Biophysical & Crystallographic Studies of DNA i-motifs

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

SARAH PRAVA GURUNG

August 2017

Declaration

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

<div align="right">Sarah Prava Gurung</div>

Acknowledgements

First of all, I would like to thank my supervisors Professor Christine Cardin, Dr. John Brazier, Professor Thomas Sorensen and Dr. Graeme Winter for all their support and patience throughout my years as a PhD student. You have really let me experience a studentship with no limitations by allowing and encouraging me to venture on with the sciences I never imagined I could be a part of.

A big thanks to Dr. James Hall for the time and energy he has put into my project since the beginning of my studies at Reading University. I would like to thank Professor David Cardin, Kane Mcquaid, Ann-Kathrin Geiger and all the technical students from the Germany exchange programme, not only for synthesising the compounds used in this work but also for the wonderful moments shared together.

Thank you to everyone at Diamond Light Source and the Research Complex at Harwell for all your help with my project and for your wonderful company. Thank you to Dr. Carina Lobley, Maria Rosa, Dr. Anil Verma, Dr. James Sandy, Dr. Juan Sanchez-Wetherby and Dr. Tina Geraki. I am also very grateful to Dr. Susana Teixeira and Dr. Leighton Coates for all their help with neutron crystallography at the Spallation Neutron Source.

As a laboratory demonstrator during my studies, I had the surprising opportunity to work with my A-level Chemistry teacher, Dr. David Wright, who had initially inspired me to take the subject further at university. I want to thank him and the late Dr. Sandy Briden, one of the bravest people I've ever known, for re-igniting the fun in Chemistry and encouraging me with my project.

A lot has happened during the past four years. Not only has this journey been exceptionally educational but it has also been extremely emotional. I would like to thank my late aunts and friends who had always cheered me on with my work. And I want to thank everyone who was there for me during those tough times. I will always remember your help.

Finally, I would like to thank my parents, Dhak and Susiela Gurung, who have made a lot of sacrifices to make sure that I received the education I wanted; back in Nepal as well as here in the UK, and to my other half, Damber Limbu, for always supporting me. Also, a huge shout-out to my brother Joe, my relatives and friends, my cat Rocket and dog Jet for keeping me sane during writing up.

Abstract

Intramolecular i-motifs of the form $C_3L_{3-8}C_3L_{3-8}C_3L_{3-8}C_3$, where C_3 denotes the cytosine stretch and L_{3-8} are "loop" regions containing any DNA base (L) including cytosine, were studied to understand the effects of loop length on i-motif stability. It contrast to the previously held notion that long-looped i-motifs are more stable, it was found that i-motif structures with short loops exhibit higher thermal stabilities and transitional pH values.

The stability of long-looped i-motifs are then shown to increase with the addition of $[Ru(phen)_2dppz]^{2+}$ (phen = 1,10-phenanthroline, dppz = dipyrido [3,2-a:2',3'-c] phenazine); a polypyridyl complex that has a potential for photodynamic therapy. Addition of the complex enhances the stability of $d(C_3T_{838})_3C_3$ but not that of $d(C_3T_{383})_3C_3$, implying that loop lengths are important in defining i-motif-ligand interactions.

The effects of loop base composition on the stability of i-motifs have also been presented. It is shown that when $d(C_3XYZ)_3C_3$ sequences are used (where X and Z are adenine, thymine and guanine, and Y is any of the four DNA bases), pyrimidine-rich sequences form more stable i-motif structures. However, when guanine is X, only two out of 12 $d(C_3XY\textbf{\textit{G}})_3C_3$ sequences were able to form i-motifs. Change in sequence direction also resulted in different thermal and pH stabilities; emphasising the role of loop base composition on not only the i-motif's stability but also in its formation.

X-ray crystallography was used to further understand the effects of loop bases on i-motif structures. The study focuses on four tetramolecular i-motifs; two of which were solved in the mid-1990's; $(d(C_4)_4$ and $d(C_3T)_4$ but have now been re-examined using improved experimental approaches. Two novel i-motif structures of $d(C_3A)_4$ and $[d(C_3A) + d(C_3T)]$ are presented.

Following the X-ray diffraction of $d(C_3T)_4$ crystals to 0.68 Å resolution (previously reported at 1.4 Å) at beamline I02 (Diamond Light Source Ltd.), a novel neutron diffraction study on the particular i-motif was conducted. Single crystal neutron diffraction was carried out at MaNDi beamline (Spallation Neutron Source) to find the distribution of the proton between the hemi-protonated cytosine$^+$·cytosine base pairs and to understand the role that H-bonded water can play in stabilising the i-motif structure.

Contents

Abbreviations Used

DNA – Deoxyribonucleic acid

RNA – Ribonucleic acid

ss – single-stranded DNA

ds – double-stranded DNA

C – Cytosine

T – Thymine

G – Guanine

A – Adenine

N – Nitrogen

H – Hydrogen

D – Deuterium

T_m – DNA melting temperature

T_{pH} – Transitional pH value

phen – 1,10-phenanthroline

dppz – dipyridophenazine

$[Ru(phen)_2dppz]^{2+}$
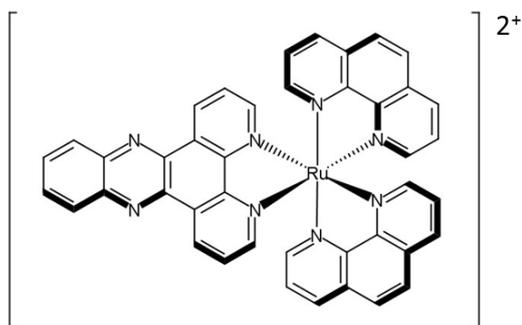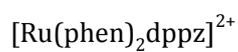
Experimental Methods

UV – Ultraviolet

CD – Circular Dichroism

srCD – Synchrotron Radiation Circular Dichroism

VD – Vapour Diffusion

SDVD – Sitting Drop Vapour Diffusion

XC – X-ray Crystallography

NC – Neutron Crystallography

MR – Molecular Replacement

TOF – Time of Flight

DNA Sequences used for Crystallographic Studies

Abbreviation	Sequence (5'→3')
C_4	CCCC
C_3T	CCCT
C_3A	CCCA
C_3G	CCCG

DNA Sequences used for Solution Studies

Group	Name	Sequence (5'→3')
C$_3$T$_x$	C$_3$T$_3$	CCCTTTCCCTTTCCCTTTCCCT
	C$_3$T$_4$	CCCTTTTCCCTTTTCCCTTTTCCCT
	C$_3$T$_5$	CCCTTTTTCCCTTTTTCCCTTTTTCCCT
	C$_3$T$_6$	CCCTTTTTTCCCTTTTTTCCCTTTTTTCCCT
	C$_3$T$_7$	CCCTTTTTTTCCCTTTTTTTCCCTTTTTTTCCCT
	C$_3$T$_8$	CCCTTTTTTTTCCCTTTTTTTTCCCTTTTTTTTCCCT
C$_3$T$_{abc}$	C$_3$T$_{338}$	CCCTTTCCCTTTCCCTTTTTTTTCCC
	C$_3$T$_{383}$	CCCTTTCCCTTTTTTTTCCCTTTCCC
	C$_3$T$_{833}$	CCCTTTTTTTTCCCTTTCCCTTTCCC
	C$_3$T$_{883}$	CCCTTTTTTTTCCCTTTTTTTTCCCTTTCCC
	C$_3$T$_{838}$	CCCTTTTTTTTCCCTTTCCCTTTTTTTTCCC
	C$_3$T$_{388}$	CCCTTTCCCTTTTTTTTCCCTTTTTTTTCCC
C$_3$**A**YZ	AAA	CCCAAACCCAAACCCAAACCC
	AAT	CCCAATCCCAATCCCAATCCC
	AAG	CCCAAGCCCAAGCCCAAGCCC
	ATA	CCCATACCCATACCCATACCC
	ATT	CCCATTCCCATTCCCATTCCC
	ATG	CCCATGCCCATGCCCATGCCC
	AGA	CCCAGACCCAGACCCAGACCC
	AGT	CCCAGTCCCAGTCCCAGTCCC
	AGG	CCCAGGCCCAGGCCCAGGCCC
	ACA	CCCACACCCACACCCACACCC
	ACT	CCCACTCCCACTCCCACTCCC
	ACG	CCCACGCCCACGCCCACGCCC
C$_3$**T**YZ	TAA	CCCTAACCCTAACCCTAACCC
	TAT	CCCTATCCCTATCCCTATCCC
	TAG	CCCTAGCCTAAGCCCTAGCCC
	TTA	CCCTTACCCTTACCCTTACCC
	TTT	CCCTTTCCCTTTCCCTTTCCC
	TTG	CCCTTGCCCTTGCCCTTGCCC
	TGA	CCCTGACCCTGACCCTGACCC
	TGT	CCCTGTCCCTGTCCCTGTCCC
	TGG	CCCTGGCCCTGGCCCTGGCCC
	TCA	CCCTCACCCTCACCCTCACCC
	TCT	CCCTCTCCCTCTCCCTCTCCC
	TCG	CCCTCGCCCTCGCCCTCGCCC
C$_3$**G**YZ	GAA	CCCGAACCCGAACCCGAACCC
	GAT	CCCGATCCCGATCCCGATCCC
	GAG	CCCGAGCCCGAGCCCGAGCCC
	GTA	CCCGTACCCGTACCCGTACCC
	GTT	CCCGTTCCCGTTCCCGTTCCC
	GTG	CCCGTGCCCGTGCCCGTGCCC
	GGA	CCCGGACCCGGACCCGGACCC
	GGT	CCCGGTCCCGGTCCCGGTCCC
	GGG	CCCGGGCCCGGGCCCGGGCCC
	GCA	CCCGCACCCGCACCCGCACCC
	GCT	CCCGCTCCCGCTCCCGCTCCC
	GCG	CCCGCGCCCGCGCCCGCGCCC

Contribution Statement

Chapter 2 – The Importance of Loop Length on i-motif Stability

DNA melting experiments were conducted by Sarah Gurung. Synchrotron radiation CD (srCD) measurements were recorded at beamline B23, Diamond Light Source Ltd. by Sarah Gurung, Dr. J. Hall and Dr. J. Brazier. Sarah Gurung and Christine Schwarz carried out the pH titration measurements. The work presented in this chapter is featured in the following publication:

Gurung, S. P., Schwarz, C., Hall, J. P., Cardin, C. J. and Brazier, J. A. The importance of loop length on the stability of i-motif structures. *Chem. Comm.* **51**, 5630-5632 (2015)

Chapter 3 – Stabilisation of Long-looped i-motifs by [Ru(phen)$_2$(dppz)]$^{2+}$

The racemic [Ru(phen)$_2$(dppz)]$^{2+}$ complex used in this study was synthesised by Professor D. Cardin, Hanna Beer and Kathrin Buchner. Λ- and Δ-[Ru(phen)$_2$(dppz)]$^{2+}$ complexes were synthesised by Kane McQuaid. DNA melting and pH titration experiments were measured by Sarah Gurung. SrCD spectra were recorded at Diamond Light Source Ltd. (beamline B23) by Sarah Gurung, Kane McQuaid, Dr. J. Hall and Dr. J. Brazier.

Chapter 4 – The Importance of Loop Composition on i-motif Formation

DNA melting experiments and pH titration measurements were performed by Sarah Gurung. SrCD measurements were taken at Diamond Light Source Ltd. (beamline B23) by Sarah Gurung, Kane McQuaid, Dr. J. Hall and Dr. J. Brazier.

Chapter 5 – The Effect of non-Cytosine Bases on i-motif Structure

Crystals were grown by Sarah Gurung. Data were collected in beamline I02 at Diamond Light Source Ltd. by Sarah Gurung, Dr. J. Hall, Prof. C. Cardin and Prof. T. Sorensen. Data for d(C$_3$T)$_4$ and [d(C$_3$A)$_4$+d(C$_3$T)$_4$] were processed by Dr. G. Winter. Studies on the structures of d(C$_4$)$_4$ and d(C$_3$A)$_4$ were made by Sarah Gurung, and d(C$_3$T)$_4$ and [d(C$_3$A)$_4$+d(C$_3$T)$_4$] by Sarah Gurung and Dr. J. Hall. (d(C$_3$A)$_4$ and [d(C$_3$A)$_4$+d(C$_3$T)$_4$] refinements ongoing).

Chapter 6 – Neutron Crystallographic Studies of d(C$_3$T)$_4$ i-motif

Crystals were grown by Sarah Gurung. Neutron data were collected at ManNDi at the Spallation Neutron Source (SNS), Oakridge National Laboratory by Sarah Gurung, Dr. J. Hall, Dr. S. Teixeira, Junhong Helen He and Dr. L. Coates. X-ray data for joint neutron + X-ray refinement were collected at SNS. Neutron data were processed by Dr. L. Coates. High-resolution d(C$_3$T)$_4$ data (from chapter 5) were collected at I02 at Diamond Light Source Ltd. and processed by Dr. G. Winter. The structures presented were solved by Sarah Gurung, Dr. S. Teixeira, Prof. C. Cardin and Dr. J. Hall.

Chapter 7 – Other i-motif Crystallisation Attempts

All i-motif systems were set up for crystal growth by Sarah Gurung at the University of Reading and the Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot. Bimolecular i-motif sequences were designed by Dr. J. Hall and Sarah Gurung. High-throughput srCD spectra were recorded at beamline B23, Diamond Light Source Ltd.

_____

Introduction

DNA (deoxyribonucleic acid) encodes the genetic information responsible for the functioning of most living organisms. The discovery of the DNA double helix in 1953[1] by Watson and Crick with the X-ray fibre diffraction data of Franklin and Wilkins immediately provided fundamental new insights into the nature of genetic events which revealed DNA as the molecule of life. RNA (ribonucleic acid), on the other hand, helps to synthesise, regulate, and process proteins; therefore it plays a fundamental role in performing functions within a cell. Transcription is the first step of gene expression which requires DNA to be copied into RNA as DNA carries the "code" for amino acid sequences. Studying the role of DNA in transcription has thus paved the way to understanding the processes of gene regulation, mutation/carcinogenesis and drug action. The advances in oligonucleotide research have led to the development of DNA binding anti-cancer drugs to prevent cell division and tumour growth.[2–4] In addition to medical research, DNA has also been a molecule of great interest in nanotechnology. Its physical and chemical properties have led to findings such as the organised self-assembly of nucleic acids for molecular nanofabrication[5] and a recent development involving molecular transport through large-diameter DNA nanopores.[6]

1.1 Native DNA Structure

DNA is a polymer made up of repeating nucleotides (Fig. 1.1A), each of which is composed of a phosphate group, a deoxypentose sugar and any one of the four distinct nucleobases (Fig. 1.1B). The nucleobases are classified as purines; adenine (A) and guanine (G), and pyrimidines; thymine (T) and cytosine (C). Individual nucleoside units are joined together as a nucleic acid chain through phosphate groups attached to the 3' and 5' positions of the sugars. The 5' and 3' labels are used to emphasise chain direction.
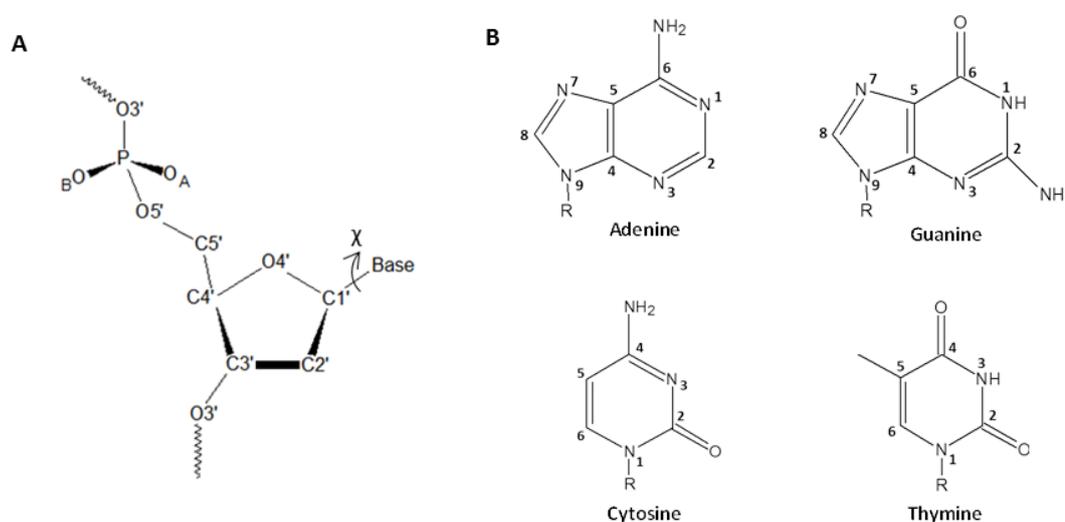
Figure 1.1: (**A**) The organization of repeating units in a polynucleotide chain. The rotable χ bond is indicated by a curved arrow. (**B**) The four bases of DNA.

_____

The bond between the sugar and the base is called the glycosidic bond. The torsion angle, χ (Fig. 1.1A), is subject to nucleotide type and the overall structure of the DNA. When χ ranges between 0° and 90°, the backbone is in *syn* conformation and if χ lies between -120° and -180°, then it is at *anti* conformation (Fig. 1.2).
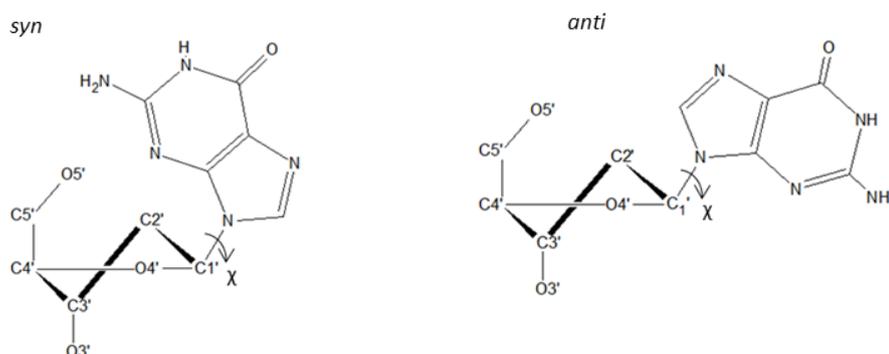
Figure 1.2: The *syn* and *anti* glyocosidic angle conformation for the deoxyguanosine nucleoside.

The 5-membered sugar ring is unable to support a planar geometry and, therefore, puckers out of the plane at certain positions. It is possible for the sugar pucker to differ from base to base if deformations in the structure are induced by external factors; such as the binding of ligands. Figure 1.3A shows the two major sugar pucker conformations; C2'-*endo* and C3'-*endo*. The pucker type is defined by the phase angle of pseudorotation, P and the maximum degree of pucker τ_m (Fig. 1.3B). P is between 140° and 185° in C2'-*endo* and between -10° and 40° in C3'-*endo* conformation.[7]
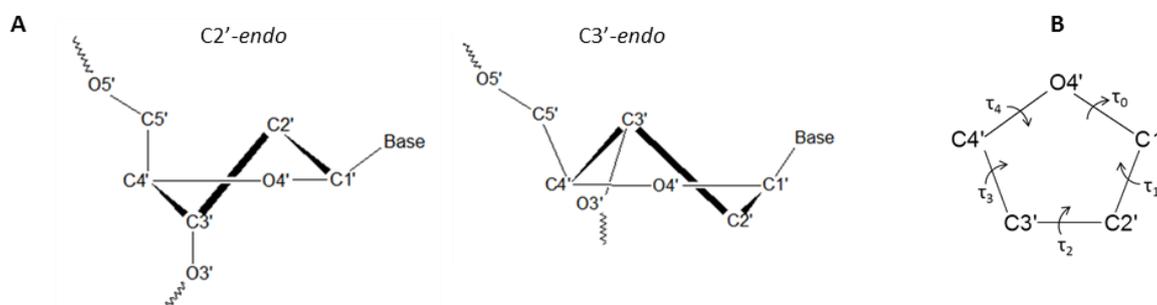
Figure 1.3: (**A**) Two major DNA sugar puckers. (**B**) The five internal torsion angles in a deoxyribose ring.

In 1952, Chargaff showed that the molar ratio of adenine was equal to thymine, and guanine was equal to cytosine.[8,9] This led to the proposal by Watson and Crick that the purine and pyrimidine bases are held together by specific hydrogen bonds to form planar base pairs. In its native form, DNA usually exists as a double helix of two antiparallel strands, and so the two bases of the base pair arise from two separate strands of DNA. The sugar groups are both attached to the bases on the same side of the base pair. The A·T base pair has two hydrogen bonds whereas the G·C base pair has three (Fig. 1.4A). In theory, the bases can pair up wherever the donor and acceptor groups meet but the Watson-Crick base pairs are the most common form as they bring maximum stability to DNA structures. However, paired mismatches are possible where the purine is flipped 180° about the glycosidic bond to give Hoogsteen base pairs[10,11] (Fig. 1.4B).
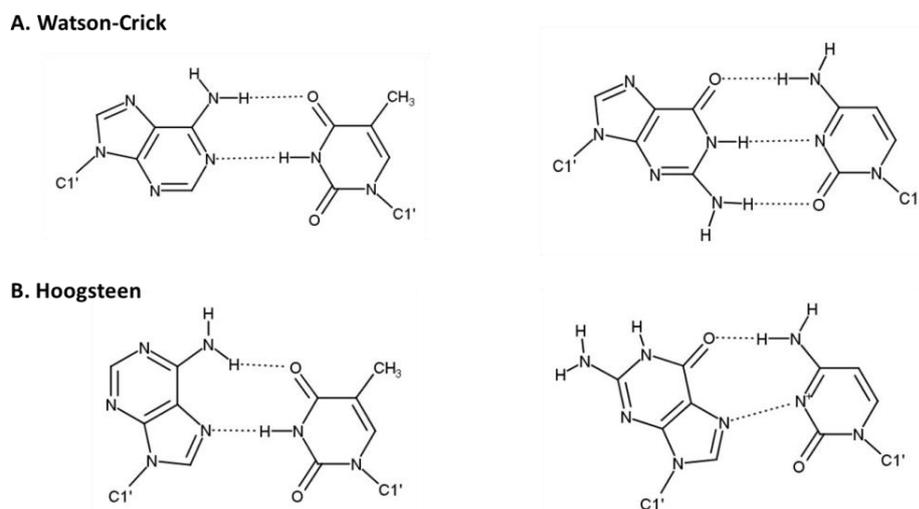
Figure 1.4: (**A**) Watson-Crick and (**B**) Hoogsteen hydrogen bonding in (left) A·T and (right) G·C base pairs.

1.1.1 Secondary DNA Structures

The three DNA double helix conformations originally determined from fibre diffraction were labelled A, B and Z-DNA (Fig. 1.5). B-DNA is the most generally accepted standard model for double helical DNA due to its dominance in biological cells.[12] The B-DNA form is 23.7Å wide and extends 34.0Å per ten base pairs of sequence. The double helix makes one complete turn about its axis every ten base pairs in solution. This frequency of twist, or helical pitch, depends largely on π–π stacking forces that each aromatic nucleobase exerts on its neighbour in the chain. The absolute configuration of the bases determines the direction of the helical curve for a given conformation. The DNA has two grooves (Fig. 1.5A) formed by the phosphate backbones which run the length of the polymer. In B-DNA, the major groove is 11.7Å wide and 8.8Å deep and the minor groove is 5.7Å wide and 7.5Å deep. The width of a groove is defined by the distance between a phosphorus and the phosphate perpendicular to it, minus the Van der Waal's radius of one phosphate (approximately 5.8Å).[7]

A-DNA and Z-DNA also form helical structures but differ significantly in their geometry and dimensions from B-DNA. Double stranded RNA adopts the A-form. Reversible transition of B to A-form has been shown to occur upon dehydration to protect the DNA from extreme desiccation of bacteria.[13] Segments of DNA that proteins methylate for transcription regulatory purposes may adopt the Z geometry,[14] in which the strands turn left-handed about the helical axis (Fig. 1.5B).[15] It is commonly believed to provide torsional strain relief (supercoiling) while DNA transcription occurs.[16] Table 1.1 summarises the differences between these three forms of DNA.
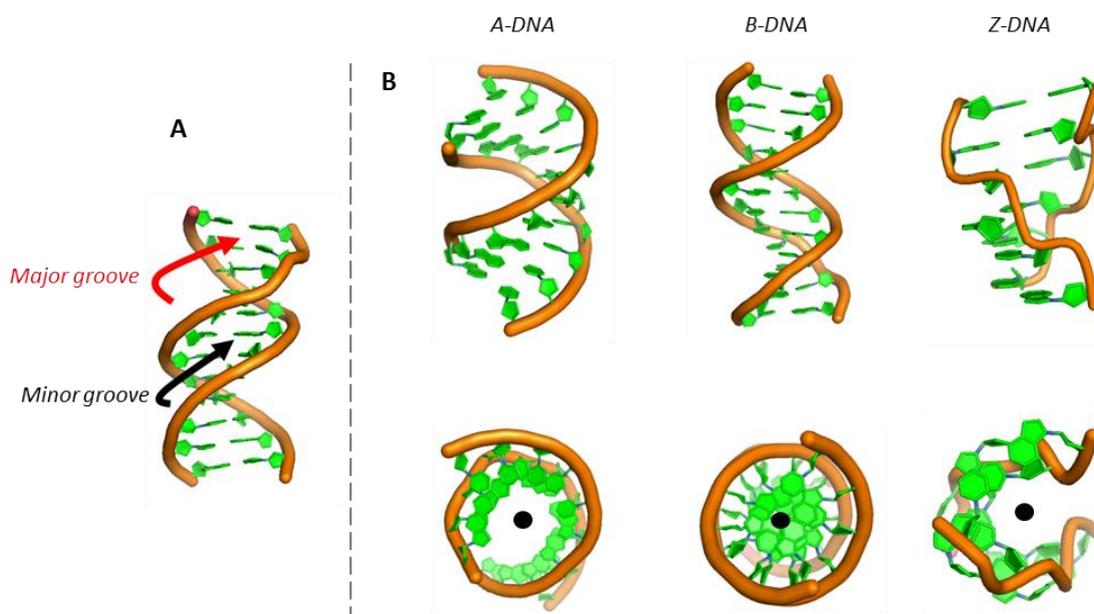
Figure 1.5: (**A**) The two grooves of a B-DNA double helix. (**B**) Views of the three conformations of the double helix DNA (top row) along and (bottom row) down the helical axes; highlighted as black dots. All nucleotides have been coloured in green and the backbone chains in orange.

Table 1.1: Average parameters of the three forms of DNA from single-crystal X-ray analyses.[17]

Geometry Attribute	A-DNA	B-DNA	Z-DNA
Helix sense	right-handed	right-handed	left-handed
Repeating helix unit (no. of base pairs)	1	1	2
Rotation per base pair (degrees)	33.6	35.9	-60.0
Base pairs per turn	10.7	10.0	12.0
Inclination of base pair to axis (degrees)	+19.0	−1.2	−9.0
Rise per base pair along axis (Å)	2.30	3.32	3.80
Pitch/turn of helix (Å)	24.6	33.2	45.6
Glycosyl angle	*anti*	*anti*	Cytosine: *anti*, Guanine: *syn*
Sugar pucker	C3'-*endo*	O1'-*endo* to C2'-*endo*	Cytosine: C2'-*endo*, Guanine: C2'-*exo* to C1'-*exo*
Diameter (Å)	23.0	23.7	18.0
Major groove width (Å)	13.0	11.7	7.6
Major groove depth (Å)	2.2	8.8	10.5
Minor groove width (Å)	11.1	5.7	4.8
Minor groove depth (Å)	2.6	7.5	7.9

Advances in DNA structural research have led to the development of DNA binding anti-cancer drugs. DNA-reactive ligands can be carcinogens as well as chemotherapeutic agents. There are three classes of DNA-reactive drugs; intercalating agents, DNA-degradative drugs and alkylating agents.[18] Nuclear DNA is the general target for all three classes of DNA-reactive drugs but cellular RNA and the A·T rich mitochondrial DNA may also be significant targets.[19] Mitochondrial DNA may

_____

be preferred for drugs that are A·T selective or drugs whose binding to DNA is restricted by nucleosomal structure.

Actively transcribed regions of DNA can also be targets for some DNA-reactive molecules. The potent carcinogenic fungal toxin aflatoxin B1 selectively reacts with such regions[20] and DNA strand cleavage agents such as bleomycin and neocarzinostatin, selectively cut within transcriptionally active regions.[21] Steroids that can induce transcription may also increase the selectivity of DNA-reactive drugs for these regions. The binding of drugs to DNA recognition motifs of DNA-binding proteins such as zinc-fingers[22] or helix-turn-helix[23] motifs would be expected to modulate transcriptional control. The therapeutic opportunities offered through DNA and associated targets are not limited to anticancer and antiviral diseases, but they also include genetic disorders that result in over or under-expression of gene products and autoimmune diseases.

1.2 Higher-order DNA Structures

DNA can adopt many different secondary structures as alternative to duplexes. Moreover, it has the ability to exist in a wide variety of forms such as hairpin, triple-stranded helices and four-stranded quadruplexes.[24] Repetitive DNA sequences have the potential to fold into such non-B DNA structures that may induce genetic instability and result in human diseases.[25] Hence, the molecular mechanism for their genetic instability has been investigated extensively. The best studied of these has been the G-quadruplex.

1.2.1 G-quadruplexes

G-rich sequences can fold into four-stranded structures known as the G-quadruplex (Fig. 1.6). DNA sequences which have the potential to form these structures are found in telomeric and gene promoter regions.[26–28] The G-quadruplex has been reported to be stable at physiological temperature, pH and ionic strength; it has been shown to exist in human cells.[29–31] Ligands binding to G-quadruplex have been used to inhibit telomerase and control the expression of oncogenes.[32] The presence of guanine rich sequences in genomic DNA means that there are always complementary cytosine-rich sequences. Such sequences can also form four-stranded structures but as intercalated motifs (i-motifs). However, much less is known about the prevalence of i-motif structures *in vivo*.

_____
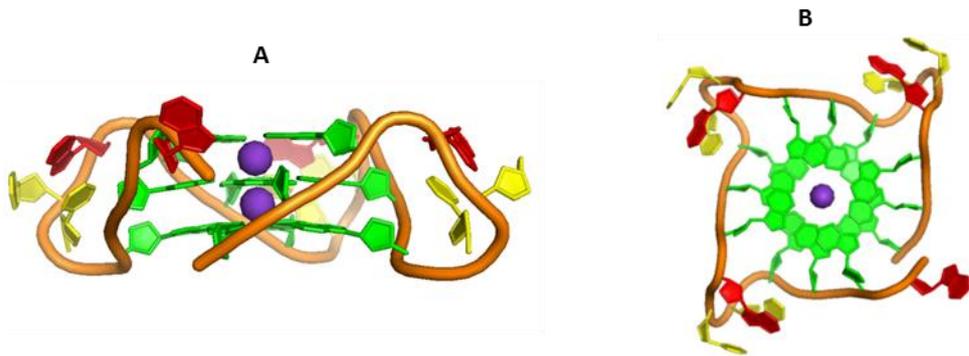
Figure 1.6: Crystal structure of the human telomeric G-quadruplex (**A**) perpendicular to and (**B**) down the helical axes (PDB ID: 1KF1).[33] Guanine, adenine and thymine nucleobases are green, red and yellow, respectively. The purple spheres denote potassium cations which help hold the G-quadruplex structure together.

1.2.2 i-motifs

Hemi-protonated cytosine base pairs had been identified in 1962[34] and formation of hairpins stabilised by cytosine$^+$·cytosine (C$^+$·C) base pairs had been proposed in 1992.[35] The identification of an intercalated tetramer was made by Gehring *et al.* in 1993 with the DNA sequence d(TCCCCC).[36] They discovered that cytosine rich sequences have the ability to form four-stranded i-motifs (Fig. 1.7).[36] Unlike the B-DNA or G-quadruplex which are held together by stacked base pairs, the i-motif is the only known DNA structure that consists of parallel-stranded duplexes held together through intercalated base pairs.

The formation of the i-motif requires protonation of cytosine bases to form C$^+$·C base pairs and *in vitro* studies show that i-motif structures are more stable at pH values lower than 7. Due to this, the hypothetical role of the i-motif in biological processes is uncertain and more effort has been put into the study of the G-quadruplex. There are still limited published investigations into the biological function of i-motif DNA and relatively few examples of i-motif binding ligands which could be used as probes in such investigations. However, several i-motif forming sequences in promoter regions of oncogenes have been found to be stable at neutral pH[37] and most recently, tracts of at least five cytosines have also been found to be stable at neutral pH, suggesting a possible biological function of the structure.[38] Stabilisation of i-motifs at pH 7 in molecular crowding conditions mimicking physiological conditions[39,40] and moreover, isolation of proteins that specifically bind to C-rich sequences[41] have all challenged the traditional assumption that i-motifs only form at acidic pH.

_____
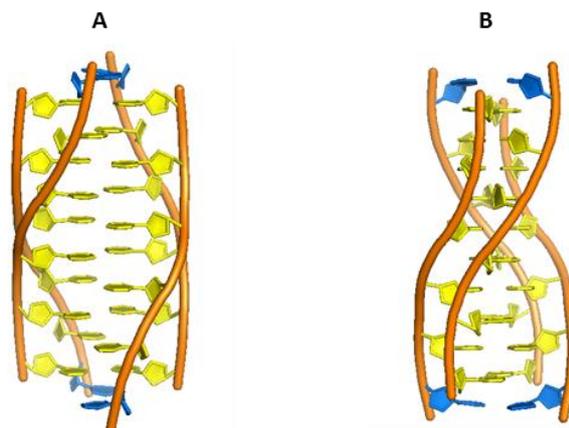
Figure 1.7: (**A**) Major and (**B**) minor groove views of the intermolecular i-motif d(TC$_5$) along the helical axis, as identified by Gehring *et al.* (PDB ID: 225D)[36] Cytosines are coloured yellow and thymines are coloured blue.

I-motifs have recently been an attractive subject for DNA nanotechnology,[42–44] with the structure being the first ever DNA molecular motor known to be driven by pH changes.[45] By altering pH, these sequences can reversibly fold and unfold to help power nanomachines.[33] They have also been used in applications such as the assembly of gold nanoparticles,[47,48] switches for logic operations[49–51] and as sensors to map pH changes in living cells.[52,53]

1.3 The DNA i-motif Structure

The building block of the i-motif structure is a base pair involving one neutral cytosine and one protonated cytosine at N3 (the C$^+$·C base pair) bonded by three hydrogen bonds (Fig. 1.8A). Based on NMR spectroscopy and theoretical calculations using the human telomeric i-motif forming sequence;[54] d[(C$_3$TA$_2$)$_3$C$_3$], as a model, it has been proposed that this hemi-protonated bond may be described as hydrogen bonds with asymmetric double-well potentials rather than a symmetric hydrogen bond with a single-well potential.[55] The formation of these hydrogen bonds produces a stronger base-pair interaction than the canonical G·C base pair.[56] Hence, the base-pairing energy (BPE) for C$^+$·C is 169.7 kJ mol^{-1}, whereas the BPEs of the canonical Watson–Crick G·C and neutral C·C base pairs are only 96.6 and 68.0 kJ mol^{-1}, respectively.
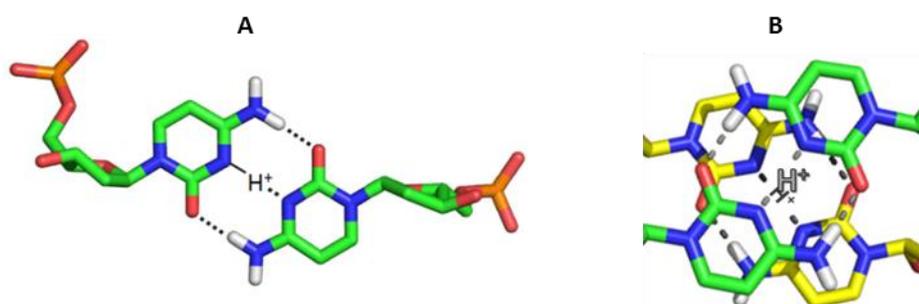
Figure 1.8: (**A**) A hemi-protonated cytosine$^+$–cytosine (C$^+$·C) base pair. (**B**) View of the C$^+$·C base pair down the helical axis. Nitrogen, oxygen, hydrogen and phosphorus have been coloured in blue, red, white and orange, respectively. Carbon is coloured in both green and yellow.

_____

There are two different forms of i-motifs; inter- and intramolecular. The latter involves cytosine tracts present in a single DNA strand. DNA bases that are not involved in cytosine intercalation are called "loop" regions. Intermolecular i-motifs can be further categorised into tetramolecular; which involves four independent DNA strands, and bimolecular; which is made up of two strands. In any case, the core consists of cytosine tracts that are arranged spatially as a tetramer composed of two parallel-stranded duplexes that are interspersed in an anti-parallel way (Fig. 1.9).
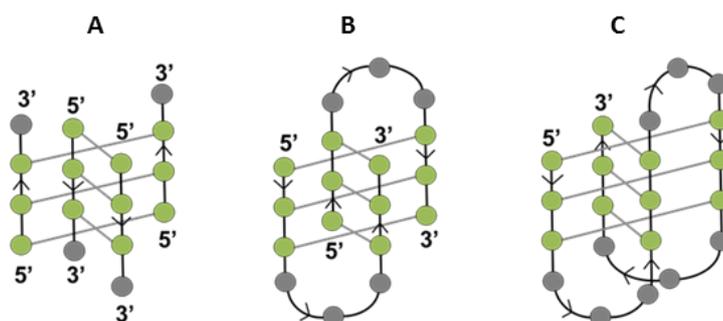
Figure 1.9: Schematic diagrams of (**A**) tetramolecular, (**B**) bimolecular and (**C**) intramolecular i-motif structures. Cytosine residues are green dots and DNA bases in the "loop" regions are grey.

The $C^+\cdot C$ distance, between two N3 atoms, is 3.1Å in comparison to 3.4Å between C-C base pairs in B-DNA. The helical twist between adjacent $C^+\cdot C$ pairs (12–16°) is also smaller than in the case of B-DNA.[57] This spatial arrangement produces the existence of two broad and flat major grooves and two extremely narrow minor grooves. This has led to the classification of i-motif structures into two groups: those with the terminal $C^+\cdot C$ base pair at the 3' end (3'E intercalation topology), and those with the terminal base pair at the 5' end (5'E intercalation topology) (Fig. 1.10).[54,58]

_____
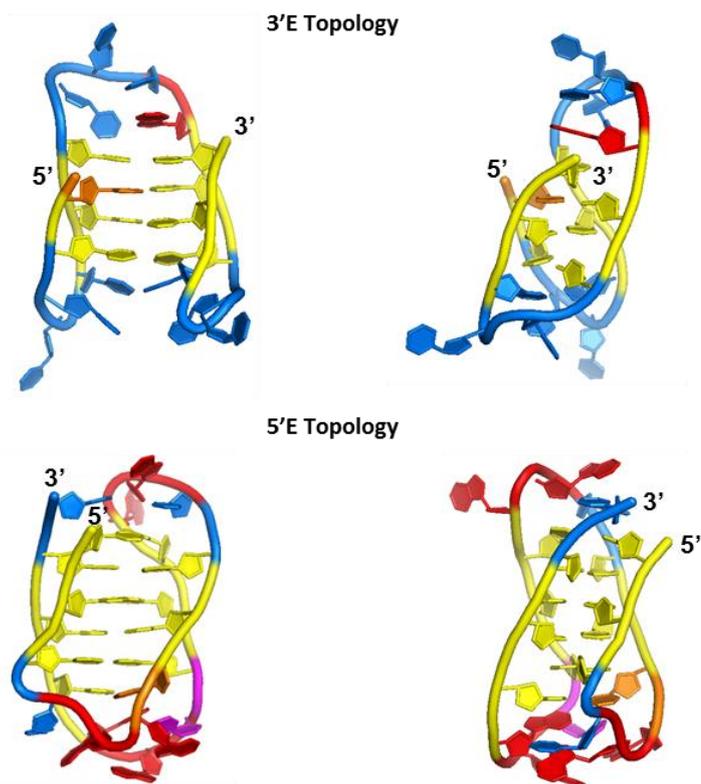
Figure 1.10: NMR structures of (top) the 3'E i-motif sequence d(5mCCT$_3$CCT$_3$ACCT$_3$CC)[59] (PDB ID: 1A83) and (bottom) the 5'E i-motif sequence d(CCCTA$_2$5mCCCTA$_2$CCCUA$_2$CCCT)[54] (PDB ID: 1EL2). Cytosines are coloured yellow, thymines are in blue, adenines in red, 5-methyl-cytosine in orange and uracil in purple.

1.3.1 Stability of the i-motif Structure in vitro

The relative stability of nucleic acid structures is often measured in terms of their melting temperature; T_m. T_m is defined as the temperature at which half of the DNA is denatured to the single-stranded state; induced by heating of the DNA sample. The transition is usually monitored using molecular absorption or circular dichroism (CD) spectroscopies. As with other nucleic acid structures, the stability of the i-motif depends on several factors such as the nucleotide sequence, ionic strength or temperature amongst many. Because the protonation of one of the cytosine bases in the C$^+$·C base pairs is an absolute requirement for the formation of the i-motif structure, the pH of the medium plays a crucial role.

As the pKa value of cytosine is around 4.6 (in pure water at 25°C), it would be expected that the formation of i-motif structures occurs at pH values lower than around 6.6.[60,61] At pH values from 4 to 7 and at 25°C, the cytosine bases are partially protonated and the DNA folds into the closed i-motif structure. In this pH range, the stability of the i-motif is a linear function of pH.[60,62] The highest stability of the i-motif structures occurs at pH values close to the pKa of cytosine. At higher pH values, the cytosine bases can deprotonate and the structure may unfold to a single-stranded form. On the other hand, if the pH value is too low (approximately below 3) all the cytosine bases are protonated

_____

and they cannot form the hydrogen bond pattern needed for the C[+]·C base pair.[62] G·C and A·T Watson–Crick base pairs are pH-independently stable in a broader pH range from 2 to 10 units. At 37°C and 150 mM NaCl, the i-motif was reported to not form at the physiological pH of 7.4 but, depending on the sequence, it may form at the ischemic pH of 6.7.[63] However, it has recently been published that intramolecular i-motif formation has been observed at neutral[38,64] or even slightly basic pH values.[65]

The folding of i-motif structures at physiological pH has also been reported in the presence of silver cations.[66] The formation of the i-motif is strongly enthalpy driven at near-neutral pH; with maximum formation at pH values of 6–7, and structure degradation below pH 5.5.[67] Mathur *et al.* studied the thermodynamic parameters associated with the formation of a 31-oligomer i-motif belonging to the c-MYC oncogene.[68] At 20°C, a ΔG of -10.4 ± 0.1 kcal mol^{-1} was observed with favourable enthalpy and unfavourable entropy at pH 5.3 in 20 mM NaCl for i-motif folding. It is possible to determine the contribution of a C[+]·C base pair to the thermodynamic parameters. Using sequences that form both intra- and intermolecular i-motif structures, Völker *et al.* determined that the i-motif enthalpy and entropy values were larger than those corresponding to the disruption of base pairs in the Watson–Crick duplex.[61]

Apart from pH, a fundamental factor affecting the structure and stability of the i-motif structure is the number of C[+]·C base pairs present in the folded structure. For a given pH value, the stability of two i-motifs differing in the number of cytosine bases in their sequences would be higher for the one showing a higher number of C[+]·C base pairs.[69] I-motif forming sequences with tracts of at least five cytosines have recently been found to be stable at neutral pH.[38] The intra versus intermolecular folding may also depend on the length of the cytosine tracts.[44] Two stretches less than seven cytosine residues favour the intermolecular form, whereas longer cytosine tracts promote the formation of intramolecular i-motif structures with high thermal stability. DNA bases other than cytosine have also been found accommodated into the i-motif core. Thymidines can form symmetrical T·T base pairs that are nearly isomorphic with C[+]·C base pairs. The NMR study of d(5mC$_2$TCTCTC$_2$)$_4$ showed that the thymine pairs at T5 positions are intercalated into the i-motif core.[70] This is in contrast to a similar sequence, d(5mCCTCACTCC)$_4$, which forms a dimer[71] where the i-motif is built by intercalation of two symmetrical hairpins held together. It can be said that the formation of T·T base pairs, in general, stabilises the i-motif structure because of extension of the i-motif core and decrease in flexibility.[72]

The length and composition of loops is another key factor in i-motif stability. The length of the loops has been shown to be critical in determining the multimeric nature of the i-motif structure. C-rich sequences with just one base in the loops tend to form mixtures of tetra- and bimolecular i-motifs,[44] whereas longer loops favour the formation of intramolecular i-motifs.[67] Intramolecular i-motif structures were classified into two groups depending on the length of the loops.[73] Class I

_____

consists of the loop sizes 2:3/4:2 with four, five or six $C^+ \cdot C$ base pairs. Class II has loop sizes 6/8:2/5:6/7 (Fig. 1.11). The midpoint of the pH-induced folding from the neutral strand for class II i-motif structures is slightly higher at pH 6.6 than that for i-motif structures identified as class I (pH 5.8–6.4). This was reasoned as being due to the presence of stabilising effects in longer loops. An i-motif showing a 12-base long loop is able to form a hairpin stabilised by Watson–Crick base pairs.[74]
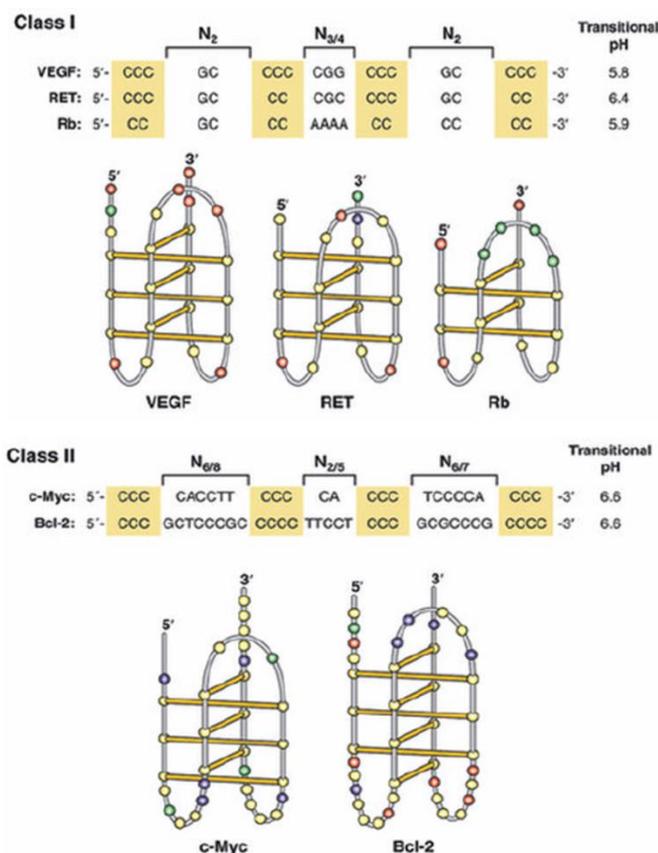
Figure 1.11: The two classes of intramolecular i-motifs. Reprinted with permission from Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **277,** 3459–3469 (2011). Copyright 2017 John Wiley and Sons.

Using NMR methods, it has been shown that protonation of adenine bases in the loop regions, when the pH is lower than 4.6, produces a conformational change involving disruption of the i-motif core.[75] The sequence studied was a mutant fragment of the human centromeric satellite III DNA; $d[C_2AT_2C_2AT_2C_2T_3C_2]$, which folds into an intramolecular i-motif structure. Protonation of adenine bases hinders the formation of an A·T base pair that extends the $C^+ \cdot C$ core at pH higher than 4.6. It has also been observed that adenine containing i-motif structures unfold at lower temperatures.[67] Thymine base pairs can occur in the loops connecting the cytosine tracts.[59] Loop interactions were of particular relevance in the dimeric structure of $d(TC_3GT_3C_2A)$[76] and $d(TCGT_3CGT_2)$.[64] Both cases showed formation of intermolecular G·T base pairs that interact to each other through their minor groove side, forming a G:T:G:T tetrad. The bases at the 5' and 3' ends of C-rich sequences that are able to form i-motif structures also have an influence on the stability of the i-motif structures. The study

_____

of the d[$C_3TA_2C_3$] and d[$C_3TA_2C_3TA_2$] sequences by DNA melting experiments suggests that the inclusion of the additional TAA segment at the 3' end produces a change in the structure. Hence, the 9-mer folds in a bimolecular structure, whereas the 12-mer sequence exists in two (bimolecular and tetramolecular) forms.[77]

The change in arrangement of neighbouring bases may also influence the stability of the i-motif structure. It was shown that the d[A_2C_4]$_4$ and d[C_4A_2]$_4$ folded into different i-motif structures.[78] The first sequence adopts a unique structure showing A·A pairs stacked to C^+·C base pairs whereas the second sequence adopts two distinct intercalation topologies (3'E and 5'E) showing well stacked adenosine bases to the adjacent C^+·C base pairs. Using X-ray crystallography, it was shown that the i-motif formed by the d(AC_3T)$_4$ sequence was further stabilized at one end by a three-base hydrogen-bonding network, in which two adenines and a thymine form four hydrogen bonds via a reverse Hoogsteen and an asymmetric A·A base pairing.[79]

As with many other nucleic acid structures, the ionic strength of the solution has an influence on the stability of the i-motif structure. Nitrogenous bases are neutral throughout a wide pH range, from 5 to 9, approximately. The same occurs with the pentose sugar, which can only lose the proton at very alkaline pH values. However, at neutral pH values, DNA is a polyanion because of the negatively charged phosphate groups. The presence of appropriate cations, such as sodium and potassium balances the negative charge. Folded conformations may produce changes in the cation atmosphere surrounding the phosphate backbone. Moreover, in the case of the i-motif, protonation at N3 produces the appearance of a positive charge at the base pairs. Therefore, variations in ionic strength can cause conformational changes and differences in the relative stability of i-motif structures. In general, it has been observed that an increase in ionic strength from 0 to 100 mM NaCl produces a destabilization of the i-motif structure.[67] The transition between single-stranded DNA and i-motif conformations is induced by a change in pH or ionic strength. Apart from the canonical i-motif structure formed at acidic pH values, the formation of partially folded structures at neutral and slightly acidic pH values (6-7) has been proposed. It has also been suggested that the partially folded species coexist with the single-stranded structure at neutral pH and room temperature.[80]

_____

1.3.2 Ligands Binding to i-motifs

The increase in interest of G-quadruplexes as therapeutic targets and the discovery of proteins specifically binding to C-rich sequences have prompted studies on i-motif stabilisation using various ligands. Moreover, the finding of three principal i-motif forming sequences; all in promoter regions, have contributed to further research on i-motif stability. These three i-motifs are, namely, the telomeric i-motif, the c-MYC promoter i-motif and the Bcl-2 promoter i-motif.

The human telomeric DNA is rich in cytosine and guanine. 85% of cancers have shown increase in telomerase activity.[81,82] Stabilisation of human telomeric G-quadruplex with a range of ligands inhibits this activity.[83] The c-MYC oncogene consists of an i-motif forming sequence in the nuclease hypersensitive element (NHE) III1 region.[84] The c-MYC gene codes for a transcription factor with a wide range of functions, including regulation of cell cycle progression and cell growth.[85] The P1 promoter region of the B-cell lymphoma-2 (Bcl-2) oncogene also consists of an i-motif forming sequence. Bcl-2 is over-expressed in some cancers but may be under-expressed in neurodegenerative diseases such as Alzheimer's and Parkinson's.[86] The Bcl-2 transcription factor hnRNP LL has recently been reported as an i-motif binding protein,[87] suggesting a biological role of the i-motif as a protein recognition site for the activation of transcription in B-cl2.

In 2000, Hurley *et. al.* reported the binding of a cationic porphyrin, TMPyP4 (Fig. 1.12), to the human telomeric sequences d(CCCAAT)$_4$ and d(AATCCC)$_4$.[88] Although there was no significant change in the i-motif melting temperature on the addition of the ligand, modelling experiments using NMR structure suggested that TMPyP4 binds to the i-motif via stacking interactions at the ends of the structure (Fig. 1.13). TMPyP4 has shown downregulated expression of c-MYC with the G-quadruplex,[89] as well as inhibition of telomerase and tumour growth.[90] Similarly, Alberti *et al.* studied the interaction of a DNA intercalating cyclic acridine dimer; BisA, with the human telomeric G-quadruplex and i-motif.[91] They found that BisA caused a significant stabilisation of i-motif at pH 6.8 with an increase in melting temperature of 33°C. BisA also stabilises the G-quadruplex and inhibits telomerase *in vitro*. These studies thus prompted an interest in i-motif stabilisation in order to investigate potential biological functions of the i-motif DNA.

Carboxyl-modified single-walled carbon nanotubes (SWNTs) can inhibit duplex formation and induce human telomeric i-motif formation, even at pH 8. Stabilisation of the human telomeric i-motif by SWNTs has been reported to inhibit telomerase, resulting in telomere uncapping, DNA damage response and apoptosis.[92] It has been proposed that the nanotubes bind to the 5'-end in the major groove of the i-motif and under molecular crowding conditions, they can also induce i-motif formation at physiological pH.[93] Ruthenium polypyridyl complexes; [Ru(bpy)$_2$(dppz)]$^{2+}$ and [Ru(phen)$_2$(dppz)]$^{2+}$ were also shown to bind to i-motif DNA.[94] Other ligands which have been shown to interact with the i-motif most recently include mitoxantrone[95] and thiazole orange[96] (Fig. 1.12).

_____

Crystal violet[97] has been proposed to interact with i-motif on the ends of the DNA structure (Figures 1.13B and 1.13C).
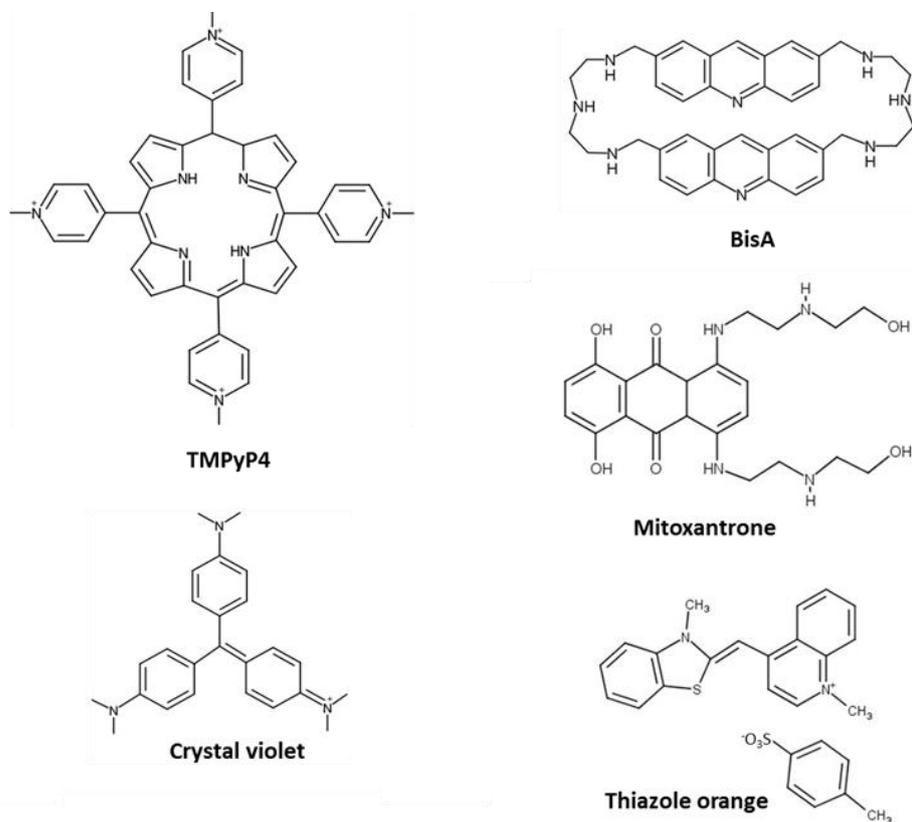
Figure 1.12: Structures of some known i-motif binding ligands.
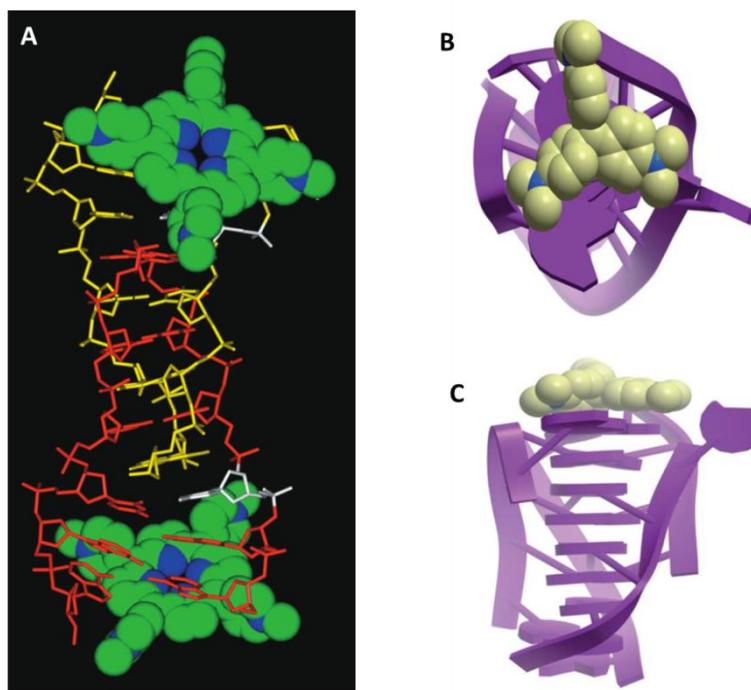
Figure 1.13: (**A**) NMR-based model of the 2:1 TMPyP4-d(AACCCC)$_4$ complex. One C$^+$-C parallel duplex is coloured yellow and the other red. Cytosine residues at C3 positions of two chains are white. TMPyP4 molecules are shown in the CPK representation, where carbon is coloured green and nitrogen in blue. [Reprinted with permission of American Chemical Society. Oleg Yu. Fedoroff et al., Cationic Porphyrins Promote the Formation of i-Motif DNA and Bind Peripherally by a Nonintercalative Mechanism, *Biochemistry*, **39** (49), pp 15083–15090 (2000)]. Molecular modelling diagrams showing views (**B**) down and (**C**) along the helical axes of the interaction of Crystal violet (CV) with an i-motif [5'-(ACCCT)$_4$-3']. CV is represented as a space-filling model showing carbon (beige) and nitrogen (blue) atoms. The i-motif is represented in ribbon form (purple). [Reprinted with permission of Royal Society of Chemistry. Dik-Lung Ma et al., Crystal violet as a fluorescent switch-on probe for i-motif: label-free DNA-based logic gate, *Analyst*, **136** (13), pp 2692-2696 (2011)]

1.4 Biophysical Methods Applied to Nucleic Acids

There are different approaches to study the formation and stability of DNA structures, including i-motifs. The first studies that demonstrated the formation of the tetrameric i-motif structure were made using Nuclear Magnetic Resonance (NMR) measurements.[60] However, the assignment of NMR spectra on more complex and relevant intramolecular oligonucleotide sequences is much more challenging. Apart from NMR, one of the first techniques used to elucidate the i-motif structure was X-ray diffraction. Ultraviolet (UV) molecular absorption spectroscopy is another useful technique to observe the changes in protonation and stacking of the nitrogenated bases involved in the formation of the i-motif. I-motif structure formation can also be detected by Circular Dichroism (CD) spectroscopy, where the spectrum only needs to show two characteristic bands to differentiate it from secondary DNA or other higher order DNA structures.

1.4.1 UV Molecular Absorption Spectroscopy

UV spectroscopy refers to absorption or reflectance spectroscopy using UV light. The atomic orbitals of the atoms involved in a bond merge to form molecular orbitals which can be occupied by electrons of different energy levels. Ground state molecular orbitals can be excited to anti-bonding molecular orbitals. Hence, absorption measures transitions from the ground state to the excited state.[98] The electrons in a molecule can be of one of three types: σ (single bond), π (multiple-bond), or n (non-bonding, usually caused by lone pairs). These electrons, together with energy in the form of light radiation, get excited from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO) and the resulting species is known as the excited state or anti-bonding state. σ-bond electrons have the lowest energy level and are the most stable electrons. They require energy to be displaced to higher energy levels. As a result, the electrons generally absorb light in the shorter wavelengths (λ/nm) of the UV light.

π-bond electrons have much higher energy levels from the ground state. These electrons are therefore relatively unstable and can be excited more easily and would require less energy for excitation. The electrons absorb energy in the UV region. Finally, n-electrons are of higher energy levels than π-electrons and can be excited by both UV and visible light. Low energy gap between the HOMO and the LUMO results in more easily excited electrons so that a sample can absorb longer wavelengths of light. Most of the absorption in UV spectroscopy occurs due to π-electron transitions or n-electron transitions. Based on the four types of transitions that can occur, the energy required for various transitions obey the following order σ-σ* > n-σ* > π-π* > n-π* (Fig.1.14).

Conjugation of π-electrons affects their energy levels. When two double bonds are conjugated, the electrons in them create four molecular orbitals; two bonding and two anti-bonding. As a result of this, the HOMO is at a higher energy state and the LUMO is at a lower energy state. In order to excite this system, the energy that would be required to excite the electrons from the HOMO to the LUMO would therefore be reduced. As a result of this reduction in energy levels, the wavelength for absorption of conjugated molecules increases.
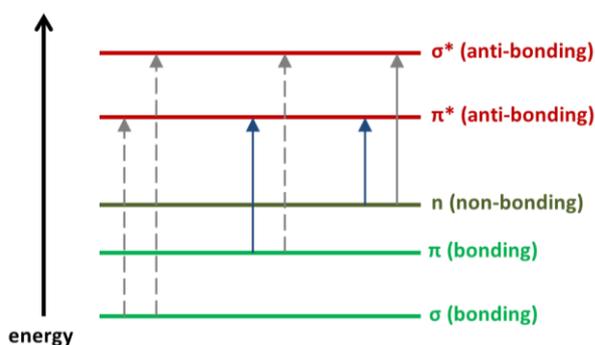
Figure 1.14: Energy level diagram showing the possible electron transitions that light causes. The blue arrows show jumps from π-π* and n-π*orbitals. The grey arrow shows the transition from n-σ* orbitals whereas the dotted arrows show jumps which absorb light outside the region of the UV-Vis spectrum.

_____

The π-conjugated systems in DNA can thus be excited using UV light. Nucleobases absorb well at 260 nm and this unique feature can be used to determine the concentrations and purity of DNA (Fig. 1.15) by using the Beer-Lambert law: $A = \varepsilon c l$, where ε is the molar absorptivity or extinction coefficient. This constant is a fundamental molecular property in a given solvent and has units of mol^{-1} dm^3 cm^{-1}, c is the concentration (mol dm^{-3}) of the absorbing species and L is the path length (cm) through the sample.
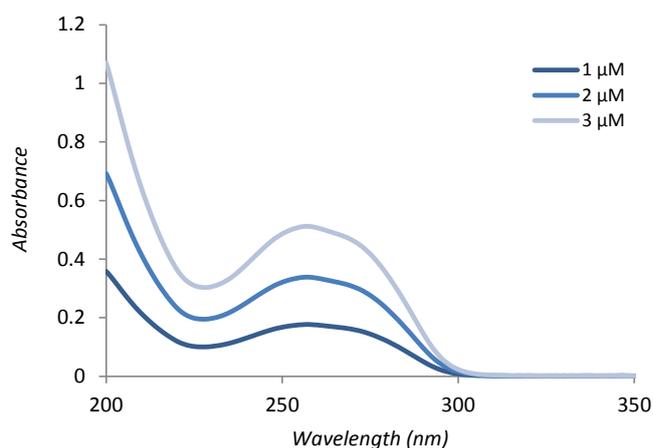
Figure 1.15: UV spectra showing increase in absorbance of a B-DNA; d(TCGGCGCCGA), with increasing single-stranded concentration.

1.4.2 Circular Dichroism Spectroscopy

Electromagnetic (EM) radiation consists of an electric and a magnetic field that oscillate perpendicular to one another towards the propagating direction. Linearly polarised light occurs when the electric field vector oscillates in only one plane whereas circularly polarised light occurs when the direction of the electric field vector rotates about its propagation direction with constant magnitude (Fig. 1.16). Circular dichroism (CD) is the difference in absorption of left-handed circularly polarised light (LCP) and right-handed circularly polarised light (RCP), and it occurs when a molecule contains one or more chiral light absorbing groups. So, CD is: $\Delta A = A(LCP) - A(RCP)$, where absorbance (A) is a function of wavelength, λ. This can also be expressed by applying the Beer-Lambert law as: $\Delta A = (\varepsilon_L - \varepsilon_R)cL$, where ε_L and ε_R are the molar extinction coefficients for LCP and RCP light. Then $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$ is the molar circular dichroism. Δε is also a function of wavelength and is referred to as the intrinsic property of a substance.

CD only occurs at wavelengths of light that can be absorbed by chiral molecules, including proteins, DNA and RNA, therefore, it can be used to differentiate the structures of biological molecules. Figure 1.17 illustrates the use of CD spectra to distinguish between A-, B- and Z-DNA. Measurements carried out in the visible and UV region of the EM spectrum monitor electronic transitions.

_____
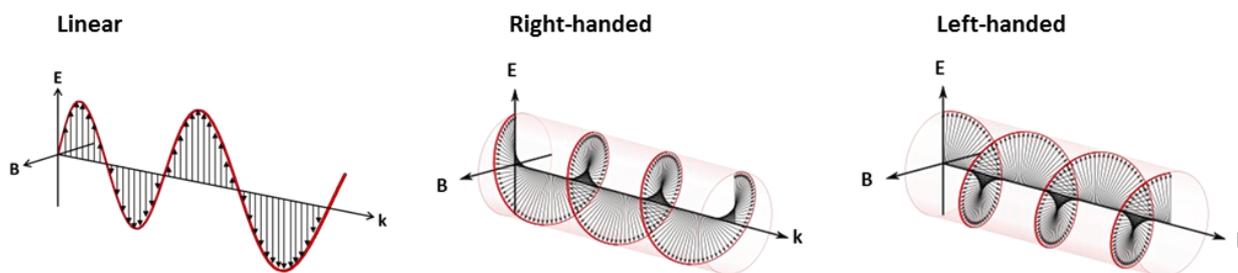
Figure 1.16: Linearly polarised light, right- and left-handed circularly polarised light. The E and B axes are the electric and magnetic fields, respectively. The k axis represents the direction of propagation. Images of right and left-handed circularly polarised light by Dave3457, distributed under a CC-BY 2.0 license.[99]
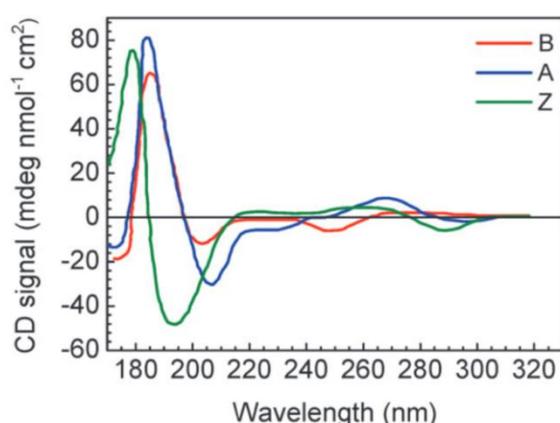
Figure 1.17: CD spectra of A-, B- and Z-DNA. Reprinted with permission from Holm, A. I. S., Nielsen, L. M., Hoffmann, S. V. & Nielsen, S. B. Vacuum-ultraviolet circular dichroism spectroscopy of DNA: a valuable tool to elucidate topology and electronic coupling in DNA. *Phys. Chem. Chem. Phys.* **12**, 9581-9596 (2010). Copyright 2017 Royal Society of Chemistry.

1.4.3 Crystallography

The goal of crystallisation is to produce a well-ordered crystal to obtain a diffraction pattern when exposed to X-rays, neutrons or electrons. This diffraction pattern can then be analysed to determine the arrangement of atoms in a crystalline solid, including that of DNA. DNA crystallisation is usually carried out in aqueous solution. The likelihood of crystallisation of a DNA is influenced by many factors, such as sample purity, pH, DNA concentration, temperature, precipitants and additives. The more homogeneous the DNA solution is, the more likely that it will crystallise. Crystallisation occurs when a metastable supersaturated solution reaches a stable lower energy state by reduction of solute concentration. There are two major stages in crystallisation: nucleation and crystal growth. Nucleation is when the solute molecules or atoms dispersed in the solvent start to gather into stable clusters of a critical size which constitute the nuclei. This size is dictated by many different factors including temperature and supersaturation. It is during the stage of nucleation when atoms or molecules arrange themselves in a manner that defines the crystal structure. Crystal growth is then

_____

the subsequent increase in volume of the nuclei after reaching the critical cluster size. It is a dynamic process occurring in equilibrium where solute molecules or atoms precipitate out of solution and dissolve back into the solution.

Vapour diffusion is the most common method of DNA crystallisation. In this method, a droplet containing purified DNA, buffer, and precipitant are allowed to equilibrate with a larger reservoir containing the same or similar buffer and precipitants but in higher concentrations. Initially, the droplet of DNA solution contains comparatively low DNA and precipitant concentrations, but as the drop and reservoir equilibrate, concentrations in the drop increase. If the appropriate crystallisation solutions are used for a given DNA, crystal growth will occur in the drop.[100,101] Vapour diffusion can be carried out in either hanging drop or sitting drop format (Fig. 1.18). Hanging drop involves a drop of DNA solution to be placed on an inverted cover slip, which is then suspended above the reservoir. Sitting drop crystallization requires the drop to be placed on a pedestal that is separated from the reservoir. Both of these methods require sealing of the environment so that equilibration between the drop and reservoir can occur.
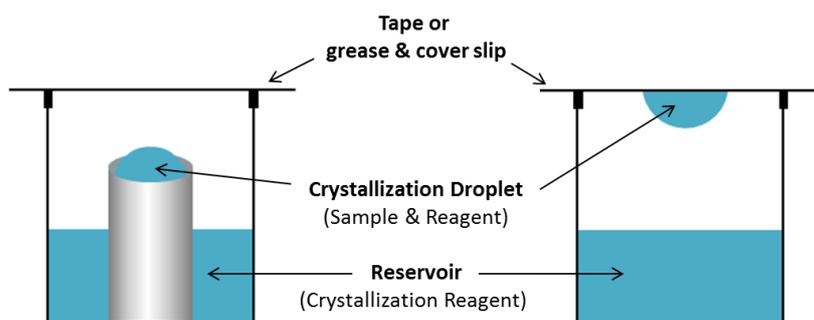
Figure 1.18: (Left) Sitting and (right) hanging drop vapour diffusion methods.

