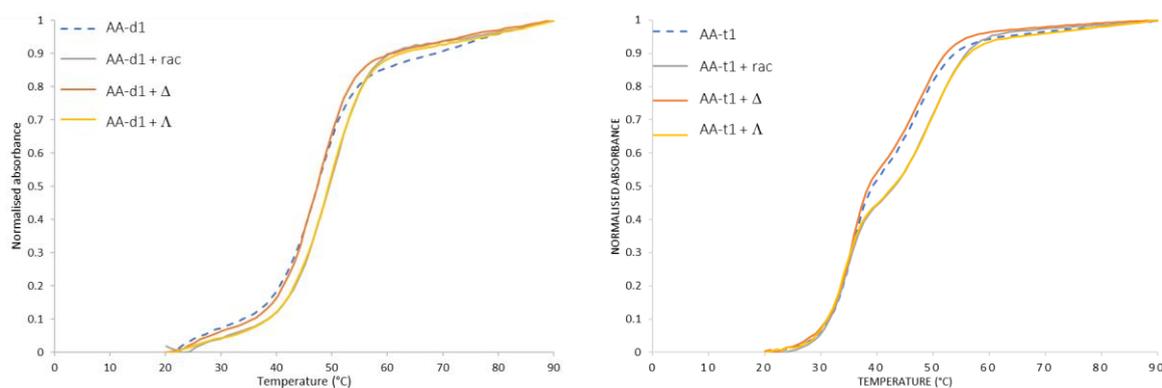
AA-d1	AA-t1
5'-AAGAAAGAAGAGA-3' 3'-TTCTTTCTTCTCT-5'	5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGA-3' 3'-TTCTTTCTTCTCT-5'
AA-t1_T	AA-t1_TA
5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGAT-3' 3'-TTCTTTCTTCTCT <b>A</b> -5'	5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGAT <b>A</b> -3' 3'-TTCTTTCTTCTCT <b>AT</b> -5'
AA-t1_TAT	AA-t1_TATA
5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGAT <b>AT</b> -3' 3'-TTCTTTCTTCTCT <b>ATA</b> -5'	5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGAT <b>ATA</b> -3' 3'-TTCTTTCTTCTCT <b>ATAT</b> -5'
AA-t1_C	AA-t1_CG
5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGAC-3' 3'-TTCTTTCTTCTCT <b>G</b> -5'	5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGAC <b>CG</b> -3' 3'-TTCTTTCTTCTCT <b>GC</b> -5'
AA-t1_CGC	AA-t1_CGCG
5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGAC <b>CGC</b> -3' 3'-TTCTTTCTTCTCT <b>GCG</b> -5'	5'- <b>TTCTTTCTTCTCT</b> -3' 5'-AAGAAAGAAGAGAC <b>CGCG</b> -3' 3'-TTCTTTCTTCTCT <b>GCGC</b> -5'

The duplex stability is determined by a number of structural factors such as the sequence length and base composition.<sup>12</sup> The extension of the duplex component of the triplex was designed with

T:A and A:T base-pair extension (AA-t1\_T/A), AA-t1-T, AA-t1-TA, AA-t1-TAT, AA-t1-TATA; and with G:C and C:G base-pair (AA-t1\_C/G), AA-t1-C, AA-t1-CG, AA-t1-CGC, AA-t1-CGCG. The sequences of the triplexes used in this study are shown in Table 3.1.

### 3.3.1 Thermal stability of DNA triplexes by $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$

Figure 3.2 shows the melting profile of AA-d1 and AA-t1, each with the racemic mixture of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ ,  $\Delta$ -enantiomer, and  $\Lambda$ -enantiomer added after annealing to ensure the binding of the TFO. The melting profile of AA-d1 in the presence of  $\Lambda$ -Ru shows increased melting stability with the melting temperature increasing from 45.5 °C without intercalator to 48.8 °C in the presence of the ruthenium complex (Table 3.2). A slightly increased melting temperature is also observable in the presence of  $\Delta$ -Ru with  $T_m$  47.7 °C. Melting profiles of DNA triplexes with rac-Ru show the contribution of both enantiomers in terms of an increased melting temperature of 49.7 °C. A similar outcome is demonstrated in the melting profile of the triplex DNA. Here, the melting profile of the duplex transition is consistent with AA-d1, although, in the presence of  $\Delta$ -Ru, there is little change in the  $T_m$  with 0.5 °C increment (Table 3.2). The triplex transition does not show a remarkable change with the significant decrease in melting temperature of -1.4 °C is observable with  $\Lambda$ -Ru (Table 3.3). The preference of  $\Lambda$ -enantiomer to stabilise triplexes more than the duplex was reported in studies of ruthenium polypyridyl complexes in RNA triplexes.<sup>13</sup>



**Figure 3.2** Normalised melting profile of AA-d1 (A) and AA-t1 (B). Samples were prepared with 20 mM sodium cacodylate buffer at ranges of pH from 6.5, 100 mM NaCl and 10 mM  $\text{MgCl}_2$  without Ru (II) complex (dashed line) and with rac-Ru (grey),  $\Delta$ -Ru (orange) and  $\Lambda$ -Ru (yellow) with [DNA triplex]:[Ru] ratio 1:1. The absorbance was recorded at  $\lambda = 260 \text{ nm}$ .

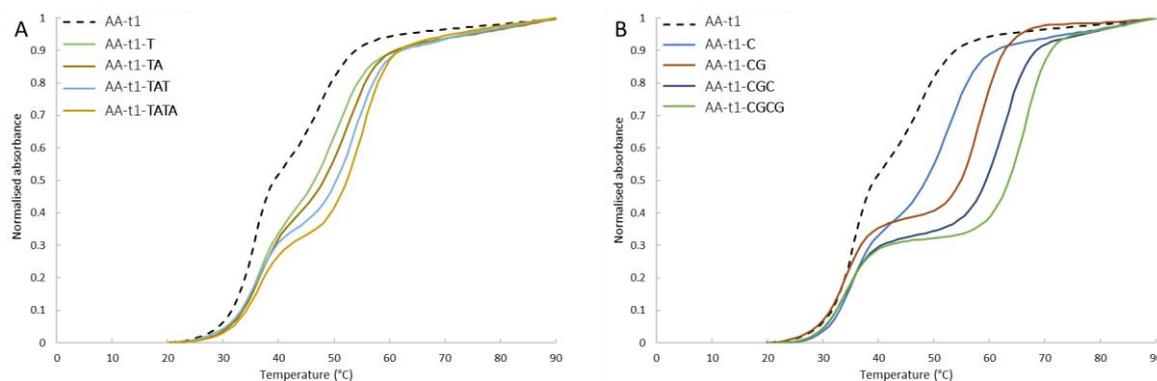
**Table 3.3.**  $\Delta T_m$  ( $^{\circ}\text{C}$ ) values of AA-d1 and AA-t1 (duplex transition and triplex transition) prepared with *rac*-Ru,  $\Delta$ -Ru, and  $\Lambda$ -Ru relative to the samples without ruthenium complexes.

DNA system		<i>rac</i> -Ru	$\Delta$ -Ru	$\Lambda$ -Ru
$T_m$ Duplex				
AA-d1	45.5	49.7	47.7	48.8
AA-t1	46	49.1	46.5	48.6
$T_m$ Triplex				
AA-t1	35	34.1	34.5	33.6

**Table 3.2** Melting temperatures ( $^{\circ}\text{C}$ ) of AA-d1 and AA-t1 (duplex transition and triplex transition) compared with samples prepared with *rac*-Ru,  $\Delta$ -Ru, and  $\Lambda$ -Ru.

DNA system	<i>rac</i> -Ru	$\Delta$ -Ru	$\Lambda$ -Ru
$\Delta T_m$ Duplex			
AA-d1	4.2	2.2	3.3
AA-t1	3.1	0.5	2.6
$\Delta T_m$ Triplex			
AA-t1	-0.9	-0.5	-1.4

The normalised melting absorbance for AA-t1 and triplexes AA-t1\_T/A and AA-t1\_C/G are shown in Figure 3.3. The hyperchromicity, the increase of absorbance of a DNA solution, is significant for the duplex section. The addition of the T:A and C:G bases affect the duplex stability, in the order AA-t1-T > AA-t1-TA > AA-t1-TAT > AA-t1-TATA, as well as AA-t1-C > AA-t1-CG > AA-t1-CGC > AA-t1-CGCG which is to be expected as a consequence of sequence extension. The difference in absorption increments for the duplex components is more evident for the AA/t1\_C/G system. This result is not surprising given that an increased content of G:C base pair leads to higher thermal stability compared to A:T base pairs (Table 3.3).



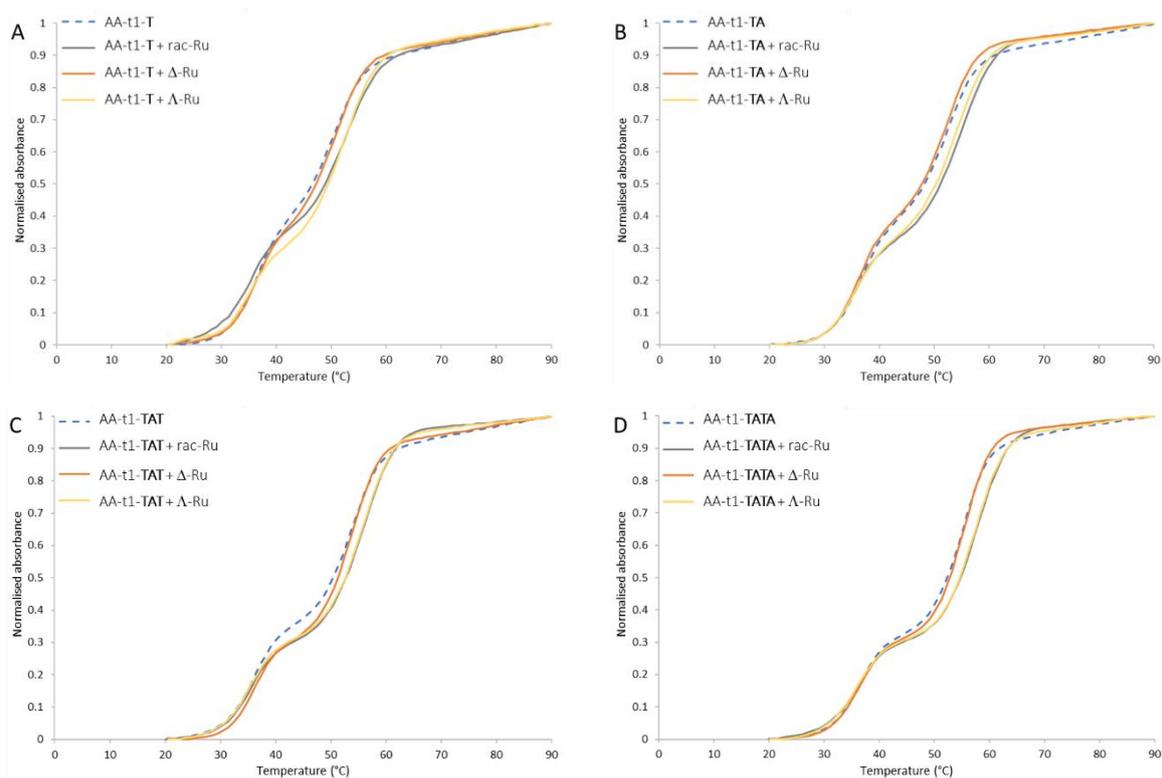
**Figure 3.3** Normalised melting profile of the duplex extended systems AA-t1\_T/A (A) and AA-t1\_C/G (B). AA-t1 is shown as dashed lines. The absorbance was recorded at  $\lambda = 260$  nm.

In regard to the triplex melting transition, AA-t1\_ **T/A** systems exhibit a triplex  $T_m$  with an average of 35.4 °C, compared to AA-t1\_ **C/G** triplexes in which the triplex component melts at slightly lower temperatures with  $T_m$  of ~33.7 °C. While the stability of the duplex with C/G extensions is more pronounced compared to the T/A extensions, the triplex result is more stabilised in the latter case with approximately a 1.7 °C difference between the two systems, thus showing that the duplex stability, in this case, does not affect the triplex formation.

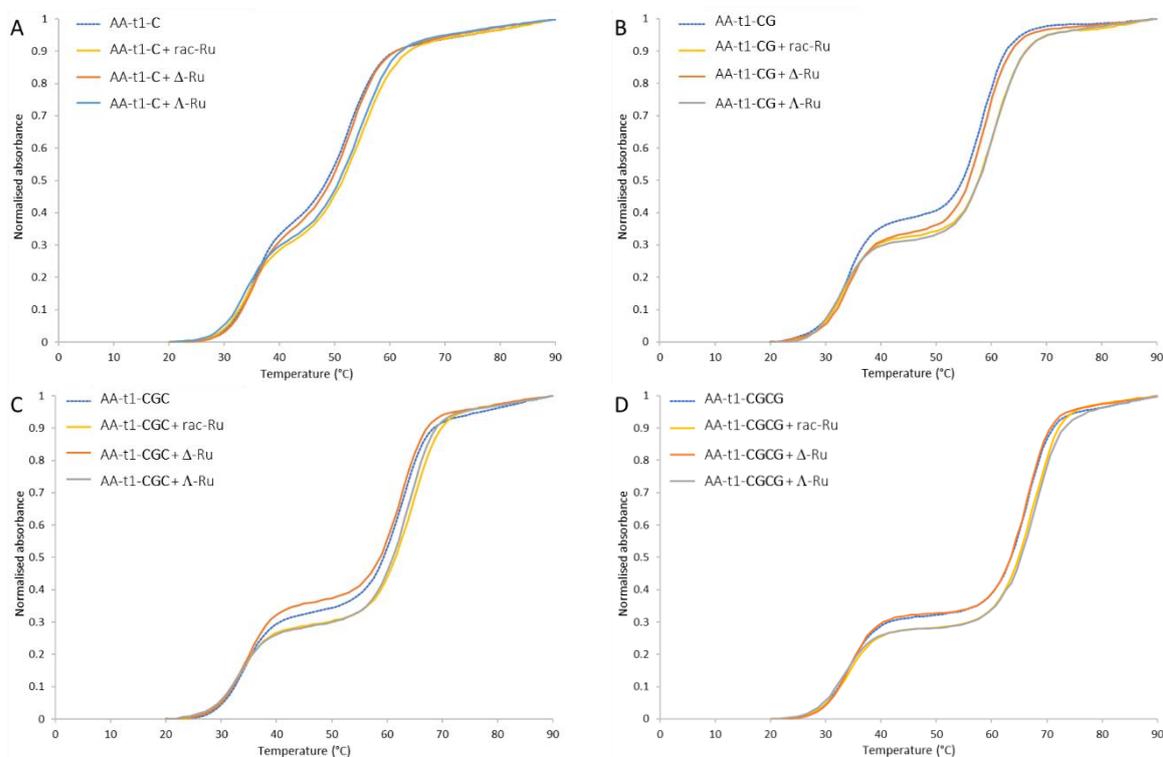
**Table 3.3** Temperature melting (°C) of the duplex and triplex transition for AA-t1\_ T/A and AA-t1\_ C/G systems.

DNA system	$T_m$ Duplex	$T_m$ Triplex
AA-t1_ <b>T</b>	49.5	35.5
AA-t1_ <b>TA</b>	52	36
AA-t1_ <b>TAT</b>	52.7	34.7
AA-t1_ <b>TATA</b>	54.6	35.5
AA-t1_ <b>C</b>	51.5	33.5
AA-t1_ <b>CG</b>	57.5	33.5
AA-t1_ <b>CGC</b>	62	34
AA-t1_ <b>CGCG</b>	65.7	33.8

To determine the binding interaction of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  with the extended AA-t1-**T/A** and AA-t1-**C/G** DNA triplexes thermal denaturation studies were performed (Figure 3.4-3.5). The metal complex was added with a ratio of 1:1 with DNA triplexes after annealing to prevent any disruption of the TFO binding after the addition of Ru(II) complexes to the solution.



**Figure 3.4** Normalised melting profile of the triplex systems AA-t1-T (A), AA-t1-TA (B), AA-t1-TAT (C) and AA-t1-TATA (D). Samples were prepared with a [DNA triplex]:[Ru] ratio of 1:1. Samples prepared without ruthenium are shown as dashed lines.



**Figure 3.5** Normalised melting profile of the triplex systems AA-t1-C (A), AA-t1-CG (B), AA-t1-CGC (C) and AA-t1-CGCG (D). Samples were prepared with a [DNA triplex]:[Ru] ratio of 1:1. Samples prepared without ruthenium are shown as dashed lines.

**Table 3.4** Temperature melting ( $^{\circ}\text{C}$ ) of the duplex transition for AA-t1, AA-t1\_ T/A and AA-t1\_C/G systems prepared with rac-Ru,  $\Delta$ -Ru, and  $\Lambda$ -Ru.

DNA system		rac-Ru	$\Delta$ -Ru	$\Lambda$ -Ru
$T_m$ Duplex				
AA-t1	46	49.1	46.5	48.6
AA-t1_T	49.5	51.6	49.7	51.9
AA-t1_TA	52	53.1	51.5	53.6
AA-t1_TAT	52.7	54.8	52.9	55.1
AA-t1_TATA	54.6	55.7	54.7	56.9
AA-t1_C	51.5	54.6	51.8	53.9
AA-t1_CG	57.5	59.6	57.7	58.9
AA-t1_CGC	62	63.1	61.5	62.6
AA-t1_CGCG	65.7	65.8	65.9	67.1

Compared to non-extended duplexes,  $\Lambda$ -Ru stabilises the duplexes component with a T/A extension, resulting in an increase in melting temperature of approximately  $2^{\circ}\text{C}$  (Table 3.4-3.5). A similar increase is observed in the presence of rac-Ru, whilst the  $\Delta$ -Ru has barely any effect on the duplex. Previous crystallographic studies show that the  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> enantiomer intercalates perpendicularly in the TA/TA base step. Hence, the increased melting stability corresponds to the addition of a TA base step, which is found in extended duplexes AA-t1-TA, AA-t1-TAT, and AA-t1-TATA.<sup>14</sup>

In the AA-t1-C/G triplex systems, the presence of rac-Ru results in a stabilisation of the duplex component with the AA-t1-C resulting in the most stabilised with a  $T_m$  increment of  $3.1^{\circ}\text{C}$ . The melting stability is mainly contributed by the  $\Lambda$ -enantiomer with a  $2.4^{\circ}\text{C}$  increment. In other cases,  $\Lambda$ -Ru stabilises the underlying duplex of AA-t1-CG and AA-t1-CGCG, while AA-t1-CGC is not affected by the presence of the metal complex. The overall  $T_m$  increment is gradually reduced as the number of C:G and G:C base steps are added.

**Table 3.5**  $\Delta T_m$  ( $^{\circ}\text{C}$ ) values of the duplex transition for AA-t1, AA-t1\_ T/A and AA-t1\_ C/G systems prepared with *rac*-Ru,  $\Delta$ -Ru, and  $\Lambda$ -Ru relative to the samples without ruthenium complexes.

DNA system	<i>rac</i> -Ru	$\Delta$ -Ru	$\Lambda$ -Ru
$\Delta T_m$ Duplex			
AA-t1	3.1	0.5	2.6
AA-t1_T	2.1	0.2	2.4
AA-t1_TA	1.1	-0.5	1.6
AA-t1_TAT	2.1	0.2	2.4
AA-t1_TATA	1.1	0.1	2.3
AA-t1_C	3.1	0.3	2.4
AA-t1_CG	2.1	0.2	1.4
AA-t1_CGC	1.1	-0.5	0.6
AA-t1_CGCG	0.1	0.2	1.4

Consistent with the triplexes without duplex extension, *rac*-Ru does not affect the stability of the extended triplexes (Table 3.6). Similarities can be shown also for  $\Delta$ -Ru where its effect on the triplex stability is not significant. The only exception is given by AA-t1\_C where the increase is of 1.2 $^{\circ}\text{C}$  (Table 3.7). The  $\Lambda$ -enantiomer instead has a lesser destabilising effect for AA-t1\_TAT, TATA, and CGCG compared to the other combinations which reported a similar temperature degree reduction. Intercalation of  $\Lambda$ -Ru could also bind within the C/G extended duplexes; however, it would require the flipping of the terminal bases as shown in crystal structures with another polypyridyl complex,  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$ .<sup>15</sup>

In general, the enantiomers contribute less to the triplex stability compared to the duplex counterpart arguing that the binding of the third strand partially disrupts the complex's intercalation, despite this being from the minor groove based on other photophysical studies.<sup>16</sup>

**Table 3.6** Temperature melting ( $^{\circ}\text{C}$ ) of the triplex transition for AA-t1, AA-t1\_ T/A and AA-t1\_ C/G systems prepared with *rac*-Ru,  $\Delta$ -Ru, and  $\Lambda$ -Ru.

DNA system		<i>rac</i> -Ru	$\Delta$ -Ru	$\Lambda$ -Ru
$T_m$ Triplex				
AA-t1	35	34.1	34.5	33.6
AA-t1_T	35.5	34.6	34.7	33.8
AA-t1_TA	36	35.1	35.4	34.6
AA-t1_TAT	34.7	34.9	35	34.1
AA-t1_TATA	35.5	35.6	35.7	34.9
AA-t1_C	33.5	33.6	34.7	31.8
AA-t1_CG	33.5	32.6	32.7	31.9
AA-t1_CGC	34	33	33.5	32.6
AA-t1_CGCG	33.8	32.9	33	33

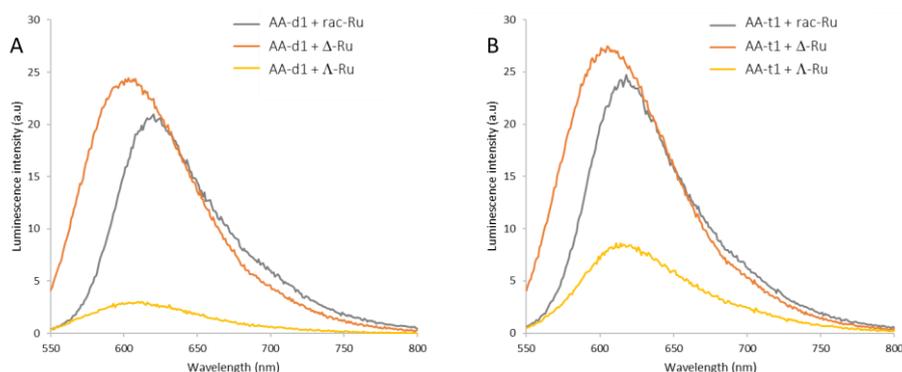
**Table 3.7**  $\Delta T_m$  ( $^{\circ}\text{C}$ ) values of the triplex transition for AA-t1, AA-t1\_ T/A and AA-t1\_ C/G systems prepared with rac-Ru,  $\Delta$ -Ru, and  $\Lambda$ -Ru relative to the samples without ruthenium complexes.

DNA system	rac-Ru	$\Delta$ -Ru	$\Lambda$ -Ru
	$\Delta T_m$ Triplex		
AA-t1	-0.9	-0.5	-1.4
AA-t1_ T	-0.9	-0.8	-1.7
AA-t1_ TA	-0.9	-0.6	-1.4
AA-t1_ TAT	0.2	0.3	-0.6
AA-t1_ TATA	0.1	0.2	-0.6
AA-t1_ C	0.1	1.2	-1.7
AA-t1_ CG	-0.9	-0.8	-1.6
AA-t1_ CGC	-1	-0.5	-1.4
AA-t1_ CGCG	-0.9	-0.8	-0.8

### 3.3.3 Determination of $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ binding in DNA triplexes

Fluorescence spectroscopy analysis was performed to identify the binding environment of the ruthenium complexes. Figure 3.6 shows fluorescence emission spectra for both AA-d1 and AA-t1 with  $\Delta$ -,  $\Lambda$ -enantiomers and the racemic mixture of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ . The concentration of the DNA was 5  $\mu\text{M}$  with a [DNA duplex] or [DNA triplex] ratio with [Ru] of 1:2. Emission experiments with ruthenium complexes on their own or in the presence of the TFO only were performed as controls and displayed negligible luminescence in aqueous solution (data not shown).

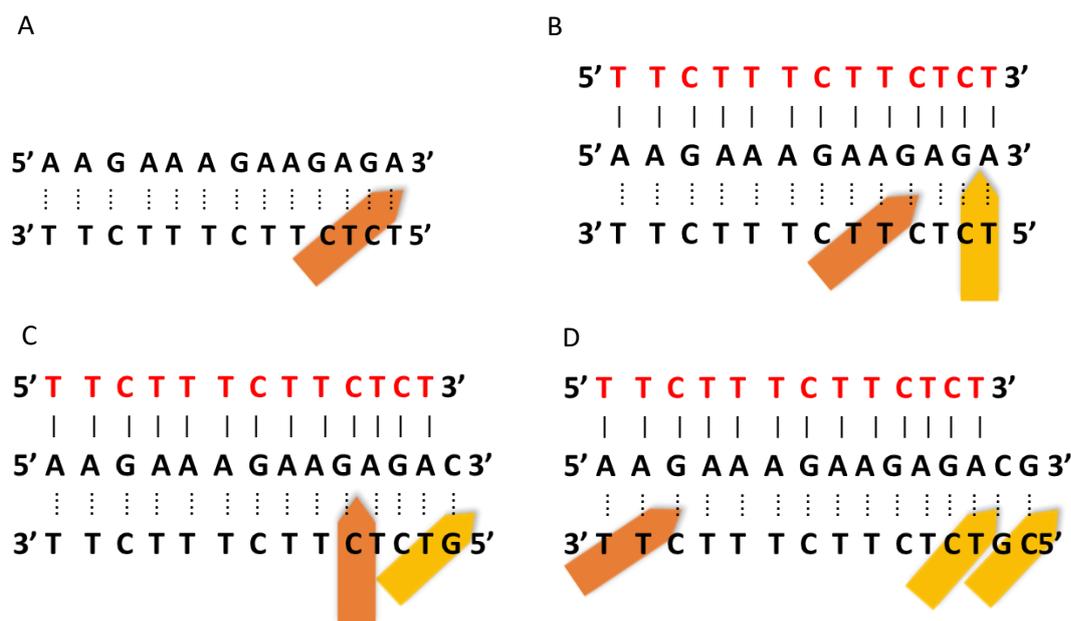
Emission experiment of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  bound to AA-d1 and AA-t1 revealed notable differences. The  $\Delta$ -enantiomer gave higher emission intensities in both the duplex and triplex structures at  $\lambda_{\text{max}} = 615$  nm, suggesting that the solvent can no longer quench the complex, independently of the binding of the TFO. In contrast, the  $\Lambda$ -enantiomer showed an approximately 2-fold enhancement in emission with the triplex over the duplex. This difference indicates that the TFO would provide additional binding sites to prevent solvent quenching of the complex.



**Figure 3.6** Fluorescence emission spectra of AA-d1 (A) and AA-t1 (B) with a ratio 2:1 with rac-Ru (grey),  $\Delta$ -Ru (orange), and  $\Lambda$ -Ru (yellow).  $\lambda_{\text{excit}} = 475$  nm

Since the thermal melting of the triplex with  $\Delta$ -Ru is enhanced, it would suggest that the complex intercalates in the minor groove of the triplex. Previous work has shown that an increased luminescence effect is attributed to DNA triplex structures, along with the increased stability of DNA triplexes, compared to the duplex DNA. This is explained by better protection of the planar dppz ligand from water quenching when intercalated into the triplets.<sup>10</sup>

Emission experiments with triplexes containing T/A and C/G duplex extensions exhibited comparable results to the triplex without duplex extension, with  $\Delta$ -Ru being more luminescent. This indicates that the solvent protection is given by the core triplex region rather than the extended duplex. It should be noted that the two ruthenium enantiomers may intercalate differently and shield distinct nitrogen groups of the dppz ligand, thus resulting in different emissions.<sup>17</sup> The crystallographic data available in the literature of the binding of  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> in the DNA molecule is not extensive, however, it is shown that  $\Delta$ -Ru is bound to the DNA molecule through an end-capping and semi-intercalation. The first required base flipping at the terminal base step A:T, and the second in a GG/CC base pairs.<sup>18</sup> Indeed, a complete intercalation is not required to emit luminescence, as the creation of a protective cavity is sufficient to shield the phenazine of  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> from the water.<sup>19</sup> While the AA-t1 sequences do not contain a GG step, the terminal base pairs A:T could accommodate the  $\Delta$ -enantiomer<sup>20</sup> (Figure 3.7). Recent data have also shown that  $\Delta$ -enantiomer of Ru (II) polypyridyl complexes exhibit a light switch effect when intercalated in DNA triplexes.<sup>21 22</sup>



**Figure 3.7** Potential binding sites of  $\Delta$ -Ru (orange), and  $\Lambda$ -Ru (yellow) in AA-d1 (A), AA-t1 (B), AA-t1-C and AA-t1-CG. Where in the presence of the complex the stability is not increased, it is suggested that a complete intercalation is not possible, thus are shown tilted.

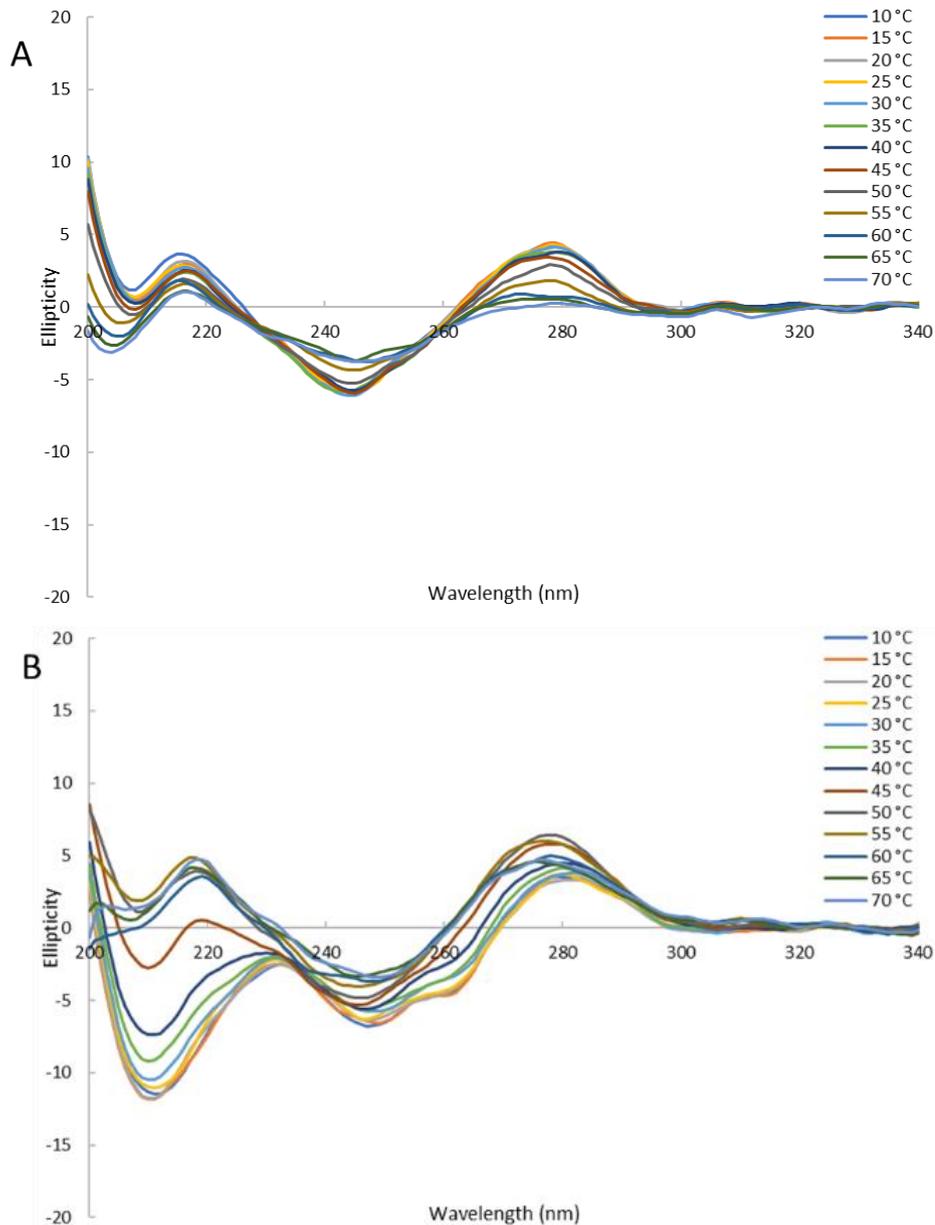
The  $\Lambda$ -Ru showed an increased fluorescence emission with the system AA-t1-C and AA-t1-CG compared to the other triplex systems with extended duplexes (Figure A3.5). This is consistent with the melting analysis where both systems showed a relative  $T_m$  increment of the duplex melting temperature. As the racemic mixture should contain both enantiomers, the fluorescence spectra of the racemic mixture result in a contribution of both  $\Delta$ -Ru and  $\Lambda$ -Ru. Indeed, the spectra shown here resemble previous fluorescence data.<sup>23</sup> Nevertheless, the fluorescence scan spectra of AA-t1-C and AA-t1-CG show a strong emission of the rac-Ru. The higher emission would translate into multiple complex binding to the DNA molecule, arguing that the binding of  $\Delta$ -Ru could potentially allow the stacking of more than one  $\Lambda$ -Ru complex. This is supported by the melting analysis of  $\Delta$ -Ru which shows the highest temperature increment for AA-t1-C compared to the other combinations. Therefore, it concludes that  $\Lambda$ -Ru has a preference for binding within the systems AA-t1-C and AA-t1-CG over the other combinations, without contributing to the triplex stability only for the AA-t1-C triplex (Figure 3.7), whilst the  $\Delta$ -Ru intercalates given the  $T_m$  increment.

### 3.3.4 Triplex formation upon $\text{Ru}(\text{phen})_2(\text{dppz})^{2+}$ binding using CD Spectroscopy

Circular dichroism measurements were carried out to investigate the structural conformation of DNA triplexes in solution. Further analysis of the triplex systems with  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  was carried out as well. CD measurements were run at wavelengths 200-340 nm at temperatures ranging from 10 °C to 70 °C with 5 °C increments to detect the transition of the triplex towards a duplex formation over increased temperature. Optically active chiral molecules, such as DNA and proteins, absorb right- and left-polarized light. The difference in absorption results in the ellipticity of the transmitted light. The degree of ellipticity as a function of wavelength provides information about the secondary structure and orientation of the chiral molecules in solution. While CD spectra of canonical duplex DNA structures have been extensively analysed, a comprehensive analysis of CD characteristics of DNA triplexes remains challenging. Factors such as sequence compositions and length, cation effects, and pH acidity can influence the CD spectra of a triple-helical structure.<sup>24,25</sup>

Prominent features are exhibited in the CD spectrum of a triple-helical structure compared to the double-stranded structures. The spectra region 200-320 nm is commonly assigned to the backbone changes of the structure and this region is sequence specific. Indeed, the split of the CD curves around 250 nm is characteristic of the sequences polypurine/polypyrimidine. Characteristic of a right-handed B-form helix, a large peak appears in the positive band in the region 270-280 nm indicative of the base stacking, instead the negative peaks in the region 240-250 nm are indicative of the DNA helicity. The CD region ranging from 210 nm to 240 nm is independent of the DNA sequence.<sup>26</sup> Negative peaks in this region are characteristic of the formation of DNA triplex, indicative of the presence of non-canonical base pairing and altered base stacking interactions within the triplex structure.<sup>27</sup> In Figure 3.8 are represented CD spectra for AA-d1 and AA-t1. Upon comparison of the two CD spectra, the negative intensity of the peaks in the range of 200-220 nm in the DNA triplex spectrum is more pronounced compared to the peaks in the duplex CD spectrum. Since the CD spectra of DNA triplexes have been recorded as a function of the temperature, it can be distinguished between the duplex formation at higher temperatures (approximately above 45 °C) and the triplex formation at lower temperatures. A correlation between the peak intensity and the temperature is indicative of the dissociation of the TFO, hence the presence of a double-stranded conformation at an increased temperature. The peaks in this re-

gion are decreased in amplitude, broadened, and shifted towards the longer wavelength confirming a structural configuration change as the temperature increases. Previous research confirming the transition duplex-triplex by altering the pH of the solution shows similar CD spectra.<sup>28</sup> Furthermore, the positive band at 280 nm and the negative at 250 nm have similar magnitudes, indicating that the TFO does not distort the overall structure.



**Figure 3.8** CD spectra of AA-d1 (A) and AA-t1 (B) at temperatures 10 °C to 70 °C with 5 °C increment.

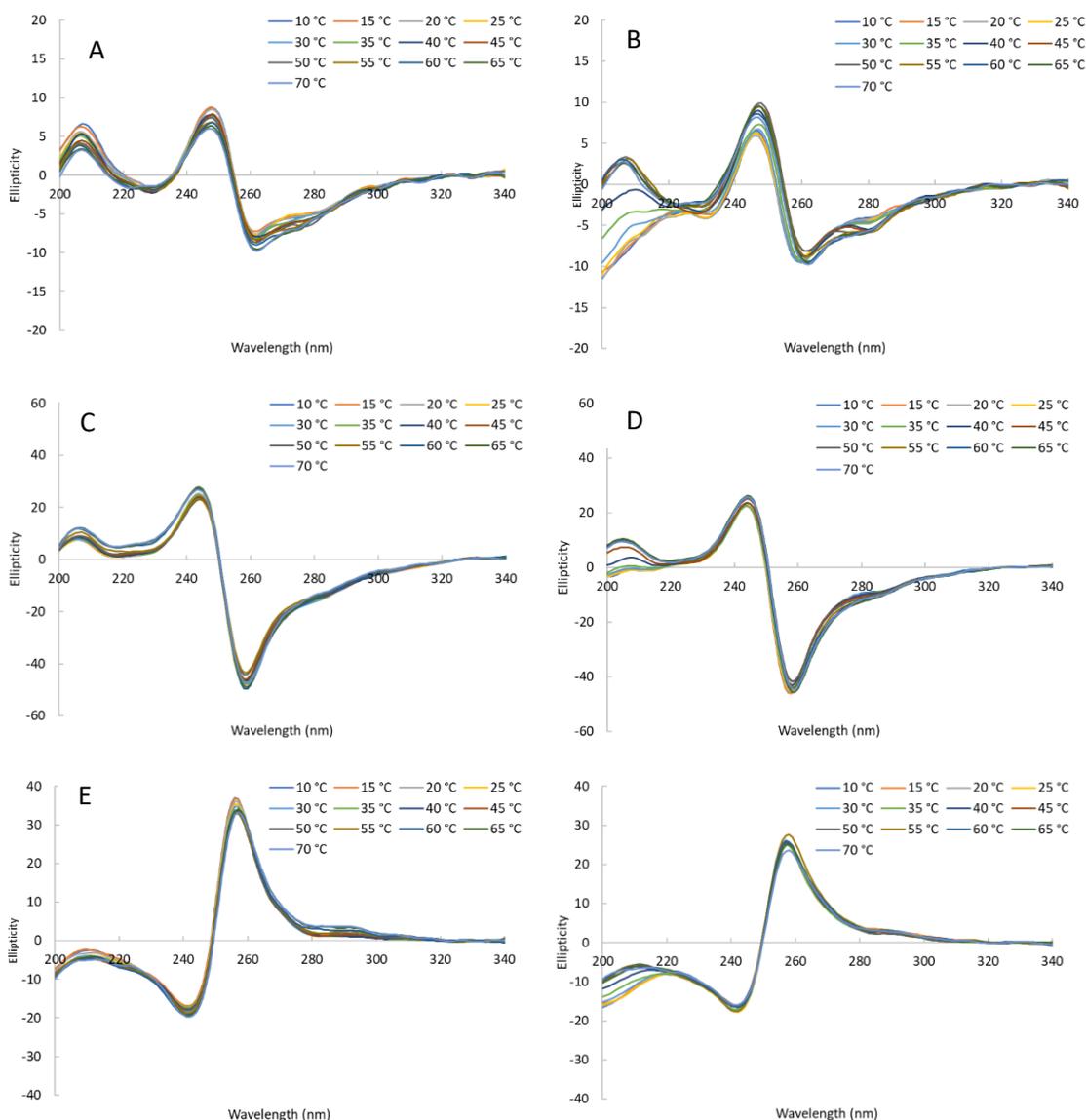
The CD spectra of DNA triplexes with the duplex extensions are shown in Figure A3.6. The peaks in the 200-220 nm region are informative of the triplex formation and confirm an altered base stacking interaction as the temperature increases. Both systems, the T/A and C/G duplex

extension, behave similarly under the right and left polarised light compared to AA-t1 triplex. However, spectra peaks in the 200-220 nm region remain negative at higher temperatures. This can be explained by the increased number of base pairs of the duplex structure.

The employment of CD showed the structural formation of the triple-helical structures as the TFO is present in the solution. The identification of the triplex signature in the 200-220 nm region is necessary to determine the DNA behaviour as ruthenium complexes are binding to the molecule. The CD spectra obtained in the presence of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  indicate a dominant influence of the complex on the observed signals. Isolating the contribution of the bound complex by subtracting the signal of the metal complexes is challenging due to the induced signals that arise from the chiral conjugation. Thus, comparisons are only feasible when examining each of the three combinations,  $\Delta$ -,  $\Lambda$ - and rac-Ru, individually in their DNA-binding interaction.

The interactions of the  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  with DNA triplexes investigated with circular dichroism spectroscopy are shown in Figures A3.6-A3.9. The absorption of the  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  shows a distinct CD spectrum indicative of the opposed chiral orientations of the enantiomers.<sup>29</sup> The two ruthenium complex enantiomers in the presence of the double-strand DNA displayed a CD spectrum with nearly perfect mirror profiles.  $\Delta$ -Ru exhibits a positive peak at 250 nm and a negative peak at 260 nm, whereas the  $\Lambda$ -Ru exhibits a negative peak at approximately 240 nm and a positive peak at 260 nm. Finally, the racemic mixture of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  shows an additive effect with a red shift of the peaks. In the three conditions, a peak is observed at 210 nm, with  $\Lambda$ -Ru being negative compared to the  $\Delta$ -Ru and rac-Ru.

Figure 3.9 shows the comparison between the CD spectra of the duplex and triplex in each of the enantiomer conditions and the combination of both. The intercalation of the ruthenium complexes in the DNA triplex structure causes variation mainly in the magnitude and wavelength of the peak at 210 nm when compared to their intercalation in the DNA duplex, confirming triple-helical formation.



**Figure 3.9** CD spectra of AA-d1 with rac-Ru (A),  $\Delta$ -Ru (C), and  $\Lambda$ -Ru (E); and AA-t1 with rac-Ru (B),  $\Delta$ -Ru (D), and  $\Lambda$ -Ru (F) at temperatures 10 °C to 70 °C with 5 °C increment.

The changes that occur by the intercalation of the ruthenium complex cause the unwinding of the structure resulting in the distortion of the positive peaks, particularly at  $\sim 250$  nm for the  $\Lambda$ -Ru for the triplexes AA-t1-C and AA-t1-CG; and  $\sim 260$  nm for the rac-Ru for triplex without duplex extension, AA-t1-T and all the AA-t11C/G triplexes. Drastic perturbation in this region is not visible in the triplexes prepared with  $\Delta$ -Ru. Compared to the effect of groove binding where the CD spectra are less perturbed, the intercalation of complexes alters the DNA molecule to accommodate

the dppz ligand between two base pairs, thus increasing the  $\pi$ - $\pi$  stacking.<sup>30</sup> This has been previously analysed for ruthenium complexes containing the dppz ligand intercalating from the minor groove.<sup>31</sup> This CD result confirms the intercalation in the same systems as previously proposed.

The intercalation from the minor groove decreases the elliptical spectra at 220 nm. This behaviour is visible in AA-d1 and AA-t1 in the presence of  $\Lambda$ -enantiomer, as well as in the extended duplexes in both enantiomers separately and mixed. The peaks are red-shifted at low temperatures, except for the triplex system AA-t1-**TATA** with the rac-Ru and  $\Lambda$ -enantiomer. Arguing that the ruthenium complex might intercalate without causing disruption of the hydrogen bonding formation (Figure A3.7 and Figure A3.9). It is suggested that the shorter and longer wavelength peaks arise from the intercalation of the complex from the major groove and minor groove, respectively.<sup>32</sup> Here, the redshifts of the peaks in the 220 nm region are reported for the  $\Lambda$ -enantiomer confirming that the intercalation is in the minor groove. On the contrary, a blue shift was observed with the  $\Delta$ -enantiomer that supports the intercalation through the major groove. Although a red shift is reported for the binding in the major groove of small complexes, there is no other structural result that confirms the major groove intercalation. Therefore, this phenomenon may be attributed to the distinct absorption properties of the  $\Delta$ - and  $\Lambda$ -enantiomers.

Finally, the location of the peaks in the positive region is indicative of the intercalation mode resulting parallel to the bases.<sup>32</sup> This is shown only for the racemic mixtures and the  $\Delta$ -enantiomer during the duplex formation. However, as previously shown both the phen or the dppz components of the  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  could interact differently with the DNA.<sup>19</sup>

### 3.4 Conclusions

Ruthenium polypyridyl complexes have attracted growing research attention as these metal complexes possess photoluminescence properties.<sup>33</sup> Particularly interesting is the intercalation of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  that display DNA mediated luminescence. Numerous studies have explored the light switch effect and binding intercalation in different DNA conformations, such as dsDNA, G-quadruplexes, *i*-motif, and DNA triplexes.<sup>34</sup> DNA triplexes with their sequence-specific recognition features, combined with the intercalation of the “photoswitch” compound have emerged as potential tools for light-responsive gene regulation and editing therapies.

In this chapter, the intercalation of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  in DNA triplexes was studied with three different spectroscopic techniques to obtain a structural description of the intercalation of polypyridyl complexes in DNA triplexes with additional duplex base pairs. The effect of the triplex stability dependent on the stability of the underlying duplex has been previously reported.<sup>35,36</sup> In this case great stability was not observed.

Additionally, the duplex DNA sections were extended to allow the metal complex to intercalate adjacent to triplex-forming region. UV-melts shows that  $\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  might not fully disrupt the triplex, however, significant stabilisation is also not obtained. This is further confirmed by the CD spectra that remain unperturbed upon intercalation of the complex. On the other hand, the high emission obtained for  $\Delta$ -Ru confirms the binding in a way that the nitrogen groups are protected from the water environment and that the protection is consistent throughout the systems. The only exception here is the increased melting temperature in the AAt-1-C, which along with the fluorescence results would support the intercalation in the minor groove in the C:G base pair as previously shown by crystallographic studies.<sup>18</sup> It is important to note that previous linear dichroism experiments confirm that the  $\Delta$ -enantiomer binds stronger and is more emissive than the  $\Lambda$ -enantiomer.<sup>37</sup>

The melting temperature increase of the extended duplex containing T/A extensions in the presence of the  $\Lambda$ -enantiomer would suggest that at higher temperatures, the complex has a preference for this site, as it is well known in DNA crystallography.<sup>38</sup> This is further confirmed by the redshift visible in the CD. This intercalation, however, has little effect on the triplex stability. Finally, when the triplexes are prepared with *rac*- $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  the increased melting stability is mostly due to the  $\Lambda$ -Ru component. Indeed, the CD spectra show structural perturbation for

AA-t1-**C** and AA-t1-**CG** in the presence of both the rac-Ru and  $\Lambda$ -Ru. This observation coupled with the increased luminescence of the rac-Ru in these systems and the increased melting stability in the presence of the  $\Delta$ -Ru would suggest a multiple complex interaction of the ruthenium complexes.

There is a growing interest in exploring metal intercalators that recognise and stabilize DNA triplex structures, as these compounds have the potential to be developed into gene targeting and editing tools.<sup>39</sup> Of particular interest are metal complexes conjugated to the ends of oligonucleotides to harness their unique photophysical and DNA binding properties and have shown promise for gene editing and antigene strategies.<sup>40</sup> As there is limited published research in this area, this study contributes to the literature by providing insights into ruthenium binding with DNA triplexes.

### 3.5 References

1. Komor, A. C. & Barton, J. K. The path for metal complexes to a DNA target. *Chemical Communications* **49**, 3617–3630 (2013).
2. Di Pietro, M. L., La Ganga, G., Nastasi, F. & Puntoriero, F. Ru(II)-dppz derivatives and their interactions with DNA: Thirty years and counting. *Applied Sciences* **11**, 1–22 (2021).
3. Olson, E. J. C. *et al.* First observation of the key intermediate in the ‘light-switch’ mechanism of [Ru(phen)<sub>2</sub>dppz]<sup>2+</sup>. *Chemtracts* **11**, 377–384 (1998).
4. Li, G., Sun, L., Ji, L. & Chao, H. Ruthenium(II) complexes with dppz: From molecular photoswitch to biological applications. *Dalton Transactions* **45**, 13261–13276 (2016).
5. Lerman, L. S. Structural considerations in the interaction of DNA and acridines. *J Mol Biol* **3**, 18–30 (1961).
6. Lincoln, P. & Nordén, B. DNA Binding Geometries of Ruthenium(II) Complexes with 1,10-Phenanthroline and 2,2′-Bipyridine Ligands Studied with Linear Dichroism Spectroscopy. Borderline Cases of Intercalation. *J Phys Chem B* **102**, 9583–9594 (1998).
7. Cardin, C. J. & Hall, J. P. Structural Studies of DNA-binding Metal Complexes of Therapeutic Importance. in *DNA-targeting Molecules as Therapeutic Agents* 198–227 (The Royal Society of Chemistry, 2018). doi:10.1039/9781788012928-00198.
8. Choi, S. D. *et al.* Binding mode of [ruthenium(II) (1,10-phenanthroline)<sub>2</sub>L]<sup>2+</sup> with poly(dT\*dA-dT) triplex. Ligand size effect on third-strand stabilization. *Biochemistry* **36**, 214–223 (1997).
9. Barton, J. K., Danishefsky, A. T. & Goldberg, J. M. Tris(phenanthroline)ruthenium(II): Stereoselectivity in Binding to DNA. *Proc. Natl. Acad. Sci. U.S.A* **106**, 2172–2176 (1984).
10. Jenkins, Y., Friedman, A. E., Turro, N. J. & Barton, J. K. Characterization of dipyridophenazine complexes of ruthenium(II): The light switch effect as a function of nucleic acid sequence and conformation. *Biochemistry* **31**, 10809–10816 (1992).
11. McQuaid, K. *et al.* Structural Studies Reveal Enantiospecific Recognition of a DNA G-Quadruplex by a Ruthenium Polypyridyl Complex. *Angewandte Chemie* **131**, 9986–9990 (2019).
12. SantaLucia, J., Allawi, H. T. & Seneviratne, P. A. Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability. *Biochemistry* **35**, 3555–3562 (1996).
13. Peng, M.-N., Zhu, Z.-Y. & Tan, L.-F. Binding Differences of Two Homochiral [Ru(bpy)<sub>2</sub>dppz]<sup>2+</sup> Complexes with poly(U)·poly(A)\*poly(U) Triplex RNA. *Inorg Chem* **56**, 7312–7315 (2017).
14. Niyazi, H. *et al.* Crystal structures of Λ-[Ru(phen)<sub>2</sub>dppz]<sup>2+</sup> with oligonucleotides containing TA/TA and AT/AT steps show two intercalation modes. *Nat Chem* **4**, 621–628 (2012).

15. Hall, J. P. *et al.* Structure determination of an intercalating ruthenium dipyridophenazine complex which kinks DNA by semiintercalation of a tetraazaphenanthrene ligand. *Proc Natl Acad Sci U S A* **108**, 17610–17614 (2011).
16. Tuite, E., Lincoln, P. & Nordén, B. Photophysical Evidence That  $\Delta$ - and  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> Intercalate DNA from the Minor Groove. *J Am Chem Soc* **119**, 239–240 (1997).
17. Cardin, C. J., Kelly, J. M. & Quinn, S. J. Photochemically active DNA-intercalating ruthenium and related complexes-insights by combining crystallography and transient spectroscopy. *Chem Sci* **8**, 4705–4723 (2017).
18. Hall, J. P. *et al.* Delta chirality ruthenium ‘light-switch’ complexes can bind in the minor groove of DNA with five different binding modes. *Nucleic Acids Res* **44**, 9472–9482 (2016).
19. Coates, C. G., McGarvey, J. J., Callaghan, P. L., Coletti, M. & Hamilton, J. G. Probing the interaction of [Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> with single-stranded DNA-what degree of protection is required for operation of the ‘light-switch effect’? *Journal of Physical Chemistry B* **105**, 730–735 (2001).
20. McKinley, A. W., Lincoln, P. & Tuite, E. M. Sensitivity of [Ru(phen)<sub>2</sub>dppz]<sup>2+</sup> light switch emission to ionic strength, temperature, and DNA sequence and conformation. *Journal of the Chemical Society. Dalton Transactions* **42**, 4081–4090 (2013).
21. Wen, B., Liu, X. & Tan, L. Binding and stabilizing effect of RNA triplex poly ( U ) · poly ( A )\* poly ( U ) by enantiomers of ruthenium ( II ) polypyridyl complex. *JBIC Journal of Biological Inorganic Chemistry* **28**, 509–517 (2023).
22. Wang, H., Liu, X. & Tan, L. Binding properties of a molecular “light switch” ruthenium(II) polypyridyl complex toward double- and triple-helical forms of RNA. *Int J Biol Macromol* **242**, 124710 (2023).
23. Hall, J. P. *et al.* X-ray crystal structure of rac-[Ru(phen)<sub>2</sub>dppz]<sup>2+</sup> with d(ATGCAT)<sub>2</sub> shows enantiomer orientations and water ordering. *J Am Chem Soc* **135**, 12652–12659 (2013).
24. Kypr, J., Kejnovská, I., Renčíuk, D. & Vorlíčková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res* **37**, 1713–1725 (2009).
25. Vorlíčková, M., Kejnovská, I., Bednářová, K., Renčíuk, D. & Kypr, J. Circular dichroism spectroscopy of DNA: From duplexes to quadruplexes. *Chirality* **24**, 691–698 (2012).
26. Pilch, D. S., Brousseau, R. & Shafer, R. H. Thermodynamics of triple helix formation: spectrophotometric studies on the d(A)<sub>10</sub>-2d(T)<sub>10</sub> and d(C + C + GGCC) triple helices. *Nucleic Acids Res* **18**, 5743–5750 (1990).
27. Khutsishvili, I., Johnson, S., Lee, H. T. & Marky, L. A. Unfolding thermodynamics of DNA intramolecular complexes involving joined triple- and double-helical motifs. *Methods Enzymol* **466**, 477–502 (2009).

28. Sugimoto, N., Wu, P., Hara, H. & Kawamoto, Y. pH and Cation Effects on the Properties of Parallel Pyrimidine Motif DNA Triplexes. *Biochemistry* **40**, 9396–9405 (2001).
29. Ardhammar, M., Lincoln, P., Rodger, A. & Nord E En, B. Absolute configuration and electronic state properties of light-switch complex [Ru(phen)<sub>2</sub>(dppz)]<sup>+</sup> deduced from oriented circular dichroism in a lamellar liquid crystal host. *Chem Phys Lett* **354**, 44–50 (2002).
30. Rehman, S. U., Sarwar, T., Husain, M. A., Ishqi, H. M. & Tabish, M. Studying non-covalent drug-DNA interactions. *Arch Biochem Biophys* **576**, 49–60 (2015).
31. Mårtensson, A. K. F. & Lincoln, P. Binding of Ru(terpyridine)(pyridine)dipyridophenazine to DNA studied with polarized spectroscopy and calorimetry. *Dalton Transactions* **44**, 3604–3616 (2015).
32. Nakabayashi, Y., Watanabe, Y., Nakao, T. & Yamauchi, O. Interactions of mixed ligand ruthenium(II) complexes containing an amino acid and 1,10-phenanthroline with DNA. *Inorganica Chim Acta* **357**, 2553–2560 (2004).
33. Li, G., Sun, L., Ji, L. & Chao, H. Ruthenium(II) complexes with dppz: From molecular photoswitch to biological applications. *Dalton Transactions* **45**, 13261–13276 (2016).
34. Sun, X. G., Cao, E. H., He, Y. J. & Qin, J. F. Spectroscopic comparison of different dna structures formed by oligonucleotides. *J Biomol Struct Dyn* **16**, 863–872 (1999).
35. Rusling, D. A., Rachwal, P. A., Brown, T. & Fox, K. R. The stability of triplex DNA is affected by the stability of the underlying duplex. *Biophys Chem* **145**, 105–110 (2009).
36. Shikiya, R. & Marky, L. A. Calorimetric unfolding of intramolecular triplexes: Length dependence and incorporation of single at → TA substitutions in the duplex domain. *Journal of Physical Chemistry B* **109**, 18177–18183 (2005).
37. Hiort, C., Lincoln, P. & Nordén, B. DNA Binding of Delta -and Lambda-[Ru(phen)<sub>2</sub>DPPZ]<sub>2</sub><sup>+</sup>. *J. Am. Chem. Soc* **115**, 3448–3454 (1993).
38. Niyazi, H. *et al.* parallel intramolecular DNA triple helix with G and T bases in the third strand stabilised by Xn<sup>2+</sup> ions". *Nat Chem* **4**, 621–628 (2012).
39. Fantoni, N. Z., Brown, T. & Kellett, A. DNA-Targeted Metallodrugs: An Untapped Source of Artificial Gene Editing Technology. *ChemBioChem* **22**, 2184–2205 (2021).
40. Flamme, M., Clarke, E., Gasser, G. & Hollenstein, M. Applications of ruthenium complexes covalently linked to nucleic acid derivatives. *Molecules* **23**, 28–31 (2018).

## CHAPTER 4- Intercalation of $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ in DNA triplex structure

### Contribution Statement

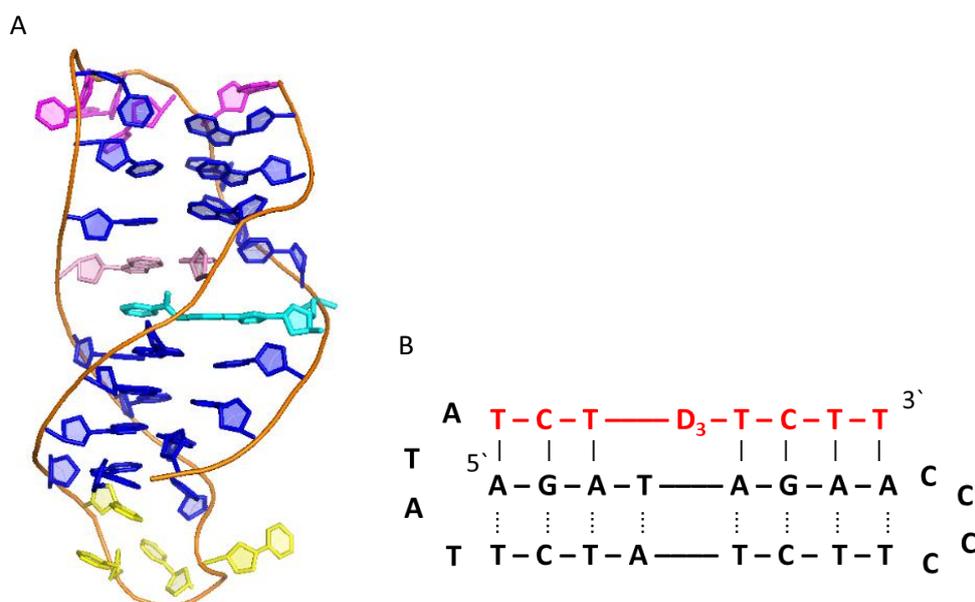
Structure **8PIP** was solved by Dr James Hall and Ahmad Abdullrahman.

**8PIP** was deposited in the Protein Data Bank with the entry title: DNA triplex structure with Polypyridyl Ruthenium Complexes.

## 4.1 Introduction

Numerous past efforts to crystallise native triplex structures resulted in partially formed triplexes. The closest existing structure shows a short triplex region as a junction between DNA region,<sup>1</sup> other X-ray structures yielded only a limited number of triplets. Extensive screening and optimization of the preliminary crystallisation conditions were conducted on the DNA triplex library illustrated in Chapter 2. As a result, a dark orange crystal appears resulting in the X-ray structure of the triplex AA-t6 d(AGATAGAACCCCTTCTATCTTATATCTTCTT).

The sequence AA-t6, including a non-canonical base D<sub>3</sub> in the TFO, was previously solved by NMR (PDB: **1WAN**), forming a stable intramolecular triplex with the polypyrimidine sequence 5'-TCTTCTT-3' binding the polypurine sequence 5'-AGAAGAA-3' in a parallel fashion. The non-canonical base D<sub>3</sub> includes a deoxyribose sugar, an imidazole moiety, and a benzamide group. The three aromatic rings of D<sub>3</sub> mimic a triplet and it's shown to insert between the T:A and T-A:T triplet through intercalation rather than standard hydrogen bonding (Figure 4.1).



**Figure 4.1** (A) NMR model of triplex **1WAN**. 7 triplets (blue), one duplex (pink), the non-canonical base D<sub>3</sub> (cyan), and the two loops CCCC (yellow) and TATA (magenta) are shown. The non-canonical base D<sub>3</sub> is shown in cyan. (B) Schematic representation of the oligonucleotide sequence is shown forming an intramolecular triplex with TFO in red.

The incorporated D<sub>3</sub> bases stack between the DNA bases without destabilising the structure, although local perturbations of helix winding are reported. The total number of triplets reported is seven, separated by the D<sub>3</sub> base and a WC base pair T<sub>4</sub>A<sub>17</sub>. The structure **1WAN** shows that the

three strands are connected by two short loops of different base compositions, 5'-CCCC-3' and 5'-TATA-3'.<sup>2</sup>

The objective of the proposed work was to crystallise the same NMR triplex forming sequence with the aim of solving a high-resolution X-ray structure. Since  $D_3$  is shown to stack like a standard nucleobase and sterically mimic a triplet, its removal was not expected to disrupt the formation of the triplex. The UV melts analysis from Chapter 2, indicates that a triplex formation is possible, though at lower temperatures. Additionally, the aim was to potentially incorporate  $[Ru(phen)_2(dppz)]^{2+}$  which has similarity with  $D_3$  as an aromatic DNA intercalator. Indeed, the dppz ligand in  $[Ru(phen)_2(dppz)]^{2+}$  can stack between DNA base pairs similar to the  $D_3$  heterocycle. Ruthenium polypyridyl complexes are reported to be sequence-specific,<sup>3</sup> as reported by  $D_3$ . Hence expecting the intercalation at the  $T_4:A_{17}$  base pair, or elsewhere within the triplex structure. Hence, co-crystallisation trials with the 31-mer oligonucleotide were set up with the racemic mixture of  $[Ru(phen)_2(dppz)]^{2+}$ . The 31-mer sequence did not crystallise as an intramolecular triplex conformation as was solved by NMR; instead, it formed a combination of intermolecular triplex and duplex structures. This result gives insight into the intercalation of *rac*- $[Ru(phen)_2(dppz)]^{2+}$  between triplex and duplex base pairs.

