Structure, sequence and ion-exchange of left-handed G-quadruplexes

Bakalar, Blaz

2019

Bakalar, B. (2019). Structure, sequence and ion-exchange of left-handed G-quadruplexes.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/106467>

<https://doi.org/10.32657/10220/47993>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**
SINGAPORE

**STRUCTURE, SEQUENCE AND ION-EXCHANGE OF
LEFT-HANDED G-QUADRUPLICES**

BAKALAR BLAŽ

School of Biological Sciences

2018

**STRUCTURE, SEQUENCE AND ION-EXCHANGE OF
LEFT-HANDED G-QUADRUPLEXES**

BAKALAR Blaž

Supervisor: Prof. PHAN Anh Tuấn

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2018

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

13 February 2019

Blaž Bakalar

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

13-Feb-2019

.....

Date

.....

Professor PHAN Anh Tuan

Authorship Attribution Statement

This thesis contains material, where I was the first author, from one published paper and one paper under peer-review.

Part of the work presented in chapter 3 is published in *Angewandte Chemie Int. Ed.*:

Bakalar, B., Heddi, B., Schmitt, E., Mechulam, Y. and Phan, A. T. (2019), A Minimal Sequence for Left-Handed G-Quadruplex Formation. *Angewandte Chemie International Edition*.

- Experiment design was done by Bakalar, B., Heddi, B. and Phan, A. T.
- Experiments were carried out by Bakalar B.
- X-ray data collection and structure determination was done by Bakalar, B., Heddi, B., Schmitt, E. and Mechulam, Y.
- Structure analysis was done by Bakalar, B. and Heddi, B.
- Initial manuscript drafting was done by Bakalar, B.
- Manuscript refinement was done by all authors.

The work presented in chapter 4 is currently submitted to *Chemical Science* for review:

Bakalar, B., Winnerdy, F. R., Heddi, B., Marchand, A., Rosu, F., Gabelica, V. and Phan, A. T. (2019), Unprecedented hour-long residence time of a cation in the left-handed G-quadruplex. *Under review with Chemical Science*

- Experiment design was done by Bakalar, B., Winnerdy, F. R., Heddi, B. and Phan, A. T.
- Experiments, *sans* Mass spec., were carried out by Bakalar, B. and Winnerdy, F. R.
- Mass spectrometry data collection and analysis was done by Marchand, A., Rosu, F. and Gabelica, V
- Initial manuscript drafting was done by Bakalar, B.
- Manuscript refinement was done by all authors.

13 February 2019

Blaž Bakalar

Acknowledgments

I would like to express my gratitude to my supervisor Prof. Phan Anh Tuấn and co-supervisor Prof. Daniela Rhodes for giving me this opportunity and their outstanding guidance during my PhD.

I would like to extend my gratitude to all of the colleagues from both labs, especially to Fernaldo, Vee Vee, Arijit, Huy, Chong Wai, Ling and Anna for their helpful hand and sound advice. For his enormous contribution during these four years, I would especially like to thank Dr. Heddi Brahim.

I would like to thank our collaborators Assoc. Prof. Yves Mechulam and Emmanuelle Schmitt from Biochemistry laboratory, École Polytechnique, Paris, France for their contributions to the structural part of the project.

I would like to thank our collaborators Dr. Valérie Gabelica and Adrien Marchand from ARNA Laboratory, Bordeaux, France for their contributions to the ion-exchange part of the project.

I would like to thank my friends back home and here in Singapore for their endless support, encouragement and all the fun times we had throughout this journey.

Lastly, I would like to thank my parents Zdenka and Ivica, my brother Igor and the rest of my family for always being by my side and helping me, even when home is far.

Table of content

STATEMENT OF ORIGINALITY	1
SUPERVISOR DECLARATION STATEMENT	2
AUTHORSHIP ATTRIBUTION STATEMENT	3
ACKNOWLEDGMENTS	4
TABLE OF CONTENT	5
TABLE OF FIGURES	7
TABLE OF TABLES	14
ABBREVIATIONS	15
ABSTRACT	16
CHAPTER 1: INTRODUCTION	17
1.1 DNA ROLE AND STRUCTURE	17
1.2 G-QUADRUPLEX FEATURES AND PROPERTIES	18
1.3 CIRCULAR DICHROISM OF G-QUADRUPLEXES.....	25
1.4 NMR OF G-QUADRUPLEXES	31
1.5 X-RAY DIFFRACTION SPECTROSCOPY OF G-QUADRUPLEXES.....	32
1.6 MASS SPECTROMETRY OF G-QUADRUPLEXES	35
1.7 ROLE OF CATIONS IN G-QUADRUPLEX FOLDING AND STUDIES.....	37
1.8 G-QUADRUPLEX ROLE IN BIOLOGICAL SYSTEMS	40
1.9 DISCOVERY OF A LEFT-HANDED G-QUADRUPLEX AND ITS PROPERTIES.....	41
CHAPTER 2: MATERIALS AND METHODS	46
2.1 SAMPLE PREPARATION AND SEQUENCES USED	46
2.2 CIRCULAR DICHROISM (CD)	48
2.3 ¹ H NMR SPECTROSCOPY	49
2.4 D ₂ O EXCHANGE NMR SPECTROSCOPY.....	49
2.5 TRACKING OF ¹⁵ NH ₄ ⁺ USING ¹ H NMR SPECTROSCOPY	49
2.6 CRYSTALLIZATION AND X-RAY CRYSTALLOGRAPHY	49
2.6.1 Block 2 structure determination	50
2.6.2 2xBlock 2 structure determination	51
2.6.3 Z-G4R structure determination	51
2.6.4 Z-G4-Bulge structure determination.....	52
2.6.5 Z-G4-2xBulge structure determination	52

2.6.6	<i>Structural analysis</i> -----	53
2.7	ELECTROSPRAY IONIZATION-MASS SPECTROMETRY	53
CHAPTER 3: A MINIMAL LEFT-HANDED MOTIF AND SEQUENCE FLEXIBILITY OF LEFT-HANDED G-QUADRUPLEXES.....		54
3.1	<i>BLOCK 1</i> AND <i>BLOCK 2</i> DISPLAYS RIGHT- AND LEFT-HANDED G-QUADRUPLEX SPECTRAL FEATURES, RESPECTIVELY	55
3.2	<i>BLOCK 2</i> FORMS A LEFT-HANDED DIMERIC G-QUADRUPLEX	57
3.3	CONNECTING TWO <i>BLOCK 2</i> MOTIFS	62
3.4	2 <i>xBLOCK 2</i> FORMS A MONOMERIC LEFT-HANDED G4 STRUCTURE.....	63
3.5	THYMINES LOOP SUBSTITUTIONS OF THE LEFT-HANDED MOTIF	66
3.6	EFFECT OF BLOCK ORDER ON LEFT-HANDED G-QUADRUPLEX FORMATION	73
3.7	EFFECT OF SPLITTING GUANINES ON THE LEFT-HANDED G-QUADRUPLEX STRUCTURE	79
3.7.1	<i>Structure of Z-G4-Bulge</i> -----	82
3.7.2	<i>Structure of Z-G4-2xBulge</i> -----	86
3.8	ANALYSIS AND DISCUSSION OF LEFT-HANDED G-QUADRUPLEX SEQUENCE AND STRUCTURE FLEXIBILITY	92
CHAPTER 4: EXTREMELY SLOW-EXCHANGE ION IN THE LEFT-HANDED Z-G4		99
4.1	Z-G4 PRESERVES LEFT-HANDED SPECTRAL AND STRUCTURAL CHARACTERISTIC IN AMMONIUM	99
4.2	OBSERVATION OF COORDINATED AMMONIUM IONS IN Z-G4 USING NMR.....	103
4.3	TRACKING OF ¹⁵ N AMMONIUM ION EXCHANGE USING ¹ H NMR	106
4.4	OBSERVATION AND TRACKING OF AMMONIUM IONS IN THE CHANNEL OF Z-G4 USING MASS SPECTROMETRY	112
4.5	DISCUSSION AND CONCLUSIONS	115
CHAPTER 5: SUMMARY AND FUTURE PERSPECTIVES		117
REFERENCES.....		120

Table of figures

FIGURE 1.1: (A) SCHEMATIC OF TWO BASE-PAIRED ANTIPARALLEL COMPLEMENTARY STRANDS OF DNA. (B) WATSON AND CRICK'S MODEL OF DNA DOUBLE HELIX [1].	17
FIGURE 1.2: AN EXAMPLE OF DIVERSE RANGE OF STRUCTURES ADOPTED BY GUANINE RICH DNA AND RNA OLIGONUCLEOTIDES.	18
FIGURE 1.3: (A) STICK REPRESENTATION OF A G-TETRAD FORMED BY FOUR PLANARY ARRANGED GUANINE BASES. HOOGSTEEN HYDROGEN BONDS ARE SHOWN AS DOTTED LINES, OXYGEN ATOMS ARE SHOWN IN RED, HYDROGENS IN GREY, NITROGEN IN DEEP BLUE, CARBONS IN CYAN AND THE SODIUM ION IS REPRESENTED AS A PURPLE SPHERE. (B) EXAMPLE OF A G-QUADRUPLEX SCHEMATIC, CONTAINING THREE STACKED G-TETRADS. GUANINE BASES ARE SHOWN AS CYAN RHOMBI.	19
FIGURE 1.4: TWO DIFFERENT SIDES OR FACES OF THE G-TETRAD AS DETERMINED BY THE HOOGSTEEN HYDROGEN BOND DIRECTIONALITY. (A) TOP VIEW OF THE "HEAD" FACE OF THE G-TETRAD. (B) TOP VIEW OF THE "TAIL" FACE OF THE G-TETRAD.	20
FIGURE 1.5: SAME-POLARITY OR HOMOPOLAR (A) AND OPPOSITE-POLARITY OR HETEROPOLAR (B) STACKING OF TWO G-TETRADS.	20
FIGURE 1.6: (A) REPRESENTATION OF INTER-TETRAD ROTATION ANGLE (Θ_{TET}) AND (B) G-TETRAD SEPARATION DISTANCE (D_{TET}).	21
FIGURE 1.7: DIFFERENT TYPES OF G-QUADRUPLEX GROOVES ADOPTED DUE TO <i>ANTI</i> AND <i>SYN</i> GUANINE GLYCOSIDIC CONFORMATIONS. (A) THE G-TETRAD CORE (LEFT) AND A SINGLE G-TETRAD WITH ANNOTATED GROOVE WIDTHS IN RESPECT TO THE GLYCOSIDIC TORSION ANGLE CONFORMATION (RIGHT). (B) <i>ANTI</i> AND <i>SYN</i> GLYCOSIDIC TORSION ANGLE CONFORMATION OF GUANINES.	22
FIGURE 1.8: STRUCTURAL ELEMENTS THAT GIVE RISE TO DIFFERING G-QUADRUPLEX STRUCTURES. (A) DIFFERENT COMBINATIONS OF BACKBONE PROGRESSION ORIENTATIONS. (B) DIFFERENT TYPES OF CONNECTING LOOPS.	23
FIGURE 1.9: EXAMPLE OF THE BULGE/DISCONTINUITY FEATURE IN A G-QUADRUPLEX, COMPARED TO A REGULAR LOOP.	24
FIGURE 1.10: UNPOLARIZED LIGHT OF A SPECIFIC WAVELENGTH IS PASSED THROUGH A LINEAR POLARIZER. LINEARLY POLARIZED LIGHT IS PASSED THROUGH A QUARTER-WAVE PLATE, MAKING IT CIRCULARLY POLARIZED. THE LIGHT WAVE PROPAGATING THROUGH SPACE IS SHOWN IN RED AND THE STRENGTH AND DIRECTION OF THE ELECTRIC FIELD VECTOR IS SHOWN AS BLACK VECTORS. IN THIS EXAMPLE THE CIRCULARLY POLARIZED LIGHT TRACES A LEFT-HANDED, COUNTERCLOCKWISE HELIX.	26
FIGURE 1.11: (A) VECTOR DEPICTION OF RELATIONSHIP BETWEEN LINEARLY POLARIZED AND LEFT/RIGHT-CIRCULARLY POLARIZED LIGHT OF EQUAL AMPLITUDES. (B) ELLIPTICALLY POLARIZED LIGHT (DASHED) IS PRODUCED BY COMBINING LEFT- AND RIGHT-CIRCULARLY POLARIZED LIGHT WITH DIFFERENT AMPLITUDES.	27
FIGURE 1.12: (A) TWO CHROMOPHORES WITH INDIVIDUAL BUT EQUAL EXCITATION STATES ENERGIES (GREEN) COUPLE AND PRODUCE TWO EXCITON-COUPLED AND DISTINCT EXCITED STATE ENERGY LEVELS (RED AND BLUE). (B) REPRESENTATION OF GEOMETRIC PARAMETERS THAT INFLUENCE THE AMPLITUDE AND SIGN OF CD SPECTRA PRODUCED BY COUPLED CHROMOPHORES. EXPECTED ABSORPTION SPECTRUM (C) AND CD SPECTRUM (D) PRODUCED BY TWO CHROMOPHORES WITH A NEGATIVE CHIRALITY ARRANGEMENT AND EXCITON SPLITTING AS SHOWN IN (B). ABSORPTION OF INDIVIDUAL EXCITON-COUPLED EXCITED STATES IS SHOWN IN RED AND BLUE, WHILE THE RESULTING ABSORPTION SPECTRUM AND CD SPECTRUM ARE SHOWN IN PURPLE. FIGURE ADAPTED FROM [44].	28
FIGURE 1.13: MOLAR CD SPECTRA OF DIFFERENT TYPES OF G-QUADRUPLEXES. PARALLEL RIGHT-HANDED (BLUE), ANTIPARALLEL RIGHT-HANDED (GREEN), HYBRID PARALLEL/ANTIPARALLEL RIGHT HANDED (MAGENTA) AND PARALLEL LEFT-HANDED G-QUADRUPLEX (RED).	30

FIGURE 1.14: (A) LABELED HYDROGENS AND HETEROATOMS OF THE GUANINE NUCLEOTIDE. THE IMINO PROTON IS SHOWN IN RED.	
(B) FOUR HOOGSTEEN BONDED GUANINE NUCLEOBASES ASSEMBLE TO FORM A TETRAD. THE HYDROGEN BONDING OF IMINO PROTONS (RED) SLOWS THEIR EXCHANGE WITH THE SOLVENT, MAKING THEM OBSERVABLE IN NMR.	31
FIGURE 1.15: (A) SCHEMATIC REPRESENTATION OF THE SITTING DROP VAPOR DIFFUSION CRYSTALLIZATION CONTAINER. (B) A SIMPLE PHASE DIAGRAM SHOWING THE ZONES THAT ARE INVOLVED IN CRYSTAL GROWTH. THE CIRCLE (1) REPRESENTS THE STARTING POSITION, BEFORE VAPOR DIFFUSION INCREASES THE CONCENTRATION OF BOTH THE PRECIPITANT AND DNA TO THE SUPERSATURATION POINT (SQUARE, 2) WHERE SPONTANEOUS HOMOGENEOUS NUCLEATION OCCURS. ONCE THE NUCLEATION SITES ARE FORMED, THE CRYSTALS GROW UNTIL THE SOLUTION REACHES THE EQUILIBRIUM POINT (RED STAR, 3) WHERE THE SAMPLE IS NEITHER SUPERSATURATED NOR UNDERSATURATED.	33
FIGURE 1.16: (A) DIFFRACTION SPOTS ARISE AS A PATTERN FROM THE (B) REGULAR, REPEATING CRYSTAL LATTICE THAT ACTS AS AN AMPLIFIER TO (C) THE CONSTRUCTIVE INTERFERENCE, PRODUCED BY ATOMS ABSORBING AND RE-SCATTERING X-RAY RADIATION, IN ACCORDANCE WITH BRAGG'S LAW.	34
FIGURE 1.17: SCHEMATIC REPRESENTATION OF MASS SPECTROMETRY (MS) PARTICLE SEPARATION, BASED ON MASS (BLUE VS RED PARTICLES) AND CHARGE (SINGLE VS DOUBLE CHARGE).	36
FIGURE 1.18: GUANOSINE GEL MELTING TEMPERATURE (T_m) DEPENDENCE ON THE TYPE OF CATION USED AND ITS RESPECTIVE IONIC RADIUS. FIGURE ADAPTED FROM [25].	38
FIGURE 1.19: SCHEMATIC REPRESENTATION OF POSSIBLE G-QUADRUPLEX FORMING LOCATIONS AND THEIR POSSIBLE EFFECTS ON CELLULAR PROCESSES. (A) G-QUADRUPLEX FORMATION IN THE PROMOTER REGION DURING TRANSCRIPTION AND (C) REPLICATION. (B) G-QUADRUPLEX FORMATION ON THE G-RICH SINGLE STRANDED OVERHANG OF TELOMERES. (D) G-QUADRUPLEX FORMATION IN MRNA. FIGURE ADAPTED FROM [132].	40
FIGURE 1.20: CD SPECTRA OF MODIFIED <i>AGRO100</i> SEQUENCES FROM TABLE 1.1 IN POTASSIUM BUFFER.	41
FIGURE 1.21: NMR SPECTRA OF MODIFIED <i>AGRO100</i> SEQUENCES FROM TABLE 1.1 IN POTASSIUM BUFFER.	42
FIGURE 1.22: STRUCTURAL FEATURES OF A LEFT-HANDED QUADRUPLEX. (A) COMPARISON OF BACKBONE PROGRESSION HANDEDNESS OF <i>Z-G4</i> (LEFT) AND A TYPICAL PARALLEL RIGHT-HANDED G-QUADRUPLEX (GREEN). (B) TOP VIEW OF LEFT-HANDED QUADRUPLEX WITH A T-CAP. THYMINE BASES ARE COLORED IN ORANGE, GUANINE BASES ARE CYAN AND THE BACKBONE IS IN GREY.	43
FIGURE 1.23: (A) STICK REPRESENTATION OF GUANINE DINUCLEOTIDE STEP WITH THE $O4'$ (YELLOW) OF THE FOLLOWING THYMINE $T7$ FORMING A HYDROGEN BOND WITH THE TWO AMINO PROTONS OF GUANINE $G5$. (B) IMINO-AMINO REGION OF THE 1H NMR SPECTRUM OF <i>Z-G4</i> BEFORE (TOP) AND AFTER D_2O EXCHANGE (BOTTOM). AMINO PROTONS ARE MARKED WITH RED ASTERISK. (C) 2D NOESY SPECTRUM SHOWING TWO DISTINCT GROUPS OF CROSS-PEAKS BETWEEN THE OUTER G-TETRAD GUANINE IMINO PROTONS AND BOTH AMINO PROTONS. FIGURE ADAPTED FROM [53].	43
FIGURE 1.24: COMPARISON OF GUANINE STACKING EFFECTS OF PARALLEL LEFT-HANDED <i>Z-G4</i> AND RIGHT-HANDED <i>T95-2T</i> (PDB CODE, 2LK7) G-QUADRUPLEXES. (A) TOP VIEW OF GUANINE STACKING WITHIN LEFT-HANDED <i>Z-G4</i> WITH A RELATIVE ROTATION OF -25.7° (LEFT) AND RIGHT-HANDED <i>T95-2T</i> WITH A RELATIVE ROTATION OF $+29.4^\circ$ (RIGHT). (B) CD SPECTRA OF <i>Z-G4</i> (LEFT) AND <i>T95-2T</i> (RIGHT).	44
FIGURE 1.25: COLOR CODED REPRESENTATION OF <i>Z-G4</i> WHERE THE TOP BLOCK, <i>BLOCK 1</i> , IS COLORED IN BLUE, THE BOTTOM BLOCK, <i>BLOCK 2</i> IN RED AND THE LOOP CONNECTING THEM IN GRAY. (A) SIDE VIEW OF THE STICK & RIBBON REPRESENTATION OF <i>Z-</i>	

G4. (B) COLOR CODED SEQUENCE AND SCHEMATIC REPRESENTATION OF Z-G4 WITH BACKBONE PROGRESSION AND NUMBERED GUANINE RESIDUES.	45
FIGURE 3.1: (A) SEQUENCE AND (B) STRUCTURE OF THE LEFT-HANDED Z-G4 STRUCTURE [53]. (C) BOTTOM VIEW OF THE Z-G4 STRUCTURE, WITH SPLIT GS IN BLOCK 2 BEING HIGHLIGHTED. THE TWO CONSTITUTIVE BLOCKS ARE COLOURED IN BLUE (BLOCK 1) AND RED (BLOCK 2) RESPECTIVELY. SOME THYMINE RESIDUES ARE NOT SHOWN FOR CLARITY.	54
FIGURE 3.2: CD SPECTRA OF BLOCK 1 (BLUE) AND BLOCK 2 (RED) WITH CORRESPONDING SEQUENCES. θ REFERS TO THE MOLAR ELIPTICITY.	55
FIGURE 3.3: CD (A) AND NMR (B) SPECTRA OF BLOCK 1 AND TWO VARIANTS WITHOUT TERMINAL TS AT THE 3' AND 5' END.	56
FIGURE 3.4: CD (A) AND NMR (B) SPECTRA OF BLOCK 2 AND VARIANTS WITHOUT TWO TS AT THE 3' END OR AN ADDITIONAL 5' THYMINE.	57
FIGURE 3.5: (A) ALIGNMENT OF BOTH BLOCK 2 SUBUNITS. THYMINE RESIDUES OF SUBUNIT 1 ARE COLOURED IN RED AND BLUE FOR SUBUNIT 2. (B) STACKING INTERACTIONS INVOLVING THYMINE RESIDUES PRESENT IN THE CRYSTAL LATTICE OF BLOCK 2.	59
FIGURE 3.6: (A) CRYSTAL STRUCTURE OF THE LEFT-HANDED G-QUADRUPLIX FORMED BY BLOCK 2. THE ELECTRON DENSITY CORRESPONDS TO THE FINAL 2MFO-FC MAP CONTOURED AT 1.0 Σ . (B) SCHEMATIC REPRESENTATION OF THE CRYSTAL STRUCTURE WITH NUMBERED GUANINE RESIDUES AND MARKED BACKBONE PROGRESSION DIRECTION. (C) RIBBON REPRESENTATION OF THE CRYSTAL STRUCTURE. (D) TOP VIEW OF THE CRYSTAL STRUCTURE.	60
FIGURE 3.7: SIDE VIEW OF SUBUNIT A (TOP) AND SUBUNIT B (BOTTOM) WITH THE HYDROGEN BOND BETWEEN DG1-O5' AND DG12-O3'.	61
FIGURE 3.8: ALIGNMENT OF BOTH BLOCK 2 SUBUNITS (IN RED AND RASPBERRY RESPECTIVELY) WITH THE Z-G4 STRUCTURE (BLUE). 3' THYMINE BASES WERE REMOVED FOR CLARITY. 5' AND 3' ENDS OF ALL 3 STRUCTURES ARE LABELED. (A) OUTER G-TETRAD VIEW AND (B) STACKING/INNER G-TETRAD VIEW. ALIGNMENT OF ALL THREE STRUCTURES GIVES AN RMSD < 1 Å.	61
FIGURE 3.9: CD (A) AND NMR (B) SPECTRA OF 2xBLOCK 2 AND VARIANTS WITHOUT 3' THYMINES OR AN ADDITIONAL 5' THYMINE.	62
FIGURE 3.10: PACKING INTERACTIONS OF 2xBLOCK 2 IN THE CRYSTAL LATTICE. T13-T19 (GREEN), T11-25 (MAGENTA) AND T5-T8 (RED) ARE INTERACTING THROUGH π - π STACKING. T2-T27 (BLUE) ARE HYDROGEN BONDING.	63
FIGURE 3.11: (A) CRYSTAL STRUCTURE OF THE LEFT-HANDED G-QUADRUPLIX FORMED BY 2xBLOCK 2 (PDB ID 6GZ6). THE ELECTRON DENSITY CORRESPONDS TO THE FINAL 2MFO-FC MAP CONTOURED AT 1.0 Σ . (B) SCHEMATIC REPRESENTATION OF THE CRYSTAL STRUCTURE WITH NUMBERED GUANINE RESIDUES AND MARKED BACKBONE PROGRESSION DIRECTION. (C) RIBBON REPRESENTATION OF THE CRYSTAL STRUCTURE.	65
FIGURE 3.12: CD (A) AND NMR (B) SPECTRA OF SINGLE THYMINE TO CYTOSINE SUBSTITUTIONS SEQUENCES FOR ALL FOUR THYMINE LOOPS IN A SINGLE BLOCK OF 2xBLOCK 2 Δ	66
FIGURE 3.13: CD SPECTRA OF 2xBLOCK 2 Δ SEQUENCES WITH DIFFERENT BLOCK CONNECTING LOOP LENGTHS (A) AND THYMINE TO ADENINE SUBSTITUTIONS (B).	67
FIGURE 3.14: CD (A) AND NMR (B) SPECTRA OF 2xBLOCK 2 Δ SEQUENCES WITH THYMINE SUBSTITUTIONS CARRIED OUT ON A SINGLE BLOCK. INCREMENTAL SUBSTITUTION OF THYMINE LOOPS, FROM ONE (BLUE) TO TWO (RED), THREE (GREEN) AND ALL FOUR (PURPLE) THYMINES SUBSTITUTED WITH CYTOSINE.	68

FIGURE 3.15: CD (A) AND NMR (B) SPECTRA OF <i>2xBLOCK 2Δ</i> SEQUENCES WITH THYMINE SUBSTITUTIONS CARRIED OUT ON A SINGLE BLOCK. INCREMENTAL SUBSTITUTION OF THYMINE LOOPS, FROM ONE (BLUE) TO TWO (RED), THREE (GREEN) AND ALL FOUR (PURPLE) THYMINES SUBSTITUTED WITH ADENINE.	69
FIGURE 3.16: CD (A) AND NMR (B) SPECTRA OF <i>2xBLOCK 2Δ</i> SEQUENCES WITH THYMINE SUBSTITUTIONS CARRIED OUT ON A SINGLE BLOCK. INCREMENTAL SUBSTITUTION OF THYMINE LOOPS, FROM ONE (BLUE) TO TWO (RED), THREE (GREEN) AND ALL FOUR (PURPLE) THYMINES SUBSTITUTED WITH TWO THYMINE RESIDUES.	69
FIGURE 3.17: CD (A) AND NMR (B) SPECTRA OF <i>2xBLOCK 2Δ</i> SEQUENCES WITH THYMINE SUBSTITUTIONS CARRIED OUT ON BOTH BLOCKS. INCREMENTAL SUBSTITUTION OF THYMINE LOOPS, FROM ONE (BLUE) TO TWO (RED), THREE (GREEN) AND ALL FOUR (PURPLE) THYMINES SUBSTITUTED WITH CYTOSINE AT BOTH EQUIVALENT POSITIONS IN EACH BLOCK.	70
FIGURE 3.18: CD (A) AND NMR (B) SPECTRA OF <i>2xBLOCK 2Δ</i> SEQUENCES WITH THYMINE SUBSTITUTIONS CARRIED OUT ON BOTH BLOCKS. INCREMENTAL SUBSTITUTION OF THYMINE LOOPS, FROM ONE (BLUE) TO TWO (RED), THREE (GREEN) AND ALL FOUR (PURPLE) THYMINES SUBSTITUTED WITH ADENINE AT BOTH EQUIVALENT POSITIONS IN EACH BLOCK.	71
FIGURE 3.19: CD (A) AND NMR (B) SPECTRA OF <i>2xBLOCK 2Δ</i> SEQUENCES WITH THYMINE SUBSTITUTIONS CARRIED OUT ON BOTH BLOCKS. INCREMENTAL SUBSTITUTION OF THYMINE LOOPS, FROM ONE (BLUE) TO TWO (RED), THREE (GREEN) AND ALL FOUR (PURPLE) THYMINES SUBSTITUTED WITH TWO THYMINE RESIDUES AT BOTH EQUIVALENT POSITIONS IN EACH BLOCK.	72
FIGURE 3.20: CD (A) AND NMR (B) DATA FOR <i>Z-G4R</i> (GREEN) AND <i>Z-G4</i> (BLACK) FOLDED IN AMMONIUM BUFFER.	74
FIGURE 3.21: (A) RIBBON REPRESENTATION OF THE LEFT-HANDED G-QUADRUPLEX FORMED BY <i>Z-G4R</i> . THYMINE BASES ARE SHOWN IN ORANGE, GUANINES IN CYAN AND Ba^{2+} IONS AS GREEN SPHERES. (B) SCHEMATIC REPRESENTATION OF THE CRYSTAL STRUCTURE WITH NUMBERED GUANINE RESIDUES AND MARKED BACKBONE PROGRESSION DIRECTION. (C) CROSS-SECTION OF THE LEFT-HANDED G-QUADRUPLEX FORMED BY <i>Z-G4R</i> WITH THE 2MFO-FC MAP CONTOURED AT 1.0 Σ	76
FIGURE 3.22: CRYSTAL PACKING INTERACTIONS OF THE <i>Z-G4R</i> G-QUADRUPLEX. INTERACTING THYMINE RESIDUES ARE LABELED WITH THEIR POSITION IN THE PRIMARY SEQUENCE AND COLOR-CODED BASED ON DISCRETE INTERACTIONS.	77
FIGURE 3.23: <i>Z-G4R BLOCK 1</i> (A) AND <i>BLOCK 2</i> (B), ALIGNED WITH OTHER SOLVED LEFT-HANDED G-QUADRUPLEXES, <i>Z-G4</i> , <i>BLOCK 2</i> , <i>2xBLOCK 2</i> , <i>Z-G4R</i> , <i>Z-G4-BULGE</i> AND <i>Z-G4-2xBULGE</i> , WHICH ARE SHOWN PARTIALLY TRANSPARENT FOR CLARITY.	78
FIGURE 3.24: CARTOON REPRESENTATION OF THYMINE ADJUSTMENTS BETWEEN <i>Z-G4R</i> (T14 AND T17) AND <i>Z-G4</i> (T4 AND T1). BLACK ARROWS SHOW THE ADJUSTMENT FROM <i>Z-G4R</i> TO <i>Z-G4</i> . THYMINES ARE COLORED ORANGE, <i>ANTI</i> GUANINES CYAN AND <i>SYN</i> GUANINES MAGENTA.	78
FIGURE 3.25: CD (A) AND NMR (B) SPECTRA OF <i>BLOCK 1</i> SEQUENCES WITH ONE BULGE (PURPLE), TWO BULGES (ORANGE), THREE BULGES (GREEN) AND FOUR BULGES (BLUE) INTRODUCED.	80
FIGURE 3.26: CD (A) AND NMR (B) SPECTRA OF MODIFIED <i>Z-G4</i> SEQUENCES WITH ONE BULGE (PURPLE), TWO BULGES (ORANGE), THREE BULGES (GREEN) OR FOUR BULGES (BLUE) INTRODUCED IN <i>BLOCK 1</i>	81
FIGURE 3.27: (A) RIBBON REPRESENTATION OF THE LEFT-HANDED G-QUADRUPLEX FORMED BY <i>Z-G4-BULGE</i> . THYMINE BASES ARE SHOWN IN ORANGE, GUANINES IN CYAN, SPERMINES IN GREEN, K^+ IONS AS PURPLE SPHERES AND $O4'$ OXYGENS AS YELLOW SPHERES. (B) SCHEMATIC REPRESENTATION OF THE CRYSTAL STRUCTURE OF <i>Z-G4-BULGE</i> WITH NUMBERED GUANINE RESIDUES AND MARKED BACKBONE PROGRESSION DIRECTION. (C) CROSS-SECTION OF THE LEFT-HANDED G-QUADRUPLEX FORMED BY <i>Z-G4-BULGE</i> WITH THE 2MFO-FC MAP CONTOURED AT 1.0 Σ	83
FIGURE 3.28: CRYSTAL PACKING INTERACTIONS OF THE <i>Z-G4-BULGE</i> G-QUADRUPLEX. INTERACTING THYMINE RESIDUES ARE LABELED WITH THEIR POSITION IN THE PRIMARY SEQUENCE AND COLOR-CODED BASED ON DISCRETE INTERACTIONS.	85