Crystallographic methods depend on beam type. X-rays are most commonly used; they principally interact with the spatial distribution of electrons in the sample. Neutrons are scattered by the atomic nuclei through strong nuclear forces. Moreover, the magnetic moment of neutrons is non-zero. They are, therefore, also scattered by magnetic fields. In many ways the principles of neutron and X-ray diffraction are similar (Fig. 1.19). The results of the two methods are frequently complementary.[102]

_____

Select a crystal and mount for X-ray/neutron study
1
Obtain unit cell geometry and preliminary symmetry
2
Measure intensity data
3
Data reduction (corrections applied to data)
4
Solve the structure: Patterson, Direct or other methods
5
Find the atoms: Fourier and difference Fourier syntheses
6
Refine the structure model
7
Interpret the results
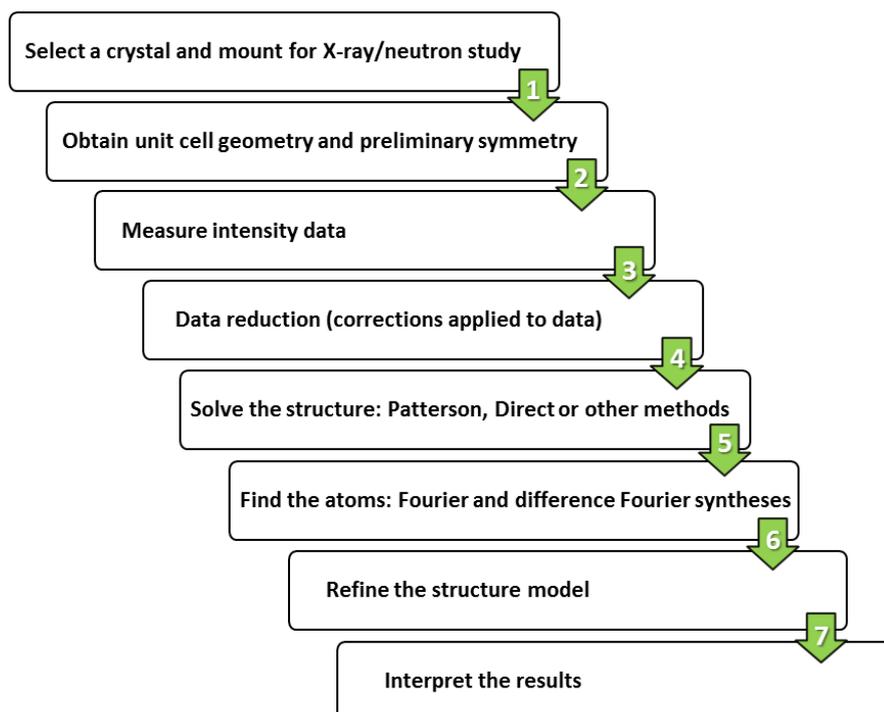
Figure 1.19: Steps involved in crystal structure determination.

(a) X-ray Diffraction

Single-crystal X-ray crystallography (XC) is one of the main methods used to determine the molecular structure of biological macromolecules.[103] A crystal is mounted on a goniometer; an instrument which can be rotated at a given angle while the crystal is hit with X-rays. Diffraction patterns[104] of regularly spaced spots are recorded on a detector (Fig. 1.20). As X-rays are a collection of separate waves, each wave can interfere with one another either constructively or destructively. Constructive interference occurs when the waves are moving in phase (ϕ) with each other (Fig. 1.21). Destructive interference occurs when the waves are out of phase and thus reduces the amplitude, $|F|$, of the resultant wave. Constructive interference results in diffraction patterns.

_____
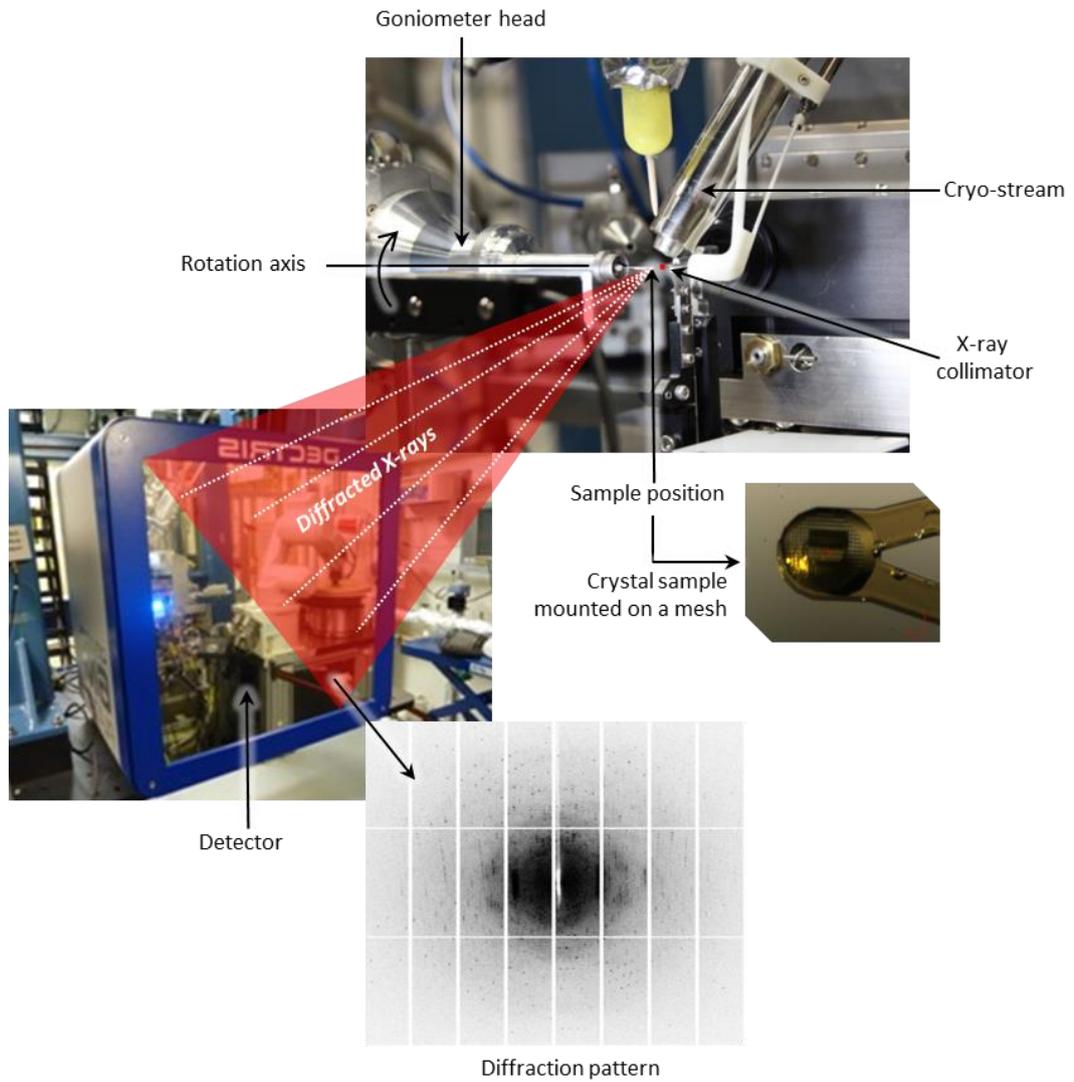
Figure 1.20: Single crystal X-ray diffraction; the end station of beamline I02 at Diamond Light Source Ltd. The crystal and diffraction pattern pictured belong to the d(CCCT)$_4$ i-motif.
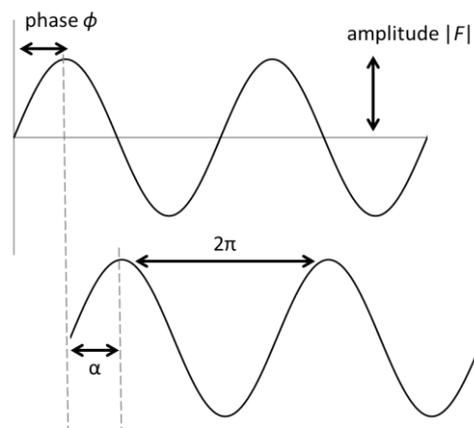
Figure 1.21: Constructive interference between two waves defines the phase angle of α, and a phase shift of 2π.

_____

After the experimental demonstration that X-rays could be diffracted by crystals, Laurence Bragg derived the Bragg equation (Fig. 1.22) which is universally used as the basis for X-ray diffraction geometry. He showed that diffraction that can be produced by an appropriate orientation of a crystal in an X-ray beam can be regarded geometrically as if it were a reflection from sets of parallel planes passing through lattice points. The angles of incidence and reflection must be equal. The incoming and outgoing beams and the normal to the reflecting planes must all lie in one plane. To define a plane, three integers; *hkl* indices, are needed to specify its orientation with respect to three unit cell edges. The overall relationship can thus be expressed as *λ = 2d$_{hkl}$sinθ*, where *d* is the distance between the planes.
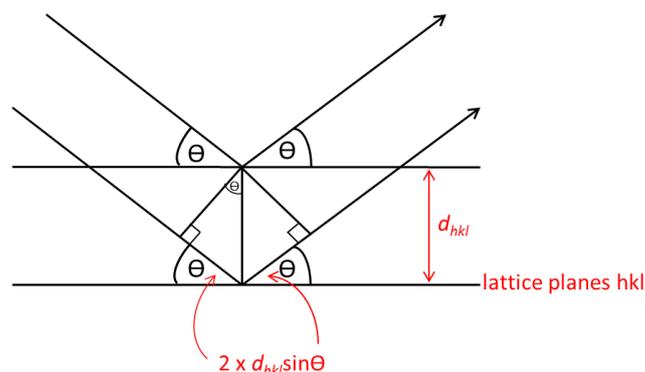
Figure 1.22: The Bragg construction for diffraction by a three-dimensional crystal structure.

In the diffraction experiment, the intensities of waves scattered from planes in the crystal are measured. The amplitude of the wave |*F$_{hkl}$*| is proportional to the square root of the intensity measured on the detector. The only relationship between the amplitudes and phases is through the molecular structure or electron density. The diffraction pattern is the forward Fourier transform of the electron density. In mathematics, this can be defined as:

$$p(xyz) = \frac{1}{V}\sum_{h,k,l} |F(hkl)| \cdot \exp[i\alpha(hkl)] \cdot \exp[-2\pi i(hk + ky + lz)]$$

where *V* is the volume of the unit cell and α$_{hkl}$ is the phase associated with the structure-factor amplitude |*F$_{hkl}$*|.[105] We can measure the amplitudes but the phases are lost in the experiment. This is the phase problem. It is this electron density equation which needs to be solved in order to reveal the contents of a crystal. Molecular replacement (MR) is a phasing method that uses a homology model.[106] A sequence identity of more than 25% is usually required. The first step of MR is thus to find the three orientation defining parameters: ϕ, ψ, χ, of the rotation function (Fig. 1.23), where only the intramolecular vectors are concerned. This is followed by determining the three positional translation function parameters. After the six parameters are applied to all atoms in the search model, the positions are then transformed onto the unknown structure.

_____
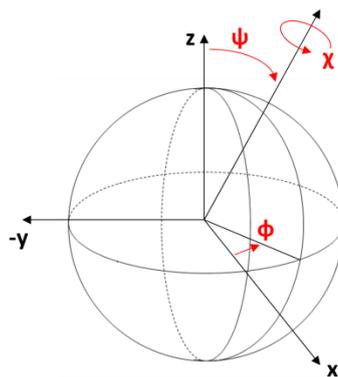
Figure 1.23: The rotation function can be specified by a latitude φ and longitude ψ through which a molecule is rotated at χ.

Then the phase $\alpha_{calc}(hkl)$ can be calculated as: $F(hkl) = |F_{obs}(hkl)| \cdot \exp[i\alpha_{calc}(hkl)]$, where F_{obs} is the observed structure factor.[105] An electron density map can be calculated through the reverse Fourier transform. The generated map is biased towards the model structure and its phase angles and any atoms in the unknown structure which do not correspond to the model will appear very weakly. To improve the electron density map of the unknown, a map showing the difference between the initial model and that obtained from MR is produced. The differences between the structure factors of the unknown and a copy of the model can be written down as:

$$F(hkl) - F_{calc}(hkl) = |F_{obs}(hkl)| \cdot \exp[i\alpha_{calc}(hkl)] - |F_{calc}(hkl)| \cdot \exp[i\alpha_{calc}(hkl)]$$

The difference (F_O-F_C) map, where F_O is the observed structure factor and F_C is the calculated structure factor, can then be expressed as:

$$P_{diff}(x) = \frac{1}{V} \sum_{hkl} (|F_{obs}(hkl)| - |F_{calc}(hkl)|) \cdot \exp[i\alpha_{calc}(hkl)] \cdot \exp[-2\pi i(hkl) \cdot x]$$

The positive densities observed in the (F_O-F_C) map are areas in the structure which is not in the search model and the negative densities are areas which are not present in the unknown structure. A ($2F_O$-F_C) map that weighs the observations further is used additionally to remove the bias from the calculated electron density map more efficiently.

The model structure obtained from MR is modified to describe the experimental data as accurately as possible. The refinement process, therefore, is carried out to obtain the best possible atomic model parameters, such as atomic positions (coordinates) and local mobility (ADP; Atomic Displacement Parameters or B-factors). Structural refinement depends on the quality of experimental data as well as the quality of the initial model used for MR. There are two ways of computing structure factor from atomic models; direct summation method and Fast Fourier Transform (*FFT*)-based method. Most macromolecular refinement programs use the *FFT*-based method as it is faster for macromolecules in comparison to small molecules.

_____

The validity of the proposed structure must then be tested by comparison of the calculated values of the amplitudes of the structure factors. This is done by calculating a *reliability index* or *R*-factor defined by:

$$R = \frac{\sum_{hkl} |F_{hkl}^{obs} - F_{hkl}^{calc}|}{\sum_{hkl} F_{hkl}^{obs}}$$

F_{hkl}^{obs} and F_{hkl}^{calc} have to be scaled with respect to each other. This formula is used for determination of both the *R* and *R_{free}* factors. The two differ in the set of reflections they are calculated from: *R* is calculated for the working set, whereas *R_{free}* is calculated for the test set. In a refined macromolecular structure, *R_{free}* is typically 2-6% higher than *R*. The *R* -factor usually ranges between 0.6 (when computed for a random model against an experimental data set) and 0.2 (for a well refined macro-molecular model at 2.5Å resolution). Small molecules of up to 1000 atoms form better-ordered crystals than larger molecules, thus the possibility of obtaining *R*-factors lower than 0.15 are higher than for larger molecules.[107]

Collaborative Computational Project, number 4 (*CCP4*)[108] and *Phenix*[109] are two program suites used for macromolecular structure determination; from data reduction to structure validation. *CCP4* is mainly used to study X-ray data whereas *Phenix* is able to refine both X-ray and neutron data simultaneously. In contrast to X-ray diffraction, neutron diffraction is used to locate protons and understand hydrogen bonding in both biological and material sciences.

(b) Neutron Diffraction

Unlike X-ray crystallography, neutrons can prevent radiation damage; allowing the same sample which has gone through neutrons to be tested for X-ray scattering as well.[110] This has paved the way for joint X-ray and neutron crystal structure refinement, including that of DNA[111,112] and proteins.[113] Neutron crystallography (NC) has an advantage over XC in that it is rather insensitive to the presence of hydrogen (H) in a structure, whereas deuterium (D), or the nuclei ^{1}H and ^{2}H, are strong scatterers for neutrons (Fig. 1.24). The greater scattering power of protons and deuterons means that the position of hydrogen in a crystal and its thermal motions can be determined with greater precision by NC. Hence, it is undesirable to work with relatively high concentration of H in a sample. The scattering intensity by H-nuclei has a large inelastic component, which creates a large continuous background that is independent of the scattering angle.

_____

	X- ray	Full D	Full H	HH/DH (0.6:0.4)
1.0 Å				
1.5 Å				
2.0 Å				
2.5 Å				
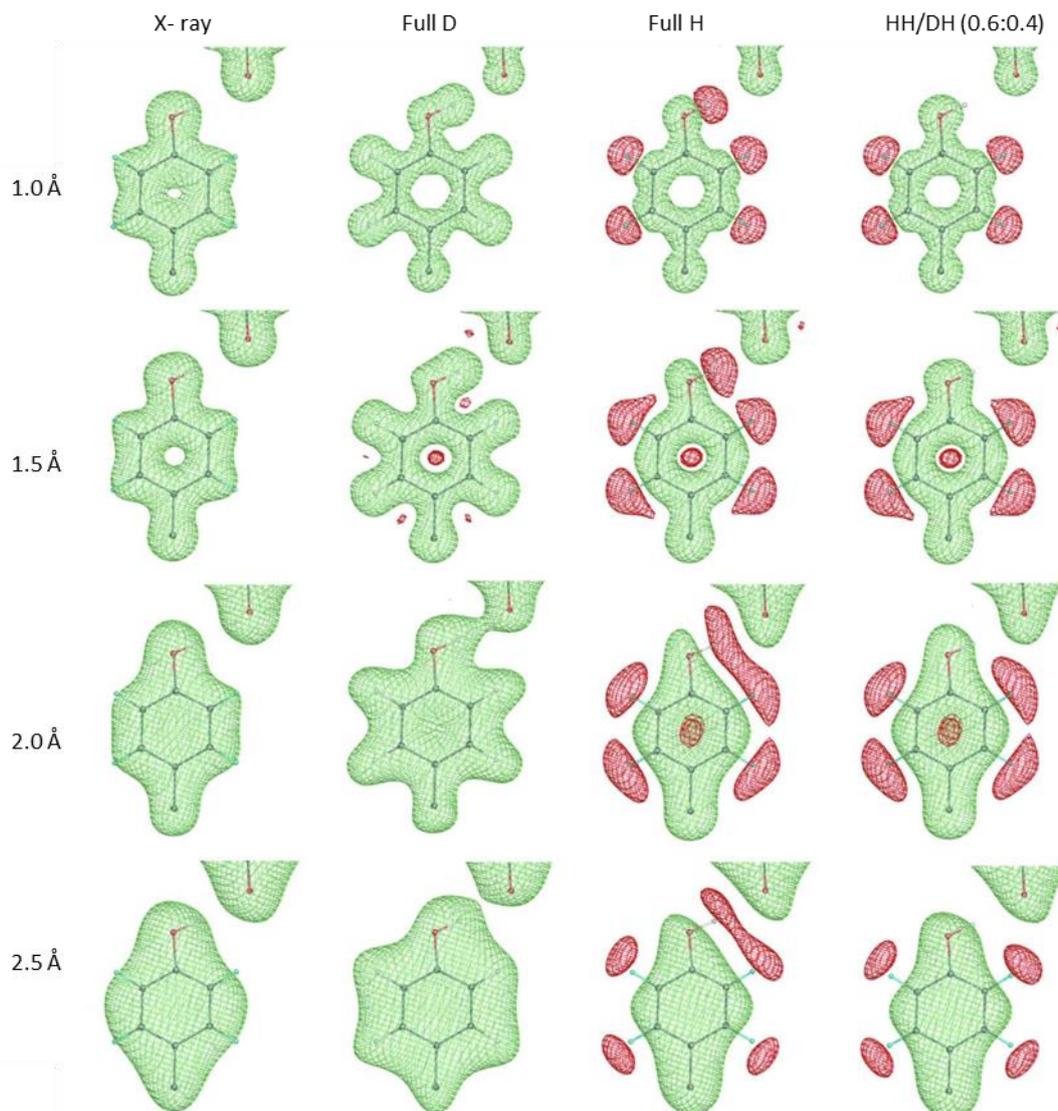
Figure 1.24: (F_{calc}, φ_{calc}) Fourier syntheses computed at different resolutions (1, 1.5, 2 and 2.5 Å) and corresponding to four different cases: X-ray, neutron fully deuterated, neutron fully hydrogenated and neutron partially deuterated, in which the hydroxyl H (HH) of tyrosine shares its site with deuterium (DH) with an occupancy ratio of 0.6:0.4. Reproduced with permission of the International Union of Crystallography. Afonine, P. V. *et al.* Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 1153–63 (2010).

The first neutron diffraction experiments were carried out in 1945 by Ernest O. Wollan using the Graphite Reactor at Oak Ridge National Laboratory. NC is used in structural biology to understand the geometry of hydrogen bonding involved in stabilising molecules and the key roles of water molecules in enzymatic reactions,[114–118] molecular recognition[119–121] and protein folding[122,123]. This is because neutrons are extremely sensitive to protons as illustrated in Figure 1.25A by a comparison of the relative scattering cross-sections. Single-crystal neutron diffraction applies the scattering of neutrons to determine the atomic structure of a material. The technique is similar to single-crystal X-

_____

ray diffraction but neutrons interact directly with the nuclei of the atoms whereas X-rays interact primarily with the electron cloud surrounding each atom.

The atomic scattering factors for neutrons do not vary directly with atomic number as they do for X-rays (Fig. 1.25B). Unlike X-ray scattering factors, neutron scattering factors cannot be calculated as there is no dependence of neutron scattering factors on Θ, the Bragg angle. This is because a nucleus is small compared with the wavelength of neutrons whereas a cloud of extranuclear electrons is of the same dimensions as the wavelength of X-rays. The scattering factor for neutrons depends on the particular isotope and some of the neutron scattering factors are negative. When X-rays are scattered by atoms, there is always a phase change of 180° between the incident and scattered waves. This is also true when most nuclei scatter neutrons, however, there are a few of which demonstrate no phase change. This is why some of the neutron scattering factors have been assigned negative; hence, the errors in a Fourier synthesis are much greater with neutrons than with X-rays.
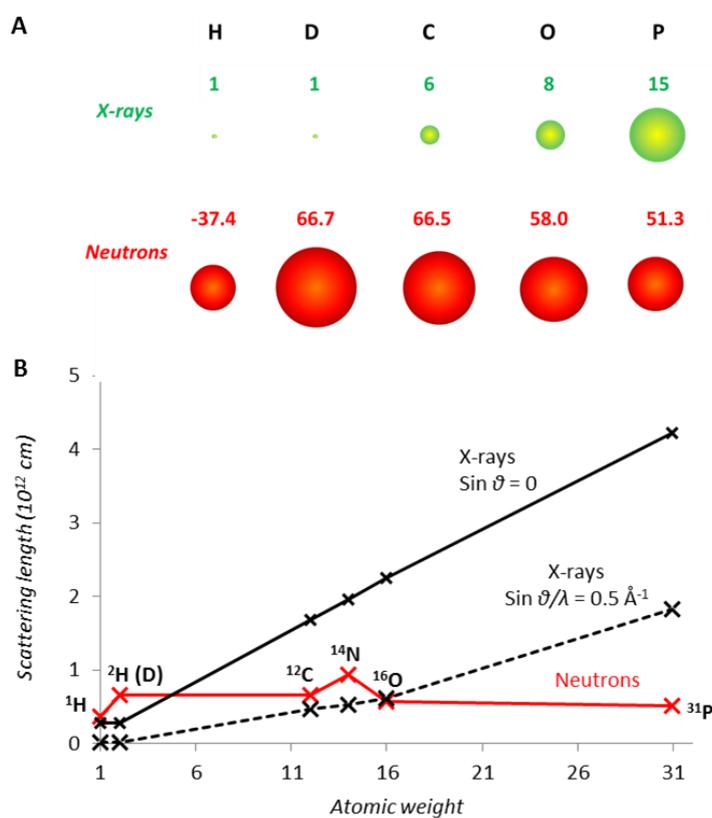
Figure 1.25: (**A**) Comparison of neutron and X-ray scattering cross-sections for elements present in DNA.[124] (**B**) Scattering length as a function of atomic weight.[125] The two different relationships of X-rays are for low and high scattering angles. As nuclei are point scattering centres, neutron scattering lengths show no angular dependence.

When neutrons are released by atomic fission, they have a very high initial velocity and a very small initial wavelength. In order to increase the wavelength, the neutrons have to be slowed down in a moderator. A beam of neutrons is weak compared with a beam of X-rays from an ordinary X-ray tube, and in order to get detectable diffraction effects, the neutron beam must be ten or twenty times

_____

as broad as an X-ray beam. Due to this low flux of neutrons, NC requires much larger crystal samples in comparison to XC, preferably of at least 0.1 mm³ in size[126] or very long exposure times for smaller crystals in order to have a measurable diffraction signal.

Neutrons produced in a spallation source have an advantage of the availability of time-of-flight (TOF) diffractometers. The spallation process provides better resolution at low d-spacings (high angle) and is good for accurate structure refinements. It gives out higher neutron flux, which translates to smaller samples and faster collection times. Typically, each proton generates 10-15 neutrons. These neutrons then pass through a moderator, often water, to reduce their energy to a level comparable with thermal neutrons from a reactor. Since the neutrons come in pulses their energy can be determined by measuring the time it takes for the neutron to hit the detector. Hence the spectra come out in TOF rather than 2θ. By recording the TOF information of a detected neutron, its wavelength can be calculated.

1.5 Investigating the Stability and Structure of DNA i-motifs

1.5.1 UV Molecular Absorption

UV molecular absorption spectroscopy is used to observe the changes in protonation and stacking of the bases involved in the formation of the i-motif. In this region, the protonation of cytosine residues produces hyperchromicity; increase in absorbance at wavelengths from 275 to 300 nm, as well a shift of the maximum wavelength from 262 nm for neutral cytosine to 275 nm for protonated cytosine (Fig. 1.26A). Hence, the stability of the i-motif as a function of temperature and pH changes can be studied by monitoring of absorbance changes in 260–300 nm range. Due to the spectral characteristics of protonated and neutral cytosine, the shape of the thermally-induced unfolding of i-motif structures monitored by molecular absorption spectroscopy is pH dependent. At pH values higher than the pKa of cytosine, the absorbance at 295 nm decreases upon unfolding of the i-motif, whereas the opposite is observed at pH values lower than the pKa of cytosine. At pH values near the pKa, no spectroscopic changes are observed upon unfolding.[67]
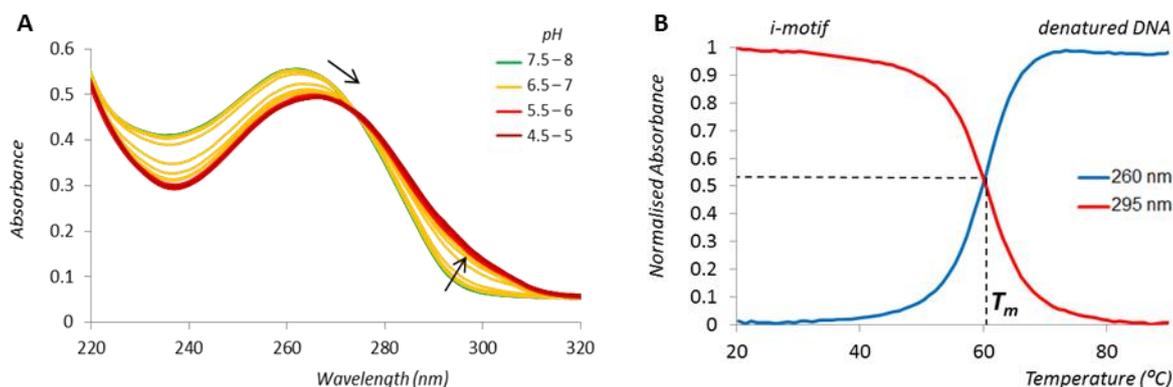
Figure 1.26: (**A**) pH titration shift of the human telomeric i-motif: d(CCCTAA)$_3$CCC, under UV light. As pH decreases, the absorbance of the i-motif shifts from 260 nm for neutral cytosine to 275 nm for protonated cytosine. (**B**) DNA melting profile of the same i-motif. The T_m value, where the two wavelengths intersect, is the melting temperature of the particular DNA.

It is possible to determine the melting temperature, T_m, of an i-motif and consequently its multimeric nature using UV light (Fig. 1.26B). If the melting of an i-motif-forming oligonucleotide shows a concentration-independent profile, the i-motif will have intramolecular pairing, whereas, if the melting spectra show a concentration-dependent profile, a bimolecular or tetramolecular association must be considered. When the i-motif is folded intermolecularly, T_m increases with the concentration.

1.5.2 Circular Dichroism Spectroscopy

CD spectroscopy of an i-motif structure shows two characteristic negative and positive bands at 260-265 and 285-290 nm, respectively (Fig. 1.27).[127] Manzini *et al.* demonstrated the pH-induced formation of the human telomeric i-motif: d[(C$_3$TA$_2$)$_3$C$_3$], found in the promoter region of the k-ras gene.[128] The stability of the i-motif with temperature or pH changes is usually studied by monitoring ellipticity around 285 nm. The light-driven DNA conformational switch of the human telomeric sequence was monitored at this wavelength to see the effects of molecular malachite green carbinol base, a light-induced hydroxide ion emitter.[129]
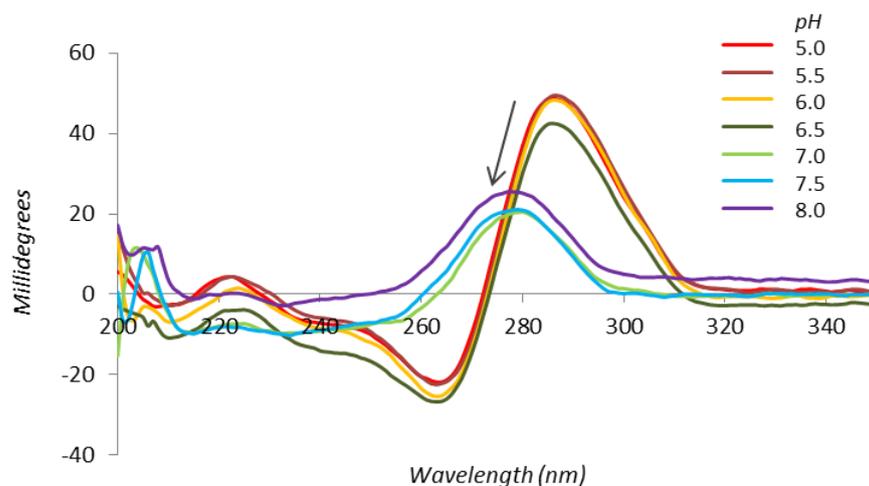
Figure 1.27: CD spectra of an i-motif from acidic to basic ionic strengths. At pH ≥ 7, the maximum shifts towards 275 nm.

Synchrotron radiation circular dichroism (srCD), has also been used for the study of i-motif structures.[130,131] The advantage of srCD is the large available photon fluxes in the vacuum UV region where absorption is strong. This allows the exploration of excitation energy for electronic coupling over a broad wavelength region. CD spectra of C-rich strands at different pH values measured by Holm *et. al.* showed that the protonation state of bases determined the folding motif and as a result the extent of electronic coupling between bases. Different electronic couplings were found depending on the protonation of bases. Single strands of all protonated bases display no coupling. I-motifs have electronic coupling within a hemi-protonated base-pair as well as nearest-neighbour coupling along the strand. Single strands of all neutral bases have nearest-neighbour couplings along the strand.

1.5.3 Crystallography

Apart from NMR, X-ray diffraction was one of the first techniques used to identify the i-motif structure. The crystal structure of $d(C_4)_4$, solved at 2.3Å resolution, revealed the formation of a four-stranded molecule composed of two intercalated duplexes.[132] Only six structures of i-motifs have been reported to date, and all are that of tetramolecular i-motifs solved using X-ray diffraction (Fig. 1.28). These molecules showed intercalated cytosine segments that were similar in their geometry, even though the sequences crystallized in different space groups.[133–135] Condensation of single molecules from solution into crystals represents a transition between distinct energetic states. In solution, the atomic interactions within the molecule dominate. In the crystalline state, however, Berger *et. al.* reported a set of additional packing interactions are formed between molecules in close contact in the lattice. This was shown in the study of the crystal structures of $d(C_3T)_4$, $d(TA_2C_3)_4$, $d(C_3A_2T)_4$ and $d(A_2C_4)_4$.[136]
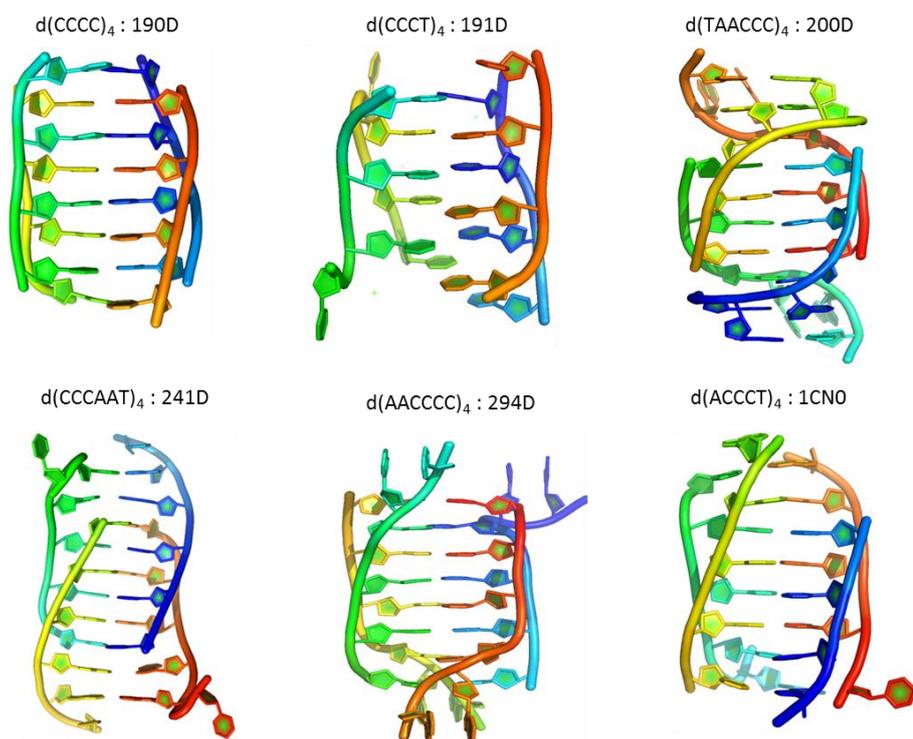
Figure 1.28: Crystal structures of six tetramolecular DNA i-motifs of the sequence; d(C$_4$)$_4$,[132] d(C$_3$T)$_4$,[134] d(TA$_2$C$_3$)$_4$,[135] d(C$_3$A$_2$T)$_4$,[133] d(A$_2$C$_4$)$_4$[58] and d(AC$_3$T)$_4$[79] with their corresponding PDB IDs.

In 2014, Ming-Hon Hou *et al.* reported an X-ray crystal structure of an i-motif tetraplex core with a parallel-duplex junction as a structural motif. The DNA crystallised contained a CCG-repeat sequence, which shows that two d[T(CCG)$_3$]A strands can associate to form a tetraplex structure with an i-motif core containing four C$^+$·C base pairs flanked by two G·G base pairs as a structural motif. CCG-containing tetraplex structures were previously reported to be stabilized by intertwining i-motifs in solution.[56,69] Figure 1.29 illustrates that the structure formed by a double hairpin closely resembles the theoretically proposed bimolecular i-motif structure. Expansion of (CCG)$_n$ (where n = 4, 8 or 16) trinucleotide sequences are associated with certain neurological diseases, including the fragile X chromosome syndrome.[137]
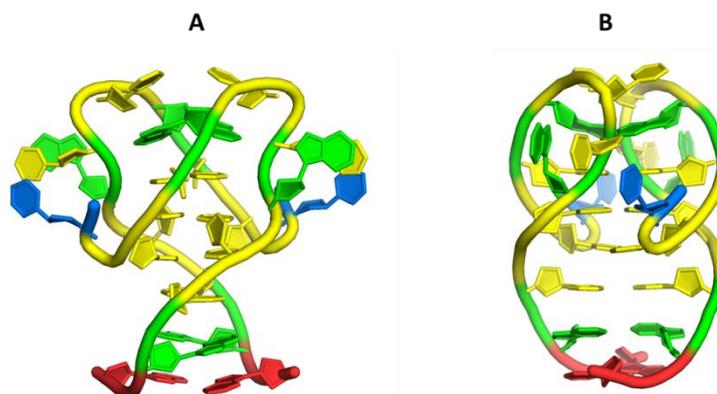
Figure 1.29: (**A**) Crystal structure of the d[T(CCG)$_3$]A DNA strands that fold into an intercalated i-motif (PDB ID: 4PQZ).[137] (**B**) View of the structure through the wide groove. Cytosine bases are yellow, guanine bases are green, thymine bases are blue and adenine bases are red.

1.6 Aims

The G-quadruplex structure has been extensively studied due to its ability to form *in vivo*.[90,138] The i-motif, on the other hand, had not been as widely studied because of the notion that its stability in acidic conditions makes it irrelevant to biological functions. However, recent findings of C-rich sequences that form i-motifs at neutral pH[38,39] and the discovery of an i-motif binding protein[41,139,] have prompted more interest in the intercalated structure. Stability of the i-motif in solution had been catagorised based on five C-rich oncogene sequences known to form the structure. A systematic study on the effect of loop length on G-quadruplex stability had been made[140] but the same approach had not been taken on i-motifs. This thesis will largely focus on the sequence-specific stability of the i-motif DNA. The four main aims are as follows:

1. Investigate the effect of loop length on the stability of i-motif structures.

Intramolecular i-motifs usually exist in the form C$_{2-5}$L$_{1-7}$C$_{2-5}$L$_{1-7}$C$_{2-5}$L$_{1-7}$C$_{2-5}$, where C$_{2-5}$ denotes the cytosine stretch and L$_{1-7}$ are "loop" regions containing any DNA base (L) including cytosine. It is known that increase in the cytosine tracts leads to increased stability of i-motifs due to the presence of more C$^+$·C base pairs.[141] The role of loops is to aid in intramolecular i-motif formation. Finding out if these loop regions contribute to structural stability could help to target C-rich sequences of specific loop lengths in the human genome for *in vivo* studies.

2. Examine the effect of [Ru(phen)$_2$dppz]$^{2+}$ on i-motif structures with altering loop lengths.

Among ligands that have been known to bind to i-motifs is a ruthenium polypyridyl complex, whose fluorescence enhances in the presence of DNA. [Ru(phen)$_2$(dppz)]$^{2+}$ (phen = 1,10-phenanthroline, dppz = dipyrido [3,2-a:2',3'-c] phenazine) has been reported to bind to both the human telomeric G-quadruplex and i-motif[142] and has been used as a probe for studying DNA properties.[143] Addition of

_____

the complex to i-motifs of differing loop lengths can provide better understanding of which loops are involved in ligand interaction.

3. Investigate the effects of loop base composition on i-motif stability.

A detailed study of loop base composition on i-motifs with a fixed number of loop length can help understand if particular DNA bases contribute more/less to the stability of intramolecular i-motif structures. The results obtained from this study can provide information on whether i-motif stability is sequence specific and to establish factors other than pH on i-motif formation.

4. Show how different DNA bases in the loop regions can contribute to stability.

Six tetramolecular i-motif structures have been reported to date, with the seventh resembling a bimolecular i-motif. Crystallisation of more i-motif structures not only offers greater insight into the four-stranded structure but also aids in design of solution works based on the geometric parameters of the structures; such as calculating which ligands have more probability of binding to the structure based on the behavior of the loop regions.

_____

The Importance of Loop Length on i-motif Stability

2.1 Introduction

Cytosine-rich nucleotide sequences can fold to form intramolecular i-motifs *in vivo*.[60] These sequences usually exist in the form $C_xL_{y1}C_xL_{y2}C_xL_{y3}C_x$, where C_x denotes the cytosine stretch and *y1*, *y2* and *y3* are known as loop regions containing any DNA base (L) including cytosine (Fig. 2.1). Similarly, guanine (G)-rich sequences form G-quadruplexes,[144] which are structures complementary to i-motifs. The G-quadruplex has been shown to exist *in vivo*.[145] The G-quadruplex structure is further stabilized by the presence of a cation, with K^+ cations sitting in a central channel between each pair of tetrads generating the most stable structures.[26,146,147]

The potential intramolecular i-motif sequence was modelled after the general G-quadruplex sequence.[148] This i-motif sequence has been defined as having four runs or tracts of three to five cytosine bases, and each tract separated by loop regions of DNA containing one to seven bases: $C_{3-5}L_{1-7}C_{3-5}L_{1-7}C_{3-5}L_{1-7}C_{3-5}$. The lengths are restrained because of the evidence to date that G-quadruplexes exist as short nucleic acid sequences[149] and also for practical reasons; including to avoid difficulty in sequence searching.[148]
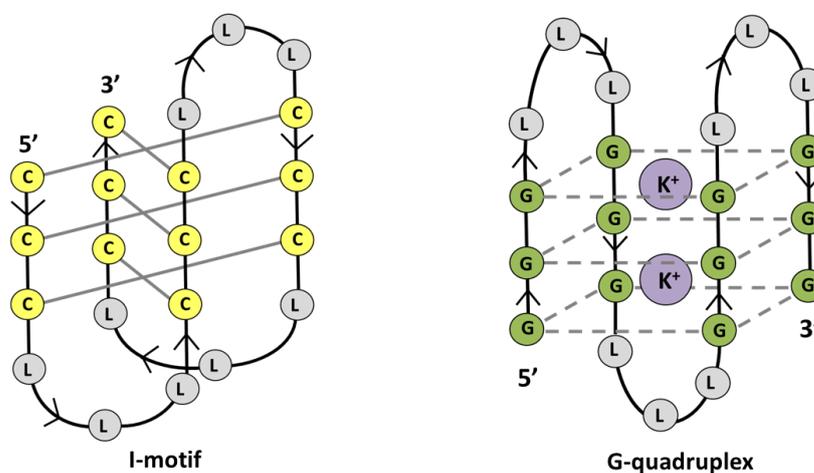
Figure 2.1: Schematic representations of an intramolecular i-motif: d[$(C_3L_3)_3C_3$] and an anti-parallel G-quadruplex: d[$(G_3L_3)_3G_3$]. Cytosine bases are yellow and guanine are green. All DNA bases in the loop regions are grey. The purple spheres represent potassium cations that help in G-quadruplex structure stabilisation.

The G-quadruplex has garnered much interest for providing a nucleic-acid-based mechanism for regulating telomere maintenance,[138,150] as well as for transcription,[151,152] replication[153,154] and translation.[155–157] Promoter G-quadruplexes have been divided into four groups[73] and solution studies show that the stability of these structures increases with increasing loop lengths.[140] In contrast, there is less data on i-motifs formed in promoter complexes. I-motifs were initially

_____

classified into two groups based on just the five i-motif forming sequences found in oncogenes.[73] It was reported that loops with less than four DNA residues have lower stabilities. The effect of loop length on the stability of G-quadruplexes has been reported using 80 different G-rich sequences.[140] However, such a systematic study had not been made on i-motifs. To initiate a similar study, the effects of different loop lengths in 12 intramolecular i-motifs were investigated.

The human telomeric i-motif DNA: d[(CCCTAA)$_3$CCC], consists of three cytosine bases per tract. The study was, hence, modelled with C-rich sequences consisting of the same number of bases in the cytosine tract. Table 2.1 lists the C-rich sequences used, which contained an increasing number of thymine bases (C_3T_X, where x is the no. of thymine residues) within the loop region so, the sequence was in the form: d(CCCT$_{(3-8)}$CCCT$_{(3-8)}$CCCT$_{(3-8)}$CCCT). C_3T_8 was added to the list to test i-motif formation beyond the suggested general sequence of d[($C_{3-5}N_{1-7}$)$_3C_{3-5}$]. DNA sequences were also designed to look into independently varying loop lengths (C_3T_{abc}, where C is cytosine, T is thymine and either a, b & c = 3 or 8). These sequences were assessed for their thermal and pH stability using UV and synchrotron radiation CD (srCD) spectroscopy. By limiting the loop bases to thymine, the possibility of interaction between the bases was limited to T·T and to ensure consistency throughout the series of oligonucleotides. All sequences used in this study proved to form i-motifs as shown in section 2.3.

Table 2.1: Sequences used in this study, where C is cytosine and T is thymine.

Group	Name	Sequence (5'→3')
C_3T_X	C_3T_3	CCCTTTCCCTTTCCCTTTCCCT
	C_3T_4	CCCTTTTCCCTTTTCCCTTTTCCCT
	C_3T_5	CCCTTTTTCCCTTTTTCCCTTTTTCCCT
	C_3T_6	CCCTTTTTTCCCTTTTTTCCCTTTTTTCCCT
	C_3T_7	CCCTTTTTTTCCCTTTTTTTCCCTTTTTTTCCCT
	C_3T_8	CCCTTTTTTTTCCCTTTTTTTTCCCTTTTTTTTCCCT
C_3T_{abc}	C_3T_{338}	CCCTTTCCCTTTCCCTTTTTTTTCCC
	C_3T_{383}	CCCTTTCCCTTTTTTTTCCCTTTCCC
	C_3T_{833}	CCCTTTTTTTTCCCTTTCCCTTTCCC
	C_3T_{883}	CCCTTTTTTTTCCCTTTTTTTTCCCTTTCCC
	C_3T_{838}	CCCTTTTTTTTCCCTTTCCCTTTTTTTTCCC
	C_3T_{388}	CCCTTTCCCTTTTTTTTCCCTTTTTTTTCCC

Intramolecular i-motif formation is typically validated using three biophysical techniques. They are: CD spectroscopy, DNA melting and pH titrations. All three methods measure the stability of an i-motif forming sequence with regards to temperature and changes in pH.[67,128] The biophysical characteristics of an intramolecular i-motif are shown in Figures 2.2 and 2.3. T_m refers to the melting temperature of an i-motif and the T_{pH} value is the pH at which half of the i-motif is unfolded.

_____
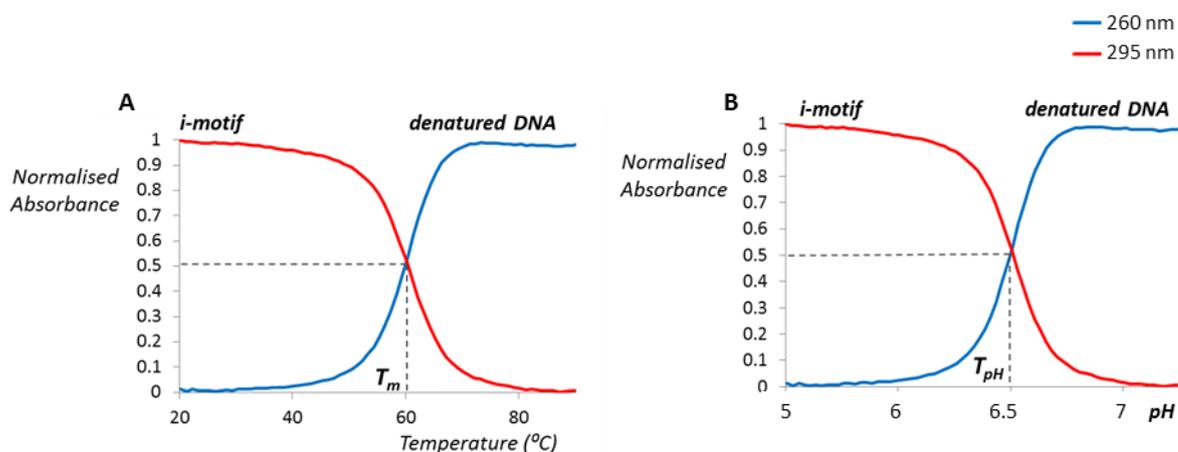
Figure 2.2: (**A**) DNA melting profiles and (**B**) pH titration spectra of the intramolecular human telomeric i-motif: d[(CCCTAA)$_3$CCC]. The T_m and T_{pH} values, where the two wavelengths intersect, are the melting temperature and transitional pH value, respectively.

In their native states, the bases of DNA absorb light in the 260 nm wavelength.[158] DNA melting involves the denaturation of DNA by increasing temperature. This results in *hyperchromicity*; the increase in absorbance in the 260 nm wavelength region with increasing number of unstacked DNA bases (Fig. 2.2A). As temperature increases, the four-stranded i-motif structure unfolds to a single-stranded (ss) sequence, resulting in an increase in absorbance due to availability of unbound nucleotides or π-*unstacking* with an aromatic residue. This effect is also seen during i-motif denaturation as a result of pH increment (Fig. 2.2B).[159]

The pK_a value of the nitrogen involved in the cytosine$^+$-cytosine (C$^+$·C) base pair is approximately 4.6 in pure water at 25°C. At pH values higher than 4.6 pK_a, the absorbance at 295 nm wavelength decreases. This is also because of the deprotonation of the cytosine bases which results in the unfolding of the i-motif into ss DNA. The opposite is observed at pH values lower than the pK_a of cytosine, therefore the transition at 295 nm under UV indicates formation/degradation of the C$^+$·C tetraplex.[67] Figure 2.3A shows pH titration spectra of the widely studied human telomeric i-motif. It shows hyperchromicity between 275 and 300 nm, and a shift of the absorbance from around 260 nm for neutral cytosine to around 275 nm for protonated cytosine.[67] Figure 2.3B shows that under circularly polarised light, an i-motif reveals a maximum at 285–290 nm at pH ≤ 5.5. At pH ≥ 7, this maximum shifts towards 275 nm wavelength.

_____
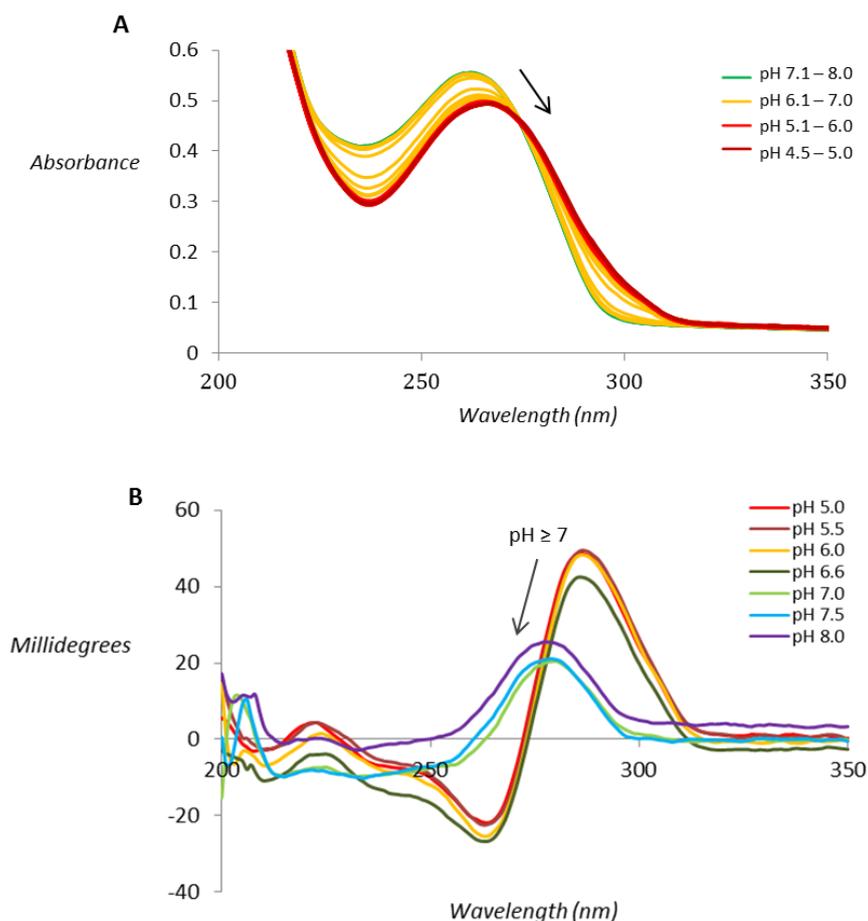
Figure 2.3: (**A**) pH titration shift of an i-motif forming sequence under UV light. As pH decreases, the absorbance of the i-motif shifts from 260 nm for neutral cytosine to 275 nm for protonated cytosine. (**B**) CD spectra of an i-motif from acidic to basic ionic strengths. At pH ≥ 7, the maximum shifts towards 275 nm.

2.2 Materials and Methods

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 provided by the manufacturer. All experiments for each sequence were run in triplicates.

2.2.1 UV Absorption Measurements

Each solution was made in 50 mM sodium cacodylate buffer with a final DNA concentration of 1 μM (ss). The pathlength of the cuvettes used was 1 cm. UV absorption spectroscopy was carried out using an Agilent Cary 100 with a temperature controlled six cell changer.

_____

(a) DNA Melting Experiments

Oligonucleotides were dissolved in 50 mM sodium cacodylate buffer of the appropriate pH to a final concentration of 1 μM (ss), and were annealed by heating to 90°C and allowing them to cool slowly to room temperature. Absorption was recorded at 260 and 295 nm at 1°C intervals between 20-90°C, with a temperature change rate of 1°C/min in a 1 cm pathlength quartz cuvette.

(b) pH Titrations

Oligonucleotides were dissolved in 50 mM sodium cacodylate at pH 8, and a spectrum recorded between 200-350 nm at 25°C. The pH was decreased in steps of 0.2 units with the addition of 2M HCl and absorption values recorded. This procedure was continued until a minimum pH of 4.5 was achieved. pH was measured in the cuvette using a Thermo Scientific Orion Star A11 pH meter equipped with a small diameter pH electrode.

2.2.2 Circular Dichroism Measurements

Oligonucleotides were dissolved in 20 mM sodium cacodylate buffer of the appropriate pH to a final concentration of 200 μM (ss), and were annealed by heating to 90°C and slowly cooling to room temperature. CD spectra were recorded at 20°C between 180 and 350 nm wavelengths with 1 nm wavelength increments in a 0.01 cm pathlength cuvette on beamline B23 at Diamond Light Source.

2.3 Results & Discussion

2.3.1 Confirmation of i-motif Formation by Circular Dichroism Spectroscopy

CD measurements were run to validate that all the C-rich oligonucleotides used in this study are i-motif forming sequences. The concentration of buffer used for srCD was lowered to 20 mM to obtain the lowest cut off measurement in the backbone region of 180-260 nm wavelengths. The 200 μM (ss) concentration of each DNA was used to read reasonable absorbance using 0.01 cm pathlength, i.e. not below or over the limit of detection in the far and near UV regions.

At pH ≤ 5.5, all sequences revealed a maximum at 285–290 nm (Fig. 2.4), which is characteristic of intramolecular i-motif structures. This maximum is retained at pH ≤ 6.5 by both C_3T_3 and C_3T_4, suggesting that the two i-motifs are more stable than the rest of the C_3T_X oligonucleotides. At pH ≥ 7, the maximum was seen to shift towards 275 nm wavelength for all oligonucleotides, which relates to the unstructured ss DNA.[160] This is because as pH increases, the intercalated $C^+ \cdot C$ core unfolds to give ss DNA, hence, increasing pH decreases i-motif stability. All six of the oligonucleotides with increasing number of thymine residues were therefore confirmed as i-motif forming sequences.

_____
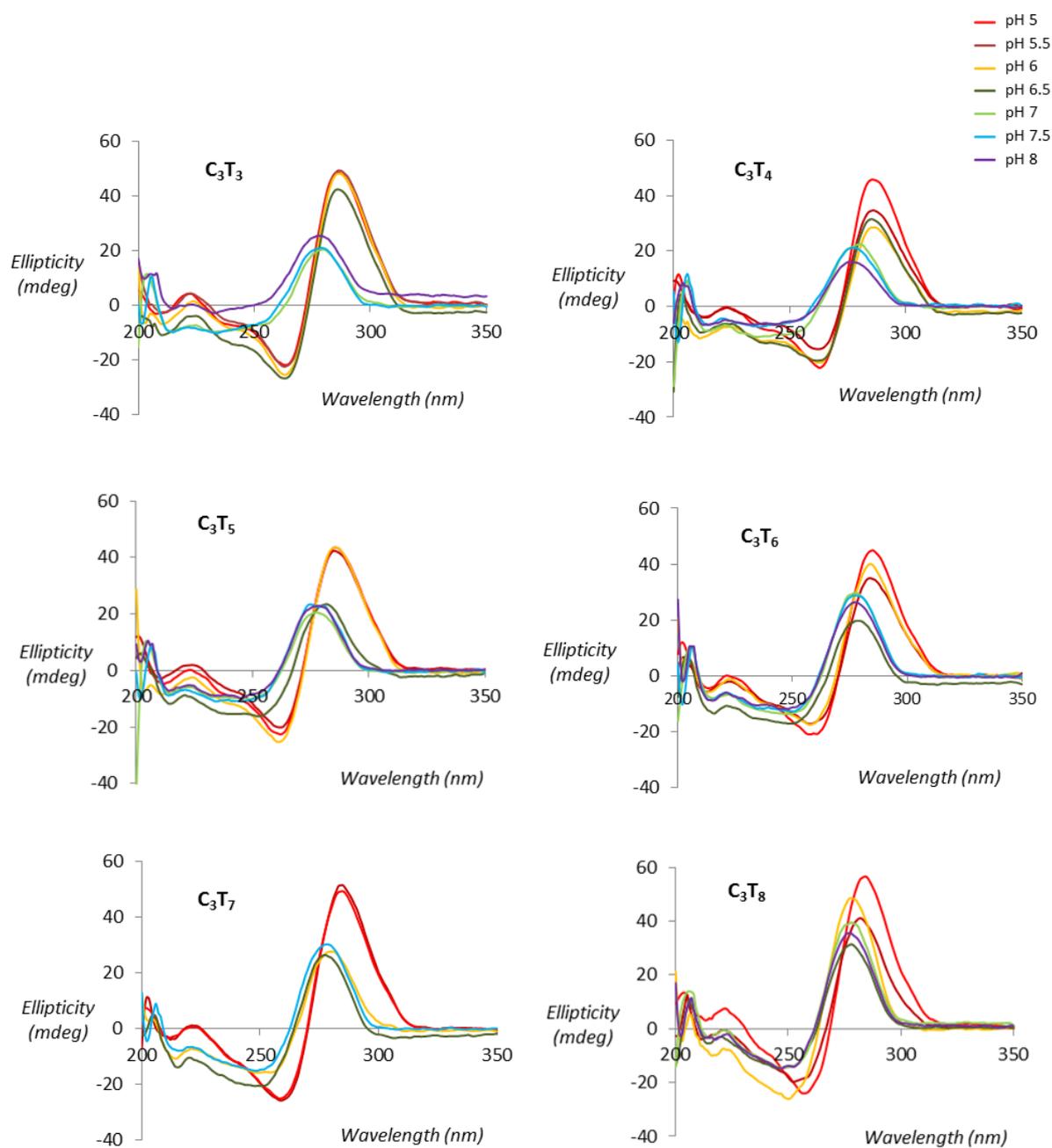
Figure 2.4: CD spectra of C_3T_x in 20 mM sodium cacodylate buffer at various pH values.

2.3.2 Analyses of i-motif Stability using UV Absorption Spectroscopy

The ionic strength of the buffer solution used in UV absorption measurements was 50 mM as studies under UV light with high levels of salt (>100 mM) are known to destabilise i-motif structures.[67] 1 cm pathlength was used in order to read reasonable absorbance, i.e. not below or over the limit of detection (0 ≤ absorbance ≤ 1.5) in the UV spectrometer.

_____

Absorbance of UV light by C_3T_X is sigmoidal to increasing temperature at 260 nm, until a maximum absorbance value is reached. Similarly, absorbance of UV light by the oligonucleotides is inversely proportional to temperature at 295 nm (Appendix Fig. A2.1A). The DNA melting temperatures in Figure 2.5 show the typical intramolecular i-motif features as expected, with dissociation of i-motif structures into ss DNA by heat. As loop length, or the number of thymine bases increases, the melting temperature decreases, except in the case of C_3T_3 and C_3T_4 where the T_m is the same at 57.0°C (to three significant figures). This supports the results obtained from CD measurements that C_3T_3 and C_3T_4 are the most stable i-motif forming sequences in the C_3T_X system.
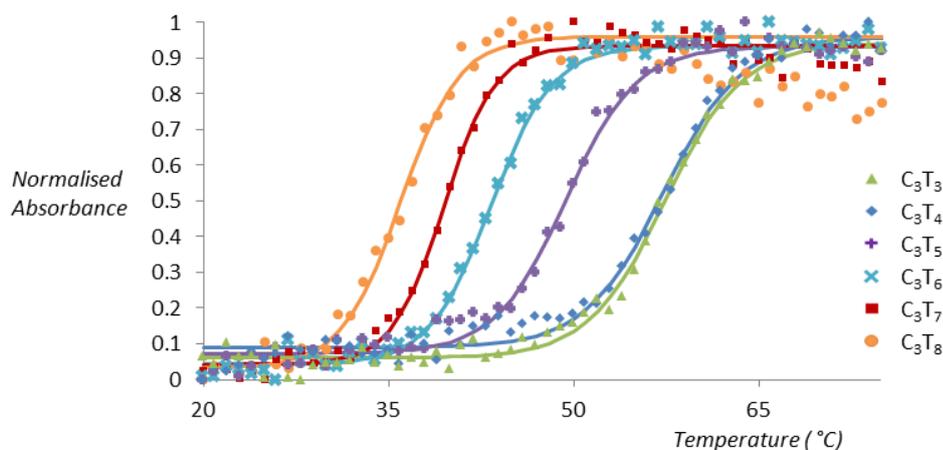
Figure 2.5: DNA melting profiles of C_3T_x in 50 mM soidum cacodylate buffer (pH 5) at 260 nm wavelength.

Figure 2.6 shows the influence of pH on i-motifs. The titrations indicate that as pH increases, the degradation of i-motifs into ss DNA is favoured. This is additionally supported by the decrease in absorbance at 295 nm which indicates the dissociation of $C^+·C$ base pairs (Appendix Fig. A2.1B). The results also support the previous notion that C_3T_3 is the most stable of the six i-motif forming sequences as its structure is maintained up to pH 6.6. However, in contrast to CD and DNA melting experiments, C_3T_4 is less stable than C_3T_3 at pH 6.4. Table 2.2 and Figure 2.7 show that increase in loop length decreases i-motif stability.
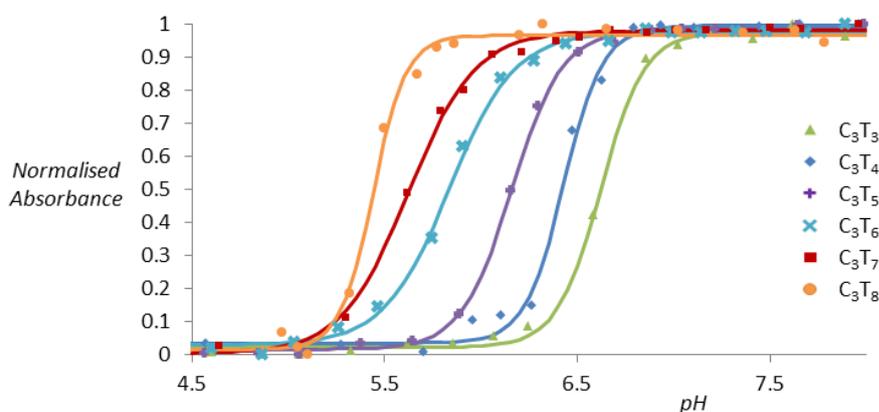
Figure 2.6: pH titration spectra (at 260 nm wavelength) of C_3T_x in 50 mM sodium cacodylate buffer.

_____

Table 2.2: Melting temperatures (T_m) and transitional pH values (T_{pH}) of i-motif forming C_3T_x oligonucleotides

Name	Sequence (5'→3')	T_m (°C)	T_{pH}
C_3T_3	CCCTTTCCCTTTCCCTTTCCCT	57.0	6.6
C_3T_4	CCCTTTTCCCTTTTCCCTTTTCCCT	57.0	6.4
C_3T_5	CCCTTTTTCCCTTTTTCCCTTTTTCCCT	49.0	6.2
C_3T_6	CCCTTTTTTCCCTTTTTTCCCTTTTTTCCCT	43.0	5.8
C_3T_7	CCCTTTTTTTCCCTTTTTTTCCCTTTTTTTCCCT	39.5	5.6
C_3T_8	CCCTTTTTTTTCCCTTTTTTTTCCCTTTTTTTTCCCT	36.0	5.4
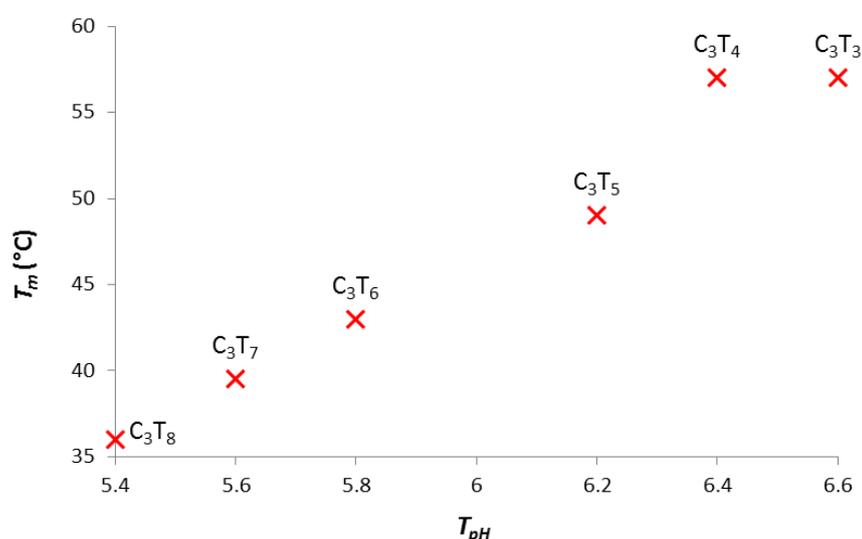
Figure 2.7: Melting temperatures (T_m) *vs.* transitional pH values (T_{pH}) of i-motif forming C_3T_x oligonucleotides

Similarly, C-rich sequences with varying loop lengths within the same sequences (C_3T_{abc}) also gave DNA melting curves and changes in pH specific to i-motifs. Figure 2.8 illustrates where the stabilities of these sequences lie between C_3T_3 and C_3T_8. Results obtained from (C_3T_{abc}) i-motifs continued to show that sequences that have more thymine bases are least stable. When compared to C_3T_3 and C_3T_8, i-motifs with two loops of three thymine residues are less stable than C_3T_3 and i-motifs with two loops of eight thymine residues are more stable than C_3T_8. The results also show that i-motif stability is dependent on the length of each loop. Figure 2.8A shows that change in the number of bases in the first and last loops, of a particular sequence length, give similar T_m values. For sequences containing 14 thymine residues, the difference in T_m between C_3T_{833} and C_3T_{338} is 1°C. For sequences containing 19 thymine residues, , the difference in T_m between C_3T_{883} and C_3T_{388} is also 1°C. However, changes in the first and last loop lengths in terms of increased number of thymine residues show that addition of more thymine bases decreases i-motif stability. C_3T_{883} is 8°C less stable than C_3T_{383} and C_3T_{838} is 10°C less stable than C_3T_{338}. Change in length of the middle loop also

_____

gives significantly different T_m values. C_3T_{383} is 15°C more stable than C_3T_{838}. Overall, i-motifs containing shorter loops are more stable than those with longer loop (Table 2.3).
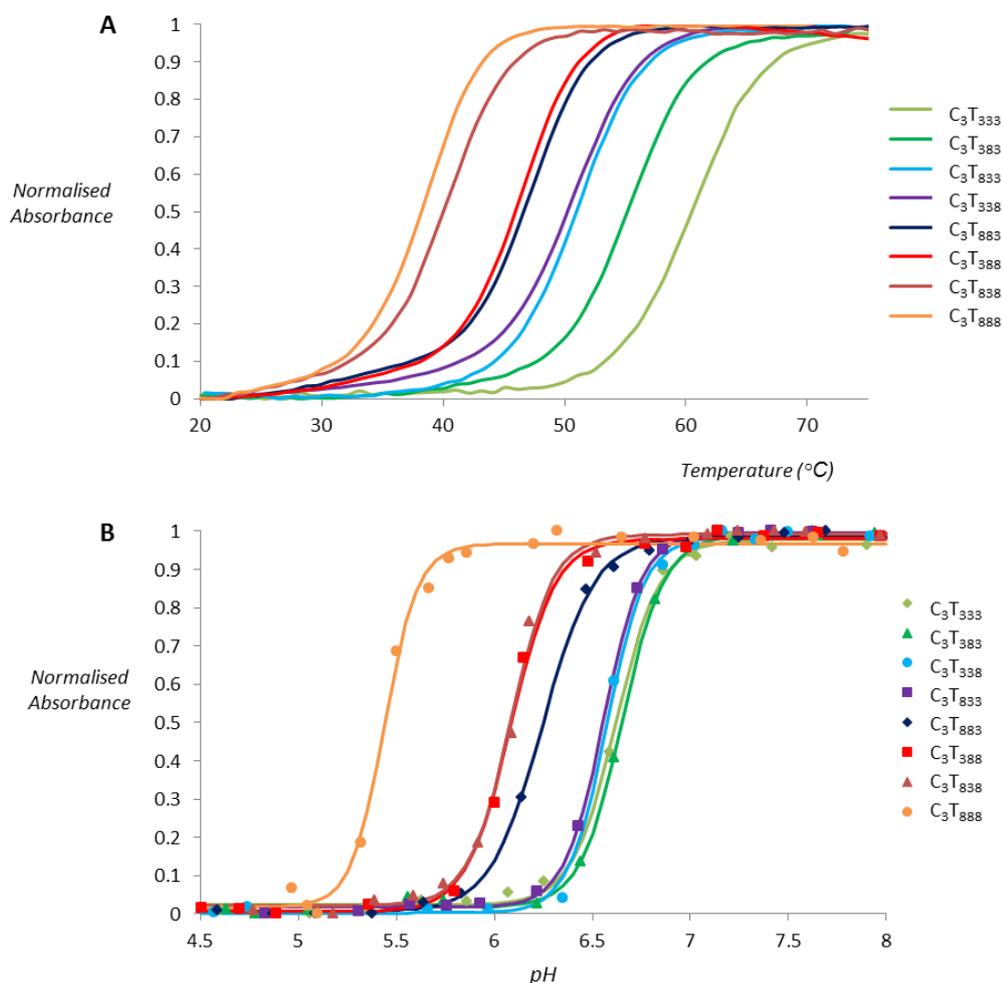
Figure 2.8: (**A**) DNA melting profiles of C_3T_3, C_3T_8 and C_3T_{abc} in 50 mM soidum cacodylate buffer (pH 5) at 260 nm wavelength. (**B**) Absorbance (at 260 nm wavelength) of C_3T_3, C_3T_8 and C_3T_{abc} in 50 mM sodium cacodylate buffer against pH.

