Structural diversity of G-rich DNA : from telomeric G-quadruplexes to G-junction motifs

Lim, Kah Wai

2013

STRUCTURAL DIVERSITY OF G-RICH DNA:

FROM TELOMERIC G-QUADRUPLEXES

TO G-JUNCTION MOTIFS

LIM KAH WAI

SCHOOL OF BIOLOGICAL SCIENCES

2013

STRUCTURAL DIVERSITY OF G-RICH DNA:

FROM TELOMERIC G-QUADRUPLEXES

TO G-JUNCTION MOTIFS

LIM KAH WAI

School of Biological Sciences

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

2013

Acknowledgments

Preface

My journey on the structures of G-quadruplex DNA began in 2007 when I was carrying out my Final Year Project for undergraduate study under the guidance of Prof. Mu (School of Biological Sciences, NTU) and Prof. Phan (School of Physical & Mathematical Sciences, NTU). Back then, we have defined the folding topology (in K^+ solution) of the two-G-tetrad basket-type fold of Form 3 human telomeric sequence, but the structural work was still largely incomplete. Intrigued by the application of NMR techniques towards high-resolution structural investigation of biomolecules, I decided to pursue a Ph.D. under the guidance of Prof. Phan.

This thesis begins with a general introduction (**Chapter 1**) on the diverse structural forms that can be adopted by DNA, together with a brief overview of the therapeutic and functional applications that arise from this diversity. The chapter concludes with the research motivation for this study. **Chapter 2** describes the various methods and techniques that were employed in this study.

The findings from this study are grouped into three major sections: human telomeric G-quadruplexes (**Chapters 3–5**), promoter G-quadruplexes (**Chapter 6**), and quadruplex–duplex hybrids (**Chapters 7–9**).

The biophysical characterization and structural work on Form 3 is presented in the first part of **Chapter 3**. Part of the NMR data for this project was recorded by Prof. Phan at Memorial Sloan-Kettering Cancer Center in the laboratory of Dr. Dinshaw Patel. UV and CD studies were performed by Dr. Samir Amrane, a former Research Fellow of the lab. I am grateful towards Prof. Bouaziz (INSERM investigator, France) for his assistance towards establishment of the structural computation protocols, which have been essential towards most of my subsequent projects. The second part of Chapter 3 describes a different G-quadruplex topology adopted by a slightly different human telomeric sequence in Na^+ solution. Veronica, an FYP student from the Department of Chemistry & Biological Chemistry (SPMS, NTU), performed a large segment of experimental and initial assignment work for this project, with additional assistance from Dr. Nerea (Instituto de Química Física "Rocasolano", Spain), who was here on a lab attachment, and Dr. Brahim Heddi, a Research Fellow of the lab. The

two G-quadruplex structures presented in this chapter build on previous findings on human telomeric G-quadruplexes, establishing the structural polymorphism of human telomeric DNA across diverse environments. I am the lead author for both projects, and have taken part in sample preparations, as well as NMR experiments and spectral analyses. I have completed the structure computations for both G-quadruplexes, as well as manuscript preparation.

The study on human telomeric DNA encompasses a diverse range of topics, including variant and long telomeric repeats, which are separately addressed in Chapter 4 and Chapter 5, respectively. The work on structures of CTAGGG repeats (**Chapter 4**) was initiated from collaboration with the Royle lab (University of Leicester, UK), who has described unusual properties arising from these sequence-variant repeats. Initial screening and NMR experiments on the series of variant sequences were performed by Jocelyn and Joefina, both FYP students from Ngee Ann Polytechnic at that time, while thermodynamic investigations were conducted by the Mergny group (INSERM, France). I was mainly involved in NMR experiments and spectral analyses, structure computations, and manuscript preparation for this project. **Chapter 5** tackles the possible structures that can be adopted by long human telomeric repeats. We showed that these sequences could adopt a (3+1) G-quadruplex structure with a long propeller loop comprising one or more TTAGGG repeats, which could potentially act as recognition elements. It was a pleasure guiding Doris (SPMS-CBC, NTU), who was working on her FYP, towards the completion of the project. My main roles here were research supervision and manuscript preparation.

Doris, during her subsequent stint in the lab, also contributed tremendously towards NMR and CD experiments for the project on hTERT promoter G-quadruplexes, which was compiled in **Chapter 6**. Initial NMR experiments were carried out by Jocelyn and Joefina. The G-rich promoter sequence was brought to our attention by Dr. Lacroix (INSERM investigator, France), who has performed extensive UV and CD analyses on the series of hTERT promoter sequences, and he was also heavily involved in manuscript preparation. In this study, the equilibrium between two G-quadruplex conformations that can be adopted by a highly G-rich hTERT promoter fragment was dissected. Structural elucidation of the two forms provides insights on how various motifs come together to stabilize distinct G-quadruplex structures. I took responsibility

on sample preparations, CD experiments, NMR experiments and spectral analyses, structure computations, and manuscript preparation for this project.

I was starting on an investigation of the possible structures that can be formed at the quadruplex–duplex interface at the very beginning of the 3' single-stranded telomeric overhang, when the idea of exploring the direct connectivity between duplex and quadruplex DNA sprang to mind. Initial results were promising, leading eventually to the establishment of five separate quadruplex–duplex constructs that cover the fundamental aspects of their compatibility. These results were summarized in **Chapter 7**, which also includes a description on the rational design of a G-junction motif mediating three orthogonally-oriented duplex stems. I performed all sample preparations, NMR experiments, and structure computations, and prepared the manuscript for this project.

A follow-up investigation on the thermal stability of these quadruplex–duplex hybrids is presented in **Chapter 8**. Zi Jian (student from NP), who was carrying out his FYP and industrial attachment, did an impeccable job collecting and characterizing CD-melting data for the two series of quadruplex–duplex hybrids. An understanding on the thermodynamic stability of these complexes should pave the way for their future design and prediction. I am the lead author for this project and was mainly in charge of research design and supervision. I carried out NMR spectral analyses and prepared the manuscript for this project.

Recognizing that quadruplex-forming motifs containing duplex stem-loop elements could potentially exist in the human genome, we elicited help from the Vladimir group (Bioinformatics Institute, ASTAR Singapore) towards the identification of these sequence motifs, the results of which are presented in **Chapter 9**. Development of the search algorithm and analysis of the sequence hits were mainly performed by Piroon, with assistance from Ghim Siong (both working under the supervision of Dr. Vladimir). Zhen Jie (student from SPMS-CBC, NTU), Yi Siang (student from NP), and Zi Jian helped with sample preparations, as well as initial NMR and CD characterization, of the sequence hits. I coordinated the research, prepared samples, performed NMR spectral analyses, and prepared the manuscript for this project.

I have written the main bulk of the manuscript for all projects presented in this study, with valuable inputs from all co-authors, especially my supervisor Prof. Phan.

A general summary of the findings from this study and future perspectives on the investigation of G-rich DNA are presented in **Chapter 11** to conclude this thesis.

Statistics for all computed structures are presented in Appendix I, while protocols for all experimental work are presented in Appendix II.

Table of Contents

List of Figures

Chapter 3   G-Quadruplex Structures of Canonical Human Telomeric Repeats

Chapter 7   Structures of Quadruplex–Duplex Hybrids & G-Junction Motifs

Chapter 8   Thermal Stability of DNA Quadruplex–Duplex Hybrids

Chapter 9   **Putative Quadruplex Motifs Containing Duplex Stem Elements in the Human Genome**

List of Tables

Chapter 9   Putative Quadruplex Motifs Containing Duplex Stem Elements in the Human Genome

Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
6-FAM	6-carboxyfluorescein
A	adenine
^{Br}G	8-bromoguanine
CD	circular dichroism
COSY	correlation spectroscopy
C	cytosine
DAPI	4',6-diamidino-2-phenylindole
DGSA	distance geometry simulated annealing
DMT	4,4'-dimethoxytrityl
DNA	deoxyribonucleic acid
ds	double-stranded
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
duplex	double helix
FID	free induction decay
G	guanine
G4	G-quadruplex
G-tetrad	G•G•G•G tetrad
HMBC	heteronuclear multiple-bond correlation
HPLC	High performance liquid chromatography
HSQC	heteronuclear single-quantum correlation
I	inosine
JR	jump-and-return
LNA	locked nucleic acid
MD	molecular dynamics
mRNA	messenger RNA
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect

NOESY	nuclear Overhauser effect spectroscopy
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PNA	peptide nucleic acid
ppm	parts per million
PQS	putative quadruplex sequence
RNA	ribonucleic acid
SELEX	systematic evolution of ligands by exponential enrichment
SLC-PQS	stem-loop–containing putative quadruplex sequence
ss	single-stranded
T	thymine
TCA	trichloroacetic acid
TDS	thermal difference spectra
TEAA	triethylamine acetate
TFA	trifluoroacetic acid
TFO	triple helix-forming oligonucleotide
T_m	melting temperature
TMS	tetramethylsilane
TOCSY	total correlation spectroscopy
triplex	triple helix
TSS	transcription start site
U	uracil
UTR	untranslated region
UV	ultraviolet

Abstract

Four-stranded G-quadruplex DNA have been implicated in cellular processes and shown to exhibit important functional properties. In particular, putative quadruplex motifs were identified in important genomic loci including telomeres and oncogenic promoters, rendering them promising therapeutic targets. G-quadruplex DNA could adopt diverse topologies, but current knowledge regarding their folding principles remains lacking. This study addresses the conformational heterogeneity of G-quadruplexes from human telomeric and promoter regions through the structural and biophysical characterization on a number of such sequences. In addition, the connectivity of duplex and quadruplex DNA is validated across the diverse geometries of the quadruplex, accompanied by a search for the existence of sequence motifs comprising both duplex and quadruplex elements in the human genome. These findings provide structural insights which could be useful towards the rational design of chemical ligands for therapeutic targeting of genomic G-rich sequences, and pave way for implementation of quadruplex motifs in DNA nanotechnology.

Chapter 1

General Introduction

1.1   Diverse Structural Forms of DNA

1.1.1   The A, B, Z of DNA

Deoxyribonucleic acid (DNA) serves as the genetic repository of the cell; it encodes instructions which direct the growth and processes of all living organisms. DNA can adopt a diverse range of structural forms under different contexts, the most common of which would be A-, B-, and Z-form double-stranded (ds) helices.

B-DNA. The canonical conformation of DNA is the B-form double helix[1-4] (duplex; Figure 1.1a, middle). It consists of two antiparallel right-handed polynucleotide chains. Each nucleotide (nt), the repeating unit, is composed of a deoxyribose ring, a phosphate group, and a nitrogenous base (Figure 1.1b). There are four different bases
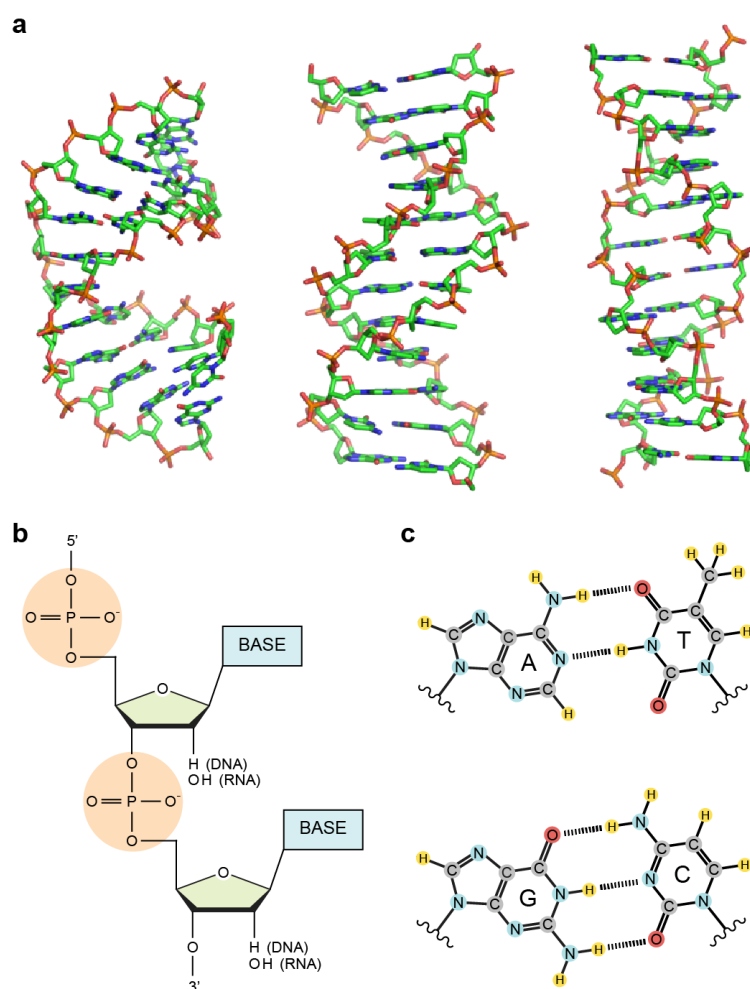
Figure 1.1 | Double helical structures of DNA. (a) Three conformations of double helical DNA, from left to right: right-handed A-DNA[5] (PDB code: 2D47) and B-DNA[4] (PDB code: 1BNA) and left-handed Z-DNA[12] (PDB code: 2ZNA). **(b)** Schematic diagram of a dinucleotide chain. Each nucleotide consists of a deoxyribose (ribose for RNA) ring (green), a nitrogenous base (blue), and a phosphate group (orange). **(c)** Purines (A/G) and pyrimidines (T/C) pair up specifically, giving rise to A•T and G•C Watson–Crick base pairs.

(the purines adenine (A) and guanine (G), and the pyrimidines thymine (T) and cytosine (C)), which pair up through Watson–Crick base pairing (A•T and G•C; Figure 1.1c) along the hydrophobic core of the helix. The phosphate backbones, which are negatively charged, line the exterior of the helix, in contact with the solvent (water). The major and minor grooves are defined by the spacing between the phosphate backbones.

A-DNA. Under low hydration conditions, G•C-rich DNA can adopt the A-form double helix[5,6] (Figure 1.1a, left), which is similarly a right-handed helix. However unlike B-DNA, in which the base pairs align in a near-perpendicular fashion to the helical axis, base pairs in A-DNA display a significant tilt. Furthermore, in A-DNA the major groove is deeper and less accessible, and the minor groove is shallower and wider, with respect to that of B-DNA. A-form double helix is the major conformation for ribonucleic acid[7] (RNA), which differs from DNA by having a hydroxyl group instead of a hydrogen at the C2' position (Figure 1.1b) and having the base uracil (U) in place of thymine. A-form double helix is also presumed to be the predominant conformation for DNA/RNA hybrids[8,9].

Z-DNA. Aside from the right-handed A- and B-form double helices, DNA is also capable of adopting left-handed Z-form double helix[10-12] (Figure 1.1a, right), named for the zigzag progression of its phosphate backbones. This higher-energy form of the double helix can be induced in alternating purine–pyrimidine sequences such as $(GC)_n$ repeats under high salt conditions, negative DNA supercoiling, or through protein binding[12]. The purines in Z-DNA adopt *syn* glycosidic conformation [see Section 1.1.4], and the base pairs are flipped over every other step. Z-DNA is believed to form transiently as a means of torsional strain relief for DNA superhelices, especially at transcription start sites[13] (TSS).

1.1.2   DNA Helical Junctions

A helical junction arises with the confluence of two or more helical segments. It exists in the form of branched architecture (joining three or more segments) or as the mediating element between two distinct helical segments.

Branched DNA structures. Three-way and four-way junctions represent helical junctions with a branched organization[14]. Such structures serve important biological

roles; they take part in DNA replication, recombination, and repair, and they also exist as architectural elements in RNA. DNA four-way junction (Figure 1.2), alternatively termed as the Holliday junction, is the key intermediate in homologous recombination[15,16]. It mediates the exchange of duplex strands between two homologous DNA molecules, bridging four helical segments. In the absence of divalent cations, DNA four-way junction assumes an extended conformation (open form) without any stacking between the four component helices (Figure 1.2a, left). Under the influence of divalent cations, it preferentially adopts the stacked X structure[17] (closed form; Figure 1.2b), which consists of the coaxial stacking of two helical pairs with an antiparallel orientation of the two continuous, non-exchanged, strands[14] (Figure 1.2a, right). This stable arrangement of duplex stems has been widely adapted in the construction of DNA nanomaterials [see Section 1.3].
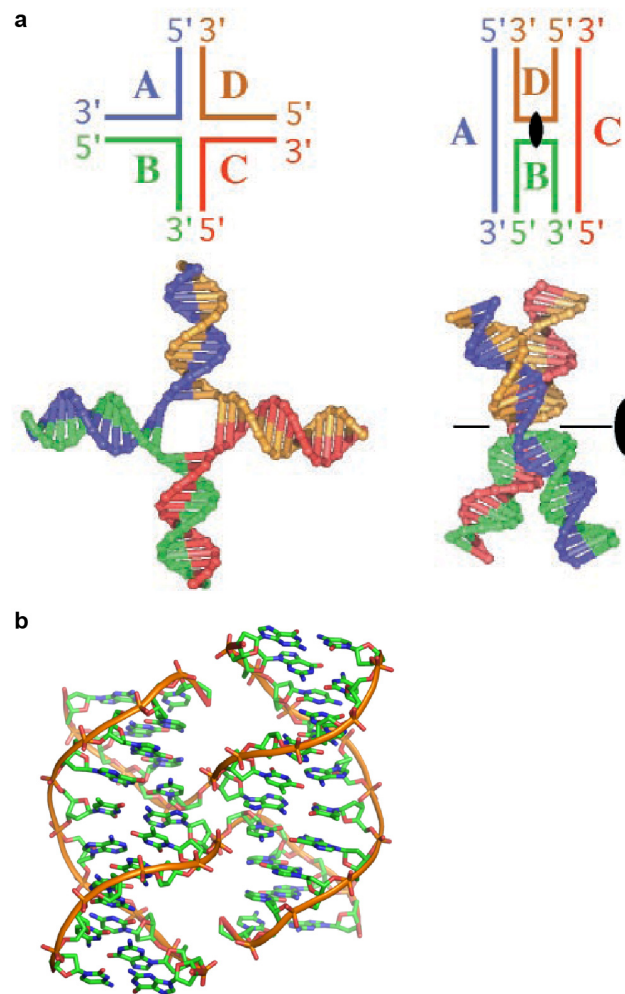
Figure 1.2 | Structures of DNA four-way junction. (a) Conversion from the open planar (left) to the stacked X (right) configuration of the four-way junction is dependent on the presence of divalent cations such as Mg^{2+}. Figure adapted from ref 17. (b) The stacked X structure[17] (PDB code: 1DCW) in ribbon representation, as viewed from the major groove side.

DNA B–Z junction. Formation of a left-handed Z-DNA helical segment has to be flanked by two B–Z junctions. It was found that stacking between the different segments, which have completely reversed handedness, can be maintained with the disruption of a base pair and the subsequent flipping out of the two bases at the junction[18] (Figure 1.3). This architecture of the B–Z junction ensures minimal disturbance towards overall base stacking, which is a major determinant of helix stability.

1.1.3 Multistranded DNA Helices

In addition to the double helices, multiple DNA strands can associate to generate multistranded complexes including three-stranded triple helices and four-stranded helical structures.

Triple helix. On the major groove edge, the projection of acceptor and donor groups from Watson–Crick base pairs permits formation of Hoogsteen hydrogen bonds[19] between the purine and a third base[20] (Figure 1.4a,b). When contiguous purines line up on the same strand of a duplex, a third strand can bind to this polypurine tract through the major groove of the duplex, giving rise to a triple helix[21,22] (triplex; Figure 1.4c). Since the discovery that short oligonucleotides could be targeted at duplex DNA sequence specifically through the establishment of triplex structures[23,24], there has
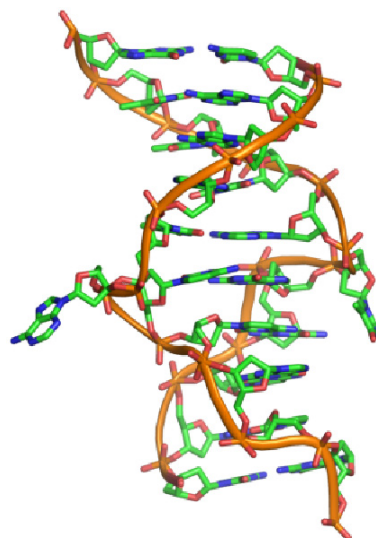
Figure 1.3 | Structure of DNA B–Z Junction. Ribbon view of a B–Z junction[18] (PDB code: 2ACJ). The top half of the helix adopts a regular right-handed twist while the bottom half adopts a left-handed twist. The zigzag progression of the backbone for Z-DNA is evident from the ribbon. At the junction, two bases constituting a base pair is disrupted and extruded outwards, with minimal disturbance towards overall base stacking of the helix.
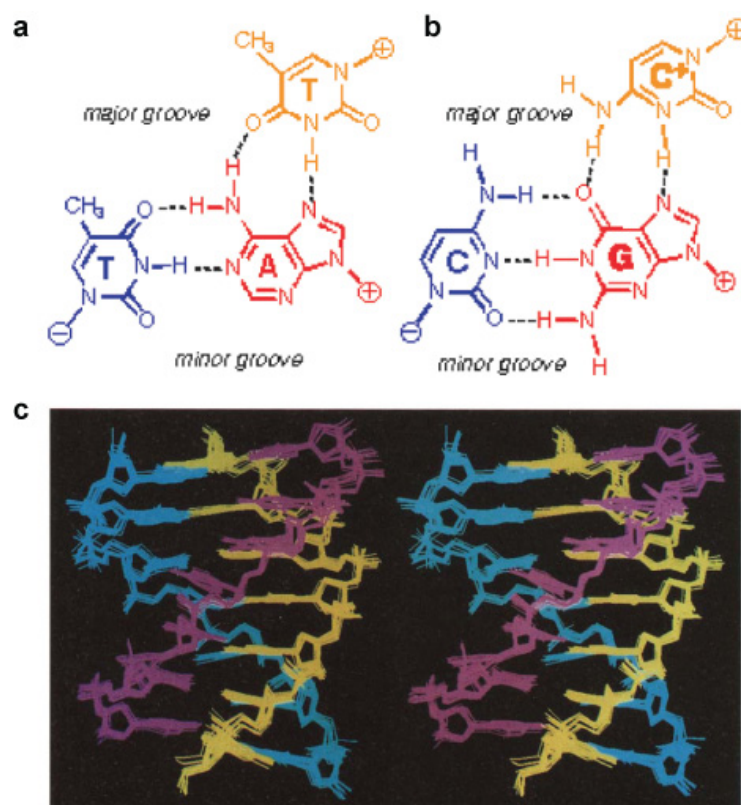
Figure 1.4 | Triple helical structures of DNA. (**a**,**b**) Examples of base triplets formed through Hoogsteen TA•T (**a**) and CG•C$^+$ (**b**) hydrogen bonds between the purine and a third base. Figure adapted from ref 25. (**c**) Stereo view of a DNA triple helix[22] (PDB code: 149D). The TFO (colored in magenta) is aligned in the same direction as the target strand (colored in yellow). Figure adapted from ref 22.

been considerable interest in the development of such triplex-forming oligonucleotides (TFO) for biotechnological and therapeutic applications[20,25] [see Section 1.2]. Assuming that all the bases adopt *anti* glycosidic conformation [see Section 1.1.4], Hoogsteen hydrogen bonds result in the same strand orientation of the TFO and the polypurine tract (Figure 1.4), whereas reverse Hoogsteen hydrogen bonds bring about an antiparallel orientation of the two[20].

Four-stranded DNA structures. Four separate DNA strands could assemble into complexes such as i-motif and G-quadruplex. The i-motif[26] comprises two parallel duplexes, each being bound by hemiprotonated C•C$^+$ base pairs (Figure 1.5a), that are interlocked in an antiparallel manner through intercalation of the base pairs (Figure 1.5b,c). The i-motif is primarily stable at acidic pH, since the pK_a of cytosine imino proton is ~4–5. The G-quadruplex, on the other hand, comprises multiple layers of G•G•G•G tetrads (G-tetrads) [see Section 1.1.4].

6
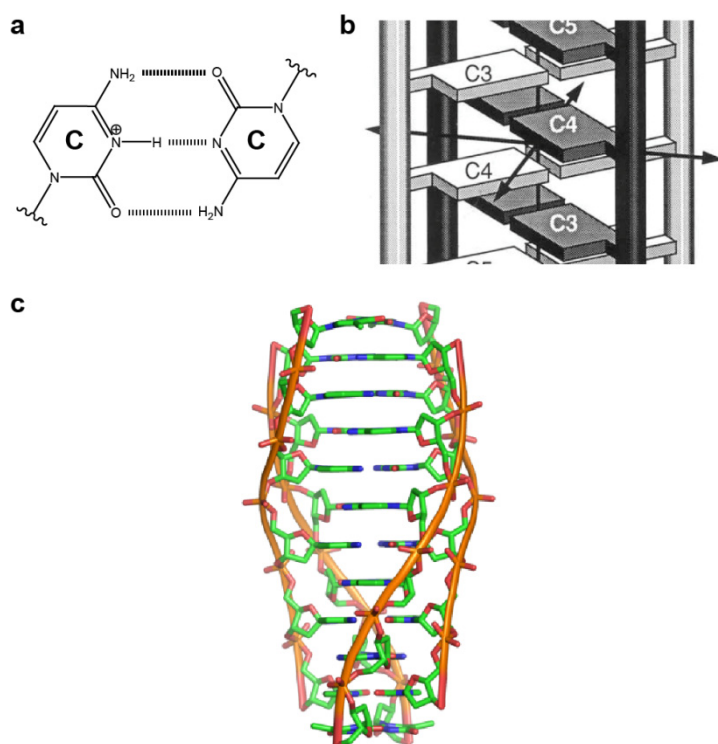
Figure 1.5 | Structure of i-motif DNA. (a) A hemiprotonated C•C$^+$ base pair. **(b)** Schematic diagram illustrating the intercalation of base pairs within i-motif. Figure adapted from ref 26. **(c)** An i-motif[26] (PDB code: 225D) in ribbon representation.

1.1.4 G-Rich DNA and the G-Quadruplex

Structural fundamentals of G-quadruplex DNA. G-rich oligonucleotides have a propensity to form G-quadruplexes[27-30] (G4; Figure 1.6a), which are stable four-stranded helical complexes arising through the multiple stacking of G-tetrads[31] (Figure 1.6b). Stability of G-quadruplexes mainly comes from cyclic Hoogsteen hydrogen bonds among the tetrad guanine bases (H1ⅢⅢO6 and H21ⅢⅢN7), coordination of monovalent cations (e.g. potassium and sodium ions) with the four central O6 atoms, and stacking interactions between the G-tetrad layers. There is a diverse range of G-quadruplex folding topologies. They can be broadly categorized based on the relative alignments of the four strands making up the core[32] (Figure 1.6c): (i) all four strands arranged in parallel, designated parallel (or propeller), (ii) three strands arranged in parallel and the remaining strand in the reverse alignment, designated (3+1) (or hybrid/mixed), (iii) two strands arranged in parallel and the remaining two in the reverse alignment, each of which has an antiparallel and a parallel neighboring strand, designated antiparallel up–up–down–down (or basket), and (iv) two strands arranged in parallel and the remaining two in the reverse alignment, each of which has antiparallel neighboring strands, designated antiparallel up–down–up–down (or chair).
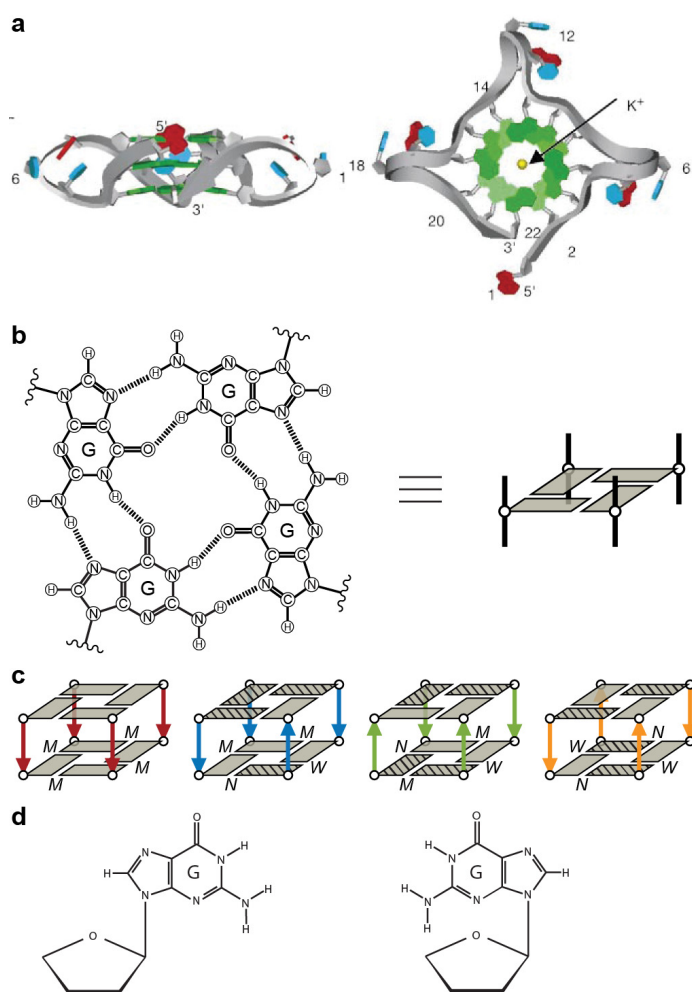
Figure 1.6 | Structures of quadruplex DNA. (a) Side (left) and top–down (right) views of a propeller-type G-quadruplex[33] (PDB code: 1KF1) in ribbon representation. Figure adapted from ref 33. **(b)** Schematic diagram of a G-tetrad. **(c)** G-tetrad core topologies with varying assortment of groove widths (*W*, *N*, and *M* stand for wide, narrow, and medium grooves, respectively). *Syn* guanine bases are shaded. **(d)** *Anti* (left) and *syn* (right) glycosidic conformations of a guanine base.

Each type of core has an associated groove dimensions (Figure 1.6c) and glycosidic conformations of the guanine residues (*anti* or *syn*; Figure 1.6d). Depending on the manner of connection between two strands, there are three general classes of loops (Figure 1.7): edgewise (or lateral)[34,35], diagonal[36], and double-chain-reversal (or propeller)[37].

Biological relevance of G-quadruplexes. The findings that G-rich sequences could adopt G-quadruplex conformations under physiological conditions[38-40] first casted the spotlight on these structures. It has since been shown that putative quadruplex sequences (PQS) are widespread in genomic DNA[41,42]; they have been identified in various genomic locations including telomeres[43] [see Section 3.1], gene promoters[44-47] [see Section 6.1], introns[48], immunoglobulin class switch regions[49], rDNAs[50],
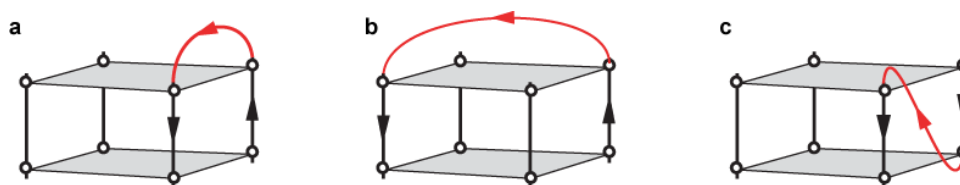
Figure 1.7 | Different classes of quadruplex loops. (**a–c**) Various manners of connection between any two strands constituting the quadruplex core give rise to different classes of loops: (**a**) edgewise, (**b**) diagonal, and (**c**) propeller. Figure adapted from ref 32.

minisatellites[51], and untranslated regions (UTRs) of mRNA[52,53]. Recently, a number of studies implicated the involvement of quadruplexes in cellular processes including recombination[54] and replication[55-58], corroborating their biological relevance[59,60]. Furthermore, induction or stabilization of quadruplex formation through the application of quadruplex-interactive compounds boasts immense therapeutic potentials[61] [see Sections 1.2, 3.1, and 6.1].

Artificial G-quadruplexes with useful properties. Synthetic G-rich oligonucleotides exhibiting interesting properties with potential applications have been generated, among which include specific binding of molecular targets[34,35,62] [see Section 1.3], catalytic property[63] [see Section 1.3], charge conductivity[64], and anticancer[65-69] and anti-HIV[70-75] activities. In addition, higher-order assemblies of G-quadruplexes have also been reported[76-79], which could contribute towards the development of DNA nanomaterials[80] [see Section 1.3].

1.2 Targeting DNA Structures for Therapeutics

The diverse structural forms that can be adopted by DNA offer a feasible route towards the targeting of specific genomic locations, whereby drugs/ligands can be individually tailored to recognize a particular gene segment for therapeutic purposes. An ideal drug should preferably recognize a unique site with minimal off-target binding, much like how regulatory RNA/proteins recognize their cellular targets.

Sequence-specific recognition. Proteins such as transcription factors recognize DNA in a sequence-specific way, thereby exerting their functions at localized positions. This sequence specificity is mainly achieved through the explicit establishments of electrostatic and hydrogen-bond interactions between functional groups present on the protein binding domain and on the major and/or minor groove edges of the base pairs. Similarly, while designing synthetic ligands or

oligonucleotides for DNA targeting, recognition of these DNA functional groups has to be maximized in order to attain sequence specificity. Different modes of ligand binding to DNA are briefly outlined below.

Groove-binders. TFO belongs to this class of ligands, which binds DNA through the major or minor groove of the duplex. To date, the majority of small-molecule groove-binders target the minor groove of duplex DNA[81] (Figure 1.8). Examples include Hoechst stains[82], 4',6-diamidino-2-phenylindole (DAPI), and pyrrole–imidazole polyamide[83]. These groove-binders could serve as fluorescent probes[84] or induce site-specific DNA cleavage[23]. In addition, they could even be used for the regulation of gene expression[85,86].

Intercalators. Small-molecule ligands that fit snugly between two adjacent base pairs are known as intercalators. They are mostly aromatic and planar in nature. Due to their tight binding to DNA, intercalators are potent inhibitors of replication and transcription. Hence some of them have been exploited as chemotherapeutic agents. Examples of DNA intercalators include acridine, ethidium bromide, and daunomycin.

Cross-linking agents. Cross-linking agents establish covalent linkages at two distinct sites on a DNA, thereby cross-linking the two positions (either on the same strand or on opposite strands of a duplex). The resultant DNA adducts could lead to mutations if they are not properly resolved by the DNA repair machinery. Cisplatin is a well-known DNA cross-linking agent used in chemotherapy.

Antigene oligonucleotides. The antigene[87] approach involves targeting of a DNA segment with oligonucleotides for transcriptional regulation of the target region. This
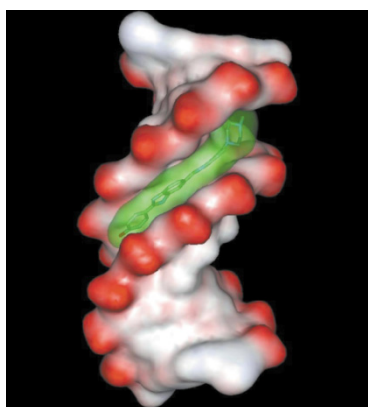
Figure 1.8 | Ligand binding to duplex DNA. Recognition of the minor groove of duplex DNA[82] (in surface representation; phosphate groups in red) by the ligand Hoechst 33258 (highlighted in green) (PDB code: 296D). Figure adapted from ref 82.

could be based on formation of triple helix[21,23,24] or through strand displacement[88], in which the antigene oligonucleotide displays a stronger affinity for the target strand as compared to its native counterpart. Binding affinity, stability, and effectiveness of antigene oligonucleotides can be enhanced with the use of modified nucleotides including peptide nucleic acid[88] (PNA) and locked nucleic acid[89,90] (LNA), or through the conjugation of strong DNA binders such as intercalators[91].

Quadruplex-binding ligands. G-quadruplex DNA have emerged as promising therapeutic targets[30,61] in recent years, as their formations have been implicated in important genomic locations including telomeres [see Section 3.1] and oncogenic promoters [see Section 6.1]. Quadruplex-binding ligands investigated thus far almost exclusively bind to G-quadruplexes through G-tetrad end-stacking[30,92] (Figure 1.9). Notable examples include derivatives of acridine[93,94] and telomestatin[95]. These ligands typically exhibit excellent selectivity against duplexes, but they often show a lack of discrimination between different quadruplexes. On the other hand, quadruplex groove-binders have also been recently explored[96,97].

1.3   Functional Properties of Different DNA Structural Forms

DNA aptamers and DNA enzymes. As befitting of its role as the genetic informational carrier, DNA is rather inert and almost always prefers to adopt the B-form double helical structure. However, it has been shown that just like RNA, DNA

Figure 1.9 | Ligand binding to quadruplex DNA through G-tetrad end stacking. Binding of an acridine derivative (in space-filling representation) to the terminal G-tetrad of the dimeric *Oxytricha* telomeric G-quadruplex[94] (PDB code: 1L1H). Figure of complex adapted from ref 92. Chemical structure of the ligand is shown on the right.

could also exhibit aptameric[62,98] and catalytic[63,98,99] properties. Many of these functional DNA molecules have been isolated through combinatorial methods such as systematic evolution of ligands by exponential enrichment[100] (SELEX), which amplify lineages of DNA molecules with the desired properties from a library of random-sequence DNA.

DNA nanomaterials and nanodevices. DNA nanotechnology[80,101] represents an emerging field with wide-ranging applications and involves the construction of nanomaterials and nanodevices out of DNA. This is possible due to the high fidelity of DNA complementary strand pairing and the robust and highly predictable nature of the DNA double helix. Current applications that have been demonstrated include the rational design of DNA tiles[102], objects[103-108], and crystals[109], construction of DNA nanodevices[110] that transit between different states[111,112], construction of DNA nanoelectronic circuits[113] and switches[64], construction of DNA frame[114] for single-molecule analysis[115,116], development of DNA sensors[117], design of DNA capsules[118,119] for drug delivery[120], and the controlled alignment of protein assembly[121] for enzymatic reaction[122].

1.4 Research Motivation

This study seeks to explore the diverse G-quadruplex structures that can be adopted by human telomeric and promoter G-rich sequences, and to characterize their properties through biophysical techniques such as NMR, UV, and CD spectroscopy. In addition, this study aims to examine the connectivity of duplex and quadruplex DNA, which could be relevant in a natural context, for instance, when G-quadruplex formation occurs in close proximity to the regular duplex.

Structural insights gained could facilitate drug design efforts or aid in future studies concerning the interactions of G-quadruplexes with biomolecules. In addition, these would contribute towards the understanding of G-quadruplex folding principles for design and prediction of such structures, as well as the structural and functional applications of these motifs.

Chapter 2

Methodology

2.1 DNA Preparation

2.2 Nuclear Magnetic Resonance Spectroscopy

2.3 Ultraviolet Spectroscopy

2.4 Circular Dichroism Spectroscopy

2.5 Structure Computation

2.1   DNA Preparation

DNA synthesis. DNA oligonucleotides in this study were chemically synthesized based on automated solid-phase synthesis[123,124] utilizing phosphoramidite chemistry[125]. This methodology enables the rapid production of custom-made oligonucleotides of any desired sequence (typically 10–40 nt with a fairly high yield) at a low cost. Mechanistic details of the synthesis have been amply described elsewhere, and herein only the general principles of the procedure shall be outlined.

Synthesis of an oligonucleotide involves the sequential addition of nucleotides onto the starting unit immobilized on a solid support. The basic building block of choice is phosphoramidite (Figure 2.1), a derivative of nucleotide much more reactive towards the formation of internucleotide linkages. The numerous functional groups of phosphoramidite are adorned with protecting groups [e.g. acid-labile 4,4'-dimethoxytrityl (DMT) on the 5'-hydroxyl groups and base-labile moieties on amino groups of the bases and on the phosphite (trivalent phosphorus) groups] to prevent undesirable reactions from happening. In contrast to the extension of nucleic acids by polymerases (which proceeds in the 5'-to-3' direction), chemical synthesis of oligonucleotides occurs in the reverse (i.e. 3'-to-5') direction. Each cycle of nucleotide addition can be broken down into four general steps: detritylation, coupling, capping, and oxidation (Figure 2.2).

Detritylation is first performed to remove the acid-labile DMT protecting group on the 5'-hydroxyl through the passage of trichloroacetic acid (TCA). The free DMT cation emits a bright orange color, and can be quantified spectrophotometrically at the beginning of each cycle to monitor the yield of the reaction. The naked 5'-hydroxyl on the immobilized chain would then be free to react with the incoming phosphoramidite during the coupling step. Immobilization of the oligonucleotide chains on a solid
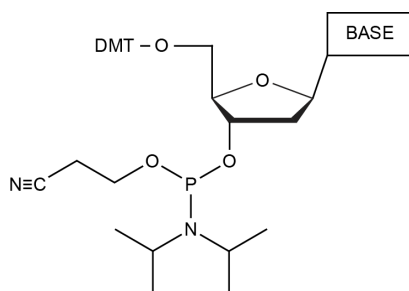
Figure 2.1 | The phosphoramidite basic building block. Schematic diagram of the phosphoramidite building block, protected at the 5'-end by a DMT group.

14
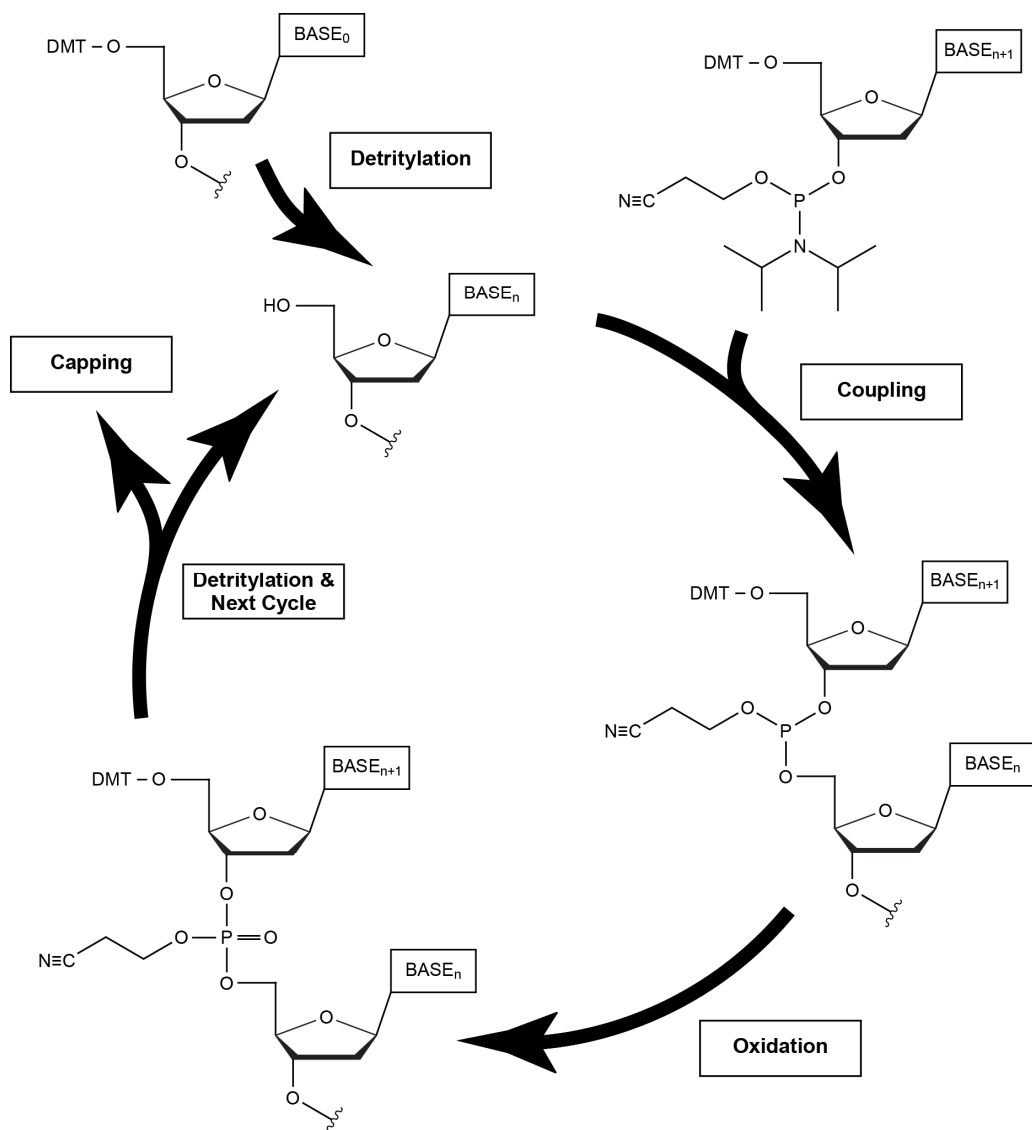
Figure 2.2 | The chemical synthesis of oligonucleotide. General work flow illustrating the chemical synthesis of oligonucleotide based on the phosphoramidite chemistry. Note that chain elongation proceeds in the reverse order with respect to the conventional numbering of oligonucleotide (from the 5'-terminal) and is initiated from the residue at the 3'-end (Base$_0$).

support (µm-sized resins) allows excess solution-phase reagents to be passed through the reaction chamber (resins trapped between filters), hence driving the reaction to near completion. Furthermore, impurities and excess reagents can be washed off thoroughly without the need for tedious purification at the end of each cycle. Successful attachment of free phosphoramidite onto the immobilized oligonucleotide chain through establishment of internucleotide linkage would extend the chain by one unit. In practice, the coupling will not reach totality, and there will be non-extended chains (or failure sequences) at the end of each coupling step. These failure sequences need to be removed from the pool of oligonucleotide chains available for the next addition cycle, lest they contribute to the oligonucleotide population at the end of

synthesis. The free 5'-hydroxyl groups of these failure sequences are sealed off by subjecting them to acetylation during the capping step. Due to the chemistry of the phosphoramidite methodology, the newly established internucleotide linkage consists of a phosphite linkage, which is unstable. This is converted to a stable pentavalent phosphate triester by treatment with iodine during the oxidation step.

The extended oligonucleotide chain would now be ready for the next cycle of phosphoramidite addition. The cycles of phosphoramidite addition are repeated till the completion of the desired oligonucleotide sequence, which then needs to be cleaved off from the solid support.

DNA de-protection. The linker that tethers the oligonucleotide chain to the solid support is base-labile, similar to the protecting groups on amino groups of the bases and on the phosphates, but differs from the acid-labile DMT groups protecting the 5'-hydroxyl groups. This orthogonal chemistry ensures that the linkers, as well as the base- and phosphate-protecting groups, are not inadvertently cleaved or removed during the phosphoramidite addition cycles. Treatment with concentrated ammonium hydroxide at 55 °C for a minimum of twelve hours will completely cleave off the oligonucleotide chains and remove the protecting groups.

DNA purification. Several techniques have been used for the purification of DNA, including high performance liquid chromatography (HPLC), purification by cartridges, and polyacrylamide gel electrophoresis (PAGE). Purification through the use of cartridges employs a hydrophobic stationary phase (reverse phase) to trap the oligonucleotide chains, as the DMT group on the 5'-terminus binds tightly to the hydrophobic solid support (polymeric resins packed in the cartridge). Failure sequences, due to the absence of the DMT group, will mostly be flushed through the cartridge unbound. Those failure sequences that bind weakly to the resins can be washed off through the passage of ammonium hydroxide (1:10). The acid-labile DMT group is cleaved from the 5'-terminus by flushing the cartridge with 2% trifluoroacetic acid (TFA). The DNA can then be eluted from the cartridge with 20% acetonitrile.

Primer on DNA sample preparation. The desired sequence is input into the DNA synthesizer (ABI 394 DNA/RNA synthesizer) and automated synthesis (in production scales of 0.2 or 1.0 μmole) of the oligonucleotide is carried out using reagents from Glen Research (Virginia, USA). Upon completion of synthesis, the resins are transferred into a glass vial together with 4 mL of concentrated ammonium hydroxide.

De-protection is carried out at 55 °C for a minimum of twelve hours. The sample is loaded three times through a cartridge (Poly-Pak™), which was pre-treated with acetonitrile (2 mL) and triethylamine acetate (TEAA) (2 mL) successively. Ammonium hydroxide/water (1:10) (3 mL) is flushed through the cartridge, followed with deionized water (2 mL). Next, 2% TFA/water (2 mL) is flushed through the cartridge, followed with deionized water (2 mL). The purified DNA is eluted from the cartridge with 20% acetonitrile/water (2 mL), and is dialyzed (using a membrane with a 1000 Da molecular weight cut-off) successively against water, target solution, and water again. For purification, the aforementioned volumes of reagent apply to DNA samples synthesized at a scale of 0.2 μmole. For synthesis in the scale of 1.0 μmole, two times the amount of reagents are used. Dry DNA powder is obtained upon freeze-drying, and is re-suspended in the buffer of choice.

2.2   Nuclear Magnetic Resonance Spectroscopy

Nuclear spin. Spin (I) is a fundamental property of nucleons (protons and neutrons) and electrons, and comes in multiples of ½. The nuclear spin defines the magnetic moment of the nucleus, which can have an integral spin, a half-integral spin, or zero spin. The nuclei most relevant for biological studies are ^1H ($I = ½$), ^{13}C ($I = ½$), ^{15}N ($I = ½$), and ^{31}P ($I = ½$). ^1H is the most commonly studied nucleus, and will be the focus of discussion henceforth.

Nuclear magnetic resonance. When subjected to an external magnetic field (B_o), ^1H exists in one of two spin states [a lower-energy +½ (or α) state and a higher-energy −½ (or β) state]. The +½ state, which is populated in a slight excess at equilibrium, can transition to the −½ state upon absorption of a photon that matches exactly the energetic difference between the two states (hence the term resonance). This energetic difference corresponds to the radio frequency band of the electromagnetic spectrum, and can be described by the equation $ν = γB$, where $ν$ is the resonance frequency at which the absorption happens, $γ$ is the gyromagnetic ratio which is specific for each nucleus type (for ^1H, $γ = 42.576$ MHz/T), and B corresponds to the field strength of B_o. Nuclear magnetic resonance (NMR) thus refers to the phenomenon whereby nuclei absorb electromagnetic radiation at a particular resonance frequency corresponding to the energetic difference between the different spin states, when they are subjected under an external magnetic field.

Bulk magnetization and free induction decay. At equilibrium, the magnetic moments of the nuclear spins in a sample are aligned in such a way that there is a slight excess pointing in the general direction of B_o, resulting in a net magnetic field along this direction. This bulk magnetization of the sample can be represented as a magnetization vector pointing in the direction of B_o (Figure 2.3). A radiofrequency pulse applied along the x-direction will generate an oscillating magnetic field along the same direction, which will lead to the rotation of the magnetization vector about the zy-plane. Such tilting of the magnetization vector away from B_o will lead to rotation (or precession) of the vector about the z-axis at the Larmor frequency, which is exactly the same as the resonance frequency. The precession of the magnetization vector will induce a current (free induction signal), recorded as a function of time, in the detection coil aligned along the xy-plane. The system will revert back to equilibrium ("relax") over time, leading to a gradual loss of the free induction signal, otherwise known as free induction decay (FID). The FID (which has a time domain) is then converted to NMR signal (which has a frequency domain) by a mathematical procedure known as the Fourier transform, generating the familiar NMR spectrum.

Electron shielding. Even though all ^{1}H nuclei in a molecule are subjected to the same B_o, the apparent magnetic field experienced by different ^{1}H is not the same. This arises due to the different electron density surrounding the ^{1}H nuclei. Electrons around the nuclei will induce a local magnetic field that opposes B_o, leading to a weaker apparent magnetic field. The greater the electron density around the ^{1}H, the stronger the local magnetic field generated, and the weaker the apparent magnetic field experienced by the nucleus. The ^{1}H will resonate at a lower frequency, and are said to be upfield-shifted. This screening effect is known as diamagnetic shielding. The converse is true; highly de-shielded nuclei experience a stronger effective magnetic
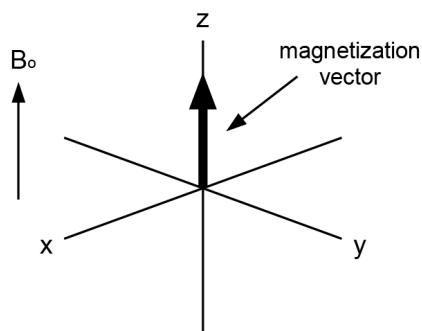
Figure 2.3 | The bulk magnetization of a sample. At equilibrium, the bulk magnetization of a sample is established in the same direction as the external magnetic field (B_o).

18

field and consequently resonate at a higher frequency, and are said to be downfield-shifted. The electron density surrounding a nucleus is to a large extent affected by the electronegativity of its neighboring atoms.

Chemical shift. The resonance frequency of ^{1}H will depend on the magnetic field strength of the spectrometer used. As such, the exact same sample recorded on spectrometers with different field strengths will yield different resonance frequencies for the same nuclei. To facilitate data interpretation/comparison, the chemical shift scale is defined. The chemical shift (δ) is described by the equation $\delta = (\nu - \nu_{ref}) / \nu_{ref}$, where ν_{ref} is the resonance frequency of an agreed reference compound, whose δ is necessarily defined to be zero. Two widely adopted reference compounds are 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and tetramethylsilane (TMS). As the chemical shift is a ratio between frequencies (each of which scales with the magnetic field), it is independent of the magnetic field strength and is universal. Typically, the chemical shift is rather small and it is a common practice to multiply it by 10^6, representing it in parts per million (ppm), i.e. $\delta_{ppm} = [(\nu - \nu_{ref}) / \nu_{ref}] \times 10^6$.

Two-dimensional NMR. Conventional one-dimensional (1D) NMR involves the Fourier transform of NMR signals recorded with respect to a single time variable, generating a spectrum with a single frequency variable. Two-dimensional (2D) NMR, on the other hand, involves the double Fourier transform of NMR signals recorded with respect to two time variables (t_1 and t_2), generating a spectrum with two frequency variables (Figure 2.4). Briefly, the work flow for the collection of 2D NMR dataset is as follows. The sample is initially excited by one or more pulses during the preparation time, followed by the evolution time t_1, during which the system is
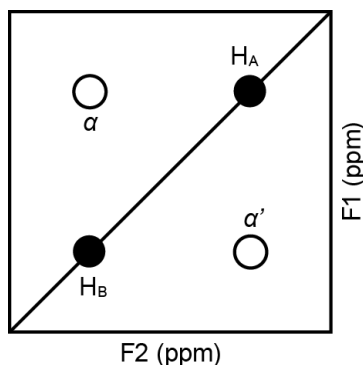
Figure 2.4 | Schematic representation of a 2D homonuclear NMR spectrum. The system consists of two hypothetical protons (H_A and H_B). The presence of interproton cross-peaks (Peak α and α') between them indicate that the two protons exhibit the kind of correlation that is detected in the particular experiment.

allowed to evolve. The sample is subsequently excited by one or more pulses during the mixing time, after which the FID is collected over the detection time t_2. This series of events is repeated multiple times with varying t_1. t_1 is incrementally increased starting from zero, and for each increment (at a fixed interval) of t_1, the corresponding FID is separately recorded and stored, until such time when enough sets of FID have been accumulated (typically 50 to 500). Each succession of preparation, evolution, mixing, and detection is defined as a pulse sequence, and the combination of pulses employed during the preparation and mixing periods will define the kind of information that is observed in a 2D spectrum.