FIGURE 3.29: <i>Z-G4-BULGE BLOCK 1 (A) AND BLOCK 2 (B)</i> , ALIGNED WITH OTHER SOLVED LEFT-HANDED G-QUADRUPLICES, <i>Z-G4</i> , <i>BLOCK 2</i> , <i>2xBLOCK 2</i> , <i>Z-G4R</i> AND <i>Z-G4-2xBULGE</i> , WHICH ARE SHOWN PARTIALLY TRANSPARENT FOR CLARITY.	86
FIGURE 3.30: (A) RIBBON REPRESENTATION OF THE LEFT-HANDED G-QUADRUPLIX FORMED BY <i>Z-G4-2xBULGE</i> . BOTH SUBUNIT A AND B OF THE ASYMMETRIC UNIT ARE SHOWN. THYMINE BASES ARE SHOWN IN ORANGE, GUANINES IN CYAN, SPERMINES IN GREEN, K^+ IONS AS PURPLE SPHERES AND $O4'$ OXYGENS AS YELLOW SPHERES. (B) SCHEMATIC REPRESENTATION OF THE CRYSTAL STRUCTURE OF <i>Z-G4-2xBULGE</i> WITH NUMBERED GUANINE RESIDUES AND MARKED BACKBONE PROGRESSION DIRECTION. (C) CROSS-SECTION OF THE LEFT-HANDED G-QUADRUPLIX FORMED BY <i>Z-G4-2xBULGE</i> WITH THE 2MFO-FC MAP CONTOURED AT $1.0 \text{ \AA}$	88
FIGURE 3.31: CRYSTAL PACKING INTERACTIONS OF THE <i>Z-G4-2xBULGE</i> G-QUADRUPLIX. INTERACTING THYMINE RESIDUES ARE LABELED WITH THEIR POSITION IN THE PRIMARY SEQUENCE AND COLOR-CODED BASED ON DISCRETE INTERACTIONS. VIEWS A AND B SHOW THE SAME MOLECULES WITH DIFFERENT ORIENTATIONS, IN ORDER TO VISUALIZE THE PACKING INTERACTIONS. .	90
FIGURE 3.32: <i>Z-G4-2xBULGE BLOCK 1 (A) AND BLOCK 2 (B)</i> EQUIVALENTS FROM BOTH SUBUNITS, ALIGNED WITH OTHER SOLVED LEFT-HANDED G-QUADRUPLICES, <i>Z-G4</i> , <i>BLOCK 2</i> , <i>2xBLOCK 2</i> , <i>Z-G4R</i> AND <i>Z-G4-BULGE</i> , WHICH ARE SHOWN PARTIALLY TRANSPARENT FOR CLARITY.	91
FIGURE 3.33: COMPARISON OF THREE DIFFERENT CONNECTING LOOP CONFIGURATIONS. (A) TT LOOP OF <i>Z-G4</i> , CONNECTING TWO INNER G-TETRADES. (B) T LOOP OF <i>Z-G4-BULGE</i> , CONNECTING TWO INNER G-TETRADES. (C) TT LOOP OF <i>2xBLOCK 2</i> , CONNECTING AN OUTER G-TETRAD TO AN INNER G-TETRAD. (D) TT LOOP OF <i>Z-G4R</i> CONNECTING THE TWO OUTER G-TETRADES AND SPANNING THE ENTIRE HEIGHT OF THE G-QUADRUPLIX.	93
FIGURE 3.34: ALIGNED <i>BLOCK 1</i> AND <i>BLOCK 2</i> EQUIVALENTS FROM THE SIX SOLVED LEFT-HANDED G-QUADRUPLICES. THERE ARE FIVE UNIQUE VERSIONS OF <i>BLOCK 1</i> AND NINE UNIQUE VERSIONS OF <i>BLOCK 2</i> . (A) ALIGNED <i>BLOCK 1</i> EQUIVALENTS WITH ANNOTATED DIFFERENCES BETWEEN THE STRUCTURES. (B) ALIGNED <i>BLOCK 2</i> EQUIVALENTS. <i>ANTI</i> GUANINE BASES ARE SHOWN IN CYAN, <i>SYN</i> GUANINES IN MAGENTA, THYMINE BASES IN ORANGE, BACKBONE IN GREY AND $O4'$ OXYGENS IN YELLOW.	94
FIGURE 3.35: SURFACE REPRESENTATIONS OF <i>BLOCK 1</i> EQUIVALENTS WITH T-CAP AND BULGE FEATURES OF <i>Z-G4 (A)</i> , <i>Z-G4R (B)</i> , <i>Z-G4-BULGE (C)</i> AND <i>Z-G4-2xBULGE (D)</i> . THYMINE BASES ARE SHOWN IN ORANGE, GUANINE BASES IN CYAN AND THE BACKBONE IN GREY. ASTERISKS DENOTE THE VACANT T-CAP POSITION.	95
FIGURE 4.1: SCHEMATIC REPRESENTATION OF THE <i>Z-G4</i> STRUCTURE. <i>ANTI</i> GUANINES ARE COLORED IN CYAN AND <i>SYN</i> GUANINES ARE COLORED IN MAGENTA. THE THREE POTASSIUM IONS ARE SHOWN AS PURPLE SPHERES COORDINATED BETWEEN THE STACKED G-TETRADES.	100
FIGURE 4.2: CD (A) AND NMR (B) SPECTRA OF <i>Z-G4</i> FOLDED IN BUFFERS WITH DIFFERENT CATIONS. NH_4^+ IS SHOWN IN BLUE, K^+ IN RED, CS^+ IN ORANGE, Na^+ IN PURPLE AND Li^+ IN GREEN.	100
FIGURE 4.3: CD (A) AND NMR (B) SPECTRA OF <i>Z-G4</i> FOLDED IN EITHER K^+ (RED) OR NH_4^+ (BLUE) BUFFER.	101
FIGURE 4.4: IMINO AND AMINO 1H NMR SPECTRA OF <i>Z-G4</i> BEFORE (A) AND TWO WEEKS AFTER (B) D_2O EXCHANGE EXPERIMENT. ONLY THE OUTER G-TETRAD IMINO PROTONS EXCHANGE AFTER TWO WEEKS, IDENTIFYING THE PROTECTED INNER G-TETRAD IMINO PROTONS.....	101
FIGURE 4.5: IMINO PEAK ASSIGNMENT OF <i>Z-G4</i> IN NH_4^+ . (A) DIRECT IMINO ASSIGNMENT USING ^{15}N LABELED OLIGONUCLEOTIDES. THE NH_4^+ IMINO SPECTRUM IS SHOWN IN BLUE, WHILE THE K^+ SPECTRUM IS SHOWN IN RED FOR COMPARISON. (B) NOESY SPECTRUM WITH IMINO-IMINO REGION ASSIGNMENTS. SEVEN IMINO PROTONS WITH DIRECT ASSIGNMENT ARE NUMBERED	

WITH BOLD AND UNDERLINED FONT, SIX IMINO PROTONS WITH NOESY ASSIGNMENT ARE NUMBERED WITH REGULAR FONT AND THE REMAINING THREE IMINO PROTONS ARE NUMBERED IN GREY.	102
FIGURE 4.6: NMR SPECTRA OF <i>Z-G4</i> FOLDED IN $^{15}\text{NH}_4^+$. (A) ^1H NMR SPECTRUM, (B) ^{15}N FILTERED NMR SPECTRUM AND (C) ^{15}N FILTERED NMR SPECTRUM WITH DECOUPLING.	103
FIGURE 4.7: DECOUPLED ^{15}N FILTERED NMR SPECTRA OF <i>Z-G4</i> FOLDED IN $^{15}\text{NH}_4^+$ BUFFER WITH EITHER (A) pH=5 OR (B) pH=6.9. PEAKS OF COORDINATED AMMONIUM IONS ARE LABELED WITH RED ASTERISK.	104
FIGURE 4.8: NOESY AND ^{15}N HSQC EXPERIMENTS CORRELATING THE POSITION OF ^1H NMR AMMONIUM ION PEAKS WITH THE POSITION OF THE AMMONIUM IONS WITHIN THE <i>Z-G4</i> STRUCTURE AT 283 K. OUTER AND INNER IMINO PROTON PEAK POSITIONS ARE LABELED IN THE NOESY SPECTRUM AS WELL AS THE POSITION OF INNER AND OUTER AMMONIUM ION. OUTER IONS ARE SHOWN IN PURPLE, WHILE THE INNER ION IS SHOW IN CYAN.	105
FIGURE 4.9: NzEx-HSQC 2D [107] SPECTRUM RECORDED WITH A 100 MS MIXING TIME AT pH = 5.1 AND 293 K. BULK, OUTER AND INNER $^{15}\text{NH}_4^+$ ION PEAKS ARE APPROPRIATELY ANNOTATED. THE CROSS-PEAKS BETWEEN THE OUTER AND BULK IONS ARE SHOWN, WITH A DASHED RED LINE CONNECTING THEM.	106
FIGURE 4.10: TRACKING OF AMMONIUM PROTON PEAK DISAPPEARANCE WITH ^{15}N FILTERED NMR. (A) AMMONIUM ^1H NMR PEAK REGION. FIVE SPECTRA ARE SHOWN REPRESENTING DIFFERENT TIME POINTS DURING THE EXPERIMENT. THE RED ARROW SHOWS THE DISAPPEARANCE OF THE INNER AMMONIUM ION PEAK WITH TIME AS THE HYDROGEN IS EXCHANGED WITH DEUTERIUM. (B) GRAPH OF INNER AMMONIUM ION AREA VERSUS TIME SHOWS A RESIDENCE TIME OF $T = 79 \pm 3$ MIN.	107
FIGURE 4.11: TRACKING OF AMMONIUM PROTON PEAK DISAPPEARANCE WITH ^{15}N FILTERED NMR. (A) AMMONIUM ^1H NMR PEAK REGION. FIVE SPECTRA ARE SHOWN REPRESENTING DIFFERENT TIME POINTS DURING THE EXPERIMENT. THE RED ARROW SHOWS THE DISAPPEARANCE OF THE INNER AMMONIUM ION PEAK WITH TIME AND IS PLOTTED ON THE GRAPH WITH $T = 55 \pm 1$ MIN (B).	108
FIGURE 4.12: TRACKING OF IMINO PROTON PEAK APPEARANCE AND DISAPPEARANCE WITH ^1H NMR. (A) INNER G-TETRAD ^1H NMR IMINO PROTON PEAK REGION. THREE SPECTRA ARE SHOWN REPRESENTING DIFFERENT TIME POINTS DURING THE EXPERIMENT. THE RED ARROW SHOWS THE APPEARANCE OF THE INNER G-TETRAD IMINO PROTON PEAK WITH TIME WHILE THE BLUE ARROW SHOWS THE DISAPPEARANCE OF THE INNER G-TETRAD IMINO PROTON PEAK WITH TIME AND BOTH ARE PLOTTED ON THE GRAPH (B).	109
FIGURE 4.13: COMPARISON OF INNER $^{15}\text{NH}_4^+$ ION PEAK POSITION IN DECOUPLED ^{15}N FILTERED NMR SPECTRA, DEPENDING ON THE CONCENTRATION OF ADDITIONALLY ADDED K^+ IONS. THE SPECTRUM FROM FIGURE 4.14 IS SHOWN IN RED, VERTICAL DASHED LINES REPRESENT DIFFERENT SPECIES, WITH THE COORDINATED IONS LISTED AT THE TOP.	110
FIGURE 4.14: TRACKING OF AMMONIUM PROTON PEAK APPEARANCE WITH ^{15}N FILTERED NMR. (A) AMMONIUM ^1H NMR PEAK REGION. FIVE SPECTRA ARE SHOWN REPRESENTING DIFFERENT TIME POINTS DURING THE EXPERIMENT. THE BLUE AND RED ARROWS SHOW THE APPEARANCE OF THE INNER $^{15}\text{NH}_4^+$ ION PEAKS WITH TIME AND IS PLOTTED ON THE GRAPH (B) WITH RESIDENCE TIMES $T = 49 \pm 2$ MIN AND $T = 44 \pm 2$ MIN, RESPECTIVELY.	111
FIGURE 4.15: TRACKING OF IMINO PROTON PEAKS APPEARANCE AND DISAPPEARANCE WITH ^1H NMR. THE TOP AND BOTTOM SPECTRA ARE OF <i>Z-G4</i> IN EITHER K^+ OR $^{15}\text{NH}_4^+$ ION. THE MIDDLE TWO SPECTRA ARE FROM THE START ($T = 3$ MIN) AND END ($T = 210$ MIN) OF THE $\text{K}^+ \rightarrow ^{15}\text{NH}_4^+$ EXPERIMENT.	112
FIGURE 4.16: NATIVE NEGATIVE-ION ELECTROSPRAY IONIZATION-MASS SPECTROMETRY DATA FOR <i>Z-G4</i> FOLDED IN K^+ (A), SUBSEQUENTLY DILUTED WITH NH_4^+ BUFFER AND TRACKED WITH TIME (B). THE 3 NH_4^+ SPECIES IS ENCIRCLED WITH A RED BOX	

WHILE THE $2 \text{NH}_4^+ + 1 \text{K}^+$ SPECIES WITH A BLUE BOX. TIME OF SPECTRUM COLLECTION IS SHOWN ON THE LEFT. THE FOUR ION SPECIES ARE MARKED; $3 \text{NH}_4^+ + 1 \text{K}^+$ SPECIES IS MARKED IN GREEN WHILE THE $2 \text{NH}_4^+ + 2 \text{K}^+$ IS MARKED IN PURPLE.....	113
FIGURE 4.17: SCHEMATIC REPRESENTATION OF THE DIFFERENT ION EXCHANGE RATES. IN ORANGE IS THE EXCHANGE FROM $^{15}\text{NH}_4^+$ TO K^+ AND IN BLUE IS THE EXCHANGE FROM K^+ TO $^{15}\text{NH}_4^+$	116

Table of tables

TABLE 1.1: TABLE OF <i>AGRO100</i> DERIVED SEQUENCES.....	41
TABLE 2.1: BUFFERS USED IN SAMPLE PREPARATION.	46
TABLE 2.2: SEQUENCES USED IN THIS WORK (5'-3') <i>SANS</i> SUBSTITUTION SEQUENCES.	47
TABLE 2.3: SEQUENCES USED IN SUBSTITUTION STUDIES (5'-3') OF <i>2xBLOCK 2A</i>	48
TABLE 2.4: DATA COLLECTION AND REFINEMENT STATISTICS FOR ALL SOLVED STRUCTURES PRESENTED IN THIS WORK, <i>BLOCK 2</i> , <i>2xBLOCK 2</i> , <i>Z-G4R</i> , <i>Z-G4-BULGE</i> AND <i>Z-G4-2xBULGE</i> , RESPECTIVELY.	50
TABLE 3.1: LIST OF <i>BLOCK 1</i> AND <i>BLOCK 2</i> SEQUENCES WITH TERMINAL-THYMINE MODIFICATIONS.	56
TABLE 3.2: DATA COLLECTION AND REFINEMENT STATISTICS FOR <i>BLOCK 2</i> (PDB ID 6FQ2).	58
TABLE 3.3: AVERAGE BACKBONE DIHEDRAL ANGLES OF <i>BLOCK 2</i> AND <i>Z-G4</i>	60
TABLE 3.4: LIST OF <i>2xBLOCK 2</i> SEQUENCES WITH TERMINAL-THYMINE MODIFICATIONS.	62
TABLE 3.5: DATA COLLECTION AND REFINEMENT STATISTICS FOR <i>2xBLOCK 2</i> (PDB ID 6GZ6).	64
TABLE 3.6: AVERAGE BACKBONE DIHEDRAL ANGLES OF <i>2xBLOCK 2</i> AND <i>Z-G4</i>	65
TABLE 3.7: RESULTS OF <i>2xBLOCK 2Δ</i> THYMINE LOOP SUBSTITUTIONS.....	73
TABLE 3.8: LIST OF SEQUENCES USED IN EXAMINING THE EFFECTS OF BLOCK ORDER ON LEFT-HANDED G-QUADRUPLEX FORMATION.	74
TABLE 3.9: DATA COLLECTION AND REFINEMENT STATISTICS FOR <i>Z-G4R</i>	75
TABLE 3.10: AVERAGE BACKBONE DIHEDRAL ANGLES OF <i>Z-G4R</i> AND <i>Z-G4</i>	76
TABLE 3.11: SEQUENCES USED FOR EXPLORING THE EFFECT OF SPLITTING GUANINE PAIRS IN <i>BLOCK 1</i> WITH THYMINE RESIDUES (SHOWN IN RED).	79
TABLE 3.12: DATA COLLECTION AND REFINEMENT STATISTICS FOR <i>Z-G4-BULGE</i>	82
TABLE 3.13: AVERAGE BACKBONE DIHEDRAL ANGLES OF <i>Z-G4-BULGE</i> AND <i>Z-G4</i> FOR COMPARISON.	84
TABLE 3.14: TABLE OF PACKING INTERACTIONS OF <i>Z-G4-BULGE</i>	85
TABLE 3.15: DATA COLLECTION AND REFINEMENT STATISTICS FOR <i>Z-G4-2xBULGE</i>	87
TABLE 3.16: AVERAGE BACKBONE DIHEDRAL ANGLES OF <i>Z-G4-2xBULGE</i> AND <i>Z-G4</i>	89
TABLE 3.17: TABLE OF PACKING INTERACTIONS OF <i>Z-G4-2xBULGE</i>	90
TABLE 3.18: AVERAGE BACKBONE DIHEDRAL ANGLES OF <i>BLOCK 2</i> , <i>2xBLOCK 2</i> AND <i>Z-G4</i>	96
TABLE 3.19: RESULTS OF <i>2xBLOCK 2Δ</i> THYMINE LOOP SUBSTITUTIONS.....	97

Abbreviations

1D	One-Dimensional
2D	Two-Dimensional
A	Adenine
ADSC	Area Detector Systems Corporations
C	Cytosine
CCD	Charge-Coupled Device
CD	Circular Dichroism
COSY	COrelated SpectroscopY
CPL	Circularly Polarized Light
D.I.T	Digital Integration Time
DNA	DeoxyriboNucleic Acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
ESI	ElectroSpray Ionization
ESI-MS	ElectroSpray Ionization-Mass Spectrometry
G	Guanine
G4	G-quadruplex
G-tetrad	G•G•G•G G-tetrad
G-tract	Guanine tract
HSQC	Heteronuclear multiple-bond correlation
IDT	Integrated DNA Technologies
IM-Q-TOF	Ion Mobility Quadrupole Time of Flight
LHG4	Left-Handed G-quadruplex
MHz	MegaHertz
min	minute
mM	milliMolar
MR	Molecular Replacement
mRNA	messenger RNA
MS	Mass Spectroscopy
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect SpectroscopY
PDB	Protein Data Bank
ppm	parts per million
RMSD	Root-Mean-Square Deviation
RNA	RiboNucleic Acid
SPM	spermine
T	Thymine
T-cap	Thymine cap
TMAA	trimethylammonium acetate
TOF	Time of Flight
Ts	Thymines
UV	Ultraviolet
XDS	X-ray Diffraction Spectroscopy

Abstract

Aside from the double-stranded helical B-DNA structure, DNA can form four-stranded G-quadruplex structures composed of stacked guanine G-tetrads. A G-tetrad is a planar arrangement of four Hoogsteen bonded guanine bases. The effects these structures have on cellular processes, as well as their potential for targeted therapeutic application, have been previously reported.

Until the recently reported, 2015 structure of *Z-G4*, all solved G-quadruplex structures adopted a right-handed helical fold. *Z-G4* is the first example of a left-handed G-quadruplex structure and it was solved by both NMR spectroscopy and X-ray crystallography.

Z-G4 is composed of four G-tetrads, arranged in two contiguous, stacked blocks with three coordinated ions between each pair of G-tetrads. The G-tetrads in *Z-G4* have a unique left-handed twist which gives *Z-G4* a distinct CD spectrum. The NMR spectrum also possesses some as well as a recognizable structural features.

We report on the discovery of five new left-handed G-quadruplex structures, including a minimal left-handed motif. They display both conserved and unique features, when compared to *Z-G4*. These structures and associated findings help pave the way towards expanding the primary sequence requirements for left-handed G-quadruplexes.

Cations are important for G-quadruplex assembly as they bind between stacked G-tetrads and coordinate with guanine carbonyl oxygens. Using NMR spectroscopy and mass spectrometry, we have observed a cation coordinated in *Z-G4* that has an exchange time on the order of half an hour at room temperature. To our knowledge, an ion with such a long exchange time has not been reported in literature previously.

Chapter 1: Introduction

1.1 DNA role and structure

One of the most significant advancements in biology came in 1953 when James Watson and Francis Crick revealed the first correct model of the double helical DNA (**Figure 1.1, B**), which became known as B-DNA [1]. Its structure was solved with significant contributions from Rosalind Franklin [2] and Maurice Wilkins [3]. In parallel with the discovery of DNA structure, a seminal paper was published by Alfred Hershey and Martha Chase proving that DNA, not proteins, was the carrier of hereditary information [4]. These two discoveries paved the way towards our modern understanding of biology.

Figure 1.1: (A) Schematic of two base-paired antiparallel complementary strands of DNA. (B) Watson and Crick's model of DNA double helix [1].

DNA is a polymer of four different nucleotides, arranged in two complementary strands (**Figure 1.1, A**) and was initially considered to be in the form of double helical B-DNA (**Figure 1.1, B**). Discoveries of other DNA secondary structures soon followed which expanded the polymorphism and structural complexity of DNA [5-7]. The conditions for the formation of alternative secondary structures are influenced by the DNA primary sequence [6-10], the local ionic conditions [11-13] and torsion stresses built up in the DNA strand [14-19]. The biological importance of alternate DNA structures was not recognized until it was shown that alternative DNA structures specifically interact with proteins and thus influence a wide variety of cellular processes [10, 20-24].

1.2 G-quadruplex features and properties

DNA and RNA G-quadruplexes are a well-studied subgroup of DNA and RNA structures, formed by short guanine rich sequences [25] that have been shown to adopt a diverse range of structure conformations [26] (**Figure 1.2**).

Figure 1.2: An example of diverse range of structures adopted by guanine rich DNA and RNA oligonucleotides.

Four guanine bases can hydrogen bond through alternative Hoogsteen base pairing to assemble a planar G-tetrad, a key component of the G-quadruplex structures (**Figure 1.3, A**). G-quadruplex structures have two or more stacked G-tetrads that comprise the core of the G-quadruplex (**Figure 1.3, B**).

Figure 1.3: (A) Stick representation of a G-tetrad formed by four planary arranged guanine bases. Hoogsteen hydrogen bonds are shown as dotted lines, oxygen atoms are shown in red, hydrogens in grey, nitrogen in deep blue, carbons in cyan and the sodium ion is represented as a purple sphere. (B) Example of a G-quadruplex schematic, containing three stacked G-tetrads. Guanine bases are shown as cyan rhombi.

The stacked G-tetrad core has been found stabilized by cations that interact with the guanine carbonyl oxygen (**Figure 1.3, A**). The Hoogsteen bonded guanine bases in the G-tetrad core differ among solved G-quadruplex structures. There are three main points of difference between solved structures; the stacking orientation between G-tetrads, the twist angle between G-tetrads and the inter-tetrad distance. A G-tetrad composed of Hoogsteen bonded guanines is asymmetric and has a “head” (denoted as H) and “tail” (denoted T) face defined by the direction of the donor-to-acceptor progression of hydrogen bonds. When viewed from the top a clockwise progression means that the “head” side is facing the viewer; the reverse side is referred to as “tail” and the hydrogen bond progression is seen to be counter-clockwise (**Figure 1.4**).

Figure 1.4: Two different sides or faces of the G-tetrad as determined by the Hoogsteen hydrogen bond directionality. **(A)** Top view of the “head” face of the G-tetrad. **(B)** Top view of the “tail” face of the G-tetrad.

This leads to three different stacking arrangements of G-tetrads; “H to H”, “H to T” and “T to T” [27, 28]. “H to T” G-tetrad stacking is known as “same-polarity” or homopolar stacking while “H to H” and “T to T” are known as opposite-polarity or heteropolar stacking (**Figure 1.5**).

Figure 1.5: Same-polarity or homopolar **(A)** and opposite-polarity or heteropolar **(B)** stacking of two G-tetrads.

The second point of diversity between stacked G-tetrads is the twist angle between G-tetrads (θ_{tet}) (**Figure 1.6, left**) which usually ranges from 10 ° to 37 ° in a clockwise, right-handed orientation [29]. Both of these factors are linked with and co-influence the inter-tetrad distance (d_{tet}), which is usually found to be between 3-3.7 Å (**Figure 1.6, right**).

Figure 1.6: (A) Representation of inter-tetrad rotation angle (θ_{tet}) and (B) G-tetrad separation distance (d_{tet}).

A source of polymorphism between solved G-quadruplex structures is the direction of backbone progression and the type and number of loops that connect G-tracts. This influences the sugar and strand orientation and molecularity of assembled G-quadruplex structures [23, 26, 30-34]. G-quadruplex structures also differ in the number of stacked G-tetrads, which is mostly determined by the number of consecutive guanine nucleotides in the four G-tracts. The usual number of stacked G-tetrads among solved G-quadruplex structures is between two and four, with three stacked G-tetrads being the most common.

Figure 1.7: Different types of G-quadruplex grooves adopted due to *anti* and *syn* guanine glycosidic conformations. **(A)** The G-tetrad core (left) and a single G-tetrad with annotated groove widths in respect to the glycosidic torsion angle conformation (right). **(B)** *Anti* and *syn* glycosidic torsion angle conformation of guanines.

Guanines comprising the G-tetrad core can be in either *syn* or *anti* glycosidic conformation which influences the width of grooves produced by the backbone (**Figure 1.7, A**). We can separate them into three categories; wide (*anti-syn*), medium (*anti-anti* or *syn-syn*) or narrow (*syn-anti*) (**Figure 1.7, A**). The arrangement of grooves in a folded G-quadruplex structure influences the surface properties such as positions and proximity of charged phosphate groups and accessibility to guanine bases, both of which can influence interactions with other biomolecules or be exploited when designing G-quadruplex binding molecules.

Figure 1.8: Structural elements that give rise to differing G-quadruplex structures. **(A)** Different combinations of backbone progression orientations. **(B)** Different types of connecting loops.

The backbone connecting the four contiguous stretches of guanines has 5' to 3' orientation of progression [23, 26, 30-34]. Depending on the relative orientations of strands to each other, there are four possible arrangements (**Figure 1.8, A**):

- One parallel arrangement (4+0) where all four the strands progress in the same direction.
- Two anti-parallel arrangements (2+2) where two strands progress in the opposite direction when compared to the other two strands. This can further be subdivided into two categories, two strands of the same directionality on diagonal corners of the G-tetrad, or on the same side of the G-tetrad.
- One hybrid arrangement (3+1) where one strand has the opposite direction of progression when compared to the other three strands.

The four backbone strands comprising the G-quadruplex can be connected with loops, depending on the molecularity of the structure. In the case of monomolecular quadruplexes we would have three loops connecting the four guanine stretches. Depending on how these

loops connect the backbone strands (**Figure 1.8, B**), we can classify loops into three categories:

- Double-chain reversal loops (propeller loops), connecting two adjacent strands with the same direction.
- Edgewise loops, connecting two adjacent strands of opposite directions.
- Diagonal loops, connecting two strands on opposite corners (strands also opposite in direction).

What types of loops a G-quadruplex has is influenced by the primary sequence and also the conditions in which the G-quadruplex is folded. The most favored type for single-residue loops seems to be the propeller type loop [23, 26, 30-34].

Another feature that has broadened the putative G-quadruplex sequence space is the discontinuity, also known as the bulge [35, 36]. Bulges allow for less restrictions when searching for putative G-quadruplex forming sequences, as guanine stretches do not have to be continuous. Besides single nucleotide bulges they can be varied both in length and composition, as well as the number of bulges per G-quadruplex [36]. They have been shown to exist in RNA G-quadruplexes as well [37, 38].

Figure 1.9: Example of the bulge/discontinuity feature in a G-quadruplex, compared to a regular loop.

Bulges differ from loops in one significant way, whereas loops connect continuous stretches of guanines, bulges disrupt the continuity of guanine stretches (**Figure 1.9**). Bulges do influence the thermal stability of G-quadruplexes, however, as they have been shown to lower their melting temperature [36].