Table 2.3: Melting temperatures (T_m) and transitional pH values (T_{pH}) of i-motif forming C_3T_{abc} oligonucleotides compared with C_3T_3 and C_3T_8

Name	Sequence (5'→3')	T_m (°C)	T_{pH}
C_3T_3	CCCTTTCCCTTTCCCTTTCCCT	57.0	6.6
C_3T_{383}	CCCTTTCCCTTTTTTTTCCCTTTCCC	55.0	6.6
C_3T_{833}	CCCTTTTTTTTCCCTTTCCCTTTCCC	51.0	6.5
C_3T_{338}	CCCTTTCCCTTTCCCTTTTTTTTCCC	50.0	6.5
C_3T_{883}	CCCTTTTTTTTCCCTTTTTTTTCCCTTTCCC	47.0	6.2
C_3T_{388}	CCCTTTCCCTTTTTTTTCCCTTTTTTTTCCC	46.0	6.1
C_3T_{838}	CCCTTTTTTTTCCCTTTCCCTTTTTTTTCCC	40.0	6.1
C_3T_8	CCCTTTTTTTTCCCTTTTTTTTCCCTTTTTTTTCCCT	36.0	5.4

_____

DNA melting experiments with different NaCl concentrations were also run in order to check for any differences in stabilities of C_3T_3 and C_3T_8 in the presence of salt. This showed a decrease in the stability of C_3T_3 by 2°C and 4°C in the presence of 50 mM and 1 M NaCl, respectively (Fig. 2.9A), whereas addition of the salt to C_3T_8 increased the long looped i-motif's stability by 11°C (Fig. 2.9B). This raises the notion that longer looped C-rich sequences have higher chances of forming i-motif structures *in vivo*.
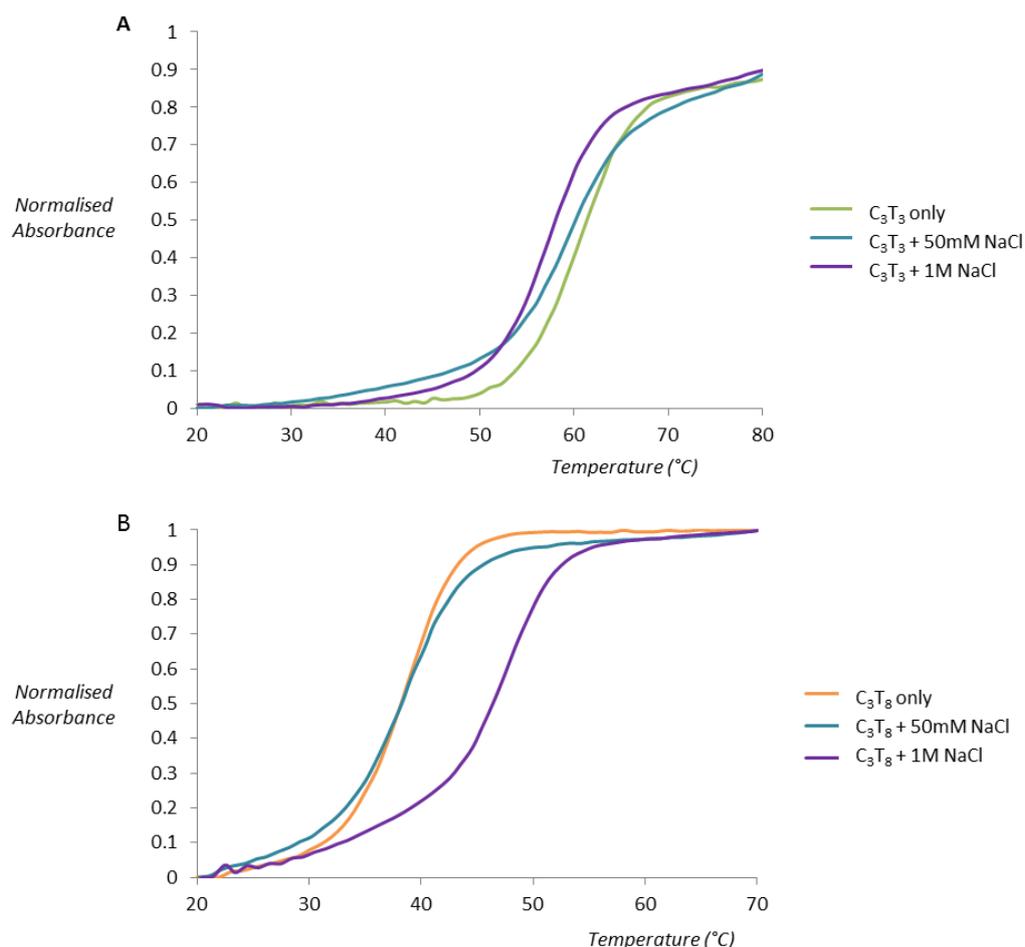
Figure 2.9: DNA melting profiles of (**A**) C_3T_3 and (**B**) C_3T_8 with increasing salt concentration. λ =260 nm.

2.4 Conclusions

DNA melting experiments of C_3T_X gave decreasing T_m values from C_3T_3 to C_3T_8, showing that the stability of the i-motif structure decreases as loop length increases; a finding which is in contrast with statements made in the literature.[73] An explanation for this observation is that as loop length increases, the structure gains more flexibility and therefore, stabilising interactions become weaker. This decrease in stability is also present in the results from the pH titration experiments, where the transitional pH decreases as loop length increases. Once again C_3T_3 shows the highest stability with T_{pH} value of 6.6 whereas C_3T_8 demonstrated the lowest T_{pH} value of 5.4.

_____

The sensitivity of i-motifs to pH was also confirmed by CD spectroscopy. The spectra obtained for all C_3T_X oligonucleotides (Fig. 2.4, pH 5) show the characteristic maximum at 285 nm and minimum at 265 nm associated with i-motif formation. Data measured at pH 6.5 shows that these peaks are observed for C_3T_3 and C_3T_4 only. This is consistent with the T_{pH} values measured using UV absorbance. CD spectroscopy data measured at pH 7.5-8 lack all the characteristic signals for an i-motif structure, confirming the absence of i-motif formation under basic conditions for all sequences. These results suggest that flexibility within loop regions is detrimental to the stability of the i-motif structure, and long loop regions are only stabilising if they can form additional intramolecular interactions that limit flexibility.

Addition of 50 mM NaCl showed a small decrease of 2°C in the stability of C_3T_3 (Fig. 2.9A). C_3T_8, on the other hand, displayed an increase in stability by 11°C in the presence of salt (Fig. 2.9B). This again suggests that stabilising interaction within long loop regions, in this case, formation of salt bridges or ordering of water between the long loops by increased ionic components has led to a stabilisation of the loop structure, and ultimately stabilisation of the i-motif structure. These results confirm that the bases in the loop region are an important consideration when determining the stability of the structure of an i-motif forming sequence.

In order to investigate the contribution from each loop region, a further series of sequences were used, in which the length of each loop was systematically altered between 3 and 8 thymine bases (C_3T_{abc}). It is clear from Table 2.3 that introducing a single long loop into a short loop i-motif sequence, has a minimal effect on the pH stability of the structure, with C_3T_{833}, C_3T_{383} and C_3T_{338} showing similar T_{pH} values to the parent sequence C_3T_3. In contrast, introducing a single short loop into a long loop sequence, results in a significant increase in the T_{pH} values, between C_3T_{888} and the C_3T_{883}, C_3T_{838} and C_3T_{388} series, suggesting that the short loop could be acting to orient the intercalating strands together to enhance stability. The thermal stability of the mixed loop structures also shows changes in stability in a loop dependent manner. Alterations in length of the central loop leads to the most significant changes in stability; with C_3T_{383} being 15°C more stable than C_3T_{838}. This finding is also supported by Reilly and co-workers who reported that change in the central loop affects intramolecular i-motif stability the most.[161]

The structures in which changes are made to the first and last looping region show similar stability irrespective of which loop is changed, although structures with two long loops are always less stable than those with two short loops. For example, taking C_3T_3 into account, the difference in T_m between the i-motif and C_3T_{338} is 7°C. C_3T_{833} is 6°C less stable than C_3T_3 whereas C_3T_{383} is only 2°C less stable than C_3T_3. These results demonstrate that the important loops for stability must be the first and last loop, with the middle loop able to accommodate a longer sequence with minimal disruption to the stability.

_____

In 2014, Hurley *et al.* reported the binding of a small molecule to the middle loop of the Bcl-2 promoter i-motif by increasing the stability of the i-motif.[41] The Bcl-2 gene is related to the evasion of apoptosis and its i-motif contains long loops of 8, 5 and 7 bases.[162] The presence of a longer sequence in the middle loop will therefore be beneficial in terms of small molecule binding, as it may provide a unique site to distinguish between different loop sequences. It is also clear that if an i-motif forming sequence contains long sequences in the first or last loop region then it must have the ability to form additional interactions to form a more stable i-motif structure. The difference in both thermal and pH stability as a function of loop length also provides the opportunity to tune the temperature and pH at which the i-motif structure undergoes conformational change and may be useful in developing more sensitive i-motif based switches for nano-devices. A recent report has shown that tuning the sequence using different length of cytosine tracts and three to four bases in the loop regions allows switching over a well-defined pH range between pH 7 and 6.5.[163] Using different length loops would allow this pH range to extend to pH 5.5, and introduce the additional parameter of heat responsive conformational change. Altering loop length would also have the added advantage of changing the distance the terminal ends of the switch can travel.

In conclusion, the results demonstrate that i-motif stability is clearly related to C-rich sequence composition. The length of the loop region is a large contributing factor to the overall stability of the i-motif structure, but in contrast to the widely held belief, it is not long loops that give the most stable structures, but short loops. In addition, the central loop is less important in providing stability to the i-motif structure, with the first and second loops shown to influence stability more. These findings suggest that in biologically relevant sequences with long loops, the sequence composition of the loops must be able to form interactions that reduce the influence of loop length and effectively mimic short loop sequences and the stability they impart. The thermal and pH responsiveness of different loop lengths may also be used to tuning the responsive nature of molecular switches based on the i-motif structural transition.

_____

Stabilisation of Long-Looped i-motifs by [Ru(phen)$_2$(dppz)]$^{2+}$

3.1 Introduction

The first i-motif binding ligand was reported in 2000 as TMPyP4, a cationic porphyrin.[88] The compound was reported to bind to tetramolecular i-motif via stacking interactions at the ends of the structure.[88] The study thus prompted an interest in i-motif stabilisation in order to investigate potential biological functions of the i-motif DNA. The discovery of an i-motif binding protein in 2014 put forward the biologically relevant notion of the i-motif as a protein recognition site.[41,87] Hurley and co-workers reported that a transcriptional factor named hnRNP LL recognises the B-cell lymphoma gene-2 (Bcl-2) i-motif to activate transcription.[87] The Bcl-2 oncoprotein prevents apoptosis when overexpressed, and this has been linked to the development of lymphocytic cancer chemoresistance.[164] The study further explained that the cytosine rich Bcl-2 sequence exists in a dynamic equilibrium between an i-motif and a hairpin structure.[41] They identified two ligands; IMC-48 (Fig. 3.1) stabilises the i-motif structure and IMC-76 destabilises the i-motif to a hairpin (Fig. 3.2). The study suggested that the binding site for IMC-48 is within the central loop region of the i-motif due to the availability of stacking interactions with thymine residues, forming a capping structure within the central loop.
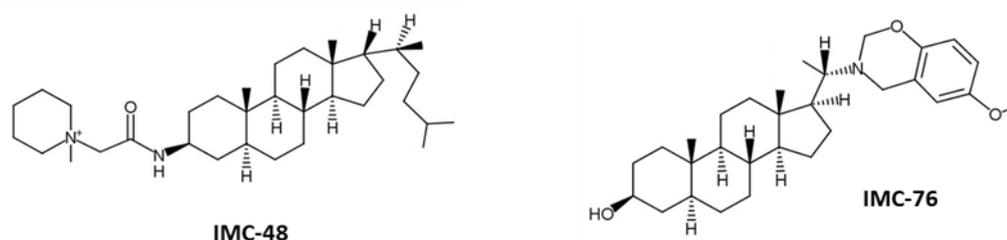
Figure 3.1: Schematic diagrams of the i-motif stabilising IMC-48 and destabilising IMC-76 ligands.

In cells with high Bcl-2 expression, the hairpin stabilising compound IMC-76 decreased Bcl-2 mRNA levels, whereas, in cells with low Bcl-2 expression, IMC-48 increased both the Bcl-2 mRNA and protein expression. These compounds are selective for Bcl-2 as experiments with c-MYC and VEGF i-motifs[84,165] resulted in no significant change in gene expression. The effect of IMC-76 on Bcl-2 expression was also demonstrated *in vivo* which gave a 20% decrease in Bcl-2 mRNA. Therefore, the group showed that IMC-76 is able to compete with hnRNP LL, resulting in a reduction in the amount of hnRNP LL-Bcl-2 i-motif complex present. This work highlighted a possible biological function of the i-motif in gene expression regulation.
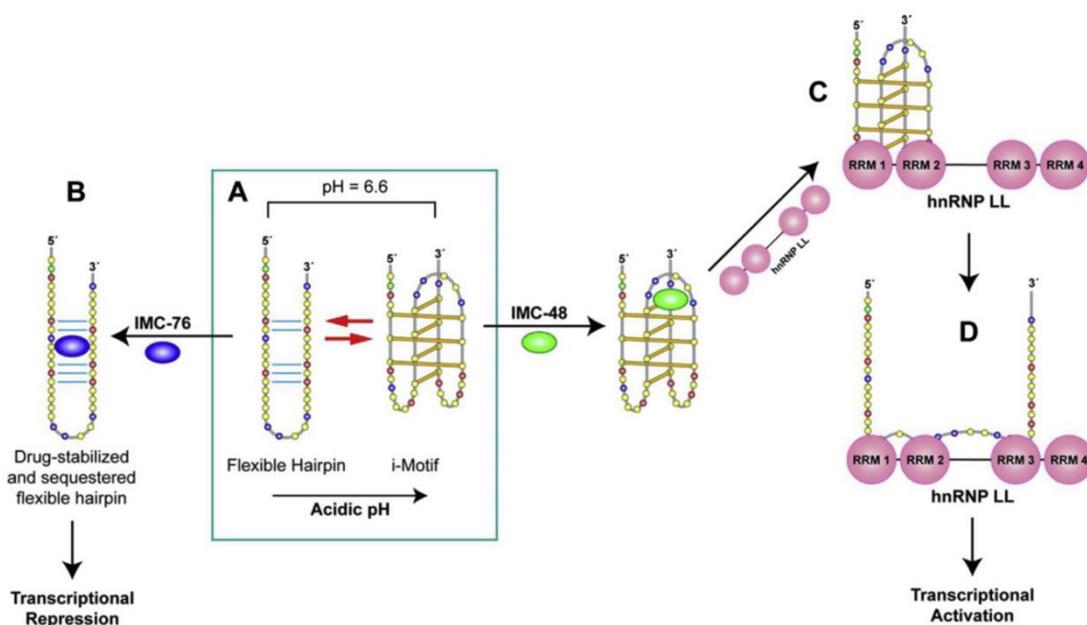
Figure 3.2: Conformational transitions and biological consequences that occur following binding of IMC-76, IMC-48, and hnRNP LL to the cytosine rich strand in the promoter of Bcl-2: (**A**) The different conformational states of the Bcl-2 promoter. Acidic pH favours i-motif whereas pH 6.6 has a mixture of hairpin and i-motif. (**B**) Addition of IMC-76 stabilises the hairpin and results in transcriptional repression. (**C**) Addition of IMC-48 stabilises the i-motif, the RNA recognition motifs (RRM) 1 and 2 of hnRNP LL bind to the first and third loops in the i-motif. (**D**) hnRNP LL–driven changes form an alternative conformation of the C-rich strand which results in transcriptional activation of Bcl-2.

In regard to potential therapeutic applications of DNA-ligand complexes, transition metal based polypyridyl complexes have been used as fluorescent probes for studying DNA properties.[166] Most of these ligands have little or no fluorescence in aqueous solution, however, upon binding to DNA, the ligand is in a hydrophobic environment and the solvent can no longer quench the intrinsic ligand fluorescence. Ruthenium polypyridyl complexes have been found to interact with single and double stranded DNA in solution and via crystallographic studies.[167,168] Λ-[Ru(phen)$_2$(p-HPIP)]$^{2+}$ has been reported to stabilise the human telomeric DNA G-quadruplex and subsequently inhibit telomerase activity.[169] [Ru(phen)$_2$(4idip)](ClO$_4$)$_2$ and [Ru(bpy)$_2$(4idip)](ClO$_4$)$_2$ have been able to induce the formation and stabilisation of telomeric G-quadruplex,[170] therefore, showing potential for photodynamic therapy using this family of complexes.[166,171]

[Ru(phen)$_2$(dppz)]$^{2+}$ (Fig. 3.3) is another polypyridyl complex whose luminescence in solution enhances in the presence of DNA.[168] A minimum of six bases of single stranded oligonucleotides are needed for this complex to exhibit the "light switch" effect.[172] This effect arises as the complex intercalates in the DNA base pairs to protect the nitrogens in dppz from water.[173] A study adding [Ru(phen)$_2$(dppz)]$^{2+}$ to the human telomeric G-quadruplex and i-motif proposed the idea that the complex may stack onto the G-quadruplex ends and non-specifically bind to the i-

_____

motif.[94] The weaker binding with the i-motif suggested that the complex may interact with the negatively charged sugar-phosphate backbone.
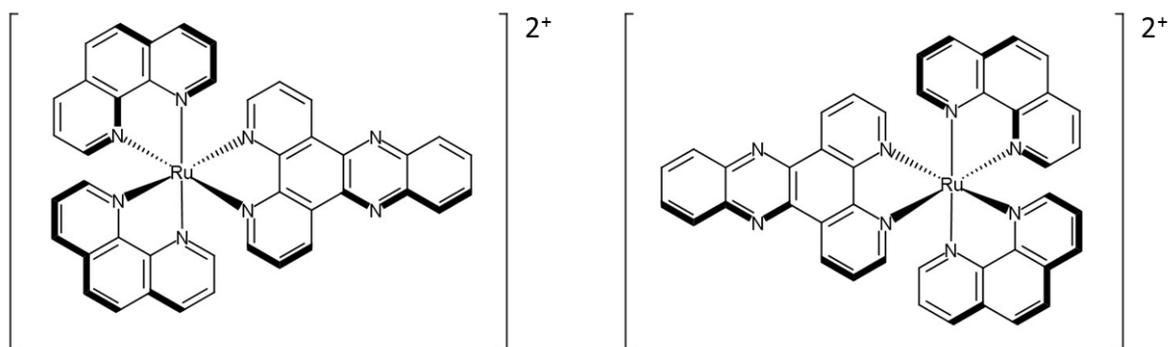
Figure 3.3: Schematic representation of of (left) Λ-[Ru(phen)$_2$dppz]$^{2+}$ and (right) Δ-[Ru(phen)$_2$dppz]$^{2+}$.

Chapter 2 covered the importance of loop length on i-motif stability. This chapter will focus on the effect of [Ru(phen)$_2$(dppz)]$^{2+}$ on i-motifs with increasing loop lengths. The i-motif forming sequences of the C$_3$T$_X$ series (where x = 3 to 8) were used to initiate the project. Although it has been suggested that in the case of an intramolecular i-motif, the complex may interact with phosphate backbones,[94] it was not clear where in the DNA structure this interaction may occur. The stable cytosine core of the intramolecular i-motif may not be prone to intercalation from ligands as suggested by the stacking of TMPyP4 at the ends of tetramolecular i-motif.[88] Instead, [Ru(phen)$_2$(dppz)]$^{2+}$ may interact with the loop regions, in similarity to the case of the Bcl-2 i-motif + IMC-48 complex. We hypothesised that the less stable i-motifs with longer loops would be more accommodating to the [Ru(phen)$_2$(dppz)]$^{2+}$ complex by providing more surface area for the ligand to interact with the i-motif.

_____

3.2 Materials and Methods

3.2.1 Luminescence Titrations

Fluorescence spectroscopy measurements were performed using a 1 cm pathlength quartz cell at 24°C. Oligonucleotides were first annealed in 50 mM sodium cacodylate buffer (pH 5.0) by heating to 90°C and slowly cooling to room temperature. A solution of $[Ru(phen)_2(dppz)]^{2+}$ (40 μM) was prepared in the same buffer. Concentrated stock solutions of the annealed i-motif forming sequences were then added until a final concentration of 80 μM ss and the emission measured at λ_{excit} = 440 nm.

3.2.2 Circular Dichroism Measurements

Oligonucleotides were dissolved with $[Ru(phen)_2(dppz)]^{2+}$ in 20 mM sodium cacodylate buffer of the appropriate pH to give a final concentration of 100 μM of both the complex and single-stranded DNA. The solutions were annealed by heating to 90°C and slowly cooling to room temperature. CD spectra were recorded at 20°C between 180 and 350 nm wavelengths with 1 nm wavelength increments in a 0.01 cm pathlength cuvette on beamline B23 at Diamond Light Source.

3.2.3 DNA Melting Experiments

UV absorption spectroscopies were carried out in triplicates using an Agilent Cary 100 with a temperature controlled six-cell changer. Oligonucleotides were dissolved in 50 mM sodium cacodylate buffer of the appropriate pH to a final concentration of 1 μM (ss) DNA to 1 μM $[Ru(phen)_2(dppz)]^{2+}$ complex, and were annealed by heating to 90°C and allowing them to cool slowly to room temperature. Absorption was recorded at 260 and 295 nm at 1°C intervals between 20-90°C, with a temperature change rate of 1°C/min in a 1 cm pathlength quartz cuvette.

3.2.4 pH Titrations

Oligonucleotides were dissolved with $[Ru(phen)_2(dppz)]^{2+}$ in 50 mM sodium cacodylate at pH 8 to give a final concentration of 1 μM each of the DNA (ss) and the complex. A spectrum was recorded between 200-350 nm at 25C. The pH was then decreased by 0.2 with the addition of 2M HCl and absorption values recorded in the same range of wavelength. This procedure was continued until a minimum pH of 4.5 was achieved. pH was measured in the cuvette using a Thermo Scientific Orion Star A11 pH meter equipped with a small diameter pH electrode.

_____

3.3 Results & Discussion

3.3.1 Determination of rac-[Ru(phen)₂(dppz)]²⁺ Binding to Long-Looped i-motifs

Emission titration experiments of *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ (Ru) with C$_3$T$_X$ in pH 5 showed that C$_3$T$_7$ and C$_3$T$_8$ gave higher luminescence intensities in comparison to when the loop length (X) is between three and six thymine residues (Appendix fig. A3.1). Figure 3.4 shows that like most polypyridyl complexes, Ru on its own showed negligible luminescence in aqueous solution but addition of the i-motifs resulted in the "light switch" effect. Figure 3.5 illustrates that the luminescence intensities of the complex with C$_3$T$_X$ at λ_{max} = 615 nm are highest for C$_3$T$_7$ and C$_3$T$_8$ than the rest of the oligonucleotides, implying that the complex may have a higher binding affinity to the i-motif when X = 7 or 8. Since the titration studies showed that longer looped i-motifs gave maximum luminescence regardless of DNA:Ru ratios, the 1:1 DNA to complex ratio was used for the following CD and UV measurements.
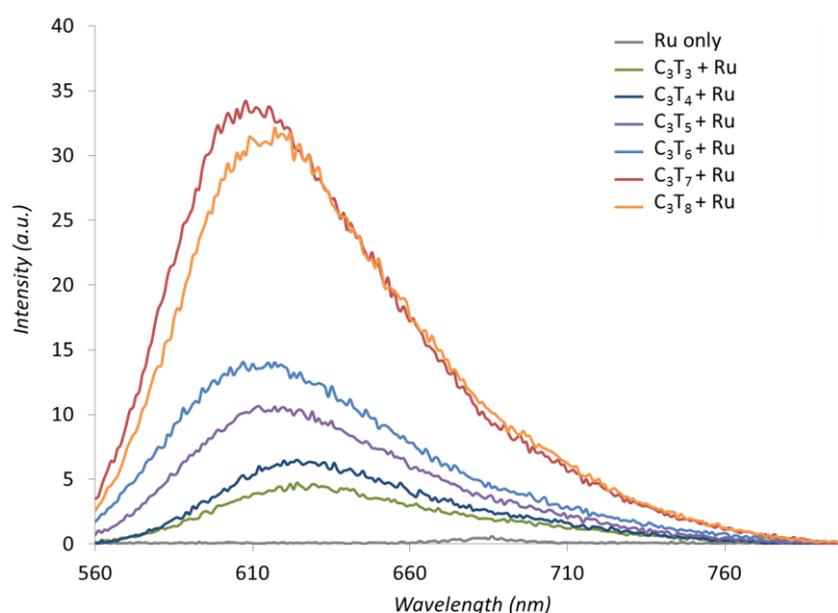
Figure 3.4: Fluorescence emission spectra of 40 µM *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ in the presence of 40 µM single-stranded C$_3$T$_X$ DNA i-motifs in 50 mM sodium cacodylate (pH 5.0). λ_{excit} = 440 nm.
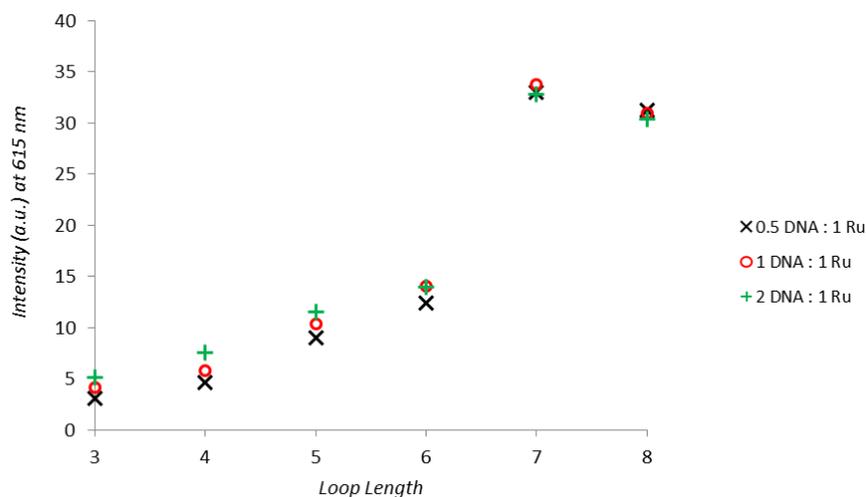
Figure 3.5: Fluorescence intensity of *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ *vs.* loop length of the C$_3$T$_X$ i-motifs at λ$_{max}$ of 615 nm. The concentration of the ruthenium complex and single-stranded C$_3$T$_X$ used were 40 µM.

3.3.2 Conservation of the i-motif Structure in the Presence of rac-[Ru(phen)$_2$(dppz)]$^{2+}$

CD spectroscopy of the C$_3$T$_X$ sequences with *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ in pH 5 revealed that the addition of the complex didn't disrupt the i-motif structures (Fig. 3.6). There is little difference in the position of the peaks between the native DNA and the DNA with added complex but the maxima and minima observed lie between 285–295 and 260-265 nm wavelengths, respectively, which are characteristic of intramolecular i-motifs. In similarity to the CD spectroscopy results of the Bcl-2 i-motif-IMC-48 ligand complex,[41] CD spectra of the C$_3$T$_X$ sequences with the ruthenium complex illustrated that i-motif stabilisation by the ligand cannot be observed in individual scans but that addition of the complex doesn't change the i-motif structures.
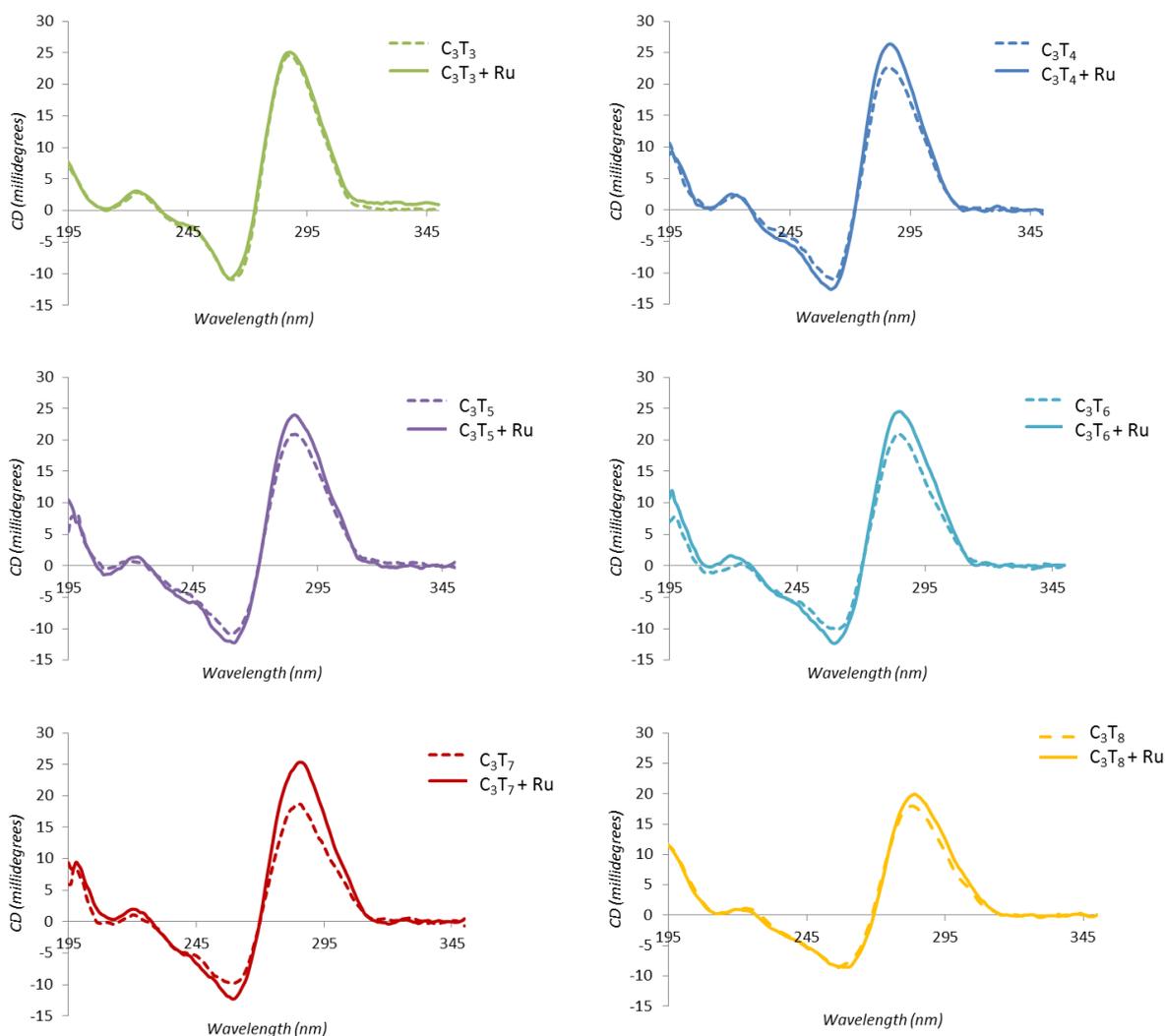
Figure 3.6: CD spectra of native C_3T_X and the oligonucletides with *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ in 20 mM sodium cacodylate buffer at pH 5.

3.3.3 Stabilisation of C_3T_7 and C_3T_8 by rac-[Ru(phen)$_2$(dppz)]$^{2+}$

Stabilisation of an i-motif by *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ can be determined by an increase of melting temperature (T_m) of the DNA in the presence of the complex. Figure 3.7 illustrates the melting curves observed for the C_3T_X i-motifs with *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ in a 1:1 DNA to complex ratio at 260 nm wavelength. In comparison to the T_m of the i-motifs on their own, the DNA melts showed no change in stability of i-motifs where X is three to six thymine residues long. However, addition of the complex to C_3T_7 and C_3T_8 increased the T_m values, hence the stability of the two i-motif structures.

_____

Chapter 2 showed that as loop length increases, the T_m value of the i-motif decreases at a rate of 4.7°C per thymine residue with an almost linear relationship (Table 3.1). When *rac*-[Ru(phen)$_2$dppz]$^{2+}$ is introduced to the i-motifs, this rate of change in temperature is similar at 4.8°C from C$_3$T$_3$ to C$_3$T$_6$. However, temperature increases from C$_3$T$_7$ to C$_3$T$_8$ by 1.0°C upon addition of the complex. Figure 3.8 demonstrates that addition of the ruthenium complex affected the stability of C$_3$T$_8$ the most, with an increment in T_m by 8.0°C. Stability of C$_3$T$_7$ also increased in the presence of the complex by 3.5°C. The T_m values recorded thus show that *rac*-[Ru(phen)$_2$dppz]$^{2+}$ helps stabilise i-motifs when X is greater than six bases.
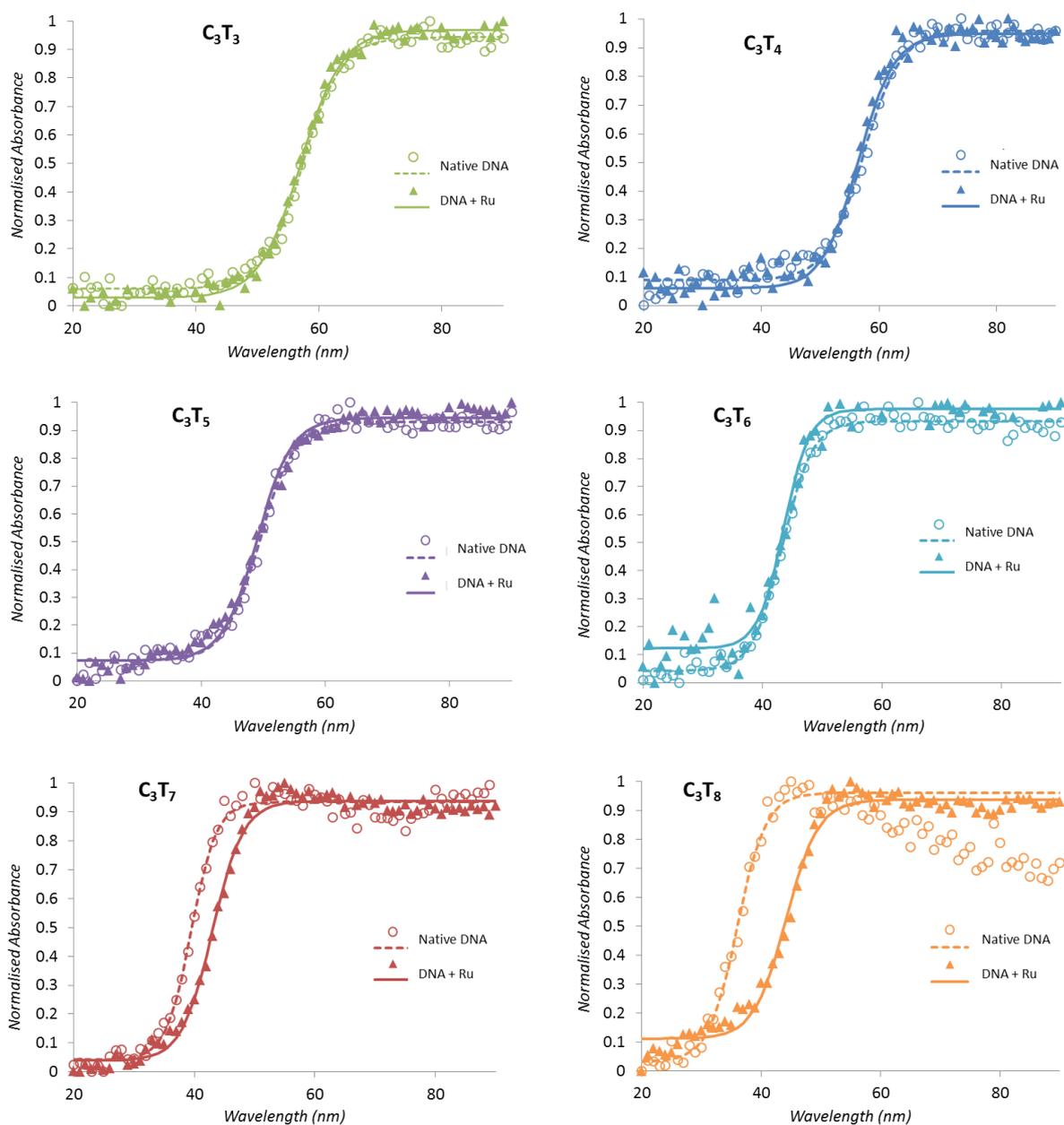
Figure 3.7: Melting curves of the six i-motif forming C$_3$T$_X$ (x = 3 to 8) sequences in their native states and in the presence of *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ (1 single-stranded DNA/1 complex). The solutions were made in 50 mM sodium cacodylate (pH 5) and the measurements recorded at 260 nm wavelength.

_____

Table 3.1: DNA melting temperatures of the i-motif forming C_3T_X sequences with and without *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ (Ru) in 50 mM sodium cacodylate (pH 5).

I-motif sequence	Melting temperature, T_m (°C)		ΔT_m (°C)
	no Ru	*with Ru*	
C_3T_3	57.0	57.0	0.0
C_3T_4	57.0	56.5	0.5
C_3T_5	49.0	49.0	0.0
C_3T_6	43.0	43.5	0.5
C_3T_7	39.5	43.0	3.5
C_3T_8	36.0	44.0	8.0
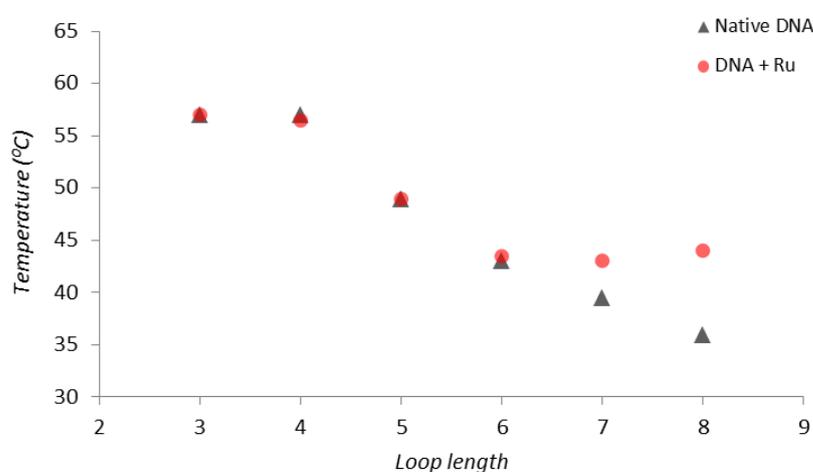
Figure 3.8: Melting temperatures of the C_3T_x i-motifs with *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ against loop length. X = no. of thymine residues in the i-motif loops.

3.3.4 The Effect of rac-[Ru(phen)₂dppz]²⁺ on i-motifs with Alternating Loop Lengths

In order to further confirm that long loops are necessary for the binding of *rac*-[Ru(phen)$_2$dppz]$^{2+}$ to i-motifs, the ligand was added to two i-motif forming sequences of alternating loop lengths; C_3T_{383} and C_3T_{838}. From the results observed above, it was suggested that if the complex only helps stabilise i-motifs with loops longer than six bases, *rac*-[Ru(phen)$_2$dppz]$^{2+}$ would increase the T_m of C_3T_{838} but not of C_3T_{383}. As expected, the DNA melts of the two oligonucleotides (Fig. 3.9) show that there is no difference in the T_m of native C_3T_{383} and that of the DNA with the ruthenium complex, whereas, there is a 6°C increment in the T_m of C_3T_{838} in the presence of the complex. This is supported by the records of measurements at both 260 and 295 nm wavelengths. This study thus supports the notion that longer looped i-motifs are required for the binding of *rac*-[Ru(phen)$_2$dppz]$^{2+}$.

A. C$_3$T$_{383}$

260 nm

295 nm

Native DNA
DNA + Ru

B. C$_3$T$_{838}$

260 nm

295 nm
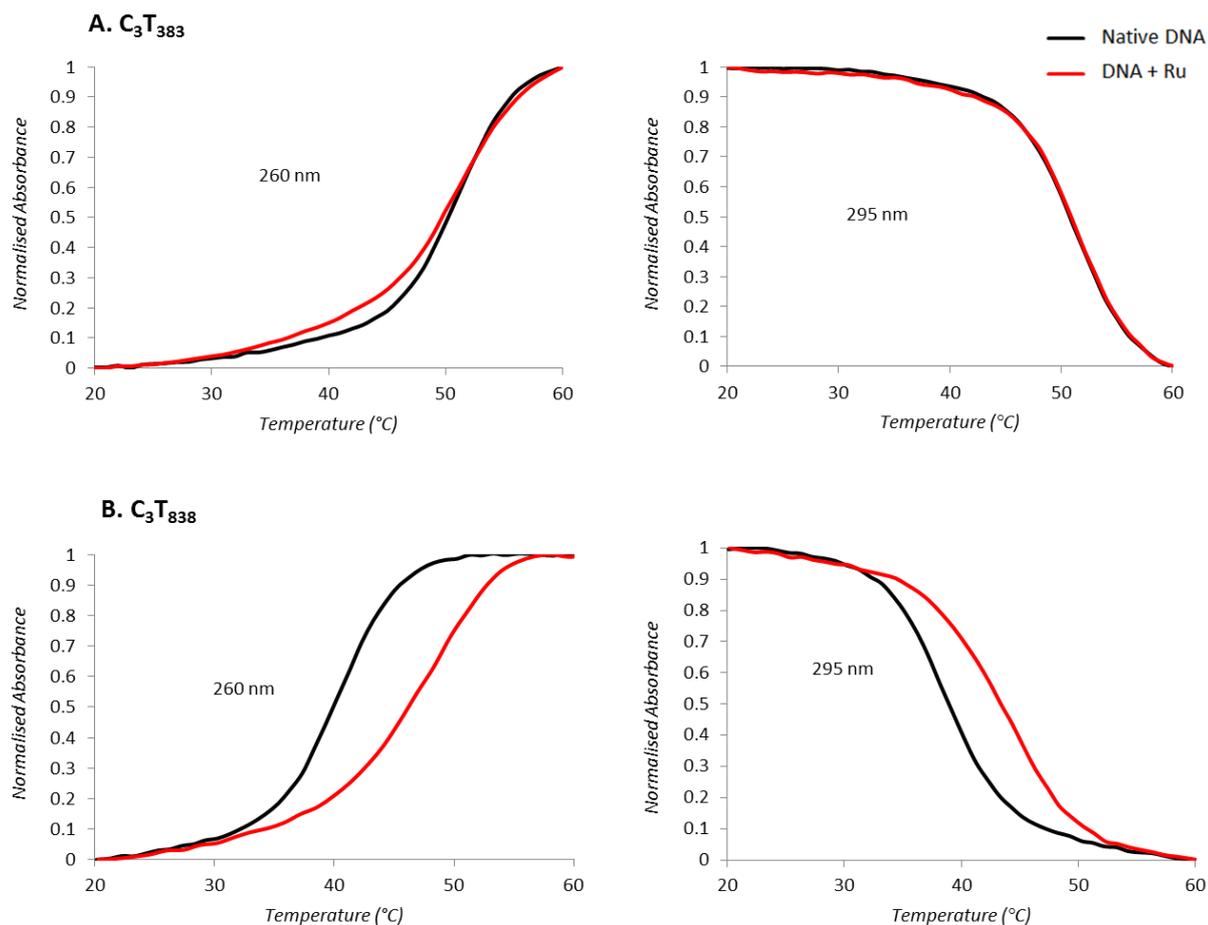
Figure 3.9: DNA melting curves of (**A**) C$_3$T$_{383}$ and (**B**) C$_3$T$_{838}$ in their native forms and in the presence of *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ (1 single-stranded DNA to 1 complex ratio) in 50 mM sodium cacodylate (pH 5).

3.3.5 The effect of [Ru(phen)$_2$dppz]$^{2+}$ Enantiomers on i-motifs

DNA melts of Λ and Δ-[Ru(phen)$_2$dppz]$^{2+}$ in the presence of the human telomeric i-motif; (C$_3$TAA)$_3$C$_3$ and C$_3$T$_{838}$ were conducted. Figure 3.10 illustrates that there is no difference in the T_m of C$_3$T$_{838}$ and the human telomeric i-motif when the two enantiomers used..

Luminescence melting of C$_3$T$_3$ with the two enantiomers showed no difference in spectra (Fig. 3.11) but the melting temperatures of C$_3$T$_{838}$ showed that there is a 1.0°C difference between use of two enantiomers and that the DNA with the Λ-enantiomer is more luminescent than the Δ. This is in contrast to the studies made with B-DNA where although both enantiomers can intercalate into the DNA structure, the Δ is known to bind more strongly and is also more emissive than its Λ counterpart.[174,175]

_____
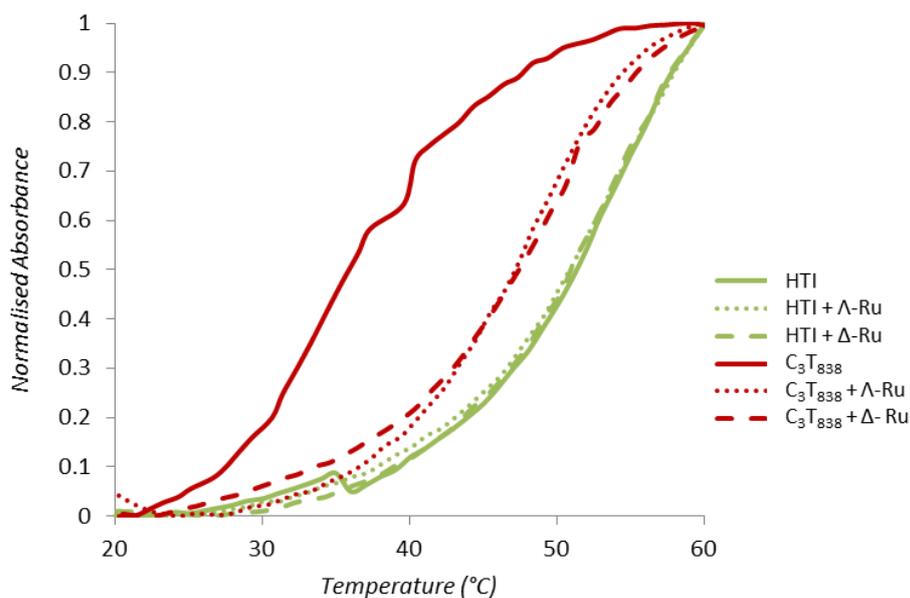
Figure 3.10: DNA melting curves of the human telomeric i-motif; d(C$_3$TAA)$_3$C$_3$ (HTI) and C$_3$T$_{838}$ in their native forms and in the presence of Λ and Δ-[Ru(phen)$_2$(dppz)]$^{2+}$ in 50 mM sodium cacodylate (pH 5).
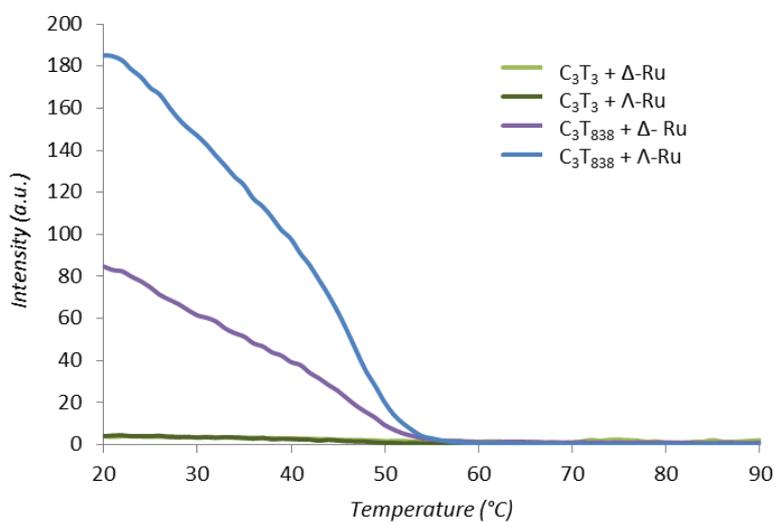
Figure 3.11: Fluorescence melting behaviours of C$_3$T$_3$ and C$_3$T$_{838}$ at 615 nm in the presence of Λ and Δ-[Ru(phen)$_2$(dppz)]$^{2+}$ in 50 mM sodium cacodylate (pH 5).

_____

3.4 Conclusions

As loop length increases, the i-motif gains more flexibility, and therefore stabilising interactions within the native structure could become weaker, enabling duplex formation.[176] Using DNA melting techniques, addition of *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ to i-motifs in the C$_3$T$_X$ series showed no change in stability when loop length was between three and six thymine residues. However, addition of the complex gave an increase in T_m of C$_3$T$_7$ and C$_3$T$_8$ by 3.5°C and 8.0°C, respectively. In similarity to the Bcl-2 i-motif stabilising and destabilising ligands; IMC-48 and IMC-76, CD spectra of the C$_3$T$_X$ oligonucleotides with the ruthenium complex showed no significant stabilisation of the i-motifs but that the i-motif structures are not disrupted by the addition of the complex. DNA melts of C$_3$T$_{383}$ and C$_3$T$_{838}$ further agree that stability of the i-motif with longer loops is enhanced by the ruthenium complex.

In chapter 2, the i-motif sequences in which changes were made to the first and last looping regions showed similar stability irrespective of which loop was changed. I-motif structures with two long loops were always less stable than those with two short loops. Changes made to the middle loop showed minimal disruption to i-motif stability. These results demonstrated that the important loops for stability must be the first and last loop. This is supported by a study made in 2015, where Sugimoto *et al.* reported that the first and third loop regions contribute to i-motif stability at neutral pH under molecular crowding conditions.[40] Preliminary studies of adding *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ showed no effect on the stability of C$_3$T$_{383}$ but a 6°C increment in the T_m value was observed for

Figure 3.12: Proposed binding scheme of [Ru(phen)$_2$(dppz)]$^{2+}$ to a long looped i-motif, in this case C$_3$T$_7$. The i-motif core is denoted by yellow spheres, the loops by grey spheres and the complex by the orange triangle.

C$_3$T$_{838}$, suggesting that the first and last loops play a role in ligand binding. The said theory is in contrast to the study of a small molecule binding to the middle loop of the i-motif structure in the Bcl-2 promoter,[41] a gene related to the evasion of apoptosis. However, in order to activate transcription, the transcriptional factor hnRNP LL was reported to recognise the Bcl-2 i-motif by binding to the first and last loops.[87] The Bcl-2 i-motif contains long loops of 8, 5 and 7.[162] Hence, the [Ru(phen)$_2$(dppz)]$^{2+}$ study needs to be repeated using C$_3$T$_{883}$ and C$_3$T$_{388}$ sequences to further confirm which loop regions are crucial in ligand binding. As *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ is known to intercalate into B-DNA structures, the two long loops may provide an intercalation site for the ligand to bind into the i-motif (Fig. 3.12).

_____

Luminescence intensities of C_3T_7 and C_3T_8 were approximately 2.4 times higher than that of the rest of the i-motifs in the C_3T_X system. This supports the DNA melting experiments that $[Ru(phen)_2(dppz)]^{2+}$ may interact with loops that are longer than six thymine bases. Luminescence melting of C_3T_3 with the two enantiomers showed no difference in spectra but the melts of C_3T_{838} showed that emission of the DNA with the Λ-enantiomer is higher than that of the Δ. Traditionally, it is the Λ-enantiomer that has been shown to be less emissive with B-DNA in solution.[175,177]

In conclusion, the results demonstrate that i-motif stability by $[Ru(phen)_2(dppz)]^{2+}$ is related to C-rich sequence composition. The length of the loop region is a large contributing factor to the overall stability of the i-motif structure. $[Ru(phen)_2(dppz)]^{2+}$ did not have an effect on short looped i-motifs which were shown to be most stable in Chapter 2. Instead, the ruthenium complex increased the stability of the less stable, long looped i-motifs. This adds to the notion that an i-motif stabilising ligand may need to be present in order for a cytosine-rich sequence to form an i-motif under physiological conditions. This concept has also been previously proposed by Hurley and co-workers using the IMC-48 ligand with the Bcl-2 cytosine-rich sequence to form an i-motif structure.[87]

The Importance of Loop Composition on i-motif Formation

4.1 Introduction
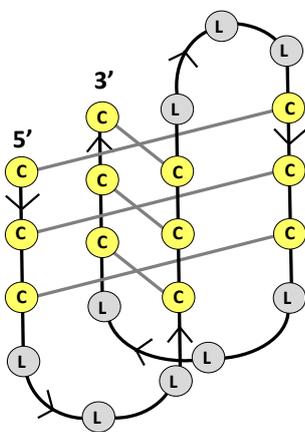
Figure 4.1: Schematic representation of an intramolecular i-motif of the sequence d[$(C_3L_3)_3C_3$], where C is cytosine and L denotes any DNA heterocyclic bases.

Intramolecular i-motif structures have a general sequence in the format: $C_{3-5}L_{1-7}C_{3-5}L_{1-7}C_{3-5}L_{1-7}C_{3-5}$, where C is cytosine and L is the loop region comprised of any DNA base including cytosine (Fig. 4.1). Chapter 2 showed that i-motifs with short loops exhibit the highest stability,[131] however, the study was made using oligonucleotides containing loops of only thymine bases. In 2015, Sugimoto *et al.* reported that i-motifs with the first and last loops that are rich in pyrimidine are more stable in comparison to those with purine.[40] The study was made using 11 oligonucleotides which showed that intrastrand i-motifs with two guanine or two thymine bases gave high stabilities. Hydrogen bonding among nucleobases have a stronger effect on the stability of the i-motif structures than the stacking of the nucleobases.

A recent study on bimolecular i-motifs with different loop lengths and composition using silver nanoclusters showed that the DNA sequence 5'-CCCC*ATAT*CCCC-3' gave the brightest emission of the fluorescent probe with a quantum yield of 0.78. However, the reversed template sequence 5'-CCCC*TATA*CCCC-3' gave the quantum yield of 0.15,[178] emphasising on the importance of the direction of base alignment. Although research on i-motif stability depending on loop length has been expanding, the role of varying DNA residues within the loop regions is still less understood. This chapter will focus on how loop composition is crucial in intramolecular i-motif structure stability using 36 C-rich sequences. The study was initially made with the idea of determining i-motif stabilities but results showed that loop base composition and directionality are both contributing factors of i-motif structure formation.

The sequences studied were limited to four tracts of three cystosine bases and three loops containing three DNA bases; modelled after the C_3T_3 sequence in chapter 2 which gave the highest stability. The sequences used were in the format; d(CCC**XYZ**CCC**XYZ**CCC**XYZ**CCC) (or C_3XYZ), where X and Z are adenine, thymine or guanine and Y is any one of the four DNA bases (Fig. 4.2, Table 4.1). Cytosine was not included for X and Z residues as this would increase the cytosine tract length and i-motif forming sequences with tracts of at least five cytosine bases have shown increased stability due to increased number of $C^+ \cdot C$ base pairs.[38] All oligonucleotides were studied using UV and srCD spectroscopies.

_____

Table 4.1: C$_3$XYZ sequences used in this study, where X and Z are adenine, thymine and guanine. Y is any of the four DNA bases.

Group	Abbreviation	Sequence 5' → 3'
C$_3$**A**YZ	AAA	CCCAAACCCAAACCCAAACCC
	AAT	CCCAATCCCAATCCCAATCCC
	AAG	CCCAAGCCCAAGCCCAAGCCC
	ATA	CCCATACCCATACCCATACCC
	ATT	CCCATTCCCATTCCCATTCCC
	ATG	CCCATGCCCATGCCCATGCCC
	AGA	CCCAGACCCAGACCCAGACCC
	AGT	CCCAGTCCCAGTCCCAGTCCC
	AGG	CCCAGGCCCAGGCCCAGGCCC
	ACA	CCCACACCCACACCCACACCC
	ACT	CCCACTCCCACTCCCACTCCC
	ACG	CCCACGCCCACGCCCACGCCC
C$_3$**T**YZ	TAA	CCCTAACCCTAACCCTAACCC
	TAT	CCCTATCCCTATCCCTATCCC
	TAG	CCCTAGCCTAAGCCCTAGCCC
	TTA	CCCTTACCCTTACCCTTACCC
	TTT	CCCTTTCCCTTTCCCTTTCCC
	TTG	CCCTTGCCCTTGCCCTTGCCC
	TGA	CCCTGACCCTGACCCTGACCC
	TGT	CCCTGTCCCTGTCCCTGTCCC
	TGG	CCCTGGCCCTGGCCCTGGCCC
	TCA	CCCTCACCCTCACCCTCACCC
	TCT	CCCTCTCCCTCTCCCTCTCCC
	TCG	CCCTCGCCCTCGCCCTCGCCC
C$_3$**G**YZ	GAA	CCCGAACCCGAACCCGAACCC
	GAT	CCCGATCCCGATCCCGATCCC
	GAG	CCCGAGCCCGAGCCCGAGCCC
	GTA	CCCGTACCCGTACCCGTACCC
	GTT	CCCGTTCCCGTTCCCGTTCCC
	GTG	CCCGTGCCCGTGCCCGTGCCC
	GGA	CCCGGACCCGGACCCGGACCC
	GGT	CCCGGTCCCGGTCCCGGTCCC
	GGG	CCCGGGCCCGGGCCCGGGCCC
	GCA	CCCGCACCCGCACCCGCACCC
	GCT	CCCGCTCCCGCTCCCGCTCCC
	GCG	CCCGCGCCCGCGCCCGCGCCC

$$\begin{matrix} X & Y & Z \end{matrix}$$
$$\begin{pmatrix} A \\ T \\ G \end{pmatrix} \begin{pmatrix} A \\ T \\ G \\ C \end{pmatrix} \begin{pmatrix} A \\ T \\ G \end{pmatrix}$$

Figure 4.2: (Top) Matrix showing DNA bases denoting to X, Y and Z loop positions in cytosine-rich sequences of the format C$_3$XYZ. (Below) Schematic representation of the intramolecular i-motif structure in the same arrangement.

_____

4.2 Materials and Methods

All oligonucleotides were purchased from Eurogentec. The concentrations of all oligonucleotides were calculated from the absorbance value at 260 nm using their extinction coefficients of the nearest neighbour model provided by the manufacturer. All experiments for each sequence were run in triplicates.

4.2.1 UV Absorption Measurements

Each solution was made in 50 mM sodium cacodylate buffer with a final DNA concentration of 1 μM (ss). The pathlength of the cuvettes used was 1 cm. UV absorption spectroscopy was carried out using an Agilent Cary 100 with a temperature controlled six cell changer.

(a) DNA Melting Experiments

Oligonucleotides were dissolved in 50 mM sodium cacodylate buffer of the appropriate pH to a final concentration of 1 μM (ss), and were annealed by heating to 90°C and allowing them to cool slowly to room temperature. Absorption was recorded at 260 and 295 nm at 1°C intervals between 20-90°C, with a temperature change rate of 1°C/min in a 1 cm pathlength quartz cuvette.

(b) pH Titrations

Oligonucleotides were dissolved in 50 mM sodium cacodylate at pH 8, and a spectrum recorded between 200-350 nm at 25°C. The pH was decreased in steps of 0.2 units with the addition of 2M HCl and absorption values recorded. This procedure was continued until a minimum pH of 4.5 was achieved. pH was measured in the cuvette using a Thermo Scientific Orion Star A11 pH meter equipped with a small diameter pH electrode.

4.2.2 Circular Dichroism Measurements

Oligonucleotides were dissolved in 20 mM sodium cacodylate buffer of the appropriate pH to a final concentration of 200 μM (ss), and were annealed by heating to 90°C and slowly cooling to room temperature. CD spectra were recorded at 20°C between 180 and 350 nm wavelengths with 1 nm wavelength increments in high-throughput cell strips of 0.02 cm pathlength at beamline B23, Diamond Light Source.

_____

4.3 Results & Discussion

4.3.1 Analyses of DNA Stability using UV Absorption Spectroscopy

The ionic strength of the buffer solution was restricted to the standard protocol of 50 mM because studies under UV light as high levels of salt (>100 mM) are known to destabilise i-motif structures.[67] For both DNA melting and pH titrations, the final concentration of each DNA used was 1 µM single-stranded (ss) in order to read reasonable absorbance with 1 cm pathlength, i.e. not below or over the limit of detection in the UV spectrometer. All C_3XYZ oligonucleotides showed an increase in absorbance at 260 nm as a result of π-unstacking and denaturation by heat. The transition at 295 nm indicating the degradation of C^+·C tetraplex with heat was observed in all melting curves at pH 5 (Appendix Figures A4.4-A4.6).

Figure 4.3 shows the DNA melting curves of C_3AYZ sequences where the first base in the loop regions is adenine. The results show that when the first and second bases in the loops are kept constant, and only the third base is changed, the sequences are more stable when Z is thymine and less stable when Z is adenine or guanine, with the exception of AGG. Out of the 12 C_3AYZ sequences, AGG gave the highest T_m of 59°C followed by ATT with 58°C. ATT and AGG are in agreement with Sugimoto and co-workers' report that the presence of two guanine and two thymine bases give high thermodynamic stabilities.[40]
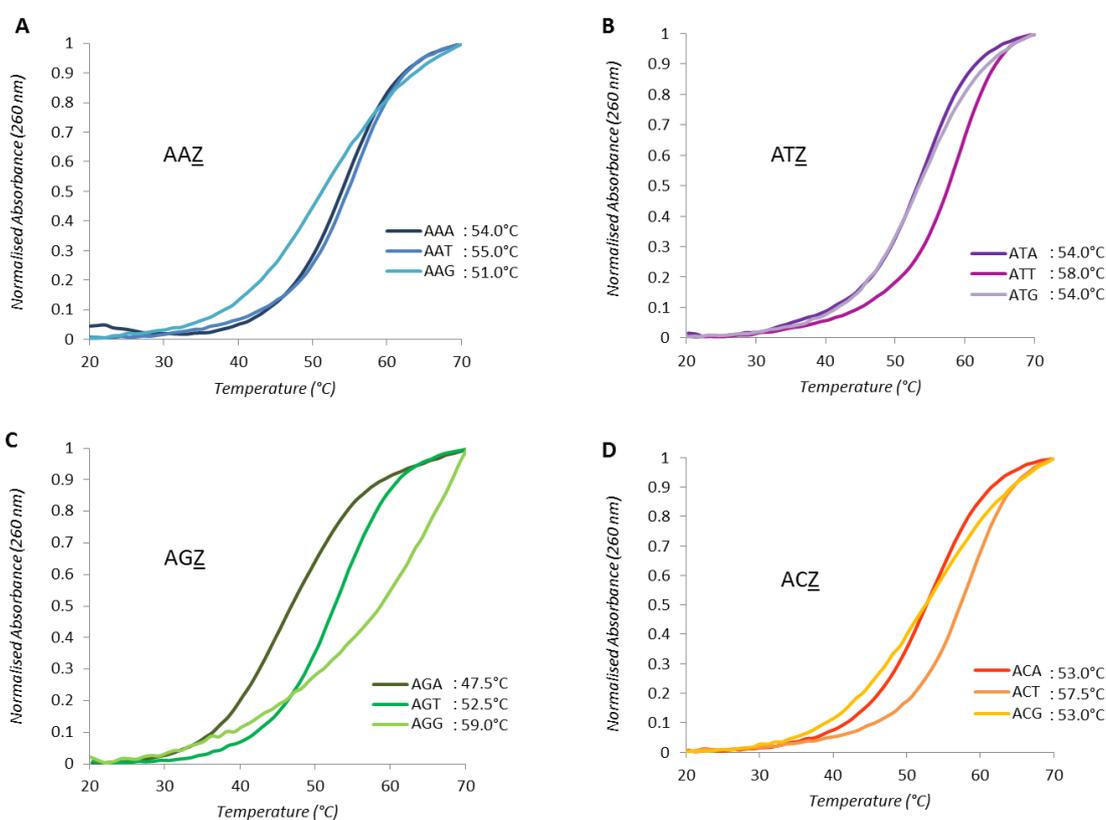
Figure 4.3: DNA melting profiles of C_3AYZ (1 µM, ss) showing the effect of the third base in the loop (Z) on thermal stability when Y is (A) adenine, (B) thymine, (C) guanine and (D) cytosine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 260 nm.

_____

Figure 4.4A compares the DNA melting curves of all C$_3$*A*YZ sequences. AGG shows the highest thermal stability and AGA the lowest with a 11.5°C difference in T_m between the two sequences. The presence of adenine bases in a 2015 study gave a decrease in T_m by approximately 7°C in comparison to when thymine or guanine was present instead.[40] Figures 4.4B-D show that changing the second base (Y) to adenine, thymine or cytosine, affects the thermal stabilities by up to 3°C, whereas when Y is guanine, the T_m value is either the highest or the lowest of its respective groups.
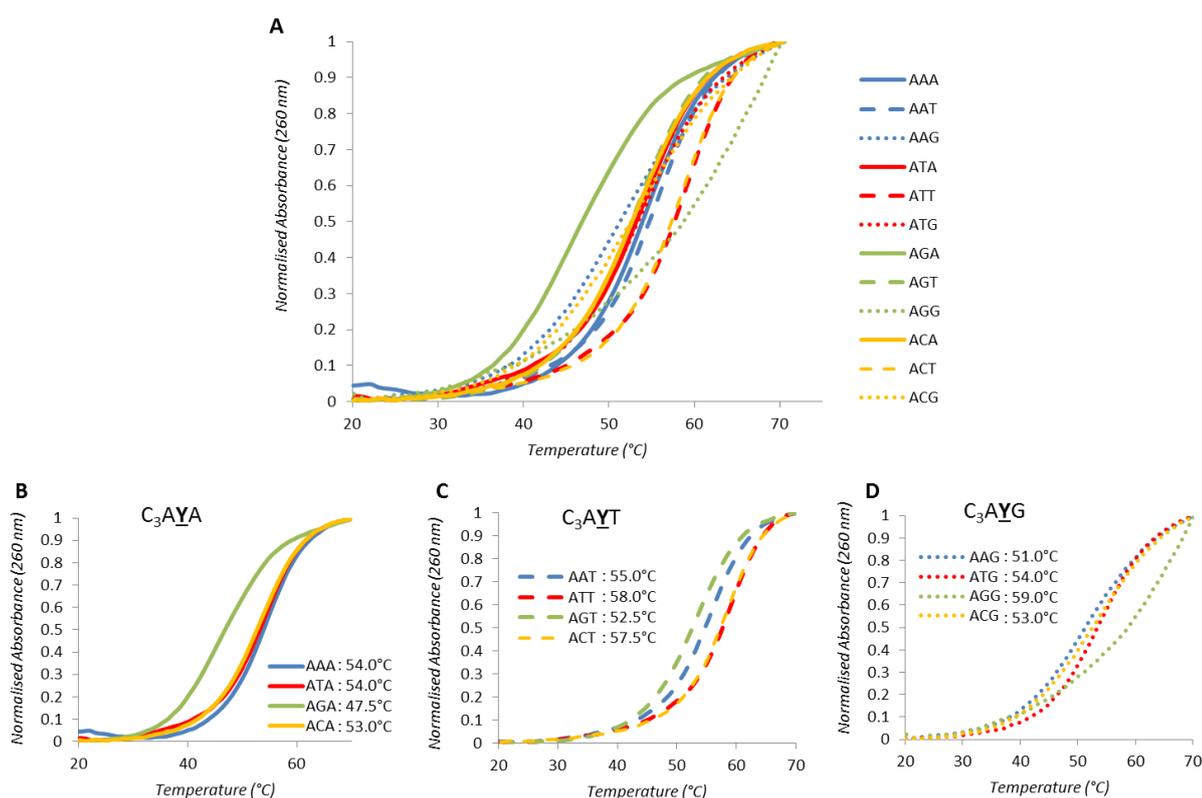
Figure 4.4: (A) DNA melting profiles (λ=260 nm) of all twelve C$_3$*A*YZ oligonucleotides (1 µM, ss) in 50 mM sodium cacodylate buffer (pH 5). Y is any of the four DNA bases and Z is adenine, thymine or guanine. The three smaller graphs show the effect on thermal stability by the middle base (Y) when Z is (**B**) adenine, (**C**) thymine and (**D**) guanine.