J-coupling and through-bond correlation. Two nuclei connected through chemical bonds will exhibit *J*-coupling, which arises from an indirect interaction between them that is mediated by the electrons constituting the bonds bridging them. When the coupling partners are of the same nucleus type (e.g. ^1H and ^1H), the system would be described as homonuclear; when the coupling partners are of different nucleus types (e.g. ^1H and ^{13}C), the system would be described as heteronuclear. *J*-coupling provides information on the connectivity within a molecule, and serves as the basis for through-bond correlation experiments including heteronuclear multiple-bond correlation (HMBC), heteronuclear single-quantum correlation (HSQC), total correlation spectroscopy (TOCSY), and correlation spectroscopy (COSY). The presence of a cross-peak between two nuclei in such experiments would indicate that the two nuclei are connected through chemical bonds.

Nuclear Overhauser effect and through-space correlation. Two nuclei close in space will exhibit dipole–dipole interactions, resulting in a mutual exchange of spin magnetization through cross-relaxation. This is known as nuclear Overhauser effect (NOE). The extent of cross-relaxation has an inverse correlation with respect to the distance between the two nuclei, according to the relationship NOE α $(1 / r^6)$. The presence of a cross-peak between two nuclei in a nuclear Overhauser effect spectroscopy (NOESY) spectrum would indicate that the two nuclei exhibit through-space correlation and are thus close in space. The stronger the intensity of the cross-peak, the closer together the two nuclei are. A good rule of thumb is that as long as two nuclei are within 5 Å apart in space, they would give rise to an NOE interaction. NOESY spectra have been widely employed for the determination of the three-dimensional (3D) structures of biological macromolecules.

20

Solvent signal suppression. Water is the solvent employed in this study. It has a concentration of ~55 M, which is around five orders of magnitude higher than a typical oligonucleotide sample (~0.2–2.0 mM). As such, if the signal from water (at ~4.7 ppm) is left unchecked, the signals of interest from the sample will be swamped by it. In this study, the jump-and-return[126] (JR) pulse sequence is typically utilized to suppress the water signal.

NMR spectroscopy of nucleic acids. Knowledge on the general distribution of ¹H chemical shifts associated with various functional groups is helpful towards the analysis of an NMR spectrum. Figure 2.5 displays the chemical shift ranges of the different proton types found in a nucleotide. As DNA is a polymer consisting of many repeating nucleotide units, protons of the same chemical type from all nucleotides will generally fall within the same chemical shift range. In order to interpret the NMR spectrum of an oligonucleotide chain, it is necessary to identify the exact chemical shifts associated with the different ¹H of the oligonucleotide. This procedure is known as the assignment of chemical shifts. Several assignment techniques employed in this study will be described below.

Primer on the assignment of ¹H chemical shifts for nucleic acids. Here assignment strategies based on site-specific isotopic labeling and through-bond/
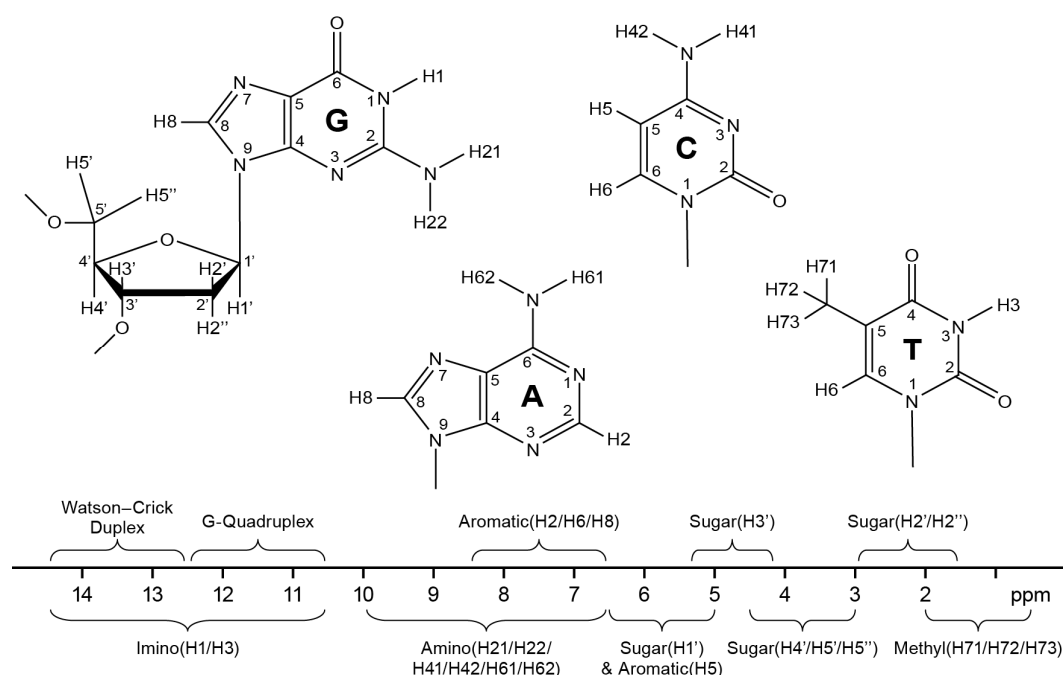
Figure 2.5 | Chemical shift distribution of the various proton types in a nucleotide. The chemical shift scale is demarcated to outline the different ranges corresponding to the various proton types (numbered as in the schematics) found in a nucleotide.

through-space connectivities will be covered. The 1D/2D NMR spectra of DNA quadruplex–duplex Construct I (5'-GGTTGGCGCGAAGCATTCGCGGGTTGG-3') [see Section 7.2] shall be used for illustration purposes henceforth.

Site-specific ^{15}N enrichment for the unambiguous assignment of imino proton. The exchangeable (or labile) imino (H1 and H3 for guanine and thymine, respectively) protons will freely exchange with H_2O when exposed, leading to a broadening (and consequently non-detection) of these proton peaks. Upon establishment of a hydrogen bond (e.g. formation of a base pair), the rate at which the imino proton exchanges with H_2O is sufficiently retarded such that the proton peak can be detected. This can also happen if the imino proton is embedded within the inner core of a folded structure and is thus protected from exchange. In this study, ^{15}N-filtered 1D experiments were carried out on site-specifically ^{15}N-enriched samples for the unambiguous assignment of imino protons. In this NMR experiment, ^{1}H is first excited, after which the spin magnetization is selectively transferred onto chemically-bonded ^{15}N and back to ^{1}H. Due to the low natural abundance of ^{15}N (0.37%), this experiment gives very poor signals. Upon site-specific enrichment of ^{15}N (2%–4%; incorporated during chemical synthesis), one residue at a time, imino proton peaks of the respective ^{15}N-enriched residues can be identified individually (Figure 2.6), leading to the unambiguous assignment of the imino protons. This labeling approach can also be extended to the assignment of H8 proton (based on its coupling with $^{15}N7$).

Site-specific ^{2}H labeling for the unambiguous assignment of H8 proton. ^{2}H (γ = 6.536 MHz/T) has a resonance frequency far away from that of ^{1}H. Hence site-specific replacement of a ^{1}H with a ^{2}H will lead to the disappearance of the corresponding
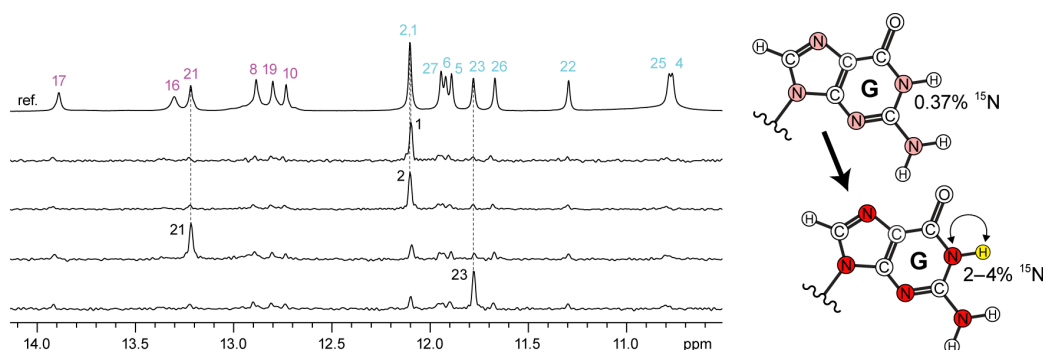
Figure 2.6 | Unambiguous assignment of imino protons through ^{15}N-enrichment. Imino proton assignment is achieved through enrichment of ^{15}N abundance from 0.37% (natural) to 2–4%, one residue at a time, and subsequent detection of the coupling between ^{1}H and ^{15}N. Imino proton peaks corresponding to Watson–Crick base pair and G-tetrad formation are labeled in magenta and cyan, respectively.

22
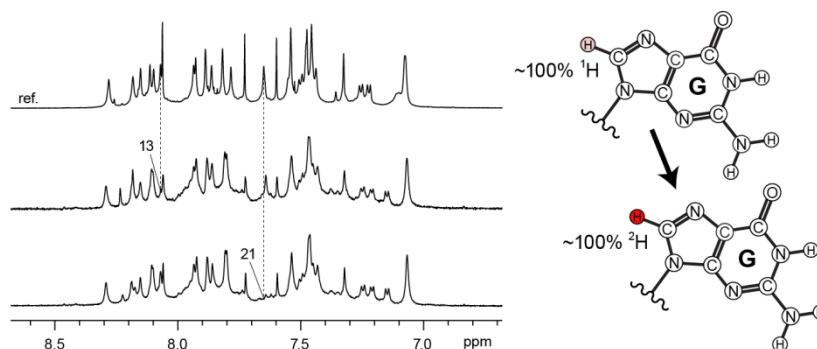
Figure 2.7 | Unambiguous assignment of H8 protons through ²H-substitution. H8 proton assignment is achieved through site-specific substitution of ¹H to ²H, one residue at a time, thereby leading to the disappearance of the corresponding ¹H signal.

proton peak from its usual chemical shift range (Figure 2.7). This strategy has been employed towards unambiguous assignments of adenine and guanine H8 protons.

JR-HMBC to determine through-bond correlation between H1 and H8 protons of guanines. In this study, JR-HMBC is conducted to establish/corroborate guanine H1 and H8 proton assignments. These two protons from a single guanine would exhibit long-range *J*-coupling with the same $^{13}C5$ (at a natural abundance of 1.1%) in the middle (Figure 2.8, inset). Hence the cross-peaks corresponding to $^{1}H1–^{13}C5$ and $^{1}H8–$ $^{13}C5$ would align at the same chemical shift in the ^{13}C dimension (Figure 2.8) and can be assigned for each intranucleotide guanine imino–H8 pair.

Establishment of spin systems. A spin system is the collection of nuclear spins from a single nucleotide. It is helpful in the assignment process to begin with the identification of all the spin systems within the oligonucleotide chain, after which their sequential order can be established based on the expected connectivity patterns between successive spin systems. For discussion purposes, here the spin system is restricted to H8/H6, H1', and H2'/H2" protons. Assignment of these protons for the
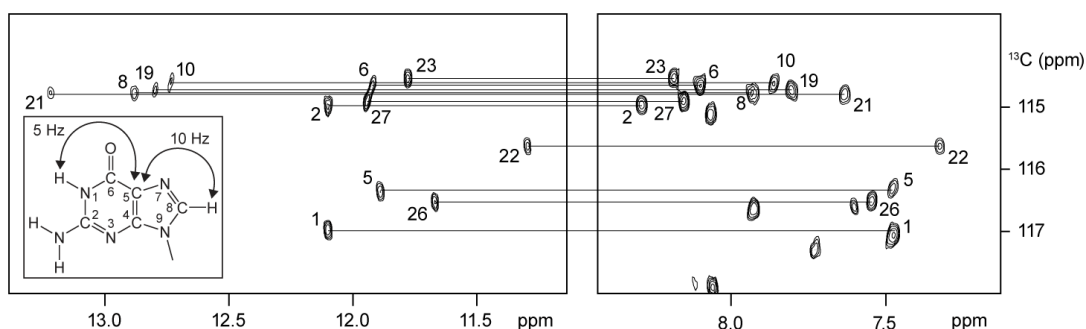
Figure 2.8 | Guanine H1 and H8 proton assignment through JR-HMBC. H1 and H8 protons from the same guanine exhibit long-range *J*-coupling with $^{13}C5$ in the middle (as shown in the inset), giving rise to two cross-peaks that are aligned at the same chemical shift in the ^{13}C dimension.

23

base-pair step G10–A11/T17–C18 of DNA quadruplex–duplex Construct I shall be described below.

Characteristic spin connectivity within a nucleotide. Due to their close proximity, spins coming from the same nucleotide can be identified fairly quickly based on the abundance of intranucleotide NOE cross-peaks (Figure 2.9). These assignments can be verified by through-bond correlations between these protons, e.g. based on cross-peaks detected in TOCSY spectrum (Figure 2.10).
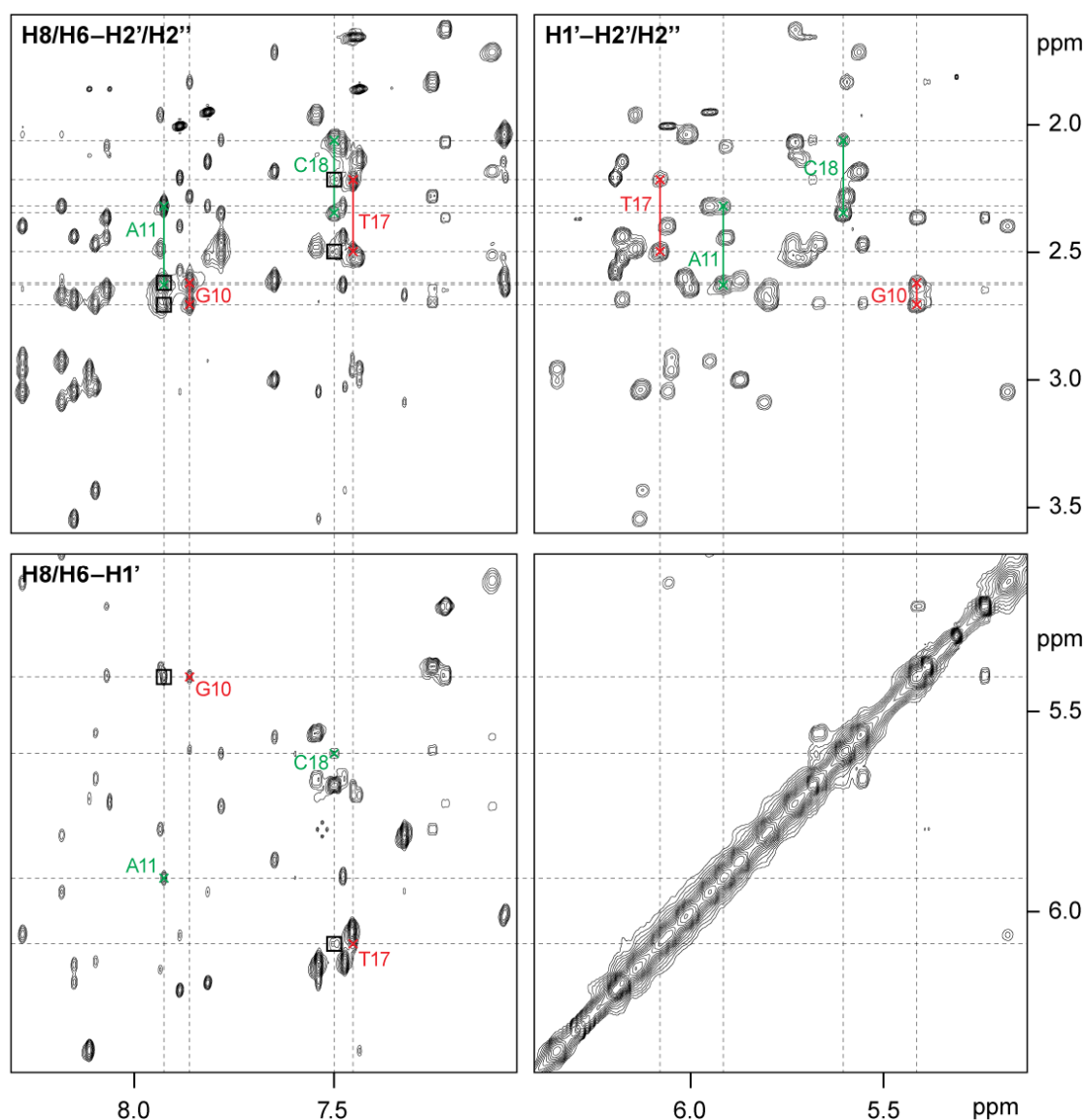
Figure 2.9 | Establishment of intranucleotide and internucleotide spin connectivity. NOESY (mixing time, 200 ms) plot highlighting the network of intra- and inter-nucleotide correlations among H6/H8, H1', and H2'/H2" protons of the dinucleotide steps G10–A11 and T17–C18. Each pair of intranucleotide H1'–H2'/H2", as well as H6/H8–H2'/H2", cross-peaks are joined by a line. Internucleotide cross-peaks (arising from spin connectivity between sequential residues) are framed in boxes. The type of proton correlations corresponding to each region of the spectrum is highlighted.
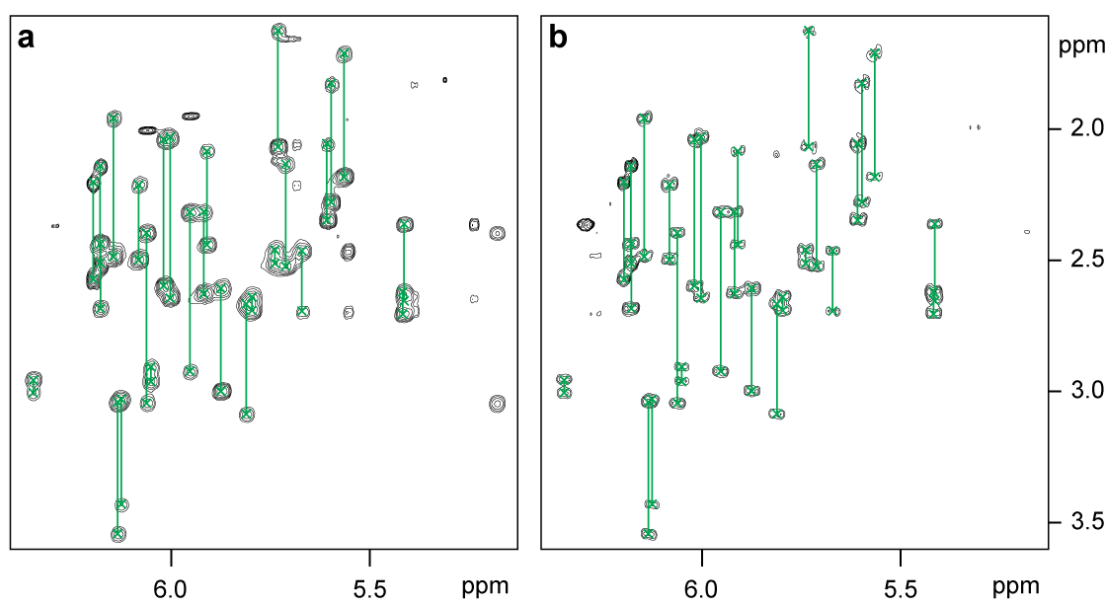
Figure 2.10 | Through-space and through-bond correlations. Plot of the (**a**) NOESY and (**b**) TOCSY spectra displaying the H1'–H2'/H2'' region. Each pair of intranucleotide H1'–H2'/H2'' correlations are joined by a line. There are altogether 27 pairs of cross-peaks, corresponding to the length of the oligonucleotide chain. Additional cross-peaks in (**a**) corresponding to internucleotide interactions are not observed in (**b**).

Characteristic spin connectivity between successive nucleotides. The progressive helical twist of an oligonucleotide chain gives rise to repetitive spin connectivity patterns between successive nucleotides, which are manifested as internucleotide NOE cross-peaks (framed in boxes in Figure 2.9). Such connectivity can be traced through the entire chain (Figure 2.11), only being disrupted when there is an abrupt turn in the
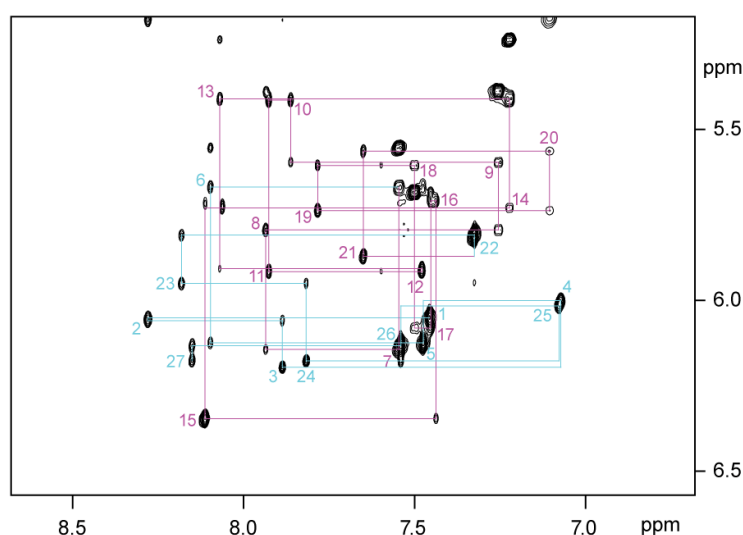
Figure 2.11 | NOE H6/H8–H1' sequential connectivity. The spin connectivity of the oligonucleotide chain can be followed from the 5'- to the 3'terminal through intra- and inter-nucleotide H8/H6–H1' NOE cross-peaks. Note the presence of four intense cross-peaks corresponding to *syn* guanine residues G1, G5, G22, and G26.

25

folded structure. Here the connectivity is illustrated with just H1', H2'/H2", and H6/H8 protons, but in practice additional connectivity patterns (either spanning the entire chain or occurring between specific dinucleotide/base-pair steps) can be traced through internucleotide through-bond/through-space correlations involving H3', H4', H5'/H5'', H1/H3, H5, and H71/H72/H73 protons as well.

NMR signatures of Watson–Crick base pairs. The imino proton involving in Watson–Crick base pair formation is hydrogen-bonded to a nitrogen atom, leading to a characteristic chemical shift ~12.5–14.5 ppm (Figures 2.5 and 2.6). Thymine H3 protons (which tend to cluster towards the 13.5–14.0 ppm end) are generally more downfield-shifted with respect to guanine H1 protons (which tend to cluster towards the 12.5–13.0 ppm end). Additionally, by forming a Watson–Crick A•T base pair, adenine H2 proton invariably approaches thymine H3 proton closely, giving rise to a signature A(H2)–T(H3) NOE cross-peak for every A•T base pair (Figure 2.12a,b). Following the same logic, two signature NOE cross-peaks corresponding to G(H1)–C(H41) and
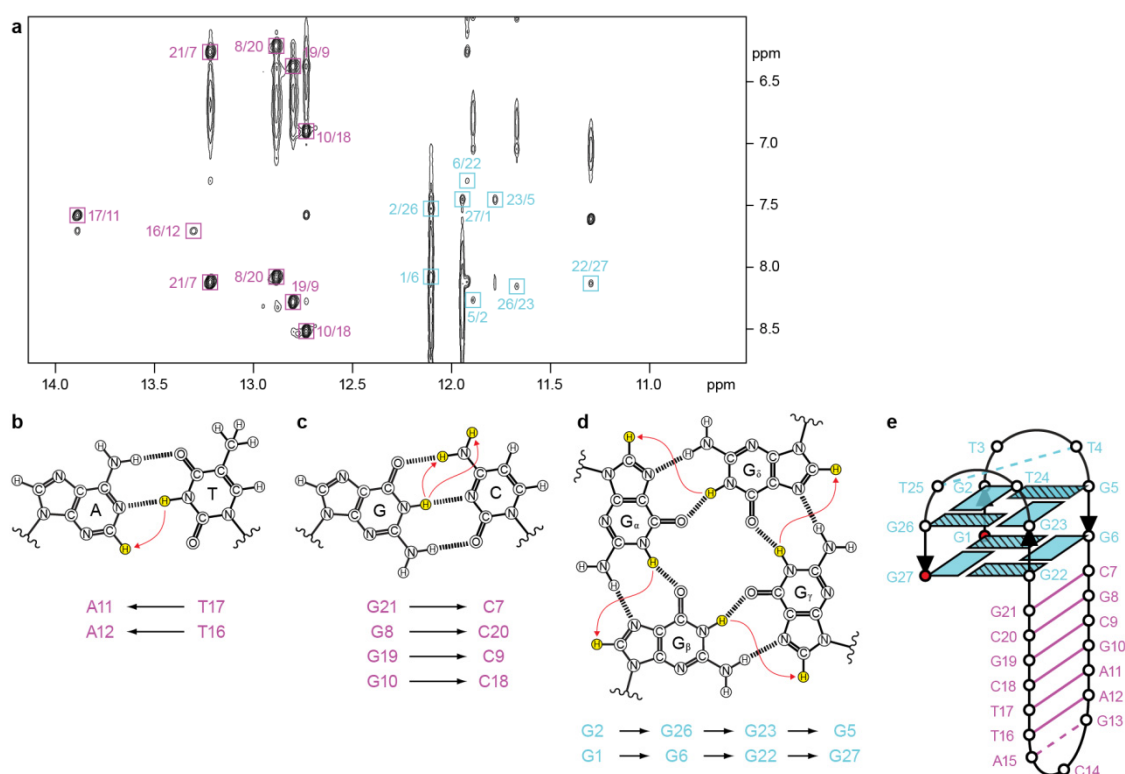
Figure 2.12 | Establishment of Watson–Crick base pairing and G-tetrad alignment. (**a**) NOESY plot showing the imino–aromatic and imino–amino region. Signature Watson–Crick base pair and G-tetrad cross-peaks are framed in magenta and cyan, respectively. (**b**) Signature T(H3)–A(H2) NOE correlation for an A•T base pair. (**c**) Signature G(H1)–C(H41/H42) NOE correlations for a G•C base pair. (**d**) Cyclic NOE connectivity patterns between H1 and H8 protons within a G-tetrad layer. (**e**) Schematic structure of quadruplex–duplex Construct I, as deduced from signature Watson–Crick base pair (**b**,**c**) and G-tetrad (**d**) cross-peaks.

G(H1)–C(H42) would be observed for every G•C Watson–Crick base pair (Figure 2.12a,c).

NMR signatures of a G-tetrad. The imino proton involving in G-tetrad formation is hydrogen-bonded to an oxygen atom, leading to a characteristic chemical shift ~10.5–12.5 ppm (Figures 2.5 and 2.6). Four imino protons partake in the establishment of a G-tetrad, resulting in four proton peaks in this region for every tetrad (assuming that all four strands constituting the quadruplex are asymmetrical). Upon forming a G-tetrad, each guanine imino proton is sufficiently close to the H8 proton of one of the neighboring guanines (but not the other guanine) to exhibit an NOE cross-peak. This leads to a cyclic imino–H8 connectivity within a G-tetrad (Figure 2.12a,d), and the directionality of the hydrogen bonds is defined following this order. Identification of this characteristic imino–H8 connectivity for each tetrad would help to establish the overall topology of the quadruplex (Figure 2.12e). For antiparallel G-quadruplexes (e.g. DNA quadruplex–duplex Construct I), some of the guanine residues would have to adopt a *syn* glycosidic conformation (in contrast to a canonical *anti* glycosidic conformation) in order to maintain the hydrogen-bond directionality of the G-tetrads. These *syn* guanine residues produce characteristic spectral patterns (e.g. very intense intranucleotide H8–H1' NOE cross-peaks; Figure 2.11) that could serve as signatures for antiparallel G-quadruplexes.

NMR structure determination for biological macromolecules. Based on data from a series of 2D NMR spectra, sufficient topological information of a biological macromolecule could be obtained towards the determination of its 3D structure. Especially useful are information from NOESY spectra, which will provide information on long-range interactions within the molecule. NOE intensity of all the assigned cross-peaks can be referenced against that of a proton–proton pair (also known as a yardstick), the distance of which does not vary, across a number of mixing times (i.e. NOE build-up). Common yardsticks employed for nucleic acids include intranucleotide cytosine H5–H6 (interatomic distance ~2.4 Å) and thymine H6–methyl (pseudo-interatomic distance ~2.95 Å) pairs. The collection of interatomic distances and additional topological information can then be subjected to structure computation [see Section 2.5] to refine the folding of the oligonucleotide chain to an ensemble of structures which would fit the dataset provided.

2.3   Ultraviolet Spectroscopy

Quantification of DNA sample. DNA bases absorb ultraviolet (UV) radiation at a maximum close to 260 nm. To measure concentration of a DNA sample, UV absorbance of the sample at 260 nm is recorded with a UV spectrophotometer. The sample is heated to 90–95 °C in a quartz cuvette (reaction volume 500 µL) so that the oligonucleotides are completely unfolded. Sample concentration can be worked out based on the formula $A = \varepsilon c l$, where A is the absorbance of the sample, ε corresponds to the molar extinction coefficient of the oligonucleotide, c stands for the oligonucleotide strand concentration, while l represents the path length of the quartz cuvette (1.0 or 0.2 cm). ε can be calculated using the online application 'UV Spectrum of DNA Calculator' (http://biophysics.idtdna.com/UVSpectrum.html).

UV-melting. Structural transition of a DNA molecule often gives rise to a change in its absorption properties. Thermal denaturation and refolding experiments have thus been employed to characterize the stability of DNA samples. UV absorbance change at 295 nm was previously shown to correlate specifically with quadruplex formation[127], and has been utilized for the characterization of quadruplex stability in this study. Analyses of UV-melting curves are performed with the assumption of a two-state transition, whereby the baselines at low temperature (corresponding to the folded state) and high temperature (corresponding to the unfolded state) are used for the normalization of transition signal to obtain the fractions of folded and unfolded structures with respect to the change in temperature. The melting temperature (T_m) is defined as the temperature for which there are equal fractions of unfolded strands and folded structures.

2.4   Circular Dichroism Spectroscopy

Ellipticity. Due to the asymmetric nature of DNA structures, when coherent left- and right-circularly polarized lights with equal intensity are passed through a sample, they are absorbed at a different extent by the sample. The initial net linear plane of light will gain in ellipticity as it passes through the sample. The resultant ellipticity varies with the wavelength of light. Hence a plot of the ellipticity across different wavelengths can be used to characterize the folding topology of a DNA structure, as G-quadruplexes having the same topology will yield CD spectra with similar characteristics.

CD signatures for various G-quadruplex topologies. CD signatures for various quadruplex topologies were extensively characterized. The all-parallel-stranded G-tetrad core produces an intense positive band around 260 nm together with a negative minimum close to 240 nm. Antiparallel G-quadruplexes, comprising at least two G-tetrads arranged in the opposite hydrogen-bond directionality, give a signature peak at ~295 nm. A detailed analysis on the spectral characteristics of different G-quadruplex topologies, as well as duplex stem elements, is presented in Section 8.2.

CD-melting. Similar to UV-melting, CD signals of a sample also change with respect to the conformation of the DNA structures. Hence the monitoring of CD signal transition at characteristic wavelengths can be utilized to follow the melting transition of a sample [see Section 8.2].

2.5   Structure Computation

Molecular mechanics. One popular approach for structure computation and simulation of molecular system is based on molecular mechanics, which describes molecular interactions using classical mechanics. In molecular mechanics, chemical bonds are represented as spring connectivity, with an associated spring (force) constant. Similarly, angular motions and dihedral rotations of chemical bonds are described by simple classical mechanics. The collection of equations and parameters describing the potential energy of the system in molecular mechanics is known as a force field.

NMR structure computation. Structure computations in this study were mostly carried out with the program XPLOR-NIH[128]. Typically, these were carried out in two stages: distance geometry simulated annealing followed by distance-restrained molecular dynamics structural refinement. In distance geometry, the molecular connectivity of the oligonucleotide chain is initially disregarded, bringing the atoms together based on the topological information and NMR restraints provided. The system is heated up to a high temperature in order to overcome potential energy barrier, and is subsequently cooled down in search of the global energy minimum that would approximate the folded structure of the chain. This coarse model was subsequently refined by the second step, which restricts the movement of atoms based on the set of NMR restraints provided. Again heating and subsequent cooling of the system is performed, and the system is then given time to equilibrate, based on molecular

dynamic simulations. The computation procedure begins with 100 conformers in the ensemble, and the 10 lowest-energy structures that best satisfy the set of NMR restraints are eventually selected.

31

G-Quadruplex Structures of Canonical Human Telomeric Repeats

3.1 G-Quadruplex & Telomere

3.2 G-Quadruplexes of Canonical Human Telomeric Repeats in Potassium and Sodium Solution

3.1   G-Quadruplex & Telomere

Telomere. The termini of linear eukaryotic chromosomes are capped by DNA–protein complexes known as telomeres[43,129-131] (Figure 3.1a). In humans, telomeric DNA consist of successive repeats (amounting to the thousands) of the sequence 5'-TTAGGG-3'/3'-AATCCC-5'[132,133]. A single-stranded (ss) 3' G-rich overhang of 100–200 nt[134-136] protrudes from the tail end of telomeric DNA (Figure 3.1b). Telomeric DNA are tightly associated with the protein complex called shelterin[137,138] (Figure 3.2a), which protects the chromosome ends from erosion and fusion events[138,139], thus maintaining the integrity of the chromosomes. The shelterin complex[137,138] comprises six subunits, namely POT1[140], Rap1[141], TIN2[142], TPP1[143-145], TRF1[146,147], and TRF2[148,149]. TRF1 and TRF2 are related proteins that attach onto the duplex portion of telomeric repeats, while POT1 binds specifically to ss TTAGGG repeats. Together, these three components ensure the localization of the shelterin complex to the telomeric ends. POT1 is linked to TRF1 and TRF2 through TIN2 and TPP1 (Figure 3.2a).

Telomeric organization. It has been proposed that the ss telomeric overhang could fold back onto the duplex segment, generating a lariat-like architecture known as the t-loop[150] (Figure 3.2b). This would occur through encroachment of the 3' overhang into the double-stranded segment and subsequent displacement of the G-rich strand to generate a D-loop[137,138]. It is believed that this concealment of the ss telomeric ends
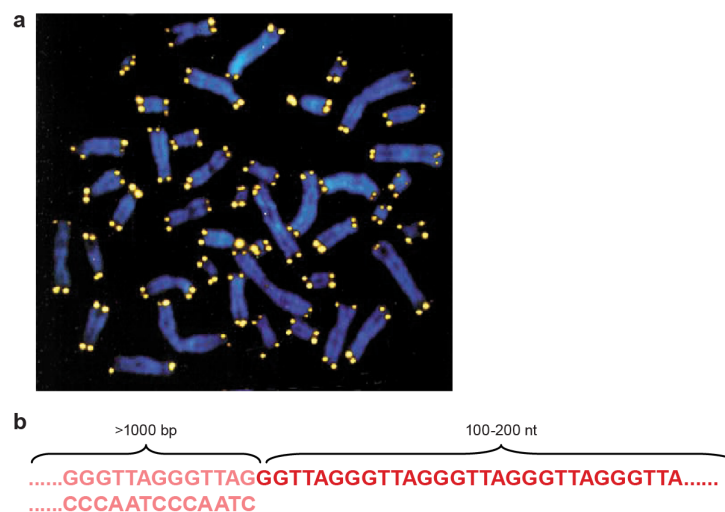
Figure 3.1 | Location and sequence composition of human telomeric DNA. a, The telomeres (yellow tips) serve as protective caps at the chromosomal ends. Figure adapted from ref 131. **b,** Human telomeric DNA are made up of successive repeats of the sequence TTAGGG[132,133], ending as a G-rich overhang[134-136].
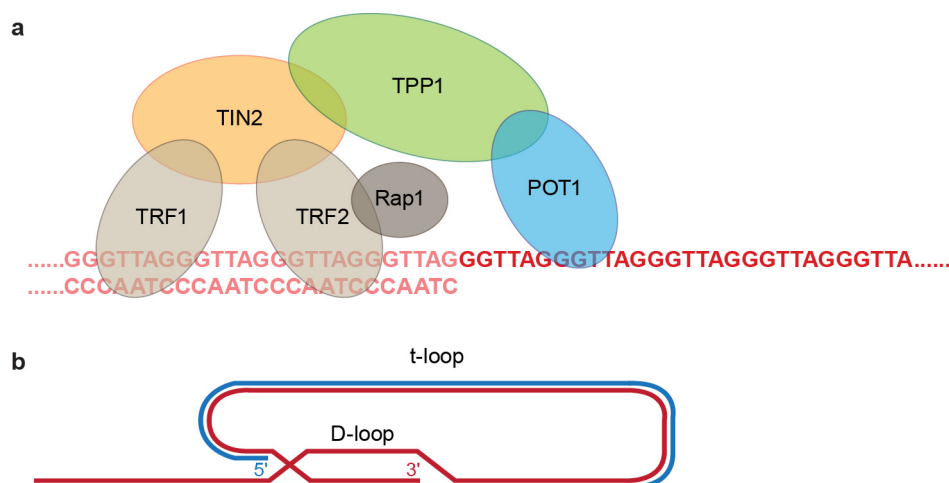
Figure 3.2 | Organization of telomeric DNA. a, The shelterin complex comprises six protein components. TRF1 and TRF2 recognize TTAGGG repeats in the double-stranded segment. POT1 recognizes telomeric DNA in the single-stranded form. **b,** Proposed architecture of the t-loop at the human telomere. The TTAGGG single-strand at the 3'-end encroaches into the double-stranded segment, generating a D-loop.

prevents them from being erroneously recognized as broken chromosomal termini by the DNA repair machinery.

Telomerase. The length of the telomere is reduced after each cell division in somatic cells[151]. Eventually, the cell undergoes apoptosis when any of the telomeres has shortened to a critical length[152,153], hence limiting the number of divisions that normal somatic cells can go through. However this mechanism is bypassed in most cancer cells through up-regulation[154,155] of the enzyme telomerase [see Section 6.1], which is a reverse transcriptase that catalyzes the addition of TTAGGG repeats onto the 3' overhang[156]. As such, down-regulation of telomerase activity[157] has been explored as a viable means to curb the growth of cancer cells. One approach involves the targeting of telomeric G-quadruplexes.

Telomeric G-quadruplexes. The G-rich ss telomeric end has an extreme preference to fold into G-quadruplexes in the unbound state, which could arise when the ss telomere-binding protein POT1 dissociates from it[158]. Adoption of the G-quadruplex conformation by telomeric DNA substrate has been demonstrated to suppress the catalytic potential of telomerase[159,160] (Figure 3.3). Hence, ligand-induced formation or stabilization of telomeric G-quadruplexes represents a promising anticancer strategy[61,92,157,161-163]. To facilitate the development of telomeric quadruplex-specific drugs, it is imperative to study possible structures that can be adopted by these telomeric repeats.
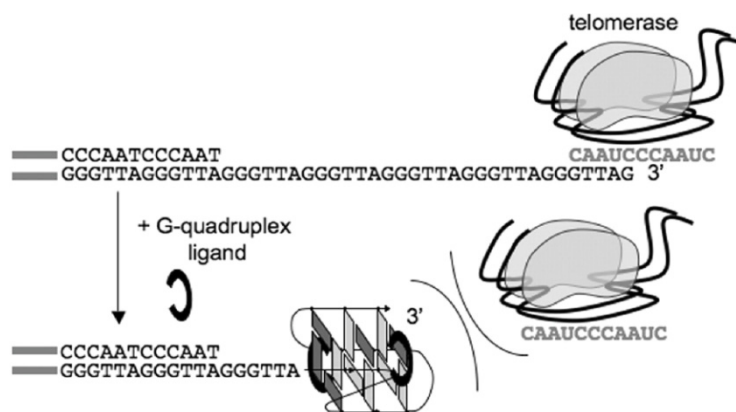
Figure 3.3 | Ligand-induced G-quadruplex formation renders the telomere inaccessible for telomerase attachment. One proposed mechanism by which the induction of telomeric G-quadruplex formation by a specific ligand exerts an inhibitory effect towards the catalytic activity of telomerase. Figure adapted from ref 160.

Previously, four-repeat human telomeric TTAGGG sequences have been extensively explored under various experimental conditions[33,162,164-190], and the high-resolution structures of four distinct G-quadruplex conformations have been resolved. In sodium solution, the d[AGGG(TTAGGG)3] sequence adopts a basket-type quadruplex structure[164] (Figure 3.4a) comprising three layers of tetrad held together by two edgewise and a diagonal loops. The identical sequence was later revealed to assume an all-parallel-stranded topology (Figure 3.4b) in a K+-containing crystal[33], with all three TTA loops taking up the propeller configuration.

As K+ ions exist at a higher concentration within the cells as compared to Na+ ions, structures in potassium solution of human telomeric sequences comprising four TTAGGG repeats were extensively studied. However, structural elucidation of these sequences was complicated by their conformational heterogeneity in K+ solution. This
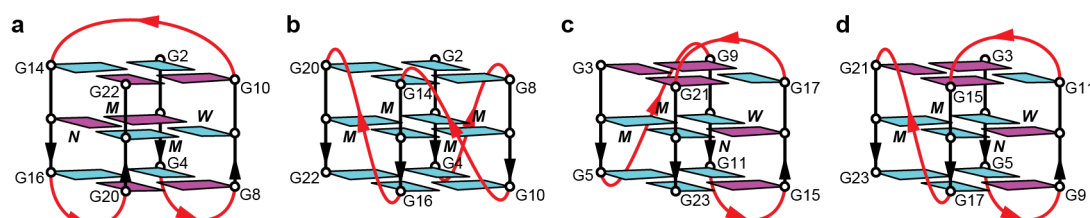
Figure 3.4 | Diverse G-quadruplex topologies of four-repeat human telomeric sequences. (**a–d**) Schematic diagrams of G-quadruplex topologies adopted by telomeric TTAGGG repeats: (**a**) basket-type structure of d[AGGG(TTAGGG)3] in sodium solution[164], (**b**) propeller-type structure of d[AGGG(TTAGGG)3] in crystallized form[33], (**c**) Form 1 (3+1) architecture adopted by d[TAGGG(TTAGGG)3] in potassium solution[167], and (**d**) Form 2 (3+1) architecture adopted by d[TAGGG(TTAGGG)3TT] in potassium solution[167]. Guanine residues in *syn* and *anti* glycosidic conformations are shown as magenta and cyan rectangles, respectively.

multiplicity has been mitigated/eliminated by modifying the nucleotides at the ends of the oligonucleotides[165-167,172-174] and/or by replacing tetrad guanine(s) with constituting guanine analogues[167,169-171], hence driving the folding of the oligonucleotides towards a single major species. By virtue of this approach, two related G-quadruplex conformations comprising the (3+1) G-tetrad scaffold were discovered[165-167,169-174] (Form 1 and 2; Table 3.1 and Figure 3.4c,d). Both conformations comprise one propeller and two lateral loops, but they vary in the procession of these loops[166].

While not fully addressed, this structural polymorphism has also been implicated in Na$^+$ solution[175,178,181], yet the basket-type form has largely been used as the sole reference structure for interpretation of studies under Na$^+$-containing environment. Knowledge on the structural disparities, if any, in the presence of Na$^+$ ions would contribute towards fundamental understanding of the cationic influence on human telomeric repeats under different contexts.

Here the NMR structures of two novel quadruplex topologies adopted by human telomeric sequences in potassium [see Section 3.2.1] and sodium [see Section 3.2.2] solution are presented. These structures further establish the structural polymorphism of human telomeric DNA across diverse environments, and build on the available pool of structural motifs for targeting of these important entities.

3.2 G-Quadruplexes of Canonical Human Telomeric Repeats in Potassium and Sodium Solution

3.2.1 A Two-Tetrad Basket-Type G-Quadruplex Formed in Potassium Solution[#]

A novel G-quadruplex topology (Form 3) of human telomeric sequences in potassium solution. The natural human telomeric sequence d[GGG(TTAGGG)$_3$T] (denoted as natural Form 3; Table 3.1) in potassium solution produced NMR H1 proton spectra showing eight major peaks from 11.4 to 11.9 ppm and two upfield-shifted peaks from 10.5 to 10.8 ppm (Figure 3.5a). This implicated its adoption of a novel G-quadruplex fold (estimated population ~60%) comprising only two G-tetrads. Substitution of G7 by 8-bromoguanine (^{Br}G), yielding the sequence d[GGGTTA(^{Br}G)GG(TTAGGG)$_2$T] (denoted as ^{Br}G7-Form 3; Table 3.1), boosted the population of this conformation to ~90% (Figure 3.5b).

Table 3.1 | Natural and modified human telomeric DNA sequences used in this study (K⁺ solution).[a]

Name	Sequence					T_m[b] (°C)
natural Form 1[c]	TA GGG TTA GGG	TTA GGG	TTA GGG			53.6
^{Br}G16-Form 1[c]	TA GGG TTA GGG	TTA G**^{Br}G**G	TTA GGG			57.5
natural Form 2[c]	TA GGG TTA GGG	TTA GGG	TTA GGG TT			47.2
^{Br}G15-Form 2[c]	TA GGG TTA GGG	TTA **^{Br}G**GG	TTA GGG TT			54.0
natural Form 3	GGG TTA GGG	TTA GGG	TTA GGG T			57.0
^{Br}G7-Form 3	GGG TTA **^{Br}G**GG	TTA GGG	TTA GGG T			62.8

[a] 8-bromoguanine (^{Br}G) residues are shown in bold. [b] Melting temperature in 20 mM K⁺ solution, measured by following 295-nm UV absorbance. [c] These sequences were studied in ref 167.

CD and UV-melting investigation of Form 3 human telomeric G-quadruplex. CD spectra of three natural human telomeric sequences (Form 1, 2, and 3) and their respective ^{Br}G-modified counterparts (Table 3.1) were measured in K⁺ solution. CD spectra of all six sequences showed positive maxima near 295 nm, indicative of antiparallel G-quadruplexes[191], as well as a shoulder band at ~270 nm. Form 3 (Figure 3.6a) showed an additional positive peak at 250 nm, and the 270-nm band has a lower intensity than those of Form 1 and 2 (3+1) G-quadruplexes (Figure 3.6b,c).

Thermal stability of these sequences was investigated by monitoring the UV absorbance at 295 nm[127] (Figure 3.6d,e). Melting temperature (T_m) of natural Form 3 was 57.0 °C in 20 mM K⁺, whereas T_m of natural Form 1 and 2 were 53.6 °C and 47.5 °C, respectively (Table 3.1). The ^{Br}G-modified sequences exhibit T_m ~4–7 °C higher than their respective natural counterparts. Again, the T_m of ^{Br}G-modified Form 3 is greater than the T_m of ^{Br}G-modified Form 1 and 2. These findings were surprising,
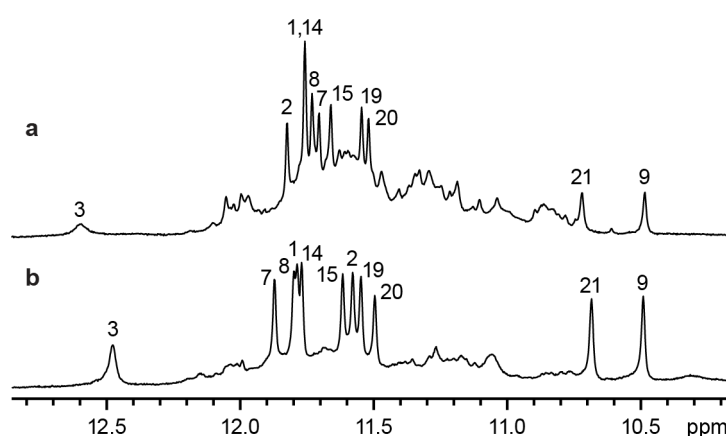
Figure 3.5 | 1D proton NMR spectra of Form 3 human telomeric G-quadruplex in potassium solution. (**a**,**b**) NMR spectra showing the H1 protons of natural (**a**) and ^{Br}G7-modified (**b**) Form 3 human telomeric sequences. Peaks from the major form are marked with their respective assignments.
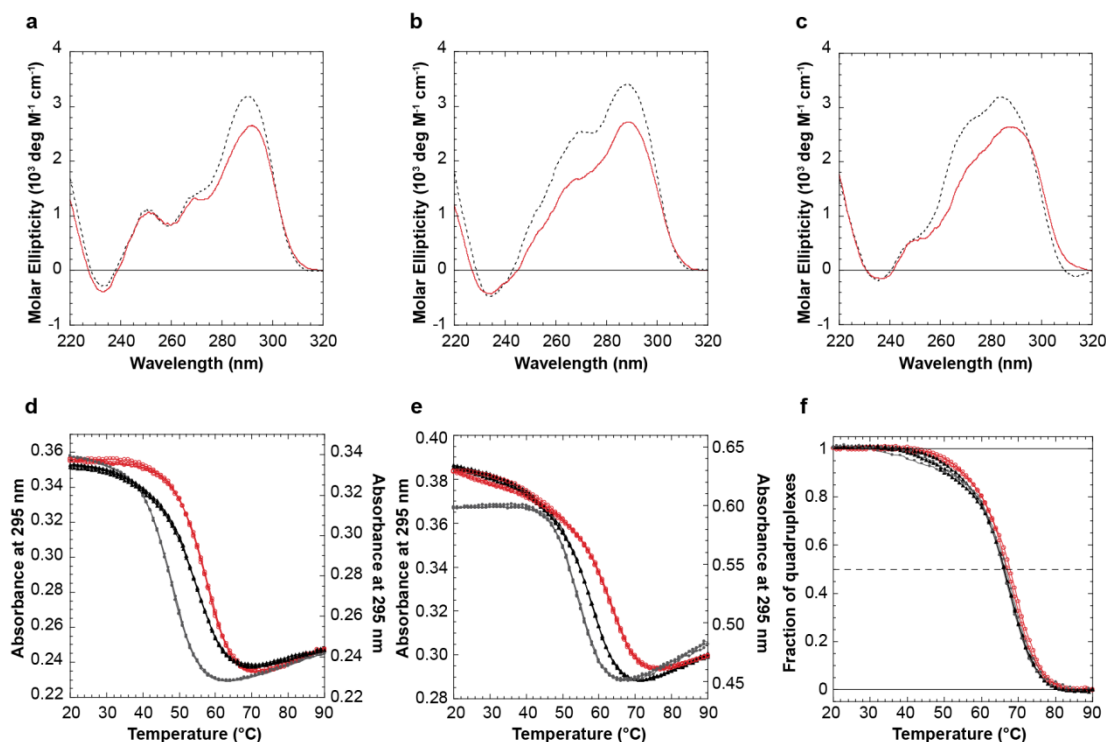
Figure 3.6 | CD and UV-melting studies of four-repeat natural and ^{Br}G-modified human telomeric sequences in potassium solution. (**a**–**c**) CD spectra of human telomeric G-quadruplex Form 3 (**a**), Form 2 (**b**), and Form 1(**c**). Black and red curves represent natural and ^{Br}G-substituted sequences, respectively. (**d**,**e**) UV-melting profiles of the natural (**d**) and ^{Br}G-substituted (**e**) sequences. The left axis corresponds to Form 1 and Form 2, while the right axis corresponds to Form 3. Form 1 and 2 are represented in continuous and dotted lines, respectively, while Form 3 is represented in red circles. (**f**) Plot of the fractions of folded ^{Br}G7-Form 3 G-quadruplex against temperature. The T_m remained unchanged at DNA strand concentrations of 250 μM (red circles), 50 μM (dotted line), and 10 μM (continuous line).

since Form 3 consists of just two G-tetrad layers, as compared to three tetrad layers for both Form 1 and 2. At a fixed concentration of K$^+$, T_m of ^{Br}G7-Form 3 remained unchanged upon up to 25-fold difference in DNA strand concentrations (10–250 μM; Figure 3.6f), indicating that it corresponds to an intramolecular G-quadruplex.

NMR spectral assignments. High NMR spectral correspondence was observed between the major form of natural and ^{Br}G7-modified Form 3 (Figure 3.5), indicating their adoption of the same G-quadruplex scaffold. This similarity facilitated the assignments for both cases; resonances of ^{Br}G7-Form 3 could be deduced based on spectral resemblance between the two, while resonances for natural Form 3 could be determined/corroborated based on the cleaner spectra of ^{Br}G7-Form 3.