## 4.2 Materials and Methods

### 4.2.1 Oligonucleotide, buffer solution, and crystallisation reagents

Oligonucleotides were purchased from Eurogentec as HPLC-purified solid. All buffer materials and chemicals were purchased from Sigma Aldrich. Crystallisation screens were sourced from Hampton Research Comp and crystallisation materials were purchased from SWISSCI.

### 4.2.2 Crystallisation

Crystals containing the oligonucleotide AA-t6 d(AGATAGAACCCCTTCTATCTTATATCTTCTT) were grown via the vapour diffusion method from a sitting drop at 4 °C. The crystal was observed in a drop with a final volume of 400 nl from an optimisation plate prepared using the dispensing robot Oryx8, Douglas Instruments Screens Version 10.00. The drop contained 1.6 mM oligonucleotide AA-t6 20 mM Sodium Cacodylate pH 5, 10 mM MgCl<sub>2</sub>, 1.6 mM *rac*-[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> was mixed in a 1:1 ratio with the crystallisation solution containing 1.8 mM LiSO<sub>4</sub>, 50 mM tris-HCL pH 7.16, 1.4 mM CuCl<sub>2</sub> and 0.6 mM spermine. Orange rods grew 15 months after the plate was moved from 18 °C to 4 °C.

### 4.2.3 Data collection, refinement, and analysis

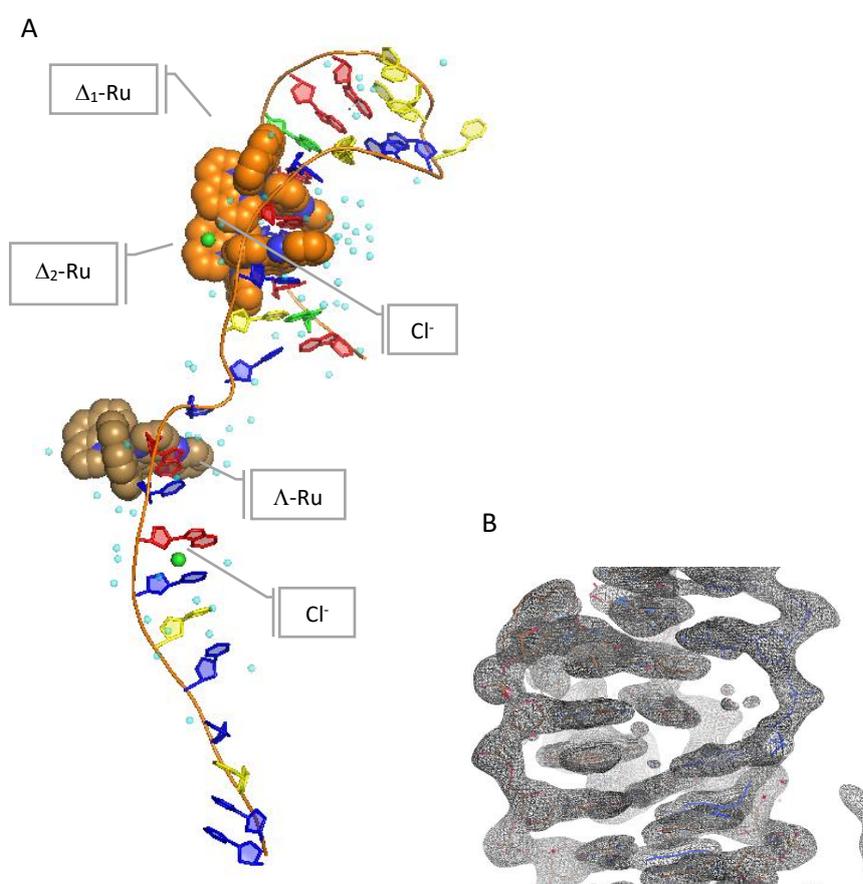
The data were collected on the in-house Rigaku Oxford Diffraction Gemini Ultra X-ray diffractometer. The data were collected using radiation with a Cu-wavelength at 1.54 Å at 100 K. The anomalous scattering of ruthenium was used by single-wavelength anomalous dispersion (SAD) methodology using the Phaser-EP pipeline in the PHENIX software package<sup>4</sup> to build an initial structure. The model, with a completeness of 98.71% was built using WinCoot<sup>5</sup> and refined using Phenix.refine from the PHENIX software package<sup>6</sup>, to give a final  $R_{work}$  0.2293 and  $R_{free}$  of 0.2585 with 5.05 % of reflections reserved for the  $R_{free}$  set. The structure was deposited in the publicly available database Protein Data Bank (PDB)<sup>7</sup> with the title DNA triplex structure with Polypyridyl Ruthenium Complexes, PDB ID **8PIP**. Data collection and refinement parameters are shown in Table 1. All derived parameters calculated using 3DNA<sup>8</sup> are available in Tables A4.1-4 in the appendix.

**Table 4.1** Crystallisation parameters and refinement statistics of **8PIP**.

<b>Crystallisation Parameters</b>	
DNA Sequence	d(AGATAGAACCCCTTCTATCTTATATCTTCTT)
Complex	<i>rac</i> -[Ru(phen) <sub>2</sub> (dppz)] <sup>2+</sup>
Crystallisation Temperature, °C	4
Growth Time (month)	15
Crystal Morphology	Rod
<b>Data Collection</b>	
Diffraction source	micro-focus sealed X-ray tube
Radiation wavelength, Å	1.541
Temperature, K	100
Exposure time, s	30
Detector	Hybrid Pixel Array Detector
Resolution, Å	1.8
<b>Data Processing</b>	
Structure Solution Method	SAD
Space group	<i>I</i> 2 2 2
a, b, c, Å	37.46, 86.15, 87.85
α, β, γ, Å	90, 90, 90
Resolution, Å	2.00 - 20.34 (2.00 - 2.05)
Total reflections	401872 (21622)
Unique reflections	9970 (717)
<i>R</i> <sub>meas</sub>	0.202 (3.732)
<i>R</i> <sub>merge</sub>	0.199 (3.67)
<i>R</i> <sub>pim</sub>	0.032 (0.677)
Mean <i>I</i> /σ	15.5 (1.0)
CC <sub>1/2</sub>	0.99 (0.84)
Completeness, %	99.9 (100)
Multiplicity	40.3 (30.2)
Average B factors, Å <sup>2</sup>	33.76
Data Collection Date	01 February 2023
*Outer shell statistics shown in parentheses	
<b>Refinement</b>	
No. Reflections	9886
<i>R</i> <sub>work</sub>	0.23
<i>R</i> <sub>free</sub>	0.26
<b>Number of components</b>	
Nucleotide	31
Ligands	3
Ions	2
Water	81
<b>rmsd</b>	
Bond Lengths, Å	0.004
Bond Angles, °	1.814

### 4.3 Results and Discussion

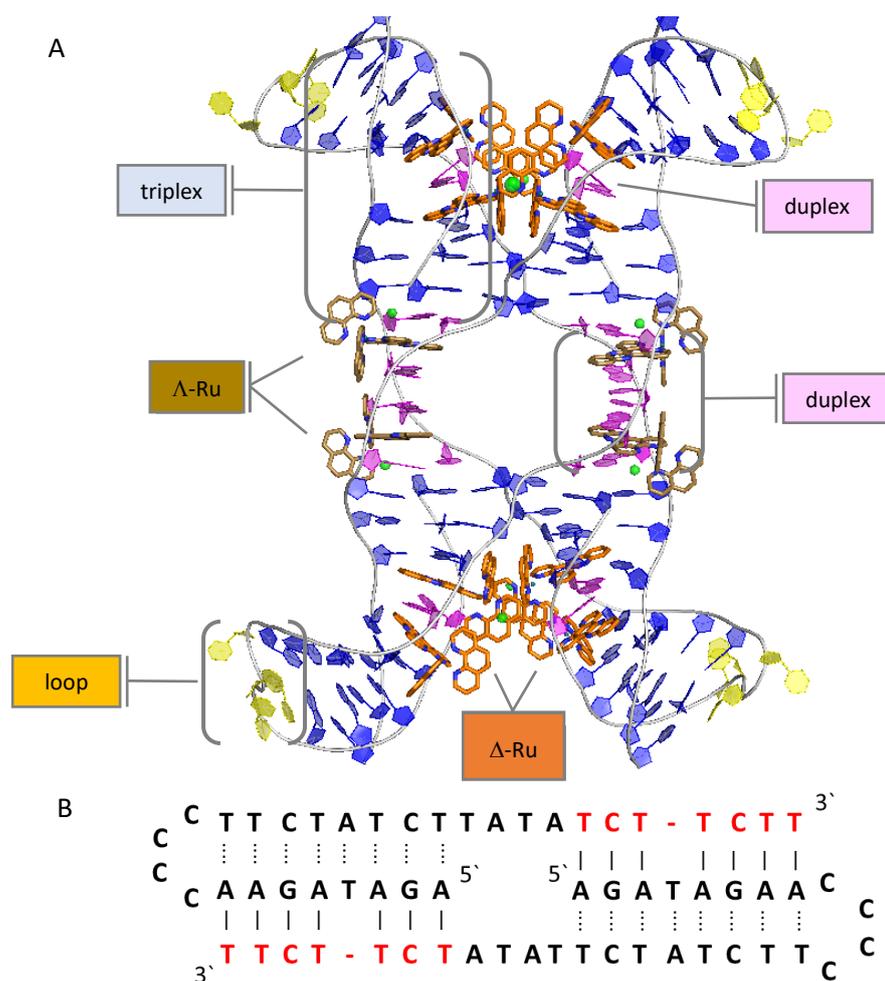
As a result of the optimisation of the screening trials, the oligonucleotide forming AA-t6 yielded a crystal structure of a hairpin structure with three  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  complexes. In the asymmetric unit, the oligonucleotide sequence adopted a hairpin conformation, with the intercalation of  $\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]$  and  $\Lambda$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]$  enantiomers, with a ratio of DNA: $\Delta$ : $\Lambda$  of 1:2:1. Additionally, the asymmetric unit two chlorine cations located around the metal complexes (Figure 4.2).



**Figure 4.2** (A) The asymmetric unit of **8PIP** contains one strand of  $d(\text{AGATAGAACCCCTTCTATCTTA-TATCTTCTT})$ , two  $\text{Cl}^-$  ions (green), two  $\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]$  (orange), one  $\Lambda$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  (brown) and 81 water molecules (cyan). (B) Representative example of electron density ( $2F_o - F_c$ ) is contoured at the  $1\sigma$  level (grey). Nucleobases adenine (red), guanine (green), thymine (blue), and cytosine (yellow) are coloured based on the nucleic acid database standard.

The asymmetric unit of **8PIP** contains a hairpin DNA stand with eight Watson-Crick (WC) base pairs. The remaining portion of the molecule 5'-TCTTCTT-3' does not self-fold but forms seven Hoogsteen base pairs, therefore part of this molecule acts as a TFO for a neighbouring hairpin (Figure 4.3). The TFO binds to a polypurine strand in a parallel fashion, however, the latter

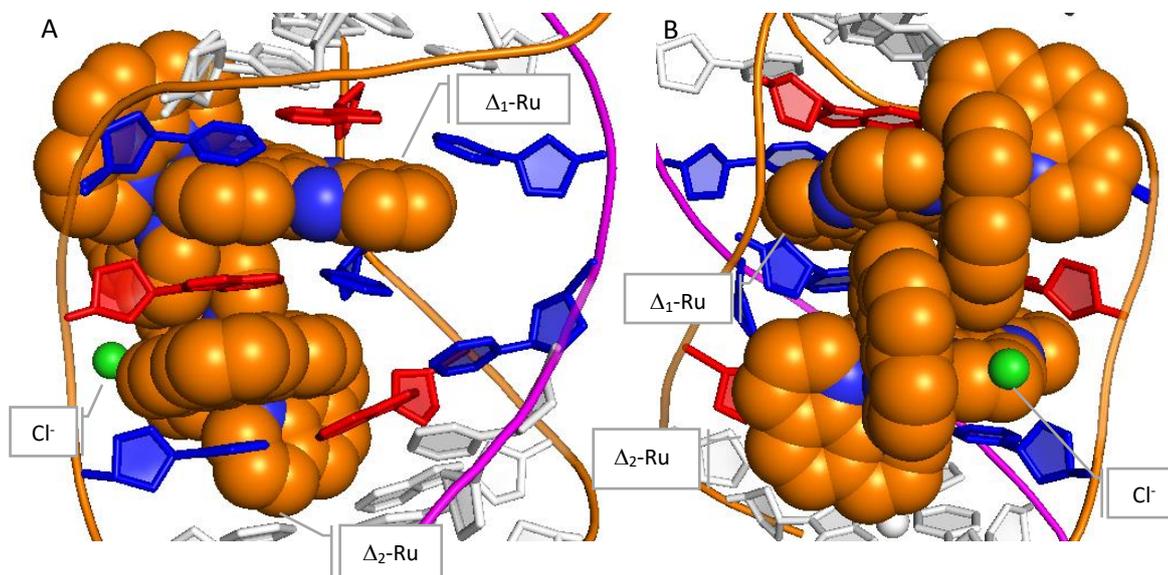
is interrupted by a thymine that can form WC bonds only with one adenine base inserted between the polypyrimidine strand of the underlying duplex strand, contrary to the purines that could donate and accept hydrogen bonds from two bases. Therefore, T<sub>4</sub> in the shown configuration is impeded from forming hydrogen bonds other than with A<sub>17</sub>. Overall, the crystallised molecule with its symmetry-related molecules presents a four-way junction conformation resulting in 16 WC base pairing, 16 unbound bases in the loop region, and the remaining 32 bases forming Hoogsteen and WC base pairing.



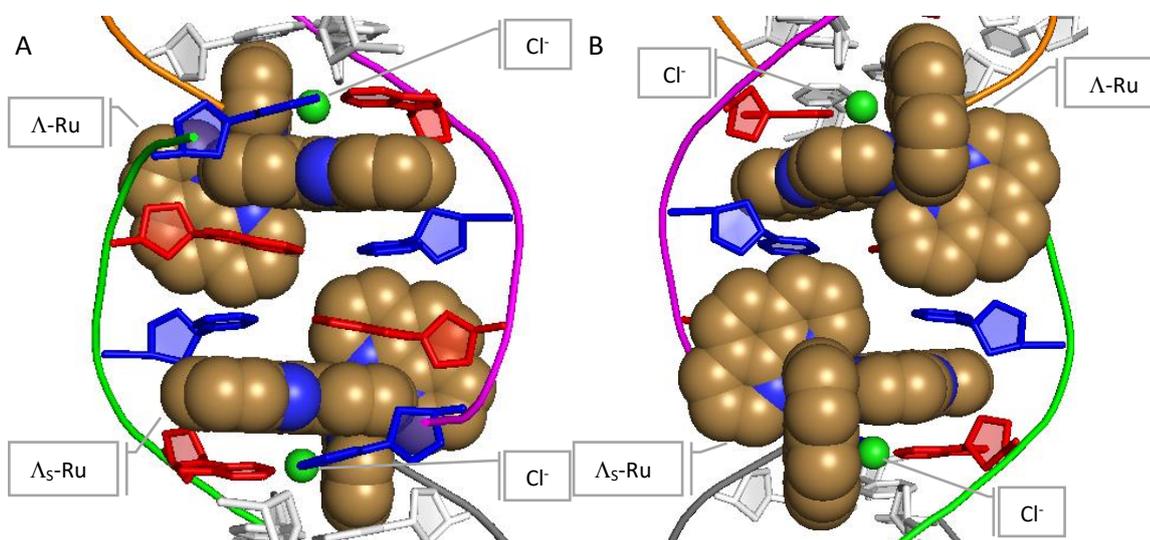
**Figure 4.3** Biological unit with other 3 symmetry-related molecules of **8PIP**. (A) 24 bases form Watson-Crick base pairing (magenta), 16 unbound bases in the loop (yellow), and 84 bases forming triplets (blue).  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)] (orange), and  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> (brown) are shown as well. (B) Schematic representation of the hairpin with the TFO sequence (red) forming an intermolecular triplex adjacent molecule.

The  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)] enantiomers stabilise this conformation by intercalating adjacent to triplet steps within the triplex region of the DNA (Figure 4.4). An additional symmetry-related mol-

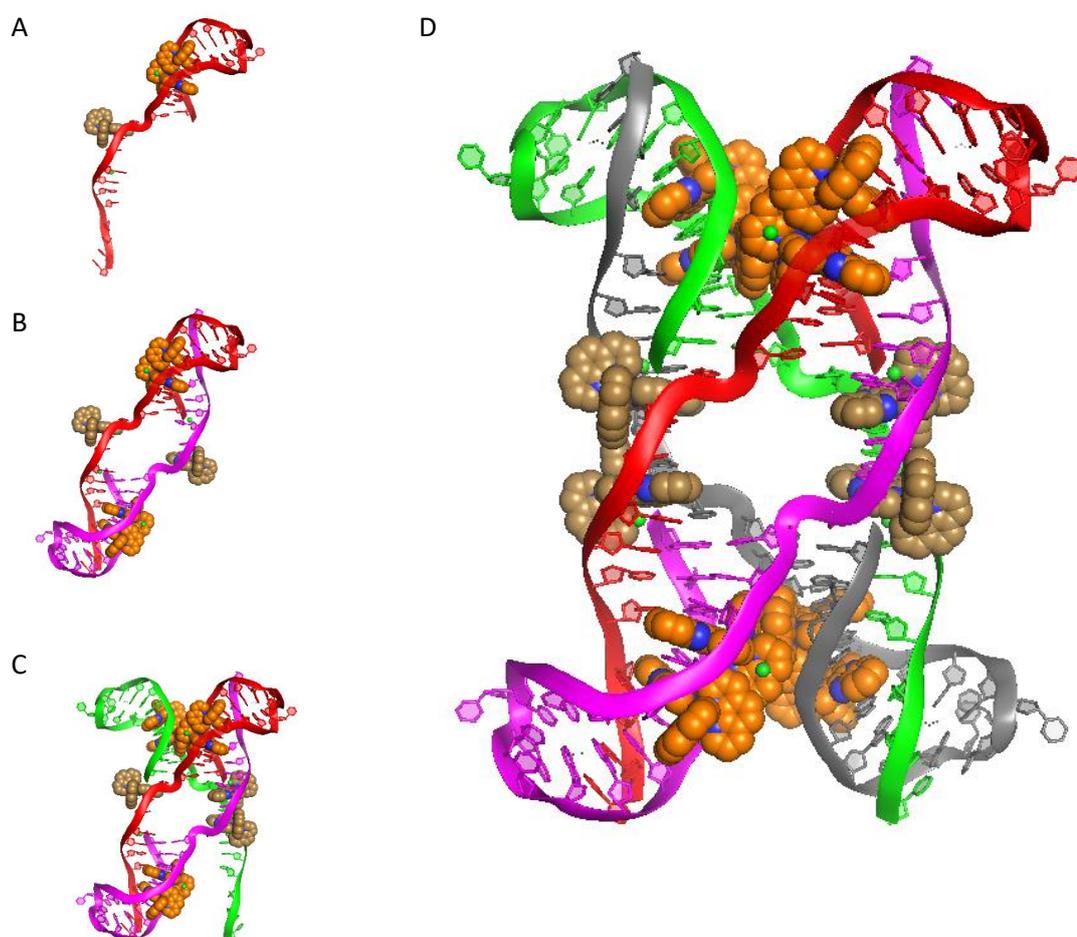
ecule contributes to the formation of WC base pairs intercalated by  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> complexes (Figure 4.5). Finally, the last symmetry molecule completes the binding with the other three molecules showing a four-way junction configuration (Figure 4.6).



**Figure 4.4** Front view (A) and back view (B) of the intercalation of the  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)] enantiomers between triplets. Adenine (red) and thymine (blue) coloured based on the nuclei acid database standard.



**Figure 4.5** Front view (A) and back view (B) of the intercalation of the  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)] enantiomers between duplexes. Adenine (red) and thymine (blue) coloured based on the nuclei acid database



**Figure 4.6** (A) The model **8PIP** (phosphate backbone in red) forms two sections of triplex bonding with the (B) first asymmetric unit (phosphate backbone in magenta). (C) The third crystallographic molecule (phosphate backbone in green) shows the Watson-Crick binding. (D) The fourth crystallographic molecule (phosphate backbone in brown) concludes the binding, resulting in a four-way junction.

The overall conformational analysis reveals that the hairpin structure **8PIP** forms with the neighbouring molecules two types of right-handed structures: four triplex and two duplex structures. The triplex region is composed of the same number of triplets (7) reported in the NMR structure **1WAN** with the twist angle in agreement with each other. Parameters independent of the backbone, such as rise, tilt, and roll have a degree of movement halved in the crystal structure (Table A4.3). While the rise discrepancy is justified by the  $\pi$ - $\pi$  stacking of the ruthenium complexes, the overall increased values reflect the dynamic nature of the **1WAN** molecule in solution. The overall twist degree of the helix rotation for the duplex, of the NMR and the crystal structures have an angle of  $32.13^\circ$  and  $27.37^\circ$ , respectively (Table A4.5). The  $\sim 4$ -degree difference is compensated by the intercalation of the TFO that introduces an additional twist. The average twist

degree of the polypurine-TFO with the polypurine-polypyrimidine, results in  $\sim 29.4^\circ$ . The reduced flexibility of the here crystallised triplex compared to the canonical DNA is consistent with both the structural analysis of NMR triplexes described in Chapter 1 and recent molecular simulations.<sup>9</sup>

Contrary to the results from NMR, the binding of the TFO to other hairpin molecules induces the TATA-loop to form WC base pairs with complementary strands, consequently giving rise to a duplex region (Figure 4.3). While the transition from hairpin to triplex is energetically favourable,<sup>10</sup> the tendency of the TFO to bind to neighbouring molecules, may be attributed to the intrinsic weakness of the Hoogsteen bonding, compared to WC bonds. Particularly, it is proposed that the transition from Hoogsteen bonding to WC requires purine bases to flip out of the TFO and undergo structural transition to then form back hydrogen bonding with the duplex.<sup>11</sup> This theory is further supported by altered atomic vibrations that participate in the bonding.<sup>12</sup> As the TFO undergoes conformational changes, the environment facilitates the formation of WC bonds further supported by the  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> in solution. These complexes have been previously demonstrated to exhibit sequence selectivity, contributing to the overall stability in this region.<sup>13</sup> Hence, it is proposed that the selectivity of  $\Lambda$ -Ru for T:A base pair has contributed to the structure stability without interfering in the formation of Hoogsteen bonds (Figure 4.3). As DNA loops have been investigated for their role in gene editing and transcription alteration,<sup>14</sup> it is worth considering their structural alteration as a result of forming highly ordered structures, such as triplexes.

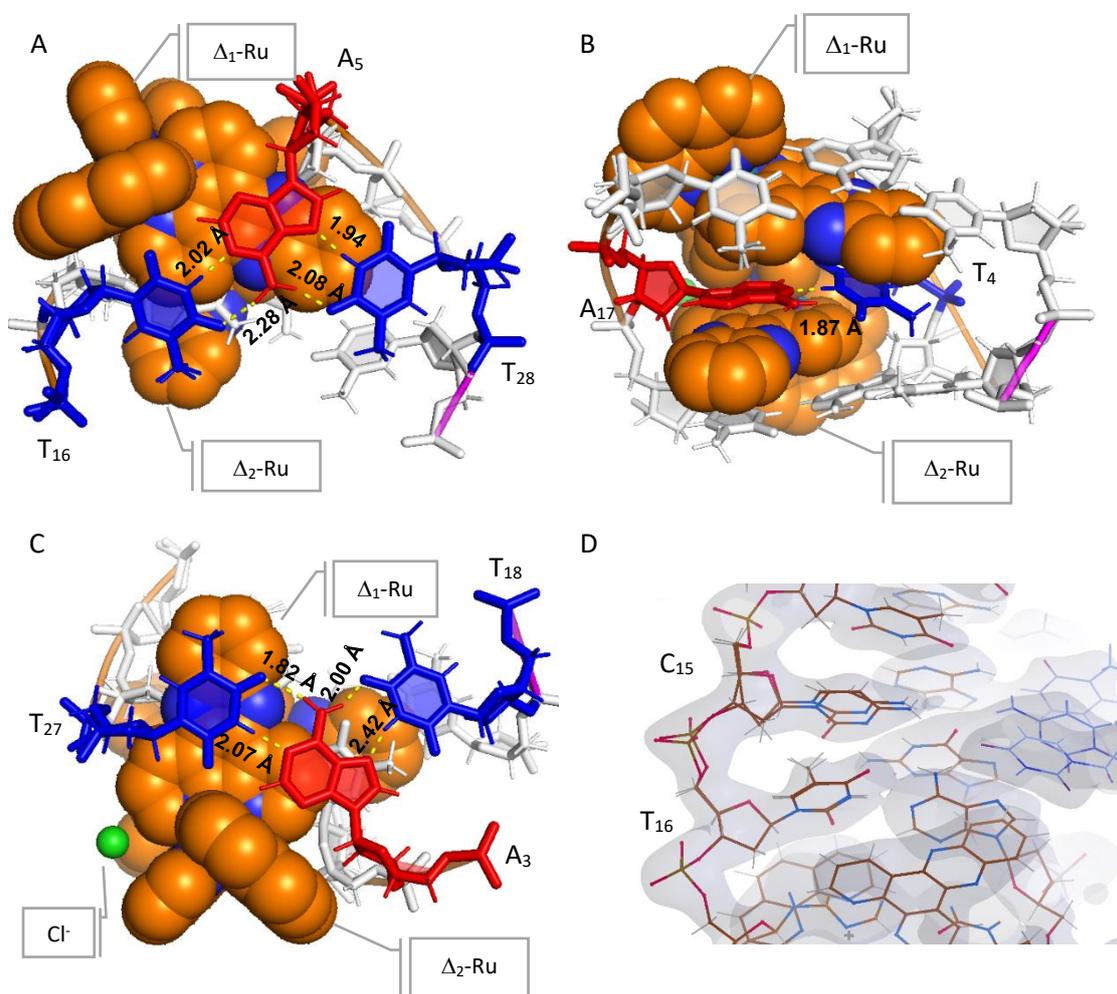
Compared to the modified TFO in **1WAN**, the removal of D<sub>3</sub> modification does not prohibit the triplex formation. This is due to the high flexibility nature of the DNA that allows the nucleotides T<sub>27</sub>T<sub>28</sub> to form Hoogsteen bonds with their respective adenine. The absence of the D<sub>3</sub> base between the two thymine gives a shorted phosphate backbone, hence the base step is required to rise to a higher degree of  $11.48^\circ$  and tilt to  $-50.45^\circ$  (Table A4.4). The intercalation of the TFO in the major groove of the duplex, results in the formation of other two grooves, each with an average width of 8 Å and 14 Å, with a minor groove of 13 Å. These values were obtained by subtracting 5.8 Å from the distance between the phosphate groups on opposing strands to take into account the van der Waals radii of the phosphate group<sup>15</sup>. The values obtained from the **8PIP** structure are consistent with the NMR model. Previous attempts to calculate the groove width in solution are inconsistent as variations are shown to be sequence-specific based on NOESY calculation,<sup>16</sup> however are generally confirmed to be lower than the canonical B-DNA.<sup>17,18</sup> Analysis of the sugar

pucker parameters, including the pucker amplitudes, phase angles, and the  $\delta$  torsion angle indicate that the triplex structure has an A-to-B conformational transition, with four base steps that have a pseudo rotation angle below  $40^\circ$ , including  $A_{22}:T_{23}$  forming WC base pairs (Figure A4.6). The average helical rise for the triplex is  $2.9 \text{ \AA}$  which is closer to the typical A-DNA ( $2.6 \text{ \AA}$ ). This was calculated by excluding the consecutive bases  $T_{27}$  and  $T_{28}$  which have an unusual rise of  $6.3 \text{ \AA}$  due to their formation of triplets with alternated base pairs. The explanation for the incongruity with the values reported for the NMR structure is given by stacking the ruthenium complexes. The helical twist shows similarity to A-DNA, with the exception of the widened  $T_{27}T_{28}$  step. And the  $\chi$ -displacement value of  $-1.87 \text{ \AA}$  is intermediate between A and B-DNA ( $-5.4 \text{ \AA}$  and  $-0.7 \text{ \AA}$ , respectively), this reflects the effect of the third strand binding. Overall, the triplex formed with the hairpin **8PIP** is a right-handed molecule with an A-to-B conformational transition caused by the presence of a TFO. The tendency of the triplex structure to adopt a configuration closer to A-DNA was confirmed by NMR analysis, showing that the polypurine strands of a DNA triplex assume an A-DNA conformation.<sup>19</sup> Conversely, studies interested in supercoiling show that the DNA bends in a B-DNA conformation.<sup>20</sup> Early crystal structure with the triplex-duplex junction shows similarity with both A-DNA and B-DNA<sup>21</sup>, however, it is clear now that DNA triplexes have unique structural configurations. The addition of this structure to the existing literature will contribute to understanding the triplex configuration, perhaps proposing a new form of DNA (T-DNA) as previously suggested.<sup>2</sup>

#### 4.3.1 The triplex region accommodates $\Delta$ -Ru enantiomers

The triplex region of the 4-way junction is composed of seven consecutive triplets divided by a duplex base pair. The two  $\Delta$ -Ru enantiomers,  $\Delta_1$ -Ru and  $\Delta_2$ -Ru, intercalate in the WC and Hoogsteen-WC base pairing with the neighbouring base pairs with the configuration: triplet- $\Delta_1$ -duplex- $\Delta_2$ -triplet (Figure 4.7). The base  $T_4$  does not have a corresponding base to pair in the TFO, thus forming WC base pair  $T_4:A_{17}$  between the two triplets which has the highest propeller value of  $24.29^\circ$  and an opening angle of  $19.74^\circ$  (Figure A4.1). All hydrogen bonding between donor and acceptor atoms of the nucleobases forming the triplets T-A:T and C-G:C have a distance between  $1.82 \text{ \AA}$  and  $2.96 \text{ \AA}$  (Figure 4.7). Hydrogen bonding is also preserved for nucleotides  $C_{15}$  and  $T_{16}$ ,

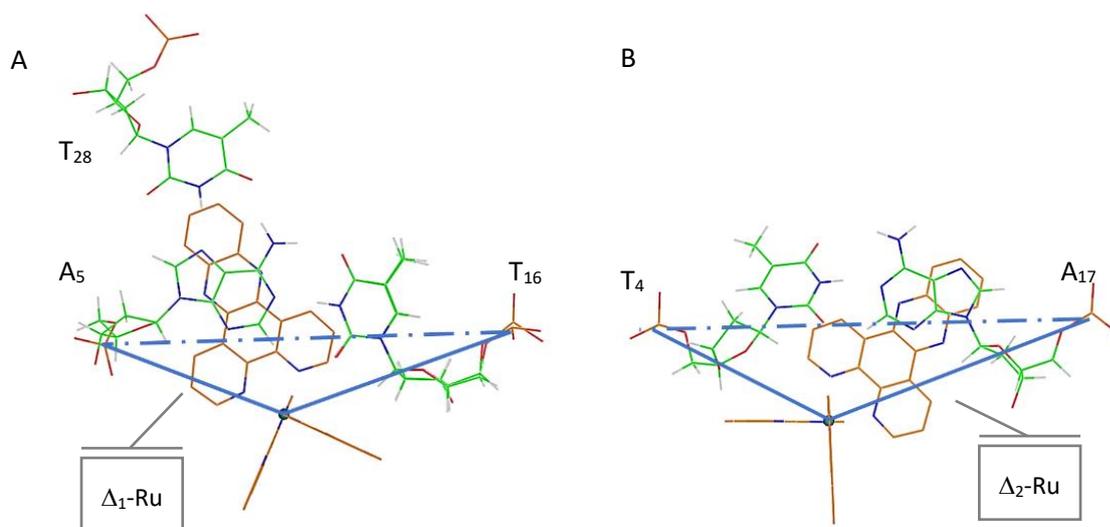
whose structures were fitted with multiple occupancies. The hydrogen bond distance corresponds to the analysis of the interatomic distances and interaction energies of triplets in the triplex.<sup>22</sup>



**Figure 4.7** View of hydrogen bonding of triplets (A, C) and WC base pair (B) with intercalation of  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup>. Adenine (red) and thymine (blue) coloured based on the nucleic acid database standard. (D) Fitting of C<sub>15</sub> and T<sub>16</sub> in multiple occupancies of electron density ( $2F_o - F_c$ ) is contoured at the  $1\sigma$  level (grey).

Two  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> complexes bind in the minor groove in adjacent base steps (Figure 4.4)  $\Delta_1$ -Ru is intercalated in T<sub>4</sub>A<sub>5</sub>/T<sub>16</sub>A<sub>17</sub> base step. The nucleotide A<sub>5</sub> forms two WC bonds with T<sub>16</sub>, buckled by 19.35°, and two Hoogsteen bonds with T<sub>28</sub>, resulting in base triplet with a buckle value of 12.83°.  $\Delta_2$ -Ru intercalates in the Watson-Crick base pair A<sub>3</sub>T<sub>4</sub>/A<sub>17</sub>T<sub>18</sub>, where A<sub>3</sub> forms Hoogsteen bonds with T<sub>27</sub> at a negative propeller parameter (-6.89°). The rise of 6.3 Å, almost double the canonical B-DNA, is due to the accommodation of the ruthenium complexes and this is consistent with previous crystal data.<sup>13</sup>

$\Delta$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> complexes cause extreme winding of the duplex with twist angles decreasing to 8.47° and 11.68° in the intercalating step site, to increase back to above 25°. Conversely, the TFO sequence remains unperturbed by the intercalation with a similar twist angle. The stability of the TFO allowing the formation of triple bases with the ruthenium complexes was previously confirmed by linear dichroism.<sup>23</sup> The intercalation mode of the two  $\Delta$ -Ru enantiomers from the minor groove in a canted orientation (Figure 4.8) is proposed as a binding mode in mismatched DNA and reported for other polypyridyl complexes, such as  $\Delta$ -[Ru(bpy)<sub>2</sub>dppz]<sup>2+</sup>.<sup>24-25</sup> The canted intercalation mode has been suggested for T-A:T sequences, primarily due to the enhanced emission observed in the light switch compound. In this specific configuration, the nitrogen atoms are shielded from water molecules, contributing to enhanced and long lifetimes emission.<sup>26</sup> The dppz axis of both  $\Delta$ -Ru enantiomers are twisted in opposite geometrical directions. The dppz ligand of  $\Delta_1$ -Ru is canted toward the P...P vector of T<sub>4</sub>A<sub>5</sub> by 62°, whilst the  $\Delta_2$ -Ru has an opposite direction, forming with its dppz vector a 64° angle with the P...P axis of A<sub>17</sub>T<sub>18</sub>. This orientation allows the dppz of  $\Delta_1$ -Ru to stack directly to the Hoogsteen base pair thus directly recognising triplet bases of the T-A:T triad.



**Figure 4.8** Canted intercalation of  $\Delta_1$ -Ru (A) and  $\Delta_2$ -Ru (B)

A variety of aromatic compounds with cationic charge such as acridines, anthraquinones, coralyne, and quinacridines have been shown to favour intercalation between the T-A:T triad, providing higher stability.<sup>27,28</sup> Another example is the aromatic base modification D<sub>3</sub> which protrudes from the major groove as part of a modified TFO and exhibits preference for intercalating in T-A:T over C-G:C.<sup>2</sup> The intercalation in T-A:T causes minimal perturbation to the triplex structure as the

formation of two hydrogen bonds is not dependent on the protonation, thus offering greater stability after intercalation. Ruthenium polypyridyl complexes have been also studied as intercalators in DNA triplexes, as well as RNA triplexes.<sup>29,30</sup> Investigation by linear and circular dichroism confirms the intercalation preference of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  for T-A:T DNA triplets.<sup>23</sup> Selectivity for this triad was also shown for the compound  $[\text{Ru}(\text{bpy})_2(11,12\text{-Br-dppz})]^{2+}$ <sup>31</sup>.

Recently, there has been an increased number of studies investigating the intercalation of various ruthenium complexes in RNA triplex structures. Notably, the  $\Delta$ -enantiomer has demonstrated a consistent effect to enhance the stability of U-A:U rich RNA triplexes.<sup>32–34</sup> The model **8PIP** further confirmed the preference of  $\Delta$ -Ru to bind in the T-A:T region over  $\Lambda$ -Ru provided to the same crystallisation solution. This study reveals the intercalation of ruthenium complexes in DNA triplexes, providing a unique example of an X-ray structure.

$\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  enantiomers are positioned  $\sim 1\text{\AA}$  offset from the P $\cdots$ P vector of the corresponding base step, with the  $\text{Ru}^{2+}$  cations approximately  $2.3\text{\AA}$  distant from the P $\cdots$ P axis. This configuration results in the distinctive stacking of the phen ligand between adjacent base pairs. Previous crystallographic results report phen ligand  $\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  stacking onto ancillary dppz group of  $\Lambda$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ , supported by Hoogsteen base pairs with the flipping base of the symmetry-related molecules.<sup>24</sup> Dppz-dppz stacking was also reported for  $\Delta$ - $[\text{Ru}(\text{phen})_2(11\text{-CN-dppz})]^{2+}$  with the phen ligand carrying a nitrile substituent.<sup>35</sup> The model **8PIP** shows the stacking of the phen ancillary groups of  $\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  stacking in complexes intercalated between adjacent base pairs, violating the Neighbour Exclusion Principle.

The neighbour exclusion model was mathematically calculated by McGhee and von Hippel in 1974 and later reviewed.<sup>36</sup> This principle states that intercalators cannot bind adjacent sites due to stereochemical constraints imposed by the sugar-phosphate backbone. Nevertheless, it was proposed that compounds, such as anthracene-spermine compounds can intercalate in adjacent base pairs from minor groove binding violating this principle.<sup>37</sup> Similarly, certain acridine ligands that bis-intercalate from both the minor and major groove violate the principle.<sup>38</sup> Other examples that show the violation of the Neighbour Exclusion Principle have been reported for naphthalene molecules in RNA, arguing that backbone differences in the RNA and DNA could play a role in accommodating planar molecules in adjacent bases.<sup>39,40</sup>

The intercalation limit was rationalised by the requirement of a C3' endo- C2'-endo pucker alternation to accommodate complex intercalation.<sup>41,42</sup> In the canonical duplex region of **8PIP**,  $\Lambda$ -Ru is

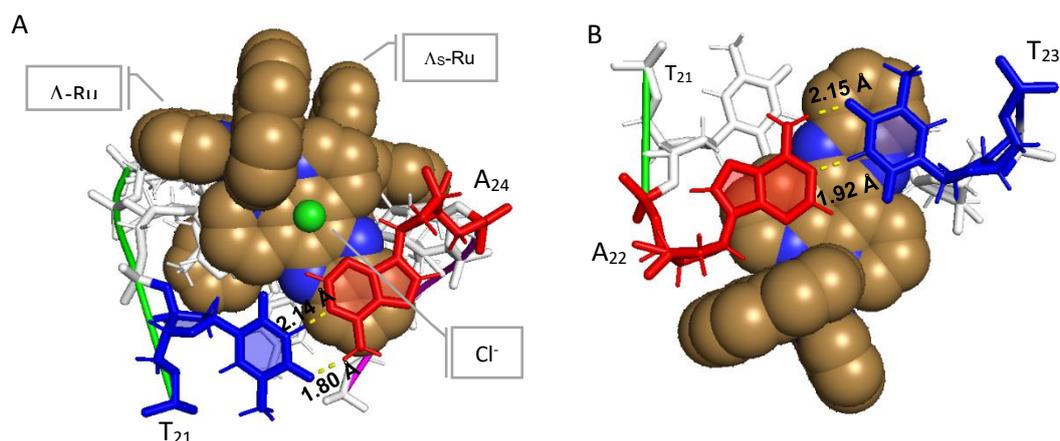
intercalated between WC base pairs that assume a C3'-endo- C2'-endo pucker configuration, thus agreeing with spectroscopic data. On the other hand, the base pairs stacking with  $\Delta_1$ -Ru and  $\Delta_2$ -Ru in the triplex region assume a C4'-exo and C1'-exo sugar configuration, except for the T<sub>4</sub> (C2'-endo) which forms a WC base pair between the two triplets (Figure A4.6). This incongruity is not unique to this intercalation; increasing structural data reveal intercalated bases assuming different stereochemical conformation. The flexibility of the sugar and phosphate backbone, as emphasised by dynamic modelling simulations, violates and adheres to the neighbour exclusion principle.<sup>43</sup> Hence, reaffirming the need to revisit the principle to include exceptions, as shown for the first time in this case.

In addition to complex-DNA stereochemical changes, factors that contribute to the intercalation of compounds in adjacent base pairs such as counterions effects, vibrational entropies, and other solvent interactions could not be used to rule out hypothetical intercalation in adjacent bases.<sup>43</sup> In the crystallised solutions the presence of counterions could have contributed to the minor groove intercalation between adjacent base pairs of the positively charged ruthenium complexes. More extensive research is needed to delineate the precise role of cations in modulating the neighbour exclusion principle.

#### 4.3.2 Intercalation of $\Lambda$ -Ru in Watson-Crick base pairs

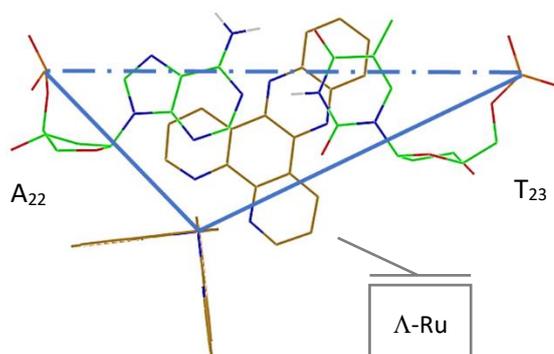
Examination of **8PIP** reveals notable distortion of the DNA backbone at the T<sub>21</sub> base. Such distortion allows the strand to bind with the adjacent molecule to form the TATA box that accommodates the  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup>. The transition duplex-triplex has been previously shown in the crystal structure **1D3R**, reporting similar strand behaviour.<sup>1</sup>

The  $\Lambda$ -Ru intercalation site is generated by a 2-fold symmetry element within the crystal, which then shows the second intercalating group,  $\Lambda_5$ -Ru.  $\Lambda$ -Ru intercalates into the T<sub>21</sub>A<sub>22</sub>/A<sub>24</sub>T<sub>23</sub> step, where T<sub>23</sub>:A<sub>24</sub> belongs to a symmetry molecule (Figure 4.9).



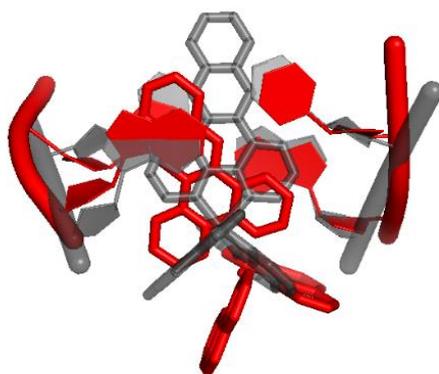
**Figure 4.9** View of T21:A24 base pair (A) and A22:T23 base pair (B) with intercalation of  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> Adenine (red) and thymine (blue) coloured based on the nuclei acid database standard, and  $\Lambda$ -Ru is shown in orange with nitrogen atoms in blue.

The long dppz axis of  $\Lambda$ -Ru lies on the P<sup>⋯</sup>P vector of A<sub>22</sub>:T<sub>23</sub> base pairs, with Ru<sup>2+</sup> 5.1 Å distant from the helical axis. The dppz axis is shifted 8.1 Å towards (P)T<sub>23</sub>A<sub>24</sub>, forming an angle of 117.54° with the P<sup>⋯</sup>P axis (Figure 4.10). Due to the intercalation of the ruthenium complex, the buckle of the base pairs T<sub>21</sub>:A<sub>24</sub> and A<sub>22</sub>:T<sub>23</sub> are 20.35° and -16.19°, respectively (Figure A4.1). The symmetry-related  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> point with its dppz ligand in opposite directions resulting in a canted orientation of the complex (Figure 4.7).



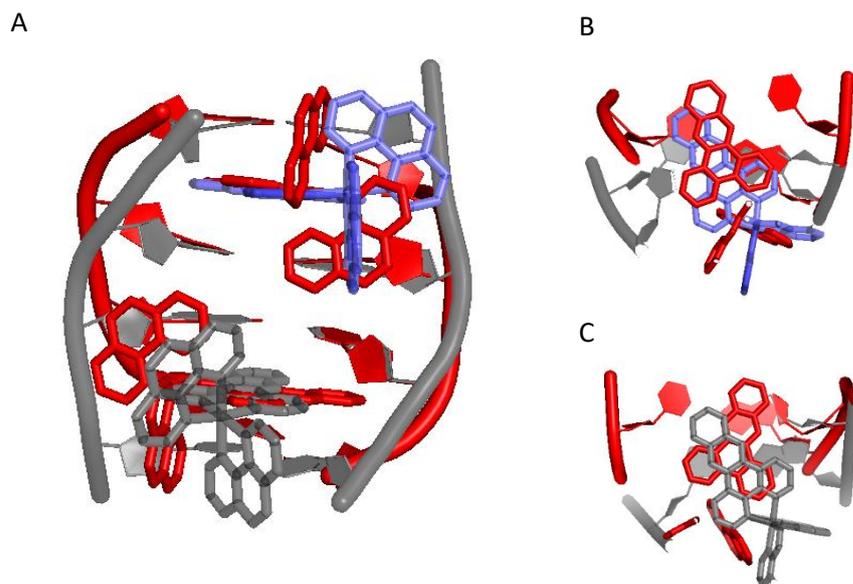
**Figure 4.10** Canted intercalation of  $\Lambda$ -Ru.

The preference of  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> for the TA/TA base step has been previously demonstrated in the **3U38** X-ray structure.<sup>13</sup> Figure 4.9 shows the superimposition of the TA/TA base step of both **8PIP** and **3U38**, showing their canted and perpendicular intercalation geometry of the  $\Lambda$ -enantiomer, respectively. Moreover, the local base pair twist is drastically reduced to nearly 15° compared to what has been reported for **3U38**. This result confirms that the  $\Lambda$ -Ru preferentially intercalates in TA/TA base step, yet multiple intercalation modes are possible.



**Figure 4.9** Superimposition of Intercalation of  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> in TA/TA step in structure **8PIP** (red) and **3U38** (grey). Rmsd < 1 Å.

Furthermore,  $\Lambda$ -Ru and  $\Lambda_S$ -Ru intercalate in adjacent base pair step TATA/TATA. The saturation of intercalation sites was previously confirmed using the intrinsic light switch emission of the complex.<sup>26</sup> The crystal structure **4JD8** shows an intercalation at every other base step with both  $\Lambda$ - and  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> enantiomers in the TG/CA and CA/TG base step, respectively.<sup>44</sup> (Figure 4.10). Upon superimposition of the two structures  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> enantiomers show both canted intercalation modes but with opposite geometries. Interestingly, the  $\Delta$ -Ru of **4JD8** is aligned on top of  $\Lambda$ -Ru, with the latter penetrating ~1 Å deeper.

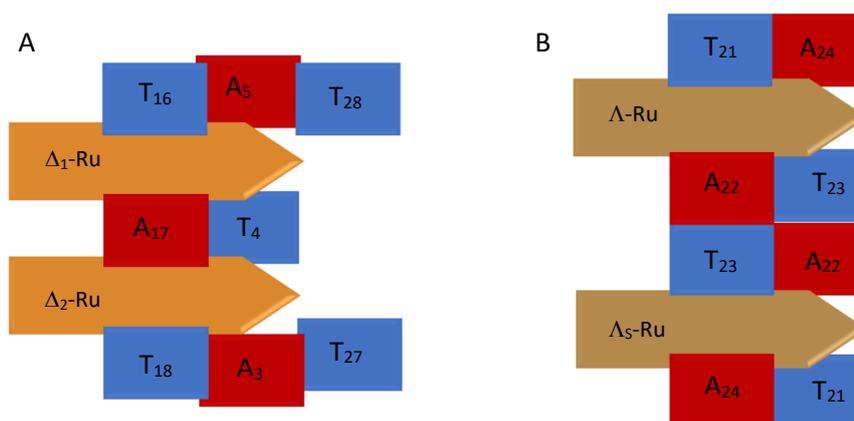


**Figure 4.10** (A) Superimposition of Intercalation of TA/TA step in model **8PIP** (red) and **4JD8** (grey) with its  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> shown in blue. (B) Superimposition of  $\Delta$ -Ru-**4JD8** (blue) and  $\Lambda$ -Ru-**8PIP**. (C) Superimposition of  $\Lambda$ -Ru-**4JD8** and  $\Lambda$ -Ru-**8PIP**. (rmsd < 1Å)

The structural configuration of the model **8PIP** shows that the intercalation of two  $\Lambda$ -Ru enantiomers separated by a DNA step is responsible for the stabilisation of the duplex region that links the triplexes in the structural configuration. Early study biophysical studies have considered the cooperative binding of multiple ligands within the same DNA molecule.<sup>45</sup> In fact, increased structural and thermodynamic evidence now supports the stability of cooperative intercalations.<sup>46</sup> Along with the discussed crystal result of the *rac*-[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> intercalating in adjacent base step ( $\Lambda$ -Ru- $\Delta$ -Ru),<sup>44</sup> this structure contributes to the literature by elucidating the intercalation of  $\Lambda$ -Ru- $\Lambda$ -Ru. Investigating the  $\Delta$ -Ru- $\Delta$ -Ru configuration would provide a comprehensive understanding of the complete set of enantiomers intercalation of [Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup> in DNA. The close intercalation of the complexes is further supported by the chloride ions that help to balance the overall charges.

## 4.4 Conclusion

The refined structure **8PIP** with its symmetries shows a complete DNA triplex with two  $\Delta$ -Ru enantiomers of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  complex intercalated within the triplets and two  $\Lambda$ -Ru between duplex base pairs (Figure 4.11). Notably, the two  $\Delta$ -Ru enantiomers are bound to adjacent steps in the triplex section of the molecule, in violation of the Neighbour Exclusion Principle.



**Figure 4.11** Schematic representation of the intercalation of the  $\Delta$ - $\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  enantiomers (orange) and  $\Lambda$ - $\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  (brown). (A)  $\Delta_1$ -Ru intercalates between triplex bonding and duplex bonding,  $\Delta_2$ -Ru between duplex bonding and triplex bonding. (B)  $\Lambda$ -Ru and  $\Lambda_5$ -Ru between WC base pairs. Thymine is shown in blue, adenine in red.

Interestingly, both enantiomers are intercalated from the sterically constricted minor groove in the TA ( $\Delta_1$ -Ru and  $\Lambda$ -Ru) and AT ( $\Delta_2$ -Ru) base step and not CG or GC. Atomic simulations suggest that the complex binds in the minor groove by reducing the total free energy and then inserting between the bases.<sup>47</sup> Even though the binding of the TFO would decrease the space in the minor groove<sup>22</sup>.

The electrostatic energy plays an important role in the intercalation of positively charged complexes, like polypyridyl complexes. Indeed, the binding strength is dependent on the type of intercalator, the position of the chromophores, and the DNA sequence.<sup>48</sup> In this context, the strong binding of  $\Delta$ -Ru<sup>49</sup> along with the base stacking preference of T-A:T, and the additional C-G:C triplets that contribute to the overall enthalpy,<sup>50</sup> could rationalise the intercalation in neighbouring base pairs. This is further supported by the T:A base pair between the triplets, potentially optimising the overall stacking interactions and enhancing the stability of the triplex.<sup>51</sup>

The groundbreaking X-ray structure provides structural insight into a DNA triplex with intercalated ruthenium polypyridyl complexes. No previous DNA-intercalator crystal structures have exhibited

violation of the neighbour exclusion principle. Consequently, an in-depth comparison with the NMR structures of triplexes is not feasible. Recognition of metal complexes of triplet bases, separated by a duplex base pair may guide the design of novel triplex-targeted binders capable of recognising insertion/deletion mutations to address genetic diseases. Previous studies have explored the recognition and deletion of point mutations by harnessing the target specificity of the TFO.<sup>52-54</sup> Overall, this work represents a major advancement in understanding the metal complex intercalation with DNA triplexes.<sup>57</sup>

The flexibility of the DNA molecule to fold and form multiple structures is further highlighted here. The transition to form a duplex region is promoted by the bending of the strand. A DNA feature that encompasses a whole other area of research.<sup>55,56</sup> Finally, the intercalation of  $\Lambda$ -Ru- $\Lambda$ -Ru adjacent in the TATA step could advance the research on intercalation of photo-therapeutic agents in the TATA-box which has biological implications as this sequence is found in the origin of replication.<sup>57,58</sup>

Overall, this novel crystal structure brings structural knowledge in the intercalation of complexes in a triplex conformation, highlights the flexibility of the DNA molecule and proposes novel binding specificities and geometries of ruthenium polypyridyl complexes, hence contributing to the ongoing research in gene editing technology.

## 4.5 References

1. Rhee, S., Han, Z., Liu, K., Miles, H. T. & Davies, D. R. Structure of a Triple Helical DNA with a Triplex–Duplex Junction. *Biochemistry* **38**, 16810–16815 (1999).
2. Wang, E., Koshlap, K. M., Gillespie, P., Dervan, P. B. & Feigon, J. Solution structure of a pyrimidine-purine-pyrimidine triplex containing the sequence-specific intercalating non-natural base D3. *J Mol Biol* **257**, 1052–1069 (1996).
3. Cardin, C. J., Kelly, J. M. & Quinn, S. J. Photochemically active DNA-intercalating ruthenium and related complexes-insights by combining crystallography and transient spectroscopy. *Chem Sci* **8**, 4705–4723 (2017).
4. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).
5. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
6. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).
7. Bernstein, F. C. *et al.* The protein data bank: A computer-based archival file for macromolecular structures. *J Mol Biol* **112**, (1977).
8. Li, S., Olson, W. K. & Lu, X.-J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res* **47**, (2019).
9. Pant, P. & Fisher, M. DNA triplex with conformationally locked sugar disintegrates to duplex: Insights from molecular simulations. *Biochem Biophys Res Commun* **532**, 662–667 (2020).
10. Völker, J., Botes, D. P., Lindsey, G. G. & Klump, H. H. Energetics of a Stable Intramolecular DNA Triple Helix Formation. *J Mol Biol* **230**, 1278–1290 (1993).
11. Vreede, J., Pérez de Alba Ortíz, A., Bolhuis, P. G. & Swenson, D. W. H. Atomistic insight into the kinetic pathways for Watson–Crick to Hoogsteen transitions in DNA. *Nucleic Acids Res* **47**, 11069–11076 (2019).
12. Strobel, J. L. & Scovell, W. M. Laser Raman spectroscopy of a complementary base pair in the Hoogsteen configuration. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **608**, 201–214 (1980).
13. Niyazi, H. *et al.* Crystal structures of  $\Lambda$ -[Ru(phen)<sub>2</sub>dppz]<sub>2</sub><sup>+</sup> with oligonucleotides containing TA/TA and AT/AT steps show two intercalation modes. *Nat Chem* **4**, 621–628 (2012).
14. Völker, J., Plum, G. E., Gindikin, V., Klump, H. H. & Breslauer, K. J. Impact of bulge loop size on DNA triplet repeat domains: Implications for DNA repair and expansion. *Biopolymers* **101**, 1–12 (2014).
15. El Hassan, M. A. & Calladine, C. R. Two distinct modes of protein-induced bending in DNA. *J Mol Biol* **282**, 331–343 (1998).

16. Asensio, J. L., Brown, T. & Lane, A. N. Solution conformation of a parallel DNA triple helix with 5' and 3' triplex–duplex junctions. *Structure* **7**, 1–11 (1999).
17. Shields, G. C., Laughton, C. A. & Orozco, M. Molecular Dynamics Simulations of the d(T·A·T) Triple Helix. *J Am Chem Soc* **119**, 7463–7469 (1997).
18. Thenmalarchelvi, R. New insights into DNA triplexes: residual twist and radial difference as measures of base triplet non-isomorphism and their implication to sequence-dependent non-uniform DNA triplex. *Nucleic Acids Res* **33**, 43–55 (2005).
19. Westhof, E. *et al.* NMR Studies of Triple-Strand Formation from the Homopurine-Homopyrimidine Deoxyribonucleotides d(GA)<sub>4</sub> and d(TC)<sub>4</sub>t. *Proc. Natl. Acad. Sci. U.S.A* **28**, 7859–7870 (1989).
20. Pyne, A. L. B. *et al.* Base-pair resolution analysis of the effect of supercoiling on DNA flexibility and major groove recognition by triplex-forming oligonucleotides. *Nat Commun* **12**, 1–12 (2021).
21. Rhee, S., Han, Z.-J., Liu, K., Miles, H. T. & Davies, D. R. Structure of a Triple Helical DNA with a Triplex-Duplex Junction. *Biochemistry* **38**, 16727–17000 (1999).
22. Peters, M., Rozas, I., Alkorta, I. & Elguero, J. DNA Triplexes: A Study of Their Hydrogen Bonds. *J Phys Chem B* **107**, 323–330 (2003).
23. Choi, S. D. *et al.* Binding mode of [ruthenium(II) (1,10-phenanthroline)<sub>2</sub>L]<sub>2</sub><sup>+</sup> with poly(dT\*dA-dT) triplex. Ligand size effect on third-strand stabilization. *Biochemistry* **36**, 214–223 (1997).
24. Hall, J. P. *et al.* Delta chirality ruthenium ‘light-switch’ complexes can bind in the minor groove of DNA with five different binding modes. *Nucleic Acids Res* **44**, 9472–9482 (2016).
25. Song, H., Kaiser, J. T. & Barton, J. K. Crystal structure of Δ-[Ru(bpy)<sub>2</sub>dppz]<sub>2</sub><sup>+</sup> bound to mismatched DNA reveals side-by-side metalloinsertion and intercalation. *Nat Chem* **4**, 615–620 (2012).
26. McKinley, A. W., Lincoln, P. & Tuite, E. M. Sensitivity of [Ru(phen)<sub>2</sub>dppz]<sub>2</sub><sup>+</sup> light switch emission to ionic strength, temperature, and DNA sequence and conformation. *Journal of the Chemical Society. Dalton Transactions* **42**, 4081–4090 (2013).
27. del Mundo, I. M. A., Vasquez, K. M. & Wang, G. Modulation of DNA structure formation using small molecules. *Biochim Biophys Acta Mol Cell Res* **1866**, 1–19 (2019).
28. Wilson, W. D., Mizan, S., Taniuos, F. A., Yao, S. & Zon, G. The interaction of intercalators and groove-binding agents with DNA triple-helical structures: The influence of ligand structure, DNA backbone modifications and sequence. *Journal of Molecular Recognition* **7**, 89–98 (1994).
29. Wang, H., Liu, X. & Tan, L. Binding properties of a molecular “light switch” ruthenium(II) polypyridyl complex toward double- and triple-helical forms of RNA. *Int J Biol Macromol* **242**, 124710 (2023).

30. Tan, L.-F. *et al.* First Ruthenium(II) Polypyridyl Complex As a True Molecular “Light Switch” for Triplex RNA Structure: [Ru(phen)<sub>2</sub>(mdpz)]<sup>2+</sup> Enhances the Stability of Poly(U)·Poly(A)\*Poly(U). *Inorg Chem* **51**, 4417–4419 (2012).
31. Wachter, E., Moyá, D. & Glazer, E. C. Combining a Ru(II) “Building Block” and Rapid Screening Approach to Identify DNA Structure-Selective “Light Switch” Compounds. *ACS Comb Sci* **19**, 85–95 (2017).
32. Tan, L., Wang, H. & Liu, X. Insight into achirality and chirality effects in interactions of a racemic ruthenium(II) polypyridyl complex and its Δ- and Λ-enantiomers with an RNA triplex. *Int J Biol Macromol* **219**, 579–586 (2022).
33. Peng, M.-N., Zhu, Z.-Y. & Tan, L.-F. Binding Differences of Two Homochiral [Ru(bpy)<sub>2</sub>dppz]<sup>2+</sup> Complexes with poly(U)·poly(A)\*poly(U) Triplex RNA. *Inorg Chem* **56**, 7312–7315 (2017).
34. Tan, L.-F., Xie, L.-J. & Sun, X.-N. Binding properties of Ru(II) polypyridyl complexes with poly(U)·poly(A)\*poly(U) triplex: the ancillary ligand effect on third-strand stabilization. *JBIC Journal of Biological Inorganic Chemistry* **18**, 71–80 (2013).
35. McQuaid, K. *et al.* Structural Studies Reveal Enantiospecific Recognition of a DNA G-Quadruplex by a Ruthenium Polypyridyl Complex. *Angewandte Chemie* **131**, 9986–9990 (2019).
36. Rocha, M. S. Revisiting the neighbor exclusion model and its applications. *Biopolymers* **93**, 1–7 (2010).
37. Rodger, A., Taylor, S., Adlam, G., Blagbrough, I. S. & Haworth, I. S. Multiple DNA Binding Modes of Anthracene-9-Carbonyl-N 1-Spermine. *Bioorg Med Chem* **3**, 861–872 (1995).
38. Rowell, K. N., Thomas, D. S., Ball, G. E. & Wakelin, L. P. G. Molecular dynamic simulations of diacridine binding to DNA: Indications that C6 diacridine can bisintercalate spanning two base pairs. *Biopolymers* **112**, 1–8 (2021).
39. Yousuf, M. *et al.* Violation of DNA neighbor exclusion principle in RNA recognition. *Chem Sci* **7**, 3581–3588 (2016).
40. Horowitz, E. D., Lilavivat, S., Holladay, B. W., Germann, M. W. & Hud, N. V. Solution structure and thermodynamics of 2',5' RNA intercalation. *J Am Chem Soc* **131**, 5831–5838 (2009).
41. Wilson, W. D. *et al.* Binding of Unfused Aromatic Cations to DNA. The Influence of Molecular Twist on Intercalation. *Annu. Rev. Biophys. Bioeng* **110**, 8292–8299 (1988).
42. Stockert, J. C. A neighbour-exclusion intercalating model for poly(ADP-ribose) binding to DNA. *J Theor Biol* **105**, 461–467 (1983).
43. Rao, S. N. & Kollman, P. A. Molecular mechanical simulations on double intercalation of 9-amino acridine into d(CGCGCGC)d(GCGCGCG): Analysis of the physical basis for the neighbor-exclusion principle. *Proceedings of the National Academy of Sciences* **84**, 5735–5739 (1987).
44. Hall, J. P. *et al.* X-ray crystal structure of rac-[Ru(phen)<sub>2</sub>dppz]<sup>2+</sup> with d(ATGCAT)<sub>2</sub> shows enantiomer orientations and water ordering. *J Am Chem Soc* **135**, 12652–12659 (2013).