Figure 4.5 shows the DNA melting curves of C$_3$*T*YZ sequences where the first base in the loop regions is thymine. In similarity to C$_3$*A*YZ, the results show that when the first and second bases in the loops are kept constant, the sequences are more stable when Z is thymine and less stable when Z is guanine; with the exception of TGG which shows an almost linear relationship between absorbance and temperature instead of the classic sigmoidal DNA melting profile. Out of the 12 C$_3$*T*YZ sequences, TCT gave the highest T_m of 61°C and TGG gave the lowest T_m of 51°C (Fig. 4.6A). Figures 4.6B-D also show that change in Y to adenine, thymine or cytosine, affects the thermal stabilities by up to 3°C. The T_m values are significantly lowered when Y is guanine, implying that the identity of the second base in the loop also contributes to thermal stability, especially when it is guanine as it has so far shown both stabilisation and destabilisation of the C-rich DNA sequences.
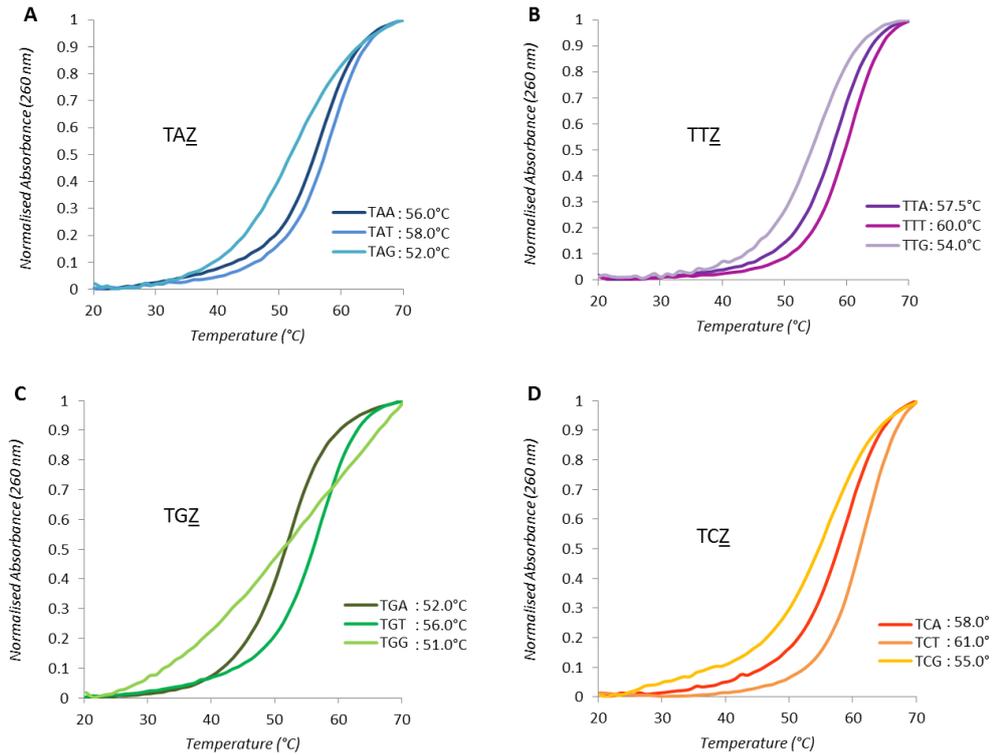
Figure 4.5: DNA melting profiles of C_3TYZ (1 µM, ss) showing the effect of varying the third base in the loop (Z) on thermal stability when Y is (A) adenine, (B) thymine, (C) guanine and (D) cytosine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 260 nm.
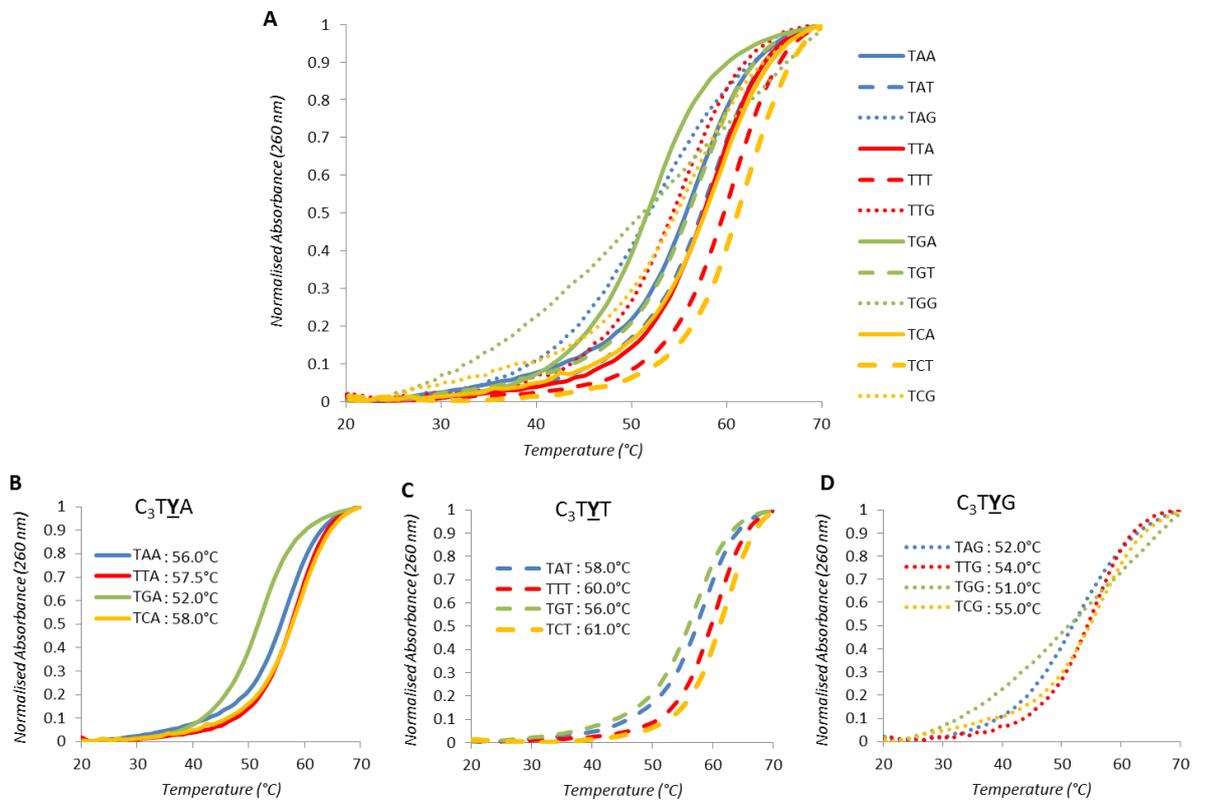
Figure 4.6: (**A**) DNA melting profiles (λ=260 nm) of C_3TYZ oligonucleotides (1 µM, ss) in 50 mM sodium cacodylate buffer (pH 5). Y is any of the four DNA bases and Z is adenine, thymine or guanine. The three smaller graphs show the effect on thermal stability by the middle base (Y) when Z is (**B**) adenine, (**C**) thymine and (**D**) guanine.

_____

In agreement to both C$_3$*A*YZ and C$_3$*T*YZ, the results obtained from C$_3$*G*YZ show that the sequences are more stable when Z is thymine but their stabilities are lowered when Z is either adenine or guanine, with the exception of GGG (Fig. 4.7). When the 12 C$_3$*G*YZ sequences are compared amongst themselves, (Fig. 4.8A), GTT and GAT display the highest stabilities with T_m of 65°C. The lowest T_m recorded out of the 12 oligonucleotides is 51°C for GCG. It is interesting that the opposite directionality of GAT gave a one of the least stable C$_3$*T*YZ sequences; TAG, T_m = 52°C. The same can be applied for TTG with a T_m of 54°C. Figures 4.8B-D show that when Z is constant, having Y as guanine instead of adenine, thymine or cytosine, affects the thermal stabilities of the C$_3$*G*YZ sequences the most.
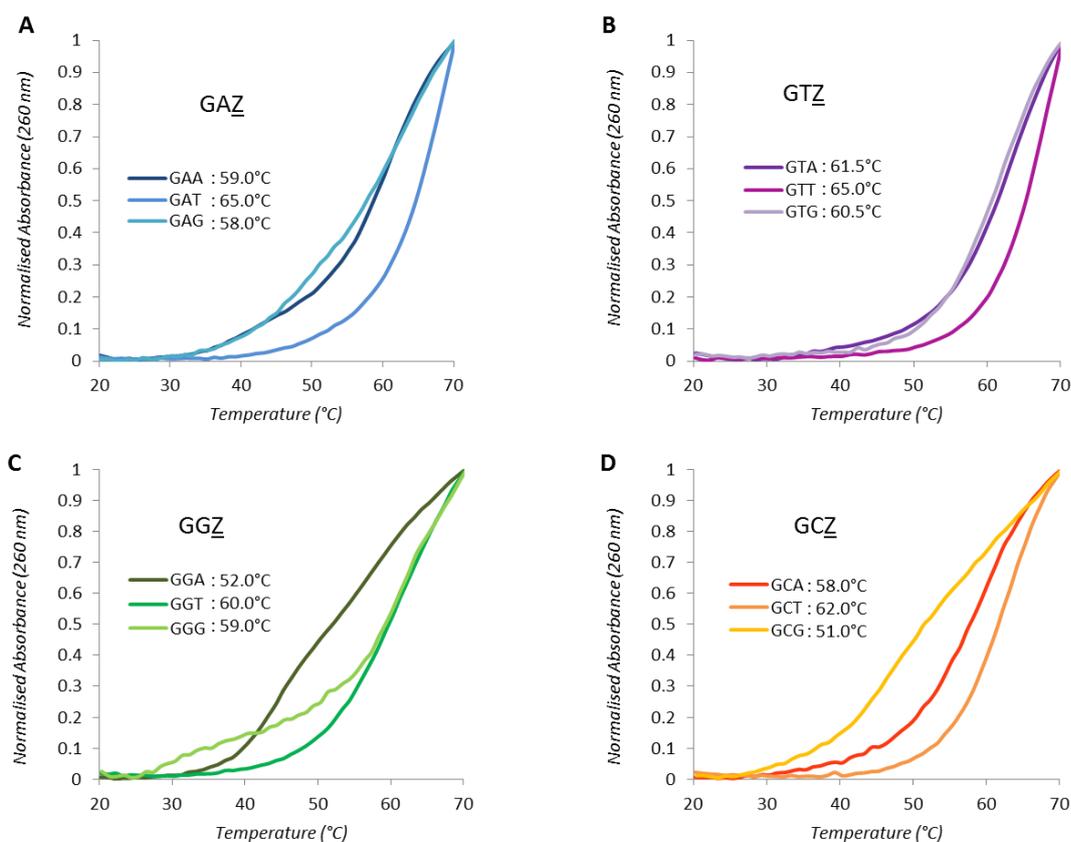
Figure 4.7: DNA melting profiles of C$_3$*G*YZ (1 μM, ss) showing the effect of varying the third base in the loop (Z) on thermal stability when Y is is (A) adenine, (B) thymine, (C) guanine and (D) cytosine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 260 nm.

Overall, for all C$_3$XYZ oligonucleotides, the middle base in the loop regions also contribute to changes in thermal stability. C$_3$XYZ is thermodynamically most stable when Y is thymine and least stable when Y is guanine, with the exception of AGG and GGG. The T_m values of C$_3$*G*YZ oligonucleotides are also significantly higher than those of C$_3$*A*YZ or C$_3$*T*YZ. In terms of thermal stabilities only, the C$_3$XYZ sequences can be categorised in the order: C$_3$*G*YZ > C$_3$*T*YZ > C$_3$*A*YZ.
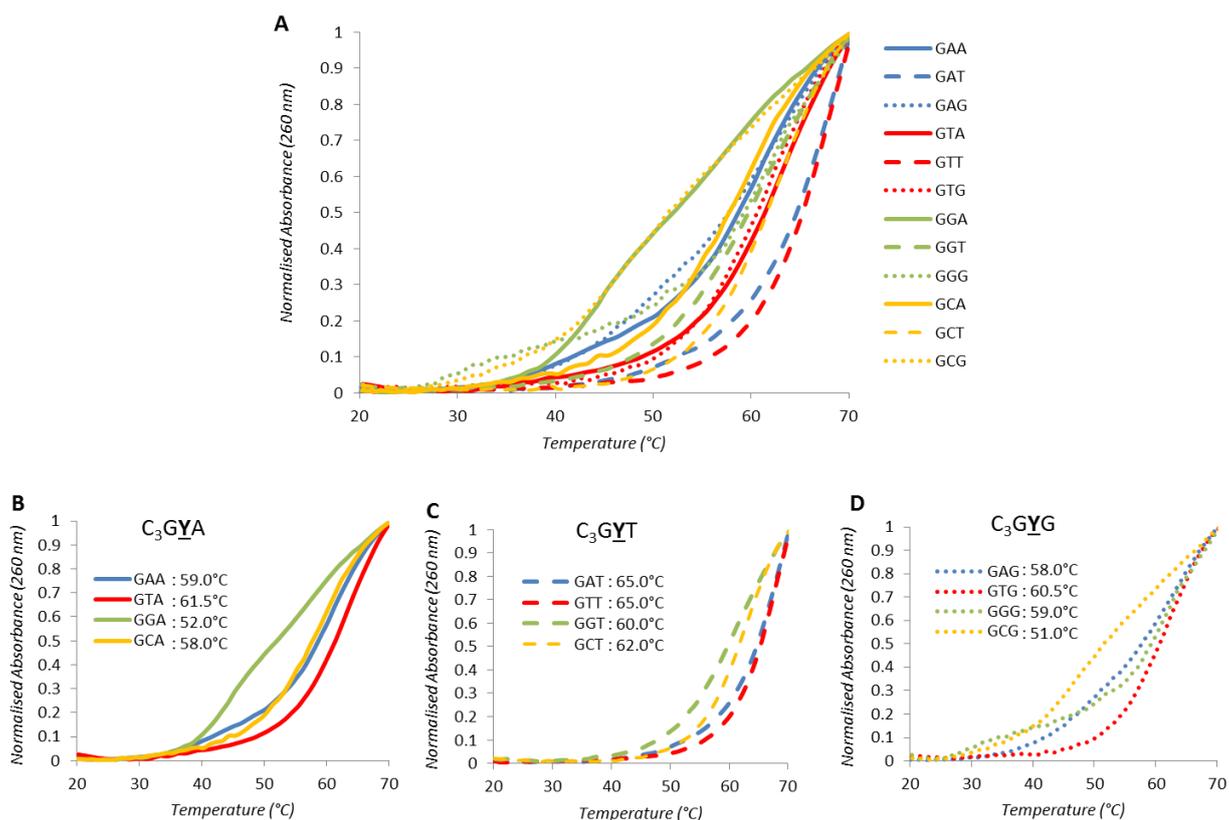
Figure 4.8: (**A**) DNA melting profiles (λ=260 nm) of all twelve C$_3$*G*YZ oligonucleotides (1 µM, ss) in 50 mM sodium cacodylate buffer (pH 5). Y is any of the four DNA bases and Z is adenine, thymine or guanine. The three smaller graphs show the effect on thermal stability by the middle base (Y) when Z is (**B**) adenine, (**C**) thymine and (**D**) guanine.

The T_m values obtained for the oligonucleotides showed concentration-independent profiles, suggesting that these potential i-motif forming sequences could be intramolecular.[128] pH titrations of all C$_3$*A*YZ and C$_3$*T*YZ sequences showed hyperchromicity between 275 and 300 nm and the shift in absorbance from around 260 nm for neutral cytosine to around 275 nm for protonated cytosine[67] (Appendix figures A4.7 to A4.9). Most of the C$_3$*G*YZ sequences, on the other hand, did not show these i-motif characteristic features. Broader peaks with flatter maxima were observed instead. Table 4.2 lists the T_m and T_{pH} values of all C$_3$XYZ sequences. Figure 4.9 shows the distribution of the C-rich sequences as a T_{pH} versus T_m graph. Seven out of the twelve C$_3$*G*YZ sequences lie on the most right-hand side of the graph, illustrating high structural stabilities.

Table 4.2: Sequences used in this study with associated melting temperatures (T_m/°C) and transitional pH (T_{pH}) values.

C$_3$AYZ				C$_3$TYZ				C$_3$GYZ			
Abbreviation	Sequence 5′ → 3′	T_m (°C)	T_{pH}	Abbreviation	Sequence 5′ → 3′	T_m (°C)	T_{pH}	Abbreviation	Sequence 5′ → 3′	T_m (°C)	T_{pH}
AAA	(CCCAAA)$_3$CCC	54.0	5.9	TAA	(CCCTAA)$_3$CCC	56.0	6.3	GAA	(CCCGAA)$_3$CCC	59.0	6.3
AAT	(CCCAAT)$_3$CCC	55.0	6.2	TAT	(CCCTAT)$_3$CCC	58.0	6.3	GAT	(CCCGAT)$_3$CCC	65.0	6.6
AAG	(CCCAAG)$_3$CCC	51.0	5.9	TAG	(CCCTAG)$_3$CCC	52.0	6.0	GAG	(CCCGAG)$_3$CCC	58.0	5.3
ATA	(CCCATA)$_3$CCC	54.0	6.2	TTA	(CCCTTA)$_3$CCC	57.5	6.2	GTA	(CCCGTA)$_3$CCC	61.5	6.6
ATT	(CCCATT)$_3$CCC	58.0	6.6	TTT	(CCCTTT)$_3$CCC	60.0	6.6	GTT	(CCCGTT)$_3$CCC	65.0	7.0
ATG	(CCCATG)$_3$CCC	54.0	6.1	TTG	(CCCTTG)$_3$CCC	54.0	6.2	GTG	(CCCGTG)$_3$CCC	60.5	6.3
AGA	(CCCAGA)$_3$CCC	47.5	5.8	TGA	(CCCTGA)$_3$CCC	52.0	6.3	GGA	(CCCGGA)$_3$CCC	52.0	5.6
AGT	(CCCAGT)$_3$CCC	52.5	5.9	TGT	(CCCTGT)$_3$CCC	56.0	6.5	GGT	(CCCGGT)$_3$CCC	60.0	6.2
AGG	(CCCAGG)$_3$CCC	59.0	5.2	TGG	(CCCTGG)$_3$CCC	51.0	5.7	GGG	(CCCGGG)$_3$CCC	59.0	5.0
ACA	(CCCACA)$_3$CCC	53.0	5.7	TCA	(CCCTCA)$_3$CCC	58.0	5.7	GCA	(CCCGCA)$_3$CCC	58.0	6.3
ACT	(CCCACT)$_3$CCC	57.5	5.9	TCT	(CCCTCT)$_3$CCC	61.0	6.6	GCT	(CCCGCT)$_3$CCC	62.0	6.6
ACG	(CCCACG)$_3$CCC	53.0	6.1	TCG	(CCCTCG)$_3$CCC	55.0	5.9	GCG	(CCCGCG)$_3$CCC	51.0	5.9
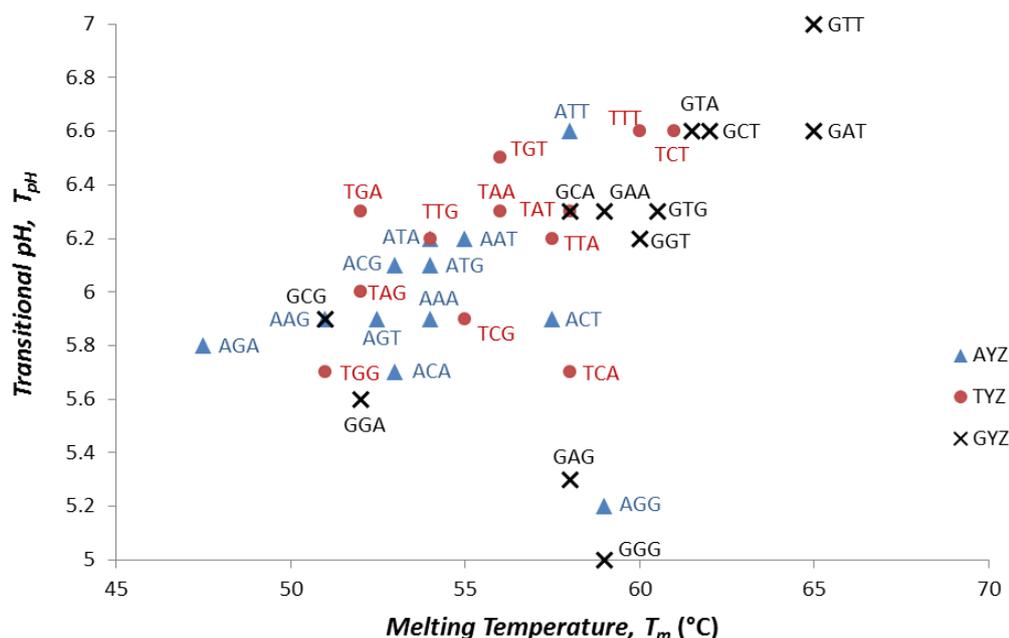
Figure 4.9: pH transition (T_{pH}) values against melting temperatures (T_m) of C_3XYZ oligonucleotides, where X and Z are adenine, guanine and thymine and Y is any of the four DNA bases.

4.3.2 i-motif Structure Determination using Circular Dichroism Spectroscopy

CD spectroscopy measurements were carried out to further investigate the effects of pH on the C_3XYZ sequences, and to consequently confirm that the oligonucleotides formed i-motif structures (Fig. 4.10 and Appendix fig. A4.11 for measurements made in pH 8). The concentration of the buffer used for srCD was lowered to 20 mM to obtain the lowest cut off measurement in the backbone region of 180-260 nm wavelengths. The final concentration of each DNA used was 200 μM (ss) to read reasonable absorbance using 0.01 cm pathlength, i.e. not below or over the limit of detection in the far and near UV regions.

C_3**A**YZ and C_3**T**YZ in pH 5 showed the i-motif characteristic maximum and minimum peaks at 285-290 and 255-260 nm respectively.[128] Ten out of the 12 oligonucleotides with guanine as the first base in the loops, however, gave a negative peak at 245 nm instead. Negative peaks at 245-250 nm wavelengths are associated to duplex DNA, hairpin and G-quadruplex structures.[127,179] Duplex DNA displays a positive peak at 260 nm wavelength but the C_3**G**YZ series either have a positive peak at 285-290 nm or a broad positive band from 260-283 nm. The broad band at about 260–280 nm together with the negative peak at around 245 nm are specific to the hairpin structure.[127] These peaks are observed for GAA, GAG and GGG. The rest of the sequences, excluding GGT and GCG, have a positive peak at 285-290 nm. The combination of a positive and negative band at 290 and 240 nm wavelengths, respectively, are observed in the CD spectra of anti-parallel G-quadruplex structures.[127,180] Only GGT and GCG display i-motif formation. Based on the oligonucleotides that

showed all the expected features of an intramolecular i-motif using both UV and CD spectroscopies, only 26 of the C-rich sequences were able to form i-motifs.
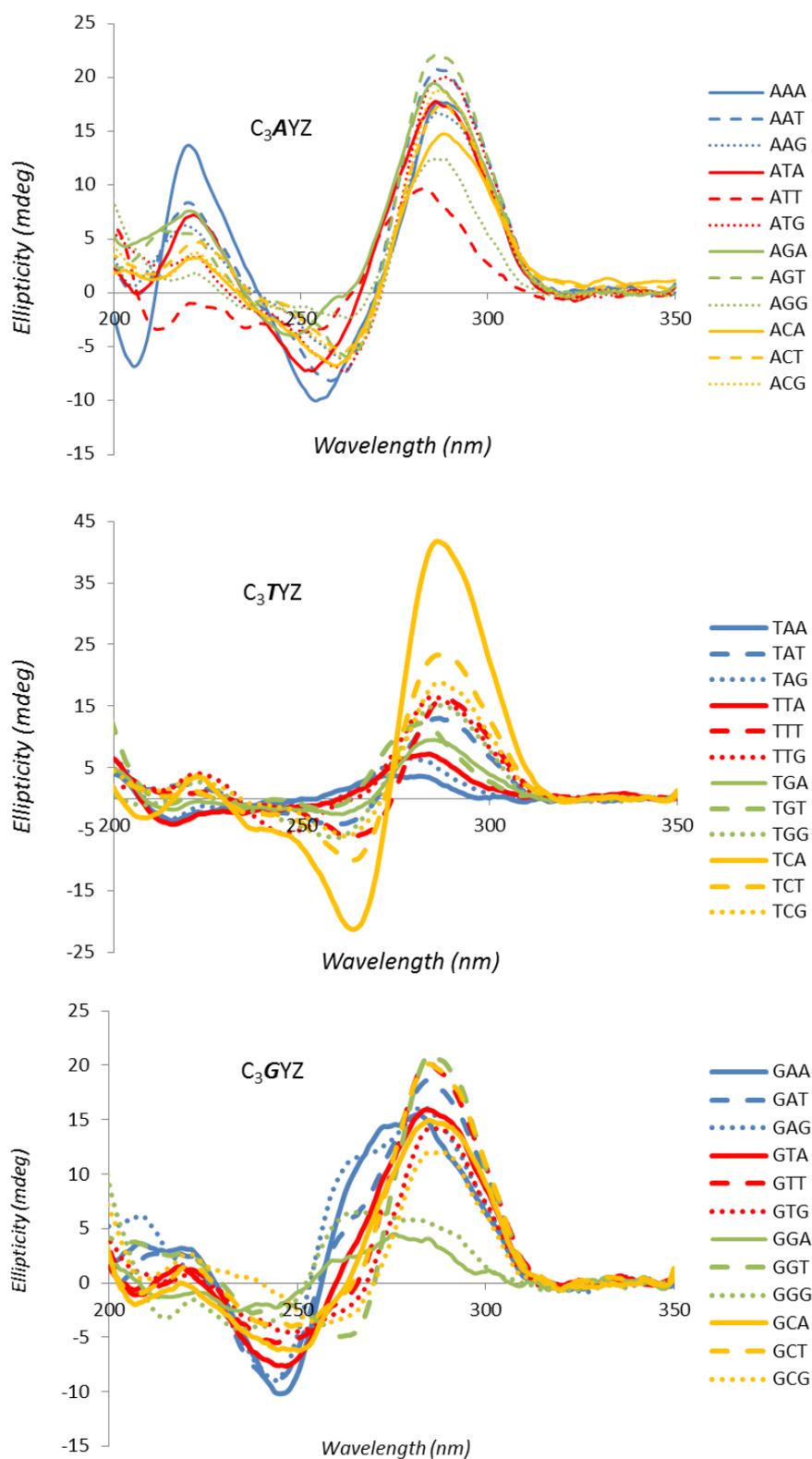
Figure 4.10: CD spectra of (top) C$_3$*A*YZ, (middle) C$_3$*T*YZ and (bottom) C$_3$*G*YZ oligonucleotides in 20 mM sodium cacodylate buffer (pH 5), 20°C. Z is adenine, thymine or guanine and Y is any of the four DNA bases.

Tables 4.3 and 4.4 list the T_m and T_{pH} values of all the sequences used in this study. In terms of thermal stability only, AGA is the least stable and TCT is the most stable i-motif forming sequence. Overall, GAT and GTT are the most stable DNA sequences. Regarding to transitional pH values, AGG is the least stable. ATT, TTT & TCT share the same T_{pH} for the most stable i-motif forming sequences whereas GTT is the most stable sequence overall. In terms of both thermal and pH stabilities, the sequences with guanine as the first base and thymine as the last base are most stable. When only i-motif forming C_3XYZ sequences are taken into account, those containing the most number of pyrimidine residues are the most stable. Hence, the most stable i-motif forming sequence is TCT.

Table 4.3: DNA melting temperatures (°C) of C_3XYZ. X, Y and Z bases are labelled blue, orange and black, respectively. The sequences that didn't show i-motif formation by circular dichroism are highlighted in pink. The lowest and highest i-motif T_m values are highlighted in yellow and green. The most thermally stable sequence is highlighted blue.

X base		Y base A			T			G			C	
	A	T	G	A	T	G	A	T	G	A	T	G
A	54.0	55.0	51.0	54.0	58.0	54.0	47.5	52.5	59.0	53.0	57.5	53.0
T	56.0	58.0	52.0	57.5	60.0	54.0	52.0	56.0	51.0	58.0	61.0	55.0
G	59.0	65.0	58.0	61.5	65.0	60.5	52.0	60.0	59.0	58.0	62.0	51.0

Z base

Table 4.4: Transitional pH values of C_3XYZ. X, Y and Z bases are labelled blue, orange and black, respectively. The sequences that didn't show i-motif formation by circular dichroism are highlighted in purple. The lowest and highest i-motif T_{pH} values are highlighted yellow and green. The sequence with the highest T_{pH} is highlighted blue.

X base		Y base A			T			G			C	
	A	T	G	A	T	G	A	T	G	A	T	G
A	5.9	6.2	5.9	6.2	6.6	6.1	5.8	5.9	5.2	5.7	5.9	6.1
T	6.3	6.3	6.0	6.2	6.6	6.2	6.3	6.5	5.7	5.7	6.6	5.9
G	6.3	6.6	5.3	6.6	7.0	6.3	5.6	6.2	5.0	6.3	6.6	5.9

Z base

_____

4.3.3 The Effect of Loop Base Directionality on i-motif Stability

Change in direction of loop base alignment gave differences in both thermal and pH stabilities of the C$_3$XYZ sequences. Moreover, seven of the i-motif forming sequences do not form i-motifs when their sequence directionality is changed. Table 4.5 lists the differences in T_m and T_{pH} values for sequences with the same base composition but opposite directionality. Sequence direction is identical when X and Z bases are the same so 12 sequences have been excluded from comparison.

Table 4.5: The differences in thermal (ΔT_m) and pH (ΔT_{pH}) stabilities when the direction of base alignment is changed. The differences are calculated by subtracting the T_m and T_{pH} values of oligonucleotides in column 2 from column 1. GGT is the only i-motif forming C$_3$**G**YZ sequence listed in the table.

Sequence 5' → 3'	Sequence 5' → 3'	ΔT_m (°C)	ΔT_{pH}
AAT	TAA	+1.0	+0.1
AAG	GAA	+8.0	+0.4
ATT	TTA	-0.5	-0.4
ATG	GTA	+7.5	+0.5
AGT	TGA	+0.5	+0.4
AGG	GGA	-7.0	+0.4
ACT	TCA	+0.5	-0.2
ACG	GCA	+5	+0.2
TAG	GAT	+13	+0.6
TTG	GTT	+11	+0.8
TGG	GGT	+9	+0.5
TCG	GCT	+7	+0.7

The lowest difference in stability were recorded for sequences where X is adenine and Z is thymine, with up to 1°C and 0.4 unit differences in T_m and T_{pH}, respectively. The largest ΔT_m and ΔT_{pH} values observed between i-motif forming sequences of the same base alignment were for TGG and GGT (9°C, 0.5 pH units). The largest differences observed overall are for the i-motif forming TAG sequence and the non-i-motif forming GAT with ΔT_m of 13°C and ΔT_{pH} of 0.6 units. This confirms that direction of sequence alignment also plays a crucial role in i-motif formation and stability.

4.4 Conclusions

Results from DNA melting and pH titration experiments of the C_3XYZ oligonucleotides show that loops that have X as guanine have the highest thermal and pH stabilities. However, only two out of the 12 C_3GYZ sequences; GGT and GCG, formed i-motif structures as revealed by CD spectroscopy. Instead, GAA, GAG and GGG gave bands that are characteristic of DNA hairpins. Hairpin structures are thermodynamically very stable, with T_m values of at least 60°C,[181] which are around the range of T_m values observed for sequences where X is guanine. The highest T_m recorded was for GAT and GTT with 65°C. Both sequences are stable to neutral or near-neutral pH. The remaining sequences had features that resembled the CD spectra of anti-parallel G-quadruplex structures. CD spectroscopy measurements need to be recorded in the presence of K^+ and Na^+ at pH 7 to confirm G-quadruplex formation. Sugimoto *et al.* reported that i-motif forming sequences rich in guanine gave significantly higher thermodynamic stabilites.[40] If the structure is not of a G-quadruplex, a hypothesis can be raised of a possible i-motif and G-quadruplex hybrid structure (Fig. 4.11). SrCD spectroscopy also revealed positive bands between 180-200 nm (Appendix fig. A4.7) that appear to relate to guanine residues, however, less information is available on the region, with none on i-motifs until now. Only 26 out of the 36 C-rich sequences were able to form i-motifs, implying that loop base composition has a vital influence on i-motif formation.
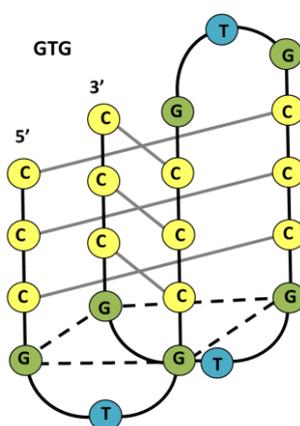
Figure 4.11: Schematic diagram of the proposed GTG intramolecular G-quadruplex-i-motif structure. Cytosine, guanine and thymine are represented by yellow, green and blue dots, respectively. The dashed lines outline possible G-tetrad formation underneath the cytosine core.

I-motif forming C_3XYZ sequences with adenine as the first base have relatively lower stabilities in comparison to the rest of the sequences. These results agree with that of Sheardy and co-workers.[182] Sequences that are rich in adenine also gave low T_m and T_{pH} values. Presence of adenine bases in loops have been shown to affect the stability of G-quadruplex structures, whose T_m values decrease when adenine is the first base in the loop regions.[183,184] The base adjacent to the cytosine tetraplex core is crucial in determining the stability of an i-motif structure,[40] and NMR studies have shown that protonation of adenine at loops disrupts the i-motif core.[75] This could explain why loops starting with adenine are less stable.

_____

T-rich sequences were the most stable i-motif forming sequences, with thymine as all X, Y and Z bases in the loop displaying higher T_m and T_{pH} values. The second base only resulted in T_m value differences greater than 3°C when Y was guanine; both stabilising and destabilising i-motif structures. TGG has a low T_m value of 51°C in comparison to AGG and GGG which have higher T_m values of 59°C. I-motifs that were rich in purines, with the exception of GGT, proved to be less stable than those rich in pyrimidines. Benabou *et. al.* also reported that addition of purines into the first and third loops destabilises the i-motif structure.[185] The i-motif structures with the lowest and highest stabilities recorded from the series were AGA and TCT, respectively. Insertion of guanine and thymine into an oligomeric C-rich sequence was reported to make the formation of i-motif unfavourable[186] but the study was made on C-rich sequences with guanine residues among the middle positions in the loop regions. The GGT i-motif structure emphasises the notion that loop base position is an important factor in i-motif formation.

This brings out the point on the importance in the direction of loop base alignment. 5'-GGT-3' has the melting temperature of 60°C whereas the T_m for 5'-TGG-3' is 51°C. Changes in melting temperatures and transitional pH values brought by opposite sequence directionality can be observed in 11 sequences altogether. GGT, for example, may be more stable than TGG due to the accessibility of the loop bases to form base pairs within the i-motif structures. The positions of the thymine residues in the first and third loops of GGT allow room for T·T base pairing directly underneath the cytosine core, therefore possibly contributing to higher stability (Fig. 4.12). Conversely, the thymine bases in TGG are quite apart for T·T base pairs to form. This strengthens the notion that although stabilities result from contributions of both stacking and hydrogen bonding interactions,[187] hydrogen bonding among bases have a much stronger effect on the stability of the i-motif structures than the stacking of the nucleobases.[40]
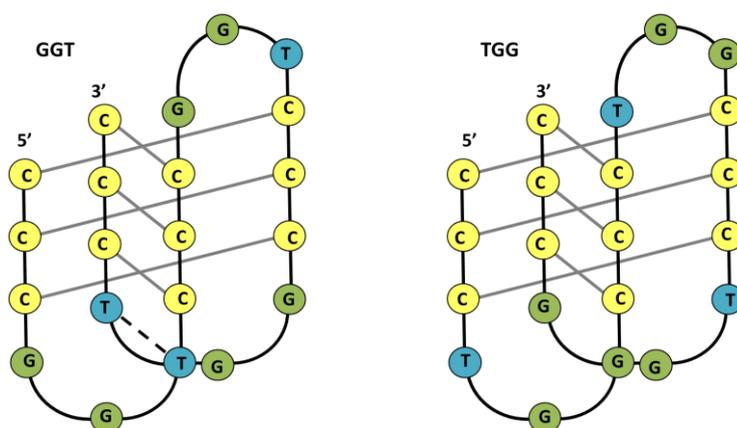
Figure 4.12: Schematic diagrams of the GGT and TGG intramolecular i-motif structures. Cytosine, guanine and thymine are represented by yellow, green and blue dots, respectively. The dashed lines outline possible T·T base pairing underneath the cytosine core.

_____

In conclusion, although the study was initially carried out to understand the effects of loop composition on i-motif stability, the experiments conducted showed that not all C-rich sequences are composed of the general i-motif sequence format form the intercalated structure. The results demonstrate that directionality of bases in loops also need to be considered for i-motif formation. I-motif forming sequences containing pyrimidine loops are more stable than those with purines. Hairpin and G-quadruplex focussed studies need to be designed in order to confirm that the C_3*G*YZ sequences do form these structures. These findings can then provide insight into designing i-motif to hairpin/G-quadruplex switches. The results can also be used for predicting stable i-motif forming sequences for use in various other studies, including crystallographic trials for intramolecular i-motif crystal structure determination.

_____

The Effect of Non-Cytosine Bases on i-motif Structure

5.1 Introduction

In 1962, Marsh *et. al.* solved the crystal structure of cytosine-5-acetic acid which showed that cytosine (C) could form a hydrogen-bonded pair with itself if the base was hemiprotonated.[34] 31 years later, a four-stranded NMR structure of d(TCCCCC)$_4$ at acidic pH was reported,[36] which revealed two parallel duplexes intercalated to each other and held by C$^+$·C base pairs. In 1994, the first crystal structure of such an intercalated DNA motif; the d(CCCC)$_4$ i-motif, was published at 2.3 Å resolution.[132] There are six tetramolecular i-motif crystal structures reported to date and all were solved between the mid- to late 1990's using X-ray crystallography. The structures of d(CCCC)$_4$,[132] d(CCCT)$_4$,[134] d(TAACCC)$_4$,[135] d(CCCAAT)$_4$,[133] d(AACCCC)$_4$[58] and d(ACCCT)$_4$[79] show intercalated cytosine segments that all crystallised with similar geometry. In this chapter, these structures are abbreviated to C$_4$, C$_3$T, TA$_2$C$_3$, C$_3$A$_2$T, A$_2$C$_4$ and AC$_3$T, respectively. Figure 5.1 illustrates that the i-motif structure has a flat and ribbon-shaped construction with very wide grooves along two sides and very narrow grooves at the ends. The DNA structures have a "zigzag" pathway of sugar-phosphate backbones. The glycosidic torsion angles (χ) are in the *anti*-range, between -120° and -180° and the sugar puckers are mixed with varying values for the phase angle of pseudorotation, *P*.

Although the geometry of the cytosine core is similar for all six i-motif structures, the non-cytosine bases in either end of the core, or the loop regions, have different folding topologies (Fig. 5.2). The 5'-TA$_2$C$_3$-3' and 5'-C$_3$A$_2$T-3' DNA sequences are different only by directionality, yet the adenine bases in the A2 and A3 positions in first sequence and thymine in the T6 position in the latter sequence are flipped away from the plane of the core. The 5'-TA$_2$ segments exhibit a novel tight loop in which the 5'- and 3'- ends of adjacent strands are in close proximity. This loop is stabilised by a Hoogsteen T1·A3 base pair which sits above the terminal C$^+$·C base pair. On the other hand, C$_3$A$_2$T consists of two A·A mismatch pairs; an asymmetric A4·A4 base pair stacking over the terminal C$^+$·C base pair, and an A5·A5 mismatch. Each mismatch is involved in an A·A·T base triple with a thymine from a symmetry-related i-motif. Similarly, the A$_2$C$_4$ structure has both adenine bases facing away from the cytosine core whereas AC$_3$T and C$_3$T have two thymine residues per i-motif molecule orientated away from the plane of the core.

Table 5.1 lists the crystallisation parameters of the six structures along with their refinement statistics from data collected using an in-house X-ray source. C$_4$ and AC$_3$T are cubic crystal systems, C$_3$T, A$_2$C$_4$ and TA$_2$C$_3$ are orthorhombic and C$_3$A$_2$T is hexagonal. Table 5.2 lists the structural parameters of the cytosine core, showing that there are slight differences in the intercalated site of the six i-motif crystals. These parameters are compared to the new i-motif crystal structures discussed in section 5.3.
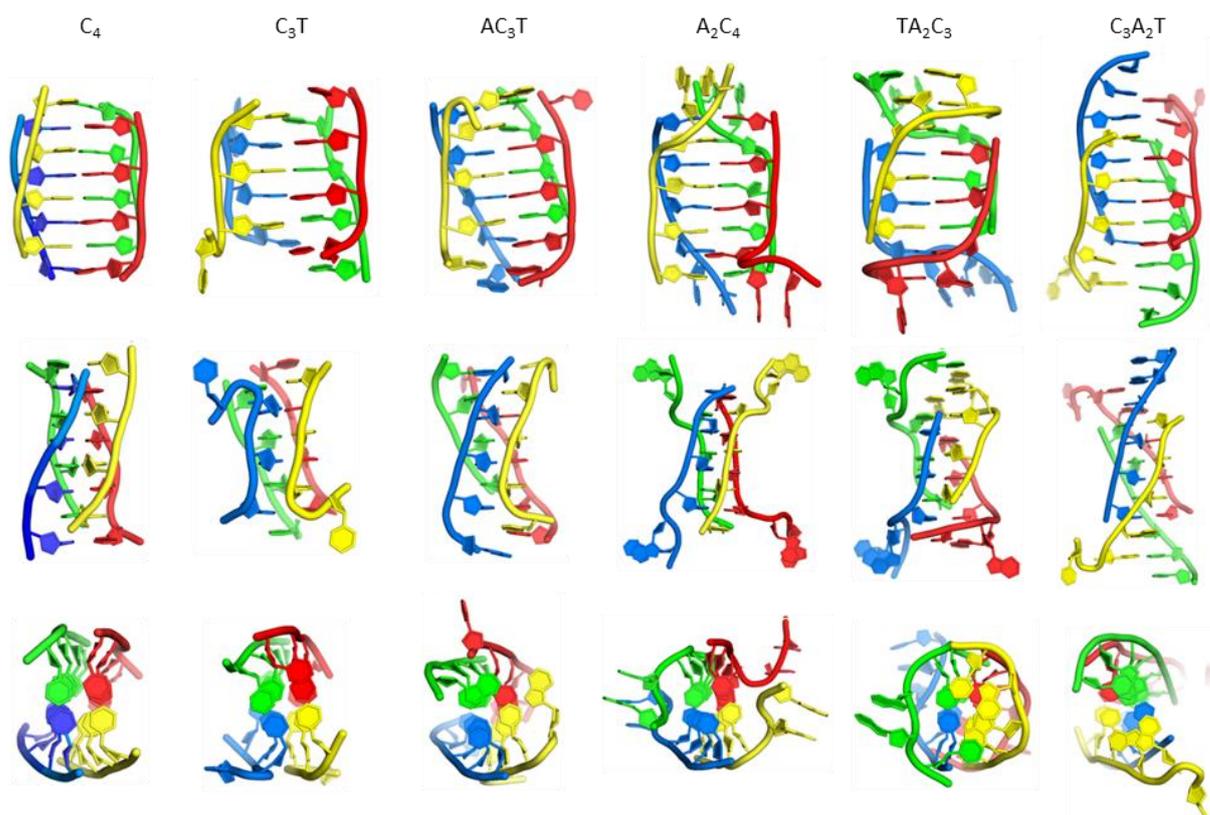
Figure 5.1: Crystal structures of six tetramolecular DNA i-motifs of the sequence; C_4,[132] C_3T,[134] TA_2C_3,[135] C_3A_2T,[133] A_2C_4[58] and AC_3T[79]. The first and seconds rows show the views of the molecules from the wide and narrow grooves (when rotated 90° to the right), respectively. The third row shows the top views of the molecules.
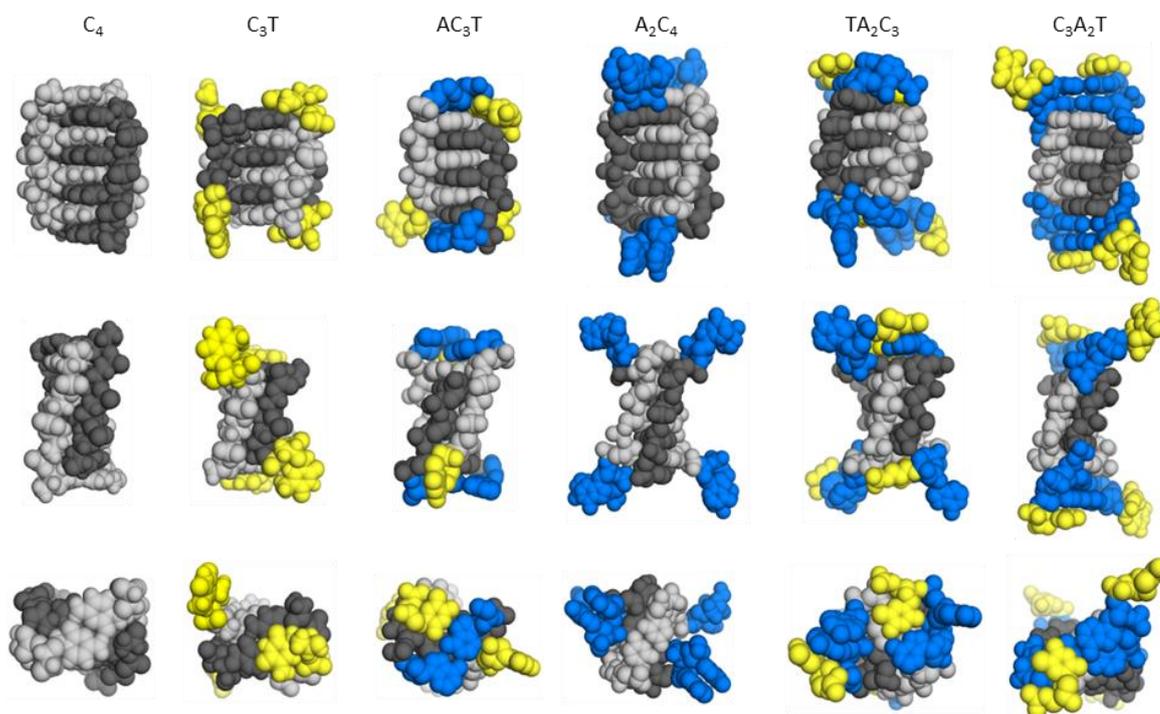
Figure 5.2: Van der Waals representation of the six crystal tetramolecular DNA i-motifs. The base pairing cytosine residues are coloured in light and dark grey, adenines in yellow and thymine residues in blue.

_____

Table 5.1: Crystallographic data of the six tetramolecular DNA i-motifs. Under the method section, VD stands for Vapour Diffusion with HD = Hanging Drop and SD = Sitting Drop techniques. ASU (refinement) stands for asymmetric unit. Data that were unavailable have been labelled N/A for non-applicable.

Crystallisation Parameters

DNA Sequence	C_4	C_3T	AC_3T	A_2C_4	TA_2C_3	C_3A_2T
PDB ID	190D	191D	1CN0	294D	200D	241D
Method	VDHD	VDHD	VDHD	VDSD	VD	VDHD
pH	5.5	6.5	7.0	7.5	6.0	6.0
Crystal size (mm)	0.3 x 0.3 x 0.3	0.5 x 0.5 x 1.0	0.3 x 0.3 x 0.1	0.3 x 0.2 x 0.1	0.4 x 0.7 x 0.3	0.6 x 0.6 x 0.3
Growth period	1 month	3 weeks	1 month	N/A	24 hours	several days

Data Collection

Temperature (K)	293	277	103	277	277	277

Crystal Data

Space Group	$I\,2\,3$	$C\,2\,2\,2_1$	$I\,2\,3$	$P\,2\,2_1\,2_1$	$F\,2\,2\,2$	$P\,6_2$
Unit-cell (Å)	a = 82.3 b = 82.3 c = 82.3	a = 28.3 b = 44.3 c = 50.5	a = 93.8 b = 93.8 c = 93.8	a = 35.9 b = 52.3 c = 76.9	a = 59.9 b = 81.3 c = 26.9	a = 32.2 b = 32.2 c = 52.5
	$\alpha = 90$ $\beta = 90$ $\gamma = 90$	$\alpha = 90$ $\beta = 90$ $\gamma = 90$	$\alpha = 90$ $\beta = 90$ $\gamma = 90$	$\alpha = 90$ $\beta = 90$ $\gamma = 90$	$\alpha = 90$ $\beta = 90$ $\gamma = 90$	$\alpha = 90$ $\beta = 90$ $\gamma = 120$

Refinement

DNA strands/ASU	8	4	4	8	2	2
Resolution (Å)	8.0 - 1.8	10.0 - 1.4	15.0 - 2.2	2.5 - N/A	12.0 - 1.85	10.0 - 1.85
Cut-off Sigma (F)	2.0	2.0	2.0	2.0	1.0	2.0
No. Reflections (Observed)	8177	5013	6914	4539	2122	2285
Completeness (%)	96.0	96.0	96.3	86.5	N/A	N/A
Software	X-PLOR	X-PLOR	X-PLOR	X-PLOR	X-PLOR	X-PLOR
R_{work}/R_{free}	0.218/NA	0.177/0.225	0.192/0.235	0.212/0.295	0.190/0.223	0.194/NA
No. Atoms						
DNA	584	296	360	920	232	232
Ligands	0	3 (Na^+)	0	0	0	0
Water	46	63	62	60	26	57
RMS Deviations						
Bond lengths (Å)	0.015	0.029	0.004	0.016	0.030	0.028
Bond angles (°)	3.700	4.000	0.574	3.740	4.000	3.700

_____

Table 5.2: The geometrical parameters of the cytosine cores in tetramolecular DNA i-motif structures.

DNA Sequence	C_4	C_3T	AC_3T	A_2C_4	TA_2C_3	C_3A_2T
Helix sense	RH	RH	RH	RH	RH	RH
Helical twist (°)	12.4	17.1	14.5	16.6	19.7	18.3
Average base stacking distance (Å)	3.10	3.10	3.20	3.20	3.13	3.15
Average P-P distance (Å)						
Narrow groove	7.10	7.02	8.10	6.33	6.70	7.09
Wide groove	15.90	16.63	15.0	16.09	16.90	17.00
Average $C^+{\cdot}C$ distance (Å)						
N-H⋯O bond	2.75	2.77	2.77	2.68	2.84	2.74
N-H⋯N bond	2.76	2.74	2.70	2.70	2.79	2.76
N-H⋯O water bond	3.00	3.00	2.94	3.10	2.87	3.08

As there was no systematic study of DNA sequences forming tetramolecular i-motifs, the folding topologies of non-cytosine bases in the loop regions could not be predicted. Consequently, there was little insight into understanding potential folding topologies in bimolecular and intramolecular i-motifs. The base adjacent to the cytosine tetraplex core is crucial in determining the stability of an i-motif structure.[40] Hence, a simple system of short tetramers; d(C_3X)$_4$, where C is cytosine and X is any DNA base including cytosine, were tested for crystal growth.

Crystallisation trials included those of the individual oligonucleotides on their own and with a mixture of the oligonucleotides. A total of seven systems were set up for crystallisation (Table 5.3). Five systems produced diffracting crystals, however, trials for [C_4 + C_3T + C_3G + C_3A] produced crystals of [C_3T + C_3A], giving four systems (two new) to study. No crystal growth was observed for C_3G and [C_4 + C_3G]. In chapter 4, 83% of C-rich sequences that had guanine as the first base in the loop failed to form i-motif structures and gave CD spectra resembling hairpin and G-quadruplex structures instead. Therefore, it is possible that the C_3G sequence does not form an i-motif.

Table 5.3: List of DNA sequences set for crystal growth.*X-ray diffraction showed that the crystals observed for the mixture of all four DNA oligonucleotides of d(C_3X)$_4$ (where X = any DNA base) were that of [C_3T + C_3A].

System	DNA Sequence 5' → 3'	Crystal growth	Diffraction observed	Structure solved
C_4	CCCC	Yes	Yes	Yes
C_3T	CCCT	Yes	Yes	Yes
C_3G	CCCG	No	N/A	N/A
C_3A	CCCA	Yes	Yes	Yes
C_4 + C_3G	CCCC, CCCG	No	N/A	N/A
C_3T + C_3A	CCCT, CCCA	Yes	Yes	Yes
C_4 + C_3T + C_3G + C_3A	CCCC, CCCT, CCCG, CCCA	Yes	Yes	No*

_____

In this chapter, the structures of C_4 and C_3T are compared to those previously published. The two novel structures of C_3A and $[C_3T + C_3A]$ will be discussed. In contrast to the methodology of the previously reported structures, data collection of all crystal systems from this project were carried out in a synchrotron and diffraction data were collected at sub-atomic resolutions.

5.2 Materials and Methods

5.2.1 Crystallisation of the i-motifs

Oligonuclotides were purchased from Eurogentec and purified by reverse phase HPLC. Crystals were grown using sitting drop vapour diffusion. Initial screening was performed using 192 crystallisation conditions from the Natrix HT Screen by Hampton Research and HELIX by Molecular Dimensions at 18°C and 4°C. Each drop contained 2μl of 2mM single-stranded oligonucleotide and 2μl of the crystallisation solution. The drop was equilibrated against 500μl of the same crystallisation condition. Crystals of C_4, C_3T and C_3A were obtained at 18°C (Appendix fig. A5.1). Table 5.4 lists the final crystallisation conditions.

Table 5.4: Crystallisation conditions of C_3X, where X = A, T or C.

DNA Sequence	Salt 1	Salt 2	Buffer	Buffer pH	Precipitant	Additive
C_4	Ammonium chloride (0.2 M)	Magnesium chloride hexahydrate (0.01 M)	HEPES sodium (0.05 M)	7.0	1,6-Hexanediol (2.5 M)	N/A
C_3T	Lithium chloride (0.04 M)	Magnesium chloride hexahydrate (0.02 M)	Sodium cacodylate trihydrate (0.04 M)	5.5	2-methyl-2, 4-pentanediol (30% v/v)	Hexammine cobalt(III) chloride (0.02 M)
C_3A	Potassium chloride (0.2 M)	Magnesium chloride hexahydrate (0.01 M)	HEPES Sodium (0.05 M)	7.0	1,6-Hexanediol (1.7 M)	N/A
$C_3A + C_3T$	Ammonium chloride (0.2 M)	Calcium chloride dihydrate (0.01 M)	TRIS hydrochloride (0.05 M)	8.5	Polyethylene glycol 4,000 (30% w/v)	N/A

_____

The C_4 and C_3T crystal structures previously solved will be compared to those grown in new conditions. The C_4 crystal reported by Chen *et. al.* was grown in 0.1 M sodium cacodylate buffer (pH 5.5) equilibrated against 20% MPD which was then soaked in 1 mM K_2PtCl_6 for 25 days to obtain a heavy atom difference Patterson map. The new C_4 crystals were grown without any heavy atoms. The crystallisation conditions for the C_3T structure solved by Kang and co-workers included 0.02 M $MgCl_2$, 0.04 M strontium cacodylate buffer at pH 6.5, 0.12 mM spermine and 10% MPD. The new C_3T crystals were grown in a more acidic pH of 5.5. Overall, two out of the four crystals in this project were grown near physiological pH. The presence of precipitant or/and additive with salt may contribute to molecular crowding and hence stabilise the DNA structure,[39,188] giving more scope for the biological role of the i-motif to be discussed.

4.2.2 Synchrotron Data Collection and Data Analysis

Data were collected on beamline I02 at Diamond Light Source. Data collection, processing and refinement statistics are given in Table 5.5. In all cases, the data were processed using XDS[189] and XSCALE with xia2.[190] The structures were solved using molecular replacement with Phaser[191] through the CCP4i[108] and PHENIX[109] interfaces. All four structures were built using Coot[192] and refined using *phenix.refine.*[193]

_____

Table 5.5: Crystallographic Data Collection and Refinement Statistics

Sequence	C_4	C_3T	C_3A	$C_3A + C_3T$
Crystallisation Parameters				
Crystal size (µm)	40 x 40 x 40	250 x 130 x 80	250 x 200 x 100	100 x 75 x 50
Growth period	2 weeks	24 hours	2 weeks	2 weeks
Data Collection				
X-ray wavelength (Å)	0.9919	0.5510	0.6888	0.6888
Exposure time (s)	0.04	0.10	0.04	0.04
No. of images	3600	1800	3600	3600
Space group	$I\ 2\ 3$	$P\ 1\ 2_1\ 1$	$C\ 1\ 2\ 1$	$P\ 1$
Unit-cell (Å)	a = 81.57	a = 26.05	a = 51.08	a = 24.19
	b = 81.57	b = 49.48	b = 34.42	b = 24.20
	c = 81.57	c = 26.03	c = 30.80	c = 24.93
	α = 90.00	α = 90.00	α = 90.00	α = 90.01
	β = 90.00	β = 114.80	β = 146.04	β = 90.02
	γ = 90.00	γ = 90.00	γ = 90.00	γ = 90.05
Data Processing *Outer shell statistics shown in parentheses				
Resolution (Å)	40.78-1.28	21.94-0.68	25.54-0.66	24.20-0.84
	(1.30-1.28)*	(1.85-0.68)	(1.80-0.66)	(2.28-0.84)
R_{merge} (I)	0.061 (2.955)	0.072 (0.795)	0.042 (0.350)	0.059 (1.428)
R_{meas} (I)	0.062 (2.995)	0.085 (0.944)	0.047 (0.451)	0.064 (1.540)
R_{pim} (I)	0.010 (0.484)	0.045 (0.504)	0.021 (0.281)	0.024 (0.574)
Total number of observations	913962	341361	268813	348627
$I/\sigma I$	32.7 (1.5)	8.0 (1.3)	20.1 (1.6)	14.7 (1.2)
$CC_{1/2}$	1 (0.567)	0.995 (0.522)	0.981 (0.876)	1 (0.473)
Completeness (%)	99.9 (99.9)	99.1 (97.0)	82.2 (8.0)	95.0 (92.2)
Multiplicity	39.1 (38.2)	3.4 (3.3)	5.9 (2.0)	7.1 (7.1)
Refinement				
No. Reflections	22691	96484	43203	46581
R_{work}/R_{free} (%)	12.48/15.08	15.46/17.45	19.85/20.77	21.17/21.69
No. Atoms				
DNA	984	968	550	596
Ligands	0	5	1	0
Water	164	208	91	162
Average B-factors (Å^2)				
DNA	25.581	6.980	10.137	9.064
Ligands	-	25.956	5.527	-
Water	40.399	30.293	19.897	21.934
Rmsd				
Bond lengths (Å)	0.0250	0.0310	0.0170	0.009
Bond angles (°)	2.4160	2.3207	1.6420	1.087

5.3 Results & Discussion

5.3.1 C₄

The unit cell parameters of the previously published C_4 (PDB ID: 190D) were reported as $a=b=c$=82.3Å. Data collected from crystals grown in new conditions gave similar unit cell parameters $a=b=c$=81.57Å and $\alpha=\beta=\gamma$=90.00° in I23 space group. The structure was solved by molecular replacement (MR) at final R_{work}/R_{free} values of 0.1248/0.1508 with 5.2% of reflections reserved for the R_{free} set. Two C_4 i-motifs are modelled in the asymmetric unit, i.e. 32 nucleotides and 164 ordered water molecules. Figure 5.3 illustrates that parallel stranded duplexes base pair with one another, resulting in two very wide grooves between each i-motif molecule. The anti-parallel strands are adjacent to each other and face the narrow grooves.
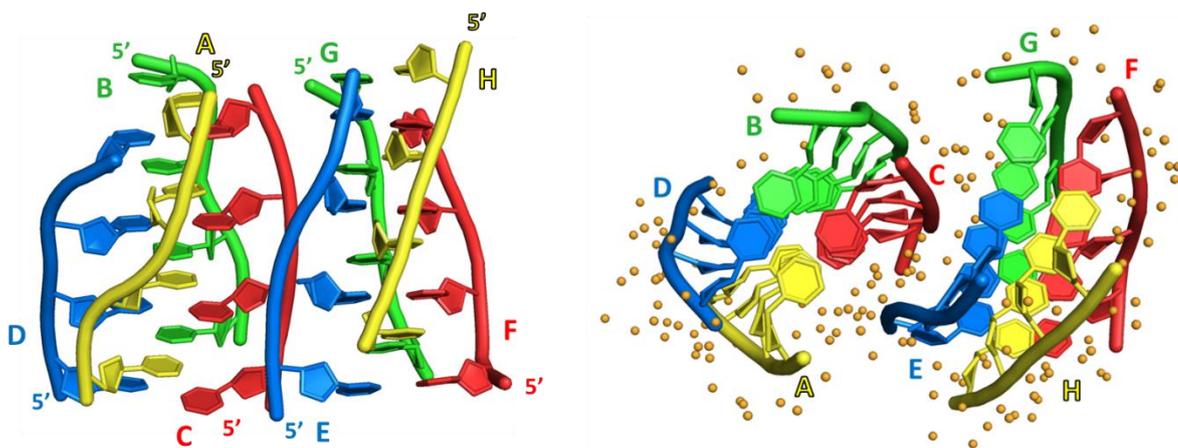
Figure 5.3: Views of the asymmetric unit of C_4 looking (left) along and (right) down the helical axes. The eight strands have been labelled A to H with chain A base pairing with B, C-D, E-F and G-H. Water molecules are orange spheres.

The average helical twist in both i-motif molecules is 13.5°, in comparison with 12.4° in 190D. The twist allows the strongest hydrogen bonding and base stacking with the lowest amount of steric or electrostatic repulsion. As the new C_4 has a higher angle of twist, this structure may be more stable than 190D. The Van der Waals representations in Fig. 5.4 show the close packing between the two sugar phosphate chains at the narrow ends leads to variability in the phosphorus-phosphorus (P-P) distances. The average intrachain P-P distance along the chains is 6.5Å and the average interchain P-P distance (between two adjacent chains) is 8.4Å. The B-DNA major groove is 11.7Å wide and the minor groove is smaller at 5.7Å.[7] In the case of C_4, the width of the major groove is 16.9Å and the minor groove is 10.8Å.
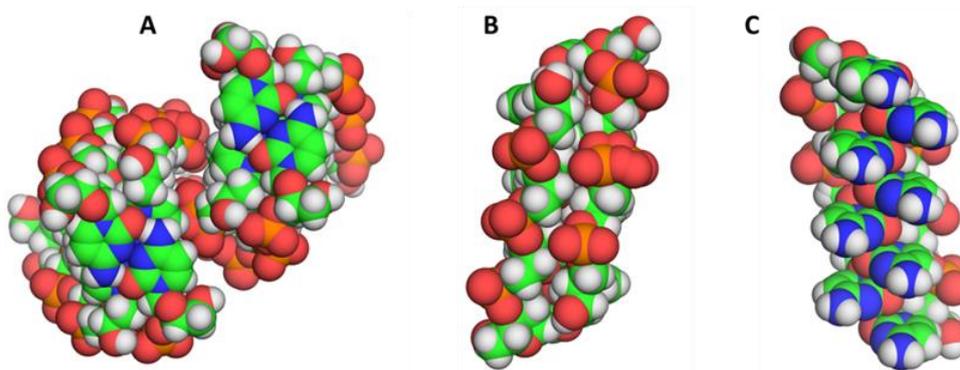
Figure 5.4: Van der Waals model of (**A**) the asymmetric unit of C$_4$ where oxygen is red, carbon is green, nitrogen is blue, hydrogen is white and phosphorus is orange. (**B**) Two chains fit in an antiparallel orientation adjacent to each other in the narrow groove. (**C**) Base stacking of the cytosine residues as viewed through the narrow groove.

The i-motif structure is held together by intercalated cytosine tracts. The C$_4$ structure shows that only the exocyclic atoms O2 and N4 are involved in stacking of the bases. The π-electron systems over the rings themselves are not directly involved in stacking. The average base-stacking distance between two cytosine residues in B-DNA is 3.4Å whereas in C$_4$, the cytosine bases are 3.1Å apart. There are three hydrogen bonds that hold the C$^+$·C base pair together (Fig. 5.5). The N3-N3 hydrogen bonds in the centre of the cytosine pairs have an average bond length of 2.77Å. The O2-N4 and N4-O2 have bond lengths of 2.81Å and 2.82Å respectively. Water molecules are hydrogen bonded to the amino group N4 and are also found near phosphate groups, mainly bridging between the two i-motif molecules.
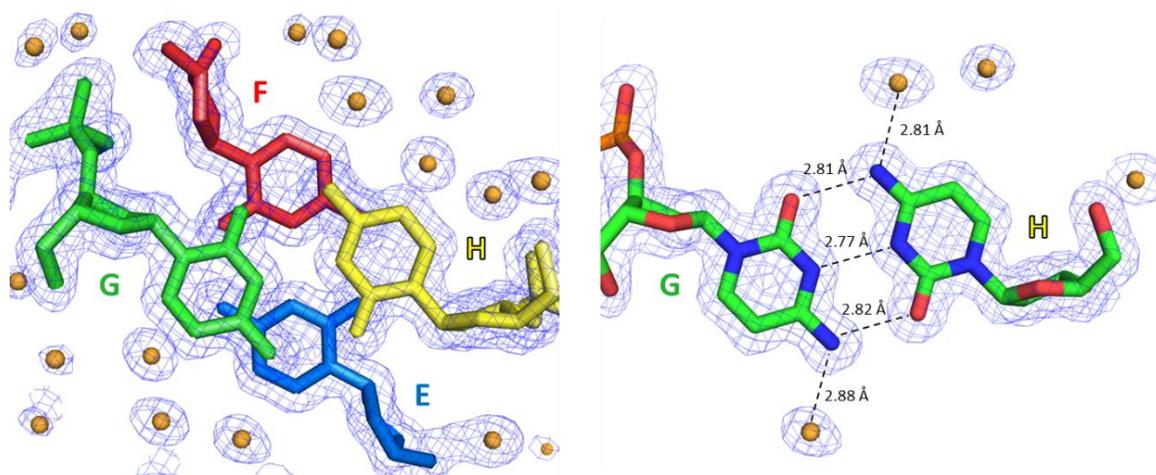
Figure 5.5: (Left) The stacking of bases in the i-motif core involves only exocyclic atoms. Residues are coloured matching to chains E to H. (Right) Bond length parameters in a C$^+$·C base pair. Carbon is green, nitrogen is blue, oxygen is red and phosphorus is orange. Water molecules are represented by orange spheres. The 2Fo-Fc electron density map is drawn at the 1σ contour level in blue.

_____

The sugar puckers in DNA/RNA structures are predominately in either C3′-endo (A-DNA or RNA) or C2′-endo (B-DNA), corresponding to the A- or B-form conformation in a duplex. Four of the sugar puckers in the C_4 i-motifs have C3'-*endo* conformation and six are in the C2'-*endo* form (Appendix Table A5.1). These parameters were calculated using the DNA analysis web server w3DNA.[194] The chi (χ) torsion angle, which characterizes the relative base/sugar orientation, has an average value of 235° (Appendix Table A5.2) meaning that the rings have high *anti* conformations.

Comparison with 190D

The principal difference between the 190D and the new C_4 structure is that the latter consists of alternate phosphate conformations (Fig. 5.6). This may be due to improved experimental approaches or the availability of a more ordered crystal. Two phosphates were modelled at 50% occupancy each in chains B (at P_3; third phosphorus towards the 3'-end), C (P_2, P_3) and E (P_1, P_2).
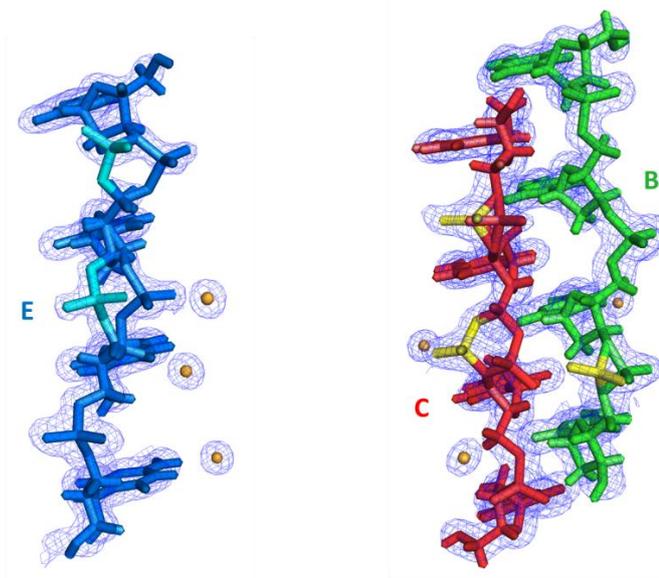
Figure 5.6: Alternative phosphate conformations in chains (left) E (cyan), (right) B and C (yellow) of the C_4 structure. 2Fo-Fc electron density map is drawn plotted in blue at 1σ contour level.

However, both structures still share similar key features of the i-motif being a low-twist, right-handed helix with an average base stacking distance of 3.1Å. There are differences in P-P distances in both the narrow and wide grooves which are largely attributed to alternative phosphate conformations in the new structure. The two structures were superimposed and Figure 5.7 illustrates the resulting difference in coordinates. Table 5.6 states the key differences between the two structures.

Both Figure 5.7 and the Table 5.6 illustrate that the two i-motif structures are similar even though 190D was crystallised in pH 5.5 and the crystal for new structure grew in neutral pH. There is a difference of 3.7Å in the width of the narrow groove. This is because the average width includes the

_____

larger P-P distances resulted by alternate conformations. Almost 3 times more water molecules were observed in the new structure, most likely due to the new C_4 data collected at higher resolution.
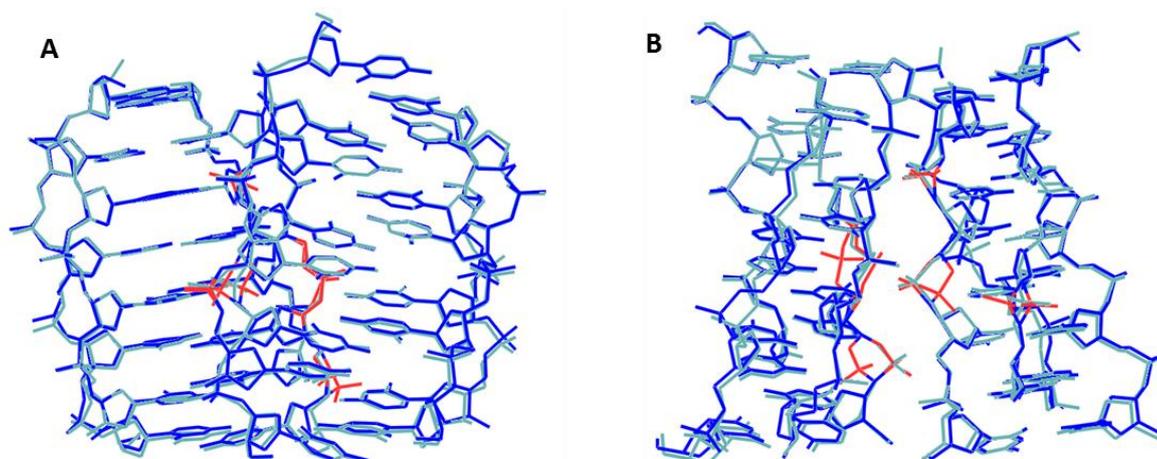
Figure 5.7: Superimposition of the new C_4 structure (blue) onto 190D (cyan), showing views of the asymmetric unit of C_4 looking into the (**A**) wide and the (**B**) narrow grooves. The parts of the structure highlighted red are the alternative phosphate conformations in the new structure.

Table 5.6: Differences in structural parameters between 190D and the new C_4 structure.

DNA Sequence	190D	C_4
Helical twist (°)	12.4	13.5
Groove width (Å)		
Minor groove	7.1	10.8
Major groove	15.9	16.5
No. of Sugar Puckers		
C4'-exo	15	17
C3'-endo	5	4
C2'-endo	4	6
C2'-exo	0	1
C1'-exo	5	1
04'-endo	3	3
Average $C^+ \cdot C$ distance (Å)		
N-H···O bond	2.75	2.82
N-H···N bond	2.76	2.77
No. of waters	58	164

_____

5.3.2 C₃T

The unit cell parameters for the previously published C_3T structure (PDB ID: 191D) were reported as *a*=28.3Å, *b*=44.3Å, *c*=50.5Å and *α*=*β*=*γ*=90.00°. The new C_3T crystal was initially identified as having the $C222_1$ orthorhombic space group and MR gave a high Translation Function Z-score (TFZ) of 10.9, suggesting that the initial model was in the correct orientation. The first round of refinement gave R_{work}/R_{free} values of 0.4874/0.5026. However, further refinement gave relatively low R_{work}/R_{free} values of 0.19/0.22 even though the data was collected to a resolution of 0.68Å.

In an attempt to improve the model, the data was then reprocessed in lower symmetries *C*2 and *P*1, in which MR with the same ensemble model gave TFZ-scores of 3.81 and 3.00, respectively. Subsequently, the data was processed and modelled with $P2_1$ symmetry with new unit cell parameters of a=26.1Å, b=49.5Å, c=26.03Å and α=γ=90.00°, β =114.8°. This gave a TFZ score of 57.4 and the first refinement gave R_{work}/R_{free} of 0.2964/0.3007. Current R-values stand at 0.1546/0.1745 with 5.2% of reflections reserved for the R_{free} set, which are more appropriate considering the quality of the data.