Unambiguous assignments of guanine (for natural Form 3; Figure 3.7) and thymine (for ^{Br}G7-Form 3; Figure 3.8) imino protons were achieved based on site-specific 2% ^{15}N-enrichment of samples[192], one residue at a time. Mutual through-bond correlations
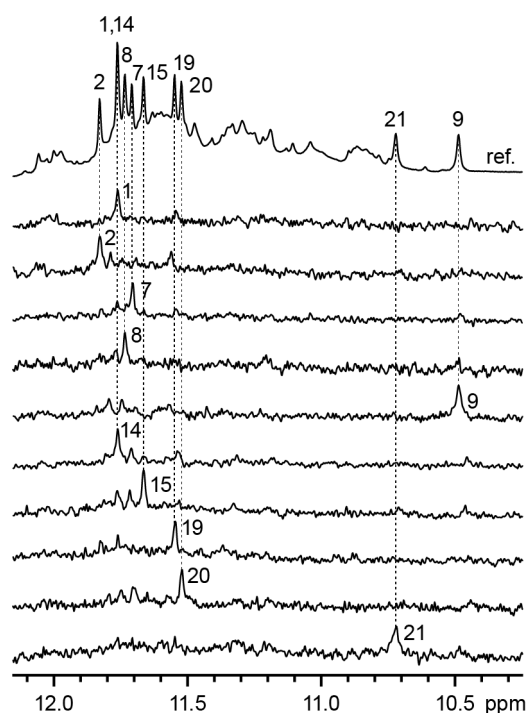
Figure 3.7 | Assignment of guanine H1 protons for natural Form 3 human telomeric sequence. Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments.

of both H1 and H8 protons of guanine to [13]C5[193] (Figure 3.9) assisted in the assignment of the H8 protons. Assignments of thymine residues were further corroborated by individual T-to-U modifications. Based on these unambiguous resonance assignments, as well as additional information from ^{13}C–^{1}H-HSQC, TOCSY, and COSY, sequential H6/H8–H1' NOE connectivity of the DNA strand could be established from the 5'- to the 3'-terminal (Figure 3.10). Spectral correspondence between natural and [Br]G7-
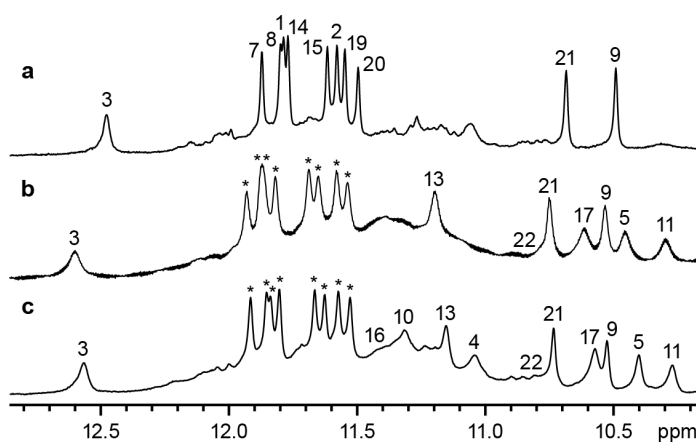
Figure 3.8 | Assignment of guanine and thymine imino protons of [Br]G7-Form 3. (a) Experimental condition: temperature, 25 °C; pH 7. **(b)** Experimental condition: temperature, 10 °C; pH 7. **(c)** Experimental condition: temperature, 10 °C; pH 5.
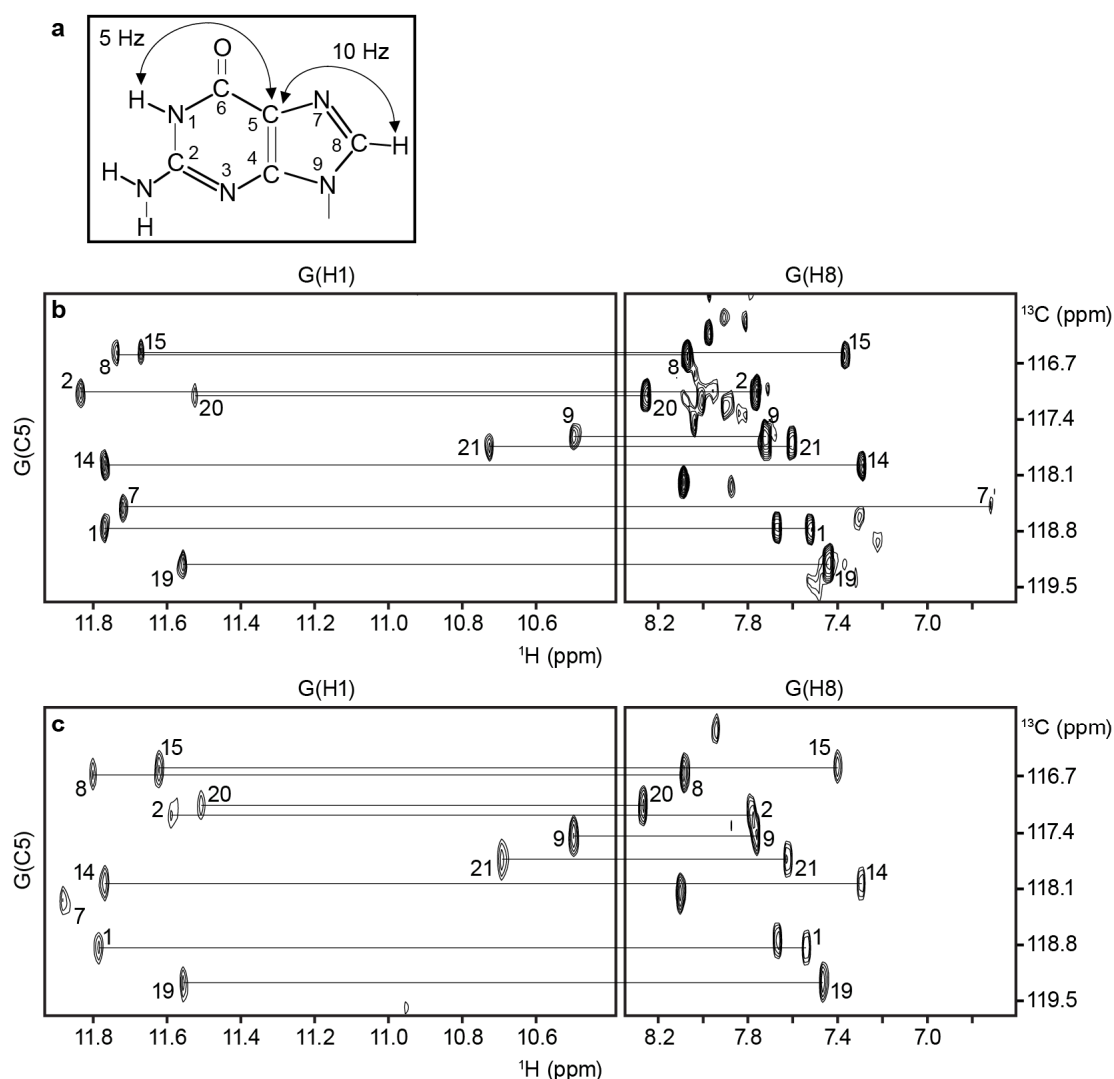
Figure 3.9 | Through-bond connectivity of guanine H1 and H8 protons for Form 3 human telomeric sequences. (**a–c**) JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to [13]C5 (**a**), for natural (**b**) and [Br]G7-modified (**c**) Form 3 human telomeric sequences.

modified Form 3 was clearly evident in Figures 3.5, 3.9, and 3.10. Intense H8–H1' NOE cross-peaks of G1, G7, G14, and G19 (Figure 3.11) pointed to their adoption of the *syn* glycosidic conformation, whereas the rest of the guanine residues assume the *anti* conformation.

A two-tetrad basket-type G-quadruplex. H1–H8 NOE connectivity within each tetrad was determined for both natural and [Br]G7-modified Form 3, revealing the adoption of a basket-type G-quadruplex (Figure 3.12). The two G-tetrads constituting the core (G2•G7•G19•G15 and G1•G14•G20•G8) display opposing directionalities, and guanines around both tetrads exhibit *syn•syn•anti•anti* glycosidic conformations. All four strands constituting the core have both an antiparallel and a parallel
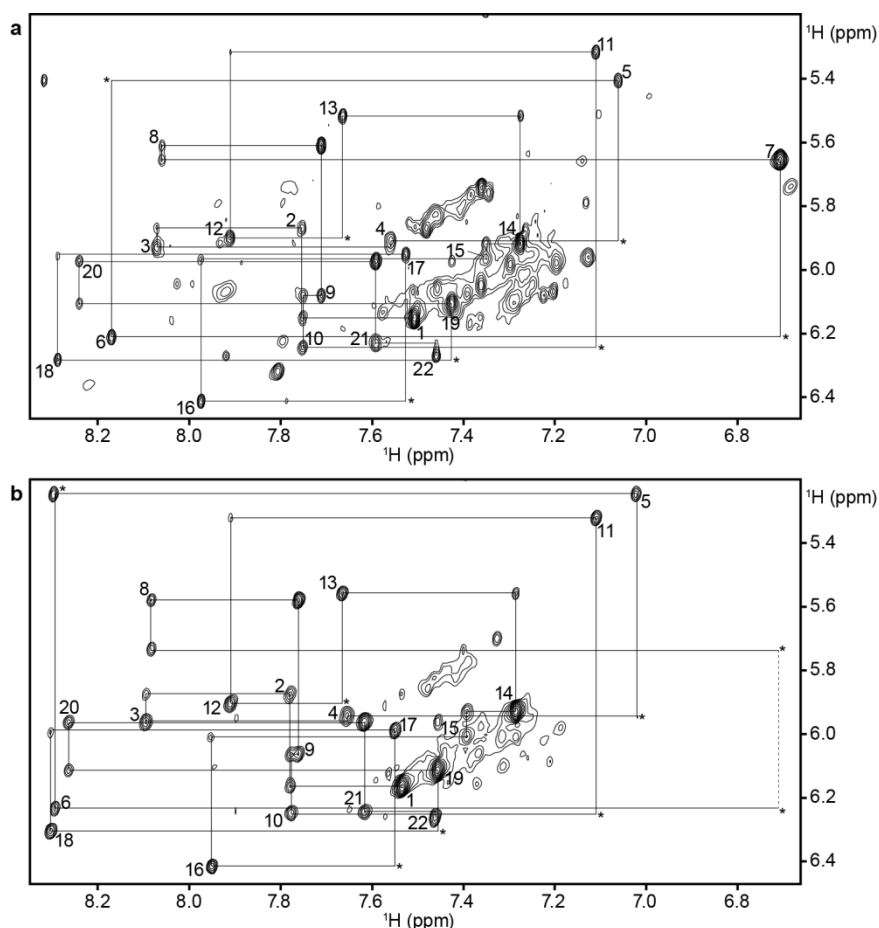
Figure 3.10 | NOESY spectra showing the sequential connectivity of Form 3 human telomeric sequences. (a,b) H6/H8–H1' NOE connectivity of natural (**a**) and [Br]G7-modified (**b**) Form 3 human telomeric sequences from G1 to T22. Residue numbers are marked on the corresponding intraresidue cross-peaks. Asterisks indicate cross-peaks that are weak or missing. The connectivity of [Br]G7-Form 3 (**b**) is disrupted at Residue 7 (indicated with a dotted line) since G7(H8) was substituted by a bromine.
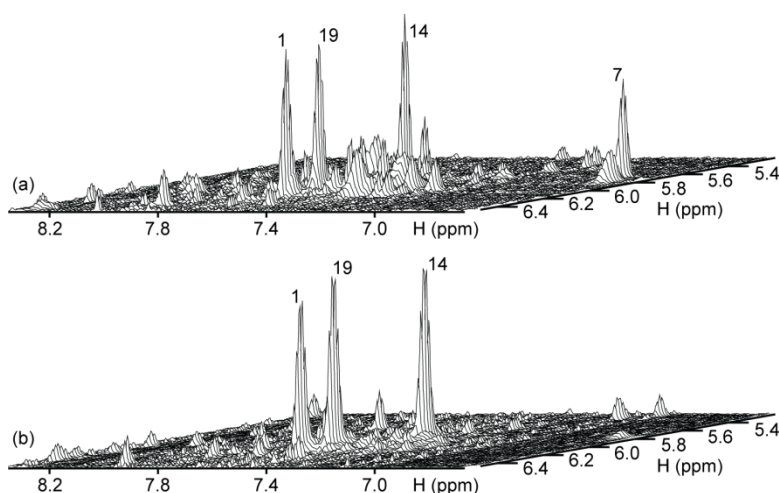
Figure 3.11 | NOESY spectra of Form 3 in a stacked plot view. (a,b) Stacked plot of the NOESY spectra highlighting the intense intraresidue H8–H1' NOE corresponding to *syn* guanine residues for the natural (**a**) and [Br]G7-modified (**b**) Form 3 human telomeric sequences. The intraresidue H8–H1' cross-peak for G7 is missing in [Br]G7-Form 3 since G7(H8) has been replaced by a bromine.
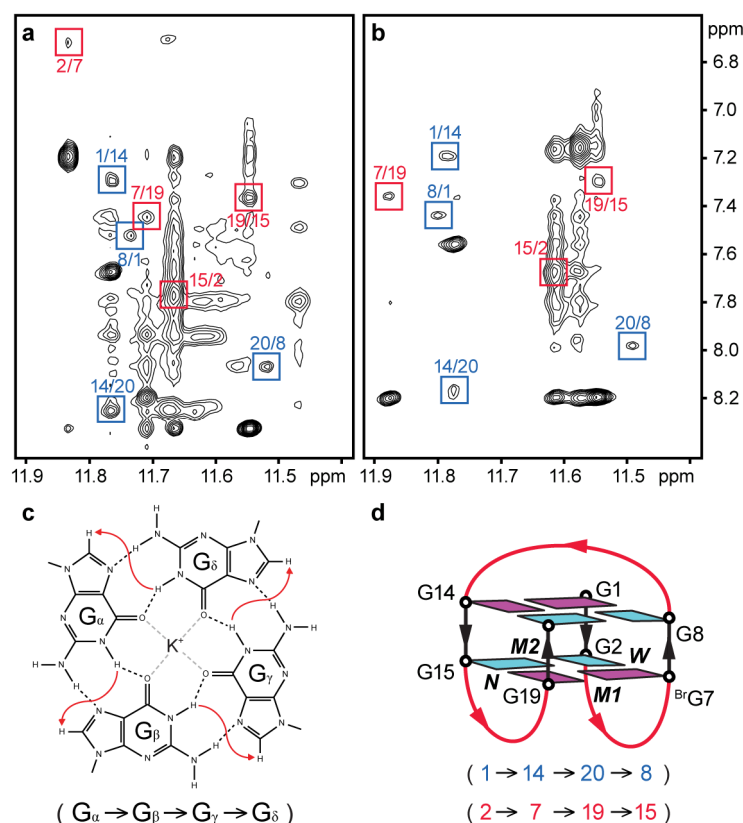
Figure 3.12 | Folding topology of Form 3 in potassium solution. (**a–c**) NOESY plot of natural (**a**) and ^{Br}G7-modified (**b**) Form 3 human telomeric sequences highlighting cross-peaks corresponding to cyclic guanine H1–H8 NOE connectivity (**c**) within a tetrad. (**d**) Schematic diagram of human telomeric G-quadruplex Form 3, which exhibits hydrogen-bond directionalities of tetrads as shown below.

neighboring strand, and the core is held up by two edgewise loops and a middle diagonal loop.

This new folding principle of Form 3 challenges the previous belief that an oligonucleotide sequence with four stretches of three consecutive guanines would naturally adopt a three-tetrad G-quadruplex structure. It is worth pointing out that even though Form 3 comprises just two tetrad layers, it actually has a higher stability than both human telomeric G-quadruplex Form 1 and 2 (Table 3.1), both of which contain three G-tetrad layers. Note however that the most stable conformation adopted by different sequences should not necessarily be the same, even if they are made up of the same repeat type (i.e. TTAGGG for human telomeric DNA).

NMR structure of Form 3 in potassium solution. The K$^+$ solution structure of Form 3 human telomeric G-quadruplex (Figure 3.13), consisting of an antiparallel (2+2) basket-type core, was computed based on NMR restraints (Appendix I). The tetrad bases of the core are structurally more converged than the nucleotides residing within
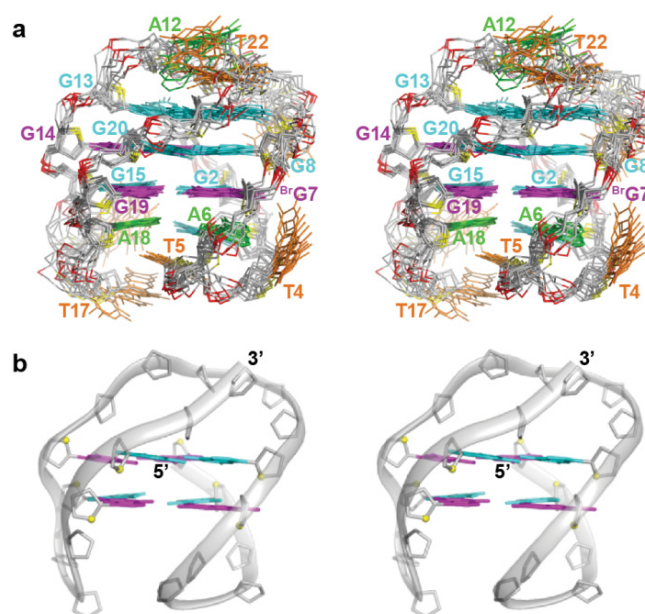
Figure 3.13 | NMR structure of Form 3 in potassium solution. (**a**) Stereo view of ten computed structures of ^{Br}G7-modified Form 3 G-quadruplex, aligned based on the tetrad core. (**b**) Stereo view of a representative conformer in ribbon representation.

the loops. The five-membered rings of guanines from the two tetrads exhibit extensive stacking (Figure 3.14d), the pattern of which is peculiar to successive G-tetrad layers having anticlockwise and clockwise hydrogen-bond directionalities[164]. The groove widths are successively wide, medium (*M1*), narrow, and medium (*M2*) (Figure 3.12d). The wide groove arises from the antiparallel arrangement of the two adjacent strands G7–G8 and G1–G2, whereas the narrow groove results from the opposite alignment of the two adjacent strands G19–G20 and G14–G15 in the reverse orientation. The medium grooves are generated from the parallel alignment of two adjacent strands, (G14–G15/G1–G2) and (G19–G20/G7–G8).

Extensive base interactions in the loops. G-tetrad core of Form 3 is capped at the top and bottom by base triples (Figures 3.13 and 3.14). Formation of the top base triple (Figure 3.14b,c), which consists of G21, G9, and G13, would account for G21(H1) and G9(H1) at around 10.5 ppm (Figure 3.5); G21(H1) establishes a hydrogen bond with G9(O6), while G9(H1) establishes a hydrogen bond with G13(O6) (Figure 3.14b,c). The base triple at the bottom comprises G3, A18, and A6 (Figure 3.14e,f), consistent with the observation of G3(H1) (Figures 3.5 and 3.8) and NOE interactions between these residues and the G2•G7•G19•G15 tetrad at the bottom. The additional establishment of a top T11•T22 base pair (Figure 3.14b) and a bottom T5•T17 base pair (Figure 3.14f)
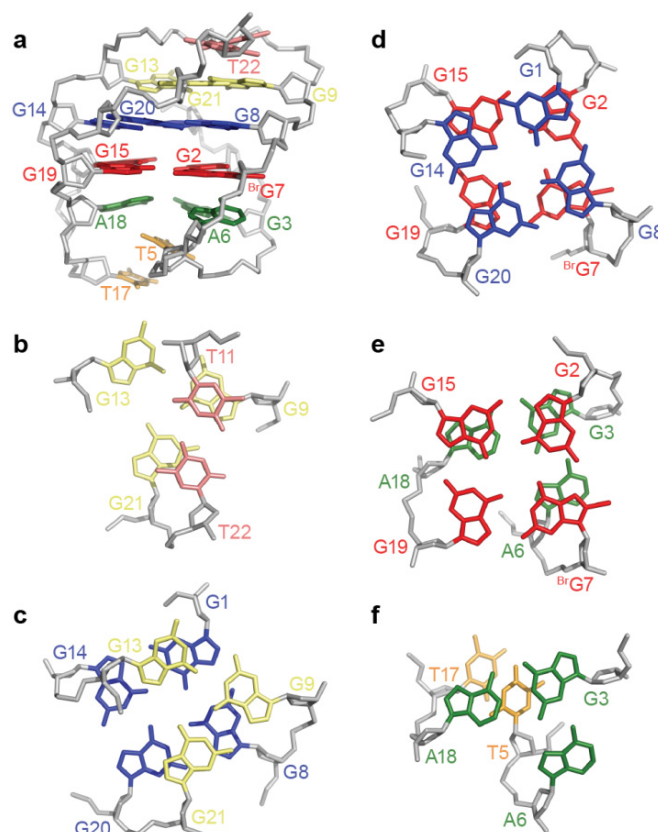
Figure 3.14 | Various stacking layers found in Form 3 human telomeric G-quadruplex. (**a**) The overall structure is stabilized by up to six stacking layers. (**b**) T11•T22 base pair and G21•G9•G13 triple. (**c**) G21•G9•G13 triple and G1•G14•G20•G8 tetrad. (**d**) G1•G14•G20•G8 and G2•^{Br}G7•G19•G15 tetrads. (**e**) G2•^{Br}G7•G19•G15 tetrad and G3•A18•A6 triple. (**f**) G3•A18•A6 triple and T5•T17 base pair.

would agree with the presence of these thymine H3 protons at ~10.5 ppm (Figure 3.8b,c). The T11•T22 base pair is further stacked over by the base of A12 (Figure 3.13).

Base pairing/stacking within loop elements was observed to build on the stability of quadruplexes[167]. In the case of Form 3 human telomeric G-quadruplex, it comprises a total of four to six layers of tetrad/base triple/base pair (Figure 3.14), which could have contributed significantly towards its stabilization. This stabilizing influence based on the stacking of loop elements was previously seen in a dimeric quadruplex adopted by the sequence d[A$_2$G$_2$T$_4$A$_2$G$_2$][194], which comprises two base pairs, two triads, and two tetrads. This novel structure of Form 3 would indicate that the major folding topology of a particular sequence is favored through maximizing the number of stabilizing interactions both in the tetrad core and the loops.

The folding of thymine (T4, T10, and T16) bases into their respective grooves (Figure 3.15) could have added onto its overall stability. Imino protons of these thymine residues, which would experience similar chemical environments, were observed
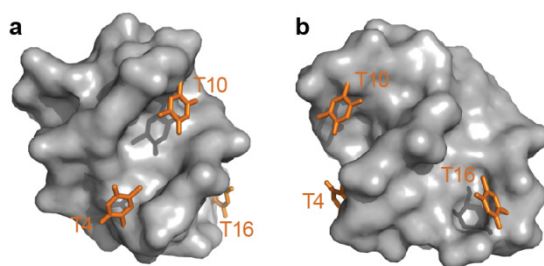
Figure 3.15 | Folding of thymine bases into the G-quadruplex grooves. (a,b) Surface view of Form 3 human telomeric G-quadruplex from the wide (**a**) and medium (*M2*) grooves (**b**). Thymine (T4, T10, and T16) bases are shown in stick representation.

around 11–11.5 ppm at pH 5 under low temperatures (Figure 3.8c). Such base conformations have also been observed in other G-quadruplex structures[164,167].

3.2.2 An Antiparallel (2+2) G-quadruplex Scaffold Formed in Sodium Solution[#]

Diverse G-quadruplex conformations in Na$^+$ solution. NMR spectra of natural human telomeric sequences comprising four TTAGGG repeats with varying dangling termini (Table 3.2) in Na$^+$ solution pointed to the adoption of multiple forms (Figure 3.16). In agreement with the previous study[164], the sequence d[AGGG(TTAGGG)$_3$]

Table 3.2 | Human telomeric DNA sequences used in this study (Na$^+$ solution).[a]

Name	Sequence	T_m[b] (°C)
4TTA01	GGG TTA GGG TTA GGG TTA GGG	–
4TTA02	GGG TTA GGG TTA GGG TTA GGG T	–
4TTA03	GGG TTA GGG TTA GGG TTA GGG TT	–
4TTA04	GGG TTA GGG TTA GGG TTA GGG TTA	–
4TTA05	A GGG TTA GGG TTA GGG TTA GGG	57.6
4TTA06	A GGG TTA GGG TTA GGG TTA GGG T	–
4TTA07	A GGG TTA GGG TTA GGG TTA GGG TT	–
4TTA08	A GGG TTA GGG TTA GGG TTA GGG TTA	–
4TTA09	TA GGG TTA GGG TTA GGG TTA GGG	–
4TTA10	TA GGG TTA GGG TTA GGG TTA GGG T	–
4TTA11	TA GGG TTA GGG TTA GGG TTA GGG TT	–
4TTA12	TA GGG TTA GGG TTA GGG TTA GGG TTA	–
4TTA13	TTA GGG TTA GGG TTA GGG TTA GGG	–
4TTA14	TTA GGG TTA GGG TTA GGG TTA GGG T	–
4TTA15	TTA GGG TTA GGG TTA GGG TTA GGG TT	–
4TTA16	TTA GGG TTA GGG TTA GGG TTA GGG TTA	44.5
4TTA16[Br22]	TTA GGG TTA GGG TTA GGG TTA **^{Br}G**GG TTA	50.8

[a] 8-bromoguanine (^{Br}G) is shown in bold. [b] Melting temperature in ~100 mM Na$^+$, measured by following 295-nm UV absorbance.

(denoted as *4TTA05*; Table 3.2) showed twelve H1 proton peaks from ~10.6–12 ppm (Figure 3.17a, black), which correspond to the establishment of the basket-type quadruplex (Figure 3.4a). Furthermore, two sequences (i.e. *4TTA06, 4TTA07*) with the same 5'-flanking end as *4TTA05* showed a major species with a highly similar spread of imino proton peaks (Figure 3.16f,g), suggesting their adoption of the same topology as the prevailing form. The other sequences mostly displayed the coexistence of two or more major species, judging from the intensity and number of H1 proton peaks. All sixteen sequences displayed CD spectra comprising two positive bands near 250 and 295 nm, together with a trough around 260–270 nm (Figure 3.18). The 295-nm peak is
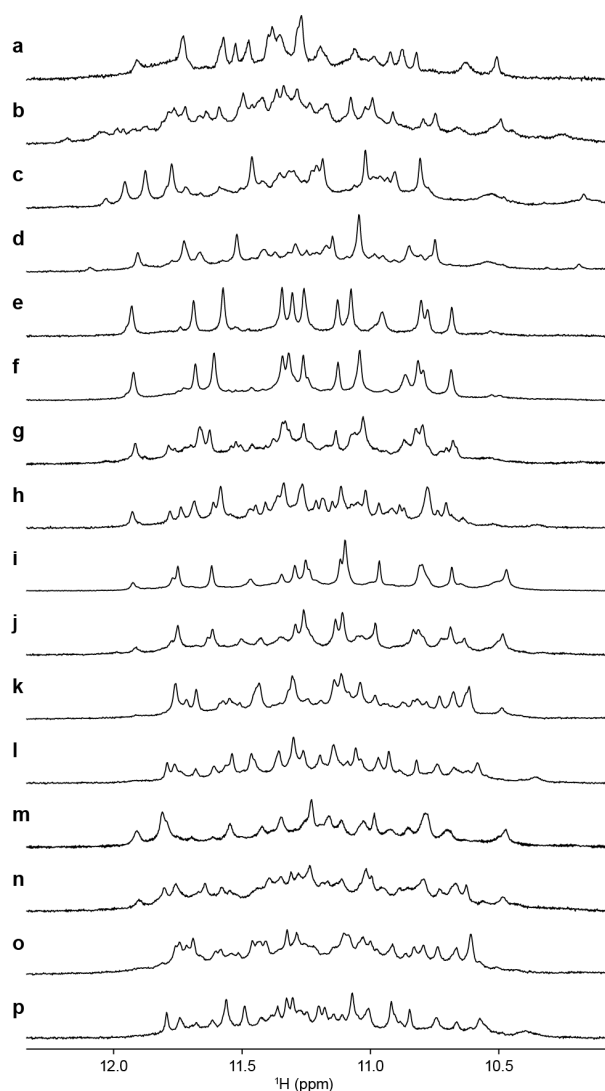
Figure 3.16 | NMR H1 proton spectra of natural human telomeric sequences comprising four TTAGGG repeats in Na$^+$ solution. (**a**) GGG(TTAGGG)$_3$, (**b**) GGG(TTAGGG)$_3$T, (**c**) GGG(TTAGGG)$_3$TT, (**d**) GGG(TTAGGG)$_3$TTA, (**e**) AGGG(TTAGGG)$_3$, (**f**) AGGG(TTAGGG)$_3$T, (**g**) AGGG(TTAGGG)$_3$TT, (**h**) AGGG(TTAGGG)$_3$TTA, (**i**) TAGGG(TTAGGG)$_3$, (**j**) TAGGG(TTAGGG)$_3$T, (**k**) TAGGG(TTAGGG)$_3$TT, (**l**) TAGGG(TTAGGG)$_3$TTA, (**m**) (TTAGGG)$_4$, (**n**) (TTAGGG)$_4$T, (**o**) (TTAGGG)$_4$TT, and (**p**) (TTAGGG)$_4$TTA.
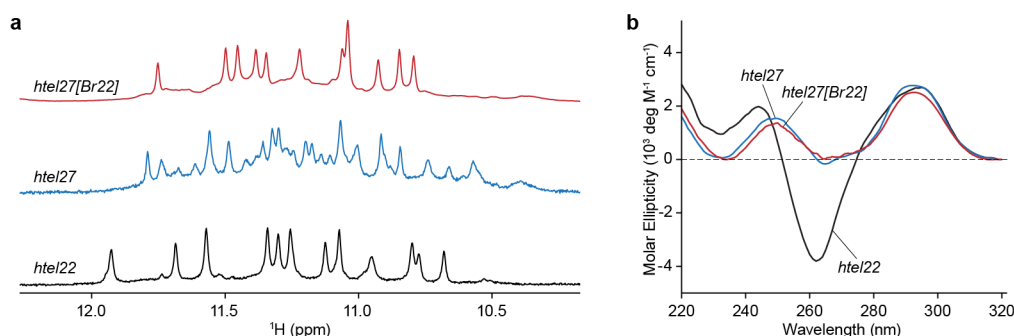
Figure 3.17 | NMR and CD spectra of human telomeric sequences comprising four TTAGGG repeats in Na$^+$ solution. (**a**,**b**) 1D NMR H1 proton (**a**) and CD (**b**) spectra of *4TTA05* (black), *4TTA16* (blue), and *4TTA16[Br22]* (red).

characteristic of opposite-polarity stacking of G-tetrads[195], suggesting that these sequences largely conform to antiparallel G-quadruplexes in Na$^+$ solution. Note that *4TTA05* produced a sharp negative trough at ~260 nm (Figure 3.17b, black), which was less intense in the other sequences.

Formation of a G-quadruplex with an alternative conformation distinct from *4TTA05*. We chose to probe the structures present in the sequence [(TTAGGG)$_4$TTA] (denoted as *4TTA16*; Table 3.2) (Figure 3.17, blue), which has the largest number of flanking nucleotides for potential interactions. Previously, ^{Br}G-for-G substitution has been utilized to drive the folding of an oligonucleotide towards a single major G-quadruplex structure by favoring the *syn* glycosidic conformation at the substituted residue[167,169,170]. Here single ^{Br}G incorporation at position G22 of *4TTA16*, giving rise to the sequence d[(TTAGGG)$_3$TTA(^{Br}G)GGTTA] (denoted as *4TTA16[Br22]*;
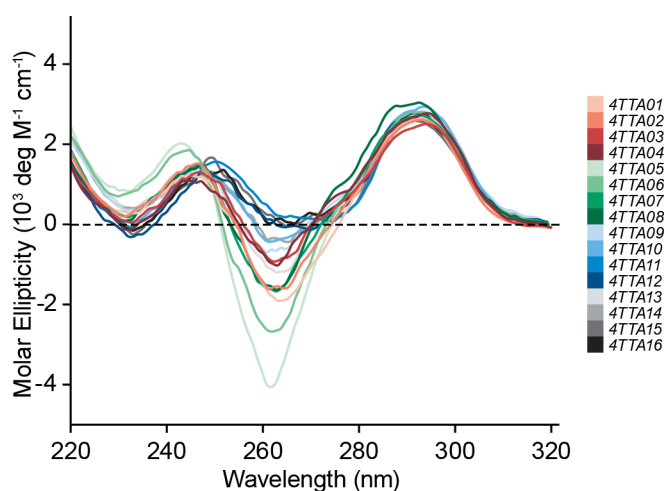
Figure 3.18 | CD spectra of four-repeat natural human telomeric sequences. CD spectra of *4TTA01, 4TTA02, 4TTA03, 4TTA04, 4TTA05, 4TTA06, 4TTA07, 4TTA08, 4TTA09, 4TTA10, 4TTA11, 4TTA12, 4TTA13, 4TTA14, 4TTA15,* and *4TTA16* in Na$^+$ solution. Color code is shown on the right.

46

Table 3.2), resulted in the emergence of a major conformation (>85%) amenable for detailed structural characterization (Figure 3.17a, red). *4TTA16[Br22]* showed twelve major imino proton peaks from ~10.8–11.8 ppm, pointing to its adoption of a three-layered quadruplex. *4TTA16[Br22]* and *4TTA16* produced highly similar CD spectra (Figure 3.17b), suggesting that *4TTA16[Br22]* could correspond to one of the pre-existing conformations in *4TTA16*, or that the core topologies of *4TTA16[Br22]* and *4TTA16* were closely related. Their CD spectra were substantially different from that of *4TTA05*, notably the absence of the sharp negative trough at ~260 nm, pointing to the adoption of an alternative folding topology. Melting analyses of the three sequences were carried out by following their UV absorbance at 295 nm[127]. Melting temperatures (T_m) of *4TTA05*, *4TTA16*, and *4TTA16[Br22]* were found to be 57.6, 44.5, and 50.8 °C, respectively, in ~100 mM Na$^+$ (Figure 3.19). The ^{Br}G-substituted oligonucleotide displayed a T_m ~6 °C higher than that of the natural counterpart, consistent with observations from previous studies[169] [see Section 3.2.1].

NMR spectral assignments. To proceed with the structural elucidation of *4TTA16[Br22]*, guanine imino (Figure 3.20a) and adenine/guanine H8 (Figure 3.20b) protons were assigned unambiguously based on site-specific 2% ^{15}N-enrichment of samples[192] and ^{2}H labeling, respectively. Assignments of H1 and H8 protons of guanines were further corroborated by through-bond correlations at natural abundance[193] (Figure 3.20c). Thymine residue assignments were also complemented with thymine-to-uracil modifications. Sequential H6/H8–H1' NOE connectivity of *4TTA16[Br22]* (Figure 3.21) was completed with the assistance of ^{13}C-^{1}H HSQC,
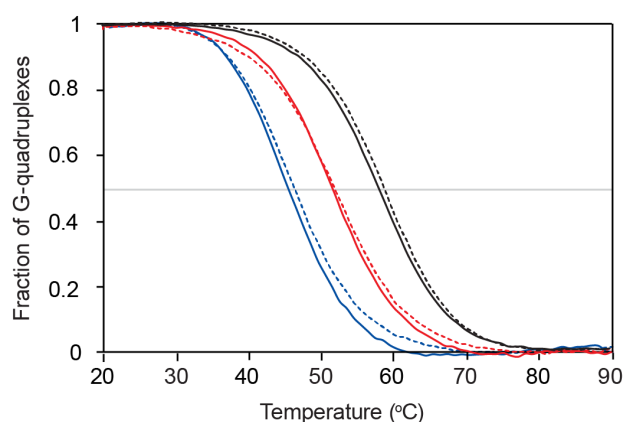
Figure 3.19 | Plot of the fractions of folded structure against temperature for intramolecular human telomeric G-quadruplexes. Heating (dotted lines) and cooling (solid lines) curves for *4TTA05* (black), *4TTA16* (blue), and *4TTA16[Br22]* (red), monitored by 295-nm UV absorbance.

Figure 3.20 | Assignment of H1 and H8 protons for *htel27[Br22]*. (a) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. **(b)** Assignment of adenine and guanine H8 protons for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. **(c)** JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to ^{13}C5. Weak/missing peaks are marked by boxes.

TOCY, and COSY. Intense intranucleotide H8–H1' NOE for G4, G5, G10, G16, and G23 (Figure 3.22) indicated that these guanine residues, together with ^{Br}G22, adopt the *syn* glycosidic conformation. All other guanine residues adopt the *anti* glycosidic conformation.

***4TTA16[Br22]* adopts a new antiparallel (2+2) G-quadruplex folding topology in Na$^+$ solution.** Having determined the assignments of guanine H1 and H8 protons for *4TTA16[Br22]*, the alignment of the tetrads could then be deduced based on characteristic cyclic H1–H8 NOE connectivity within individual tetrads (Figure 3.23a,b). The top G-tetrad (G6•G24•G16•G10) is aligned in the opposite hydrogen-bond directionality as compared to the other tetrads, (G5•G11•G17•G23) and (G4•G12•G18•^{Br}G22), and the glycosidic conformations of guanines are *anti•anti•syn•syn* for all three tetrads. The central positioning of (G5•G11•G17•G23) in the tetrad core agrees with D$_2$O exchange experiments indicating that the H1 protons from this G-tetrad were the most protected (Figure 3.24). Joining the corners of
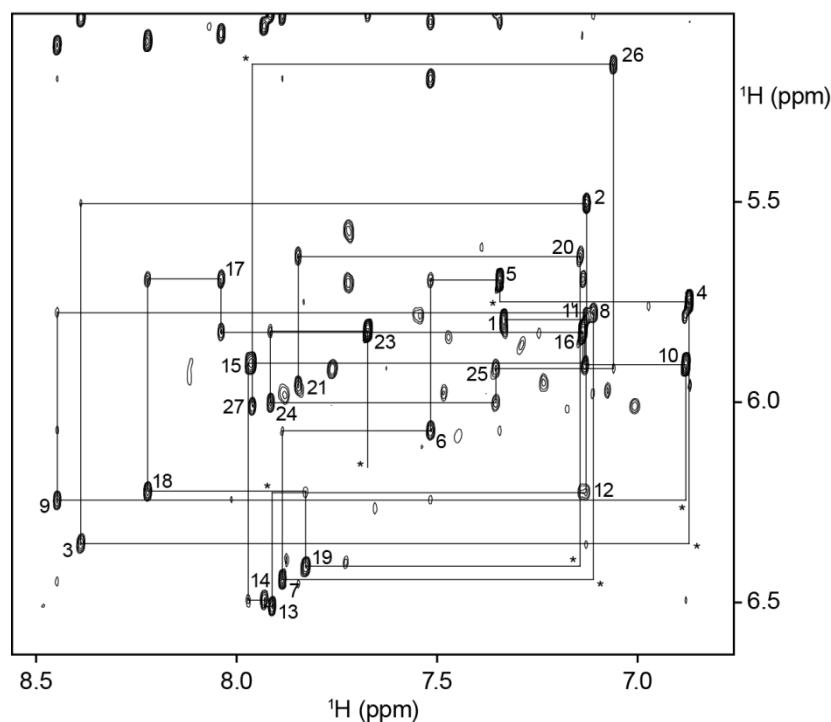
Figure 3.21 | NOESY spectrum showing the sequential connectivity of *4TTA16[Br22]*. NOE H6/H8–H1' sequential connectivity of *4TTA16[Br22]* from T1 to A27. Residue numbers are marked on intraresidue cross-peaks. Asterisks indicate weak/missing cross-peaks. Sequential connectivity is disrupted at position 22 since G22(H8) has been substituted by a bromine.

the tetrad core by the linking loops, an antiparallel up–up–down–down (or (2+2)) G-tetrad core[32] with a successive loop arrangement edgewise–propeller–edgewise was derived (Figure 3.23c).
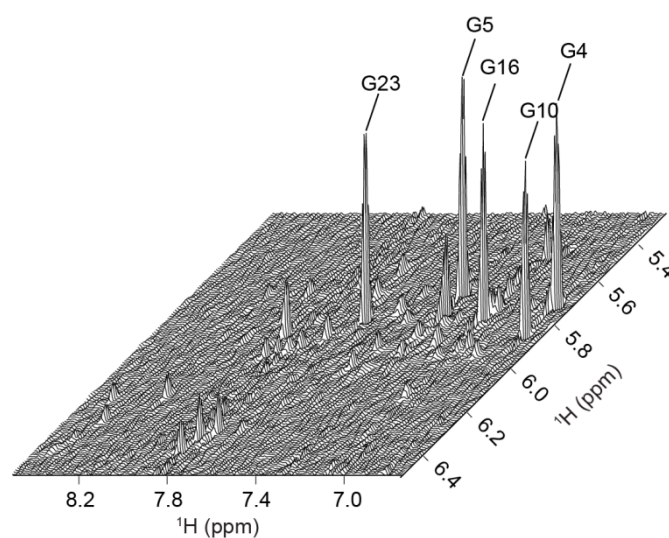
Figure 3.22 | NOESY spectrum of *4TTA16[Br22]* in a stacked plot view. Stacked plot of the NOESY spectrum highlighting the intense intraresidue H8–H1' NOE corresponding to the *syn* guanine residues.
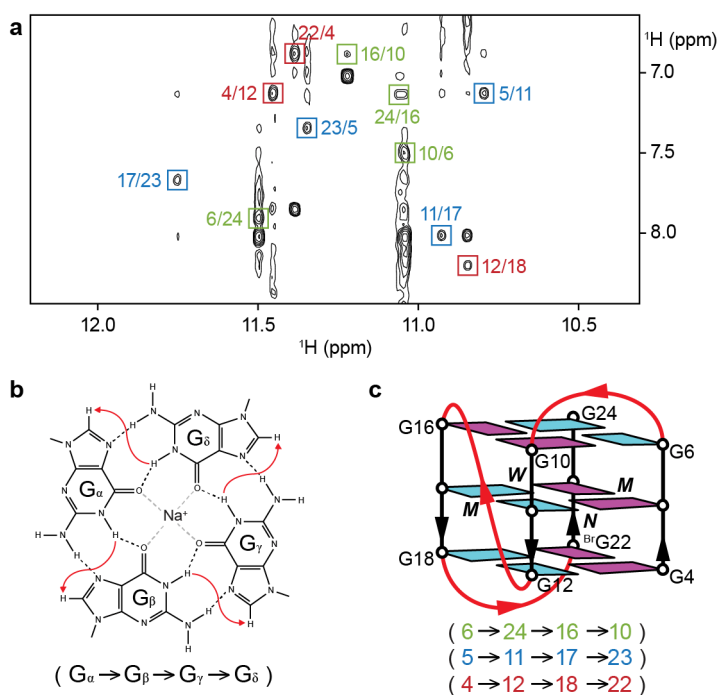
Figure 3.23 | Folding topology of *4TTA16[Br22]* in Na⁺ solution. (**a**,**b**) NOESY plot (**a**) of *4TTA16[Br22]* highlighting cross-peaks corresponding to the characteristic H1–H8 NOE connectivity within a tetrad (**b**). (**c**) Schematic structure of *4TTA16[Br22]* human telomeric G-quadruplex, which exhibits hydrogen-bond directionalities of tetrads as shown below.

Structure of *4TTA16[Br22]* G-quadruplex in sodium solution. Na⁺ solution structure for *4TTA16[Br22]* (Figure 3.25) was computed utilizing NMR restraints (Appendix I). The groove widths are successively narrow, medium, wide, and medium. The first (edgewise) loop crosses the top of the narrow groove, the second (propeller) loop traverses across one of the medium grooves, while the third (edgewise) loop completes a turn over the bottom of the wide groove. The 5'- and 3'-flanking
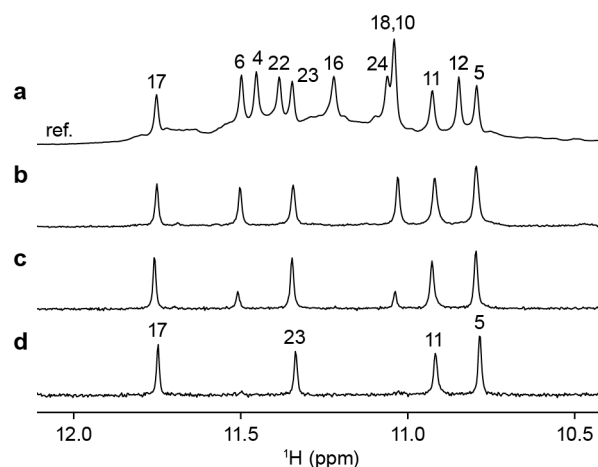
Figure 3.24 | Solvent exchange rate for imino protons of *4TTA16[Br22]*. Imino proton spectrum of *4TTA16[Br22]* (**a**) after 5 min in D₂O, (**b**) after 2 h in D₂O, and (**c**) after 5 h in D₂O.

50

Figure 3.25 | NMR structure of *4TTA16[Br22]* in sodium solution. (**a**) Stereo view of ten computed structures of *4TTA16[Br22]* G-quadruplex, aligned based on the tetrad core. (**b**) Stereo view of a representative conformer in ribbon representation.

TTA nucleotides take part in hydrogen-bond and stacking interactions at the bottom and the top of the structure, respectively. There is a Watson–Crick A9•T25 base pair capping the top of the tetrad core. A15 from the propeller loop also stacks onto the top tetrad, potentially establishing interactions with loop elements at the top. T14 from the same loop projects into the medium groove, whereas T13 flips out of the propeller loop. At the bottom, A3 and T20 establishes a Watson–Crick base pair, with A21 completing a triple platform. These stabilizing interactions of the loop and terminal residues would agree with the observation of additional broad imino proton peaks at ~12.2–14.0 ppm. The establishment of base pair stacking interactions across the ends of the tetrad core could have provided considerable contribution towards the adoption of the major conformation of *4TTA16[Br22]*, consistent with previous observations regarding other human telomeric G-quadruplexes in potassium solution[165,167,170,173,174]. In a longer sequence context, the various structural forms could exist in a state of equilibrium, with possible interconversion among them.

Novel features of the *4TTA16[Br22]* G-quadruplex. We have shown that in Na⁺ solution, *4TTA16[Br22]* assumes a novel antiparallel (2+2) folding topology (Figure 3.23c). Although the tetrad cores of *4TTA16[Br22]* (this work) and *4TTA05*[164] possess the same relative strand orientations, they differ in the relative orientation of tetrads: anticlockwise–clockwise–clockwise for *4TTA16[Br22]* versus clockwise–

anticlockwise–clockwise for *4TTA05* (Figure 3.4a). Loop arrangement of *4TTA16[Br22]* (edgewise–propeller–edgewise) is also different from that of *4TTA05* (edgewise–diagonal–edgewise). To date, antiparallel (2+2) G-quadruplexes that have been investigated largely consist of edgewise and/or diagonal loops[164,196-199], while quadruplexes containing propeller loops mostly belong to the all-parallel-stranded[33,200] or (3+1)[165,167,170,173,174,201-203] core topology. Antiparallel (2+2) G-quadruplexes with a propeller loop[204], such as that formed by *4TTA16[Br22]*, was so far not seen under the context of human telomeric repeats.

Conclusion

G-quadruplex structures adopted by human telomeric DNA have been actively explored as potential therapeutic targets. Previous structural studies have concurred on the polymorphic nature of these repeats under the influence of potassium ions, and the high-resolution structures of four distinct G-quadruplex conformations, all having a three-layered G-tetrad core, have been elucidated under different experimental conditions.

In this study, two novel intramolecular G-quadruplex folding topologies of canonical human telomeric sequences comprising four TTAGGG repeats were presented. The two-G-tetrad basket-type folding topology comprising extensive base interactions in the loop elements adopted by Form 3 in K^+ solution [see Section 3.2.1] revealed a new folding principle with regards to the establishment of stable quadruplex structures. On the other hand, the antiparallel (2+2) G-quadruplex folding topology adopted by *4TTA16[Br22]* (and presumably a subpopulation of *4TTA16*) in Na^+ solution [see Section 3.2.2] established the polymorphic nature of human telomeric repeats beyond the influence of K^+ ions. These two additions to the ensemble of telomeric G-quadruplex structures, with their unique loop architectures, would contribute to the available pool of structural motifs for targeting of these important entities.

G-Quadruplex Structures of Variant Human Telomeric Repeats

4.1　Variant Human Telomeric Repeats

Sequence variants within human telomeric DNA. The majority of human telomeric DNA are made up of canonical TTAGGG repeats[132,133]. Oftentimes, they are interspersed with repeats comprising sequence variants[133,205] (e.g. TCAGGG, TGAGGG, CTAGGG, TTGGGG, TAGGG, and TTAGG; variations underlined), especially towards the inner end of the chromosomal tip. These variant arrays are believed to have arisen randomly within the telomeric repeats, and their potential functions, if any, remain poorly understood. It has been reported that occurrence of a short contiguous repeat of the particular variant, CTAGGG, could lead to extreme localized instability at the telomeres of male germline and somatic cells[206]. Furthermore, it was shown that CTAGGG arrays exhibit a higher binding efficiency to the shelterin component POT1 as compared to canonical TTAGGG repeats. Here an investigation on the potential G-quadruplex structures that could be adopted by this sequence variant was performed. These aberrant structures could be responsible for the instability associated with this particular variant, as they could result in abnormalities during replication of the telomeres.

4.2　Sequence Variant d[(CTAGGG)$_n$] Prefers a G-Quadruplex Structure Comprising a G•C•G•C Tetrad[#]

Variant human telomeric sequences comprising four CTAGGG repeats adopt a novel G-quadruplex fold in potassium solution. The variant human telomeric sequence d[AGGG(CTAGGG)$_3$] (denoted as *22CTA*, Table 4.1) showed an NMR imino proton spectrum (Figure 4.1) with different signatures from those of four-repeat canonical human telomeric TTAGGG sequences, pointing to the adoption of a novel

Table 4.1 | Representative canonical and variant human telomeric DNA sequences used in this study.[a]

Name	Sequence	T_m[b] (°C)
22wt	A GGG TTA GGG TTA GGG TTA GGG	57.6
22CTA	A GGG CTA GGG CTA GGG CTA GGG	44.5
23CTA	A GGG CTA GGG CTA GGG CTA GGG C	50.8

[a] Sequence-variants are underlined. [b] Melting temperature in ~100 mM K$^+$, measured by following 295-nm UV absorbance.

[#] Results from this study were published in Lim *et al.*, *NAR* (2009).
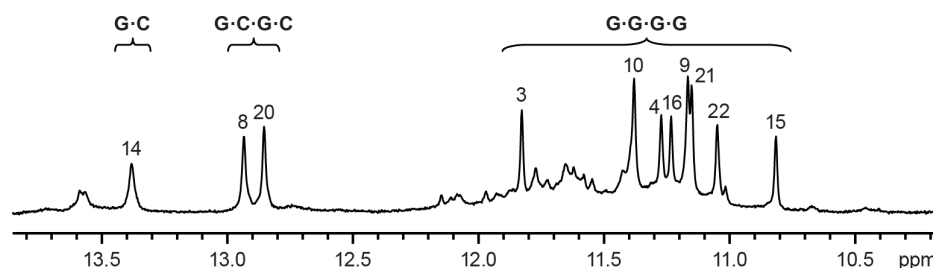
Figure 4.1 | 1D NMR spectrum of *22CTA* in K⁺ solution. NMR spectrum showing the H1 protons of *22CTA*, which are categorized based on their participation in the formation of G•C base pair, G•C•G•C tetrad, or G•G•G•G tetrad. Peaks from the major form are marked with their respective assignments.

quadruplex conformation. Eight major peaks were observed from 10.8–11.9 ppm, corresponding to the establishment of a two-tetrad quadruplex. In addition, three major peaks were observed from 12.8–13.4 ppm, indicating the existence of three G•C Watson–Crick base pairs. H1 proton spectra of variant human telomeric CTAGGG sequences (comprising four repeats) with different combinations of dangling CTA ends were screened (Figure 4.2), with many of them showing similar spectral patterns consisting of eight (or possibly more) peaks from 10.8–11.9 ppm and several major peaks from 12.8–13.6 ppm. This suggested that the majority of these sequences adopt in K⁺ solution a predominant structure comprising a G-tetrad core and multiple G•C Watson–Crick base pairs. In particular, the sequence *23CTA* (Table 4.1), which has an extra C at the 3'-terminal of *22CTA*, showed spectral characteristics closely approximating those of *22CTA* (Figure 4.2g).

UV-melting and CD studies. UV-melting experiments were carried out by following the 295-nm UV absorbance[127] of the DNA oligonucleotides. At a constant K⁺ concentration, T_m of *22CTA* remained mostly unchanged upon up to 200-fold difference in DNA strand concentration (5–200 µM), supporting its intramolecular nature. In 100 mM KCl, *22CTA* showed a slightly lower T_m as compared to the wild-type sequence *22wt* (Table 4.1). The variant *23CTA* sequence exhibited a slightly higher T_m as compared to *22CTA*.

CD spectra of *22CTA* and *22wt* (Figure 4.3) were recorded in 100 mM KCl. The CD spectrum of *22CTA* exhibited a sharp negative minimum near 260 nm together with a positive band around 295 nm, indicative of antiparallel G-quadruplexes[195]. The CD spectrum of its canonical counterpart *22wt* differed significantly, showing a huge positive peak at around 295 nm, as well as a negative trough near 235 nm.