45. Franklin, S. J., Treadway, C. R. & Barton, J. K. A Reinvestigation by Circular Dichroism and NMR: Ruthenium(II) and Rhodium(III) Metallointercalators Do Not Bind Cooperatively to DNA. *Inorg Chem* **37**, 5198–5210 (1998).
46. Biver, T., Cavazza, C., Secco, F. & Venturini, M. The two modes of binding of Ru(phen)2dppz2+ to DNA: Thermodynamic evidence and kinetic studies. *J Inorg Biochem* **101**, 461–469 (2007).
47. Mukherjee, A., Lavery, R., Bagchi, B. & Hynes, J. T. On the Molecular Mechanism of Drug Intercalation into DNA: A Simulation Study of the Intercalation Pathway, Free Energy, and DNA Structural Changes. *J Am Chem Soc* **130**, 9747–9755 (2008).
48. Medhi, C., Mitchell, J. B. O., Price, S. L. & Tabor, A. B. Electrostatic factors in DNA intercalation. *Biopolymers* **52**, 84–93 (1999).
49. Haq, I. *et al.* Interaction of Delta- and Lambda-[Ru(phen)2DPPZ]2+ with DNA: A Calorimetric and Equilibrium Binding Study. *J. Am. Chem. Soc* **117**, 4788–4796 (1995).
50. Soto, A. M. & Marky, L. A. Thermodynamic Contributions for the Incorporation of GTA Triplets within Canonical TAT/TAT and C+GC/C+GC Base-Triplet Stacks of DNA Triplexes. *Biochemistry* **41**, 12475–12482 (2002).
51. Soto, A. M., Loo, J. & Marky, L. A. Energetic Contributions for the Formation of TAT/TAT, TAT/CGC+, and CGC+/CGC+ Base Triplet Stacks. *Biopolymers* **124**, 14355–14363 (2002).
52. Chin, J. Y. *et al.* Correction of a splice-site mutation in the beta-globin gene stimulated by triplex-forming peptide nucleic acids. *Proceedings of the National Academy of Sciences* **105**, 13514–13519 (2008).
53. Félix, A. J., Ciudad, C. J. & Noé, V. Correction of the aprt Gene Using Repair-Polypurine Reverse Hoogsteen Hairpins in Mammalian Cells. *Mol Ther Nucleic Acids* **19**, 683–695 (2020).
54. Wang, Y. *et al.* Stable recognition of TA interruptions by triplex forming oligonucleotides containing a novel nucleoside. *Biochemistry* **44**, 5884–5892 (2005).
55. Werner, M. H., Gronenborn, A. M. & Clore, G. M. Intercalation, DNA Kinking, and the Control of Transcription. *Science (1979)* **271**, 778–784 (1996).
56. Hall, J. P. *et al.* Structure determination of an intercalating ruthenium dipyridophenazine complex which kinks DNA by semiintercalation of a tetraazaphenanthrene ligand. *Proc Natl Acad Sci U S A* **108**, 17610–17614 (2011).
57. Hunter, C. A. Sequence-dependent DNA Structure. *J Mol Biol* **230**, 1025–1054 (1993).
58. Yella, V. R. & Bansal, M. DNA structural features of eukaryotic TATA-containing and TATA-less promoters. *FEBS Open Bio* **7**, 324–334 (2017).

## CHAPTER 5 - Major groove intercalation of ruthenium polypyridyl complexes

### Contribution Statement

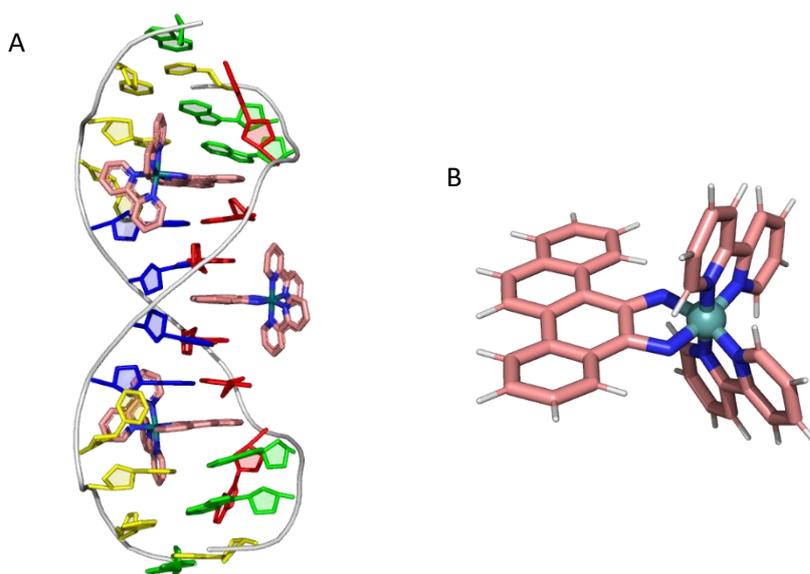
Structure **8RER** was solved by Dr James Hall and Ahmad Abdullrahman.

**8RER** was deposited in the Protein Data Bank with the entry title: Major groove intercalation with Polypyridyl Ruthenium complex.

## 5.1 Introduction

Octahedral ruthenium(II) complexes have been the subject of extensive studies aimed at elucidating their binding mode and DNA sequence specificity.<sup>1</sup> The helical structure of DNA provides two potential locations of these complexes: minor and major grooves, which differing size and shape.<sup>2</sup> Electrostatic calculation suggests that the minor groove, due to its narrow shape, exhibits a negative charge that would attract the positively charged complex, particularly in the context of TA sequences.<sup>3,4</sup>

Ongoing controversy persists regarding the preference of ruthenium polypyridyl complexes for major or minor groove intercalation. Initial suggestions reported that the preference depends on the enantioselectivity of the complex. NMR data showed that  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  intercalates from either the minor and major groove for the  $\Lambda$ -enantiomer and  $\Delta$ -enantiomer, respectively.<sup>5</sup> Furthermore, luminescence spectroscopy comparison of both enantiomers supported the intercalation of  $\Delta$ - $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  from the major groove, evident by an increased luminescence.<sup>6</sup> However, later spectroscopic studies, including linear and circular dichroism, appear to favour the intercalation polypyridyl complexes from the minor groove.<sup>7</sup>



**Figure 5.1** (A) X-ray structure of 2011 with  $\Delta$ - $[\text{Rh}(\text{bpy})_2(\text{chrysi})]^{3+}$  intercalating in the minor and major groove. Nucleobases are coloured with guanine in green, adenine in red, cytosine in yellow, and thymine in blue (B) Structure of the rhodium complex  $\Delta$ - $[\text{Rh}(\text{bpy})_2(\text{chrysi})]^{3+}$ . Carbon atoms in pink, nitrogen in blue, and rhodium atom in cyan.

The increasing number of crystal structures provides further evidence supporting the intercalation of ruthenium complexes from the minor grooves. Major groove intercalations are less explored in the literature, with only one structure (**2011**) showing the intercalation of rhodium complexes in both the minor and major grooves<sup>8</sup> (Figure 5.1). The metal complex  $\Delta$ -[Rh(bpy)<sub>2</sub>(chrysi)]<sup>3+</sup> (bpy = 2,2'-bipyridine, dppz = dipyrrophenazine; chrysi = chrysene-5,6-quinonediimine) is shown intercalated in the major groove into the central AT step.

During the screening of ruthenium complexes intercalated in DNA triplexes, additional systems were introduced to the initial screening library. Triplex AA-t7 consists of three 7-mer long sequences that fold to form a triple-helical structure. This sequence was previously studied as an intramolecular triplex and the NMR structure was solved and deposited as an intermolecular configuration (PDB: **149D**). In contrast to the AA-t3 from the triplex library, which has the same strand length, AA-t7 folds in a parallel fashion. Additionally, various ruthenium polypyridyl complexes commonly studied in Cardin's research group at the University of Reading were included in the screening. These complexes usually exhibit chemical alterations in their moieties, which results in various intrinsic binding constants to DNA.<sup>9-11</sup> The primary motivation behind this inclusion was to determine which alteration of the polypyridyl complex would result in bound within a triplex structure. Through wide screening, the metal complex [Ru(TAP)<sub>2</sub>(11-CN-dppz)]<sup>2+</sup> (TAP=1,4,5,8-tetraazaphenanthrene; dppz=dipyrrophenazine) crystallised with duplex DNA, while the TFO was not included in the crystal packaging. This unexpected outcome revealed surprising intercalation in the major groove, as discussed in this chapter.

## 5.2 Materials and Methods

### 5.2.1 Oligonucleotide, buffer solution, and crystallisation reagents

Oligonucleotides were purchased from Sigma Aldrich (Merck) as HPLC-purified solid and used without further purification. All buffer materials and chemicals were purchased from Sigma Aldrich. Crystallisation screens were sourced from Hampton Research, and the materials for crystallisation were obtained from SWISSCI.

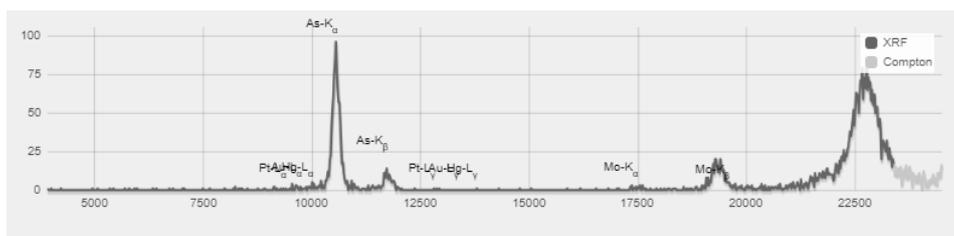
### 5.2.2 Crystallisation

Crystallisation was performed using vapour diffusion from sitting drop, which contained a solution of oligonucleotides  $d_1$ (GAATAGG),  $d_2$ (CCTATTC), and  $d_{TFO}$ (CTTGTCC) at a final concentration of 600  $\mu$ M, 20 mM Sodium Cacodylate pH 5.5, 100 mM NaCl, 10 mM  $MgCl_2$  and 1.6 mM  $[Ru(TAP)_2(11-CN-dppz)]^{2+}$ . The solution was mixed 1:1 ratio with a solution containing 80.6 mM potassium chloride, 40 mM sodium cacodylate trihydrate pH 5.86, 55% -2-Methyl-2,4-pentenediol, 120 mM Spermine tetrahydrochloride. The total drop volume of 400nL was mixed and dispensed by the robot Oryx8, Douglas Instruments Screens Version 10.00. The crystallisation was attempted at 18 °C, but plates later moved at 4°C where crystals subsequently grew.

### 5.2.3 Data collection, refinement, and analysis

Data were collected on I03 at Diamond Light Source Ltd from a flash-cooled crystal using radiation with a wavelength of 0.9762 Å. The beam size used of 80x20  $\mu$ m was sufficient to collect data with oscillation of 0.10° at 100 K obtaining 3600 images. Data were collected on I03 at Diamond Light Source Ltd from a flash-cooled crystal using radiation with a wavelength of 0.9762 Å. The beam size used of 80x20  $\mu$ m was sufficient to collect data with oscillation of 0.10° at 100 K obtaining 3600 images.

An X-ray fluorescence (XRF) scan was performed to examine the metal content of the crystal bulk (Figure 5.2). The wavelength was selected at 24.5 KeV to reach the photon energies of the  $K\alpha_1$  characteristic of the ruthenium atom (ca. 22.6 KeV).



**Figure 5.2** X-ray fluorescence scan of the crystal. Intense peaks are observed at low energy corresponding to the arsenic atom present in the buffer sodium cacodylate solution and at high intensity reflecting the presence of ruthenium atoms.

Two intense peaks of the XRF spectrum were observed. The first peak at lower energy corresponds to the  $K\alpha_1$  of the arsenic element (10,543.72 eV), which is present in the crystallisation solution (sodium cacodylate). The emission at higher energies confirms the presence of the ruthenium complex within the crystal bulk.<sup>12</sup>

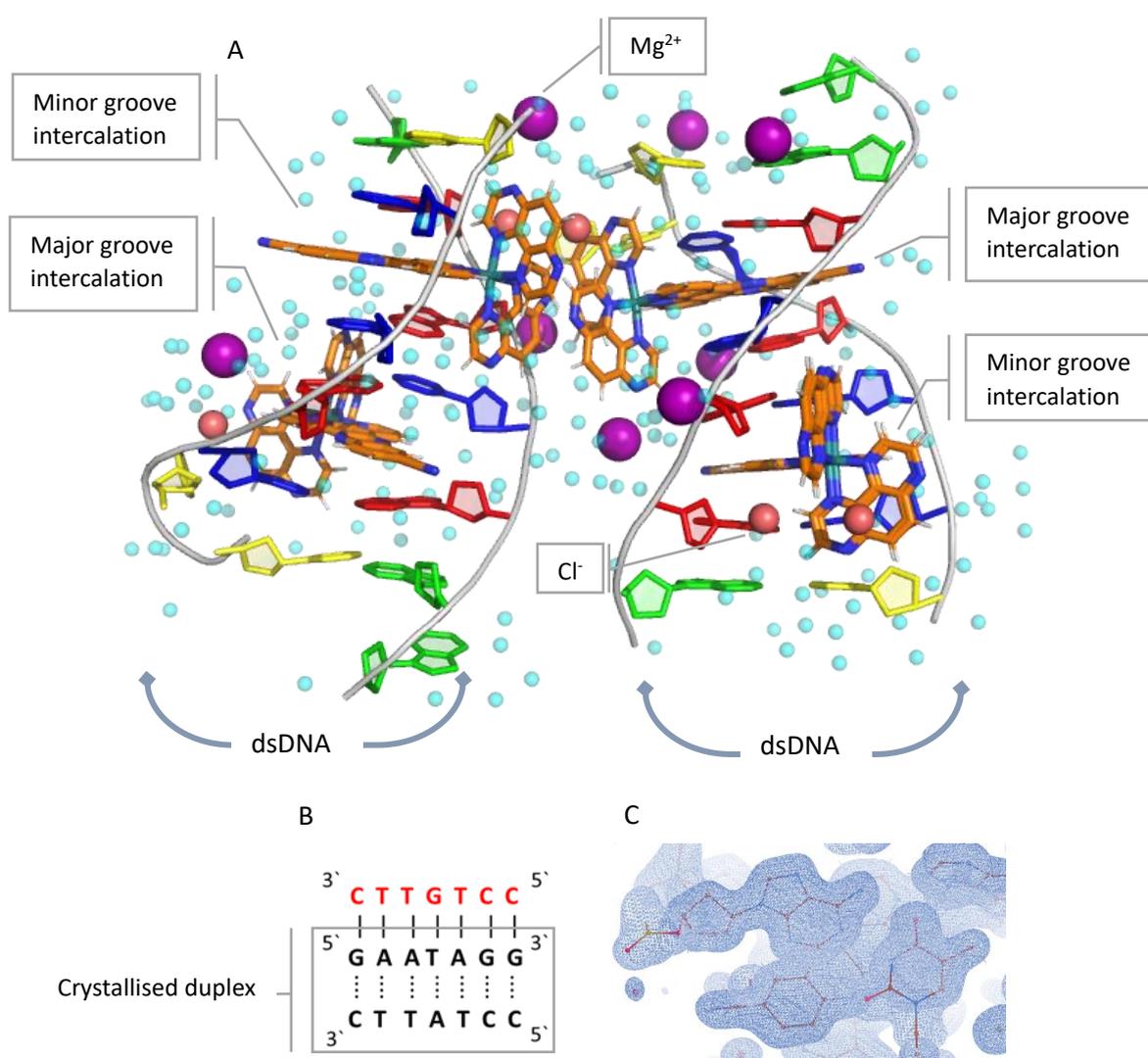
The resulting data were processed using DIALS and DIALS.SCALE to integrate and merge the peaks from all collected images, through the Xia2 pipeline<sup>13</sup>; giving 33977 unique reflections to a resolution of 1.20 Å. The phasing was determined using single-wavelength anomalous dispersion (SAD) with the anomalous scattering of the ruthenium using the SHELX/D/E<sup>14</sup> pipeline in CCP4.<sup>15</sup> The model was built by hand with WinCoot<sup>16</sup> and refined using the PHENIX software package.<sup>17</sup> The structure was deposited in the publicly available database Protein Data Bank (PDB)<sup>18</sup> with the title “DNA triplex structure with Polypyridyl Ruthenium Complexes”, PDB ID **8RER**. Data collection and refinement parameters are shown in Table 5.1. All derived parameters calculated using 3DNA<sup>19</sup> are available in Tables 5.1-4 in the appendix.

**Table 5.1** Crystallisation parameters and refinement statistics of **8RER**.

<b>Crystallisation Parameters</b>	
DNA Sequence	d <sub>1</sub> (GAATAGG), d <sub>2</sub> (CCTATTC)
Complex	<i>rac</i> -[Ru(TAP) <sub>2</sub> (11-CN-dppz)] <sup>2+</sup>
Crystallisation Temperature, °C	4
Growth Time (month)	16
Crystal Morphology	rod
<b>Data Collection</b>	
Diffraction source	I03
Radiation wavelength, Å	0.97623
Temperature, K	100
Exposure time, s	0.004
Detector	Hybrid Pixel Array Detector
Resolution, Å	1.20 (32.68)
<b>Data Processing</b>	
Structure Solution Method	SAD
Space group	<i>P</i> 4 1
a, b, c, Å	32.68, 32.68, 103.98
α, β, γ, Å	90, 90, 90
Resolution, Å	1.20 (32.68)
Total reflections	18794 (23826)
Unique reflections	1736 (1710)
<i>R</i> <sub>meas</sub>	0.029 (3.666)
<i>R</i> <sub>merge</sub>	0.028 (3.494)
<i>R</i> <sub>pim</sub>	0.008 (1.092)
Mean I/σ	0.828
CC <sub>1/2</sub>	0.995 (0.700)
Completeness, %	100
Multiplicity	13.7 (11.0)
Average B factors, Å <sup>2</sup>	12.380
Data Collection Date	27-Sep-23
*Outer shell statistics shown in parentheses	
<b>Refinement</b>	
No. Reflections	32448 (2197)
<i>R</i> <sub>work</sub>	0.1715
<i>R</i> <sub>free</sub>	0.2025
<b>Number of components</b>	
Nucleotide	28
Complex	4
Ligands	14
Water	207
<b>rmsd</b>	
Bond Lengths, Å	0.006
Bond Angles, °	2.27

## 5.3 Results and Discussion

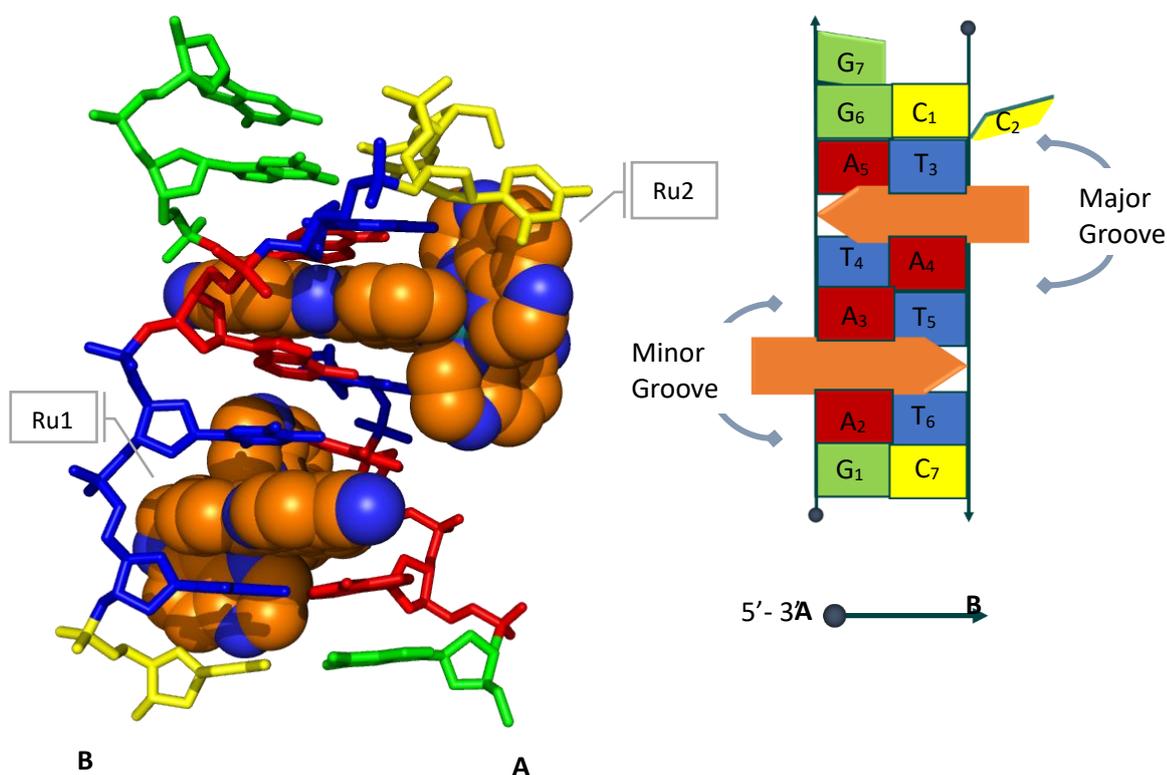
Following an extensive crystallisation screening aimed at investigating the structural details of the intercalation of ruthenium complexes within DNA triplexes, the resulting crystal structure reveals a distinctive preference for  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$  to intercalate in the minor and major grooves of the underlying duplex of the triplex. Indeed, in the crystallisation solution, three strands were included with a ratio of 1:1:1, however, only the duplex component of the triplex was found crystallised (Figure 5.3).



**Figure 5.3** (A) The asymmetric unit of **8RER** contains two right-handed DNA structures. Nucleobases are coloured as per convention with guanine in green, adenine in red, cytosine in yellow and thymine in blue. 8  $\text{Mg}^{2+}$  ions (purple), 6  $\text{Cl}^-$  ions, and 207 water molecules (cyan) are shown. For each DNA two  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$  complexes (orange) are intercalating in the major and minor groove. (B) Schematic representation of AA-t7 from which the only the underlying duplex (black) crystallised and not the TFO (red). (C) Representative example of electron density ( $2F_o - F_c$ ) is contoured at the  $1\sigma$  level (grey).

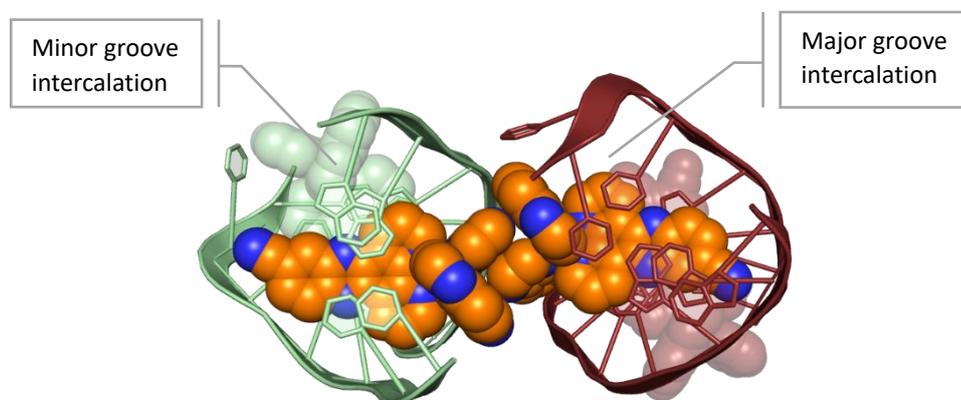
The crystal structure **8RER** shows the presence of two double-strand DNA molecules in the asymmetric unit. The structural arrangement demonstrates an opposite orientation of duplexes within the crystal packaging. As both DNA structures are similar (rmsd = 0.12 Å), only one copy will be discussed. The ions shown in the solvent, 8 Mg<sup>2+</sup> and 6 Cl<sup>-</sup> ions are predominantly localised in the region of the TAP ligands stacking, suggesting a potential role in stabilising the proximity of the DNA duplexes. Finally, the crystal structure is hydrated with 207 water molecules.

Two ruthenium complexes  $\Lambda$ -[Ru(TAP)<sub>2</sub>(11-CN-dppz)]<sup>2+</sup> are intercalated through the dppz ligand in each DNA molecule, for a total of four complexes. Figure 5.4 illustrates one biological unit observed in the crystal structure **8RER**. The first [Ru(TAP)<sub>2</sub>(11-CN-dppz)]<sup>2+</sup> complex (Ru1) intercalates in the AA/TT base step from the minor groove, while the second ruthenium complex (Ru2) is stacked in the TA/TA base step from the major groove of the DNA molecule.



**Figure 5.4** Structural (right) and schematic representation (left) of the biological unit with strand A and B.  $\Lambda$ -[Ru(TAP)<sub>2</sub>(11-CN-dppz)]<sup>2+</sup> complexes (orange) are intercalating in the major and minor groove. Nucleobases are coloured as per convention with guanine in green, adenine in red, cytosine in yellow and thymine in blue.

The TAP ligands, one emerging from the minor groove and the other from the major groove of the adjacent DNA molecule show a stacking interaction (Figure 5.5).

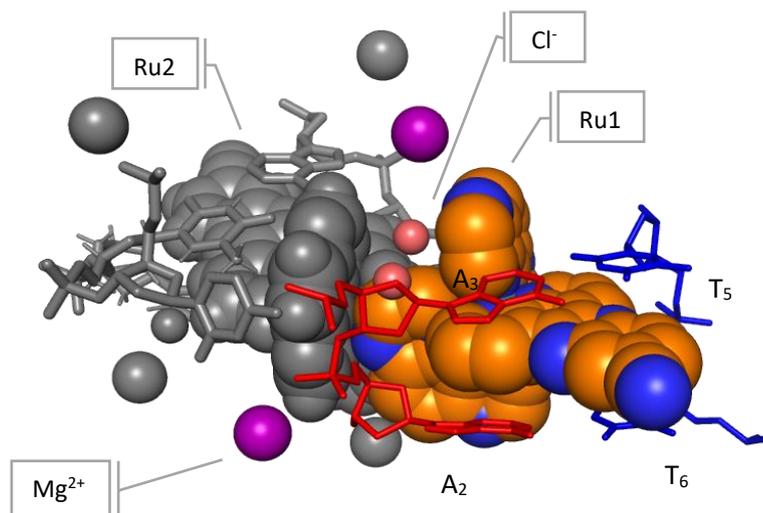


**Figure 5.5** Stacking of the TAP ligands belonging to ruthenium complexes that intercalate from the minor and major groove from each duplex (green and red).

The conformational analysis shows that the sugar pucker assignment corresponds to a B-DNA configuration (Table A5.5). Nevertheless, the analysis of the helical rise results in  $\sim 2.89$  Å which is closer to the A-DNA form, which typically exhibits an average helical rise of 2.5 Å. (Table A5.2) This has been calculated excluding the base step that reports higher rise values due to the intercalation of the ruthenium complexes. The intercalation of the ruthenium complexes results in an increased helical rise of the base pair steps, with 6.7 Å and 7.1 Å for the minor and major grooves, respectively.

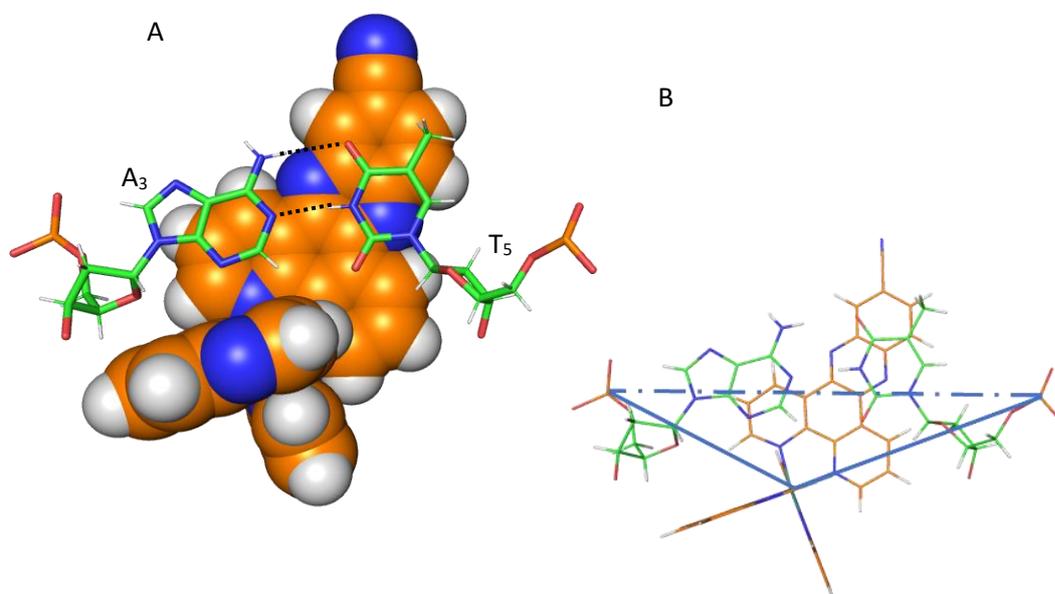
### 5.3.3 Minor groove intercalation

In the (A) $A_2A_3/T_5T_6$ (B) base step, Ru1 intercalates in a classical mode, with the CN-dppz intercalating from the minor groove and TAP ligand stacking with the TAP-Ru2 from the second DNA molecule (Figure 5.6). The base pairs buckle away from the CN-dppz ligand by  $9.96^\circ$  (Table A5.1); and unwind with a base step twist of  $17.4^\circ$ , which is lower compared to the average twist angle of B-DNA (around  $36^\circ$ ). The overall configuration is coordinated by  $Mg^{2+}$  and  $Cl^-$  cations.



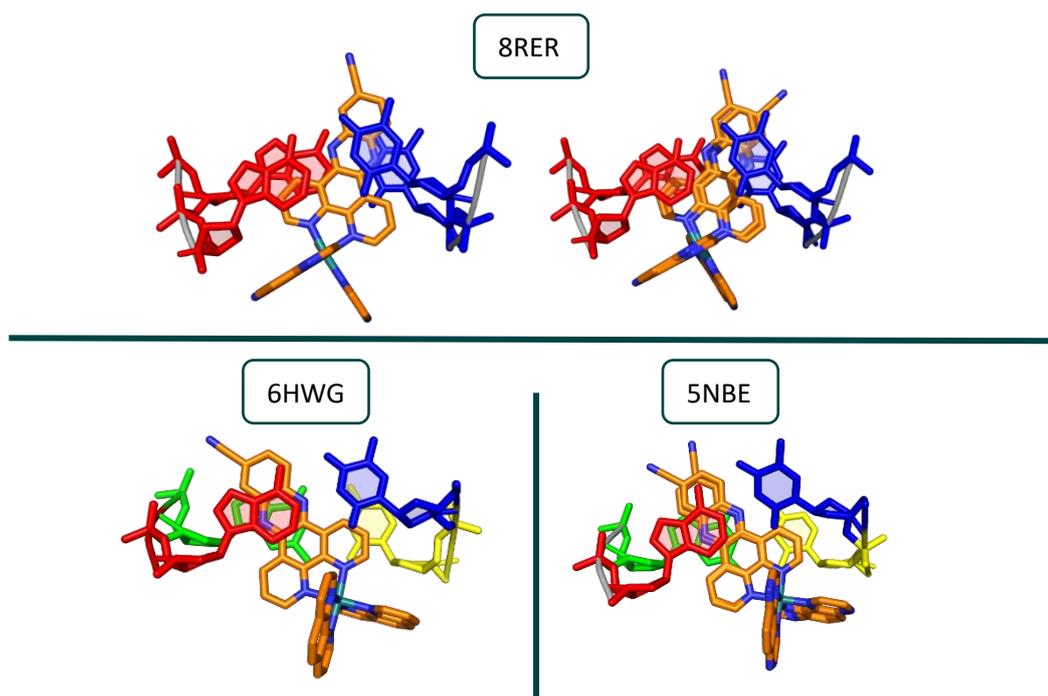
**Figure 5.6** Minor groove intercalation of Ru1 (orange) in the  $A_2A_3/T_5T_6$ . Major groove intercalation of Ru2 in the second duplex shown in grey. Adenine and thymine are coloured red and blue, respectively.  $Mg^{2+}$  ions (purple)  $Cl^-$  (pink) are shown as coordinating within the intercalation.

The orientation of Ru1 is slightly canted towards the pyrimidines  $T_5T_6$ , with the CN-dppz ligand forming an angle of  $80^\circ$  with the vector between the opposing phosphate groups at this step (Figure 5.7). The CN-dppz ligand is deeply inserted protruding in the major groove, and the  $Ru^{2+}$  atom is 2.6 Å distant from the DNA helix.



**Figure 5.7** (A) Stacking of Ru1 (orange) on the A:T base pair. Schematic representation showing Ru1 canted towards the thymine. Hydrogen bonds in dashed line between nitrogen atoms (blue) and oxygen (red).

The high-resolution crystal structure shows  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$  intercalating in the minor groove of the DNA duplexes with two distinct characteristics. In one model, the 11-CN substitution is oriented toward the major groove, while in the second model, the ruthenium complex is fitted after occupancy refinement in the two mirrored orientations, with the substituent directed towards the major groove and the phosphate backbone (Figure 5.8). A similar complex disorder is observed in the previously deposited crystal structure, **5NBE**. Moreover,  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$  reported crystallised in two duplex DNA (**5NBE**, **6HWG**) shows intercalation in the TC/GA base step with an angled orientation towards the GC bases, while in **8RER** the stacking is shown with the TT base. It is proposed that the polarisation induced by the cyano group favours the  $\pi$ - $\pi$  stacking of the dppz moiety with the high electron-rich guanine.<sup>20</sup> In the structure described here, T-stacking of the CN-dppz and the disordered orientation of one complex together suggests unspecific binding.



**Figure 5.8** Comparison of the minor groove intercalation of  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$  in the structure **8RER**, **6HWG** and **6NBE**. Multiple occupancies are reported in the structure **8RER** and **6NBE**. Nucleobases are coloured with guanine in green, adenine in red, cytosine in yellow and thymine in blue.

Table 5.2 shows X-ray structures of  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  both with and without dppz substitutions (Br, Cl, CN, and Me) intercalated in duplex DNA. These structures show the tendency of the dppz moiety for angled intercalation modes within the TC/GA and CC/GG steps, oriented to-

wards the guanine. This orientation often leads to the flipping out of the adenine. The only exceptions are reported for the internal base step TA/AT in the structures **3UYB**, **4M3I**, and **4M3V** where the complex assumes a perpendicular intercalation mode and adopts multiple occupancies. Conversely,  $\Lambda$ -[Ru(TAP)<sub>2</sub>(dppz)]<sup>2+</sup> intercalating in the terminal TT/AA base step is reported in the structure **5ET2** where the dppz ligand is canted and  $\pi$ - $\pi$  stacking with one adenine while the with 5' or 3' are flipped out. **8RER** structure shows unprecedented stacking stability of the cyano-substituted dppz with the pyrimidine in the TT/AA base step, contributing to the increasing understanding of the diverse intercalation modes of these complexes.

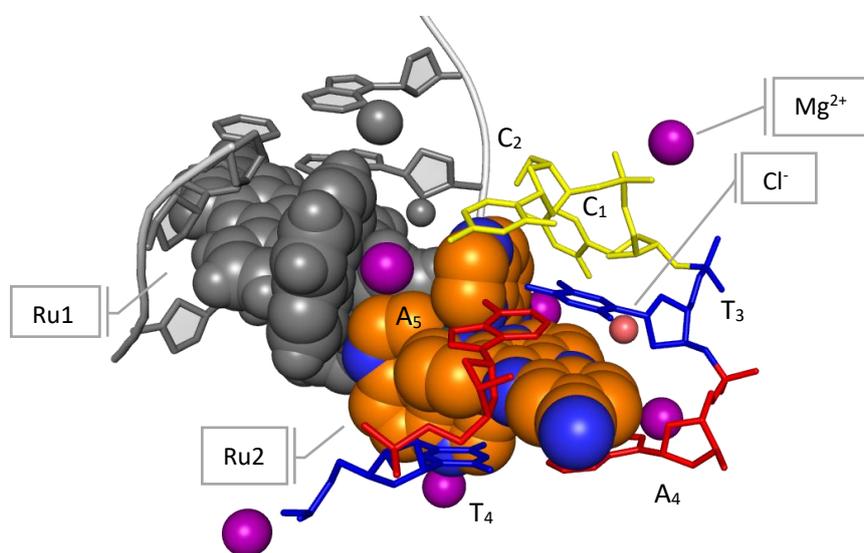
**Table 5.2** DNA structures deposited in the Protein Data Bank containing [Ru(TAP)<sub>2</sub>(dppz)]<sup>2+</sup> and its substituted variants.

Sequence	Intercalator	Intercalation mode	PDB ID	Ref.
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (11-CN-dppz)] <sup>2+</sup>	Canted to G	5NBE	20
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (11-CN-dppz)] <sup>2+</sup>	Canted to G	6HWG	21
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (11,12-CN2-dppz)] <sup>2+</sup>	Canted to G	6R6D	20
CCGGACCCGG	$\Lambda$ -[Ru(TAP) <sub>2</sub> (11,12-CN2-dppz)] <sup>2+</sup>	Canted to G	6G8S	20
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (11-Br-dppz)] <sup>2+</sup>	Canted to G	6GLD	20
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (11-Cl-dppz)] <sup>2+</sup>	Canted to G	4III	22
CCGGATCCGG	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz-11,12-(F)2)] <sup>2+</sup>	Canted to G	4MS5	23
CCGGATCCGG	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz-{Me2})] <sup>2+</sup>	Canted to G	4E95	23
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz-{Me2})] <sup>2+</sup>	Canted to G	4E8S	24
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz-{Me2})] <sup>2+</sup>	Canted to G	4X1A	24
CCGGCTCCGG	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz-{Me2})] <sup>2+</sup>	Canted to G	4E8X	23
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz-11-Me)] <sup>2+</sup>	Canted to G	4X18	24
CCGGTACCGG	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz-{Me2})] <sup>2+</sup>	Perpendicular	4M3I	25
CCGGTACCGG	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz-{Me2})] <sup>2+</sup>	Perpendicular	4M3V	25
CCGGGCCCGG	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	5IWJ	26
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	3QF8	23
CCGGAGCCGG	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	5IP8	27
XCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	4R8J	28
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	3QRN	29
TCGGCICCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	5IU5	30
TCGGTACCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Perpendicular	3UYB	31
TCGGCGCCIA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to I	4QIO	32
TTGGCGCCAA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to A	5ET2	32
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	4LTF	33
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	4LTH	33
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	4LTJ	33
TCGGCGCCGA	$\Lambda$ -[Ru(TAP) <sub>2</sub> (dppz)] <sup>2+</sup>	Canted to G	4LTL	33

Nevertheless, it should be acknowledged that the occurrence of flipping bases is commonly observed in the terminal step of the structure. Consequently, it is plausible that this arrangement is influenced by the crystal packing, a phenomenon commonly encountered in crystallographic studies. Notably, a preference for minor groove intercalation is observed for the  $\Lambda$ -enantiomer over the  $\Delta$ -enantiomer. The **8RER** structure confirms the enantioselectivity of the  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$  for the  $\Lambda$  chiral counterpart.

### 5.3.2 Major groove intercalation

Figure 5.9 shows Ru2 complex intercalated into the (A) $T_4A_5/T_3A_4$ (B) step from the major groove with roll angle of  $-11.14^\circ$  (Table A5.4). The CN-dppz moiety of Ru2 is intercalated symmetrically into the major groove, with the  $\text{Ru}^{2+}$  positioned  $8.75 \text{ \AA}$  away from the helical axis (Figure 5.9). The negative roll associated with this base step suggests that the major groove opens, upon intercalation. Compared to the minor groove intercalation, which results in positive roll, the intercalation of the complex seems to compress the depths of the major groove (Table A5.6). These values are consistent with the major intercalation of  $\Delta$ - $[\text{Rh}(\text{bpy})_2(\text{chrysi})]^{3+}$  observed in the structure **201I**.<sup>34</sup>

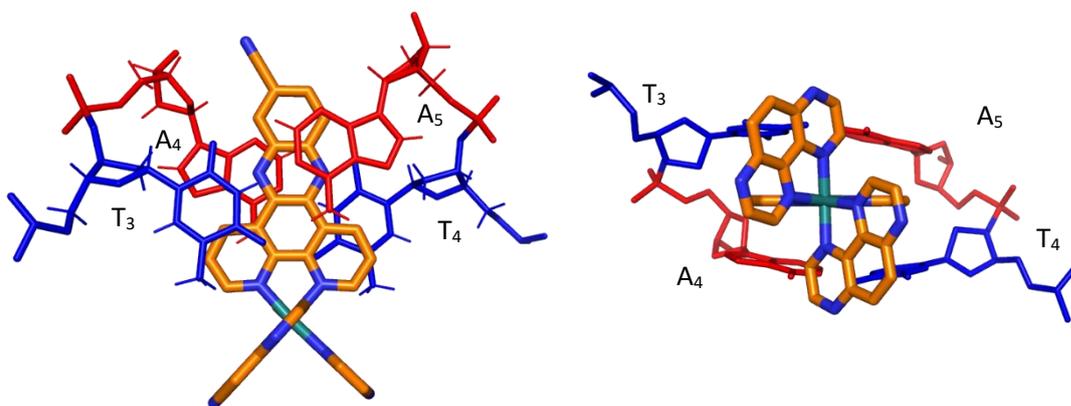


**Figure 5.9** Major groove intercalation of Ru2 (orange) in the  $T_4A_5/T_3A_4$ . Minor groove intercalation of Ru1 in the second duplex shown in grey. Adenine and thymine are coloured red and blue, respectively. Terminal cytosine  $C_1$  and  $C_2$  are shown in yellow.  $\text{Mg}^{2+}$  ions (purple)  $\text{Cl}^-$  (pink) are shown as coordinating within the intercalation.

The significant difference in twist angle between  $T_3A_4$  and  $T_4A_5$  ( $19.6^\circ$  versus  $40.1^\circ$  respectively), along with a slide difference ( $-0.47$  and  $0.75$ ) enables favourable stacking of the dppz on

both (A)A<sub>4</sub> and (B)A<sub>5</sub>. Perpendicular stacking with complexes positioned in multiple occupancies is also seen in other structures (Table 5.2), with  $\pi$ - $\pi$  stacking with multiple nucleobases due to multiplex occupancies of the complex. Similar diffraction outcomes are reported as well for  $\Delta$ -[Rh(bpy)<sub>2</sub>(chrysi)]<sup>3+</sup> as major groove intercalator in the **2011** structure. This suggests that the high DNA twisting of the DNA at the T<sub>4</sub>A<sub>5</sub>/T<sub>3</sub>A<sub>4</sub> base steps supports a stable perpendicular intercalation from the major groove. As a result of these arrangements, the dppz moiety forms weak hydrogen interactions with both A<sub>4</sub> and A<sub>5</sub>, with the 11-CN group close to A<sub>4</sub> (H4') at a hydrogen distance of 2.9 Å.

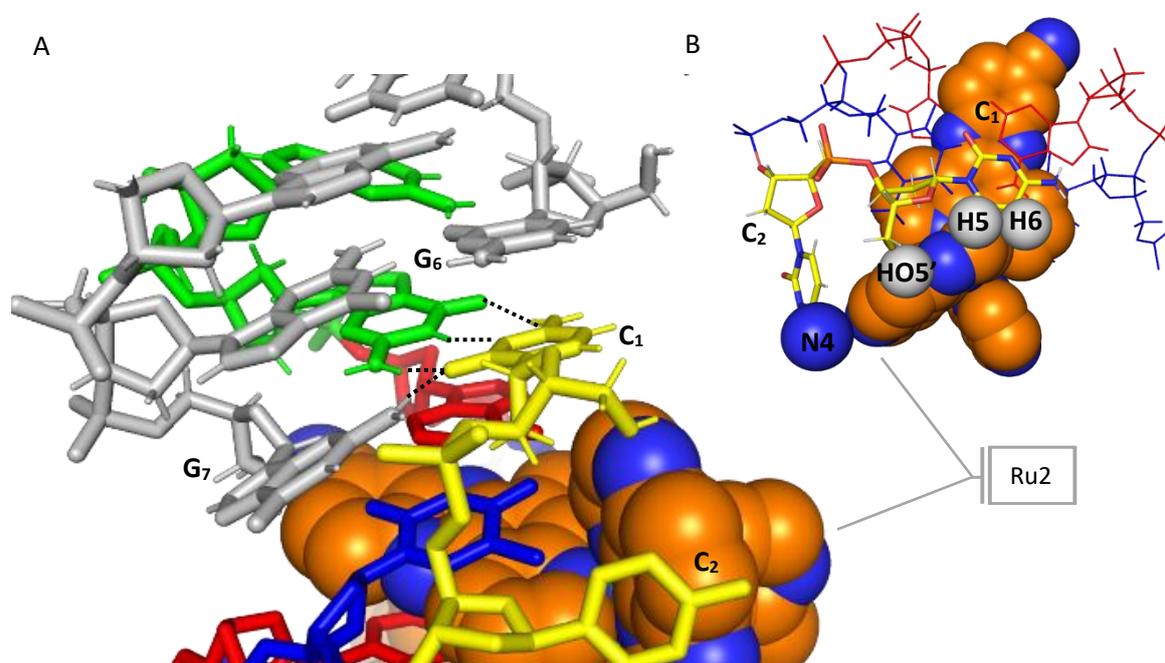
The helical twist variation between the minor and major groove intercalation steps (Table A.5.3) indicated conformational variability of the DNA structure. Previous, NMR studies revealed that binding in the major groove of the phen ligand of [Ru(phen)<sub>3</sub>]<sup>2+</sup> induces distortions in the DNA structure, contrary to binding in the minor groove, where such were less pronounced.<sup>35</sup> This suggests that the major groove is prone to structural alterations after complex interactions. Nevertheless, the major groove width reported in **8RER** (24.9 Å) matches the canonical B-DNA, as expected the intercalation of the complex does not impact the groove width.



**Figure 5.10** Major groove intercalation of Ru2 (orange) in the T<sub>4</sub>A<sub>5</sub>/T<sub>3</sub>A<sub>4</sub>. Ru2 stacks on A4 and A5 due to twisting of base pairs. Adenine and thymine are coloured red and blue, respectively.

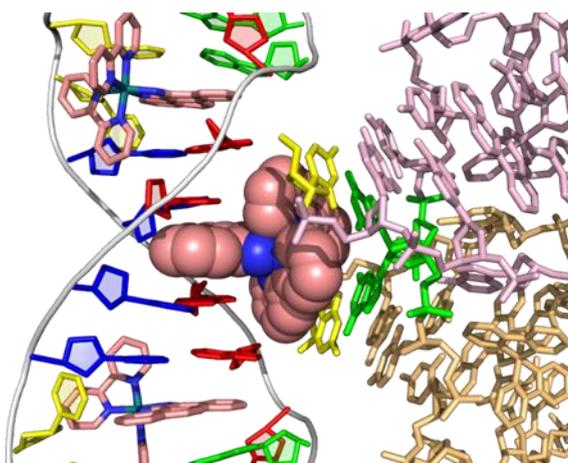
In the **8RER** model, the (B)C<sub>2</sub> is flipped out, enabling the formation of hydrogen bonds between (A)G<sub>6</sub> and (B)C<sub>1</sub>, as well as one hydrogen bond with (A)G<sub>7</sub> of the symmetry-related molecules (Figure 5.10). It is argued that the absence of mismatched bases could favour major groove intercalation of complexes.<sup>34</sup> Flipping bases in crystal structures were reported in several DNA structures bound to polypyridyl complexes. Conformational variations of base pairs in crystal

structures often play a role in accommodating the crystal packing, optimising interactions for stable arrangements. To date, the limited number of polypyridyl complexes reported is not sufficient to provide a definitive conclusion about the sequence effect on the major groove intercalation. However, this conformation allows additional  $\pi$ -stacking of the terminal cytosine ( $C_1$  and  $C_2$ ) on the TAP ligand. Close contacts between  $HO5'$ ,  $H5$ , and  $H6$  from  $C_1$  and  $N4$  of  $C_2$  with the TAP ligand are shown in Figure 5.11.



**Figure 5.11** (A)  $C_1$  forming hydrogen bond with  $G_6$  and  $G_7$  from symmetry related molecule (grey) (B)  $N4$  from  $C_1$  and  $HO5'$ ,  $H5$  and  $H6$  from  $C_2$  in contact with TAP ligand. Nucleobases are with guanine in green, adenine in red, cytosine in yellow and thymine in blue.

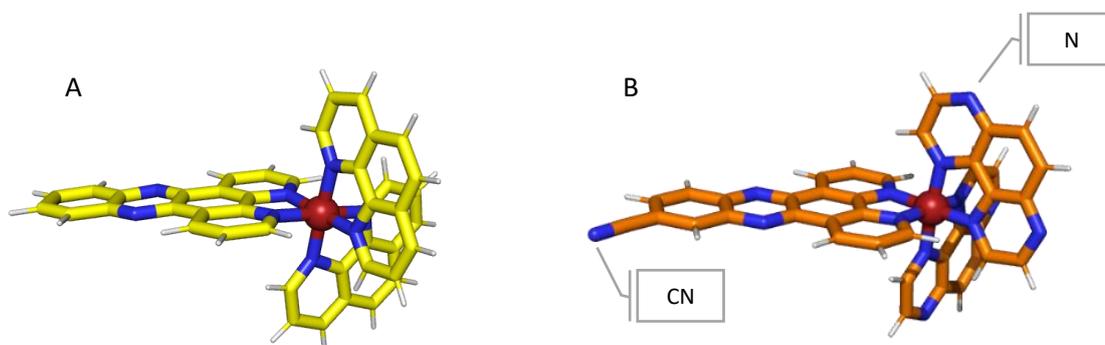
Additionally, the stacking of the TAP ligands of  $Ru1$  and  $Ru2$  described above may contribute to the enhanced stability of the major groove intercalation. This arrangement resembles the  $[Rh(bpy)_2(chrysi)]^{3+}$  complex, which intercalates into the major groove with both bpy ligands stacking on the terminal G:C base pairs of two neighbouring duplexes (Figure 5.12). Thus,  $[Ru(TAP)_2(11-CN-dppz)]^{2+}$  and  $[Rh(bpy)_2(chrysi)]^{3+}$  represent examples of major groove intercalation geometry facilitated by polypyridyl ligands.



**Figure 5.12** Intercalation of  $[Rh(bpy)_2(chrysi)]^{3+}$  in the major groove in the **2O1I** structure and stacking of *bpy* ligand with G:C base pairs from two different molecules (pink and orange). Nucleobases are with guanine in green, adenine in red, cytosine in yellow and thymine in blue.

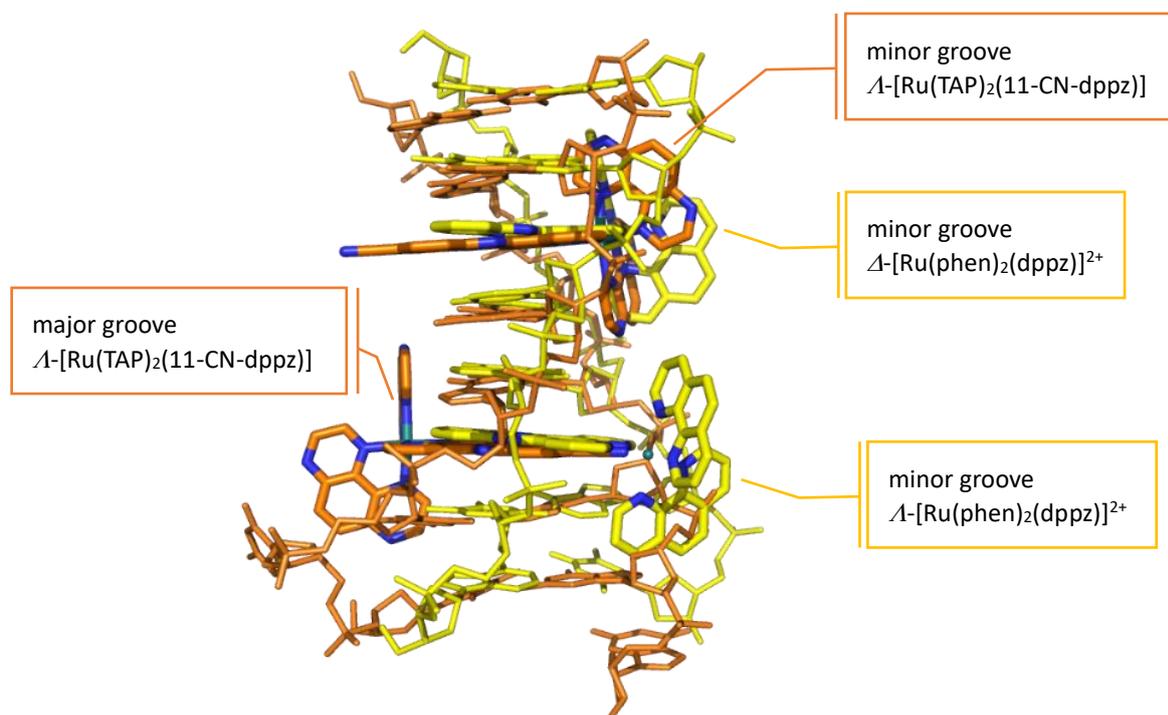
Previous analysis of interaction energies involving aromatic groups emphasised the requirement for a substantial number of  $\pi$ -stacking interactions to promote major groove intercalations. Conversely, minor groove intercalation is characterised by an elevated number of  $\pi$ -stacking with aromatic groups, attributed to the narrower width of the groove.<sup>36,37</sup> Steric hindrance imposed by large metal complexes contributes to the number of interactions with the nucleobases further stabilising major groove intercalation.<sup>37</sup> In the **8RER** structure, the major groove width does not impose steric hindrance on the CN-dppz ligand, suggesting that the moiety size is unlikely to contribute to the preference for the major groove. Nevertheless, the proximity of the CN group with  $A_5(H4')$  would suggest that the intercalation is governed by electronics between the complex and the DNA base pairs, thus promoting major groove intercalation.

The polypyridyl complex  $[Ru(phen)_2(dppz)]^{2+}$ , explored in the previous chapters as a “Light switch” candidate, and  $[Ru(TAP)_2(11-CN-dppz)]^{2+}$  are isostructural, differing only in the substitution with two nitrogen atoms on the phen ligand, forming the TAP ligand (TAP = 1,4,5,8-tetraazaphenanthrene) and the nitrile substitution (CN) at position 11 of the dppz ligand (Figure 5.13).



**Figure 5.13** Comparison of (A) Crystal structure of  $\text{Ru}(\text{phen})_2(\text{dppz})^{2+}$  and (B)  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$ . Nitrogen in blue and ruthenium in red.

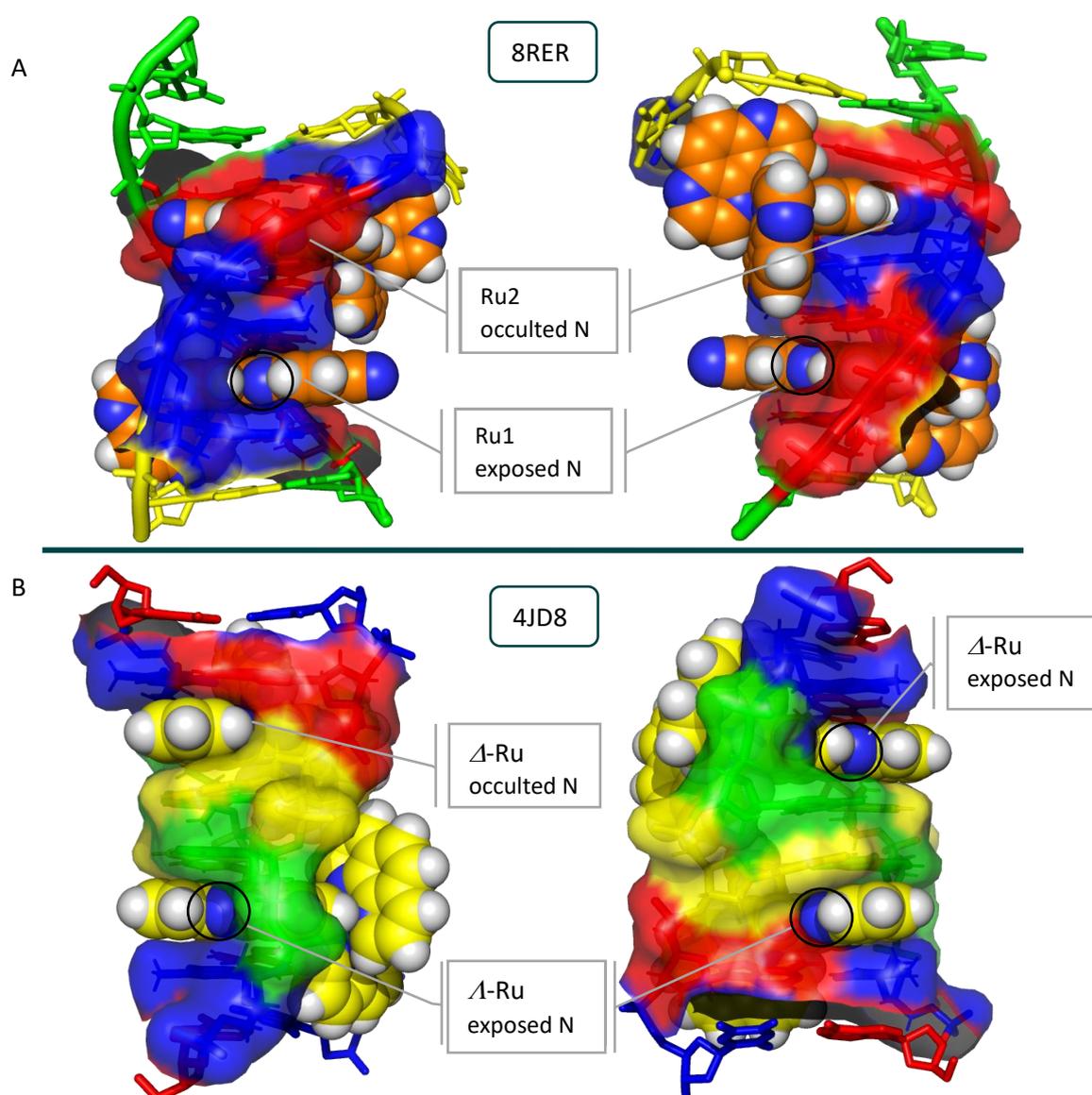
Intercalation of ruthenium complexes in alternated base steps was previously reported, as seen in **4JD8**, where both  $\Lambda$ - and  $\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  ( $\Lambda$ -Ru;  $\Delta$ -Ru) intercalated in the TG/CA base steps.<sup>38</sup> Figure 5.14 illustrates the superimposition of the **4JD8** and **8RER** (rmsd with all DNA atoms of 1.15 Å), showing overall similarity in DNA geometry.



**Figure 5.14** Superimposition of crystal structure **4JD8** (yellow) with **8RER** (orange).

The canted geometries of the  $\Delta$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  and  $\Lambda$ - $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$  show distinct degrees of offset for the dppz group from the P-P vector ( $65^\circ$  and  $80^\circ$ , respectively). However, the intercalation of  $\Lambda$ - $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ , in contrast to major groove intercalation, exhibits slightly different changes in the DNA structure attributed to the stacking interaction of cytosine on the TAP ligand.

The luminescent enhancement upon DNA intercalation is attributed to the shielding from the aqueous solvent interactions. The minor groove intercalation in a canted mode of  $\Delta$ -[Ru(TAP)<sub>2</sub>(11-CN-dppz)]<sup>2+</sup> shows phenazine nitrogen exposed to the solvent, compared to the canted orientation of where one phenazine nitrogen is solvent accessible (Figure 5.15). This variation can be rationalised by the distinct angle of intercalation relative to the phosphate backbone. Interestingly, the perpendicular major groove intercalation of  $\Delta$ -[Ru(TAP)<sub>2</sub>(11-CN-dppz)]<sup>2+</sup> results in both phenazine nitrogen protected from water in the solvent, compared to  $\Delta$ -Ru and Ru1, thus hypothesising that the major groove provides a shielded environment, rendering the phenazine nitrogen less accessible to the surrounding aqueous solvent.



**Figure 5.15** (A) Structure **8RER** Phenazine nitrogen of Ru2 blocked by DNA, while in Ru1 exposed in the **8RER** structure. (B) Structure **4JD8**, one phenazine of  $\Delta$ -Ru nitrogen exposed, in  $\Delta$ -Ru are both exposed to the solvent. The phenazine nitrogen are indicated by a black circle.