1.3 Circular dichroism of G-quadruplexes

Nucleic acids absorb light in the near- and far-UV spectral region. The far-UV contributions, below the wavelength of 200 nm, are mainly composed of phosphate group contributions and sugar (ribose in RNA and deoxyribose in DNA) $\sigma \rightarrow \sigma^*$ transitions [39-42]. Due to strong background absorbance of molecular oxygen and many co-solutes like chloride ions and buffers in the far-UV region, the study of nucleic acids in this spectral region is limited to brilliant light sources like synchrotrons or very specific sample preparation. Nucleobases, however, also strongly absorb in the near-UV region between 220-320 nm, due to characteristic $\pi \rightarrow \pi^*$ transitions [40-43], where contributions from molecular oxygen and many commonly used co-solutes is minimal. The absorbance of nucleic acid solutions at 260 nm is commonly used for determining their solution concentration using UV spectrometers. This is calculated using the measured sample absorbance at 260 nm, path-length (in cm) and the sequence specific extinction coefficient at 260 nm (ϵ_{260} in $M^{-1} cm^{-1}$) to obtain strand molar concentration.

UV Circular dichroism (CD) spectroscopy is an extension of UV spectroscopy. Non-polarized UV light of a specific wavelength is passed through a linear polarizer and a quarter-wave plate which converts the linearly polarized incident light into circularly polarized light (**Figure 1.10**). The circularly polarized light (CPL) can be either left- or right-handed. The electric field vector of each traces out a left- or right-handed helix, respectively, while propagating through space and this makes CPL itself chiral.

Figure 1.10: Unpolarized light of a specific wavelength is passed through a linear polarizer. Linearly polarized light is passed through a quarter-wave plate, making it circularly polarized. The light wave propagating through space is shown in red and the strength and direction of the electric field vector is shown as black vectors. In this example the circularly polarized light traces a left-handed, counterclockwise helix.

A chiral and absorbing sample interacts with circularly polarized light (CPL) in a unique way. Left- and right-CPL beams are absorbed to a different extent due to $\epsilon_L \neq \epsilon_R$. The difference in absorbance of the left- and right-CPL component produces $\Delta A = A_L - A_R$ and is directly related to $\Delta\epsilon = \epsilon_L - \epsilon_R$. This difference in absorbance is the circular dichroism that commercial spectrometers are designed to detect when they pass periodically oscillating left- and right-CPL through a sample and record the difference in absorbance of each.

Figure 1.11: (A) Vector depiction of relationship between linearly polarized and left/right-circularly polarized light of equal amplitudes. (B) Elliptically polarized light (dashed) is produced by combining left- and right-circularly polarized light with different amplitudes.

Combining two oppositely circularized light beams of equal amplitude produces a linearly polarized light beam (**Figure 1.11, A**). After a light beam of periodically oscillating left- and right-circularly polarized rays passes through a chiral absorbing sample, the amplitude of one component is reduced. As a result, combining the two circularly polarized components produces an elliptically polarized light beam (**Figure 1.11, B**). Ellipticity (Θ) is calculated from arctangent of the short divided by the long axis of the ellipse, $\Theta = \arctan(b/a)$. Ellipticity (Θ), measured in millidegrees is usually given when recording CD spectra. Most commonly, this is converted to molar ellipticity which calibrates the CD spectrum to sample concentration and has $10^3 \text{ deg M}^{-1} \text{ cm}^{-1}$ as units.

Figure 1.12: (A) Two chromophores with individual but equal excitation states energies (green) couple and produce two exciton-coupled and distinct excited state energy levels (red and blue). (B) Representation of geometric parameters that influence the amplitude and sign of CD spectra produced by coupled chromophores. Expected absorption spectrum (C) and CD spectrum (D) produced by two chromophores with a negative chirality arrangement and exciton splitting as shown in (B). Absorption of individual exciton-coupled excited states is shown in red and blue, while the resulting absorption spectrum and CD spectrum are shown in purple. Figure adapted from [44].

When two chromophores are close in space and their relative positions to each other are well defined they no longer have individual excited states, rather they produce a degenerate exciton-coupled state (**Figure 1.12, A**) [44, 45]. This state has two absorption levels, one redshifted (red) and the other an equal amount blueshifted (blue) in energy when compared to the individual excited states (green) (**Figure 1.12, A**). The amplitude of the CD signal produced by coupled chromophores depends on the interchromophoric distance r_{12} and the angles between the chromophores (**Figure 1.12, B**) while the sign of the CD spectrum depends on the angle denoted as gamma (γ). Spatially defined and coupled chromophores with a negative, counterclockwise chirality (negative γ) will preferentially absorb photons of left-CPL at the redshifted energy level and photons of right-CPL at the blueshifted energy level (**Figure 1.12, D**). The resulting CD spectrum (purple) has a dip at higher wavelength (lower energy)

and a peak at lower wavelength (higher energy). For chromophores with positive, clockwise chirality the opposite is the case.

The guanine base has two well defined and separated $\pi \rightarrow \pi^*$ absorption bands at 279 nm and 248 nm [42]. Guanine bases assemble in planar G-tetrads, which in turn stack to produce the core of the G-quadruplex structure. Guanine bases are very constrained in their movement and rotation once assembled in the G-quadruplex. Their absorption bands become coupled to produce distinct exciton-coupled absorption bands (**Figure 1.12, A**) that have preferential absorbance for left- and right-CPL thus producing a CD signal.

In G-quadruplexes the interchromophoric distance (r_{12}) and gamma (γ) (**Figure 1.12, B**) can be represented by G-tetrad distance (d_{tet}) and G-tetrad twist (Θ_{tet}), respectively [27, 46]. The contribution of nucleobases in loops to the CD spectrum is usually small as their position is not as constrained relative to G-tetrads. The contribution of phosphate and sugar backbone is also small in the near-UV region, as their absorption bands are below 200 nm.

Besides the G-tetrad distance and G-tetrad twist, the polarity of stacked G-tetrads, which can be same- or opposite-polarity, is the third factor that has a major influence on the CD spectral shape in the 220-320 nm region [27, 47, 48]. The polarity of stacked quartets and the G-tetrad twist angle, not the relative orientations of the backbone strands, is likely to be the determining factor for the shape of G-quadruplex CD spectra [27].

A G-quadruplex with a right-handed G-tetrad twist (positive twist angle) and same-polarity stacking of G-tetrads is commonly referred to as a right-handed parallel G-quadruplex [48, 49]. It produces a CD spectrum with a peak at 260 nm and a dip at 240 nm (**Figure 1.13, blue**). This is a result of the 248 nm guanine absorption band being exciton-coupled, positive chirality and same-polarity stacking of guanines.

A G-quadruplex with a right-handed G-tetrad twist (positive twist angle) and opposite-polarity stacking of G-tetrads is commonly referred to as a right-handed antiparallel G-quadruplex [48, 50, 51]. It produces a CD spectrum with a dip at 260 nm and a peak at 240 nm with an additional peak at 290 nm (**Figure 1.13, green**). The 260 nm dip and 240 nm peak are a result of the 248 nm guanine absorption band being exciton-coupled, positive chirality and opposite-polarity stacking of guanines. The opposite-polarity stacking of guanines inverts the CD spectrum around the 248 nm absorption band as compared to same-polarity stacking of

guanines. The additional 290 nm peak is due to the 279 nm absorption band of guanine [27, 48].

Figure 1.13: Molar CD spectra of different types of G-quadruplexes. Parallel right-handed (blue), antiparallel right-handed (green), hybrid parallel/antiparallel right handed (magenta) and parallel left-handed G-quadruplex (red).

The third group of observed G-quadruplexes are hybrids of the two classes above as they have both opposite- and same-polarity stacking of G-tetrads, while still retaining the right-handed G-tetrad twist (positive twist angle) [32, 48, 52]. As a result their CD spectra produce a dip at 240 nm and a peak at 260 nm similar to right-handed parallel G-quadruplexes, while also having a 290 nm peak characteristic of right-handed antiparallel G-quadruplexes (**Figure 1.13, magenta**)[32, 48].

Recently a fourth group of G-quadruplexes has emerged. A G-quadruplex with a left-handed G-tetrad twist (negative twist angle) and same-polarity stacking of G-tetrads was observed [53] (Chapter 1.9, page 41). Named Z-G4 it is a left-handed parallel G-quadruplex and has a CD spectrum with a dip at 260 nm and a peak at 240 nm (**Figure 1.13, red**). This is a result of the 248 nm guanine absorption band being exciton-coupled, negative chirality and same-polarity stacking of guanines. Its CD spectrum is mirroring the CD spectra observed for right-handed parallel G-quadruplex structures.

1.4 NMR of G-quadruplexes

NMR is a very convenient tool for the study of G-quadruplex structures [54]. The most important proton, especially in the initial screenings of samples is the guanine H1 or imino proton (**Figure 1.14, A**). When a G-quadruplex is formed, guanines assemble into tetrads using Hoogsteen hydrogen bonds [10, 55]. The hydrogen bonding of the guanine imino proton to the O6 carbonyl oxygen of the neighboring guanine base slows down the exchange of imino protons with the solvent, making them visible on NMR time scales (**Figure 1.14, B**). They appear as sharp peaks between 10.5 - 12.5 ppm in the 1D ^1H NMR spectrum, which is very convenient, as Watson-Crick base paired imino protons appear between 12.5 – 14.5 ppm. This makes the G-quadruplex imino signature easily distinguishable from other assembled DNA structures and also enables differentiation between G-quadruplex structures. As long as there is only one G-quadruplex structure in solution, in almost all cases (one of the exceptions being a G-quadruplex with an abasic tetrad [56-58]) imino protons show up in multiples of four, representing the number of assembled tetrads.

Figure 1.14: (A) Labeled hydrogens and heteroatoms of the guanine nucleotide. The imino proton is shown in red. (B) Four Hoogsteen bonded guanine nucleobases assemble to form a tetrad. The hydrogen bonding of imino protons (red) slows their exchange with the solvent, making them observable in NMR.

Another notable protons in the context of this work are the two amino protons H22 and H21 (**Figure 1.14**). In a tetrad, one of them is hydrogen bonded to the N7 nitrogen of the neighboring guanine residue, while the other is not, making the amino group rotate. Due to this and the exposure to the solvent, where hydrogens are exchanged with water molecules, the amino group is usually not visible in NMR. Only when the second proton of the amino group is hydrogen bonded or the rotation slowed down, can both become visible in NMR experiments [53, 59]. This can be useful as a probe for local structure and ligand binding [59], as well as a fingerprint of specific structure formation [53].

Other hydrogen protons (aromatic protons of nucleobases and aliphatic protons of sugars and nucleobases) are very useful for solving G-quadruplex structures using a multitude of 2D NMR experiments, where cross peaks between protons indicate their proximity through bonds (COSY-based pulse sequences) [60] or through space (NOE-based pulse sequences) [54]. Since G-quadruplexes are synthesized as short oligonucleotides, partial site specific labeling can be employed to enrich a nucleotide at a specific position with ^{15}N and ^{13}C NMR active nuclei [61]. Recording NMR spectra that use ^{15}N or ^{13}C filtering, peaks in the 1D spectrum can be unambiguously assigned to a specific residue in the primary sequence. Once spectra have been assigned, the information about the presence and strength of cross peaks in NOESY spectra with different delay times can be used as restraints in energy minimization software such as XPLOR-NIH [62] together with parametrized force fields such as AMBER [63]. The end result is usually an ensemble of 10 structures that have the lowest energy, based on the provided cross peak restraints, which can be deposited to PDB¹ as a solved structure.

1.5 X-ray diffraction spectroscopy of G-quadruplexes

X-ray diffraction spectroscopy (XDS) has been successfully applied to solve many G-quadruplex structures (over 100 per Nucleic Acid Database²). The principles for solving G-

¹ <https://www.rcsb.org/>

² <http://ndbserver.rutgers.edu/>

quadruplex structures using XDS are identical to those employed for solving structures of other biomolecules such as proteins [64], RNA [65] and DNA [66].

Samples preparation is similar as for other techniques with special attention being given to sample concentration (usually between 0.5 – 3 mM), sample homogeneity and buffer used. When studying DNA, phosphate buffer is commonly used but it is generally avoided when crystallizing DNA molecules as it has a propensity to co-crystallize with divalent cations before the sample crystallizes, thus producing false-positives [67]. It is commonly replaced with the methyl arsonate buffer (cacodylate buffer). The sample is then mixed with commercial screens which have been optimized to screen for many crystallization conditions, which have been successfully employed in the past [68]. Most commonly the hanging or sitting drop vapor diffusion methods are used in a sealed container (**Figure 1.15, A**). The volatile components of the crystallization screen diffuse between the sample drop and the reservoir. This causes the concentration of the precipitating agent in the screen to rise, supersaturating the sample. If the conditions are right, nucleation sites will form, around which the sample crystal will grow (**Figure 1.15, B**). Optimizing the crystallization conditions usually involves controlling the speed of this process; if it is too fast, the crystal packing can be irregular, leading to badly diffracting crystals or precipitation of the sample.

Figure 1.15: (A) Schematic representation of the sitting drop vapor diffusion crystallization container. (B) A simple phase diagram showing the zones that are involved in crystal growth. The circle (1) represents the starting position, before vapor diffusion increases the concentration of both the precipitant and DNA to the supersaturation point (square, 2) where spontaneous homogeneous nucleation occurs. Once the nucleation sites are formed, the crystals grow until the solution reaches the equilibrium point (red star, 3) where the sample is neither supersaturated nor undersaturated.

Once crystals of sufficient size and quality have been obtained, they are shot with X-rays, either on in-house X-ray light sources or on synchrotrons, generating a series of diffraction images, each containing many diffraction spots (**Figure 1.16, A**). Diffraction spots arise when a specific set of circumstances are satisfied:

- Regular crystal packing and uniformity across unit cells (**Figure 1.16, B**). This is necessary since the crystal acts as an amplifier, where each unit cell generates the diffraction signal. The more unit cells in a crystal, the stronger the diffraction signal will be and the higher the possible resolution of the data obtained.
- Diffraction spots are a product of constructive interference (**Figure 1.16, C**). The diffracted incident X-rays interfere destructively everywhere except where Bragg's law is satisfied:

$$n\lambda = 2d \sin \theta$$

In essence Bragg's law tells us at which angle (θ) and plane distance (d), the crests and valleys from one X-ray scatterer will match up with the crests and valleys of the other X-ray scatterer (being in phase), thus giving rise to constructive interference (**Figure 1.16, C**) and producing a diffraction spot.

Figure 1.16: (A) Diffraction spots arise as a pattern from the (B) regular, repeating crystal lattice that acts as an amplifier to (C) the constructive interference, produced by atoms absorbing and re-scattering X-ray radiation, in accordance with Bragg's law.

To collect the full set of reflections, the crystal is rotated with the diffraction data usually collected every 0.1-0.5 ° on the newest EIGER counters [69]. Based on the diffraction patterns of the collected images, a list of likely unit cell symmetries with corresponding cell dimensions and angles is generated (indexing) [70]. Using this parameters, the data from all the images is

integrated, where the intensities and Miller indices are assigned to each reflection, followed by scaling and merging that adjusts the intensities of symmetry connected diffraction spots and reduces the data, based on the selected space group [71]. From this data, the Patterson function [72, 73] can be employed in order to generate a Patterson map, a crucial step in solving the structure, using only the Miller indices and intensities without the phase information.

Since G-quadruplexes are usually not large structures, rarely exceeding 20 kDa, the method most commonly employed to solve the structure from the data obtained is Molecular Replacement (MR) [74-76]. This works particularly well for G-quadruplexes, as MR requires a known similar structure. The known structure is placed in the unit cell, determined previously from the data obtained and its Patterson map is generated. The newly generated Patterson map is then compared to the Patterson map of the collected data using programs like PHASER [77]. Based on the map overlap, the known structure is rotated and translated, a new map generated and compared against the Patterson map of the collected data. This is repeated until a good fit is found, which means the known structure is properly placed in the unit cell and represents an initial solution to the collected data. The final solution is obtained after many steps of refinement, where ions/residues/ligands/waters are placed/removed in the unit cell, until a good fit with the electron density map is obtained, using programs like PHENIX [78]. The quality of the fit is determined by using “quality of fit parameters” R_{free} and R_{work} [79]. The optimization is finished when these two numbers cannot be improved upon further, after which the structures is deposited to a structural database such as PDB³ for final validation & publication.

1.6 Mass spectrometry of G-quadruplexes

Mass spectrometry (MS) at its core relies on Newton’s second law of motion and the Lorentz force law. If we put a particle of certain mass in an electromagnetic field, not much will happen. However, if the same particle is also charged, the magnetic field will accelerate

³ <https://www.rcsb.org/>

and/or deflect the particle in a predictable fashion, solely depended on the particle's mass-to-charge ratio (m/z). More massive particles will experience less of an acceleration/deflection (**Figure 1.17**, blue sphere), while conversely lighter particles will experience more (**Figure 1.17**, red sphere), in the same electromagnetic field. Charge has the opposite effect on the acceleration/deflection of particles, where particles with a higher charge (**Figure 1.17**, double charge) will experience a greater electromagnetic force, while particles with less charge (**Figure 1.17**, single charge) will experience a lesser force. Thus a particle with more charge will accelerate/deflect more in the same electromagnetic field. Together, mass-to-charge ratio (m/z) of a particle completely governs its behavior in a known electromagnetic field.

Figure 1.17: Schematic representation of mass spectrometry (MS) particle separation, based on mass (blue vs red particles) and charge (single vs double charge).

To properly achieve this, mass spectrometers have to desolvate and charge the sample being analyzed. This process gives rise to three types of MS; fragmenting [80-82], non-fragmenting [83, 84] and native (where 3D structure is preserved) [84-86], each with a multitude of uses and implementations.

For G-quadruplexes MS has been employed to analyze coordinated ions [87-89], ligand binding [90-92], formation [93, 94] and structural changes [87, 92, 93]. Since G-quadruplexes already carry a charge due to the phosphate backbone, desolvation and ionization conditions

can be tailored for non-destructive sample preparation [85]. For these purposes ElectroSpray Ionization (ESI), a soft ionization technique, has emerged as most suitable [84, 85]. After the samples are prepared they are separated in a negative-ion mode, due to the negative charge DNA carries natively. The separation of ionized species can be performed on an Ion Mobility Quadrupole Time-of-Flight (IM-Q-TOF) mass spectrometer [95]. In this implementation of MS, ions are accelerated by an electric field of known strength, down a tube of known length, towards the detector where time of impact is recorded [96, 97]. From there, time of flight can be calculated and the m/z ratio of the ion determined. The advantage of this implementation is the preservation of the native structure [98] and optimized sensitivity [99], making it suitable for detection of events related to G-quadruplexes.

1.7 Role of cations in G-quadruplex folding and studies

Ion coordination with assembled DNA structures has been extensively studied using NMR methods and NMR active nuclei $^{23}\text{Na}^+$ [100-103], $^{39}\text{K}^+$ [103, 104], $^{87}\text{Rb}^+$ [105] and $^{15}\text{NH}_4^+$ [106-108]. G-quadruplex structures have proven particularly useful in this respect, as dG-O6 atoms in assembled G-tetrads are uniquely positioned for coordination of dehydrated cations [109, 110].

The effect of ionic radius and charge of the coordinated ion on the stability of assembled G-quadruplex structures has been well-documented [111-114]. We have known for half a century that the stability of folded guanosine gels correlates with the nature and ionic radius of the cation [115], where K^+ has been shown to be most favorable and follows the trend $\text{K}^+ \gg \text{Rb}^+ > \text{Ba}^{2+} > \text{NH}_4^+ > \text{Na}^+ \gg \text{Li}^+$ (**Figure 1.18**) [25, 115]. In the last two decades similar results have been observed for folded G-quadruplex structures [30, 116-121].

Figure 1.18: Guanosine gel melting temperature (T_m) dependence on the type of cation used and its respective ionic radius. Figure adapted from [25].

$^{15}\text{NH}_4^+$ ammonium ion in particular provides unique advantages for studying G-quadruplex ion coordination using NMR, due to relatively slow exchange of bound ions [107, 108, 122-124]. Hydrogen is a spin $\frac{1}{2}$ nucleus and is the most sensitive NMR active nucleus due to its high gyromagnetic ratio. The presence of four hydrogen nuclei in the ammonium ion is thus very advantageous since the coordinated ammonium ion can be observed with 1D ^1H and 2D ^1H - ^1H NOESY experiments. $^{15}\text{NH}_4^+$ ammonium observations can be further substantiated with the use of 1D ^{15}N -filtered ^1H experiment and 2D ^1H - ^{15}N HSQC experiment. This is possible due to ^{15}N isotope of nitrogen being an NMR active, spin $\frac{1}{2}$ nucleus.

Coordinated ions do not stay static within a binding pocket but exchange between binding pockets as well as with the bulk ions present in the solvent [107, 108, 123-126]. The exchange rates observed for $^{15}\text{NH}_4^+$ vary between millisecond scale [123, 124] to over a second [122]. Sub millisecond residence times have also been observed for the Na^+ ion [100, 101].

Ions coordinated within the G-quadruplex can also be determined using native negative-ion electrospray ionization-mass spectrometry (ESI-MS) [85, 127, 128]. Mild conditions of ionization are able to preserve the native structure of folded G-quadruplex species in the gas phase and also determine the type and number of ions coordinated [129]. Due to experimental constraints though, use of buffers containing K^+ at physiological ionic strengths for native electrospray ionization is not possible as they are nonvolatile. K^+ ions have to be removed or diluted before running the experiment due to clusters and numerous adducts formation, making the mass spectra noisy [89]. In order to maintain consistent ionic strength, ammonium acetate (NH_4OAc) and trimethylammonium acetate (TMAA) are commonly used as both are volatile [129]. Buffers containing NH_4^+ ions are particularly useful for G-quadruplex studies using native ESI-MS, since the ammonium ions are able to coordinate between G-tetrads of folded G-quadruplex structures with high affinity [88, 94, 130, 131]. In order to examine the coordination of nonvolatile cations like K^+ , samples can be folded in K^+ and subsequently diluted with TMAA to preserve ionic strength, as the trimethylammonium cation is too bulky to fit between G-quadruplex G-tetrads [89] and does not exchange with the coordinated K^+ ions.

1.8 G-quadruplex role in biological systems

The role of G-quadruplexes (G4) in cellular processes has been expanding in the last decade [132]. Their capability to fold into stable structures from a very diverse subset of guanine-containing sequences[26] has recently been expanded even further. The latest algorithms for evaluating the likelihood of G-quadruplex formation incorporate guanine-richness of the primary sequence [133] alongside previously established patterns like $d(G_{3+} N_{1-7} G_{3+} N_{1-7} G_{3+} N_{1-7} G_{3+})$ [134]. This flexibility is based on experimental evidence of more than 230 solved G-quadruplex structures in Nucleic Acid Database⁴. Using CD and NMR spectroscopies together with established spectral features of G-quadruplex forming sequences, fast an accurate determination of G-quadruplex folding can be achieved [27, 135]. While *in vitro* evidence of G-quadruplex formation has been undisputed for a long time, G-quadruplex role in cells has gathered increasing evidence recently with studies reporting on a widening array of G-quadruplex interacting proteins [136] as well as *in situ* G-quadruplex visualization with specific G-quadruplex antibodies [137, 138]. G-quadruplexes are increasingly likely to play an important role in many cellular processes (**Figure 1.19**), such as transcription [139, 140], translation [141, 142], replication [143-145] and telomere stability [146, 147].

Figure 1.19: Schematic representation of possible G-quadruplex forming locations and their possible effects on cellular processes. (A) G-quadruplex formation in the promoter region during transcription and (C) replication. (B) G-quadruplex formation on the G-rich single stranded overhang of telomeres. (D) G-quadruplex formation in mRNA. Figure adapted from [132].

⁴ <http://ndbserver.rutgers.edu/>

1.9 Discovery of a left-handed G-quadruplex and its properties

G-quadruplex backbone has a handedness of progression which is closely linked with the G-tetrad stacking twist. Despite their polymorphic nature, almost all G-quadruplexes adopt a right-handed G-tetrad stacking twist. Recently our laboratory reported on the first observed case of a natural left-handed G-quadruplex structure, named *Z-G4*, with some unique features [53]. The sequence of *Z-G4* was obtained from a well-known anti-cancer aptamer *AGRO100* (also known as *AS1411*) [148] which was also shown to have potent anti-HIV activity [149, 150].

AGRO100 was shown to adopt many different conformations in vitro and we did a systematic study to obtain single-conformation adopting sequences by systematically replacing guanines with thymines in the *AGRO100* primary sequence [151]. Sequences relevant for this manuscript are summarized in **Table 1.1**.

Table 1.1: Table of *AGRO100* derived sequences.

Name	Sequence	Number of nucleotides
<i>AGRO100</i>	GGTGGTGGTGGTTGTGGTGGTGGTGG	26
<i>A1003</i>	TGGTGGTGGTGGTTGTGGTGGTGGTGGT	28
<i>AT27</i>	TGGTGGTGGTGGTTGTGGTGGTGGTGT	28

The first modification to *AGRO100* primary sequence was attaching an extra thymine on both the 5' and 3' end of the primary sequence, which helps prevent quadruplex dimerization, thus obtaining the sequence named *A1003*. This helped clear up both CD and NMR spectra, shown in **Figure 1.20** and **Figure 1.21**, respectively.

Figure 1.20: CD spectra of modified *AGRO100* sequences from **Table 1.1** in potassium buffer.

A1003 has 17 guanines, one more than required for a G-quadruplex with four G-tetrads. Each of the 17 guanines was systematically replaced with a thymine residue, obtaining a list of sequences with 16 guanines each. Since four G-tetrads require 16 guanines to form, they were more likely to produce a single-conformation structure. Additional guanines are more likely to accommodate multiple G-quadruplex structures. Of particular note is the sequence named *AT27*, where G27 is replaced with a thymine. *AT27* produces a clean NMR imino spectrum (**Figure 1.21**, bottom) indicating a single-conformation, while its CD spectrum (**Figure 1.20**, right) was very unusual, compared to CD spectra of known G-quadruplexes [27]. The structure of *AT27* was solved by both solution NMR and X-ray crystallography [53] and it was later renamed to *Z-G4*. The structure of *Z-G4* showed it is a left-handed parallel G-quadruplex, with a negative, counter-clockwise G-tetrad twist, which was explaining the unusual CD spectrum. After observing the first natural left-handed G-quadruplex structure, *Z-G4*, we analyzed its primary sequence and unique structural and spectral characteristics in order to find properties that might drive the left-handed folding and possibly determine the sequence space of a putative left-handed G-quadruplex motif.

Figure 1.21: NMR spectra of modified *AGRO100* sequences from **Table 1.1** in potassium buffer.

The *Z-G4* structure is composed of two blocks with a total of four layers. The two continuous blocks stack on the 5'-5' interface and are connected by thymine loop residues. The two things that stood out in the structure of *Z-G4* were the backbone progression (**Figure 1.22, A**), which was shown to be left-handed and the thymine loop orientations as they were flipped up towards the outer G-tetrad face to form a feature named the T-cap (orange) (**Figure 1.22, B**).

Figure 1.22: Structural features of a left-handed quadruplex. **(A)** Comparison of backbone progression handedness of Z-G4 (left) and a typical parallel right-handed G-quadruplex (green). **(B)** Top view of left-handed quadruplex with a T-cap. Thymine bases are colored in orange, guanine bases are cyan and the backbone is in grey.

Thymine residues flipping up and over the outer G-tetrad produce a characteristic NMR feature due to a hydrogen bond formation between the O4' oxygen of thymine and amino protons of outer G-tetrad guanines (**Figure 1.23, A**). The formation of this hydrogen bond slows down the rotation of the amino protons of outer G-tetrad guanine residues enough, to be observable on the NMR time scale (**Figure 1.23, B**) [59]. The two respective amino protons are found around 9.3 and 4.5 ppm and also produce cross-peaks in with imino protons of outer G-tetrad guanines in 2D NOESY spectra (**Figure 1.23, C**).

Figure 1.23: **(A)** Stick representation of guanine dinucleotide step with the O4' (yellow) of the following thymine T7 forming a hydrogen bond with the two amino protons of guanine G5. **(B)** Imino-amino region of the ^1H NMR spectrum of Z-G4 before (top) and after D_2O exchange (bottom). Amino protons are marked with red asterisk. **(C)** 2D NOESY spectrum showing two

distinct groups of cross-peaks between the outer G-tetrad guanine imino protons and both amino protons. Figure adapted from [53].

Imino protons are found between 11.8 and 10.6 ppm (**Figure 1.23, B**). They form two distinct groups of eight peaks each, with the downfield group (11.8-11.2 ppm) containing all eight imino peaks from the two outer G-tetrads, and the upfield group (11.0-10.6 ppm) containing all eight imino peaks from the two inner G-tetrads.

The unique shape of the *Z-G4* CD spectrum together with NMR spectra of amino protons at 9.5 ppm and two distinct groups of eight imino protons are spectral characteristics that can be utilized for quick and easy identification of structures that are similar to *Z-G4* and likely also left-handed.

The left-handed nature of *Z-G4* has already been exploited to design a targeted enantioselective ligand [152], showing great promise for targeting this new feature of G-quadruplex structures.

Figure 1.24: Comparison of guanine stacking effects of parallel left-handed *Z-G4* and right-handed *T95-2T* (PDB code, **2LK7**) G-quadruplexes. (A) Top view of guanine stacking within left-handed *Z-G4* with a relative rotation of -25.7° (left) and right-handed *T95-2T* with a relative rotation of $+29.4^\circ$ (right). (B) CD spectra of *Z-G4* (left) and *T95-2T* (right).

The change of handedness in strand progression of *Z-G4*, compared to parallel right-handed G-quadruplexes, is also reflected in the twist angle of stacked guanine bases (**Figure 1.24, A**) and by extension, the G-tetrad stacking angle. While the magnitude of the twist angle is relatively similar between the left- and right-handed G-quadruplex structures, the twist angle is of the opposite direction and is mirroring the parallel right-handed G-quadruplex. The change in the twist angle directionality explains the unique CD spectrum of *Z-G4*, which is also mirroring the CD spectrum of right-handed parallel G-quadruplexes (**Figure 1.24, B**).

The *Z-G4* structure is a four-layer parallel left-handed G-quadruplex composed of two continuous blocks that stack on the 5'-5' interface and are connected by thymine loop residues (**Figure 1.25, A**). Each block is composed of two G-tetrads with partial 5/6-member ring stacking (**Figure 1.25, B**). The sequences corresponding to each blocks differ slightly. The 5'-end *Block 1* (blue) is composed of four G-doublets separated by thymine residues d(TGG)₄TT while the 3'-end *Block 2* (red) has three G-doublets and two separated, single guanines dG(TGG)₃TGTT, G15 and G26 respectively, that stack on each other in the *Z-G4* structure (**Figure 1.25, B**).