NMR spectral assignments. For *22CTA*, unambiguous assignments for H1 and H8 protons from guanines were achieved based on site-specific 2% ^{15}N-enrichement of samples[192], one residue at a time, and site-specific ^{2}H labeling[207], respectively. These assignments were also corroborated by through-bond correlations between them[193] (Figure 4.4). Assignments for thymine nucleotides were validated with thymine-to-
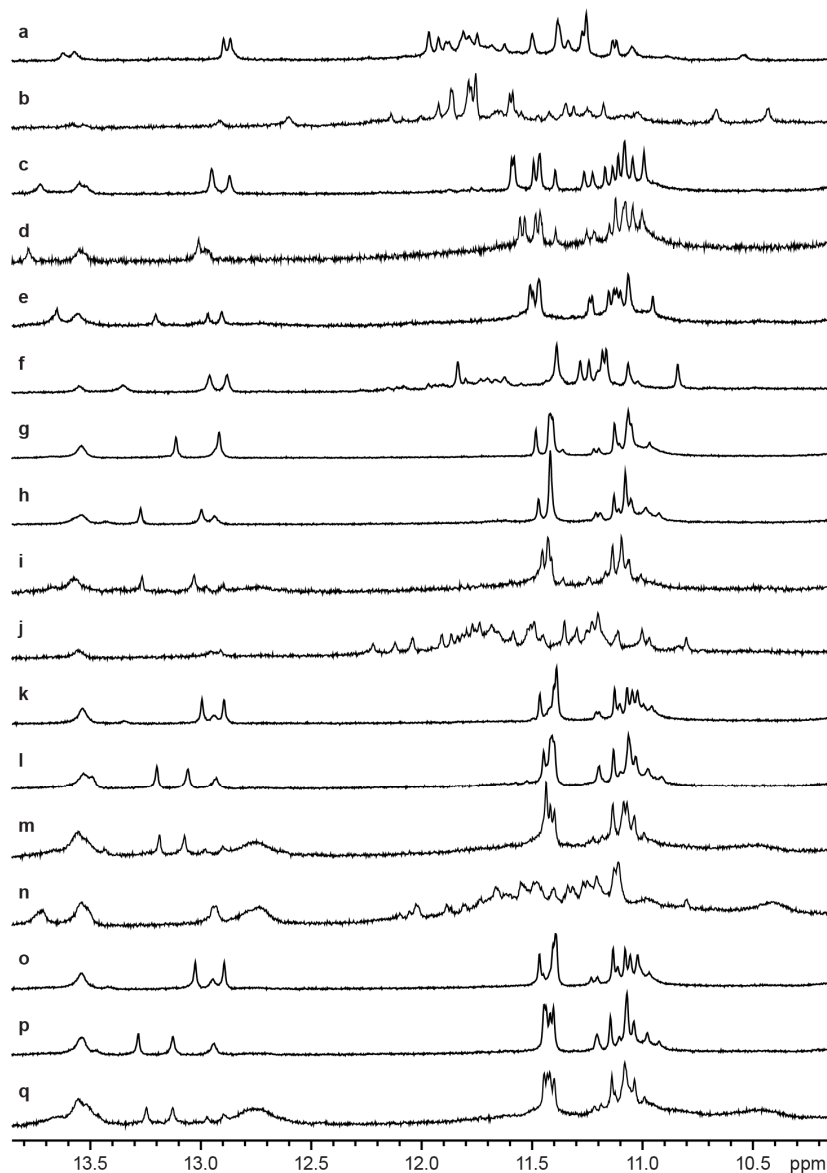
Figure 4.2 | NMR spectra showing the H1 protons of human telomeric variants comprising four CTAGGG repeats in potassium solution. (**a**) d[GG(CTAGGG)₃C], (**b**) d[GGG(CTAGGG)₃], (**c**) d[GGG(CTAGGG)₃C], (**d**) d[GGG(CTAGGG)₃CT], (**e**) d[GGG(CTAGGG)₃CTA], (**f**) d[AGGG(CTAGGG)₃], (**g**) d[AGGG(CTAGGG)₃C], (**h**) d[AGGG(CTAGGG)₃CT], (**i**) d[AGGG(CTAGGG)₃CTA], (**j**) d[TAGGG(CTAGGG)₃], (**k**) d[TAGGG(CTAGGG)₃C], (**l**) d[TAGGG(CTAGGG)₃CT], (**m**) d[TAGGG(CTAGGG)₃CTA], (**n**) d[(CTAGGG)₄], (**o**) d[(CTAGGG)₄C], (**p**) d[(CTAGGG)₄CT], and (**q**) d[(CTAGGG)₄CTA]. Most of them adopt predominantly the same chair-type intramolecular G-quadruplex, as suggested by the number of major tetrad-forming H1 protons from 10.8–11.9 ppm, as well as the two major downfield-shifted sharp imino signature peaks.
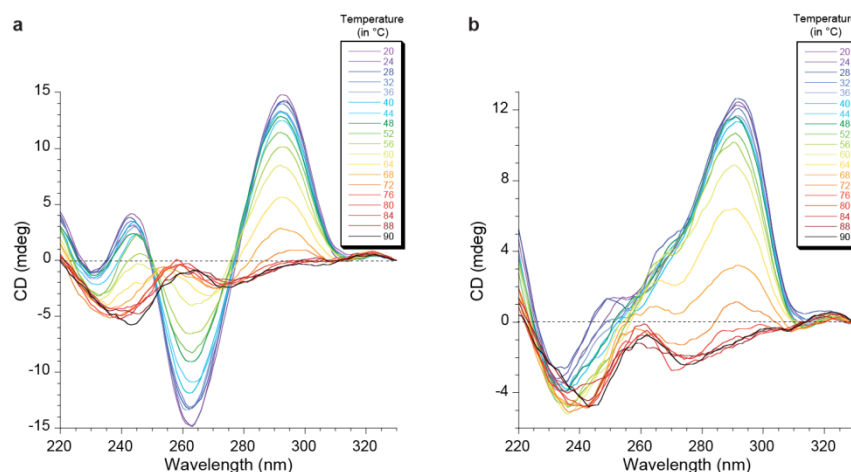
Figure 4.3 | CD spectra of four-repeat canonical and variant human telomeric sequences in potassium solution. (**a**,**b**) CD of *22CTA* (**a**) and *22wt* (**b**) were measured across different temperatures. Oligonucleotide concentration: 4 μM; buffer consisted of 10 mM lithium cacodylate and 100 mM KCl.
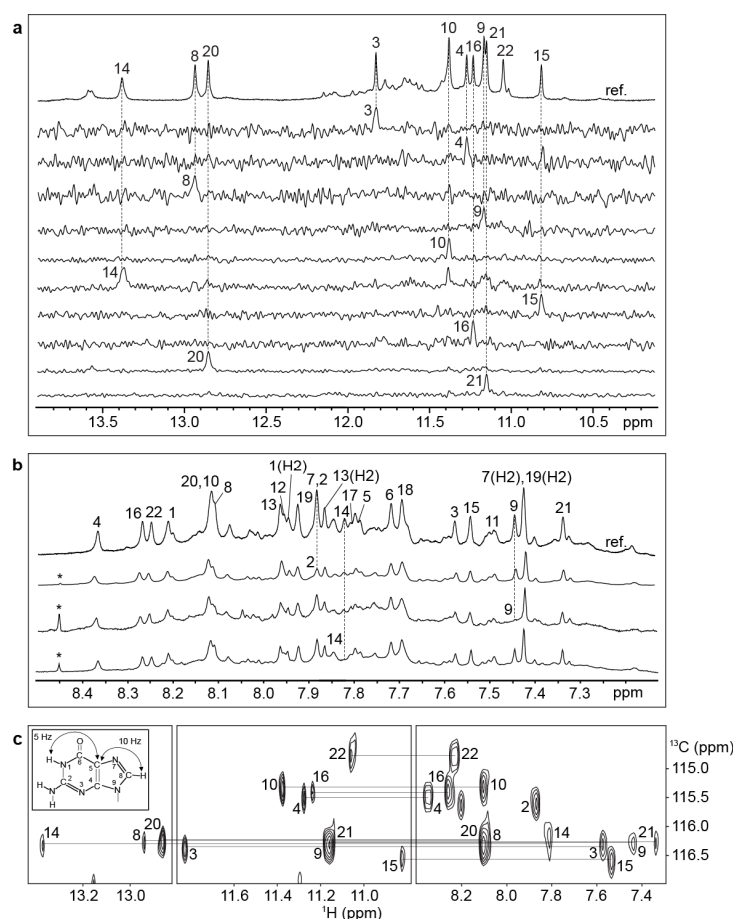
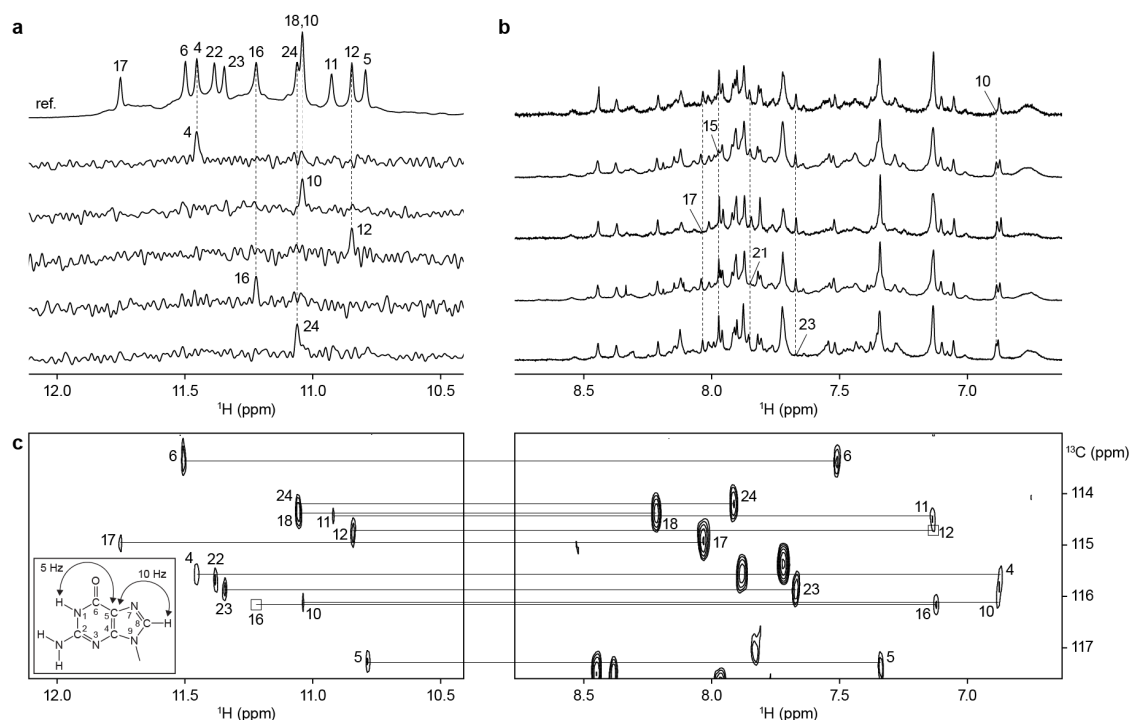**Figure 4.4 | Assignment of guanine H1 and H8 protons for *22CTA*.** (**a**) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. (**b**) Assignment of H8 protons for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. The asterisks indicate a contamination peak. (**c**) JRHMBC experiments showing long-range connectivity of guanine H1 and H8 protons mediated by their mutual through-bond correlations to [13]C5.

57

uracil modifications. Based on these unambiguous resonance assignments, as well as additional information from through-bond correlation experiments (^{13}C–^{1}H-HSQC, ^{13}C–^{1}H-HMBC, TOCSY, and COSY), NOE sequential connectivity of H8/H6–H1' for the DNA strand could be established from the 5'- to the 3'-terminal (Figure 4.5). Intense intranucleotide H8–H1' NOE of G3, G9, G15, and G21 indicated their adoption of the *syn* glycosidic conformation, while all other residues assume the anti glycosidic conformation. These results were consistent with results from G-to-^{Br}G substitutions of these four guanines, which gave rise to NMR imino proton spectra with similar spectral patterns (Figure 4.6).

A chair-type G-quadruplex comprising a G•C•G•C tetrad. The H1–H8 NOE connectivity within each tetrad (Figure 4.7a,b) was determined for *22CTA*, revealing the adoption of a two-G-tetrad chair-type G-quadruplex (Figure 4.7d). The two G-tetrads constituting the core, G3•G10•G15•G22 and G4•G21•G16•G9, possess opposing hydrogen-bond directionalities, and the guanines around both tetrads display glycosidic conformations of *syn•anti•syn•anti*. All four strands making up the core has two antiparallel neighboring strands, giving rise to two wide and two narrow grooves. There are three edgewise loops in total; the first and third loop double back at the narrow grooves over the top of the structure, whereas the second loop bends across the wide groove at the bottom.

Signature G(H1)–C(H41)/G(H1)–C(H42) NOE cross-peaks for G•C Watson–Crick base pairs were observed between three guanine–cytosine pairs, (G8 and C17), (G20
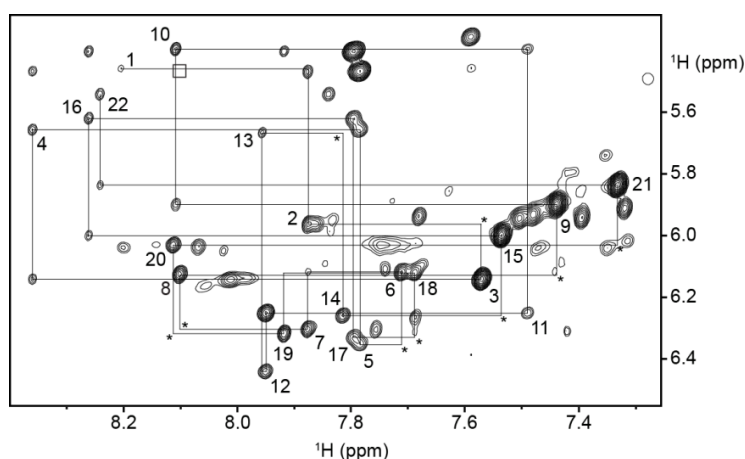
Figure 4.5 | NOESY spectrum showing the sequential connectivity of *22CTA*. H6/H8–H1' NOE connectivity of *22CTA* from A1 to G22. Residue numbers are marked on the corresponding intraresidue cross-peaks. Asterisks indicate cross-peaks that are weak or missing. A very weak NOE cross-peak (framed in a box; not observed at the current threshold) between G8(H8) and C5(H5) supports the formation of a slipped G8•C17•G20•C5 tetrad.
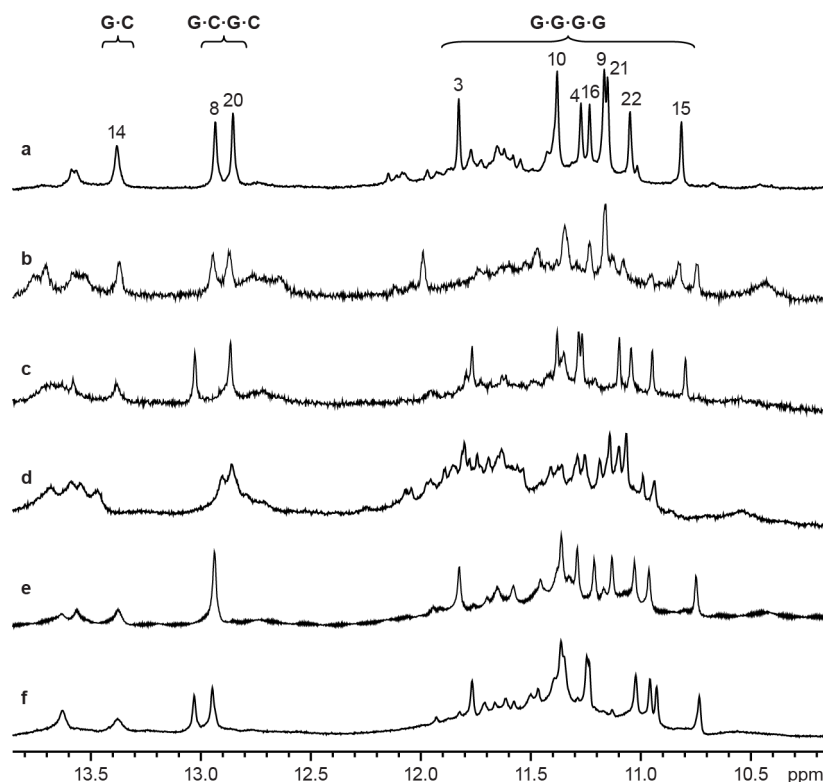
Figure 4.6 | H1 proton spectra of ^{Br}G-modified *22CTA* sequences. (a) d[AGGG(CTA GGG)₃], **(b)** d[AG(^{Br}G)G(CTAGGG)₃], **(c)** d[AGGGCTAG(^{Br}G)G(CTAGG G)₂], **(d)** d[AGGGCTAGGGCTAG(^{Br}G)GCTAGGG], **(e)** d[AGGG(CTAGGG)₂CTAG(^{Br}G) G], and **(f)** d[AGGGCTAG(^{Br}G)GCTAGGGCTAG(^{Br}G)G]. ^{Br}G substitution of *syn*-guanine(s) for *22CTA* maintains the same general fold as the unmodified sequence, as suggested by the number of major tetrad-forming imino protons and the presence of the two downfield-shifted sharp imino signature peaks.

and C5), and (G14 and C11) (Figure 4.7a; left panel). These indicated the formation of G8•C17, G20•C5, and G14•C11 base pairs, over the wide grooves of *22CTA*. The two former base pairs, situated across the top, are further aligned to give a G•C•G•C tetrad (Figure 4.7c). The placement of G4•G21•G16•G9 tetrad was supported by the slower solvent exchange rate exhibited by these imino protons as compared to those from the G8•C17•G20•C5 and G3•G10•G15•G22 tetrads.

NMR structure of *22CTA* G-quadruplex in potassium solution. The K⁺ solution structure of *22CTA* variant human telomeric G-quadruplex (Figure 4.8), which consists of an antiparallel (2+2) chair-type core, was computed based on NMR restraints (Appendix I). The five-membered rings of guanines from the two tetrads exhibit extensive stacking, the pattern of which is peculiar to successive G-tetrad layers having anticlockwise and clockwise hydrogen-bond directionalities.

Two distinct arrangements of G•C•G•C tetrad, which arises through association of two G•C base pairs across their major grooves, were observed in prior studies: a direct
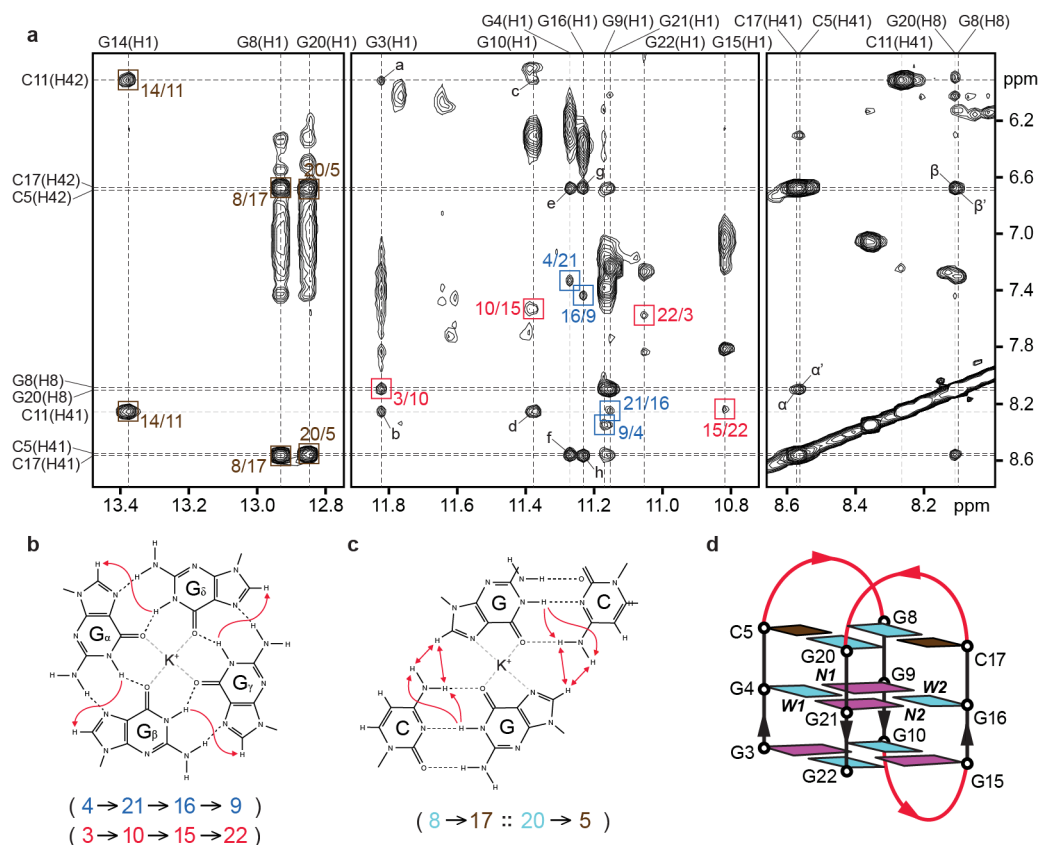
Figure 4.7 | Folding topology of *22CTA* in potassium solution. (a) NOESY plot of *22CTA* highlighting the characteristic cross-peaks corresponding to guanine H1–H8 NOE connectivity **(b)** around each tetrad, as well as G(H1)–C(H41)/G(H1)–C(H42) for G•C Watson–Crick base pair **(c)**. NOE cross-peaks a to h arise from interactions between tetrad H1 protons and cytosine amino protons. Cross-peaks α, α', β, and β' establish the slipped alignment **(c)** of G8•C17•G20•C5. **(d)** Schematic structure of *22CTA* variant human telomeric G-quadruplex.

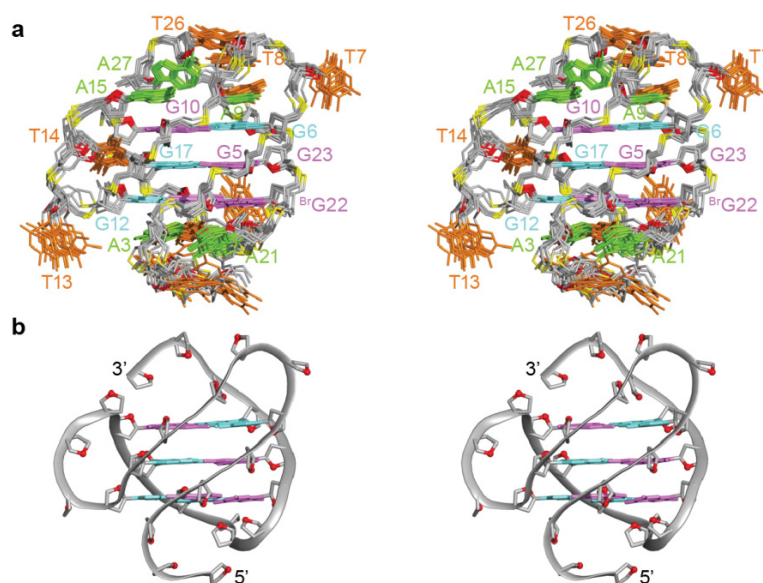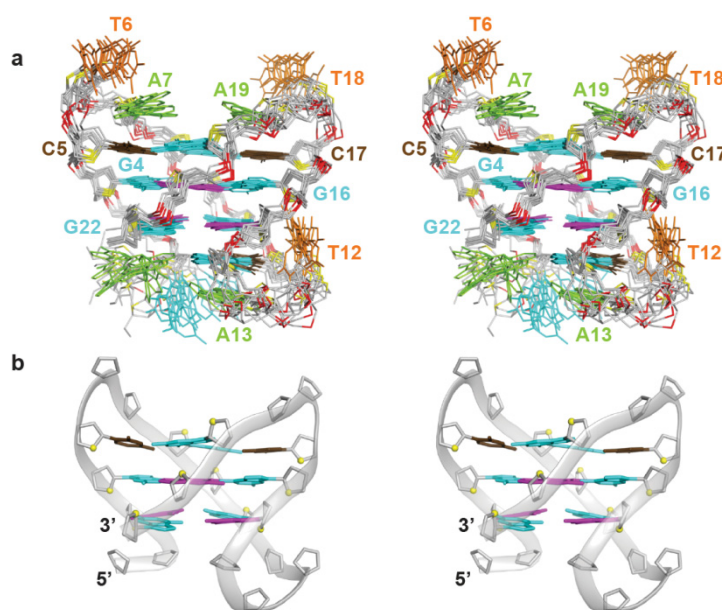**Figure 4.8 | NMR structure of *22CTA* in potassium solution. (a)** Stereo view of ten computed structures of *22CTA* G-quadruplex, aligned based on the tetrad core. **(b)** Stereo view of a representative conformer in ribbon representation.

alignment (Figure 4.9a,c) in Na[+] solution[208-210] and a slipped alignment (Figure 4.9b,d) in K[+] solution[211]. The slipped alignment was posited to accommodate a K[+] ion in the middle of the tetrad through its coordination to the two G•C base pairs[211] (Figures 4.7c and 4.9b). The adoption of either arrangement can be distinguished based on the expected NOE connectivity patterns (Figure 4.9c,d). For the direct alignment, G(H8) from one base pair would be at a distance of ~3 Å and ~5 Å to C(H5) and C(H41)/C(H42) from the adjacent base pair, respectively (Figure 4.9c). It follows that the first correlation, e.g. G(H8)–C(H5), should give rise to a strong NOE, whereas the second correlation pair, e.g. G(H8)–C(H41) and G(H8)–C(H42), should give rise to weak NOEs. The situation is reversed for the slipped alignment. In this case, G(H8) from one base pair would be closer to C(H41)/C(H42) (~3 Å) than C(H5) (~5 Å) from the adjacent base pair (Figure 4.9d). The top G8•C17•G20•C5 tetrad in *22CTA* was determined to adopt the slipped alignment (Figures 4.7, 4.8, and 4.10) in K[+] solution. This was verified through the strong NOE correlations of G8(H8)–C5(H41), G20(H8)–C17(H41), G8(H8)–C5(H42), and G20(H8)–C17(H42) (Figure 4.7a, right panel), as well as the very weak NOE correlations of G20(H8)–C17(H5) and G8(H8)– C5(H5) (Figure 4.5). NOEs from G8(H1) and G20 (H1) to G16(H1) and G4(H1),
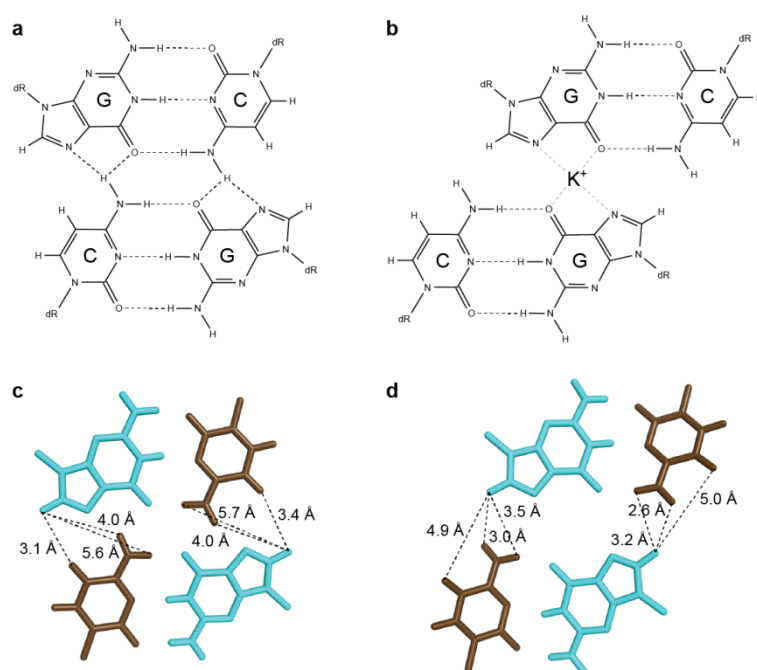
Figure 4.9 | Two possible arrangements of a G•C•G•C tetrad. The two G•C base pairs are paired up across the major groove edges, either in (**a**, **c**) a direct alignment or (**b**, **d**) a slipped alignment. Interproton distances that can be used to distinguish between the two alignments are shown by black dotted lines in (**c**) and (**d**). Coordinates in (**c**) and (**d**) are taken from structures deposited in the PDB (PDB codes 1A8N and 1A8W, respectively).
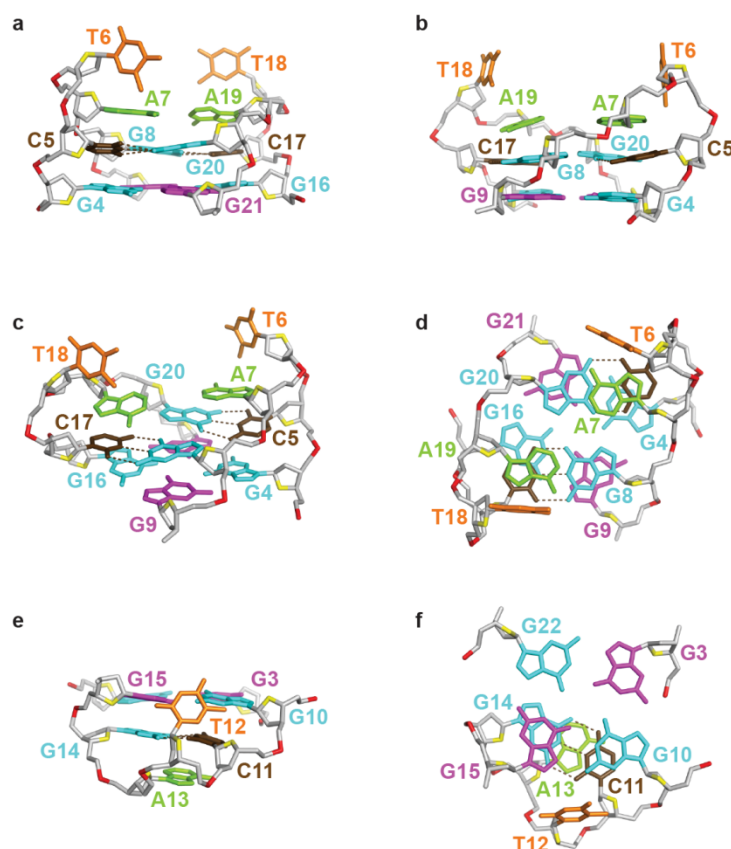
Figure 4.10 | Detailed view of *22CTA* variant human telomeric G-quadruplex. (a–d) The top segment comprising G8•C17•G20•C5 and G4•G21•G16•G9 tetrads. (**a**) View from the wide groove. (**b**) View from the narrow groove. (**c**) Top view at an angle. (**d**) View from the top down. (**e**,**f**) The bottom segment comprising G3•G10•G15•G22 tetrad and G14•C11 base pair. (**e**) View from the wide groove. (**f**) Top–down view.

respectively, define the relative arrangement of G4•G21•G16•G9 and G8•C17•G20•C5 tetrads.

Over at the top of the structure, G8•C17•G20•C5 tetrad is further stacked over by two adenine (A7 and A19) bases (Figure 4.10a–d). This was supported by NOEs between protons from C5 and C17 to those from A7 and A19, respectively. Bases of the two thymines T6 and T18 are flipped out of their respectively loops (Figure 4.10a–d), consistent with a dearth of NOEs from their protons to those from the neighboring residues. The top segment of the overall structure displayed a high degree of quasi-symmetry[212,213] (Figures 4.8 and 4.10), which is evident in the NMR spectra; the two halves constituting the top, G4–G9 and G16–G21, display a number of spectral similarities such as NOE patterns and proton chemical shifts (Figures 4.1, 4.4, 4.5, and 4.7).

At the bottom half, G14•C11 base pair is positioned beneath the G3•G10•G15•G22 tetrad, right below G15 and G10 (Figure 4.10e,f). This arrangement was defined based

on NOE cross-peaks between their protons [e.g. G10(H1)–G14(H1)], as well as the slower solvent exchange rate of G10(H1) with respect to the H1 protons of the other three guanines from the same G-tetrad. A13 from the bottom loop is placed below the G14•C11 base pair, and the loop is completed by T12, which folds into the hydrophobic groove.

A robust G-quadruplex scaffold. In the structure of *22CTA*, an antiparallel (2+2) chair-type core comprising two G-tetrad layers is packed between a G•C•G•C tetrad and a G•C base pair. This robust scaffold seems to constitute the predominant species in potassium solution for the bulk of four-repeat variant human telomeric CTAGGG sequences having various dangling termini. This is in stark contrast to the situation involving the canonical human telomeric TTAGGG counterparts, in which a diversity of structural forms have been revealed[162,165-190]. Preliminary NMR spectral analyses and modeling of *23CTA* were also performed, suggesting that it could adopt a similar G-quadruplex scaffold as *22CTA*, with an additional G14•C11•G2•C23 established at the bottom of the G-quadruplex (instead of just a single G14•C11 base pair as in *22CTA*). This structural model of *23CTA* would be in line with its higher T_m as compared to *22CTA* (Table 4.1). Such structures comprising two G•C•G•C tetrads sandwiching two G-tetrads were observed previously[211]. The stable G-quadruplex structures of *22CTA* and *23CTA* indicate that establishment of G•C•G•C tetrads could improve stability of a G-quadruplex[208-211,214], as opposed to the debilitating influence observed for other tetrads of guanine derivatives[215].

Conclusion

Human telomeric DNA comprises successive arrays of canonical TTAGGG sequence, with intermittent occurrence of sequence variants, especially so towards the sub-telomeric region. One particular variant repeat type, CTAGGG, was found to lead to extreme telomere instability, suggesting its potential formation of unusual structures.

The high-resolution NMR structure in K$^+$ solution adopted by one particular four-repeat variant human telomeric CTAGGG sequence, *22CTA*, was presented in this study. It corresponds to a quadruplex scaffold comprising an antiparallel two-G-tetrad core flanked on the two ends by a G•C•G•C tetrad and a G•C base pair. The same robust scaffold was implicated in a series of four-repeat CTAGGG sequences having various dangling termini, suggesting its prevalence in this particular repeat type. This

robust G-quadruplex scaffold, displaying unique structural features, could serve as a novel target for the design of ligands specifically targeting human telomeric DNA.

65

G-Quadruplex Structures of
Long Human Telomeric Repeats

5.1   Long Human Telomeric Repeats

Structural models of long human telomeric DNA. Human telomeric DNA comprises successive arrays of the TTAGGG sequence. Myriad studies have revealed the structural heterogeneity of these arrays, with the high-resolution G-quadruplex structures of at least five distinct conformations elucidated under various experimental conditions [see Chapter 3]. This ensemble of structures could exist in a dynamic equilibrium at the telomere, and an understanding on their structural dynamics and/or potential interactions would be valuable towards drug design efforts targeting telomeric G-quadruplexes.

Structures arising from longer human telomeric sequences (i.e. comprising five or more TTAGGG repeats) are the subjects of more recent research[186,189,216-223], but structural characterization of these larger complexes remains challenging. Previous studies have yielded structural insights on long human telomeric repeats across several fronts. It was observed that long sequences comprising five or more human telomeric repeats preferentially fold into G-quadruplexes at the 3'-end[217]. Architectures of eight-repeat TTAGGG sequences[33,218,219,223], which correspond to the minimum length required to fold into two G-quadruplex units, have also been explored, largely through molecular dynamics simulations. In this case, different models involving either the tight packing[33,219] or partial stacking[218,223] (with interactions between the loop elements) of the two quadruplex blocks have been put forward. The possible extension of these packing principles to three quadruplex blocks and beyond has further been proposed[33,219,223]. Furthermore, distribution of multiple quadruplex blocks in a consecutive[221] and/or beads-on-a-string[189,222] arrangement has been experimentally observed. On the other hand, alternative G-quadruplex folding based on long-range interactions of G-tracts (which could potentially occur due to the formation of t-loop[150]) has also been put forward[220,224].

In this study, G-quadruplex formation in sequences spanning five to seven TTAGGG repeats was investigated. It was found that these sequences could all assume the (3+1) G-tetrad scaffold by accommodating one or more TTAGGG repeat(s) within the propeller loop. A Watson–Crick duplex could be induced within the loop upon the introduction of a complementary strand, implicating the possibility for the long loop to act as a recognition site.

5.2   Human Telomeric G-Quadruplexes Harboring a Long Loop[#]

Inosine substitution reduces the multiplicity of conformations in potassium solution for five-repeat human telomeric sequences. In potassium solution, natural five-repeat human telomeric TTAGGG sequences possessing various dangling termini produced NMR imino proton spectra that pointed to the existence of multiple G-quadruplex species (examples shown in Figure 5.1a–e). We decided to proceed with detailed analyses on the sequence d[TAGGG(TTAGGG)₄TT] (denoted as *31htel*; Table 5.1), which harbors the same terminal sequences as d[TAGGG(TTAGGG)₃TT] (denoted as *25htel*; Table 5.1). *25htel* was demonstrated to adopt the (3+1) G-quadruplex scaffold in previous NMR studies[166,167].

Replacement of a guanine by its analogue inosine was previously applied to disrupt the guanine at that particular position from taking part in G-quadruplex formation[225,226] [see Section 6.2]. This strategy was adapted here to disfavor an entire G-tract from participating in G-tetrad formation by substituting the middle G with inosine to generate a GIG tract. The modification was individually implemented on the G-tracts for *31htel*. The intensity and number of peaks across imino proton region of the NMR spectra indicated a reduction in the multiplicity of existing conformations for all five inosine-substituted sequences with respect to *31htel* (Figure 5.1f–j). In addition, no sharp H1 proton peaks were observed at 13–14 ppm[227], the characteristic chemical shift range of inosine H1 protons, indicating that the inosines did not participate in the formation of tetrads. These results corroborated the assumption that the GIG tract was
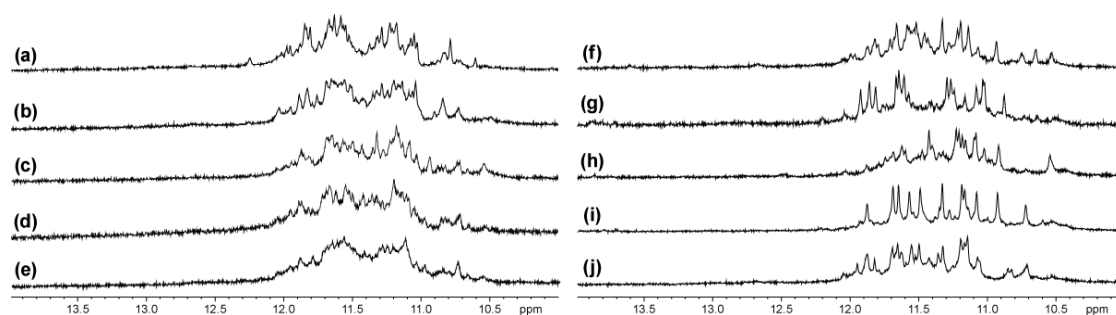
Figure 5.1 | NMR spectra showing the H1 protons of natural and inosine-substituted human telomeric sequences comprising five TTAGGG repeats in potassium solution. (a) d[TAGGG(TTAGGG)₄], **(b)** d[TAGGG(TTAGGG)₄T], **(c)** d[TAGGG(TTAGGG)₄TT], **(d)** d[TAGGG(TTAGGG)₄TTA], **(e)** d[(TTAGGG)₅TTA], **(f)** d[TAGIG(TTAGGG)₄TT], **(g)** d[TAGGGTTAGIG(TTAGGG)₃TT], **(h)** d[TAGGGTTAGGGTTAGIG(TTAGGG)₂TT], **(i)** d[TAGGG(TTAGGG)₂TTAGIGTTAGGGTT], and **(j)** d[TAGGG(TTAGGG)₃TTAGIGTT].

Table 5.1 | Representative natural and modified four- to seven-repeat human telomeric DNA sequences used in this study.[a]

Name	Sequence			T_m[b] (°C)
25htel	TA GGG TTA GGG TTA GGG	TTA	GGG TT	62.0
31htel	TA GGG TTA GGG TTA GGG	TTA GGG TTA	GGG TT	–
31htel[I10]	TA GGG TTA **GIG** TTA GGG	TTA GGG TTA	GGG TT	–
31htel[I22]	TA GGG TTA GGG TTA GGG	TTA **GIG** TTA	GGG TT	49.8
37htel	TA GGG TTA GGG TTA GGG	TTA GGG TTA GGG TTA	GGG TT	–
37htel[I22/I28]	TA GGG TTA GGG TTA GGG	TTA **GIG** TTA **GIG** TTA	GGG TT	47.7
43htel	TA GGG TTA GGG TTA GGG	TTA GGG TTA GGG TTA GGG TTA	GGG TT	–
43htel[I22/I28/I34]	TA GGG TTA GGG TTA GGG	TTA **GIG** TTA **GIG** TTA **GIG** TTA	GGG TT	46.5
43htel[T22/T28/T34]	TA GGG TTA GGG TTA **Br**G**GG	TTA G**T**G TTA G**T**G TTA G**T**G TTA	GGG TT	–

a Modified residues are shown in bold. *b* Melting temperature in ~100 mM K⁺, measured by following 295-nm UV absorbance.

excluded from tetrad core formation in these modified sequences.

Individual replacement of the first, third, and fifth G-tracts by GIG failed to favor a predominant species (Figure 5.1f,h,j). On the other hand, individual inosine substitution at the second and fourth G-tracts (denoted as *31htel[I10]* and *31htel[I22]*, respectively; Table 5.1) both favored a single major conformation (Figure 5.1g,i). In particular, *31htel[I22]* displayed especially high spectral correspondence to *25htel* (see below) and was investigated further.

Spectral correspondence between *31htel[I22]* and *25htel.* *31htel[I22]* displayed eleven major peaks in the NMR imino proton spectrum (Figure 5.1i). The assignment of these protons were unambiguously determined based on site-specific 2% ¹⁵N-enrichment of samples[192] (Figure 5.2b), one residue at a time, and was found to resemble those of *25htel* (Figure 5.2a), all the way until the third G-tract. Three remaining H1 proton peaks belong to the last G-tract, in line with exclusion of the fourth tract (G21–I22–G23) from participating in G-tetrad core formation. A close correspondence was observed between the H1 protons of the fifth G-tract from *31htel[I22]* and those of the fourth G-tract from *25htel* (Figure 5.2a,b, underlined). Furthermore, spectral correspondence between the two sequences was detected in their NOESY spectra, including similar NOE patterns, as well as the existence of five intense intranucleotide H8–H1' NOE corresponding to *syn* guanines residues (Figure 5.3a). The slower solvent exchange rate displayed by the H1 protons from G4, G10, G16, and G28 (Figure 5.2c) pointed to their middle positioning in the tetrad core.

G-quadruplexes comprising a long propeller loop. All these observations converged on the argument that *25htel* and *31htel[I22]* assume the identical (3+1)
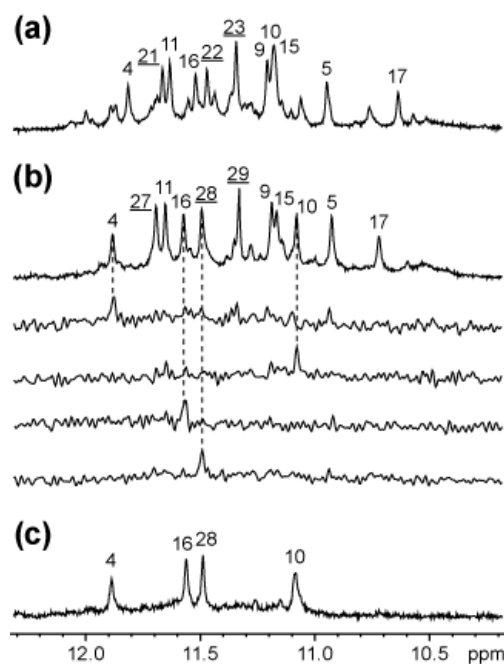
Figure 5.2 | 1D NMR spectral correspondence between *25htel* and *31htel[I22]* in K⁺ solution. (a,b) NMR spectra of *25htel* (**a**) and *31htel[I22]* (**b**) in potassium solution, showing close spectral correspondence between the two. Peaks from the major form are marked with their respective assignments. Note that the last G-tracts (underlined) utilized for G-tetrad core formation are different. Assignment of guanine H1 protons for *31htel[I22]* is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. (**c**) NMR spectrum of *31htel[I22]* in ²H₂O for 1 h at 25 °C.

quadruplex scaffold. Instead of proceeding with the thorough assignment process through various through-bond and through-space NMR experiments, an alternative approach was adopted to validate the folding topology of *31htel[I22]*. Herein, NOESY experiments were carried out on a select list of site-specific ²H-substituted samples[207]
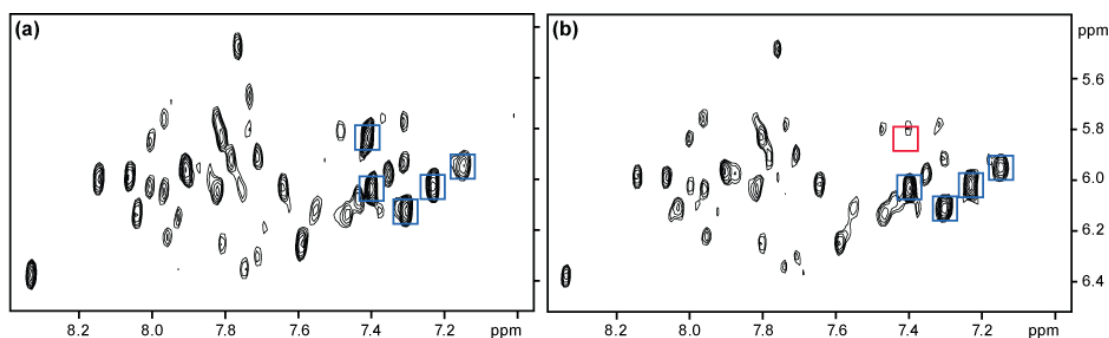
Figure 5.3 | NOESY spectra displaying the H6/H8–H1' NOE connectivity for five-repeat modified human telomeric sequence *31htel[I22]* and a site-specific deuterated sample. (**a,b**) NOESY spectra of the unlabeled d[TAGGG(TTAGGG)₂TTAGIGTTAGGGTT] (**a**) and the site-specific ²H-labeled d[TAGGGTTAG(#G)GTTAGGGTTAGIGTTA GGGTT] (**b**) sequence. The ²H-labeled guanine (at the H8 position) is marked by a hash sign; I is inosine. *Syn* guanine intraresidue cross-peaks (five in **a**; four in **b**) are framed in blue, while the missing cross-peak in **b** is framed in red.

69

to verify the alignment of a tetrad. Specifically, the middle G4•G10•G16•G28 tetrad was picked as the subject for validation, since cross-peaks corresponding to imino–H8 NOE connectivity for this tetrad layer were well-resolved (Figure 5.4b). Upon substitution by ^{2}H for individual H8 protons, the respective cross-peaks corresponding to imino–H8 correlation with the adjacent guanine disappeared from the NOESY plot covering the typical chemical shift ranges for H1/H8 protons (Figure 5.4c–f). Identity for all four cross-peaks corresponding to the cyclic NOE connectivity within the G4•G10•G16•G28 tetrad (Figure 5.4a) was thus established in an unambiguous manner. The strong intranucleotide H8–H1' NOE cross-peak corresponding to *syn* G10 also vanished upon substitution of G10(H8) by ^{2}H (Figure 5.3b), further corroborating the adoption by *31htel[I22]* of the same scaffold as *25htel*. *25htel* comprises a 3-nt (TTA) propeller loop (Figure 5.5a), while *31htel[I22]* comprises a 9-nt (TTAGGGTTA) propeller loop (Figure 5.5b).
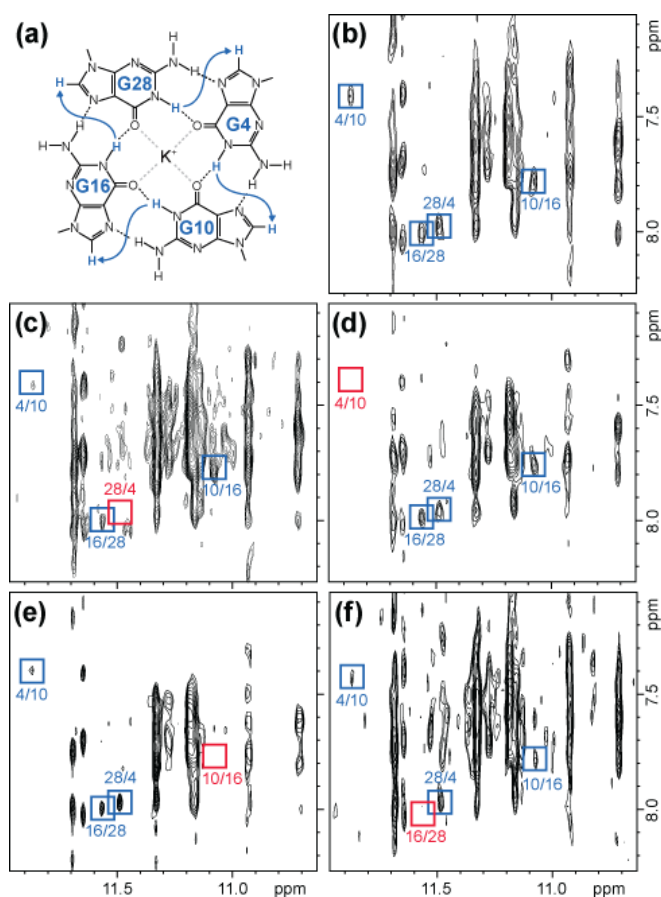
Figure 5.4 | Determination of the hydrogen-bond directionality of the middle tetrad for *31htel[I22]*. (**a**–**f**) Four cross-peaks corresponding to cyclic H1–H8 NOE connectivity within the tetrad (**a**) can be identified in the NOESY plot of *31htel[I22]* (**b**). Substitution of the H8 protons of G4 (**c**), G10 (**d**), G16 (**e**), and G28 (**f**) with ^{2}H, one residue at a time, resulted in the disappearance of the respective cross-peaks, verifying the G-tetrad alignment.
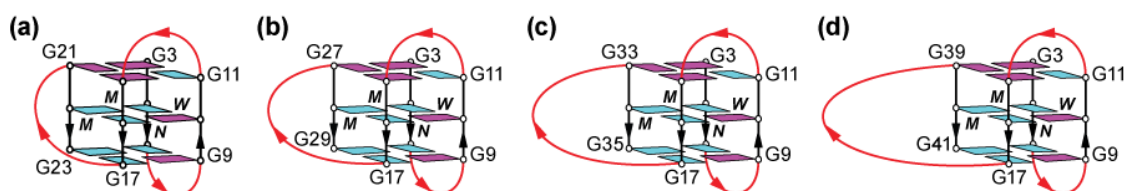
Figure 5.5 | Form 2 (3+1) G-quadruplexes formed by long human telomeric repeats in potassium solution. (**a**) *25htel* with a 3-nt propeller loop, (**b**) *31htel[I22]* with a 9-nt propeller loop, (**c**) *37htel[I22/I28]* with a 15-nt propeller loop, and (**d**) *43htel[I22/I28/I34]* with a 21-nt propeller loop.

Implementation of the GGG-to-GIG substitution approach on the six-repeat sequence *37htel* (Table 5.1), which comprises the same dangling termini as *31htel[I22]* and *25htel*, somewhat pared down the conformational heterogeneity, but considerable multiplicity was still observed across these single-inosine-substituted sequences (data not shown). Exclusion of the fourth and fifth G-tracts from participating in G-tetrad core formation through the concurrent incorporation of a middle inosine in both G-tracts, giving rise to the sequence *37htel[I22/I28]* (Table 5.1), produced a major conformation (~70%). The imino proton peak distribution patterns of *37htel[I22/I28]* resembled those of *31htel[I22]* and *25htel* to some degree. ^{Br}G-for-G replacement at G15 of *25htel* was previously demonstrated to drive its folding towards the Form 2 (3+1) G-quadruplex topology[167]. Accordingly, incorporation of ^{Br}G in the corresponding position in each of *25htel*, *31htel[I22]*, and *37htel[I22/I28]* led to a cleaner NMR spectrum (Figure 5.6). Highly similar spread of the imino proton peaks in all three sequences suggested that all of them adopt the Form 2 (3+1) G-quadruplex with variable propeller loop lengths (Figure 5.5).

The possible adoption of an even longer propeller loop in seven-repeat sequences was investigated, in this case through triple GIG replacement of the fourth, fifth, and sixth G-tracts in the sequence *43htel* to give *43htel[I22/I28/I34]* (Table 5.1). Again, structural polymorphism was greatly suppressed in *43htel[I22/I28/I34]* as compared to *43htel*, though considerable multiplicity of conformations remained (data not shown). Incorporation of ^{Br}G at the corresponding Form 2-favoring position similarly resulted in the emergence of a major species (Figure 5.6d). Similar imino proton peak distribution patterns of this seven-repeat sequence as the aforementioned sequences comprising fewer repeats implicated its adoption of the same quadruplex scaffold harboring a 21-nt loop (Figure 5.5d).
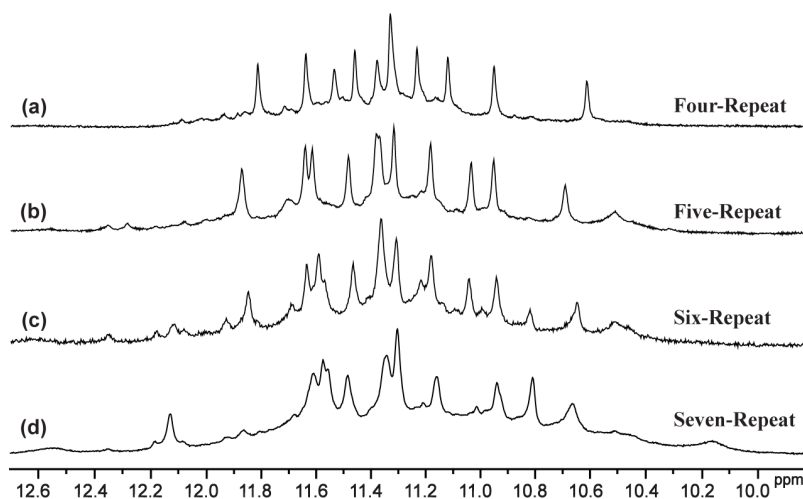
Figure 5.6 | Imino proton spectra of ^{Br}G-modified human telomeric DNA comprising four to seven sequence repeats. (**a**) d[TAGGGTTAGGGTTA^{Br}GGGTTAGGGTT], (**b**) d[TAGGGTTAGGGTTA^{Br}GGGTTAGIGTTAGGGTT, (**c**) d[TAGGGTTAGGGTTA^{Br}GGG (TTAGIG)$_2$TTAGGGTT, and (**d**) d[TAGGGTTAGGGTTA^{Br}GGG (TTAGIG)$_3$TTAGGGTT.

UV-melting studies. Stability of the series of Form 2 (3+1) human telomeric G-quadruplexes with varying propeller loop lengths was investigated by UV-melting experiments. The melting profiles of *31htel[I22]* remained largely unchanged upon up to 50-fold difference in DNA strand concentrations (from 4–200 µM), consistent with an intramolecular folding of the G-quadruplex. At ~100 mM K$^+$, T_m of *31htel[I22]* was found to be 49.8 °C, considerably below that of *25htel* (T_m = 62.0 °C). This trend of diminishing stability with larger loop size has been observed in previous studies[180,228-232]. On the other hand, increasing length of the propeller loop in *37htel[I22/I28]* (T_m = 47.7 °C) and *43htel[I22/I28/I34]* (T_m = 46.5 °C) brought about a smaller drop in melting temperatures (Table 5.1). It should be pointed out however that the T_m for both of these sequences contained more considerable contributions from minor species.

A robust G-quadruplex scaffold. The (3+1) G-quadruplex topology was repeatedly observed within the context of human telomeric sequences under various guises[165-167,169-174,220,224]. These include the intramolecular formation of Form 1[165,167,169,170,172,173] and Form 2[166,167,174] G-quadruplexes by human telomeric TTAGGG sequences comprising four repeats in potassium solution, as well as the putative formation of quadruplexes within the t-loop architecture[220,224]. In the Form 2 arrangement, two edgewise loops are followed by a propeller loop[166,167,174]. Here we showed that more than one TTAGGG repeats can be nested within the propeller loop in this scaffold (Figure 5.5). This long loop architecture could possibly apply to the

Form 1 arrangement as well, given that the adaptability of structural motifs between different G-quadruplexes has been observed previously[233]. Looking back, this could be the situation happening in the inosine-substituted five-repeat sequence *31htel[I10]* (Table 5.1 and Figure 5.1g), in which the second G-tract was modified to GIG.

Long propeller loop as a recognition motif. We have shown that human telomeric sequences comprising five to seven TTAGGG repeats are capable of adopting (3+1) G-quadruplexes harboring up to three TTAGGG repeats within a single propeller loop (Figure 5.5). Partial secondary structures could arise within the long loop, and these could potentially serve as recognition motifs, either existing within the natural telomere context, or to be exploited for ligand design. To validate the latter proposition, a seven-repeat sequence containing a modified loop (denoted as *43htel[T22/T28/T34]*; Table 5.1) was titrated against a 15-nt oligonucleotide (mimic for a putative ligand) having a complementary sequence to the modified loop region of *43htel[T22/T28/T34]*. H1 proton peaks corresponding to Watson–Crick base pair formation emerged from 12.3–13.8 ppm on addition of the short strand (Figure 5.7a,b). This indicated that the oligonucleotide successfully homed in on the long propeller loop, which projects out from the pre-existing G-quadruplex, and subsequently induced duplex formation (Figure 5.7c). As such, this targeting strategy could in principle be extended to the design of a chemical ligand.
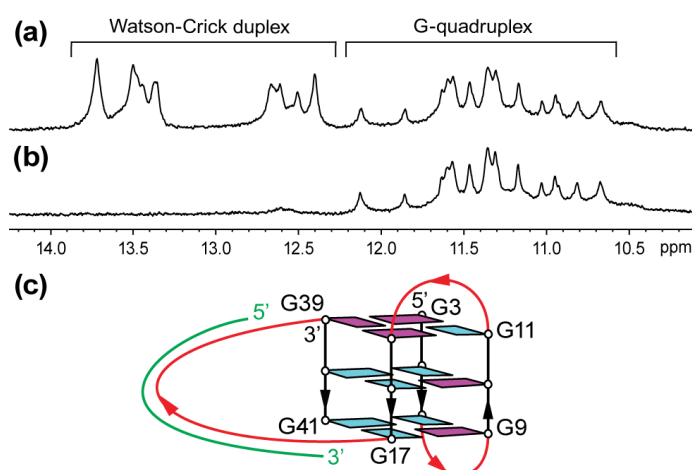
Figure 5.7 | Long loop of quadruplex as a targeting site. (**a**,**b**) 1D NMR spectra of free (**b**) and bound (**a**) *43htel[T22/T28/T34]* in K$^+$ solution. Sequence of the targeting oligonucleotide is d[CACTAACACTAACAC]. (**c**) Schematic diagram illustrating induction of a duplex within a long loop (colored in red) of a (3+1) quadruplex upon its specific binding by a complementary oligonucleotide (colored in green).