The photophysical and photochemical behaviours of these complexes were previously compared.<sup>39</sup>  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  complex emits luminescence both in water and in organic solvents, with its luminescence quenched by the guanine. Additionally, upon binding of the guanine-rich sequences, a photooxidising effect is observed.<sup>40</sup> The CN substitution on the dppz does not interfere with the intercalation and its phototoxicity potential has been investigated.<sup>41</sup> Moreover, the CN moiety protrudes in the solvent grooves (both minor and major grooves in **8RER**) rendering this accessible to the solvent.<sup>42</sup> The intercalation of  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  with various substitutions on the distal ring of the dppz ligand, was examined through X-ray crystallography.<sup>24</sup> In particular,  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$  was reported intercalating on crystal structures of duplexes and G-quadruplexes.<sup>20,42</sup> While this complex may lack the characteristic “light switch” effect it is worth exploring its potential as a DNA-intercalating compound with photooxidising properties.

The photooxidising property of the  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  is attributed to the formation of a covalent linkage between the guanine amino group and the pyrazine carbons C2/C7 of the DNA bound-TAP ligand.<sup>43</sup> Such linkages have not been reported within the adenine or pyrimidine bases, with intercalation limited to non-covalent intercalation. Consequently, the observed intercalation mode between TA bases would be expected to lack photooxidation activity compared to the absence of the covalent link with guanine. While it is proposed that linking with adenine can occur with free TAP fragments,<sup>44</sup> the **8RER** structure provides an AT-rich DNA model for further spectroscopic analysis with the ruthenium complex  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  bearing a CN substitution, thereby advancing the development of targeted photooxidising applications.<sup>45</sup>

## 5.4 Conclusion

A second finding resulting from the screening of DNA triplex systems with ruthenium complexes was the unexpected crystallisation of a duplex structure rather than the designed triplex structure when using  $[\text{Ru}(\text{TAP})_2(11\text{-CN-dppz})]^{2+}$ . While the TFO did not incorporate into the crystal lattice, the ruthenium complex exhibited a novel binding mode. The substituted dppz ligand resulted in intercalation not only into the expected minor groove but also into the major groove in the same duplex. The **8RER** model showed a novel  $\pi$ -stacking arrangement of the TAP ligand with multiple nucleobases from the major groove. The addition of a cyano group at position 11 contributed to the major groove intercalation. Binding preferences were previously observed in methyl substitutions, suggesting that dppz substitutions could play a role in binding preferences.<sup>24</sup> Nonetheless, the minor and major groove intercalation do not significantly perturb the overall DNA duplex structure, with deviations from the regular B-DNA geometry only observed at the metal complex binding sites. With the currently limited structural data available, it remains unclear whether the cationic conditions used played a role in stabilising this major groove intercalation and/or contributed to the displacement of the TFO.<sup>46</sup>

Compared to  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  which was found to successfully intercalate and crystallised within a DNA triplex structure (demonstrated in Chapter 4), the visible light absorption of  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  is governed by metal-to-ligand charge transfer (MLCT) transitions localised on the TAP ligands, regardless of the surrounding environment.<sup>47</sup> Moreover,  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  exhibits emissive properties, with an emission sensitive to the binding environment. Specifically, its emission is sensitive to the binding environment. Specifically, the emission is quenched upon binding with guanine; however, outside of such sites, it displays emission characteristics. Although this renders  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  less favourable as a light-switch complex, the known photocleavage activity could be harnessed as a photoactivated DNA triplex cleavage agent.<sup>48,49</sup>

Studies involving the conjugation of the TFO with groove binders have been previously explored, however, the TFO was found dissociated with the binder alone localised in the minor groove.<sup>50</sup> Other examples involve modified bases within the TFO itself, such as the non-natural base D3, which intercalates into the major groove as part of the TFO.<sup>51</sup> As observed in Chapter 4, the removal of the D3 base facilitated the accommodation of  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$  from the minor groove in the same position. Polypyridine complexes linked to TFO have also been developed as

probes to selectively target and cleave the underlying duplex DNA.<sup>52</sup> Since the objective is to obtain a system that recognises both the TFO and the duplex, the **8RER** structure demonstrates major groove intercalation by  $[\text{Ru}(\text{TAP})_2(\text{dppz})]^{2+}$ , enabling future studies of triplex stabilisation and metal complex intercalation.

## 5.5 References

1. Cardin, C. J., Kelly, J. M. & Quinn, S. J. Photochemically active DNA-intercalating ruthenium and related complexes-insights by combining crystallography and transient spectroscopy. *Chem Sci* **8**, 4705–4723 (2017).
2. Oguey, C., Foloppe, N. & Hartmann, B. Understanding the Sequence-Dependence of DNA Groove Dimensions: Implications for DNA Interactions. *PLoS One* **5**, e15931 (2010).
3. Rohs, R. *et al.* The role of DNA shape in protein–DNA recognition. *Nature* **461**, 1248–1253 (2009).
4. Honig, B. & Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* (1979) **268**, 1144–1149 (1995).
5. Barton, J. K., Goldberg, J. M., Kumar, C. V & Turro, N. J. Binding Modes and Base Specificity of Tris(phenanthroline)ruthenium(II) Enantiomers with Nucleic Acids: Tuning the Stereoselectivity. *Bleomycin. Chemical, Biochemical and Biological Aspects* **108**, 2081–2088 (1986).
6. Dupureur, C. M. & Barton, J. K. Structural Studies of  $\Lambda$ - and  $\Delta$ -[Ru(phen)<sub>2</sub>dppz]<sub>2</sub><sup>+</sup> Bound to d(GTCGAC)<sub>2</sub>: Characterization of Enantioselective Intercalation. *Inorg Chem* **36**, 33–43 (1997).
7. Brabec, V. & Kasparkova, J. Ruthenium coordination compounds of biological and biomedical significance. DNA binding agents. *Coord Chem Rev* **376**, 75–94 (2018).
8. Pierre, V. C., Kaiser, J. T. & Barton, J. K. Insights into finding a mismatch through the structure of a mispaired DNA bound by a rhodium intercalator. *Proceedings of the National Academy of Sciences* **104**, 429–434 (2007).
9. Ortmans, I., Elias, B., Kelly, J. M., Moucheron, C. & Kirsch-DeMesmaeker, A. [Ru(TAP)<sub>2</sub>(dppz)]<sub>2</sub><sup>+</sup>: a DNA intercalating complex, which luminesces strongly in water and undergoes photo-induced proton-coupled electron transfer with guanosine-5'-monophosphate. *Dalton Trans.* 668–676 (2004).
10. Mihailovic, A. *et al.* Exploring the Interaction of Ruthenium(II) Polypyridyl Complexes with DNA Using Single-Molecule Techniques. *Langmuir* **22**, 4699–4709 (2006).
11. Liu, X.-W. *et al.* Synthesis, characterization, DNA-binding and photocleavage of complexes [Ru(phen)<sub>2</sub>(6-OH-dppz)]<sub>2</sub><sup>+</sup> and [Ru(phen)<sub>2</sub>(6-NO<sub>2</sub>-dppz)]<sub>2</sub><sup>+</sup>. *J Inorg Biochem* **99**, 2372–2380 (2005).
12. Rodríguez, T., Sepúlveda, A., Carreras, A., Castellano, G. & Trincavelli, J. Structure of the Ru, Ag and Te L X-ray emission spectra. *J Anal At Spectrom* **31**, 780–789 (2016).
13. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* **69**, 1260–1273 (2013).

14. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479–485 (2010).
15. Collaborative Computational Project, N. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760–763 (1994).
16. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
17. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).
18. Bernstein, F. C. *et al.* The protein data bank: A computer-based archival file for macromolecular structures. *J Mol Biol* **112**, (1977).
19. Li, S., Olson, W. K. & Lu, X.-J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res* **47**, (2019).
20. McQuaid, K., Hall, J. P., Brazier, J. A., Cardin, D. J. & Cardin, C. J. X-ray Crystal Structures Show DNA Stacking Advantage of Terminal Nitrile Substitution in Ru-dppz Complexes. *Chemistry – A European Journal* **24**, 15859–15867 (2018).
21. McQuaid, K. T. Structural Insights into the DNA Binding Properties of Ruthenium Polypyridyl Complexes, Ph.D. Thesis. (University of Reading, 2019).
22. Hall, J. P., Beer, H., Buchner, K., Cardin, D. J. & Cardin, C. J. Preferred orientation in an angled intercalation site of a chloro-substituted  $\Lambda$ -[Ru(TAP)<sub>2</sub>(dppz)]<sup>2+</sup> complex bound to d(TCGGCGCCGA)<sub>2</sub>. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 1–8 (2013).
23. Hall, J. P. X-ray Crystallographic Studies of DNA-Bound Polypyridyl Ruthenium Complexes, Ph.D. Thesis. (University of Reading, 2014).
24. Hall, J. P., Beer, H., Buchner, K., Cardin, D. J. & Cardin, C. J. The Structural Effect of Methyl Substitution on the Binding of Polypyridyl Ru-dppz Complexes to DNA. *Organometallics* **34**, 2481–2486 (2015).
25. Niyazi, H., Teixeira, S., Mitchell, E., Forsyth, T. & Cardin, C. X-ray crystal structure of the ruthenium complex [Ru(TAP)<sub>2</sub>(dppz-{Me<sub>2</sub>})]<sup>2+</sup> bound to d(CCGGTACCGG). *Protein Data Bank* (2018).
26. Souter, J. E., Hall, J. P. & Cardin, C. J.  $\Lambda$ -[Ru(TAP)<sub>2</sub>(dppz)]<sup>2+</sup> bound to d(CCGGGCCCCGG). *Protein Data Bank* (2017).
27. Souter, J. E., Hall, J. P. & Cardin, C. J.  $\Lambda$ -Ru(TAP)<sub>2</sub>dppz bound to d(CCGGCTCCGG). *Protein Data Bank* (2017).
28. Hall, J. P. *et al.* Monitoring one-electron photo-oxidation of guanine in DNA crystals using ultrafast infrared spectroscopy. *Nat Chem* **7**, 961–967 (2015).

29. Hall, J. P. *et al.* Structure determination of an intercalating ruthenium dipyrrophenazine complex which kinks DNA by semiintercalation of a tetraazaphenanthrene ligand. *Proc Natl Acad Sci U S A* **108**, 17610–17614 (2011).
30. Hall, J. P. & Cardin, C. J. Lambda-[Ru(TAP)2(dppz)]<sup>2+</sup> bound to d(TCGGCICCGA)<sub>2</sub>. *Protein Data Bank* (2017).
31. Niyazi, H. N., Texeira, S. C. M., Mitchell, E. P., Cardin, J. C. & Forsyth, V. T. X-ray crystal structure of the ruthenium complex [Ru(tap)2(dppz)]<sup>2+</sup> bound to d(TCGGTACCGA). *Protein Data Bank* (2013).
32. Keane, P. M. *et al.* Inosine Can Increase DNA's Susceptibility to Photo-oxidation by a Ru II Complex due to Structural Change in the Minor Groove. *Chemistry – A European Journal* **23**, 10344–10351 (2017).
33. Hall, J. P. *et al.* Controlled Dehydration of a Ruthenium Complex–DNA Crystal Induces Reversible DNA Kinking. *J Am Chem Soc* **136**, 17505–17512 (2014).
34. Franco, D., Vargiu, A. V. & Magistrato, A. Ru[(bpy)2(dppz)]<sup>2+</sup> and Rh[(bpy)2(chrysi)]<sup>3+</sup> Targeting Double Strand DNA: The Shape of the Intercalating Ligand Tunes the Free Energy Landscape of Deintercalation. *Inorg Chem* **53**, 7999–8008 (2014).
35. Eriksson, M., Leijon, M., Hiort, C., Norden, B. & Graeslund, A. Binding of delta and lambda-[Ru(phen)3]<sup>2+</sup> to [d(CGCGATCGCG)]<sub>2</sub> Studied by NMR. *Biochemistry* **33**, 5031–5040 (1994).
36. Gil, A., Branchadell, V. & Calhorda, M. J. A theoretical study of methylation and CH/π interactions in DNA intercalation: methylated 1,10-phenanthroline in adenine–thymine base pairs. *RSC Adv* **6**, 85891–85902 (2016).
37. Sánchez-González, Á., Castro, T. G., Melle-Franco, M. & Gil, A. From groove binding to intercalation: unravelling the weak interactions and other factors modulating the modes of interaction between methylated phenanthroline-based drugs and duplex DNA. *Physical Chemistry Chemical Physics* **23**, 26680–26695 (2021).
38. Hall, J. P. *et al.* X-ray crystal structure of rac-[Ru(phen)2dppz]<sup>2+</sup> with d(ATGCAT)<sub>2</sub> shows enantiomer orientations and water ordering. *J Am Chem Soc* **135**, 12652–12659 (2013).
39. Ortmans, I., Elias, B., Kelly, J. M., Moucheron, C. & Kirsch-DeMesmaeker, A. [Ru(TAP)2(dppz)]<sup>2+</sup>: a DNA intercalating complex, which luminesces strongly in water and undergoes photo-induced proton-coupled electron transfer with guanosine-5'-monophosphate. *Dalton Trans.* **1**, 668–676 (2004).
40. Lecomte, J.-P., Kirsch-De Mesmaeker, A., Feeney, M. M. & Kelly, J. M. Ruthenium(II) Complexes with 1,4,5,8,9,12-Hexaazatriphenylene and 1,4,5,8-Tetraazaphenanthrene Ligands: Key Role Played by the Photoelectron Transfer in DNA Cleavage and Adduct Formation. *Inorg Chem* **34**, 6481–6491 (1995).

41. Mari, C., Rubbiani, R. & Gasser, G. Biological evaluation of nitrile containing Ru(II) polypyridyl complexes as potential photodynamic therapy agents. *Inorganica Chim Acta* **454**, 21–26 (2017).
42. McQuaid, K. *et al.* Structural Studies Reveal Enantiospecific Recognition of a DNA G-Quadruplex by a Ruthenium Polypyridyl Complex. *Angewandte Chemie International Edition* **58**, 9881–9885 (2019).
43. Jacquet, L., Davies, R. J. H., Kirsch-De Mesmaeker, A. & Kelly, J. M. Photoaddition of Ru(tap)2(bpy)2+ to DNA: A New Mode of Covalent Attachment of Metal Complexes to Duplex DNA. *J Am Chem Soc* **119**, 11763–11768 (1997).
44. Knapp, C., Lecomte, J.-P., Kirsch-De Mesmaeker, A. & Orellana, G. Photoinduced electron transfer from nucleotides to DNA intercalating viologens A study by laser-flash photolysis and spectroelectrochemistry. *J Photochem Photobiol B* **36**, 67–76 (1996).
45. Keane, P. M. & Kelly, J. M. Transient absorption and time-resolved vibrational studies of photophysical and photochemical processes in DNA-intercalating polypyridyl metal complexes or cationic porphyrins. *Coord Chem Rev* **364**, 137–154 (2018).
46. Rhee, S., Han, Z., Liu, K., Miles, H. T. & Davies, D. R. Structure of a Triple Helical DNA with a Triplex–Duplex Junction. *Biochemistry* **38**, 16810–16815 (1999).
47. Ambrosek, D., Loos, P.-F., Assfeld, X. & Daniel, C. A theoretical study of Ru(II) polypyridyl DNA intercalators. *J Inorg Biochem* **104**, 893–901 (2010).
48. Li, G., Sun, L., Ji, L. & Chao, H. Ruthenium(II) complexes with dppz: From molecular photoswitch to biological applications. *Dalton Transactions* **45**, 13261–13276 (2016).
49. Da Ros, T., Spalluto, G., Boutorine, A. S., Bensasson, R. V. & Prato, M. DNA-Photocleavage Agents. *Curr Pharm Des* **7**, (2001).
50. Novopashina, D. S. *et al.* Sequence-Specific Conjugates of Oligo(2'-O-methylribonucleotides) and Hairpin Oligocarboxamide Minor-Groove Binders: Design, Synthesis, and Binding Studies with Double-Stranded DNA. *Chem Biodivers* **2**, 936–952 (2005).
51. Wang, E., Koshlap, K. M., Gillespie, P., Dervan, P. B. & Feigon, J. Solution structure of a pyrimidine-purine-pyrimidine triplex containing the sequence-specific intercalating non-natural base D3. *J Mol Biol* **257**, 1052–1069 (1996).
52. Lauria, T. *et al.* A Click Chemistry Approach to Developing Molecularly Targeted DNA Scissors. *Chemistry – A European Journal* **26**, 16782–16792 (2020).

## CHAPTER 6 - Crystallisation of a DNA quadruplex from a TFO

### Contribution Statement

Structure **8RMH** was collected by Dr Christian Orr, and solved by Dr Kamel El Omari, Dr James Hall, and Ahmad Abdullrahman. Crystal grown by Ahmad Abdullrahman.

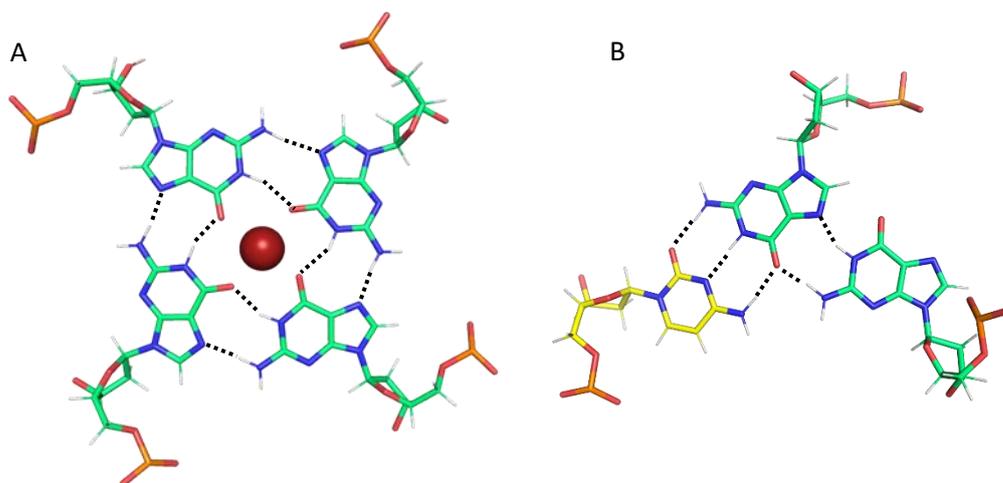
**8RMH** was deposited in the Protein Data Bank with the entry title: Crystal structure of parallel G-quadruplex containing T-tetrads and TG-octaplet.

## 6.1 Introduction

DNA triplexes can adopt either a parallel or antiparallel configuration, depending on the orientation of the third strand (TFO) relative to the polypurine strand of the duplex. In various studies, parallel and antiparallel DNA triplexes have been investigated for their stability, dynamics, and biological relevance.<sup>1</sup> In parallel triplexes, C-G:C triplets rely on low pH for protonation and stability, while antiparallel triplets with triplet hydrogen bonding in A-A:T, G-G:C and T-A:T motifs are stable at neutral pH without the protonation restriction.<sup>2,3</sup>

Antiparallel triplexes have shown greater stability under neutral pH conditions compared to the parallel triplexes; however, results vary depending on factors such as the origin of the third strand, either inter- or intramolecular, ionic, and nucleobase content.<sup>4-6</sup> Molecular modelling showed the effect of the sugar distortion that energetically favours the formation of reverse-Hoogsteen bonding for antiparallel triplexes.<sup>7</sup> Notably, nonisostericity, defined as residual twist and radial difference, between the parallel triplets, causes conformational effects and disruption of hydrogen bonds in parallel triplexes. This effect is less pronounced in antiparallel triplexes, conferring them higher stability advantages. Structural changes are further altered by the base composition of the underlying duplex forming Hoogsteen and non-Hoogsteen hydrogen bonds.<sup>8,9</sup>

Triplex-forming oligonucleotides can contain G-rich repeats that can result in the formation of secondary structures. Oligonucleotides with long tracts of guanine tend to form G-quadruplex conformations.<sup>10</sup> G-Quadruplexes (G4) are nucleic acid secondary structures formed by  $\pi$ - $\pi$  stacking of a minimum of two G-quartets. Each G-quartet consists of four guanines bound in a cyclical fashion. This nucleobase arrangement (also called G-tetrads) is formed through Hoogsteen hydrogen bonding. Specifically, N1 atoms of one guanine form a Hoogsteen hydrogen bond with O6 of adjacent guanine, while the N7 atoms form a Watson-Crick hydrogen bond with the N2 atom of another neighbouring guanine Figure 6.1.



**Figure 6.1** Structure of G-quartet (**1KF1**) (A) and G-G:C triplet (**135D**) (B). Guanine shown in green, and cytosine in yellow. Hydrogen bonds in dashed line between nitrogen atoms (blue) and oxygen (red). Potassium in red.

This arrangement is further supported by cations, centrally coordinated in the stacking core of the G-quartet. Recent extensive biophysical and structural studies have focused on these secondary structures due to their association with a variety of biological functions. G-rich sequences capable of forming G-quadruplexes are found near telomeric ends as well as in promoters that play a role in cancer growth and proliferation.<sup>11</sup>

To prevent self-association of the TFO, chemical base modifications have been developed with the substitution of the N7 atom, necessary for the G-quartet formation, with 7-deazaguanine and 6-thio-7deazaguanine, however, the triplex formation was not observed. Alternative successful approaches involve the insertion of monomers to prevent the formation of quadruplexes, this includes Twisted Intercalating Nucleic Acids (TINA) that enhance stability formation of DNA triplexes while inhibiting G-quadruplex assembly.<sup>12</sup>

In Chapter 2, a library of DNA triplexes with different triplex motifs was designed for crystallisation analysis. Among them, the AA-t3 system is composed of a 7-mer short duplex and a G-rich TFO that binds in an antiparallel configuration. During crystallisation trials, the unbound  $d_{\text{TFO}}$  oligonucleotide underwent a self-assembly process resulting in the formation of a quadruplex structure. Here, the characterisation of the crystal structure formed from the AA-t3 TFO strand and analysis of novel structural features is presented.

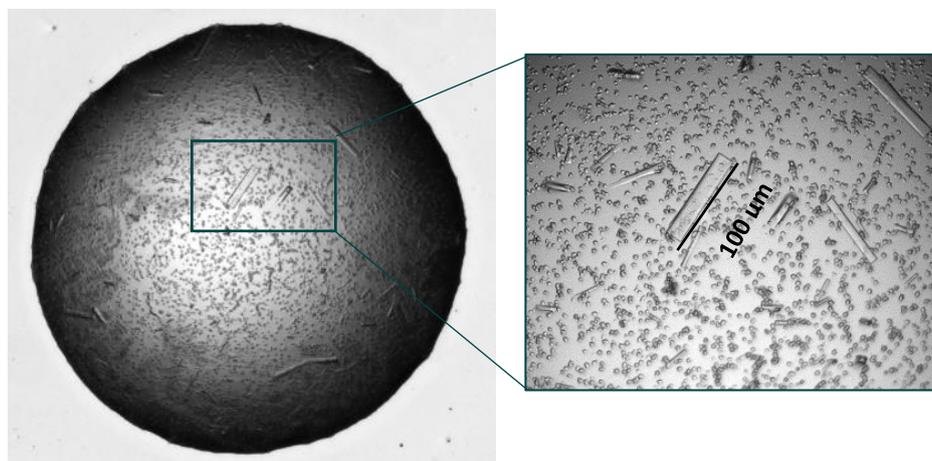
## 6.2 Materials and Methods

### 6.2.1 Oligonucleotide, buffer solution, and crystallisation reagents

Oligonucleotides were sourced from Eurogentec Ltd as HPLC-purified solids. All buffer components were purchased from Sigma Aldrich. For crystallisation, screens were purchased from Hampton Research Comp. and plates were provided by Diamond Light Source from MiTeGen LLC.

### 6.2.2 Crystallisation

The solution that resulted in the crystallisation of the oligonucleotide  $d_{TFO}(TGGTGGT)$  was prepared along with  $d_1(AGGAGGA)$  and  $d_1(TCCTCCT)$  with a ratio of 1:1:1. The DNA at a final concentration of 1.8 mM was annealed with 0.02 M Sodium Cacodylate pH 5.5, 0.1 M NaCl, 0.01 mM  $MgCl_2$  at 90 °C for 10 min and left cooling overnight. The annealed mixture was mixed at an equal volume of 100 nl with the screen condition containing 0.2 M Potassium chloride, 0.05 M Sodium cacodylate trihydrate pH 5.9, 10% w/v Polyethylene glycol 4,000 & 0.01 M Magnesium chloride hexahydrate. The final sitting volume was 200 nl and dispensed with the Mosquito LCP instrument. The plates were left at 20 °C and after 69 days white rod crystals reaching 100-130  $\mu m$  long appeared along with microcrystals Figure 6.2.

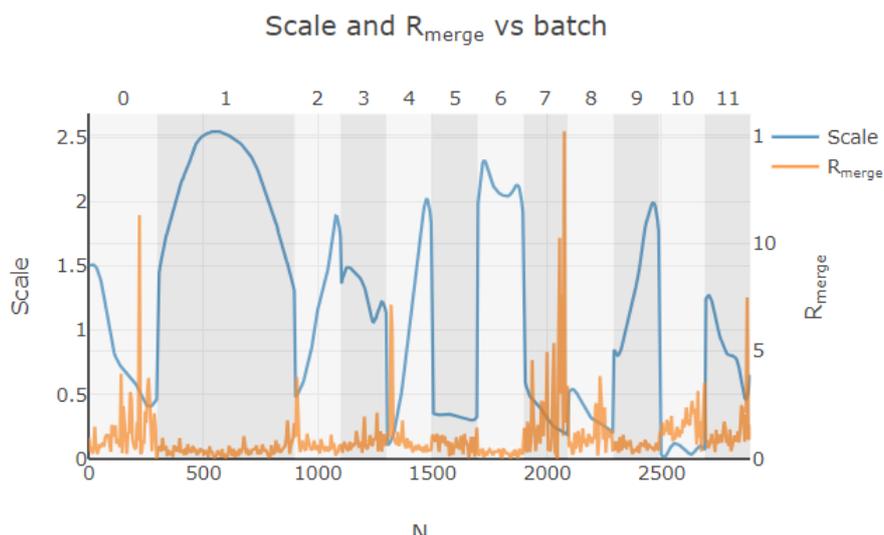


**Figure 6.2** Sitting drop from 2-drop MiTeGen plates containing rod crystals of 8RMH and microcrystals with example of 100  $\mu m$  crystal. Crystals were grown at 20 °C for 69 days.

### 6.2.3 Data collection, refinement, and analysis

The data were collected as a result of the initial screening at room temperature at VMXi beamline at Diamond Light Source.<sup>13</sup> The in-situ data collection with grid scan setting at 10-micron steps, 100% transmission using radiation with a wavelength of 0.774 Å resulted in a resolution and diffraction pattern that indicates similarity with DNA diffraction crystals. Since the VMXi beamline allows the collection of high-quality data from multiple crystals, the crystal condition was repeated to cover a full plate of 96 drops.

Data from multiple crystals were thus collected on VMXi with a point collection setting with beam rotation of 60° at 5% transmission using radiation at a wavelength of 0.774 Å with an exposure time of 0.0018° per frame. The automatic processing of data using Xia2<sup>14</sup> with the DIALS database<sup>15</sup> was used to index, integrate, and scale the multiple datasets collected resulting in 12 Multiplex Data. From the multiplex report, the  $R_{\text{merge}}$  statistics shown in Figure 6.3 indicate that the quality data collected for certain crystals was drastically higher. As the  $R_{\text{merge}}$  values are indicative of the data quality, datasets 0 and 7 were removed before reprocessing. The failed attempts to obtain a final structure using the space group indicated that the model used in molecular replacement could be incorrect. Re-indexing to the space group  $P1$  failed as well to successfully resolve the data.

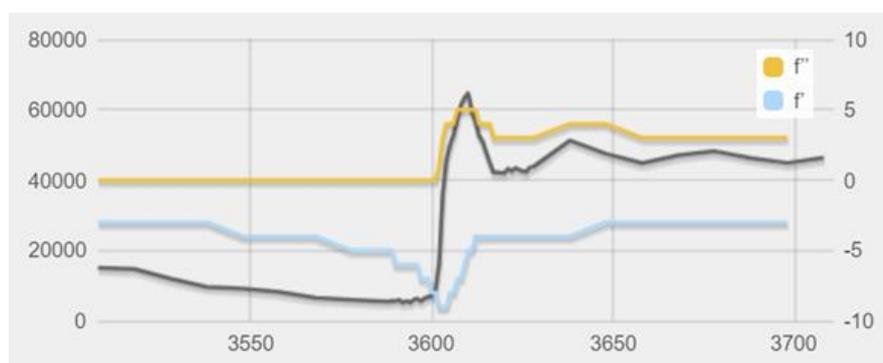


**Figure 6.3** Result from processing data. Comparison of the Scale and  $R_{\text{merge}}$  for the 12 datasets merged as part of the Xia2 multiplex processing.

Crystals also grew in crystallisation plates prepared at 4 °C with the same drop composition. Further data from crystals of AA-t3 were collected on beamline I03 at Diamond Light Source using a radiation wavelength of 0.946 Å from the flash-cooled crystal at 100K. 3600 images were

generated from collection with an oscillation angle of  $0.1^\circ$  per frame. Data were processed using DIALS<sup>15</sup> through the Xia2 pipeline<sup>14</sup> which resulted in a higher resolution limit of  $1.18 \text{ \AA}$  and a lower percentage of completeness. The most likely space group suggested was  $P 4_2 2_1 2$ . The data were processed with ACORN system<sup>16</sup> from CCP4 package<sup>17</sup>. The direct method through ACORN gave an initial density map that showed the configuration of the TFO in a quadruplex configuration. Although the map was good for an initial model fitting, refinement could not proceed and only resulted in a poor-quality map with very high  $R_{\text{factors}}$  ( $>45\%$ ).

The crystal AA-t3 was prepared without the inclusion of metals with atomic above 20 (heavy metals). However, in X-ray fluorescence (XRF) scan at longer wavelengths (10 KeV) the emission of potassium atoms illustrated in Figure 6.4 shows a peak at 3609.0 eV. The anomalous scattering factors  $f''/f'$  were determined as 5.77 and  $-9.73$  electrons, respectively (Figure 6.4).



**Figure 6.4** K-Edge Scan of the AA-t3 crystal. Peaks are shown for  $f'$ :  $-9.73 \text{ e}$  (blue),  $f''$ :  $5.77 \text{ e}$  (yellow), and emission peak at  $3610.0 \text{ eV}$  (black).

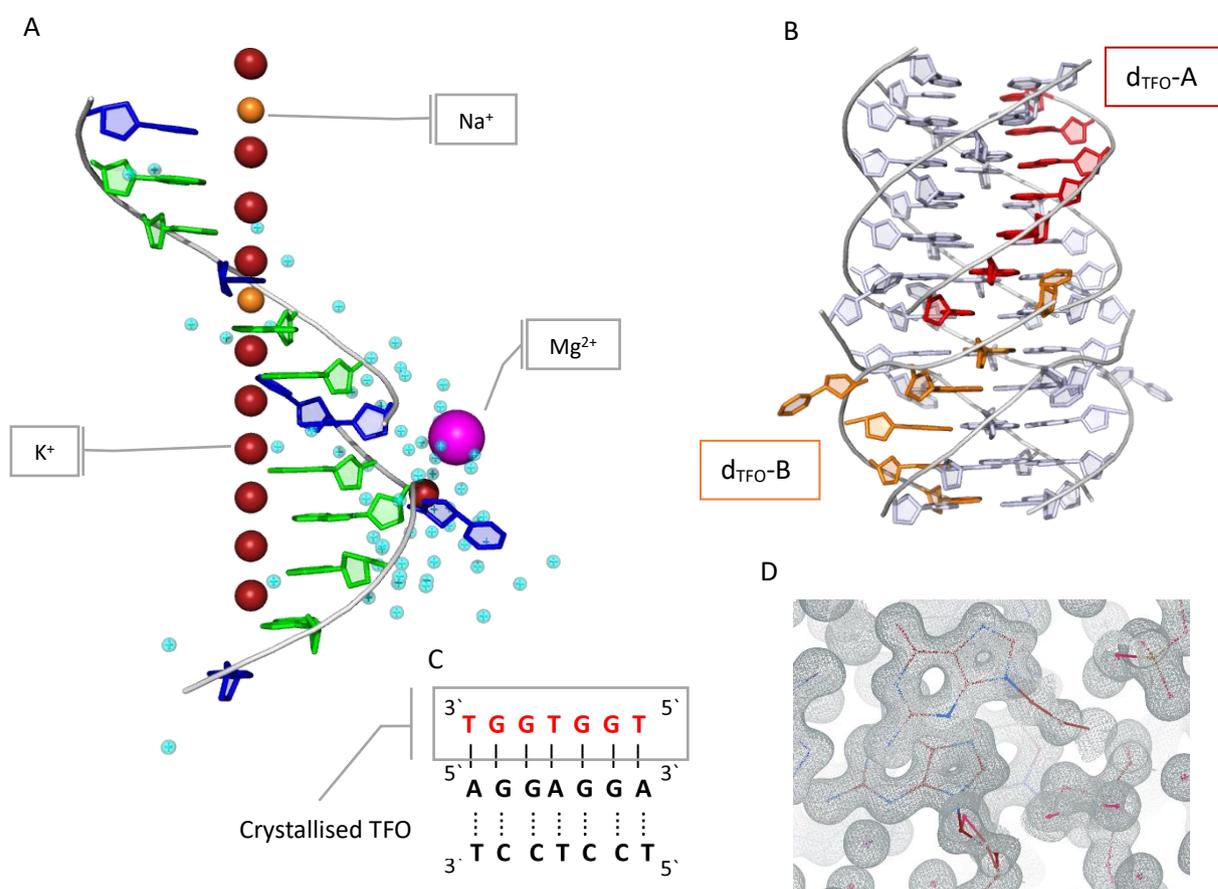
The phasing was determined by single-wavelength anomalous diffraction (SAD) using the anomalous scattering of potassium. This was possible at the Long-wavelength Beamline I23 at Diamond Light Source which has access to lower energies around the K-edge.<sup>18</sup> Data were collected from several crystals at a wavelength of  $1.1271 \text{ \AA}$  with the Pilatus 12M detector. Datasets were then merged using XSCALE included in the Xials package<sup>14</sup> giving 12480 unique reflections. The data were phased using ACORN within the CCP4 suite<sup>17</sup> The final space group  $P 4_2 2_1 2$ , confirmed the data processed at I03, where half of the asymmetric unit was visible in the electron density map. The final model was built using Coot<sup>19</sup> and refined with PHENIX software package.<sup>20</sup> The data and final coordinates are deposited in the Protein Data Bank<sup>21</sup> (PDB: **8RMH**). The data collection and refinement statistics for **8RMH** are in Table 6.1.

**Table 6.1** Data processing and refinement statistics for *8RMH*.

<b>Crystallisation Parameters</b>	
DNA Sequence	d <sub>TFO</sub> (TGGTGGT)
Crystallisation Temperature, °C	20
Growth Time (month)	2.2
Crystal Morphology	rod
<b>Data Collection</b>	
Diffraction source	I23
Radiation wavelength, Å	1.1271
Temperature, K	100
Exposure time, s	0.3
Detector	Pilatus 12M
Resolution, Å	1.15 (40.48)
<b>Data Processing</b>	
Structure Solution Method	SAD
Space group	<i>P</i> 4 2 <sub>1</sub> 2
a, b, c, Å	40.43, 40.43, 40.47
α, β, γ, Å	90, 90, 90
Resolution, Å	1.15 (40.48)
Total reflections	1581764 (62267)
Unique reflections	12480 (589)
<i>R</i> <sub>meas</sub>	0.400 (7.865)
<i>R</i> <sub>merge</sub>	0.399 (7.828)
<i>R</i> <sub>pim</sub>	0.035 (0.734)
Mean I/σI	0.353
CC <sub>1/2</sub>	1.00 (0.567)
Completeness, %	100
Multiplicity	126.7 (105.7)
Average B factors, Å <sup>2</sup>	8.52
Data Collection Date	18-Sep-23
*Outer shell statistics shown in parentheses	
<b>Refinement</b>	
No. Reflections	12352 (1167)
<i>R</i> <sub>work</sub>	0.1514
<i>R</i> <sub>free</sub>	0.1900
<b>Number of components</b>	
Nucleotide	14
Ligands	14
Water	60
<b>rmsd</b>	
Bond Lengths, Å	0.0011
Bond Angles, °	1.28

### 6.3 Results and Discussion

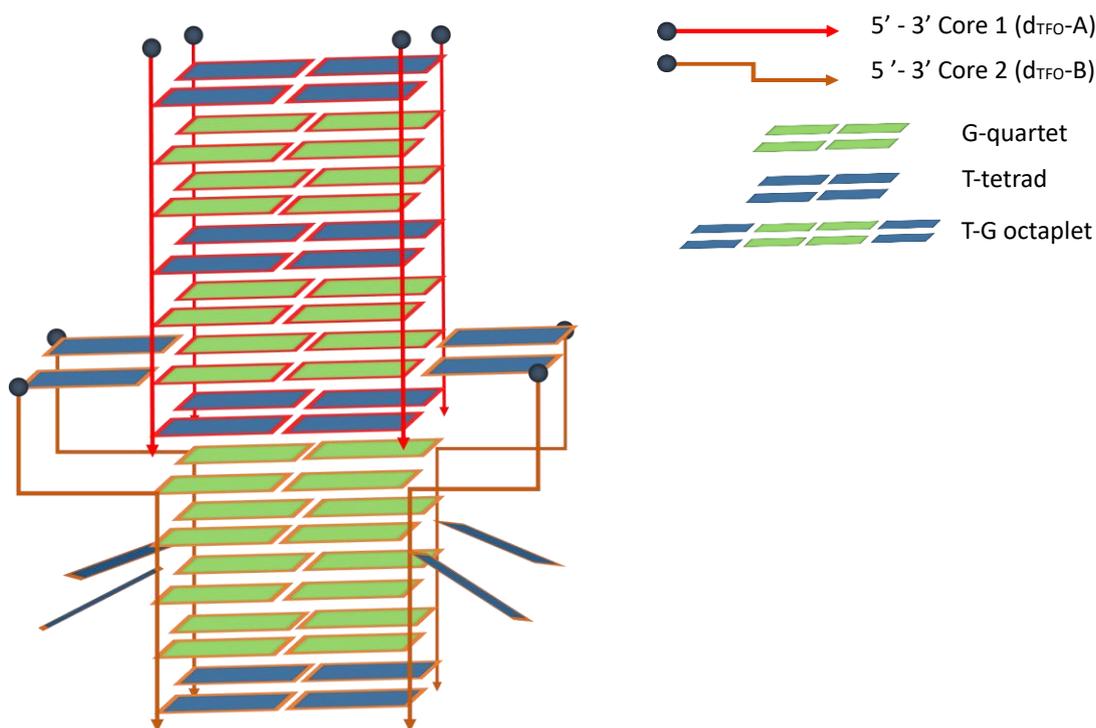
Despite the sample prepared in the presence of three oligonucleotides, which would have formed an antiparallel triplex, the crystallisation events yielded crystals in which only one of the three strands was present. The  $d_{TFO}$  sequence self-assembled and formed a peculiar parallel G-quadruplex structure. Figure 6.5A shows the asymmetric unit containing two  $d_{TFO}$  strands binding at a single base pair only coordinated with 11  $K^+$  ions, two  $Na^+$  ions and one  $Mg^{2+}$  ion, hydrated with 79 water molecules. The two strands in the asymmetric unit  $d_{TFO-A}$  and  $d_{TFO-B}$  bind other symmetry-related molecules forming a G-quadruplex.



**Figure 6.5** (A) The asymmetric unit of **8RMH** contains two  $d_{TFO}(TGGTGGT)$ , 11  $K^+$  ions (red), two  $Na^+$  ions (orange) and one  $Mg^{2+}$  ion (magenta) and 79 water molecules (cyan). Guanine shown in green, and thymine in blue. (B) Two strands  $d_{TFO-A}$  (red) and  $d_{TFO-B}$  (orange) in the asymmetric unit bind to six symmetry-related molecules (grey). (C) Schematic representation of AA-t3 from which only the TFO (red) was crystallised. (D) Representative example of electron density ( $2F_o-F_c$ ) is contoured at the  $1\sigma$  level (grey).

### 6.3.1 The self-assembly of triplex-forming oligonucleotide

The biological assembly shows a quadruplex structure with two dimers, Core 1 and Core 2, stacked on top of each other and connected by non-canonical pairing in the interface region (Figure 6.6). The guanine  $G_2$  from  $d_{TFO-A}$  forms two groups of non-canonical hydrogen bonds, the first with thymine  $T_7$  at the 3'-end of  $d_{TFO-B}$  molecule generating a  $G_2:T_7$  non canonical base pair.  $G_2$  then forms quartets with guanine from symmetry-related molecules resulting in a T-G octaplet.

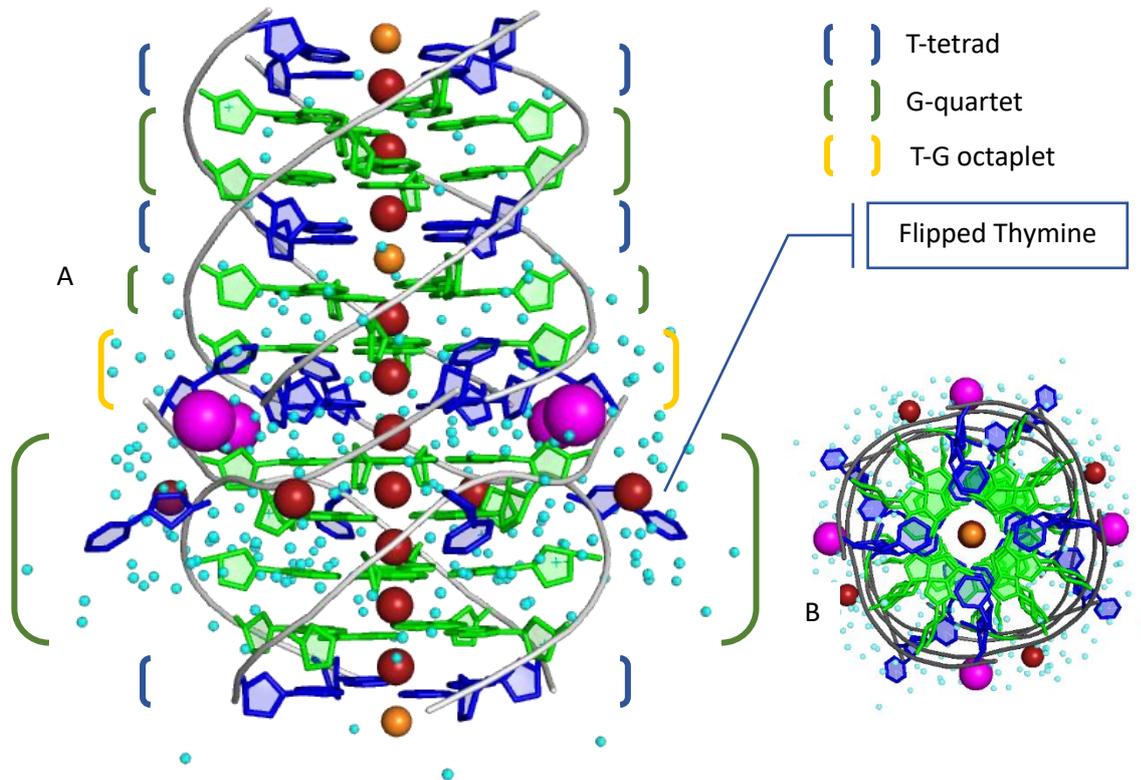


**Figure 6.6** Schematic representation of the quadruplex structure

In Core 1 two G-quartets are stacked between a two T-tetrad. Additionally, a third G-quartet is positioned on top of a G-quartet which  $G_2$  forms a T-G octaplet with  $T_1$  in the interface region. The second core instead shows four consecutive G-quartets stacking between  $T_7$ -tetrad and  $T_1$ -tetrad from  $d_{TFO-A}$ ; thus,  $d_{TFO-B}$  will only form one T-tetrad, while its  $T_4$  shows a flipping position (Figure 6.7).

The total 12 quartets of the two cores have identical polarity, defined by the hydrogen bonding donor to acceptor direction. In addition, the model contains 10 potassium ions well-ordered in the core of the quadruplexes along with three sodium ions. Other four potassium ions and four magnesium ions can be found around the interface region further stabilising the assembly. The

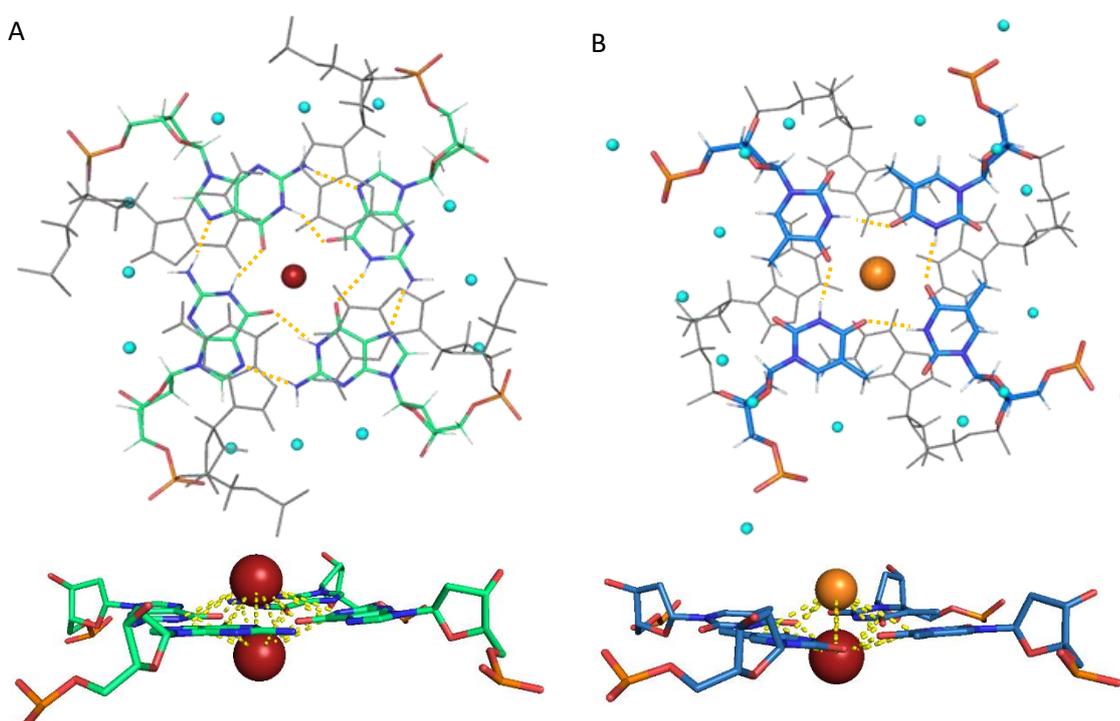
high number of ions contributes to the overall hydration of the crystal structure resulting in a total of 316 water molecules, mostly localised externally of the quadruplex in the presence of  $K^+$  and  $Mg^{2+}$  ions.



**Figure 6.7** (A) Model **8RMH** contains seven  $d_{TFO}$  strands forming seven G-quartets, three T-tetrads, one T-G octaplet and four thymine flipped out. 10  $K^+$  ions (red) and two  $Na^+$  ions (orange) are localised in the quadruplex. (B) Schematic representation of AA-t3 from which only the TFO (red) was crystallised. Four  $K^+$  ions (red) and four  $Mg^{2+}$  ions (magenta) are localised outside the structure. The total number of water molecules (cyan) are 316. Guanine shown in green, and thymine in blue.

G-rich sequences form G-quadruplexes with a variety of conformation and morphology depending on their base content. Commonly studied are intramolecular G4, composed of a unimolecular strand that folds assuming parallel, anti-parallel, or hybrid backbone arrangement.<sup>22</sup> Other conformations are possible as G-rich strands can assume an intermolecular folding through dimerization or the tetramerization of four strands. The latter conformation is described here with four  $d_{TFO}$ (TGGTGGT) forming an interstrand tetramolecular quadruplex. Although it is in discussion whether they are biologically relevant, these structures are commonly used as a model of telomeric ends *in vitro* as they present a slow dissociation rate.<sup>23</sup>

Finally, the glycosidic bond of the guanine can adopt two conformations depending on the topology of the quadruplex, *syn* or *anti* with the torsion angle  $\chi$  -90 to 90; 90 to 180, and -90 to -180, respectively. Both the guanine and thymine in the quadruplex core of the model adopt an *anti* conformation in agreement with the parallel topology<sup>24</sup> (Table A6.4). In Figure 6.8 are illustrated G-quartets and T-tetrads from the model **8RMH**. The groove width is visibly larger in the Core 2 region with the overall groove of the quadruplex being on average 10.5 Å wide, in accordance with the medium width described for parallel quadruplexes.<sup>25,26</sup>



**Figure 6.8** (A) Structure of G-quartet stacked on G-quartet (grey) and coordinated in dashed lines by  $K^+$  ions (red). (B) Structure of T-quartet stacked on G-quartet and coordinated in yellow dashed lines by  $K^+$  ion (red) and  $Na^+$  ion (orange). Guanine shown in green, thymine in blue, water molecules in cyan.

Quadruplex structures are not limited to G-rich sequences, but at least one non G-tetrad may be accommodated. In fact, a variety of G-quadruplexes non-entirely formed by G-tetrads are reported for A-tetrads, C-tetrads, T-tetrads as well as for U-tetrads in RNA structures.<sup>27</sup> The crystal structure **6A85** is an example of a DNA G-quadruplex composed of a combination of all the non-G-tetrads.<sup>28</sup>

In the **8RMH** model, T-tetrads are formed at both extremities of the quadruplex, as well as stacked between G-quartets. A sequence similar to **8RMH**, with the T<sub>7</sub> substituted with a C, 5'-

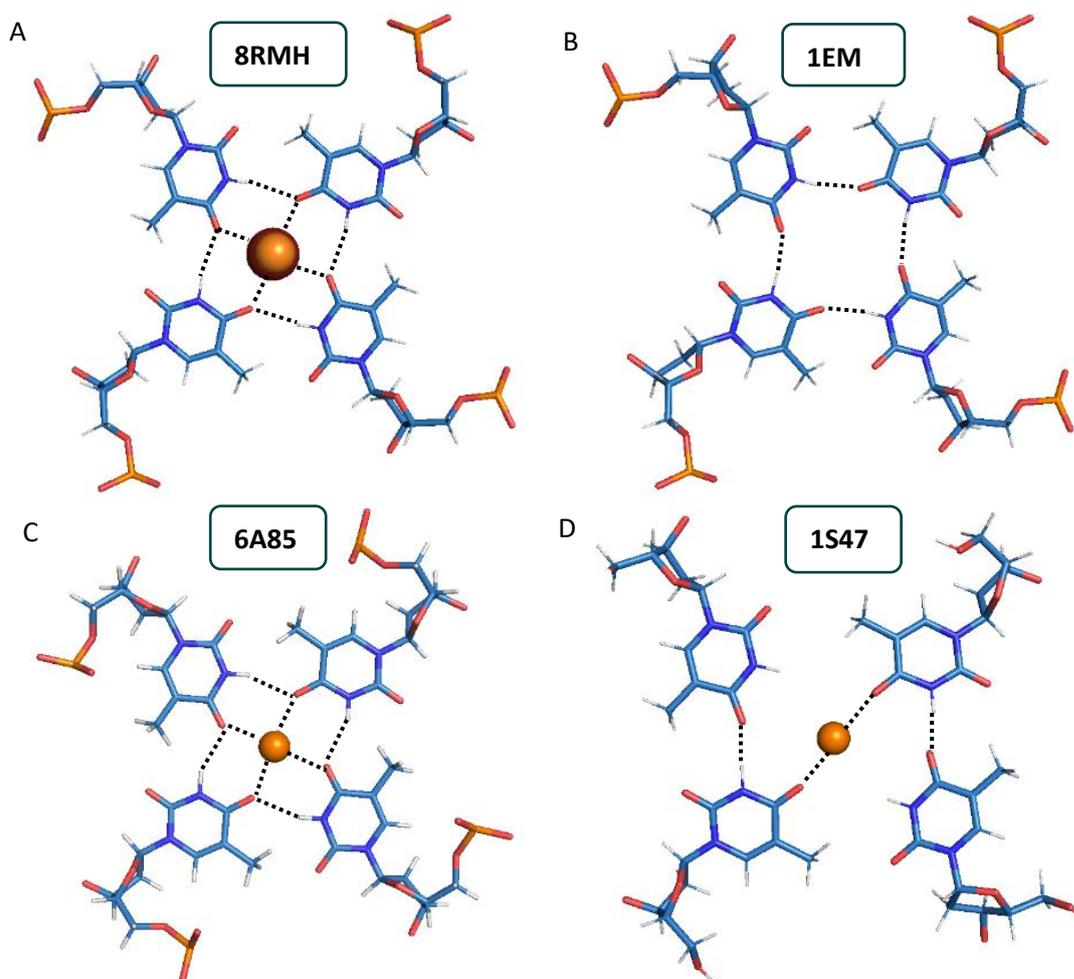
TGGTGGC-3' (**1EMQ**) was found to form a quadruplex structure. The NMR structure **1EMQ** shows that none of the bases flips out, instead reports T-tetrads stacking between G-quartets and at the terminal position.<sup>29</sup> Whereas, if the central T<sub>4</sub> is removed the final sequence 5'-TGGGGT-3' gives rise to DNA quadruplexes whose crystal structures **1S47** and **1S45**, show more conformational similarity with **8RMH**, though with T-tetrads visible only the 5'-end.<sup>30</sup>

Each thymine forms a hydrogen bond with the N3 and O4 atoms of adjacent thymine with an average of 2.3 Å, agreeing with the NMR **1EMQ** and X-ray solved structure **6A85**. Comparison with the T-tetrad from **1S47** shows a different thymine distortion, with similar N3 and O4 distances for two thymine only. Figure 6.9 shows a comparison with the different T-tetrads stabilised by monovalent ions Na<sup>+</sup> and K<sup>+</sup>. The NMR structure **1EMQ** does not show any metal ion since Na<sup>+</sup> or K<sup>+</sup> are difficult to detect through nuclear magnetic resonance. Kinetic studies show that the addition of non-G quartets can have implications on the association rate of the quadruplex with A-quartets resulting in an increased association rate than T-tetrad at the terminal position. However, additional T-tetrads at the terminal 5' or 3'-end have increased thermodynamic stability slowing the dissociation rate.<sup>23</sup> **8RMH**, showing T-tetrads at the end of the quadruplex could explain the similarity between TGGGT sequences in forming quadruplexes as the fluorescein is attached at either the 5' or 3'-end in FRET analysis.<sup>31</sup>

### 6.3.2 Stacking of a novel octaplet

As intramolecular G-quadruplexes are mostly studied because of their biological role,<sup>32</sup> thymine is often found to flip out facilitating the G-quartet formation and contributing to the stability of the structure.<sup>33</sup> Due to the extensive polymorphism among G-quadruplexes, here the comparison is limited to structures containing the palindromic TGGTGGT motif, excluding protein-bound complexes, as shown in Table 6.2. Current structures deposited in the Protein Data Bank<sup>21</sup> show diverse secondary structures with the TGGTGGT motif, including quadruplexes forming only intramolecular conformations reported to date. **8RMH** shows thymine T<sub>4</sub>-d<sub>TFO</sub>-B flipping out allowing the consecutive stacking of 4 G-quartets, like other structures containing a similar sequence where thymine frequently occupies the solvent allowing G-quartet stacking. In some cases, thymine can be found forming non-canonical base pairs, T:T or G:T (**4U5M** and **6JCD**, respectively). Additionally, triplet formation is being reported in both **6QJO** and the **1S47** structure

previously described, with T-G-T and T-triad, respectively. Thermodynamic analysis of unimolecular quadruplexes indicates that flipping out of thymine is influenced by molecular crowding, the rigidity of the quadruplex core, and the water environment.<sup>34</sup>

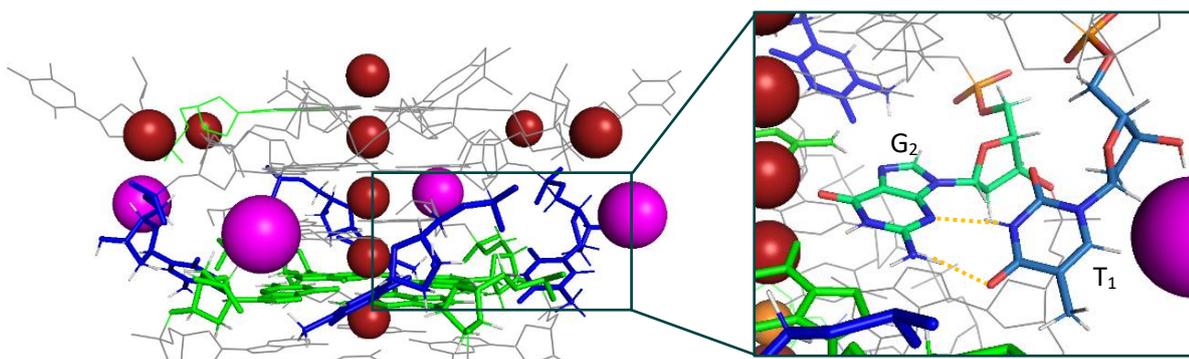


**Figure 6.9** Comparison of T-tetrads obtained from **8RMH** (A), **1EMQ** (B), **6A85** (C) and **1S47** (D). Na<sup>+</sup> ion (orange) coordinates the X-ray structures **6A85**, **1S45** and in **8RMH**, in the latter K<sup>+</sup> ion (red) is shown behind the sodium. Hydrogen bonds in dashed line between nitrogen atoms (blue) and oxygen (red).

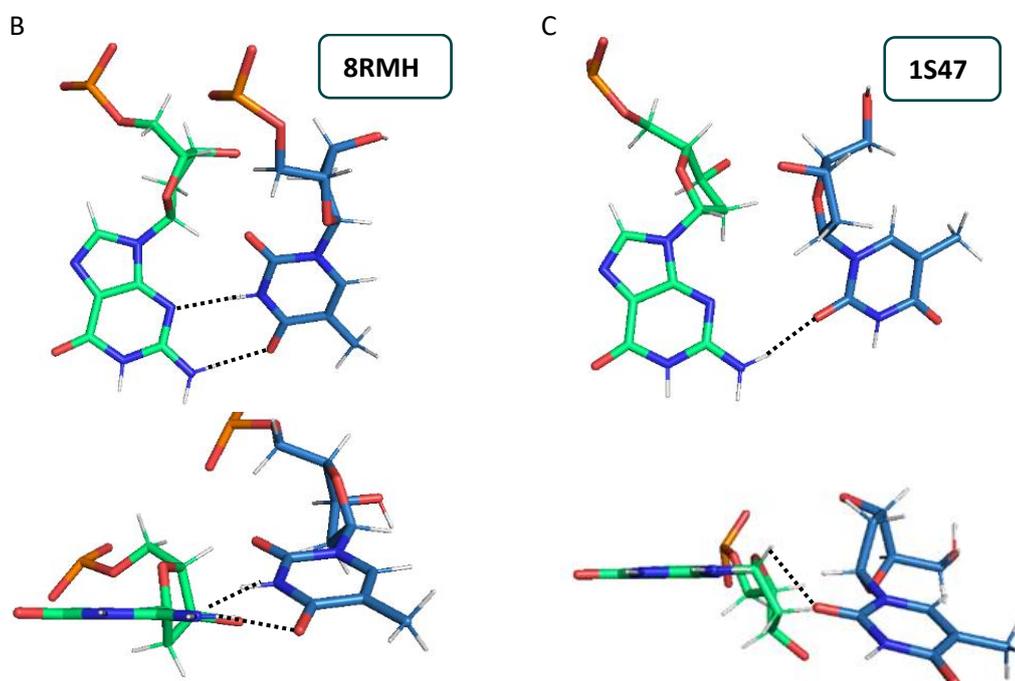
**Table 6.2** Structures deposited in the Protein Data Bank containing the DNA motif TGGTGGT.

Sequence	Secondary structure	Method	PDB ID	Year	Ref.
TCCTCCTTTTTAGGAGGATTTTT <b>TGGTGGT</b>	Intramolecular Triplex	NMR	134D	1993	35
TCCTCCTTTTTAGGAGGATTTTT <b>TGGTGGT</b>	Intramolecular Triplex	NMR	135D	1993	35
TCCTCCTTTTTAGGAGGATTTTT <b>TGGTGGT</b>	Intramolecular Triplex	NMR	136D	1933	35
<b>TGGTGGT</b> GGTGGTTGTGGTGGTGGTGT	Intramolecular G4	NMR	2MS9	2014	36
<b>TGGTGGT</b> GGTGGTTGTGGTGGTGGTGGT	Intramolecular G4	NMR	2N3M	2015	36
TGGTTGTGGTTTGGTGT <b>TGGTGGT</b>	Intramolecular G4	NMR	6JCD	2019	37
GGTTGGTGTGGTTGGTTGTGGT <b>TGGTGGT</b> G	Intramolecular G4	NMR	6JCE	2019	38
<b>TGGTGGT</b> GGTGGTTGTGGTGGTGGTTGT	Intramolecular G4	NMR	6KVB	2019	39
GGTGTGTGTGTGTGT <b>TGGTGGT</b> GGTG	Intramolecular G4	NMR	7D5F	2020	40
<b>TGGTGGT</b> GGTGGTTGTGGTGGTGGTGT	Intramolecular G4	X-ray diffraction	4U5M	2014	41
GTGGT <b>TGGTGGT</b> G	Intramolecular G4	X-ray diffraction	6FQ2	2018	42
GTGGT <b>TGGTGGT</b> GTTGTGGTGGTGGTGT	Intramolecular G4	X-ray diffraction	6GZ6	2018	42
GGTTGGTGTGGTTGGTGT <b>TGGTGGT</b> GGTG	Intramolecular G4	X-ray diffraction	6QJO	2019	43
GGTGTGT <b>TGGTGGT</b> GTGGTGGTGGTGT	Intramolecular G4	X-ray diffraction	7D5D	2020	44
GGTGTGTGTGTGGTGT <b>TGGTGGT</b> GGTGT	Intramolecular G4	X-ray diffraction	7D5E	2020	44
GGT <b>TGGTGGT</b> GTGTTGGTGGTGGTGTG	Intramolecular G4	X-ray diffraction	7DFY	2020	45
CGCT <b>TGGTGGT</b> TCGA	Duplex DNA	X-ray diffraction	8EPB	2022	46
ACGCT <b>TGGTGGT</b> TCGA	Duplex DNA	X-ray diffraction	8EPF	2023	46

In the interface region the thymine at the 3' is flipped inward with a  $\gamma$  torsion of  $-69.5^\circ$  and its sugar assumes a *syn* conformation (Table A6.4). T<sub>7</sub>-d<sub>TFO</sub>-B is twisted to  $106.47^\circ$  to roll negatively ( $-108^\circ$ ) and forms non-canonical hydrogen bonds with G<sub>2</sub>-d<sub>TFO</sub>-A with a propeller and opening of  $33.73^\circ$  and  $102.64^\circ$ , Table A6.1-2. The hydrogen bonds are reported involving the (T) O4:N2 (G) and (T) N3:N3 (G) with a distance of 2.1 Å and 1.8 Å, respectively. The nitrogen N2 of the guanine then donates electrons to N7 allowing the formation of G-quartets (Figure 6.10). The structure **1S47**, previously described as possessing structural similarities with **8RMH** has shown one hydrogen bond between the (T) O2:N2 (G) atoms (Figure 6.11).