Figure 1.25: Color coded representation of *Z-G4* where the top block, *Block 1*, is colored in blue, the bottom block, *Block 2* in red and the loop connecting them in gray. **(A)** Side view of the stick & ribbon representation of *Z-G4*. **(B)** Color coded sequence and schematic representation of *Z-G4* with backbone progression and numbered guanine residues.

Chapter 2: Materials and methods

2.1 Sample preparation and sequences used

DNA oligonucleotide samples were ordered from Integrated DNA Technologies (IDT) with standard desalting. Sample purity was measured with ESI-MS and was >99%. Further purification, if necessary, was performed using dialysis, successively against water, 5 mM LiCl solution and water after which the pH was checked to be above pH=6. Samples were flash frozen in liquid nitrogen, lyophilized and dissolved to the desired concentration in a buffer with appropriate cation. If necessary pH was corrected to be pH=6.5-7, unless otherwise stated. The buffers used in this study are listed in **Table 2.1**.

Table 2.1: Buffers used in sample preparation.

Buffer	Buffer composition
Potassium buffer (NMR)	70 mM KCl, 20 mM KPi pH=6.9, 10% D ₂ O, 20 μM DSS
Potassium buffer (crystallization)	100 mM KCl, 10 mM potassium cacodylate pH=6.6
Ammonium buffer (NMR)	100 mM ammonium acetate pH=6.9, 10% D ₂ O, 20 μM DSS
Ammonium buffer (crystallization)	100 mM ammonium acetate pH=6.9
¹⁵ N Ammonium buffer (NMR)	100 mM ¹⁵ N ammonium chloride, 20 mM deuterated Tris pH=6.9, 10% D ₂ O, 20 μM DSS
Sodium buffer (NMR)	100 mM sodium chloride, 20 mM deuterated Tris pH=6.9, 10% D ₂ O, 20 μM DSS
Lithium buffer (NMR)	100 mM Lithium chloride, 20 mM deuterated Tris pH=6.9, 10% D ₂ O, 20 μM DSS
Caesium buffer (NMR)	100 mM caesium chloride, 20 mM deuterated Tris pH=6.9, 10% D ₂ O, 20 μM DSS

Prior to experiments, DNA samples were either heated at ~95°C for 5 minutes and slowly cooled down to room temperature over-night, or left at room temperature for 2 or more days before performing experiments. For crystallization, samples were centrifuged for 10 minutes at 20,000 rpm to remove any possible precipitate. A list of samples used with corresponding sequences and extinction coefficients are listed in **Table 2.2** and **Table 2.3**, where **Table 2.3** lists the sequences used specifically for the substitution analysis of *2xBlock 2A*.

Table 2.2: Sequences used in this work (5'-3') *sans* substitution sequences.

Name	Sequence	Nucleotides	ϵ_{260} (M ⁻¹ cm ⁻¹)
Z-G4	TGGTGGTGGTGGTTGTGGTGGTGGTGT	28	264800
Block 1	TGGTGGTGGTGGTT	14	133600
Block 1 Δ	TGGTGGTGGTGG	12	115800
Δ Block 1 Δ	GGTGGTGGTGG	11	108300
Block 2	GTGGTGGTGGTGT	14	133600
Block 2 Δ	GTGGTGGTGGTG	12	117000
5'T-Block 2	TGTGGTGGTGGTGT	15	141100
2xBlock 2	GTGGTGGTGGTGTGTGGTGGTGGTGT	27	257900
2xBlock 2 Δ	GTGGTGGTGGTGTGTGGTGGTGGTG	26	249400
5'T-2xBlock 2	TGTGGTGGTGGTGTGTGGTGGTGGTGT	28	265400
2xBlock 2-1T loop	GTGGTGGTGGTGTTGTGGTGGTGGTGT	26	249800
2xBlock 2-3T loop	GTGGTGGTGGTGTTTGTGGTGGTGGTGT	28	266000
2xBlock 2-4T loop	GTGGTGGTGGTGTTTGTGGTGGTGGTGT	29	274100
2xBlock 2-13A	GTGGTGGTGGTGATGTGGTGGTGGTGT	27	262400
2xBlock 2-14A	GTGGTGGTGGTGAGTGGTGGTGGTGT	27	263800
2xBlock 2-13,14A	GTGGTGGTGGTGAAAGTGGTGGTGGTGT	27	266300
Z-G4R	GTGGTGGTGGTGTGGTGGTGGTGGTT	27	257900
Block 1-Bulge	GGTGTTGTGGTGG	12	117000
Block 1-2xBulge	GGTGTTGTGTGGTGG	13	125700
Block 1-3xBulge	GGTGTTGTGTGTGG	14	134400
Block 1-4xBulge	GTGTGTGTGTGTGTGTG	15	143100
Z-G4-Bulge	GGTGTTGTGGTGGTGGTGGTGGTGGTT	27	257900
Z-G4-2xBulge	GGTGTTGTGTGTGGTGGTGGTGGTGGTT	28	265400
Z-G4-3xBulge	GGTGTTGTGTGTGTGGTGGTGGTGGTT	29	275300
Z-G4-4xBulge	GTGTGTGTGTGTGTGTGGTGGTGGTGGTT	30	284000

Table 2.3: Sequences used in substitution studies (5'-3') of 2xBlock 2Δ.

	Name	Sequence	Nucleotides	ϵ_{260} (M ⁻¹ cm ⁻¹)
2xBlock 2Δ with single block substitutions	2xBlock 2 T2-C	G C GGTGGTGGTGTGTGGTGGTGGTG	26	247300
	2xBlock 2 T5-C	GTGG C GGTGGTGTGTGGTGGTGGTG	26	247300
	2xBlock 2 T8-C	GTGGTGG C GGTGTGTGGTGGTGGTG	26	247300
	2xBlock 2 T11-C	GTGGTGGTGG C GTTGTGGTGGTGGTG	26	247300
	2xBlock 2 T2-C	G C GGTGGTGGTGTGTGGTGGTGGTG	26	247300
	2xBlock 2 T2,8-C	G C GGTGG C GGTGTGTGGTGGTGGTG	26	245200
	2xBlock 2 T2,5,8-C	G C GG C GG C GGTGTGTGGTGGTGGTG	26	243100
	2xBlock 2 T2,5,8,11-C	G C GG C GG C GG C GTTGTGGTGGTGGTG	26	241000
	2xBlock 2 T2-A	G A GGTGGTGGTGTGTGGTGGTGGTG	26	253900
	2xBlock 2 T2,8-A	G A GGTGG A GGTGTGTGGTGGTGGTG	26	258400
	2xBlock 2 T2,5,8-A	G A GG A GG A GGTGTGTGGTGGTGGTG	26	262900
	2xBlock 2 T2,5,8,11-A	G A GG A GG A GG A GTTGTGGTGGTGGTG	26	267400
	2xBlock 2 T2-TT	G T TGGTGGTGGTGTGTGGTGGTGGTG	27	257500
	2xBlock 2 T2,5-TT	G T TGG T TGGTGGTGTGTGGTGGTGGTG	28	265600
	2xBlock 2 T2,5,8-TT	G T TGG T TGG T TGGTGTGTGGTGGTGGTG	29	273700
	2xBlock 2 T2,5,8,11-TT	G T TGG T TGG T TGG T TGTTGTGGTGGTGGTG	30	281800
2xBlock 2Δ with both blocks substitutions	2xBlock 2 T5,19-C	GTGG C GGTGGTGTGTGG C GGTGGTG	26	245200
	2xBlock 2 T2,8,16,22-C	G C GGTGG C GGTGTG C GGTGG C GGTG	26	241000
	2xBlock 2 T2,5,8,16,19,22-C	G C GG C GG C GGTGTG C GG C GG C GGTG	26	236800
	2xBlock 2 T2,5,8,11,16,19,22,25-C	G C GG C GG C GG C GTTG C GG C GG C GG C G	26	232600
	2xBlock 2 T2,16-A	G A GGTGGTGGTGTG A GGTGGTGGTG	26	258400
	2xBlock 2 T2,8,16,22-A	G A GGTGG A GGTGTG A GGTGG A GGTG	26	267400
	2xBlock 2 T2,5,8,16,19,22-A	G A GG A GG A GGTGTG A GG A GG A GGTG	26	276400
	2xBlock 2 T2,5,8,11,16,19,22,25-A	G A GG A GG A GG A GTTG A GG A GG A GG A G	26	285400
	2xBlock 2 T2,16-TT	G T TGGTGGTGGTGTG T TGGTGGTGGTG	28	265600
	2xBlock 2 T2,5,16,19-TT	G T TGG T TGGTGGTGTG T TGG T TGGTGGTG	30	281800
	2xBlock 2 T2,5,8,16,19,22-TT	G T TGG T TGG T TGGTGTG T TGG T TGG T TGGTG	32	298000
	2xBlock 2 T2,5,8,11,16,19,22,25-TT	G T TGG T TGG T TGG T TGTTG T TGG T TGG T TGG T TG	34	314200

2.2 Circular dichroism (CD)

CD spectra were recorded at 20°C on a JASCO-815 spectropolarimeter using a 1-cm path-length quartz cuvette. The DNA concentrations were 3–10 μM. For each experiment, an average of three scans was recorded between 220 to 320 nm at 100 nm/min scanning speed, 1 nm data pitch, 1 nm band-width and 1 second digital integration time (D.I.T). Buffer solution was used for baseline correction for all spectra and the data was zero-corrected at 320 nm. Molar ellipticity of CD spectra was calculated using the DNA concentration derived from the sample absorbance at 260 nm and the sample extinction coefficient calculated at 260 nm.

2.3 ^1H NMR spectroscopy

Experiments were performed on 600/800 MHz Bruker spectrometers equipped with cryoprobe, at 25°C with DNA concentrations 0.1–0.2 mM, unless otherwise stated. All spectra were processed and analyzed using TopSpin 2.1 (Bruker).

2.4 D_2O exchange NMR spectroscopy

Previously prepared NMR samples were flash frozen in liquid nitrogen and subsequently lyophilized. Just before the start of the experiment, an amount of D_2O was added that was equal to the previous sample volume.

2.5 Tracking of $^{15}\text{NH}_4^+$ using ^1H NMR spectroscopy

A sample of Z-G4 was folded in either 100 mM $^{15}\text{NH}_4^+/\text{K}^+$ and subsequently dialyzed against H_2O for 10 minutes in order to remove most of the $^{15}\text{NH}_4^+/\text{K}^+$ ions in solution. It was quickly flash frozen in liquid nitrogen and lyophilized. Just before the start of the experiment, the sample was re-suspended in a buffer containing the opposite ion, either 100 mM $\text{K}^+ / ^{15}\text{NH}_4^+$ ions. The sample was immediately tracked with NMR where we interchangeably recorded both ^1H and ^{15}N filtered ^1H NMR spectra over 3 hours.

2.6 Crystallization and X-ray crystallography

Crystallization assays were performed with 0.8-2 mM solution of sample in potassium or ammonium buffer (**Table 2.1**) using the sitting-drop method. Initial screening was done with the Natrix commercial screens (Hampton Research). Crystals were fished directly out of screening plates and flash frozen in liquid nitrogen. Data collection and refinement statistics for all structures presented in this work are summarized in **Table 2.4**.

Table 2.4: Data collection and refinement statistics for all solved structures presented in this work, *Block 2*, *2xBlock 2*, *Z-G4R*, *Z-G4-Bulge* and *Z-G4-2xBulge*, respectively.

Data collection					
	<i>Block 2</i>	<i>2xBlock 2</i>	<i>Z-G4R</i>	<i>Z-G4-Bulge</i>	<i>Z-G4-2xBulge</i>
Space group	P3 ₂	P2 ₁ 2 ₁ 2 ₁	P3 ₁ 12	P12 ₁ 1	P12 ₁ 1
Cell dimensions					
a	30.56	29.57	31.30	29.22	29.10
b	30.56	37.63	31.30	40.04	39.81
c (Å)	53.66	54.28	118.15	30.38	54.97
α	90.0	90.0	90.0	90.0	90.0
β	90.0	90.0	90.0	115.87	91.28
γ (°)	120.0	90.0	120.0	90.0	90.0
Resolution (Å)	26.46 - 2.31	30.93 - 2.01	59.2 - 1.59	40.4 – 0.972	29.1 – 1.30
R _{sym}	0.031 (0.230)	0.089 (0.668)		0.112 (0.94)	0.093 (0.797)
I / σI	16.4 (3.4)	18.3 (3.95)	12.5 (1.18)	8.67 (1.52)	6.86 (1.07)
Completeness (%)	97.7 (96.9)	98.9 (93.6)	98.5 (96.9)	100 (100)	95 (93)
Redundancy	2.8 (2.3)	13.0 (13.4)	19.3 (17)	6.5 (5.9)	6.0 (4.9)
CC1/2 (%)	100 (93)	100 (96)	99.8 (98.9)	100 (53)	100 (73.4)
Refinement					
Resolution (Å)	26.46 – 2.31	30.93 – 2.01	10.0 – 1.6	26.3 – 1.18	29.1 – 1.30
Number of reflections	2405	4285	9206	20678	29721
Twin fraction	0.51 (k, h, -l)				
R _{work} / R _{free}	0.196/0.227	0.185/0.231	0.185/0.214	0.124/0.168	0.184/0.215
Number of atoms					
DNA	506	569	569	569	1172
K ⁺	3	3		3	8
Ba ²⁺			2		
Spermine				56	56
Waters	12	10	21	158	210
B-factors (Å ²)					
	DNA 44.5 K ⁺ 27.4 Water 26.6	DNA 31.0 K ⁺ 22.1 Water 31.0	DNA Ba ²⁺ Water	DNA 10.5 K ⁺ 6.1 SPM 22.7 Water 33.27	DNA 14.5 K ⁺ 10.4 SPM 33.1 Water 27.9
R.M.S.D					
Bond lengths (Å)	0.004	0.008	0.008	0.017	0.007
Bond angles (°)	0.697	0.997	0.998	1.27	1.03

2.6.1 Block 2 structure determination

The Block 2 sequence (concentration, 0.8 mM) folded in 100 mM potassium chloride, 10 mM potassium cacodylate buffer (pH 6.6), was crystallized in a solution containing 80 mM potassium chloride, 40 mM sodium cacodylate (pH=6.0) and spermine (12 mM), using 55% (v/v) 2-Methyl-2,4-pentanediol as a precipitating agent. The X-ray data to 2.3 Å resolution was obtained on a RIGAKU in-house light source ($\lambda = 1.54$ Å) using a Pilatus 200K detector. Data

displayed apparent P622 symmetry. However, for this point group, the cell size was not sufficient to accommodate one *Block 2* molecule, indicative of twinning. Molecular replacement using a *Block 2* model derived from the *Z-G4* structure (PDB ID 4U5M) was successful in the lower symmetry space groups P3₂12, P3₂21 and P3₂. The best refinement statistics were obtained with the P3₂ space group with two NCS-related molecules in the asymmetric unit. Refinement was performed in PHENIX [153] with the twin law k, h, -l and a twinning fraction close to 0.5 (**Table 2.4**) in agreement with nearly perfect hemihedral twinning ($R_{\text{free}}=0.227$). The final model forms a stacked dimer containing 2 x 12 residues of DNA, 3 potassium ions and 12 water molecules.

2.6.2 2xBlock 2 structure determination

The 2xBlock 2 sequence (concentration, 2 mM) folded in 100 mM potassium chloride, 10 mM potassium cacodylate buffer (pH 6.6), was crystallized in a solution containing 80 mM potassium chloride, 20 mM barium dichloride, 40 mM sodium cacodylate (pH=7.0) and spermine (12 mM), using 40% (v/v) 2-Methyl-2,4-pentanediol as a precipitating agent. For data collection, crystals were directly flash frozen in liquid nitrogen. Data were collected from a single crystal at the Berkeley beamline ($\lambda = 0.979 \text{ \AA}$; Berkeley, USA) equipped with an ADSC Q315 CCD detector, and processed using XDS [154]. Data collection statistics are summarized in **Table 2.4**. Molecular replacement using a *Block 2* model derived from the *Z-G4* structure (PDB ID 4U5M) was successful in the space group P2₁2₁2₁ using PHASER [155] and data to 2 Å resolution. The model was then refined and completed with missing residues as well as ions and water molecules, by several cycles of manual building using Coot [156] and energy minimization using PHENIX [153] ($R_{\text{free}}=0.225$). The final model contains all 27 residues of DNA, 3 potassium ions and 10 water molecules.

2.6.3 Z-G4R structure determination

The Z-G4R sequence (concentration, 0.8 mM) folded in 100 mM ammonium acetate buffer (pH 6.9), was crystallized in a solution containing 80 mM sodium chloride, 20 mM barium dichloride, 40 mM sodium cacodylate (pH=7.0) and spermine (12 mM), using 40% (v/v) 2-Methyl-2,4-pentanediol as a precipitating agent. For data collection, crystals were directly

flash frozen in liquid nitrogen. Data were collected from a single crystal at the SOLEIL PX2 beamline ($\lambda = 0.979 \text{ \AA}$; SOLEIL, France) equipped with an EIGER X 9M detector, and processed using XDS [154]. Data collection statistics are summarized in **Table 2.4**. Molecular replacement using a *Block 2* model derived from the Z-G4 structure (PDB ID 4U5M) was successful in the space group $P3_112$ using PHASER [155] and data to $1.6 \text{ \AA}$ resolution. The model was then refined and completed with missing residues as well as ions and water molecules, by several cycles of manual building using Coot [156] and energy minimization using PHENIX [153] ($R_{\text{free}}=0.214$). The final model contains all 27 residues of DNA, 2 barium ions and 21 water molecules.

2.6.4 Z-G4-Bulge structure determination

The Z-G4-Bulge sequence (concentration, 0.8 mM) folded in 100 mM potassium chloride, 10 mM potassium cacodylate buffer (pH 6.6), was crystallized in a solution containing 80 mM potassium chloride, 40 mM sodium cacodylate (pH=6.0) and spermine (12 mM), using 55% (v/v) 2-Methyl-2,4-pentanediol as a precipitating agent. For data collection, crystals were directly flash frozen in liquid nitrogen. Data were collected from a single crystal at the SOLEIL PX1 beamline ($\lambda = 0.979 \text{ \AA}$; SOLEIL, France) equipped with an PILATUS 6M detector, and processed using XDS [154]. Data collection statistics are summarized in **Table 2.4**. Molecular replacement using a *Block 2* model derived from the Z-G4 structure (PDB ID 4U5M) was successful in the space group $P12_11$ using PHASER [155] and data to $1.18 \text{ \AA}$ resolution. The model was then refined and completed with missing residues as well as ions and water molecules, by several cycles of manual building using Coot [156] and energy minimization using PHENIX [153] ($R_{\text{free}}=0.168$). The final model contains all 27 residues of DNA, 3 potassium ions two spermine molecules and 158 water molecules.

2.6.5 Z-G4-2xBulge structure determination

The Z-G4-2xBulge sequence (concentration, 0.8 mM) folded in 100 mM potassium chloride, 10 mM potassium cacodylate buffer (pH 6.6), was crystallized in a solution containing 80 mM potassium chloride, 40 mM sodium cacodylate (pH=7.0) and spermine (12 mM), using 60% (v/v) 2-Methyl-2,4-pentanediol as a precipitating agent. For data collection, crystals were

directly flash frozen in liquid nitrogen. Data were collected from a single crystal at the SOLEIL PX1 beamline ($\lambda = 0.979 \text{ \AA}$; SOLEIL, France) equipped with an PILATUS 6M detector, and processed using XDS [154]. Data collection statistics are summarized in **Table 2.4**. Molecular replacement using a *Block 2* model derived from the *Z-G4* structure (PDB ID 4U5M) was successful in the space group $P12_11$ using PHASER [155] and data to $1.3 \text{ \AA}$ resolution. The model was then refined and completed with missing residues as well as ions and water molecules, by several cycles of manual building using Coot [156] and energy minimization using PHENIX [153] ($R_{\text{free}}=0.215$). The final model contains all two subunits of DNA with 28 residues each, 8 potassium ions four spermine molecules and 210 water molecules.

2.6.6 Structural analysis

Structures were analyzed using open-source Pymol 1.8 [157] and in-house scripts. Measurements of backbone dihedral angles (α to ζ) were performed using Curves+ [158].

2.7 Electrospray ionization-mass spectrometry

A sample of *Z-G4* was folded in 90 mM TMAA pH = 7, 10 mM KCl to a concentration of 0.1 mM, heated to 90° , slowly cooled overnight and left for 7 days. Before measurements the sample was diluted 10x by either 100 mM TMAA pH = 7 or 100 mM NH_4OAc pH = 7. In both cases, after the 10x dilution, the effective concentration of K^+ in solution was 1 mM, while the sample concentration was 0.01 mM.

100 mM TMAA buffer was used when measuring the sample behavior in K^+ as TMAA does not coordinate between tetrads, due to its bulky nature, and is added to maintain ionic strength.

100 mM NH_4OAc buffer was used when tracking the conversion from K^+ to NH_4^+ . The effective concentration of the NH_4^+ ion after the 10x dilution is 90 mM making the K^+/NH_4^+ ratio 1/90.

Samples were injected in the negative-ion ESI-MS with softest possible experimental tuning, that preserved the most NH_4^+ ion species. Samples were tracked through time by periodical injections and the recorded data was cleaned of non-specific adducts.

The Z-G4 structure is composed of two continuously connected blocks of parallel left-handed two-layered G-quadruplexes (**Figure 3.1, A and B**). The sequences of the two blocks differ slightly, whereas the 5'-end block (*Block 1*, T(GGT)₄T) is composed of four continuous G-doublets, the 3'-end block (*Block 2*, GT(GGT)_{3GTT) has three G-doublets and two separated single guanines (underlined). In the Z-G4 structure, the two split Gs in *Block 2* (G15 and G26) stack upon each other in the same way as G-doublets (**Figure 3.1, C**) which is facilitated by the backbone circling all the way around the G-tetrad core. We examine which of these two sequences, *Block 1* or *Block 2*, might infer the left-handed helicity.}

3.1 *Block 1* and *Block 2* displays right- and left-handed G-quadruplex spectral features, respectively

The *Block 1* and *Block 2* sequences were folded in potassium buffer and their CD and NMR spectra were recorder (**Figure 3.2**). The CD spectrum (**Figure 3.2, A**) of *Block 1* (blue) has a positive peak at 260 nm and a negative peak at 240 nm, characteristic of parallel right-handed G-quadruplexes [48, 49, 159]. The CD spectrum of *Block 2* (red) has a negative peak at 270 nm and a positive peak at 250 nm, similar to that of the left-handed G-quadruplex Z-G4 [53].

Figure 3.2: CD spectra of *Block 1* (blue) and *Block 2* (red) with corresponding sequences. θ refers to the molar ellipticity.

The NMR imino proton spectra (**Figure 3.2, B**) of *Block 1* (blue) and *Block 2* (red) both show peaks between 11.8 to 10.5 ppm, indicating the formation of G-quadruplexes. There is no

clear grouping of imino proton peaks and sharp amino proton peaks observed for *Block 1* as was observed for *Z-G4*. The distribution of imino protons of *Block 2*, with peaks being separated into two groups, is highly similar to that observed for *Z-G4*. A particular feature of the *Z-G4* structure is the formation of hydrogen bonds between O4' of the capping thymine and the guanine amino protons of the adjacent G-tetrad, resulting in sharp amino protons at 9-10 ppm [59]. Such protons are also observed for *Block 2* (**Figure 3.2, B**). Altogether, these results indicate that *Block 2* forms a left-handed G-quadruplex.

Table 3.1: List of *Block 1* and *Block 2* sequences with terminal-thymine modifications.

Name	Sequence	Number of nucleotides
Z-G4	TGGTGGTGGTGGTTGTGGTGGTGGTGT	28
Block 1	TGGTGGTGGTGGTT	14
Block 1Δ	TGGTGGTGGTGG	12
ΔBlock 1Δ	GGTGGTGGTGG	11
Block 2	GTGGTGGTGGTGT	14
Block 2Δ	GTGGTGGTGGTG	12
5'T-Block 2	T GTGGTGGTGGTGT	15

Since thymine can have an influence on dimerization we checked their influence on both *Block 1* and 2. Three sequences of *Block 1*, *Block 1Δ* and *ΔBlock 1Δ* (**Figure 3.3**) do not give either a clear grouping of imino proton peaks or sharp amino proton peaks (**B**). The three variants of the ***Block 1*** sequence exhibit a right-handed parallel G-quadruplex CD spectrum with a peak at 260 nm and a trough at 240 nm.

Figure 3.3: CD (**A**) and NMR (**B**) spectra of *Block 1* and two variants without terminal Ts at the 3' and 5' end.

Adding a 5'-thymine to the *Block 2* sequence removes both the left-handed CD signal and NMR imino proton signals (**Figure 3.4, A and B**), indicating that dimerization through 5'-5' stacking might be disrupted with the 5' thymine addition. The variant of the *Block 2* sequence without two Ts at the 3'-end, *Block 2 Δ* , display very similar CD and NMR spectra (**Figure 3.4, A and B**), indicating the formation of the same structure.

Figure 3.4: CD (A) and NMR (B) spectra of *Block 2* and variants without two Ts at the 3'-end or an additional 5' thymine.

3.2 *Block 2* forms a left-handed dimeric G-quadruplex

The *Block 2* sequence and a variant without two Ts at the 3'-end (**Table 3.1**) display very similar CD and NMR spectra (**Figure 3.4**), indicating the formation of the same structure. The latter sequence was successfully crystallized and its structure solved (**Table 3.2**), containing two stacked *Block 2* molecules in the asymmetric unit.

Table 3.2: Data collection and refinement statistics for *Block 2* (PDB ID **6FQ2**).

Data collection	
Space group	P3 ₂
Cell dimensions	
a, b, c (Å)	30.56, 30.56, 53.66
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	26.46 - 2.31
R _{sym}	0.031 (0.230)
I / σI	16.4 (3.4)
Completeness (%)	97.7 (96.9)
Redundancy	2.8 (2.3)
CC1/2 (%)	100 (93)
Refinement	
Resolution (Å)	26.46 – 2.31
Number of reflections	2405
Twin fraction	0.51 (k, h, -l)
R _{work} / R _{free}	0.196/0.227
Number of atoms	
DNA	506
K ⁺	3
Waters	12
B-factors (Å ²)	
	DNA 44.5 K ⁺ 27.4 Water 26.6
R.M.S.D.	
Bond lengths (Å)	0.004
Bond angles (°)	0.697

[a] values in parentheses are for the highest resolution shell

The two blocks in the asymmetric unit are superimposed with an RMSD of 0.77 Å for 253 atom pairs, reflecting adjustment of the thymine residues in the two blocks (**Figure 3.5, A**), probably due to packing interactions in the crystal involving different thymine residues for Subunits A (T2 and T8) and B (T5 and T11) (**Figure 3.5, B**). Four out of eight thymine residues are engaged in the crystal packing interactions. The structure was deposited in PDB databank with accession number **6FQ2**.

Figure 3.5: (A) Alignment of both *Block 2* subunits. Thymine residues of Subunit 1 are coloured in red and blue for Subunit 2. (B) Stacking interactions involving thymine residues present in the crystal lattice of *Block 2*.

In our structure, the asymmetric unit is composed of two G-quadruplex (Subunits 1 and 2) stacked on 5'-5' interface. Each G-quadruplex unit is composed of two-layered parallel-stranded structure (**Figure 3.6**) with a left-handed helicity. The two G-tetrads are formed by G1•G10•G7•G4 and G3•G12•G9•G6 with identical hydrogen-bond directionalities. The stacking interface of the subunits is 5/6 ring [28], with G1 and G10 from one subunit stacking above G10 and G1 from the other subunit, respectively (**Figure 3.6, B**).

Figure 3.6: (A) Crystal structure of the left-handed G-quadruplex formed by *Block 2*. The electron density corresponds to the final 2mFo-Fc map contoured at 1.0 σ . (B) Schematic representation of the crystal structure with numbered guanine residues and marked backbone progression direction. (C) Ribbon representation of the crystal structure. (D) Top view of the crystal structure.

Thymines cap the outer G-tetrads of both subunits which allows for hydrogen bond formation between the O4' sugar of thymines and amino proton of outer G-tetrad guanines. Stacking between two consecutive guanines in the G-tetrad is partial 5/6-member ring with a twist mean angle value of $(-25.2 \pm 0.6)^\circ$. Dihedral backbone angles are summarized in **Table 3.3**, consistent with the values found in *Z-G4*. Three K^+ ions are found between the G-tetrads, 12 molecules of water are located around the grooves and on top of G-tetrads.

Table 3.3: Average backbone dihedral angles of *Block 2* and *Z-G4*.

Backbone dihedral	<i>Block 2</i> ^[a]	<i>Z-G4</i> ^[b]
ϵ	$(321 \pm 2)^\circ$	$(316 \pm 2)^\circ$
ζ	$(82 \pm 1)^\circ$	$(86 \pm 3)^\circ$
$\alpha+1$	$(159 \pm 4)^\circ$	$(161 \pm 6)^\circ$
$\beta+1$	$(232 \pm 6)^\circ$	$(226 \pm 5)^\circ$
$\gamma+1$	$(53 \pm 2)^\circ$	$(54 \pm 3)^\circ$

[a] Average values and standard deviations obtained from the dinucleotide steps of consecutive guanines. [b] Values reported in [53].

In both subunits, an intramolecular hydrogen bond could be formed between backbone oxygens of the first and last residues (**Figure 3.7**), likely further stabilizing the folded structure.

Figure 3.7: Side view of subunit A (top) and subunit B (bottom) with the hydrogen bond between dG1-O5' and dG12-O3'.

Overall the structural features are very similar to those found in Z-G4 (**Figure 3.8**). Comparison of two *Block 2* (red and raspberry) subunits with the 3' half of Z-G4 (blue) shows a high similarity with an all-atom RMSD < 1 Å.

Figure 3.8: Alignment of both *Block 2* subunits (in red and raspberry respectively) with the Z-G4 structure (blue). 3' thymine bases were removed for clarity. 5' and 3' ends of all 3 structures are labeled. (A) Outer G-tetrad view and (B) stacking/inner G-tetrad view. Alignment of all three structures gives an RMSD < 1 Å.