Conclusion

G-quadruplex structures of human telomeric repeats have been extensively studied towards the rational design of quadruplex-interactive compounds for anticancer treatment. Several high-resolution G-quadruplex structures adopted by human telomeric sequences comprising four TTAGGG repeats were elucidated, revealing a high degree of conformational heterogeneity. However to date, high-resolution structures of long human telomeric repeats remain elusive. Various techniques have been applied to characterize the structures and dynamics of these long repeats, which will be crucial towards the understanding of telomere biology.

In this study, guanine-to-inosine substitutions have been employed to mitigate the multiplicity of conformations present in long human telomeric repeats. A (3+1) G-quadruplex harboring a long propeller loop was isolated from a five-repeat sequence based on this approach. A cost-effective method was implemented to validate the fold of the quadruplex. This folding principle was shown to apply to longer telomeric repeats, whereby one or more TTAGGG repeats can be accommodated within the long propeller loop. Such a long propeller loop could serve as a recognition motif, either within the natural telomere context, or to be exploited for ligand design. More work needs to be done to reveal yet more possible structures that can be adopted by long human telomeric repeats, and to understand the various interplays among these complex structural entities.

Structures of Promoter G-Quadruplexes

6.1   G-Quadruplex & Gene Transcription

6.2   G-Quadruplexes in the hTERT Promoter

6.1   G-Quadruplex & Gene Transcription

Enrichment of G-rich sequences in gene promoters. Recent bioinformatics analyses have revealed that putative quadruplex sequences (PQS) of the generic form $G_{3+}N_{L1}G_{3+}N_{L2}G_{3+}N_{L3}G_{3+}$ are prevalent in the human genome[41,42]. It has further been shown that these motifs are highly enriched in the promoter regions immediately upstream of TSS[47] and are strongly associated with specific functional classes of genes[46]. In particular, PQS are overrepresented in oncogenic promoters (e.g. c-myc, c-kit, bcl-2, etc.). These observations led to the proposition of a direct regulatory role of PQS in gene transcription[47,234].

Putative role of promoter quadruplexes as transcriptional regulators. The first experimental evidence of the transcriptional regulatory role of quadruplexes came from studies on quadruplex formations in the oncogenic promoter of *c-myc*[45] (Figure 6.1); stabilization of quadruplex formation within the nuclease hypersensitivity element III$_1$ (NHE III$_1$) by the quadruplex-selective small molecule TMPyP4 led to a downregulation of *c-myc* transcription. Development of specific ligands for the targeting of promoter quadruplexes has thus become an attractive anticancer strategy[44,45,61,235], which would involve the modulation of gene expression at the transcription stage.

Promoter G-quadruplexes. G-quadruplex formations by the promoter sequences of various oncogenes have been characterized, including c-myc[44,45,225,228,236], c-kit[237-243], bcl-2[244], VEGF[245], HIF-1α[246], PDGF-A[247], retinoblastoma[248], KRAS[249], and
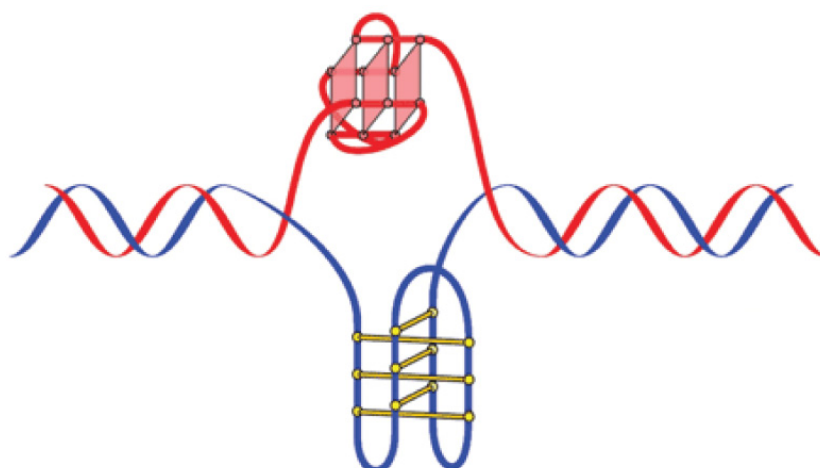
Figure 6.1 | Schematic diagram illustrating the putative simultaneous formation of i-motif and G-quadruplex. Formation of these non-canonical DNA structures might take part in the regulation of gene transcription. Figure adapted from ref 234.

RET[250,251]. Among these, high-resolution structures of c-myc[225,236], c-kit[239,241-243], bcl-2[244], and RET[251] promoter quadruplexes have been characterized. Most of these G-quadruplex structures conform to the propeller-type all-parallel-stranded topology. This is consistent with an overabundance of short linkers of 1–2 nt, which has a propensity to adopt the double-chain-reversal configuration[180,228,231,252], between consecutive guanine tracts for these motifs.

Telomerase inhibition as anticancer strategy. The enzyme telomerase catalyzes the addition of TTAGGG repeats onto the 3' ss overhang of the telomere. This action is performed by the reverse transcriptase component hTERT, which utilizes the RNA component (hTERC) as a template. Telomerase usually remains quiescent in normal somatic cells. However, it is found to be undesirably up-regulated in most cancer cells[154], helping to maintain their proliferative capacity. Hence inhibition of telomerase represents a promising anticancer strategy[157]. One avenue of research involves the rational design of ligands specifically targeting human telomeric G-quadruplexes [see Section 3.1], which have been shown to prevent telomere elongation by telomerase[159]. Direct targeting of the catalytic or RNA component of telomerase has been explored as a possible route for telomerase inhibition. For instance, there were attempts to knock down telomerase activity through the application of PNAs specific for hTERC. Another approach would be to selectively target quadruplexes in the promoter of hTERT, which has been found to exhibit a high propensity to fold into quadruplexes, thereby suppressing the expression of hTERT.

6.2   G-Quadruplexes in the hTERT Promoter[#]

High potential for G-quadruplex formation within the hTERT promoter. The potential of both the coding and noncoding strands of the hTERT promoter (defined as the region up to 240 nt upstream of the first exon) for G-quadruplex formation (at least four patches of ≥3 successive guanines) was examined with the G4P calculator[46], based on a 40-nt scanning window and a 1-nt frame shift. The noncoding strand yielded a G4P score of 34.17%, indicative of a high concentration of putative quadruplex-forming motifs. One particular segment, lying at a position –20 to –110 nt with respect to the transcription start site (TSS), comprises altogether fourteen G-tracts

---

[#] Results from this study were published in Lim *et al., JACS* (2010).

interspersed with linkers of 1–13 nt. Successive runs of four G-tracts (giving eleven sequences in total) were investigated for G-quadruplex formation, out of which five sequences exhibited the formation of stable quadruplexes having melting temperatures (T_m) of 52 °C to 65 °C at a low-salt environment containing only 10 mM K^+.

Adoption of two predominant G-quadruplex conformations by the 20-nt fragment *20gtert*. One particular fragment d[AGGGGAGGGGCTGGGAGGGC] (denoted as *20gtert*; Table 6.1), located at a position −41 to −60 nt upstream of the TSS, has a high G-content of 70% and gave rise to very stable G-quadruplex structures (T_m of 55 °C and 78 °C at 10 mM and 100 mM K^+, respectively). Judging from the intensity and number of H1 proton peaks at 10.9–12 ppm in the NMR spectrum, *20gtert* assumes two predominant quadruplex forms in potassium solution (Figure 6.2g). The NMR spectrum was fairly well-resolved to make further structural characterization possible. UV-melting studies were performed on *20gtert* under different K^+ concentrations, illustrating the decreasing folded fraction and stability of the G-quadruplexes with decreasing salt concentration. In addition, oligonucleotide concentration-independent melting profiles of *20gtert* (from 0.5–100 μM; Figure 6.3b) at a fixed K^+ environment indicated that these G-quadruplexes are intramolecular in nature.

Driving the folding of *20gtert* towards disparate G-quadruplex conformations. Judicious substitutions of guanines to inosines on *20gtert* (Table 6.1) could drive the folding topology of the oligonucleotide towards the two separate G-quadruplex conformations; individual replacement of G5 (*20gtert[I5]*) or G7 (*20gtert[I7]*) with inosine favored one conformation (denoted as Form 1; Figure 6.2a,b), while

Table 6.1 | Natural and inosine-substituted hTERT promoter sequence *20gtert* used in this study. Estimated compositions of the different conformations at 25 °C are listed.

Name	Sequence[a]	Population (%)		
		Form 1	Form 2	others
20gtert	A GGGG A GGGG CT GGG A GGG C	>40	<40	<20
20gtert[I2]	A **I**GGG A GGGG CT GGG A GGG C	–	>85	<15
20gtert[I5]	A GGG**I** A GGGG CT GGG A GGG C	>90	–	<10
20gtert[I7]	A GGGG A **I**GGG CT GGG A GGG C	>90	–	<10
20gtert[I10]	A GGGG A GGG**I** CT GGG A GGG C	–	>90	<10
20gtert[I5/I7]	A GGG**I** A **I**GGG CT GGG A GGG C	>90	–	<10
20gtert[I2/I10]	A **I**GGG A GGG**I** CT GGG A GGG C	–	>95	< 5
20gtert[+2]	AA GGGG A GGGG CT GGG A GGG CC	>50	<40	<10

[a] Inosine (I) residues are shown in bold.
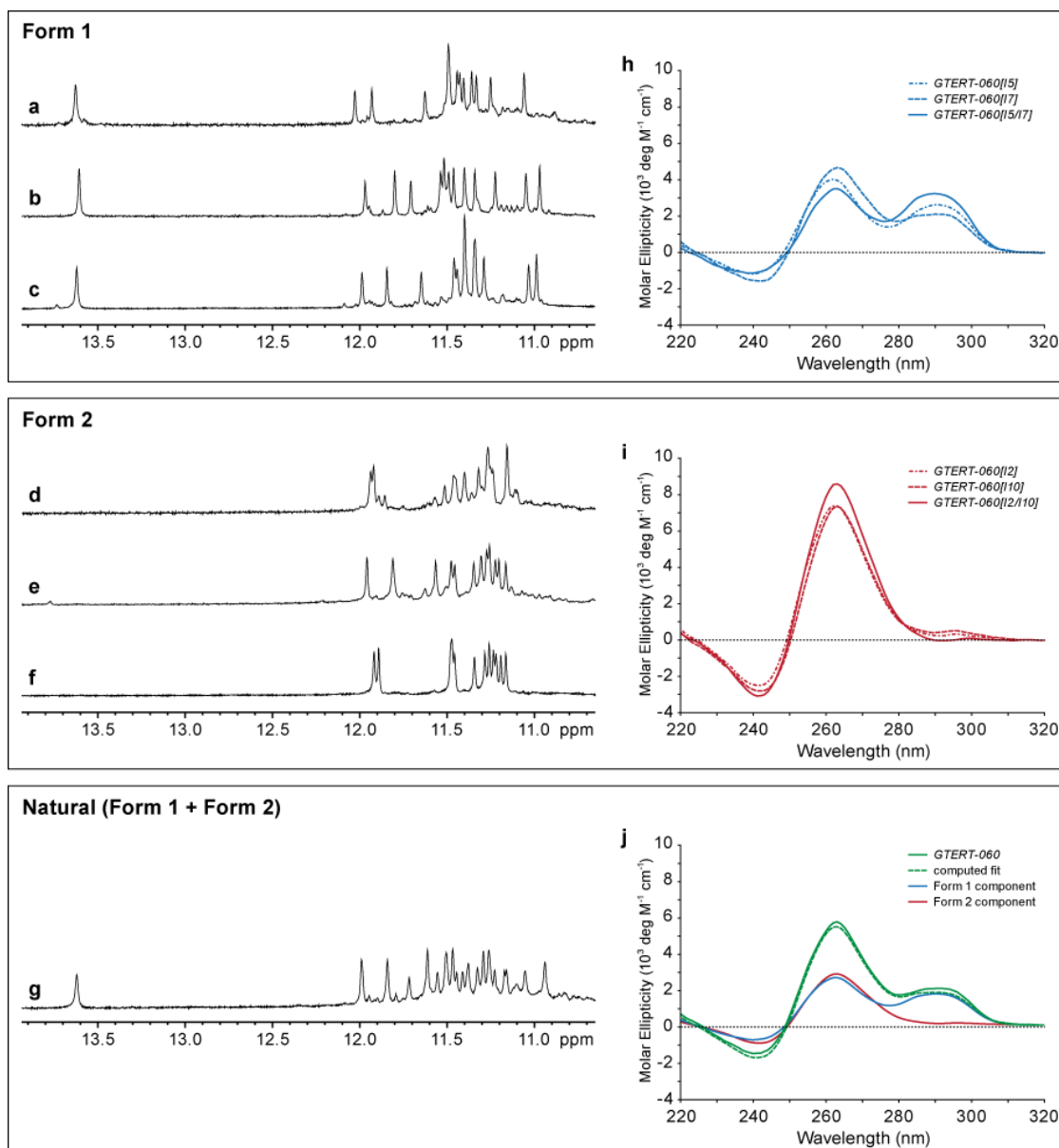
Figure 6.2 | NMR and CD spectra of *20gtert* in potassium solution. (**a–g**) 1D NMR H1 proton spectra of *20gtert[I5]* (**a**), *20gtert[I7]* (**b**), *20gtert[I5/I7]* (**c**), *20gtert[I2]* (**d**), *20gtert[I10]* (**e**), *20gtert[I2/I10]* (**f**), and *20gtert* (**g**). (**h–j**) CD spectra of Form 1 (**h**), Form 2 (**i**), and *20gtert* (**j**). In (**j**), CD spectrum of *20gtert* (solid line) is approximated by a computed fit (dotted line), which comprises both Form 1(blue) and 2 (red) components.

modification of G2 (*20gtert[I2]*) or G10 (*20gtert[I10]*) with inosine favored the other conformation (denoted as Form 2; Figure 6.2d,e). When applied in concert, i.e. (G5 and G7; *20gtert[I5/I7]*) or (G2 and G10; *20gtert[I2/I10]*), Form 1 and 2 were retained as the respective predominant species (Figure 6.2c,f), with *20gtert[I2/I10]* producing a superbly clean NMR spectrum. The reasoning behind these modifications will be described below with regards to the G-quadruplex structures.
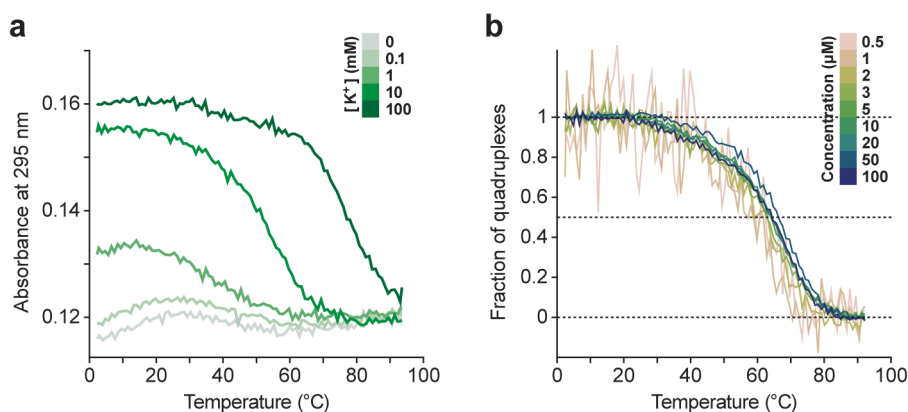
Figure 6.3 | UV-melting analyses of *20gtert* in K⁺ solution. (**a**) UV-melting of *20gtert*, as monitored through the absorbance at 295 nm, across various K⁺ concentrations. (**b**) A plot of fractions of folded structures of *20gtert* against temperature, measured across various oligonucleotide strand concentrations.

Twelve major H1 protons were seen from 10.9–12.0 ppm in the NMR spectra of Form 1 (Figure 6.2a–c), pointing to the establishment of a quadruplex core comprising three G-tetrad layers. An additional H1 proton could be seen at ~13.6 ppm at 25 °C. As opposed to the tetrad imino protons, this proton peak was broadened at 45 °C, suggesting its participation in the formation of an A•T base pair. The same signature peak for Form 1 was also observed for an expanded sequence *22gtert* (Table 6.1), supporting the possible existence of Form 1 G-quadruplex under a longer natural sequence context.

Twelve major imino proton peaks were also observed from 11.1–12.0 ppm for Form 2 (Figure 6.2d–f), pointing again to the establishment of a quadruplex core comprising three G-tetrad layers. The absence of the signature H1 proton in Form 2, coupled with a distinct distribution pattern of the tetrad imino proton peaks, implied the adoption of a G-quadruplex structure with a distinct folding topology from that of Form 1.

CD characterization of Form 1 and 2 G-quadruplexes. In potassium solution, Form 1 and 2 displayed rather distinct CD signatures at 25 °C (Figure 6.2h,i). Form 1 exhibited positive bands near 295 and 260 nm, a trough straddling them at ~280 nm, as well as a negative band near 240 nm. Form 2 displayed a positive maximum near 260 nm and a negative band around 240 nm. Previous studies have concurred that the peak at 260 nm reflects quadruplexes comprising stacked tetrads that are arranged in the same orientation, while the peak at 295 nm reflects quadruplexes in which two adjacent tetrads possess alternate polarities[195]. The CD signatures of both G-

quadruplexes were consistent with Form 2 assuming a propeller-type core topology and Form 1 adopting a (3+1) quadruplex scaffold (see below).

At 25 °C, CD spectrum of *20gtert* in potassium solution (solid green line, Figure 6.2j) suggested that it consists of both Form 1 and 2 components; there is a positive peak close to 260 nm, a negative trough around 240 nm, and a shoulder peak spanning across 280–300 nm. This would align with the observations from NMR, which indicated the coexistence of Form 1 and 2 for *20gtert* in K^+ solution (Figure 6.2g). The relative populations of both species in *20gtert* could be estimated based on a linear fit to its CD spectrum, C = αA + (1-α)B. Here C represents the composite CD spectrum, A/B corresponds to the component CD spectrum, while α and (1-α) define the proportions of the two species. Relative populations of ~50–80% and ~20–50% were obtained for Form 1 and 2, respectively (Figures 6.2j and 6.4), subjected to the component spectra used in the fitting procedure. This was in good agreement with the observations from NMR (Figure 6.2g and Table 6.1).

Interconversion between two G-quadruplex forms. Equilibrium CD spectra of *20gtert* were measured across the temperature range 5–90 °C (Figure 6.5a) at increments of 5 °C, in ~100 mM K^+. The equilibrium was shifted towards Form 2 with increasing temperature (exceeding 25 °C), consistent with observations from NMR spectra (5–45 °C; Figure 6.6). The conversion (from Form 1 to Form 2) could be tracked by the signal change at 295 nm (Figure 6.5b, blue curve), which comprises minimal contribution from Form 2. The 260-nm band, on the other hand, comprises
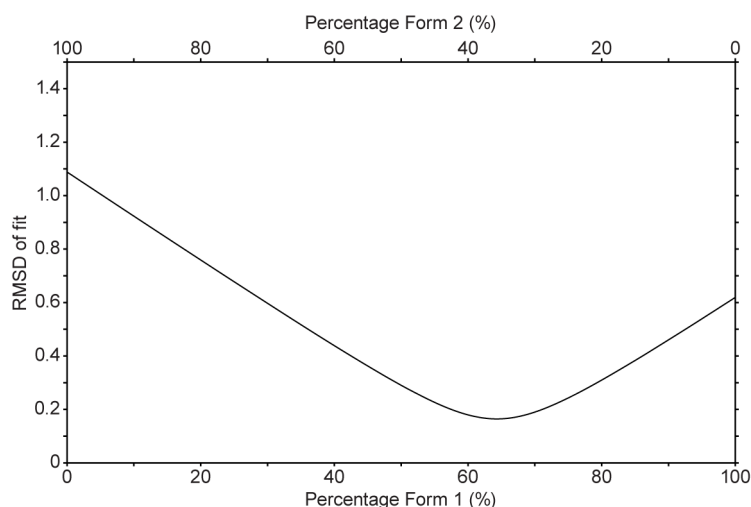
Figure 6.4 | RMSD of the computed fit based on varying proportions of Form 1 and 2 against CD spectrum of *20gtert*. The lowest RMSD of 0.165 is obtained with relative populations of Form 1 and 2 at 64% and 36%, respectively.
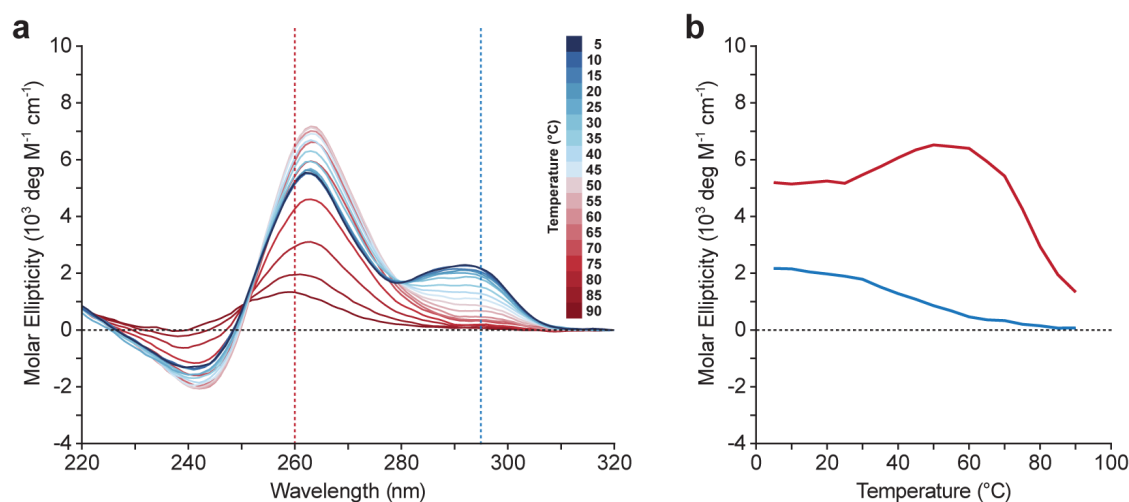
Figure 6.5 | CD-melting profile of *20gtert*. (a) Equilibrium CD spectra of *20gtert* across the temperature range from 5 to 90 °C. **(b)** Peak intensity at 260 nm (red) and 295 nm (blue) can be used to monitor the interconversion between the two forms.

signals from both conformations. The signal intensity rose initially from 25 °C to 50 °C, corresponding to an increase in the proportion of Form 2. At temperatures greater than 50 °C, the melting process overcame the conversion of Form 1 to Form 2, leading to a continuous drop in signal intensity all the way to 90 °C (Figure 6.5b, red curve). This temperature-dependency of the CD signals could potentially be applied for
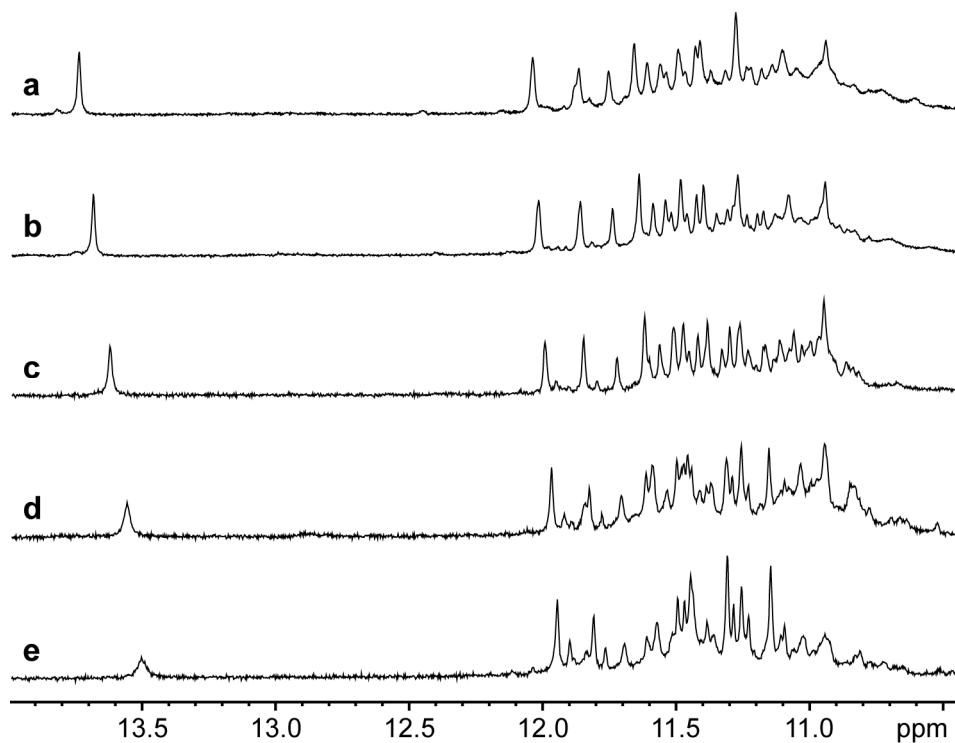
Figure 6.6 | 1D NMR imino proton spectra of *20gtert* across various temperatures. (a) 5°C, **(b)** 15°C, **(c)** 25°C, **(d)** 35°C, and **(e)** 45°C.

the dissection of other sequences that could adopt multiple structural forms (with distinct CD signatures).

The equilibrium was again shifted towards the propeller-type G-quadruplex (Form 2) under water-depleted condition, which could be generated with the introduction of the crowding agent polyethylene glycol (PEG). The CD spectra of the samples showed varying degree of conversion over time (Figure 6.7), commensurate with the compositions of PEG added (from 10 to 40% v/v). Total conversion to Form 2 was nearly achieved at 30% PEG. These observations were in agreement with previous findings, which demonstrated that water-depleted (or molecular crowding) condition steers a conversion towards the propeller-type G-quadruplex conformation[190,253].

NMR spectral assignments. Structural characterization was carried out for Form 1 (*20gtert[I5]*) and Form 2 (*20gtert[I2/I10]*) hTERT promoter G-quadruplexes. For both oligonucleotides, unambiguous assignments for H1 and H8 protons of guanines were achieved based on site-specific 2% ^{15}N-enrichment[192] and ^{2}H labeling[207] of samples, respectively, coupled with through-bond correlations[193] (Figures 6.8 and 6.9). The NOE sequential H8/H6–H1' connectivities for both sequences were subsequently completed from A1 to C20 (Figures 6.10 and 6.11).

Form 1 corresponds to a (3+1) G-quadruplex. For *20gtert[I5]*, the strong intranucleotide H8–H1' NOE cross-peaks for G2, G8, G9, G13, and G17 (Figure 6.10) pointed to their adoption of the *syn* glycosidic conformation. The rest of the guanine residues adopt the *anti* glycosidic conformation. Cyclic H1–H8 NOE connectivity
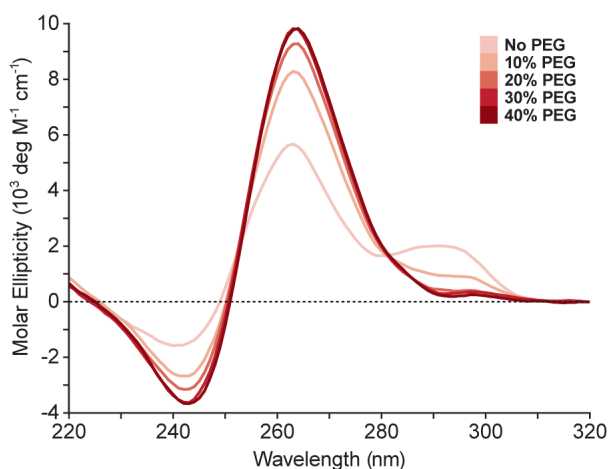
Figure 6.7 | CD spectra of *20gtert* under water-depleted conditions. The equilibrium of *20gtert* between Form 1 and 2 (pink) is steered towards Form 2 alone (maroon) with increasing composition of PEG added.
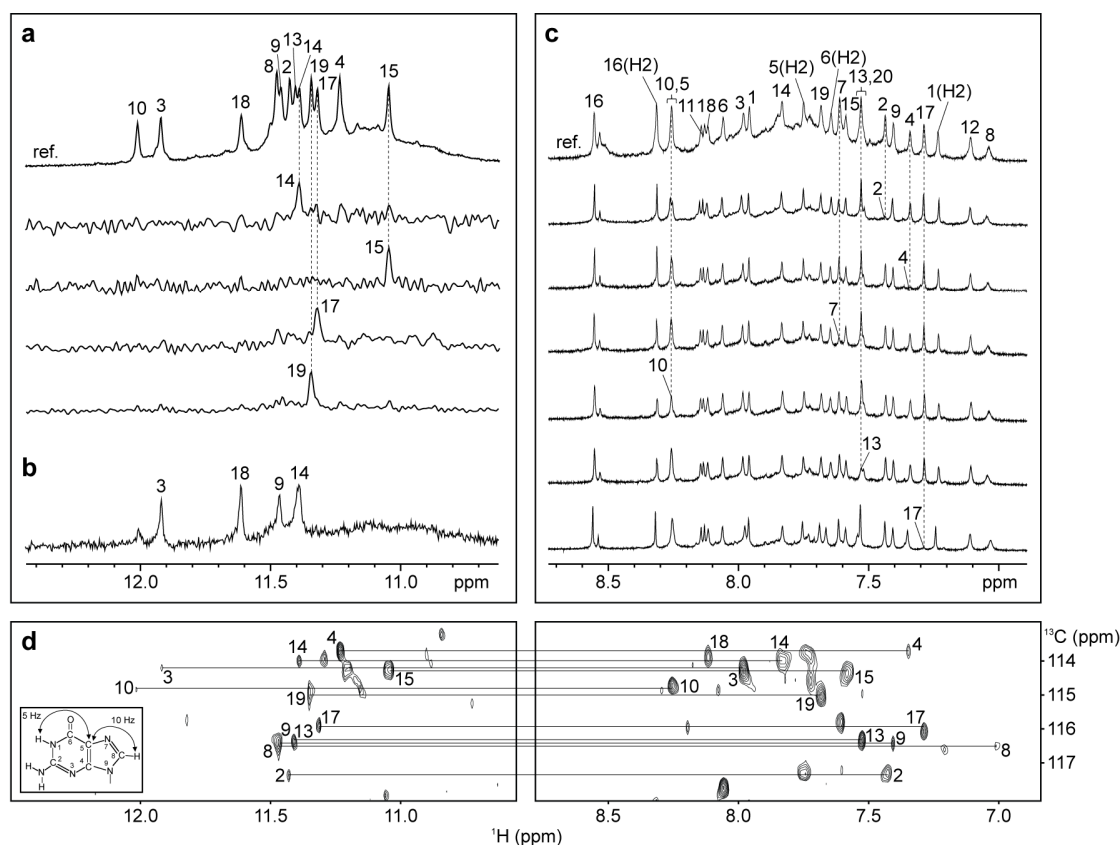
Figure 6.8 | Assignment of guanine H1 and H8 protons for *20gtert[I5]* (Form 1). (a) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. (**b**) H1 proton spectra in [2]H$_2$O for 30 min at 25°C. (**c**) Guanine H8 proton assignment for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. (**d**) JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to [13]C5.

patterns within individual tetrads were determined (Figure 6.12a), revealing the formation of G2•G17•G13•G10, G3•G9•G14•G18, and G4•G8•G15•G19 tetrads, whose hydrogen-bond directionalities are respectively anticlockwise–clockwise–clockwise. Accordingly, guanines constituting the top tetrad are arranged with glycosidic conformations of *syn•syn•syn•anti*, whereas those making up the other two tetrads are arranged with glycosidic conformations of *anti•syn•anti•anti*. The alignment of the core strands corresponded to a (3+1) G-tetrad scaffold (Figure 6.12c), with the loops being successively lateral–lateral–propeller. The first edgewise loop connects across a wide groove at the bottom, the second edgewise loop doubles back at the top across a narrow groove, while the third propeller loop (A16) mediates two parallel strands over a medium groove.
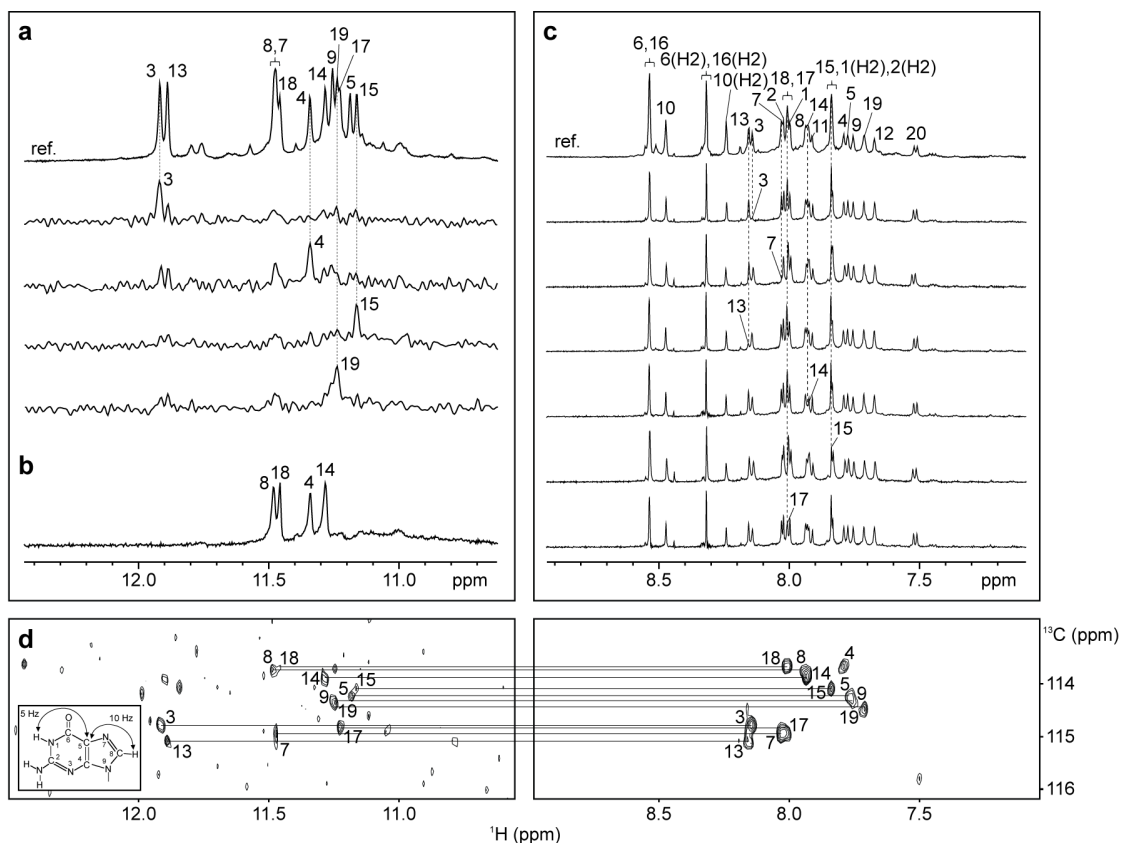
Figure 6.9 | Assignment of guanine H1 and H8 protons for *20gtert[I2/I10]* (Form 2). (**a**) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. (**b**) H1 proton spectra in [2]H$_2$O for 30 min at 25°C. (**c**) Guanine H8 proton assignment for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. (**d**) JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to [13]C5.

K$^+$ solution structure of *20gtert[I5]* G-quadruplex (Figure 6.13) was computed utilizing NMR restraints (Appendix I). The G-tetrad scaffold is the same as those observed for (3+1) human telomeric G-quadruplexes in potassium solution[165-167,169-174]. The central positioning of G3•G9•G14•G18 in the tetrad core was supported by the slower solvent exchange rate of these imino protons as compared to those from the other two tetrads (Figure 6.8b).

The signature H1 proton peak around 13.6 ppm arises from the formation of reverse Hoogsteen A1•T12 base pair at the top. T12 is positioned right over G10 and G13, supported by NOE cross-peaks between G10 and T12 (for both base and sugar protons), and between G13 (imino proton) and T12 (sugar protons). The alignment of A1 was determined by NOE cross-peaks between its H8 proton and the imino protons of G2 and G10. The bases of C11 (from the top edgewise loop) and A16 (from the

85
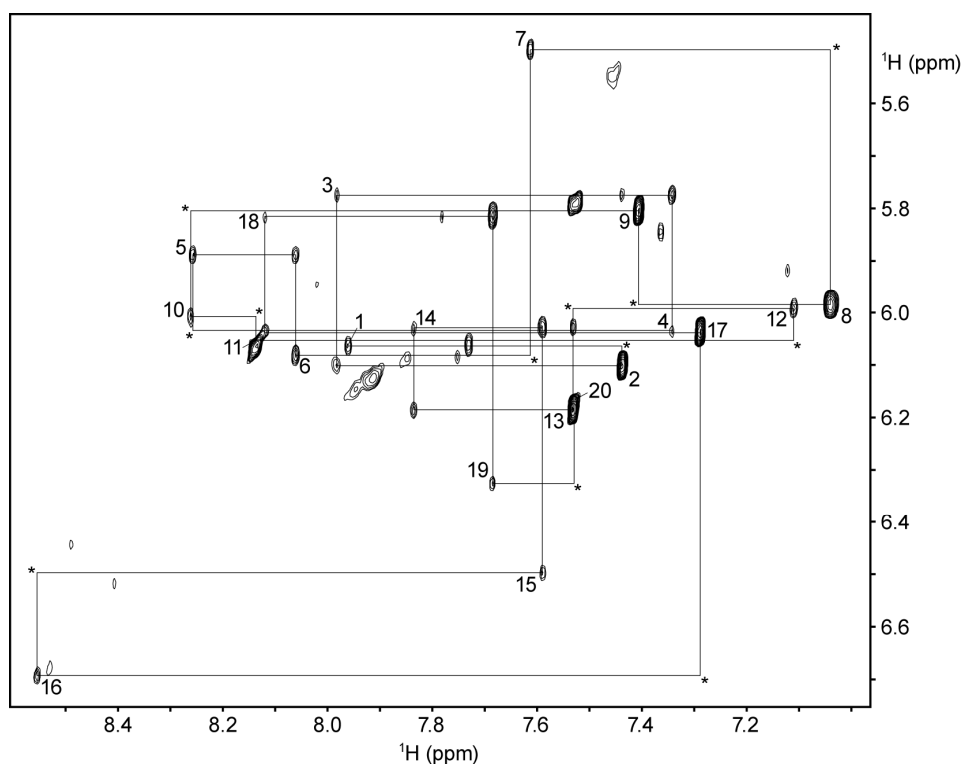
Figure 6.10 | NOESY spectrum showing the sequential connectivity of *20gtert[I5]*. H6/H8–H1' NOE connectivity of *20htert[I5]* from A1 to C20. Residue numbers are marked on the corresponding intraresidue cross-peaks. Asterisks indicate cross-peaks that are weak or missing.
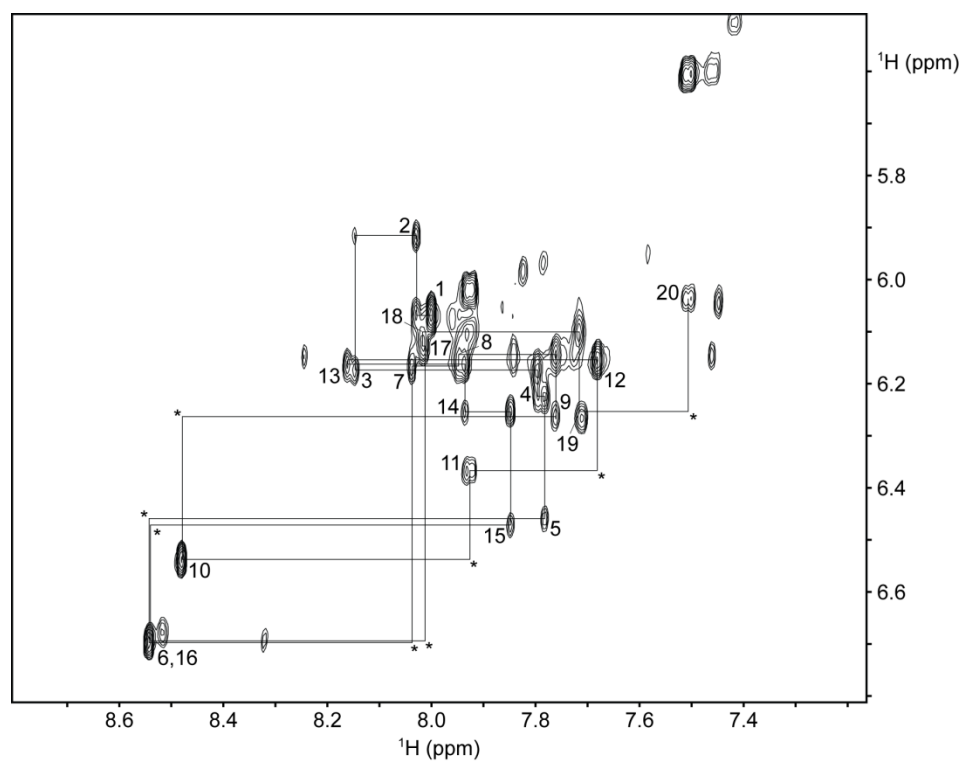
Figure 6.11 | NOESY spectrum showing the sequential connectivity of *20gtert[I2/I10]*. H6/H8–H1' NOE connectivity of *20htert[I2/I10]* from A1 to C20. Residue numbers are marked on the corresponding intraresidue cross-peaks. Asterisks indicate cross-peaks that are weak or missing.

86
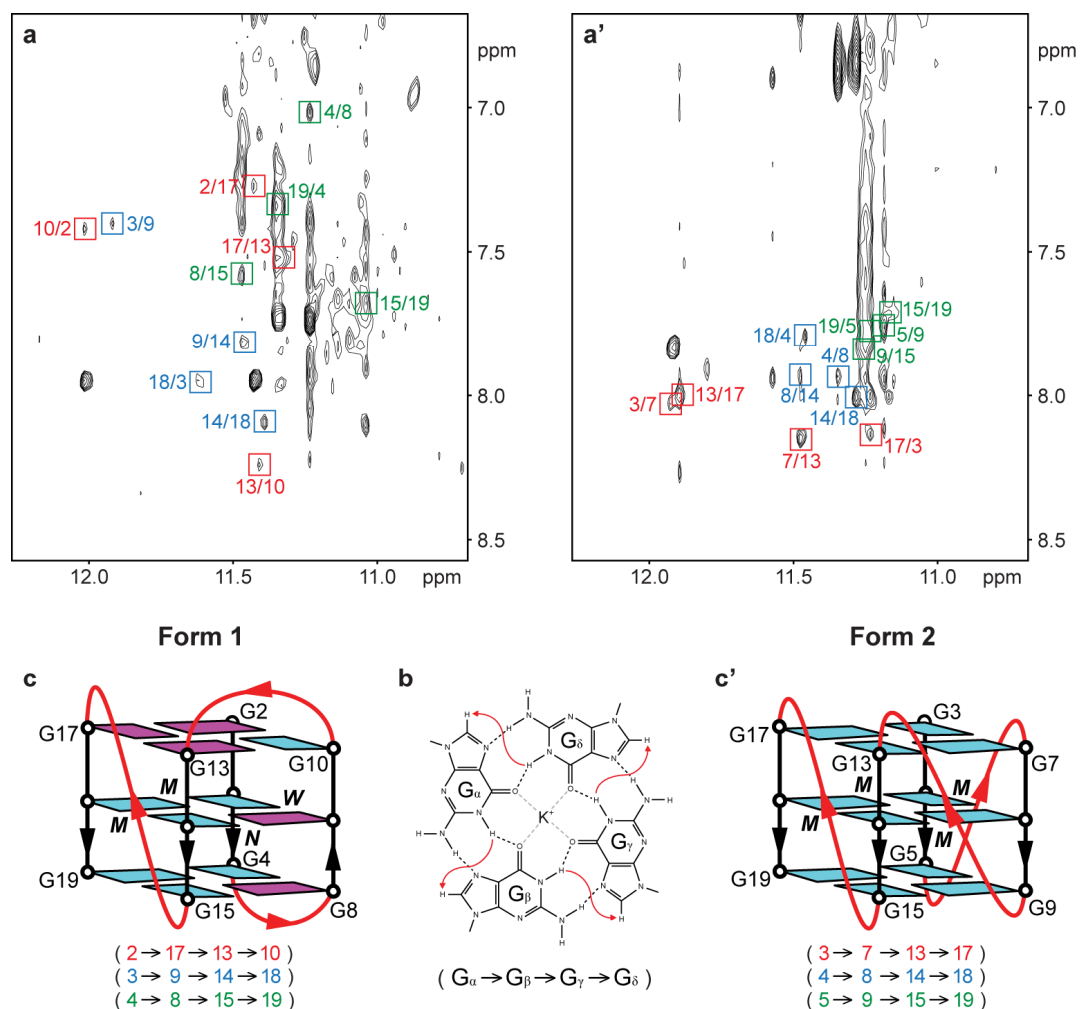
Figure 6.12 | Folding topologies of Form 1 and 2 hTERT promoter G-quadruplexes in potassium solution. (**a**,**b**) NOESY plots of *20gtert[I5]* (**a**) and *20gtert[I2/I10]* (**a'**) highlighting cross-peaks corresponding to the characteristic guanine H1–H8 NOE connectivity (**b**) within individual tetrad. (**c**) Schematic structures of Form 1 (**c**) and 2 (**c'**) hTERT promoter G-quadruplexes, which exhibit hydrogen-bond directionalities of tetrads as shown below.

propeller loop) are projected away from their respective loops, as supported by a dearth of NOEs between these bases and their nearby residues. The structures of the I5–A6–G7 loop and C20 below the tetrad core are less well-defined, perhaps reflecting some degree of conformational dynamics for these residues.

NOESY spectrum of *20gtert[I7]* indicated that it assumes the same (3+1) G-quadruplex folding topology as *20gtert[I5]*; excellent spectral correspondence was observed between the two, including similar H6/H8–H1' NOE sequential connectivity, as well as the existence of five *syn* intranucleotide H8–H1' NOE correlations.

Form 2 corresponds to a propeller-type G-quadruplex. Judging from the moderate intensity of intranucleotide H8–H1' NOE correlations (Figure 6.11), guanine residues of *20gtert[I2/I10]* all adopt the *anti* glycosidic conformation. Cyclic H1–H8

87
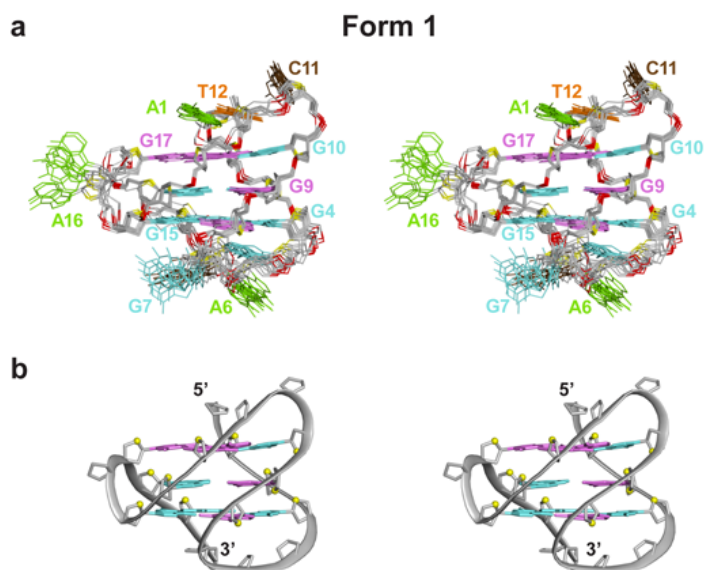
Figure 6.13 | NMR structure of *20gtert[I5]* (3+1) G-quadruplex (Form 1) in potassium solution. (a) Stereo view of ten computed structures of *20gtert[I5]* G-quadruplex, aligned based on the tetrad core. **(b)** Stereo view of a representative conformer in ribbon representation.

NOE connectivity patterns within individual tetrads were determined (Figure 6.12a′), revealing the formation of G3•G7•G13•G17, G4•G8•G14•G18, and G5•G9•G15•G19 tetrads, all of which possess the same clockwise hydrogen-bond directionality. Parallel alignment of the four strands gives rise to a propeller-type tetrad core (Figure 6.12c′), which is linked by three double-chain-reversal loops. All four grooves have medium widths.

K$^+$ solution structure of *20gtert[I2/I10]* G-quadruplex (Figure 6.14) was computed utilizing NMR restraints (Appendix I). The all-parallel-stranded tetrad core can be compared to that adopted by human telomeric propeller-type G-quadruplex observed in a crystallized state[33]. The central positioning of G4•G8•G14•G18 in the tetrad core was supported by the slower solvent exchange rate of these imino protons as compared to those from the other two tetrads (Figure 6.9b).

The bases of A1 and C20 are stacked over either end of the tetrad core, while the base of I2 collapses onto the groove. A1 and T12 could potentially establish an A•T Watson–Crick base pair at the top. I10 is projected into the groove, possibly interacting with the core bases lining the groove. C11 from the same loop is flipped outwards, in contact with the solvent. The near-identical arrangement of the two single-nucleotide propeller loops A6 and A16 were supported by the closely matched spectral characteristics of the two fragments (G3–G9 and G13–G19) constituting these
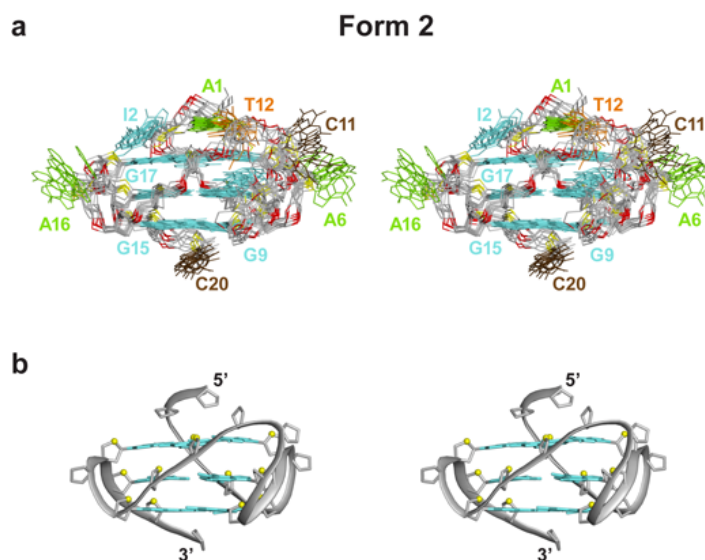
Figure 6.14 | NMR structure of *20gtert[I2/I10]* all-parallel-stranded G-quadruplex (Form 2) in potassium solution. (**a**) Stereo view of ten computed structures of *20gtert[I2/I10]* G-quadruplex, aligned based on the tetrad core. (**b**) Stereo view of a representative conformer in ribbon representation.

edges of the tetrad core, including NOE patterns and chemical shifts (Figures 6.9 and 6.11).

The moderate intensity of intranucleotide H8–H1' correlations in the NOESY spectrum of *20gtert[I10]* indicated that the guanines all adopt an *anti* conformation, corroborating its adoption of the propeller-type architecture.

Formation of G-quadruplexes in the hTERT promoter. The hTERT promoter comprises a high concentration of putative quadruplex-forming motifs. The segment on the noncoding strand positioned at −20 to −110 nt with respect to the TSS exhibits a high propensity to form G-quadruplexes, with five out of eleven four-G-tract sequences from this region capable of forming stable G-quadruplexes (T_m ~52–65 °C) under a low-salt environment (10 mM K$^+$). In a long sequence context, various G-quadruplexes could possibly arise from different combinations of G-tracts, including the involvement of G-tracts that are not immediately following one another. However, one might still expect that the rapid kinetics of G-quadruplex formation from four successive G-tracts should lead to the predominant populations on the folding pathway[254,255].

Coexistence of two interconvertible quadruplex forms. *20gtert*, a hTERT promoter fragment having a high G-content, was shown to exist in a state of equilibrium between two predominant intramolecular quadruplex forms in potassium

solution; Form 1 corresponds to a (3+1) G-quadruplex while Form 2 corresponds to an all-parallel-stranded G-quadruplex. These two quadruplex conformations exhibit similar energy level, being interconvertible in solution. The all-parallel-stranded Form 2 G-quadruplex is favored both at high temperatures[255] (Figures 6.5 and 6.6) and under water-depleted (or molecular crowding) condition (Figure 6.7), which emulates the intracellular environment[256].

Robust G-quadruplex structural elements. Form 1 hTERT promoter (3+1) G-quadruplex is stabilized by several robust structural elements. The core element, a (3+1) G-tetrad scaffold, was first shown in four-repeat *Tetrahymena* telomeric sequences in sodium solution[37]. It has since been revealed in various contexts, including bcl-2 promoter[244] and human telomeric repeats[165-167,169-174]. The overall architecture is further stabilized by a reverse Hoogsteen A•T capping the top, together with a single-nucleotide propeller loop (Figure 6.13), a robust G-quadruplex structural element[180,228,231,252].

The stability of Form 2 propeller-type hTERT promoter G-quadruplex arises from the adoption of two robust single-nucleotide propeller loops over opposite sides of the quadruplex core. These two linkers fit snugly across the all-parallel-stranded core, locking them into a compact arrangement. The three-nucleotide middle propeller loop further interacts with bases lining the groove, contributing to the stability of the overall structure.

A sizeable count of quadruplex-forming motifs were identified in the human genome[41,42]. Many of these sequences, in particular those found in oncogenic promoters[47], contain successive guanines interspersed with short linkers of one or two nt, which display a propensity to adopt the propeller configuration[180,228,231,252]. There is a consensus that the presence of these short loop motifs would drive the folding of a sequence towards the propeller-type quadruplex topology. In the case of *20gtert*, a considerable population (>40%) of the oligonucleotide still adopts the Form 1 (3+1) quadruplex structure, in spite of the existence of these short loop motifs. This would agree with our previous observations that the establishment of the major G-quadruplex form is affected to a large extent through the formation of stabilizing loop interactions[196] [see Section 3.2.1].

Favoring/disfavoring separate G-quadruplex conformations through inosine substitutions. Successive runs of four or more guanines within a G-tract, as is the case

for *20gtert* at the first two G-tracts, introduce flexibility in the establishment of a three-G-tetrad quadruplex structure; at least two distinct combinations of guanine triplet from the G-tract could partake in the establishment of tetrad core. In this study, appropriate substitutions of guanines to inosines[37,225,226,257], either in isolation or in concert, were employed to disfavor the guanines at the particular positions from taking part in G-tetrad formation[226], thus tilting the equilibrium towards the alternative form. Specifically, Form 1 (3+1) G-quadruplex was favored by inosine incorporation at G5 and/or G7, whereas Form 2 all-parallel-stranded G-quadruplex was favored by inosine incorporation at G2 and/or G10 (Figure 6.2 and Table 6.1).

CD signals from coexisting G-quadruplex conformations. CD spectrum of *20gtert* in potassium solution displays a sharp positive peak close to 260 nm, a negative trough near 240 nm, and a shoulder stretching from 280–300 nm (Figure 6.2j). We have shown that this arises from a mixture of two predominant species, Form 1 (3+1) and Form 2 all-parallel-stranded G-quadruplexes, which could be selectively favored over one another through judicious inosine substitutions (Figure 6.2 and Table 6.1). In addition, we have dissected the contributions of both forms towards the mixture based on a linear fit to its CD spectrum using component CD spectra of the individual conformations (Figures 6.2j and 6.4).