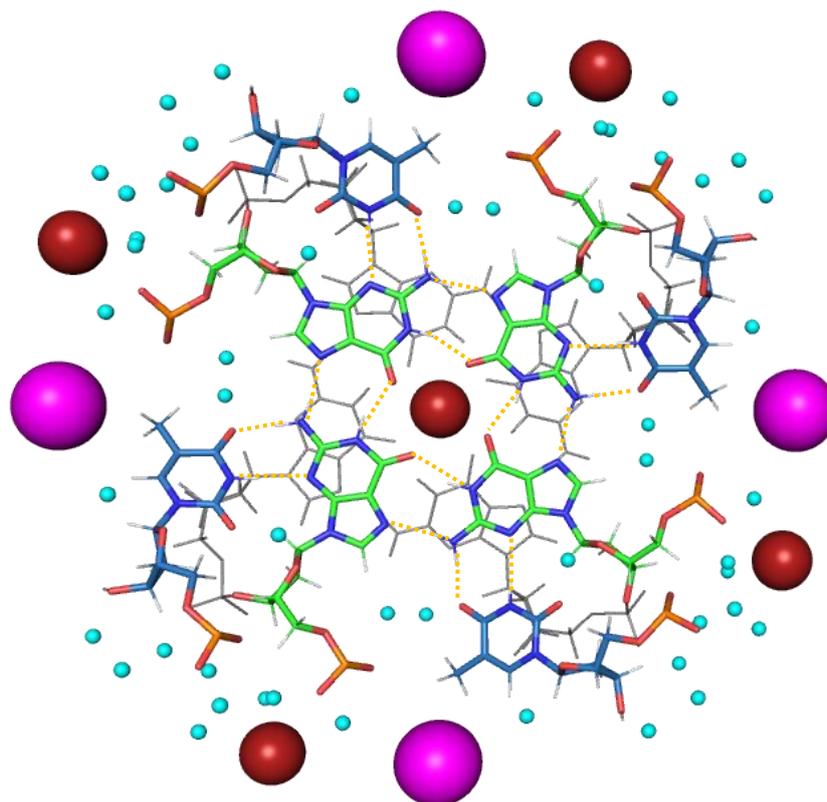


**Figure 6.10** Flipping inward of thymine in the interface of the structure **8RMH** with highlight to the  $T_7:G_2$  base pair. Guanine shown in green and thymine in blue. Hydrogen bonds in dashed line between nitrogen atoms (blue) and oxygen (red).



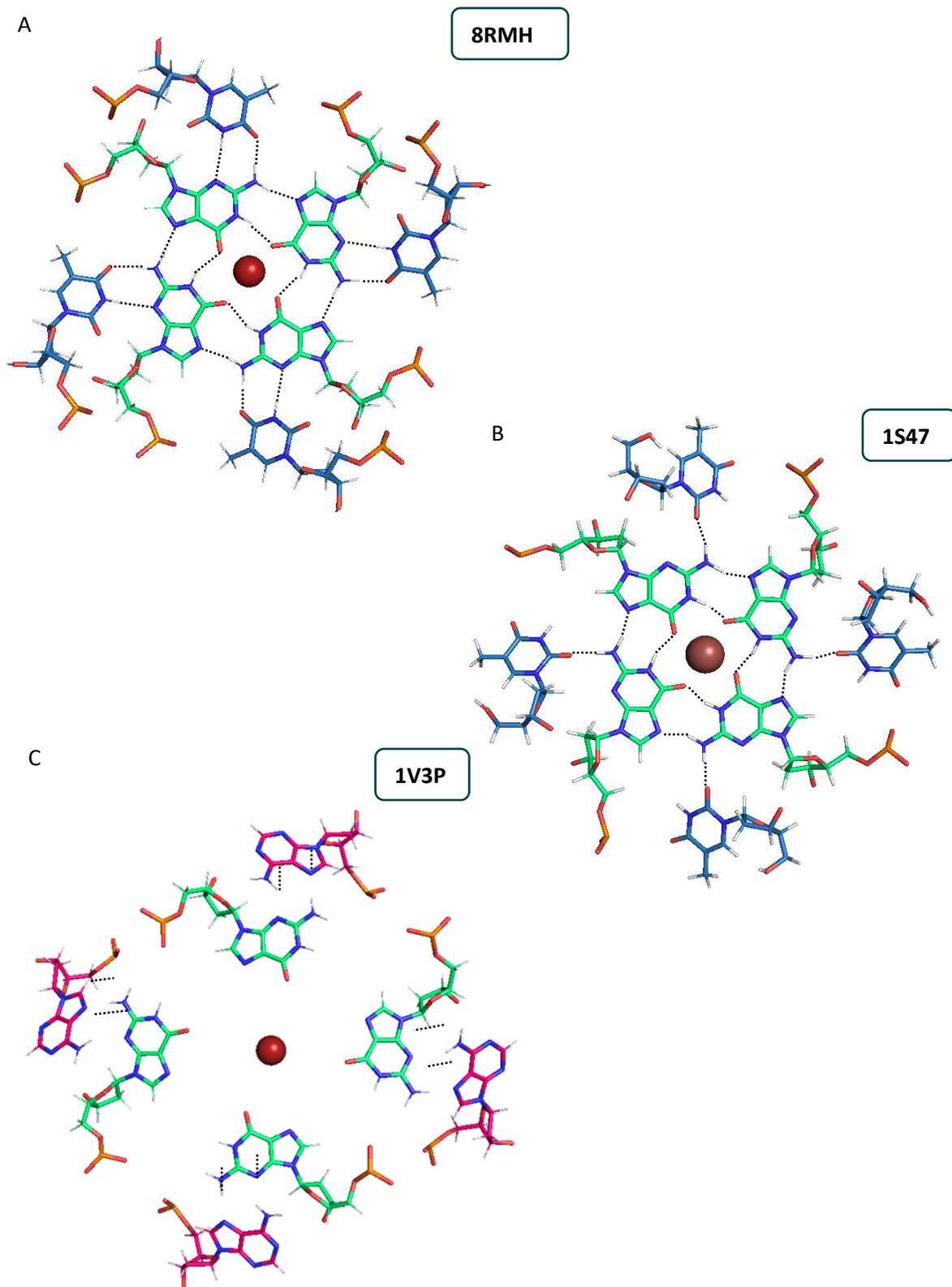
**Figure 6.11** Comparison of  $T:G$  base pair from structure **1VTT** (A), **8RMH** (B) and **1S47** (C). Guanine shown in green and thymine in blue. Hydrogen bonds in dashed line between nitrogen atoms (dark blue) and oxygen (red).

The binding of thymine  $T_7$  of guanines involved in the G-quartet results in nucleobases organised to form an octaplet configuration  $G(:T)-G(:T)-G(:T)-G(:T)$  (Figure 6.12). A similar configuration was reported in the **1S47** structure, although in this case, the thymine forms one hydrogen bond only, making the octaplet in the **8RMH** structure more stable (Figure 6.13) Nevertheless, both structures require a coordination atom to stabilise the octad organisation.



**Figure 6.12** Structure of the octaplex stacking on G-quartet (grey) and coordinated by K<sup>+</sup> ion (red) and stabilised by Mg<sup>2+</sup> ion (magenta). Guanine shown in green, thymine in blue, water molecules in cyan.

Previous crystallographic data show the formation of an octaplex formation within DNA duplex **1V3P**<sup>47</sup> (Figure 6.13). Authors show that at low K<sup>+</sup> levels, four intercalated DNA duplexes can assemble into an octaplex structure involving the non-canonical A:G base pairs, in which the guanine pairing is mediated by water molecules. As the ionic conditions are changed, the guanine forms two identical G-quadruplexes, thus splitting the octaplex. This octaplex is different from the one reported in the **8RMH** structure, where bases interact directly via hydrogen bonds and are coordinated by a K<sup>+</sup> ion without visible coordination by bridging water. Hence, it would be more appropriate to define this arrangement as the first octaplex X-ray structure since all the bases are connected forming an octaplex configuration via hydrogen bonds. Although, the role of the octaplex has not been fully elucidated structures show that this base configuration seems to play a role as an interface between two stacking quadruplexes, confirming an alternative model of quadruplex dimer stacking.<sup>48</sup>



**Figure 6.13** Comparison of octaplets from structure **8RMH** (A), **1S47** (B) and **1V3P** (C). Guanine shown in green, thymine in blue and adenine in magenta. Hydrogen bonds in dashed line between nitrogen atoms (blue) and oxygen (red).

### 6.3.2 Metal ions coordination

The crystallisation trials resulted in a crystal model that shows ions localised in the G-quadruplex core ( $K^+$ ,  $Na^+$ ) and counterions surrounding the structure ( $K^+$ ,  $Mg^{2+}$ ). The important role of monovalent cations in stabilising the G-quadruplex structure has been extensively studied. The cation stabilises the quadruplex in the order  $K^+ > Ca^{2+} > Na^+ > Mg^{2+} > Li^+$ .<sup>49</sup> The major focus has been on  $K^+$  cations as these are commonly found in higher concentrations in the intracellular environment.<sup>10</sup> Kinetic studies show that the presence of ions such as  $K^+$  and  $Na^+$  promotes the folding and stability of the quadruplexes.<sup>50</sup> The small ion radius of  $Na^+$  and  $K^+$  (0.95 Å and 1.33 Å, respectively), contributes significantly to the quadruplex stability.<sup>49</sup> Numerous crystal structures show these ions are found bound in the core of the quadruplex, by coordinating and forming non-polar bonds with O6 of the guanine O4 of thymine as shown in the **8RMH** structure, while  $Na^+$  ions are found coordinating with the T-tetrads only. Nevertheless, ions with a much larger ionic radius are reported as well, such as Tl (1.40 Å) which coordinate G-quartets as seen in the structure **1S47**. Potassium ions are also found in the solvent favouring a hydrated environment to bind the flipping base. Flipping of the bases is dependent on the  $K^+$  content, directly contributing to the final quadruplex topology.<sup>51</sup>  $Mg^{2+}$  ions are located at the intersection between the two strands reducing the electrostatic repulsion between the negatively charged phosphate backbones, thus allowing the T:G base pair formation.  $Mg^{2+}$  counterions are not found condensed elsewhere since the shape of the quadruplex is sufficient to diminish the electrostatic repulsion. This is supported by theoretical studies addressing counterion condensation on cylinder shapes.<sup>52</sup>

Furthermore, solution studies using similar high concentrations of  $K^+$  and  $Mg^{2+}$  ions as used in the crystallisation condition for **8RMH** showed that these cations can promote higher order association of G-quadruplexes.<sup>53</sup> The UV-vis analysis for AA-t3 reported in Chapter 2 did not show a two-transition melting profile as for other systems. Arguing that triplex formation did not happen in solution, even in the absence of  $K^+$ . The UV melting profile resembles the one obtained in the analysis of supramolecular structures.

The low concentration of  $MgCl_2$  compared to the high concentration of both  $NaCl$  and  $KCl$ , did not inhibit the binding of the other ions within the stacked G-quartets and T-tetrads.<sup>49</sup> These observations confirm that internal cations and external counterions have different effects on the overall quadruplex stability and that the final stability of the quadruplex structure is primarily determined by the internally bound ions.<sup>54</sup>

## 6.4 Conclusion

The high-resolution X-ray structure of a G-quadruplex obtained in this study is attributed to the self-assembly of the G-rich  $d_{TFO}$  sequence under high ionic conditions used during the crystallisation setup. Since all three strands were included in the crystallisation trial, it is hypothesised that the major contribution to the self-assembly is the heterogeneous content of ions that destabilised triplex structures and supported the formation of a stacked G-quartet core. Potassium ions are indeed well known to destabilise antiparallel triplex formation and favour the formation of G-quadruplexes, as discussed in Chapter 2.

Potassium and ion interaction with the DNA enables reversible structural transitions between hairpins, triplexes, and quadruplexes as demonstrated through kinetic studies.<sup>55 56</sup> G-triplexes formation was indeed considered in the presence of divalent ions, such as  $Mg^{2+}$  and  $Ca^{2+}$  along with the monovalent ions  $K^+$  and  $Na^+$ ,<sup>57</sup> although temperatures (37 °C) compared to the room temperatures used in this study. Quadruplexes and triplexes could also co-exist as G-quadruplex scaffolds for DNA bulges<sup>58</sup> and triplex bulges<sup>59</sup>. Substitution with non-guanine bases at the 3-end and 5' end was proposed to overcome the  $K^+$  effect<sup>60</sup>, however contrary to what was suggested the thymine, at each end of  $d_{TFO}$  in this instance formed T-tetrads further stabilising the quadruplex structure. Instead, extensions of the TFO with loops to connect the duplex and form an intramolecular triplex structure were reported for the same sequence Table 6.2.

Finally, **8RMH** shows the versatility of thymine within the G-quadruplexes, which has three structural behaviours: 1. formation of T-tetrads at both extremities and stacked between G-quartets 2. flipped-out occupying solvent channels, and 3. groove binding to G-quartets that together form a novel thymine-guanine octuplet motif.

This structure expands the structural repertoire of intermolecular G-quadruplexes, showing thymine-guanine octuplet whose biological significance remains to be fully elucidated. Inhibition of gene transcription through quadruplex disruption using DNA triplexes was previously reported, demonstrating the interplay between G-quadruplex and triplexes.<sup>61</sup> Nevertheless, controlling the TFO self-association and formation of the quadruplex due to its G-rich strand remains a challenge. This work suggests that more attention should be directed towards the TFO selection and the ionic condition to obtain triple-helical structures for therapeutic purposes.

## 6.5 References

1. Vasquez, K. M. & Glazer, P. M. Triplex-forming oligonucleotides: principles and applications. *Q Rev Biophys* **35**, 89–107 (2002).
2. James, P. L., Brown, T. & Fox, K. R. Thermodynamic and kinetic stability of intermolecular triple helices containing different proportions of C<sup>+</sup>·GC and T·AT triplets. *Nucleic Acids Res* **31**, 5598–5606 (2003).
3. Chandler, S. P. & Fox, K. R. Specificity of Antiparallel DNA Triple Helix Formation. *Biochemistry* **35**, 15038–15048 (1996).
4. Shchyolkina, A. K. *et al.* Parallel purine-pyrimidine-purine triplex: experimental evidence for existence. *FEBS Lett* **367**, 81–84 (1995).
5. Liquier, J. *et al.* Parallel and antiparallel G\*G.C base triplets in pur\*pur.pyr triple helices formed with (GA) third strands. *J Biomol Struct Dyn* **19**, 527–534 (2001).
6. Šponer, J. *et al.* Stabilization of the purine•purine•pyrimidine DNA base triplets by divalent metal cations. *J Biomol Struct Dyn* **16**, 139–143 (1998).
7. Dagneaux, C. *et al.* Parallel and Antiparallel A\*A-T Intramolecular Triple Helices. *Nucleic Acids Res* **24**, 4506–4512 (1996).
8. Goldsmith, G., Rathinavelan, T. & Yathindra, N. Selective preference of parallel DNA triplexes is due to the disruption of Hoogsteen hydrogen bonds caused by the severe nonisostericity between the G\*GC and T\*AT triplets. *PLoS One* **11**, 1–29 (2016).
9. Thenmalarchelvi, R. & Yathindra, N. New insights into DNA triplexes: residual twist and radial difference as measures of base triplet non-isomorphism and their implication to sequence-dependent non-uniform DNA triplex. *Nucleic Acid Research* **33**, 43–55 (2005).
10. Bhattacharyya, D., Mirihana Arachchilage, G. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front Chem* **4**, 1–14 (2016).
11. Spiegel, J., Adhikari, S. & Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem* **2**, 123–136 (2020).
12. Doluca, O., Boutorine, A. S. & Filichev, V. V. Triplex-Forming Twisted Intercalating Nucleic Acids (TINAs): Design Rules, Stabilization of Antiparallel DNA Triplexes and Inhibition of G-Quartet-Dependent Self-Association. *ChemBioChem* **12**, 2365–2374 (2011).
13. Sanchez-Weatherby, J. *et al.* VMXi: a fully automated, fully remote, high-flux in situ macromolecular crystallography beamline. *J Synchrotron Radiat* **26**, 291–301 (2019).
14. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* **69**, 1260–1273 (2013).
15. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85–97 (2018).

16. Yao, J. X., Dodson, E. J., Wilson, K. S. & Woolfson, M. M. ACORN: A review. *Acta Crystallogr D Biol Crystallogr* **62**, 901–908 (2006).
17. Yao, J.-X. ACORN in CCP4 and its applications. *Acta Cryst* **58**, 1941–1947 (2002).
18. El Omari, K. *et al.* Experimental phasing opportunities for macromolecular crystallography at very long wavelengths. *Commun Chem* **6**, 1–11 (2023).
19. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
20. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).
21. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).
22. Lightfoot, H. L., Hagen, T., Tatum, N. J. & Hall, J. The diverse structural landscape of quadruplexes. *FEBS Lett* **593**, 2083–2102 (2019).
23. Mergny, J.-L. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res* **33**, 81–94 (2005).
24. Tang, C.-F. & Shafer, R. H. Engineering the Quadruplex Fold: Nucleoside Conformation Determines Both Folding Topology and Molecularity in Guanine Quadruplexes. *J Am Chem Soc* **128**, 5966–5973 (2006).
25. Lim, K. W. & Phan, A. T. Structural Basis of DNA Quadruplex–Duplex Junction Formation. *Angewandte Chemie* **125**, 8728–8731 (2013).
26. Harrell Jr, W. A. *Quadruplex Nucleic Acids*. (The Royal Society of Chemistry, 2006).
27. Escaja, N., Mir, B., Garavís, M. & González, C. Non-G Base Tetrads. *Molecules* **27**, 1–14 (2022).
28. Liu, H. *et al.* High-resolution DNA quadruplex structure containing all the A-, G-, C-, T-tetrads. *Nucleic Acids Res* **46**, 11627–11638 (2018).
29. Patel, P. K. & Hosur, R. V. NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. *Nucleic Acids Res* **27**, 2457–2464 (1999).
30. Caceres, C. A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl<sup>+</sup>/Na<sup>+</sup> ions. *Nucleic Acids Res* **32**, 1097–1102 (2004).
31. Merkina, E. E. & Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophys J* **89**, 365–373 (2005).
32. Qin, Y. & Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90**, 1149–1171 (2008).

33. Rachwal, P. A., Findlow, I. S., Werner, J. M., Brown, T. & Fox, K. R. Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res* **35**, 4214–4222 (2007).
34. Fujimoto, T., Nakano, S., Sugimoto, N. & Miyoshi, D. Thermodynamics-Hydration Relationships within Loops That Affect G-Quadruplexes under Molecular Crowding Conditions. *J Phys Chem B* **117**, 963–972 (2013).
35. Radhakrishnan, I. & Patel, D. J. Solution structure of a purine·purine·pyrimidine DNA triplex containing G·GC and T·AT triples. *Structure* **1**, 135–152 (1993).
36. Chung, W. J. *et al.* Structure of a left-handed DNA G-quadruplex. *Proceedings of the National Academy of Sciences* **112**, 2729–2733 (2015).
37. Truong, T. H. A., Winnerdy, F. R. & Phan, A. T. An Unprecedented Knot-like G-Quadruplex Peripheral Motif. *Angewandte Chemie International Edition* **58**, 13834–13839 (2019).
38. Winnerdy, F. R. *et al.* NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes. *Nucleic Acids Res* **47**, 8272–8281 (2019).
39. Maity, A., Winnerdy, F. R., Chang, W. D., Chen, G. & Phan, A. T. Intra-locked G-quadruplex structures formed by irregular DNA G-rich motifs. *Nucleic Acids Res* **48**, 3315–3327 (2020).
40. Das, P. *et al.* Bulges in left-handed G-quadruplexes. *Nucleic Acids Res* **49**, 1724–1736 (2021).
41. Chung, W. J. *et al.* Structure of a left-handed DNA G-quadruplex. *Proceedings of the National Academy of Sciences* **112**, 2729–2733 (2015).
42. Bakalar, B., Heddi, B., Schmitt, E., Mechulam, Y. & Phan, A. T. A Minimal Sequence for Left-Handed G-Quadruplex Formation. *Angewandte Chemie International Edition* **58**, 2331–2335 (2019).
43. Winnerdy, F. R. *et al.* NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes. *Nucleic Acids Res* **47**, 8272–8281 (2019).
44. Das, P. *et al.* Bulges in left-handed G-quadruplexes. *Nucleic Acids Res* **49**, 1724–1736 (2021).
45. Das, P., Winnerdy, F. R., Maity, A., Mechulam, Y. & Phan, A. T. A novel minimal motif for left-handed G-quadruplex formation. *Chemical Communications* **57**, 2527–2530 (2021).
46. Zhang, C. *et al.* Engineering DNA Crystals toward Studying DNA–Guest Molecule Interactions. *J Am Chem Soc* **145**, 4853–4859 (2023).
47. Kondo, J., Adachi, W., Umena, S., Sunami, T. & Takenaka, A. Crystal structures of a DNA octaplex with I-motif of G-quartets and its splitting into two quadruplexes suggest a folding mechanism of eight tandem repeats. *Nucleic Acids Res* **32**, 2541–2549 (2004).

48. Ferreira, R. *et al.* Synthesis and Structural Characterization of Stable Branched DNA G-Quadruplexes Using the Trebler Phosphoramidite. *ChemistryOpen* **1**, 106–114 (2012).
49. Harding, C. C., Watson, T., Corregan, M. & Bailey, C. Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG). *Proc. Natl. Acad. Sci. U.S.A* **31**, 833–841 (1992).
50. Gray, R. D. & Chaires, J. B. Kinetics and mechanism of K<sup>+</sup>- and Na<sup>+</sup>-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res* **36**, 4191–4203 (2008).
51. Marchand, A. & Gabelica, V. Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Res* **44**, 10999–11012 (2016).
52. Manning, G. S. Counterion Condensation on Charged Spheres, Cylinders, and Planes. *J Phys Chem B* **111**, 8554–8559 (2007).
53. Shankaraswamy, J., Tyagi, S., Singh, A., Miyoshi, D. & Saxena, S. Metal sensitive and DNA concentration dependent structural rearrangement of short oligonucleotide into large supra-structures. *J Biomol Struct Dyn* **37**, 2211–2218 (2019).
54. Kim, B. G., Shek, Y. L. & Chalikian, T. V. Polyelectrolyte effects in G-quadruplexes. *Biophys Chem* **184**, 95–100 (2013).
55. Gray, R. D., Li, J. & Chaires, J. B. Energetics and Kinetics of a Conformational Switch in G-Quadruplex DNA. *J Phys Chem B* **113**, 2676–2683 (2009).
56. Hou, X.-M. *et al.* Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res* **45**, 11401–11412 (2017).
57. Jiang, H.-X. *et al.* Divalent cations and molecular crowding buffers stabilize G-triplex at physiologically relevant temperatures. *Sci Rep* **5**, 1–11 (2015).
58. Ngoc Nguyen, T. Q., Lim, K. W. & Phan, A. T. Duplex formation in a G-quadruplex bulge. *Nucleic Acids Res* **48**, 10567–10575 (2020).
59. Bing, T. *et al.* Triplex-quadruplex structural scaffold: a new binding structure of aptamer. *Sci Rep* **7**, 1–10 (2017).
60. Svinarchuk, F., Paoletti, J. & Malvy, C. An Unusually Stable Purine(Purine-Pyrimidine) Short Triplex. *Journal of Biological Chemistry* **270**, 14068–14071 (1995).
61. Teng, Y., Tateishi-Karimata, H. & Sugimoto, N. C-Rich Sequence in a Non-Template DNA Strand Regulates Structure Change of G-Quadruplex in a Template Strand during Transcription. *Bull Chem Soc Jpn* **92**, 572–577 (2019).

## CHAPTER 7 - Summary and Future Work

### 7.1 Summary

The structural characterization of DNA triplexes and metal complex interaction is crucial for the study of gene-targeting molecules. The work in this thesis provides a deeper understanding of the interaction between DNA triplexes and ruthenium complexes, crucial for harnessing their stability, photoluminescence, and cleavage properties.<sup>1</sup> The study also provides the first insight into how polypyridyl ruthenium complexes bind triple-helical structures.

The first aim of this work was to lay the ground for crystallisation screening by generating a library of triplexes based on sequences available in the literature. Given the inherent challenges in controlling the crystallisation process,<sup>2</sup> factors influencing the solution stability of DNA triplexes were explored. These findings provided guidance for further optimisation of the crystallisation process. Intramolecular triplexes exhibited overall higher stability, with UV melting curves displaying characteristic triplex DNA features. In some cases, the triplex transition coincided with the duplex transition, indicating the presence of both conformations at similar temperatures. In contrast, triplex formed by three strands except for the 13-mer TFO triplex, showed less stable melting profiles, particularly for short sequences. Whilst not displaying clear melting transitions in solution, short sequences were included in the crystallisation screening, as they previously resulted in diffracting crystals.<sup>3,4</sup> Despite the limited depth in the UV melting analysis, it serves as an initial step towards developing a nucleic acid crystallisation screen, tailored for DNA triplexes. Future work should involve a more thorough UV melting analysis, further optimising conditions by varying DNA concentration or cell path, to mimic the highly concentrated crystallisation sample; increasing the number of triplexes, and giving careful consideration to cations in obtained crystals.

The solution stability was conducted also for  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  with DNA triplexes. Here the aim was to examine the binding site adjacent to the TFO-binding region of metal complexes. By harnessing the photoluminescence effect typical of this ruthenium compound, coupled with CD spectroscopy and UV-vis melting experiments, potential binding sites were identified. The increased luminescence was observed in the triplex compared to the duplex. The  $\Delta$ -enantiomer was protected from the aqueous solution emitting luminescence regardless of the presence of the TFO and duplex extension. Of particular interest was the extension with C:G base pair where

it is suggested to favour the intercalation of the ruthenium complex, resulting in increased melting. Interestingly, the racemic form in the solution showed increased luminescence, indicating a configuration involving multiple enantiomers binding. As will be shown with crystallography, extensions with TA base pair are expected to accommodate ruthenium complexes, therefore it is expected to obtain a similar outcome when the extension includes TA. This was shown with the CD spectra giving significant changes in the polarised light. While this study provides additional CD spectra profiles to recognise the triplex signature with intercalating polypyridyl complexes, further studies are essential to address the in-solution binding of triplex considering other metal complexes and sequence alteration. Testing different polypyridyl complexes is important to determine the intercalation of metal complexes to DNA triplexes. Despite the interpretation challenges posed by the strong absorption of ruthenium complexes being overcome through comparisons, certain details might still be unclear. Consequently, there is a possibility that valuable insights, which could contribute additional information, have not been thoroughly explored.

The crystallisation of the system AA-t4, initially shown stable at low temperatures in solution compared to other intramolecular triplexes, yielded a stable triplex structure with  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  during the crystallisation process. Despite being designed as an intramolecular structure, the crystallised outcome shows an intermolecular structure, with the TFO originating from symmetry-related molecules. The intercalation of  $[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$  shows a novel intercalation mode within T-A:T triplets. The complexes were intercalated in neighbouring base steps separated by a TA base pair, thus violating the neighbour exclusion limit. Isothermal titration calorimetry could be employed to determine the binding affinity of the  $\Delta$  enantiomers and evaluate the DNA pocket's capacity to accommodate more than one complex. The outcomes of such investigations, coupled with the crystal structure result, will provide a deeper understanding of the molecular interactions beyond the crystalline state. This could offer a promising approach for more precise and targeted applications in the manipulation of genetic sequences.<sup>5</sup>

Such applications were previously explored as a method for recognising and binding specific genes and demonstrated promising results.<sup>6</sup> Since the TFO binds in the major groove of a duplex, understanding how the ruthenium complex can bind from the major groove is crucial for supporting the design of novel TFO-ligand tools. Major groove binding of metal complexes has been a subject of prior investigation.<sup>7</sup> To date, the literature lacks substantial information about

the metal complex binding in the duplex major groove, while increasing results report intercalation from the minor groove. The second X-ray structure obtained from the screening shows  $[\text{Ru}(\text{TAP})_2(\text{CN-dppz})]^{2+}$  intercalated in the duplex DNA, resulting in the first dppz X-ray structure intercalated in the major groove. Considering the role of nitrogen phenazine in influencing the light switch effect of the complexes, major groove intercalation holds the potential for shielding specific ligand moieties. Future research should focus on biophysical studies, such as fluorescent titrations and Job plots experiments with polypyridyl complexes to determine the potential impact of major groove intercalation on the light switch effect. The role of CN substitution in the major groove intercalation was suggested,<sup>8</sup> however additional spectroscopic data are necessary for confirmation. Previous spectroscopic studies have shown that extended dppz ligands intercalating from the minor groove in the T-A:T triplex interfere sterically with the TFO.<sup>9</sup> This was confirmed in the triplex structure presented in this thesis. Therefore, it can be hypothesised that the CN substitution prevents the TFO from binding in the major groove while allowing the complex to intercalate. However, to support this hypothesis and determine the stability of the triplex with and without  $[\text{Ru}(\text{TAP})_2(\text{CN-dppz})]^{2+}$  complexes, further UV melting studies and solution evidence are required.

The structural analysis indicates a clear binding affinity to TA base pairs, similar to previous crystal structures.<sup>10</sup> However, it is crucial to recognise that while the structure reveals the crystallisation state it cannot establish the binding preferences of TA-rich sequences over CG-rich regions.<sup>11</sup> The crystal packaging requirements may influence the observed binding preference, especially if multiple binding modes are possible, with the crystal structure favouring the one that optimises the crystal packaging. It is also important to acknowledge that other factors, such as cations and DNA sequence, contribute to the overall stability of these interactions. Further studies of solution behaviour and dynamic properties would allow a more comprehensive, physiological understanding beyond the crystalline state.

The crystallisation of the intermolecular triplex AA-t3 revealed interesting outcomes where the G-rich TFO self-assembled into a quadruplex. The quadruplex composed of G-tetrads and T-tetrads, features a T-G octaplet that connects two quadruplex dimers. The exact role of T-G octaplet conformation remains debatable; however, insights from other reported structures suggest a role of a connector with more secondary structures.<sup>12,13</sup> This unexpected finding underlies the need for careful TFO design and highlights the flexibility of DNA in forming supramolecular

structures. Gel electrophoresis techniques with crowding agents could be used to study the behaviour of supramolecular DNA assemblies in environments that resemble physiological conditions. Furthermore, this crystallisation result explains anomalous melting data observed during the initial screening.

Overall, the crystal structures obtained – one showing TFO self-assembly of a G-quadruplex rather than a triplex, and another exhibiting simultaneous minor and major groove metal intercalation in duplex without the TFO - confirm the intrinsic instability of the triplexes under the crystallisation conditions. The one sequence that successfully formed a crystal was a longer, intramolecular triplex motif. This suggests that long sequences may favour the crystallisation of triplexes, though their larger size renders them more difficult to crystallise. Additionally, the role of metal cations should be taken into consideration.<sup>14</sup> While, divalent cations are needed to compensate for the negatively charged phosphate backbone, here, magnesium ions were observed to coordinate within the major groove upon complex intercalation and support the T-G octaplet formation. However, cation varieties in crystallisation solutions failed to stabilise triplexes, instead promoting secondary structures like G-quadruplexes. Therefore, it is suggested to carefully optimise the cations in the crystallisation screen, along with considering the concentration of the strands. Additionally, a deliberate approach could involve increasing the concentration of the TFO, potentially enhancing its binding in the major groove.

In conclusion, the findings reveal sequence and geometry preferences of intercalated ruthenium complexes supporting intercalation in triple-helical structures. This body of work contributes to the understanding of the structure and stability of DNA triplex interactions with metal complexes. These insights offer guidance for the rational design of triplex-forming oligonucleotides for gene editing technologies.

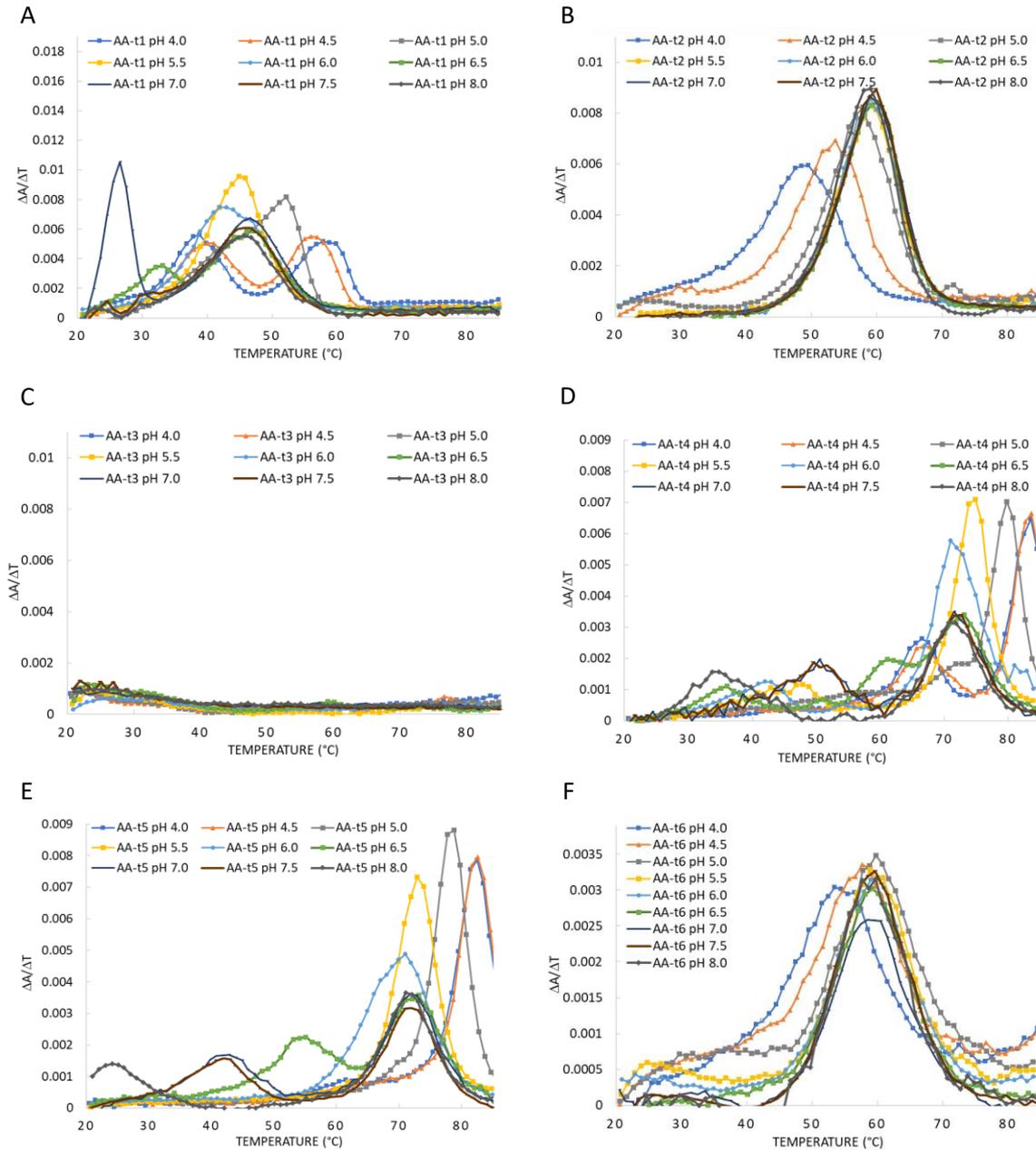
## 7.1 References

1. Łęczkowska, A. & Vilar, R. Interaction of metal complexes with nucleic acids. *Annual Reports Section 'A' (Inorganic Chemistry)* **109**, 299 (2013).
2. McPherson, A. & Cudney, B. Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr F Struct Biol Commun* **70**, 1445–1467 (2014).
3. Hall, J. P. *et al.* X-ray crystal structure of rac-[Ru(phen)2dppz]2+ with d(ATGCAT)2 shows enantiomer orientations and water ordering. *J Am Chem Soc* **135**, 12652–12659 (2013).
4. Hall, J. P. *et al.* Guanine Can Direct Binding Specificity of Ru–dipyridophenazine (dppz) Complexes to DNA through Steric Effects. *Chemistry - A European Journal* **23**, 4981–4985 (2017).
5. Vasquez, K. Chromosomal mutations induced by triplex-forming oligonucleotides in mammalian cells. *Nucleic Acids Res* **27**, 1176–1181 (1999).
6. Zuin Fantoni, N. *et al.* Development of Gene-Targeted Polypyridyl Triplex-Forming Oligonucleotide Hybrids. *ChemBioChem* **21**, 3563–3574 (2020).
7. Molphy, Z. *et al.* A phosphate-targeted dinuclear Cu(II) complex combining major groove binding and oxidative DNA cleavage. *Nucleic Acids Res* **46**, 9918–9931 (2018).
8. McQuaid, K., Hall, J. P., Brazier, J. A., Cardin, D. J. & Cardin, C. J. X-ray Crystal Structures Show DNA Stacking Advantage of Terminal Nitrile Substitution in Ru-dppz Complexes. *Chemistry – A European Journal* **24**, 15859–15867 (2018).
9. Choi, S. D. *et al.* Binding mode of [ruthenium(II) (1,10-phenanthroline)2L]2+ with poly(dT\*dA-dT) triplex. Ligand size effect on third-strand stabilization. *Biochemistry* **36**, 214–223 (1997).
10. Cardin, C. J., Kelly, J. M. & Quinn, S. J. Photochemically active DNA-intercalating ruthenium and related complexes-insights by combining crystallography and transient spectroscopy. *Chem Sci* **8**, 4705–4723 (2017).
11. Ortmans, I., Elias, B., Kelly, J. M., Moucheron, C. & Kirsch-DeMesmaeker, A. [Ru(TAP)2(dppz)]2: a DNA intercalating complex, which luminesces strongly in water and undergoes photo-induced proton-coupled electron transfer with guanosine-5'-monophosphate. *Dalton Trans.* **1**, 668–676 (2004).
12. Sato, Y., Mitomi, K., Sunami, T., Kondo, J. & Takénaka, A. DNA Octaplex Formation with an I-Motif of Water-Mediated A-Quartets: Reinterpretation of the Crystal Structure of d(GCGAAAGC). *The Journal of Biochemistry* **140**, 759–762 (2006).
13. Kondo, J., Adachi, W., Umena, S., Sunami, T. & Takenaka, A. Crystal structures of a DNA octaplex with I-motif of G-quartets and its splitting into two quadruplexes suggest a folding mechanism of eight tandem repeats. *Nucleic Acids Res* **32**, 2541–2549 (2004).
14. Cardin, C. J. Interactions Between Metal Ions and DNA. in *Structure and Bonding* vol. 182 203–237 (2019).

# Appendix

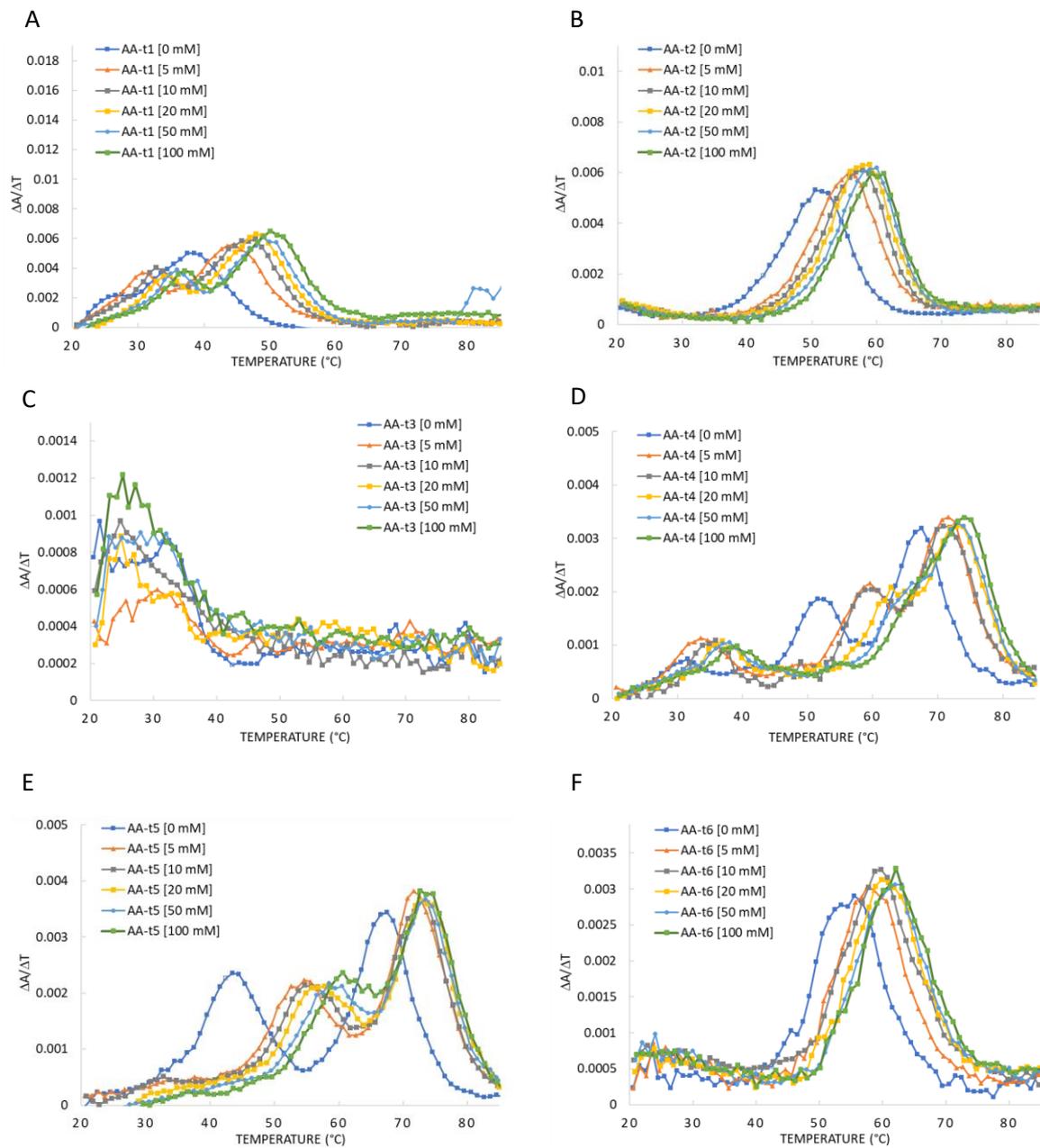
## Chapter 2

Figure A2.1



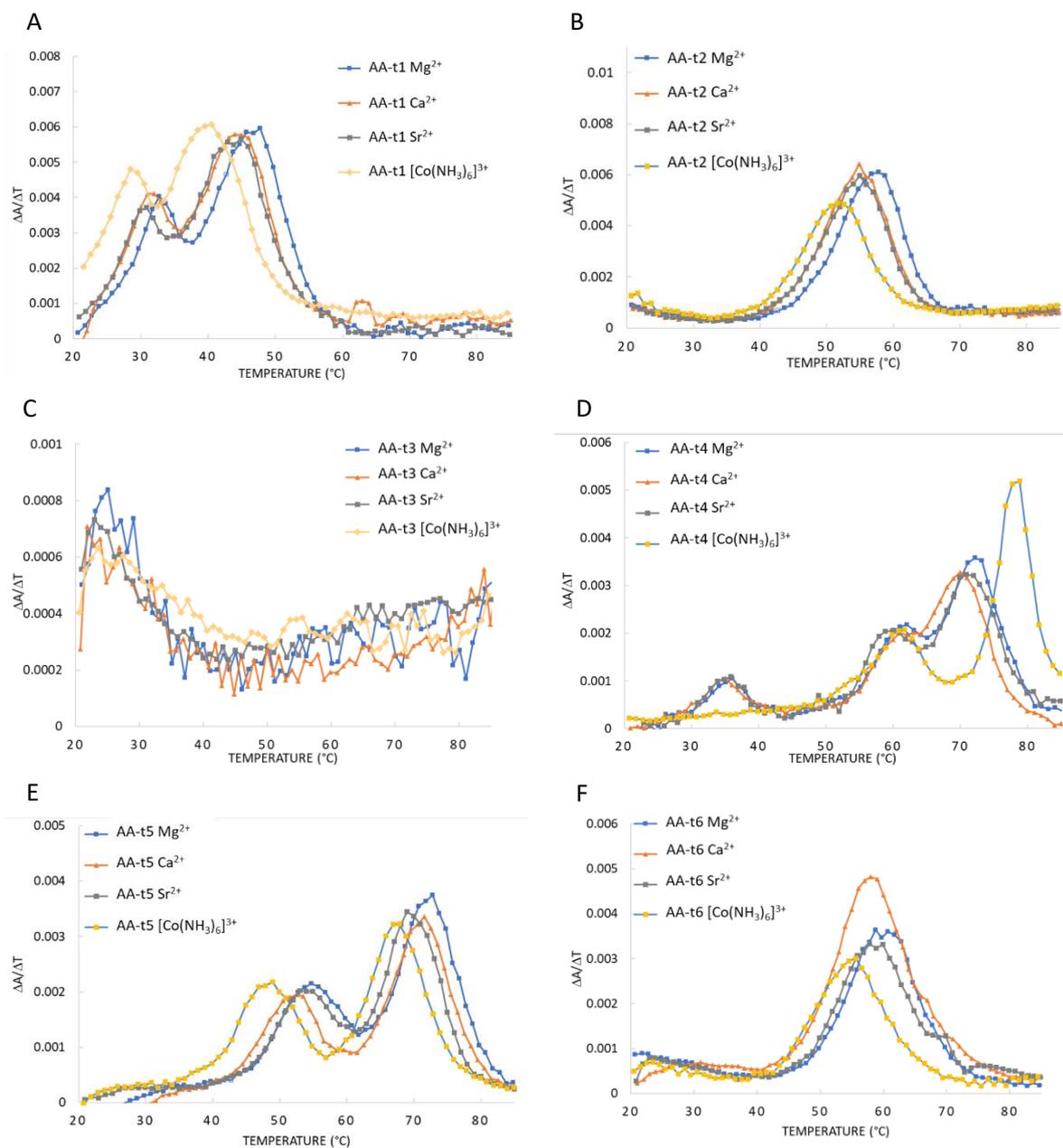
**Figure A2.1** First derivative of AA-t1 (A), AA-t2 (B), AA-t3, (C) AA-t4 (D), AA-t5 (E), AA-t6 (F). Samples were prepared with 20 mM sodium cacodylate buffer at ranges of pH from 4.0 to 8.0 with a 0.5 pH unit increment, 100 mM NaCl, and 10 mM  $\text{MgCl}_2$ .

Figure A2.2



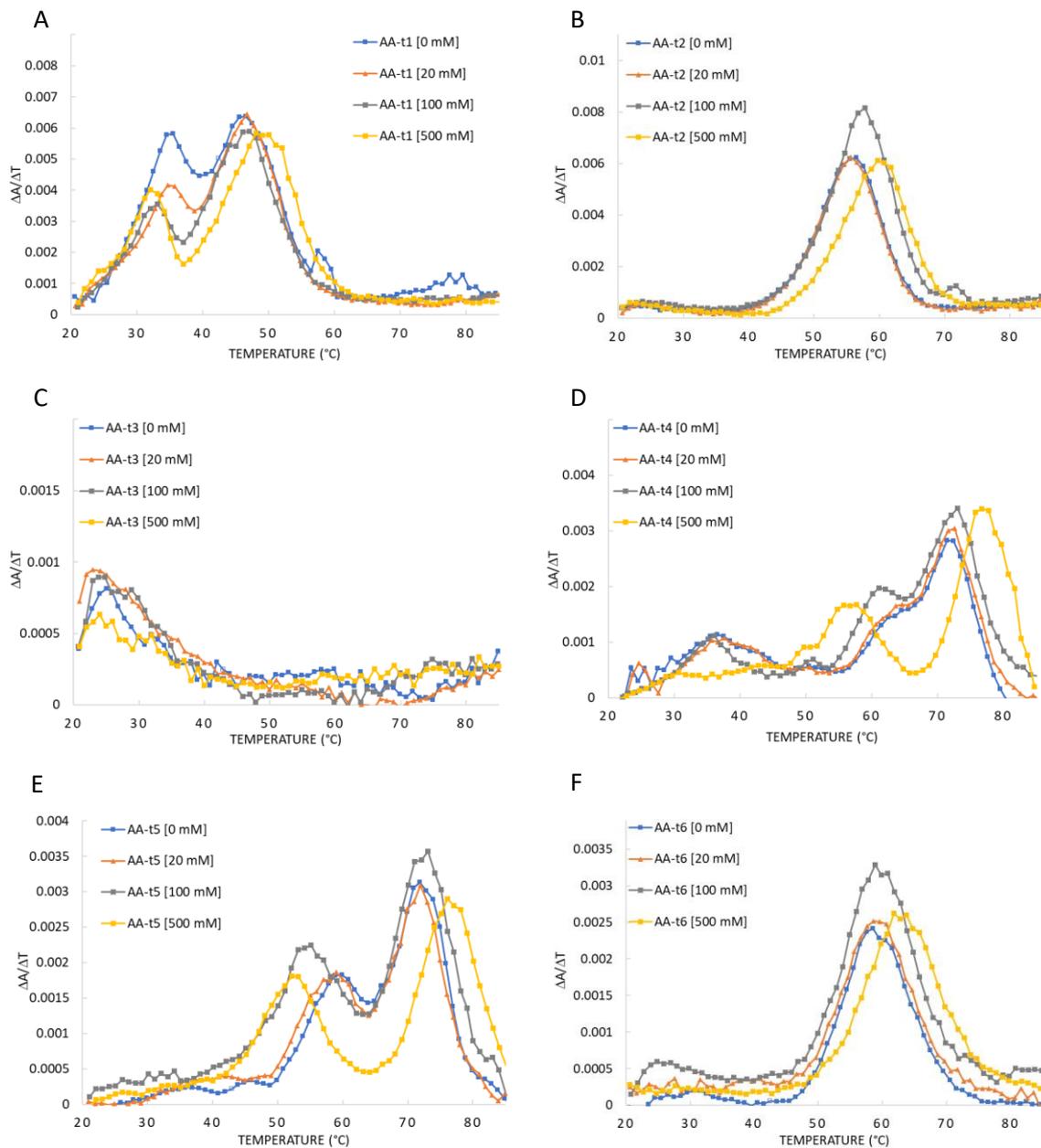
**Figure A2.2** First derivative of AA-t1 (A), AA-t2 (B), AA-t3, (C) AA-t4 (D), AA-t5 (E), AA-t6 (F). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t1, AA-t4), pH 5 (AA-t2), pH 5.5 (AA-t3, AA-t5, AA-t6); 100 mM NaCl and MgCl<sub>2</sub> at a concentration of 0 mM, 5mM, 10mM, 20 mM, 50 mM and 100 mM.

Figure A2.3



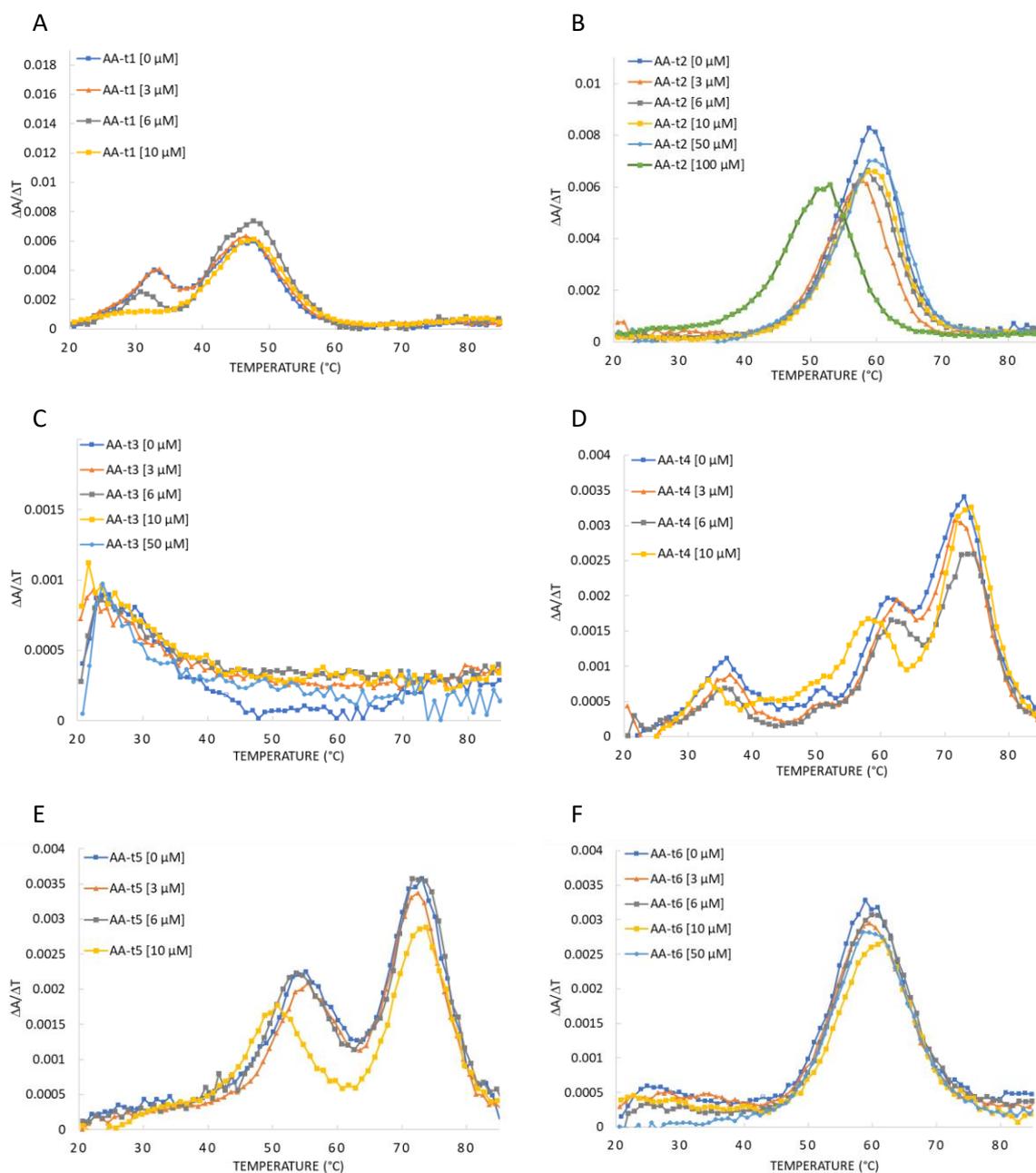
**Figure A2.3** First derivative of AA-t1 (A), AA-t2 (B), AA-t3, (C) AA-t4 (D), AA-t5 (E), AA-t6 (F). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t1, AA-t4), pH 5 (AA-t2), pH 5.5 (AA-t3, AA-t5, AA-t6); 100 mM NaCl and 10 mM  $\text{MgCl}_2$ , 10 mM  $\text{CaCl}_2$ , 10mM  $\text{SrCl}_2$  or 0.1 mM of  $[\text{Co}(\text{NH}_3)_6]\text{Cl}_3$ .

Figure A2.4



**Figure A2.** First derivative of AA-t1 (A), AA-t2 (B), AA-t3 (C) AA-t4 (D), AA-t5 (E), AA-t6 (F). Samples were prepared with 20 mM sodium cacodylate buffer at pH 6.5 (AA-t1, AA-t4), pH 5 (AA-t2), pH 5.5 (AA-t3, AA-t5, AA-t6); 100 mM NaCl and a concentration of 0 mM, 5 mM, 10 mM, 20 mM, 50 mM and 100 mM MgCl<sub>2</sub>.

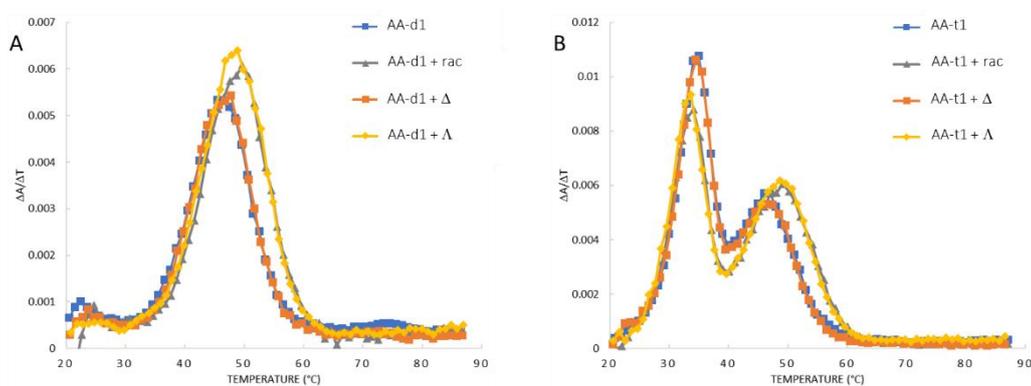
Figure A2.5



**Figure A2.5** Normalised melting profile of AA-t1 (A), AA-t2 (B) and AA-t3 (C). Samples were prepared with sodium cacodylate buffer at pH 6.5 (AA-t1), pH 5.0 (AA-t2) and pH 5.5 (AA-t3); 100 mM NaCl, 10 mM MgCl<sub>2</sub> and different concentration of spermine (3 μM, 6 μM, 10 μM, 50 μM, 100 μM and 500 μM). Precipitation was observed at 100 μM in AA-t1 and AA-t3, and at 500 μM in all systems failed to give melting transition and have been not reported.

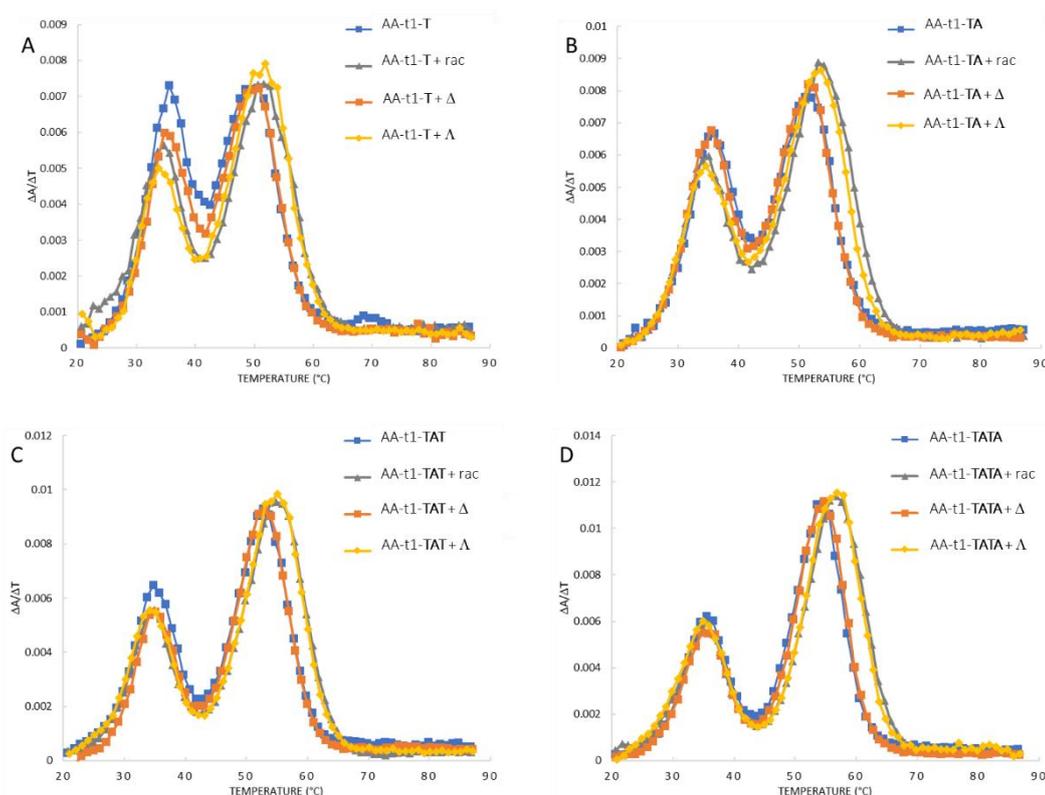
## Chapter 3

Figure A3.1



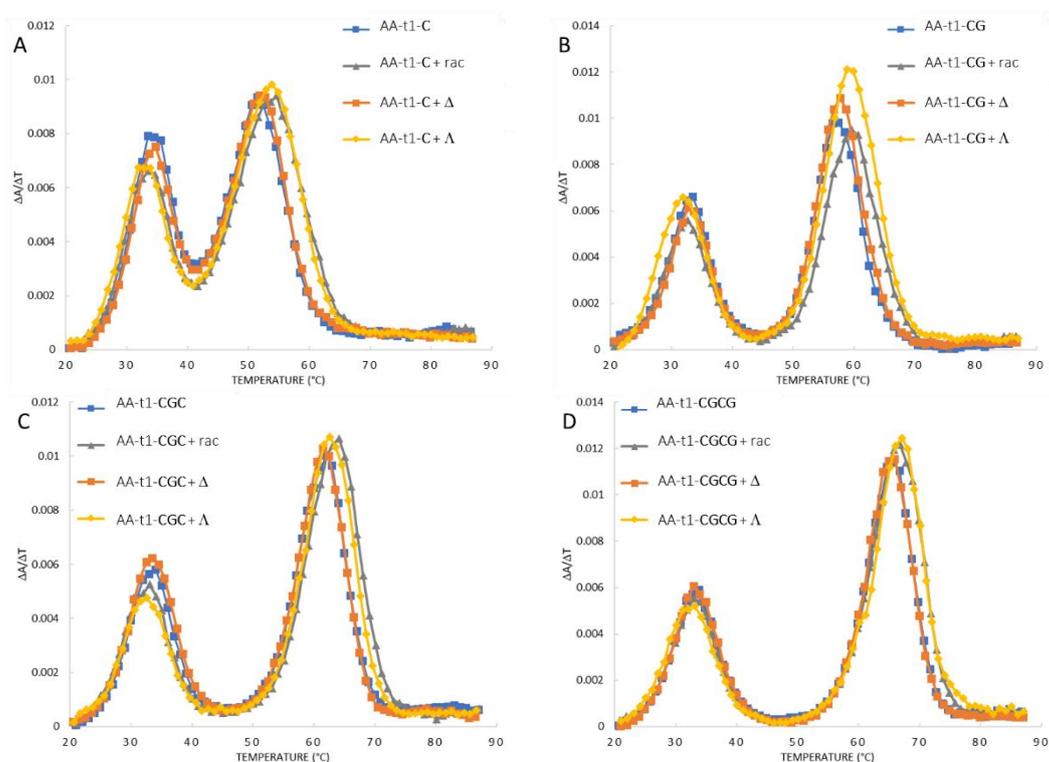
**Figure A3.1** First derivative of normalised UV melting profiles of AA-d1 (A) AA-t1 (B). Samples were prepared without Ru (II) complex (blue), and with rac-Ru (grey),  $\Delta$ -Ru (orange), and  $\Lambda$ -Ru (yellow). [DNA duplex] or [DNA triplex] ratio with [Ru] of 1:2. Absorbance recorded at  $\lambda = 260$  nm.