3.3 Connecting two *Block 2* motifs

Since *Block 2* requires dimerization through 5'-5' stacking in order to form a stable structure, we further investigated directly linking two *Block 2* sequences with a TT loop (**Table 3.4**).

Table 3.4: List of *2xBlock 2* sequences with terminal-thymine modifications.

Name	Sequence	Number of nucleotides
Z-G4	TGGTGGTGGTGGTTGTGGTGGTGGTGT	28
2xBlock 2	GTGGTGGTGGTGGTTGTGGTGGTGGTGT	27
2xBlock 2Δ	GTGGTGGTGGTGGTTGTGGTGGTGGTG	26
5'-2xBlock 2	T GTGGTGGTGGTGGTTGTGGTGGTGGTGT	28

The new *2xBlock 2* sequence interestingly preserves both the shape and intensity of the left-handed CD signal (**Figure 3.9, A**), while its NMR spectrum shows 16 imino proton peaks and sharp amino peaks (**Figure 3.9, B**), indicating the formation of a single, left-handed G4 structure. Presence of a 5'-T residue in *5'-2xBlock 2* does not prevent structure formation based on CD and NMR spectra, suggesting that strand progression from 5' as well as 3' end can be accommodated (**Figure 3.9**). Removing the 3' thymine from *2xBlock 2* does not impact the CD and NMR characteristics suggesting that the 3' thymine is not necessary for G-quadruplex structure formation and that *2xBlock 2* and *2xBlock 2Δ* form a very similar structure.

Figure 3.9: CD (**A**) and NMR (**B**) spectra of *2xBlock 2* and variants without 3' thymines or an additional 5' thymine.

3.4 2xBlock 2 forms a monomeric Left-handed G4 structure

We have successfully crystallized 2xBlock 2 in potassium buffer and solved its structure. The data collection and refinement statistics can be found in **Table 3.5** and in **Table 2.4** on page 50, where they can be compared to all of the solved structures. The structure was deposited in PDB databank with accession number **6GZ6**. The crystal packing is similar to previous observations for *Z-G4* and *Block 2*; the interactions are mediated by thymine residues in the T-cap and the block-connecting-loop either by π -stacking or engaging in direct thymine-thymine hydrogen bond formation (**Figure 3.10**). Eight out of eleven thymine residues are engaged in crystal lattice packing interactions.

Figure 3.10: Packing interactions of 2xBlock 2 in the crystal lattice. T13-T19 (green), T11-25 (magenta) and T5-T8 (red) are interacting through π - π stacking. T2-T27 (blue) are hydrogen bonding.

Table 3.5: Data collection and refinement statistics for *2xBlock 2* (PDB ID **6GZ6**).

Data collection	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
a, b, c (Å)	29.57, 37.63, 54.28
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	30.93 - 2.01
R _{sym}	0.089 (0.668)
I / σI	18.3 (3.95)
Completeness (%)	98.9 (93.6)
Redundancy	13.0 (13.4)
CC1/2 (%)	100 (96)
Refinement	
Resolution (Å)	30.93 – 2.01
Number of reflections	4285
R _{work} / R _{free}	0.185/0.231
Number of atoms	
DNA	569
K ⁺	3
Waters	10
B-factors (Å ²)	
	DNA 31.0 K ⁺ 22.1 Water 31.0
R.M.S.D	
Bond lengths (Å)	0.008
Bond angles (°)	0.997

[a] values in parentheses are for the highest resolution shell

The structure exhibits very similar features as were observed for *Block 2* dimer with a similar 5/6 ring stacking arrangement, a hydrogen bond between O4' sugars of T-cap thymine loops and amino protons of outer G-tetrad guanines. The consecutive guanine stacking within each block is partial 5/6-member ring with a twist mean angle value of $(-25.4 \pm 1)^\circ$. The backbone dihedral angles are similar to previous observations and can be found in **Table 3.6**.

Table 3.6: Average backbone dihedral angles of *2xBlock 2* and *Z-G4*.

Backbone dihedral	<i>2xBlock 2</i>	<i>Z-G4</i> ^[b]
ϵ	$(323 \pm 3)^\circ$	$(316 \pm 2)^\circ$
ζ	$(84 \pm 4)^\circ$	$(86 \pm 3)^\circ$
$\alpha+1$	$(164 \pm 6)^\circ$	$(161 \pm 6)^\circ$
$\beta+1$	$(222 \pm 4)^\circ$	$(226 \pm 5)^\circ$
$\gamma+1$	$(51 \pm 6)^\circ$	$(54 \pm 3)^\circ$

[a] Average values and standard deviations obtained from consecutive guanine dinucleotide steps. [b] Values reported in [53].

A potassium ion is present between each pair of G-tetrads (**Figure 3.11, A and C**) for a total of three ions coordinated between G-tetrads. Compared to *Z-G4*, where the loop between *Block 1* and *2* connects two neighbouring inner G-tetrads, the connecting loop of *2xBlock 2* spans 3 G-tetrads, from the outer G-tetrad of one block, to the inner G-tetrad of the 2nd block, and does not connect guanines that stack on each other (**Figure 3.11, B and C**). Neither *Z-G4* nor *2xBlock 2* possess the hydrogen bond observed between dG1-O5' and dG12-O3' of *Block 2*, indicating that it is likely an additional feature and not critical for left-handed G-quadruplex formation.

Figure 3.11: (A) Crystal structure of the left-handed G-quadruplex formed by *2xBlock 2* (PDB ID **6GZ6**). The electron density corresponds to the final 2mFo-Fc map contoured at 1.0 σ . (B) Schematic representation of the crystal structure with numbered guanine residues and marked backbone progression direction. (C) Ribbon representation of the crystal structure.

3.5 Thymine loop substitutions of the left-handed motif

To determine the sequence flexibility of left-handed quadruplexes we carried out a systematic thymine substitution analysis, while relying on established CD and NMR spectral characteristic features of solved left-handed quadruplexes. For these reasons we chose *2xBlock 2Δ* as our testing scaffold since it is very similar to the solved dimer structure of *Block 2* but does not require dimerization like *Block 2*. The NMR spectrum of *2xBlock 2Δ* is also cleaner with 16 imino peaks (**Figure 3.9, B**) as compared to *Block 2* (**Figure 3.4, B**) which has more than 16 peaks. This would make interpreting the influence of substitutions on the structure more difficult.

Figure 3.12: CD (A) and NMR (B) spectra of single thymine to cytosine substitutions sequences for all four thymine loops in a single block of *2xBlock 2Δ*.

5' block of the *2xBlock 2Δ* structure has four thymines that form the T-cap feature, T2, T5, T8 and T11, respectively. We have checked the effect of single T → C substitutions at all four thymine positions within a single block and found that they are all equivalent and introduce minimal disturbances both in CD and NMR spectra (**Figure 3.12, A and B**). The only substitution that differs slightly is when we replaced the thymine at position 2 to cytosine (**Figure 3.12, blue**). The intensity of CD signal for this sequence is slightly reduced and smaller peaks appear between the two main peak groupings in the imino region of the NMR spectrum.

Figure 3.13: CD spectra of *2xBlock 2Δ* sequences with different block connecting loop lengths (**A**) and thymine to adenine substitutions (**B**).

The block connecting loop of *2xBlock 2Δ* can be varied both in length (**Figure 3.13, A**) and composition (**Figure 3.13, B**), with only the loop of length one (blue) showing significant deviation from the expected CD spectrum. The composition of the loop does not seem to influence the observed CD spectrum as thymine nucleotides can be substituted with adenine nucleotides with minimal deviation in CD spectral shape.

In order to better translate the thymine T-cap substitution results from *2xBlock 2Δ* to the *Block 2* minimal left-handed motif, the substitutions were carried out either on a single block (**Figure 3.14, Figure 3.15, Figure 3.16**) or on both blocks simultaneously at both equivalent thymine positions, 2T and 16T for example (**Figure 3.17, Figure 3.18, Figure 3.19**). Three types of substitutions were carried out:

- T → C (**Figure 3.14, Figure 3.17**)
- T → A (**Figure 3.15, Figure 3.18**)
- T → TT (**Figure 3.16, Figure 3.19**).

Thymine to cytosine substitution is testing *2xBlock 2Δ* sensitivity to potential changes in pyrimidine hydrogen bonding as well as possible disruptions due to G-C base pairing. Thymine to adenine substitution is testing for *2xBlock 2Δ* sensitivity to the introduction of a bulkier purine base in the T-cap structure. Thymine to two thymines substitution is testing for *2xBlock 2Δ* sensitivity to loop elongation and potential T-cap and backbone disruptions. The list of all sequences used for substitution analysis can be found in **Table 2.3** on page 48.

Single block substitutions to either C (**Figure 3.14**), A (**Figure 3.15**) or TT (**Figure 3.16**) preserve some left-handed CD and NMR spectral characteristic even if all four substitutions are carried out, showing great robustness of *2xBlock 2Δ* to both loop composition as well as loop length.

Figure 3.14: CD (**A**) and NMR (**B**) spectra of *2xBlock 2Δ* sequences with thymine substitutions carried out on a single block. Incremental substitution of thymine loops, from one (blue) to two (red), three (green) and all four (purple) thymines substituted with cytosine.

The left-handed CD spectra (**Figure 3.14, A**), imino and amino NMR (**Figure 3.14, B**) peaks are well preserved for one (blue), two (red) or three (green) $T \rightarrow C$ substitutions. Only when all four $T \rightarrow C$ substitutions (purple) are carried out do smaller imino peaks significantly increase in intensity and we see the appearance of Watson-Crick imino proton peaks at 13 ppm suggesting G-C base-pairing. However the sharp imino peak at 9.5 ppm is still present and the CD spectrum primarily still assumes a left-handed shape, indicating that the primary folded structure is still likely to be a left-handed G-quadruplex.

Figure 3.15: CD (A) and NMR (B) spectra of *2xBlock 2 Δ* sequences with thymine substitutions carried out on a single block. Incremental substitution of thymine loops, from one (blue) to two (red), three (green) and all four (purple) thymines substituted with adenine.

Similar observations can be made for T $\rightarrow$ A substitutions. The left-handed CD spectra (**Figure 3.15, A**), imino and amino NMR (**Figure 3.15, B**) peaks are well preserved for one (blue), two (red) or three (green) substitutions. After all four thymines in a single block are substituted to adenine residues (purple), a broad hump appears in the imino NMR region. However, the sharp imino protons are still present in the same positions and the CD spectrum is the same as is observed for one adenine substitutions, indicating that the primary folded structure is still likely to be a left-handed G-quadruplex.

Figure 3.16: CD (A) and NMR (B) spectra of *2xBlock 2 Δ* sequences with thymine substitutions carried out on a single block. Incremental substitution of thymine loops, from one (blue) to two (red), three (green) and all four (purple) thymines substituted with two thymine residues.

The CD spectra (**Figure 3.16, A**) of all four constructs, with one (blue), two (red), three (green) or four (purple) T → TT substitutions, are the most different to the one observed for *2xBlock 2Δ*, as both their shape and intensity is perturbed. However the shape of the CD curves still seems to exhibit the characteristic left-handed 270 nm dip and 245 nm peak. The NMR spectra (**Figure 3.16, B**) also reflect this observation as no clear imino peak grouping is present while all four constructs have 16 imino peaks, indicating a single G-quadruplex structure formation. The presence of sharp imino peaks at 9.5 ppm and the primarily left-handed CD spectral shape of all four constructs would indicate that the primary folded structure is still likely to be a left-handed G-quadruplex for all four constructs.

Substitutions in both blocks at equivalent thymine positions to C (**Figure 3.17**), A (**Figure 3.18**) or TT (**Figure 3.19**) do not preserve the left-handed spectral characteristics as well as single block substitutions. 1-2 substitutions can be made in each block, or 3-4 total for *2xBlock 2Δ*, before spectral characteristics of left-handed G-quadruplexes are lost.

Figure 3.17: CD (**A**) and NMR (**B**) spectra of *2xBlock 2Δ* sequences with thymine substitutions carried out on both blocks. Incremental substitution of thymine loops, from one (blue) to two (red), three (green) and all four (purple) thymines substituted with cytosine at both equivalent positions in each block.

The left-handed CD spectra (**Figure 3.17, A**), imino and amino NMR (**Figure 3.17, B**) peaks are well preserved for a single T → C substitution in each block (blue). Two substitutions in each block (red), four in total, preserve the imino peak distribution and sharp amino peaks, but there are many imino peaks as well as the presence of G-C base pair imino peaks at 13 ppm, suggesting multiple structural conformations. This is reflected in the CD spectrum (red), where we can only observe hints of the left-handed CD spectrum, likely due to additional

contributions from multiple different folds present in the sample. In the case of three (green) or four (purple) T → C substitutions in each block, all left-handed CD and NMR spectral characteristics are lost. In the NMR spectrum, we can observe only Watson-Crick G-C base pair imino protons for both constructs, with only a slight hump present in the Hoogsteen imino region at 10.8 ppm. NMR results correlate with the observed CD spectra of both constructs as they are similar to reported B-DNA CD spectra. Only one or two T → C substitutions in each block would seem to be accommodated by the *2xBlock* 2Δ structure, before the loss of left-handed G-quadruplex spectral characteristics.

Figure 3.18: CD (A) and NMR (B) spectra of *2xBlock* 2Δ sequences with thymine substitutions carried out on both blocks. Incremental substitution of thymine loops, from one (blue) to two (red), three (green) and all four (purple) thymines substituted with adenine at both equivalent positions in each block.

Similar observations can be made for T → A substitutions (**Figure 3.18**). The left-handed CD spectra (**Figure 3.18, A**), imino and amino NMR (**Figure 3.18, B**) peaks are well preserved for a single T → A substitution in each block (blue). Two substitutions in each block (red), four in total, preserve the imino peak distribution and sharp amino peaks, but there are many imino peaks, suggesting multiple G-quadruplex conformations. This is reflected in the CD spectrum, where we can only observe hints of the left-handed CD spectrum, likely due to additional contributions from multiple different folded structures present in the sample. In the case of three (green) or four (purple) T → A substitutions in each block, all left-handed CD and NMR spectral characteristics are lost. In the NMR spectrum, we can observe only broad imino protons for both constructs. The CD spectra of both constructs show a very strong signal characteristic of right-handed parallel G-quadruplexes. Only one or two T → C substitutions

in each block would seem to be accommodated by the *2xBlock 2Δ* structure, before the loss of left-handed G-quadruplex spectral characteristics.

Figure 3.19: CD (A) and NMR (B) spectra of *2xBlock 2Δ* sequences with thymine substitutions carried out on both blocks. Incremental substitution of thymine loops, from one (blue) to two (red), three (green) and all four (purple) thymines substituted with two thymine residues at both equivalent positions in each block.

The NMR spectrum (**Figure 3.19, B**) of imino and amino peaks of one T → TT substitution in each block (blue), is the only one out of the four constructs that shows any imino proton peaks. There are many imino peaks present in the 11.8-10.5 ppm region, suggesting multiple G-quadruplex conformations but there is also a small amino peak present. This is reflected in the CD spectrum (**Figure 3.19, A**), where we can only observe hints of the left-handed CD spectrum, likely due to additional contributions from multiple different folds present in the sample. The constructs with two (red), three (green) or four (purple) T → TT substitution in each block do not form a stable structure as there are no observable imino or amino peaks in the NMR spectrum and they also lack any characteristic CD spectrum.

Table 3.7: Results of *2xBlock 2Δ* thymine loop substitutions.

Connecting loop modifications					
TT	TA	AT	AA	TTT	TTTT
✓	✓	✓	✓	✓	✓
Single block position modifications					
T2 to C	T5 to C	T8 to C	T11 to C		
✓	✓	✓	✓		
Substitution in single block					
		T to C	T to A	T to TT	
1 substitution		✓	✓	✓	
2 substitutions		✓	✓	✓	
3 substitutions		✓	✓	✓	
4 substitutions		✓	✓	✓	
Substitution in both blocks					
		T to C	T to A	T to TT	
2 substitution		✓	✓	✗	
4 substitutions		—	—	✗	
6 substitutions		✗	✗	✗	
8 substitutions		✗	✗	✗	

✓ - Left-handed G4 observed by NMR and CD
 — - Left-handed G4 suggested by NMR (CD patterns unclear)
 ✗ - Left-handed G4 not observed

Based on these results we can summarise that out of the 8 thymine loops that are part of the T-cap feature of *2xBlock 2Δ*, the left-handed structure can be retained while replacing 3-4 thymines with either C, A or TT residues. The block connecting loop can be varied both in length and composition without disruption in the CD spectra.

If we extend these results to the minimal left handed motif, *Block 2*, out of the four thymine loops that are part of the T-cap feature, the left-handed *Block 2* structure can likely be retained while substituting 1-2 thymines with A, C or TT residues. The results are summarized in **Table 3.7** while the complete list of sequences studied can be found in **Table 2.3**.

3.6 Effect of block order on left-handed G-quadruplex formation

Considering the sequence of *Z-G4* produces a structure with two discrete blocks, we checked what effect the 5' to 3' order of the blocks has on spectral characteristics and if the left-handed structure would be preserved. A sample with the reversed order of the two blocks,

named *Z-G4R* (**Table 3.8**), was prepared in ammonium buffer⁶ and its CD and NMR spectra were measured.

Table 3.8: List of sequences used in examining the effects of block order on left-handed G-quadruplex formation.

Name	Sequence	Number of nucleotides
<i>Z-G4</i>	TGGTGGTGGTGGTTGTGGTGGTGGTGT	28
<i>Z-G4R</i>	GTGGTGGTGGTGGTTGGTGGTGGTGGTT	27

Examining the CD spectrum of *Z-G4R* in ammonium buffer (**Figure 3.20, A**), we can notice that it is not exactly the same as that of *Z-G4* in ammonium buffer, with *Z-G4R* having a slight increase in intensity. Otherwise both sequences exhibit the characteristic left-handed CD spectrum with a trough at 270 nm and a peak at 245 nm [53]. Examining the NMR spectrum of *Z-G4* and *Z-G4R* (**Figure 3.20, B**) we can observe 16 imino peaks, indicating a formation of a single structure. We notice that imino peaks group into two distinct groups with each group having 8 peaks. A feature of left-handed G-quadruplex structure is the formation of hydrogen bonds between O4' of the capping thymines and the guanine amino protons of the adjacent G-tetrad, resulting in sharp amino protons at 9-10 ppm [59]; such protons are also observed for *Z-G4* and *Z-G4R* in ammonium buffer (**Figure 3.20, B**). Altogether, these results indicate that both *Z-G4* and *Z-G4R* form a left-handed G-quadruplex in the ammonium buffer.

Figure 3.20: CD (**A**) and NMR (**B**) data for *Z-G4R* (green) and *Z-G4* (black) folded in ammonium buffer.

⁶Effects of ions are discussed in chapters 1.7 on page 25 and 4.1 on page 90.

Z-G4R folded in ammonium buffer was successfully crystallized and its structure solved using *Block 2* from *Z-G4* (PDB ID **4U5M**) as a model for molecular replacement. The space group of the solved structure is $P3_112$ and the asymmetric unit contains a single molecule of *Z-G4R*, two barium ions and 21 water molecules. The data collection and refinement statistics can be found in **Table 3.9** and in **Table 2.4** on page 50, where they can be compared to all of the solved structures.

Table 3.9: Data collection and refinement statistics for *Z-G4R*.

Data collection	
Space group	$P3_112$
Cell dimensions	
a, b, c (Å)	31.32, 31.32, 118.32
α , β , γ (°)	90.0, 90.0, 120.0
Resolution (Å)	39.4 - 1.59
Completeness (%)	98.5 (96.9)
Refinement	
Resolution (Å)	10.0 – 1.6
Number of reflections	9020
$R_{\text{work}} / R_{\text{free}}$	0.185/0.214
Number of atoms	
DNA	569
Ba^{2+}	2
Waters	21
R.M.S.D	
Bond lengths (Å)	0.008
Bond angles (°)	0.998

[a] values in parentheses are for the highest resolution shell

The solved structure of *Z-G4R* is composed of two blocks which stack on the 5'-5' interface with a 5/6 ring overlap [28] where the first and last guanines, G1 and G25 respectively, are stacking (**Figure 3.21, B**). This is similar to *2xBlock 2*, where G1 and G24 are stacking while *Z-G4* is different and has G12 and G15 stacking. Even though the final guanine in the *Z-G4R* sequence G25, is part of the middle G-tetrad, the *Z-G4R* structure shows that strand progression can be accommodated without disrupting the structure, as G25 is followed by two thymine residues T26 and T27 (**Figure 3.21, A**).

Figure 3.21: (A) Ribbon representation of the left-handed G-quadruplex formed by *Z-G4R*. Thymine bases are shown in orange, guanines in cyan and Ba²⁺ ions as green spheres. (B) Schematic representation of the crystal structure with numbered guanine residues and marked backbone progression direction. (C) Cross-section of the left-handed G-quadruplex formed by *Z-G4R* with the 2mFo-Fc map contoured at 1.0 σ .

Thymines cap the outer G-tetrads of both blocks in *Z-G4R*. This creates the T-cap feature, observed in all solved left-handed G-quadruplex structures and also facilitates the hydrogen bond formation between the O4' sugar of thymines and amino protons of outer G-tetrad guanines, slowing their rotation and making them observable with NMR.

Compared to *Z-G4*, where the loop between *Block 1* and 2 connects two neighbouring inner G-tetrads, the connecting loop of *Z-G4R* spans 4 G-tetrads, from the outer G-tetrad of one block, to the outer G-tetrad of the 2nd block, and does not connect guanines that stack on each other as in *Z-G4*. Each block is composed of two G-tetrads stacked with a partial 5/6-member ring overlap and a left-handed helicity with a mean twist angle of $(-25.6 \pm 0.9)^\circ$. Dihedral backbone angles are summarized in **Table 3.10**.

Table 3.10: Average backbone dihedral angles of *Z-G4R* and *Z-G4*.

Backbone dihedral	<i>Z-G4R</i> ^[a]	<i>Z-G4</i> ^[b]
ϵ	$(317 \pm 5)^\circ$	$(316 \pm 2)^\circ$
ζ	$(83 \pm 2)^\circ$	$(86 \pm 3)^\circ$
$\alpha+1$	$(163 \pm 8)^\circ$	$(161 \pm 6)^\circ$
$\beta+1$	$(229 \pm 7)^\circ$	$(226 \pm 5)^\circ$
$\gamma+1$	$(51 \pm 4)^\circ$	$(54 \pm 3)^\circ$

[a] Average values and standard deviations obtained from the dinucleotide steps of consecutive guanines. [b] Values reported in [53].

The two G-tetrads of *Block 2* equivalent in *Z-G4R* are formed by G1•G10•G7•G4 and G3•G12•G9•G6 with identical hydrogen-bond directionalities, while the two G-tetrads

comprising the 3' *Block 1* are formed by G15•G24•G21•G18 and G16•G25•G22•G19 with identical hydrogen bond directionality. A Ba²⁺ ion is found coordinated between the two G-tetrads that comprise each block, while the third binding site, between the stacked blocks, does not have a coordinated ion (**Figure 3.21, A**).

Packing interactions can be divided into two types; thymine-thymine hydrogen bonding of T2-T2 residues (**Figure 3.22**, blue) and thymine π -stacking of T8-T8 residues (**Figure 3.22**, magenta) and T11-T13 residues (**Figure 3.22**, red). The 3' TT-tail has both types of packing interactions (**Figure 3.22**, green) with a 3' TT-tail of a neighboring G-quadruplex. Six out of twelve thymine residues are engaged in crystal lattice packing.

Figure 3.22: Crystal packing interactions of the *Z-G4R* G-quadruplex. Interacting thymine residues are labeled with their position in the primary sequence and color-coded based on discrete interactions.

Block 1 equivalent, which is on the 3' end of the *Z-G4R* sequence, superimposes to the *Block 1* of *Z-G4* with an RMSD of 2.77 Å for 214 atom pairs (**Figure 3.23, A**). The *Block 2* equivalents of *Z-G4* and *Z-G4R* superimpose with an RMSD of 0.22 Å for 209 atom pairs. In *Z-G4R*, *Block 2*

stays conserved when compared to solved structures of *Block 2*, *2xBlock2* and *Z-G4* (**Figure 3.23, B**), the higher RMSD of *Block 1* is mostly due to multiple thymine adjustments.

Figure 3.23: *Z-G4R Block 1 (A)* and *Block 2 (B)*, aligned with other solved left-handed G-quadruplexes, *Z-G4*, *Block 2*, *2xBlock2*, *Z-G4R*, *Z-G4-Bulge* and *Z-G4-2xBulge*, which are shown partially transparent for clarity.

Thymine bases T14 and T17 in *Z-G4R* and their *Z-G4* equivalents T1 and T4 switch roles (**Figure 3.24**). In *Z-G4* T1 is part of the T-cap feature and T4 flipped out. In *Z-G4R*, however, T14 (T1 in *Z-G4*) is flipped out while T17 (T4 in *Z-G4*) is part of the T-cap feature. The adjustments are related to differences in strand progression as *Block 1* now connects with a TT-loop to *Block 2* on the 5' end of the primary sequence (**Figure 3.21, B**).

Figure 3.24: Cartoon representation of thymine adjustments between *Z-G4R* (T14 and T17) and *Z-G4* (T4 and T1). Black arrows show the adjustment from *Z-G4R* to *Z-G4*. Thymines are colored orange, *anti* guanines cyan and *syn* guanines magenta.

The block connecting TT-loop in *Z-G4R* changes the glycosidic conformation of *Block 1* G15 to *anti* (**Figure 3.23, A**), whereas the *Z-G4* equivalent G2 is in the *syn* glycosidic conformation (magenta). All guanines of *Z-G4R* are thus observed in the *anti* glycosidic conformation, while *Z-G4* has one guanine, G2, in the *syn* conformation.

3.7 Effect of splitting guanines on the left-handed G-quadruplex structure

The only difference in the primary sequence of *Block 1* and *Block 2* is the split guanine pair in *Block 2* (see page 54 for details). In *Block 1* all four pairs of guanines are continuous. In order to try and replicate the split guanine feature of *Block 2*, we have introduced an additional thymine residues in the *Block 1* sequence to progressively split up guanine pairs. This resulted in four sequences with 1-4 additional thymine residues (**Table 3.11**).

Table 3.11: Sequences used for exploring the effect of splitting guanine pairs in *Block 1* with thymine residues (shown in red).

Name	Sequence	Number of nucleotides
<i>Z-G4</i>	TGGTGGTGGTGGTGTGGTGGTGGTGT	28
<i>Block 1</i>	GGTGGTGGTGG	11
<i>Block 1-Bulge</i>	GGTGTGTGGTGG	12
<i>Block 1-2xBulge</i>	GGTGTGTGTGTGG	13
<i>Block 1-3xBulge</i>	GGTGTGTGTGTGTG	14
<i>Block 1-4xBulge</i>	GTTGTGTGTGTGTG	15
<i>Z-G4-Bulge</i>	GGTGTGTGGTGGTGTGGTGGTGGTGT	27
<i>Z-G4-2xBulge</i>	GGTGTGTGTGTGGTGTGGTGGTGGTGT	28
<i>Z-G4-3xBulge</i>	GGTGTGTGTGTGTGTGTGGTGGTGGTGT	29
<i>Z-G4-4xBulge</i>	GTTGTGTGTGTGTGTGTGGTGGTGGTGT	30

All four *Block 1-Bulge* sequences were folded in potassium buffer and their CD and NMR spectra were measured (**Figure 3.25**). The CD spectrum (**Figure 3.25, A**) of *Block 1-Bulge* (purple) has two peaks, a smaller one at 290 nm and a larger one at 260 nm, indicating a formation of a right-handed G-quadruplex structure. The NMR spectrum shows 6 imino peaks in the 12-10.5 ppm Hoogsteen region and is in agreement with the CD data, indicating a single G-quadruplex structure formation. We can conclude that the introduction of a single thymine residue to split continuous guanines can be tolerated and a G-quadruplex can be formed. The

CD spectrum and the lack of any sharp amino peaks in the 10-9 ppm NMR region, however, suggests that the G-quadruplex formed is not left-handed.

Figure 3.25: CD (A) and NMR (B) spectra of *Block 1* sequences with one bulge (purple), two bulges (orange), three bulges (green) and four bulges (blue) introduced.

The CD spectra (**Figure 3.25, A**) of *Block 1-2xBulge* (orange), *Block 1-3xBulge* (green) and *Block 1-4xBulge* (blue) do not display features characteristic of G-quadruplex formation. Their corresponding NMR spectra (**Figure 3.25, B**) also lack observable imino peaks that would indicate a formation of a structure. Additional thymine residues splitting the continuous guanines are likely too disruptive for the formation of a folded structure as was observed for *Block 1* (**Figure 3.3**, page 56) and *Block 1-Bulge* (**Figure 3.25**, purple).

Since none of the four *Block 1* bulge constructs formed a left-handed G-quadruplex on their own, we added the *Block 2* sequence on the 3' end, due to its intrinsic left-handed propensity. The two blocks were connected with a single thymine loop (**Table 3.11**) in contrast with previous structures where two thymine residues connected the two blocks. The four new constructs are similar to the *Z-G4* sequence but also differ from it by having a single instead of two thymine loop, having 1-4 extra bulges in *Block 1* and by not having a 5' terminal thymine. All four sequences were folded in potassium buffer and their CD and NMR spectra were recorded (**Figure 3.26**). CD spectra of three constructs, *Z-G4-Bulge* (purple), *Z-G4-2xBulge* (orange) and *Z-G4-3xBulge* (green) show a familiar left-handed G-quadruplex spectral profile (**Figure 3.26, A**) with a trough at 270 nm and a peak at 245 nm. The fourth construct, *Z-G4-4xBulge* (blue), does not display a CD spectrum that is distinct for G-quadruplex structures.

Figure 3.26: CD (A) and NMR (B) spectra of modified Z-G4 sequences with one bulge (purple), two bulges (orange), three bulges (green) or four bulges (blue) introduced in *Block 1*.