The C_3T structure has the intercalated cytosine core characteristic of all i-motifs. The average P-P intrachain distance is 6.45Å and the average interchain P-P distance is 6.37Å. The major groove is 16.77Å wide and the minor groove is 9.17Å wide. The average cytosine base-stacking distance is 3.1Å. The N3-N3 hydrogen length is 2.80Å. The O2-N4 and N4-O2 bonds are 2.82Å and 2.77Å respectively. Out of the 12 cytidine sugar rings in the asymmetric unit, four of them have C3'-*endo*, six C2'-*endo*, eight C4'-*exo*, four C1'-*exo* and two O4'-*endo* puckers (Appendix; Table A5.3). The glycosidic angle χ has an average value of 235.6° (Appendix; Table A5.4) so, the rings have high *anti* conformation.

While C_4 is held entirely by C⁺·C base pairs the thymine residues in C_3T can adopt two conformations. Figure 5.8A shows that one thymine residue remains stacked parallel to the plane of the core whereas the other is flipped away from the plane of the core where it is involved in lattice interactions. The O4-N3 bond length in the T·T base pair is 2.84Å and the N3-O2 bond is 2.79Å long (Fig. 5.9). The thymine residues play an essential role in building the lattice. A twofold screw axis is found perpendicular to the broad groove at the end of the molecule; this results in forming a T·T base pair holding molecules together along the helical axis. The other two thymine residues are not stacked in the molecule but have an orientation parallel to the helical axis and perpendicular to the wide groove. These thymine residues form stabilising interactions[34,195] that build up the lattice (Fig. 5.10).

_____

The torsion angle α around the phosphate group adopts two different sets of values, one associated with the sugar phosphate chains that are straight and another associated with those in which the phosphates are rotated away from the centre of the molecule. These bent phosphate groups are stabilised by water bound to NH groups on the sides of the cytosine core. Water molecules help stabilise the cytosine core by base pairing with the N4 atoms in cytosine (Fig. 5.9). The N4 to water distance is approximately 2.9Å. They are hence found bridging between the i-motif molecules. Water molecules also stabilise the T·T base pair by interacting with both O2 and O4 atoms in the thymine residues. There are 3.7 times more water molecules in the new structure, which is highly likely, the result of improved data collection methods.
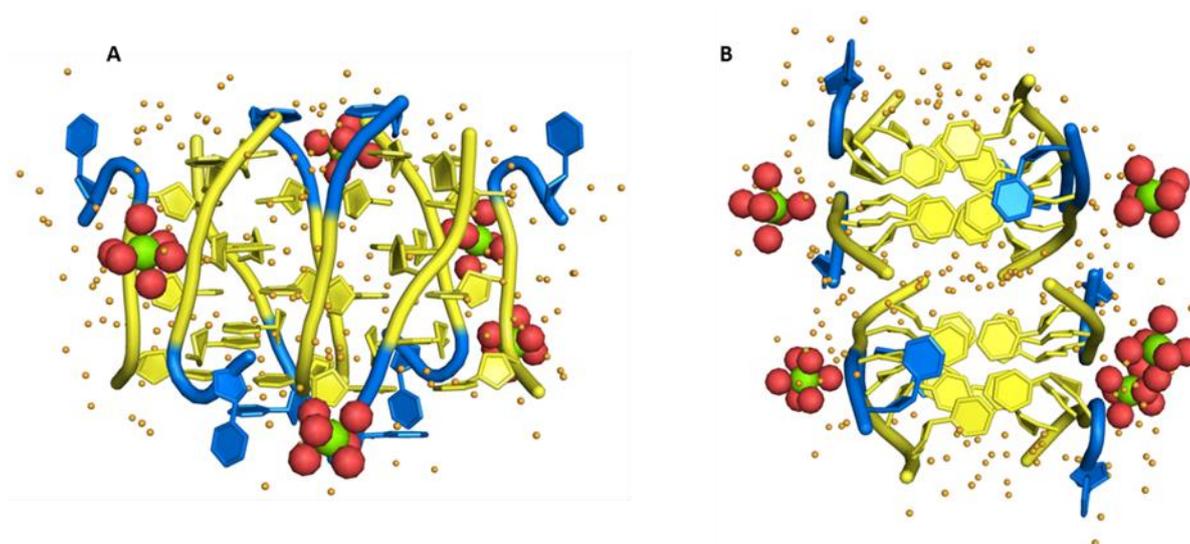
Figure 5.8: Views of the asymmetric unit of C_3T looking (**A**) along and (**B**) down the helical axes. Cytosine and thymine bases have been coloured yellow and blue respectively. The green and red spheres are magnesium with six coordinated water molecules. The orange spheres represent water molecules.
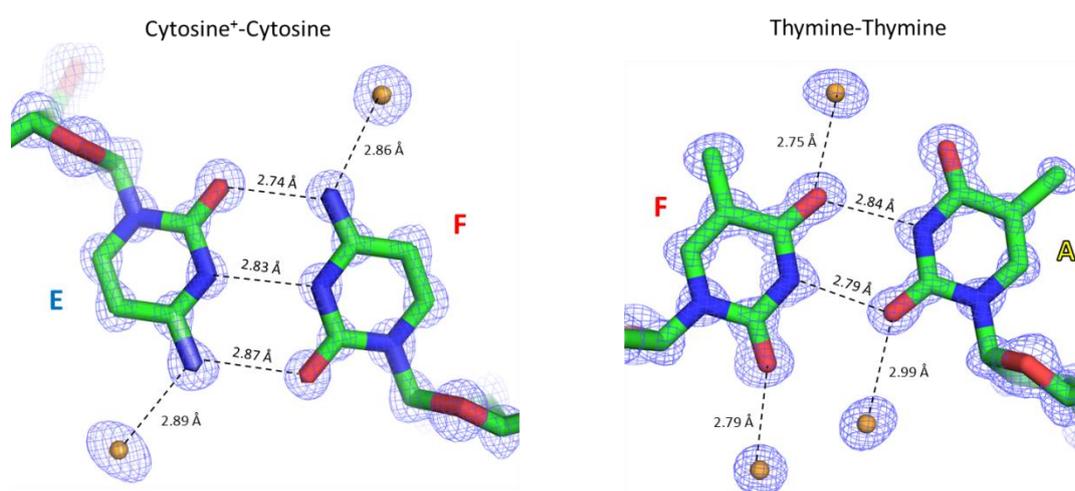
Figure 5.9: (Left) The $C^+·C$ and (right) T-T base pairs. Carbon is green, nitrogen is blue, oxygen is red and phosphorus is orange. Water molecules are represented by orange spheres. The electron density map is drawn blue at the 1σ contour level.

_____

wide groove	narrow groove	core ends
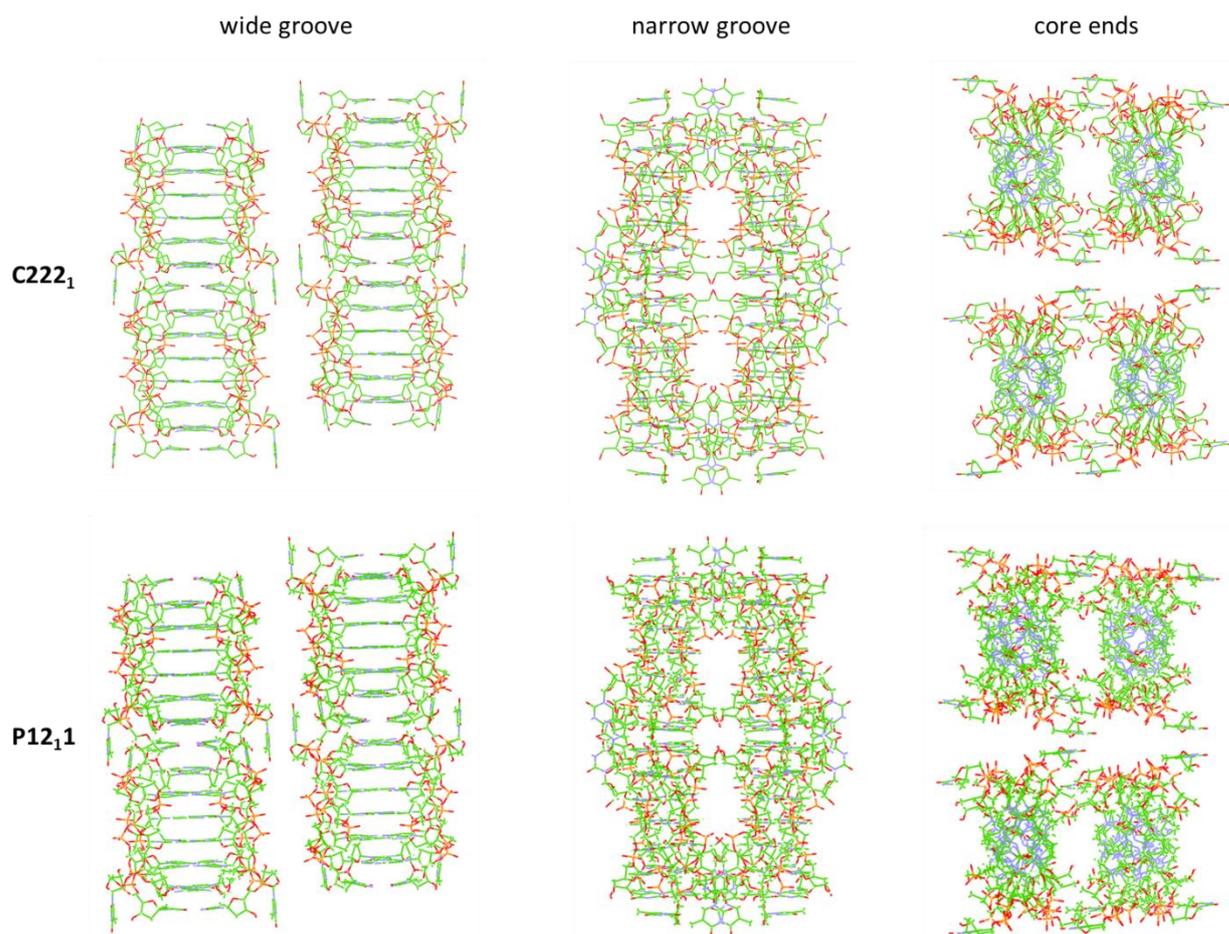
Figure 5.10: Lattice of 191D (*C*222₁) and the new C₃T (*P*2₁) showing the packing of molecules as seen through the wide and narrow grooves (along the helical axes) and down the helical axes. The i-motif molecules are held together by the stacking of flipped thymine residues upon each other.

The new C_3T molecule is flat with a right-handed twist of 19.11° whereas 191D has a helical twist of 17.1°. The asymmetric unit consists of 32 nucleotides, 208 water molecules and five magnesium ions. The crystal used to obtain 191D was grown in conditions including 20mM $MgCl_2$ but only three Na^+ were assigned as metal ions around the i-motifs. In the new structure, magnesium ions have been identified in the lattice surrounded by distorted octahedra of water molecules and phosphate oxygens in their coordination sphere (Fig. 5.11).
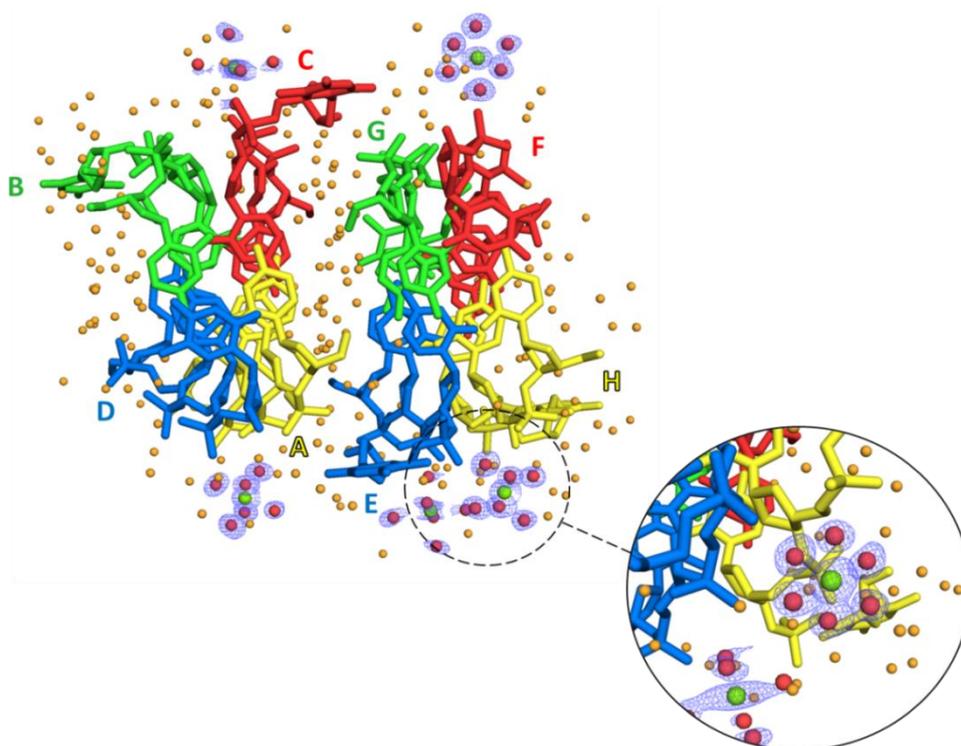
Figure 5.11: View of the asymmetric unit of C$_3$T down the helical axes. The eight strands have been labelled A to H with chain A base pairing with B, C-D, E-F and G-H. The orange and red spheres represent water molecules. Magnesium atoms are green spheres. 2Fo-Fc electron density map surrounding [Mg(H$_2$O)$_6$]$^{2+}$ is drawn at 1σ contour level.

There are five [Mg(H$_2$O)$_6$]$^{2+}$ complexes near the phosphate backbones. Figure 5.12 shows the bond length distances between the Mg^{2+} ion and the six waters coordinated to the metal. The first neutron structure of magnesium chloride hexahydrate reports an average Mg-O bond length of 2.08Å,[196,197] which is slightly longer than the Mg-O distance of 2.01Å reported here. The distances between the oxygen from a phosphate group (depending which one is closer), and its nearest Mg^{2+} are listed in Table 5.7. The PO–Mg^{2+} distances are also reported.
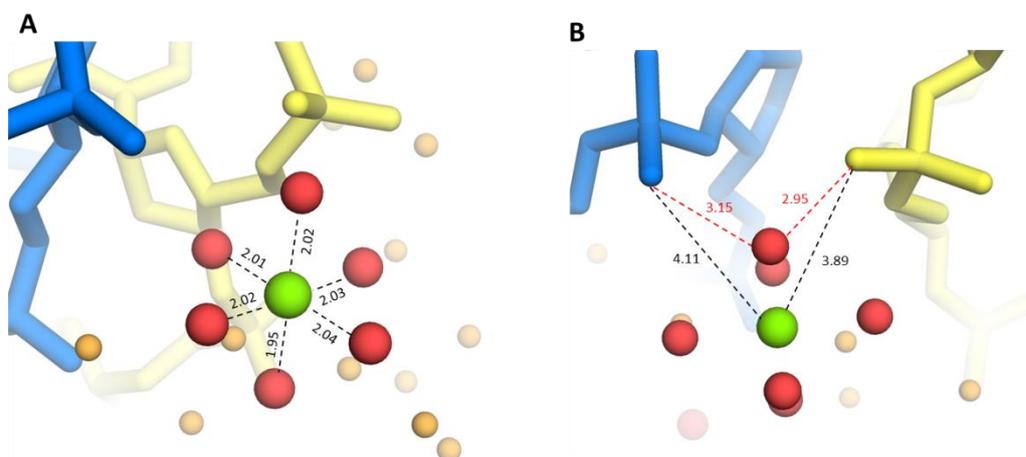
Figure 5.12: (A) Bond length distances of the [Mg(H$_2$O)$_6$]$^{2+}$ complex in Å. **(B)** Distance measurements between the complex and its neighbouring phosphate groups. Chains A and D are coloured yellow and blue respectively.

_____

Table 5.7: Distance measurements (Å) between $[Mg(H_2O)_6]^{2+}$ and its neighbouring phosphate oxygens.

Chain neighbouring complex	PO-H$_2$O	PO-Mg^{2+}
A	4.11	3.12
C	4.22	3.02
E	4.15	2.87
F	3.68	2.80
H + E	3.89	2.95

The positively charged magnesium complexes near the negative phosphate groups may have contributed to disorder in the structure. In contrast to 191D, positive densities were observed in the new difference map for multiple phosphate group conformations. Again, this could be due to the new crystallisation conditions or better X-ray source that we were able to see details of disorder which are not reported in 191D.

Positive densities for alternative phosphate conformations were observed in the Fo-Fc map throughout half of the chains in the asymmetric unit (Fig. 5.13). Simulated annealing was also applied to lower high B-factors in order to lower R-factors. However, only anisotropic refinement with single backbone conformations was successful in lowering the R-factors. High-resolution data of the new C$_3$T has allowed the modelling of hydrogen atoms in some sugar rings, however hydrogen atoms could not be located in water molecules (Fig. 5.13, D).
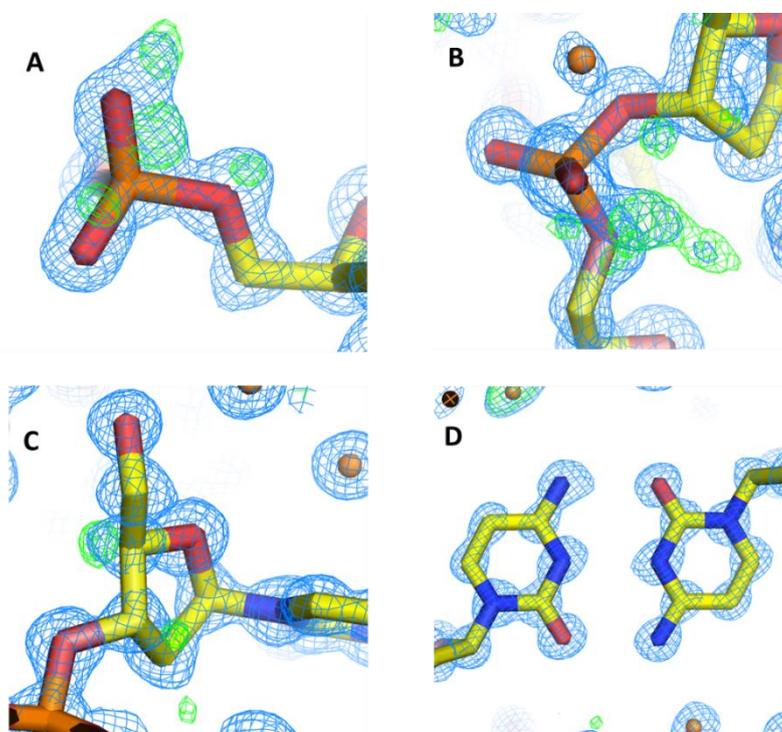
Figure 5.13: 2Fo-Fc (blue) and Fo-Fc (green for positive and red for negative) electron density maps, drawn at 1σ and 3σ contour levels, respectively. (**A-B**) The Fo-Fc map suggests alternative phosphate conformations. (**C**) The Fo-Fc map suggests hydrogen occupancies in the sugar rings. (**D**) C$^+$·C base pair. Carbon is yellow, oxygen is red and phosphorus is orange. Water molecules are orange spheres.

_____

Comparison with 191D

When C_3T and 191D are superimposed, the root mean square deviation between the atomic positions is 0.0284Å, showing that they are similar in terms of atomic position even though the first was crystallised in pH 5.5 and 191D in pH 6.5. Superimposition of one of the cores (chains A-D) onto 191D gives an rms deviation of 1.1035 Å so, the i-motif structures alone are very similar to each other. Table 5.8 and Figure 5.14 illustrate the difference in coordinates and structural parameters between the two i-motifs.
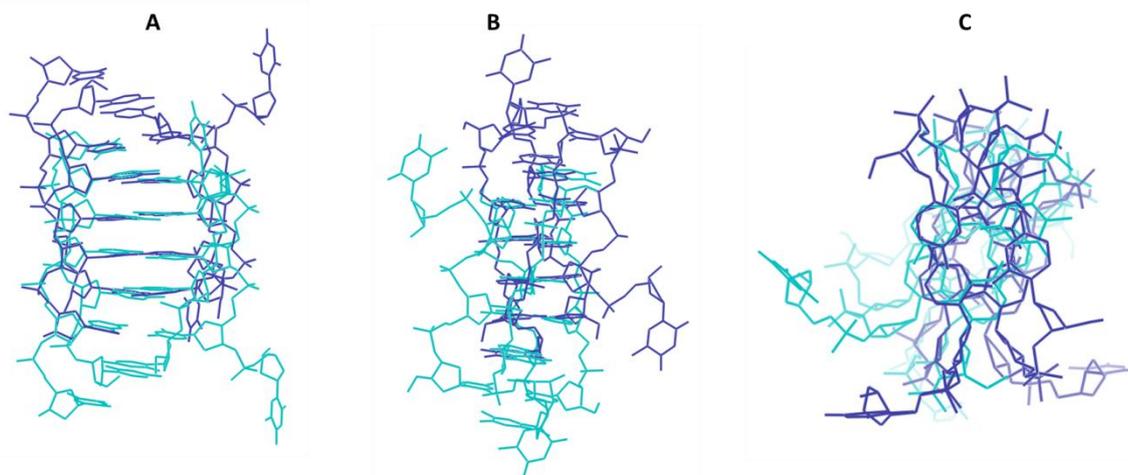
Figure 5.14: The new C_4T structure (blue) and 191D (cyan) superimposed from chain A. Views of the asymmetric unit of C_4 looking into the (**A**) wide and (**B**) narrow grooves and (**C**) down the helical axes.

Table 5.8: Differences in structural parameters between 191D and the new C_3T structure.

DNA Sequence	191D	C_3T
Space group	$C222_1$	$P2_1$
Helical twist (°)	17.1	19.1
Groove width (Å)		
Minor groove	7.02	9.17
Major groove	16.63	16.77
No. of Sugar Puckers		
C3'-endo	6	4
C2'-endo	2	6
Average $C^+ \cdot C$ distance (Å)		
O2-N4	2.77	2.80
N3-N3	2.74	2.80
N-H···O water bond (Å)	3.00	2.88
Average T-T distance (Å)		
O4-N3	2.81	2.79
N3-O2	2.84	2.79
No. of waters	56	208
No. of metal ions	3 Na$^+$	5 Mg^{2+}

_____

5.3.3 C_3A

The C_3A crystal has cell parameters of ($a=b=24.3$, $c=51.1$, $\alpha=\beta=\gamma=90$) in the tetragonal $I422$ space group. Although MR using a single strand of C_3A as an initial model was successful with a high TFZ-score of 14.7, the refinement process failed with an error relating to a missing sigFP column even though the column was present. The sigFP column is related to the amplitude of the wave $|F_{hkl}|$ which is proportional to the square root of the intensity measured on the detector. The same error was observed for refinement on data processed with a lower symmetry of $I4$. Therefore, data was processed in P1 space group with MR carried out using 8 strands of C_3A.

The highest possible symmetry for the particular dataset was later found to be $C2$. This was confirmed using the macromolecular phasing software; SHELX C/D/E. The data was then processed with xia2 in $C2$ space group. At the current stage of refinement, R_{work}/R_{free} is 0.1985/0.2077 with 4.9% of reflections reserved for the R_{free} set. The asymmetric unit of C_3A consists of four 5'-CCCA-3' single strands or 16 nucleotides (Fig. 5.15) and 91 water molecules. Positive density for a K^+ ion was observed. Although not tested during data collection, X-ray fluorescence could have been used on the crystal to confirm the presence of the element. The current C_3A structure has been modelled as having a single potassium ion which is 5.08Å from both of its nearest phosphate oxygens of chains A and B.
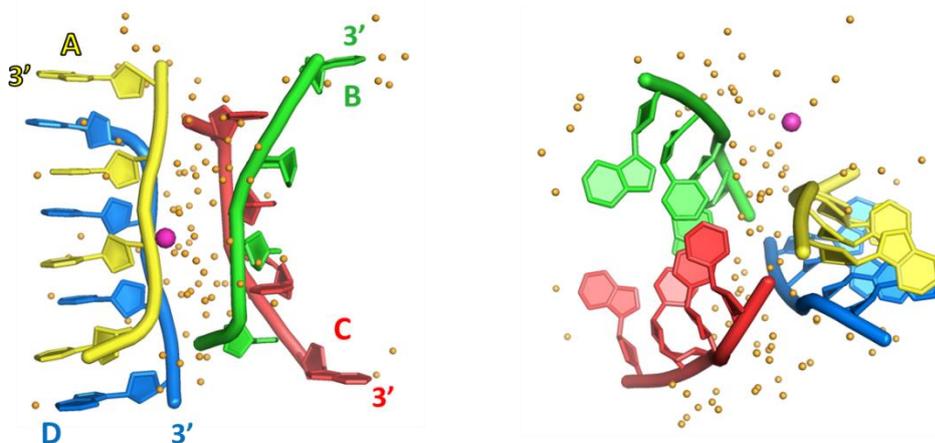
Figure 5.15: Views of the asymmetric unit of C_3A (left) along and (right) down the helical axes. Orange spheres represent water molecules and the purple sphere is K^+.

C_3A consists of a cytosine core similar to all i-motif structures seen previously. Chain A undergoes base pairing with chain D and B base pairs with C by symmetry related i-motif building the lattice (Fig. 5.16). The average P-P intrachain distance is 6.51Å and the average interchain P-P distance is 6.73Å. The average P-P distance across the wide groove is 16.82Å and the distance between the phosphorus atoms in the narrow groove is 8.12Å. In the $C^+\cdot C$ base pair (Fig. 5.17), the N3-N3 hydrogen length is 2.78Å. The O2-N4 and N4-O2 have bond lengths of 2.83Å. The glycosidic

_____

angle χ has an average value of 233.9° (Appendix; Table A3) with the rings possessing high *anti* conformations.
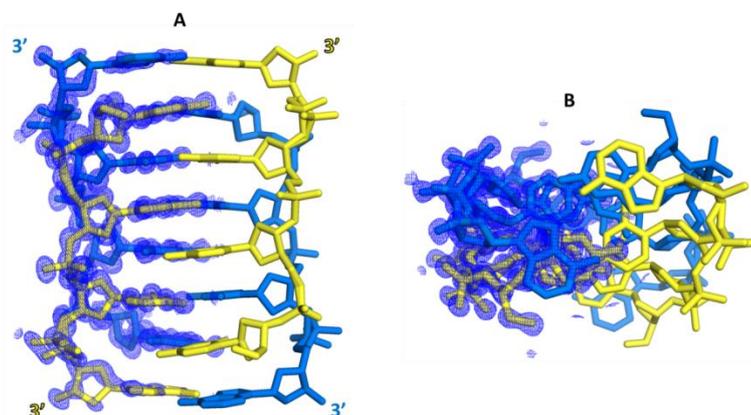
Figure 5.16: Chains A (yellow) and D (blue) base pair with each other to form a symmetry related i-motif of C$_3$A. The 2Fo-Fc electron density map is drawn in blue at the 1σ contour level. (**A**) View of the i-motif along and (**B**) down the helical axes.
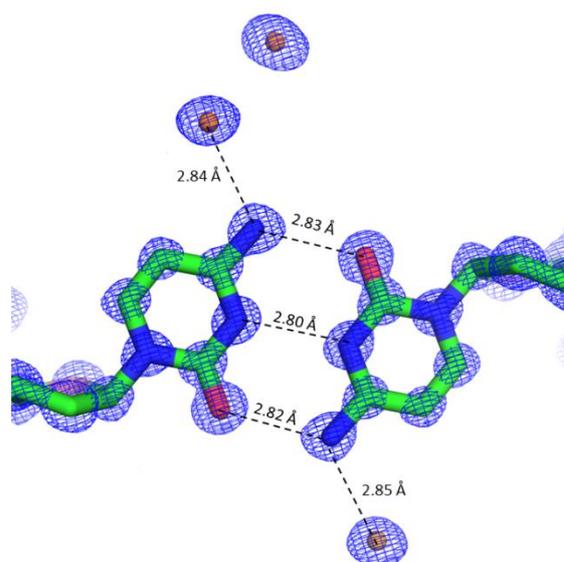
Figure 5.17: A view of the 2Fo-Fc electron density map, drawn in blue at the 1σ contour level, surrounding the C$^+$·C base pairs. Carbon is green, nitrogen is blue and oxygen is red. Water molecules are represented by orange spheres.

Positive densities for alternative DNA backbone conformations were also observed for C$_3$A (Fig. 5.18) where the structure was fitted with multiple phosphate group occupancies. In addition to the C$^+$·C core, the i-motifs are also stabilised by adenine·adenine base pairs (Fig. 5.19). The C1'·C1' distance is 12.01Å, suggesting that the base pair belongs to the *cis* Watson-Crick/Watson-Crick geometric family, where nitrogen at the N6 position of one adenine base pairs with N7 of another andenine.[198] The average N6-N7 bond distance is 4.87Å. Figure 5.20 shows that the water molecules are distributed around the wide and narrow grooves and away from the core, suggesting the stabilising role of water in the i-motif structure.
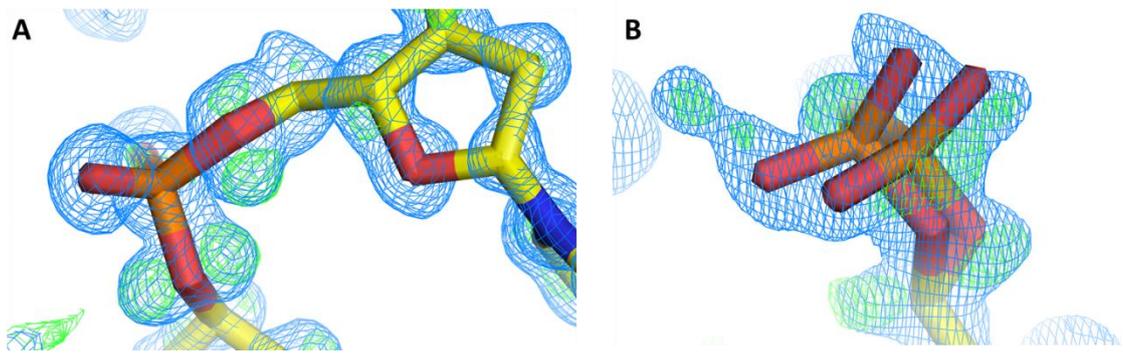
Figure 5.18: 2Fo-Fc (blue) and Fo-Fc (green for positive and red for negative) electron density maps of C_3A, drawn in 1σ and 3σ contour levels, respectively. (**A**) The Fo-Fc map suggests hydrogen occupancies in the sugar rings. (**B**) $C^+ \cdot C$ base pair. Carbon is yellow, oxygen is red and phosphorus is orange.
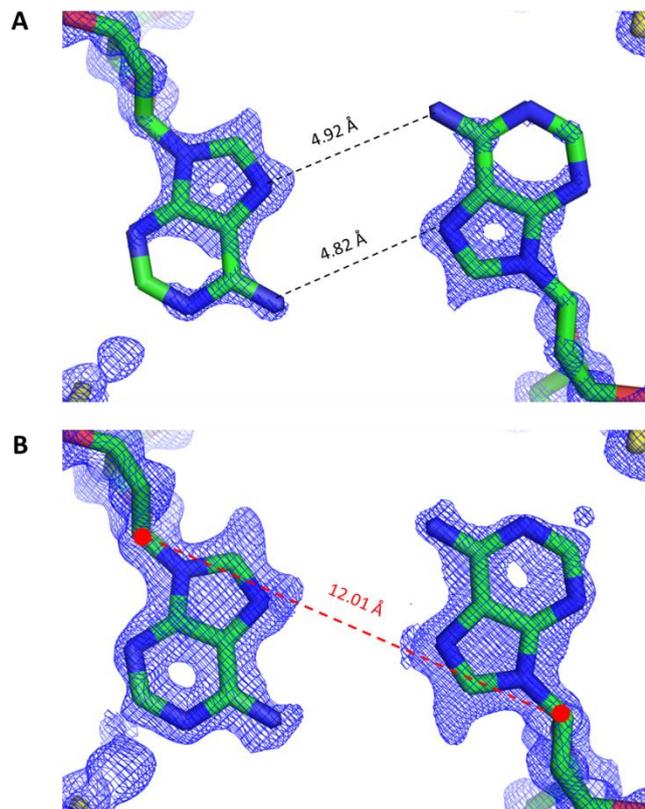
Figure 5.19: (**A**) The A·A base pair. (**B**) Distance between the C1' carbon atoms of two adenines. Carbon is green, nitrogen is blue and oxygen is red. The 2Fo-Fc electron density map is drawn in blue at the 1σ contour level.

_____
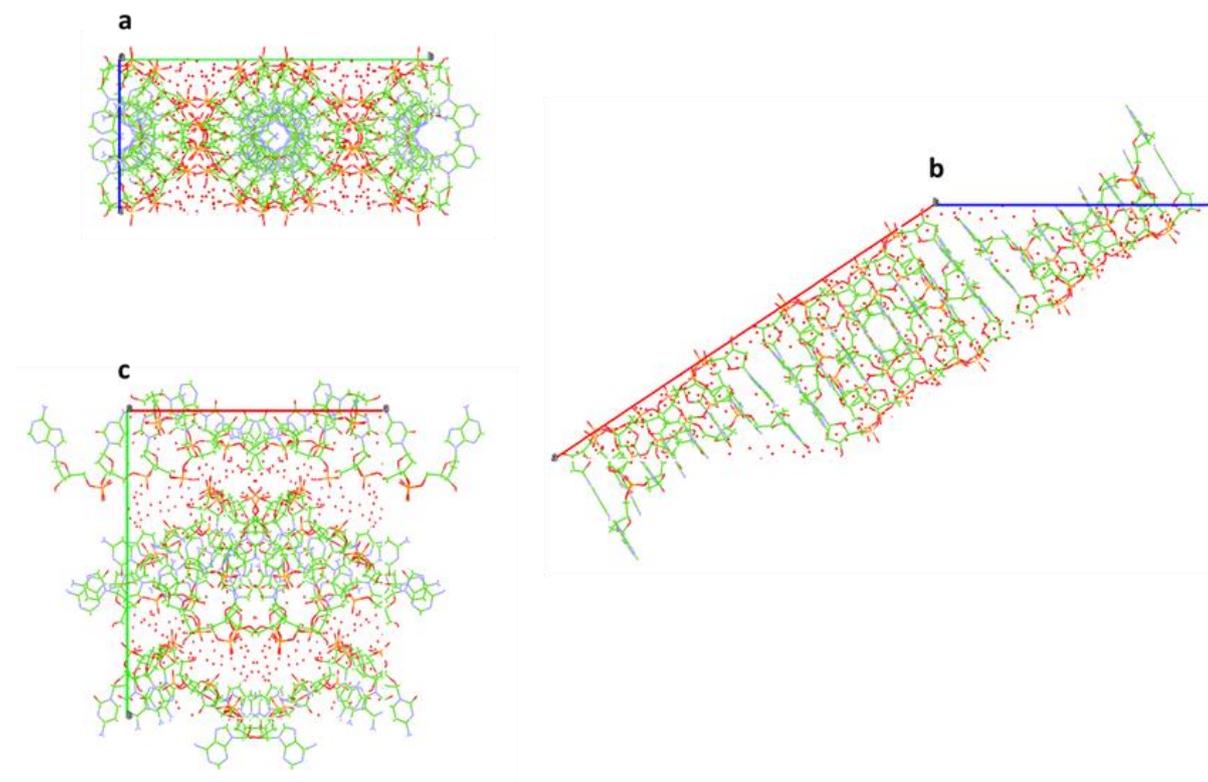
Figure 5.20: Packing diagrams of C₃A as seen down the a, b and c axes. Carbon is green, nitrogen is blue and oxygen is red. The red dots are water molecules in the solvent channels.

_____

5.3.4 C₃A + C₃T

The data obtained for [C$_3$A + C$_3$T] was initially modelled in a tetragonal $P42_12$ space group but the map obtained after MR could not distinguish between adenine and thymine residues. The collected data was hence processed in $P1$ space group. The R$_{work}$/R$_{free}$ currently stands at 0.2117/0.2169 with 4.9% of reflections reserved for the R$_{free}$ set. The asymmetric unit of [C$_3$A + C$_3$T] consists of 32 nucleotides and 162 water molecules. While viewing through the b axis, one of the i-motifs exposes it narrow groove whereas the other its major groove (Fig. 5.21).
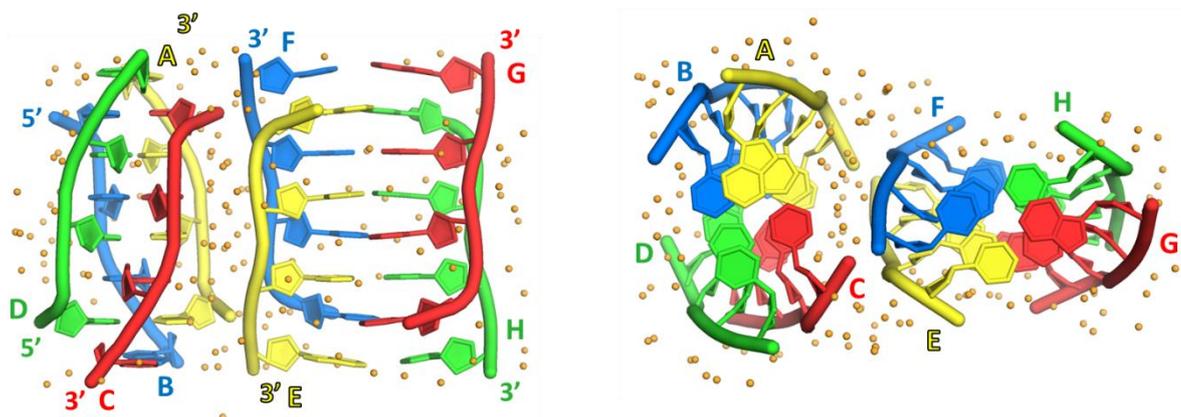
Figure 5.21: Views of the asymmetric unit of C$_3$A + C$_3$T (**A**) along and (**B**) down the helical axes. Chain A base pairs with chain D, B-C, E-H and F-G. Orange spheres represent water molecules.

The [C$_3$A + C$_3$T] structure is right-handed with a helical twist of 19.57°. The average P-P intrachain distance is 6.59Å and the average interchain P-P distance is 6.71Å. The average P-P distance across the wide groove is 16.93Å and the distance between the phosphorus atoms in the narrow groove is 7.90Å. The average base-stacking distance is 3.10Å, in similarity to all i-motif structures visited in this chapter. The N3-N3 hydrogen length in the C$^+$·C base pair is 2.79Å (Fig. 5.22). The O2-N4 and N4-O2 have bond lengths of 2.83Å. The glycosidic angle χ has an average value of 236.1° (Appendix; Table A5.7).
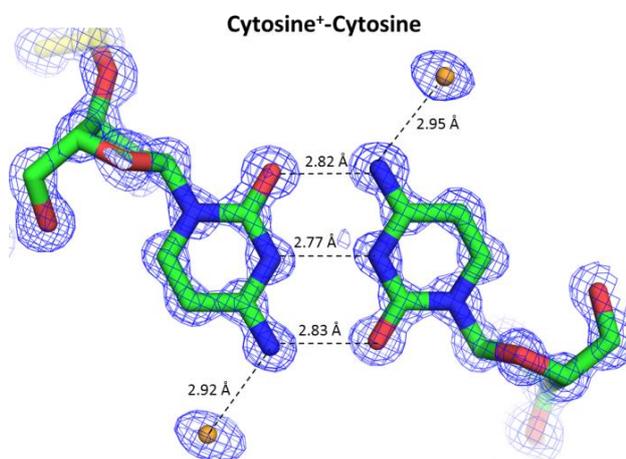
Figure 5.22: The C$^+$·C base pairs in [C$_3$A + C$_3$T]. Carbon is green, nitrogen is blue and oxygen is red. Water molecules are orange spheres. 2Fo-Fc electron density map is drawn in blue at 1σ contour level.

_____

The ends of the core are held together by Watson-Crick A-T base pairing, but the densities observed for the A·T base pair region, both before and after refinement, suggested that each base can be modelled in either densities. Figure 5.23 shows that addition of either base in any order in the A·T base pair is acceptable, i.e. it could exist as T·A base pair as well. The difference in R_{work}/R_{free} between the two fittings was 1.38/0.15%. A strategy was undertaken where alternative base conformations were added with 50% occupancy for each of the base, however, the structure building software; Coot,[192] only allows for multiple conformations of the same base. A second strategy was applied where each chain was duplicated with 50% occupancy for C_3A and 50% for C_3T but this resulted in an increase in R-values due to clashes of over 200 units. Refinement was also carried out with just adenine or thymine, the individual base on its own, which increased the R-values and resulted in negative densities, therefore, dismissing the structure formation possibilities of either C_3A or C_3T. The current structure is fitted as C_3A for chains A, B, E and G and C_3T for chains C, D, F and H.
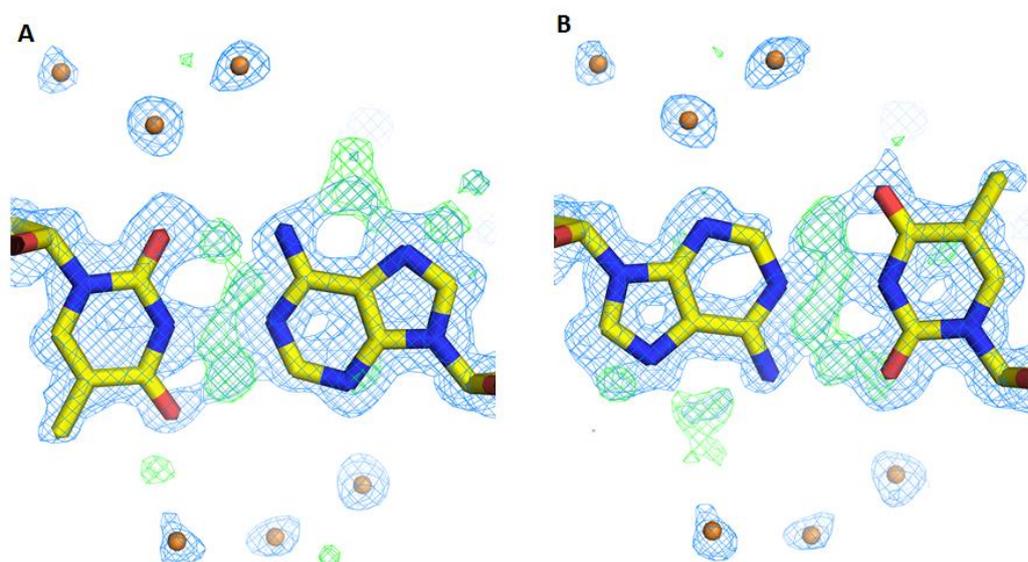
Figure 5.23: 2Fo-Fc electron density map surrounding the A·T base pairs in [C_3A + C_3T], drawn in blue at 1σ contour level. (A) Refinement output after an A·T base pair fit. (B) Swap in the bases pre-second refinement in the same density. Carbon is yellow, nitrogen is blue and oxygen is red. Water molecules are orange spheres.

Even though the assignment of the A·T base pairs is still in question, the structure is 75% complete with regards to the i-motif molecules. The output would still be the same even if the assignment of two bases were to switch. The average N6-O2 bond length distance in the A·T base pair can then be measured as 3.13Å and N1-N3 as 2.78Å (Fig. 5.24). Figure 5.25 shows Van der Waals representations of the i-motif. Figure 5.26 shows the packing diagrams of the [C_3A + C_3T] structure. In similarity to the last three i-motif structures discussed, the solvent in this structure is distributed within the grooves of the DNA, highlighting the importance of bridging water in i-motif stability. The nearest water molecule is 2.91Å far from both N3 in adenine and O4 in thymine in the A·T/T·A base pair.

Adenine-Thymine
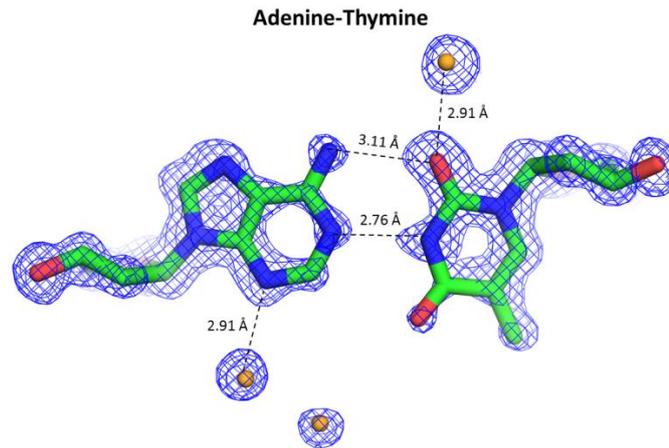
Figure 5.24: 2Fo-Fc electron density map surrounding the A·T base pair between chains B and C of [$C_3A + C_3T$], drawn in blue at 1σ contour level. Carbon is green, nitrogen is blue and oxygen is red. Water molecules are represented by orange spheres.
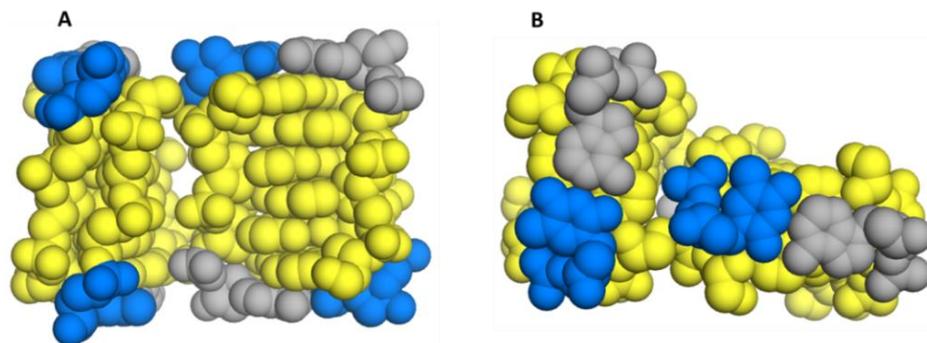
Figure 5.25: Van der Waals representation of the asymmetric unit of [$C_3A + C_3$]. (**A**) along and (**B**) down the helical axes. Yellow spheres represent the cytosine core, blue is thymidine and grey is adenine.
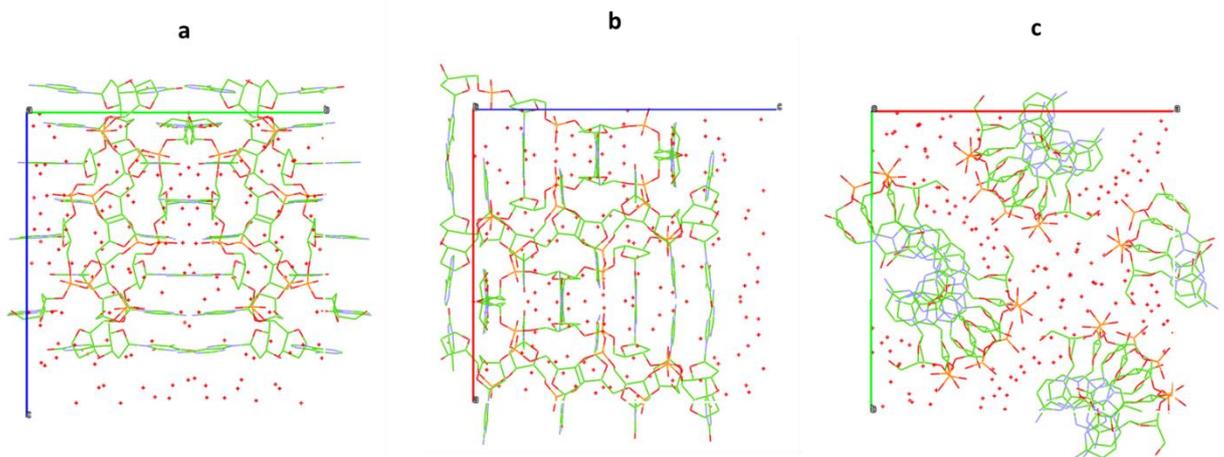
**Figure 5.26:** Packing diagrams of [$C_3A + C_3T$] as seen down the a, b and c axes. Carbon is green, nitrogen is blue and oxygen is red. The red dots are water molecules.

_____

5.4 Conclusions

All i-motifs presented have a close base stacking distance of 3.10Å in comparison to the more familiar B-DNA stacking of 3.4Å. The base stacking in i-motifs only involves the exocyclic residues which may be associated with a larger overlap of the local π-electron clouds. The cytosine core for all i-motifs studied here are similar with the hemi-protonated N3-N3 bond distance in the $C^+ \cdot C$ base pairs of 2.77Å. The P-P distances are varied for all structures (Table 5.9).

Thymidines are able to form stacking interactions so, C_3T was the only structure in which a nucleotide was able to rotate away from the cytosine core to build up a lattice.[195] The remaining three i-motifs have contained structures in comparison, where all bases undergo base pairing. The $[C_3T + C_3A]$ i-motif is more closely packed than the C_3A structure as the former has the shorter inter-chain distance and narrow groove, which is where two i-motif cores face each other. Therefore, $[C_3T + C_3A]$ may be more stable than C_3A as a Watson-Crick A·T base pair is more stabilising than a *cis*-Watson-Crick/Watson-Crick A·A base pair.[198]

Table 5.9: Phosphate-phosphate distances in tetramolecular i-motifs.

Average P-P distance (Å)	$C_3T + C_3A$	C_3A	C_3T	C_4
Intra-chain	6.59	6.51	6.45	6.50
Inter-chain	6.71	6.73	6.37	8.40
Narrow groove	7.90	8.12	9.17	10.80
Wide groove	16.93	16.82	16.77	16.90

As there was no systematic study of DNA sequences forming tetramolecular i-motifs before, it was not possible to predict which non-cytosine base flips away from the central core. Looking at these results, it can be suggested that the intramolecular loop regions may not even form from anti-parallel base pairing as it is believed to, but instead, the structure could form from parallel strands base pairing on either sides of the cytosine core.

Although data collection at high resolutions was able to locate hydrogens in the sugar rings and in some cases, the bases, there were difficulties in refinement (in lowering the R_{work}/R_{free} values). The F_o-F_c maps of all four crystals showed positive densities for alternative DNA chain conformations. This positional disorder, however, did not exceed more than two equally plausible conformations of an affected residue. The modelled temperature factors for majority of the atoms were decreased using simulated annealing and anisotropic refinement was only performed in later stages as it obscures disorder.

Substitutional disorder[100] was observed for $[C_3T + C_3A]$, where the same electron density site could be occupied by both A and T bases. Two strategies were used to solve the problem: adding alternative base conformations with 50% occupancy for each of the base, and duplicating each chain

_____

with 50% occupancy for each of the CCCA and CCCT sequences. However, the first strategy failed due to restrictions in the structure building software and the second approach resulted in an increase in R-values by ~0.25 due to clashes of over 200 units. Either way, the proposed structure is relevant as it stands due to the possibilities of having AT or TA base pairs on both ends of the cytosine cores.

C_4 and C_3A crystals were grown in pH 7.0 whereas the $[C_3A + C_3T]$ crystal was grown in pH 8.5. The 190D C_4 crystal was grown in pH 5.5 and its structure was similar to the new C_4 structure. This implies that i-motif structures may form in physiological conditions. In terms of sugar puckering and base pair distances, no major structural differences were observed for the previously published and newly modelled structures of C_3T and C_3A. The new structure of C_3T was solved in $P2_1$ space group in contrast to $C222_1$ in 191D. It also features the presence of $Mg(H_2O)_6^{2+}$. More water molecules, which contribute heavily to the stability of the DNA backbones, were found in both structures in comparison to previous reports. The novel structures of $[C_3T + C_3A]$ and C_3A have been presented.

Due to the very limited number of i-motif crystal structures reported to date, the four structures in this study could only be compared with each other or to the six tetramolecular structures present in the Nucleic Acid Database[199] (NDB). The crystal structure of C_3A shows that A·A base pairs occur on either side of the cytosine core. This is in contrast to the folding of all adenine bases away from the cytosine core in the A_2C_4 structure. NMR studies have shown that protonation of adenine at loops disrupts the i-motif core[75] and chapter 4 showed that adenine-rich i-motifs form less stable i-motif structures. This may explain why the presence of more than two adenine bases does not result in A·A base pairs. CD spectroscopy measurements also revealed that 83% of C-rich sequences that had guanine as the first base in the loops formed either DNA hairpins or G-quadruplex structures instead of i-motifs. This may explain why no crystals were obtained for C_3G as crystallisation conditions need to be optimised for hairpin/G-quadruplex formation instead. These results can, therefore, be used to design i-motif forming sequences for bimolecular and intramolecular i-motif crystallisation.

Gehring *et. al.* stated that van der Waals stabilisation between the sugar phosphate backbones across the narrow grooves and the opposite orientation of the carbonyl and amino group dipoles are two major contributing factors to i-motif stability.[36] The water molecules further add to this stabilisation as portrayed by its bridging role; connecting individual CCCX chains in all of the four structures. The presence of $Mg(H_2O)_6^{2+}$ in C_3T also displays stabilisation of the i-motif with the help of ligands. Overall, it can be said that the water ligands, in both cases, help stabilise the structure through phosphate interaction. However, high-resolution data was not able to locate hydrogen atom positions in water molecules so, water to C^+·C base pair distances could not be measured as accurately as hoped for.

_____

Neutron Crystallographic Studies of the d(C$_3$T)$_4$ i-motif

6.1 Introduction

Hemi-protonation of the nitrogen (N) at the N3 position in cytosine (C) is responsible for the formation of the intercalated i-motif structure so, C-rich oligomers are more prone to form i-motifs under acidic conditions. This unique property of the i-motif has enabled it to be the first DNA "molecular motor" driven by pH changes,[42] making it an attractive subject in DNA nanotechnology.

The location of the proton between the hemi-protonated C$^+$·C base pairs in an i-motif crystal has never been determined. It is not clear if this proton resides on a single base pair or if it is shared equally between the two cytosine bases. Hydrogen bonding in the hemi-protonated N⋯H$^+$⋯N moiety has, therefore, been described as either a symmetric hydrogen bond with a single-well potential (Fig. 6.1A) or an asymmetric hydrogen bond defining a double-well potential with a delocalized proton that oscillates between the two wells (Fig 6.1B).[55,200]
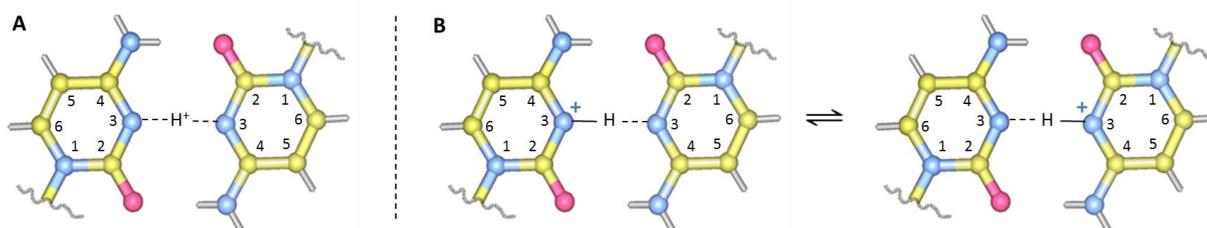
Figure 6.1: (A) Symmetric and **(B)** asymmetric N⋯H$^+$⋯N hydrogen-bonding in the hemi-protonated C$^+$·C base pairs of DNA i-motif. The symmetric system involves 50% of H$^+$ occupancy contributed by each cytosine whereas the asymmetric form involves 100% hydrogen occupancy from only one protonated cytosine.

NMR experimental data and quantum chemical (QM) calculations have shown that proton transfer between two cytosines in the human telomeric i-motif; d(CCCTAA)$_3$CCC, is strongly dependent on the distance between the two N3 atoms.[55] Lieblein *et al.* observed that for N3-N3 (in N⋯H$^+$⋯N) distances smaller than 2.5Å, the H$^+$ proton is shared between the two cytidines and is located precisely in the midpoint between the nitrogen atoms so, initially they assumed a symmetric C$^+$·C base pair. However, they also showed that with increasing N–N distance the proton localizes at one nitrogen atom and undergoes fluctuations in a typical double-well potential. Moreover, H$^+$ localization is slightly higher at "zigzagging" or diagonally opposite cytosine residues (Fig. 6.2). This was attributed to small structural asymmetries in the π-stacking of the hemi-protonated base pairs as well as from electrostatic repulsion.
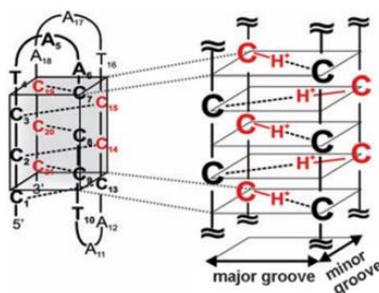
Figure 6.2: Distribution of proton population at cytidine residues in the human telomeric i-motif; d(CCCTAA)$_3$CCC. Reproduced with permission. Lieblein, A. L., Krämer, M., Dreuw, A., Fürtig, B. & Schwalbe, H. The nature of hydrogen bonds in C···H$^+$···C DNA base pairs. *Angew. Chemie - Int. Ed.* **51,** 4067–4070 (2012). Copyright 2017 John Wiley and Sons.

Another important form of hydrogen bonding observed in all four i-motif crystal structures from the previous chapter was the H-bonding between water molecules and the cytosine N4 amino groups in the major grooves. Diffraction data of C$_4$ and C$_3$T revealed densities for 3 times as many water molecules than in the structures of the same i-motif systems published in 1994.[132,134] In the new structures, more water molecules are found near phosphate groups in the minor grooves, illustrating their role in bridging i-motif molecules in both grooves. In the case of C$_3$T, water groups from MgCl$_2$(H$_2$O)$_6$ were also located near phosphate groups. Hence, the position of solvation water and ion distribution was found to be important for the structural stability of the i-motif. Although sub-Ångstrom data collected for three out of the four i-motif systems in the last chapter demonstrated the significance of water molecules in i-motif stability, it was not possible to position atoms in water molecules.

Scattering of X-rays is proportional to electron density; the scattering at Bragg angle $\theta = 0$ from an atom is given by its atomic number so X-ray scattering increases across and down the Periodic Table (Table 6.1).[201] This means that it is very hard to detect hydrogen (H) accurately in macromolecular crystallography using X-rays as it only has one electron. Neutron scattering by nuclei, on the other hand, has no simple pattern of scattering power in terms of the position of elements in the Periodic Table. Scattering length and scattering cross-section are used instead for neutrons and these factors vary from element to element (Table 6.1).[202] The contribution of H atoms is half that of the heavier atoms and has a negative magnitude, whereas deuterium (D) atoms are in-phase neutron scatterers with the scattering length of 6.7 femtometres.[203] The contribution of D to neutron scattering is of the same order of magnitude as the contribution of atoms such as carbon, nitrogen or oxygen.[113] This is why it is essential for crystals for neutron crystallographic studies to be either grown in D$_2$O or exchanged in the medium.

_____

Atom	X-ray (electrons)	Neutron (femtometres)
^{1}H	1	-3.74
^{2}H (D)	1	6.67
^{12}C	6	6.65
^{14}N	7	9.37
^{16}O	8	5.80
^{31}P	15	5.13

Application of neutron crystallography can characterize the position of protons and the orientation of water molecules. Both X-ray and neutron data sets can be collected from the same crystal under the same conditions and refined simultaneously. This concept was pioneered in small-molecule crystallography[205] and subsequently in macromolecular crystallography.[206] Following the data collection of d(C$_3$T)$_4$; abbreviated to C$_3$T, at 0.68 Å resolution using X-rays, the system was studied using neutron diffraction. This chapter will focus on two main aims:

1. To find the location (or distribution) of the proton between the hemi-protonated C$^+$·C base pairs

2. To find the role that H-bonded water can play in stabilizing the i-motif structure

Regarding to the first aim, two methods were applied to infer the localisation of the proton in the C$^+$·C base pair in the structure refinement step. The first was to add 50% occupancies of two D$^+$ protons from both cytosine residues in the base pair; following the symmetric hydrogen-bonding scheme with a single-well potential. Secondly, 100% D$^+$ occupancy from only one protonated cytosine was modelled in to mimic the asymmetric H-bonding scheme.

6.2 Materials and Methods

The d(C$_3$T) oligonucleotide was purchased from Eurogentec and purified by reverse phase HPLC. C$_3$T crystals for neutron diffraction were grown in H$_2$O using sitting drop vapour diffusion at 18°C. Table 6.2 lists the differences in C$_3$T growth conditions for neutron and the X-ray studies made in chapter 5. The DNA was mixed with crystallisation reagent in a 1:1 ratio and equilibrated against a reservoir of the same reagent. In contrast to the C$_3$T crystallisation reagent used for the X-ray study in chapter 4, the reagent for NC studies did not contain any LiCl as lithium has a negative magnitude of -1.90 femtometres.[207]

The crystals grew within 24 hours with dimensions of approximately 250 x 130 x 80 μm. Once the crystals grew to their individual maximum volume, the mother liquor was extracted and the crystals were soaked in the same crystallisation reagent made in D$_2$O. The mother liquor was exchanged with fresh DNA and D$_2$O reagent every other day until a crystal volume of approximately ~0.1 mm^3 was achieved (Fig. 6.3). Each crystal was drawn into quartz capillaries with 1.5 to 2.0 mm

_____

inner diameter attached to a 1.0 mL pipette. Approximately 30 uL of the mother liquor was left in one end of each capillary. The ends of the capillaries were sealed with beeswax.

A C_3T crystal of 0.096 mm³ volume was first used for time-of-flight (TOF) neutron diffraction data collection at 293 K. Data were recorded to 1.80 Å resolution using the MaNDi instrument[208,209] at the Spallation Neutron Source (SNS) in Oakridge National Laboratory (ORNL). The MaNDi beamline collects Laue diffraction data. In contrast to Bragg's condition where diffraction occurs when path length = nλ, the Laue criterion is that diffraction occurs when the scattering vector is the reciprocal vector. Each diffraction spot as a result of Laue exposure corresponds to a different diffracting plane and a different wavelength.

Table 6.3 lists the crystallographic data collection statistics. The ω angle was fixed at 90° for data collection. The crystal was held static for each image and was rotated by 20° on φ between images. A total of ten images were collected which were processed and integrated using the *Mantid* package.[210] *LAUENORM* from the *LAUEGEN* package[211] was then used for wavelength normalization of the Laue data and scaling between Laue diffraction images.

Table 6.2: Crystallisation conditions of C_3T crystals for X-ray and neutron diffraction studies. The crystals were grown using sitting drop vapour diffusion at 18 °C.

Diffraction type	X-ray	Neutron
Crystallisation Reagent		
Solvent	H_2O	H_2O
Salts	40 mM LiCl, 20 mM $MgCl_2$	20 mM $MgCl_2$
Buffer	40 mM sodium cacodylate (pH 5.5)	40 mM sodium cacodylate (pD 5.1)
Precipitant	30% MPD	30% MPD
Additive	20 mM $[Co(NH_3)_6]Cl_3$	20 mM $[Co(NH_3)_6]Cl_3$
Crystallisation Parameters		
Method	Sitting Drop Vapour Diffusion (SDVD)	Crystal growth using SDVD followed by feeding in DNA + D_2O buffer once every two days
Temperature (°C)	18	18
Crystal size (μm)	250 x 130 x 80	850 x 750 x 150
Growth period	24 hours	1 month

X-ray diffraction data were collected from the same crystal at 293 K using an in-house source at ORNL. The data were processed using the *XDS*[189] and *SCALA* from the *CCP4* suite.[108] The initial neutron data refinement values at the same resolution were 0.25/0.29. The *Phenix* suite[109] was used to refine the neutron and X-ray data both individually as well as simultaneously. Model building was done using the molecular graphics software *Coot*.[192]

_____
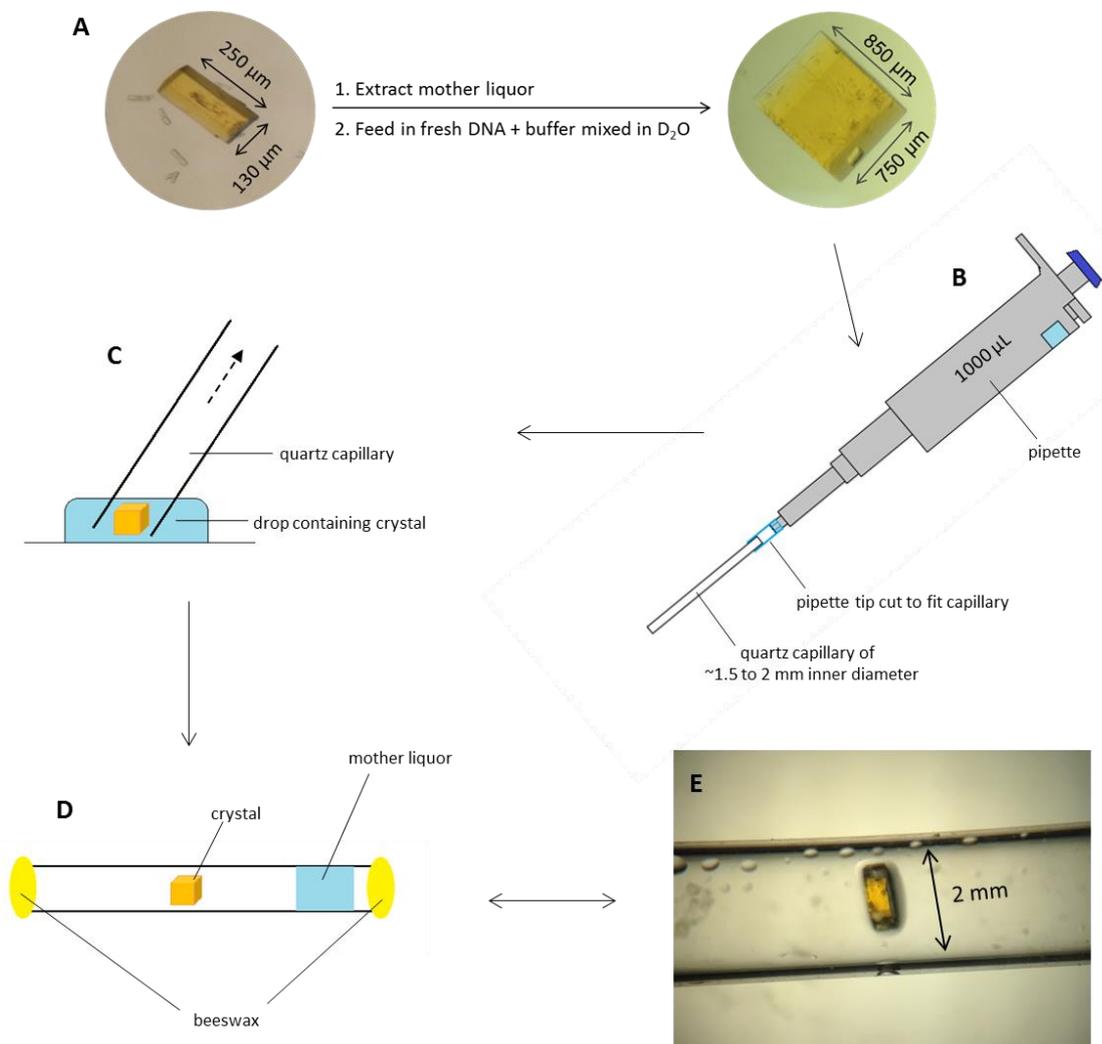
Figure 6.3: The process of preparing d(C$_3$T)$_4$ crystals for neutron data collection. (**A**) C$_3$T crystals are grown in H$_2$O then the mother liquor was exchanged with DNA and crystallisation reagent made in D$_2$O. (**B**) Pipette tips were customised to fit to quartz capillaries of 1.5-2 mm inner diameter. (**C**) Crystals are transferred into capillaries via pipette suction. (**D**) A neutron diffraction sample with mother liquor in one end of the capillary to prevent the crystal from drying out, sealed with wax on either ends. (**E**) View of a 0.096 mm^3 crystal inside a capillary with 2 mm inner diameter.

_____

Table 6.3: Data-collection statistics of C_3T.

Data Collection		
Diffraction type	X-ray	Neutron
Diffraction source	Rigaku MicroMax-007 HF, ORNL	MaNDi, Spallation Neutron Source, ORNL
Wavelength (Å)	1.54	2-4
Temperature (K)	293	293
Detector	R-AXIS IV^{++}	40 SNS Anger cameras
Crystal-to-detector distance (mm)	100	450
Rotation range per image (°)	1	0
No. of Images collected	150	10
Exposure time per image	30 seconds	24 hrs
Crystal Data		
Space Group	P 1 2_1 1	P 1 2_1 1
Unit-cell (Å)	a = 26.31 b = 50.12 c = 26.30	a = 26.31 b = 50.12 c = 26.30
	α = 90.0 β = 114.8 γ = 90.0	α = 90.0 β = 114.8 γ = 90.0
Data Processing *Outer shell statistics shown in parentheses		
Resolution (Å)	25.06-1.80 (1.90-1.80)*	13.64-1.79 (1.86-1.79)
R_{merge} (I)	0.083 (0.197)	0.087 (0.217)
R_{meas} (I)	0.101 (0.238)	0.105 (0.263)
R_{pim} (I)	0.056 (0.132)	0.057 (0.145)
No. of unique reflections	5809 (741)	4494 (438)
I/σI	57 (3.3)	13.8 (5.5)
Completeness (%)	93.6 (89.9)	76.98 (74.11)
Multiplicity	3.1 (3.2)	2.64 (2.42)

_____

6.3 Results and Discussion

6.3.1 The Distribution of Proton in the Hemi-Protonated C$^+$·C Base Pair

Since the C$_3$T crystal used for diffraction studies was grown in H$_2$O and the buffers only exchanged after crystal growth was observed; partial deuteration of the DNA was expected. Refinement of C$_3$T was carried out against three datasets; X-ray data alone, neutron data only and joint X-ray + neutron data. This section will focus on results obtained from neutron data only and joint X-ray + neutron refinements. The C$_3$T X-ray structure from chapter 5 was used as the starting model for both neutron and joint refinements. The X-ray crystal was not perdeuterated so the presence of exchangeable H/D sites was accounted for during refinement. *Phenix* automatically determines these sites and includes them for constrained occupancy refinement, ensuring that the sum of the occupancies for each H and D pair is equal to one.[113] Nonexchangeable sites remained modelled as H. Partial deuteration was observed in the structure with positive densities for D$^+$ seen in six out of the 12 C$^+$·C base pairs in the two C$_3$T i-motif molecules. The F$_o$-F$_c$ nuclear map shows that a proton lies in the centre of the N3-N3 base pair (Fig. 6.4).
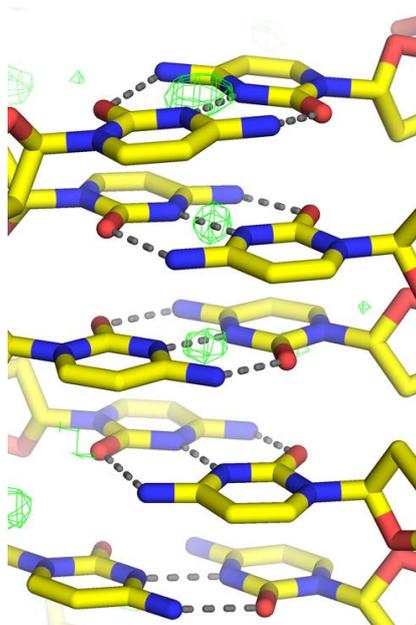
Figure 6.4: F$_o$-F$_c$ neutron scattering length density map (green) of C$_3$T contoured at 3.0σ locates D$^+$ protons between the two nitrogens in the N3 positions of Cytosine$^+$·Cytosine base pairs. The carbon, oxygen and nitrogen atoms have been coloured yellow, red and blue, respectively.