Figure A3.2



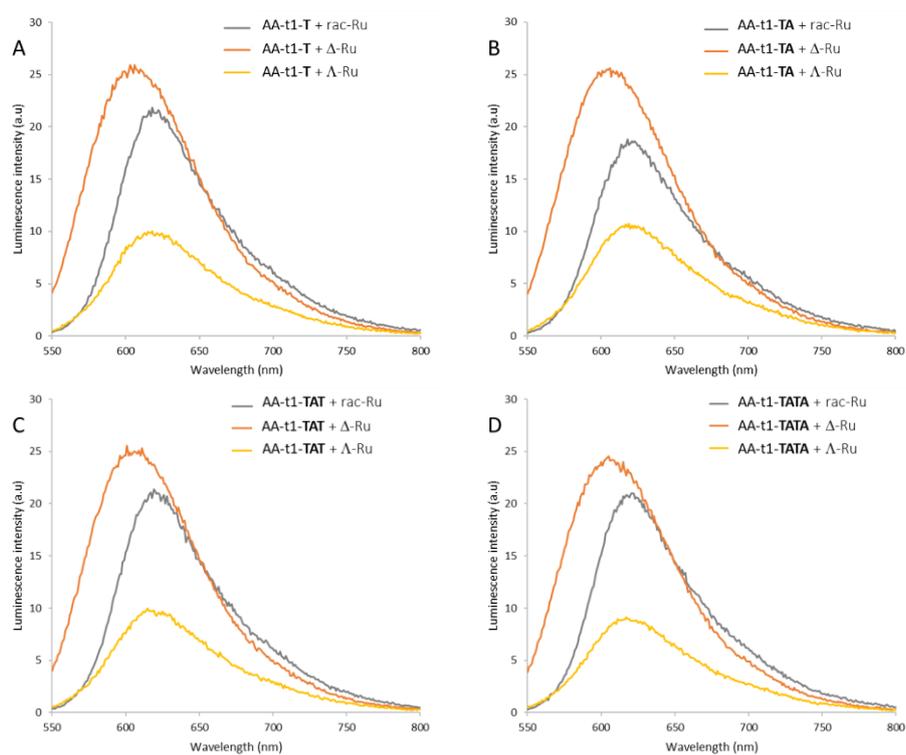
**Figure A3.2** First derivative of normalised melting profile of the triplex systems AA-t1-T (A), AA-t1-TA (B), AA-t1-TAT (C), and AA-t1-TATA (D). Samples were prepared without Ru (II) complex (blue), and with rac-Ru (grey),  $\Delta$ -Ru (orange), and  $\Lambda$ -Ru (yellow). [DNA triplex]:[Ru] ratio of 1:2. Absorbance recorded at  $\lambda = 260$  nm.

Figure A3.3



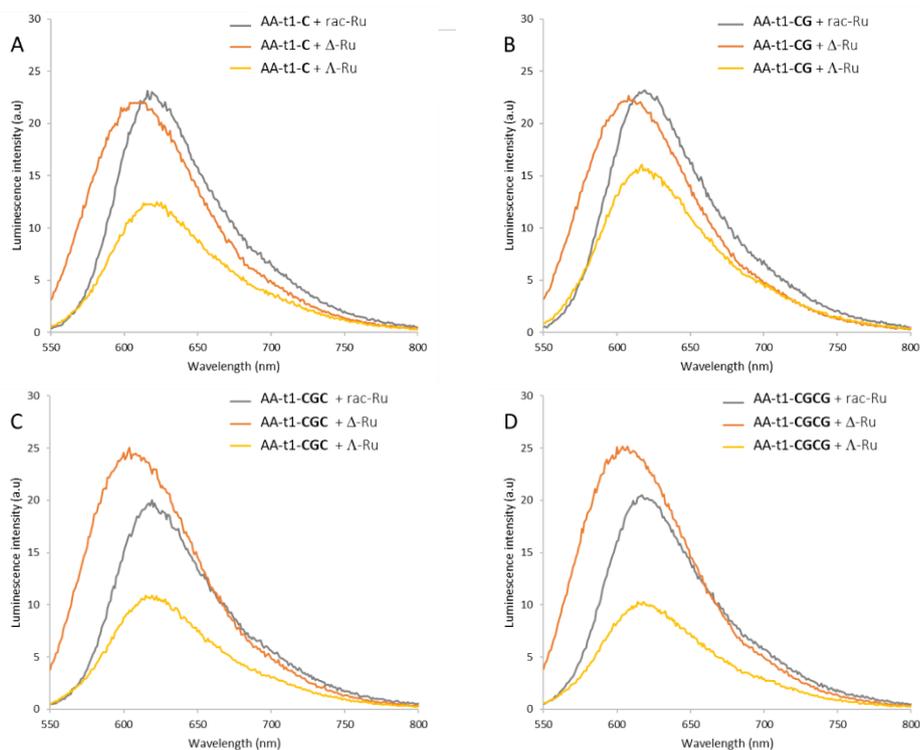
**Figure A3.3** First derivative of normalised melting profile of the triplex systems AA-t1-C (A), AA-t1-CG (B), AA-t1-CGC (C), and AA-t1-CGCG (D). Samples were prepared without Ru (II) complex (blue), and with *rac*-Ru (grey),  $\Delta$ -Ru (orange), and  $\Lambda$ -Ru (yellow). [DNA triplex]:[Ru] ratio of 1:2. Absorbance recorded at  $\lambda = 260$  nm.

Figure A3.4



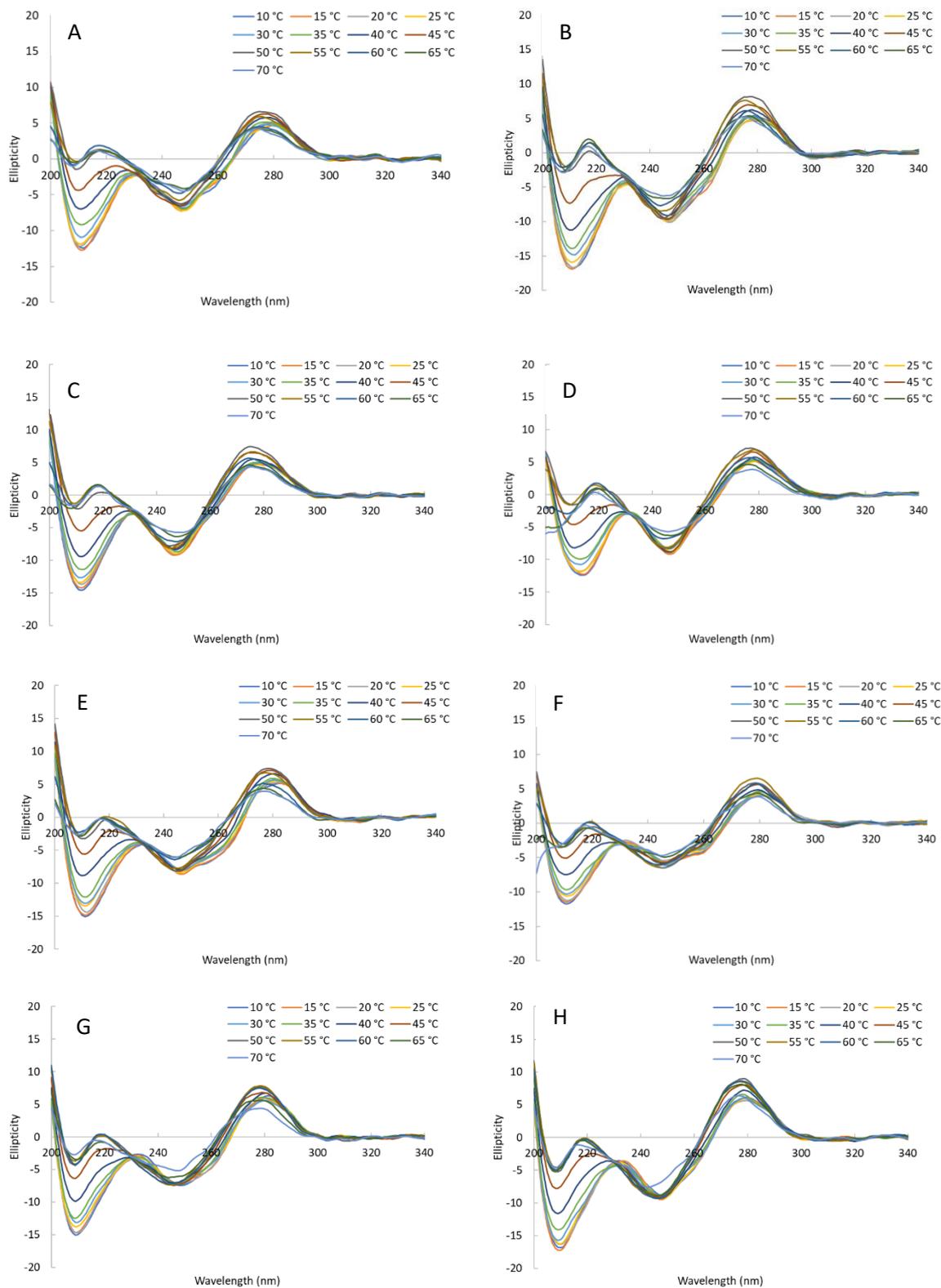
**Figure A3.4** Fluorescence emission spectra of AA-t1-T (A), AA-t1-TA (B), AA-t1-TAT (C), and AA-t1-TATA (D) with a ratio 2:1 with rac-Ru (grey), Δ-Ru (orange), and Λ-Ru (yellow).  $\lambda_{excit} = 475 \text{ nm}$

Figure A3.5



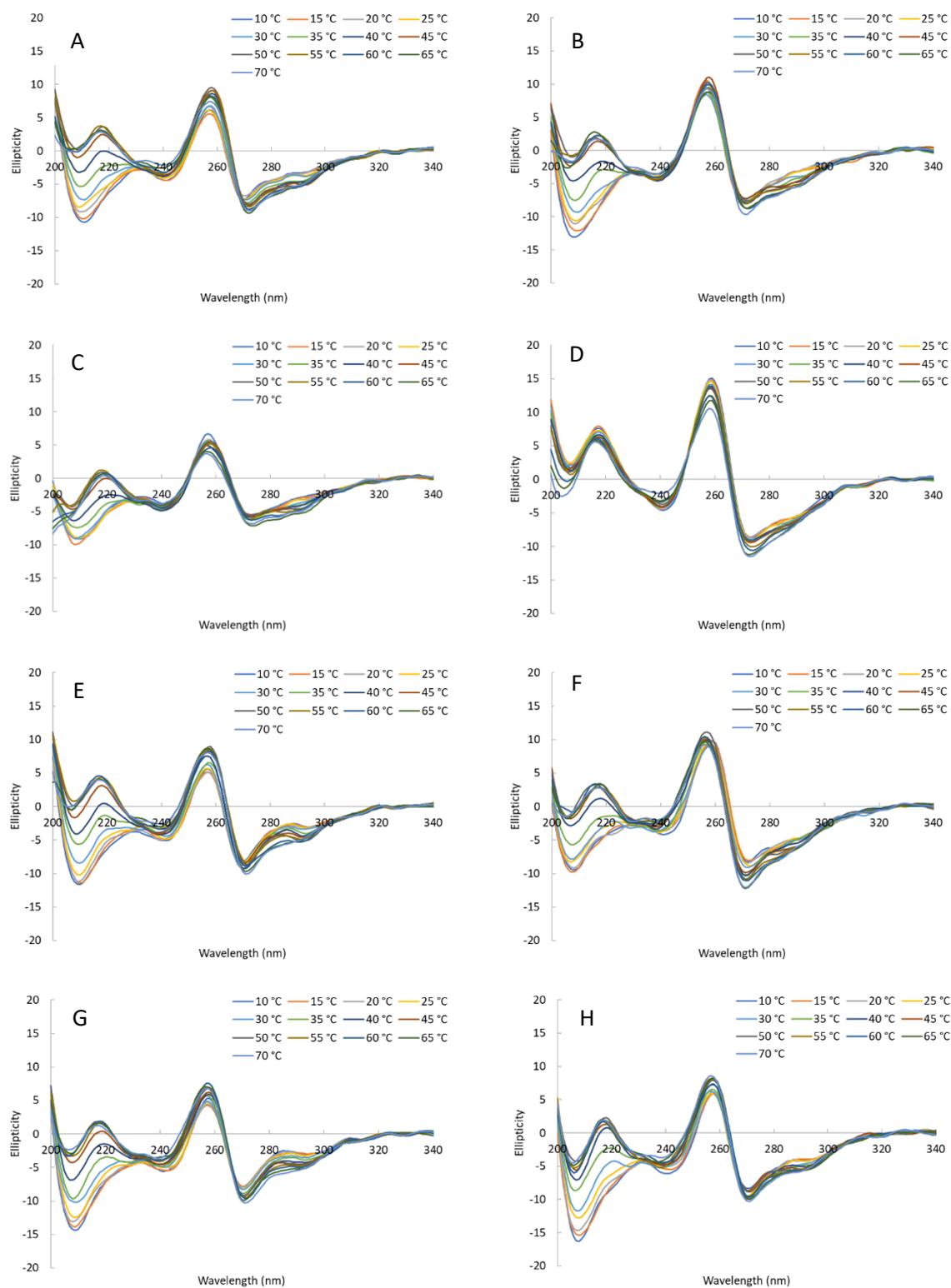
**Figure A3.5** Fluorescence emission spectra of AA-t1-C (A), AA-t1-CG (B), AA-t1-CGC (C) and AA-t1-CGCG (D) with a ratio 2:1 with rac-Ru (grey), Δ-Ru (orange), and Λ-Ru (yellow).  $\lambda_{excit} = 475 \text{ nm}$

Figure A3.6



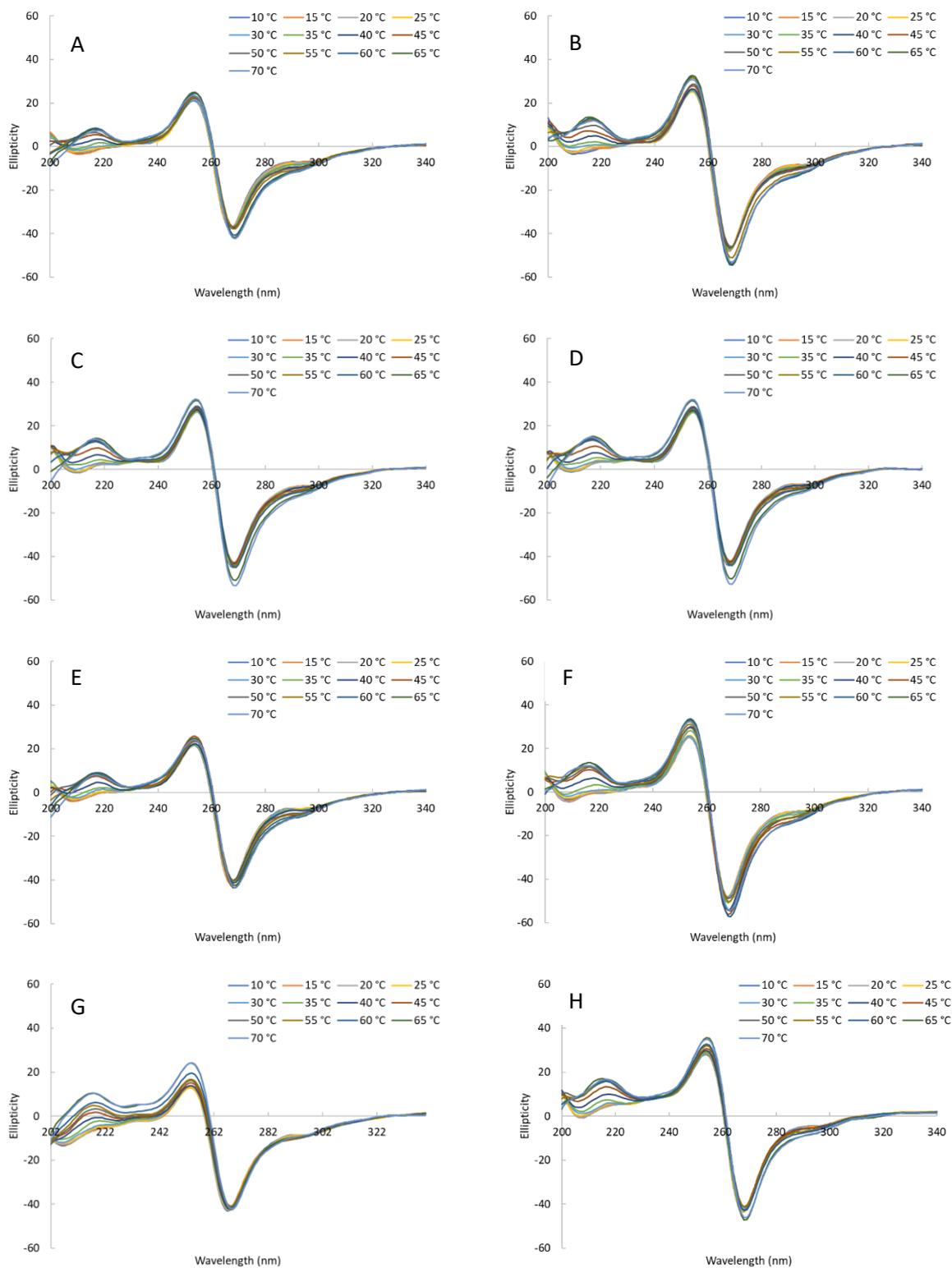
**Figure A3.6** AA-t1-T (A), AA-t1-TA (B), AA-t1-TAT (C), AA-t1-TATA (D), AA-t1-C (E), AA-t1-CG (F), AA-t1-CGC (G) and AA-t1-CGCG (H). Spectra collected at temperatures 10 °C to 70 °C with a 5-degree increment.

Figure A3.7



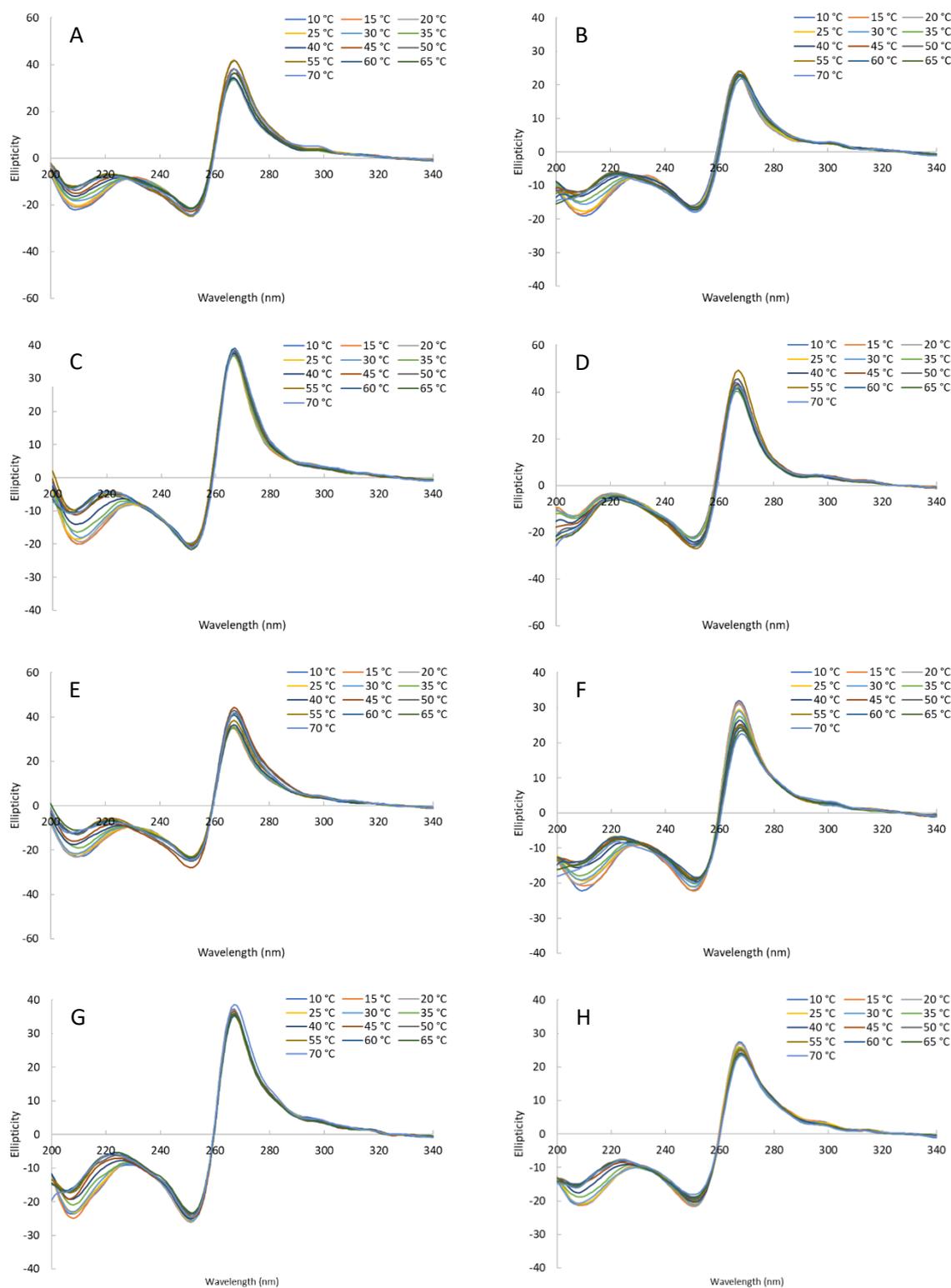
**Figure A3.7** AA-t1-T (A), AA-t1-TA (B), AA-t1-TAT (C), AA-t1-TATA (D), AA-t1-C (E), AA-t1-CG (F), AA-t1-CGC (G) and AA-t1-CGCG (H) with  $\text{rac-}[\text{Ru}(\text{phen})_2(\text{dppz})]^{2+}$ . Spectra collected at temperatures 10 °C to 70 °C with a 5-degree increment.

Figure A3.8



**Figure A3.8** AA-t1-T (A), AA-t1-TA (B), AA-t1-TAT (C), AA-t1-TATA (D), AA-t1-C (E), AA-t1-CG (F), AA-t1-CGC (G) and AA-t1-CGCG (H) with  $\Delta$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup>. Spectra collected at temperatures 10 °C to 70 °C with a 5-degree increment.

Figure A3.9



**Figure 3.9** AA-t1-T (A), AA-t1-TA (B), AA-t1-TAT (C), AA-t1-TATA (D), AA-t1-C (E), AA-t1-CG (F), AA-t1-CGC (G) and AA-t1-CGCG (H) with  $\Lambda$ -[Ru(phen)<sub>2</sub>(dppz)]<sup>2+</sup>. Spectra collected at temperatures 10 °C to 70 °C with a 5-degree increment.

## Chapter 4

Table A4.1

**Local base-pair parameters for structure 8PIP**

<b>Base Pair</b>	<b>Shear (Å)</b>	<b>Stretch (Å)</b>	<b>Stagger (Å)</b>	<b>Buckle (°)</b>	<b>Propeller (°)</b>	<b>Opening (°)</b>
<b>Pu-Py</b>						
<b>A1:T20</b>	-0.11	-0.05	-0.21	-14.18	-9.34	0.16
<b>G2:C19</b>	-0.12	-0.13	0.43	9.02	-2.88	0.47
<b>A3:T18</b>	0.21	-0.01	0.13	12.8	-13.21	6.4
<b>T4:A17</b>	0.11	-0.02	-0.06	-8.42	24.29	19.74
<b>A5:T16</b>	0.17	-0.04	-0.41	-19.35	-14.01	3.04
<b>G6:C15</b>	-0.09	0.1	-0.35	-3.44	-11.13	7.23
<b>A7:T14</b>	0.55	-0.11	0.11	3.95	-11.85	6.95
<b>A8:T15</b>	0.17	-0.04	0.45	16.65	-14.23	12.68
<b>T21:A24</b>	-0.01	-0.24	0.04	20.35	-1.16	4.23
<b>A22:T23</b>	0.24	-0.13	-0.23	-16.19	9.69	2.01
<b>T22:A24</b>	-0.24	-0.13	-0.23	16.19	9.69	2.01
<b>A24:T21</b>	0.01	-0.24	0.04	-20.35	-1.16	4.23
<b>Pu-TFO</b>						
<b>A1-T25</b>	0.41	-3.58	-0.63	10.7	-5.3	64.54
<b>G2-C26</b>	-0.06	-3.32	-0.09	8.71	18.39	65.45
<b>A3-T27</b>	0.39	-3.23	-0.18	6.26	16.18	66.24
<b>A5-T28</b>	0.63	-3.54	-0.65	12.83	-6.89	68.71
<b>G6-C29</b>	0.21	-3.23	-0.57	10.91	5.87	69.7
<b>A7-T30</b>	-0.03	-2.87	-0.61	9.96	8.06	69.06
<b>A8-T31</b>	-0.01	-2.85	-0.56	16.1	13.06	67.96

Table A4.2

**Local base-pair step parameters for structure 8PIP**

<b>Step</b>	<b>Shift (Å)</b>	<b>Slide (Å)</b>	<b>Rise (Å)</b>	<b>Tilt (°)</b>	<b>Roll (°)</b>	<b>Twist (°)</b>
<b>Pu-Py</b>						
AG/CT	0.44	0.6	2.83	-6.43	11.05	27
GA/TC	0.59	-0.25	3.24	4.11	10.78	29.13
AT/AT	0.27	-0.35	6.34	-13.45	17.28	11.68
TA/TA	-1.1	0.33	6.35	16.22	15.16	8.47
AG/CT	0.28	0.38	3	-3.99	12.19	25.53
GA/TC	0.64	-0.37	3	-2.45	6.28	34.22
AA/TT	0.53	-0.68	2.92	-1.57	3.46	29.3
AT/AT	0	0.2	2.77	0	6.78	14.97
TA/TA	----	----	----	----	----	----
<b>Pu-TFO</b>						
AG/CT	0.37	-0.11	2.93	0.29	-1.09	23.86
GA/TC	0.12	-0.34	3.2	1.26	-13.14	32.2
AG/CT	0.15	0.41	3.05	2.12	-3.08	24.91
GA/TC	-0.58	-0.34	3.16	0.12	-4.61	32.41
AA/TT	-0.33	0.01	3.05	-1.5	-4.65	32.47

Table A4.3

**Local base step parameters for structure 8PIP**

<b>Step</b>	<b>Shift (Å)</b>	<b>Slide (Å)</b>	<b>Rise (Å)</b>	<b>Tilt (°)</b>	<b>Roll (°)</b>	<b>Twist (°)</b>
A/G	0.55	0.45	3.17	6.27	13.56	27.52
G/A	1	0.1	3.03	8.16	8.22	30.7
A/T	-0.05	-0.31	6.23	-20.33	38.68	17.5
T/A	-1.38	-0.18	6.11	9.78	-5.96	1.22
A/G	0.39	0.13	3.04	7.05	10.71	28.67
G/A	1.19	-0.47	3.12	4.82	5.73	33.98
A/A	0.64	-0.33	3.07	8.23	4.2	31.48
A/C	-9.84	-5.55	-7.29	41.42	104.95	-49.57
C/C	0.86	-0.02	3.83	4.99	7.95	32.62
C/C	13.4	-3.5	6.39	-114.45	43.33	159.94
C/C	-16.58	-3.38	-0.3	-75.81	55.49	-131.41
C/T	10.16	1.26	12.24	-86.54	2.25	100.65
T/T	-0.51	-1.01	2.72	11.14	1.07	26.07
T/C	-0.12	-0.34	2.8	9.73	5.97	33.87
C/T	-0.15	0.63	2.96	14.97	12.58	21.3
T/A	0.9	0.97	6.51	-18.55	37.01	13.11
A/T	-0.6	-0.32	6.44	2.11	-1.84	5.33
T/C	-0.2	-0.66	3.36	0.16	12.96	27.54
C/T	-0.34	0.74	2.49	19.11	8.5	26.2
T/T	-15.25	-0.91	2.97	-11.56	-15.73	-92.77
T/A	-2.82	1.3	7.13	-26.01	-0.86	26.24
A/T	-0.51	0.21	2.74	14.86	6.54	15.73
T/A	3.22	0.94	7.37	-11.83	-12.7	28.35
A/T	-1.9	-1	3.24	5.01	-2.97	16.34
T/C	-0.54	-0.35	2.72	9.44	-9.17	26.67
C/T	-0.34	-0.22	3.58	6.27	15.51	28.98
T/T	0.7	-0.42	11.48	-50.45	17.24	23.06
T/C	-0.44	-1.17	2.98	4.18	-3.72	28.33
C/T	0.43	-0.52	3.22	5.2	2.58	34.81
T/T	0.11	-0.65	3.13	5.79	-0.76	32.18
T/T	28.89	-4.66	-0.81	18.14	124.28	-109.48
T/A	-2.82	1.3	7.13	-26.01	-0.86	26.24
A/T	-0.51	0.21	2.74	14.86	6.54	15.73
T/A	3.22	0.94	7.37	-11.83	-12.7	28.35
A/T	16.1	9.04	2.92	1.29	19.16	-163.22
T/C	-0.54	-0.35	2.72	9.44	-9.17	26.67
C/T	-0.34	-0.22	3.58	6.27	15.51	28.98
T/T	0.7	-0.42	11.48	-50.45	17.24	23.06
T/C	-0.44	-1.17	2.98	4.18	-3.72	28.33
C/T	0.43	-0.52	3.22	5.2	2.58	34.81
T/T	0.11	-0.65	3.13	5.79	-0.76	32.18

Table A4.4

**Local base-pair helical parameters for structure 8PIP**

step	X-disp (Å)	Y-disp (Å)	helical rise (Å)	Inclination (°)	Tip (°)	helical Twist (°)
<b>Pu-Py</b>						
AG/CT	-0.88	-2.04	2.7	22.15	12.89	29.83
GA/TC	-2.49	-0.33	3.01	20.45	-7.8	31.28
AT/AT	-10.66	-8.28	2.62	49.13	38.25	24.78
TA/TA	-8.43	12.54	1.74	47.38	-50.68	23.74
AG/CT	-1.84	-1.43	2.82	25.62	8.39	28.52
GA/TC	-1.48	-1.41	2.84	10.55	4.12	34.86
AA/TT	-1.97	-1.33	2.79	6.8	3.09	29.54
AT/AT	-3.34	0	2.61	24.45	0	16.42
TA/TA	---	---	---	---	---	---
<b>Pu-TFO</b>						
AG/CT	0.46	0.53	2.94	2.11	-1.09	27.58
GA/TC	-1.2	1.11	3.08	18.92	-14.6	32.97
AA/TT	---	---	---	---	---	---
AG/CT	-1.14	0.88	2.99	7.39	-1.05	29.13
GA/TC	-0.79	-0.61	3.18	5.53	-5.38	34.94
AA/TT	-0.79	0.34	3.03	3.77	-7.9	32.5

Table A4.5

**Local base-pair helical parameters for structure 8PIP**

Base	Pucker amplitude (°)	Pseudorotation phase angle (°)	Puckering	DNA type
A <sub>1</sub>	32.3	184.8	C3'-exo	A-DNA
G <sub>2</sub>	35.2	104.6	O4'-endo	B-DNA
A <sub>3</sub>	39.5	47.6	C4'-exo	A-B DNA transition
T <sub>4</sub>	35	144.6	C2'-endo	B-DNA
A <sub>5</sub>	33.5	136.1	C1'-exo	B-DNA
G <sub>6</sub>	33.9	122.3	C1'-exo	B-DNA
A <sub>7</sub>	39.3	60	C4'-exo	A-B DNA transition
A <sub>8</sub>	42.3	129	C1'-exo	B-DNA
T <sub>13</sub>	29	51.8	C4'-exo	A-B DNA transition
T <sub>14</sub>	26.3	101.5	O4'-endo	B-DNA
C <sub>15</sub>	37.7	24.9	C3'-endo	A-DNA
T <sub>16</sub>	35.8	45.1	C4'-exo	A-B DNA transition
A <sub>17</sub>	37.7	131.7	C1'-exo	B-DNA
T <sub>18</sub>	35.7	108.2	C1'-exo	B-DNA
C <sub>19</sub>	33.5	81.3	O4'-endo	B-DNA
T <sub>20</sub>	32	34.6	C3'-endo	A-DNA
T <sub>21</sub>	34.9	166.1	C2'-endo	B-DNA
A <sub>22</sub>	29.3	16.1	C3'-endo	A-DNA
T <sub>23</sub>	31.9	34.5	C3'-endo	A-DNA
A <sub>24</sub>	35	206.3	C3'-exo	A-DNA
T <sub>25</sub>	38	144.1	C2'-endo	B-DNA
C <sub>26</sub>	40.7	54	C4'-exo	A-B DNA transition
T <sub>27</sub>	26.4	51.9	C4'-exo	A-B DNA transition
T <sub>28</sub>	33.8	119	C1'-exo	B-DNA
C <sub>29</sub>	34.2	117.8	C1'-exo	B-DNA
T <sub>30</sub>	34.7	139.5	C1'-exo	B-DNA
T <sub>31</sub>	34.2	123.9	C1'-exo	B-DNA

## Chapter 5

Table A5.1

### Local base-pair parameters for structure 8RER

Base Pair	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
(A) G-C	-0.22	-0.14	-0.08	-1.88	-4.58	0.2
(A) A-T	0.09	-0.19	0.16	10.12	-10.11	4.1
(A) A-T	-0.13	-0.2	0.21	-6.15	12.17	6.61
(A) T-A	-0.08	-0.17	0.22	13.17	5.82	3.8
(A) A-T	0.17	-0.15	-0.21	-19.49	3.84	3.13
(A) G-C	-0.31	-0.15	0.37	16.56	-4.42	-1.71
(B) G-C	-0.19	-0.12	-0.1	-2.47	-4.26	0.45
(B) A-T	0.07	-0.16	0.13	9.96	-8.84	4.46
(B) A-T	-0.14	-0.12	0.19	-6.8	13.47	5.65
(B) T-A	-0.09	-0.16	0.21	13.06	6.35	3.74
(B) A-T	0.17	-0.11	-0.21	-19.42	3.36	3.26
(B) G-C	-0.3	-0.16	0.35	16.26	-4.15	-2.01

Table A5.2

### Local base-pair step parameters for structure 8RER

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
(A) GA/TC	0.51	-0.13	2.97	-1.96	1.74	33.18
(A) AA/TT	-0.65	0.61	6.74	-9.37	7.21	17.58
(A) AT/AT	-0.11	-0.59	2.97	-0.45	-0.63	20.62
(A) TA/TA	0.5	0.32	7.13	4.4	-9.48	40.62
(A) AG/CT	0.58	1.93	2.73	-7.49	3.53	47.71
(A) GG/CC	----	----	----	----	----	----
(B) GA/TC	0.49	-0.11	2.98	-2.06	1.81	32.77
(B) AA/TT	-0.75	0.63	6.73	-9.7	7.35	17.39
(B) AT/AT	-0.1	-0.65	2.96	-0.49	-1.24	20.79
(B) TA/TA	0.47	0.32	7.12	4.59	-9.11	40.6
(B) AG/CT	0.57	1.96	2.75	-7.42	3.06	47.9

Table A5.3

**Local base step parameters for structure 8RER**

step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
(A) G/A	0.85	-0.06	3.06	5.89	0.16	34.99
(A) A/A	-0.68	0.66	6.78	-16.7	19.31	18.84
(A) A/T	-0.35	-0.47	2.98	7.29	-3.45	19.61
(A) T/A	0.75	0.17	6.92	-12.89	-11.14	40.08
(A) A/G	0.25	1.92	3.05	9.11	-0.9	45.33
(A) G/G	4.84	2.72	3.03	12.46	6.7	68.47
(A) G/C	-5.08	-1.56	3.44	128.26	-87.9	-30.53
(B) C/C	7.27	1.13	13.59	-99.92	-1.78	87.99
(B) C/T	-8.74	-0.86	-9.78	113.7	-52.54	-38.04
(B) T/A	-0.24	0.47	7.34	-21.57	-6.85	41.28
(B) A/T	-0.15	-0.69	2.94	8.4	1.48	22.45
(B) T/T	0.63	0.56	6.7	0.98	-4.15	16.32
(B) T/C	-0.19	-0.2	2.85	9.87	3.04	31.05
(B) C/G	16.05	-13.4	-22.69	-4.66	3.38	-122.31
(C) G/A	0.78	-0.04	3.07	5.76	0.55	34.65
(C) A/A	-0.86	0.69	6.77	-17.51	19.39	17.87
(C) A/T	-0.37	-0.58	2.97	7.3	-4.48	20.28
(C) T/A	0.71	0.2	6.91	-12.63	-11.25	40.18
(C) A/G	0.24	1.91	3.06	9.04	-1.06	45.31
(C) G/G	4.79	2.71	3.01	12.23	7.7	67.81
(C) G/C	-5.17	-1.45	3.21	129.43	-87.16	-24.24
(D) C/C	7.31	1.22	13.55	-100.1	-1.5	88.69
(D) C/T	-8.89	-0.82	-9.59	112.92	-53.56	-39.43
(D) T/A	-0.22	0.45	7.32	-21.67	-6	41.14
(D) A/T	-0.17	-0.7	2.93	8.52	1.35	22.21
(D) T/T	0.64	0.59	6.69	0.89	-3.95	16.85
(D) T/C	-0.22	-0.18	2.87	9.93	2.74	30.62

Table A5.4

**Local base-pair helical parameters for structure 8RER**

step	X-disp (Å)	Y-disp (Å)	helical rise (Å)	Inclination (°)	Tip (°)	helical Twist (°)
(A) GA/TC	-0.49	-1.19	2.92	3.05	3.45	33.28
(A) AA/TT	-4.80	-6.61	6.10	20.81	27.04	21.16
(A) AT/AT	-1.40	0.20	2.99	-1.75	1.24	20.63
(A) TA/TA	2.61	0.30	6.91	13.39	-6.21	41.88
(A) AG/CT	2.12	-1.19	2.74	4.32	9.18	48.38
(A) GG/CC	----	----	----	----	----	----
(B) GA/TC	-0.47	-1.19	2.93	3.21	3.65	32.88
(B) AA/TT	-4.87	-6.64	6.09	21.26	28.05	21.20
(B) AT/AT	-1.31	0.08	3.00	-3.43	1.36	20.83
(B) TA/TA	2.53	0.39	6.91	-12.88	-6.50	41.81
(B) AG/CT	2.18	-1.18	2.75	3.74	9.07	48.53

Table A5.5

**Local base-pair helical parameters for structure 8RER**

Base	Pucker amplitude (°)	Pseudorotation phase angle (°)	Puckering	DNA type
(A) G	38.9	196.1	C3'-exo	B-DNA
(A) A	35.2	66.3	C4'-exo	B-DNA
(A) A	35.9	171.6	C2'-endo	B-DNA
(A) T	36.1	157.4	C2'-endo	B-DNA
(A) A	45.9	160.5	C2'-endo	B-DNA
(A) G	35.8	181.5	C3'-exo	B-DNA
(A) G	23.6	33.3	C3'-endo	A-DNA
(B) C	45	144.4	C2'-endo	B-DNA
(B) C	34.2	141.4	C2'-endo	B-DNA
(B) T	37.6	139.9	C1'-exo	B-DNA
(B) A	30.9	173.7	C2'-endo	B-DNA
(B) T	42.1	128	C1'-exo	B-DNA
(B) T	36.2	107.7	O4'-endo	B-A DNA transition
(B) C	36.2	46.7	C4'-exo	B-DNA

Table A5.6

**Local base step parameters for structure 8RER**

step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
(A) G/A	0.85	-0.06	3.06	5.89	0.16	34.99
(A) A/A	-0.68	0.66	6.78	-16.7	19.31	18.84
(A) A/T	-0.35	-0.47	2.98	7.29	-3.45	19.61
(A) T/A	0.75	0.17	6.92	-12.89	-11.14	40.08
(A) A/G	0.25	1.92	3.05	9.11	-0.9	45.33
(A) G/G	4.84	2.72	3.03	12.46	6.7	68.47
(A) G/C	-5.08	-1.56	3.44	128.26	-87.9	-30.53
(B) C/C	7.27	1.13	13.59	-99.92	-1.78	87.99
(B) C/T	-8.74	-0.86	-9.78	113.7	-52.54	-38.04
(B) T/A	-0.24	0.47	7.34	-21.57	-6.85	41.28
(B) A/T	-0.15	-0.69	2.94	8.4	1.48	22.45
(B) T/T	0.63	0.56	6.7	0.98	-4.15	16.32
(B) T/C	-0.19	-0.2	2.85	9.87	3.04	31.05
(B) C/G	16.05	-13.4	-22.69	-4.66	3.38	-122.31
(C) G/A	0.78	-0.04	3.07	5.76	0.55	34.65
(C) A/A	-0.86	0.69	6.77	-17.51	19.39	17.87
(C) A/T	-0.37	-0.58	2.97	7.3	-4.48	20.28
(C) T/A	0.71	0.2	6.91	-12.63	-11.25	40.18
(C) A/G	0.24	1.91	3.06	9.04	-1.06	45.31
(C) G/G	4.79	2.71	3.01	12.23	7.7	67.81
(C) G/C	-5.17	-1.45	3.21	129.43	-87.16	-24.24
(D) C/C	7.31	1.22	13.55	-100.1	-1.5	88.69
(D) C/T	-8.89	-0.82	-9.59	112.92	-53.56	-39.43
(D) T/A	-0.22	0.45	7.32	-21.67	-6	41.14
(D) A/T	-0.17	-0.7	2.93	8.52	1.35	22.21
(D) T/T	0.64	0.59	6.69	0.89	-3.95	16.85
(D) T/C	-0.22	-0.18	2.87	9.93	2.74	30.62

## Chapter 6

Table A6.1

**Local base-pair parameters for structure 8RMH**

Base Pair	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
(A) T+T	0.82	3.14	-0.27	-0.75	-9.45	-89.61
(A) G+G	-1.56	-3.46	0.09	-4.85	5.08	89.78
(A) G+G	1.59	3.4	-0.03	6.41	-4.12	-89.75
(A) T+T	-0.72	-3.34	-0.01	-1.47	-0.14	89.99
(A) G+G	1.72	3.49	0.01	5.12	-2.14	-89.87
(A) G+G	-1.5	-3.54	0.05	-2.71	2.85	89.93
(A) T+T	0.19	3.8	-0.15	-0.37	-4.6	-89.91
(B) G+G	1.5	3.57	-0.08	-3.05	-1.28	-89.95
(B) G+G	-1.53	-3.55	0.07	-2.57	3.24	89.93
(B) G+G	-1.58	-3.53	0.1	-4.67	5.45	89.77
(B) G+G	1.52	3.54	0.03	3.69	-0.7	-89.94
(B) T+T	-0.62	-2.99	0.1	0.28	3.66	89.94

Table A6.2

**Local base step parameters for structure 8RMH**

step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
(A) T/G	-0.22	-0.56	3.51	-1.68	3.67	35.27
(A) G/G	-0.64	-0.89	3.29	-1.74	2.46	26.26
(A) G/T	-0.4	-1.69	3.63	-2.59	-2.21	25.25
(A) T/G	-0.7	-0.79	3.17	1.22	4.18	34.65
(A) G/G	-0.77	-0.92	3.33	-1.12	0.38	25.85
(A) G/T	-0.33	-1.45	3.26	2.21	-1.08	30.97
(A) T/T	-1.39	-4.94	-36.43	-3.01	5.42	67.28
(B) T/G	-0.88	-0.75	3.45	-2.73	3.63	35.8
(B) G/G	-0.55	-0.69	3.27	2.31	1.42	19.58
(B) G/T	-13.56	-6.67	-8.87	103.86	94.56	-74.99

Table A6.3

**Local base-pair step parameters for structure 8RMH**

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
(A) GT/TG	0.15	0.39	-3.49	1.19	-2.6	-35.31
(A) GT/TG	0.62	0.53	-3.45	1.93	-2.56	-35.83
(A) GG/GG	0.8	1.5	-3.48	2.63	-1.03	-53.92
(A) TG/GT	0.23	1.02	-3.19	-1.57	0.77	-30.93
(A) TG/GT	0.29	1.19	-3.62	1.83	1.56	-25.26
(A) TG/GT	-0.15	-0.34	-3.13	-1.19	-0.59	28.09
(A) GT/TG	0.5	0.56	-3.18	-0.86	-2.96	-34.65
(B) GG/GG	0.54	0.65	-3.33	0.8	-0.27	-25.87
(B) GG/GG	0.45	0.63	-3.29	1.23	-1.74	-26.3
(B) GG/GG	0.42	0.93	-3.41	0.87	2.06	-32.93
(B) GG/GG	0.39	0.49	-3.27	-1.64	-1.01	-19.56

Table A6.4

**Main chain and  $\chi$  torsion angles for structure 8RMH**

	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\chi$
(A) anti-T	---	---	58.7	142.5	-174.9	-92.3	-119.9
(A) anti-G	-71.4	-169.2	47.7	131.4	179.9	-98.2	-108.8
(A) anti-G	-64.9	177.9	51.3	120.5	179.6	-91.2	-120.9
(A) anti-T	-63.5	179	51.8	120.5	-167.4	-119.4	-131
(A) anti-G	-34	-176.4	22.7	148.9	-176.4	-104.8	-105.1
(A) anti-G	-60.3	174.9	19.3	131.7	-171.6	-104.5	-114.2
(A) anti-T	-69.9	177.4	48.2	102.2	---	---	-128.8
(B) anti-T	---	---	54	146.7	-173.6	-113.4	-112.7
(B) anti-G	-58.5	-158.9	33.8	146.4	-173	-97.1	-96.6
(B) anti-G	-66.8	170.2	47.6	125.5	-85.6	178.7	-108.7
(B) anti-T	73.3	175.6	53.3	149.5	-91.2	-60.8	-129.7
(B) anti-G	-54.8	-123.5	50.9	145.4	-169.9	-92.4	-133.3
(B) anti-G	-73.9	-165.5	41.6	141.4	-145.2	-163.7	-102.6
(B) syn-T	71.2	148.5	-69.5	140.9	---	---	62.8

## Publications from this work

Dalla Pozza, M., Abdullrahman, A., Cardin, C. J., Gasser, G. & Hall, J. P. Three's a crowd - stabilisation, structure, and applications of DNA triplexes. *Chem Sci* **13**, 10193–10215 (2022).

Cite this: *Chem. Sci.*, 2022, 13, 10193

## Three's a crowd – stabilisation, structure, and applications of DNA triplexes

Maria Dalla Pozza, †<sup>a</sup> Ahmad Abdullrahman, †<sup>b</sup> Christine J. Cardin, \*<sup>c</sup> Gilles Gasser \*<sup>a</sup> and James P. Hall \*<sup>b</sup>

DNA is a strikingly flexible molecule and can form a variety of secondary structures, including the triple helix, which is the subject of this review. The DNA triplex may be formed naturally, during homologous recombination, or can be formed by the introduction of a synthetic triplex forming oligonucleotide (TFO) to a DNA duplex. As the TFO will bind to the duplex with sequence specificity, there is significant interest in developing TFOs with potential therapeutic applications, including using TFOs as a delivery mechanism for compounds able to modify or damage DNA. However, to combine triplexes with functionalised compounds, a full understanding of triplex structure and chemical modification strategies, which may increase triplex stability or *in vivo* degradation, is essential – these areas will be discussed in this review. Ruthenium polypyridyl complexes, which are able to photooxidise DNA and act as luminescent DNA probes, may serve as a suitable photophysical payload for a TFO system and the developments in this area in the context of DNA triplexes will also be reviewed.

Received 28th March 2022

Accepted 2nd August 2022

DOI: 10.1039/d2sc01793h

rsc.li/chemical-science

### 1. Introduction

DNA is the carrier of genetic information in all cellular systems and in many viruses. As the carrier of genetic material, it directs

its own replication during the cell division process and the transcription of complementary molecules of RNA. One of the defining features of DNA is its structural flexibility. DNA can adopt a wide range of higher order structures including the duplex, G-quadruplex, i-motif and Holliday Junction, all of which either have a confirmed or suspected role in gene regulation and/or transcription processes<sup>1</sup> and have been investigated in the context of ligand targeting.<sup>2</sup>

The DNA triplex is of particular interest due to its potential for exploitation in the targeting of therapeutics to specific DNA sequences. The triplex is formed when a DNA duplex is joined by a third strand, which binds in the major groove of the duplex to form a three-stranded assembly. Research efforts have

<sup>a</sup>Chimie ParisTech, PSL University, CNRS, Institute of Chemistry for Life and Health Sciences, Laboratory for Inorganic Chemical Biology, F-75005 Paris, France. E-mail: gilles.gasser@chimieparistech.psl.eu; Web: [www.gassergroup.com](http://www.gassergroup.com)

<sup>b</sup>Department of Pharmacy, Chemistry and Pharmacy Building, University of Reading, Whiteknights Campus, Reading, Berkshire, RG6 6AD UK. E-mail: james.hall@reading.ac.uk

<sup>c</sup>Department of Chemistry, University of Reading, Whiteknights, Reading, RG6 6AD, UK. E-mail: c.j.cardin@reading.ac.uk

† These authors have contributed equally to this work.



Maria Dalla Pozza received her M.Sc. degree in Chemistry in 2017 from the University of Padua (Italy). After gaining experience in the industrial sector, she started a PhD in 2020 in the group of Prof. Gilles Gasser at Chimie ParisTech, PSL University (France) in co-supervision with Prof. Christine Cardin (University of Reading, UK). Her PhD is funded by the ETN European consortium Nature led

by Prof. Andrew Kellert (Dublin City University, Ireland). She is currently investigating the use of metal complexes to cleave DNA.



Ahmad Abdullrahman carried out his bachelor's studies in biotechnology at University of Urbino Carlo Bo (2016) and master's in medical biotechnology at University of Turin (2018) where he developed interests in nucleic acids technology and chemoresistance. He is currently a PhD student at the University of Reading under the supervision of Dr James Hall. His research focuses on structure

determination of DNA triplexes using X-ray crystallography.



increasingly focussed on TFO modification, to aid delivery *in vivo*, reduce or prevent degradation by nucleases and increase triplex stability. However, to fully understand the structure of the DNA triplex and how modifications and bound ligands can affect its stability, it is important to first examine the structure of the DNA duplex.

The most common and best-known form of DNA is the B-DNA form, characterized by two polynucleotide strands with a right-handed helical twist about a long axis to form a double helix, bound together by hydrogen bonds and further stabilised by  $\pi$ -stacking between adjacent bases. This winding generates two grooves: the major one is wide and deep, while the minor groove is narrow (Fig. 1). This structure has been widely characterised by X-ray diffraction and occurs at high humidity and with a variety of DNA counterions including  $\text{Na}^+$ , which serves to balance the negative charge of the phosphate backbone.<sup>3,4</sup> The most significant characteristic of B-DNA is the possibility to accommodate only two types of naturally occurring base pairs

(*i.e.*, adenine–thymine A–T and cytosine–guanine C–G). In B-DNA both base pairs can be replaced by each other without altering the position of the sugar–phosphate backbone, although runs of A–T base pairs are known to have a narrower minor groove. Similarly, the double helix is not disturbed by swapping the partners (*i.e.*, changing a C:G with a G:C or a T:A with a A:T). However, different combinations of bases lead to the formation of non-Watson–Crick base pairs with a significant distortion of the double helix. Since the variation of pairing causes distortions, DNA is a molecule able to adopt different non-canonical structures whilst exposed to physiological and non-physiological conditions. When the relative humidity is reduced to 75%, the B-DNA changes conformation, adopting the so-called A-DNA form, which presents a wider and flatter right-handed helix compared to the B-DNA form (Fig. 1). In contrast to the right-handed form, Z-DNA is a left-handed analogue which has a deep minor groove and a shallow but wide major groove.<sup>5,6</sup> Z-DNA is formed as a function of DNA sequence and contains long sections of alternating purine–pyrimidine bases, most commonly as GC repeat units.

In addition to these, DNA can also form other non-canonical structures as a function of sequence, which are especially stable in the crowded intracellular environment. These arrangements were demonstrated to play a role in different biological processes such as replication, transcription, translation and reverse translation.<sup>7</sup> Three strands of DNA can form a triplex structure, which was initially predicted to exist in 1953 by Pauling and co-workers and subsequently observed by Rich and co-workers after mixing poly U and poly A ribonucleotides in a 2 : 1 ratio.<sup>8,9</sup> Triplex formation has been identified both *in vitro* and *in vivo*,<sup>10</sup> as will be discussed in Section 5. Tetraplex structures, known as G-quadruplexes, have also been observed in G-rich strands. They are formed in sequences containing multiple guanine tracts within a G-rich sequence and are bound together by Hoogsteen hydrogen bonding.<sup>11</sup> G-Quadruplexes have interestingly been observed in many different locations, correlated with genomic regions that play a functional role such as



*Christine Cardin has been Professor of Crystallography since 2005, specialising in nucleic acid structures and other supra-molecular assemblies. A graduate of the University of Oxford, her research career started in organometallic chemistry and bio-inorganic chemistry, leading to a lectureship in Trinity College Dublin in 1977. She was elected to Fellowship in 2002, for her work in developing X-ray crystal-*

*lography in Ireland, including the installation of the first diffractometer in the Republic of Ireland. In 2016 she was awarded the RSC Cornforth Medal for collaboration with her former Irish colleagues. She has co-authored more than 170 papers.*



*Gilles Gasser started his independent scientific career at the University of Zurich (Switzerland) in 2010 before moving to Chimie ParisTech, PSL University (Paris, France) in 2016 to take a PSL Chair of Excellence. Gilles was the recipient of several fellowships and awards including the Alfred Werner Award from the Swiss Chemical Society, an ERC Consolidator Grant, the European Bio-*

*Inorganic Chemistry (EuroBIC) medal and the Pierre Fabre Award for therapeutic innovation from the French Société de Chimie Thérapeutique. Gilles' research interests lay in the use of metal complexes in different areas of medicinal and biological chemistry.*



*James Hall began his independent scientific career in the Department of Pharmacy, University of Reading in 2017 and was promoted to Associate Professor of Biomolecular Structure and Educational Technology in 2021. James is interested in nucleic acid structure and function at a molecular level and is a co-author of over 45 crystal structures published in the Protein Data Bank. James*

*also undertakes pedagogic and educational research and development, with a particular interest in incorporating virtual reality methods within teaching.*



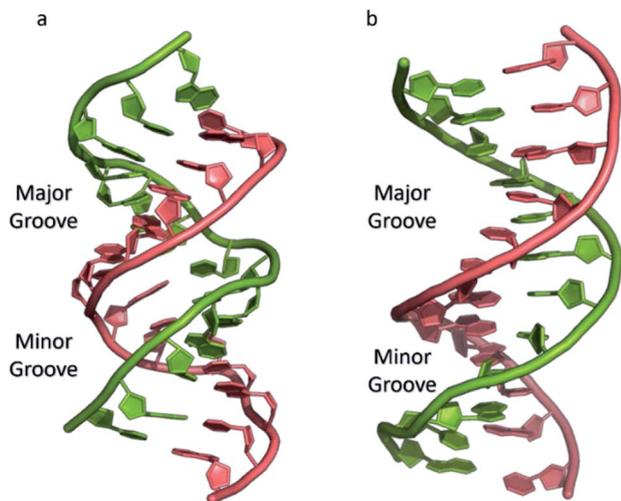


Fig. 1 (a) A-DNA, (b) B-DNA.

replication origin sites, telomeres and promoter regions.<sup>12</sup> Another type of tetraplex structure is the intercalated motif (i-motif), formed between C-rich strands in acidic conditions. C-Rich sequences are found in telomeres, and in promoter regions of many human genes, indicating a probable role in biological processes.<sup>13</sup> Finally, the cruciform structure is formed by intra-strand base pairing of inverted repeat sequences. It can be either a four-way junction or a three-way junction depending on the number of hairpins present (Fig. 2).<sup>14</sup> In this review, we will focus on DNA triplex structures, discussing their structural characteristics, stability, and their potential applications. The DNA triplex has been investigated for decades as a very promising tool in gene editing, but development has been challenging, due both to the low thermal stability of the structure, and the poor cellular uptake of the triplex-forming oligonucleotides. The possible biological application of triplexes and approaches to mitigate their limitations will be covered in this review. The application and interaction of ruthenium

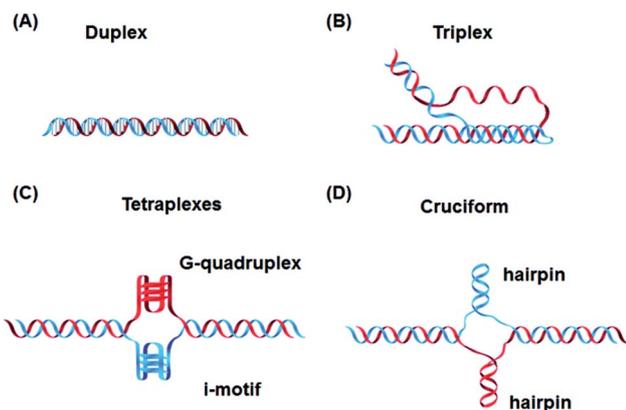


Fig. 2 Canonical DNA structure and non-canonical structures including (A) duplex, (B) triplex, (C) G-quadruplex and i-motif and (D) hairpin. Reprinted from H. Tateishi-Karimata and N. Sugimoto, *Chem. Commun.*, 2020, 56, 2379.

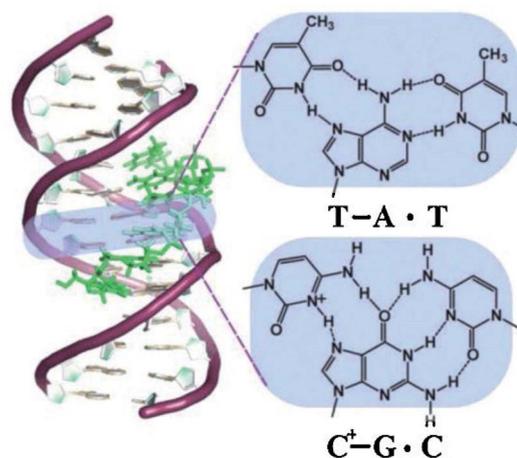


Fig. 3 Schematic drawing of a triplex forming oligonucleotide that specifically recognises a DNA sequence, with the TFO binding in the major groove of the DNA duplex.<sup>179</sup> Reprinted from *Coord. Chem. Rev.*, 257, Tarita Biver, Stabilisation of non-canonical structures of nucleic acids by metal ions and small molecules, 2765–2783, Copyright (2013), with permission from Elsevier.

polypyridyl complexes with DNA triplexes will also be discussed, to explore potential future therapeutic applications in areas such as photodynamic therapy (PDT). Ruthenium polypyridyl complexes possess useful properties which are particularly suitable for biological applications, as presented in Section 6 of this review. Indeed, Ru-based compounds have been intensively studied in the last decades as antiparasitic, antimicrobial or anticancer drug candidates.<sup>15,16</sup> In particular, ruthenium polypyridyl compounds have attracted much interest.<sup>17</sup> Their ability to absorb light *via* a metal-to-ligand charge transfer (MLCT) process among other charge transfers have made them very interesting tools for photodynamic therapy (PDT).<sup>18,19</sup> Therefore, we suggest that the intrinsic triplexes' sequence-specific binding properties combined with the phototoxicity of ruthenium derivatives can be exploited together to obtain breakthrough tools in gene editing technology.

## 2. Type of triplexes

Triplex structures can be formed by DNA, RNA or hybrids of the two. This review focuses on DNA triplexes, so RNA-containing triplexes will not be considered here. DNA triplexes can be grouped based on the origin of the third strand. Intermolecular triplexes are formed between a double-stranded DNA (dsDNA) and an independent molecule termed the triplex-forming oligonucleotide (TFO). If the third strand is part of a single strand which also contains the dsDNA, the triplex is referred to as an intramolecular triplex. The hydrogen bonds between the two helices of DNA are typically Watson–Crick bonds, whereas the bonds between the duplex and TFO are either Hoogsteen or reverse-Hoogsteen bonds (Fig. 3). The directionality of the TFO can be either parallel or anti-parallel to the DNA strand which forms the hydrogen bonds.