NMR spectra (**Figure 3.26, B**) of *Z-G4-Bulge* (purple), *Z-G4-2xBulge* (orange) and *Z-G4-3xBulge* (green) have 16 imino proton peaks arranged into two groups of eight in the Hoogsteen 11.8 to 10.5 ppm ^1H NMR region, indicating a formation of a single structure with four G-tetrads. We can also observe amino proton peaks around 9.3 ppm for all three constructs, indicating slower rotation of the guanine amino group [160], one of the characteristics of left-handed G-quadruplex structures[53, 59]. The fourth construct, *Z-G4-4xBulge* (blue), which has all four of the guanine pairs in *Block 1* split by thymine, does not have clear imino or amino proton peaks, indicating that it does not fold into a single stable left-handed G-quadruplex structure.

3.7.1 Structure of *Z-G4-Bulge*

A sample of *Z-G4-Bulge* folded in potassium buffer was successfully crystallized (paragraph 2.6.4, page 52 for methods details) and its structures solved to a resolution of 1.18 Å, using *Block 2* of *Z-G4* as a model for molecular replacement. The space group of the solved structures is P12₁1 and the asymmetric unit contains a single G-quadruplex, 3 potassium ions, two spermine molecules and 158 waters. The data collection and refinement statistics can be found in **Table 3.12** and in **Table 2.4** on page 50, where they can be compared to all of the solved structures.

Table 3.12: Data collection and refinement statistics for *Z-G4-Bulge*.

Data collection	
Space group	P12 ₁ 1
Cell dimensions	
a, b, c (Å)	29.22, 40.04, 30.38
α, β, γ (°)	90.0, 115.87, 90.0
Resolution (Å)	40.4 – 0.972
R _{sym}	0.112 (0.94)
I / σI	8.67 (1.52)
Completeness (%)	100 (100)
Redundancy	6.5 (5.9)
CC1/2 (%)	100 (53)
Refinement	
Resolution (Å)	26.3 – 1.18
Number of reflections	20678
R _{work} / R _{free}	0.124/0.168
Number of atoms	
DNA	569
K ⁺	3
Spermine	56
Waters	158
B-factors (Å ²)	
	DNA 10.5 K ⁺ 6.1 Spermine 22.7 Water 33.27
R.M.S.D	
Bond lengths (Å)	0.017
Bond angles (°)	1.27

[a] values in parentheses are for the highest resolution shell

The solved structure of *Z-G4-Bulge* is composed of two blocks which stack on the 5'-5' interface with a 5/6 ring overlap [28] (**Figure 3.27, B**). The stacking is similar to that observed in *Z-G4*, where two guanines connected by the TT-loop, G12 and G15, also stack. Length of the block connecting loop of *Z-G4* and *Z-G4-Bulge* is different, with *Z-G4* having two and *Z-G4-Bulge* having one thymine residue in the loop (**Table 3.11**). The thymine loop connecting the blocks connects two neighboring, inner G-tetrads, as was previously observed for *Z-G4*. This is different compared to the two other structures already examined in this work, *2xBlock 2* and *Z-G4R*, where the connecting loop spans three and four G-tetrads, respectively.

Figure 3.27: (A) Ribbon representation of the left-handed G-quadruplex formed by *Z-G4-Bulge*. Thymine bases are shown in orange, guanines in cyan, spermines in green, K⁺ ions as purple spheres and O4' oxygens as yellow spheres. (B) Schematic representation of the crystal structure of *Z-G4-Bulge* with numbered guanine residues and marked backbone progression direction. (C) Cross-section of the left-handed G-quadruplex formed by *Z-G4-Bulge* with the 2mFo-Fc map contoured at 1.0 σ .

The two G-tetrads of the 3' *Block 2* equivalent are formed by G14•G23•G20•G17 and G16•G25•G22•G19 with identical hydrogen-bond directionalities, while the two G-tetrads comprising the 5' *Block 1* are formed by G1•G11•G8•G4 and G2•G12•G9•G6 with identical hydrogen bond directionality. A K⁺ ion is found coordinated between each pair of stacked G-tetrads for a total of three potassium ions. Two spermine molecules are found on top of each outer G-tetrad face between thymine residues comprising the T-cap (**Figure 3.27, A and C**).

Thymines cap the outer G-tetrads of both blocks in *Z-G4-Bulge*. This creates the T-cap feature, observed in all solved left-handed G-quadruplex structures and also facilitates the hydrogen bond formation between the O4' sugar of thymines and amino protons of outer G-tetrad guanines, slowing their rotation and making them observable with NMR. The T-cap structure of *Z-G4-Bulge* does, however, differ compared to other structures presented. Where usually

four thymines would be found in a complete T-cap, we observe only three thymine residues capping the 5' block of *Z-G4-Bulge*. This suggests that a formation of a complete T-cap structure might not be essential for the formation of left-handed G-quadruplexes.

Both blocks are composed of two G-tetrads stacked with a partial 5/6-member ring overlap and a left-handed helicity with a mean twist angle of $(-26.7 \pm 2)^\circ$ for consecutive guanines. Dihedral backbone angles of continuous guanine pairs are summarized in **Table 3.3** and do not differ significantly from previous values for left-handed G-quadruplexes.

Table 3.13: Average backbone dihedral angles of *Z-G4-Bulge* and *Z-G4* for comparison.

Backbone dihedral	<i>Z-G4-Bulge</i>	<i>Z-G4</i> ^[b]
ϵ	$(312 \pm 1)^\circ$	$(316 \pm 2)^\circ$
ζ	$(88 \pm 3)^\circ$	$(86 \pm 3)^\circ$
$\alpha+1$	$(172 \pm 2)^\circ$	$(161 \pm 6)^\circ$
$\beta+1$	$(228 \pm 3)^\circ$	$(226 \pm 5)^\circ$
$\gamma+1$	$(47 \pm 6)^\circ$	$(54 \pm 3)^\circ$

[a] Average values and standard deviations obtained from the dinucleotide steps of consecutive guanines. [b] Values reported in [53].

The crystal packing interactions of *Z-G4-Bulge* are extensive, as 9 out of 11 thymine residues engage in π -stacking (**Table 3.14**). The bulge T5 is π -stacking with a T-cap T18 (**Figure 3.28**, blue), T-cap residues T7 and T21 together with the block connecting loop T13 triple π -stack (green), the 3' tail residues T26 and T27 stack with T-cap residues T15 and T24, respectively (red). Second to last thymine, T26, also completes the T-cap feature of a neighboring G-quadruplex by occupying the vacant spot.

Table 3.14: Table of packing interactions of *Z-G4-Bulge*.

	<i>Z-G4-Bulge</i>
Bulge and T-cap thymine π - π stacking	T5-T18
T-cap/connecting loop/T-cap thymine π - π - π stacking	T7-T13-T21
T-cap and 3' tail π - π stacking	T15-T26
T-cap and 3' tail π - π stacking	T24-T27
3' tail completes T-cap	T26

Figure 3.28: Crystal packing interactions of the *Z-G4-Bulge* G-quadruplex. Interacting thymine residues are labeled with their position in the primary sequence and color-coded based on discrete interactions.

Block 1 equivalent, which is on the 5' end of the *Z-G4-Bulge* sequence, superimposes to the *Block 1* of *Z-G4* with an RMSD of 2.54 Å for 191 atom pairs (**Figure 3.29, A**). The higher RMSD of *Block 1* equivalent in *Z-G4-Bulge* is mostly due to sequence dissimilarities, as the core of the structure is conserved with an RMSD of 0.178 Å for 175 atom pairs. The *Block 2* equivalent of *Z-G4* and *Z-G4-Bulge* superimpose with an RMSD of 0.175 Å for 197 atom pairs. In *Z-G4-Bulge*, the *Block 2* equivalent stays conserved when compared to other solved structures of *Block 2*, *2xBlock2*, *Z-G4* and *Z-G4R* (**Figure 3.29, B**).

Figure 3.29: *Z-G4-Bulge Block 1 (A) and Block 2 (B)*, aligned with other solved left-handed G-quadruplexes, *Z-G4*, *Block 2*, *2xBlock 2*, *Z-G4R* and *Z-G4-2xBulge*, which are shown partially transparent for clarity.

The 5' terminal thymine residue T1 in *Z-G4* forces G2 to *syn* conformation in order to accommodate strand progression [53]. In *Z-G4-Bulge* the 5' thymine is absent and the first guanine is found in the *anti* glycosidic conformation. All guanines of *Z-G4-Bulge* are thus observed in the *anti* glycosidic conformation, while *Z-G4* has one guanine, the aforementioned G2 in the *syn* conformation.

The thymine T5 of *Z-G4-Bulge* (**Figure 3.29, A**) folds as an outwards projecting bulge and does not disrupt the two surrounding guanine residues significantly, when aligned and compared with other *Block 1* structures. It resembles propeller loops of right-handed parallel G-quadruplexes [36, 49, 161, 162] and has the opposite backbone directionality when compared to the rest of the residues in *Z-G4-Bulge* (**Figure 3.27, B**).

3.7.2 Structure of *Z-G4-2xBulge*

A sample of *Z-G4-2xBulge* folded in potassium buffer was successfully crystallized (paragraph 2.6.5, page 52 for methods details) and its structures solved to a resolution of 1.3 Å, using *Block 2* of *Z-G4* as a model for molecular replacement. The space group of the solved structures is P12₁1 and the asymmetric unit contains two G-quadruplexes, eight potassium ions, four spermine molecules and 210 waters. The data collection and refinement statistics

can be found in **Table 3.15** and in **Table 2.4** on page 50, where they can be compared to all of the solved structures.

Table 3.15: Data collection and refinement statistics for *Z-G4-2xBulge*.

Data collection	
Space group	P12 ₁ 1
Cell dimensions	
a, b, c (Å)	29.10, 39.81, 54.97
α , β , γ (°)	90.0, 91.28, 90.0
Resolution (Å)	29.1 – 1.30
R _{sym}	0.093 (0.797)
I / σ I	6.86 (1.07)
Completeness (%)	95 (93)
Redundancy	6.0 (4.9)
CC1/2 (%)	100 (73.4)
Refinement	
Resolution (Å)	29.1 – 1.30
Number of reflections	29721
R _{work} / R _{free}	0.184/0.215
Number of atoms	
DNA	1172
K ⁺	8
Spermine	56
Waters	210
B-factors (Å ²)	
	DNA 14.5 K ⁺ 10.4 Spermine 33.1 Water 27.9
R.M.S.D	
Bond lengths (Å)	0.007
Bond angles (°)	1.03

[a] values in parentheses are for the highest resolution shell

The solved structure of *Z-G4-2xBulge* is composed of two blocks which stack on the 5'-5' interface with a 5/6 ring overlap [28]. The stacking is similar to that observed in *Z-G4* and *Z-G4-Bulge*, where the last guanine in *Block 1*, G13 and first guanine of *Block 2*, G15 stack and are connected by the block connecting loop (**Figure 3.30, B**). The difference between the block connecting loops of *Z-G4* and *Z-G4-2xBulge* is the length of the loop with *Z-G4* having two and *Z-G4-2xBulge* having one thymine residue, respectively. The thymine loop connecting the blocks connects two neighboring, inner G-tetrads, as was previously observed for *Z-G4*. This

is different compared to two other structures in this work, *2xBlock 2* and *Z-G4R*, where the connecting loop spans three and four G-tetrads, respectively.

Figure 3.30: (A) Ribbon representation of the left-handed G-quadruplex formed by *Z-G4-2xBulge*. Both subunit A and B of the asymmetric unit are shown. Thymine bases are shown in orange, guanines in cyan, spermines in green, K⁺ ions as purple spheres and O4' oxygens as yellow spheres. (B) Schematic representation of the crystal structure of *Z-G4-2xBulge* with numbered guanine residues and marked backbone progression direction. (C) Cross-section of the left-handed G-quadruplex formed by *Z-G4-2xBulge* with the 2mFo-Fc map contoured at 1.0 σ .

Thymines cap the outer G-tetrads of both blocks in *Z-G4-2xBulge* (Figure 3.30, A). This creates the T-cap feature, observed in all solved left-handed G-quadruplex structures and also facilitates the hydrogen bond formation between the O4' sugar of thymines and amino protons of outer G-tetrad guanines, slowing their rotation and making them observable with NMR. The T-cap structure of *Z-G4-2xBulge* is similar to the one observed for *Z-G4-Bulge*. Where usually four thymine would be found in a complete T-cap, we observe only three thymine residues capping the 5' block of *Z-G4-2xBulge*. This suggests that a formation of a

complete T-cap structure might not be essential for the formation of left-handed G-quadruplexes.

The two G-tetrads of the 3' *Block 2* equivalent are formed by G15•G24•G21•G18 and G17•G26•G23•G20 with identical hydrogen-bond directionalities, while the two G-tetrads comprising the 5' *Block 1* equivalent are formed by G1•G12•G8•G4 and G2•G13•G10•G6 with identical hydrogen bond directionality (**Figure 3.30, B**). A K⁺ ion is found coordinated between each pair of stacked G-tetrads for a total of three potassium ions in each of the subunits with a further two potassium ions found next to Subunit B (**Figure 3.30, A**). Four spermine molecules (green), two for each G-quadruplex of the asymmetric unit, are found on the outer G-tetrad face between thymine residues comprising the T-cap.

Both blocks are composed of two G-tetrads stacked with a partial 5/6-member ring overlap and a left-handed helicity with a mean twist angle of $(-27 \pm 2)^\circ$ for consecutive guanines. Dihedral backbone angles of continuous guanine pairs are summarized in **Table 3.16** and do not differ significantly from previous values for left-handed G-quadruplexes.

Table 3.16: Average backbone dihedral angles of *Z-G4-2xBulge* and *Z-G4*.

Backbone dihedral	<i>Z-G4-2xBulge</i>	<i>Z-G4</i> ^[b]
ϵ	$(317 \pm 2)^\circ$	$(316 \pm 2)^\circ$
ζ	$(86 \pm 3)^\circ$	$(86 \pm 3)^\circ$
$\alpha+1$	$(162 \pm 7)^\circ$	$(161 \pm 6)^\circ$
$\beta+1$	$(228 \pm 6)^\circ$	$(226 \pm 5)^\circ$
$\gamma+1$	$(51 \pm 4)^\circ$	$(54 \pm 3)^\circ$

[a] Average values and standard deviations obtained from the dinucleotide steps of consecutive guanines. [b] Values reported in [53].

The crystal packing interactions of *Z-G4-2xBulge* are extensive and similar to *Z-G4-Bulge*, with 11 out of 12 thymine residues engage in π -stacking (**Table 3.17**). The bulge T5 is π -stacking with a T-cap T18 (**Figure 3.31**, blue) while the bulge T9 is π -stacking with the T-cap T3 (magenta). T-cap residues T7 and T22 together with the block connecting loop T14 triple π -stack (green), the 3' tail residues T27 and T28 π -stack with T-cap residues T16 and T25, respectively (red). Second to last thymine, T27, also completes the T-cap feature of a neighboring G-quadruplex by occupying the vacant spot (**Figure 3.31, B**).

Table 3.17: Table of packing interactions of *Z-G4-2xBulge*.

	<i>Z-G4-2xBulge</i>
Bulge and T-cap thymine π - π stacking	T5-T19
T-cap/connecting loop/T-cap thymine π - π - π stacking	T7-T14-T22
T-cap and Bulge thymine π - π stacking	T3-T9
T-cap and 3' tail π - π stacking	T16-T27
T-cap and 3' tail π - π stacking	T25-T28
3' tail completes T-cap	T27

Figure 3.31: Crystal packing interactions of the *Z-G4-2xBulge* G-quadruplex. Interacting thymine residues are labeled with their position in the primary sequence and color-coded based on discrete interactions. Views A and B show the same molecules with different orientations, in order to visualize the packing interactions.

The two *Block 1* equivalents, which are on the 5' end of both *Z-G4-2xBulge* subunits in the asymmetric unit, superimpose to *Block 1* of *Z-G4* with an RMSD of 2.55 Å for 191 atom pairs (**Figure 3.32, A**). The higher RMSD of *Block 1* equivalent in *Z-G4-2xBulge* is mostly due to sequence dissimilarities, as the core of the structure is conserved with an RMSD of 0.178 Å for 175 atom pairs. *Block 1* equivalents from both subunits in the asymmetric unit superimpose with each other with an RMSD of 1.13 Å for 273 atom pairs, reflecting the adjustment of the T9 thymine bulge, likely due to crystal packing (**Figure 3.32, A**).

Both *Block 2* equivalents in *Z-G4-2xBulge* superimpose with an RMSD of 0.52 Å for 234 atom pairs with the *Z-G4* equivalent *Block 2*, reflecting the conserved nature of the minimal left-handed motif (**Figure 3.32, B**). In *Z-G4-2xBulge*, the *Block 2* equivalent stays conserved when compared to solved structures of *Block 2*, *2xBlock2*, *Z-G4*, *Z-G4R* and *Z-G4-Bulge* (**Figure 3.32, B**).

Figure 3.32: *Z-G4-2xBulge* Block 1 (A) and Block 2 (B) equivalents from both subunits, aligned with other solved left-handed G-quadruplexes, *Z-G4*, *Block 2*, *2xBlock 2*, *Z-G4R* and *Z-G4-Bulge*, which are shown partially transparent for clarity.

The 5' terminal thymine residue T1 in *Z-G4* forces G2 to *syn* conformation in order to accommodate strand progression. In *Z-G4-2xBulge* the 5' thymine is absent and the first guanine is found in the *anti* glycosidic conformation, as is also the case for *Z-G4-Bulge*. All guanines of *Z-G4-2xBulge* are thus observed in the *anti* glycosidic conformation. *Z-G4* remains the only left-handed G-quadruplex structure observed so far that has the G2 guanine or equivalent in the *syn* conformation.

The thymines T5 and T9 of *Z-G4-2xBulge* (Figure 3.32, A) fold as an outwards projecting bulge and do not disrupt the two surrounding guanine residues significantly, when aligned and compared with other *Block 1* structures. Thymine T9 adopts two slightly different orientations in each subunit, likely due to crystal packing interaction constraints and suggests that the bulges might have flexibility in the left-handed G-quadruplexes. The position and orientation of T5 is highly convergent between the two subunits. Both T5 and T9 bulges resembles propeller loops of right-handed parallel G-quadruplexes [36, 49, 161, 162] and have the opposite backbone directionality when compared to the rest of the residues in *Z-G4-2xBulge* (Figure 3.30, B).

3.8 Analysis and discussion of left-handed G-quadruplex sequence and structure flexibility

The split-guanines feature of *Block 2* and *2xBlock 2* sequences represents the only major difference in the primary sequence when compared to the sequence of *Block 1* (**Table 3.1**, page 56). The sequence analysis of *Z-G4* together with new experimental evidence of the solved *Block 2* (**Figure 3.6**, page 60) and *2xBlock 2* (**Figure 3.11**, page 65) structures shows that a shorter motif of 12-nt (GTG)₄ is essential and sufficient for the formation of a G-quadruplex with a left-handed helicity.

The motif deviates significantly from the previously proposed model sequences, such as G₃₊N₁₋₇G₃₊N₁₋₇G₃₊N₁₋₇G₃₊, used for finding putative G-quadruplex forming sequences [134, 163]. The repetitive nature of the minimal left-handed motif sequence makes the sequence occurrence in the human genome highly abundant, with a two-order of magnitude enrichment and over 10,000 hits. A random 12-mer DNA sequence would produce 360 hits in a haploid human genome by chance, so the increase is very substantial. *Z-G4*, without 5' and 3' thymine residue, produced 8 hits (1/200,000 odds for a random 25-mer sequence), which is also a substantial increase over chance [53]. The need of the minimal motif for 5'-5' block stacking in order to form a stable structure, based on all the solved structures, does however reduce the possible left-handed G-quadruplex forming sequence space.

The *Z-G4R* structure shows that the order of the blocks in the primary sequence is interchangeable, as the minimal left-handed motif of *Block 2* is able to drive left-handed structure formation from both the 5' (*Z-G4R*) and 3' end (*Z-G4*). With block connecting loops (**Figure 3.33**) from *Z-G4* (**A**), *Z-G4-Bulge* (**B**), *2xBlock 2* (**C**) and *Z-G4R* (**D**) we demonstrate that the block connecting loop is able to span two, three or four G-tetrads. The flexibility of its connectivity would suggest that the left-handed helicity is likely not conveyed from the minimal left-handed motif *Block 2*, to *Block 1* through the connecting loop.

Figure 3.33: Comparison of three different connecting loop configurations. **(A)** TT loop of Z-G4, connecting two inner G-tetrads. **(B)** T loop of Z-G4-Bulge, connecting two inner G-tetrads. **(C)** TT loop of 2xBlock 2, connecting an outer G-tetrad to an inner G-tetrad. **(D)** TT loop of Z-G4R connecting the two outer G-tetrads and spanning the entire height of the G-quadruplex.

The block connecting loop length and composition analysis performed on the 2xBlock 2 sequence (**Figure 3.13**, page 67), also suggests that the blocks do not necessarily need to be extremely close in the primary sequence, in order to stack and form a left-handed G-quadruplex.

Since the split guanines in *Block 2* seemed like a feature that might contribute to the left-handed twist, we tried to recreate this feature in the *Block 1* sequence (paragraph 3.7, page 79). The four pairs of continuous guanines were progressively split with thymine residues, ranging from one split guanine pair to all four guanine pairs being split (**Table 3.11**, page 79). Only one construct formed a structure, likely a right-handed hybrid G-quadruplex based on NMR and CD spectra (**Figure 3.25**, page 80). The three constructs with two, three or four split guanine pairs did not form a structure. What proved very interesting, however, is when we attached a *Block 2* sequence on the 3' end of all four constructs. Three out of four constructs,

with one, two or three split guanine pairs respectively, not only folded into a single G-quadruplex structure, but exhibited very clear spectral properties of left-handed G-quadruplexes (**Figure 3.26**, page 81). The sequence of *Block 1-3xBulge* equivalent in *Z-G4-3xBulge* only has one continuous guanine pair, while the rest are single guanines (5'-GGT GTG T GTG T GTG-3'). That *Z-G4-3xBulge* is able to fold into a single G-quadruplex structure is interesting by itself as traditionally such sequences are not considered to form G-quadruplexes due to distinct lack of G-tracts. Solving the structures of *Z-G4-Bulge* (**Figure 3.27**, page 83) and *Z-G4-2xBulge* (**Figure 3.30**, page 88) confirmed that *Block 1-Bulge* and *Block 1-2xBulge* equivalents were indeed left-handed and based on CD and NMR spectral similarities, *Z-G4-3xBulge* is also very likely left-handed (**Figure 3.26**, page 81).

Figure 3.34: Aligned *Block 1* and *Block 2* equivalents from the six solved left-handed G-quadruplexes. There are five unique versions of *Block 1* and nine unique versions of *Block 2*. (A) Aligned *Block 1* equivalents with annotated differences between the structures. (B) Aligned *Block 2* equivalents. *Anti* guanine bases are shown in cyan, *syn* guanines in magenta, thymine bases in orange, backbone in grey and O4' oxygens in yellow.

Aligning *Block 1* and *Block 2* equivalents of all six left-handed structures (**Figure 3.34**) we can observe some interesting patterns. In the six solved structures there is a total of nine unique *Block 2* equivalents (**Figure 3.34, B**). We can notice that the *Block 2* motif is conserved, especially in the guanine core and backbone regions. Only the thymine bases of the T-cap exhibit minor changes in position and orientation. There is a total of five unique *Block 1* equivalents across the six left-handed structures solved so far (**Figure 3.34, A**). The guanine core is conserved in much the same way as we observe for *Block 2* guanine bases. We can notice that the thymines and the backbone is far less conserved, both due to differences in sequence and structural features. *Z-G4 Block 1* is the only one with a *syn* guanine residue, G2 (magenta). All of the other guanines in both blocks are found in the *anti* glycosidic conformation.

Figure 3.35: Surface representations of *Block 1* equivalents with T-cap and bulge features of *Z-G4* (A), *Z-G4R* (B), *Z-G4-Bulge* (C) and *Z-G4-2xBulge* (D). Thymine bases are shown in orange, guanine bases in cyan and the backbone in grey. Asterisks denote the vacant T-cap position.

Z-G4 also has the only *Block 1* with an outward projecting thymine loop (T4), due to T1 taking its place in the T-cap feature (**Figure 3.35, A**). The T4 loop does however resemble the bulges observed in *Z-G4-Bulge* (**C**) and *Z-G4-B2xBulge* (**D**) as its sugar is oriented in the opposite direction than the sugars of the other thymine and guanine residues (**Figure 3.34, A**). The bulges T9 of *Z-G4-2xbulge* (**Figure 3.35, D**) and T5 of both *Z-G4-Bulge* (**C**) and *Z-G4-2xBulge* (**D**) surprisingly do not disrupt the surrounding guanine residues. They project out and resemble propeller loops of right-handed parallel G-quadruplexes. The final point of difference between *Block 1* structures is the absence of the final thymine from the T-cap feature in *Z-G4-Bulge* (**Figure 3.35, C**) and *Z-G4-2xBulge* (**D**). This makes them the only two blocks that do not have four thymine residues in the T-cap feature. In the crystal packing of both structures, the second-to-last thymine residue, T26 (**Figure 3.28**, page 85) and T27 (**Figure 3.31**, page 90) respectively, does occupy the vacant T-cap position of a neighboring G-quadruplex, suggesting it might be energetically favorable to do so. Out of the five *Block 1* equivalents, the one from *Z-G4R* most closely resembles the conserved *Block 2* (**Figure 3.35, B**).

Table 3.18: Average backbone dihedral angles of *Block 2*, *2xBlock 2* and *Z-G4*.

Backbone dihedral	<i>Block 2</i>	<i>2xBlock 2</i>	<i>Z-G4R</i>	<i>Z-G4-Bulge</i>	<i>Z-G4-2xBulge</i>	<i>Z-G4</i> ^[b]
ϵ ^[a]	(321 ± 2)°	(323 ± 3)°	(317 ± 5)°	(312 ± 1)°	(317 ± 2)°	(316 ± 2)°
ζ ^[a]	(82 ± 1)°	(84 ± 4)°	(83 ± 2)°	(88 ± 3)°	(86 ± 3)°	(86 ± 3)°
$\alpha+1$ ^[a]	(159 ± 4)°	(164 ± 6)°	(163 ± 8)°	(172 ± 2)°	(162 ± 7)°	(161 ± 6)°
$\beta+1$ ^[a]	(232 ± 6)°	(222 ± 4)°	(229 ± 7)°	(228 ± 3)°	(228 ± 6)°	(226 ± 5)°
$\gamma+1$ ^[a]	(53 ± 2)°	(51 ± 6)°	(51 ± 4)°	(47 ± 6)°	(51 ± 4)°	(54 ± 3)°
Guanine twist	(−25.2 ± 0.6)°	(−25.4 ± 1)°	(−25.6 ± 0.9)°	(−26.7 ± 2)°	(−27 ± 2)°	(−25.2 ± 0.6)°

[a] Average values and standard deviations obtained from the dinucleotide steps of consecutive guanines.

[b] Values reported in [53].

Comparing the dihedral angles of all six left-handed structures (**Table 3.18**), we can confirm that the backbone orientation and progression for consecutive guanine dinucleotides is conserved, as the differences between structures and within structures are small. The guanine twist also stays roughly the same, with only *Z-G4-Bulge* and *Z-G4-2xBulge* having a slightly less acute angle. This might explain the slightly more intense CD spectrum as compared to *Z-G4*.

The sequence flexibility we have demonstrated in our *2xBlock 2* substitution analysis (chapter 3.5, page 66), where three to four of eight T-cap thymines can be substituted without loss of

left-handed spectral properties, significantly increases the putative left-handed G-quadruplex sequence space. Especially interesting are single-block substitutions where all four thymine residues can be substituted with C, A or TT residues, while still retaining left-handed spectral characteristics. Substitutions performed in both blocks at equivalents positions (T2 and T16 for a single substitution for example) is less robust as left-handed CD and NMR spectral features start getting distorted even at one or two substitutions (**Table 3.19**).

Table 3.19: Results of *2xBlock 2Δ* thymine loop substitutions.

Connecting loop modifications					
TT	TA	AT	AA	TTT	TTTT
✓	✓	✓	✓	✓	✓
Single block position modifications					
T2 to C	T5 to C	T8 to C	T11 to C		
✓	✓	✓	✓		
Substitution in single block					
		T to C	T to A	T to TT	
1 substitution		✓	✓	✓	
2 substitutions		✓	✓	✓	
3 substitutions		✓	✓	✓	
4 substitutions		✓	✓	✓	
Substitution in both blocks					
		T to C	T to A	T to TT	
2 substitution		✓	✓	✗	
4 substitutions		—	—	✗	
6 substitutions		✗	✗	✗	
8 substitutions		✗	✗	✗	

✓ - Left-handed G4 observed by NMR and CD
 — - Left-handed G4 suggested by NMR (CD patterns unclear)
 ✗ - Left-handed G4 not observed

These results would seem to suggest, that the unsubstituted block in *2xBlock 2* with single block substitutions, is able to drive left-handed folding of the block with all four substitutions carried out. With the observations that *Block 1* or its equivalents in *Z-G4*, *Z-G4R*, *Z-G4-Bulge* and *Z-G4-2xBulge* either form a right-handed G-quadruplex or no structure at all, we can assume that the *Block 2* minimal left-handed motif drives left-handed folding of otherwise non-folding sequences or even G-quadruplexes that are otherwise found to be right-handed. This also substantially increases the putative sequence space for left-handed G-quadruplex formation in genomes, as one of the blocks can possess substantial variability.

A substantial melting analysis was performed using UV and/or CD spectroscopies on all of the samples that formed a single structure and are presented in this work. However in all of the cases, even with the slowest achievable temperature ramp rates (0.1 °C/min), the samples exhibited a large hysteresis between the cooling and heating cycles. The highest observed hysteresis, 9 °C, was recorded for *2xBlock 2 T2,16-A* and the lowest for the two sequences with bulges *Z-G4-Bulge* and *Z-G4-2xBulge* at 1.5 °C. Previously reported *Z-G4* structure exhibits the same behavior and has a melting temperature of 45 °C [53]. Due to the transition not occurring at thermal equilibrium and the folding of left-handed sequences being kinetically driven, the melting data cannot be faithfully compared with previously published data [164]. This is also the reason behind samples being left on the bench for an extended period of time before measurement. However for a rough comparison to previously published data, the melting temperatures of left-handed G-quadruplexes ranged from 60 °C (*Z-G4-Bulge*) to as low as 37 °C in the case of *Z-G4-3xBulge*, when considering only the first heating cycle. The majority ranged between 45-55 °C with a 2-5 °C hysteresis. The measured melting temperatures, especially for the samples above 50 °C, are on par with many of the right-handed G-quadruplexes, believed to exist *in vivo* [52, 165, 166].