Two modelling methods were applied to infer the localisation of the proton in the C$^+$·C base pair:

1. *Asymmetric D-bonding*: 100% D$^+$ occupancy from only one protonated cytosine (Fig. 6.5A)

2. *Symmetric D-bonding*: 50% occupancies of two D$^+$ protons from each cytosine in the base pair (Fig 6.5B)

_____

Both approaches did not produce negative density in the location of the proton, initially suggesting that either type of D-bonding in the $C^+ \cdot C$ base pair could be correct. However, in contrast to the first approach, modelling in two protonated cytosines pushes the residues away from each other's plane by a 25.9° angle of deviation (Fig. 6.6B) with an N3-N3 distance of 2.76Å. The atoms used to measure the angle are labelled in Fig. 6.5B; where the $\angle_{deviation}$ = 180° - $\angle$(N3-N3-C6). $C^+ \cdot C$ base pairs are planar with a maximal deviation of 18° and an N3–N3 distance of 2.6– 2.8Å.[36] The angle of deviation is 14.1° when only one cytosine contributes to 100% D^+ occupancy (Fig. 6.6A). The N3-N3 distance is 2.62Å when this approach is used. The first model best portrays the high-resolution C_3T structure, where the $C^+ \cdot C$ base pairs are planar. Therefore, the first modelling method, i.e. the asymmetric deuterium bonding, is correct.
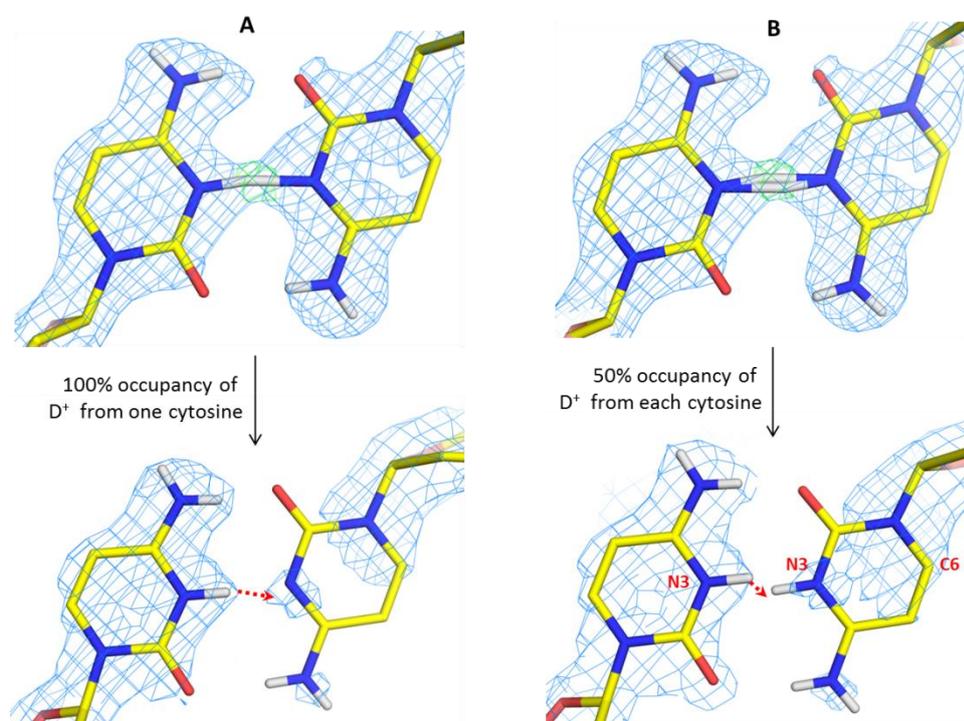
Figure 6.5: The two strategies applied for finding out the distribution of D^+ in the hemi-protonated $C^+ \cdot C$ base: (**A**) 100% of D^+ occupancy from only one cytosine. (**B**) 50% D^+ occupancies from both cytosines. $2F_o$-F_c neutron scattering length density maps are drawn in blue, contoured at 1.0σ. The green F_o-F_c map is contoured at 3.0σ. Carbon, oxygen, nitrogen and deuterium atoms are yellow, red, blue and white, respectively. Red arrows indicate the shift in cytosine planes.

_____
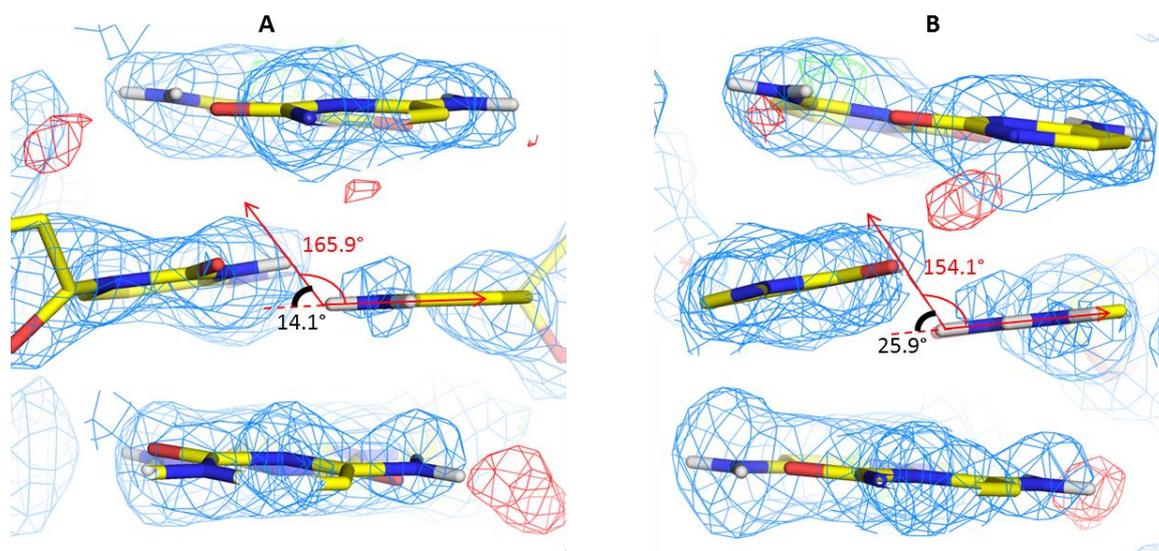
Figure 6.6: Results of modelling two protonated cytosine residues in a hemi-protonated C$^+$·C base pair. One cytosine lies (**A**) 165.9° ($\angle_{deviation}$ = 14.1°) away from the plane of another cytosine when asymmetric modelling is applied and (**B**) 154.1° ($\angle_{deviation}$ = 25.9°) away when symmetric modelling is applied. 2F$_o$-F$_c$ neutron scattering length density maps are drawn in blue, contoured at 1.0σ. The positive F$_o$-F$_c$ map is contoured at 3.0σ in green and the negative F$_o$-F$_c$ map is contoured at 3.0σ in red. (N3-N3-C6) angles measured are directed by red arrows and angles of deviation are in black.

The asymmetric model and observation of deuterium atoms in the i-motif C$^+$·C base pairs for the first time has allowed new bond length parameters to be recorded; the N3-D and O2-D bond distances (Fig. 6.7A). The N3-D bond length, in particular, can be important in finding other possible atoms to substitute the H$^+$ proton which is essential for i-motif formation. I-motif folding has been studied for nanotechnology applications but only under acidic conditions. The newly measured bond distance may provide insight into possible use of other cations to help in the formation of an i-motif at physiological pH.

The O4-D bond length in the symmetry-related T·T wobble base pair of C$_3$T has also now been determined, including the distances between the atoms in both base pairs and the nearest water molecules. Figure 6.7B illustrates the tight network of the two thymine bases with two water molecules to form an almost rectangular grid in the same plane. This adds to the conclusion of chapter 5 that the presence of a base pair on either side of the cytosine core is important in forming a stable i-motif structure. This also correlates to the solution studies in chapter 4 where the most stable intramolecular i-motif forming sequences were found to be rich in thymine; d(C$_3$ATT)$_3$C$_3$, d(C$_3$TTT)$_3$C$_3$ and d(C$_3$TCT)$_3$C$_3$.
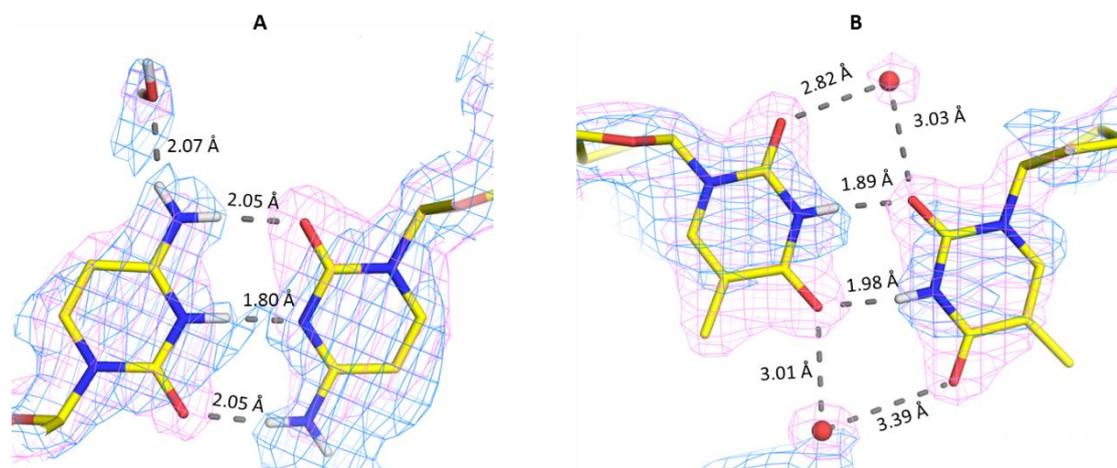
Figure 6.7: New bond length parameters in C_3T revealed by neutron diffraction. (**A**) The N3-D and O4-D bond distances in a hemi-protonated cytosine[+]·cytosine base pair. (**B**) The O4-D bond length in the thymine-thymine base pair. $2F_o$-F_c neutron and electron scattering length density maps are drawn in blue and pink, respectively; both contoured at 1.0σ.

6.3.2 The Role of Hydorgen-Bonded Water in Stabilising the i-motif Structure

The hydrogen atoms of water molecules appear differently in X-ray and neutron maps and at different resolutions, with only subatomic resolution X-ray maps showing H-atom positions for very well ordered water molecules.[113] D_2O molecules can appear as various shapes in nuclear maps. The spherical-shaped waters (Fig. 6.8A) appear when the D atoms are smeared in space owing to random rotation around the O-atom position. Figure 6.8B shows a fixed water molecule having a rotational degree of freedom around a D-O bond axis and Fig. 6.8C shows a density suggesting fully resolved D_2O.

The joint X-ray + neutron refinement of the C_3T structure supports the X-ray structure in that the position of the water molecules aid in the stability of the i-motif. H_2O/D_2O molecules can be seen to bridge the two i-motif molecules in Figure 6.9 by hydrogen-bonding to the C[+]·C base pairs. Two $[Mg(H_2O)_6]^{2+}$ ligands were located in the joint X-ray + neutron structure but the positions of H/D in the waters surrounding magnesium could not be determined.
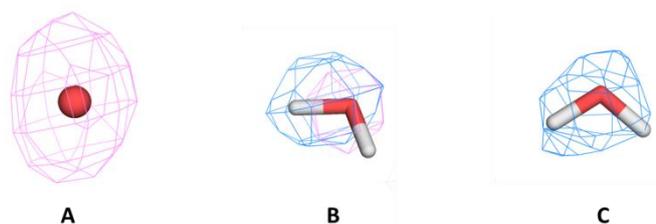
Figure 6.8: Water molecules of (**A**) spherical, (**B**) cylindrical and (**C**) boomerang shaped densities. $2F_o$-F_c X-ray (pink) and nuclear (blue) maps have been contoured at 1.0σ.

_____
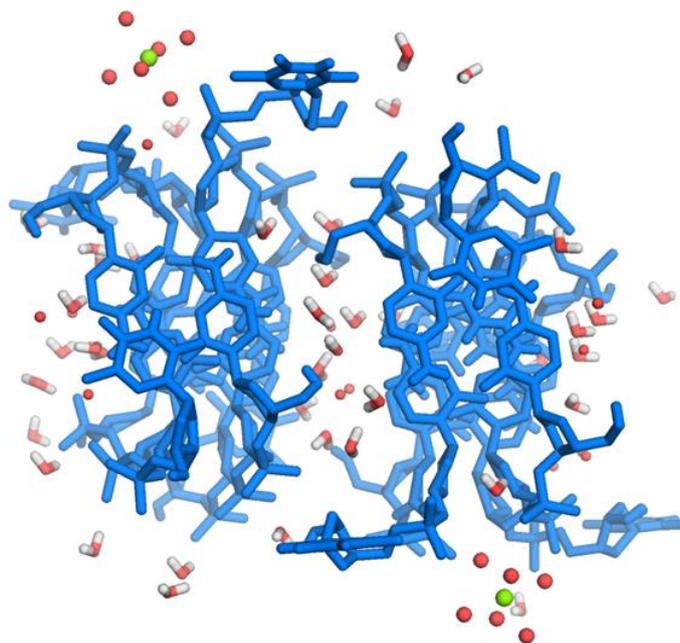
Figure 6.9: Asymmetric unit of the d(C$_3$T)$_4$ structure. The two i-motif molecules are represented by blue sticks, the oxygen and H/D by red and white sticks, respectively. Water molecules for which the H/D atoms could not be observed are denoted by red spheres. Magnesium ions from the [Mg(H$_2$O)$_6$]$^{2+}$ ligands are represented by green spheres.

6.4 Conclusions

Refinement of neutron data using the two methods in Figure 6.1 revealed that both routes position D$^+$ deuterons without resulting in negative density; i.e. both methods are plausible. The difference in R$_{work}$/R$_{free}$ between the two strategies is quite insignificant with % difference of 0.26/1.71. However, C$^+$·C base pairs are planar with a maximal deviation of 18° and an N3–N3 distance of 2.6– 2.8Å.[36] Although both approaches give N3-N3 distances in this region, the symmetric D-bonding method gives a deviation of 25.9° whereas the asymmetric route shows the two cytosine residues 14.1° apart from each other. It can be argued that the cytidines in the symmetric method may not be in plane with one another due to restraints in the refinement programme. *Phenix* developers have been contacted for help regarding the issue and to consequently confirm that the observations made using the symmetric approach are not a result of refinement restraints. However, comparison of the two methods to the high-resolution C$_3$T X-ray data from chapter 5 indicates that the symmetric D-bonding method is incorrect. Moreover, Lieblein *et al.* reported that for N3-N3 distances smaller than 2.5Å, the H$^+$ proton is shared between the two cytidines and is located precisely in the midpoint between the N3 atoms.[55] Both neutron and joint X-ray + neutron refinements gave an average N3-N3 distance of 2.62Å, implying that the i-motif has asymmetric H-bonding with a double-well potential.

_____

The refinements against joint X-ray + neutron data, and neutron data on its own, gave similar results, which was expected because the *Phenix* software does a Z-H normalization (where Z is any atom), by extending the Z-H bond towards the neutron expected length, therefore, weighing the refinement of H/D towards the neutron data.[109] Nuclear and electron densities (on their own) show slight differences in atomic positions including $C^+ \cdot C$ base pair distances. This is because the neutron data shows where the nuclei are. So, unlike X-ray data, it shows the inter-nuclear distances; resulting in bonds that are 0.1-0.2Å longer. Superimposition of the 0.68Å X-ray structure onto the structure derived from neutron data alone illustrates that the two i-motif structures are similar. The C_3T crystal used for neutron diffraction was partially deuterated and Figure 6.10 shows that addition of deuterium doesn't significantly change the i-motif structure.
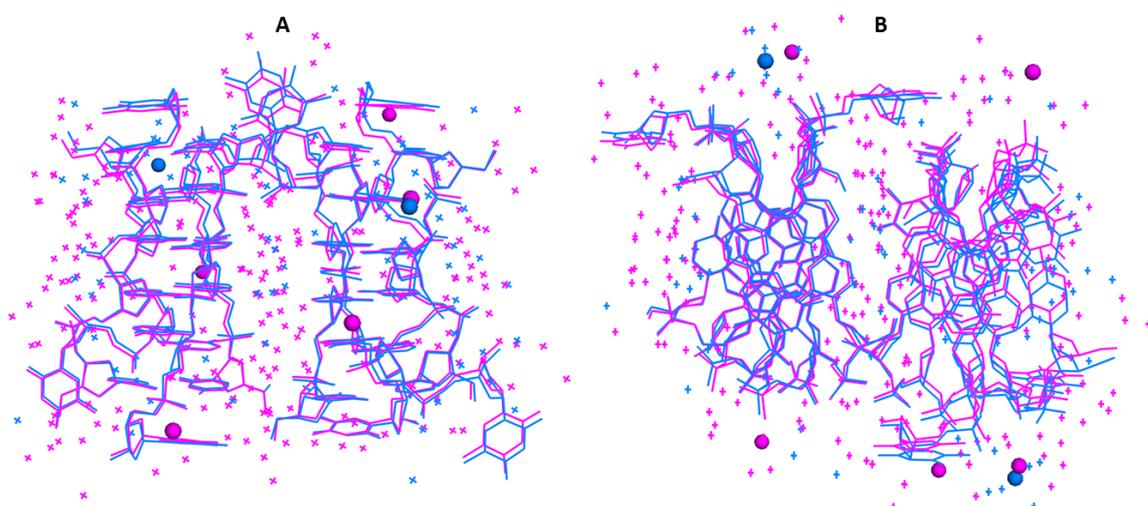
Figure 6.10: Superimposed X-ray (pink) and neutron (blue) crystal structures of C_3T.
The water molecules are represented by crosses.

Table 6.4 compares the current *R* values from the joint refinement with refinements using X-ray and neutron data alone. The joint refinement significantly improved the neutron R_{work}/R_{free} values by 5.39/7.58% from refinement carried out against neutron data alone. The X-ray *R* factors, on the other hand, are similar.

Table 6.4: R_{work}/R_{free} (%) values of individual and joint X-ray + Neutron refinements.

X-ray only	Neutron only	Joint X-ray + Neutron	
		X-ray	*Neutron*
18.04/22.45	31.79/38.37	18.10/22.55	26.40/30.79

The neutron data obtained is complete to 77% in comparison to the 76% average data completeness for all neutron structures currently in the PDB. Partial deuteration of C_3T also contributed to challenges in refinement. Cancellation effects by the negative neutron scattering length of hydrogen could have significantly limited the interpretability of the density map around bonds involving H/D atoms. For example, not all hydrogen positions in water molecules could be

_____

located; with none located in $[Mg(H_2O)_6]^{2+}$ ligands. Positive densities for D^+ were also seen for only six out of the 12 $C^+ \cdot C$ base pairs in the two C_3T i-motif molecules.

In conclusion, neutron crystallography is able to reveal H/D atom positions from macromolecular diffraction data. Collection of neutron data allowed new bond distances; N3-D and O2-D, in the hemi-protonated $C^+ \cdot C$ base pair to be measured for the first time. I-motif folding in DNA nanotechnology has only been studied under acidic conditions. The results obtained can thus be useful in finding other possible cations to replace the proton in N3···H$^+$···N3. The interest in reversible formation of the i-motif structure as a switchable scaffold in DNA nanotechnology has been expressed by the use of various cations including gold nanoparticles and Cu^{2+}.[47,48,179] Ag^+, in particular, has been reported to fold an i-motif forming sequence at physiological pH through possible N3–Ag^+–N3 base pairing.[66,212]

An i-motif forming sequence can potentially have three structural outputs; single strand, hairpin and i-motif.[179] The newly measured bond distance can, therefore, provide insight into possible use of other cations to help achieve the three structures using the same oligonucleotide. The newly determined bond length parameters of the T-T base pair further supports the notion that the cytosine core must have base pairs on either of its side in order to form a stable i-motif structure.

_____

Other i-motif Crystallisation Attempts

7.1 Introduction

Following the crystal structure determination of two new tetramolecular i-motif systems, $d(C_3A)_4$ and $[d(C_3A) + d(C_3T)]$, crystallisation trials of the highly diffracting tetramolecular $d(C_3T)_4$ i-motif in the presence of silver and Tilorone ligands were carried out. The two ligands have been reported to bind to intramolecular i-motifs in solution.[66,96] Crystallisation attempts of native bimolecular and unimolecular/intramolecular i-motifs were also conducted. The bimolecular i-motif sequence $d(C_3TA_2C_3)_2$ was designed after the unimolecular human telomeric i-motif (HTi) sequence $d(C_3TAA)_3C_3$. An extra 5'-TATA-3' overhang was added to the same sequence to initiate some form of base pairing. The oligonucleotides used for unimolecular i-motif crystallisation include the HTi and the HIF-1αS-CG sequences. Overexpression of the HIF-1α gene is associated with oral squamous cell carcinoma and its promoter sequence i-motif structure in solution shows stability near neutral pH.[37] The rest of the oligonucleotides were selected from solution studies in Chapters 2, 3 and 4 based on their stabilities.

Table 7.1: List of DNA sequences set for crystal growth.

I-motif Structure	System	DNA Sequence 5'→3'
Tetramolecular	C_3T	CCCT
Bimolecular	$C_3TA_2C_3$	CCCTAACCC
	TATA	CCCTAACCCTATA
Intramolecular	HTi	CCCTAACCCTAACCCTAACCC
	HIF-1αS-CG	TCCCGCCCCCTCTCCCCTCCCC
	TCG	CCCTCGCCCTCGCCCTCGCCC
	TCT	CCCTCTCCCTCTCCCTCTCCC
	AGA	CCCAGACCCAGACCCAGACCC
	C_3T_4	CCCTTTTCCCTTTTCCCTTTTCCC
	C_3T_5	CCCTTTTTCCCTTTTTCCCTTTTTCCC
	C_3T_6	CCCTTTTTTCCCTTTTTTCCCTTTTTTCCC
	C_3T_7	CCCTTTTTTTCCCTTTTTTTCCCTTTTTTTCCC
	C_3T_8	CCCTTTTTTTTCCCTTTTTTTTCCCTTTTTTTTCCC
	C_3T_{838}	CCCTTTTTTTTCCCTTTCCCTTTTTTTTCCC

7.2 Materials and Methods

Oligonuclotides were purchased from Eurogentec and purified by reverse phase HPLC. Sitting drop vapour diffusion was the only technique used because crystallisation trials for tetramolecular i-motifs were negative with the hanging drop method. Initial screening was performed using the readily available 192 crystallisation conditions from the NATRIX HT Screen by Hampton Research and HELIX by Molecular Dimensions at 20°C and 4°C. Each drop contained 2 μl of single-stranded oligonucleotide with/without ligand or salt (Table 7.2) and 2μl of the crystallisation solution. The drop was equilibrated against 500 μl of the same crystallisation condition. Tables 7.2 to 7.5 list the systems used for crystallisation attempts. All bimolecular i-motif sequences were annealed in 20 mM sodium cacodylate buffer at pH 5 by heating the solutions to 90°C and letting them cool down to room temperature. The unimolecular i-motifs were prepared in various pH strengths of 20 mM sodium cacodylate buffer as well as 20 mM sodium acetate buffer as per the human telomeric G-quadruplex crystallisation condition.[33] Unimolecular oligonucleotides were both annealed and non-annealed before setting up for vapour diffusion. Systems which gave crystals/spheroids of unimolecular i-motifs are listed in Table 7.5.

Table 7.2: List of tetramolecular i-motif systems set for crystal growth. DNA concentrations are for single-stranded oligonucleotides. Crystals which diffracted are highlighted in green.

System in Sitting Drop		Screen in Reservoir	Temperature (°C)	Crystal Growth	Condition No.	Diffraction Observed
DNA	Ligand					
4 mM C_3T	4 mM $AgNo_3$	HELIX	20	Yes	2-35	No
4 mM C_3T	2 mM $AgNo_3$	HELIX	20	Yes	2-35	No
4 mM C_3T	4 mM $AgNo_3$	NATRIX	20	No	-	-
4 mM C_3T	2 mM $AgNo_3$	NATRIX	20	Yes	A1	Yes
2 mM C_3T	2 mM Tilorone	HELIX	20	No	-	-
2 mM C_3T	1 mM Tilorone	HELIX	20	No	-	-
2 mM C_3T	2 mM Tilorone	NATRIX	20	Yes	A12	Yes
2 mM C_3T	1 mM Tilorone	NATRIX	20	No	-	-

Table 7.3: List of bimolecular i-motif systems set for crystal growth. All DNA solutions were prepared in 20mM sodium cacodylate buffer at pH 5 and annealed prior to crystallisation set-up. DNA concentrations are for single-stranded oligonucleotides. Crystals which diffracted are highlighted in green.

System in Sitting Drop (DNA in buffer only)	Screen in Reservoir	Temperature (°C)	Crystal Growth	Condition No.	Diffraction Observed
1 mM (ss) $C_3TA_2C_3$	NATRIX	20	No	-	-
1 mM (ss) $C_3TA_2C_3$	HELIX	20	No	-	-
1 mM (ss) $C_3TA_2C_3$	NATRIX	4	No	-	-
1 mM (ss) TATA	NATRIX	20	Yes	A12	Yes
1 mM (ss) TATA	HELIX	20	Yes	2-2	No

_____

Table 7.4: List of unimolecular i-motif systems set for crystal growth. All DNA solutions were prepared in either 20mM sodium acetate or 20mM sodium cacodylate buffer, unless unstated. All DNA solutions were also annealed prior to setting up unless stated. DNA concentrations are for single-stranded oligonucleotides. Systems where crystal growth was observed are highlighted in yellow.

DNA System in Sitting Drop				Screen	Temperature
DNA	*Buffer (20mM)*	*Ligand*	*Salt (20mM)*		**(°C)**
0.5 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	20
0.5 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	20
1 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	20
1 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	20
1.5 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	20
1.5 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	20
2 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	20
2 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	20
0.5 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	4
0.5 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	4
1 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	4
1 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	4
1.5 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	4
1.5 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	4
2 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	4
2 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	4
1 mM HTi	Na cacodylate, pH 5	1 mM AgNo$_3$	NaCl	HELIX	20
1 mM HTi	Na cacodylate, pH 5	1 mM AgNo$_3$	NaCl	NATRIX	20
1 mM HTi (NOT ANNEALED)	Na cacodylate, pH 5	-	-	HELIX	20
1 mM HTi (NOT ANNEALED)	Na cacodylate, pH 5	-	-	NATRIX	20
1.5mM HIF-1αS-CG	Na acetate, pH 4.5	-	NaCl	HELIX	20
1.5mM HIF-1αS-CG	Na acetate, pH 4.5	-	NaCl	NATRIX	20
0.75 mM TCG	Na cacodylate, pH 5	-	-	HELIX	20
0.75 mM TCG	Na cacodylate, pH 5	-	-	NATRIX	20
0.75 mM TCT	Na cacodylate, pH 5	-	-	HELIX	20
0.75 mM TCT	Na cacodylate, pH 5	-	-	NATRIX	20
0.75 mM AGA	Na cacodylate, pH 5	-	-	HELIX	20
0.75 mM AGA	Na cacodylate, pH 5	-	-	NATRIX	20
1 mM C$_3$T$_7$ (NOT ANNEALED)	-	-	-	NATRIX	20
1 mM C$_3$T$_7$ (NOT ANNEALED)	-	-	-	HELIX	20
1 mM C$_3$T$_8$ (NOT ANNEALED)	-	-	-	NATRIX	20
1 mM C$_3$T$_8$ (NOT ANNEALED)	-	-	-	HELIX	20
1 mM C$_3$T$_7$	Na acetate, pH 5	-	NaCl	HELIX	20
1 mM C$_3$T$_7$	Na acetate, pH 5	-	NaCl	NATRIX	20
1 mM C$_3$T$_8$	Na acetate, pH 5	-	NaCl	HELIX	20

_____

Table 7.4 continued.

DNA System in Sitting Drop				Screen	Temperature (°C)
DNA	*Buffer (20mM)*	*Ligand*	*Salt (20mM)*		
1 mM C_3T_8	Na acetate, pH 5	-	NaCl	NATRIX	20
1 mM C_3T_7	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	HELIX	20
1 mM C_3T_7	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	NATRIX	20
1 mM C_3T_7 (NOT ANNEALED)	Na acetate, pH 4.5	-	-	HELIX	20
1 mM C_3T_7 (NOT ANNEALED)	Na acetate, pH 4.5	-	-	NATRIX	20
0.5 mM C_3T_8	Na cacodylate, pH 5	0.5 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	NATRIX	20
0.5 mM C_3T_8	Na cacodylate, pH 5	0.5 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	HELIX	20
0.5 mM C_3T_8	Na cacodylate, pH 5	-	-	NATRIX	20
0.5 mM C_3T_8	Na cacodylate, pH 5	-	-	HELIX	20
1 mM C_3T_{838}	Na cacodylate, pH 5	-	-	NATRIX	20
1 mM C_3T_{838}	Na cacodylate, pH 5	-	-	HELIX	20
1 mM C_3T_6	Na cacodylate, pH 5	-	-	NATRIX	20
1 mM C_3T_6	Na cacodylate, pH 5	-	-	HELIX	20
1 mM C_3T_5	Na cacodylate, pH 5	-	-	NATRIX	20
1 mM C_3T_5	Na cacodylate, pH 5	-	-	HELIX	20
1 mM C_3T_4	Na cacodylate, pH 5	-	-	NATRIX	20
1 mM C_3T_4	Na cacodylate, pH 5	-	-	HELIX	20
1 mM C_3T_{838}	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	NATRIX	20
1 mM C_3T_{838}	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	HELIX	20
1 mM C_3T_6	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	NATRIX	20
1 mM C_3T_6	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	HELIX	20
1 mM C_3T_5	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	NATRIX	20
1 mM C_3T_5	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	HELIX	20
1 mM C_3T_4	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	NATRIX	20
1 mM C_3T_4	Na cacodylate, pH 5	1 mM *rac*-[Ru(phen)$_2$dppz]$^{2+}$	-	HELIX	20

_____

Table 7.5: List of unimolecular i-motif systems where crystal growth was observed. All DNA solutions were prepared in either 20mM sodium acetate or 20mM sodium cacodylate buffer. All DNA solutions were also annealed prior to setting up at 20°C. DNA concentrations are for single-stranded oligonucleotides. Systems which diffracted are highlighted in red. Ru is *rac*-[Ru(phen)$_2$dppz]$^{2+}$.

DNA System in Sitting Drop				Screen	Crystallisation Condition No.
DNA	*Buffer (20mM)*	*Ligand*	*Salt (20mM)*		
1 mM HTi	Na acetate, pH 4.5	-	NaCl	HELIX	1-8, 1-48
1 mM HTi	Na acetate, pH 4.5	-	NaCl	NATRIX	E3, F6, F7, F8, **F9**, F10
1.5mM HIF-1αS-CG	Na acetate, pH 4.5	-	NaCl	HELIX	1-48
1.5mM HIF-1αS-CG	Na acetate, pH 4.5	-	NaCl	NATRIX	F1
0.75 mM TCG	Na cacodylate, pH 5	-	-	NATRIX	A9
0.75 mM TCT	Na cacodylate, pH 5	-	-	NATRIX	**G8**
1 mM C$_3$T$_4$	Na cacodylate, pH 5	1 mM Ru	-	HELIX	2-23

7.3 Results & Discussion

7.3.1 Tetramolecular i-motifs with ligands

Crystals of 4 mM C$_3$T (ss) in 2 mM AgNo$_3$ mixture were obtained in Natrix A1 in 4 days. The crystallisation conditions include 0.01 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate (pH 5.6) and 1.8 M Lithium sulphate monohydrate. However, diffraction results gave the structure of the native DNA only, with no presence of silver ions.

Crystals of 2 mM C$_3$T (ss) with 2 mM Tilorone were obtained in Natrix A12 within two weeks of set-up. The reagent conditions include 0.01 M Magnesium sulfate heptahydrate, 0.05 M Sodium cacodylate trihydrate (pH 6.0) and 1.8 M Lithium sulfate monohydrate. Diffraction data were collected at 1.5Å resolution. The i-motif structure could be fitted in but the modelling of Tilorone on either ends of the DNA structure, as suggested by the difference map, gave high *R*-values, indicating that the overall model was over-fitted thus unreliable. A better ordered crystal that can diffract at higher resolution needs to be collected.

7.3.2 Bimolecular i-motifs

Two sequences; d(C$_3$TA$_2$C$_3$)$_2$ and d(C$_3$TA$_2$C$_3$TATA)$_2$, based on the human telomeric i-motif, were initially checked for thermal stability. The TATA segment was added to the latter sequence to help initiate base pairing. DNA melts under UV showed that d(C$_3$TA$_2$C$_3$TATA)$_2$ is 11°C more stable than the d(C$_3$TA$_2$C$_3$)$_2$ sequence (Fig. 7.1). Crystal growth was observed after three weeks of set-up but for only the d(C$_3$TA$_2$C$_3$TATA)$_2$ sequence (Fig. 7.2). Crystals of the longer oligonucleotide were collected from HELIX 2-2 and Natrix A12. The HELIX 2-2 conditions include 10 % v/v MPD, 0.005 M Spermine and 0.05 M HEPES (pH 6.5) whereas the Natrix A12 condition is made of 0.01 M Magnesium sulphate, 0.05 M Sodium cacodylate trihydrate (pH 6.0) and 1.8 M Lithium sulphate. Diffraction could only be observed for the crystal grown in Natrix A12 and data were collected to 2.1Å resolution.

_____
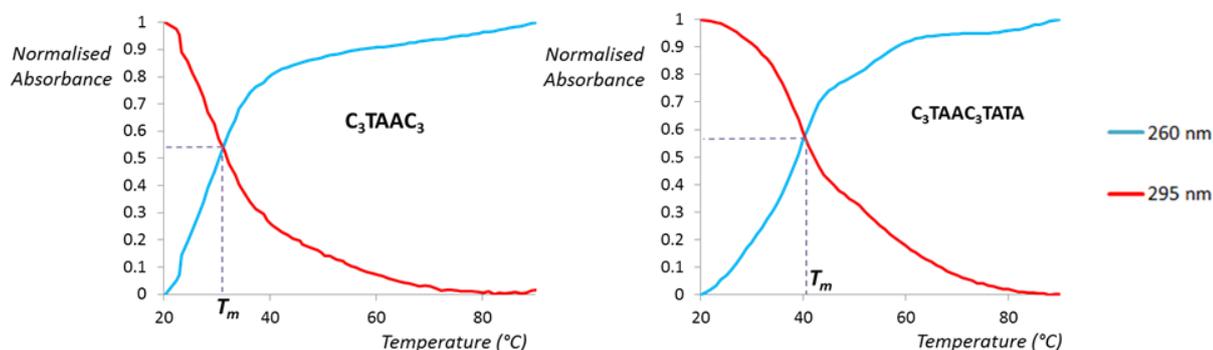
Figure 7.1: DNA melting profiles of d(C$_3$TA$_2$C$_3$)$_2$ and d(C$_3$TA$_2$C$_3$TATA)$_2$ in 50 mM sodium cacodylate buffer (pH 5).
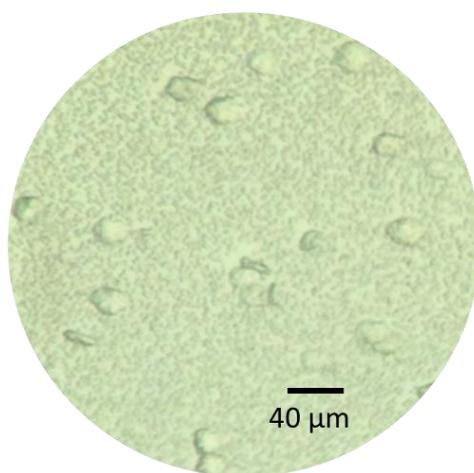
Figure 7.2: Crystals of d(C$_3$TA$_2$C$_3$TATA)$_2$.

Following data processing, the i-motif core was modelled in the resulting electron density map but the loop regions could not be recognised. Therefore, three more oligonucleotides based on the d(C$_3$TA$_2$C$_3$TATA)$_2$ sequence were purchased and set for crystal growth (Tables 7.6 and 7.7). Cytosine residues at positions C1, C3 and C7 were brominated to identify if the particular nucleobases form the i-motif core or the loop. Similarly, the thymine residue at position T4 was substituted with uracil (U) to check if the base lies within the loop or the intercalated centre. Crystals of TATA-1C3 were obtained in Natrix condition A12, however, they did not diffract. Optimisation of the crystallisation condition is to be carried out in order to grow more ordered crystals.

Table 7.6: List of bimolecular DNA sequences set for crystal growth.

I-motif Structure	System	DNA Sequence 5'→3'
Bimolecular	TATA-7CC	CCCTAA(5-Br-C)CCTATA
	TATA-1C3	(5-Br-C)C(5-Br-C)TAACCCTATA
	TATA-4AA	CCC(5-Br-U)AACCCTATA

_____

Table 7.7: List of modified d(C$_3$TA$_2$C$_3$TATA)$_2$ bimolecular i-motif sequences set for crystal growth. All DNA solutions were prepared in 20mM sodium cacodylate buffer at pH 5 and annealed prior to crystallisation set-up. DNA concentrations are for single-stranded oligonucleotides. Systems where crystal growth was observed are highlighted in green.

System in Sitting Drop	Screen in Reservoir	Temperature	Crystal Growth
1 mM (ss) TATA-7CC	NATRIX	20°C	No
1 mM (ss) TATA-7CC	HELIX	20°C	No
1 mM (ss) TATA-1C3	NATRIX	20°C	Yes but no diffraction observed
1 mM (ss) TATA-1C3	HELIX	20°C	No
1 mM (ss) TATA-4AA	NATRIX	20°C	No
1 mM (ss) TATA-4AA	HELIX	20°C	No

7.3.3 Unimolecular i-motifs

High-throughput srCD spectroscopy at beamline B23 was used to test which readily available crystallisation conditions in the Natrix HT screen would show i-motif formation. Instead of crystals, spheroids with irregular sides were observed after six months of set-up. Out of the 96 conditions, spheroids of the HTi oligonucleotide were obtained in Natrix conditions E3 and F6 to F10. SrCD spectroscopy also showed formation of the human telomeric i-motif in the given crystallisation conditions (Fig. 7.3). The information obtained from CD spectroscopy can thus be used to predict conditions for intramolecular i-motif crystallisation.

Diffraction data could only be collected from the crystal grown in Natrix F9 and to 15.1Å resolution at best. Natrix F9 consists of 0.012M Sodium chloride, 0.08M Potassium chloride, 0.04M Sodium cacodylate trihydrate (pH 6.0), 50% v/v(+/-)-2-Methyl-2,4-pentanediol and 0.012M Spermine tetrahydrochloride.

Additionally, crystallisation trials were also made on three of the intramolecular i-motif forming C$_3$XYZ sequences; AGA, TCG and TCT. These sequences were randomly selected from a range of different stabilities, and as expected from solution studies which showed that AGA is the least stable of the three, crystal growth was observed for TCG and TCT (Fig. 7.4). Diffraction, however, was only observed for TCT to 20.0Å resolution. The TCT i-motif was grown in Natrix G8 which consists of 0.08M Sodium chloride, 0.04M Sodium cacodylate trihydrate (pH 7.0), 30% v/v (+/-)-2-Methyl-2,4-pentanediol and 0.012M Spermine tetrahydrochloride.

_____
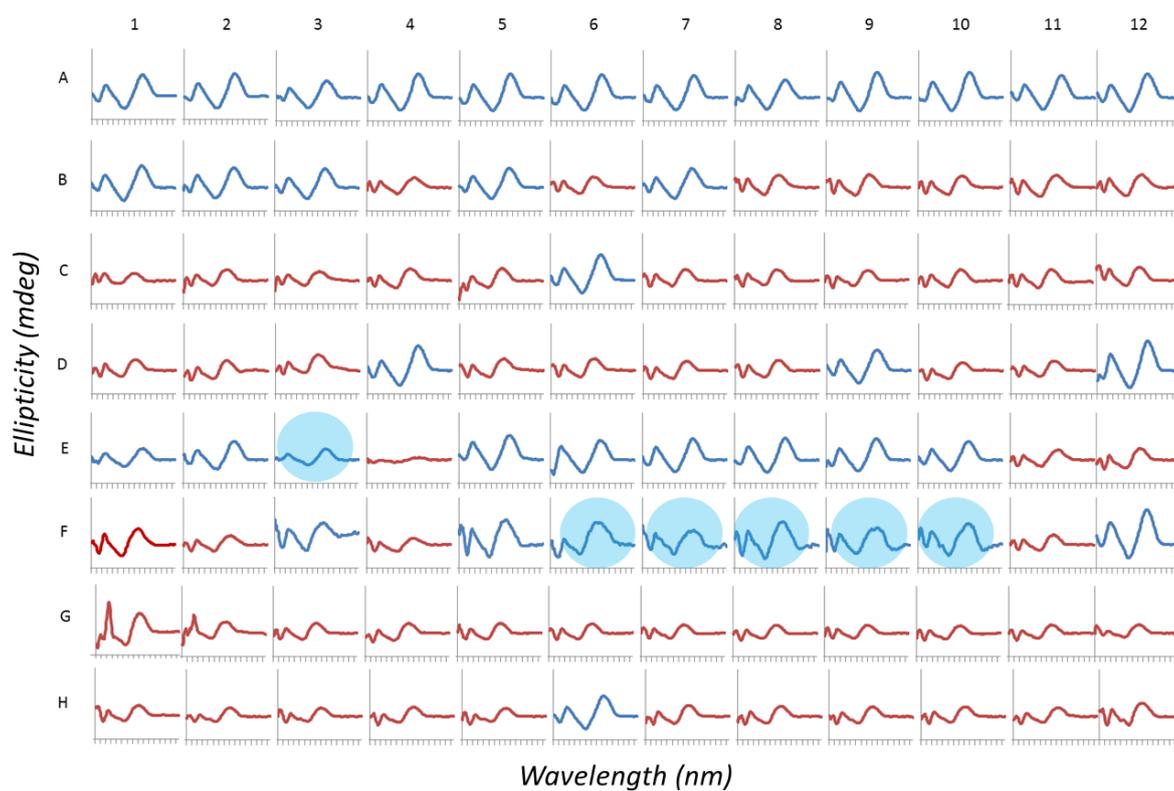
Figure 7.3: Molar ellipticity *vs.* wavelength graphs of the human telomeric i-motif; d(C$_3$TAA)$_3$C$_3$ in the Natrix HT screen. I-motif formation is illustrated by blue spectra and the conditions in which crystal growth were observed are highlighted with blue circles.
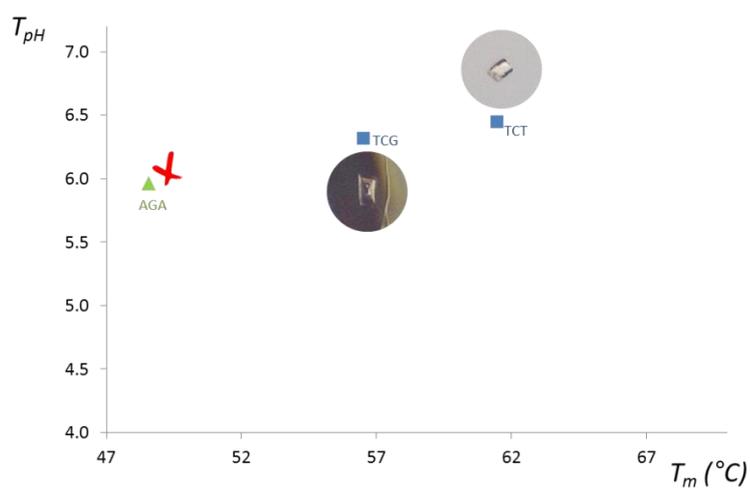
Figure 7.4: Graph showing the melting temperatures and transitional pH values of three C$_3$XYZ sequences. Crystal growth was observed for d(C$_3$TCG)$_3$C$_3$ and d(C$_3$TCT)$_3$C$_3$ but not for d(C$_3$AGA)$_3$C$_3$.

_____

7.4 Conclusions

Although crystals of C_3T were obtained when set up with $AgNo_3$ in a 2:1 DNA to ligand system, diffraction data showed no presence of silver cations. Crystals of 2mM C_3T (ss) with 2mM Tilorone grown at pH 6, on the other hand, gave diffracted to 1.5Å resolution. The i-motif structure could be fitted in but the modelling of Tilorone on either ends of the DNA structure, as suggested by the difference map, gave high R-values indicating that the overall model was over-fitted thus unreliable. The refinement results suggest that data at higher resolution need to be collected. Hence, more ordered crystals that can diffract to resolutions higher than 1.5Å need to be collected.

Crystalliation set up of two possible bimolecular i-motif sequences; $d(C_3TA_2C_3)_2$ and $d(C_3TA_2C_3TATA)_2$, gave crystals of only the latter sequence. The sequence is also more stable than $d(C_3TA_2C_3)_2$ in solution. The i-motif core was modelled in the resulting electron density map but the loop regions could not be recognised. Two oligonucleotides based on the $d(C_3TA_2C_3TATA)_2$ sequence with brominated cytosine residues and one uracil residue were purchased and set for crystal growth, however, diffraction data could not be obtained. The Natrix A12 condition from which the $d(C_3TA_2C_3TATA)_2$ crystals were used for X-ray diffraction contains no cryoprotectant and data were collected at the standard temperature of 100 Kelvin. This may have therefore affected the resolution of the data collected. The crystallisation condition needs to be optimised with cryoprotectant for future data collections.

High-throughput srCD spectroscopy was used to check which Natrix conditions gave intramolecular i-motif formation in solution. The conditions in which crystal growth of the HTi sequence were observed matched the CD spectroscopy results suggesting that the two techniques can be quite complementary. Spheroids with irregular sides, rather than crystals, were obtained after six months of set-up and diffraction data could only be collected to 15.1Å resolution.

Crystallisation trials made on three of the intramolecular i-motif forming C_3XYZ sequences; AGA, TCG and TCT gave crystals of TCG and TCT. These sequences were randomly selected from a range of different stabilities, and solution studies from Chapter 4 showed that AGA is the least stable of the three. Therefore, DNA melts and pH titrations can also be used to determine or design stable i-motif sequences for crystal growth trials. However, diffraction spots were only observed for TCT to an extremely low resolution of 20.0Å. The intramolecular i-motif crystallisation conditions need to be optimised further to obtain more ordered and better diffracting crystals to locate the atomic positions accurately.

_____

Summary and Future Work

Increase in number of cytosine residues in C-tracts give high i-motif stabilities.[38] It has recently been reported that sequences with at least six cytosine tracts can form intramolecular i-motifs.[213] This thesis has shown that loop length and base composition are also important factors that affect i-motif stability. Forty-seven C-rich sequences were tested, with all but ten sequences having the ability to form i-motif structures. CD spectra of the ten sequences resembled DNA hairpin and anti-parallel G-quadruplex formation, emphasising the importance of loop base composition in i-motif formation. C-rich sequences with high guanine and thymine content have been reported to form more stable structures[40] but it is now known that loops that are rich in pyrimidines form thermally more stable i-motifs whereas those containing guanine may fold to form other higher-order DNA structures instead. Moreover, the direction of loop base alignment was also shown to affect i-motif formation.

Change in sequence directionality has been shown to affect i-motif stability[178] but its effect on structural conformation had not been known. As i-motifs have always been a structure of interest in DNA nanotechnology, finding a sequence that can change between an i-motif and a G-quadruplex structure under the influence of pH or salt can contribute to a new form of "DNA switch", which would garner more interest on the i-motif in biological applications. Although the stated idea may seem implausible, it is important to note that formation of the G-quadruplex and i-motif have always been proposed and shown to occur complementary to the same genetic position,[73,214,215] and mutually exclusive formation of the two structures has been reported.[216] This indicates that the formation of each structure may be dependent on another upon the unfolding of duplex DNA. The G-quadruplex in retinoblastoma susceptibility genes (Rb), whose complementary C-rich sequence forms an i-motif, adopts an antiparallel structure,[214] adding more interest to the G-quadruplex formation of C_3GXY sequences observed in CD spectra.

Group II i-motifs containing less than four bases in the loop regions were reported to have low stabilities,[73] however, tests on intramolecular i-motifs with increasing thymine residues showed that i-motifs with shorter loop lengths are more stable. Altering loop lengths within individual i-motifs showed that the first and last loops (in i-motifs containing three loop regions) are important in defining i-motif stability. The intercalated structures have significantly lower melting temperatures when they have longer loops (from 57°C in C_3T_3 to 36°C in C_3T_8), however, these structures can be stabilised in the presence of [Ru(phen)$_2$dppz]$^{2+}$. The ruthenium polypyridyl complex did not significantly affect the thermal stabilities of C_3T_3 to C_3T_6 (only by 0.5°C for C_3T_4 and C_3T_6 individually) but it increased the T_m values of C_3T_7 and C_3T_8 by 3.5°C and 8°C, respectively. Interaction between the complex and the two structures was confirmed by luminescence titrations; where luminescence intensities of rac-[Ru(phen)$_2$dppz]$^{2+}$ at λ_{max} = 615 nm were highest for C_3T_7 and C_3T_8. This implies that the complex has higher binding affinity to long-looped i-motifs. Furthermore,

_____

experiments using enantiomers of the complex showed that the Λ-enantiomer is more luminescent than the Δ in the presence of an i-motif. The opposite is observed in B-DNA where Δ is known to bind more strongly and is also more emissive than its Λ counterpart.[174,175] Tests with different Λ-polypyridyl complexes need to be taken to find out if this family of complexes uniquely bind to DNA i-motifs.

Experiments can be carried out on the C_3T_X oligonucleotides using other known i-motif binding ligands, to investigate if the ligands favour binding to longer looped i-motifs. I-motif formation under physiological conditions have been shown by molecular crowding,[39,217] ligand[66,95] and protein binding.[41,87,218] If i-motif formation of C-rich sequences with longer loops are shown to occur through ligand binding, then targeting such sequences in the human genome can open doors to more potential biological functions of the four-stranded structure.

The X-ray crystal structure of C_3A shows that A·A base pairs occur on either side of the cytosine core. This is in contrast to the folding of all adenine bases away from the core in the A_2C_4 structure. NMR studies have shown that protonation of adenine at loops disrupts the i-motif core[75] and chapter 4 showed that adenine-rich i-motifs form less stable i-motif structures. This information can thus be used to design better sequences for i-motif crystallisation. In comparison to the previously solved structures of C_4 and C_3T, the new crystal structures showed approximately three times as many water molecules and DNA backbone disorders. These differences may be attributed to the use of different instruments. Three out of the four C_3X sequences crystallised in pH 7 and 8.5, showing more potential of i-motif formation under physiological conditions.

Neutron diffraction of only three DNA structures have been reported to date, making C_3T the third DNA and the first i-motif structure to be studied using neutron crystallography. Refinement methods applied to the study of the i-motif showed that occupancy of the D^+ deuteron is contributed by only one cytosine; confirming that the C^+·C base pair has asymmetric H-bonding. The new distance parameters obtained from neutron studies can be used to find other possible cations to replace the proton in N3···H^+···N3. As an i-motif forming sequence can potentially have three structural outputs; single strand, hairpin and i-motif,[179] the newly measured bond distance can provide insight into possible use of other cations to help achieve the three structures from the same sequence.

Following the crystal structure determination of two new tetramolecular i-motifs, crystallisation trials of bimolecular and unimolecular i-motifs have been conducted. Diffraction data of the human telomeric seuquence $d(C_3TA_2)_3C_3$ and the bimolecular sequence $d(C_3TA_2C_3TATA)_2$ were collected at 15.1 and 2.1Å resolutions, respectively. However, optimisation of the crystallisation conditions is to be carried out in order to grow more ordered crystals. Additionally, crystallisation trials were also made on three of the intramolecular i-motif forming C_3XYZ sequences; AGA, TCG and TCT. As expected from solution studies in Chapter 4, crystal growth was observed for TCG and TCT but diffraction was only observed to 15.1Å resolution.

_____

The information obtained from UV and CD spectroscopic experiments can be used to optimise better crystallising conditions for both inter- and intramolecular i-motifs. C-rich sequences that can form very stable i-motifs can be designed and predicted for crystal growth. The results obtained agree that solution studies and crystal growth can be complementary techniques. The results can also be used to create crystallisation reagents that are more suited to forming bimolecular and intramolecular i-motif crystals.

References

1. Watson, J. D. & Crick, F. H. C. A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 737–738 (1953).

2. Alderden, R. a, Hall, M. D. & Hambley, T. W. The Discovery and Development of Cisplatin. *J. Chem. Educ.* **83,** 728–734 (2006).

3. Rosenberg, B. *et al.* Platinum compounds: a new class of potent antitumour agents. *Nature* **222,** 385–386 (1969).

4. Weiss, R. B. The anthracyclines: will we ever find a better doxorubicin? *Semin. Oncol.* **19,** 670–86 (1992).

5. Goodman, R. P. *et al.* Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science (80-.).* **310,** 1661–1665 (2005).

6. Krishnan, S. *et al.* Molecular transport through large-diameter DNA nanopores. *Nat. Commun.* **7,** 12787 (2016).

7. Neidle, S. *DNA Structure and Recognition*. (1994).

8. Elson, D. & Chargaff, E. On the desoxyribonucleic acid content of sea urchin gametes. *Experientia* **8,** 143–5 (1952).

9. Chargaff, E., Lipshitz, R. & Green, C. Composition of the desoxypentose nucleic acids of four genera of sea-urchin. *J. Biol. Chem.* **195,** 155–60 (1952).

10. Hoogsteen, K. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **12,** 822–823 (1959).

11. Nikolova, E. N. *et al.* Transient Hoogsteen Base Pairs in Canonical Duplex DNA. *Nature* **470,** 498–502 (2011).

12. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423,** 145–150 (2003).

13. Whelan, D. R. *et al.* Detection of an en masse and reversible B- to A-DNA conformational transition in prokaryotes in response to desiccation. *J. R. Soc. Interface* **11,** (2014).

14. Rich, A. & Zhang, S. Z-DNA: the long road to biological function. *Nat. Rev. Genet.* **4,** 566–572 (2003).

15. Mitsui, Y. *et al.* Physical and enzymatic studies on poly d(I-C)-poly d(I-C), an unusual double-helical DNA. *Nature* **228,** 1166–9 (1970).

16. Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y.-G. & Kim, K. K. Crystal structure of a junction between B-

_____

DNA and Z-DNA reveals two extruded bases. *Nature* **437,** 1183–1186 (2005).

17.   Dickerson, R. E. *et al.* The anatomy of A-, B-, and Z-DNA. *Science (80-.).* **216,** 475–485 (1982).

18.   Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmond, M. H. & Waring, M. J. *The Molecular Basis of Antibiotic Action.* (John Wiley & Sons, 1981).

19.   Frantz, M.-C. & Wipf, P. Mitochondria as a target in treatment. *Environ. Mol. Mutagen.* **51,** 462–75 (2010).

20.   Irvin, T. R. & Wogan, G. N. Quantitation of aflatoxin B1 adduction within the ribosomal RNA gene sequences of rat liver DNA. *Proc. Natl. Acad. Sci. U. S. A.* **81,** 664–8 (1984).

21.   Beckmann, R. P. P., Agostino, M. J. J., McHugh, M. M., Sigmund, R. D. & Beerman, T. A. Assessment of Preferential Cleavage of an Actively Transcribed Retroviral Hybrid Gene in Murine Cells by Deoxyribonuclease I, Bleomycin, Neocarzinostatin, or Ionizing Radiation. *Biochemistry* **26,** 5409– 541 (1987).

22.   Klug, A. & Rhodes, D. Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harb. Symp. Quant. Biol.* **52,** 473–82 (1987).

23.   Ohlendorf, D. H., Anderson, W. F., Lewis, M., Pabo, C. O. & Matthews, B. W. Comparison of the structures of cro and lambda repressor proteins from bacteriophage lambda. *J. Mol. Biol.* **169,** 757– 69 (1983).

24.   Choi, J. *et al.* Conformational changes of non-B DNA. *Chem. Soc. Rev.* **40,** 5893 (2011).

25.   Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J. E. & Wells, R. D. The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair (Amst).* **5,** 1161–70 (2006).

26.   Sen, D. & Gilbert, W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **344,** 410–4 (1990).

27.   Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **350,** 718–720 (1991).

28.   Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342,** 825–829 (1989).

29.   Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5,** 182–6 (2013).

30.   Huppert, J. L. Structure, location and interactions of G-quadruplexes. *FEBS J.* **277,** 3452–8 (2010).

31.   Rodriguez, R. *et al.* Small molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. …* **8,** 301–310 (2012).

_____

32.   Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters : a novel anticancer strategy ? *Nat Rev Drug Discov.* **10,** 261–275 (2011).

33.   Parkinson, G. N., Lee, M. P. H. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417,** 876–80 (2002).

34.   Marsh, R. E., Bierstedt, R., Eichhorn, E. L. & IUCr. The crystal structure of cytosine-5-acetic acid. *Acta Crystallogr.* **15,** 310–316 (1962).

35.   Ahmed, S. & Henderson, E. Formation of novel hairpin structures by telomeric C-strand oligonucleotides. *Nucleic Acids Res.* **20,** 507–11 (1992).

36.   Gehring, K., Leroy, J.-L. & Gueron, M. A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature* **363,** 561–565 (1993).

37.   Brazier, J. A., Shah, A. & Brown, G. D. I-motif formation in gene promoters: unusually stable formation in sequences complementary to known G-quadruplexes. *Chem Commun* **48,** 10739–10741 (2012).

38.   Wright, E. P., Huppert, J. L. & Waller, Z. A. E. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* **45,** 2951–2959 (2017).

39.   Rajendran, A., Nakano, S. & Sugimoto, N. Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chem Commun* **46,** 1299–1301 (2010).

40.   Fujii, T. & Sugimoto, N. intrastrand i-motif structures at neutral pH †. *Phys. Chem. Chem. Phys.* **17,** 16719–16722 (2015).

41.   Kendrick, S. *et al.* The Dynamic Character of the BCL2 Promoter i‐Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure. *J. Am. Chem. Soc.* **136,** 4161–4171 (2014).

42.   Dong, Y., Yang, Z. & Liu, D. DNA nanotechnology based on i-motif structures. *Acc. Chem. Res.* **47,** 1853–1860 (2014).

43.   Liu, D. & Balasubramanian, S. A Proton-Fuelled DNA Nanomachine. *Angew. Chemie - Int. Ed.* **42,** 5734–5736 (2003).

44.   Li, T. & Famulok, M. I-motif-programmed functionalization of DNA nanocircles. *J. Am. Chem. Soc.* **135,** 1593–9 (2013).

45.   Choi, J. & Majima, T. Reversible conformational switching of i-motif DNA studied by fluorescence spectroscopy. *Photochem. Photobiol.* **89,** 513–522 (2013).

46.   Wang, Z.-G., Elbaz, J. & Willner, I. DNA machines: bipedal walker and stepper. *Nano Lett.* **11,** 304–9

(2011).

47.   Wang, W. *et al.* Use of the interparticle i-motif for the controlled assembly of gold nanoparticles. *Langmuir* **23,** 11956–9 (2007).

48.   Sharma, J., Chhabra, R., Yan, H. & Liu, Y. pH-driven conformational switch of "i-motif" DNA for the reversible assembly of gold nanoparticles. *Chem. Commun. (Camb).* 477–9 (2007). doi:10.1039/b612707j

49.   Elbaz, J., Wang, Z.-G., Orbach, R. & Willner, I. pH-stimulated concurrent mechanical activation of two DNA 'tweezers';. A 'SET-RESET' logic gate system. *Nano Lett.* **9,** 4510–4 (2009).

50.   Yang, Y. *et al.* An electrochemically actuated reversible DNA switch. *Nano Lett.* **10,** 1393–7 (2010).

51.   Li, T., Ackermann, D., Hall, A. M. & Famulok, M. Input-dependent induction of oligonucleotide structural motifs for performing molecular logic. *J. Am. Chem. Soc.* **134,** 3508–16 (2012).

52.   Modi, S. *et al.* A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nat. Nanotechnol.* **4,** 325–30 (2009).

53.   Surana, S., Bhat, J. M., Koushika, S. P. & Krishnan, Y. An autonomous DNA nanomachine maps spatiotemporal pH changes in a multicellular living organism. *Nat. Commun.* **2,** 340 (2011).

54.   Phan, A. T., Gueron, M. & Leroy, J.-L. The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.* **299,** 123–44 (2000).

55.   Lieblein, A. L., Krämer, M., Dreuw, A., Fürtig, B. & Schwalbe, H. The nature of hydrogen bonds in cytidine···H +··· cytidine DNA base pairs. *Angew. Chemie - Int. Ed.* **51,** 4067–4070 (2012).

56.   Yang, B. & Rodgers, M. T. Base-Pairing Energies of Proton-Bound Heterodimers of Cytosine and Modi fi ed Cytosines: Implications for the Stability of DNA i‐Motif Conformations. *J. Am. Chem. Soc.* **136,** 282–290 (2014).

57.   Bloomfield, V. A., Crothers, D. M. & Tinoco, I. J. *Nucleic acids: Structures, Properties, and Functions.* (University Science Books, 2000).

58.   Cai, L. *et al.* Intercalated cytosine motif and novel adenine clusters in the crystal structure of the Tetrahymena telomere. *Nucleic Acids Res.* **26,** 4696–705 (1998).

59.   Han, X., Leroy, J. & Gue, M. An Intramolecular i-Motif : the Solution Structure and Base-pair Opening Kinetics of d(5mCCT3CCT3ACCT3CC). *J. M* **278,** 949–965 (1998).

60.   Leroy, J. L., Guéron, M., Mergny, J. L. & Hélène, C. Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res.* **22,** 1600–6 (1994).

61.   Völker, J., Klump, H. H. & Breslauer, K. J. The energetics of i-DNA tetraplex structures formed

_____

intermolecularly by d(TC5) and intramolecularly by d[(C5T3)3C5]. *Biopolymers* **86,** 136–147 (2007).

62. Khan, N. *et al.* Solution equilibria of the i-motif-forming region upstream of the B-cell lymphoma-2 P1 promoter. *Biochimie* **89,** 1562–72 (2007).

63. Zager, R. A., Schimpf, B. A., Gmur, D. J. & Burke, T. J. Phospholipase A2 activity can protect renal tubules from oxygen deprivation injury. *PNAS* **90,** 8297–301 (1993).

64. Escaja, N. *et al.* A minimal i-motif stabilized by minor groove G:T:G:T tetrads. *Nucleic Acids Res.* **40,** 11737–47 (2012).

65. Zhou, J. *et al.* Formation of i-motif structure at neutral and slightly alkaline pH. *Mol. Biosyst.* **6,** 580–586 (2010).

66. Day, H. A., Huguin, C. & Waller, Z. A. E. Silver cations fold i-motif at neutral pH. *Chem. Commun.* **49,** 7696–7698 (2013).

67. Mergny, J.-L., Lacroix, J. L., Han, X., Leroy, J.-L. & Helen, C. Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif. *J. Am. Chem. Soc.* **117,** 8887–8898 (1995).

68. Mathur, V., Verma, A., Maiti, S. & Chowdhury, S. Thermodynamics of i-tetraplex formation in the nuclease hypersensitive element of human c-myc promoter. *Biochem. Biophys. Res. Commun.* **320,** 1220–1227 (2004).

69. Fojtík, P. & Vorlíčková, M. The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH. *Nucleic Acids Res.* **29,** 4684–90 (2001).

70. Canalia, M. & Leroy, J.-L. [5mCCTCTCTCC] $_4$: An i-Motif Tetramer with Intercalated T•T Pairs. *J. Am. Chem. Soc.* **131,** 12870–12871 (2009).

71. Canalia, M. & Leroy, J. L. Structure, internal motions and association-dissociation kinetics of the i-motif dimer of d(5mCCTCACTCC). *Nucleic Acids Res* **33,** 5471–81 (2005).

72. Lieblein, A. L., Fürtig, B. & Schwalbe, H. Optimizing the Kinetics and Thermodynamics of DNA i-Motif Folding. *ChemBioChem* **14,** (2013).

73. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **277,** 3459–3469 (2011).

74. Benabou, S. *et al.* Solution equilibria of cytosine- and guanine-rich sequences near the promoter region of the n-myc gene that contain stable hairpins within lateral loops. *Biochim. Biophys. Acta* **1840,** 41–52 (2014).

75. Nonin-Lecomte, S. & Leroy, J. L. Structure of a C-rich strand fragment of the human centromeric satellite III: a pH-dependent intercalation topology. *J. MoL Biol* **309,** 491–506 (2001).

_____

76.   Gallego, J., Chou, S.-H. & Reid, B. R. Centromeric pyrimidine strands fold into an intercalated motif by forming a double hairpin with a Novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer. *J. Mol. Biol.* **273,** 840–856 (1997).

77.   Kaushik, M., Prasad, M., Kaushik, S., Singh, A. & Kukreti, S. Structural transition from dimeric to tetrameric i-motif, caused by the presence of TAA at the 3'-end of human telomeric C-rich sequence. *Biopolymers* **93,** 150–160 (2010).

78.   Esmaili, N. & Leroy, J. L. i-motif solution structure and dynamics of the d(AACCCC) and d(CCCCAA) tetrahymena telomeric repeats. *Nucleic Acids Res.* **33,** 213–24 (2005).

79.   Weil, J. *et al.* Stabilization of the i -motif by intramolecular adenine–adenine–thymine base triple in the structure of d(ACCCT). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 422–429 (1999).

80.   Choi, J., Kim, S., Tachikawa, T., Fujitsuka, M. & Majima, T. pH-induced intramolecular folding dynamics of i-motif DNA. *J. Am. Chem. Soc.* **133,** 16146–53 (2011).

81.   Mergny, J.-L., Riou, J.-F., Mailliet, P., Teulade-Fichou, M.-P. & Gilson, E. Natural and pharmacological regulation of telomerase. *Nucleic Acids Res.* **30,** 839–65 (2002).

82.   Kim, N. W. *et al.* Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science (80-.).* **266,** 2011–2015 (1994).

83.   Balasubramanian, S. & Neidle, S. G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **13,** 345–53 (2009).

84.   Simonsson, T., Pribylova, M. & Vorlickova, M. A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures. *Biochem. Biophys. Res. Commun.* **278,** 158–166 (2000).

85.   González, V. & Hurley, L. H. The c-MYC NHE III(1): function and regulation. *Annu. Rev. Pharmacol. Toxicol.* **50,** 111–129 (2010).

86.   Bar-Am, O., Weinreb, O., Amit, T. & Youdim, M. B. Regulation of Bcl-2 family proteins, neurotrophic factors, and APP processing in the neurorescue activity of propargylamine. *FASEB J.* **19,** 1899–1901 (2005).

87.   Kang, H.-J., Kendrick, S., Hecht, S. M. & Hurley, L. H. The Transcriptional Complex Between the BCL2 i‑Motif and hnRNP LL Is a Molecular Switch for Control of Gene Expression That Can Be Modulated by Small Molecules. *J. Am. Chem. Soc.* **136,** 4172–4185 (2014).

88.   Fedoroff, O. Y. *et al.* Cationic porphyrins promote the formation of i-motif DNA and bind peripherally by a nonintercalative mechanism. *Biochemistry* **39,** 15083–15090 (2000).

89.   Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl.*

Acad. Sci. U. S. A. **99,** 11593–8 (2002).

90. Grand, C. L. *et al.* The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.* **1,** 565–73 (2002).

91. Alberti, P. *et al.* Interaction of an Acridine Dimer with DNA Quadruplex Structures. *J. Biomol. Struct. Dyn.* **193,** 505–513 (2001).

92. Chen, Y. *et al.* Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres. *Nat. Commun.* **3,** 1074 (2012).

93. Zhao, C., Ren, J. & Qu, X. Single-Walled Carbon Nanotubes Binding to Human Telomeric i-Motif DNA Under Molecular-Crowding Conditions: More Water Molecules Released. *Chem. - A Eur. J.* **14,** 5435–5439 (2008).

94. Shi, S. *et al.* Molecular 'light switch' for G-quadruplexes and i-motif of human telomeric DNA: [Ru(phen)2(dppz)]2+. *Dalt. Trans.* **39,** 2490–2493 (2010).

95. Wright, E. P. *et al.* Mitoxantrone and Analogues Bind and Stabilize i-Motif Forming DNA Sequences. *Sci. Rep.* **6:39456,** (2016).

96. Sheng, Q. *et al.* Identification of new DNA i-motif binding ligands through a fluorescent intercalator displacement assay. *Org. Biomol. Chem.* **15,** 5669–5673 (2017).

97. Ma, D.-L. *et al.* Crystal violet as a fluorescent switch-on probe for i-motif: label-free DNA-based logic gate. *Analyst* **136,** 2692–2696 (2011).

98. Skoog, D. A. ., Holler, J. F. & Crouch, S. R. *Principles of Instrumental Analysis.* (Thomson Brooks/Cole., 2007).

99. Dave3457. Circular Polarization. *Wikimedia Commons* (2010). Available at: https://en.wikipedia.org/wiki/Circular_polarization. (Accessed: 8th August 2017)

100. Rhodes, G. *Crystallography Made Crystal Clear, Third Edition: A Guide for Users of Macromolecular Models.* (Academic Press, 2006).

101. McRee, D. *Practical Protein Crystallography.* (Academic Press, 1993).

102. Hamilton, W. C. & IUCr. Comparison of X-ray and neutron diffraction structural results: a study in methods of error analysis. *Acta Crystallogr. Sect. A Cryst. Physics, Diffraction, Theor. Gen. Crystallogr.* **25,** 194–206 (1969).

103. Bragg, W. H. & Bragg, W. L. The Reflection of X-rays by Crystals. *Proc. R. Soc. London A Math. Phys. Eng. Sci.* **88,** (1913).

_____

104.   Bragg, W. L. The Diffraction of Short Electromagnetic Waves by a Crystal. *Proc. Camb. Philol. Soc.* **17,** 43 (1913).

105.   Clegg, W. *Crystal Structure Determination.* (Oxford University Press, 1998).

106.   Rossmann, M. G. *et al.* Molecular replacement – historical background. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 1360–1366 (2001).