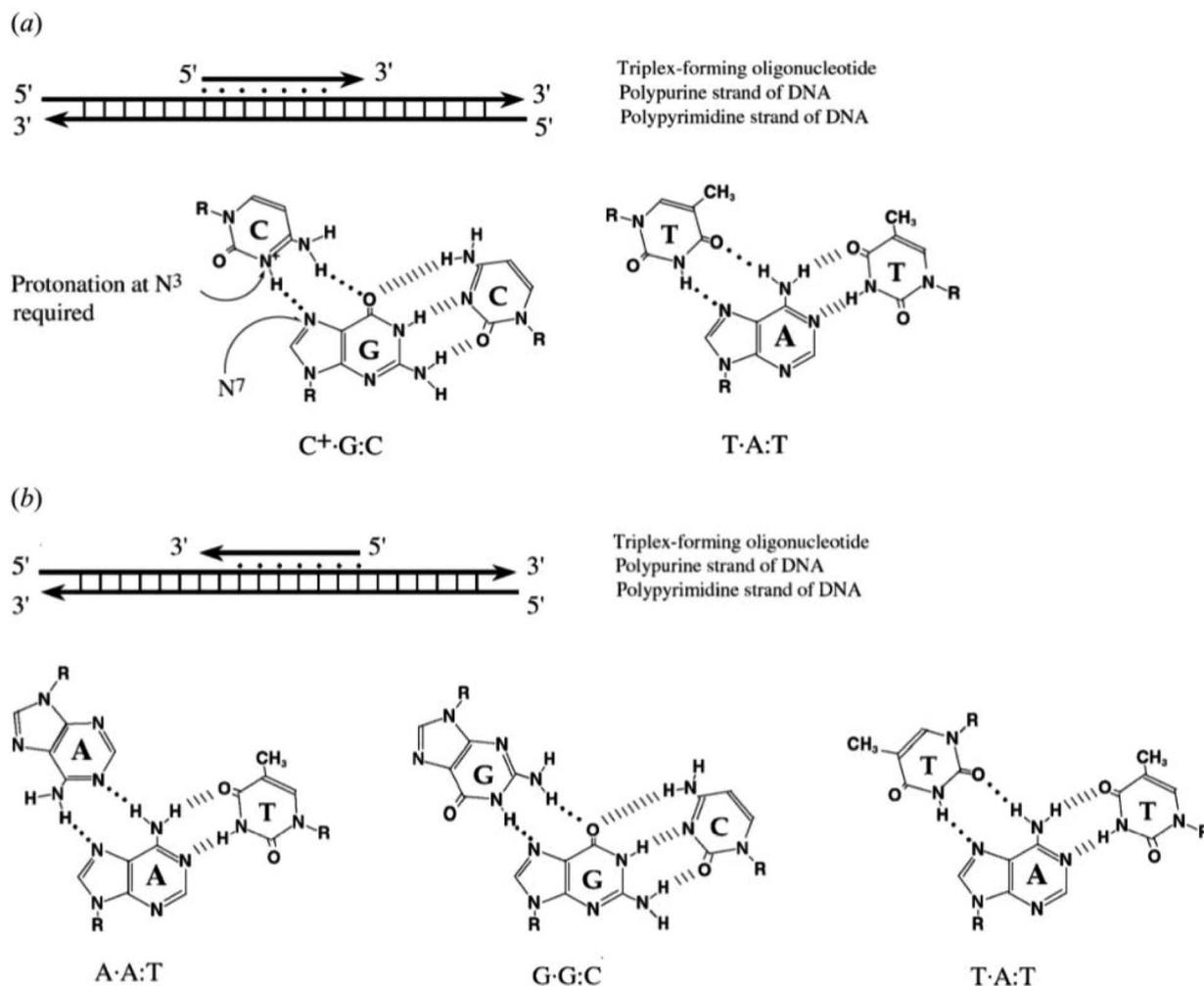


Fig. 4 Intermolecular triplexes and canonical base triplets. (a) Polypyrimidine triplexes Y-R:Y (b) polypurine triplexes R-R:Y<sup>96</sup> Reprinted from K. M. Vasquez and P. M. Glazer Triplex-forming oligonucleotides: principles and applications, *Q. Rev. Biophys.*, 35, 89–107, copyright 2002, with permission from Cambridge University Press.

## 2.1 Intermolecular DNA triplexes

To explain the possible combinations of intermolecular DNA triplexes, a close examination of the sequence of the triplex-forming species is required (Fig. 4). In a polypyrimidine TFO that consists entirely of pyrimidines, the thymine will bind to the adenosine T·A:T or cytosine binds to guanine C·G:C, forming a triplex. The cytosine, however, requires a protonation of the N3 atom to ensure the second Hoogsteen bond with the guanine. Therefore, these parallel triplexes require a mildly acidic environment.<sup>20</sup> However, there is a limit to protonation that, if not respected, will result in charge repulsion between the adjacent cytosines.<sup>21</sup> When a TFO contains only purine bases, adenine binds to adenine (A·A:T) or guanine binds to guanine (G·G:C) with reverse-Hoogsteen hydrogen bonds. In contrast, a polypurine TFO forms a triple-helix by binding the duplex with an anti-parallel conformation.<sup>22</sup> Additionally, in the anti-parallel conformation, it is also possible to have T·A:T steps within the DNA triplex.<sup>23</sup>

The base identity plays a key role in determining the local and overall twist angle of the DNA triplex. This residual twist is calculated based on the measurement of the angle between the two carbon atoms of the adjacent Hoogsteen base pairs and the base of interest.<sup>24</sup> G·G:C triplets have the effect of increasing twist within the triplex, with an average increase of 10.6° per step, whilst T·A:T steps reduce the twist by the same value, with the overall twist angle of the helix being maintained at 30°. This is lower than the average twist for a B-DNA duplex of ca. 34° and therefore suggests that the binding of a TFO induces a slight unwinding of the duplex. This results in significant distortion after each A·T bond of the duplex within the polypurine triplexes. By contrast, the polypyrimidine triplex has much less backbone distortion and a higher number of hydrogen bonds between the TFO and the duplex, compared with polypurine. This reduction in distortion is one possible reason why parallel triplexes are generally more stable than antiparallel helices.<sup>22</sup>



Table 1 DNA-containing structures of triplexes deposited in the Protein Data Bank<sup>31</sup>

Intramolecular or intermolecular	Triplex type	Nucleic acid type	Method	PDB ID	Year	Reference
<b>DNA only</b>						
Intramolecular	Antiparallel	DNA	NMR	134D	1993	34
Intramolecular	Antiparallel	DNA	NMR	135D	1993	34
Intramolecular	Antiparallel	DNA	NMR	136D	1993	34
Intramolecular	Antiparallel	DNA	NMR	177D	1994	35
Intermolecular	Parallel	DNA	NMR	149D	1994	34
Intermolecular	Parallel	DNA	X-Ray diffraction	208D	1995	36
Intramolecular	Parallel	DNA	NMR	1AT4	1997	37
Intramolecular	Parallel	DNA	NMR	1D3X	1998	38
Intramolecular	Parallel	DNA	NMR	1BCB	1998	39
Intramolecular	Parallel	DNA	NMR	1BCE	1998	39
Intermolecular	Parallel	DNA	X-Ray diffraction	1D3R	1999	40
Intermolecular	Parallel	DNA	NMR	1BWG	1999	41
Intramolecular	H-DNA H-Y5 isomer	DNA	NMR	1B4Y	1999	42
Intermolecular	G-Triplex	DNA	NMR	2MKM	2014	43
Intermolecular	G-Triplex	DNA	NMR	2MKO	2014	43
<b>Modified DNAs</b>						
Intermolecular	P-Form	DNA + PNA	X-Ray diffraction	1PNN	1995	44
Intramolecular	Parallel	DNA + 1-(2-deoxy-beta-D-ribofuranosyl)-4-(3-benzamido) phenylimidazole	NMR	1WAN	1996	45
Intramolecular	Parallel	DNA + N7-glycosylated guanine	NMR	1GN7	1997	46
Intramolecular	Parallel	DNA + 1-propynyl deoxyuridine in third strand	NMR	1P3X	1998	47
Intramolecular	Parallel	DNA + LNA	NMR	1W86	2004	48
Intermolecular	P-Form	PNA	X-Ray diffraction	1XJ9	2005	49
Intramolecular	Antiparallel	DNA + TINA intercalator	NMR	6QHI	2019	50

## 2.2 Intramolecular DNA triplexes

In addition to the intermolecular DNA triplexes, where the TFO is an external oligo, the triplex can be formed by one DNA strand which folds back on itself, to form an intramolecular assembly. These are commonly referred to as H-DNA (hinged DNA), as their stability depends on the presence of acidic pH and negative superhelical stress. H-DNAs may be formed under supercoiled conditions with a mirror repeat polypurine-polypyrimidine sequence and the base motifs are the same as in the intermolecular triplexes with a pyrimidine third strand. Moreover, an intramolecular triplex composed with bases of pyrimidine-purine-purine in the DNA stretches, and a non-mirror repeat, is defined as \*H-DNA.<sup>25</sup>

## 2.3 G-triplexes, R-DNA and PNA

It is also possible to form a triplex from G bases – the G-triplex, which contains a strand rich in guanine bases, and can be formed as an intermediate during the formation of a DNA G-quadruplex.<sup>26,27</sup> Using fluorescence resonance energy transfer (FRET), it was determined that G-triplexes can assume both parallel and anti-parallel topologies.

A parallel DNA triplex may also be formed during homologous recombination and assists the recruitment of the

homologous sequences. During the formation of the recombinant DNA (R-DNA), a complex with Rec-A may be formed, leading to a triplex with an extended rise distance of 5.1 Å, compared to a standard rise distance of 3.4 Å.<sup>28</sup>

Peptide nucleic acids, PNA, are modified oligonucleotides that contain a polyamide chain, instead of the sugar-phosphate backbone.<sup>29</sup> Whilst the bases retain the canonical Watson-Crick pairing scheme, the PNA backbone lacks the negative charge associated with a phosphate backbone and therefore PNA can form a highly stable triplex with one or more DNA strands with reduced electrostatic repulsion. The binding directionality respect of the ds-DNA molecule can be both parallel, or anti-parallel forming a stable D-loop, *i.e.*, forming a momentary triple strand with one of the DNA strands.<sup>30</sup> This DNA triplex can be seen as a triple-helix assembly, the stability of which can be increased by the incorporation of synthetic modifications, which will be discussed later in this article.

## 3. Structural analysis of triplexes

At time of writing, structural characterizations of triplexes are limited. Only 32 structures, with the majority solved using NMR, have been published in the Protein Data Bank.<sup>31</sup> This includes triplexes composed of hybrids of DNA-RNA, DNA-PNA



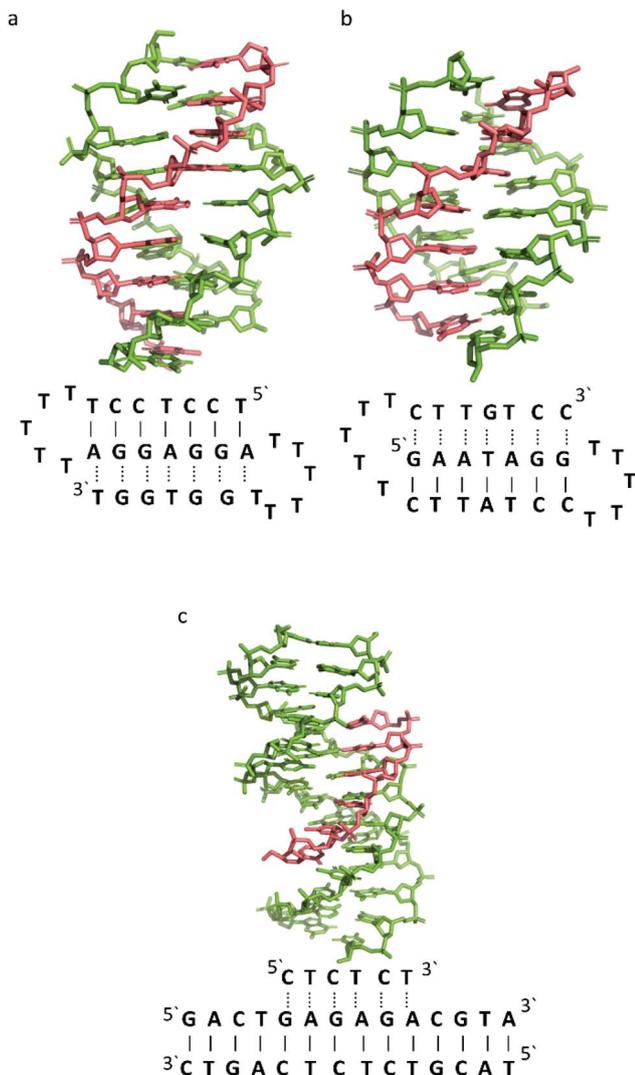


Fig. 5 3D Representation of and schematic diagram of (a) triplex A (intramolecular antiparallel, PDB ID 134D), (b) triplex B (intramolecular parallel, PDB ID 149D) (c) triplex C (intermolecular parallel, PDB ID 1BWG). The TFO is displayed in red and the DNA duplex is in green. In the schematic diagrams, Watson–Crick hydrogen bonding is displayed using lines with Hoogsteen bonds illustrated in dashed lines.

and RNA–RNA triplexes, some of which contain modified bases, sugars or intercalators, and excluding any structures which contain proteins. Only four of the structures published,

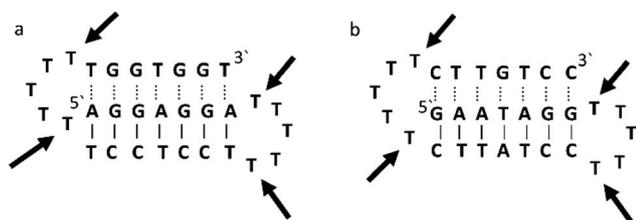


Fig. 6 Schematic representation of (a) triplexes A (PDB ID 134D) and (b) triplex B (PDB ID 149D). The arrows indicate the four thymine that are reported in the analysis, but do not bind to any complementary base.

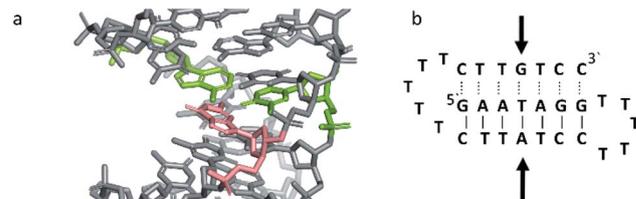


Fig. 7 (a) 3D and (b) schematic representations of the G–T:A triplet of the triplex B. Green indicate the duplex bases, guanine and adenine, while the orange base is the guanine of the TFO.

containing DNA, have been determined using X-ray diffraction. DNA-containing triplex structures obtained by X-ray analysis were either formed with protein nucleic acid (PNA), intercalators, or as a result of DNA overlap with only a small number of bases forming Hoogsteen bonds and therefore do not represent a full and complete true DNA triplex, unlike several of the structures solved using NMR. Structure determinations of DNA-containing triplexes are summarized in Table 1.

To better illustrate the structural influence of the binding of a TFO to a DNA duplex to yield a triplex, three DNA triplexes, triplex A, B and C, were selected. The three structures were chosen as examples of triple helix structures which did not contain intercalators, other small molecules or chemical modifications. As the structures were solved using NMR, they are representative of triplex species in solution (Fig. 5).

### 3.1 Similarity with B-DNA

B-DNA is a right-handed form of the double helix, with 10.1 base pairs in each turn<sup>3</sup> and a helix diameter of 20 Å.<sup>32</sup> Fibre diffraction data, obtained from X-ray studies, show that the average value of the helical twist per base pair is 36.1°, but that this can vary from 24° to 51°. The distance between bases (rise) is 3.4 Å per base pair. Whilst B-DNA is the most frequently encountered DNA conformation in physiological conditions,<sup>33</sup> others are possible, including A- and Z-forms, and are promoted by both sequence and changes in the DNA microenvironment.

The B-DNA structure forms two grooves, a minor and major with a width of ~5.7 Å and ~11.7 Å, respectively. The value is obtained by subtracting 5.8 Å from the distance between the phosphate groups on opposing strands, which is the van der Waals radius of one phosphate group.<sup>51</sup>

The DNA triplex possesses significant similarity in structure to the B-form duplex. The base rise distance remains consistent at 3.3 Å and the twist value of triplex A is also similar to a standard B-form duplex at *ca.* 34°. Triplex A contains a polypurine TFO, as illustrated in Fig. 5.

Triplex A is an intramolecular triplex constructed from a single oligonucleotide. However, the loop positions could not be assigned due to disorder and are therefore not included in the structural coordinates. Whilst loop bases may not form hydrogen bonds with each other or with the TFO, and therefore disorder within this region is expected, the T bases indicated by arrows in Fig. 6 adopt T–T wobble pairs, indicating two hydrogen bonds are present between the first T bases in each



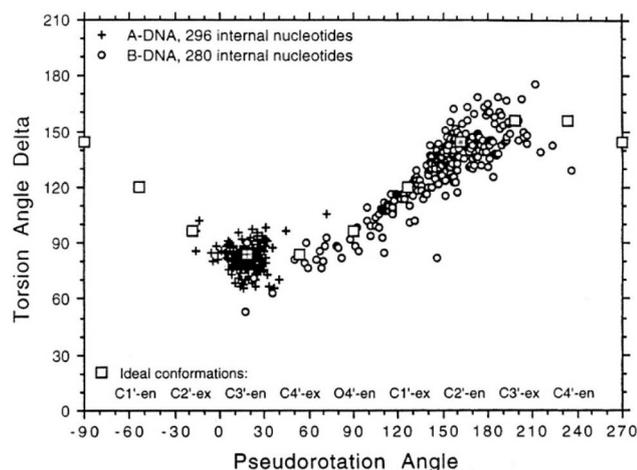


Fig. 8 Representation of sugar rings of B-DNA (circle) and A-DNA (crosses) based on pseudorotation and torsion angle.<sup>180</sup> Reproduced with permission from R. E. Dickerson, *International Tables for X-ray Crystallography*, Volume F: *Macromolecular Crystallography*, ed. M. G. Rossmann, E. Arnold (International Union of Crystallography, Chester, U.K. (2001)).

loop. The average base pair twist at this pair is  $32^\circ$  which is slightly reduced compared to the average helical twist value for B-DNA ( $36.1^\circ$ ). However, other than this there is no significant perturbation to the duplex part of the triplex structure compared to B-DNA, highlighting that the interaction of the TFO-region in the major groove does not significantly alter the structure of the template duplex.

Triplex B is also an intramolecular triplex but with a TFO composed of purine bases that bind the duplex strand in a parallel arrangement (Fig. 5). Whilst the overall structure shows little difference with that of triplex A, which adopts an antiparallel arrangement, local distortions can be observed in individual base triplets. The most significant of these is in the central step within the triplex, as indicated by arrows in Fig. 7b. At this step, the G base in the triplex strand is unable to form a proper binding interaction with the T-A base pair (a T base would be needed for this to occur), as illustrated in Fig. 7. Whilst this mismatch of bases would be expected to reduce the overall stability of the triplex assembly, individual sites of mismatched bases do not necessarily prevent triplex formation.

Perturbation of other derived parameters within the structure, including changes in the propeller and buckle value either side of the mismatch site, indicate that this single step of instability may result in an overall reduction of stability or rigidity across the triplex assembly<sup>6</sup> despite an overall twist value of  $30.8^\circ$ , which is reduced compared to that found for B-DNA.

In triplex C the TFO is a purine-rich hexamer oligonucleotide that binds in the major groove of a 13 base-pair duplex, forming a parallel triplex assembly (Fig. 5). In contrast to triplexes A and B, triplex C is an intermolecular assembly and the length of the TFO is shorter than the duplex to which it is bound. As illustrated in the schematic representation of the structure (Fig. 5), the TFO forms both TA  $\times$  T and CG  $\times$  C+ triplets, with charge

neutralization of the C+ bases by the phosphate backbone being expected to confer greater stability on the assembly.<sup>41</sup>

Whilst the triplex section of triplex C is structurally similar to A and B, this structure gives insight into the structure of the helix either side of the TFO. Whilst the overall twist angle per step within the triplex region is maintained at *ca.*  $33^\circ$ , the remaining duplex steps display much greater variability, with twist angles ranging from  $29\text{--}45^\circ$  per step.

The reduced twist angle common to triplex structures raises the question of whether the duplex component is closer to B-DNA or A-DNA in conformation, the latter of which is characterized by a reduced twist of *ca.*  $32^\circ$  per base in combination with a dominant C3'-endo sugar pucker for the ribose ring in the bases.

To determine whether the DNA triplex has an A or B conformation, the angle values needed are the backbone sugar torsion  $\delta$ , the glycosyl torsion  $\chi$  or the pseudorotation angle of sugar rings P...P.<sup>52</sup> Typically, the duplex can adopt the A- or B-form, and this is dependent on the sugar pucker adopted in each nucleotide. An A-form is adopted when the dominant sugar pucker is C3'-endo, with a pseudorotation value of between  $-30^\circ$  and  $40^\circ$ , while a wider range of pseudorotation values can be indicative of the B-DNA conformation. Indeed, the B-conformation is not limited to the C2'-endo pucker, where the majority of the nucleotides can be found, but can adopt several other forms including C4'-exo, O4'-endo, C1'-exo, C3'-exo and C4'-endo<sup>53</sup> (Fig. 8). The dominant sugar pucker can be used to assign the overall conformation of the helix and is particularly important for the development of ligands designed to target specific steps, as a change in sugar pucker will change the spatial arrangement of atoms around the binding site, potentially changing the mode of interaction by the ligand. The overall conformation of the duplex component of the triplex can be assigned to a conformation using the pseudorotation value (P) for each base.<sup>54</sup>

In triplex A, the P value for the bases forming the duplex lie within the range of  $100\text{--}160^\circ$ , indicating a majority B-DNA conformation. Whilst the terminal bases in the duplex lie outside of this range, this could be because of torsional stress placed on the structure due to the folding of the loops, which have not been presented in the coordinates for the structure. The TFO strand however, displays much less variation in the sugar pucker values. Whilst these again indicate a B-like conformation, the majority of values are either *ca.*  $176^\circ$  or are within a range of  $50\text{--}70^\circ$ , indicating less flexibility in the TFO strand compared to the duplex. This trend, of an overall B-DNA conformation for the duplex matched with less variation in the P for the TFO, is observed in all three structures. However, there are exceptions such as the central G:T-A triplet in Triplex B, which has a (P) value of  $19.0^\circ$ , indicating an A-like C3'-endo sugar pucker. This may be a pucker which is sequence dependent, or could indicate torsional strain in the TFO which is corrected in the central step by the adoption of this unusual pucker.

Due to the relatively small number of DNA triplex structures available, it is not yet possible to identify the expected structural variation as a function of sequence. However, a better



understanding of the structural variation expected for the DNA triplex may assist with the development of ligands designed to bind to specific sites within the assembly, to understand the distinctive behavior of triple helices more structural analysis is essential.

## 4. Stability of DNA triplexes

DNA triplexes are inherently less stable than their duplex counterparts in part due to the increased negative charge density from the phosphate backbones, which increases repulsion between the strands. However, multiple factors can affect the stability of a triplex assembly including the presence and concentration of monovalent or divalent cations, pH and temperature. Additionally, triplex hybridization can be promoted by the presence of molecular crowding and chromatin accessibility in the biological environment.

Efforts have been made to increase triplex stability through chemical modification of the base, sugar, or phosphate

backbone of DNA.<sup>55,56</sup> Base modifications have been the focus of extensive synthetic efforts due to pH being one significant factor which can negatively affect triplex stability. Modification has also focused on changing the phosphate and sugar within the nucleotides to enhance resistance to nucleases in the cell, in order to reduce degradation, and to enhance the ability of the TFO to enter and bind in the major groove of the duplex.

Finally, the use of ligands, such as intercalators or groove binders, has been explored as one approach to increase triplex stability without chemical modification to the TFO, although this is a secondary effect of targeting the triplex assembly with such a molecule.

### 4.1 Cations and anion enhancement of DNA triplex stability

The cellular microenvironment exercises direct control over triplex stability and activity at the molecular level. Considering the intense negative charge of a structure that is formed by three strands of DNA, a high concentration of multivalent cations will

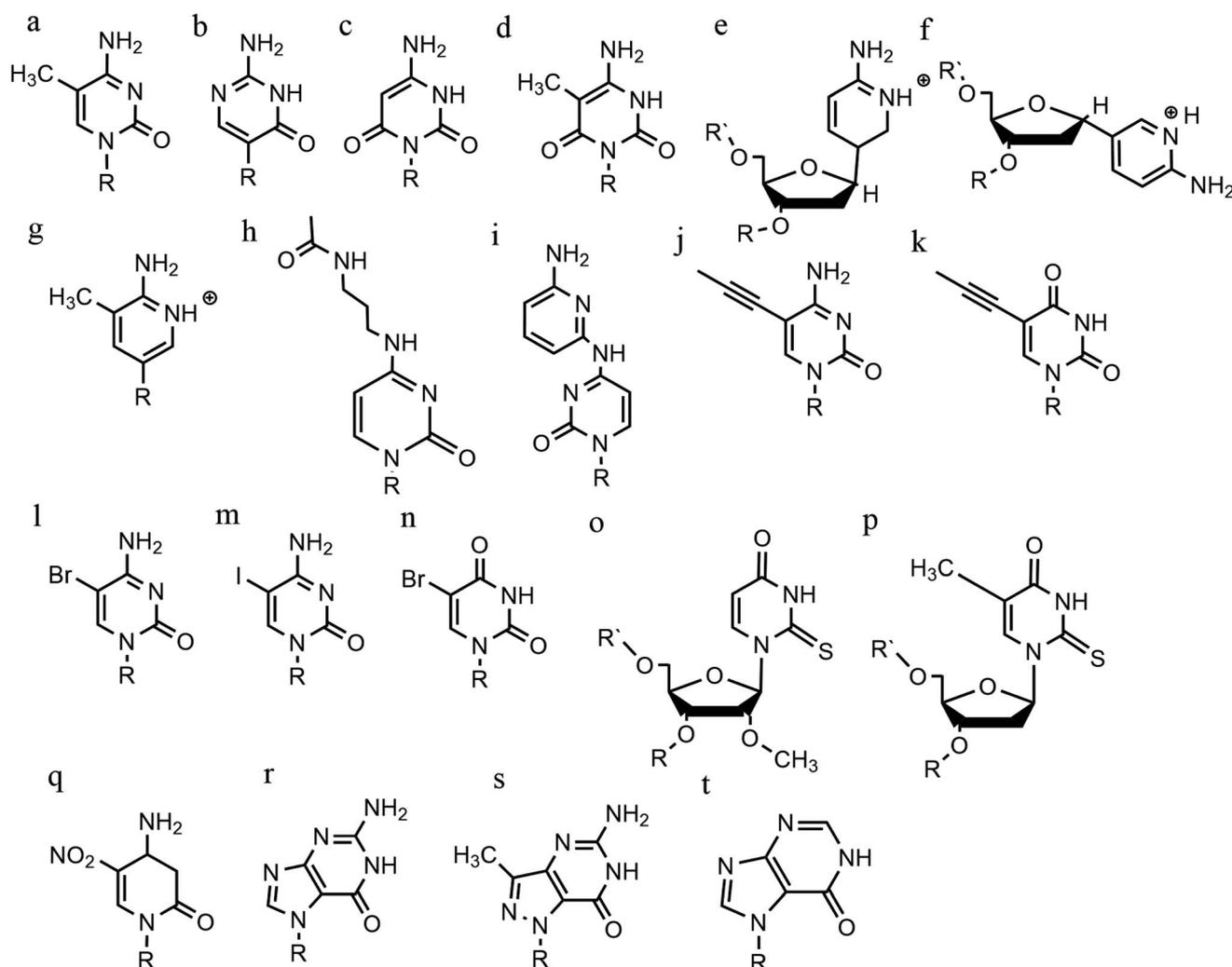


Fig. 9 Base modifications in parallel triplexes. (a) 5-Methyl-cytosine, (b) 2'-O-methyl-pseudoisocytidine, (c) 6-oxo-cytosine, (d) 5-methyl-6-oxo-cytosine, (e)  $\alpha$ -AP, (f)  $\beta$ -AP, (g) 2'-aminoethoxy-thymine, (h)  $N^4$ -3-acetamidopropyl-cytosine, (i)  $N^4$ -6-aminopyridinyl-cytosine, (j) 5-propynyl-cytosine, (k) 5-propynyl-uracil, (l) 5-bromo-cytosine, (m) 5-iodo-cytosine, (n) 5-bromo-uridine, (o) 2'-O-methyl-2-thio-uridine, (p) 2-thio-thymidine, (q) 6-amino-5-nitropyridin-2-one, (r) N7-glycosilated-guanine, (s) P1-guanine, (t) inosine.



mitigate the repulsion.<sup>57</sup> Generally, it is agreed that the formation of intermolecular triplexes with a polypurine sequence requires divalent cations<sup>58</sup> such as  $Mg^{2+}$ , whereas for the intramolecular assembly, sodium ions are sufficient. It has also been demonstrated that the inclusion of  $Mg^{2+}$  can contribute to an increase in stability of reverse Hoogsteen bonds, resulting in an increased thermal stability for intramolecular triplexes.<sup>59</sup>

Several cations can increase triplex stability. For divalent cations, the order of stabilisation is  $Mg^{2+} > Mn^{2+} > Ca^{2+} > Ba^{2+}$ , which can be attributed to the ionic radius of each ion – the smaller the radius is, the greater the alignment between nucleotides and hence the greater the stability of the triplex assembly.<sup>60</sup>

In contrast, monovalent ions, such as the physiological concentration of  $K^+$ , reduce the propensity of a G-rich strand to form a triplex. The presence of molecular crowding conditions (which are often simulated *in vitro* by using high concentrations of polyethylene glycol, such as PEG 200) can also affect the formation of triplexes, and with a G-rich strand in the presence of  $Ca^{2+}$ , the formation of a G-triplex is promoted with endothermic energy.<sup>61</sup> Molecular crowding conditions can also promote triplex formation and change the effect on stability of adding monovalent ions. For example, in the absence of crowding conditions the addition of  $K^+$  has been demonstrated to increase triplex stability as a function of  $K^+$  concentration. However, in crowding conditions the addition of  $K^+$  actually reduces the stability of the triplex assembly.<sup>62</sup>

Using a crowding agent along with ions to simulate the environment in which triplexes might be found, short triplexes tend to stack together and form a highly condensed structure.<sup>63</sup> Since this effect was also observed with duplexes, it has been argued that DNA triplexes may affect the genome structure with modification at a chromosome level.<sup>64</sup>

## 4.2 Base modifications

An increasing number of oligonucleotide analogues have been developed to obtain TFOs with increased stability (both of the resulting triplex and increased resistance to degradation by nucleases) and enable greater selectivity of targeting towards specific structures or DNA sequences.<sup>55,65</sup>

**4.2.1 Base modifications in parallel triplexes.** Parallel triplex stability can be increased when the sequence contains a greater number of  $C^+-G:G$  triads rather than T–A:T steps, but the observed stability is still pH-dependent, with an optimal pH below 6.2.<sup>66</sup> Indeed, the protonation of the cytosine bases will provide a second hydrogen bond between the N-3 of cytosine itself and the N-7 of guanine, favouring a Hoogsteen bond and consequently the triplex formation in mildly acidic conditions.<sup>67</sup>

However, a series of C bases in a tract will result in lower triplex stability, due to the proximity of multiple charges from the protonated bases, which require more acidic conditions to stabilise.<sup>21</sup> This has prompted researchers to focus on cytosine analogues that support pH-independent triplex formation *e.g.*, neutral cytosines with two hydrogen donor groups, or analogues that protonate more easily. To reduce pH-dependency,

modifications to cytosine have been explored, with the aim of increasing triplex stability in a wider pH range. The methylation of cytosine in the TFO, in 5-methyl-cytosine, contributes to the base stacking, increasing stability (Fig. 9a).<sup>55,66</sup> In recent calculations, it was demonstrated that the methylation of C also shifts the pKa from 4.6 (for cytidine) to 4.9.<sup>68</sup> The 2'-O-methyl-pseudoisocytidine (Fig. 9b) promotes the formation of triplexes in a neutral environment in a TFO that will recognize GC-tracts, exemplified using poly(GC). However, this modification is not widely used because the synthesis is highly challenging, even though it is an excellent candidate for increasing the stability of parallel triplexes. The stabilising effect of this pseudo-isocytosine is reported in intramolecular triplexes with a loop composed of only two bases.<sup>69</sup> The incorporation of 6-oxo-cytosine can increase triplex stability to above pH 7, however, when compared with triplexes containing protonated or methylated cytosines in more acidic conditions, the stability decreases (Fig. 9c). The analogue 6-oxo-cytosine can be further modified by the addition of a methyl group in position 5, obtaining 5-methyl-6-oxo-cytosine (Fig. 9d), which can also promote the stability of the DNA triplex. The use of glycerol linkers combined with 6-oxocytosine has been proposed as a modification, which reduces the steric interaction between the 6-carbonyl and the sugar, increasing the stability of the triplex in comparison to the stability observed with no linker present. The absence of glycerol linkers particularly reduces the stability of the triplex if it contains a G-tract.<sup>70</sup> The incorporation of 2-aminopyrimidine (AP) can promote increased triplex stability at physiological pH without protonation due to the low basicity of the modified base. AP can be incorporated in the TFO as  $\beta$  and  $\alpha$ -anomers, the first cytosine anomer has a lower pH dependency due to its pKa of 6.5, resulting in stable triplexes (Fig. 9e and f).<sup>71,72</sup> Unsurprisingly, the addition of a 5-methyl-cytosine in the same TFO containing the  $\beta$ -AP does not form triplexes because of the unfavourable steric interaction. An alternative is a combination of the methylated version of the 2-aminopyrimidine with the 2'-aminoethoxy-thymine (Fig. 9g) reaching a binding affinity at pH 9.0.<sup>73</sup>

In terms of base recognition, modified oligonucleotides play a crucial role in enhancing sequence-specific recognition. In the case of parallel triplexes, a TFO with  $N^4$ -3-acetamidopropyl-cytosine can recognise a GC base pair, by forming a more stable triplex due to the increased chain flexibility (Fig. 9h) and has higher stability than the equivalent TFO containing only cytosine.<sup>74</sup> A similar example reported is  $N^4$ -6-aminopyridinyl-cytosine, which can recognise pyrimidine base interruptions in a polypurine sequence (Fig. 9i).<sup>75</sup> The addition of a propynyl group can increase triplex hydrophobicity and consequently stacking interaction. Another example reported is 5-propynyl-cytosine, which replaced the cytosine, but when the propynyl group was attached to uracil, the TFO with 5-propynyl-uracil is more favourable for the stability of parallel triplex compared to the 5-propynyl-cytosine (Fig. 9j and k).<sup>55,76</sup>

Cytosine analogues containing bromine or iodine atoms at position 5 have also been explored, obtaining 5-bromo-cytosine and 5-iodo-cytosine respectively, but the incorporation of these into a TFO, by replacement of cytosine, actually reduced triplex



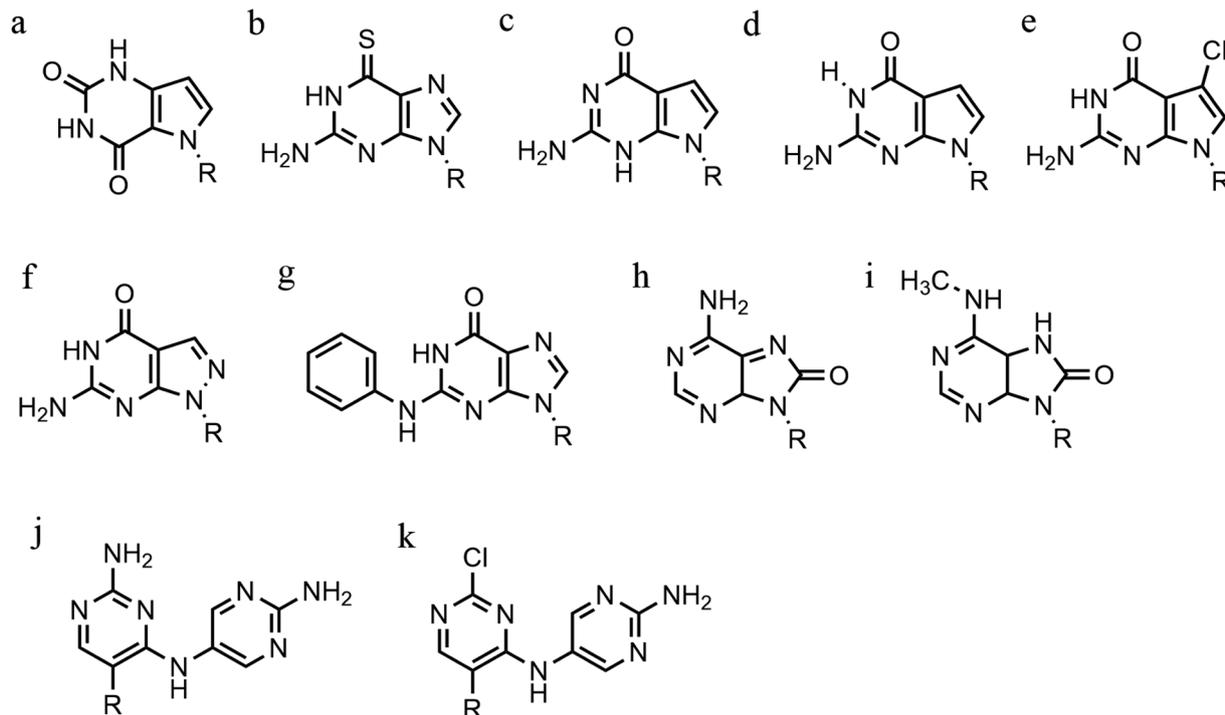


Fig. 10 Base modification for anti-parallel triplexes. (a) 7-deaza-xanthine, (b) 6-thioguanine, (c) 9-deaza-guanine, (d) 7-deaza-guanine, (e) 7-chloro-7-deaza-guanine, (f) 8-aza-7-deaza-guanine, (g) PhdG, (h) 8-oxo-adenine, (i)  $N^6$ -methyl-8-oxo-adenine, (j) AY-d(Y-NH<sub>2</sub>), (k) AY-d(Y-Cl).

stability (Fig. 9l and m). Instead, the substitution of thymine by a 5-bromo-uridine (Fig. 9n) enabled the formation of triplex at room temperature. The inability to obtain triplex structures with 5-halocytosine derivatives is explained by their lower  $pK_a$  and the requirement of protonation.<sup>77</sup> Some studies show that the use of modification on both uracil and thymine in a TFO, such as 2'-O-methyl-2-thio-uridine and 2-thio-thymidine increase the stability of a DNA parallel triplex and the reason is the stacking properties of the 2-thiocarbonyl on the 5' of the upper thiouracil base and the nitrogen atom of the 3' of the lower pyrimidine (Fig. 9o and p). Additionally, it is emphasized that a TFO that includes thiocarbonyl moieties recognizes a base mismatch, a key feature for antibody therapies.<sup>78</sup> A recently published study proposed a TFO containing 6-amino-5-nitropyridin-2-one that overcomes the need for protonation, by acting as an uncharged mimic which can form a parallel triplex, with *in vitro* evidence demonstrating that this approach shows promise (Fig. 9q). The modified nucleobase was included in the TFO through an enzymatic process at physiological pH, relying on the thermodynamic stability of 6-amino-5-nitropyridin-2-one compared to other mismatched bases. Additionally, the modified TFO enhanced protection to the DNA from nucleases.<sup>79</sup> The modification of purine bases has also been explored, although this has received less attention than the pyrimidines. A substitution of N7-glycosylated-guanine or P1-guanine with a cytosine has a remarkable impact on the triplex stability when in the presence of a G-tract (Fig. 9r and s).<sup>46,80,81</sup> If guanine is converted into inosine by removal of the guanine- $N^2$  amino group, then this is able to recognise a GC base pair and form a triplex

structure. Additionally, the absence of the amino function give space to an unusual bond between the carbonyl group of the modified base and the CH of the guanine of the duplex, resulting in a higher electrostatic stability (Fig. 9t).<sup>82</sup>

**4.2.2 Base modification in anti-parallel triplexes.** The principal concern when working with anti-parallel triplexes is the competitive formation of a G-quadruplex structure due to large numbers of guanine residues. It has been reported that the physiological level of  $K^+$  (over 100 mM) will stabilize the formation of quadruplexes rather than triplexes. Therefore, the aim of chemical modification is to produce analogues which will prevent quadruplex formation whilst promoting the formation of a parallel triplex. To stabilise an antiparallel DNA triplex, it was proposed to replace thymine with a 7-deaza-xanthosine (Fig. 10a). The introduction of this modification will reduce the likelihood of the oligo assuming a G-quadruplex structure, since the N7 needed as a hydrogen donor in the modified guanine is absent.<sup>83</sup> The essential role of potassium suggests, 6-thio-guanine should prevent the formation of quadruplexes due to the very weak electron pair donor properties of the S lone pairs to  $K^+$  ions, compared to the carbonyl group (Fig. 10b).<sup>84,85</sup> Other examples of analogues that prevent  $K^+$  coordination are 9-deaza-guanine, 7-deaza-guanine and 7-chloro-7-deaza-guanine, and although they will form triple helical structures, there is no sign of significantly increased triplex stability (Fig. 10c–e).<sup>86–88</sup> Instead, a purine modification in parallel triplexes that can form triplexes in G-rich TFO at physiological [ $K^+$ ] is 8-aza-7-deaza-guanine (PPG) (Fig. 10f).



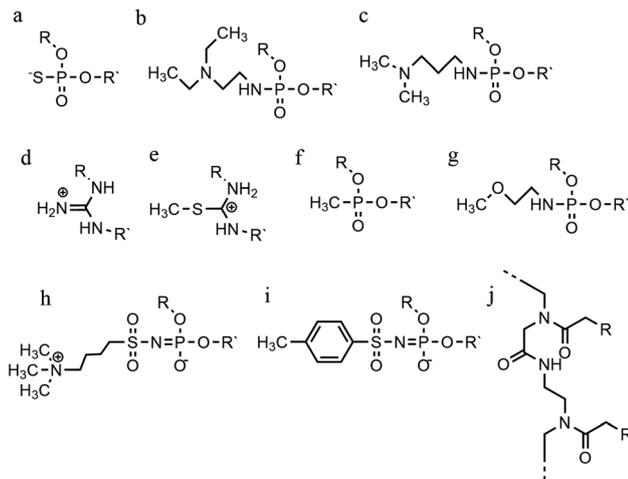


Fig. 11 Phosphate backbone modifications. (a) Phosphorothioates, (b) DEED, (c) DMAP, (d) guanidino, (e) methylthiourea, (f) methyl-phosphonates, (g) PNHME, (h) azido-phosphoramidate, (i) tosyl sulfonyl phosphoramidite, (j) PNA.

Furthermore, modified TFOs containing this modification were used in cells to generate triplex-induced mutations and to cause double-strand breaks (DSBs) that will lead to cell death. These results show the modification forming a stable triplex, preventing G-quartet formation and inducing gene modification, editing and cell apoptosis.<sup>89</sup> Another, more recent, modification included in antiparallel triplex DNA is the product from a synthesis of a guanine derivative *N*<sup>2</sup>-phenyl-2'-deoxyguanosine (PhdG), which was shown to form a stable and selective triplex with the GC base pair (Fig. 10g). As a drawback, as more PhdG bases are introduced, there is an increased likelihood that the structure could assume a higher order.<sup>90</sup> Finally, to support the triplex formation in the presence of high  $[K^+]$ , the protonation of the backbone is often used as an alternative approach,<sup>91</sup> which will be discussed in the next section.

The modification of adenine has also been reported as a potential route to enhancing triplex stability. The purine analogue 8-oxo-adenine forms stable Hoogsteen bonds with a G:C Watson-Crick base pairing (Fig. 10h). Additionally, an *N*<sup>6</sup>-methyl-8-oxo-adenine binds a purine sequence improving the triple helical stability (Fig. 10i).<sup>92</sup> The 8-NH<sub>2</sub> modification of the 8-amino-purine creates a stable interaction either with cytosine or guanine. Therefore, numerous 8-amino-purine derivatives were tested in DNA triplexes, demonstrating that, regardless of structural alterations to the chemical structure, antiparallel triplexes are found to be more stable in physiological pH conditions.<sup>93</sup>

Pyrimidine derivatives have been exploited to stabilize antiparallel triplexes. The incorporation of a cytosine nucleoside containing an amino-pyrimidine unit AY-d(Y-NH<sub>2</sub>) or AY-d(Y-Cl), results in stable triplexes able to recognise the inverted G:C instead of the canonical C:G, or T:A instead A:T with a duplex (Fig. 10j and k).<sup>94</sup>

### 4.3 Phosphate backbone modification

An alternative strategy to promote triplex stability is to focus on modifications to the phosphate backbone of the

oligonucleotide.<sup>95</sup> In general, TFOs are more likely to form a self-associated structure when the backbone is neutral or cationic, due to a decrease of the electrostatic repulsion between the three anionic strands. Modifications have been designed to promote a higher affinity between the TFO and the duplex strands whilst also increasing TFO nuclease resistance, which is important for the longevity of a TFO strand inside a cellular environment.

A significant number of backbone modifications have been explored in the context of DNA triplexes. One of the first to be produced, the phosphorothioate modification (S-oligos), included a substitution to one of the non-bridging oxygen atoms in the phosphate group, replacing the O with S. This modification presents a significant drawback when applied *in vivo*, as TFOs containing this modification tend to bind proteins non-specifically. So, whilst this modification does confer nuclease resistance to the TFO, increasing longevity in the cell, it still maintains the negatively charged backbone, which is thought to reduce triplex stability (Fig. 11a).<sup>96</sup>

The formation of positively charged backbones have been explored, with the incorporation of cationic amine groups into the DNA backbone, including groups such as *N,N*-diethyl-ethylenediamine (DEED) or *N,N*-dimethyl-aminopropylamine (DMAP) (Fig. 11b and c). These modifications increase the binding affinity of the TFO *in vitro* and make them increasingly nuclease-resistant.<sup>96</sup> Increasingly complex changes in the backbone modification have been also proposed. Guanidino and methylthiourea are some examples of a complete substitution of the phosphate group with a cationic linked nucleoside, resulting in more stable triplex oligomers even though an increasing proportion of T and A in parallel triplexes decreases the melting temperature (Fig. 11d and e).<sup>97,98</sup>

Options to obtain non-ionic alternatives are available as well, such as methyl-phosphonates, phosphotriesters and non-phosphate hydrazide derivatives (Fig. 11f). However, the lack of charge makes them highly insoluble, and they are therefore less suited to *in vivo* applications.

Different phosphoramidate-linkage modified TFOs have been proposed to bind the dsDNA efficiently by enhancing DNA stability.<sup>99</sup> An example of a phosphoramidate-modified oligonucleotide is methoxyethylphosphoramidate (PNHME) for pyrimidine with the  $\alpha$ -anomeric configuration (Fig. 11g).<sup>100</sup> Whilst backbone repulsion is decreased, the triplex is only formed at pH 7 or lower and therefore this process is still protonation dependent.

Recently, an increasing number of studies have incorporated zwitterionic modifications in oligonucleotides, yielding thermostable triplexes. For example, an azidosulfonyl ammonium

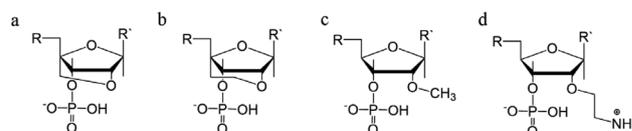


Fig. 12 Sugar backbone modifications. (a) LNA, (b) ENA, (c) 2'-OMe, (d) 2'-AE.



salt can be used, instead of  $\text{TsN}_3$ , to give a zwitterionic derivative (see Fig. 11h). As a consequence, the duplex formation is less dependent on ionic strength. While this change increased the hydrophobicity of the molecule compared to the unmodified DNA, stable parallel triplexes form at a pH optimum of 5, and only when the modification is at the 3' end. Furthermore, the presence of a tosylsulfonyl phosphoramidite (Ts) can be exploited as a negatively charged phosphate (Fig. 11i). Both modifications, when introduced into the same oligonucleotide, form stable parallel DNA triplexes and show promise for *in vivo* applications, especially as nuclease resistance and cellular uptake were increased compared to non-modified oligonucleotides.<sup>101</sup>

A more drastic modification of the phosphodeoxyribose backbone features the use of peptide nucleic acids, PNA (Fig. 11j). A PNA strand was conceived as a triplex-forming oligonucleotide, able to bind to a dsDNA due to its neutral charge. Subsequently, it appeared that two PNA strands, where the phosphate backbone is substituted by units of *N*-(2-aminoethyl) glycine, form remarkably stable triplexes when binding the unmodified TFO. The high stability of the triple-helical structure arises primarily from the neutral charge, drastically reducing backbone repulsion. Molecular dynamics simulations confirm that PNA backbones provide additional flexibility to the triplex and in some cases can assume A-type

conformations.<sup>102,103</sup> An alteration of PNA was proposed with an arginine instead of glycine, forming the G-PNA. This modification has overcome the solubility issue of PNA.<sup>104</sup> Two other modifications reported are olefinic peptide nucleic acids (OPA) and oxy-PNA. These alternatives seek to improve the cellular uptake rather than the triplex stability itself.<sup>105,106</sup> Alternatively, PNA can contain ligands with coordinated metal ions instead of nucleobases. The outcome is a stable triplex in solution experiments, due to the strength of coordinative bonds compared to hydrogen bonds, but which is reduced by the steric interactions of the metallo-complex and the triplexes.<sup>107</sup> The use of PNA has been explored in a number of different areas including cellular uptake,<sup>108</sup> regulation of gene expression,<sup>109</sup> interruption of the RNA polymerase and inhibition of translation and activation of DNA repair system.<sup>95</sup> It shows great potential as a future therapeutic, and work in this area is ongoing to address some of the challenges associated with its use, such as cellular delivery.

#### 4.4 Sugar modification

Sugar modifications focus on the sugar pucker conformations that will influence the ability of the TFO to form a stable structure. The most common approach used to increase the stability of the triplex restricts the range of sugar conformations, relying on the use of bridged nucleic acids (BNA).<sup>110</sup> The

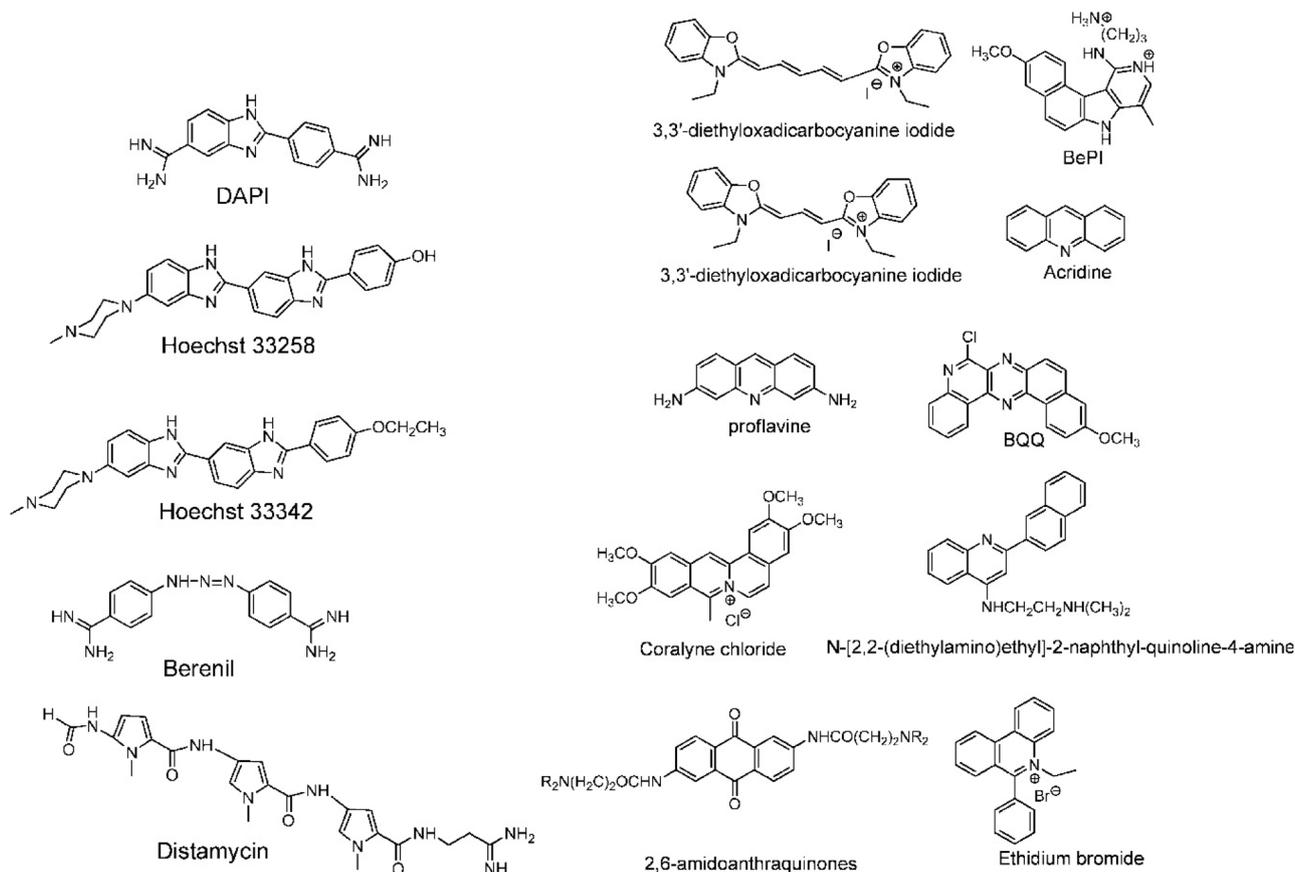


Fig. 13 (Left) DNA triplex groove binders and (right) DNA triplex intercalators. Adapted with permission from D. P. Arya, *Acc. Chem. Res.*, 2011, **44**, 134–146. Copyright 2011 American Chemical Society.



puckering characteristics of the sugar ring allows the ribofuranose structure to assume a range of conformations but, once it was realised that the C3'-endo configuration is more likely to stabilise a triplex, a range of modifications were explored, with the aim of promoting this conformation.<sup>48,111</sup> The first generation of BNA is locked nucleic acid (LNA), which consists of a 2'-O, 4'-C methylene bridge that restricts the sugar backbone movement (Fig. 12a) and promotes the formation of an A-form duplex in the binding partner of the LNA strand. This reduced flexibility enhances the stability and selectivity of the TFO strand. It has been reported that including short LNA residues in pyrimidine-motif triplexes will enhance stability due to the significant puckering amplitude.<sup>48</sup> However, the modification can only be included once in every 2–3 nucleotides; a TFO composed of only LNA modifications does not form triple helices.

A second modification, with an ethylene link (ethylene-bridged nucleic acid, ENA, Fig. 12b) instead of methylene, was proposed to overcome this incorporation limit and allows for the production of fully modified TFOs able to form a triplex.<sup>48,70</sup> This modification is less restrictive, and therefore allows for a greater variation in the observed LNA sugar pucker, giving more flexibility to accommodate the third strand, which can be composed fully of ENA.<sup>112</sup>

Alternative strategies to modify the sugar component of the TFO without imposing a locked conformation are the addition of an ammonium group to the sugar, 2'-O-methylribose (2'-OMe) (Fig. 12c)<sup>113</sup>, or a protonated aminoethyl group at C3'-endo, 2'-O-aminoethylribose (2'-AE) (Fig. 12d). Both modifications bias the sugar pucker towards C3'-endo, favouring the A-form conformation, improving the stability of the TFO towards nucleases and enhancing triplex stability.<sup>114</sup>

Continued development of nucleotide analogues, modified phosphate backbone and sugar-based variants is ongoing. When evaluating a modified TFO, an ideal candidate forms DNA triplexes with a high association rate and remain thermostable, both *in vitro* and *in vivo*. Thus far, modifications have typically been investigated singularly *i.e.*, candidate TFO strands have contained one modification, although this can be at multiple sites within a single strand. Future development should therefore focus on combining modifications to provide a successful outcome in terms of triplex stability and biological function. Indeed, for cellular applications, it must be taken into

consideration that the TFO or DNA triplex must initially be delivered into the cell and therefore the hydrophobicity properties must be considered and carefully balanced. The main challenges, however, are to stabilize the triplex at physiological pH, maximise nuclease resistance and finally promote specificity in sequence targeting.