Changes in the relative population of the two conformations can be monitored through NMR and CD spectra (Figures 6.5 and 6.6). We have shown that variations in the experimental conditions (temperature and introduction of crowding agent), together with alterations to the sequence of *20gtert*, could affect the equilibrium between the two forms. A direct characterization on the interconversion between them was greatly facilitated through the availability of CD spectra for the isolated conformations.

CD signatures of Form 1 (two positive peaks near 260 and 295 nm) and Form 2 (huge positive signal around 260 nm) were in line with those from other (3+1) and propeller-type G-quadruplexes, respectively[169,172,195,196,258]. The only discrepancy is that the 260-nm peak has a higher intensity than the 295-nm peak for Form 1, while the opposite trend was reported in the past for other (3+1) G-quadruplexes[169,172,196,258]. This could have arisen simply due to differences in base stacking within the loops and/or across the loops and the core.

Conclusion

Telomerase, an enzyme which catalyzes the addition of TTAGGG repeats onto the ss telomeric overhang, is erroneously up-regulated in most cancer cells. Hence the knockdown of its activity has been envisioned as a potential anticancer treatment. Various approaches have been explored towards this end, ranging from the direct targeting of its catalytic and RNA components, to the induction/stabilization of telomeric G-quadruplexes. Yet another avenue of research would involve the specific targeting of G-quadruplexes that are formed within the promoter of its catalytic subunit hTERT, which was found to be extremely G-rich, and the subsequent repression of hTERT expression.

The present work demonstrated potential G-quadruplex formation within hTERT promoter sequences, focusing on a 20-nt fragment (*20gtert*) situated at –41 and –60 nt upstream of the TSS. In K^+ solution, *20gtert* was shown to coexist as two distinct G-quadruplex forms, the high-resolution structures of which were elucidated using NMR. The equilibrium and interconversion of the two conformations were dissected with NMR and CD through a combination of sequence modifications and alterations to experimental condition.

The two coexisting conformations, a (3+1) and a propeller-type G-quadruplexes, shed light on how disparate structural elements would come together to favor a particular structure. Detailed structural/biochemical studies can be extended to other G-rich sequences within the hTERT promoter, in the hope of identifying a suitable therapeutic target for rational ligand design.

Chapter 7

Structures of Quadruplex–Duplex Hybrids and G-Junction Motifs

7.1   Duplex & Quadruplex DNA

7.2   Structural Basis of DNA Quadruplex–Duplex Junction Formation

7.1   Duplex & Quadruplex DNA

Biological relevance and functional applications of duplex and quadruplex.
The canonical conformation of DNA is the B-form double helix[1-3] (or duplex), which
consists of the antiparallel pairing of two complementary strands. B-DNA is
ubiquitous in biology, being implicated in almost all chromosomal processes. It is also
widely involved in molecular biology and biotechnological applications, e.g.
molecular cloning and polymerase chain reaction (PCR). Structural variants of duplex
DNA include the four-way Holliday junction[14,15] [see Section 1.1.2], which mediates
the exchange of duplex strands during genetic recombination. B-DNA and the
Holliday junction have been adapted in the design of DNA nanomaterials[80] and hold
promise for wide-ranging nanotechnological applications.

Apart from the duplex structure, DNA is capable of adopting alternative forms
including the G-quadruplex[27-30]. The G-quadruplex is a helical complex constructed
through the multiple stacking of G•G•G•G tetrads[31]. A multitude of folding topologies
have been described for the G-quadruplex, giving rise to a humongous structural
diversity of these complexes. The formation of G-quadruplex was implicated in
cellular processes including recombination[54] and replication[55-58], and in a natural
context, this would occur in close proximity to the regular duplex[59]. On the other hand,
quadruplexes were demonstrated to exhibit important functional properties [see Section
1.1.4], and they were also incorporated in DNA nanoarchitectures[259] and
nanodevices[260].

To date, there is a lack of knowledge with regards to how a quadruplex can form
next to regular duplex DNA. An understanding on the compatibility between the two
disparate conformations could potentially lead to novel approaches for the targeting of
genomic G-rich sequences[30,235,261]. In addition, knowledge on duplex–quadruplex
connectivity could pave the way for the integration G-rich motifs in DNA structural
engineering[262] and nanotechnology[80].

7.2   Structural Basis of DNA Quadruplex–Duplex Junction Formation

Duplex and quadruplex geometry. Canonical B-DNA[1-3] consists of two
antiparallel right-handed polynucleotide chains. The four bases A, T, G, and C are

---

[#] Results from this study were published in Lim & Phan, *Angew.Chemie* (2013).
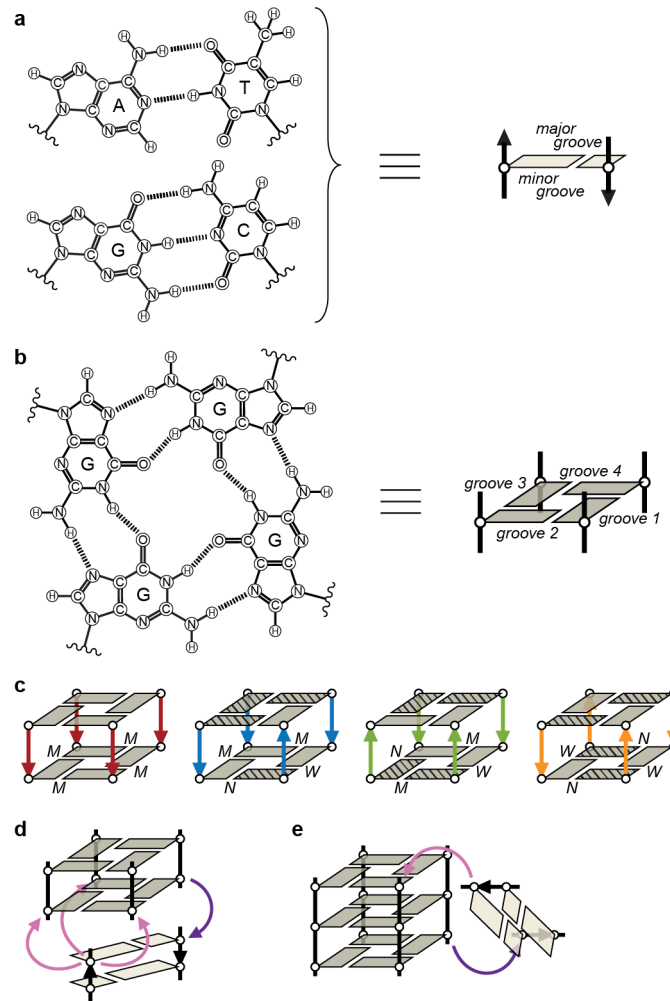
Figure 7.1 | Geometry of duplex and quadruplex DNA and their connectivity. (a) Watson–Crick A•T and G•C base pairs. (**b**) The G•G•G•G tetrad. (**c**) Four distinct G-quadruplex core topologies. (**d**,**e**) Different possibilities for the coaxial (**d**) and perpendicular (**e**) arrangement of quadruplex and duplex helix.

packed along the hydrophobic core through Watson–Crick base pairing (Figure 7.1a). The phosphate backbones have a strand separation (henceforth defined as the distance between matching sets of phosphates) of ~18 Å, and the relative spacing between them defines the major and minor groove. The building block of G-quadruplex[27-30] is the G-tetrad[31] (Figure 7.1b), which is held together by Hoogsteen hydrogen bonds[19] and cation coordination through its middle. Multiple G-tetrads stack together to give rise to a G-quadruplex. The four strands making up the G-tetrad core can be aligned in four different relative orientations (Figure 7.1c), and there are a variety of ways to join them across the corners with linkers (known as loops). These different possibilities result in a diverse range of folding topologies. The different types of core have distinct combinations of groove widths, and there are altogether three types of groove: wide grooves are defined by two antiparallel strands, with strand separations of ~19 Å;

medium grooves are defined by two parallel strands, with strand separations of ~16 Å; narrow grooves are defined by two antiparallel strands (having an alternate orientation as that of the wide grooves), with strand separations of ~12 Å.

Design of DNA constructs. To tackle the connectivity between duplex and quadruplex helix, the possible connection of a duplex hairpin over the various dimensions of a quadruplex was explored. The reference duplex hairpin (5'-CGCGAAGCATTCGCG-3') was arbitrarily designed (with a GCA loop[263]) and its structure was validated with NMR (Figure 7.2–7.5 and Appendix I). Assuming that one strand of the duplex is anchored to one corner of the tetrad core, there would be at least three ways to join the remaining duplex strand onto the tetrad core to produce a coaxial arrangement of the duplex and quadruplex helix (Figure 7.1d). On the other hand, orthogonal orientation of the duplex and quadruplex helix through a sideway connection (Figure 7.1e) might also be possible. For any particular connection between the duplex and the quadruplex, as long as the arrangement is geometrically correct and

Figure 7.2 | Assignment of guanine H1 and H8 protons for the reference duplex hairpin. (**a**) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. (**b**) H1 proton spectrum at 5°C. (**c**) Assignment of guanine H8 protons for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions.
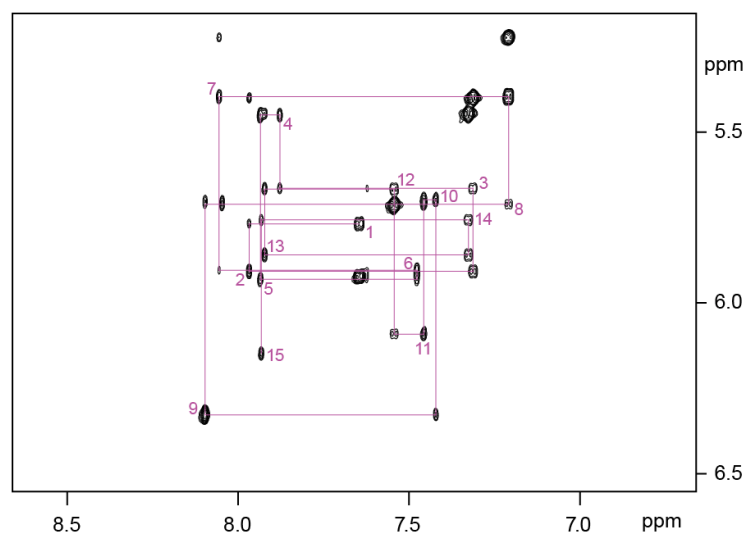
Figure 7.3 | NOESY spectrum showing the sequential connectivity of the reference duplex hairpin. H6/H8–H1' NOE connectivity of the reference duplex hairpin from C1 to G15. Residue numbers are marked on the corresponding intraresidue cross-peaks.

Figure 7.4 | Folding topology of the reference duplex hairpin in K$^+$ solution. (**a**) NOESY plot of the reference duplex hairpin highlighting the characteristic cross-peaks corresponding to T(H3)–A(H2) for A•T Watson–Crick base pair (**b**) and G(H1)–C(H41)/G(H1)–C(H42) pair for G•C Watson–Crick base pair (**c**). (**d**) Schematic structure of the reference duplex hairpin.
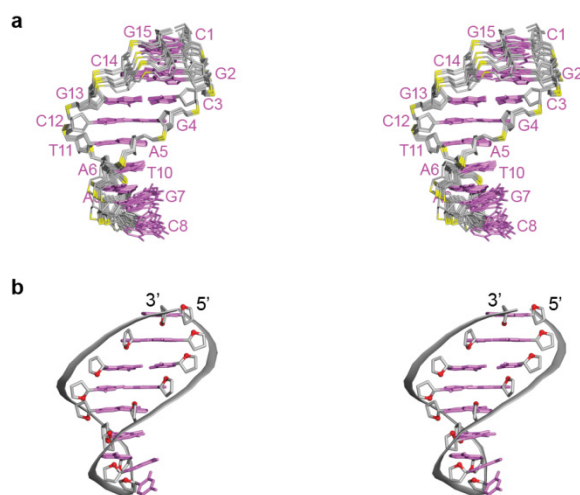
Figure 7.5 | NMR structure of the reference duplex hairpin in potassium solution. (a) Stereo view of ten computed structures of the reference duplex hairpin. (**b**) Stereo view of a representative conformer in ribbon representation.

does not lead to unfavorable energetic penalty, a stable quadruplex–duplex junction could be formed in principle. We assimilated our findings into the elucidation of the solution structures (buffer contains 20 mM KCl in 20 mM potassium phosphate, pH 7.0) for five representative constructs that cover the fundamental modes of connection between quadruplex and duplex DNA (Figures 7.6 and 7.7). In each design, strategic positioning of the hairpin and additional stabilizing motifs resulted in the adoption of a single predominant quadruplex–duplex hybrid, and structural characterizations were carried out based on rigorous spectral assignment approaches.

Compatibility of a duplex across a quadruplex wide groove. A duplex minor groove has a strand separation of ~18 Å while a quadruplex wide groove has a strand separation of ~19 Å. In addition, the strand directionality across both groove types is the same. This pointed to the possibility of a seamless connection between the two, with a gradual transition at the junction. Indeed, a stable quadruplex–duplex construct was established upon incorporation of the reference duplex hairpin onto the wide groove of a chair-type quadruplex[35] (Construct I; Figures 7.6a,7.8–7.11 and Appendix I). Continuous base stacking was observed across the quadruplex–duplex interface, giving rise to a tight junction. The compatibility between these two groove types was shown to apply across antiparallel quadruplexes with different topologies (Figure 7.12).
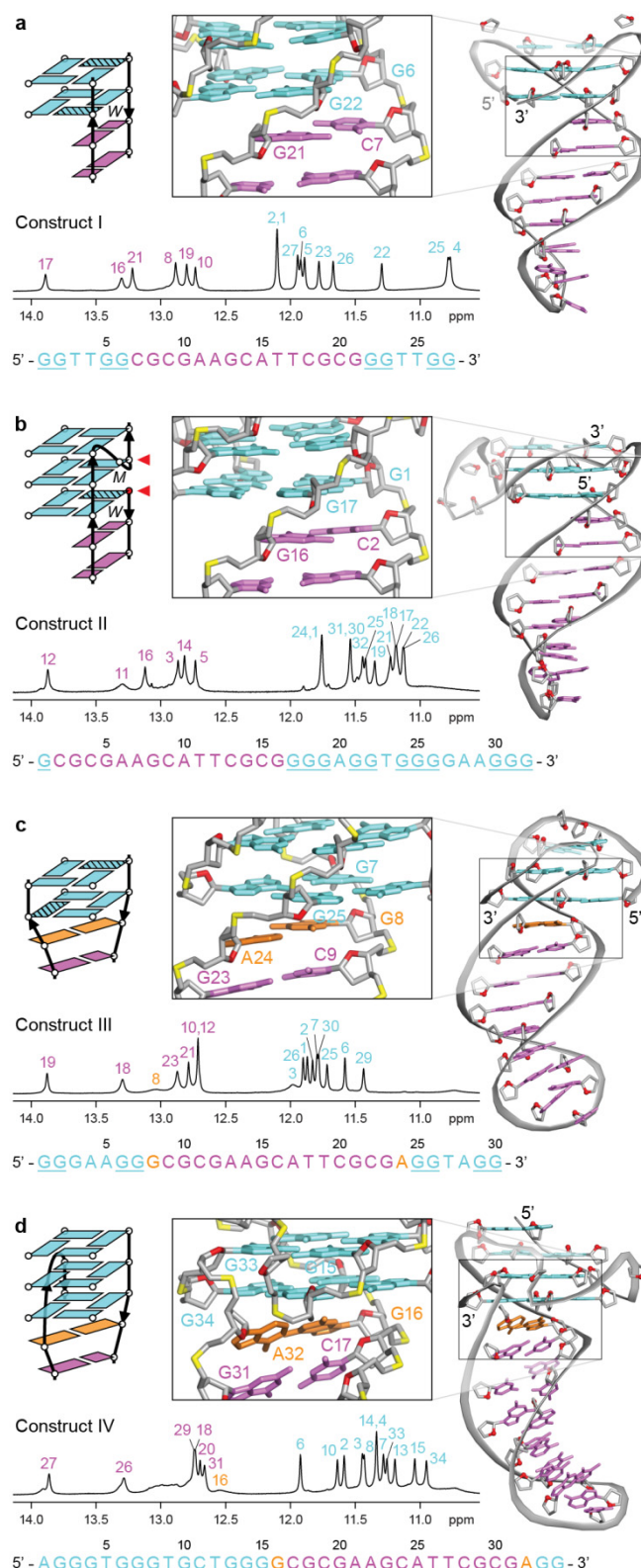
Figure 7.6 | Coaxial attachment of a duplex onto a quadruplex. For each construct, the sequence and 1D NMR H1 spectrum are shown together with a schematic illustrating the attachment strategy. (**a**) Continuous extension of a duplex from the wide groove. (**b**) Accommodation of a duplex across the medium-made-wide groove through a nick. (**c**) Stretching a duplex across the diagonal corners with an adaptor G•A base pair. (**d**) Sidestepping the unfavorable attachment of a duplex onto the narrow groove with a snapback approach.
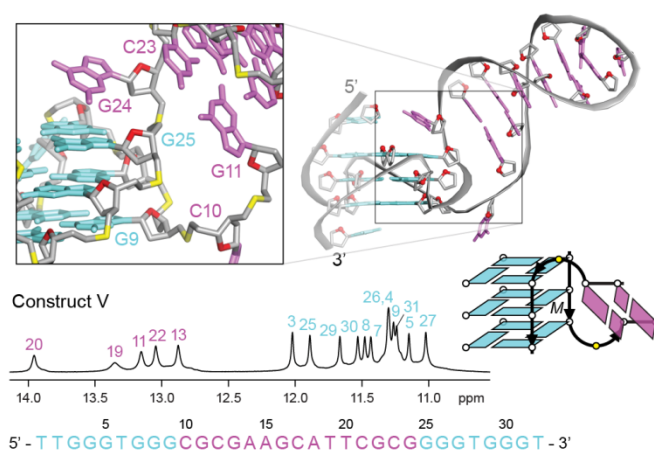
Figure 7.7 | Orthogonal attachment of a duplex onto a quadruplex. Sideway connection of a duplex onto a quadruplex, wherein the base pair at the junction is disrupted.
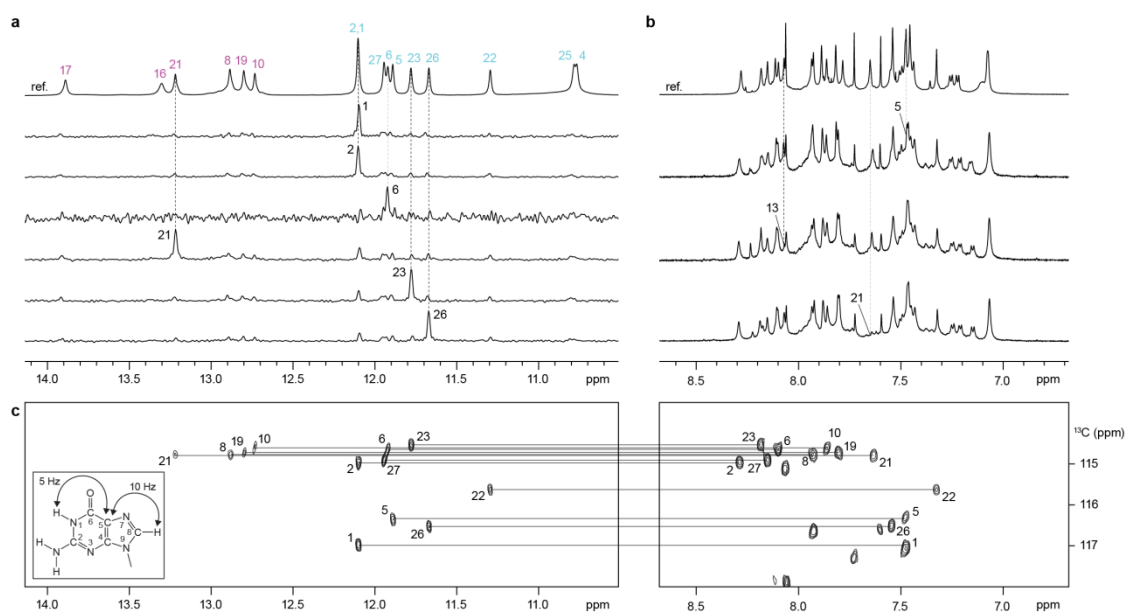
Figure 7.8 | Assignment of guanine H1 and H8 protons for quadruplex–duplex Construct I. (a) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. **(b)** Assignment of guanine H8 protons for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. **(c)** JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to ^{13}C5.
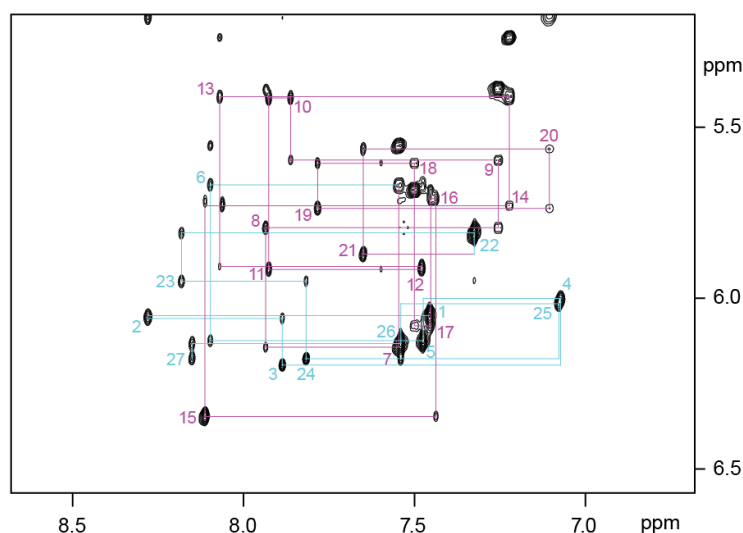
Figure 7.9 | NOESY spectrum showing the sequential connectivity of quadruplex–duplex Construct I. H6/H8–H1' NOE connectivity of Construct I from G1 to G27. Intraresidue cross-peaks are marked with the respective residue numbers.
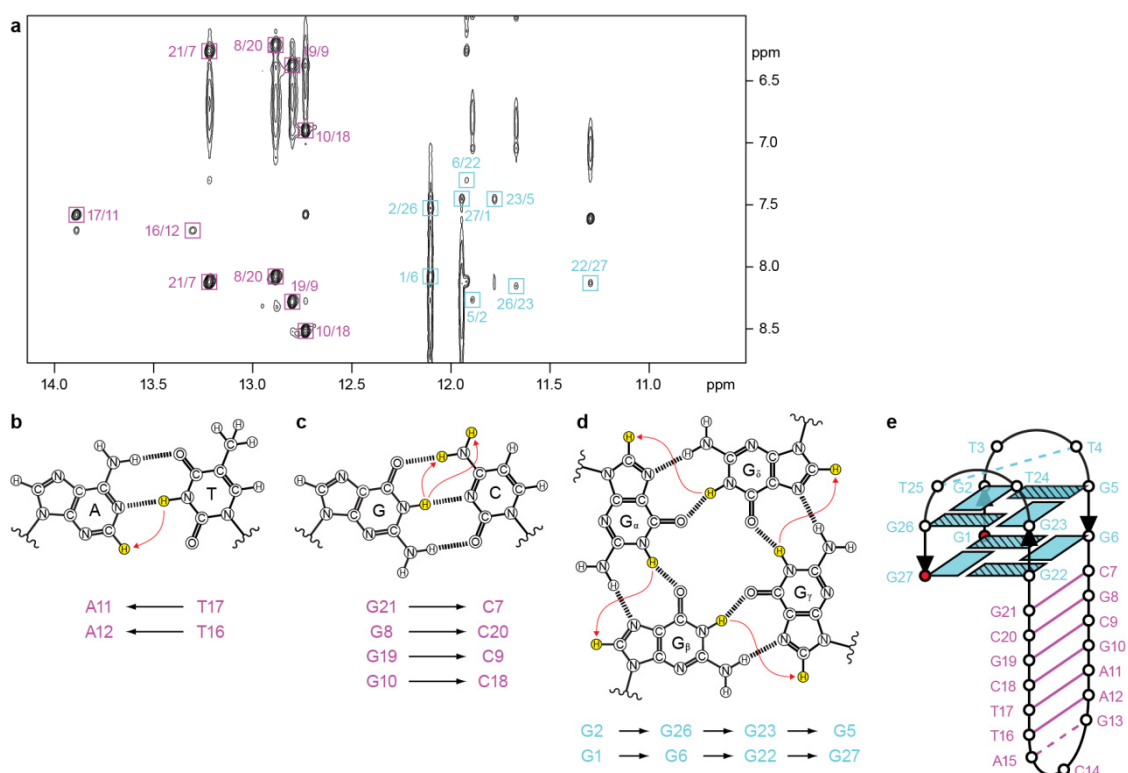
Figure 7.10 | Folding topology of quadruplex–duplex Construct I in K⁺ solution. (**a**) NOESY plot of Construct I highlighting the characteristic cross-peaks corresponding to T(H3)–A(H2) for A•T Watson–Crick base pair (**b**) and G(H1)–C(H41)/G(H1)–C(H42) for G•C Watson–Crick base pair (**c**), as well as guanine H1–H8 NOE connectivity (**d**) around each tetrad. (**e**) Schematic structure of quadruplex–duplex Construct I.
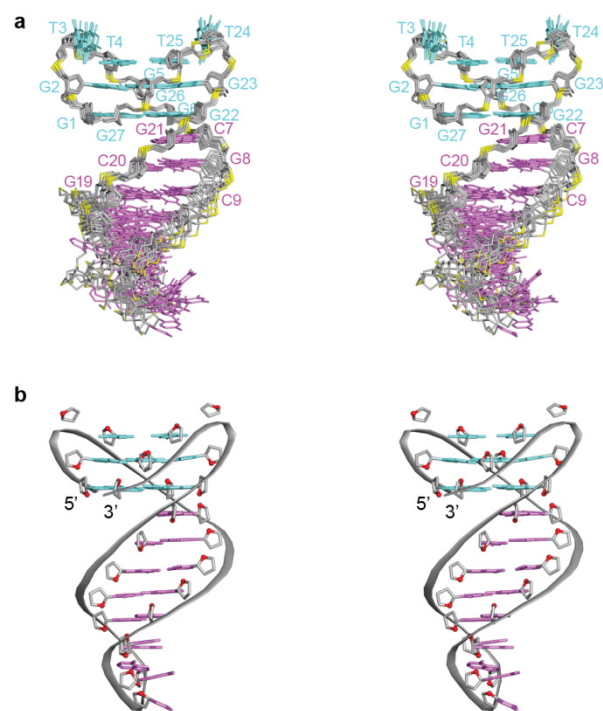
Figure 7.11 | NMR structure of quadruplex–duplex Construct I in potassium solution. (**a**) Stereo view of ten computed structures of Construct I, aligned based on the tetrad core. (**b**) Stereo view of a representative conformer in ribbon representation.

Generating a medium-made-wide groove through a nick. The medium groove is defined by two strands that are aligned in parallel, with a strand separation of ~16 Å. Hence the coaxial attachment of a regular duplex directly onto a medium groove would be geometrically impossible. Previously, a nick in the medium groove was observed to allow the interlocking of two quadruplex subunits[75]. This approach could be adapted here for the coaxial attachment of a duplex onto a quadruplex medium groove (Construct II; Figures 7.6b,7.13–7.16 and Appendix I). Introduction of a nick within the G-column would generate a strand discontinuity in the quadruplex core, thereby allowing a duplex strand to anchor onto the terminal G-tetrad in the reverse strand orientation[225], either at the 5'- or 3'-end (Figure 7.17). A wide groove (strand separation ~18–19 Å) is generated locally at the medium groove edge of the quadruplex (i.e. a medium-made-wide groove), and continuous stacking of bases was observed across the quadruplex and duplex segments, similar as in Construct I.

102
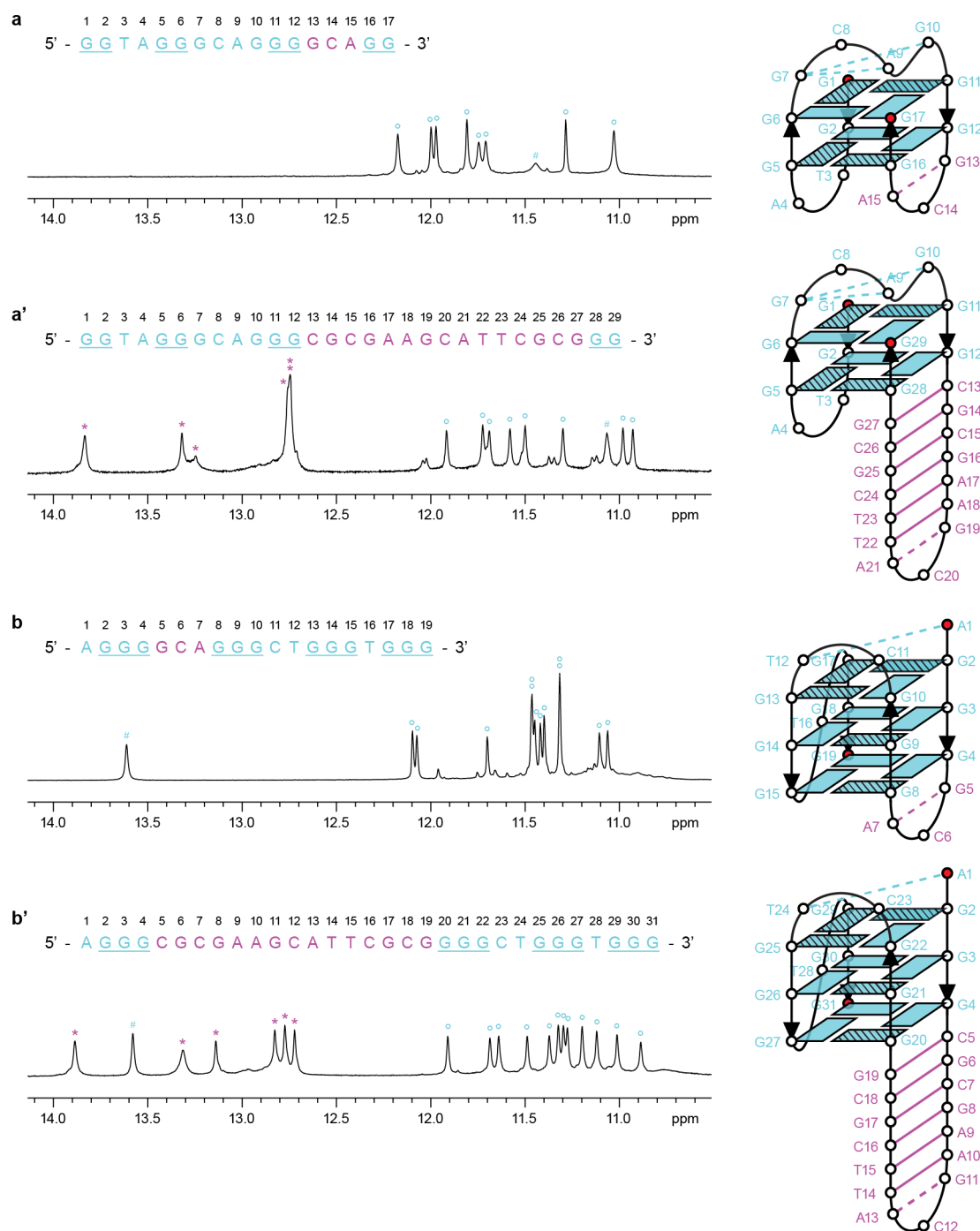
Figure 7.12 | Compatibility of duplex over the wide grooves of antiparallel G-quadruplexes. (**a**,**b**) A duplex can be attached seamlessly onto wide grooves of basket[226] (**a**) and (3+1)[233] (**b**) G-tetrad cores to generate the corresponding duplex–quadruplex constructs (**a'** and **b'**, respectively). In each panel, sequence and H1 proton spectrum of the oligonucleotide are shown together with the schematic diagram for the putative fold adopted by the oligonucleotide.
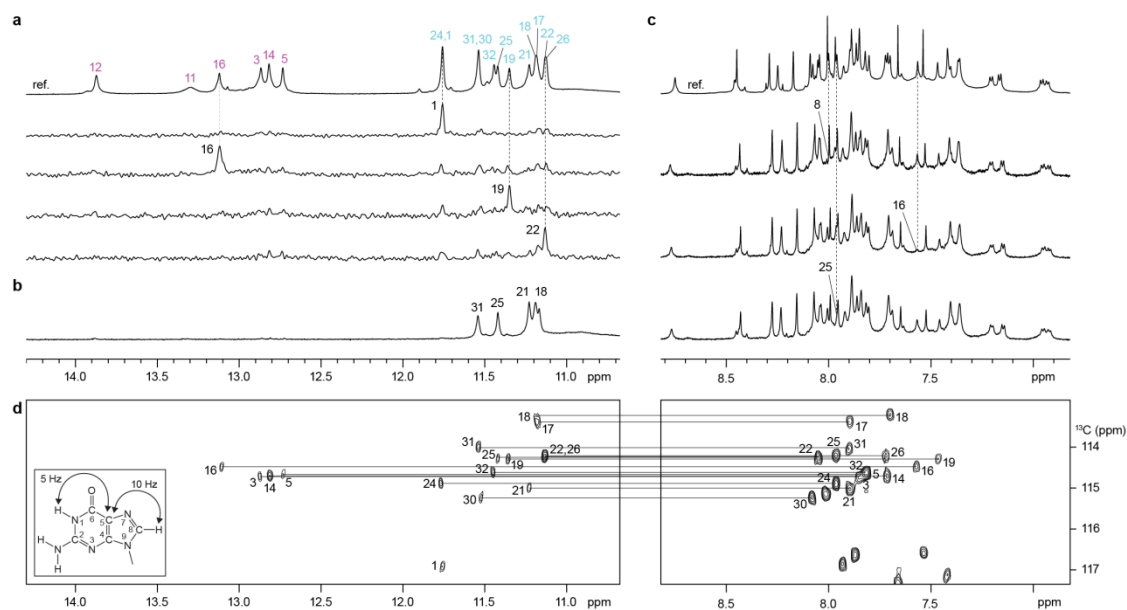
Figure 7.13 | Assignment of guanine H1 and H8 protons for Construct II. (a) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. (**b**) H1 proton spectra in [2]H$_2$O for 2 h at 25°C. (**c**) Assignment of guanine H8 protons for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. (**d**) JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to [13]C5.
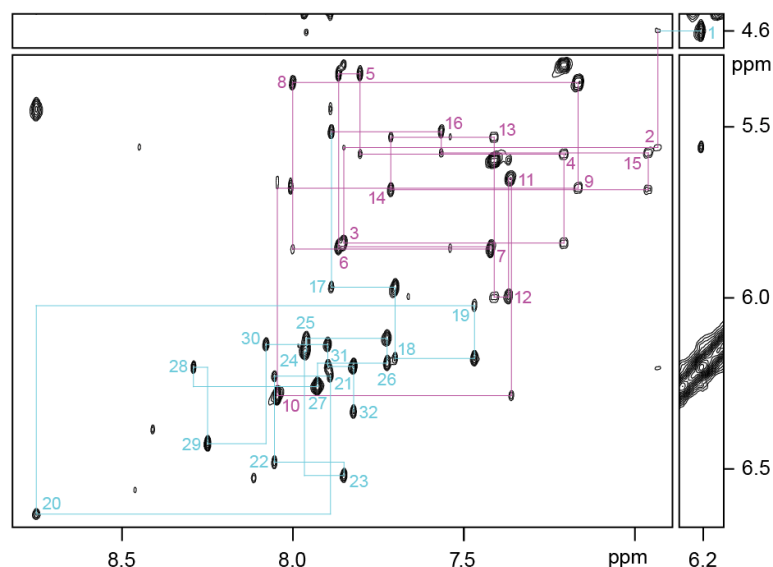
Figure 7.14 | NOESY spectrum showing the sequential connectivity of Construct II. H6/H8–H1' NOE connectivity of Construct II from G1 to G32. Residue numbers are marked on the corresponding intraresidue cross-peaks.
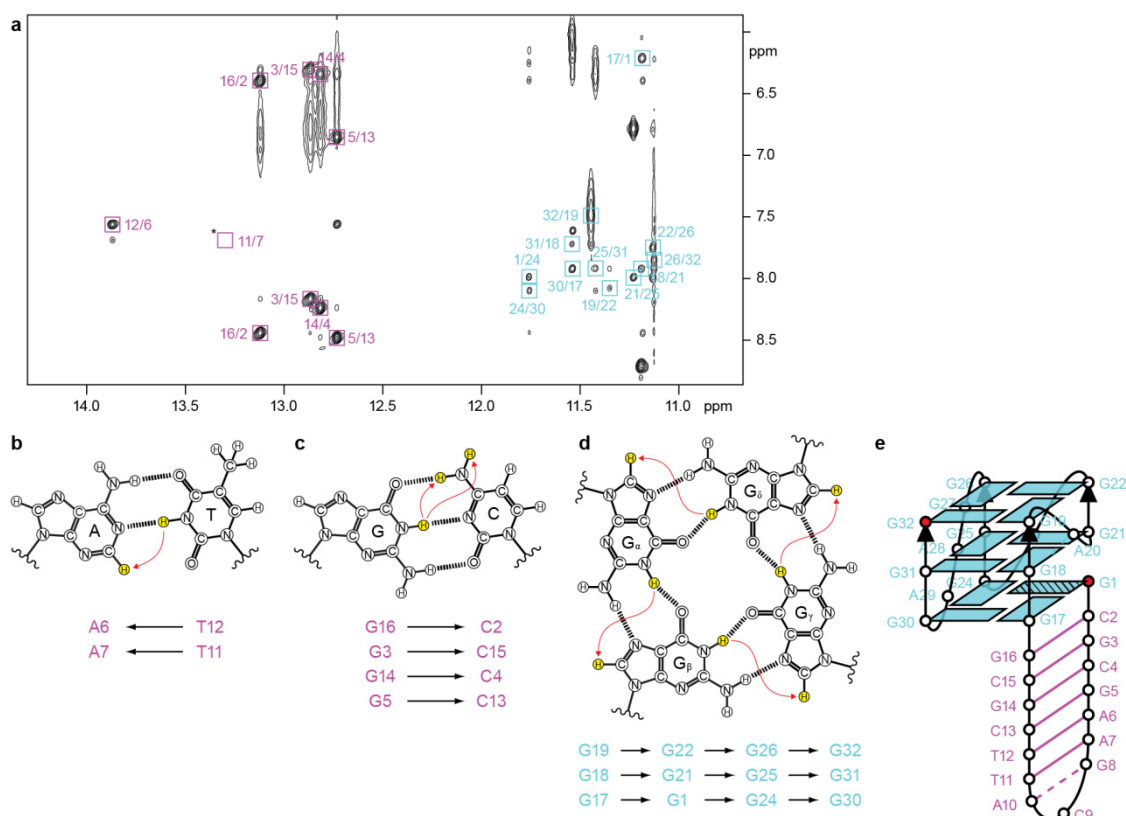
Figure 7.15 | Folding topology of quadruplex–duplex Construct II in K⁺ solution. (a) NOESY plot of Construct II highlighting the characteristic cross-peaks corresponding to T(H3)–A(H2) for A•T Watson–Crick base pair (**b**) and G(H1)–C(H41)/G(H1)–C(H42) for G•C Watson–Crick base pair (**c**), as well as guanine H1–H8 NOE connectivity (**d**) around each tetrad. (**e**) Schematic structure of quadruplex–duplex Construct II.
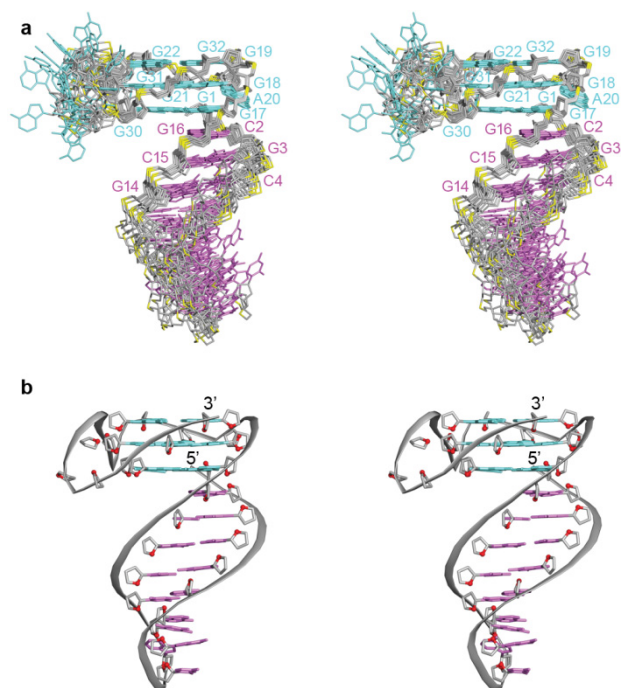
Figure 7.16 | NMR structure of quadruplex–duplex Construct II in potassium solution. (a) Stereo view of ten computed structures of Construct II, aligned based on the tetrad core. (**b**) Stereo view of a representative conformer in ribbon representation.
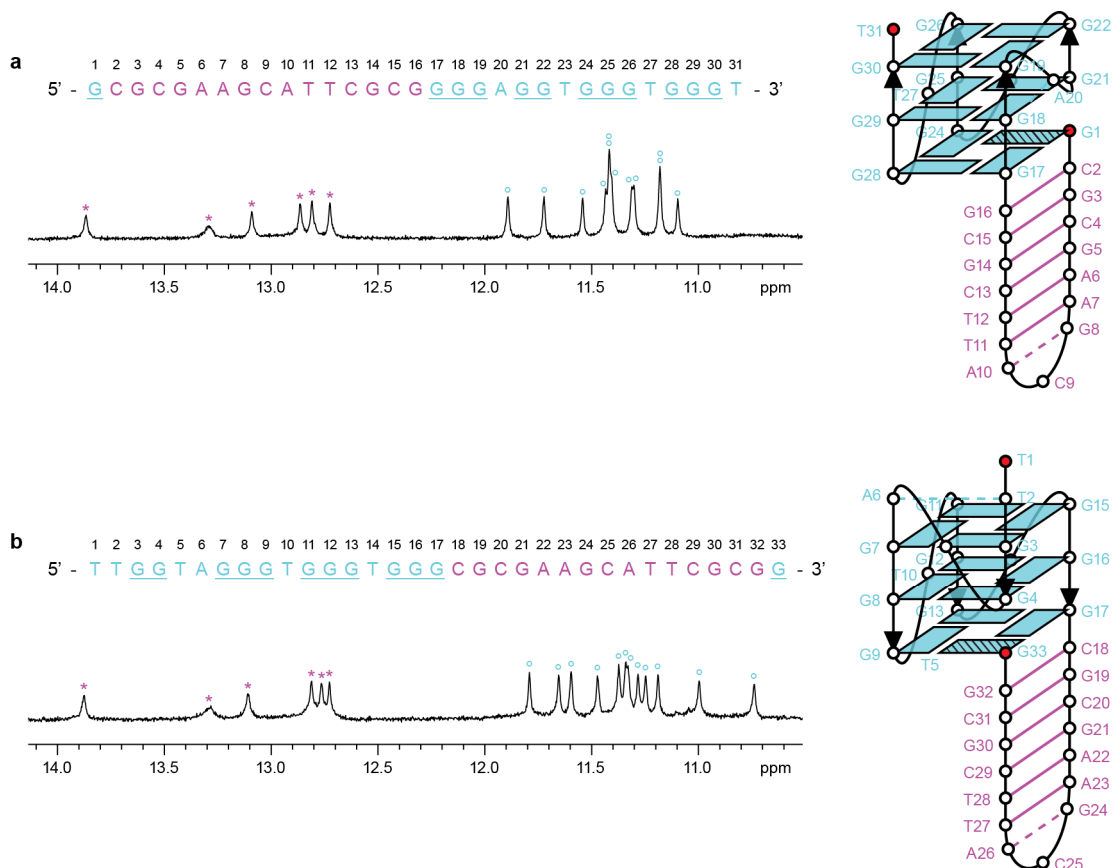
Figure 7.17 | Introduction of a nick on the G-column to mediate insertion of a loose duplex strand onto the terminal G-tetrad. (**a**,**b**) The terminal G-tetrad of a quadruplex can serve as a foothold for the 5'- (**a**) and 3'- (**b**) terminals of a loose duplex strand. In each panel, sequence and H1 proton spectrum of the oligonucleotide are shown together with the schematic diagram for the putative fold adopted by the oligonucleotide.

Connecting a duplex over diagonal corners of a tetrad. Phosphate groups on the diagonal corners of a tetrad are separated by >20 Å. Correspondingly, in order to for a duplex to span across this dimension of the quadruplex (extending from a strand separation of ~18 Å), some sort of adaptor would be required to mediate the expansion in strand separation. A comprehensive list of non-canonical base pairs was described previously[264], among which include purine–purine Watson–Crick mismatch base pairs. One particular base pair, G•A Watson–Crick mismatch, would fit the expansion in strand separation, and can be formed in an antiparallel strand orientation. Furthermore, it has been observed previously as a structural motif in G-quadruplexes [see Section 3.2.1]. Indeed, incorporation of the G•A Watson–Crick mismatch at the junction allowed the connection of the reference duplex hairpin across the diagonal corners of an antiparallel quadruplex (Construct III; Figures 7.6c,7.18–7.21 and Appendix I). Due to an associated twist in helical progression, the duplex helical axis is displaced with respect to the quadruplex helical axis.
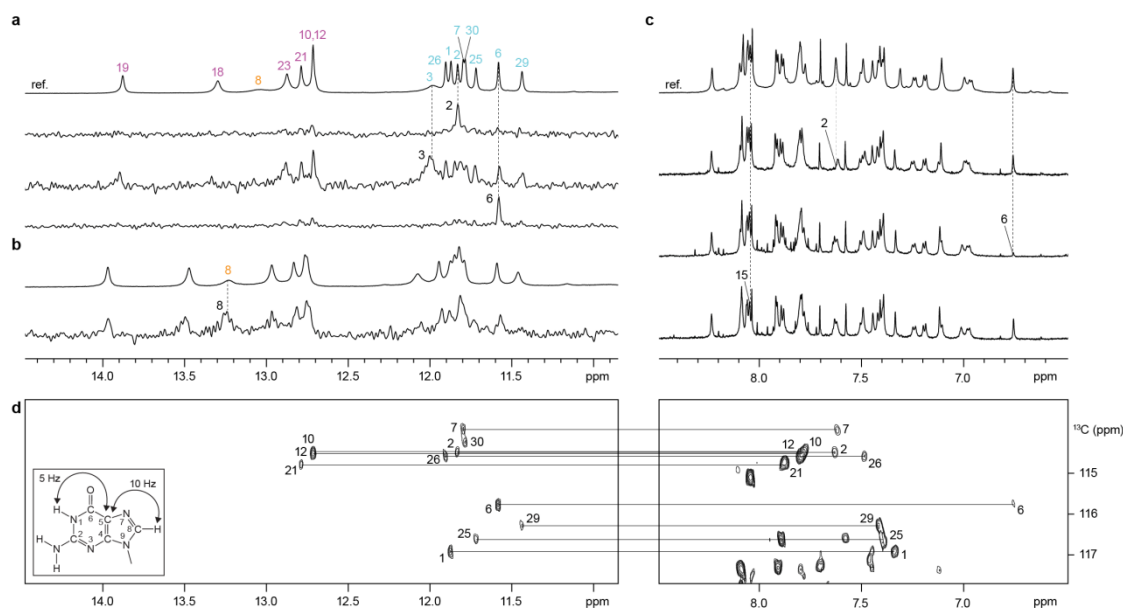
Figure 7.18 | Assignment of guanine H1 and H8 protons for Construct III. (**a**) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. (**b**) H1 proton spectrum at 5°C and the assignment of G8(H1). (**c**) Assignment of guanine H8 protons for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. (**d**) JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to [13]C5.
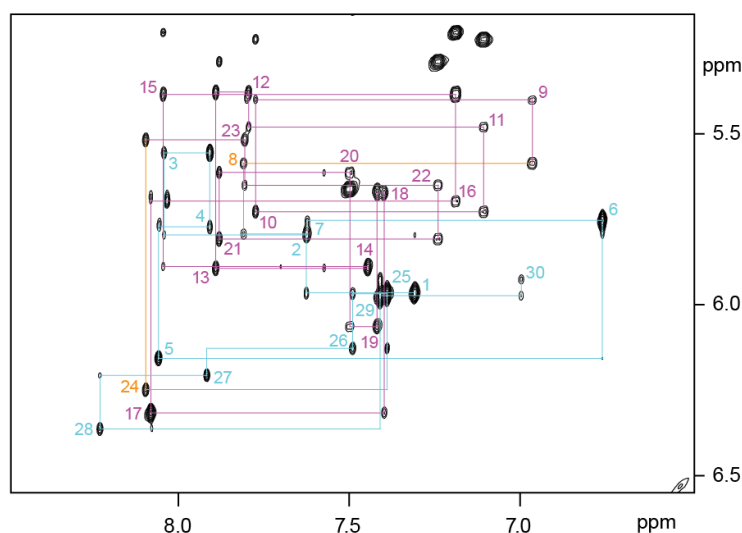
Figure 7.19 | NOESY spectrum showing the sequential connectivity of Construct III. H6/H8–H1' NOE connectivity of Construct III from G1 to G30. Residue numbers are marked on the corresponding intraresidue cross-peaks.
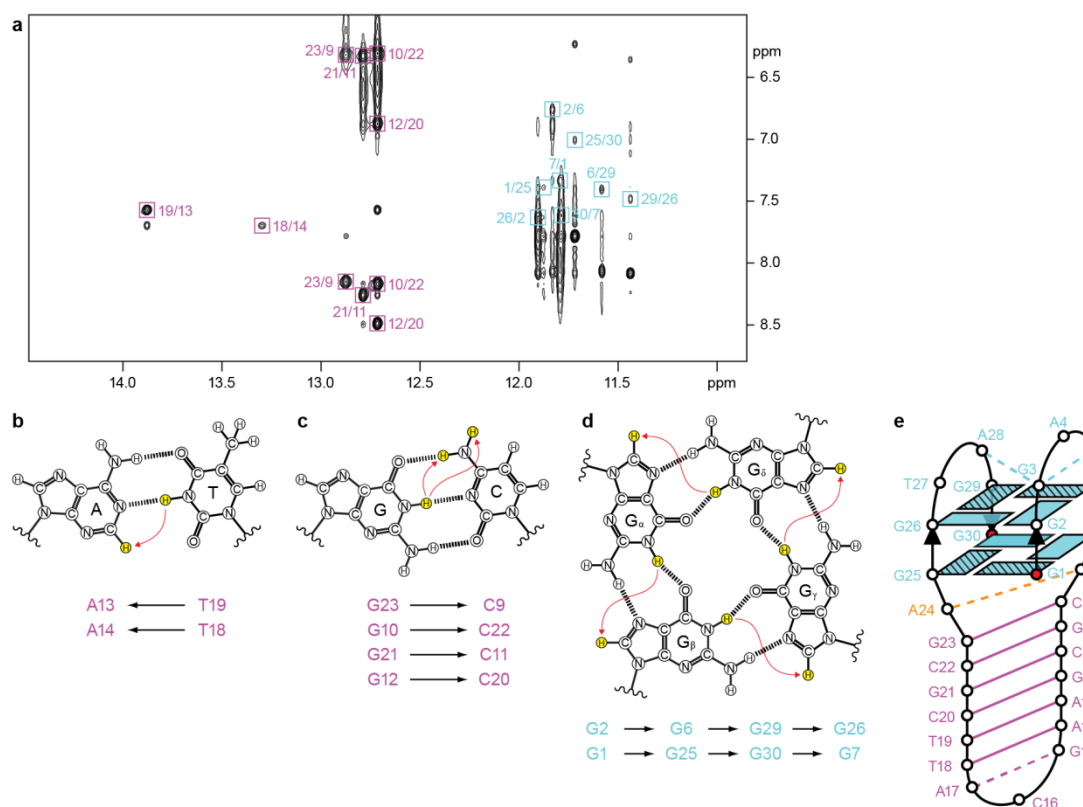
Figure 7.20 | Folding topology of quadruplex–duplex Construct III in K⁺ solution. (**a**) NOESY plot of Construct III highlighting the characteristic cross-peaks corresponding to T(H3)–A(H2) for A•T Watson–Crick base pair (**b**) and G(H1)–C(H41)/G(H1)–C(H42) for G•C Watson–Crick base pair (**c**), as well as guanine H1–H8 NOE connectivity (**d**) around each tetrad. (**e**) Schematic structure of quadruplex–duplex Construct III.
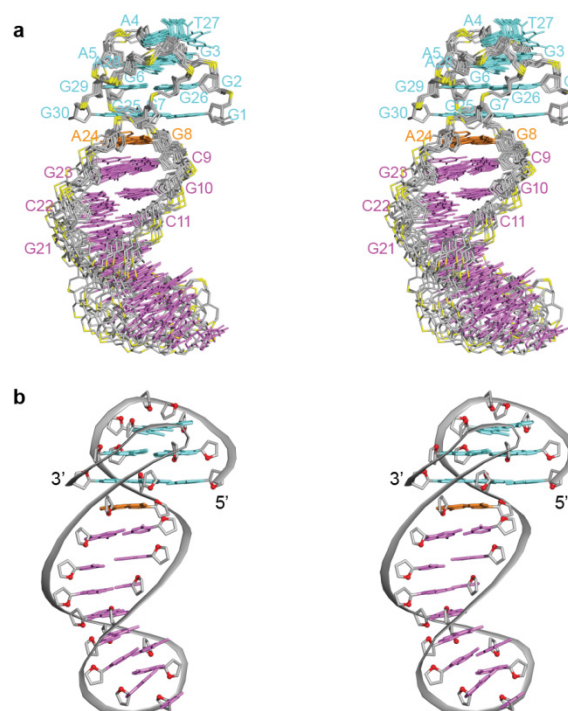
Figure 7.21 | NMR structure of quadruplex–duplex Construct III in potassium solution. (**a**) Stereo view of ten computed structures of Construct III, aligned based on the tetrad core. (**b**) Stereo view of a representative conformer in ribbon representation.

Alternative instances of duplex compatibility across the diagonal corners of a quadruplex involving a different context/approach were also described previously[225,265].