Chapter 4: Extremely slow-exchange ion in the left-handed Z-G4

The work presented in this chapter is currently submitted to *Chemical Science* for review:

Bakalar, B., Winnerdy, F. R., Heddi, B., Marchand, A., Rosu, F., Gabelica, V. and Phan, A. T. (2019), Unprecedented hour-long residence time of a cation in the left-handed G-quadruplex. *Under review with Chemical Science*

Monovalent and divalent ions interact and help stabilize G-quadruplex structures [25, 109, 110, 115]. They are especially important in the assembly of G-quadruplexes where planar G-tetrads, composed of four Hoogsteen-bonded guanine bases, stack to form the core of the G-quadruplex structure. Ions have been found coordinated between stacked G-tetrads or in the plane of a G-tetrad, depending on their respective ionic radius and charge [30, 50, 109, 167]. Previous studies using the ^{15}N labeled ammonium ion have shown that coordinated ions do not stay static in the coordination site. They dynamically exchange between binding sites as well as the bulk ions present in the solvent [107, 108, 122-126]. Exchange rates observed vary between sub millisecond [123, 124] to over a second [122]. In this study we report the first observation of an extremely slow-exchange ions K^+ and NH_4^+ , with an exchange time constant on the order of an hour, in a novel left-handed G-quadruplex motif.

4.1 Z-G4 preserves left-handed spectral and structural characteristic in ammonium

Z-G4 has four G-tetrads arranged in two stacked blocks (see paragraph 1.9 on page 41 for details). Three potassium ions have been found in between each pair of stacked G-tetrads of the solved structure [53]. The two outer ions are found coordinated in the middle of the two Z-G4 blocks. The inner potassium ion is located between the two 5'-5' stacked blocks.

Figure 4.1: Schematic representation of the Z-G4 structure. *Anti* guanines are colored in cyan and *syn* guanines are colored in magenta. The three potassium ions are shown as purple spheres coordinated between the stacked G-tetrads.

G-quadruplexes have previously been observed to have different structure or not fold at all, depending on the cation present in solution [52, 106, 111-114]. Besides folding Z-G4 in the potassium buffer, we have tested folding Z-G4 in solutions with other cations to see which cations are able to either change or preserve the unique left-handed structure and its CD and NMR spectral features (**Figure 4.2**). Besides K^+ only NH_4^+ was able to produce characteristic imino protons which indicate the formation of a G-quadruplex structure (**Figure 4.2, B**).

Figure 4.2: CD (A) and NMR (B) spectra of Z-G4 folded in buffers with different cations. NH_4^+ is shown in blue, K^+ in red, Cs^+ in orange, Na^+ in purple and Li^+ in green.

The CD spectrum of Z-G4 folded in either K^+ only NH_4^+ (**Figure 4.3, A**) has a dip at 270 nm and a peak at 245 nm, consistent with previous observations for left-handed G-quadruplexes. The NMR spectrum of Z-G4 folded in either K^+ or NH_4^+ (**Figure 4.3, B**) further shows that a single

G-quadruplex structure is formed, with a similar imino peak distribution indicating structural similarity irrespective of the ion used to fold the *Z-G4* sequence.

Figure 4.3: CD (A) and NMR (B) spectra of *Z-G4* folded in either K^+ (red) or NH_4^+ (blue) buffer.

We can also observe amino protons at 9 ppm (**Figure 4.3, B**), another characteristic feature of the *Z-G4* structure, that arises from the additional hydrogen bond of the amino proton with the O4' oxygen of a neighboring thymine residue [53].

Figure 4.4: Imino and amino 1H NMR spectra of *Z-G4* before (A) and two weeks after (B) D_2O exchange experiment. Only the outer G-tetrad imino protons exchange after two weeks, identifying the protected inner G-tetrad imino protons.

Since *Z-G4* folded in K^+ has two inner G-tetrads and two outer G-tetrads, we performed a D_2O exchange experiment (**Figure 4.4**) with *Z-G4* folded in NH_4^+ buffer. We can observe the outer G-tetrad imino protons disappear immediately as they are exchanged with the D_2O solvent,

while inner G-tetrad imino protons can be observed for more than two weeks. This is consistent with previous observations of *Z-G4* folded in K^+ [53].

We can reasonably conclude that *Z-G4* folded in NH_4^+ also has two blocks with 5'-5' stacking which keeps the two inner, stacked G-tetrad imino protons protected from solvent exchange with D_2O . Since we observe the inner imino protons even after weeks in D_2O this suggest that the two blocks that compose *Z-G4* very rarely unstack (or open up) to allow for exchange of the inner imino protons with the D_2O solvent.

Figure 4.5: Imino peak assignment of *Z-G4* in NH_4^+ . **(A)** Direct imino assignment using ^{15}N labeled oligonucleotides. The NH_4^+ imino spectrum is shown in blue, while the K^+ spectrum is shown in red for comparison. **(B)** NOESY spectrum with imino-imino region assignments. Seven imino protons with direct assignment are numbered with bold and underlined font, six imino protons with NOESY assignment are numbered with regular font and the remaining three imino protons are numbered in grey.

To further corroborate that the structure of *Z-G4* stays the same, irrespective of ion, we have performed a partial imino proton assignment (**Figure 4.5, A**). Together with a NOESY spectrum (**Figure 4.5, B**), D_2O exchange and CD data, we can confidently conclude that the fold and general structure of *Z-G4* is not influenced by switching the coordinated ions from K^+ to NH_4^+ . While the imino protons do shift slightly (**Figure 4.5, A**), they stay in the same relative position to each other with the same NOESY connectivity (**Figure 4.5, B**).

4.2 Observation of coordinated ammonium ions in Z-G4 using NMR

Folding Z-G4 in $^{15}\text{NH}_4^+$ produces a new set of peaks around 7-6.5 ppm (**Figure 4.6, A**). Performing a ^{15}N filtered (**Figure 4.6, B**) and decoupled (**Figure 4.6, C**) experiment we can observe 2 peaks, a downfield peak at 6.87 ppm and an upfield peak at 6.57 ppm. The ratio between them is roughly 2:1, with the downfield peak being the bigger of the two. The chemical shift between the two peaks is 0.3 ppm, while the splitting due to ^{15}N - ^1H J-coupling is 74 Hz and no overlap is observed even without ^{15}N decoupling (**Figure 4.6, B**).

Figure 4.6: NMR spectra of Z-G4 folded in $^{15}\text{NH}_4^+$. (**A**) ^1H NMR spectrum, (**B**) ^{15}N filtered NMR spectrum and (**C**) ^{15}N filtered NMR spectrum with decoupling.

Adjusting the pH of the ammonium buffer to pH=5 slows down the proton exchange between the ammonium ion and bulk solvent. When the proton exchange is slow enough, the bulk ammonium peak becomes visible in NMR (**Figure 4.7, B**). We can still observe the ammonium ions coordinated within the Z-G4 structure (red asterisk).

Figure 4.7: Decoupled ^{15}N filtered NMR spectra of Z-G4 folded in $^{15}\text{NH}_4^+$ buffer with either (A) pH=5 or (B) pH=6.9. Peaks of coordinated ammonium ions are labeled with red asterisk.

While all previous observations of ammonium peaks coordinated within the G-quadruplex structure have been downfield of the bulk ammonium peak [106-108, 122-126], we observe them upfield of the bulk ammonium peak (**Figure 4.7, A**). This suggests that ammonium protons receive more electron shielding from the B_0 magnetic field in the Z-G4 structure compared to bulk ammonium protons in the H_2O solvent.

Observing the coordinated ammonium ions with a roughly 2:1 peak ratio (**Figure 4.7, A**) correlates with our expectations for the Z-G4 structure as it has 3 potential binding sites between the four stacked G-tetrads.

We have correlated the position of the ammonium ions within the G-quadruplex structure with their corresponding proton peaks in ^1H NMR by using 2D NOESY and ^{15}N HSQC experiments (**Figure 4.8**). NOESY gives off-diagonal cross-peaks for protons that are close in space allowing us to determine which imino protons are close to the ammonium ion. ^{15}N HSQC correlates nuclei of different types which are separated by one bond, in our case the ^{15}N and ^1H nuclei of the ammonium ion.

Figure 4.8: NOESY and ^{15}N HSQC experiments correlating the position of ^1H NMR ammonium ion peaks with the position of the ammonium ions within the Z-G4 structure at 283 K. Outer and inner imino proton peak positions are labeled in the NOESY spectrum as well as the position of inner and outer ammonium ion. Outer ions are shown in purple, while the inner ion is shown in cyan.

The upfield ammonium ion peak at 6.57 ppm (**Figure 4.8**, light blue line) has NOESY cross-peaks only with the inner imino protons and the NOESY cross-peaks correlate with a single ^{15}N HSQC cross-peak at the same ^1H ppm. It is reasonable to conclude that this ammonium ion occupies the inner binding site, between the two blocks of Z-G4 (**Figure 4.8**, light blue sphere).

For the downfield ammonium ion peak at 6.87 ppm (**Figure 4.8**, purple line) we can observe NOESY cross-peaks with imino protons from both outer and inner G-tetrads. The NOESY cross-peaks correlate with two slightly separated ^{15}N HSQC cross-peaks (**Figure 4.8**, purple line). Based on the imino proton assignment we can conclude the position of both outer ions within the Z-G4 structure and they are appropriately labeled in the schematic (**Figure 4.8**, purple spheres).

Figure 4.9: NzEx-HSQC 2D [107] spectrum recorded with a 100 ms mixing time at pH = 5.1 and 293 K. Bulk, outer and inner $^{15}\text{NH}_4^+$ ion peaks are appropriately annotated. The cross-peaks between the outer and bulk ions are shown, with a dashed red line connecting them.

A sample of Z-G4 folded in 100 mM $^{15}\text{NH}_4^+$ pH = 5.1 was also examined using the NzEx-HSQC 2D pulse program [107] with a 100 ms mixing time at 293 K. The additional mixing time allows for $^{15}\text{NH}_4^+$ ions to exchange between the bulk and binding sites which produces cross-peaks. In **Figure 4.9** we can observe that only the outer ions produces cross-peaks with the bulk $^{15}\text{NH}_4^+$. The inner ion does not produce any cross-peaks, with either the outer ions or the bulk $^{15}\text{NH}_4^+$, suggesting its exchange time might be outside the scope of measurement. Further modifications of mixing time length and temperature did not produce results to the contrary.

4.3 Tracking of ^{15}N ammonium ion exchange using ^1H NMR

Ions coordinated within right-handed G-quadruplexes have been shown to be very dynamic. Exchanges of ions between binding sites as well as between a binding site and the bulk solvent

have been observed [106-108, 122-126]. We tested the behavioral dynamics of ions coordinated in *Z-G4*, a left-handed G-quadruplex.

A sample of *Z-G4* was folded in 100 mM $^{15}\text{NH}_4^+$ with standard 10% D_2O in solution. It was quickly flash frozen in liquid nitrogen and lyophilized. Just before the start of the experiment, the sample was re-suspended in 100% D_2O . The sample was immediately tracked with NMR where we recorded ^{15}N filtered NMR spectra over 3 hours.

Figure 4.10: Tracking of ammonium proton peak disappearance with ^{15}N filtered NMR. **(A)** Ammonium ^1H NMR peak region. Five spectra are shown representing different time points during the experiment. The red arrow shows the disappearance of the inner ammonium ion peak with time as the hydrogen is exchanged with deuterium. **(B)** Graph of inner ammonium ion area versus time shows a residence time of $\tau = 79 \pm 3 \text{ min}$.

In this experiment, we observed the outer ammonium ion proton peaks disappear immediately, while the inner ammonium ion peak remained observable for a prolonged period of time (**Figure 4.10, A**). Tracking the inner ammonium ion with time (**Figure 4.10, B**), we detected a decrease in the area of the peak which closely fit an exponential decay curve with a time constant of $\tau = 79 \pm 3 \text{ min}$. This exchange time is three orders of magnitude longer than previous longest residence time reported for ions coordinated within G-quadruplex structures [122].

Since hydrogen in $^{15}\text{NH}_4^+$ might be replaced with deuterium in a non-trivial way we checked the behavior of $^{15}\text{NH}_4^+$ displacement versus K^+ . A sample of *Z-G4* was folded in 100 mM $^{15}\text{NH}_4^+$ and subsequently dialyzed against H_2O for 10 minutes in order to remove most of the $^{15}\text{NH}_4^+$ ions in solution. It was quickly flash frozen in liquid nitrogen and lyophilized. Just before the

start of the experiment, the sample was re-suspended in a buffer containing 100 mM K⁺ ions. The sample was immediately tracked with NMR where we interchangeably recorded both ¹H and ¹⁵N filtered NMR spectra over 3 hours.

Figure 4.11: Tracking of ammonium proton peak disappearance with ¹⁵N filtered NMR. (A) Ammonium ¹H NMR peak region. Five spectra are shown representing different time points during the experiment. The red arrow shows the disappearance of the inner ammonium ion peak with time and is plotted on the graph with $\tau = 55 \pm 1 \text{ min}$ (B).

As seen from **Figure 4.11, A**, the peak of the inner ammonium ion at 6.48 ppm is shifted slightly upfield. The shift of the inner ¹⁵NH₄⁺ ion peak from 6.57 to 6.48 ppm is likely to be due to both outer sites being occupied by the K⁺ ion (**Figure 4.13**). The graph in **Figure 4.11, B**, tracks the area of the inner ammonium ion peak over the course of the experiment and shows that the ion does indeed exchange but does so in a slow-exchange manner. The residence time observed for the inner ammonium ion is $\tau = 55 \pm 1 \text{ min}$.

This slow-exchange behavior is also mirrored by observing the imino protons (**Figure 4.12, A**), where we can see them starting ($t = 3 \text{ min}$) in the same positions as observed for Z-G4 folded in NH₄⁺, while the final spectrum looks like the spectrum of Z-G4 folded in K⁺.

Figure 4.12: Tracking of imino proton peak appearance and disappearance with ^1H NMR. (A) Inner G-tetrad ^1H NMR imino proton peak region. Three spectra are shown representing different time points during the experiment. The red arrow shows the appearance of the inner G-tetrad imino proton peak with time while the blue arrow shows the disappearance of the inner G-tetrad imino proton peak with time and both are plotted on the graph (B).

Tracking the area of two isolated peaks over the course of the experiment is shown in the graph of **Figure 4.12, B**. The red curve represents an appearing imino peak with $\tau = 54.2 \pm 0.4$ min, while the blue curve shows a disappearing imino peak with $\tau = 53.6 \pm 0.4$ min. The residence times for both match the rate observed for the disappearance of the inner ammonium peak (**Figure 4.11, B**) within experimental error. This represents a first experimental observation of a long-lived ion in a biologically relevant structure (to our knowledge).

Figure 4.13: Comparison of inner $^{15}\text{NH}_4^+$ ion peak position in decoupled ^{15}N filtered NMR spectra, depending on the concentration of additionally added K^+ ions. The spectrum from Figure 4.14 is shown in red, vertical dashed lines represent different species, with the coordinated ions listed at the top.

Since we wanted to see if the unprecedented slow-exchange is specific to the NH_4^+ ion, or if it is a property of the binding site and the Z-G4 structure itself, we conducted the experiment in reverse. We prepared a sample of Z-G4 folded in K^+ and followed the same procedure as in the previous experiment, before re-suspending it in 100 mM $^{15}\text{NH}_4^+$. This means that there are K^+ ions coordinated within Z-G4 at the start, with an excess of $^{15}\text{NH}_4^+$ in solution. We proceeded to track the sample by interchangeably recording both ^1H and ^{15}N filtered NMR spectra over 3 hours.

Figure 4.14, A, shows the gradual appearance of the $^{15}\text{NH}_4^+$ inner ion peaks. There are two peaks slightly lower than the previously observed 6.57 ppm. The peak at 6.52 ppm likely corresponding to a Z-G4 structure having one $^{15}\text{NH}_4^+$ and one K^+ outer ion, while the peak at 6.48 ppm likely has two K^+ outer ions (**Figure 4.13**). This is due to higher binding affinity of K^+ , making it very difficult to prepare a sample where excess K^+ would be removed while Z-G4 would still stay folded. We have observed a similar shift of the inner $^{15}\text{NH}_4^+$ ion peak in the previous experiment as well.

By tracking the increase in area of both inner ion peaks over time (**Figure 4.14, B**) we can observe a residence time of $\tau = 49 \pm 2$ min and $\tau = 44 \pm 2$ min, respectively. This is very similar to the residence times observed in two previous experiments, $^{15}\text{NH}_4^+ \rightarrow ^{15}\text{ND}_4^+$ and $^{15}\text{NH}_4^+ \rightarrow \text{K}^+$. The discrepancy between the three experiments is likely due to K^+ ions increasing the exchange rate of the $^{15}\text{NH}_4^+$ ion, an effect which was previously observed and reported [168]. It is reasonable to conclude that the slow exchange rate is a characteristic of the inner binding site and Z-G4 itself. Since both K^+ and $^{15}\text{NH}_4^+$ give the same slow-exchange rate results, it is likely not due to ion characteristics such as charge, ionic radius or the ability to form hydrogen bonds.

Figure 4.14: Tracking of ammonium proton peak appearance with ^{15}N filtered NMR. (A) Ammonium ^1H NMR peak region. Five spectra are shown representing different time points during the experiment. The blue and red arrows show the appearance of the inner $^{15}\text{NH}_4^+$ ion peaks with time and is plotted on the graph (B) with residence times $\tau = 49 \pm 2$ min and $\tau = 44 \pm 2$ min, respectively.

The heterogeneity of $\text{K}^+ \rightarrow ^{15}\text{NH}_4^+$ is also reflected in the imino ^1H NMR spectra (**Figure 4.15**) and complicates data extraction as the spectrum does not have clear isolated peaks appearing/disappearing. The binding affinity for K^+ ions is much higher than the affinity for NH_4^+ ions [25, 30]. This is also reflected in the $t = 210$ minute spectrum, as we can see it much more closely resembles the spectrum of Z-G4 in K^+ with many additional broad peaks. This is in line with ^{15}N filtered NMR spectral observations (**Figure 4.14**) that there are species with different bound ions present in solution. We were not able to remove enough of K^+ ions in order to obtain an imino spectrum of sufficient quality for information extraction, while keeping the Z-G4 structure folded.

Figure 4.15: Tracking of imino proton peaks appearance and disappearance with ^1H NMR. The top and bottom spectra are of Z-G4 in either K^+ or $^{15}\text{NH}_4^+$ ion. The middle two spectra are from the start ($t = 3$ min) and end ($t = 210$ min) of the $\text{K}^+ \rightarrow ^{15}\text{NH}_4^+$ experiment.

4.4 Observation and tracking of ammonium ions in the channel of Z-G4 using mass spectrometry

Using mass spectrometry (MS) to track ion exchange is possible due to NH_4^+ and K^+ having different mass, 18 and 39 g/mol, respectively and due to advancements which allow for non-destructive sample ionization. A folded G-quadruplex molecule with coordinated K^+ or NH_4^+ ions will be detected at a different m/z ratio, depending on the number of each ion coordinated within the G-quadruplex.

A sample of Z-G4 was prepared in a K^+ buffer and subsequently diluted with the NH_4^+ buffer just prior to injection to native negative-ion ESI-MS. The sample was tracked with time and the subsequent data was cleaned of non-specific adducts and plotted in **Figure 4.16**. We can observe Z-G4 folded in K^+ (**Figure 4.16, A**) has one major peak, a 3 K^+ species, and a minor peak of 0 K^+ species which is likely unfolded. Diluting the sample with the NH_4^+ buffer (**Figure 4.16, B**), we can immediately notice that the 3 K^+ species disappears. The most dominant species at $t = 2$ min, is the 2 NH_4^+ + 1 K^+ species (blue box), indicating the fast exchange of the two outer ions. Through subsequent spectra we can see this species slowly disappear and the slow emergence of the 3 NH_4^+ species (red box).

Figure 4.16: Native negative-ion electrospray ionization-mass spectrometry data for Z-G4 folded in K^+ (A), subsequently diluted with NH_4^+ buffer and tracked with time (B). The $3 NH_4^+$ species is encircled with a red box while the $2 NH_4^+ + 1 K^+$ species with a blue box. Time of spectrum collection is shown on the left. The four ion species are marked; $3 NH_4^+ + 1 K^+$ species is marked in green while the $2 NH_4^+ + 2 K^+$ is marked in purple.

The slow emergence of the $3 NH_4^+$ species is in agreement with our NMR observations (paragraph 4.3, page 106) where we also observed the fast exchange of the outer ions followed by a very slow exchange of the inner ion.

Even after a prolonged amount of time, significant populations of four ion species can be observed (Figure 4.16, B), namely the $3 NH_4^+ + 1 K^+$ (green) and $2 NH_4^+ + 2 K^+$ (purple) species. This is interesting as four ion species are not observed for Z-G4 folded in potassium sample (Figure 4.16, A). This might indicate that significant population exist where the ammonium

ion might be interacting with *Z-G4* in a way that is distinct from normal coordination of cations between G-tetrads.

As was discussed in paragraph 1.7 on page 37, the reverse experiment $\text{NH}_4^+ \rightarrow \text{K}^+$ is not possible due to the K^+ ion not being volatile and the excess of K^+ needed to perform the experiment.

4.5 Discussion and conclusions

Folding Z-G4 in buffers containing cations other than K^+ (**Figure 4.2**, page 100) we can notice that only NH_4^+ was able to preserve both CD and NMR spectral characteristics observed for left-handed G-quadruplexes (**Figure 4.3**, page 101). D_2O exchange experiment (**Figure 4.4**, page 101) showed further that the structure of Z-G4 in NH_4^+ is very similar to the solved structure of Z-G4 in K^+ with two outer G-tetrads and two inner G-tetrads and 5'-5' block stacking. Folding Z-G4 in $^{15}NH_4^+$ we observed two $^{15}NH_4^+$ peaks, at 6.87 and 6.57 ppm, with a 2:1 ratio (**Figure 4.6**, page 103). We were able to correlate the position of the ions within Z-G4 with their corresponding $^{15}NH_4^+$ peak (**Figure 4.8**, page 105) and identified two outer ions and an inner ion, between the 5'-5' stacked blocks of Z-G4. The two outer $^{15}NH_4^+$ ions are found at 6.87 ppm, while the inner $^{15}NH_4^+$ ion is found at 6.57 ppm. Using the NzEx-HSQC pulse program that has an additional mixing time allowing for ion movement, we could only observe exchange cross-peaks between the outer $^{15}NH_4^+$ ions and the bulk $^{15}NH_4^+$ ions (**Figure 4.9**, page 106).

The ions in the inner binding site have a very slow-exchange rate with the outer binding sites. We were able to track:

- $^{15}NH_4^+ \rightarrow ^{15}ND_4^+$ (**Figure 4.10**, page 107)
- $^{15}NH_4^+ \rightarrow K^+$ (**Figure 4.11**, page 108 and **Figure 4.12**, page 109)
- and $K^+ \rightarrow ^{15}NH_4^+$ (**Figure 4.14**, page 111)

The exchange rate of the ion in the inner binding site was measured at:

- $^{15}NH_4^+ \rightarrow ^{15}ND_4^+$ $\tau = 79 \pm 3$ min
- $^{15}NH_4^+ \rightarrow K^+$ $\tau = 55 \pm 1$ min and $\tau = 54 \pm 1$ min
- $K^+ \rightarrow ^{15}NH_4^+$ $\tau = 49 \pm 2$ min and $\tau = 44 \pm 2$ min

The similarity of the three exchange rates also suggests that the slow exchange-rate is not due to ion specificities but might be linked with the structure of Z-G4. The discrepancy between the experimentally observed residence times is likely due to K^+ increasing the exchange rate of $^{15}NH_4^+$, as was previously reported [168]. The outer ions were observed to exchange much faster in a similar time-frame as previous studies have reported [107, 108, 124].

These findings were further corroborated using mass spectrometry (**Figure 4.16**, page 113), where a $K^+ \rightarrow NH_4^+$ experiment was performed and we also observe a very fast exchange of two NH_4^+ ions, followed by a much slower exchange of the third NH_4^+ ion.

The very slow-exchange rate of the inner binding site is novel (**Figure 4.17**). No matter which of the two ions, NH_4^+ or K^+ , occupies it initially, the residence time is three orders of magnitude longer than the previous observation for a G-quadruplex [122]. To our knowledge, this is the first documented observation of such a long residence time for an ion within a biological structure.

Figure 4.17: Schematic representation of the different ion exchange rates. In orange is the exchange from $^{15}NH_4^+$ to K^+ and in blue is the exchange from K^+ to $^{15}NH_4^+$.

If the primary way of exchange from the inner binding site was by the 5'-5' stacked blocks opening up, the exchange would likely be much slower than observed, based on the slow disappearance of the imino proton peaks in the D_2O exchange experiment. It is therefore reasonable to conclude the exchange happens by the movement of the inner ion to one of the outer binding sites.

Chapter 5: Summary and future perspectives

G-quadruplexes are DNA and RNA structures formed by short, guanine-rich sequences are very likely to play a crucial part in many cellular processes such as transcription regulation [139, 140] and telomere dynamics [137, 146]. In vitro experiments have shown that G-quadruplex structures possess a surprising amount of heterogeneity arising from differences in the number of guanine G-tetrads, orientation of strand progression, loop length and the way loops connect guanine repeats [32, 36, 50, 169-172].

Until recently, all natural G-quadruplex forming sequences have had a right-handed G-tetrad twist. Our laboratory has previously reported on the first observed case of a natural left-handed G-quadruplex structure [53], *Z-G4*, opening the door for exploring this new and interesting subgroup of G-quadruplexes. We report on a new but related left-handed G-quadruplex structure that represents a minimal left-handed motif. Together with four more left-handed structures we began to establish what structural, CD and NMR features are conserved and which ones differ. These results establish the defining characteristics of left-handed G-quadruplexes while enabling further exploration of left-handed G4 potential significance.

The short 12 nt sequence of the minimal left-handed G4 greatly expands the possible sequence space due to being so short and is found highly enriched in the human genome, owing to its repetitive nature. On the other hand the apparent need of left-handed G-quadruplexes for 5'-5' block stacking, when forming a stable structure, limits the putative sequence space. A substantial thymine T-cap substitution analysis was performed on one of the new left-handed structures to try and expand the putative sequence space of left-handed G-quadruplexes. The results show moderate sequence flexibility of the minimal left-handed motif. Together these results show substantial flexibility for left-handed G4 formation *in vitro*, which opens up the door to further bioinformatic explorations of the human and other genomes in search of putative left-handed G4 sequences and their possible biological significance.

Besides having a left-handed strand progression and G-tetrad twist, an interesting aspect of left-handed G-quadruplex structures is the so-called T-cap, where four thymines in the loops

are flipped upward and cover the outer G-tetrad of the G-quadruplex. Thymines of the T-cap are constrained in movement when compared to regular loop nucleobases and could be utilized as potential targeting sites for ligand design. These two features combine to raise some interesting opportunities as most of today's small molecules [173-176] and proteins [136, 177] that bind to G-quadruplexes, recognize the outer G-tetrad and/or the grooves created by the loops. These two features that are similar amongst many G-quadruplex structures, make it very difficult to find ligands that bind specifically to a subgroup or a specific G-quadruplex. In this respect, left-handed G-quadruplexes might have a distinct advantage as their unique features have already been exploited for targeted ligand design [152]. Future work on potential specific peptide/protein binders will need to be carried out as this avenue might provide both biological significance as well as potential for specific targeting of the left-handed G-quadruplex motif. Together with the bioinformatic analysis of putative left-handed motif localization within genomes, these two research avenues could present potential for construction of novel DNA switches for gene expression regulation.

Cations have a big effect on folded G-quadruplex structures, as they are coordinate between tetrads and interact with guanine carbonyl oxygens [109, 110]. The ionic radius, charge and potential for hydrogen bonding in the case of NH_4^+ , has a direct effect on both the type of folding of a G-quadruplex as well as on the G-quadruplex stability [30, 115, 116].

Besides folding the left-handed Z-G4 in solution with K^+ ions, it was successfully folded into a stable left-handed G-quadruplex in the presence of NH_4^+ ions. Z-G4 folded in NH_4^+ preserves all the CD and NMR spectral characteristics as were observed for the K^+ structure. We performed ion exchange experiments, $^{15}\text{NH}_4^+ \rightarrow ^{15}\text{ND}_4^+$, $^{15}\text{NH}_4^+ \rightarrow \text{K}^+$ and $\text{K}^+ \rightarrow ^{15}\text{NH}_4^+$, to characterize the exchange rate of coordinate cations using NMR and ESI-MS. Our results point to an inner binding site, between both 5'-5' stacked blocks of Z-G4, with unique properties. Irrespective of the type of coordinated ion, the exchange rate of the inner ion is extremely slow, as the residence time observed was longer than 44 minutes. This is three orders of magnitude greater than the previous longest observation for an ion in a G-quadruplex [122] and to our knowledge, the longest observation for any biological structure. We have observed and measured this slow exchange rate with NMR as well as ESI-MS and both techniques show very similar results.

The results of NzEx-HSQC experiment [107] suggest that the two outer binding sites are similar in regard to their $^{15}\text{NH}_4^+$ ion exchange rates to previous reports as we can observe cross-peaks between the outer $^{15}\text{NH}_4^+$ ion and the bulk $^{15}\text{NH}_4^+$ ion [122-124]. Exploring the exchange kinetics of the outer ions further will be necessary in order to determine the unique structural feature(s) that might contribute to the slow-exchange rate of the inner ion. A complimentary approach is to explore the ion exchange behavior of other solved left-handed G-quadruplex structures which all have their own unique features, for example, the differences between their respective T-cap features. This will help narrow down the possible structural features that might be causing this novel behavior of the inner coordinated ion. Similarly, further exploration of right-handed G-quadruplex structures will need to be performed, to see if such long residence times of ions might have previously been missed.