#### 4.5 DNA triplexes intercalators and groove binders

A completely different approach to DNA triplex enhancement that does not require chemical modification or solutes is the noncovalent intercalation of a small molecule stabiliser. The latter are molecules, widely studied over the years, which are able to specifically bind DNA triplexes, since they can provide tools to enhance triplex stability and support biological applications. As we have seen, triplex structures are less stable than the duplexes. Specifically, the need for cytosine protonation in the pyrimidine third strand leads to limited triplex stability at physiological pH. For these reasons, intercalation by molecules able to selectively stabilize the triplex structure is of great interest.<sup>96</sup>

For example, the common duplex DNA binder ethidium bromide (EtBr, Fig. 13) can also stabilize a C-G:C structure with a triplex-specific stabilizing effect, due to the electrostatic repulsion between ethidium and cytosine. However, the stabilization of the triplex with ligands will also depend on the concentration of the chosen ligand. It has been reported that two molecules of either EtBr or acridine orange (AO, Fig. 13) in 10-base pair long triplex will stabilise the structure, while a third molecule leads to destabilisation, highlighting that the effect of concentration must be carefully balanced.<sup>115</sup> Also, the increase of stability, measured as the increase in the triplex melting temperatures, depends on the DNA sequence. The melting temperatures of the 15-mer triplexes were obtained from the hyperchromicity observed at 260 nm upon thermal denaturation. A larger increase in melting temperatures for sequences having A-tract duplex structures was observed by UV spectroscopy, using a ratio of 2 : 1 pyrimidine to purine strand. This large thermal stabilizing effect on dTn·dAn·dTn triplexes is partly due to the intercalators that break up the intrinsic A-tract structure of the underlying duplex.<sup>116</sup> In fact, the intrinsically rigid and highly propeller-twisted structure of A-tract DNA disfavors triplex formation.<sup>117</sup> Propidium iodide (PI, Fig. 13) has been reported as a potent stabiliser of the parallel triple helix, with association constant similar to that of PI binding to duplex DNA.<sup>118</sup> PI was shown to increase the parallel triplex stability after intercalation of three molecules into the triplex, with melting temperature increasing from 21.4 up to 44.4 °C in different media such as Na phosphate buffer, pH 7 and NaCl.<sup>119</sup>

Other DNA triplex binding intercalators include indolo-carbazole and benzopyridoquinoline derivatives. These provide additional stacking interactions with the pyrimidine strand of the Watson-Crick double helix, resulting in a very efficient and specific stabilizing effect on triple helices and/or in inducing triple helix formation under physiological conditions.<sup>99</sup>

Another class of intercalators able to stabilize the triple helices is the twisted intercalating nucleic acids (TINA) (Fig. 14). These nucleic acids are characterised by the ability to twist

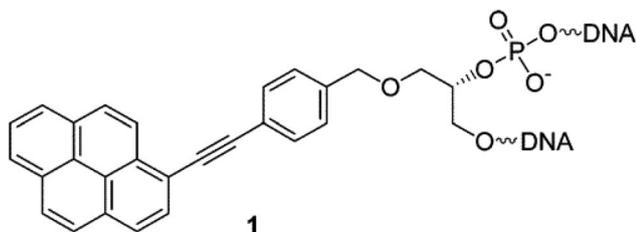


Fig. 14 Example of structure of a TINA intercalating unit.<sup>181</sup> Reprinted with permission from I. Géci, V. V. Filichev and E. B. Pedersen, *Bioconjug. Chem.*, 2006, 17, 950–957. Copyright 2006 American Chemical Society.



around a triple bond. This twisting promotes intercalation within double stranded DNA in order to form triplex DNA. Moreover, it has been demonstrated that these oligonucleotides can discriminate between matched and mismatched sequences of DNA.<sup>120–122</sup>

In this context, it is worth noting that intercalators usually have a stabilising effect on DNA triplexes, whereas minor groove binders will generally destabilise the assembly (Fig. 13). Nevertheless, some aminoglycosides were tested as triplex binders and it was shown that neomycin selectively recognises the triplex Watson–Hoogsteen groove and stabilises it without any effect on dsDNA. This very interesting selectivity may be related to the shape complementarity to the triplex Watson–Hoogsteen groove (the groove formed between the TFO and DNA strand which does not bind to the TFO).<sup>123</sup> Other minor groove binders that are well exploited are netropsin, spermine and cyclopolyamines.<sup>124</sup> Psoralen has also been used as it can intercalate efficiently between bases and can provide a covalent linkage by forming an adduct on photoreaction with the stacked pyrimidine.<sup>125–127</sup>

Other reported groove binders are Hoechst 33258, Berenil, DAPI and distamycin A (Fig. 13), however, their stabiliser ability as well as the triplex stability is lower than with neomycin. In this area almost no structural characterisation of triplex–ligand systems has taken place and therefore this is an area which could be the subject of future focus to understand the DNA triplex–ligand molecular interaction.<sup>99,124</sup>

## 5. Applications based on biomolecular approaches

The ability to form a three-stranded complex based on base–base recognition can be exploited to develop biotechnologies suited for diagnosis, prognosis, or disease treatment. Indeed, a modified TFO included in a dsDNA is considered as a potential future for genetic medicine, exploiting sequence-specificity to target genes for manipulation. TFOs have proven to be useful tools, able to alter gene expression and cause genome modification in mammalian cells.<sup>128</sup> However, several limitations must be overcome to improve their therapeutic value. Often, these applications are restricted because of the low-affinity binding *in vivo* conditions, as well as TFO stability and integrity during cellular uptake. Numerous attempts have been made to modify oligonucleotides and improve these characteristics, as discussed earlier in this article.<sup>96</sup>

The ability of a TFO to inhibit a transcription was demonstrated for the first time with the human *c-myc* protooncogene in HeLa cells. This protooncogene plays a crucial role in normal cell proliferation and programmed cell death. In particular, *c-myc* gene expression is present in cancer cells at an increased level compared to normal cells.<sup>129</sup> Specifically, after entering the nucleus, TFOs bind to the DNA duplex at the target sequence to form the triple helix, which prevents the polymerase and other transcription factors from initiating transcription. This results in the inhibition of mRNA synthesis from the *c-myc* promoter, demonstrating that the administration of the TFO to the cells can influence the transcription of the *c-myc* gene.<sup>130</sup>

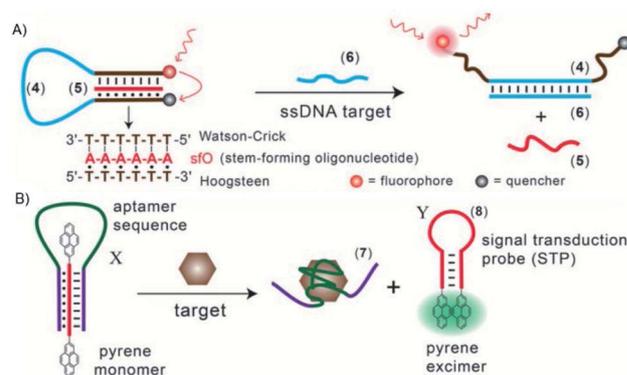


Fig. 15 (A) Optical sensor based on hairpin triplex structure (4) of a target gene (6) by the reconfiguration of a fluorophore/quencher-modified triplex DNA hairpin structure and the release of the stem forming oligonucleotide (5). (B) A triplex DNA hairpin moiety (X) containing an aptamer sequence used as an optical aptasensor that binds the target (7) with subsequent formation of a hairpin excited structure (8). Reprinted and adapted from *Triplex DNA Nanostructures: From Basic Properties to Applications* Y. Hu, A. Cecconello, A. Idili, F. Ricci, and I. Willner, pages 15210–15233, Copyright (2017), *Angew. Chem.*

A therapeutic application that was proposed relied on the ability of the TFOs to bind a duplex structure related to the Friedreich's ataxia gene. The formation of the triplex structure stalls the RNA polymerase and decreases the frataxin protein level, which causes the disease. The GAA triplet repeat, which is responsible for the neurodegenerative disease, folds back, forming a triplex structure with the polypurine strand. In this case, disfavouring the formation of the triplex structure could be the key to restore the FXN gene transcription, and therefore regenerate the normal frataxin protein level.<sup>131,132</sup>

Since the TFO should form a DNA triplex along a gene of interest, it is useful to direct a site-specific mutation. Indeed, a psoralen-modified TFO directed to the supF reporter gene, along with UV irradiation in order to allow the cross-linking of the psoralen to the DNA, resulted in a 100-fold increase of mutations, in which 70% are TA to AT transversions. In mammalian cells, chromosomal mutations have been enhanced tenfold after targeting specific genes. Moreover, triplex formation creates a helical distortion to trigger DNA repair by different pathways, *i.e.* involving the nucleotide excision repair (NER) system or homologous recombination (HR).<sup>96</sup>

In addition to induced mutagenesis, another role of the DNA triplex is genome modification based on the recombination strategy. Triplex technology was used to determine whether interstrand cross-links (ICL) could be repaired through homologous recombination (HR). Indeed, a green fluorescent protein reporter forms a triplex with the psoralen-TFO and intercalates through the specific ICL sequence by confirming the HR effect.<sup>133</sup> Moreover, targeting a specific gene sequence could be used for deleting or replacing sequences on chromosomes. Therefore, a DNA break that happens during the formation of the triplex, stimulates the recombination. To support this notion, a simian virus 40 (SV40) shuttle vector was modified to present psoralen-TFO, then inoculated in human



cells, resulting in DNA damage. As consequence, a mutation is induced in a NER/XPA dependent manner.<sup>134,135</sup> A result obtained with luciferase reporter assays shows that p53 was transactivated when a triplex-forming sequence, introduced *via* plasmid, was formed close to the p53 target sequence.<sup>136</sup>

As reported above in this review, one of the major problems related to TFO application *in vivo* is the instability of the triplex at neutral pH, due to the requirement of cytosine protonation to form the triplex, which is not possible at physiological pH. Different strategies have been studied, such as walled nanotubes (SWNT), to stabilize C-G:C triplexes under physiological conditions. Such studies may facilitate the application of nanomaterials in the artificial control of gene expression and biosensing.<sup>137</sup> Another interesting and very recent approach proposes to modify the TFOs with the nucleobase 6-amino-5-nitropyridin-2-one (*Z*), which acts as uncharged replacement for the protonated cytosine. By using this method, Rusling obtained stable and selective triplex formation stable at neutral or even slightly basic pH.<sup>79</sup>

Triplex DNA structures were also used as structure-switching units to trigger a signal, following the recognition of specific targets such as proteins, antibodies, small molecules and pH.<sup>138</sup> For example, fluorophore/quencher pair molecular beacons are exploited as optical switches to detect pathogens and genetic disorders. These tools can be used with triplex structures. Indeed, a hairpin triplex helix, functionalised with a fluorophore in one edge, and the quencher in the other edge, is reconfigured in an open structure after recognition of the target. The target recognition leads to the opening of the triplex structure and to an increase in fluorescence, due to the spatial separation of the fluorophore and the quencher that were adjacent when the hairpin triplex structure was formed. This idea was applied in the design of a bimolecular triplex helix stem for the analysis of a DNA single strand. The stem containing a T-A·T triplex incorporating a poly-T DNA and a poly-A peptide nucleic acid (PNA) strand was used to increase the stability of a molecular beacon. In this case, after recognition of the target, *i.e.* the single strand of DNA, the formation of the DNA duplex leads to the opening of the triplex structure with an increase in the fluorescence signal (Fig. 15A).<sup>139</sup>

Triplex-based hairpins have also been exploited with a luminescent pair to obtain a sensing platform. This system is characterised by the presence of a pyrene excimer pair attached to the two edges of a linear triplex forming oligonucleotide. Once the hairpin portion of the triplex recognises the analyte, the hairpin is opened and folds around the target molecule. The stem with the pyrenes is thus released, and able to fold into another hairpin structure, causing the contact of the luminescent pair (Fig. 15B). This results in the emission of the pyrene excimer at 485 nm. The emission level is then proportional to the concentration of the target species. This sensing platform has been used for the detection of thrombin, ATP or L-arginamide. All these methods exploit the presence of an anti-thrombin/anti-ATP/anti-L-arginamide aptamer sequence in the triplex-based hairpin. Indeed, many sensors can be designed, but their efficacy depends on reliable opening of the hairpin triplex structure after recognition

of the analyte. This could be affected by low sensitivity, so strategies to stabilise the target-recognition sequence are required.<sup>140</sup>

Triplexes can be exploited to detect a specific duplex sequence. The duplex assembly is recognised by a suitable TFO sequence folded into a hairpin loop and containing a fluorophore/quencher pair in proximity to each other. In the presence of the duplex target sequence, the fluorophore and the quencher are separated by the opening of the hairpin structure, leading to an increased fluorescence of the system. This fluorescence increase depends on the concentration of the duplex analyte. This method was applied to detect cancer cells and also non-DNA targets, like the NF- $\kappa$ B p50 transcription factor.<sup>141,142</sup>

Beside the application of triplexes as molecular beacons, triplexes have also been applied as functional units for electrochemical sensors. Electrodes have been functionalised with programmed, redox-labeled DNA structures to obtain a probe attached to the electrode surface. The concept is based on the fact that, when the analyte is present, the binding in between the triplex and the target sequence leads to the formation of a duplex structure. This complex displaces the redox label from the electrode surface, suppressing the electrochemical signal produced by the probe itself. In this way, a quantitative determination of the analyte (*i.e.* DNA, proteins, small molecules, metal ions) is obtained by controlling the voltammetric response.<sup>143</sup> This method has been applied for the analysis of sequence-specific double strands, adenosine, transcription factors and to detect HIV-1 strains.<sup>144,145</sup>

Similarly, triplexes have been used also as pH probes, exploiting the ability of the oligonucleotides to change the duplex/triplex ratio depending on pH. At around pH 5.0, we have already seen that cytosine bases are protonated, permitting the formation of a parallel triplex structure. This concept has been applied in the development of a construct formed by a long strand with two arms capable of bridging a fluorophore/quencher-functionalised strand *via* the formation of the C-G duplex. In neutral conditions, the fluorophore and the quencher are separated in the medium used. In acidic conditions, the protonation of the cytosines promotes the formation of a triplex structure, causing the proximity of the fluorophore/quencher pair and leading to the decrease of fluorescence intensity.<sup>146</sup>

Another application of the pH dependence of the duplex/triplex structure is the control of aggregation/disaggregation of nanostructures driven by the equilibrium between triplex formation and dissociation. In one example, this equilibrium was used to switch the aggregation/disaggregation of gold nanoparticles (NPs), in a reversible process. The nanoparticles were functionalised with nucleic acids that were partially self-complementary. In neutral conditions, the NPs are separated while in acidic conditions (pH 5.0) the formation of a triplex C<sup>+</sup>-G-C structure leads to NP aggregation. When the system is neutralised, the triplex structures were dissociated and the nanoparticles disaggregated.<sup>147</sup>

In the biomedical field, the trigger release of loads is an important objective that has aroused interest. Stimuli-



responsive microcapsules loaded with a substrate and stabilised by DNA shells have been used to specifically release a cargo. Elegantly, the microcapsules are released after enzymatic digestion of the DNA shells. In this context, triplexes have been attached to the microcapsules and used as pH-responsive carriers. For example, QD-loaded  $\text{CaCO}_3$  micro-particles, coated with poly(allylamine hydrochloride) (PAH) polyelectrolyte and functionalised with nucleic acid composites containing the caged triplex sequences, were used. The DNA-stabilised  $\text{CaCO}_3$  core was dissolved by adding EDTA. At pH 5.0, the triplex structure is formed, with a subsequent separation of the microcapsules and the release of the QD loads.<sup>148</sup>

Overall, all these findings represent very intriguing and promising steps in the application of TFOs in the biomedical field.

## 6. DNA triplex and related interactions with metal complexes

Transition metal complexes have been investigated in the last decades for a large range of healthcare applications, including diagnosis and treatment of various diseases. Several characteristics are appealing for study with nucleic acids, such as the positive charge, the ability to coordinate directly to Lewis base sites on DNA, the possibility to undergo redox reactions with DNA and to generate reactive oxygen species – an attribute particularly relevant for photodynamic therapy (PDT) – make

these systems exceptionally attractive for the development of new therapeutics.<sup>149</sup>

Since the serendipitous discovery of cisplatin and its ability to covalently bind duplex DNA,<sup>150</sup> many metal complexes have been studied to obtain compounds with less side effects than cisplatin and an improved and more selective toxicity towards cancer cells. In parallel, other approaches to the use of metal complexes for targeting DNA in different ways have been developed.<sup>151–153</sup> Very interestingly, metal compounds can also be exploited with non-canonical DNA structures, to stabilise these structures and/or to functionalise them for a specific application, as presented in this section.

Early attempts were made to introduce Ag(I)-based complexes as artificial nucleosides to stabilise DNA triplexes through metal complexation. The incorporated Ag(I) complex significantly stabilised the DNA duplex and triplex by introducing a pair of pyridine nucleobases in the middle of the sequence. The nitrogen of the pyridyl complex coordinates with Ag(I) at the centre of the triplex, stabilising the triplex structure.<sup>154</sup> Although it is not an independent molecule that intercalates in the DNA triplex, it is noteworthy that  $\text{OsO}_4^-$  bipyridine stabilises the triplexes by protecting the thymine from being disrupted. In the study, it was observed that intercalation caused a thymine base to flip out of the DNA helix. When the complex was added, the thymine was protected from this disruption.<sup>155,156</sup>

The ability to specifically recognise a non-conventional DNA structure is a very powerful tool to increase specificity in targeting biomolecular sites. For example, tetracationic supramolecular helicates such as  $[\text{Fe}_2\text{L}_3]^{4+}$ , formed from  $\text{Fe}^{2+}$  ions wrapped by three bis-pyridylimine organic strands, were used in a new approach for synthetic DNA recognition. Intriguingly, one of the compounds ( $\text{L} = \text{C}_{25}\text{H}_{20}\text{N}_4$ ) recognised a three-way junction in duplex DNA, giving a unique hydrophobic binding site characterised by a triangular shape. The structure was determined by X-ray crystallography. This result gave information on the existence of DNA binding modes of metal-based drugs that differ from the most common ones (*i.e.*, covalent bond, intercalation, major groove binding, minor groove binding and sugar-phosphate backbone binding).<sup>157</sup>

Bulges are sites of DNA where one or more nucleotides are not paired within the double helix. These unpaired nucleotides arise after replication and recombination errors or after carcinogen-induced DNA damage. They are believed to play an important role in various diseases such as cancer, Alzheimer and muscular dystrophy. Thus, DNA sequences containing a bulge are an important target for developing potential therapeutic drugs. Also, small molecules able to target DNA bulges are particularly interesting for their use as potential therapies. The interaction of the above-cited compound  $[\text{Fe}_2(\text{C}_{25}\text{H}_{20}\text{N}_4)_3]^{4+}$  with bulged DNA was studied by DNA melting temperature and gel electrophoresis assays to evaluate the binding affinity of this helicate for various DNA bulges. Both enantiomers of the compound bind to bulges containing two or more unpaired nucleotides. Moreover, this compound had higher binding affinity for bulges containing unpaired pyrimidines and/or flanking pyrimidines. It is suggested that the bulge allows the

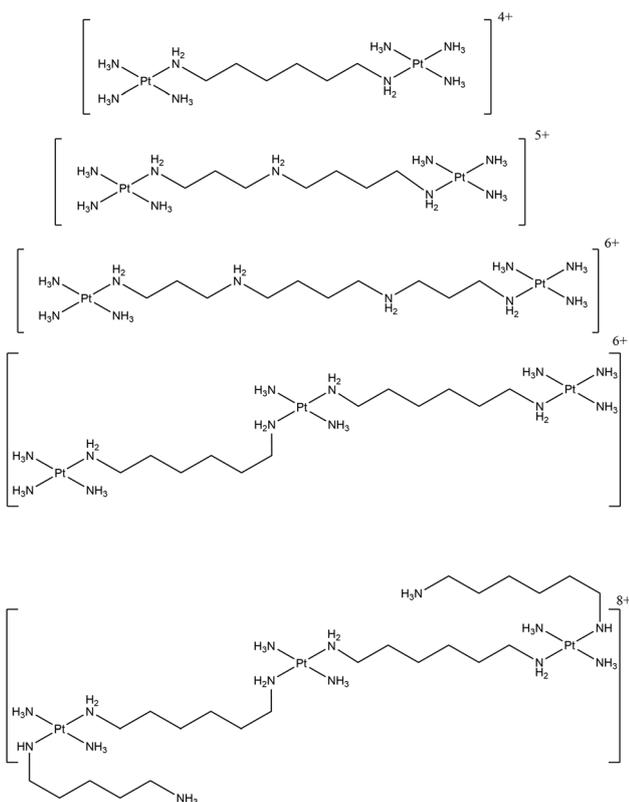


Fig. 16 Chemical structures of the various SI-PPCs.



triangular prismatic motif necessary to accommodate the helicate. This is an example of another uncommon DNA structure that is specifically recognised by  $[\text{Fe}_2\text{L}_3]^{4+}$  supramolecular helicates.<sup>158</sup> Brabec and co-workers described a class of dinuclear  $\text{Fe}^{\text{II}}$  triplex-forming metallohelices able to specifically recognize and stabilise DNA bulges of different size and composition. The compounds preferably bind the DNA bulges instead of double-strand DNA. Their binding affinity showed to be dependent on the individual metallohelices, the bulge size and the bases present in the bulge loop. In particular, pyrimidine-containing bulges are preferred compared to the purine-containing ones. These compounds were shown to have the ability to stabilise the bulge containing sequences. In fact, an increased thermal stability was obtained with DNA bulges containing three or more unpaired adenines or two unpaired thymines, indicating a stabilising effect.<sup>159</sup>

A range of antitumour substitution-inert polynuclear platinum complexes (SI-PPCs) have been studied as small molecules able to recognise, bind and stabilise the triplex structure of DNA and RNA (Fig. 16).

This class of compounds bind DNA through noncovalent interactions, in particular by “phosphate clamp”, a mode of DNA–ligand recognition different from the intercalative or minor groove binding. They had the ability to inhibit DNA synthesis by DNA polymerase when the DNA sequences used are prone to form pyrimidine- and purine-motif triplex DNAs. It was suggested that these compounds act as very effective stabilisers of triplex DNA and that they can play a stabilising role in triple-helical DNA. The results from a *Taq* DNA polymerase assay showed that the pyrimidine-rich template used for the experiment does not permit the primer extension when the SI-PPCs compounds are present. This indicates that the compounds stabilise or form a DNA topology that impedes DNA polymerisation. Interestingly, the formation of the DNA triple helix is not

stopped in the absence of the compounds and a displacement of TO (which intercalates with high-affinity in triplex structures) takes place when the SI-PPCs are present. This indicates the ability of the Pt-derivatives to form a complex with triple-helical DNA. It was suggested that the ability to stabilise the triplex structure plays a crucial role in the cytotoxicity of this class of compounds. This is particularly important since nucleotide sequences able to form a triplex structure are present in natural DNA, preferentially near regulatory regions.<sup>160</sup>

Moreover, the ability of these class of compounds to inhibit the reverse transcription in RNA template prone to form a triplex structure was described. In particular, the ability of a class of SI-PPCs to inhibit DNA synthesis by reverse transcriptase was evaluated. A purine-rich primer and a pyrimidine-rich RNA template able (TFT) or non-able (SST) to form triplex structures were annealed together and the reverse transcriptase activity was checked by several biophysical techniques. UV melting studies were used to prove that the TFT annealed with the primer formed a triplex structure, showing a biphasic transition in the melting curve, characteristic of a triplex structure. Moreover, the primer extension was allowed in a reverse transcriptase assay without of SI-PPCs, proving that the triple helix formation does not hamper the reverse transcriptase to extend the primer. On the contrary, in the presence of SI-PPCs, the reverse transcriptase ability to extend the primer annealed with the RNA templates was reduced. This inhibition, related to the presence of the platinum compounds, depends on the charge of the compounds and on their size. Moreover, the inhibiting activity in TFT was higher than in SST, suggesting that SI-PPCs can preferentially recognise, stabilise and inhibit the reverse transcription in RNA template prone to triplex formation rather than in SST. Overall, the ability to bind nucleic acids and inhibit protein–RNA triplex interaction is a very promising extension of the biological activity of this class of compounds.<sup>161</sup>

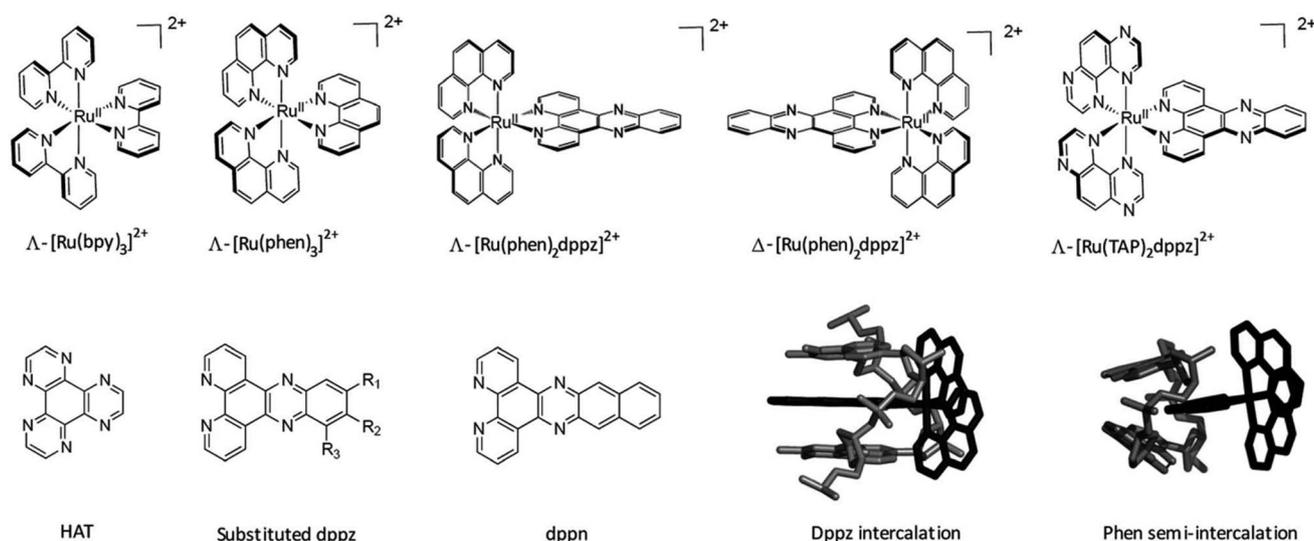


Fig. 17 Important ruthenium complexes and binding modes<sup>174</sup> Reproduced from Cardin C. J., Kelly J. M. & Quinn S. J. Photochemically active DNA-intercalating ruthenium and related complexes—insights by combining crystallography and transient spectroscopy. *Chem. Sci.* **8**, 4705–4723 (2017).



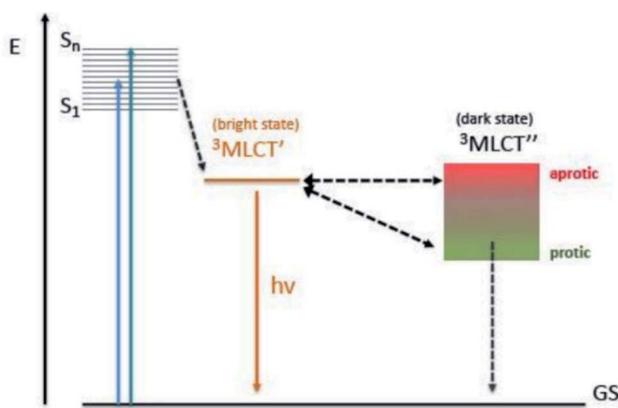


Fig. 18 Jablonski diagram indicating the electronic transition from the excited to the ground state, depending on the solvent.<sup>182</sup> Reproduced from Di Pietro M. L., La Ganga G., La Nastasi F. & Puntoriero F. Ru(II)-dppz derivatives and their interactions with DNA: thirty years and counting. *Appl. Sci.* **11**, (2021).

A wide range of octahedral ruthenium(II) complexes have been investigated for potential biomedical uses, making use of the slow rate of ligand exchange for this electron configuration, multiple and accessible oxidation states, positive charge, and ability to mimic iron in the physiological environment.<sup>162</sup> Ruthenium complexes have been associated with reduced side effects in clinical trials when compared to drugs containing other metals, such as platinum.<sup>163</sup> Ru(II) polypyridyl complexes are notable for their favourable photophysical and photochemical properties, such as visible light absorption (lower energy than 400 nm) due to metal-to-ligand charge transfer (MLCT),<sup>164</sup> and particularly important for the application of such compounds in PDT. This medical technique is based on the use of an ideally non-toxic molecule, called photosensitizer (PS), which is activated by light to produce singlet oxygen with a lifetime in metabolically healthy cells of  $\sim 3 \mu\text{s}$  just at the site of irradiation, obtaining therefore a high spatial and temporal selective treatment.<sup>165,166</sup> Indeed, by varying the ligand set, Ru-based complexes can be tailored not only to obtain desired photophysical and photochemical properties in the PDT application window, but also to improve their DNA binding.<sup>167</sup> Different Ru-polypyridyl compounds have been studied for their ability to intercalate in the DNA by  $\pi$ - $\pi$  interaction between the aromatic ligands and DNA  $\pi$ -stack (Fig. 17). In the next section, we will present some examples, from the very large range already known, of Ru complexes with interesting photophysical and photochemical properties for the application in DNA binding studies, with special attention to actual or potential triplex DNA binding.<sup>168</sup>

A series of Ru(II) complexes with the 1,12-diazaperylene (DAP) ligand of the type  $[\text{Ru}(\text{bpy})_2(\text{DAP})]^{2+}$ ,  $[\text{Ru}(\text{bpy})(\text{DAP})_2]^{2+}$ ,  $[\text{Ru}(\text{bpy})(\text{DAP})_3]^{2+}$  (bpy = 2,2'-bipyridine) was shown to intercalate into calf thymus DNA. The DAP ligand is characterized by an extended  $\pi$ -system and a large surface area to improve the DNA intercalation. Because of the lack of water solubility of the bis- and tris-DAP species, thermal denaturation experiments

were performed only with the more water-soluble compound  $[\text{Ru}(\text{bpy})_2(\text{DAP})]^{2+}$ , showing that this compound can stabilise calf thymus DNA with an efficiency comparable to that of ethidium bromide. Moreover, photocleavage of pUC18 supercoiled plasmid was observed in the presence of  $[\text{Ru}(\text{bpy})_2\text{DAP}]^{2+}$  after irradiation with  $\lambda > 395 \text{ nm}$  for 30 min. The absence of photocleavage in a deoxygenated water environment demonstrated that the  $^1\text{O}_2$  species is involved in the photoreactivity with DNA.<sup>167</sup> As long ago as 1990 the “light-switch” effect of the compound  $[\text{Ru}(\text{bpy})_2(\text{dppz})]^{2+}$  (dppz = dipyrido[3,2-*a*:2',3'-*c*]phenazine) was demonstrated by Barton and co-workers, describing this compound as a highly sensitive spectroscopic reporter of double helical DNA. They demonstrated that this compound displays luminescence only when intercalated into the duplex structure *via* the planar aromatic ligand dppz. It was shown that after intercalation between DNA base pairs the compound displays an intense luminescence activity, quenched in aqueous solution.<sup>169,170</sup> An accepted explanation is that  $[\text{Ru}(\text{bpy})_2(\text{dppz})]^{2+}$  has a non-emissive (dark) MLCT low-lying excited state involving the phenazine moiety of the dppz ligand, and another emissive (bright) MLCT state related to the bpy part of the dppz ligand. In aqueous solution the dark state is favoured being at lower energy compared to the bright state. On the contrary, when intercalated into DNA, the dark state gets closer in energy to the bright state, allowing thermal population and increasing the emission (Fig. 18). The DNA duplex in which the  $[\text{Ru}(\text{bpy})_2\text{dppz}]^{2+}$  is intercalated prevents the quenching effect of the aqueous solution, resulting in a luminescence effect. Further investigation has demonstrated that after binding the DNA *via* intercalation, also the  $[\text{Ru}(\text{bpy})_2\text{dppz}]^{2+}$  compound can trigger the photocleavage of pUC18 plasmid DNA in presence of  $\text{O}_2$  ( $\lambda_{\text{irr}} > 455 \text{ nm}$ , 15 min).<sup>171</sup> In 1992, the light-switch effect of both  $[\text{Ru}(\text{bpy})_2\text{dppz}]^{2+}$  and  $[\text{Ru}(\text{phen})_2(\text{-dppz})]^{2+}$  was reported as function of the nucleic acid sequence and conformation. Indeed, the strongest luminescence effect was observed when the greatest amount of overlap between the nucleic acid structure and the complex was involved, such as when one of these complexes intercalates into triple helices. In fact, an increased luminescence was observed when the two compounds were bound to the triple helical assembly, permitting the dppz ligand to be better shielded from water by the extended surface area of the triplex. Subsequently, a detailed analysis by Choi *et al.* using separated  $\Lambda$  and  $\Delta$  enantiomers showed that both compounds can bind to a poly(dT  $\times$  dA-dT) triplex, displaying an increased luminescence compared to the duplex, assumed to be due to the larger surface area of the triplex that better protects the intercalating ligand dppz from water. This better protection and higher luminescence give a useful diagnostic of triplex formation. At the time of these solution studies, there was no clear structural evidence for any binding mode of these complexes to nucleic acids. Despite the third strand, access for intercalation is possible *via* the major groove, as has been proposed.<sup>172</sup> Detailed studies with separate enantiomers have elucidated by linear and circular dichroism that the Ru complexes with dppz and dppn (dppn = benzodipyrido[3,2-*a*:2',3'-*c*]phenazine) as ligands are able to intercalate between the nucleobases of a T-A:T triplex in the minor groove.



These authors made a detailed study of the bound chromophore orientation, and concluded that, especially for the  $\Delta$  complexes, the triplex binding mode had a close resemblance to that seen with duplexes. Very interestingly, the stabilisation of the third strand is related to the nature of the third phenanthroline, showing a stabilizing effect that increases in the order phen < dppn < dppz (phen = 1,10-phenanthroline). Intriguingly, the stabilising effect is not related to the size of the ligand.<sup>173</sup>

At the time of that publication, no structural data on duplex binding by these compounds was available. The later demonstration that the dppz chromophore intercalated exclusively from the minor groove implies that this would also be true with triplexes.<sup>174</sup> Thereafter, numerous studies have confirmed these interesting features, demonstrating the possible value of this class of compounds as photoluminescent probe for bioanalysis and application in PDT.<sup>164</sup>

Ru(II) complexes linked to triplex forming oligonucleotides could be used as photosensitisers in site-specific damaged DNA, as demonstrated by H el ene and co-workers. In fact, the complex  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  attached to the oligonucleotide and intercalated in the DNA formed a stable triplex. Different behaviours were observed between the two enantiomers of the compound, in fact the luminescence of the  $\Delta$  enantiomer linked to HIV-T oligonucleotide increased by 6–10 times, while no enhancement was observed with the  $\Lambda$  enantiomer. The  $\Delta$  enantiomer of the compound  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  linked to the 5'-phosphate group of the oligonucleotide by phenanthroline binds the DNA duplex in a sequence-specific way. The proposed mechanism is the formation of the triplex and the intercalation of the dppz ligand into the DNA molecule, leading to the stabilisation of the structure and to an enhancement of the fluorescence. Once again, the photophysical properties of ruthenium compounds such as the ability to photocleave, long-distance electron transfer and luminescence can be exploited for application in antigene-therapy or as photosensitiser for photodamage of the DNA by triple helix formation.<sup>175</sup> Indeed, the  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  complex is reported to successfully bind the DNA double helix, so that this property can be exploited to stabilise the triplex by conjugation of the complex to the 5'-end of a TFO. Importantly, the triplex formed by a TFO functionalised with  $[\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$  showed an increased stability by thermal denaturation compared to the triplex formed by the same unmodified oligonucleotide, with a  $\Delta T_m = 12$  °C. This indicates that the unmodified oligonucleotide forms less stable triplexes than the nucleotide decorated with the ruthenium complex.<sup>168</sup> The strong aromatic character of the dppz ligand allows for the intercalation both in duplex and triplex DNA, lying parallel to the triplex bases and intercalating into the minor groove of the triplex. Notably, the whole triplex structure is stabilised by the intercalation of the Ru-dppz complex bound to the TFO.<sup>176</sup>

Therefore, Ru polypyridyl derivatives are of great interest to obtain a stabilising effect on triplexes and to selectively cleave DNA by exploiting the high binding specificity of TFO and the photophysical properties of the ruthenium derivatives linked to the TFO.

A library of Ru(II) complexes with halogenated dppz ligands was screened against several biological molecules, such as proteins, ssDNA, dsDNA, DNA triplexes and DNA G-quadruplexes to understand the main factors influencing luminescent behaviour. It was proposed that (i) intercalation in the DNA structure of these compounds mainly depends on the changes of the halogenated substituent on the dppz ligand, (ii) the luminescence is increased in the presence of DNA structures but not in the presence of hydrophobic non-DNA structures such as BSA (iii) the  $\pi$  stacking surface area influences the luminescence. Indeed, after studying a panel of different substituents on the dppz ligand, more luminescence effect was detected with the compound  $[\text{Ru}(\text{bpy})_2\text{dppz-11,12-Br}]^{2+}$  in the A-T:A triplex and in intrastrand G-quadruplexes compared to intercalation into the DNA duplex. The authors suggest that large Br atoms in positions 11 and 12 prevent the complex from fully intercalating in the DNA duplex, causing the phenazine N atoms to be partially exposed to water, resulting in increased luminescence quenching. The luminescence was enhanced by 89 $\times$  in the presence of DNA triplexes compared to that in buffer alone. Moreover, this compound has also shown a 2.8 $\times$  higher luminescence when bound to G-quadruplexes compared to DNA triplexes, confirming that the  $\pi$  stacking surface area plays an important role in increasing the luminescence.<sup>177</sup> More structural studies are required to understand if this effect is due to the structure itself or to the DNA sequences.

DNA triplexes have also been used as part of an Enhanced Chemiluminescence (ECL) biosensor approach to detect the presence of adenosine in serum (Fig. 18). The ECL based on  $[\text{Ru}(\text{bpy})_3]^{2+}$  complexes are used to detect a large number of analytes with different percentages of selectivity and sensitivity. Those characteristics change based on different elements that are part of the ruthenium complexes. Nevertheless, the advantage is to work with an approach that completely avoids radioactive labels with the limit of detection that is low and simple to use. To quantify the presence of adenosine in serum, the method is based on an aptamer, attached on the surface of a gold electrode with an ECL signal marker composed of  $[\text{Ru}(\text{bpy})_3]^{2+}$  forming the first DNA strand. The other strand used as a quenching probe binds a ferrocene carboxylic acid (FcA) at the 5' end. A complex is formed with a third strand, complementary to the quencher, and coralyne chloride as binder. This complex is stable until the concentration of the adenosine increases. At this point the first strand assumes a hairpin configuration generating an intense luminescence due to the ruthenium complex and the absence of the FcA activity. This technique based on a DNA triplex has a more sensitive adenosine detection compared to the DNA duplex-based sensor.<sup>178</sup>

## 7. Conclusions

DNA triplexes are non-canonical structures that together with other unusual configurations, such as *i*-motif or quadruplexes, are part of the molecular biology field that is yet to be exploited. Triplexes possess large diversity in terms of stability, distortion, and environmental conditions required for the formation. In



order to exploit the DNA triplex for biological uses, numerous ligands have been designed over the years to functionalise these structures and enhance their stability in physiological conditions. Whilst multiple metal-based compounds have been developed to interact with DNA triplexes, Ru(II) polypyridyl compounds are of significant interest due to their photo-physical, electronic and biological properties.<sup>164</sup> Many interesting and promising results have been obtained. However, investigations that cover the role of ruthenium complexes in DNA triplexes are very limited. More studies are required to overcome the difficulties related to their chemical and cellular properties and increase the possibility of medical applications.

## Author contributions

MDP and AA contributed equally to this work. The literature review was undertaken by MDP and AA with input from CJC, GG and JPH. MDP and AA led the writing of the manuscript with input from CJC, GG and JPH.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We are grateful for financial support from the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie Innovative Training Network Nature-ETN, H2020-MSCA-ITN-2019-861381), the ERC Consolidator Grant PhotoMedMet to G. G. (GA 681679), the program "Investissements d' Avenir" launched by the French Government and implemented by the ANR with the reference ANR-10-IDEX-0001-02 PSL (G. G.).

## References

- M. Kaushik, S. Kaushik, K. Roy, A. Singh, S. Mahendru, M. Kumar, S. Chaudhary, S. Ahmed and S. Kukreti, *Biochem. Biophys. Rep.*, 2016, **5**, 388–395.
- H. Tateishi-Karimata and N. Sugimoto, *Chem. Commun.*, 2020, **56**, 2379–2390.
- R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Nature*, 1980, **287**, 755–758.
- B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, in *Molecular Biology of the Cell*, Garland Science, 4th edn., 2002.
- D. Voet and J. G. Voet, *Biochemistry*, Pedersen, New York, 2nd edn, 1995, vol. 1.
- D. W. Ussery, *eLS*, 2002, 1–7.
- N. Saini, Y. Zhang, K. Usdin and K. S. Lobachev, *Biochimie*, 2013, **95**, 117–123.
- L. Pauling and R. B. Corey, *Arch. Biochem. Biophys.*, 1956, **65**, 164–181.
- G. Felsenfeld, D. R. Davies and A. Rich, *J. Am. Chem. Soc.*, 1957, **79**(8), 2023–2024.
- M. D. Frank-Kamenetskii and S. M. Mirkin, *Annu. Rev. Biochem.*, 1995, **64**, 65–95.
- J. Spiegel, S. Adhikari and S. Balasubramanian, *Trends Chem.*, 2020, **2**, 123–136.
- D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627–8637.
- J. A. Brazier, A. Shah and G. D. Brown, *Chem. Commun.*, 2012, **48**, 10739–10741.
- H. Inagaki and H. Kurahashi, in *Brenner's Encyclopedia of Genetics*, 2nd edn, 2013, vol. 2, pp. 241–243.
- A. Mandal, *IJRES*, 2020, **8**, 17–20.
- C. S. Allardyce and P. J. Dyson, *Platinum Met. Rev.*, 2001, **45**, 62–69.
- A. C. Hachey, D. Havrylyuk and E. C. Glazer, *Curr. Opin. Chem. Biol.*, 2021, **61**, 191–202.
- A. Notaro and G. Gasser, *Chem. Soc. Rev.*, 2017, **46**, 7317–7337.
- S. Monro, K. L. Colón, H. Yin, J. R. III, P. Konda, S. Gujar, R. P. Thummel, L. Lilge, C. G. Cameron and S. A. McFarland, *Chem. Rev.*, 2019, **119**, 797–828.
- L. Petraccone, E. Erra, C. A. Mattia, V. Fedullo, G. Barone and C. Giancola, *Biophys. Chem.*, 2004, **110**, 73–81.
- J. Volker and H. H. Klump, *Biochemistry*, 1994, **33**, 13502–13508.
- B. J. Boehm, C. Whidborne, A. L. Button, T. L. Pukala and D. M. Huang, *Phys. Chem. Chem. Phys.*, 2018, **20**, 14013–14023.
- M. Mills, P. B. Arimondo, L. Lacroix, T. Garestier, H. Klump and J.-L. Mergny, *Biochemistry*, 2002, **41**, 357–366.
- R. Thenmalarchelvi and N. Yathindra, *Nucleic Acid Res.*, 2005, **33**, 43–55.
- Y. Kohwi and Y. Panchenko, *Genes Dev.*, 1993, **7**, 1766–1778.
- W. Li, X.-M. Hou, P.-Y. Wang, X.-G. Xi and M. Li, *J. Am. Chem. Soc.*, 2013, **135**, 6423–6426.
- X.-M. Hou, Y.-B. Fu, W.-Q. Wu, L. Wang, F.-Y. Teng, P. Xie, P.-Y. Wang and X.-G. Xi, *Nucleic Acids Res.*, 2017, **45**, 11401–11412.
- V. B. Zhurkin, G. Raghunathan, N. B. Ulyanov, R. D. Camerini-Otero and R. L. Jernigan, *J. Mol. Biol.*, 1994, **239**, 181–200.
- P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497–1500.
- A. Gupta, A. Mishra and N. Puri, *J. Biotechnol.*, 2017, **259**, 148–159.
- H. M. Berman, J. Westbrook, F. Zukang, G. Gilliland, T. N. Bhat, H. Weissig, L. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
- A. H.-J. Wang, T. Hakoshima, G. van der Marel, J. H. van Boom and A. Rich, *Cell*, 1984, **37**, 321–331.
- E. S. Baker and M. T. Bowers, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 1188–1195.
- L. I. Radhakrishnan, D. J. Patell, C. Biochemistry and B. Program, *Structure*, 1993, **1**, 135–152.
- I. Radhakrishnan and D. J. Patel, *J. Mol. Biol.*, 1994, **241**, 600–619.
- L. Van Meervelt, D. Vlieghe, A. Dautant, B. Gallois, G. Précigoux and O. Kennard, *Nature*, 1995, **374**, 742–744.



- 37 J. P. Bartley, T. Brown and A. N. Lane, *Biochemistry*, 1997, **36**, 14502–14511.
- 38 L. M. Tarköy, A. K. Phipps, P. Schultze and J. Feigon, *Biochemistry*, 1998, **37**, 5810–5819.
- 39 J. L. Asensio, A. N. Lane and T. Brown, *Nucleic Acids Res.*, 1998, **26**, 3677–3686.
- 40 S. Rhee, Z.-J. Han, K. Liu, H. T. Miles and D. R. Davies, *Biochemistry*, 1999, **38**, 16810–16815.
- 41 J. L. Asensio, T. Brown and A. N. Lane, *Structure*, 1999, **7**, 1–11.
- 42 M. J. P. Van Dongen, J. F. Doreleijers, G. A. Van Der Marel, J. H. Van Boom, C. W. Hilbers and S. S. Wijmenga, *Nat. Struct. Biol.*, 1999, **6**, 854–859.
- 43 L. Cerofolini, J. Amato, A. Giachetti, V. Limongelli, E. Novellino, M. Parrinello, M. Fragai, A. Randazzo and C. Luchinat, *Nucleic Acids Res.*, 2014, **42**, 13393–13404.
- 44 L. Betts, J. A. Josey, J. M. Veal and S. R. Jordan, *Science*, 1995, **270**, 1838.
- 45 E. Wang, K. M. Koshlap, P. Gillespie, P. B. Dervan and J. Feigon, *J. Mol. Biol.*, 1996, **257**, 1052–1069.
- 46 K. M. Koshlap, P. Schultze, H. Brunar, P. B. Dervan and J. Feigon, *Biochemistry*, 1997, **36**, 2659–2668.
- 47 A. K. Phipps, M. Tarköy, P. Schultze and J. Feigon, *Biochemistry*, 1998, **37**, 5820–5830.
- 48 J. J. Sørensen, J. T. Nielsen and M. Petersen, *Nucleic Acids Res.*, 2004, **32**, 6078–6085.
- 49 B. Petersson, B. B. Nielsen, H. Rasmussen, I. K. Larsen, M. Gajhede, P. E. Nielsen and J. S. Kastrup, *J. Am. Chem. Soc.*, 2005, **127**, 1424–1430.
- 50 M. Garavis, P. J. B. Edwards, O. Doluca, V. V. Filichev and C. Gonzalez, *TINA-conjugated antiparallel DNA triplex*, [https://www.wwpdb.org/pdb?id=pdb\\_00006qhi](https://www ww p d b . o r g / p d b ? i d = p d b _ 0 0 0 0 6 q h i), accessed 17 July 2021.
- 51 S. Neidle, *DNA Structure and Recognition*, 1994.
- 52 M. Esguerra, L. Nilsson and A. Villa, *Nucleic Acids Res.*, 2014, **42**, 11329–11338.
- 53 R. E. Dickerson and H.-L. Ng, *Proc. Natl. Acad. Sci.*, 2001, **98**, 6986–6988.
- 54 X.-J. Lu and W. K. Olson, *Nat. Protoc.*, 2008, **3**, 1213–1227.
- 55 J. Robles, A. Grandas, E. Pedrosa, F. Luque, R. Eritja and M. Orozco, *Curr. Org. Chem.*, 2002, **6**, 1333–1368.
- 56 D. A. Rusling, T. Brown and K. R. Fox, DNA Recognition by Triple Helix Formation, *Sequence specific DNA Binding Agents*, 2006.
- 57 V. A. Malkov, O. N. Voloshin, V. N. Soyfer and M. D. Frank-Kamenetskii, *Nucleic Acids Res.*, 1993, **21**, 585–591.
- 58 F. Svinarchuk, J. Paoletti and C. Malvy, *J. Biol. Chem.*, 1995, **270**, 14068–14071.
- 59 S. Kaushik, M. Kaushik, F. Svinarchuk, C. Malvy, S. Fermanjian and S. Kukreti, *Biochemistry*, 2011, **50**, 4132–4142.
- 60 N. Sugimoto, P. Wu, H. Hara and Y. Kawamoto, *Biochemistry*, 2001, **40**, 9396–9405.
- 61 H.-X. Jiang, Y. Cui, T. Zhao, H.-W. Fu, D. Koirala, J. A. Punnoose, D.-M. Kong and H. Mao, *Sci. Rep.*, 2015, **5**, 9255.
- 62 Y. Teng, H. Tateishi-Karimata, T. Ohyama and N. Sugimoto, *Molecules*, 2020, **25**, 1–13.
- 63 G. Chen and S.-J. Chen, *Phys. Biol.*, 2011, **8**, 066006.
- 64 R. Goobes, O. Cohen and A. Minsky, *Nucleic Acids Res.*, 2002, **30**, 2154–2161.
- 65 A. Ohkubo, K. Yamada, Y. Ito, K. Yoshimura, K. Miyauchi, T. Kanamori, Y. Masaki, K. Seio, H. Yuasa and M. Sekine, *Nucleic Acids Res.*, 2015, **43**, 5675–5686.
- 66 C. E. Carr, R. Ganugula, R. Shikiya, A. M. Soto and L. A. Marky, *Biochimie*, 2018, **146**, 156–165.
- 67 A. Ono, P. O. P. Ts'o and L. sing Kan, *J. Org. Chem.*, 1992, **57**, 3225–3230.
- 68 E. L. Jones, A. J. Mlotkowski, S. P. Hebert, H. B. Schlegel and C. S. Chow, *J. Phys. Chem. A*, 2022, **126**, 1518–1529.
- 69 T. M. Chin, S. B. Lin, S. Y. Lee, M. L. Chang, A. Y. Y. Cheng, F. C. Chang, L. Pasternack, D. H. Huang and L. S. Kan, *Biochemistry*, 2000, **39**, 12457–12464.
- 70 K. R. Fox, T. Brown and D. A. Rusling, in *DNA-targeting Molecules as Therapeutic Agents*, Royal Society of Chemistry, London, 2018, pp. 1–32.
- 71 P. Bates, C. A. Loughton, T. C. Jenkins, D. C. Capaldi, P. D. Roselt, C. B. Reese and S. Neidle, *Nucleic Acids Res.*, 1996, **24**, 4176–4184.
- 72 D. L. Chen and L. W. McLaughlin, *J. Org. Chem.*, 2000, **65**, 7468–7474.
- 73 C. Lou, M. Shelbourne, K. R. Fox and T. Brown, *Chem.–Eur. J.*, 2011, **17**, 14851–14856.
- 74 C. Y. Huang, C. D. Cushman and P. S. Miller, *J. Org. Chem.*, 1993, **58**, 5048–5049.
- 75 C. Y. Huang, G. Bi and P. S. Miller, *Nucleic Acids Res.*, 1996, **24**, 2606–2613.
- 76 C. Y. Huang, C. D. Cushman and P. S. Miller, *J. Org. Chem.*, 1993, **58**, 5048–5049.
- 77 T. J. Povsic and P. B. Dervan, *J. Am. Chem. Soc.*, 1989, **111**, 3059–3061.
- 78 I. Okamoto, K. Seio and M. Sekine, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 3334–3336.
- 79 D. A. Rusling, *Nucleic Acids Res.*, 2021, **49**, 7256–7266.
- 80 J. S. Koh and P. B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 1470–1478.
- 81 I. Radhakrishnan, D. J. Patel, E. S. Priestley, H. M. Nash and P. B. Dervan, *Biochemistry*, 1993, **32**, 11228–11234.
- 82 J. Marfurt, S. P. Parel and C. J. Leumann, *Nucleic Acids Res.*, 1997, **25**, 1875–1882.
- 83 J. F. Milligan, S. H. Krawczyk, S. Wadwani and M. D. Matteucci, *Nucleic Acids Res.*, 1993, **21**, 327–333.
- 84 W. M. Olivas and L. J. Maher, *Nucleic Acids Res.*, 1995, **23**, 1936–1941.
- 85 V. M. Marathias, M. J. Sawicki and P. H. Bolton, *Nucleic Acids Res.*, 1999, **27**, 2860–2867.
- 86 T. S. Rao, A. F. Lewis, R. H. Durland and G. R. Revankar, *Tetrahedron Lett.*, 1993, **34**, 6709–6712.
- 87 F. Raynaud, U. Asseline, V. Roig and N. T. Thuong, *Tetrahedron*, 1996, **52**, 2047–2064.
- 88 Y. Aubert, L. Perrouault, C. Hélène, C. Giovannangeli and U. Asseline, *Bioorg. Med. Chem.*, 2001, **9**, 1617–1624.
- 89 F. A. Rogers, J. A. Lloyd and M. K. Tiwari, *Artif. DNA PNA XNA*, 2014, **5**, 1–9.



- 90 Y. Taniguchi, M. Miyazaki, N. Matsueda, L. Wang, H. Okamura and S. Sasaki, *Chem. Pharm. Bull.*, 2018, **66**, 624–631.
- 91 J. Dagle and D. L. Weeks, *Nucleic Acids Res.*, 1996, **24**, 2143–2149.
- 92 P. S. Miller, P. Bhan, C. D. Cushman and T. L. Trapane, *Biochemistry*, 1992, **31**, 6788–6793.
- 93 A. Aviñó, E. Cubero, C. González, R. Eritja and M. Orozco, *J. Am. Chem. Soc.*, 2003, **125**, 16127–16138.
- 94 Y. Taniguchi, Y. Magata, T. Osuki, R. Notomi, L. Wang, H. Okamura and S. Sasaki, *Org. Biomol. Chem.*, 2020, **18**, 2845–2851.
- 95 P. E. Nielsen, *Appl. Biochem. Biotechnol., Part B*, 2004, **26**, 233–248.
- 96 K. M. Vasquez and P. M. Glazer, *Q. Rev. Biophys.*, 2002, **1**, 89–107.
- 97 D. P. Arya and T. C. Bruice, *Proc. Natl. Acad. Sci.*, 1999, **96**, 4384–4389.
- 98 A. Blaskó, E. E. Minyat, R. O. Dempcy and T. C. Bruice, *Biochemistry*, 1997, **36**, 7821–7831.
- 99 C. Escudé and J.-S. Sun, *Top. Curr. Chem.*, 2005, 109–148.
- 100 B. W. Sun, F. Geinguenaud, E. Taillandier, M. Naval, A. Laurent, F. Debart and J. J. Vasseur, *J. Biomol. Struct. Dyn.*, 2002, **19**, 1073–1081.
- 101 Y. Su, M. Bayarjargal, T. K. Hale and V. V. Filichev, *Beilstein J. Org. Chem.*, 2021, **17**, 749–761.
- 102 S. C. Brown, S. A. Thomson, J. M. Veal and D. G. Davis, *Science*, 1994, **265**, 777–780.
- 103 G. C. Shields, C. A. Loughton and M. Orozco, *J. Am. Chem. Soc.*, 1998, **120**, 5895–5904.
- 104 P. Zhou, M. Wang, L. Du, G. W. Fisher, A. Waggoner and D. H. Ly, *J. Am. Chem. Soc.*, 2003, **125**, 6878–6879.
- 105 R. Schütz, M. Cantin, C. Roberts, B. Greiner, E. Uhlmann and C. Leumann, *Angew. Chem., Int. Ed.*, 2000, **39**, 1250–1253.
- 106 M. Kuwahara, M. Arimitsu and M. Sisido, *J. Am. Chem. Soc.*, 1999, **121**, 256–257.
- 107 D. R. Jayarathna, H. D. Stout and C. Achim, *Inorg. Chem.*, 2018, **57**, 6865–6872.
- 108 U. Koppelhus and P. E. Nielsen, *Adv. Drug Delivery Rev.*, 2003, **55**, 267–280.
- 109 M. Pooga, T. Land, T. Bartfai and Ü. Langel, *Biomol. Eng.*, 2001, **17**, 183–192.
- 110 A. Soler-Bistué, A. Zorreguieta and M. E. Tolmasky, *Molecules*, 2019, **24**, 1–17.
- 111 T. Højland, S. Kumar, B. R. Babu, T. Umemoto, N. Albæk, P. K. Sharma, P. Nielsen and J. Wengel, *Org. Biomol. Chem.*, 2007, **5**, 2375–2379.
- 112 K. Morita, M. Takagi, C. Hasegawa, M. Kaneko, S. Tsutsumi, J. Sone, T. Ishikawa, T. Imanishi and M. Koizumi, *Bioorg. Med. Chem.*, 2003, **11**, 2211–2226.
- 113 M. Shimizu, A. Konishi, Y. Shimada, H. Inoue and E. Ohtsuka, *FEBS Lett.*, 1992, **302**, 155–158.
- 114 M. R. Alam, A. Majumdar, A. K. Thazhathveetil, S. T. Liu, J. L. Liu, N. Puri, B. Cuenoud, S. Sasaki, P. S. Miller and M. M. Seidman, *Biochemistry*, 2007, **46**, 10222–10233.
- 115 A. K. Shchylolkina and O. F. Borisova, *FEBS Lett.*, 1997, **419**, 27–31.
- 116 K. Sandström, S. Wärmländer, J. Bergman, R. Engqvist, M. Leijon and A. Gräslund, *J. Mol. Recognit.*, 2004, **17**, 277–285.
- 117 K. Sandstrom, S. Warmlander, A. Graslund and M. Leijon, *J. Mol. Biol.*, 2002, 737–748.
- 118 W. D. Wilson, F. A. Tanius, S. Mizan, S. Yao, A. S. Kiselyov, G. Zon and L. Strekowski, *Biochemistry*, 1993, **32**, 10614–10621.
- 119 O. F. Borisova, A. X. Shchylolkina, E. N. Timofeev, S. Y. Tsybenko, A. D. Mirzabekov and V. L. Florentiev, *J. Biomol. Struct. Dyn.*, 1995, **13**, 015–027.
- 120 V. V. Filichev and E. B. Pedersen, *J. Am. Chem. Soc.*, 2005, **127**, 14849–14858.
- 121 M. Paramasivam, S. Cogoi, V. V. Filichev, N. Bomholt, E. B. Pedersen and L. E. Xodo, *Nucleic Acids Res.*, 2008, **36**, 3494–3507.
- 122 I. Gécí, V. V. Filichev and E. B. Pedersen, *Bioconjugate Chem.*, 2006, **17**, 950–957.
- 123 D. P. Arya, L. Micovic, I. Charles, R. L. Coffee, B. Willis and L. Xue, *J. Am. Chem. Soc.*, 2003, **125**, 3733–3744.
- 124 D. P. Arya, *Acc. Chem. Res.*, 2011, **44**, 134–146.
- 125 J. Song, Z. Intody, M. Li and J. H. Wilson, *Gene*, 2004, **324**, 183–190.
- 126 Y. H. Ping and T. M. Rana, *Biochemistry*, 2005, **44**, 2501–2509.
- 127 Y. Liu, R. S. Nairn and K. M. Vasquez, *Nucleic Acids Res.*, 2009, **37**, 6378–6388.
- 128 M. P. Knauert and P. M. Glazer, *Hum. Mol. Genet.*, 2001, **10**, 2243–2251.
- 129 C. A. Spencer and M. Groudine, *Adv. Cancer Res.*, 1991, **56**, 1–48.
- 130 E. H. Postel, S. J. Flint, D. J. Kesslert and M. E. Hogant, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 8227–8231.
- 131 M. R. Rajeswari, *J. Biosci.*, 2012, **37**, 519–532.
- 132 H. Ruan and Y. H. Wang, *J. Mol. Biol.*, 2008, **383**, 292–300.
- 133 K. Nakanishi, F. Cavallo, E. Brunet and M. Jasin, *Methods Mol. Biol.*, 2011, **745**, 283–291.
- 134 A. F. Faruqi, H. J. Datta, D. Carroll, M. M. Seidman and P. M. Glazer, *Mol. Cell. Biol.*, 2000, **20**, 990–1000.
- 135 B. S. Thoma, M. Wakasugi, J. Christensen, M. C. Reddy and K. M. Vasquez, *Nucleic Acid Res.*, 2005, **33**, 2993–3001.
- 136 M. Brázdová, V. Tichý, R. Helma, P. Bažantová, A. Polášková, A. Krejčí, M. Petr, L. Navrátilová, O. Tichá, K. Nejedlý, M. L. Bennink, V. Subramaniam, Z. Bábková, T. Martínek, M. Lexa and M. Adámik, *PLoS One*, 2016, 1–25.
- 137 Y. Du and X. Zhou, *Chem. Rec.*, 2013, **13**, 371–384.
- 138 Y. Hu, A. Ceconello, A. Idili, F. Ricci and I. Willner, *Angew. Chem., Int. Ed.*, 2017, **56**, 15210–15233.
- 139 P. Conlon, C. J. Yang, Y. Wu, Y. Chen, K. Martinez, Y. Kim, N. Stevens, A. A. Marti, S. Jockusch, N. J. Turro and W. Tan, *J. Am. Chem. Soc.*, 2008, **130**, 336–342.
- 140 J. Huang, Z. Zhu, S. Bamrungsap, G. Zhu, M. You, X. He, K. Wang and W. Tan, *Anal. Chem.*, 2010, **82**, 10158–10163.
- 141 D. Xi, X. Wang, S. Ai and S. Zhang, *Chem. Commun.*, 2014, **50**, 9547–9549.



- 142 Y. Chen, K. Murayama, H. Kashida, Y. Kamiya and H. Asanuma, *Chem. Commun.*, 2020, **56**, 5358–5361.
- 143 A. Patterson, F. Caprio, A. Valle, D. Moscone, K. W. Plaxco, G. Palleschi and F. Ricci, *Anal. Chem.*, 2010, **82**, 9109–9115.
- 144 V. A. Online, N. Sun, Q. Guo, J. Shao, B. Qiu, Z. Lin, K. Y. Wong and G. Chen, *Anal. Methods*, 2014, **6**, 3370–3374.
- 145 J. Zhang, P. Chen, X. Wu, J. Chen, L. Xu, G. Chen and F. Fu, *Biosens. Bioelectron.*, 2011, **26**, 2645–2650.
- 146 A. Porchetta, A. Idili, A. Valle and F. Ricci, *Nano Lett.*, 2015, **15**, 4467–4471.
- 147 C. Xiong, C. Wu, H. Zhang and L. Ling, *Spectrochim. Acta, Part A*, 2011, **79**, 956–961.
- 148 W. C. Liao, M. Riutin, W. J. Parak and I. Willner, *ACS Nano*, 2016, **10**, 8683–8689.
- 149 R. K. Sodhi, *Cancer Ther. Oncol. Int. J.*, 2019, **14**, 1–8.
- 150 B. Rosenberg, in *Cisplatin*, Elsevier, 1980, pp. 9–20.
- 151 C. Moucheron, *New J. Chem.*, 2009, **33**, 235–245.
- 152 A. C. Komor and J. K. Barton, *Chem. Commun.*, 2013, **49**, 3617.
- 153 M. Flamme, C. Figazzolo, G. Gasser and M. Hollenstein, *Metallomics*, 2021, **13**(4), mfab016.
- 154 K. Tanaka, Y. Yamada and M. Shionoya, *J. Am. Chem. Soc.*, 2002, **124**, 8802–8803.
- 155 A. K. Shchvolkina, *Nucleic Acids Res.*, 2001, **29**, 986–995.
- 156 E. Paleček, *Methods Enzymol.*, 1992, **212**, 139–155.
- 157 A. Oleksi, A. G. Blanco, R. Boer, I. Usón, J. Aymamí, A. Rodger, M. J. Hannon and M. Coll, *Angew. Chem.*, 2006, **118**, 1249–1253.
- 158 J. Malina, M. J. Hannon and V. Brabec, *FEBS J.*, 2014, **281**, 987–997.
- 159 O. Hrabina, J. Malina, P. Scott and V. Brabec, *Chem.–Eur. J.*, 2020, **26**, 16554–16562.
- 160 J. Malina, N. P. Farrell and V. Brabec, *Angew. Chem*, 2018, **130**, 8571–8575.
- 161 J. Malina, N. P. Farrell and V. Brabec, *Inorg. Chem.*, 2020, **11**, 15135–15143.
- 162 S. Y. Lee, C. Y. Kim and T.-G. Nam, *Drug Des., Dev. Ther.*, 2020, 5375–5392.
- 163 L. Zeng, P. Gupta, Y. Chen, E. Wang, L. Ji, H. Chao and Z. S. Chen, *Chem. Soc. Rev.*, 2017, **46**, 5771–5804.
- 164 S. Zhang, Y. Ding and H. Wei, *Molecules*, 2014, **19**, 11933–11987.
- 165 C. Mari, V. Pierroz, S. Ferrari and G. Gasser, *Chem. Sci.*, 2015, **6**, 2660–2686.
- 166 A. E. O'Connor, W. M. Gallagher and A. T. Byrne, *Photochem. Photobiol.*, 2009, **85**, 1053–1074.
- 167 J. D. Knoll and C. Turro, *Coord. Chem. Rev.*, 2015, **282–283**, 110–126.
- 168 M. Flamme, E. Clarke, G. Gasser and M. Hollenstein, *Molecules*, 2018, **23**, 28–31.
- 169 A. E. Friedman, J. Chambron, J. Sauvage, N. J. Turro and J. K. Barton, *J. Am. Chem. Soc.*, 1990, **112**, 4960–4962.
- 170 R. M. Hartshorn and J. K. Barton, *J. Am. Chem. Soc.*, 1992, **114**, 5919–5925.
- 171 L. Hu, Z. Bian, H. Li, S. Han, Y. Yuan, L. Gao and G. Xu, *Anal. Chem.*, 2009, **81**, 9807–9811.
- 172 Y. Jenkins, J. K. Barton, A. E. Friedman and N. J. Turro, *Biochemistry*, 1992, **31**, 10809–10816.
- 173 S. D. Choi, M. S. Kim, S. K. Kim, P. Lincoln, E. Tuite and B. Nordén, *Biochemistry*, 1997, **36**, 214–223.
- 174 C. J. Cardin, J. M. Kelly and S. J. Quinn, *Chem. Sci.*, 2017, **8**, 4705–4723.
- 175 G. N. Grimm, A. S. Boutorine, P. Lincoln, B. Nordén and C. Hélène, *ChemBioChem*, 2002, **3**, 324–331.
- 176 G. Li, L. Sun, L. Ji and H. Chao, *Dalton Trans.*, 2016, **45**, 13261–13276.
- 177 E. Wachter, D. Moyá and E. C. Glazer, *ACS Comb. Sci.*, 2017, **19**, 85–95.
- 178 S. Ye, H. Li and W. Cao, *Biosens. Bioelectron.*, 2011, **26**, 2215–2220.
- 179 T. Biver, *Coord. Chem. Rev.*, 2013, **257**, 2765–2783.
- 180 R. E. Dickerson, *International Tables for Crystallography*, Wiley, 2012, p. 610.
- 181 I. Géci, V. V. Filichev and E. B. Pedersen, *Bioconjugate Chem.*, 2006, **17**, 950–957.
- 182 M. L. Di Pietro, G. La Ganga, F. Nastasi and F. Puntoriero, *Appl. Sci.*, 2021, **11**, 3038.