Sidestepping unfavorable duplex–quadruplex connectivity across the narrow groove through a snapback. Another possibility of coaxial duplex attachment onto a quadruplex core would involve their connection across the narrow groove of the quadruplex. However, as opposed to the favorable extension of duplex from the wide groove, duplex extension from the narrow groove would be energetically costly. Even though the narrow groove has an antiparallel strand orientation, this happens in the alternate polarity with respect to that of the duplex. Furthermore, the two duplex strands would have to shrink down to ~12 Å from ~18 Å at a single step, which is too abrupt a change. A snapback motif was seen in an RNA quadruplex–duplex junction[266] and a promoter G-quadruplex[239] to mediate strand progression to this side of the tetrad core. This snapback approach was successfully adapted here to sidestep the unfavorable connectivity across the narrow groove (Construct IV; Figures 7.6d, 7.22–7.25 and Appendix I). In this case, the duplex strand is
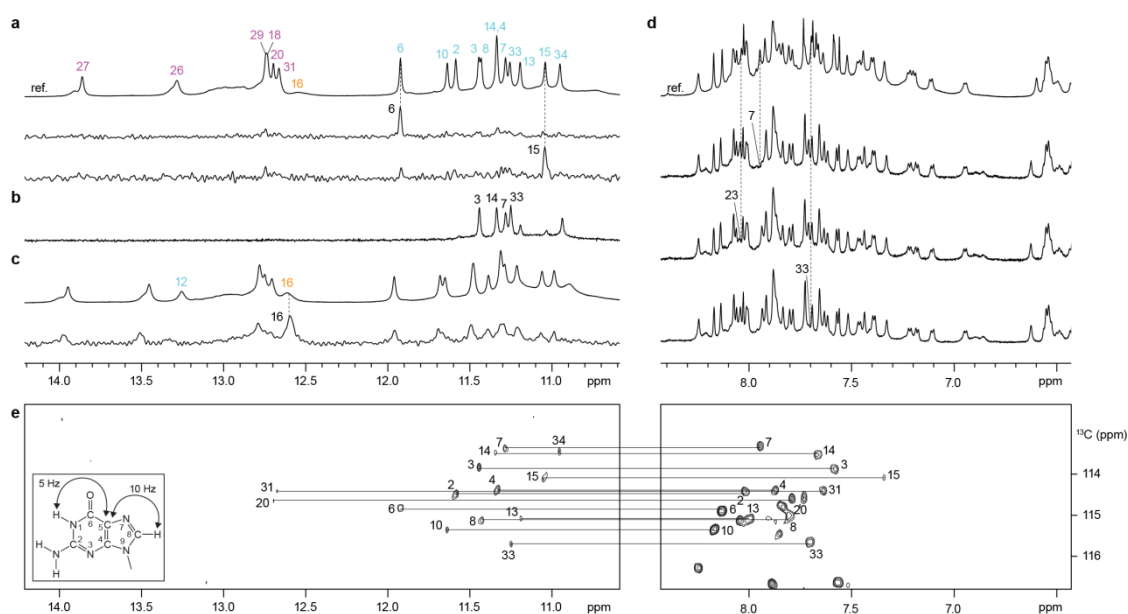
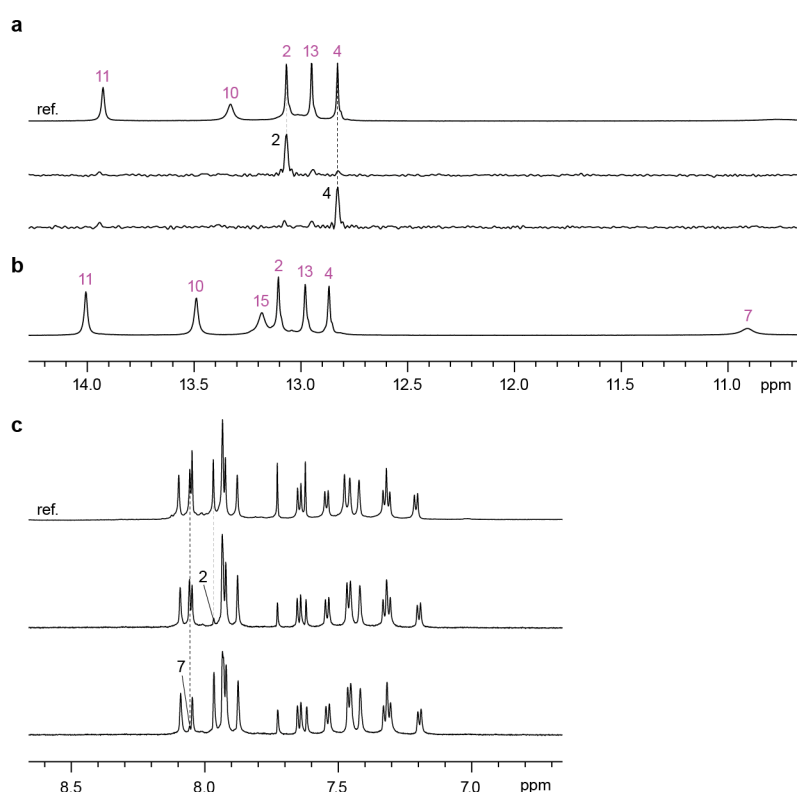**Figure 7.22 | Assignment of guanine H1 and H8 protons for Construct IV. (a)** Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. **(b)** H1 proton spectra in ^{2}H$_2$O for 2 h at 25°C. **(c)** H1 proton spectrum at 5 °C and the assignment of G16(H1). **(d)** Assignment of guanine H8 protons for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. **(e)** JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to ^{13}C5.
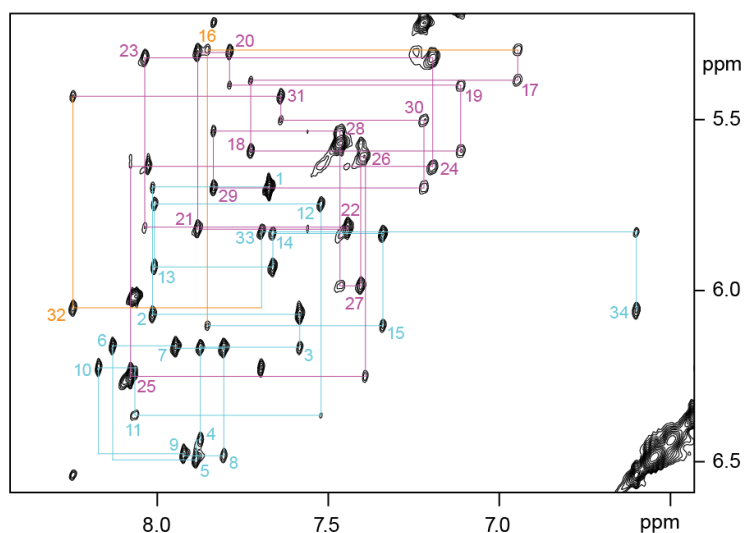
Figure 7.23 | NOESY spectrum showing the sequential connectivity of Construct IV. H6/H8–H1' NOE connectivity of Construct IV from G1 to G30. Residue numbers are marked on the corresponding intraresidue cross-peaks.
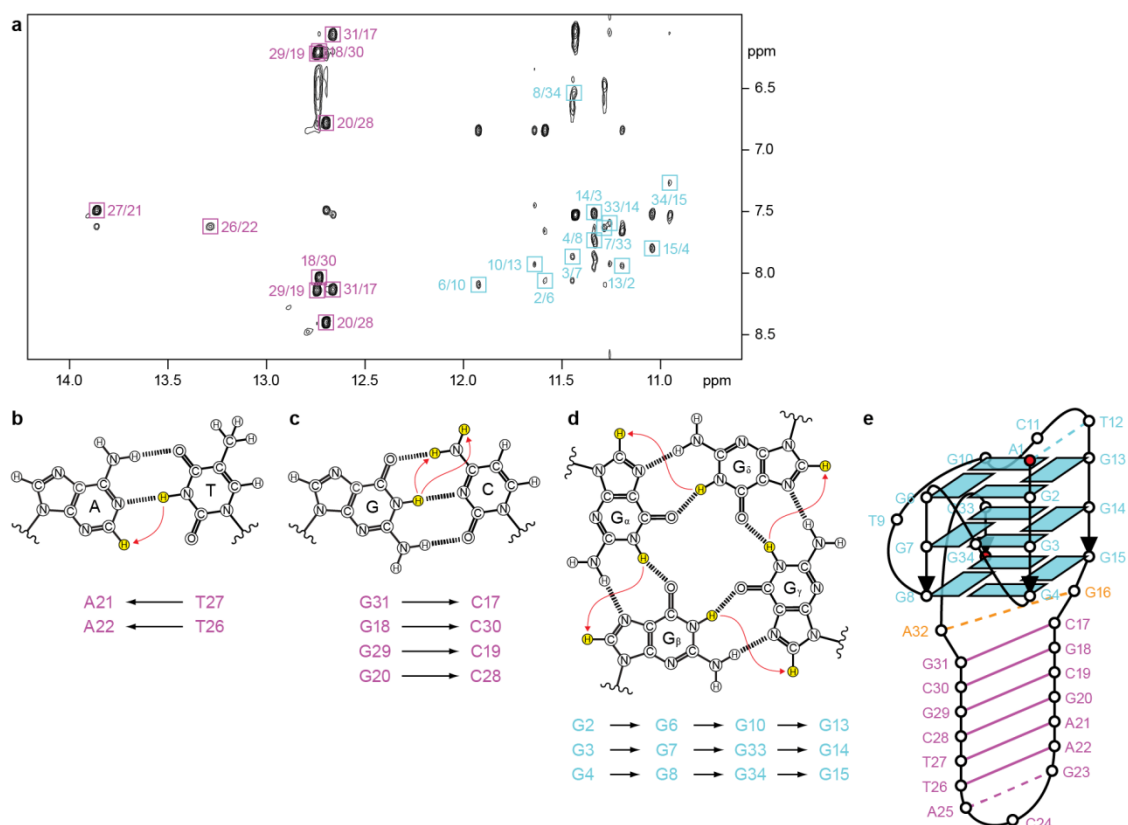
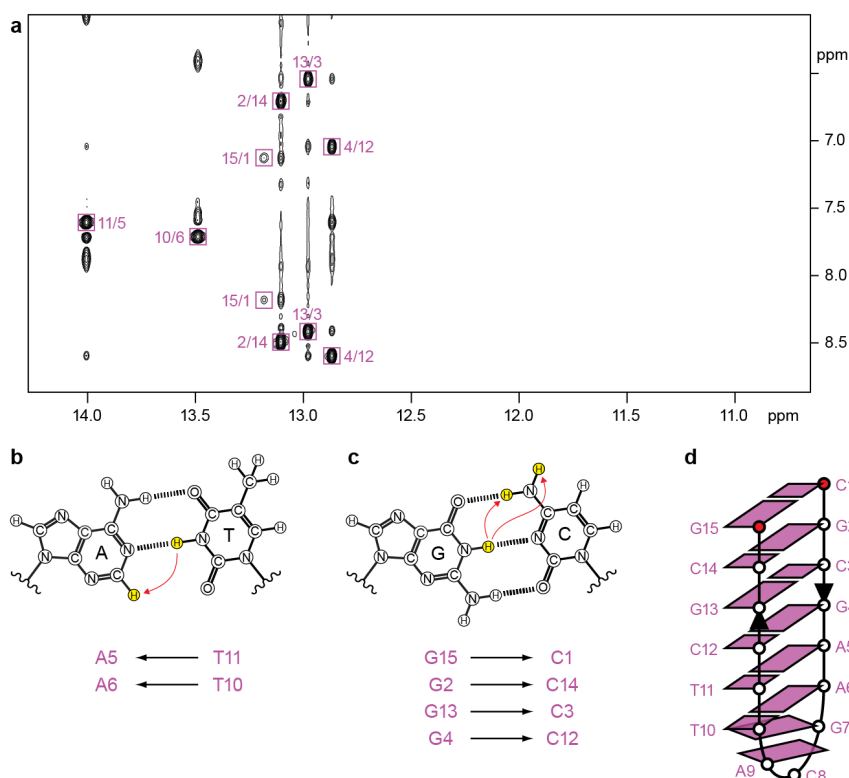**Figure 7.24 | Folding topology of quadruplex–duplex Construct IV in K⁺ solution.** (**a**) NOESY plot of Construct IV highlighting the characteristic cross-peaks corresponding to T(H3)–A(H2) for A•T Watson–Crick base pair (**b**) and G(H1)–C(H41)/G(H1)–C(H42) for G•C Watson–Crick base pair (**c**), as well as guanine H1–H8 NOE connectivity (**d**) around each tetrad. (**e**) Schematic structure of quadruplex–duplex Construct IV.

110
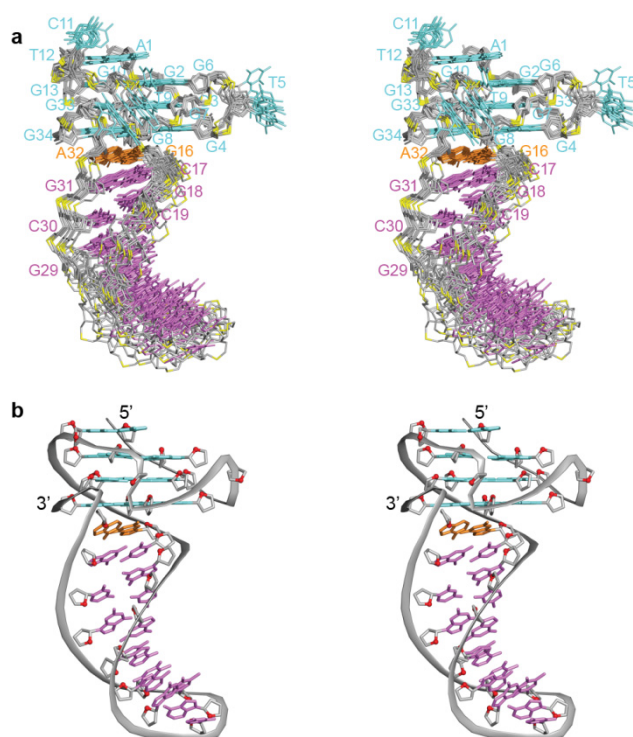
Figure 7.25 | NMR structure of quadruplex–duplex Construct IV in potassium solution.
(**a**) Stereo view of ten computed structures of Construct IV, aligned based on the tetrad core.
(**b**) Stereo view of a representative conformer in ribbon representation.

allowed to snap onto the same tetrad core edge over the diagonal ends. Note the intercalation of a G•A Watson–Crick mismatch at the connection point. Stacking of bases across the junction in this construct is less favorable as compared to the other coaxial arrangements (Construct I–III).

Sideway connection of a duplex onto a quadruplex. The connection of a duplex in an orthogonal helical orientation with respect to a quadruplex would involve the sideway attachment of the duplex helix (Figure 7.1e). This orientation of the helices was not immediately intuitive, given that base stacking would be absent at the junction. We have shown that this orthogonal connectivity could be achieved through incorporation of the reference duplex hairpin as the middle double-chain-reversal loop of a propeller-type G-quadruplex[267] (Construct V; Figures 7.7, 7.26–7.29 and Appendix I), In this robust structure, the base pair at the junction is disrupted, with the two bases flipping out of the helix. Previously, the extrusion of bases from a helix was observed in the B–Z junction[18] [see Section 1.1.2], which allows continuous base stacking

111
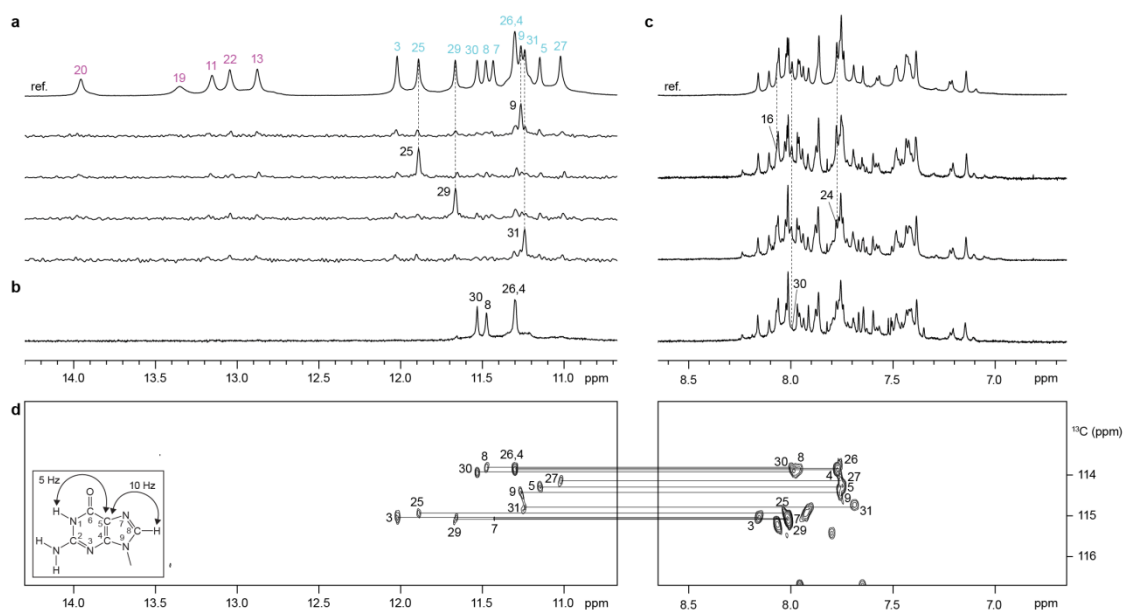
Figure 7.26 | Assignment of guanine H1 and H8 protons for Construct V. (a) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. **(b)** H1 proton spectra in ^2H_2O for 2 h at 25°C. **(c)** Assignment of guanine H8 protons for the reference spectrum (ref.) is achieved through site-specific ^2H substitution at the marked positions. **(d)** JRHMBC experiments showing long-range connectivity between guanine H1 and H8 protons, mediated by their mutual through-bond correlations to $^{13}C5$.
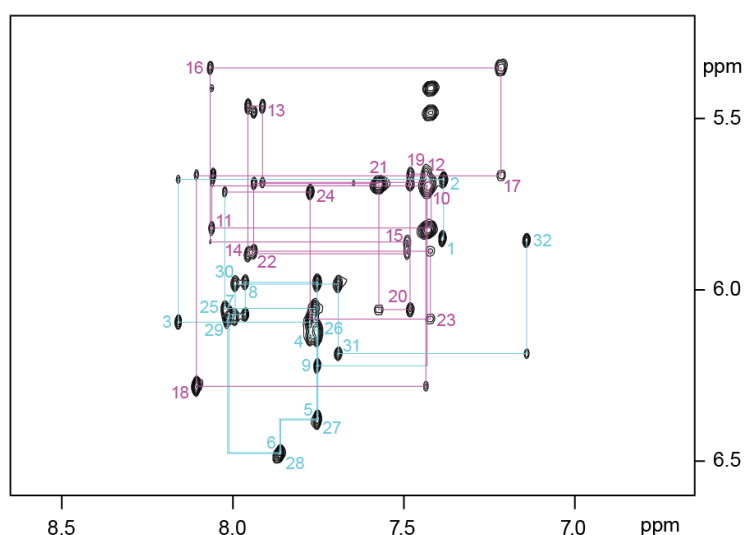
Figure 7.27 | NOESY spectrum showing the sequential connectivity of Construct V. H6/H8–H1' NOE connectivity of Construct V from G1 to T32. Residue numbers are marked on the corresponding intraresidue cross-peaks.
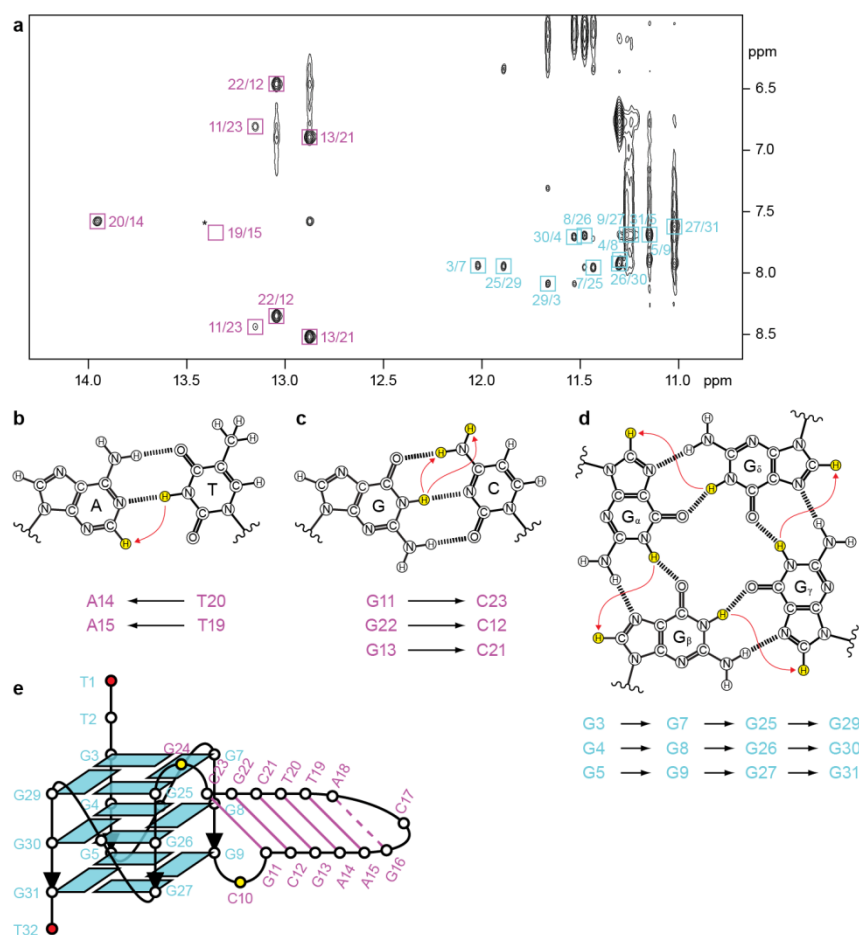
Figure 7.28 | Folding topology of quadruplex–duplex Construct V in K⁺ solution. (**a**) NOESY plot of Construct V highlighting the characteristic cross-peaks corresponding to T(H3)–A(H2) for A•T Watson–Crick base pair (**b**) and G(H1)–C(H41)/G(H1)–C(H42) for G•C Watson–Crick base pair (**c**), as well as guanine H1–H8 NOE connectivity (**d**) around each tetrad. (**e**) Schematic structure of quadruplex–duplex Construct V.

between right-handed B-DNA and left-handed Z-DNA. In the present case, disruption of the base pair allows maximal base stacking to be achieved separately at the quadruplex and duplex domains.

Quadruplex–duplex hybrids. The five quadruplex–duplex constructs presented (Construct I–V; Figures 7.6 and 7.7) epitomize the principles underlying the fusion of a duplex and a quadruplex (or the adoption of a duplex stem within a quadruplex loop[265,268]). Various quadruplex–duplex hybrid constructs can be conceived based on the adaptation/combination of these junction types. Note however that more exotic quadruplex–duplex junction variants should not be dismissed.
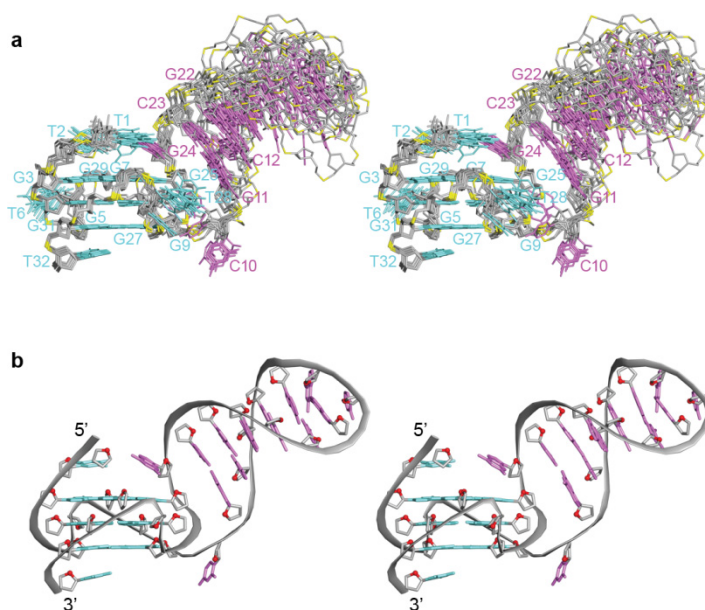
Figure 7.29 | NMR structure of quadruplex–duplex Construct V in potassium solution. (a) Stereo view of ten computed structures of Construct V, aligned based on the tetrad core. **(b)** Stereo view of a representative conformer in ribbon representation.

Connecting multiple duplex helices onto a single G-quadruplex. Since a quadruplex could encompass numerous loops, we further explored the simultaneous conjointment of multiple duplex stems onto a common quadruplex hub[269] (Figure 7.30a). This could be achieved through the attachment of individual duplex helix onto the tetrad core based on connection strategies that are mutually inclusive. Adhering to this principle, three duplex helices were successfully connected onto a common tetrad scaffold (Figures 7.30–7.34). Beginning with the all-parallel-stranded core topology, a nick is first introduced on Strand 2, allowing the docking of the 5' terminal (highlighted in red) onto the top G-tetrad layer. Consequently, Hairpin I (green) juts out of Edge A (originally of medium groove width) in a coaxial manner. The returning strand of Hairpin I joins up with Strand 1 and traverses across the medium groove to complete the disjointed G-column on Strand 2. At this point, Strand 2 is oriented in parallel with Strand 3, which permits the lateral protrusion of Hairpin II (magenta) from Edge B. Similarly, a sideway-projecting Hairpin III (orange) straddles the medium groove (Edge C) flanked by Strand 3 and Strand 4. Association of the twelve guanines into the tetrad core brought about an orthogonal orientation of all three stems. This contrasts the three-way junction[270], which consists of two end-to-end stacked hairpins with a sideway-projecting arm.

114

Figure 7.30 | Construction of a G-junction mediating three hairpin arms. (a) Schematic diagram illustrating the incorporation of three separate duplex stems onto a G-tetrad scaffold. **(b,c)** Ribbon view of the G-junction model from the side **(b)** and from the top **(c)**. **(d)** Sequence and NMR 1D H1 proton spectrum of the G-junction construct.

Figure 7.31 | Assignment of guanine H1 protons for the G-junction motif. (a) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. **(b)** H1 proton spectra in ^{2}H$_2$O for 2 h at 25°C.

115

Figure 7.32 | NOESY spectrum showing the sequential connectivity of the G-junction motif. H6/H8–H1' NOE connectivity of the G-junction motif from G1 to T51. Residue numbers are marked on the corresponding intraresidue cross-peaks.

**Biological and therapeutic implications.** Algorithms of the generic form $G_{X1}N_{L1}G_{X2}N_{L2}G_{X3}N_{L3}G_{X4}$ have been implemented in the search for quadruplex-forming motifs in the human genome[41,42,261], in which the G-tract (X) and intervening loop (L) was largely restricted to a length of 2–5 and 7 nt, respectively. The facile establishment of robust quadruplex–duplex hybrids [see Chapter 8] calls for a more permissive algorithm with respect to the loop lengths. On the other hand, current exploration of genomic G-quadruplexes as therapeutic targets[30,61,235] largely focuses on development of planar ligands exhibiting selectivity for G-tetrads over canonical Watson–Crick base pairs. Knowledge of duplex–quadruplex compatibility would contribute towards the design of quadruplex-interactive compounds with improved specificity. In addition, this could lead to alternative targeting approaches, ranging from quadruplex–duplex junction binders[266] to antigene targeting[87,271].
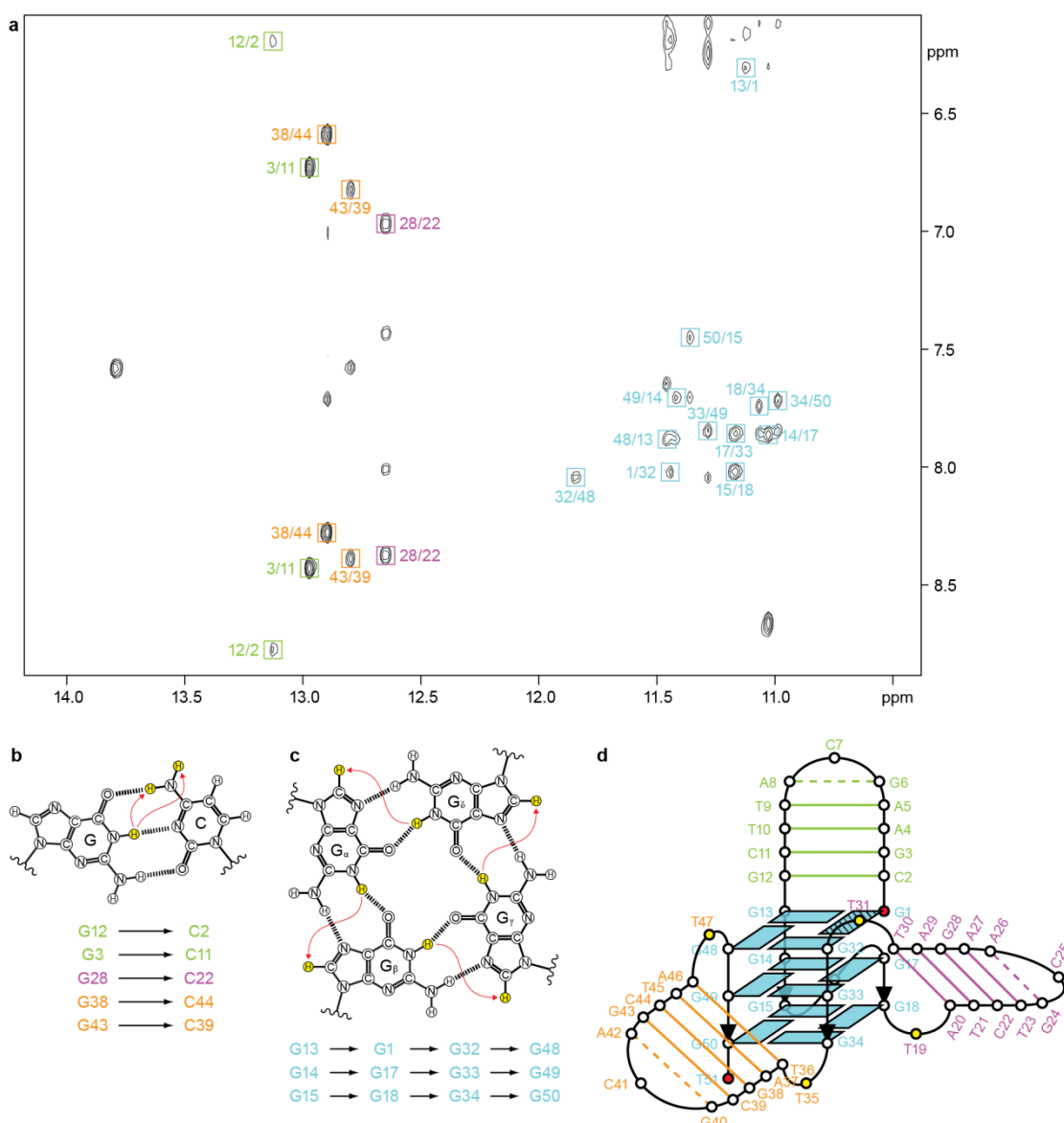
Figure 7.33 | Folding topology of the G-junction motif in potassium solution. (**a**) NOESY plot of the G-junction motif highlighting the characteristic cross-peaks corresponding to T(H3)–A(H2) for A•T Watson–Crick base pair (**b**) and G(H1)–C(H41)/G(H1)–C(H42) for G•C Watson–Crick base pair (**c**), as well as guanine H1–H8 NOE connectivity (**d**) around each tetrad. (**e**) Schematic structure of the G-junction motif.

G-quadruplex as an organizational platform. In this study, the quadruplex folding topology was controlled through strategic placement of duplex stems and ancillary stabilizing motifs. Whereas short G-rich sequences generally lack specificity, judicious introduction of duplex segments shepherds the folding towards the desired core topology, and establishment of the quadruplex structure determines the projection of the duplex stem. Similar cooperativity of duplex and quadruplex components can be adapted for the design of interactive DNA assembly[272]. The versatility of the approach was established with a range of quadruplex–duplex constructs having coaxial or

117
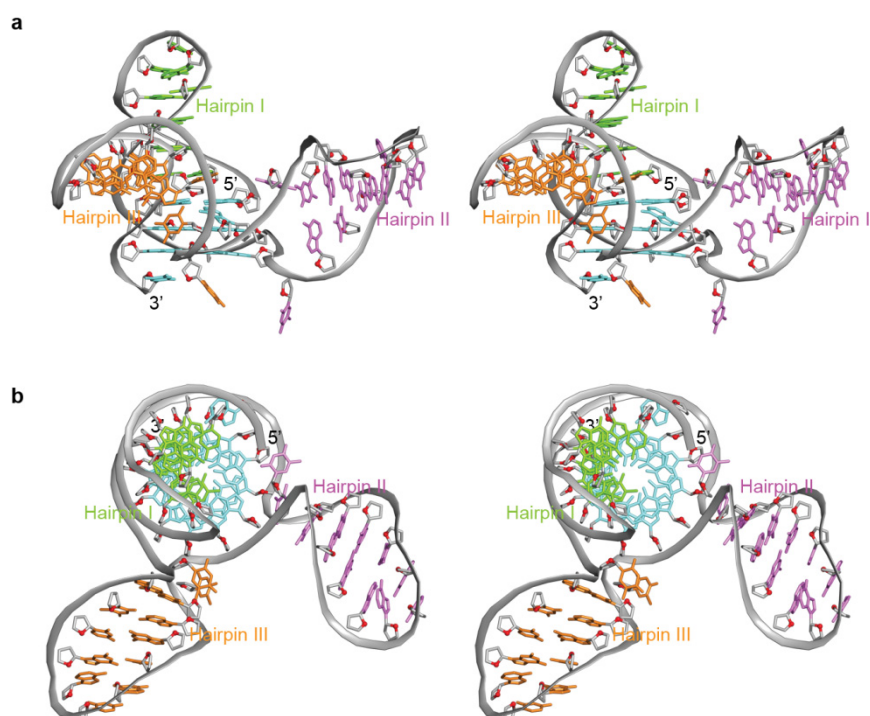
Figure 7.34 | Structural model of the G-junction motif in potassium solution. (**a**,**b**) Stereo view of the G-junction model from the side (**a**) and from the top (**b**).

orthogonal projection of the duplex arm. Furthermore, the modularity of quadruplex–duplex junctions was attested by the creation of a G-junction, an assemblage of duplex arms through a common G-quadruplex core. Thanks to the conformational diversity of quadruplex structures, myriad G-junction motifs can potentially be constructed. Random sequence oligonucleotides generated from combinatorial libraries in the past have yielded numerous enzymes and aptamers comprising G-rich elements[98], implicating that quadruplex could act as an organizational platform for these oligonucleotides. A comprehensive understanding of duplex–quadruplex connectivity would facilitate structural engineering of such complexes[262] utilizing the G-junction motifs.

Integration of G-junction motifs in DNA nanotechnology. By virtue of its stable G-quadruplex core, the G-junction would qualify as an immobile junction suitable for DNA nanotechnology[80] [see Section 1.3]. To date, G-rich elements are typically avoided in the sequence design stage for the construction of DNA nanomaterials, due to a poor understanding on their folding preferences. With an improved level of control over quadruplex folding, it should now be feasible to integrate G-junction motifs in the design of DNA nanomaterials. This would expand the current repertoire of structural motifs, which have largely been based on variants of the Holliday

118

junction stacked X structure[17] [see Section 1.1.2] (e.g. DX/TX/PX motifs[80]). For instance, the G-junction motifs could serve as robust elements at vertexes of nanoarchitectures, as chain terminators, or as a universal adaptor to snap multiple duplexes together[269]. On the other hand, functional G-rich elements (e.g. DNA aptamers[98], enzymes[98], sensors[117], and nanowires[79]) could now be incorporated in DNA nanoconstructs in a more precise manner. On a side note, cationic dependence of quadruplex structure could allow added control on the construction of DNA nanomaterials.

Conclusion

Aside from B-form duplex, DNA can adopt diverse structural forms including the G-quadruplex. Quadruplex DNA has been implicated in biological processes and shown to exhibit important functional properties. To date, the connectivity of duplex and quadruplex DNA was not well understood. This poses a hindrance towards the incorporation of G-quadruplex structural elements in biotechnological and DNA nanotechnological applications, which are heavily reliant on duplex DNA.

In this study, the coaxial and orthogonal connectivity between duplex and quadruplex DNA was explored across the diverse geometry of a quadruplex. The NMR structures of five representative quadruplex–duplex constructs were elucidated, each of which illustrates a distinct approach to attach a duplex stem onto a quadruplex core. Furthermore, the simultaneous attachment of multiplex stems onto a single quadruplex core was demonstrated, validating the modularity of these quadruplex–duplex junctions.

Equipped with an understanding on how a quadruplex can be juxtaposed next to a duplex, it should now be possible to integrate quadruplex elements in the design of DNA nanoconstructs. In the natural chromosomal environment, quadruplex formation should occur in close proximity to duplex elements. Hence the quadruplex–duplex junctions presented should provide a glimpse of the structural contexts existing at these important transitional regions. Perhaps more importantly, stable formation of these quadruplex–duplex hybrids, which have not been thoroughly examined, prompts for a re-evaluation of quadruplex-forming motifs in the human genome, and also points to alternative therapeutic strategies that can be steered towards genomic G-rich sequences.

Chapter 8

Thermal Stability of
DNA Quadruplex–Duplex Hybrids

8.1   Thermodynamic Studies on G-Quadruplex DNA

8.2   Thermal Stability of DNA Quadruplex–Duplex Hybrids

8.1   Thermodynamic Studies on G-Quadruplex DNA

Formation of quadruplex–duplex hybrids. The canonical conformation of DNA is the B-form duplex, which consists of two antiparallel right-handed chains associating through Watson–Crick base pairs (bp). On the other hand, the G-quadruplex[28-30] is a helical complex constructed through the multiple stacking of G•G•G•G tetrads[31], stabilized by Hoogsteen hydrogen bonds[19]. The G-quadruplex can adopt a diverse range of structures; the four strands making up the core can be arranged in four different relative orientations, the linkers (or loops) can connect the four strands in different manners, and the G-tetrad core can consist of two or more tetrad layers[30]. We have performed a systematic analysis on the incorporation of a duplex stem across the various geometries of a quadruplex core[273] [see Section 7.2], demonstrating the facile juxtaposition of these two structural elements. It is thus important to establish thermodynamic descriptions regarding their conjointment, as that would facilitate future design and prediction of quadruplex–duplex hybrids[79,273-276], which could find applications in therapeutics and nanotechnology.

Long non-ordered loops destabilize G-quadruplex structures. Previous thermodynamic studies[254,277] on quadruplexes comprising loops of varying compositions shed light on the influences of loop length and sequence on quadruplex conformation and stability[180,228-232,252,278,279]. A loop consisting of a single nucleotide shows a propensity towards the propeller (or double-chain-reversal) configuration[180,228,231,252], and is especially stable. In addition, an increase in loop length seems to reduce the stability of a quadruplex[180,228-232]. However, the long loops that have been investigated mostly do not contain any explicit stabilizing interactions, which could have been the cause of their destabilizing nature. Here we set out to determine the effects of stem-loop incorporation on quadruplex stability. To this end, two classes of quadruplex–duplex hybrids with contrasting orientation of the duplex stem with respect to the tetrad core were investigated. Using NMR[280] and CD[281] spectroscopy, we outline the influences that stem-loop length and its sequence, particularly the base pair steps proximal to the junction, exert on the stability of these complexes. We also highlight the consequences of having bulges at the quadruplex–duplex junction.

8.2   Thermal Stability of DNA Quadruplex–Duplex Hybrids

Design of quadruplex–duplex hybrids. The series of duplex stems (Table 8.1) and quadruplex–duplex hybrids (Tables 8.2 and 8.3) investigated were based on structures elucidated previously[263,273] [see Section 7.2]. Construct I consists of a chair-type G-quadruplex[35], in which the duplex stem is adjoined to the tetrad core across the wide groove through continuous base stacking[273] (Figure 8.1a). Construct II consists of an all-parallel-stranded G-quadruplex[267,282], wherein the duplex stem projects laterally outwards of the tetrad core as the middle propeller loop[273] (Figure 8.1b). For Construct II, a linker residue was added onto each of the two contact points between the quadruplex and duplex based on prior knowledge that a base pair at this junction type would have to be broken up upon complex formation[273] [see Section 7.2]. The length and sequence of the duplex stem were varied in a systematic manner for both constructs. Bulges[283,284] were further introduced at the quadruplex–duplex junction to determine the effects that they exert on the adoption of the desired G-quadruplex topologies. The experiments were performed with a buffer consisting of 10 mM KCl in lithium phosphate, pH 7.0.

NMR and CD validation of quadruplex–duplex hybrids with variable stem length. 1D NMR proton spectra of the skeletal structures of Construct I (denoted as *W2C-0bp*; Table 8.2) and Construct II (denoted as *M2C-0bp*; Table 8.3), which do not contain any duplex stem, indicated the presence of a single conformation. *W2C-0bp* showed ten imino proton peaks at ~11–12.4 ppm (Figure 8.2b), in line with the adoption of a two-tetrad quadruplex with a capping T•T base pair[35]. *M2C-0bp* displayed twelve H1 proton peaks from ~11–12.1 ppm (Figure 8.2d), in line with the adoption of a three-tetrad quadruplex[267,282]. In contrast, imino protons of Watson–Crick base pairs typically appear at ~12.5–14 ppm (Figure 8.3), as evident from the

Table 8.1 | Sequences of reference duplex hairpins used in this study.

Name	Sequence	$T_m{}^a$ (°C)
3bp	CGC GCA GCG	–
4bp	CGCG GCA CGCG	–
5bp	CGCGA GCA TCGCG	–
6bp	CGCGAA GCA TTCGCG	–
11bp	CGCGAATCTGT GCA ACAGATTCGCG	–
11bp	CGCGCGCGCGC GCA GCGCGCGCGCG	–
16bp	CGCGAATCTGTGATAC GCA GTATCACAGATTCGCG	–

a Melting temperature in 10 mM K⁺, determined from molar ellipticity at 250 nm. This value is an average of two readings.

123

Table 8.2 | Sequences of Construct I (*W2C*) used in this study.

	Name	Sequence[a,b]			$T_m{}^c$ (°C)
Duplex Stem Length/ Sequence Variation	*0bp*	GGTTGG	GCA	GGTTGG	36.2
	1bp	GGTTGG	C GCA G	GGTTGG	39.0 [d]
	2bp	GGTTGG	CG GCA CG	GGTTGG	38.2
	3bp	GGTTGG	CGC GCA GCG	GGTTGG	41.6
	4bp	GGTTGG	CGCG GCA CGCG	GGTTGG	43.0
	5bp	GGTTGG	CGCGA GCA TCGCG	GGTTGG	43.5
	6bp	GGTTGG	CGCGAA GCA TTCGCG	GGTTGG	42.6
	11bp	GGTTGG	CGCGAATCTGT GCA ACAGATTCGCG	GGTTGG	43.9
	11bp[CG]	GGTTGG	CGCGCGCGCGC GCA GCGCGCGCGCG	GGTTGG	44.6
	16bp	GGTTGG	CGCGAATCTGTGATAC GCA GTATCACAGATTCGCG	GGTTGG	43.7
Loop Length Reference	*1T*	GGTTGG	T	GGTTGG	– [e]
	2T	GGTTGG	TT	GGTTGG	– [e]
	3T	GGTTGG	TTT	GGTTGG	– [e]
	5T	GGTTGG	TTTTT	GGTTGG	– [e]
	15T	GGTTGG	T_{15}	GGTTGG	– [e]
	25T	GGTTGG	T_{25}	GGTTGG	– [e]
	35T	GGTTGG	T_{35}	GGTTGG	– [e]
Base Pair Variation/ Bulge at Junction	*6bp[G7/C21]*	GGTTGG	<u>G</u>GCGAA GCA TTCGC<u>C</u>	GGTTGG	35.4
	6bp[T7/A21]	GGTTGG	<u>T</u>GCGAA GCA TTCGC<u>A</u>	GGTTGG	32.2
	6bp[A7/T21]	GGTTGG	<u>A</u>GCGAA GCA TTCGC<u>T</u>	GGTTGG	36.5
	6bp[C8/G20]	GGTTGG	C<u>C</u>CGAA GCA TTCG<u>G</u>G	GGTTGG	45.7
	6bp[T8/A20]	GGTTGG	C<u>T</u>CGAA GCA TTC<u>A</u>AG	GGTTGG	43.2
	6bp[A8/T20]	GGTTGG	C<u>A</u>CGAA GCA TTC<u>T</u>TG	GGTTGG	42.1
	6bp[+T7]	GGTTGG	**T**CGCGAA GCA TTCGCG	GGTTGG	– [e]
	6bp[+T21]	GGTTGG	CGCGAA GCA TTCGCG**T**	GGTTGG	– [e]
	6bp[+T7/+T22]	GGTTGG	**T**CGCGAA GCA TTCGCG**T**	GGTTGG	24.5 [d]

[a] Modified residues are underlined. [b] Bulges are shown in bold. [c] Melting temperature in 10 mM K⁺, determined from molar ellipticity at 295 nm. This value is an average of two readings. [d] These melting temperatures consist of considerable contributions from extraneous conformation(s). [e] The melting temperatures of these oligonucleotides were not determined.
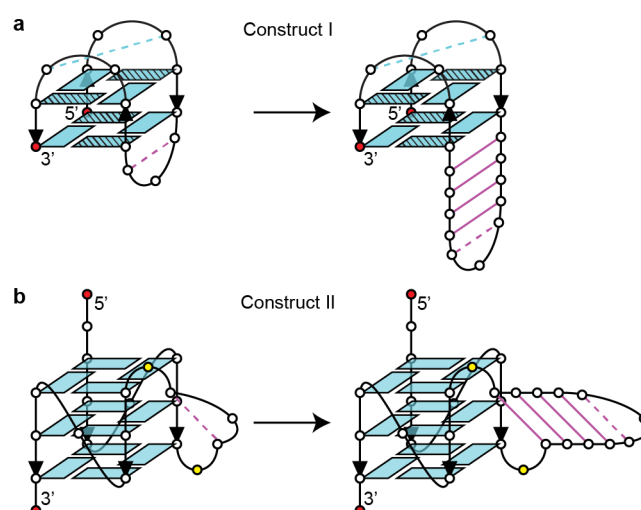
Figure 8.1 | Basis of design for the two duplex–quadruplex hybrid constructs. Schematic diagrams illustrating the basis of duplex stem incorporation for (**a**) Construct I and (**b**) construct II. Quadruplex and duplex components are colored in cyan and magenta, respectively. S*yn* guanine bases are shaded. Watson–Crick and non-canonical base pairs are represented as solid and dotted lines, respectively. 5'- and 3'-terminal residues are colored in red circles, while bulges are colored in yellow circles.

Table 8.3 | Sequences of Construct II (*M2C*) used in this study.

	Name	Sequence[a,b]			T_m[c] (°C)
Duplex Stem Length/ Sequence Variation	*0bp*	TTGGGTGGG	GCA	GGGTGGGT	62.2
	1bp	TTGGGTGGG	C GCA G	GGGTGGGT	60.9
	2bp	TTGGGTGGG	CG GCA CG	GGGTGGGT	60.8
	3bp	TTGGGTGGG	CGC GCA GCG	GGGTGGGT	65.3
	4bp	TTGGGTGGG	CGCG GCA CGCG	GGGTGGGT	64.4
	5bp	TTGGGTGGG	CGCGA GCA TCGCG	GGGTGGGT	63.6
	6bp	TTGGGTGGG	CGCGAA GCA TTCGCG	GGGTGGGT	63.0
	11bp	TTGGGTGGG	CGCGAATCTGT GCA ACAGATTCGCG	GGGTGGGT	66.6
	11bp[CG]	TTGGGTGGG	CGCGCGCGCGC GCA GCGCGCGCGCG	GGGTGGGT	67.9
	16bp	TTGGGTGGG	CGCGAATCTGTGATAC GCA GTATCACAGATTCGCG	GGGTGGGT	66.8
Loop Length Reference	*1T*	TTGGGTGGG	T	GGGTGGGT	83.2
	2T	TTGGGTGGG	TT	GGGTGGGT	76.5
	3T	TTGGGTGGG	TTT	GGGTGGGT	70.7
	[T]3T[T]	TTGGGTGGG	**T**T$_3$**T**	GGGTGGGT	63.8
	[T]5T[T]	TTGGGTGGG	**T**T$_5$**T**	GGGTGGGT	59.2
	[T]7T[T]	TTGGGTGGG	**T**T$_7$**T**	GGGTGGGT	56.6
	[T]9T[T]	TTGGGTGGG	**T**T$_9$**T**	GGGTGGGT	53.7
	[T]11T[T]	TTGGGTGGG	**T**T$_{11}$**T**	GGGTGGGT	51.8
	[T]13T[T]	TTGGGTGGG	**T**T$_{13}$**T**	GGGTGGGT	50.8
	[T]15T[T]	TTGGGTGGG	**T**T$_{15}$**T**	GGGTGGGT	48.8
	[T]25T[T]	TTGGGTGGG	**T**T$_{25}$**T**	GGGTGGGT	42.6
	[T]35T[T]	TTGGGTGGG	**T**T$_{35}$**T**	GGGTGGGT	41.6
Base Pair Variation/Bulge at Junction	*6bp[G11/C26]*	TTGGGTGGG	<u>G</u>GCGAA GCA TTCGC<u>C</u>	GGGTGGGT	63.4
	6bp[T11/A26]	TTGGGTGGG	<u>T</u>GCGAA GCA TTCGC<u>A</u>	GGGTGGGT	60.2
	6bp[A11/T26]	TTGGGTGGG	<u>A</u>GCGAA GCA TTCGC<u>T</u>	GGGTGGGT	60.8
	6bp[C12/G25]	TTGGGTGGG	C<u>C</u>CGAA GCA TTCG<u>G</u>G	GGGTGGGT	–[d]
	6bp[T12/A25]	TTGGGTGGG	C<u>T</u>CGAA GCA TTCG<u>A</u>G	GGGTGGGT	–[d]
	6bp[A12/T25]	TTGGGTGGG	C<u>A</u>CGAA GCA TTCG<u>T</u>G	GGGTGGGT	–[d]
	6bp[[C10]	TTGGGTGGG	**C**CGCGAA GCA TTCGCG**T**	GGGTGGGT	63.0
	6bp[G10]	TTGGGTGGG	**G**CGCGAA GCA TTCGCG**T**	GGGTGGGT	63.5
	6bp[A10]	TTGGGTGGG	**A**CGCGAA GCA TTCGCG**T**	GGGTGGGT	63.0
	6bp[C26]	TTGGGTGGG	**T**CGCGAA GCA TTCGCG**C**	GGGTGGGT	62.6
	6bp[G26]	TTGGGTGGG	**T**CGCGAA GCA TTCGCG**G**	GGGTGGGT	63.9
	6bp[A26]	TTGGGTGGG	**T**CGCGAA GCA TTCGCG**A**	GGGTGGGT	63.9
	6bp[C10/G26]	TTGGGTGGG	**C**CGCGAA GCA TTCGCG**G**	GGGTGGGT	62.0
	6bp[G10/C26]	TTGGGTGGG	**G**CGCGAA GCA TTCGCG**C**	GGGTGGGT	58.6
	6bp[-T10/-T26]	TTGGGTGGG	**C**GCGAA GCA TTCGC**G**	GGGTGGGT	63.0
	6bp[-T10]	TTGGGTGGG	**C**GCGAA GCA TTCGCG**T**	GGGTGGGT	62.5
	6bp[-T26]	TTGGGTGGG	**T**CGCGAA GCA TTCGC**G**	GGGTGGGT	63.1
	6bp[+T11]	TTGGGTGGG	**TT**CGCGAA GCA TTCGCG**T**	GGGTGGGT	62.3
	6bp[+T26]	TTGGGTGGG	**T**CGCGAA GCA TTCGCG**TT**	GGGTGGGT	62.3
	6bp[+T11/+T27]	TTGGGTGGG	**TT**CGCGAA GCA TTCGCG**TT**	GGGTGGGT	61.3
	6bp[+T11T12]	TTGGGTGGG	**TTT**CGCGAA GCA TTCGCG**T**	GGGTGGGT	61.3
	6bp[+T26T27]	TTGGGTGGG	**T**CGCGAA GCA TTCGCG**TTT**	GGGTGGGT	61.0
	6bp[+T11T12/+T28T29]	TTGGGTGGG	**TTT**CGCGAA GCA TTCGCG**TTT**	GGGTGGGT	59.4

[a] Modified residues are underlined. [b] Bulges are shown in bold. [c] Melting temperature in 10 mM K$^+$, determined from molar ellipticity at 262 nm. This value is an average of two readings. [d] The melting temperatures of these oligonucleotides were not determined.
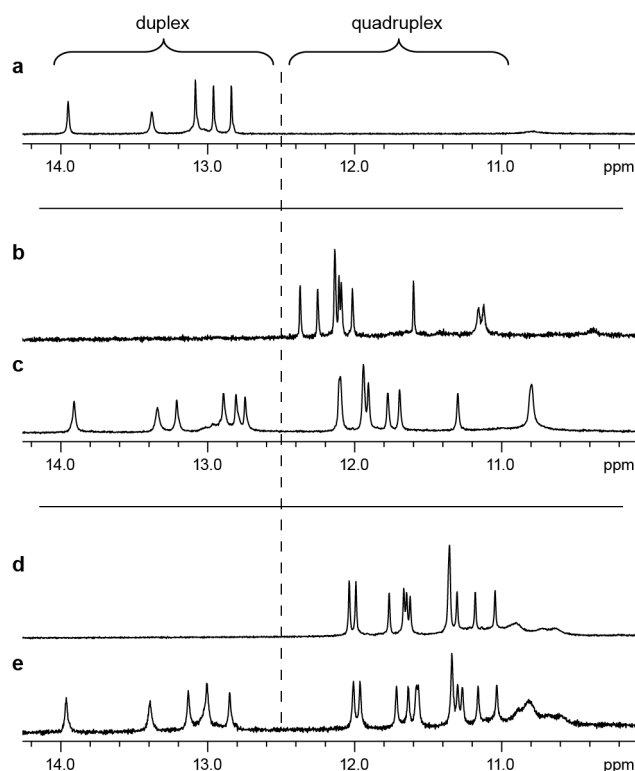
Figure 8.2 | 1D NMR imino proton spectra of representative duplex, quadruplex, and quadruplex–duplex hybrid constructs. (a) The reference duplex hairpin (*dx-6bp*), **(b)** Construct I without a duplex stem (*W2C-0bp*), **(c)** Construct I with the reference duplex stem incorporated (*W2C-6bp*), **(d)** Construct II without a duplex stem (*M2C-0bp*), and **(e)** Construct II with the reference duplex stem incorporated (*M2C-6bp*).

spectrum of a reference duplex stem with six base pairs (denoted as *dx-6bp*; Table 8.1 and Figure 8.2a). Incorporation of the same duplex stem in the respective constructs (giving *W2C-6bp* and *M2C-6bp*, respectively; Tables 8.2 and 8.3) yielded quadruplex–duplex hybrids with the expected topologies; the quadruplex–duplex hybrids displayed
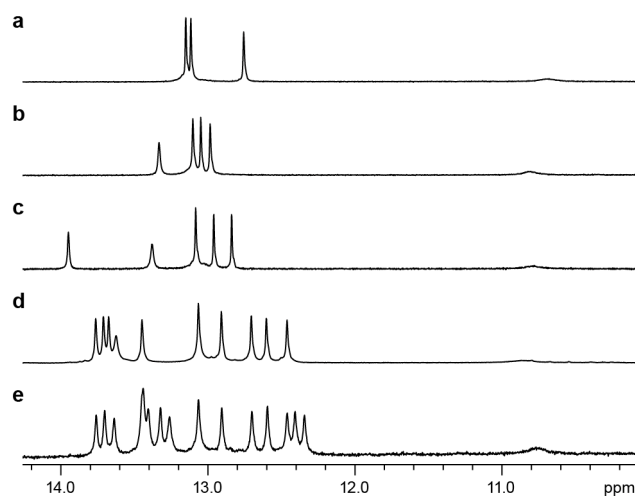
Figure 8.3 | 1D NMR imino proton spectra of reference duplex hairpins with variable stem length. (a) *dx-4bp*, **(b)** *dx-5bp*, **(c)** *dx-6bp*, **(d)** *dx-11bp*, and **(e)** *dx-16bp*.