References

1. Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature* 171, 737-738.
2. Franklin, R. E., and Gosling, R. G. (1953) The structure of sodium thymonucleate fibres. I. The influence of water content, *Acta Crystallographica* 6, 673-677.
3. Wilkins, M., Gosling, R., and Seeds, W. (1951) Physical studies of nucleic acid: Nucleic acid: An extensible molecule?, *Nature* 167, 759.
4. Hershey, A. D., and Chase, M. (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage, *The Journal of general physiology* 36, 39-56.
5. Fletcher, T. M., Sun, D., Salazar, M., and Hurley, L. H. (1998) Effect of DNA Secondary Structure on Human Telomerase Activity, *Biochemistry* 37, 5536-5541.
6. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., Van Boom, J. H., van der Marel, G., and Rich, A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution, *Nature* 282, 680.
7. Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) Helix formation by guanylic acid, *Proceedings of the National Academy of Sciences of the United States of America* 48, 2013-2018.
8. Rich, A., and Zhang, S. (2003) Z-DNA: the long road to biological function, *Nature Reviews. Genetics* 4, 566-572.
9. Guéron, M., and Leroy, J.-L. (2000) The i-motif in nucleic acids, *Current Opinion in Structural Biology* 10, 326-331.
10. Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, *Nature* 334, 364.
11. Feigon, J., Wang, A., Van Der Marel, G. A., Van Boom, J. H., and Rich, A. (1985) Z-DNA forms without an alternating purine-pyrimidine sequence in solution, *Science* 230, 82-84.
12. Drew, H., Takano, T., Tanaka, S., Itakura, K., and Dickerson, R. E. (1980) High-salt d(CpGpCpG), a left-handed Z[prime] DNA double helix, *Nature* 286, 567-573.
13. Thamann, T. J., Lord, R. C., Wang, A. H., and Rich, A. (1981) The high salt form of poly (dG-dC) · poly (dG-dC) is left-handed Z-DNA: Raman spectra of crystals and solutions, *Nucleic Acids Research* 9, 5443-5458.
14. Nordheim, A., and Rich, A. (1983) Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences, *Nature* 303, 674-679.
15. Kouzine, F., Liu, J., Sanford, S., Chung, H.-J., and Levens, D. (2004) The dynamic response of upstream DNA to transcription-generated torsional stress, *Nature Structural and Molecular Biology* 11, 1092.
16. Bryant, Z., Stone, M. D., Gore, J., Smith, S. B., Cozzarelli, N. R., and Bustamante, C. (2003) Structural transitions and elasticity from torque measurements on DNA, *Nature* 424, 338.
17. Nordheim, A., and Rich, A. (1983) The sequence (dC-dA) _n X (dG-dT) _n forms left-handed Z-DNA in negatively supercoiled plasmids, *Proceedings of the National Academy of Sciences of the United States of America* 80, 1821-1825.
18. Allemand, J. F., Bensimon, D., Lavery, R., and Croquette, V. (1998) Stretched and overwound DNA forms a Pauling-like structure with exposed bases, *Proceedings of the National Academy of Sciences of the United States of America* 95, 14152-14157.

19. Kohwi, Y., and Kohwi-Shigematsu, T. (1988) Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA, *Proceedings of the National Academy of Sciences of the United States of America* 85, 3781-3785.
20. Weintraub, H. (1983) A dominant role for DNA secondary structure in forming hypersensitive structures in chromatin, *Cell* 32, 1191-1203.
21. Hopkins, R. C. (1981) Deoxyribonucleic acid structure: a new model, *Science* 211, 289-291.
22. Mitchell, P. J., and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins, *Science* 245, 371-378.
23. Patel, D. J., Phan, A. T. n., and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics, *Nucleic Acids Research* 35, 7429-7455.
24. Paeschke, K., Bochman, M. L., Garcia, P. D., Cejka, P., Friedman, K. L., Kowalczykowski, S. C., and Zakian, V. A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs, *Nature* 497, 458.
25. Guschlbauer, W., Chantot, J.-F., and Thiele, D. (1990) Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomere DNA, *Journal of Biomolecular Structure and Dynamics* 8, 491-511.
26. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure, *Nucleic Acids Research* 34, 5402-5415.
27. Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A., and Spada, G. P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures, *Organic & Biomolecular Chemistry* 8, 2683-2692.
28. Lech, C. J., Heddi, B., and Phan, A. T. (2013) Guanine base stacking in G-quadruplex nucleic acids, *Nucleic Acids Research* 41, 2034-2046.
29. Reshetnikov, R. V., Kopylov, A. M., and Golovin, A. V. (2010) Classification of g-quadruplex DNA on the basis of the quadruplex twist angle and planarity of g-quartets, *Acta Naturae* 2, 72-81.
30. Davis, J. T. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry, *Angewandte Chemie* 43, 668-698.
31. Phan, A. T., Kuryavyi, V., and Patel, D. J. (2006) DNA architecture: from G to Z, *Current Opinion in Structural Biology* 16, 288-298.
32. Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L., and Patel, D. J. (2006) Structure of the human telomere in K⁺ solution: an intramolecular (3+ 1) G-quadruplex scaffold, *Journal of the American Chemical Society* 128, 9963-9970.
33. Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes, *Current Opinion in Structural Biology* 19, 239-250.
34. Phan, A. T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences, *FEBS Journal* 277, 1107-1117.
35. Do, N. Q., Mukundan, V. T., and Phan, A. T. (2011) HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges, *Nucleic Acids Research* 39, 8984-8991.
36. Mukundan, V. T., and Phan, A. T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences, *Journal of the American Chemical Society* 135, 5017-5028.

37. Martadinata, H., and Phan, A. T. (2014) Formation of a stacked dimeric G-quadruplex containing bulges by the 5'-terminal region of human telomerase RNA (hTERC), *Biochemistry* 53, 1595-1600.
38. Huang, H., Suslov, N. B., Li, N.-S., Shelke, S. A., Evans, M. E., Koldobskaya, Y., Rice, P. A., and Piccirilli, J. A. (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore, *Nature Chemical Biology* 10, 686.
39. Teng, N. N., Itzkowitz, M. S., and Tinoco Jr, I. (1971) Calculation of the rotational strengths of mononucleosides, *Journal of the American Chemical Society* 93, 6257-6264.
40. Zaloudek, F., Novros, J. S., and Clark, L. B. (1985) The electronic spectrum of cytosine, *Journal of the American Chemical Society* 107, 7344-7351.
41. Clark, L. B. (1990) Electronic spectrum of the adenine chromophore, *Journal of Physical Chemistry* 94, 2873-2879.
42. Clark, L. B. (1994) Electronic spectra of crystalline guanosine: transition moment directions of the guanine chromophore, *Journal of the American Chemical Society* 116, 5265-5270.
43. Lorentzon, J., Fuelscher, M. P., and Roos, B. O. (1995) Theoretical study of the electronic spectra of uracil and thymine, *Journal of the American Chemical Society* 117, 9265-9273.
44. Berova, N., Bari, L. D., and Pescitelli, G. (2007) Application of electronic circular dichroism in configurational and conformational analysis of organic compounds, *Chemical Society Reviews* 36, 914-931.
45. Harada, N., and Nakanishi, K. (1983) *Circular dichroic spectroscopy: exciton coupling in organic stereochemistry*, Univ Science Books.
46. Gottarelli, G., Palmieri, P., and Spada, G. (1990) On the exciton optical-activity of the 4-stranded helix of poly (G), *Gazzetta Chimica Italiana* 120, 101-107.
47. Gottarelli, G., Masiero, S., and Spada, G. (1998) The use of CD spectroscopy for the study of the self-assembly of guanine derivatives, *Enantiomer* 3, 429-+.
48. Randazzo, A., Spada, G. P., and da Silva, M. W. (2012) Circular dichroism of quadruplex structures, In *Quadruplex Nucleic Acids*, pp 67-86, Springer.
49. Parkinson, G. N., Lee, M. P., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature* 417, 876.
50. Lim, K. W., Ng, V. C. M., Martín-Pintado, N., Heddi, B., and Phan, A. T. n. (2013) Structure of the human telomere in Na⁺ solution: an antiparallel (2+ 2) G-quadruplex scaffold reveals additional diversity, *Nucleic Acids Research* 41, 10556-10562.
51. Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A., and Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution, *Proceedings of the National Academy of Sciences of the United States of America* 90, 3745-3749.
52. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution, *Nucleic Acids Research* 34, 2723-2735.
53. Chung, W. J., Heddi, B., Schmitt, E., Lim, K. W., Mechulam, Y., and Phan, A. T. (2015) Structure of a left-handed DNA G-quadruplex, *Proceedings of the National Academy of Sciences of the United States of America* 112, 2729-2733.

54. Adrian, M., Heddi, B., and Phan, A. T. (2012) NMR spectroscopy of G-quadruplexes, *Methods* 57, 11-24.
55. Hoogsteen, K. (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine, *Acta Crystallographica* 16, 907-916.
56. Esposito, V., Martino, L., Citarella, G., Virgilio, A., Mayol, L., Giancola, C., and Galeone, A. (2009) Effects of abasic sites on structural, thermodynamic and kinetic properties of quadruplex structures, *Nucleic Acids Research* 38, 2069-2080.
57. Školáková, P., Bednářová, K., Vorlíčková, M., and Sagi, J. (2010) Quadruplexes of human telomere dG3 (TTAG3) 3 sequences containing guanine abasic sites, *Biochemical and biophysical research communications* 399, 203-208.
58. Heddi, B., Martín-Pintado, N., Kari, Teuku Mahfuzh A., Serimbetov, Z., and Phan, A. T. (2015) G-quadruplexes with (4n - 1) guanines in the G-tetrad core: formation of a G-triad-water complex and implication for small-molecule binding, *Nucleic Acids Research* 44, 910-916.
59. Adrian, M., Winnerdy, F. R., Heddi, B., and Phan, A. T. (2017) Rotation of guanine amino groups in G-quadruplexes: a probe for local structure and ligand binding, *Biophysical Journal* 113, 775-784.
60. Rance, M., Sørensen, O., Bodenhausen, G., Wagner, G., Ernst, R., and Wüthrich, K. (1983) Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering, *Biochemical and biophysical research communications* 117, 479-485.
61. Phan, A. T., and Patel, D. J. (2002) A site-specific low-enrichment 15N, 13C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids, *Journal of the American Chemical Society* 124, 1160-1161.
62. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *Journal of magnetic resonance* 160, 65-73.
63. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham III, T. E., Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers, *Biophysical Journal* 92, 3817-3829.
64. Chayen, N. E., and Saridakis, E. (2008) Protein crystallization: from purified protein to diffraction-quality crystal, *Nature methods* 5, 147.
65. Ke, A., and Doudna, J. A. (2004) Crystallization of RNA and RNA-protein complexes, *Methods* 34, 408-414.
66. Hollis, T. (2007) Crystallization of protein-DNA complexes, In *Macromolecular Crystallography Protocols*, pp 225-237, Springer.
67. Luft, J. R., Newman, J., and Snell, E. H. (2014) Crystallization screening: the influence of history on current practice, *Acta crystallographica. Section F, Structural biology communications* 70, 835-853.
68. Ducruix, A., and Giegé, R. (1999) *Crystallization of nucleic acids and proteins: a practical approach*, Practical approach series.
69. Johnson, I., Bergamaschi, A., Billich, H., Cartier, S., Dinapoli, R., Greiffenberg, D., Guizar-Sicairos, M., Henrich, B., Jungmann, J., and Mezza, D. (2014) Eiger: a single-photon counting x-ray detector, *Journal of Instrumentation* 9, C05032.
70. Sauter, N. K., Grosse-Kunstleve, R. W., and Adams, P. D. (2004) Robust indexing for automatic data collection, *Journal of Applied Crystallography* 37, 399-409.

71. Kabsch, W. (2010) Xds, *Acta Crystallographica Section D: Biological Crystallography* 66, 125-132.
72. Patterson, A. L. (1935) A direct method for the determination of the components of interatomic distances in crystals, *Zeitschrift für Kristallographie-Crystalline Materials* 90, 517-542.
73. Brünger, A. T. (1990) Extension of molecular replacement: a new search strategy based on Patterson correlation refinement, *Acta Crystallographica Section A* 46, 46-57.
74. Rossmann, M. G. (1990) The molecular replacement method, *Acta Crystallographica Section A* 46, 73-82.
75. Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood, *Acta Crystallographica Section D: Biological Crystallography* 57, 1373-1382.
76. DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., and Axelrod, H. L. (2011) Improved molecular replacement by density-and energy-guided protein structure optimization, *Nature* 473, 540.
77. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography* 40, 658-674.
78. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., and Grosse-Kunstleve, R. W. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallographica Section D: Biological Crystallography* 66, 213-221.
79. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates, *Proteins: Structure, Function, and Bioinformatics* 12, 345-364.
80. Harvey, D. J., Bateman, R. H., and Green, M. R. (1997) High-energy collision-induced fragmentation of complex oligosaccharides ionized by matrix-assisted laser desorption/ionization mass spectrometry, *Journal of mass spectrometry* 32, 167-187.
81. Dančík, V., Addona, T. A., Clauser, K. R., Vath, J. E., and Pevzner, P. A. (1999) De novo peptide sequencing via tandem mass spectrometry, *Journal of computational biology* 6, 327-342.
82. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics, *Nature* 422, 198.
83. Laiko, V. V., Baldwin, M. A., and Burlingame, A. L. (2000) Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry, *Analytical chemistry* 72, 652-657.
84. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science* 246, 64-71.
85. Goodlett, D. R., Camp, D. G., Hardin, C. C., Corregan, M., and Smith, R. D. (1993) Direct observation of a DNA quadruplex by electrospray ionization mass spectrometry, *Biological Mass Spectrometry* 22, 181-183.
86. Morsa, D., Gabelica, V., and De Pauw, E. (2011) Effective temperature of ions in traveling wave ion mobility spectrometry, *Analytical chemistry* 83, 5775-5782.

87. Vairamani, M., and Gross, M. L. (2003) G-Quadruplex Formation of Thrombin-Binding Aptamer Detected by Electrospray Ionization Mass Spectrometry, *Journal of the American Chemical Society* 125, 42-43.
88. Balthasart, F., Plavec, J., and Gabelica, V. (2013) Ammonium Ion Binding to DNA G-Quadruplexes: Do Electrospray Mass Spectra Faithfully Reflect the Solution-Phase Species?, *Journal of The American Society for Mass Spectrometry* 24, 1-8.
89. Marchand, A., and Gabelica, V. (2014) Native Electrospray Mass Spectrometry of DNA G-Quadruplexes in Potassium Solution, *Journal of The American Society for Mass Spectrometry* 25, 1146-1154.
90. Zhou, J., and Yuan, G. (2007) Specific recognition of human telomeric G-quadruplex DNA with small molecules and the conformational analysis by ESI mass spectrometry and circular dichroism spectropolarimetry, *Chemistry* 13, 5018-5023.
91. Gabelica, V., Baker, E. S., Teulade-Fichou, M. P., De Pauw, E., and Bowers, M. T. (2007) Stabilization and structure of telomeric and c-myc region intramolecular G-quadruplexes: the role of central cations and small planar ligands, *J Am Chem Soc* 129, 895-904.
92. Hofstadler, S. A., and Griffey, R. H. (2001) Analysis of Noncovalent Complexes of DNA and RNA by Mass Spectrometry, *Chemical Reviews* 101, 377-390.
93. Li, H., Liu, Y., Lin, S., and Yuan, G. (2009) Spectroscopy probing of the formation, recognition, and conversion of a G-quadruplex in the promoter region of the bcl-2 oncogene, *Chemistry* 15, 2445-2452.
94. Rosu, F., Gabelica, V., Houssier, C., Colson, P., and De Pauw, E. (2002) Triplex and quadruplex DNA structures studied by electrospray mass spectrometry, *Rapid Communications in Mass Spectrometry* 16, 1729-1736.
95. McCullough, B. J., Kalapothakis, J., Eastwood, H., Kemper, P., MacMillan, D., Taylor, K., Dorin, J., and Barran, P. E. (2008) Development of an ion mobility quadrupole time of flight mass spectrometer, *Analytical chemistry* 80, 6336-6344.
96. Wolff, M., and Stephens, W. (1953) A pulsed mass spectrometer with time dispersion, *Review of Scientific Instruments* 24, 616-617.
97. Wiley, W., and McLaren, I. H. (1955) Time-of-flight mass spectrometer with improved resolution, *Review of Scientific Instruments* 26, 1150-1157.
98. Scarff, C. A., Thalassinou, K., Hilton, G. R., and Scrivens, J. H. (2008) Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements, *Rapid Communications in Mass Spectrometry* 22, 3297-3304.
99. Tang, K., Shvartsburg, A. A., Lee, H.-N., Prior, D. C., Buschbach, M. A., Li, F., Tolmachev, A. V., Anderson, G. A., and Smith, R. D. (2005) High-sensitivity ion mobility spectrometry/mass spectrometry using electrodynamic ion funnel interfaces, *Analytical chemistry* 77, 3330-3339.
100. Deng, H., and Braunlin, W. H. (1996) Kinetics of sodium ion binding to DNA quadruplexes, *Journal of Molecular Biology* 255, 476-483.
101. Xu, Q., Deng, H., and Braunlin, W. H. (1993) Selective localization and rotational immobilization of univalent cations on quadruplex DNA, *Biochemistry* 32, 13130-13137.
102. Braunlin, W., Anderson, C. F., and Record Jr, M. T. (1986) ²³Na-nmr investigations of counterion exchange reactions of helical DNA, *Biopolymers: Original Research on Biomolecules* 25, 205-214.

103. Stein, V., Anderson, C., and Record, M. (1993) Investigating differences in the interaction of cations with a 160 basepair calf thymus DNA and a 20 base pair oligonucleotide using Na-23 NMR, In *Biophysical Journal*, pp A267-A267.
104. Wu, G., Wong, A., Gan, Z., and Davis, J. T. (2003) Direct detection of potassium cations bound to G-quadruplex structures by solid-state 39K NMR at 19.6 T, *Journal of the American Chemical Society* 125, 7182-7183.
105. Ida, R., and Wu, G. (2005) Solid-state 87Rb NMR signatures for rubidium cations bound to a G-quadruplex, *Chemical Communications*, 4294-4296.
106. Schultze, P., Hud, N. V., Smith, F. W., and Feigon, J. (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G(4)T(4)G(4)), *Nucleic Acids Research* 27, 3018-3028.
107. Hud, N. V., Schultze, P., Sklenar, V., and Feigon, J. (1999) Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex, *Journal of Molecular Biology* 285, 233-243.
108. Hud, N. V., Schultze, P., and Feigon, J. (1998) Ammonium ion as an NMR probe for monovalent cation coordination sites of DNA quadruplexes, *Journal of the American Chemical Society* 120, 6403-6404.
109. Pinnavaia, T., Marshall, C. L., Mettler, C. M., Fisk, C. L., Miles, H. T., and Becker, E. D. (1978) Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate, *Journal of the American Chemical Society* 100, 3625-3627.
110. Arnott, S., Chandrasekaran, R., and Marttila, C. M. (1974) Structures for polyinosinic acid and polyguanylic acid, *Biochemical Journal* 141, 537-543.
111. Miyoshi, D., Nakao, A., and Sugimoto, N. (2003) Structural transition from antiparallel to parallel G-quadruplex of d (G 4 T 4 G 4) induced by Ca 2+, *Nucleic Acids Research* 31, 1156-1163.
112. Hardin, C. C., Henderson, E., Watson, T., and Prosser, J. K. (1991) Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates, *Biochemistry* 30, 4460-4472.
113. Kankia, B. I., and Marky, L. A. (2001) Folding of the thrombin aptamer into a G-quadruplex with Sr2+: stability, heat, and hydration, *Journal of the American Chemical Society* 123, 10799-10804.
114. Gaynutdinov, T. I., Neumann, R. D., and Panyutin, I. G. (2008) Structural polymorphism of intramolecular quadruplex of human telomeric DNA: effect of cations, quadruplex-binding drugs and flanking sequences, *Nucleic Acids Research* 36, 4079-4087.
115. Chantot, J.-F., and Guschlbauer, W. (1969) Physicochemical properties of nucleosides 3. Gel formation by 8-bromoguanosine, *FEBS Letters* 4, 173-176.
116. Hardin, C. C., Perry, A. G., and White, K. (2000) Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids, *Biopolymers: Original Research on Biomolecules* 56, 147-194.
117. Keniry, M. A. (2000) Quadruplex structures in nucleic acids, *Biopolymers: Original Research on Biomolecules* 56, 123-146.
118. Simonsson, T. (2001) G-quadruplex DNA structures variations on a theme, *Biological Chemistry* 382, 621-628.
119. Gilbert, D. E., and Feigon, J. (1999) Multistranded DNA structures, *Current Opinion in Structural Biology* 9, 305-314.

120. Williamson, J. R. (1994) G-Quartet Structures in Telomeric DNA, *Annual Review of Biophysics and Biomolecular Structure* 23, 703-730.
121. Sen, D., and Gilbert, W. (1992) [10] Guanine quartet structures, In *Methods in Enzymology*, pp 191-199, Elsevier.
122. Šket, P., and Plavec, J. (2007) Not All G-Quadruplexes Exhibit Ion-Channel-like Properties: NMR Study of Ammonium Ion (Non)movement within the d(G3T4G4)₂ Quadruplex, *Journal of the American Chemical Society* 129, 8794-8800.
123. Podbevsek, P., Sket, P., and Plavec, J. (2008) Stacking and not solely topology of T3 loops controls rigidity and ammonium ion movement within d(G4T3G4)₂ G-quadruplex, *Journal of the American Chemical Society* 130, 14287-14293.
124. Podbevšek, P., Hud, N. V., and Plavec, J. (2007) NMR evaluation of ammonium ion movement within a unimolecular G-quadruplex in solution, *Nucleic Acids Research* 35, 2554-2563.
125. Šket, P., Črnugelj, M., and Plavec, J. (2005) Identification of mixed di-cation forms of G-quadruplex in solution, *Nucleic Acids Research* 33, 3691-3697.
126. Sket, P., Kozminski, W., and Plavec, J. (2012) Is There Any Proton Exchange Between Ammonium Ions localized Within the d(G3T4G4)₂ Quadruplex?, *Acta Chimica Slovenica* 59, 473-477.
127. Fukushima, K., and Iwahashi, H. (2000) 1: 1 Complex of guanine quartet with alkali metal cations detected by electrospray ionization mass spectrometry, *Chemical Communications*, 895-896.
128. Manet, I., Francini, L., Masiero, S., Pieraccini, S., Spada, G. P., and Gottarelli, G. (2001) An ESI-MS and NMR study of the self-assembly of guanosine derivatives, *Helvetica Chimica Acta* 84, 2096-2107.
129. Yuan, G., Zhang, Q., Zhou, J., and Li, H. (2011) Mass spectrometry of G-quadruplex DNA: Formation, recognition, property, conversion, and conformation, *Mass Spectrometry Reviews* 30, 1121-1142.
130. Gros, J., Rosu, F., Amrane, S., De Cian, A., Gabelica, V., Lacroix, L., and Mergny, J.-L. (2007) Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes, *Nucleic Acids Research* 35, 3064-3075.
131. Baker, E. S., Bernstein, S. L., Gabelica, V., De Pauw, E., and Bowers, M. T. (2006) G-quadruplexes in telomeric repeats are conserved in a solvent-free environment, *International Journal of Mass Spectrometry* 253, 225-237.
132. Rhodes, D., and Lipps, H. J. (2015) G-quadruplexes and their regulatory roles in biology, *Nucleic Acids Research* 43, 8627-8637.
133. Bedrat, A., Lacroix, L., and Mergny, J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter, *Nucleic Acids Research* 44, 1746-1759.
134. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome, *Nucleic Acids Research* 33, 2908-2916.
135. da Silva, M. W. (2007) NMR methods for studying quadruplex nucleic acids, *Methods* 43, 264-277.
136. Fry, M. (2007) Tetraplex DNA and its interacting proteins, *Frontiers in Bioscience* 12, 4336-4351.
137. Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells, *Nature Chemistry* 5, 182.

138. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei, *Proceedings of the National Academy of Sciences of the United States of America* 98, 8572-8577.
139. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, *Proceedings of the National Academy of Sciences of the United States of America* 99, 11593-11598.
140. Cogoi, S., and Xodo, L. E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription, *Nucleic Acids Research* 34, 2536-2549.
141. Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation, *Nature Chemical Biology* 3, 218.
142. Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., and Kurreck, J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif, *RNA* 14, 1290-1296.
143. Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., and Lemaitre, J.-M. (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs, *Nature Structural and Molecular Biology* 19, 837.
144. Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M. P., Foiani, M., and Nicolas, A. (2011) G-quadruplex-induced instability during leading-strand replication, *The EMBO Journal* 30, 4033-4046.
145. Ribeyre, C., Lopes, J., Boulé, J.-B., Piazza, A., Guédin, A., Zakian, V. A., Mergny, J.-L., and Nicolas, A. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo, *PLoS genetics* 5, e1000475.
146. Rhodes, D., and Giraldo, R. (1995) Telomere structure and function, *Current Opinion in Structural Biology* 5, 311-322.
147. Biffi, G., Tannahill, D., and Balasubramanian, S. (2012) An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2, *Journal of the American Chemical Society* 134, 11974-11976.
148. Bates, P. J., Laber, D. A., Miller, D. M., Thomas, S. D., and Trent, J. O. (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer, *Experimental and Molecular Pathology* 86, 151-164.
149. Perrone, R., Butovskaya, E., Lago, S., Garzino-Demo, A., Pannecouque, C., Palu, G., and Richter, S. N. (2016) The G-quadruplex-forming aptamer AS1411 potently inhibits HIV-1 attachment to the host cell, *International Journal of Antimicrobial Agents* 47, 311-316.
150. Metifiot, M., Amrane, S., Mergny, J. L., and Andreola, M. L. (2015) Anticancer molecule AS1411 exhibits low nanomolar antiviral activity against HIV-1, *Biochimie* 118, 173-175.
151. Ngoc Qyang, D. (2012) New DNA motifs: Structure and potential activity against HIV and cancer, *PhD thesis*.
152. Zhao, A., Zhao, C., Ren, J., and Qu, X. (2016) Enantioselective targeting left-handed Z-G-quadruplex, *Chemical Communications* 52, 1365-1368.

153. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallographica. Section D, Biological Crystallography* 66, 213-221.
154. Kabsch, W. (2010) XDS, *Acta Crystallographica. Section D, Biological Crystallography* 66, 125-132.
155. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography* 40, 658-674.
156. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallographica. Section D-Biological Crystallography* 66, 486-501.
157. Alexander, N., Woetzel, N., and Meiler, J. (2011) bcl::Cluster : A method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System, *IEEE International Conference on Computational Advances in Bio and Medical Sciences 2011*, 13-18.
158. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+, *Nucleic Acids Research* 37, 5917-5929.
159. Kypr, J., Kejnovska, I., Renciuik, D., and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA, *Nucleic Acids Research* 37, 1713-1725.
160. Iwahashi, H., and Kyogoku, Y. (1978) Restricted rotation of the amino group of nucleic acid base derivatives and a study of structures of base pairs by the use of the phenomena, *Nucleic Acids Res* 1, s385-s390.
161. Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., and Sasisekharan, V. (1992) Hairpin and parallel quartet structures for telomeric sequences, *Nucleic Acids Research* 20, 4061-4067.
162. Phan, A. T., Kuryavyi, V., Ma, J. B., Faure, A., Andreola, M. L., and Patel, D. J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase, *Proceedings of the National Academy of Sciences of the United States of America* 102, 634-639.
163. Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA, *Nucleic Acids Research* 33, 2901-2907.
164. Rachwal, P. A., and Fox, K. R. (2007) Quadruplex melting, *Methods* 43, 291-301.
165. Mergny, J.-L., and Lacroix, L. (2003) Analysis of thermal melting curves, *Oligonucleotides* 13, 515-537.
166. Guedin, A., Gros, J., Alberti, P., and Mergny, J.-L. (2010) How long is too long? Effects of loop size on G-quadruplex stability, *Nucleic Acids Research* 38, 7858-7868.
167. Pinnavaia, T. J., Miles, H. T., and Becker, E. D. (1975) Letter: Self-assembled 5'-guanosine monophosphate. Nuclear magnetic resonance evidence for a regular, ordered structure and slow chemical exchange, *Journal of the American Chemical Society* 97, 7198-7200.
168. Šket, P., Črnugelj, M., Koźmiński, W., and Plavec, J. (2004) 15 NH 4+ ion movement inside d (G 4 T 4 G 4) 2 G-quadruplex is accelerated in the presence of smaller Na+ ions, *Organic & Biomolecular Chemistry* 2, 1970-1973.

169. Dolinnaya, N. G., Ogloblina, A. M., and Yakubovskaya, M. G. (2016) Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery, *Biochemistry (Moscow)* 81, 1602-1649.
170. Lim, K. W., and Phan, A. T. (2013) Structural basis of DNA quadruplex-duplex junction formation, *Angewandte Chemie* 52, 8566-8569.
171. Heddi, B., and Phan, A. T. (2011) Structure of human telomeric DNA in crowded solution, *Journal of the American Chemical Society* 133, 9824-9833.
172. Hu, L., Lim, K. W., Bouaziz, S., and Phan, A. T. (2009) Giardia telomeric sequence d(TAGGG)₄ forms two intramolecular G-quadruplexes in K⁺ solution: effect of loop length and sequence on the folding topology, *Journal of the American Chemical Society* 131, 16824-16831.
173. Monchaud, D., and Teulade-Fichou, M.-P. (2008) A hitchhiker's guide to G-quadruplex ligands, *Organic & Biomolecular Chemistry* 6, 627-636.
174. Dixon, I. M., Lopez, F., Tejera, A. M., Estève, J.-P., Blasco, M. A., Pratviel, G., and Meunier, B. (2007) A G-quadruplex ligand with 10000-fold selectivity over duplex DNA, *Journal of the American Chemical Society* 129, 1502-1503.
175. Haider, S. M., Parkinson, G. N., and Neidle, S. (2003) Structure of a G-quadruplex–ligand complex, *Journal of Molecular Biology* 326, 117-125.
176. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P., and Phan, A. T. (2014) Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC3, *Angewandte Chemie* 126, 1017-1020.
177. Heddi, B., Cheong, V. V., Martadinata, H., and Phan, A. T. (2015) Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide–quadruplex complex, *Proceedings of the National Academy of Sciences of the United States of America* 112, 9608-9613.