107.   International Union of Crystallography. Online Dictionary of Crystallography. Available at: http://reference.iucr.org/dictionary/R_factor. (Accessed: 9th August 2017)

108.   Project, C. C. & Collaborative Computational Project, N. 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763 (1994).

109.   Adams, P. D. *et al.* PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

110.   Piccoli, P. M. B., Koetzle, T. F. & Schultz, A. J. Single Crystal Neutron Diffraction for the Inorganic Chemist – A Practical Guide. *Comments Inorg. Chem.* **28,** 3–38 (2007).

111.   Langan, P., Li, X., Hanson, B. L., Coates, L. & Mustyakimov, M. Synthesis, capillary crystallization and preliminary joint X-ray and neutron crystallographic study of Z-DNA without polyamine at low pH. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **62,** 453–456 (2006).

112.   Leal, R. M. F. *et al.* Combined neutron and X-ray diffraction studies of DNA in crystals and solutions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 1244–1248 (2010).

113.   Afonine, P. V *et al.* Joint X-ray and neutron refinement with phenix.refine. *Acta Cryst* **66,** 1153–1163 (2010).

114.   Blakeley, M. P. *et al.* Quantum model of catalysis based on a mobile proton revealed by subatomic x-ray and neutron diffraction studies of h-aldose reductase. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 1844–8 (2008).

115.   Blakeley, M. P. *et al.* Comparison of hydrogen determination with X-ray and neutron crystallography in a human aldose reductase-inhibitor complex. *Eur. Biophys. J.* **35,** 577–83 (2006).

116.   Blum, M.-M. M. *et al.* Preliminary time-of-flight neutron diffraction study on diisopropyl fluorophosphatase (DFPase) from Loligo vulgaris. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **63,** 42–45 (2007).

117.   Tuan, H.-F. F., Erskine, P., Langan, P., Cooper, J. & Coates, L. Preliminary neutron and ultrahigh-resolution X-ray diffraction studies of the aspartic proteinase endothiapepsin cocrystallized with a gem-diol inhibitor. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **63,** 1080–1083 (2007).

118.   Coates, L. *et al.* The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray

diffraction. *J. ...* **130,** 7235–7237 (2008).

119. Katz, A. K. *et al.* Locating active-site hydrogen atoms in D-xylose isomerase: time-of-flight neutron diffraction. *PNAS* **103,** 8342–7 (2006).

120. Meilleur, F., Snell, E. H., van der Woerd, M. J., Judge, R. A. & Myles, D. A. A. A quasi-Laue neutron crystallographic study of D-xylose isomerase. *Eur. Biophys. J.* **35,** 601–9 (2006).

121. Kovalevsky, A. Y. *et al.* Hydrogen location in stages of an enzyme-catalyzed reaction: time-of-flight neutron structure of D-xylose isomerase with bound D-xylulose. *Biochemistry* **47,** 7595–7 (2008).

122. Niimura, N. & Bau, R. Neutron protein crystallography: beyond the folding structure of biological macromolecules. *Acta Crystallogr. A.* **64,** 12–22 (2008).

123. Blakeley, M. P., Langan, P., Niimura, N. & Podjarny, A. Neutron crystallography: opportunities, challenges, and limitations. *Curr. Opin. Struct. Biol.* **18,** 593–600 (2008).

124. National Institute of Standards and Technology (NIST). A New Neutron Imaging Facility at BT-6 for the Non-Destructive Analysis of Working Fuel Cells. Available at: https://www.ncnr.nist.gov/AnnualReport/FY2003_html/RH2/. (Accessed: 24th November 2016)

125. Blakeley, M. P., Hasnain, S. S. & Antonyuk, S. V. Sub-atomic resolution X-ray crystallography and neutron crystallography: promise, challenges and potential. *IUCrJ* **2,** 464–74 (2015).

126. International Atomic Energy Agency. Measurement of residual stress in materials using neutrons. in (2003).

127. Kypr, J., Kejnovská, I., Renciuk, D. & Vorlícková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37,** 1713–25 (2009).

128. Manzini, G., Yathindra, N. & Xodo, L. E. Evidence for intramolecularly folded i-DNA structures in biologically relevant CCC-repeat sequences. *Nucleic Acids Res.* **22,** 4634–4640 (1994).

129. Liu, H. *et al.* Light-driven conformational switch of i-motif DNA. *Angew. Chem. Int. Ed. Engl.* **46,** 2515–7 (2007).

130. Holm, A. I. S. *et al.* Vacuum-ultraviolet circular dichroism spectroscopy of DNA: a valuable tool to elucidate topology and electronic coupling in DNA. *Phys. Chem. Chem. Phys.* **12,** 9581 (2010).

131. Gurung, S. P., Schwarz, C., Hall, J. P., Cardin, C. J. & Brazier, J. a. The importance of loop length on the stability of i-motif structures. *Chem. Commun.* **51,** 5630–5632 (2015).

132. Chen, L., Cai, L., Zhang, X. & Rich, A. Crystal structure of a four-stranded intercalated DNA: d(C4). *Biochemistry* **33,** 13540–6 (1994).

133. Berger, I. *et al.* Extension of the four-stranded intercalated cytosine motif by adenine.adenine base

pairing in the crystal structure of d(CCCAAT). *Struct. Biol.* **2,** 416–425 (1995).

134.   Kang, C. *et al.* Crystal structure of intercalated four-stranded d (C3T) at 1 . 4 A resolution. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11636–11640 (1994).

135.   Kang, C. *et al.* Stable loop in the crystal structure of the intercalated four-stranded cytosine-rich metazoan telomere. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 3874–8 (1995).

136.   Berger, I., Cai, L., Chen, L. & Rich, A. Energetics of the lattice: Packing elements in crystals of four-stranded intercalated cytosine-rich DNA molecules. *Biopolymers* **44,** 257–267 (1997).

137.   Chen, Y. W., Jhan, C. R., Neidle, S. & Hou, M. H. Structural basis for the identification of an i-motif tetraplex core with a parallel-duplex junction as a structural motif in ccg triplet repeats. *Angew. Chemie - Int. Ed.* **53,** 10682–10686 (2014).

138.   Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **12,** 847–54 (2005).

139.   Cui, Y. *et al.* Molecular population dynamics of DNA structures in a bcl-2 promoter sequence is regulated by small molecules and the transcription factor hnRNP LL. *Nucleic Acids Res.* **42,** 5755–5764 (2014).

140.   Guédin, A., Gros, J., Alberti, P. & Mergny, J.-L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **38,** 7858–68 (2010).

141.   Keane, P. M. *et al.* Long-lived excited states in i-motif DNA studied by picosecond time-resolved IR spectroscopy. *Chem. Commun. (Camb).* **50,** (2013).

142.   Shi, S. *et al.* Molecular 'light switch' for G-quadruplexes and i-motif of human telomeric DNA: [Ru(phen)2(dppz)]2+. *Dalt. Trans.* **39,** 2490 (2010).

143.   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).

144.   Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* **48,** 2013–8 (1962).

145.   Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* **19,** 414–422 (2009).

146.   Hardin, C. C., Watson, T., Corregan, M. & Bailey, C. Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG). *Biochemistry* **31,** 833–841 (1992).

147.   Miura, T., Benevides, J. M. & Thomas, G. J. A Phase Diagram for Sodium and Potassium Ion Control of Polymorphism in Telomeric DNA. *J. MoL Biol* **248,** 233–238 (1995).

_____

148.   Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33,** 2901–2907 (2005).

149.   Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **126,** 16405–15 (2004).

150.   Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8572–7 (2001).

151.   Leonard, A. C. & Méchali, M. DNA replication origins. *Cold Spring Harb. Perspect. Biol.* **5,** a010116 (2013).

152.   Gilbert, D. M. Replication origins run (ultra) deep. *Nat. Struct. Mol. Biol.* **19,** 740–2 (2012).

153.   Sarkies, P., Reams, C., Simpson, L. J. & Sale, J. E. Epigenetic Instability due to Defective Replication of Structured DNA. *Mol. Cell* **40,** 703–713 (2010).

154.   Sarkies, P. *et al.* FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res.* **40,** 1485–98 (2012).

155.   Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513,** 65–70 (2014).

156.   Brown, V. *et al.* Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107,** 477–87 (2001).

157.   Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40,** 4727–41 (2012).

158.   Michelson, A. M. *The Chemistry of Nucleosides and Nucleotides*. (Academic Press, London, 1963).

159.   Cherepanov, a V & de Vries, S. Binding of nucleotides by T4 DNA ligase and T4 RNA ligase: optical absorbance and fluorescence studies. *Biophys. J.* **81,** 3545–59 (2001).

160.   Wild, N. Diamond key facts. *Oxford Mail* (2010).

161.   Reilly, S. M., Morgan, R. K., Brooks, T. a & Wadkins, R. M. Effect of Interior Loop Length on the Thermal Stability and pKa of i-Motif DNA. *Biochemistry* **54,** 1364–1370 (2015).

162.   Kendrick, S., Akiyama, Y., Hecht, S. M. & Hurley, L. H. The i-motif in the bcl-2 P1 promoter forms an unexpectedly stable structure with a unique 8:5:7 loop folding pattern. *J. Am. Chem. Soc.* **131,** 17667–76 (2009).

163.   Nesterova, I. V. & Nesterov, E. E. Rational design of highly responsive pH sensors based on DNA i-Motif. *J. Am. Chem. Soc.* **136,** 8843–8846 (2014).

164.   Reed, J. C. *et al.* BCL-2 family proteins: regulators of cell death involved in the pathogenesis of

_____

cancer and resistance to therapy. *J. Cell. Biochem.* **60,** 23–32 (1996).

165.   Guo, K., Gokhale, V., Hurley, L. H. & Sun, D. Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res.* **36,** 4598–4608 (2008).

166.   Gill, M. R. & Thomas, J. A. Ruthenium(II) polypyridyl complexes and DNA--from structural probes to cellular imaging and therapeutics. *Chem. Soc. Rev.* **41,** 3179–3192 (2012).

167.   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.* **108,** 17610–17614 (2011).

168.   Niyazi, H. *et al.* Crystal structures of L-[Ru(phen)2 dppz]2+ with oligonucleotides containing TA/TA and AT/AT steps show two intercalation modes. *Nat. Chem.* **4,** 621–628 (2012).

169.   Yu, Q. *et al.* Chiral ruthenium(II) polypyridyl complexes: stabilization of g-quadruplex DNA, inhibition of telomerase activity and cellular uptake. *PLoS One* **7,** e50902 (2012).

170.   Yu, Q. *et al.* Ruthenium(II) polypyridyl complexes as G-quadruplex inducing and stabilizing ligands in telomeric DNA. *Metallomics* **5,** 222–31 (2013).

171.   Poon, C.-T. *et al.* An amphiphilic ruthenium(II)-polypyridyl appended porphyrin as potential bifunctional two-photon tumor-imaging and photodynamic therapeutic agent. *J. Inorg. Biochem.* **104,** 62–70 (2010).

172.   Coates, C. G., Mcgarvey, J. J., Callaghan, P. L., Coletti, M. & Hamilton, J. G. Probing the Interaction of [Ru (phen) 2 (dppz)] 2 + with Single-Stranded DNA - What Degree of Protection Is Required for Operation of the ' Light-Switch Effect '? *J. Phys. Chem. B.* **105,** 730–735 (2001).

173.   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).

174.   Hiort, C., Lincoln, P. & Norden, B. DNA binding of .DELTA.- and .LAMBDA.-[Ru(phen)2DPPZ]2+. *J. Am. Chem. Soc.* **115,** 3448–3454 (1993).

175.   Lincoln, P., Broo, A. & Nordén, B. Diastereomeric DNA-Binding Geometries of Intercalated Ruthenium(II) Trischelates Probed by Linear Dichroism: [Ru(phen) 2 DPPZ] 2+ and [Ru(phen) 2 BDPPZ] 2+. *J. Am. Chem. Soc.* **118,** 2644–2653 (1996).

176.   Kumar, N., Sahoo, B., Varun, K. a S., Maiti, S. & Maiti, S. Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic Acids Res.* **36,** 4433–4442 (2008).

177.   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).

_____

178. Li, T. *et al.* Effects of the i-motif DNA loop on the fluorescence of silver nanoclusters. *RSC Adv.* **6,** 22839–22844 (2016).

179. Day, H. A., Wright, E. P., Macdonald, C. J., Gates, A. J. & Waller, Z. A. E. Reversible DNA i-motif to hairpin switching induced by copper(II) cations. *Chem. Commun.* **51,** 14099–14102 (2015).

180. Karsisiotis, A. I. *et al.* Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed. Engl.* **50,** 10645–8 (2011).

181. Senior, M. M., Jones, R. A. & Breslauer, K. J. Influence of loop residues on the relative stabilities of DNA hairpin structures (thermodynamics). *PNAS* **85,** 6242–6246 (1988).

182. McKim, M., Buxton, A., Johnson, C., Metz, A. & Sheardy, R. D. Loop Sequence Context Influences the Formation and Stability of the i-Motif for DNA Oligomers of Sequence (CCCXXX)4, where X = A and/or T, under Slightly Acidic Conditions. *J. Phys. Chem. B* **120,** 7652–7661 (2016).

183. Guédin, A., Alberti, P. & Mergny, J. L. Stability of intramolecular quadruplexes: Sequence effects in the central loop. *Nucleic Acids Res.* **37,** 5559–5567 (2009).

184. Yang, Y. *et al.* Influence of tetra(ethylene glycol) (EG 4) substitution at the loop region on the intramolecular DNA i -motif. *Macromolecules* **45,** 2643–2647 (2012).

185. Benabou, S. *et al.* Understanding the effect of the nature of the nucleobase in the loops on the stability of the i-motif structure. *Phys. Chem. Chem. Phys.* **18,** 7997–8004 (2016).

186. Geinguenaud, F. *et al.* Parallel self-associated structures formed by T,C-rich sequences at acidic pH. *Biochemistry* **39,** 12650–8 (2000).

187. Turner, D. H., Sugimoto, N., Kierzek, R. & Dreikert, S. D. Free Energy Increments for Hydrogen Bonds in Nucleic Acid Base Pairs. *J. Am. Chem. Soc.* **109,** 3783–3785 (1987).

188. Cui, J., Waltman, P., Le, V. H. & Lewis, E. a. The effect of molecular crowding on the stability of human c-MYC promoter sequence I-motif at neutral pH. *Molecules* **18,** 12751–12767 (2013).

189. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 125–32 (2010).

190. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst* **43,** 186–190 (2010).

191. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

192. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 486–501 (2010).

193. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 352–367 (2012).

_____

194.   Zheng, G., Lu, X.-J. & Olson, W. K. Web 3DNA--a web server for the analysis, reconstruction, and
        visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **37,** W240-6 (2009).

195.   Frederick, C. A., Coll, M., Van der Marel, G. A., Van Boom, J. H. & Wang, A. H. J. Molecular structure of
        cyclic deoxydiadenylic acid at atomic resolution. *Biochemistry* **27,** 8350–8361 (1988).

196.   Agron, P. A., Busing, W. R. & IUCr. Magnesium dichloride hexahydrate, MgCl2.6H2O, by neutron
        diffraction. *Acta Crystallogr. Sect. C Cryst. Struct. Commun.* **41,** 8–10 (1985).

197.   Hellenbrandt, M. The Inorganic Crystal Structure Database (ICSD) – Present and Future.
        Crystallogr. Rev. **10,** 17–22 (2004).

198.   Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson±Crick base pairs and their associated
        isostericity matrices. *Nucleic Acids Res.* **30,** 3497–3531 (2002).

199.   Berman, H. M. *et al.* The Nucleic Acid Database. *Acta Crystallogr. D. Biol. Crystallogr.* **58,** 889–98
        (2002).

200.   Leroy, J.-L., Gehring, K., Kettani, A. & Gudron, M. Acid Multimers of Oligodeoxycytidine Strands:
        Stoichiometry, Base-Pair Characterization, and Proton Exchange Properties? *Biochemistry* **32,**
        6019–6031 (1993).

201.   Schmidtmann, M. *et al.* Determining hydrogen positions in crystal engineered organic molecular
        complexes by joint neutron powder and single crystal X-ray diffraction. *CrystEngComm* **16,** 1232–
        1236 (2014).

202.   Woińska, M., Grabowsky, S., Dominiak, P. M., Woźniak, K. & Jayatilaka, D. Hydrogen atoms can be
        located accurately and precisely by x-ray crystallography. *Sci. Adv.* **2,** e1600192 (2016).

203.   *Structure from Diffraction Methods*. (John Wiley & Sons, 2014).

204.   Sears, V. F. *Neutron Scattering Lengths and Cross Sections*. *Neutron News* **3,** 26–37 (Taylor & Francis
        Group, 1992).

205.   Coppens, P., Boehme, R., Price, P. F., Stevens, E. D. & IUCr. Joint X-ray and neutron data refinement
        of structural and charge density parameters. *Acta Crystallogr. Sect. A* **37,** 857–863 (1981).

206.   Wlodawer, A., Miller, M. & Sjölin, L. Active site of RNase: neutron diffraction study of a complex
        with uridine vanadate, a transition-state analog. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 3628–31 (1983).

207.   Sears, V. F. Special Feature: Neutron scattering lengths and cross sections.

208.   Coates, L. *et al.* The macromolecular neutron diffractometer (MaNDi) at the Spallation Neutron
        Source, Oak Ridge: enhanced optics design, high-resolution neutron detectors and simulated
        diffraction. *J. Appl. Crystallogr.* **43,** 570–577 (2010).

_____

209. Coates, L. *et al.* The Macromolecular Neutron Diffractometer MaNDi at the Spallation Neutron Source. *J. Appl. Crystallogr.* **48,** 1302–1306 (2015).

210. Arnold, O. *et al.* Mantid - Data Analysis and Visualization Package for Neutron Scattering and micro-SR Experiments. (2014). doi:10.1016/j.nima.2014.07.029

211. Campbell, J. W. *et al.* LAUEGEN version 6.0 and INTLDM. *J. Appl. Crystallogr.* **31,** 496–502 (1998).

212. Xie, W. Y., Huang, W. T., Li, N. B. & Luo, H. Q. Design of a dual-output fluorescent DNA logic gate and detection of silver ions and cysteine based on graphene oxide. *Chem. Commun.* **48,** 82–84 (2012).

213. Benabou, S. *et al.* i-motif structures in long cytosine-rich sequences found upstream of the promoter region of the SMARCA4 gene. *Biochimie* **140,** 20–33 (2017).

214. Xu, Y. & Sugiyama, H. Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* **34,** 949–54 (2006).

215. Phan, A. T. & Mergny, J.-L. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* **30,** 4618–25 (2002).

216. Cui, Y., Kong, D., Ghimire, C., Xu, C. & Mao, H. Mutually Exclusive Formation of G‑Quadruplex and i‑Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **55,** 2291–2299 (2016).

217. Takahashi, S. & Sugimoto, N. Effect of pressure on the stability of G-quadruplex DNA: thermodynamics under crowding conditions. *Angew. Chem. Int. Ed. Engl.* **52,** 13774–8 (2013).

218. Bai, X. *et al.* Enhanced Binding Affinity for an i-Motif DNA Substrate Exhibited by a Protein Containing Nucleobase Amino Acids. *J. Am. Chem. Soc.* **139,** 4611–4614 (2017).

_____

Appendix

A1. Chapter 1 – N/A

A2. Chapter 2
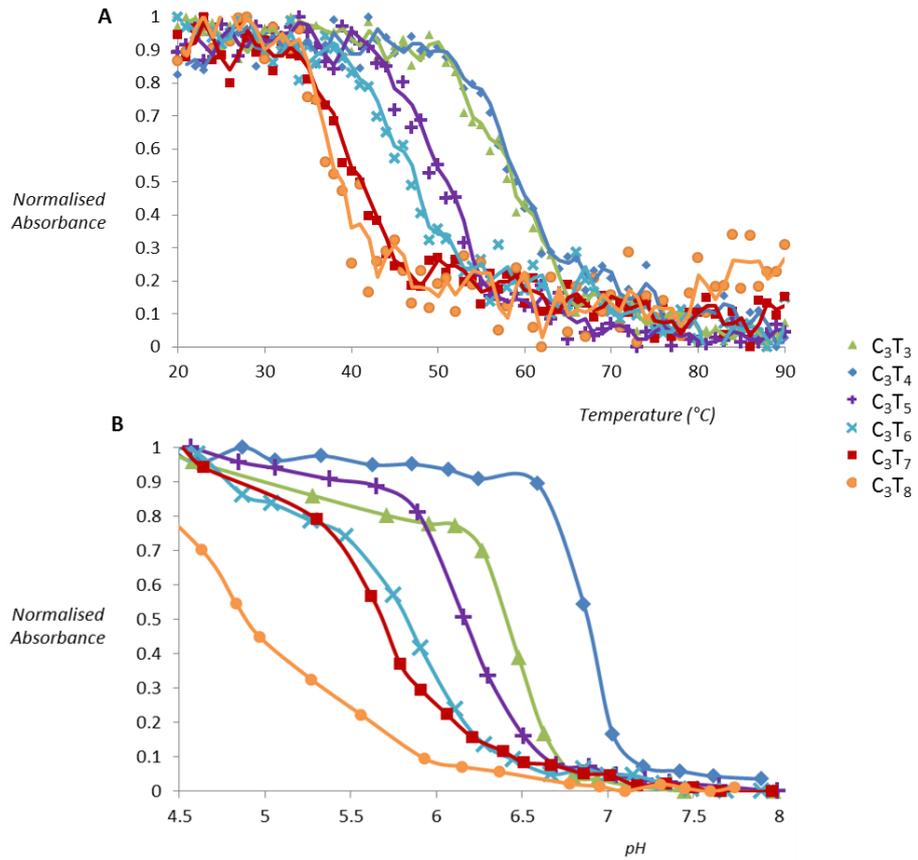
Figure A2.1: (**A**) DNA melting profiles (pH 5) and (**B**) effect of pH on C_3T_x i-motifs in 50 mM sodium cacodylate buffer at 295 nm wavelength.
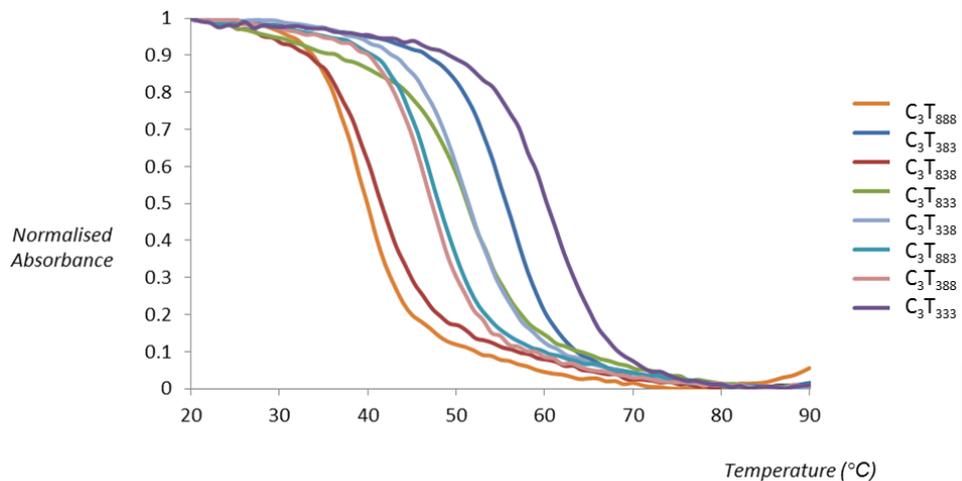
Figure A2.2: DNA melting profiles of C_3T_{abc} in 50 mM soidum cacodylate buffer (pH 5) at 295 nm wavelength.

A3. Chapter 3
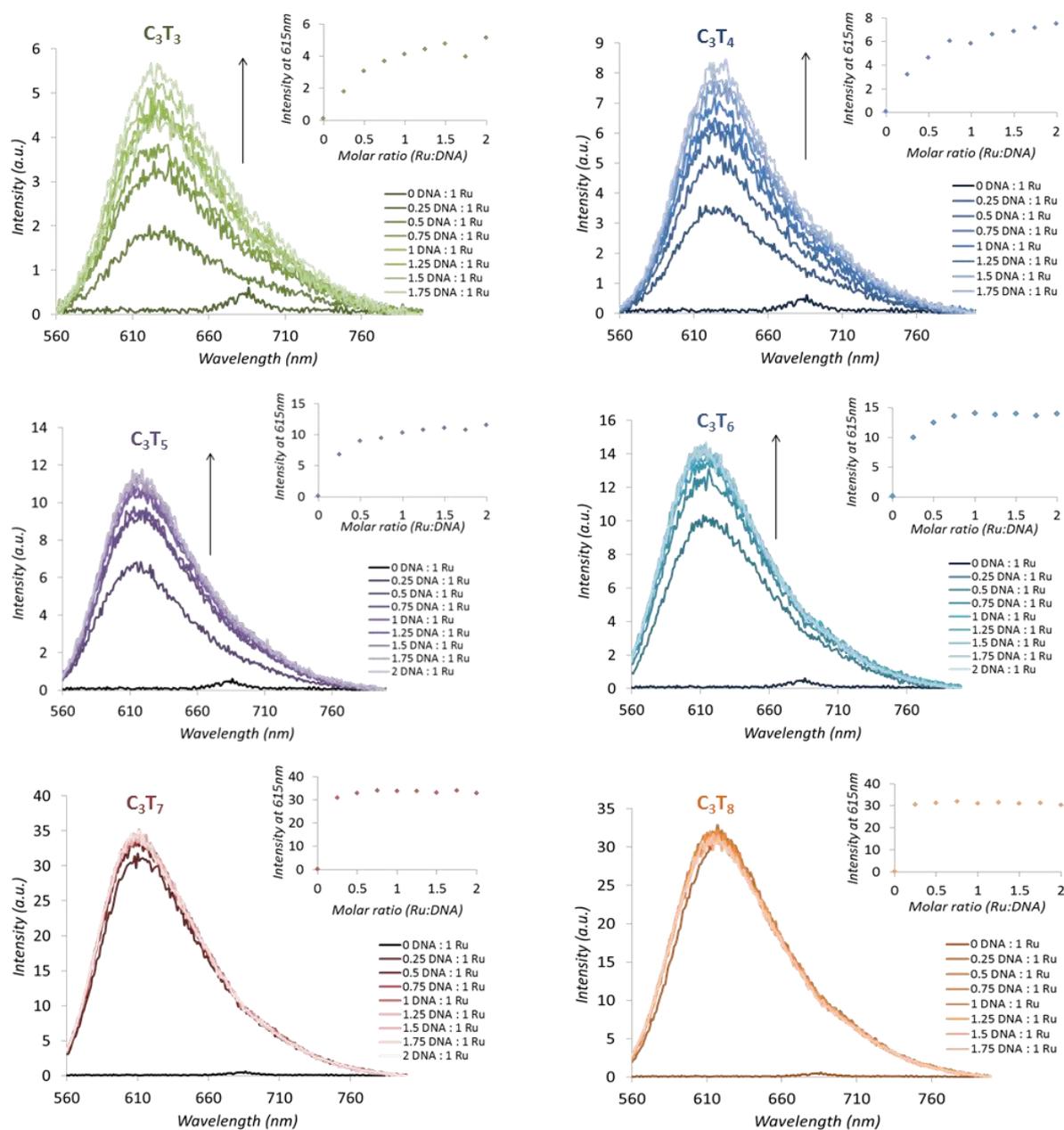
Figure A3.1: Fluorescence emission spectra of *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ (40 µM) upon the addition of C$_3$T$_X$ i-motifs (0–80 µM) in sodium cacodylate (50 mM, pH 5.0). Inset: fluorescence intensity of the complex *vs.* the i-motifs at λ$_{max}$ of 615 nm.

_____
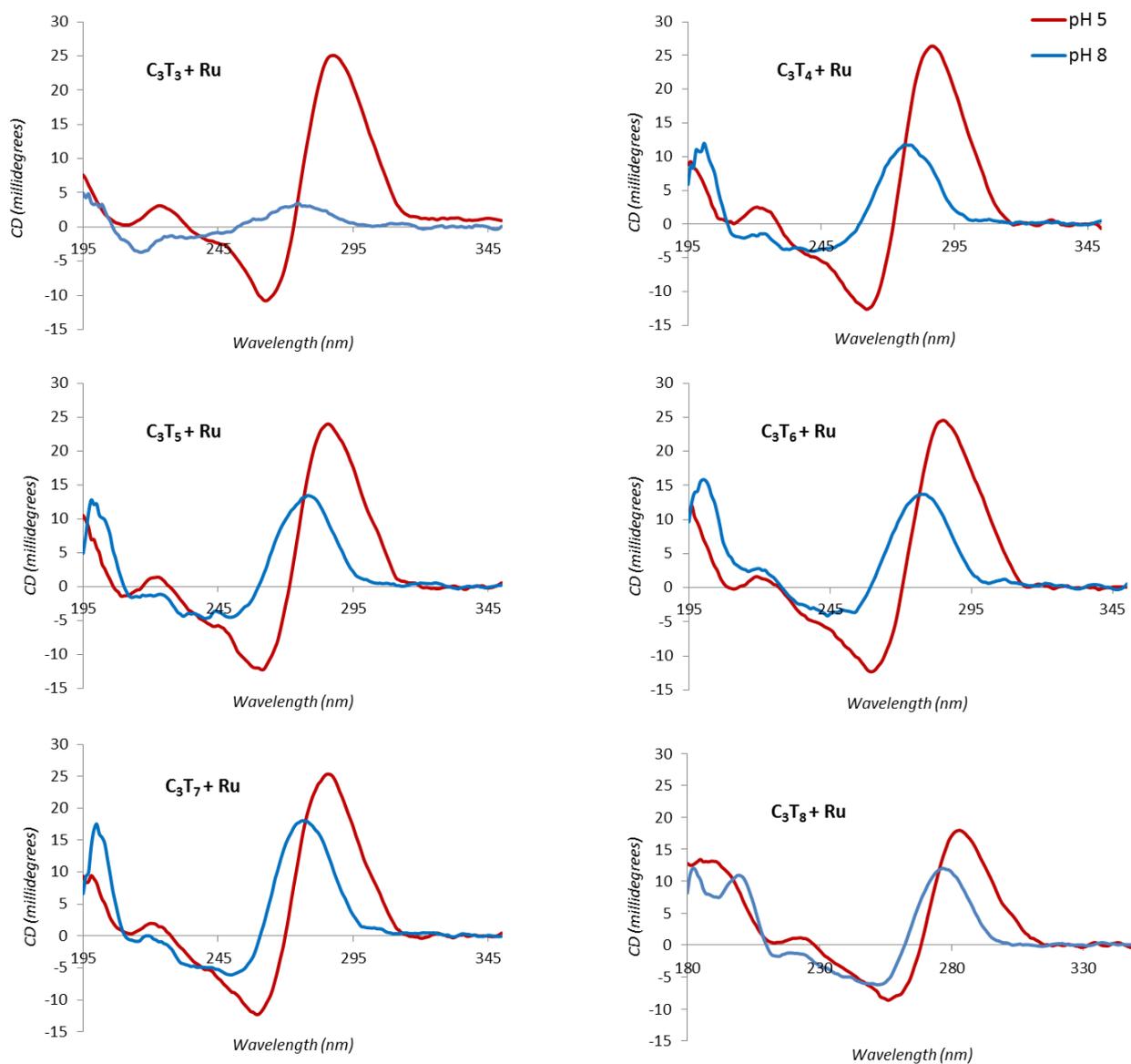
Figure A3.2: CD spectra of d(C$_3$T$_X$)$_3$C$_3$ with *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ in 20 mM sodium cacodylate buffer at pH 5 and 8.

_____
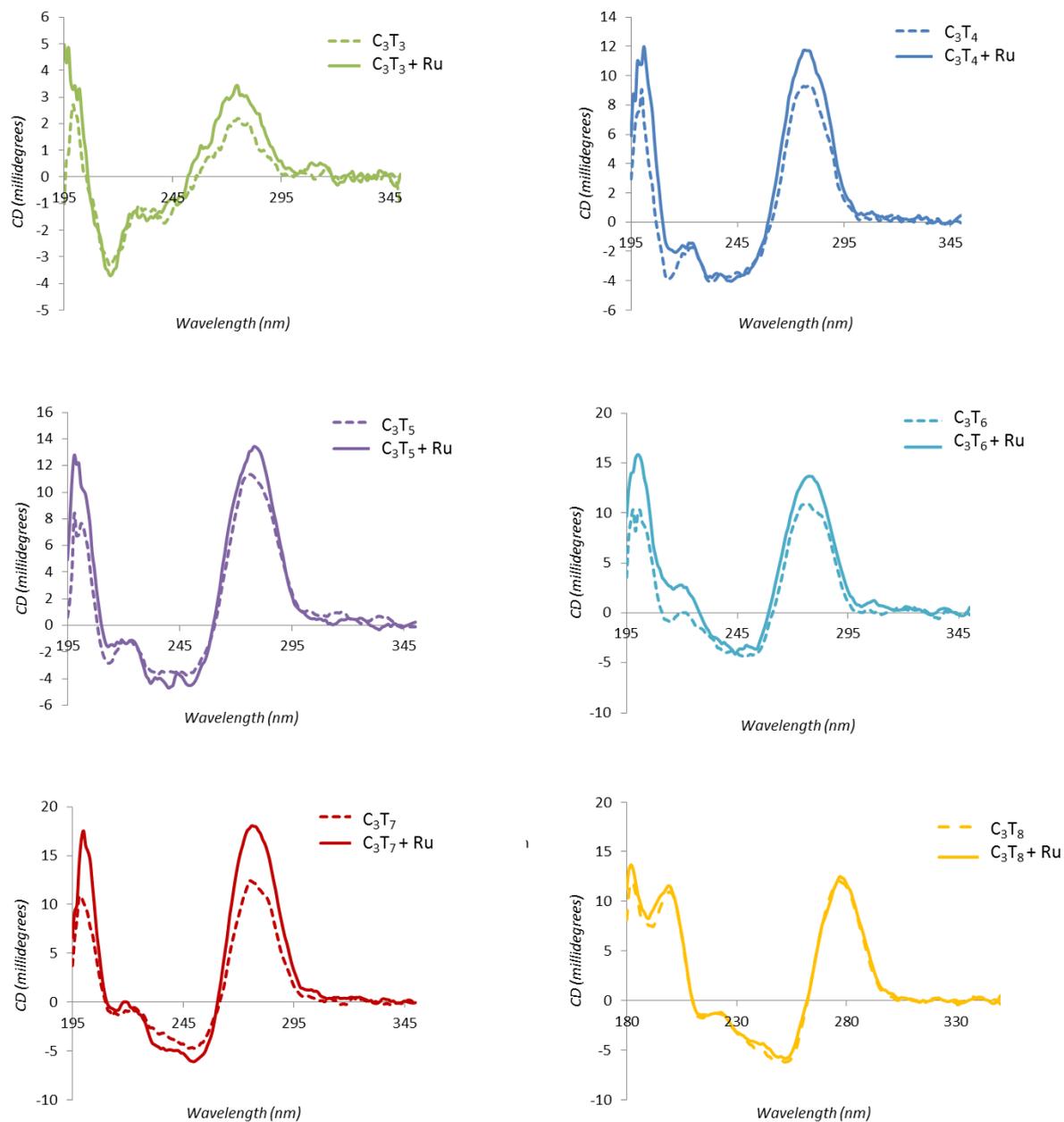
Figure A3.3: CD spectra of d(C$_3$T$_X$)$_3$C$_3$ with *rac*-[Ru(phen)$_2$(dppz)]$^{2+}$ in 20 mM sodium cacodylate buffer at pH 8.

_____
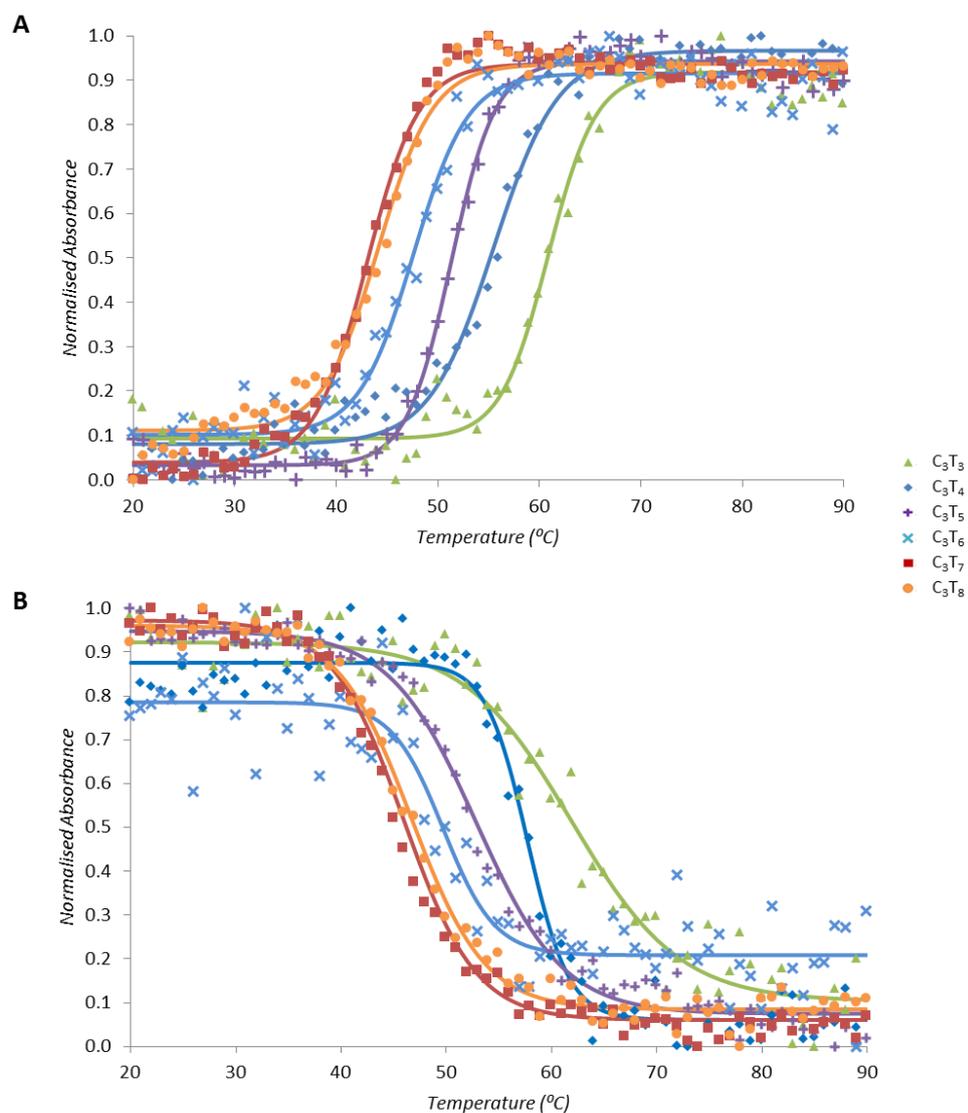
Figure A3.4: Melting curves of the six i-motif forming C_3T_x (x = 3 to 8) sequences with _rac_-[Ru(phen)$_2$(dppz)]$^{2+}$ in 50 mM sodium cacodylate (pH 5) at (A) 260 nm and (B) 295 nm wavelengths.

_____

A. C₃T₃
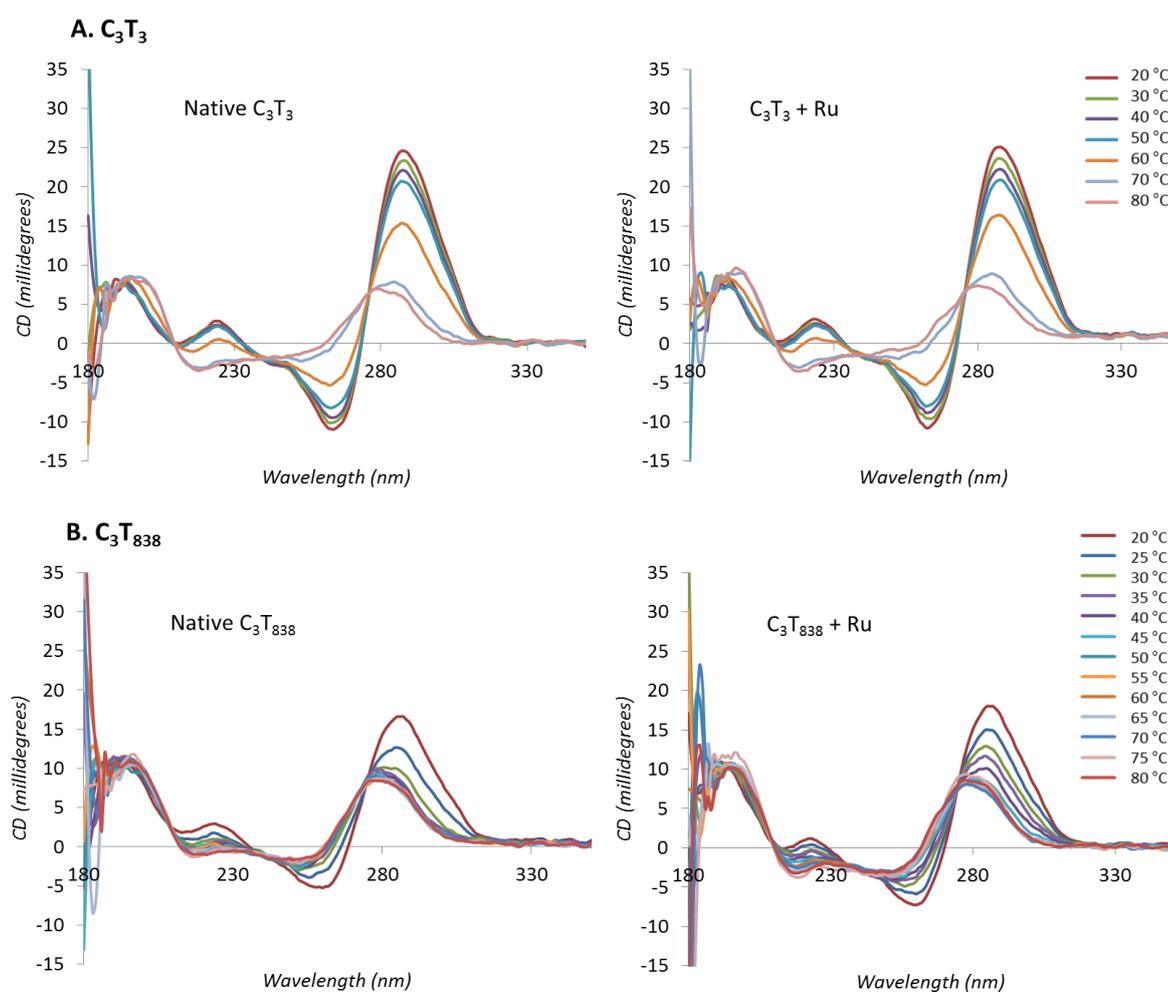
B. C₃T₈₃₈

Figure A3.5: CD melts of (top) native C₃T₃ and (bottom) the oligo in the presence of *rac*-[Ru(phen)₂(dppz)]²⁺ in 50 mM sodium cacodylate (pH 5).

_____
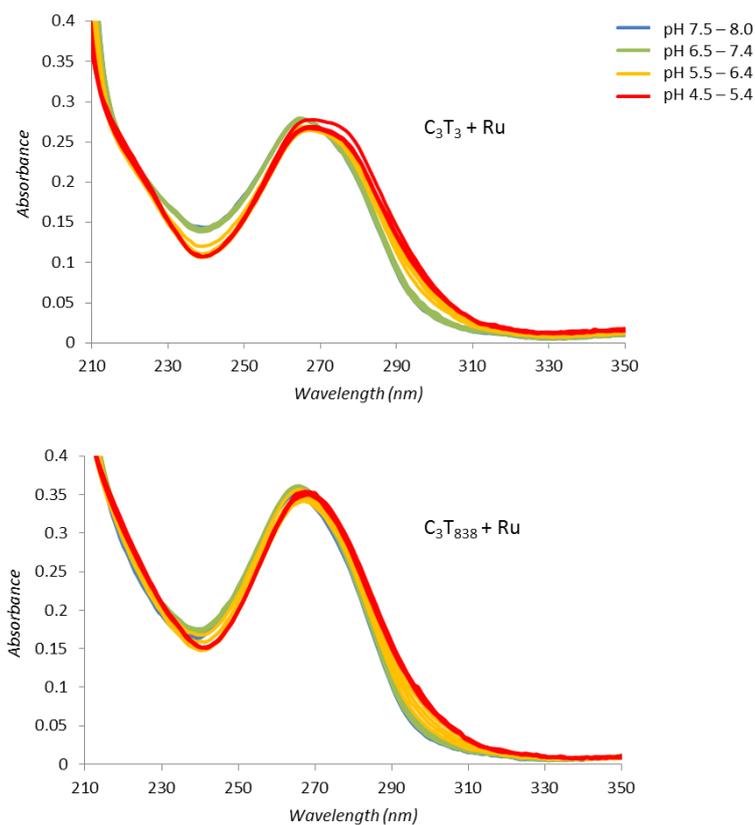
Figure A3.6: pH titration shift of (top) C_3T_3 and (bottom) C_3T_{838} with _rac_-[Ru(phen)$_2$(dppz)]$^{2+}$ in 20 mM sodium cacodylate.

A4. Chapter 4
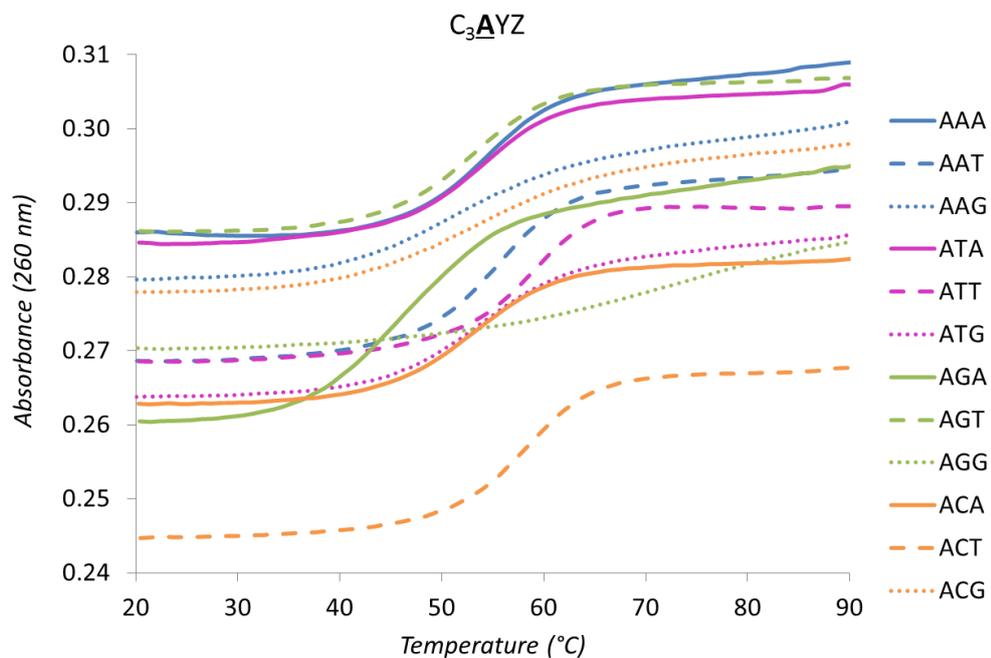
Figure A4.1: DNA melting profiles (RAW data) of $C_3\underline{A}YZ$ (1 µM, ss) where A is adenine, Y is either one of the four DNA bases, and Z is adenine, guanine or thymine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 260 nm.

_____
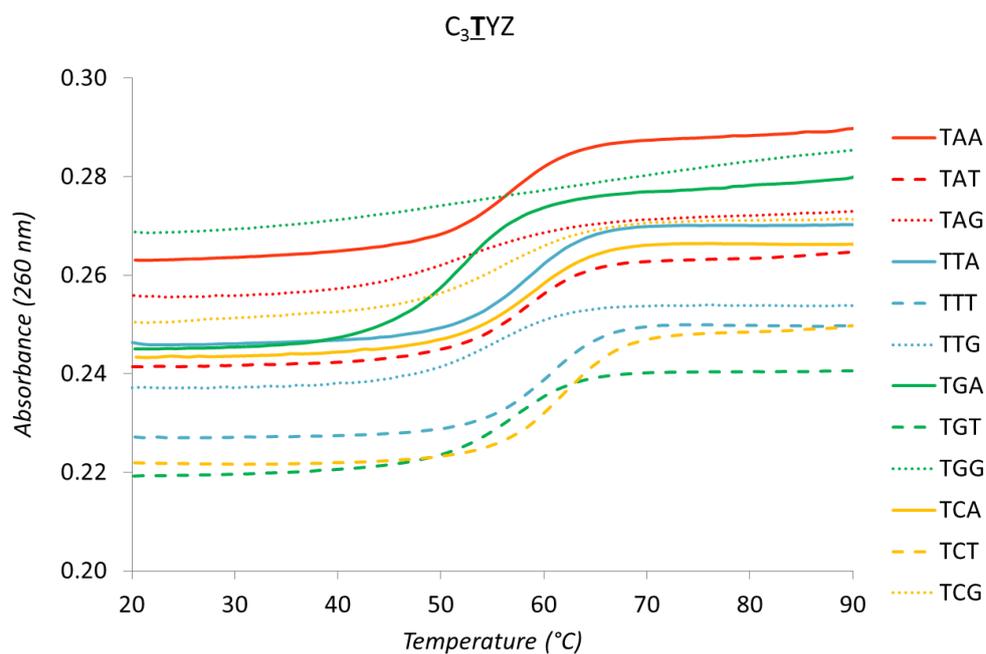
Figure A4.2: DNA melting profiles (RAW data) of C$_3$**T**YZ (1 µM, ss) where T is thymine, Y is either one of the four DNA bases, and Z is adenine, guanine or thymine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 260 nm.
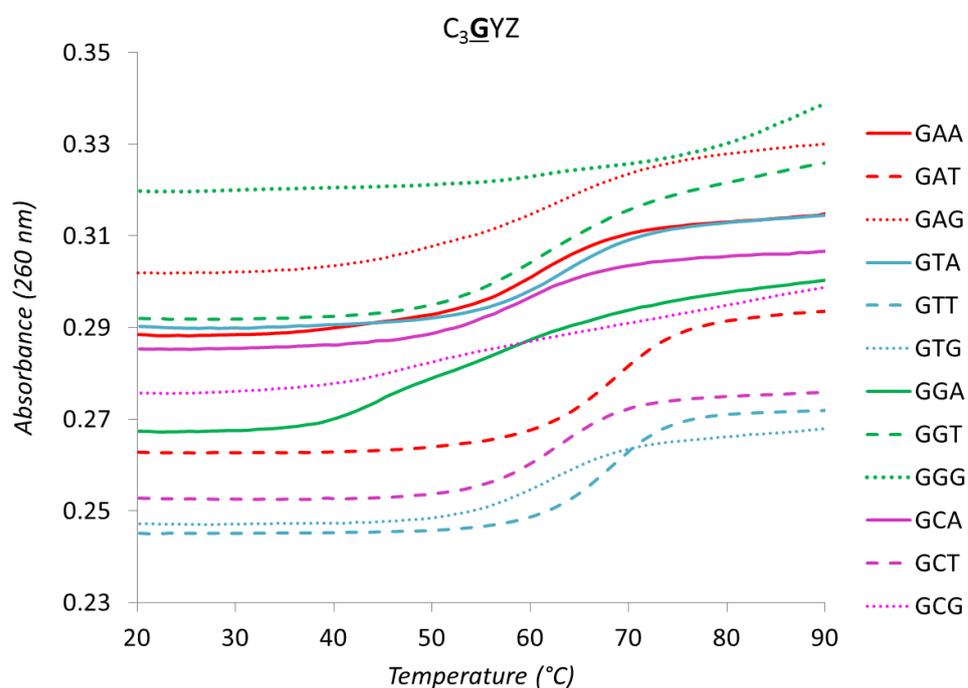
Figure A4.3: DNA melting profiles (RAW data) of C$_3$**G**YZ (1 µM, ss) where G is guanine, Y is either one of the four DNA bases, and Z is adenine, guanine or thymine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 260 nm.

_____
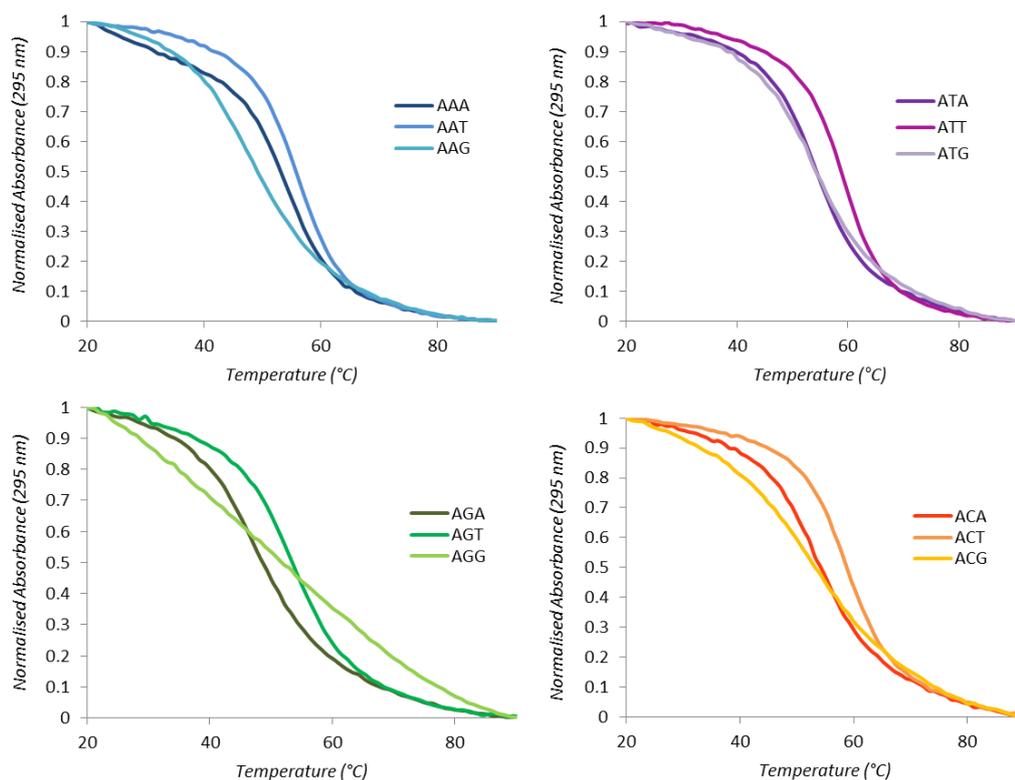
Figure A4.4: DNA melting profiles of C$_3$*A*YZ (1 μM, ss) showing the effect of varying the third base in the loop (Z) on thermal stability when Y is is (A) adenine, (B) thymine, (C) guanine and (D) cytosine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 295 nm.
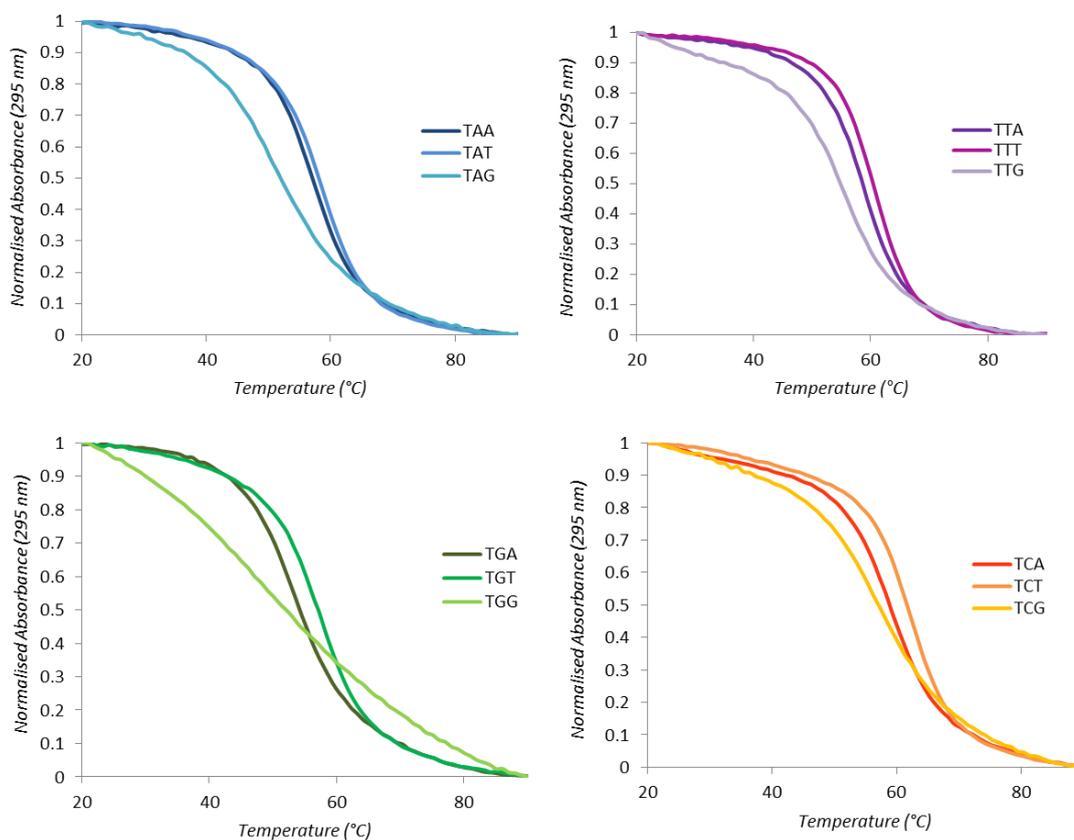
Figure A4.5: DNA melting profiles of C$_3$*T*YZ (1 μM, ss) showing the effect of varying the third base in the loop (Z) on thermal stability when Y is is (A) adenine, (B) thymine, (C) guanine and (D) cytosine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 295 nm.

_____
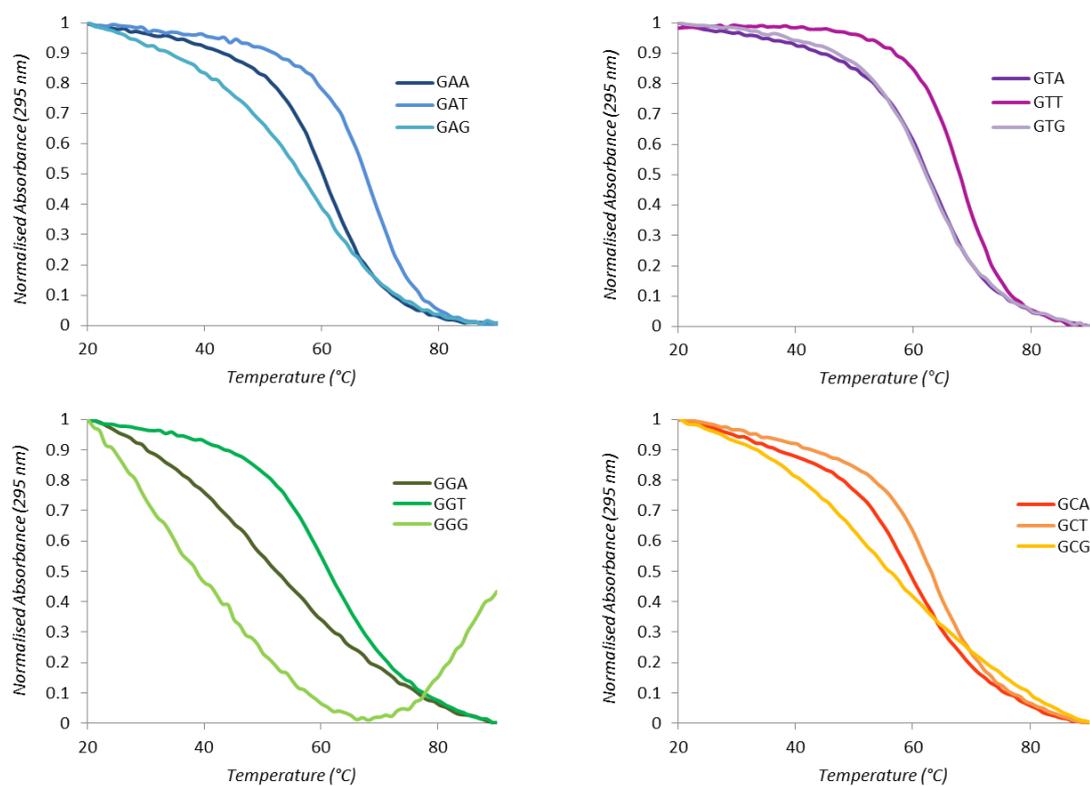
Figure A4.6: DNA melting profiles of C$_3$**G**YZ (1 µM, ss) showing the effect of varying the third base in the loop (Z) on thermal stability when Y is is (A) adenine, (B) thymine, (C) guanine and (D) cytosine. Samples were made in 50 mM sodium cacodylate buffer (pH 5) and the absorbance recorded at λ = 295 nm.

_____
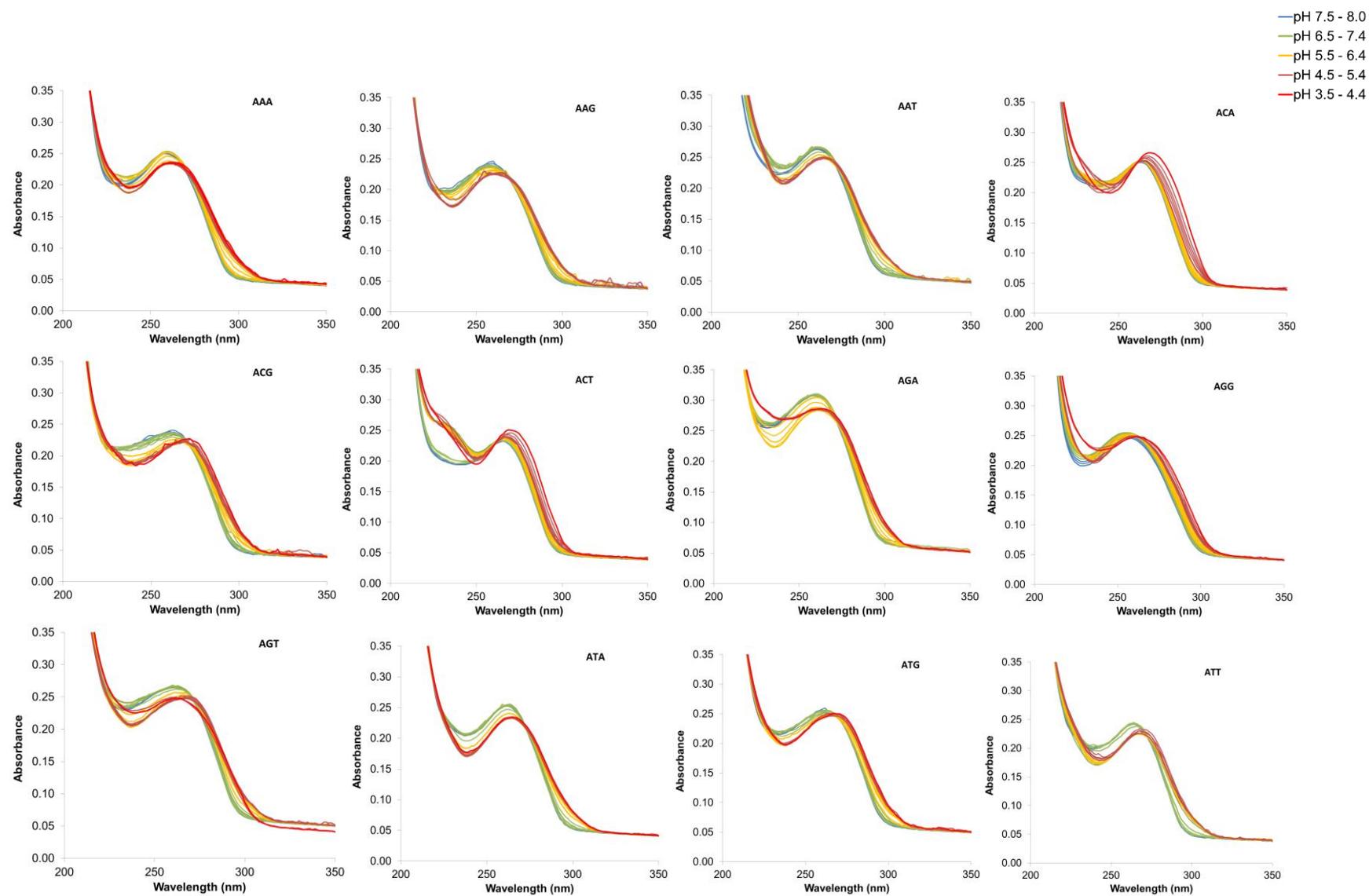
Figure A4.7: pH titration profiles of C$_3$$A$YZ, 25°C.

_____
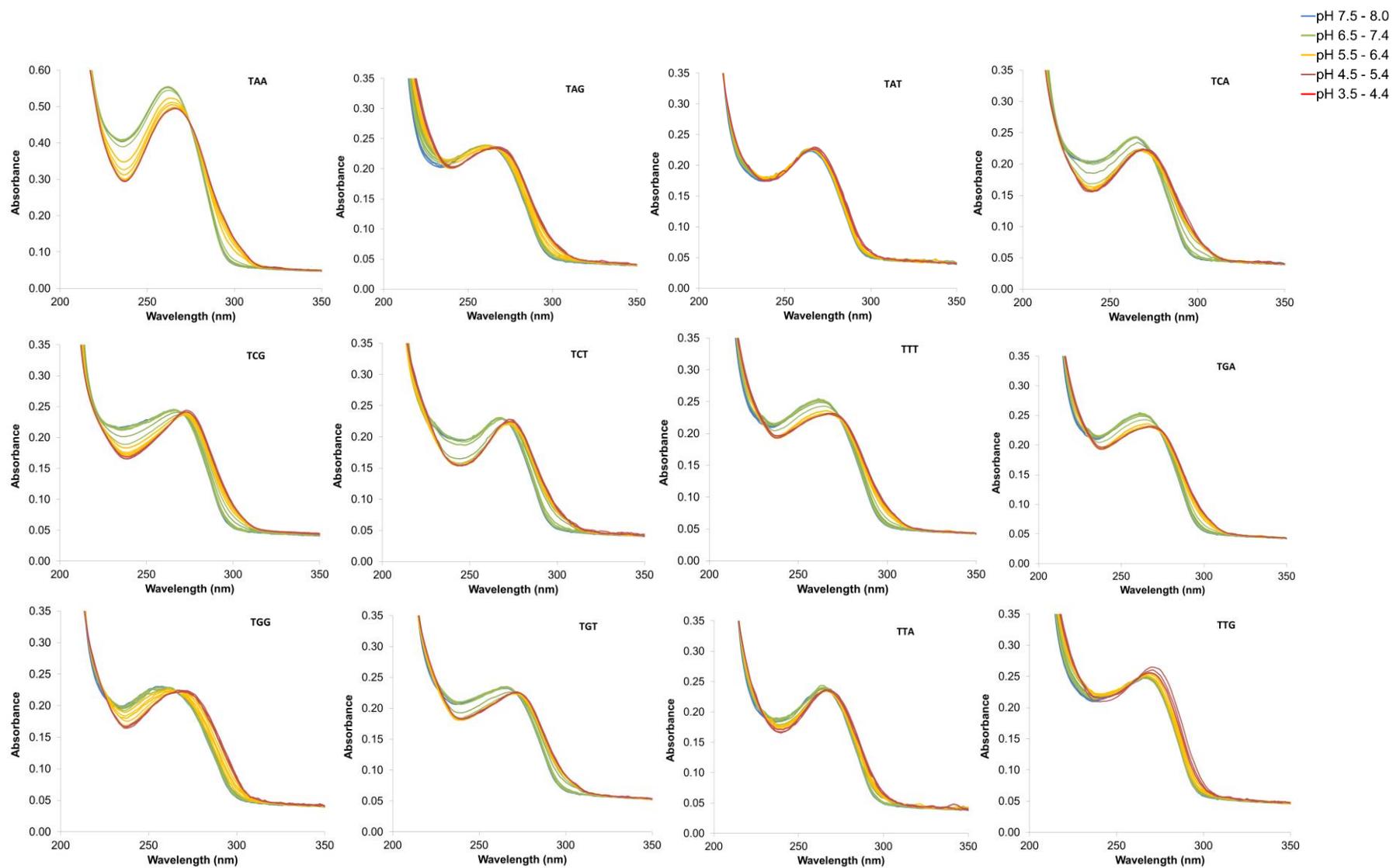
Figure A4.8: pH titration profiles of C$_3$*T*YZ, 25°C.

_____
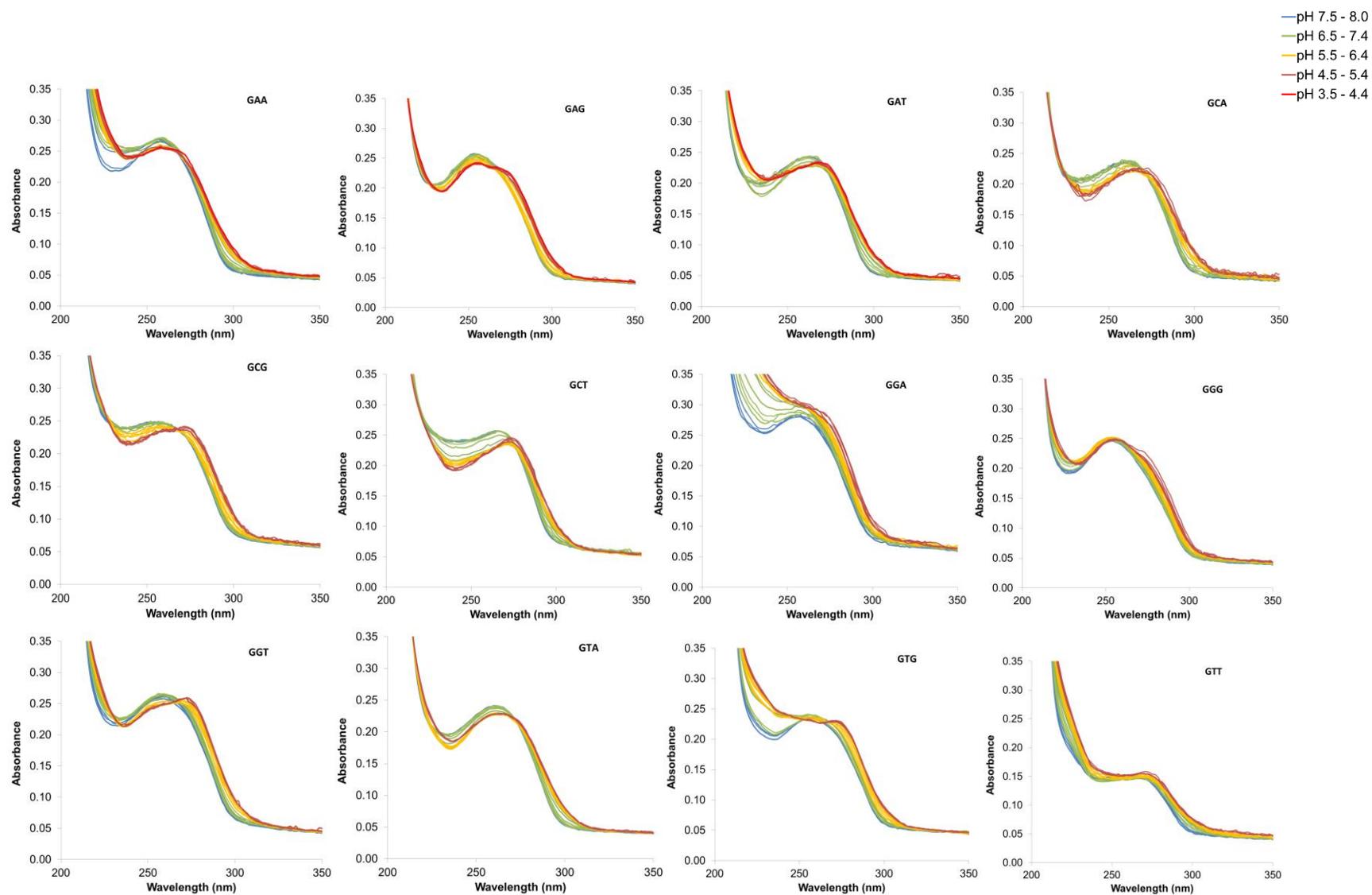
Figure A4.9: pH titration profiles of C$_3$**G**YZ, 25°C.
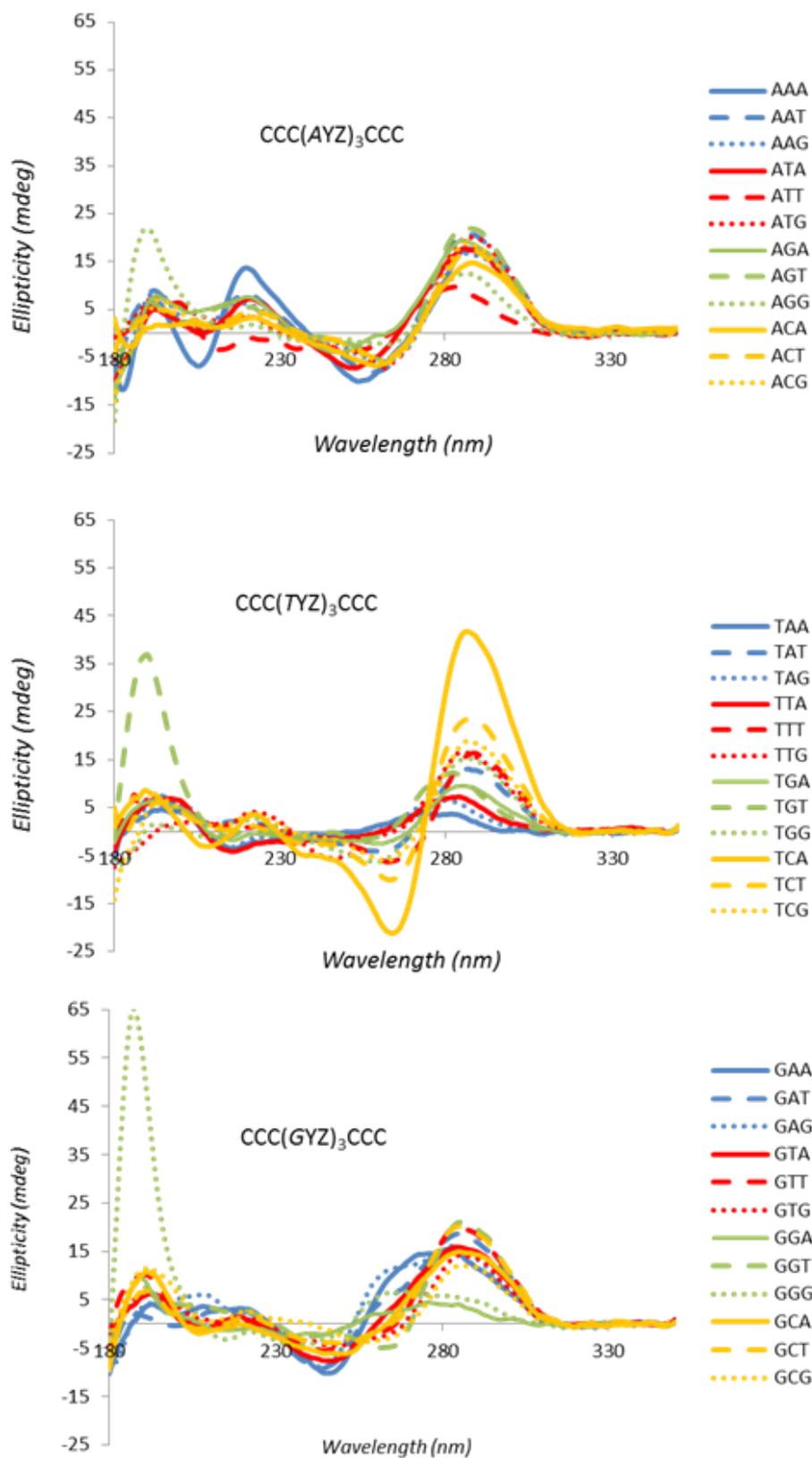
Figure A4.10: Synchrotron Radiation Circular Dichroism spectra of C_3XYZ oligonucleotides in pH 5, recorded between 180 to 350 nm wavelengths at 20°C.
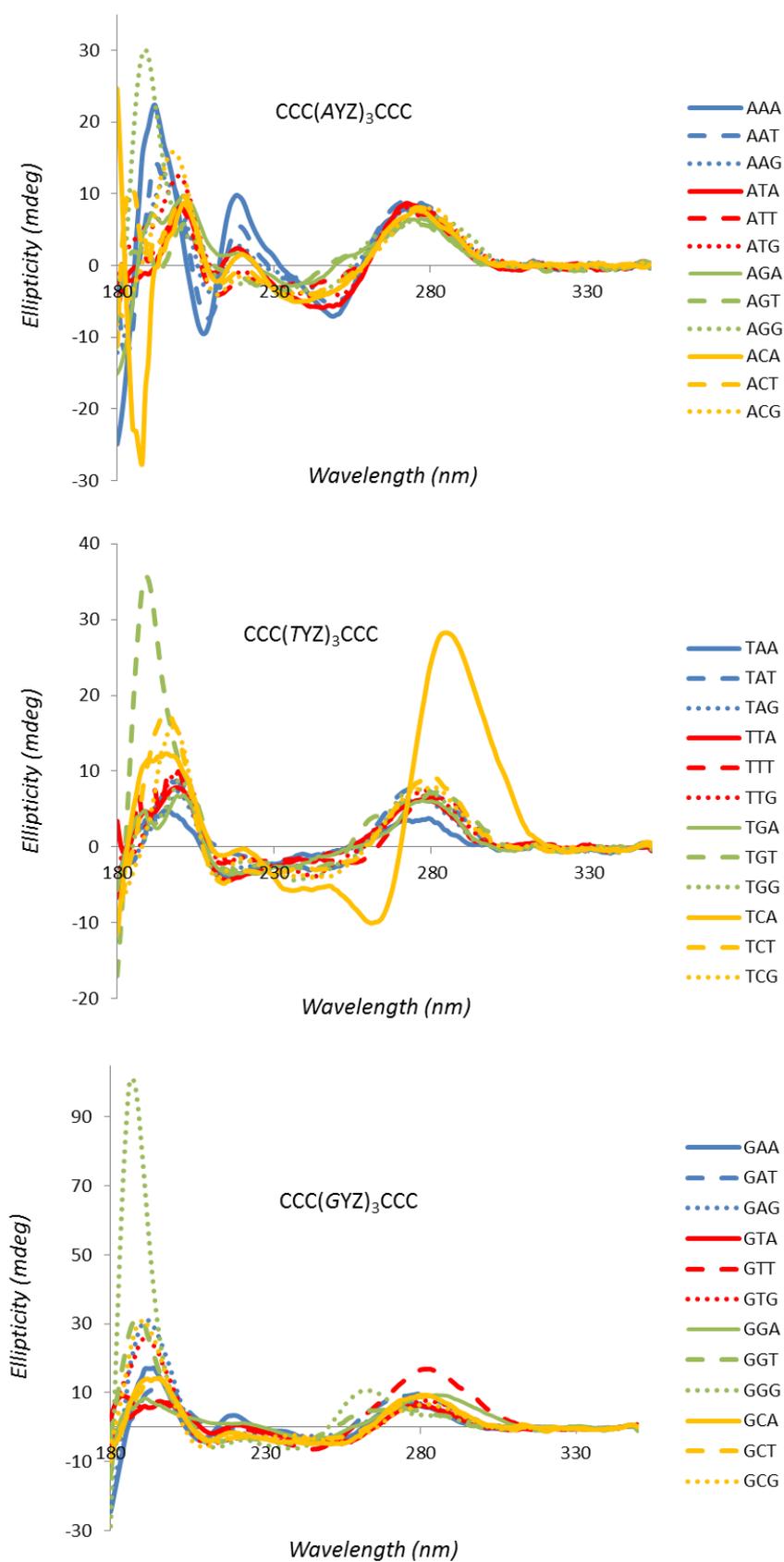
Figure A4.11: Synchrotron Radiation Circular Dichroism spectra of C$_3$XYZ oligonucleotides in pH 8, recorded between 180 to 350 nm wavelengths at 20°C.

A5. Chapter 5
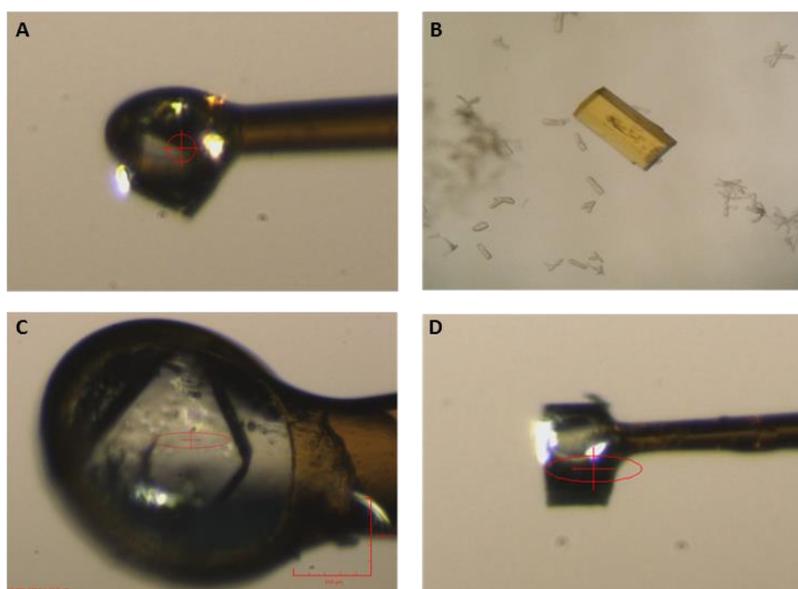
Figure A5.1: Crystals of (**A**) C_4, (**B**) C_3T, (**C**) C_3A and (**D**) [C_3T + C_3A].
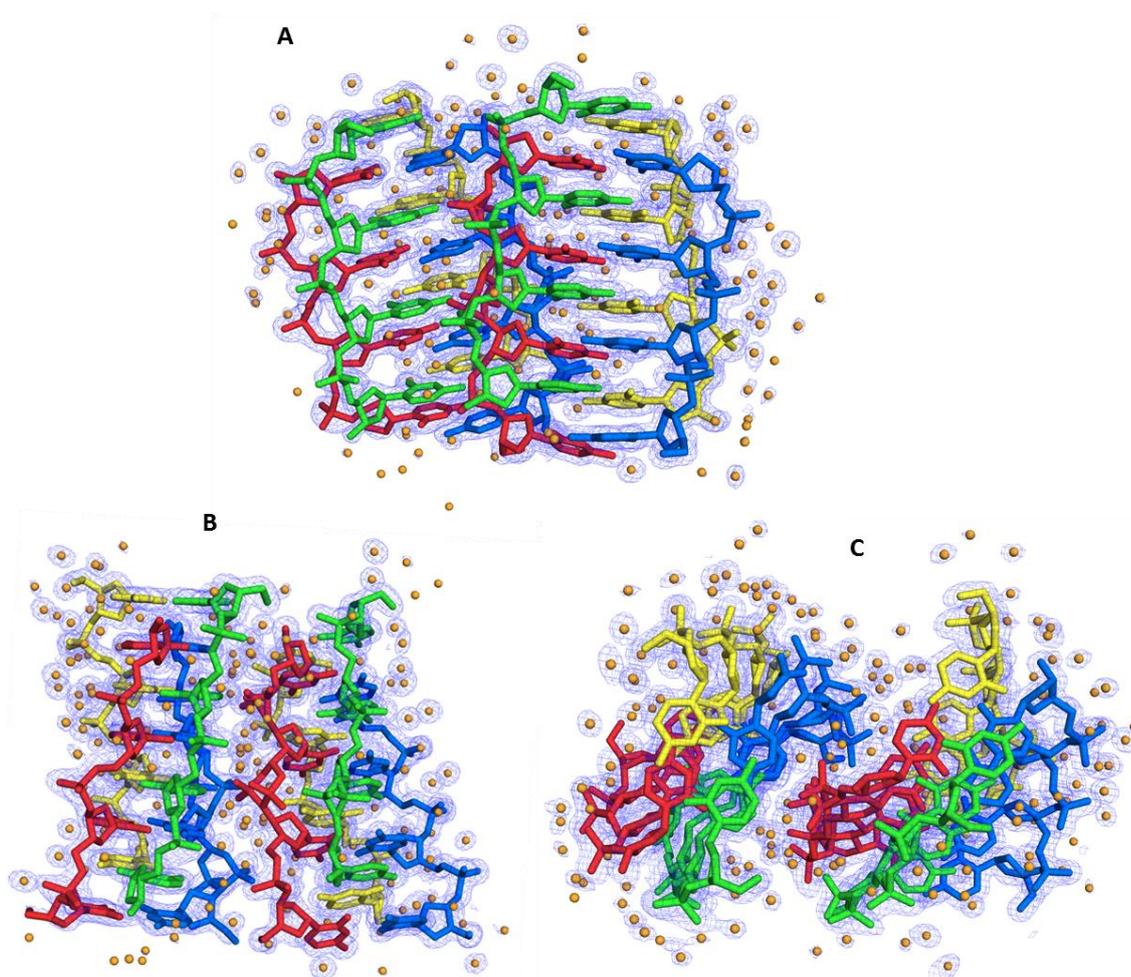
Figure A5.2: Views of the asymmetric unit of C_4 looking into the (**A**) wide and (**B**) narrow grooves and (**C**) down the helical axes. The 2Fo-Fc electron density map is drawn in blue at the 1σ contour level. The chains coloured green base pairs with the ones in yellow and the blue chains base pair with red. The orange spheres represent water molecules.

_____
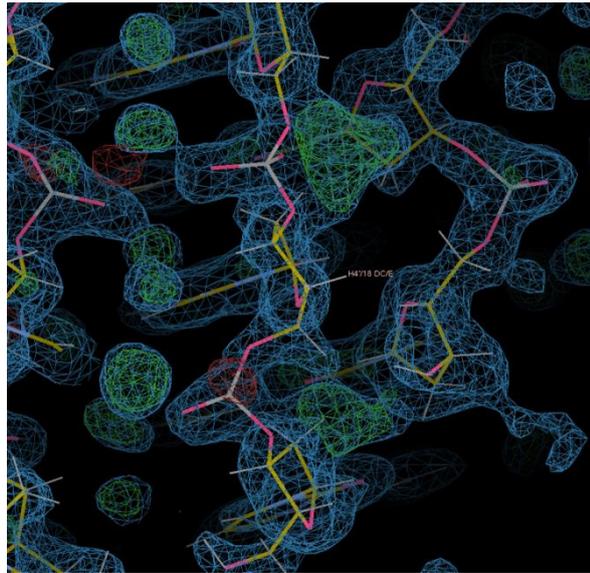
Figure A5.3: 2Fo-Fc (blue) and Fo-Fc (green and red) electron density maps of C_4, drawn at 1σ and 3σ contour levels, respectively. The Fo-Fc map suggests alternative phosphate conformations in chain E. Carbon is yellow, oxygen is pink, phosphorus is grey and hydrogen in white.
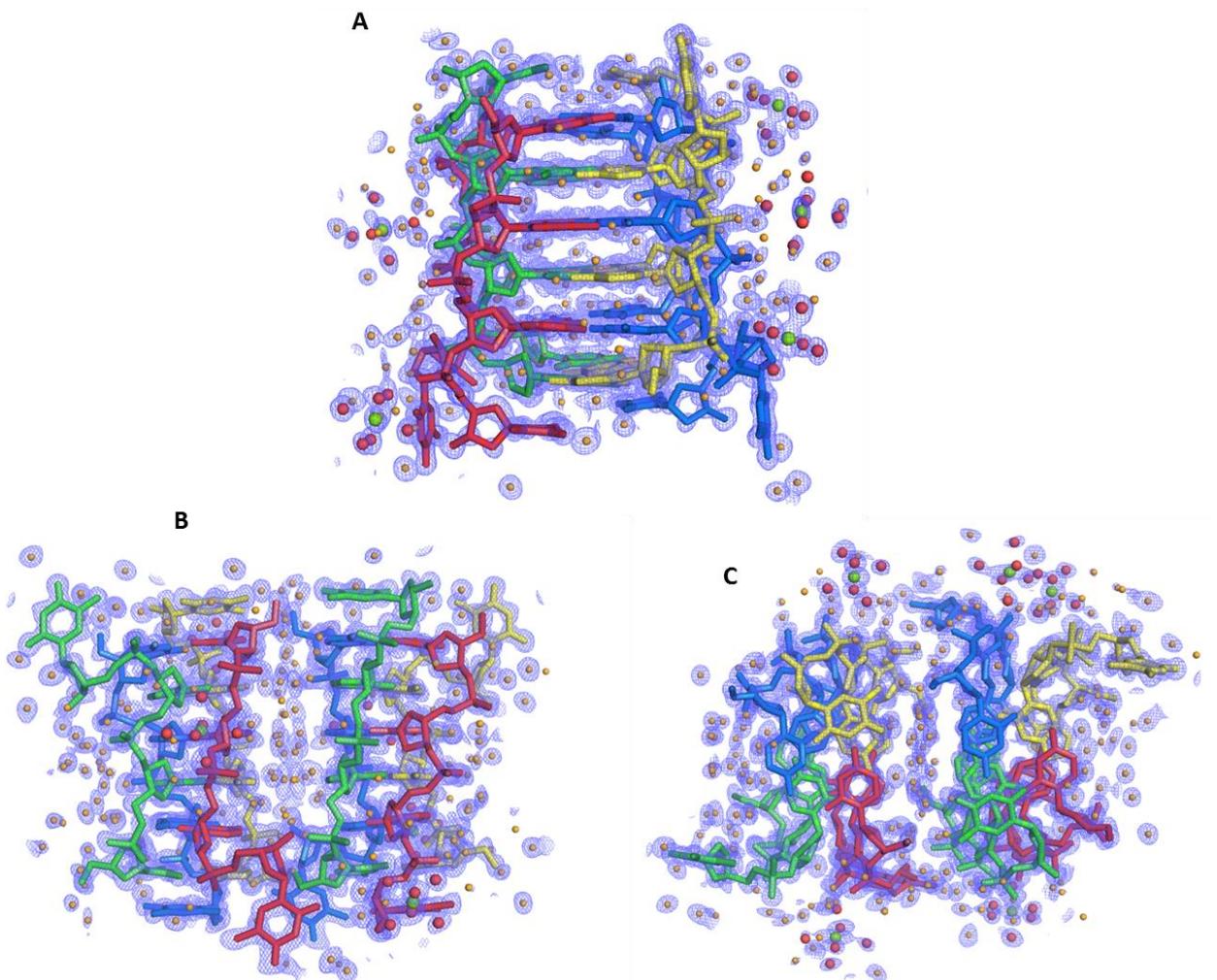
Figure A5.4: Views of the asymmetric unit of C_3T looking into the (**A**) wide and (**B**) narrow grooves and (**C**) down the helical axes. The 2Fo-Fc electron density map is drawn in blue at the 1σ contour level.

_____
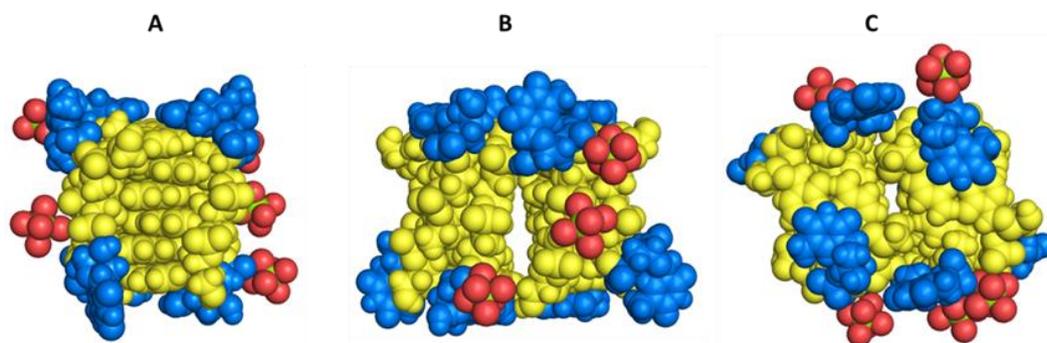
Figure A5.5: Van der Waals model of the asymmetric unit of C_3T as seen through the (**A**) major and (**B**) minor grooves. (**C**) View of the i-motif down the helical axes. Yellow spheres are cytosine residues, blue are thymines and green + red are $[Mg(H_2O)6]^{2+}$ molecules.
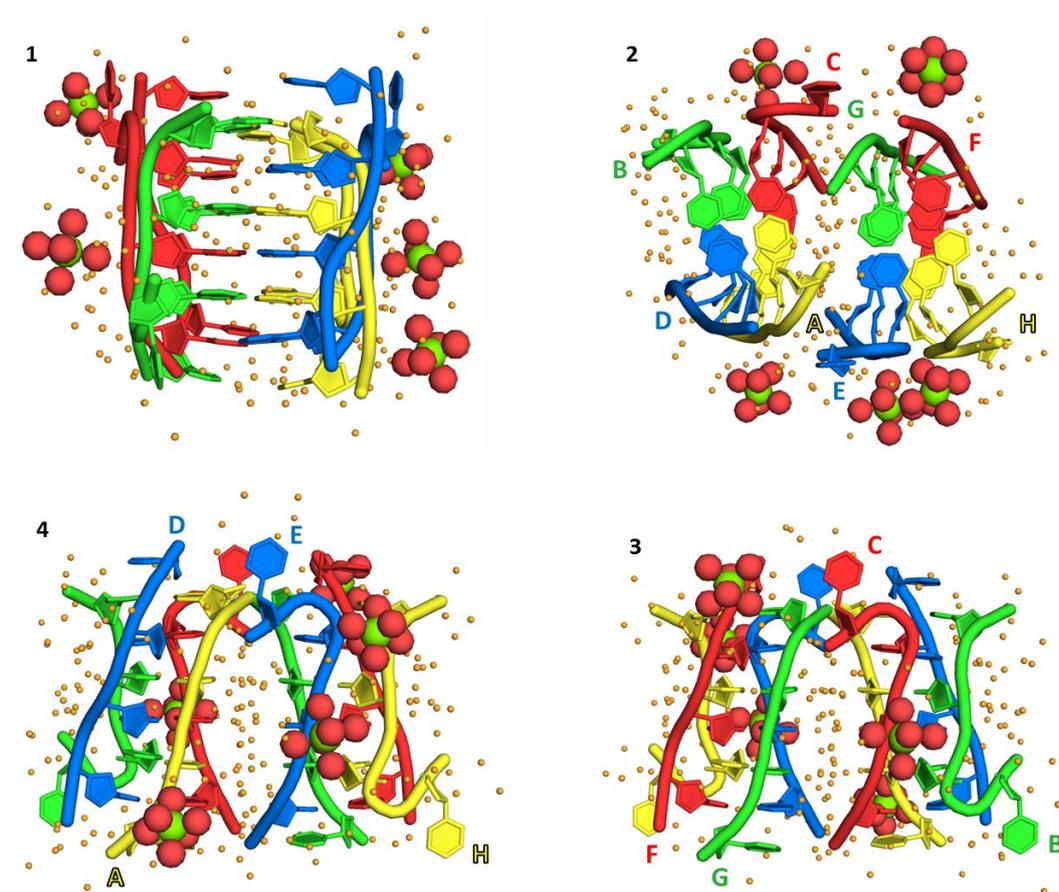
Figure A5.6: Views of the asymmetric unit of C_3T looking (1) into the wide groove, (2) down the helical axes and (3, 4) through the narrow grooves. The eight strands have been labelled A to H with chain A base pairing with B, C-D, E-F and G-H. The orange and red spheres represent water molecules and the green spheres are Mg^{2+}.
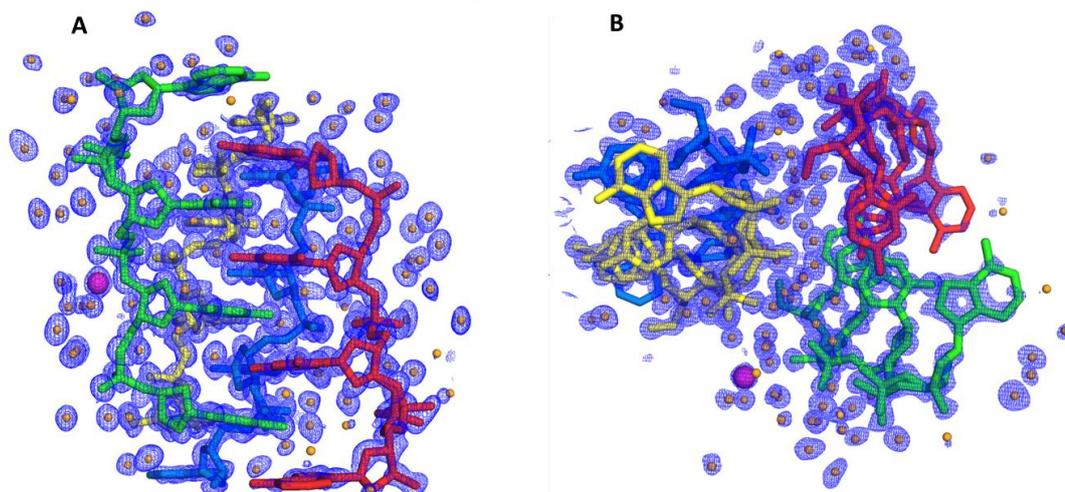
Figure A5.7: Views of the asymmetric unit of C₃A (**A**) along and (**B**) down the helical axes. The 2Fo-Fc electron density map is drawn in blue at the 1σ contour level. Orange spheres represent water and the purple sphere is K^+.
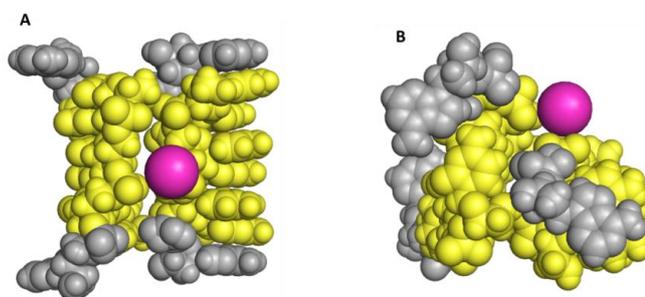
Figure A5.8: Van der Waals representation of the asymmetric unit of C₃A (**A**) along and (**B**) down the helical axes. The yellow spheres are cytosine, grey are adenine and purple is potassium.
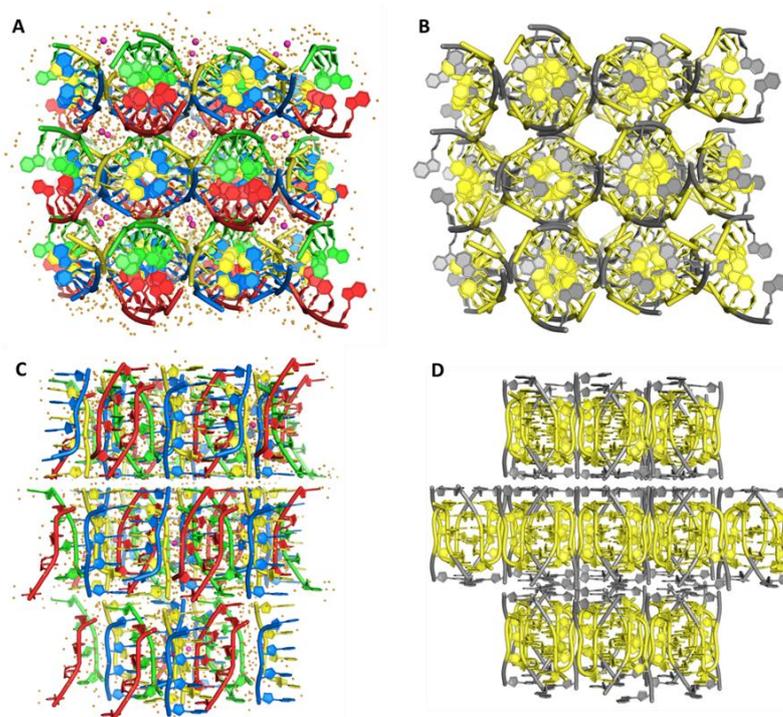
Figure A5.9: Lattice of C₃A as seen down the a and b axes. (**A, C**) Each of the four strands are coloured individually to show how the chains are packed. The orange dots are water molecules. (**B, D**) Cytosine residues are represented in yellow and adenine in grey.

_____
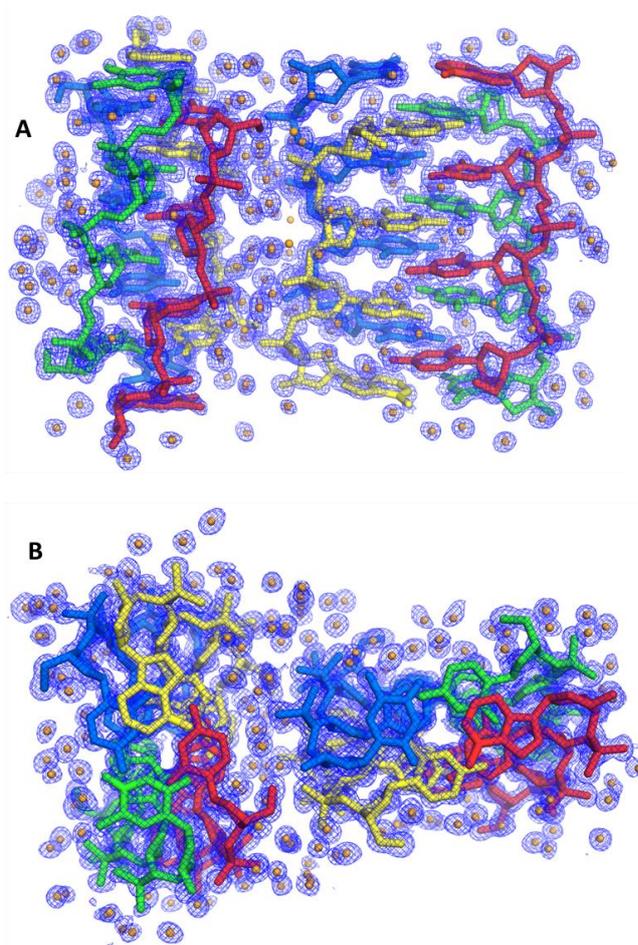
Figure A5.10: Views of the asymmetric unit of C_3A + C_3T (**A**) along and (**B**) down the helical axes. The 2Fo-Fc electron density map is drawn in blue at the 1σ contour level. Orange spheres represent water molecules.
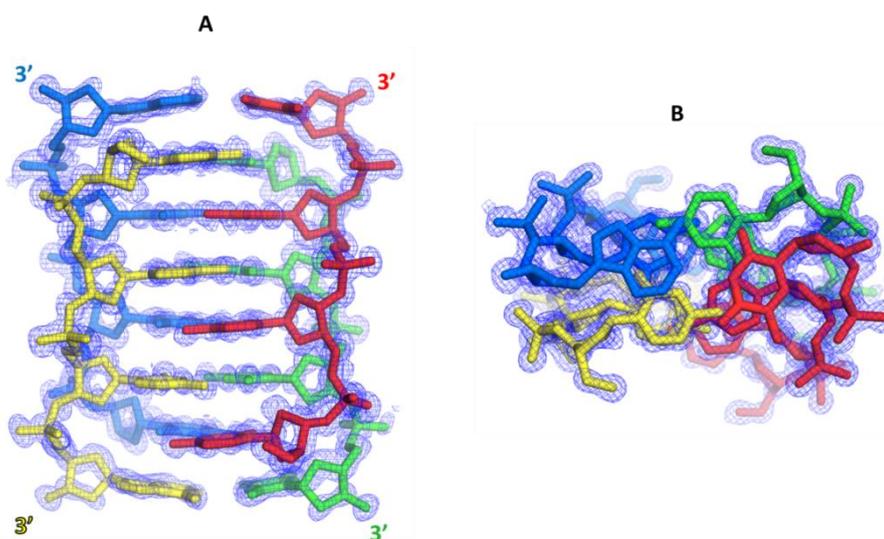
Figure A5.11: A single [C_3A + C_3T] i-motif surrounded by electron density map; drawn in blue at 1σ contour level. Views of the i-motif (**A**) along and (**B**) down the helical axis.

_____
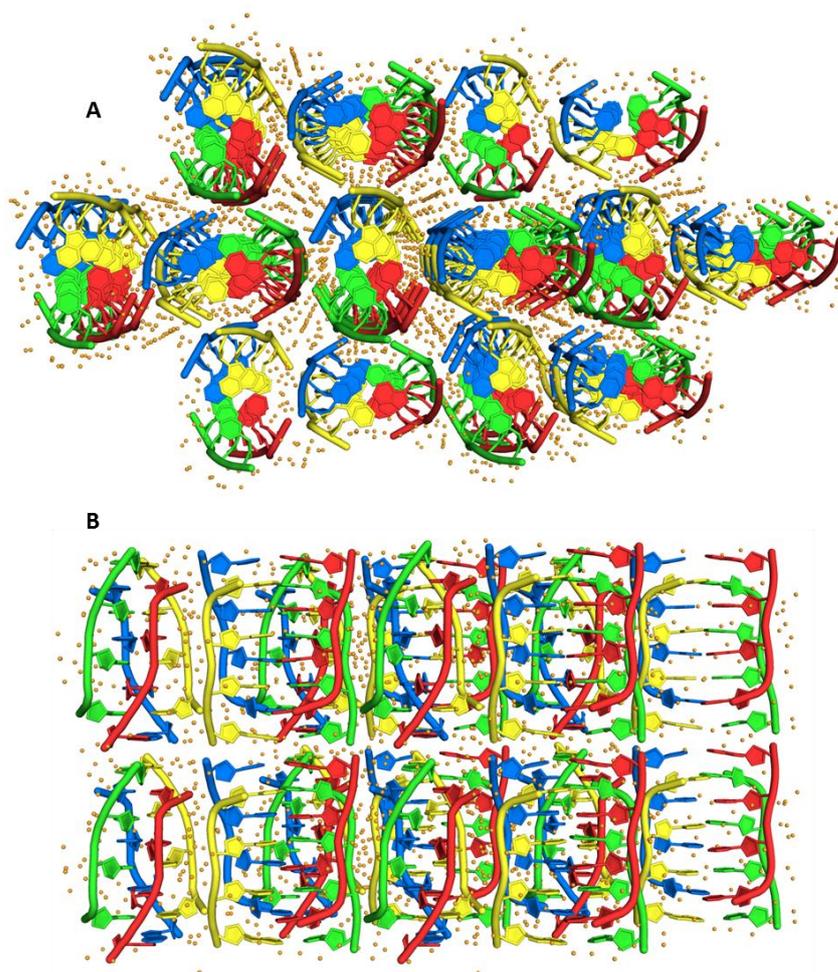
Figure A5.12: Lattice of [C$_3$A + C$_3$T] as seen down the (**A**) a and (**B**) b axes. The orange dots are water molecules.

_____

Table A5.1: Sugar pucker pesudorotation angles, sugar pucker type and closest DNA conformation for C_4.
C_X is the cytosine position in 5'-$C_1C_2C_3C_4$-3'.

Chain	Base	Phase angle of pseudorotation	Sugar Pucker	Closest DNA conformation
A	C_1	31.0	C3'-endo	A-DNA
	C_2	42.0	C4'-exo	-
	C_3	148.0	C2'-endo	B-DNA
	C_4	65.2	C4'-exo	-
B	C_1	164.7	C2'-endo	B-DNA
	C_2	72.0	O4'-endo	-
	C_3	54.8	C4'-exo	-
	C_4	51.7	C4'-exo	-
C	C_1	160.1	C2'-endo	B-DNA
	C_2	53.6	C4'-exo	-
	C_3	43.7	C4'-exo	-
	C_4	45.6	C4'-exo	-
D	C_1	25.4	C3'-endo	A-DNA
	C_2	31.4	C3'-endo	A-DNA
	C_3	37.6	C4'-exo	-
	C_4	166.0	C2'-endo	B-DNA
E	C_1	-6.5	C2'-exo	-
	C_2	63.0	C4'-exo	-
	C_3	117.6	C1'-exo	-
	C_4	91.4	O4'-endo	-
F	C_1	164.4	C2'-endo	B-DNA
	C_2	70.5	C4'-exo	-
	C_3	62.5	C4'-exo	-
	C_4	43.9	C4'-exo	-
G	C_1	158.6	C2'-endo	B-DNA
	C_2	90.4	O4'-endo	-
	C_3	58.7	C4'-exo	-
	C_4	39.4	C4'-exo	-
H	C_1	28.9	C3'-endo	A-DNA
	C_2	38.4	C4'-exo	-
	C_3	40.4	C4'-exo	-
	C_4	54.6	C4'-exo	-

Table A5.2: Conformational analysis of d(C$_4$)$_4$ as calculated by DNATCO version 2.2.

step_ID	NtC	CANA	δ	ε	ζ	α1	β1	γ1	δ1	χ	χ1	confal	nearest_NtC	Δδ	Δε	Δζ	Δα1	Δβ1	Δγ1	Δδ1	Δχ	Δχ1	sδ	sε	sζ	sα1	sβ1	sγ1	sδ1	sχ	sχ1
A_DC1_DC2	AA02	AAA	84.3	200	278.3	302.2	178	60.6	78.6	230	237.4	76	AA02	-3.4	-2.5	3.8	8.7	16.7	6.4	-8.9	-13.9	-7.3	94.1	98.6	97.2	74.6	46.6	73	61.7	55.7	83.6
A_DC2_DC3	NANT	NAN	78.6	181.7	280.7	193.7	179.4	184.2	136.4	237.4	235.8	0	BB03	-65.9	5.5	5.2	28.2	15.2	11.4	-9.6	-1.7	4	0	90.3	91.8	17.4	56.1	66.6	66.9	99.2	95.2
A_DC3_DC4	NANT	NAN	136.4	212.5	288.6	133.3	165.8	179.8	78.5	235.8	246.7	0	BB03	-8.1	36.3	13.1	-32.2	1.6	7	-67.5	-3.3	14.9	68	1.2	58	10.3	99.4	85.8	0	97.1	50.9
B_DC5_DC6	NANT	NAN	148.9	242.6	267	81.5	200.8	202.7	85	239.1	243.2	0	BA10	13.7	42.9	48.9	-23.3	-28.6	8.3	-2.9	-18.7	48.2	62.5	0	1.9	44.3	20.8	86.4	96.3	46.1	0
B_DC6_DC7	NANT	NAN	85	191.8	271.5	163	168.1	193	91.7	243.2	233.6	0	AA01	2.8	-2.2	-19.3	14.1	-24.6	11.2	4.6	38.2	45.4	90	98.5	13.9	58.5	7.7	50.3	75.2	0	0
B_DC7_DC8	NANT	NAN	91.7	135.7	36.9	148.4	127.8	160.5	79.1	233.6	230.3	0	AA01	9.5	-58.3	106.1	-0.5	-64.9	-21.3	-8	28.6	42.1	29.7	0	0	99.9	0	8.3	42.3	0	0
C_DC9_DC10	NANT	NAN	147.1	280.3	233.1	62.1	225.3	199.3	77.8	237.9	235.7	0	BA13	5	48.9	37.6	-12.1	-6.2	3.3	-11.5	-28.3	36.5	88.4	0	4.4	70	90.2	97.4	59.9	10.3	0
C_DC10_DC11	AA02	AAA	77.8	190.2	275.7	289.2	177.9	69.7	71.2	235.7	238.9	64	AA02	-9.9	-12.3	1.2	-4.3	16.6	15.5	-16.3	-8.2	-5.8	59.5	70.2	99.7	93.1	47	15.8	19.8	81.6	89.3
C_DC11_DC12	NANT	NAN	71.2	181.2	277.7	191.6	199.5	163.7	81.1	238.9	232.2	0	AA01	-11	-12.8	-13.1	42.7	6.8	-18.1	-6	33.9	44	19.6	59.3	40.3	0	82.2	16.6	61.6	0	0
D_DC13_DC14	AA02	AAA	82.4	193.9	277.1	300.3	175.5	62.9	80	229.1	232.9	72	AA02	-5.3	-8.6	2.6	6.8	14.2	8.7	-7.5	-14.8	-11.8	86.2	84.1	98.7	83.6	57.6	56	71	51.5	62.6
D_DC14_DC15	AA02	AAA	80	190.7	282	298.7	185.7	58.8	80.6	232.9	237.6	73	AA02	-7.7	-11.8	7.5	5.2	24.4	4.6	-6.9	-11	-7.1	73.1	72.2	89.4	90.1	19.6	85	74.8	69.3	84.4
D_DC15_DC16	NANT	NAN	80.6	194.6	291	245.8	134.2	162.7	146.2	237.6	237.8	0	BB13	-62.6	8.8	0.4	29.4	30.7	1.8	-0.5	-14.3	18.9	0	75.1	100	5.3	4.6	98.7	99.9	46.4	26.3
E_DC17_DC18	AA02	AAA	104.2	228.6	280.9	306.5	166.8	62.6	77.7	227.9	237.2	58	AA02	16.5	26.1	6.4	13	5.5	8.4	-9.8	-16	-7.5	23.7	20.4	92.2	52	92	58.2	55.7	46.1	82.8
E_DC18_DC19	NANT	NAN	77.7	178.9	282.2	195.1	188.9	173.5	118.6	237.2	235.3	0	AA01	-4.5	-15.1	-8.6	46.2	-3.8	-8.3	31.5	32.2	47.1	76.1	48.3	67.6	0	94.1	68.6	0	0	0
E_DC19_DC20	NANT	NAN	118.6	211	293.2	162.7	141.5	175.9	103.4	235.3	238	0	BB03	-25.9	34.8	17.7	-2.8	-22.7	3.1	-42.6	-3.8	6.2	1.9	1.7	37	98.3	27.5	97	0	96.2	89
F_DC21_DC22	NANT	NAN	144.6	234.2	272.2	111	186.9	190.7	85.6	233	238.9	0	BA10	9.4	34.5	54.1	6.2	-42.5	-3.7	-2.3	-24.8	43.9	80.2	4.7	0	94.4	3.1	97.1	97.6	25.6	0
F_DC22_DC23	AA02	AAA	85.6	199.8	273.1	292.1	171.8	66.9	77.5	238.9	236.3	80	AA02	-2.1	-2.7	-1.4	-1.4	10.5	12.7	-10	-5	-8.4	97.7	98.3	99.6	99.2	73.9	29	54.4	92.7	78.9
F_DC23_DC24	NANT	NAN	77.5	191.3	278.6	171.3	175	176.3	87.8	236.3	228.4	0	AA01	-4.7	-2.7	-12.2	22.4	-17.7	-5.5	0.7	31.3	40.2	74.3	97.7	45.5	25.8	26.6	84.7	99.3	0	0
G_DC25_DC26	NANT	NAN	139.1	213.3	291.2	142.8	164.2	182.7	95.3	232.5	239.1	0	BB03	-5.4	37.1	15.7	-22.7	0	9.9	-50.7	-6.6	7.3	84.3	0	45.7	32.3	100	73.6	0	88.9	85
G_DC26_DC27	NANT	NAN	95.3	194.4	282	161.1	161	179.3	87.6	239.1	236.8	0	AA01	13.1	0.4	-8.8	12.2	-31.7	-2.5	0.5	34.1	48.6	9.9	99.9	66.4	66.9	1.4	96.6	99.7	0	0
G_DC27_DC28	NANT	NAN	87.6	188.7	276.3	170.4	184.3	173.2	81.9	236.8	228.9	0	AA01	5.4	-5.3	-14.5	21.5	-8.4	-8.6	-5.2	31.8	40.7	67.5	91.4	32.9	28.7	74.2	66.7	69.5	0	0
H_DC29_DC30	AA02	AAA	81	202.3	279.9	298.1	181.6	62.4	80	229	233.2	72	AA02	-6.7	-0.2	5.4	4.6	20.3	8.2	-7.5	-14.9	-11.5	78.9	100	94.4	92.1	32.3	59.7	71	51.1	64.1
H_DC30_DC31	AA02	AAA	80	194.6	282.3	297.7	179.5	60.9	79.1	233.2	237.1	75	AA02	-7.7	-7.9	7.8	4.2	18.2	6.7	-8.4	-10.7	-7.6	73.1	86.4	88.6	93.4	40.4	70.9	65.1	70.7	82.3
H_DC31_DC32	AA02	AAA	79.1	196.4	284.7	286.5	179.7	75.8	74.4	237.1	233.6	62	AA02	-8.6	-6.1	10.2	-7	18.4	21.6	-13.1	-6.8	-11.1	67.6	91.7	81.3	82.7	39.6	2.8	35.2	86.9	66.1

Table A5.3: Sugar pucker pseudorotation angles, sugar pucker type and closest DNA conformation for C_3T.

Chain	Base	Phase angle of pseudorotation	Sugar Pucker	Closest DNA conformation
A	C_1	160.0	C2'-endo	B-DNA
	C_2	108.1	C1'-exo	-
	C_3	86.7	O4'-endo	-
B	C_1	165.4	C2'-endo	B-DNA
	C_2	42.7	C4'-exo	-
	C_3	40.9	C4'-exo	-
C	C_1	31.8	C3'-endo	A-DNA
	C_2	17.6	C3'-endo	A-DNA
	C_3	40.8	C4'-exo	-
D	C_1	177.4	C2'-endo	B-DNA
	C_2	109.9	C1'-exo	-
	C_3	58.5	C4'-exo	-
E	C_1	158.3	C2'-endo	B-DNA
	C_2	108.9	C1'-exo	-
	C_3	87.7	O4'-endo	-
F	C_1	166.2	C2'-endo	B-DNA
	C_2	42.9	C4'-exo	-
	C_3	40.8	C4'-exo	-
G	C_1	32.8	C3'-endo	A-DNA
	C_2	17.2	C3'-endo	A-DNA
	C_3	41.4	C4'-exo	-
H	C_1	176.1	C2'-endo	B-DNA
	C_2	109.2	C1'-exo	-
	C_3	55.3	C4'-exo	-

Table A5.4: Conformational analysis of C_3T as calculated by DNATCO version 2.2.

step_ID	NtC	CANA	δ	ε	ζ	α1	β1	γ1	δ1	χ	χ1	confal	nearest_NtC	Δδ	Δε	Δζ	Δα1	Δβ1	Δγ1	Δδ1	Δχ	Δχ1	sδ	sε	sζ	sα1	sβ1	sγ1	sδ1	sχ	sχ1
A_DC1_DC2	NANT	NAN	138.5	210.9	290	136.5	166.7	197.1	111.2	234.5	239.6	0	BB03	-6	34.7	14.5	-29	2.5	24.3	-34.8	-4.6	7.8	80.9	1.7	51.3	15.8	98.4	15.8	0	94.4	83.1
A_DC2_DC3	NANT	NAN	111.2	204.7	292.8	166.6	147.4	176.3	93.5	239.6	239.2	0	AA01	29	10.7	2	17.7	-45.3	-5.5	6.4	34.6	51	0	69.4	97.9	42.9	0	84.7	57.6	0	0
A_DC3_DC4	AA02	AAA	93.5	196.2	279.6	291	178.3	58.9	84.1	239.2	229	81	AA02	5.8	-6.3	5.1	-2.5	17	4.7	-3.4	-4.7	-15.7	83.7	91.1	95	97.6	45.3	84.4	93.2	93.5	43.6
B_DC5_DC6	AA02	AAA	80.7	188.8	276.1	301.4	181.8	61.9	88	232.1	232	71	AA02	-7	-13.7	1.6	7.9	20.5	7.7	0.5	-11.8	-12.7	77.2	64.5	99.5	78.5	31.6	63.5	99.8	65.6	58.1
B_DC6_DC7	AA02	AAA	88	209.4	284.5	292.8	175.9	62.7	79.8	232	231.4	75	AA02	0.3	6.9	10	-0.7	14.6	8.5	-7.7	-11.9	-13.3	100	89.5	82	99.8	55.8	57.4	69.7	65.1	55.2
B_DC7_DC8	AA02	AAA	79.8	191.3	288	295.9	182.6	60.5	79.4	231.4	241.3	71	AA02	-7.9	-11.2	13.5	2.4	21.3	6.3	-8.1	-12.5	-3.4	71.9	74.6	69.6	97.8	28.9	73.7	67.1	62.3	96.2
C_DC9_DC10	NANT	NAN	151.8	210	291.3	155.5	154.3	180.9	110	245	239	0	BB03	7.3	33.8	15.8	-10	-9.9	8.1	-36	5.9	7.2	73.1	2.1	45.3	80.3	78.3	81.5	0	91	85.4
C_DC10_DC11	NANT	NAN	110	202.7	274.3	156.3	160.4	180.9	81.4	239	242.4	0	AA01	27.8	8.7	-16.5	7.4	-32.3	-0.9	-5.7	34	54.2	0	78.6	23.7	86.3	1.2	99.6	64.6	0	0
C_DC11_DC12	NANT	NAN	81.4	200.7	65.8	63.8	172.8	48.2	143.9	242.4	221.9	0	NS05	-72.2	-41.7	-11.3	0.6	-4.2	-15.3	6.9	5.2	-26.7	0	0	23	99.8	51.1	8.7	0	79.8	0
D_DC13_DC14	NANT	NAN	148.2	276.4	246.2	73.3	217.3	189.4	79.4	222.5	232	0	BA13	6.1	45	50.7	-0.9	-14.2	-6.6	-9.9	-43.7	32.8	83.2	0.9	0	99.8	58.2	90	68.4	0	1.8
D_DC14_DC15	AA02	AAA	79.4	202.1	279.8	294.5	181.9	54	73.8	232	237.7	75	AA02	-8.3	-0.4	5.3	1	20.6	0.2	-13.7	-11.9	-7	69.5	100	94.6	99.6	31.3	100	31.9	65.1	84.8
D_DC15_DC16	NANT	NAN	73.8	198.6	62.6	72.5	161.6	53.5	142.8	237.7	232.3	0	NS05	-79.8	-43.8	-14.5	9.3	-15.4	-10	5.8	0.5	-16.3	0	0	8.9	56.7	0	35.2	2.6	99.8	0
E_DC17_DC18	NANT	NAN	139.2	211.6	289.2	135.6	167.1	198.5	110.3	234.5	239	0	BB03	-5.3	35.4	13.7	-29.9	2.9	25.7	-35.7	-4.6	7.2	84.8	1.5	55.1	14	97.9	12.7	0	94.4	85.4
E_DC18_DC19	NANT	NAN	110.3	204.3	293.6	166	146.8	177.2	93.2	239	238.7	0	AA01	28.1	10.3	2.8	17.1	-45.9	-4.6	6.1	34	50.5	0	71.3	95.9	45.4	0	89.1	60.6	0	0
E_DC19_DC20	AA02	AAA	93.2	197.2	279.5	291.5	177.4	59.3	83.8	238.7	228.6	81	AA02	5.5	-5.3	5	-2	16.1	5.1	-3.7	-5.2	-16.1	85.2	93.6	95.1	98.5	49.2	81.9	92	92.1	41.8
F_DC21_DC22	AA02	AAA	80.8	189	275.8	301.3	181.4	62.1	89.2	232.7	232.7	72	AA02	-6.9	-13.5	1.3	7.8	20.1	7.9	1.7	-11.2	-12	77.7	65.3	99.7	79	33.1	62	98.3	68.4	61.6
F_DC22_DC23	AA02	AAA	89.2	209.5	282.6	295.1	175.8	61.2	79.9	232.7	230.5	77	AA02	1.5	7	8.1	1.6	14.5	7	-7.6	-11.2	-14.2	98.8	89.2	87.8	99	56.2	68.7	70.3	68.4	50.7
F_DC23_DC24	AA02	AAA	79.9	191.1	288.8	295.3	182.6	61.6	78.4	230.5	241	69	AA02	-7.8	-11.4	14.3	1.8	21.3	7.4	-9.1	-13.4	-3.7	72.5	73.8	66.6	98.8	28.9	65.7	60.4	58.1	95.5
G_DC25_DC26	NANT	NAN	152.3	209.4	293.1	156.4	154.1	180.8	109	244.7	238.8	0	BB03	7.8	33.2	17.6	-9.1	-10.1	8	-37	5.6	7	70	2.5	37.4	83.4	77.5	81.9	0	91.8	86.1
G_DC26_DC27	NANT	NAN	109	201.9	275.7	156.7	159.1	180	83.1	238.8	242.5	0	AA01	26.8	7.9	-15.1	7.8	-33.6	-1.8	-4	33.8	54.3	0	82	29.9	84.9	0	98.2	80.6	0	0
G_DC27_DC28	NANT	NAN	83.1	200.6	64.5	63.6	173.6	48.1	144.1	242.5	221.6	0	NS05	-70.5	-41.8	-12.6	0.4	-3.4	-15.4	7.1	5.3	-27	0	0	16.1	99.9	64.4	8.4	0	79.1	0
H_DC29_DC30	NANT	NAN	148.1	277.4	245.2	72	216.6	190.2	80	223	231.5	0	BA13	6	46	49.7	-2.2	-14.9	-5.8	-9.3	-43.2	32.3	83.7	0.7	0	98.8	55.1	92.2	71.6	0	2
H_DC30_DC31	AA02	AAA	80	202.1	280	293.8	181.6	54.3	73.6	231.5	237.3	75	AA02	-7.7	-0.4	5.5	0.3	20.3	0.1	-13.9	-12.4	-7.4	73.1	100	94.2	100	32.3	100	30.8	62.8	83.2
H_DC31_DC32	NANT	NAN	73.6	199.2	62	73.1	161.6	53.6	143.1	237.3	232	0	NS05	-80	-43.2	-15.1	9.9	-15.4	-9.9	6.1	0.1	-16.6	0	0	7.2	52.6	0	35.9	1.7	100	0

_____

Table A5.5: Conformational analysis of C$_3$A as calculated by DNATCO version 2.2.

step_ID	NtC	CANA	δ	ε	ζ	α1	β1	γ1	δ1	χ	χ1	confal	nearest_NtC	Δδ	Δε	Δζ	Δα1	Δβ1	Δγ1	Δδ1	Δχ	Δχ1	sδ	sε	sζ	sα1	sβ1	sγ1	sδ1	sχ	sχ1
A_DC1_DC2	NANT	NAN	148.7	250.8	253.2	92.4	209.5	195	83.1	232.9	235.7	0	BA10	13.5	51.1	35.1	-12.4	-19.9	0.6	-4.8	-24.9	40.7	63.4	0	13.1	79.4	46.8	99.9	90.2	25.3	0
A_DC2_DC3	AA02	AAA	83.1	187.4	270.8	301.1	176.2	59.1	80.8	235.7	233.9	76	AA02	-4.6	-15.1	-3.7	7.6	14.9	4.9	-6.7	-8.2	-10.8	89.4	58.7	97.3	80	54.4	83.2	76.1	81.6	67.5
A_DC3_DA4	AA02	AAA	80.8	200.2	278	314	186.4	36.2	86.7	233.9	245.1	66	AA02	-6.9	-2.3	3.5	20.5	25.1	-18	-0.8	-10	0.4	77.7	98.8	97.6	19.7	17.8	8.3	99.6	73.9	99.9
B_DC1_DC2	NANT	NAN	149.2	252.7	252	89.3	211.5	196.7	82.5	232.1	235.7	0	BA10	14	53	33.9	-15.5	-17.9	2.3	-5.4	-25.7	40.7	61.2	0	15	69.7	54.1	98.9	87.7	23.2	0
B_DC2_DC3	AA02	AAA	82.5	187.1	271.5	299.9	175.4	62	86	235.7	233.1	77	AA02	-5.2	-15.4	-3	6.4	14.1	7.8	-1.5	-8.2	-11.6	86.7	57.5	98.2	85.3	58	62.7	98.6	81.6	63.6
B_DC3_DA4	NANT	NAN	86	216.6	252.7	49.2	215	284.4	134.4	233.1	250.6	0	BB02	-54.4	23.1	6.6	17.5	20.9	-13	-15.3	-19.3	-2.8	0	13.5	87.6	48.1	49.4	65	15.4	42.1	98.3
C_DC1_DC2	NANT	NAN	149.4	258.2	248.4	83.1	215.2	199.2	83.2	231.9	235.9	0	BA13	7.3	26.8	52.9	8.9	-16.3	3.2	-6.1	-34.3	36.7	76.8	18.5	0	82.5	49	97.6	86.6	3.6	0
C_DC2_DC3	AA02	AAA	83.2	187.8	270.1	302.3	175.5	59.1	82.2	235.9	234	77	AA02	-4.5	-14.7	-4.4	8.8	14.2	4.9	-5.3	-8	-10.7	89.8	60.4	96.2	74.1	57.6	83.2	84.3	82.4	68
C_DC3_DA4	AA02	AAA	82.2	202.5	275.2	315.2	186.8	34.5	101.6	234	248.7	58	AA02	-5.5	0	0.7	21.7	25.5	-19.7	14.1	-9.9	4	85.2	100	99.9	16.2	16.8	5.1	29.8	74.3	94.8
D_DC1_DC2	NANT	NAN	149.5	256.5	249.8	84.5	214.2	198.5	82.9	232.5	235.5	0	BA10	14.3	56.8	31.7	-20.3	-15.2	4.1	-5	-25.3	40.5	59.9	0	19	53.9	64.2	96.5	89.4	24.2	0
D_DC2_DC3	AA02	AAA	82.9	187.4	271.1	301	176.5	58.9	81.6	235.5	233.4	77	AA02	-4.8	-15.1	-3.4	7.5	15.2	4.7	-5.9	-8.4	-11.3	88.5	58.7	97.7	80.4	53.1	84.4	80.9	80.8	65.1
D_DC3_DA4	NANT	NAN	81.6	199.6	277.3	317.7	192.8	24.1	113.9	233.4	265	0	AB01	-4.6	6.7	-6.1	18.6	13.1	-30.9	-27.9	12.3	10.4	76.3	90.7	89.6	34.1	60.3	0	0	72.8	75.9

_____

Table A5.6: Sugar pucker pesudorotation angles, sugar pucker type and closest DNA conformation for [C$_3$A + C$_3$T].
C$_X$ is the cytosine position in 5'-C$_1$C$_2$C$_3$A$_4$-3'and 5'-C$_1$C$_2$C$_3$T$_4$-3'.

Chain	Base	Phase angle of pseudorotation	Sugar Pucker	Closest DNA conformation
A	C$_1$	166.4	C2'-endo	B-DNA
	C$_2$	57.8	C4'-exo	-
	C$_3$	37.5	C4'-exo	-
	A$_4$	29.2	C3'-endo	A-DNA
B	C$_1$	167.0	C2'-endo	B-DNA
	C$_2$	59.3	C4'-exo	-
	C$_3$	36.9	C4'-exo	-
	A$_4$	28.5	C3'-endo	A-DNA
C	C$_1$	166.8	C2'-endo	B-DNA
	C$_2$	58.3	C4'-exo	-
	C$_3$	37.3	C4'-exo	-
	T$_4$	24.2	C3'-endo	A-DNA
D	C$_1$	166.9	C2'-endo	B-DNA
	C$_2$	59.3	C4'-exo	-
	C$_3$	37.7	C4'-exo	-
	T$_4$	23.8	C3'-endo	A-DNA
E	C$_1$	167.6	C2'-endo	B-DNA
	C$_2$	58.9	C4'-exo	-
	C$_3$	38.1	C4'-exo	-
	A$_4$	28.8	C3'-endo	A-DNA
F	C$_1$	166.6	C2'-endo	B-DNA
	C$_2$	58.9	C4'-exo	-
	C$_3$	37.6	C4'-exo	-
	T$_4$	28.5	C3'-endo	A-DNA
G	C$_1$	165.5	C2'-endo	B-DNA
	C$_2$	58.7	C4'-exo	-
	C$_3$	37.7	C4'-exo	-
	A$_4$	25.4	C3'-endo	A-DNA
H	C$_1$	167.1	C2'-endo	B-DNA
	C$_2$	58.1	C4'-exo	-
	C$_3$	37.3	C4'-exo	-
	T$_4$	25.8	C3'-endo	A-DNA

Table A5.7: Conformational analysis of [C$_3$A + C$_3$T] as calculated by DNATCO version 2.2.

step_ID	NtC	CANA	δ	ε	ζ	α1	β1	γ1	δ1	χ	χ1	confal	nearest_NtC	Δδ	Δε	Δζ	Δα1	Δβ1	Δγ1	Δδ1	Δχ	Δχ1	sδ	sε	sζ	sα1	sβ1	sγ1	sδ1	sχ	sχ1
A_DC0_DC1	NANT	NAN	144	215.9	289.8	149	162.9	161.3	86.3	233.6	237.2	0	BB03	-0.5	39.7	14.3	-16.5	-1.3	-11.5	-59.7	-5.5	5.4	99.9	0	52.3	55	99.6	66.2	0	92.1	91.5
A_DC1_DC2	AA02	AAA	86.3	189.4	268.1	301.2	178.3	60.1	84.1	237.2	237.6	80	AA02	-1.4	-13.1	-6.4	7.7	17	5.9	-3.4	-6.7	-7.1	99	67	92.2	79.5	45.3	76.6	93.2	87.3	84.4
A_DC2_DA3	AA02	AAA	84.1	198.8	271.7	304.2	171.3	58.7	81.6	237.6	232.7	83	AA02	-3.6	-3.7	-2.8	10.7	10	4.5	-5.9	-6.3	-12	93.4	96.9	98.5	64.2	76	85.6	80.9	88.7	61.6
B_DC1_DC2	NANT	NAN	144.2	217.1	287.8	146	165.2	163.8	86.3	235.1	236.8	0	BB03	-0.3	40.9	12.3	-19.5	1	-9	-59.7	-4	5	99.9	0	61.9	43.4	99.8	77.6	0	95.8	92.7
B_DC2_DC3	AA02	AAA	86.3	189.2	267.8	301.5	178.5	59.5	83.8	236.8	237.6	80	AA02	-1.4	-13.3	-6.7	8	17.2	5.3	-3.7	-7.1	-7.1	99	66.2	91.5	78.1	44.5	80.6	92	85.8	84.4
B_DC3_DA4	AA02	AAA	83.8	196.3	274	301.8	174.3	59.3	81.7	237.6	233.1	82	AA02	-3.9	-6.2	-0.5	8.3	13	5.1	-5.8	-6.3	-11.6	92.3	91.4	100	76.6	62.9	81.9	81.5	88.7	63.6
C_DC1_DC2	NANT	NAN	144.4	216.7	287.5	146.1	166.3	163.5	85.1	234	236.6	0	BB03	-0.1	40.5	12	-19.4	2.1	-9.3	-60.9	-5.1	4.8	100	0	63.3	43.8	98.9	76.3	0	93.2	93.2
C_DC2_DC3	AA02	AAA	85.1	189.2	268.2	302.2	178.8	59	84.1	236.6	238.1	80	AA02	-2.6	-13.3	-6.3	8.7	17.5	4.8	-3.4	-7.3	-6.6	96.5	66.2	92.4	74.6	43.2	83.8	93.2	85.1	86.4
C_DC3_DT4	AA02	AAA	84.1	196.6	274	300.9	172.3	62.8	81.4	238.1	230	79	AA02	-3.6	-5.9	-0.5	7.4	11	8.6	-6.1	-5.8	-14.7	93.4	92.2	100	80.9	71.8	56.7	79.7	90.3	48.3
D_DC1_DC2	NANT	NAN	144	216.9	287.5	146.7	165.8	163.5	85.5	234.7	236.4	0	BB03	-0.5	40.7	12	-18.8	1.6	-9.3	-60.5	-4.4	4.6	99.9	0	63.3	46	99.4	76.3	0	94.9	93.8
D_DC2_DC3	AA02	AAA	85.5	189.2	267.8	300.7	177.2	60.7	84.9	236.4	236.1	80	AA02	-2.2	-13.3	-6.7	7.2	15.9	6.5	-2.6	-7.5	-8.6	97.5	66.2	91.5	81.8	50	72.3	96	84.3	78
D_DC3_DT4	AA02	AAA	84.9	198.6	268.6	305.9	170.3	60.1	81.9	236.1	230.6	79	AA02	-2.8	-3.9	-5.9	12.4	9	5.9	-5.6	-7.8	-14.1	95.9	96.5	93.3	55.2	80.1	76.6	82.6	83.2	51.2
H_DC1_DC2	NANT	NAN	144.2	217	287.1	145.8	166.9	163.5	84.9	234.7	236.4	0	BB03	-0.3	40.8	11.6	-19.7	2.7	-9.3	-61.1	-4.4	4.6	99.9	0	65.3	42.7	98.2	76.3	0	94.9	93.8
H_DC2_DC3	AA02	AAA	84.9	188.6	268.1	303.6	178.2	58.3	84.3	236.4	238.4	80	AA02	-2.8	-13.9	-6.4	10.1	16.9	4.1	-3.2	-7.5	-6.3	95.9	63.7	92.2	67.4	45.7	87.9	94	84.3	87.5
H_DC3_DT4	AA02	AAA	84.3	197.5	273.3	302.9	171.8	61.2	81.5	238.4	231	81	AA02	-3.4	-5	-1.2	9.4	10.5	7	-6	-5.5	-13.7	94.1	94.3	99.7	71	73.9	68.7	80.3	91.2	53.2
G_DC1_DC2	NANT	NAN	143.8	217.4	287	145.5	168	163.3	85.2	233.7	235.6	0	BB03	-0.7	41.2	11.5	-20	3.8	-9.5	-60.8	-5.4	3.8	99.7	0	65.7	41.6	96.5	75.4	0	92.4	95.7
G_DC2_DC3	AA02	AAA	85.2	188.7	267.6	302.4	178.2	59.2	84.2	235.6	237.8	79	AA02	-2.5	-13.8	-6.9	8.9	16.9	5	-3.3	-8.3	-6.9	96.7	64.1	91	73.6	45.7	82.5	93.6	81.2	85.2
G_DC3_DA4	AA02	AAA	84.2	197.4	273.1	302.7	172.8	59.8	81.7	237.8	232	82	AA02	-3.5	-5.1	-1.4	9.2	11.5	5.6	-5.8	-6.1	-12.7	93.7	94.1	99.6	72.1	69.6	78.6	81.5	89.3	58.1
E_DC1_DC2	NANT	NAN	144.4	216.6	287.1	145.3	167.2	164	84.9	234.4	236.5	0	BB03	-0.1	40.4	11.6	-20.2	3	-8.8	-61.1	-4.7	4.7	100	0	65.3	40.8	97.8	78.5	0	94.2	93.5
E_DC2_DC3	AA02	AAA	84.9	188.5	268.6	301.1	178.8	60.2	84.7	236.5	237.7	80	AA02	-2.8	-14	-5.9	7.6	17.5	6	-2.8	-7.4	-7	95.9	63.3	93.3	80	43.2	75.9	95.3	84.7	84.8
E_DC3_DA4	AA02	AAA	84.7	198.2	271.9	302.4	171.7	60.4	81.5	237.7	231.2	82	AA02	-3	-4.3	-2.6	8.9	10.4	6.2	-6	-6.2	-13.5	95.3	95.8	98.7	73.6	74.4	74.5	80.3	89	54.2
F_DC1_DC2	NANT	NAN	143.5	218	285	142	169.9	166	85.5	233.8	236.4	0	BB03	-1	41.8	9.5	-23.5	5.7	-6.8	-60.5	-5.3	4.6	99.4	0	75.1	29.7	92.2	86.6	0	92.7	93.8
F_DC2_DC3	AA02	AAA	85.5	188.7	268	301.7	178.2	59.9	84.3	236.4	237.3	80	AA02	-2.2	-13.8	-6.5	8.2	16.9	5.7	-3.2	-7.5	-7.4	97.5	64.1	91.9	77.1	45.7	77.9	94	84.3	83.2
F_DC3_DT4	AA02	AAA	84.3	198.3	272.4	303.1	171.2	61.1	81.3	237.3	231	81	AA02	-3.4	-4.2	-2.1	9.6	9.9	6.9	-6.2	-6.6	-13.7	94.1	96	99.1	70	76.5	69.4	79.1	87.6	53.2

Publications

Contributions stated for thesis author (SPG) only.

From this work:

1.   Gurung, S. P., Schwarz, C., Hall, J. P., Cardin, C. J. & Brazier, J. A. The importance of loop length on the stability of i-motif structures. *Chem. Commun.* **51,** 5630–5632 (2015).

>   SPG collected the data, performed the analysis, and helped write the supplementary information section for the manuscript.

Other publications:

2.   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).

>   SPG set plates for DNA crystallisation and prepared samples for both time-resolved infrared spectroscopy and X-ray crystallography.

3.   Keane, P. M. *et al.* Long-Lived Excited-State Dynamics of i-Motif Structures Probed by Time-Resolved Infrared Spectroscopy. *Chemphyschem,* **17,** 1281–1287 (2016).

>   SPG helped develop sample preparation methodology for TRIR spectroscopy.

4.   Hall, J. P. *et al.* Guanine Can Direct Binding Specificity of Ru–dipyridophenazine (dppz) Complexes to DNA through Steric Effects. *Chem. - A Eur. J.* **23,** 4981–4985 (2017).

>   SPG conducted and analysed DNA melts of the sequences d((5-BrC)GG(5-BrC)G(5Br-C)(5-BrC)G) and d(CGGCGCCG).

5.   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. *Chem. - A Eur. J.* **23,** 10344–10351 (2017).

>   SPG grew crystals of d(TCGGCGCCIA)$_2$ and collected the X-ray data.