126

imino proton peaks in both duplex and quadruplex ranges, and the peak distribution patterns resembled those of the individual duplex and quadruplex components (Figure 8.2c,e). Complexes with a stem length of up to sixteen base pairs were investigated for both constructs (Tables 8.2 and 8.3). For the oligonucleotides harboring a duplex stem with two or more base pairs, the duplex peaks built up incrementally as the number of base pair was increased while the quadruplex peaks retained similar distribution patterns (Construct I, Figure 8.4; Construct II, Figure 8.5), indicating that these quadruplex–duplex hybrids adopted the desired topologies as the predominant conformation.

CD spectra of the oligonucleotides concurred with their respective topologies. The CD spectrum of *dx-6bp* showed a positive band near 275 nm and a trough close to 250 nm (Figure 8.6a, orange), approximating that of a related duplex[285]. *W2C-0bp* showed two positive maxima around 245 nm and 295 nm (Figure 8.6a, green), typical of a chair-type G-quadruplex[281,286]. *M2C-0bp* exhibited a single maximum at 262 nm together with a negative minimum around 240 nm (Figure 8.6a, violet), indicative of a
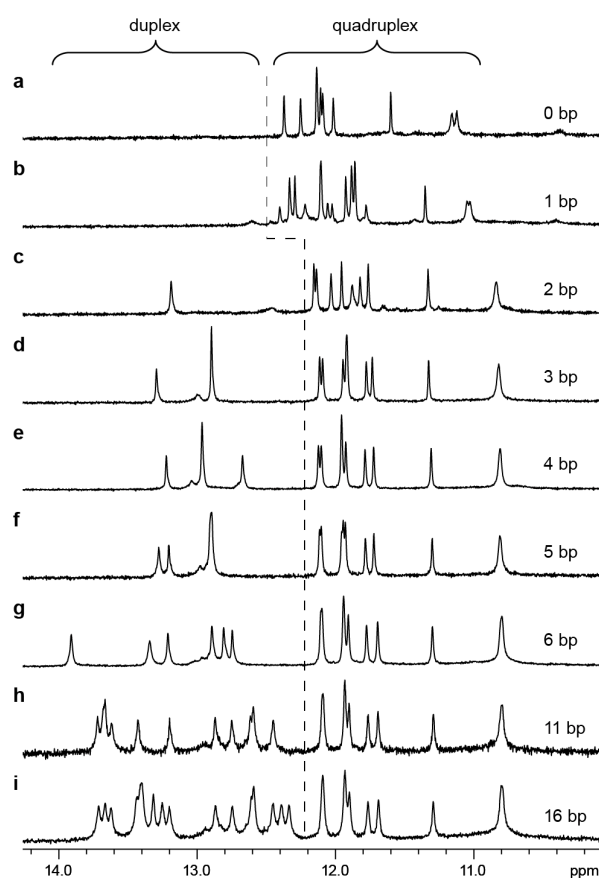
Figure 8.4 | 1D NMR imino proton spectra of Construct I with variable stem length. (a) *W2C-0bp*, (**b**) *W2C-1bp*, (**c**) *W2C-2bp*, (**d**) *W2C-3bp*, (**e**) *W2C-4bp*, (**f**) *W2C-5bp*, (**g**) *W2C-6bp*, (**h**) *W2C-11bp*, and (**i**) *W2C-16bp*.
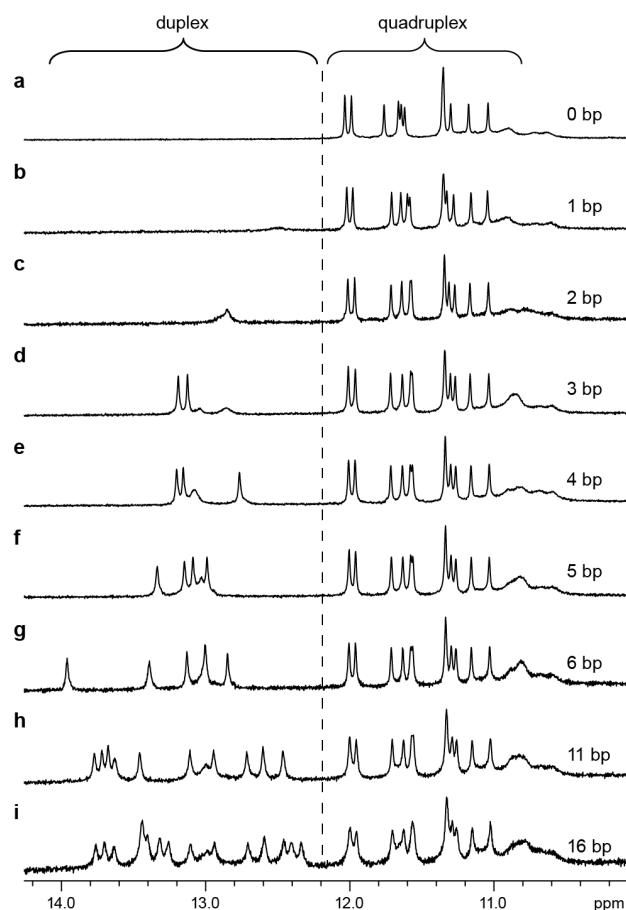
Figure 8.5 | 1D NMR imino proton spectra of Construct II with variable stem length. (**a**) *M2C-0bp*, (**b**) *M2C-1bp*, (**c**) *M2C-2bp*, (**d**) *M2C-3bp*, (**e**) *M2C-4bp*, (**f**) *M2C-5bp*, (**g**) *M2C-6bp*, (**h**) *M2C-11bp*, and (**i**) *M2C-16bp*.

propeller-type G-quadruplex[180,267,281,282]. CD spectra of the quadruplex–duplex hybrids (Construct I, Figure 8.7; Construct II, Figure 8.8) corresponded well with the sum of their component duplex (Figure 8.9) and quadruplex (Figure 8.6a) spectra, as exemplified by *W2C-6bp* (Figure 8.6b) and *M2C-6bp* (Figure 8.6c).
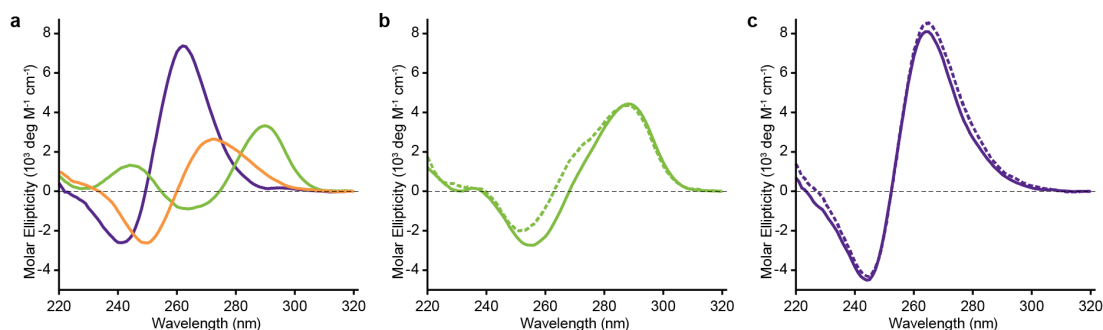
Figure 8.6 | CD spectra of representative duplex, quadruplex, and quadruplex–hybrid constructs. (**a**) CD spectra of *dx-6bp* (orange), *W2C-0bp* (green), and *M2C-0bp* (violet). (**b**) CD spectrum of *W2C-6bp* (solid green line) resembles the sum of the CD spectra of *W2C-0bp* and *dx-6bp* (dotted green line). (**c**) CD spectrum of *M2C-6bp* (solid violet line) resembles the sum of the CD spectra of *M2C-0bp* and *dx-6bp* (dotted violet line).

128
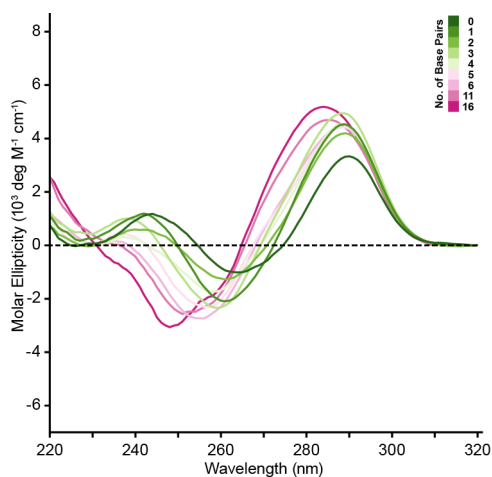
Figure 8.7 | CD spectra of Construct I with variable stem length. (**a**) *W2C-0bp*, (**b**) *W2C-1bp*, (**c**) *W2C-2bp*, (**d**) *W2C-3bp*, (**e**) *W2C-4bp*, (**f**) *W2C-5bp*, (**g**) *W2C-6bp*, (**h**) *W2C-11bp*, and (**i**) *W2C-16bp*. Color code is shown on the top right.
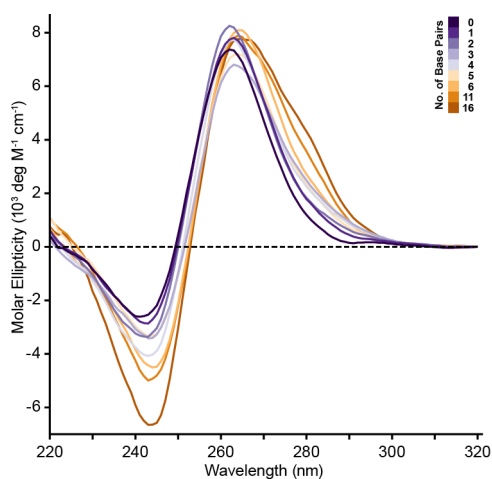
Figure 8.8 | CD spectra of Construct II with variable stem length. (**a**) *M2C-0bp*, (**b**) *M2C-1bp*, (**c**) *M2C-2bp*, (**d**) *M2C-3bp*, (**e**) *M2C-4bp*, (**f**) *M2C-5bp*, (**g**) *M2C-6bp*, (**h**) *M2C-11bp*, and (**i**) *M2C-16bp*. Color code is shown on the top right.
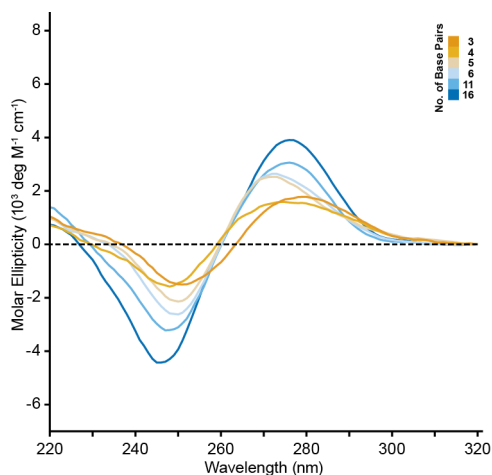
Figure 8.9 | CD spectra of reference duplex hairpins with variable stem length. (**a**) *M2C-0bp*, (**b**) *M2C-1bp*, (**c**) *M2C-2bp*, (**d**) *M2C-3bp*, (**e**) *M2C-4bp*, (**f**) *M2C-5bp*, (**g**) *M2C-6bp*, (**h**) *M2C-11bp*, and (**i**) *M2C-16bp*. Color code is shown on the top right.

CD-melting of duplex, quadruplex, and quadruplex–duplex hybrids. CD-melting of *dx-6bp*, *W2C-0bp*, and *M2C-0bp* were first performed. *dx-6bp* showed considerable variations at around 250 nm and 276 nm, and there was minimal change at ~262 nm (Figure 8.10a). *W2C-0bp* displayed a large change at ~290–295 nm and minimal change at around 276 nm (Figure 8.10b). *M2C-0bp* displayed the largest variation at 262 nm and almost no variation at around 250 nm (Figure 8.10c). Previous studies suggested that the 260-nm peak reflects quadruplexes in which two adjacent
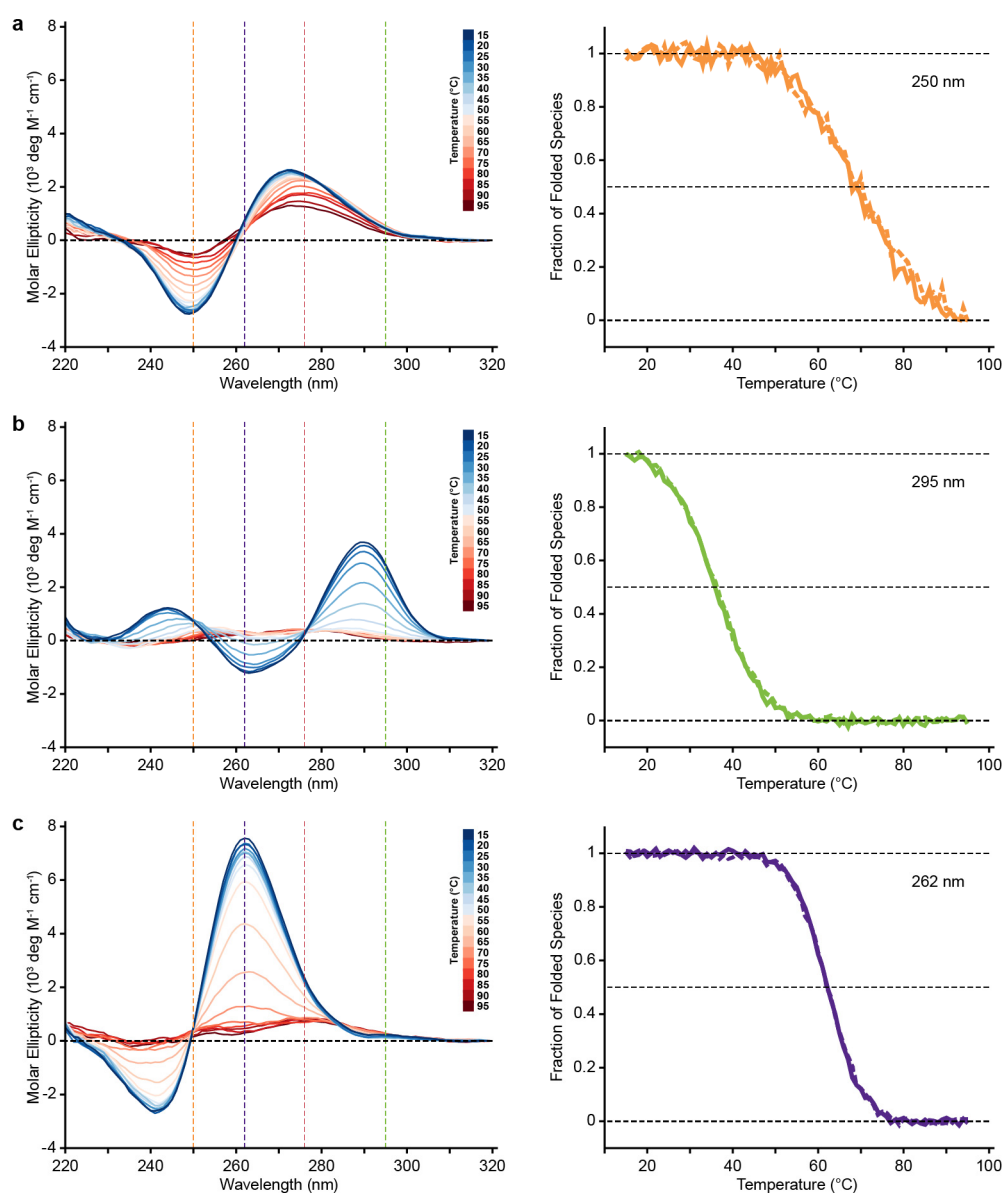
Figure 8.10 | CD-melting analyses of reference duplex and quadruplex constructs. On the left: CD spectra of (**a**) *dx-6bp*, (**b**) *W2C-0bp*, and (**c**) *M2C-0bp* from 15 to 95 °C (color code shown on the top right). On the right: Folding process of the different structural forms could be monitored through the peaks at (**a**) 250 nm (orange; hairpin segment), (**b**) 295 nm (green; Construct I), and (**c**) 262 nm (violet; Construct II). Folding and unfolding curves are represented as solid and dotted lines, respectively.
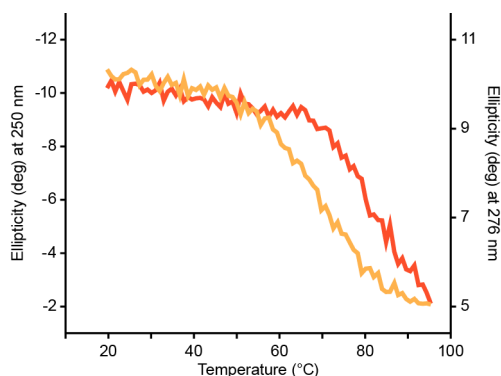
Figure 8.11 | CD-melting profiles of *dx-6bp* as monitored at different wavelengths. The signals at 250 nm and 276 nm are colored in yellow and orange, respectively. Only the cooling curve is shown here.

tetrads align with the same hydrogen-bond directionality, whereas the 295-nm peak reflects quadruplexes in which two adjacent tetrads possess alternate polarities[195]. The large variation shown by *W2C-0bp* at 295 nm coincides with a region of minimal change for *dx-6bp*, making it an ideal wavelength to trace the melting of the quadruplex segment of Construct I. Following the same logic, melting process of the quadruplex segment of Construct II could be monitored readily at 262 nm. In principle, melting process of the duplex segment could be followed through the 250-nm and 276-nm peaks. However, melting profiles of *dx-6bp* at these two wavelengths differed markedly (Figure 8.11), suggesting that they could have reflected distinct transition events of the hairpin stem. As such, subsequent melting analyses of the quadruplex–duplex hybrids were carried out for the quadruplex segment only, at 295 nm and 262 nm, respectively, for Construct I and Construct II. CD-melting of the series of reference duplex hairpins (Table 8.1) displayed minimal signal variation at these two wavelengths with respect to the quadruplex core (Figure 8.12), hence the melting temperatures determined at these two particular wavelengths should give a fair estimate of the stability of the respective quadruplex components. Note that for quadruplex–duplex hybrids with different sequence compositions of the duplex stem, monitoring the changes at alternative wavelengths might have been more suitable.

Effects of duplex stem length and sequence on stability of quadruplex–duplex hybrids. Folding curves for Construct I and Construct II (derived from molar ellipticity at 295 and 262 nm, respectively) with variable duplex stem lengths indicated a general positive correlation between duplex stem length and stability of a quadruplex–duplex hybrid (Figure 8.13). Melting temperature of Construct I increased from 36.2 °C (without any Watson–Crick base pair) to 41.6 °C (with three Watson–
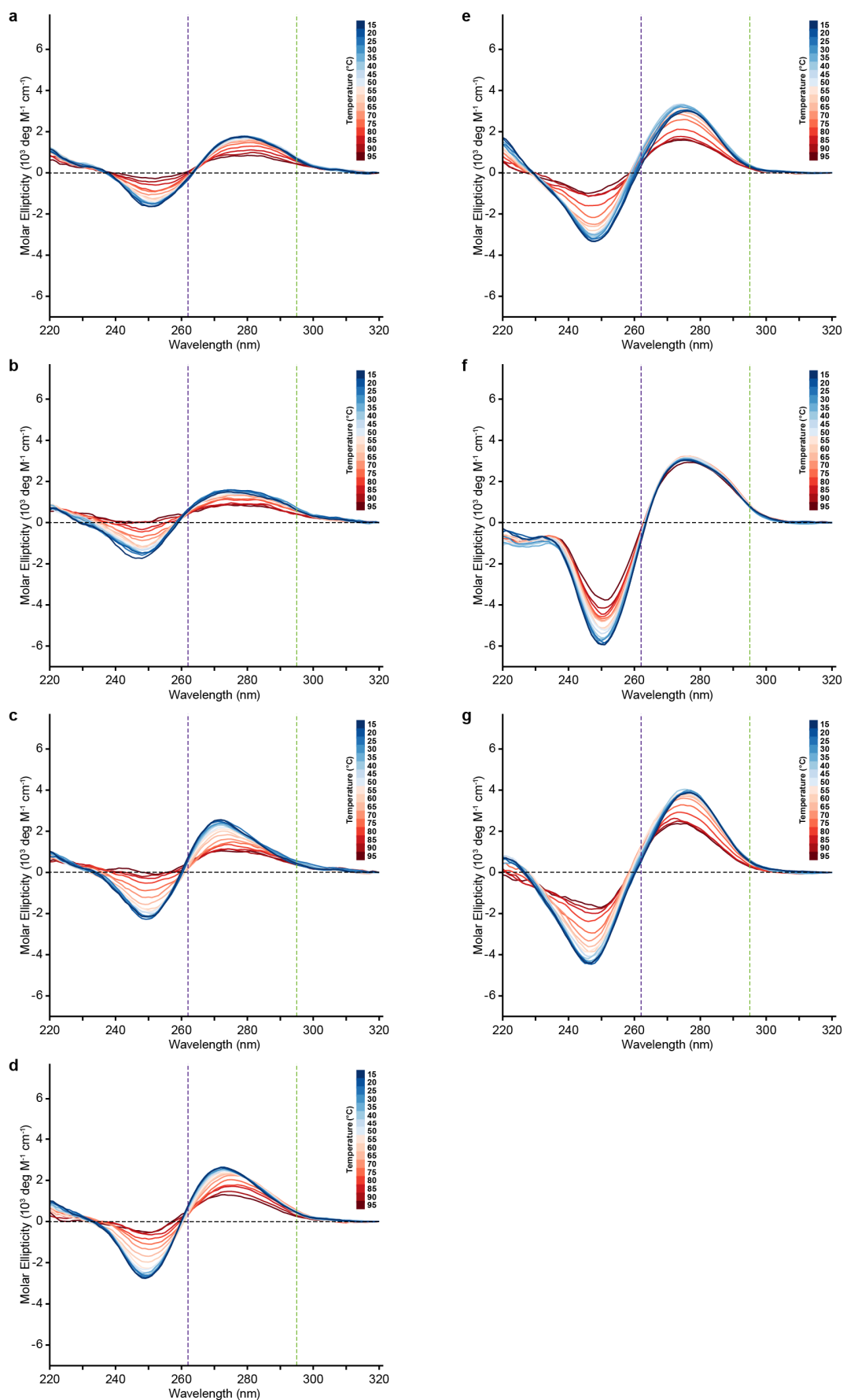
Figure 8.12 | CD-melting analyses of reference duplex hairpins. (a) *dx-3bp*, **(b)** *dx-4bp*, **(c)** *dx-5bp*, **(d)** *dx-6bp*, **(e)** *dx-11bp*, **(f)** *dx-11bp[CG]*, and **(g)** *dx-16bp*.

132
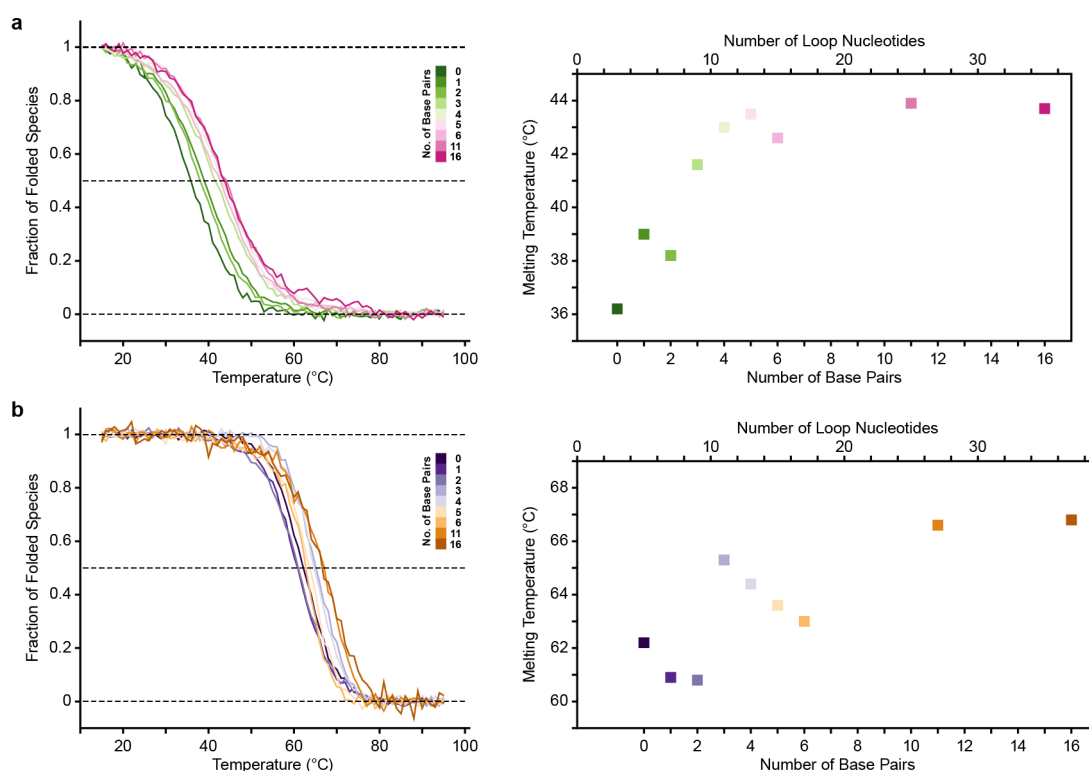
Figure 8.13 | Stability of quadruplex–duplex hybrids comprising different stem lengths.
(**a**) Folding curves (left; color code shown on the top right) and melting temperatures (right),
as monitored at 295 nm, for Construct I across various stem lengths. (**b**) Folding curves (left;
color code shown on the top right) and melting temperatures (right), as monitored at 262 nm,
for Construct II across various stem lengths.

Crick base pairs), eventually leveling off between 41–44 °C at longer duplex stem
lengths (Table 8.2 and Figure 8.13a). Melting temperature of Construct II showed a
marginal increase from 62.2 °C (without any Watson–Crick base pair) to ~63–65 °C at
intermediate duplex stem lengths (with three to six Watson–Crick base pairs),
followed by a modest increase (up to 5 °C) at long duplex stem lengths (Table 8.3 and
Figure 8.13b). This is in stark contrast to the trend observed for Construct II comprising
non-ordered loops (i.e. poly-dT), which showed a continuous decrease in melting
temperature (from 63.8 °C at T_5 to 41.6 °C at T_{37}) with increasing loop length[180,228-232]
(Table 8.3, Figures 8.14 and 8.15). Similar comparisons for Construct I could not be
made, as additional minor species were formed on top of the desired G-quadruplex
topology with a non-ordered loop as short as five nucleotides (Table 8.2 and Figure
8.16). Modification of the 11-bp duplex stem to poly-d(CG) led to an increase in
melting temperature of ~1 °C for both quadruplex–duplex constructs (Tables 8.2 and 8.3,
Figure 8.17).
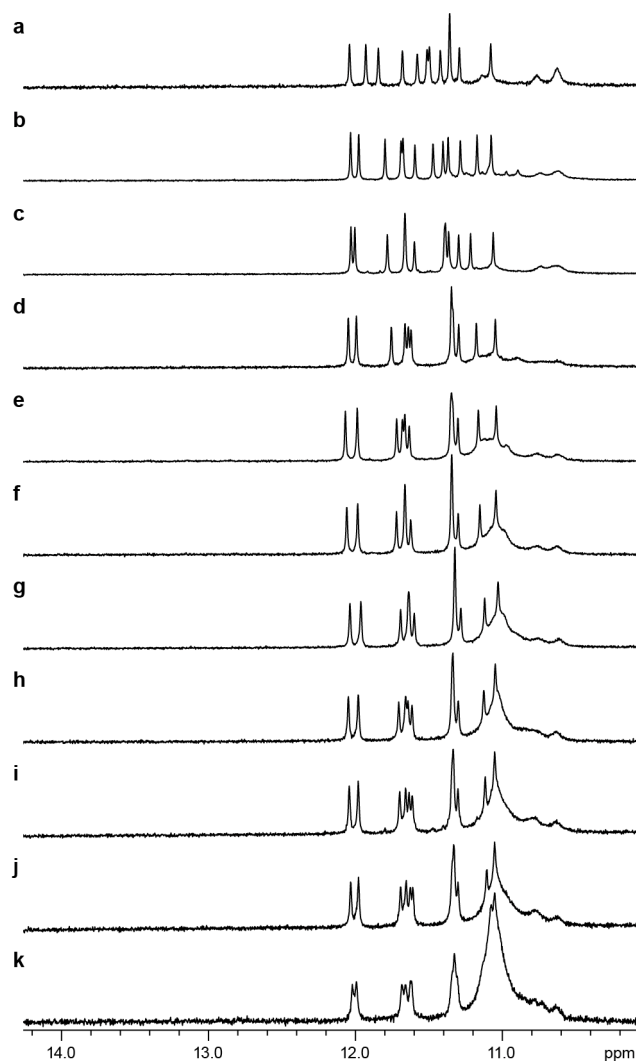
Figure 8.14 | 1D NMR imino proton spectra of Construct II comprising non-ordered loops of variable length. (**a**) *M2C-1T*, (**b**) *M2C-2T*, (**c**) *M2C-3T*, (**d**) *M2C-[T]3T[T]*, (**e**) *M2C-[T]5T[T]*, (**f**) *M2C-[T]7T[T]*, (**g**) *M2C-[T]9T[T]*, (**h**) *M2C-[T]11T[T]*, (**i**) *M2C-[T]13T[T]*, (**j**) *M2C-[T]15T[T]*, and (**k**) *M2C-[T]25T[T]*.

Effects of base pair composition at the junction on stability of quadruplex–duplex hybrids. The base pair step from the 6-bp stem immediately adjoining the quadruplex–duplex junction (i.e. C7•G21 for *W2C-6bp*; C11•G26 for *M2C-6bp*) was substituted like-for-like (with G•C, T•A, or A•T; Tables 8.2 and 8.3) to determine the changes that arise from base pair variability at the junction. The predominant conformation for both constructs remained the same, as indicated by NMR data (Figures 8.18 and 8.19). For Construct I, C•G at the junction led to the highest stability (with a melting temperature of 42.6 °C), G•C/A•T at the junction gave a melting temperature of ~36 °C, while T•A at the junction resulted in the least stable complex (with a drop in melting temperature exceeding 10 °C when compared against C•G) (Table 8.2). The disparity in stability is less pronounced in the case of Construct II,

134
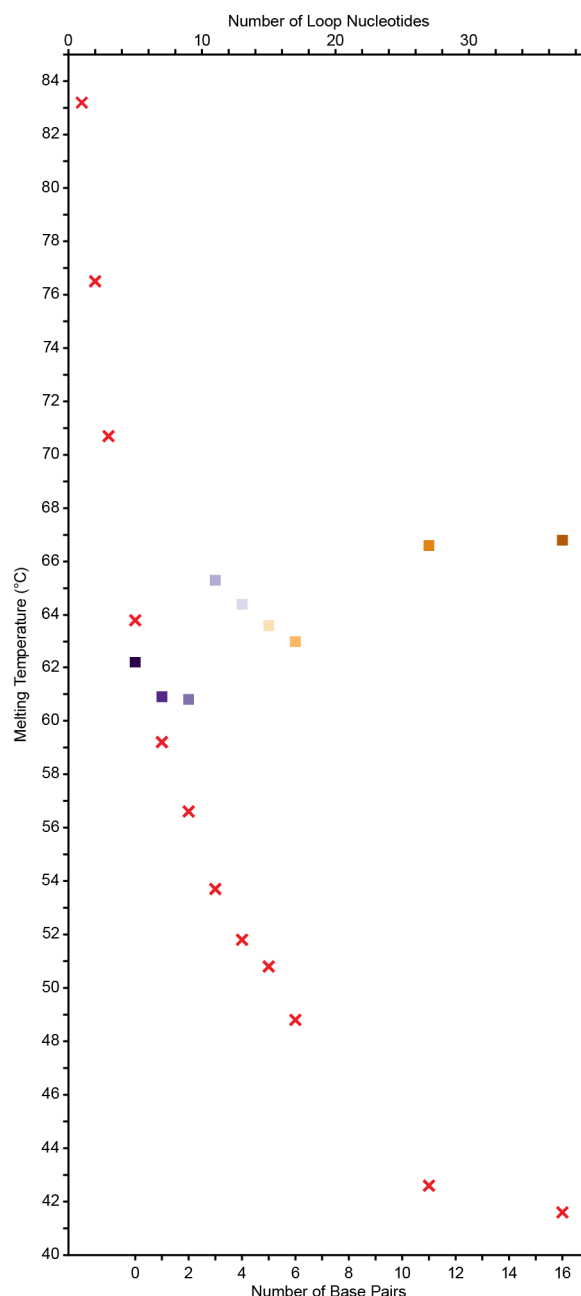
Figure 8.15 | Stability of Construct II across various stem and non-ordered loop lengths. Melting temperatures, as monitored at 262 nm, for Construct II across various stem lengths (color-coded as in Figure 8.13) and non-ordered loop lengths (in red crosses).

with C•G/G•C at the junction displaying a slightly higher melting temperature (~3 °C) as compared to either of T•A/A•T (Table 8.3).

Effects of bulges at the quadruplex–duplex junctions. We followed with an investigation regarding the effects of bulges at the quadruplex–duplex junctions (Tables 8.2 and 8.3) on the conformation and stability of the resulting complexes. For Construct I, a single bulge at the 5'- (Figure 8.20a) or 3'- (Figure 8.20b) end of the duplex stem adjoining the quadruplex was detrimental to adoption of the desired G-
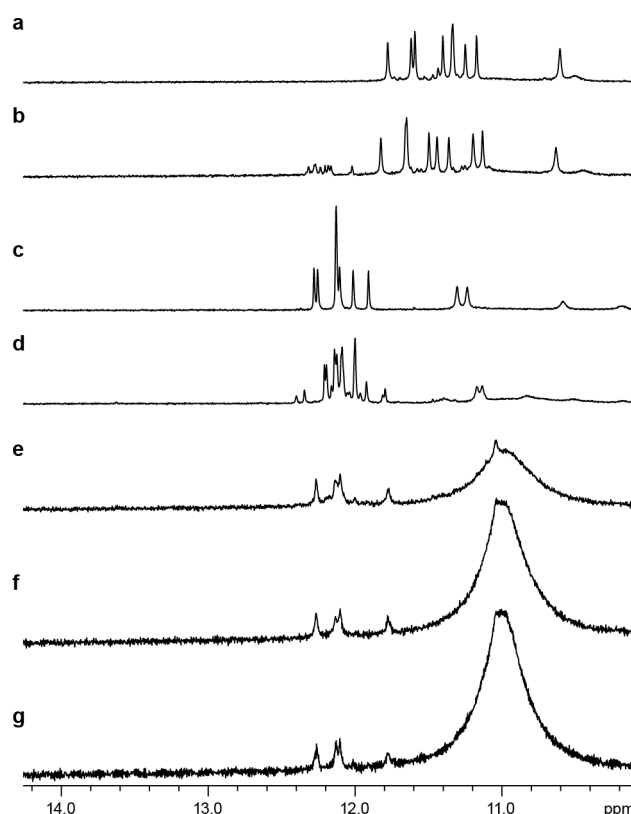
Figure 8.16 | 1D NMR imino proton spectra of Construct I comprising non-ordered loops of variable length. (**a**) *W2C-1T*, (**b**) *W2C-2T*, (**c**) *W2C-3T*, (**d**) *W2C-5T*, (**e**) *W2C-15T*, (**f**) *W2C-25T*, and (**g**) *W2C-35T*. Note that *W2C-1T* and *W2C-2T* are included for illustrative purposes and have most likely adopted an alternative (but related) folding topology, since one nucleotide and two nucleotides (to a lesser extent), respectively, are too short to bridge across the wide groove of the original construct.

quadruplex topology, with the latter posing a less severe effect. Presence of single-nucleotide bulges on both sides of the junction somewhat restored the adoption of the desired G-quadruplex topology (Figure 8.20c), but led to a drastic drop in stability (decrease in melting temperature of >16 °C with respect to *W2C-6bp*; Table 8.2). On the other hand, for Construct II, there was minimal emergence of minor conformation even with bulges of up to three nucleotides on both sides of the junction (Figures
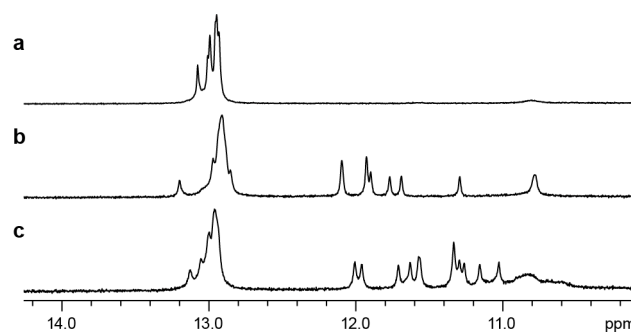
Figure 8.17 | 1D NMR imino proton spectra of duplex and quadruplex–duplex constructs with poly-d[CG] stem. (**a**) *dx-11bp[CG]*, (**b**) *W2C-11bp[CG]*, and (**c**) *M2C-11bp[CG]*.
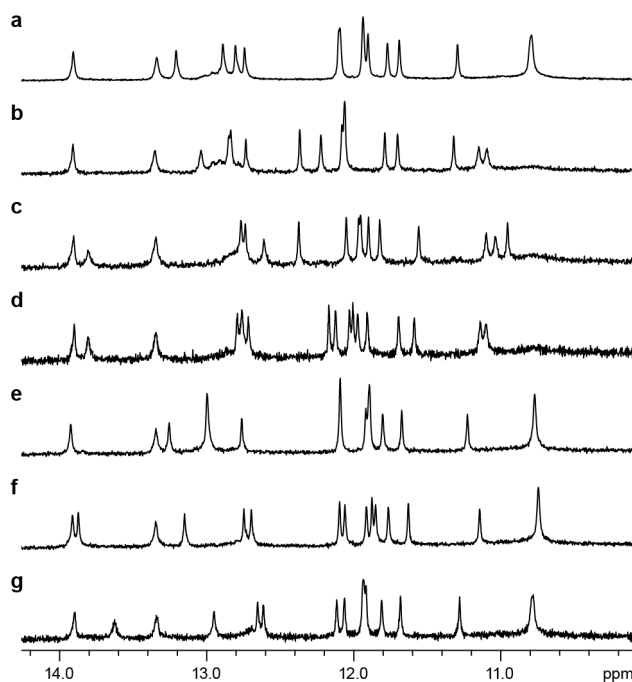
Figure 8.18 | 1D NMR imino proton spectra of base pair variants of *W2C-6bp* near the quadruplex–duplex junction. (**a**) *W2C-6bp*, (**b**) *W2C-6bp[G7/C21]*, (**c**) *W2C-6bp[T7/A21]*, (**d**) *W2C-6bp[A7/T21]*, (**e**) *W2C-6bp[C8/G20]*, (**f**) *W2C-6bp[T8/A20]*, and (**g**) *W2C-6bp [A8/T20]*.

8.21–8.22). The stability of Construct II comprising bulges of various sizes and sequence compositions did not seem to deviate much from that of *M2C-6bp* (Table 8.3).
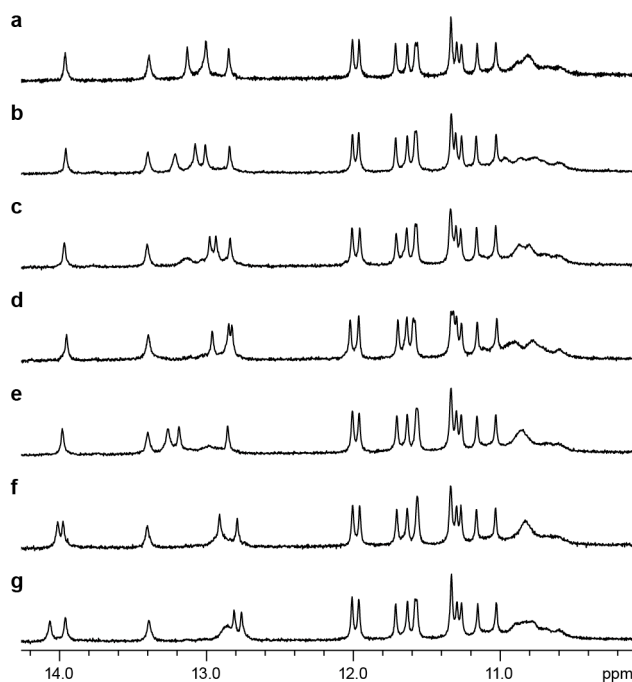
Figure 8.19 | 1D NMR imino proton spectra of base pair variants of *M2C-6bp* near the quadruplex–duplex junction. (**a**) *M2C-6bp*, (**b**) *M2C-6bp[G7/C21]*, (**c**) *M2C-6bp[T7/A21]*, (**d**) *M2C-6bp[A7/T21]*, (**e**) *M2C-6bp[C8/G20]*, (**f**) *M2C-6bp[T8/A20]*, and (**g**) *M2C-6bp [A8/T20]*.
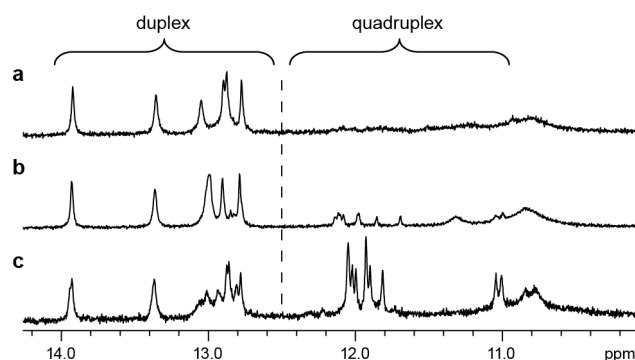
Figure 8.20 | 1D NMR imino proton spectra of Construct I with single-nucleotide bulges at the 5'- and/or 3'-ends of the duplex stem adjoining the quadruplex. (**a**) *W2C-6bp[+T7]*, (**b**) *W2C-6bp[+T21]*, and (**c**) *W2C-6bp[+T7/+T22]*.

Design and prediction of quadruplex–duplex hybrids. Previously, we elaborated the compatibility of quadruplex and duplex DNA with the design and structural elucidation of five representative quadruplex–hairpin constructs[273], which brought forward the conceptualization of a myriad of quadruplex–duplex hybrids and G-junction motifs. Here we investigated the thermal stability of two representative classes of quadruplex–duplex hybrids, either with continual stacking between the duplex and quadruplex segments (Construct I) or with orthogonal orientation of the
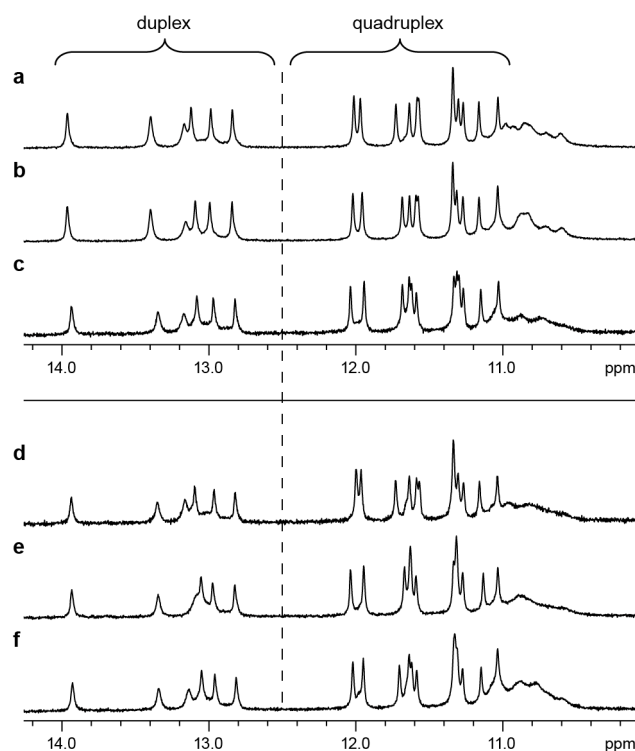
Figure 8.21 | 1D NMR imino proton spectra of Construct II with bulges at the 5'- and 3'-ends of the duplex stem adjoining the quadruplex. (**a**) *M2C-6bp[+T11]*, (**b**) *M2C-6bp[+T26]*, (**c**) *M2C-6bp[+T11/+T27]*, (**d**) *M2C-6bp[+T11T12]*, (**e**) *M2C-6bp[+T26T27]*, and (**f**) *M2C-6bp[+T11T12/+T28T29]*.

138
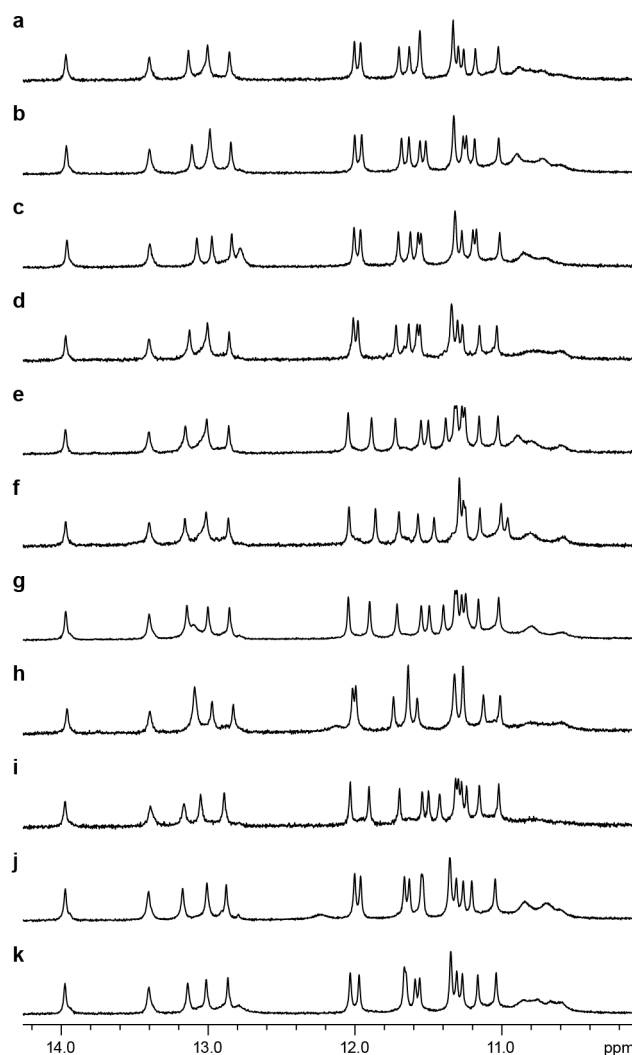
Figure 8.22 | 1D NMR imino proton spectra of Construct II with variable bulge compositions. (**a**) *M2C-6bp[+T11]*, (**b**) *M2C-6bp[+T26]*, (**c**) *M2C-6bp[+T11/+T27]*, (**d**) *M2C-6bp[+T11T12]*, (**e**) *M2C-6bp[+T26T27]*, and (**f**) *M2C-6bp[+T11T12/+T28T29]*.

duplex stem with respect to the quadruplex core (Construct II). These findings should be applicable to quadruplex–duplex hybrids comprising similar arrangement of duplex and quadruplex helices. Note however that quadruplex–duplex hybrids having alternative forms of connectivity at the junction (e.g. with a duplex stem adjoining the diagonal corners of a tetrad) might exhibit different behaviors and would warrant further studies. Insights gained from this study should contribute towards the design and prediction of these motifs (e.g. in the sequence design stage for DNA nanomaterials[80] or in the search for putative quadruplex sequences in the human genome[41,42]).

Duplex elements serve a guiding role in the establishment of quadruplex–duplex hybrids. G-quadruplexes have been demonstrated to exist in a multitude of

conformations[30]. To date, folding principles governing their formation have not been fully understood. Design of quadruplex topology through control over the glycosidic conformations of the guanine bases constituting the tetrad core has been reported previously[167,169-171,202,287]. In a recent study, duplex elements have been utilized to pre-orient three G-rich strands for the induction of a trimolecular G-quadruplex[288]. Similarly, in the development of quadruplex–duplex hybrids, we couple the integration of duplex stems as quadruplex loops with the strategic placement of auxiliary structural elements to drive G-quadruplex formation towards the desired topology[273] [see Section 7.2]. In the case of Construct I, this involves direct extension of the duplex stem from the tetrad core, as the geometry of the duplex stem matches favorably across the wide groove of G-quadruplexes. Facile formation of the duplex stem restricts the conformational space that the G-tracts could explore and guides them into the proper spatial orientation whereby the desired topology can then be attained. This guiding role of the duplex stem contributed substantially towards the adoption of the desired G-quadruplex topology; considerable minor species emerged with a non-ordered loop (in place of the duplex stem) that is just five-nucleotide-long. In the case of Construct II, the presence of the duplex stem is not as crucial in driving the G-quadruplex to a single topology, presumably since the GGGTGGG motif has an extreme propensity to assume a parallel-stranded arrangement of the two G-tracts via a double-chain-reversal loop[204,228,289,290].

Longer loops do not necessarily lead to less stable G-quadruplex structures. Prior investigations on the effects of loop length towards the conformation and stability of G-quadruplexes generally agreed that short loops (≤2 nt) show a preference towards the propeller configuration[180,228,231,252], and that longer loops tend to have a destabilizing influence on G-quadruplex structures[180,228-232]. Note that in these studies the long loops that have been investigated do not contain any explicit stabilizing interactions, which might account for their destabilizing nature. It has generally been accepted that a stable quadruplex should have loops that are no more than seven nucleotides long[42]. Recently we described the structure of a parallel G-quadruplex comprising a nine-nucleotide central propeller loop, in which interactions between sections of the long loop and capping elements of the G-tetrad core contribute towards its overall stability[290]. Here we show that a long loop (in the form of a duplex stem) by itself could also exert a stabilizing influence on a quadruplex. For both constructs, a

140

general increase in melting temperature was observed with increasing duplex stem length of two or more base pairs (at which point the duplex stem is established judging from signature Watson–Crick imino proton NMR peaks). This upward trend in melting temperature becomes more gradual with increasing duplex stem length and eventually levels off (after ~4–6 bp for Construct I; after ~5–10 bp for Construct II).

Contexts at the junctions influence the stability and uniformity of quadruplex–duplex hybrids. We have shown that variability at the base pair step immediately adjoining the quadruplex–duplex junction could alter the melting temperature of the complex to a significant extent, which differs in the case of Construct I and Construct II. For Construct II, the outcome is of a more predictable nature; G•C Watson–Crick base pairs are slightly more stable than A•T in general. On the other hand, the large differences in stability observed for Construct I suggest that other factors could be in play at the stacked quadruplex–duplex interface. This observation provided a gauge on the extent of fluctuations in stability associated with changes in base pair composition near the junction.

We have also explored the addition of bulges onto both sides of the quadruplex–duplex junction. For a junction involving continual stacking of bases, as is the case in Construct I, a single bulge at either side of the junction is detrimental to adoption of a single quadruplex conformation for the quadruplex–duplex hybrid. It is interesting to note that a bulge at the 3'-end of the duplex stem poses a less severe effect than does a bulge at the 5'-end. Presence of single-nucleotide bulges on both sides of the junction, which could either be flipped out[284] or stacked[265,283] at the quadruplex–duplex interface, restored the desired G-quadruplex topology as the major species, but is damaging to the stability of the complex nonetheless. In contrast, in the case of propeller duplex stem-loop (Construct II), bulges (up to three nucleotides on both sides of the quadruplex–duplex junction) do not seem to adversely affect formation of the desired G-quadruplex topology. This could partly have been offset by the robust nature of the all-parallel-stranded G-tetrad core. The number of bulge nucleotides on either side of the junction dictates the flexibility and projection of the propeller duplex stem-loop and could potentially be utilized to fine-tune the relative orientation of the duplex and quadruplex helices in the design for DNA nanodevices.

Stacking (or non-stacking) at the quadruplex–duplex interface. We have examined two contrasting modes of quadruplex–duplex connectivity: one that involves

the continual stacking of bases at the interface (Construct I), and another that involves the disruption of base stacking across the duplex and quadruplex segments (Construct II). The presence/absence of stacking (or the coupling/non-coupling between the duplex and quadruplex segments) is reflected in the NMR and CD spectra: (1) for Construct I base pair addition/substitution at the junction resulted in substantial variations to chemical shifts of the tetrad H1 protons, whereas for Construct II the chemical shifts of tetrad H1 protons did not vary much with base pair alterations at the junction; (2) for Construct I the CD spectrum of the quadruplex–duplex hybrid did not match closely to the sum of the respective duplex and quadruplex component spectra at ~250–280 nm, whereas for Construct II the CD spectrum of the quadruplex–duplex hybrid matches closely to the sum of the respective duplex and quadruplex component spectra. These differences in stacking/non-stacking at the quadruplex–duplex interface are also translated as distinctive trends in stability/uniformity of adopted G-quadruplex topology with regards to changes in composition at the junction. Construct I is affected to a greater extent by alterations at the quadruplex–duplex interface (wild fluctuations in melting temperature with variation of base pair at the junction; introduction of bulges extremely detrimental to adoption of the desired G-quadruplex topology), whereas Construct II appears to be more refractory to such changes (slight changes in melting temperature with variation of base pair at the junction; introduction of bulges more or less inconsequential to adoption of the desired G-quadruplex topology).

Conclusion

DNA has the capacity to adopt several distinct structural forms such as duplex and quadruplex helices, which have been implicated in cellular processes and shown to exhibit important functional properties. Quadruplex–duplex hybrids, generated from the juxtaposition of these two structural elements, could find applications in therapeutics and nanotechnology. It is thus important to establish an understanding on the stability of these hybrid complexes.

We have investigated the thermal stability of two classes of quadruplex–duplex hybrids: Construct I involves the coaxial conjointment of duplex and quadruplex helices with continual base stacking across the two components; Construct II involves the orthogonal connection of duplex and quadruplex helices with no base stacking between the two components. For both constructs, stability of the quadruplex generally

increases with the length of the stem-loop incorporated, up to a point, in contrast to quadruplexes comprising non-structured loops, which showed a continuous decrease drop in stability with increasing loop length. Depending on the context, alterations to the quadruplex–duplex junction could substantially affect the conformation and stability of these structures.

These findings pave the way for future investigations on more complex quadruplex–duplex hybrid complexes, either involving variant junction types or increasing number of duplex stems. A better thermodynamic understanding on these complex structures will in the future design and prediction of these motifs.

Chapter 9

Putative Quadruplex Motifs Containing Duplex Stem Elements in the Human Genome

9.1   G-Quadruplex-Forming Motifs in the Human Genome

Biological relevance of G-quadruplex. Guanine-rich sequences have a propensity to fold into G-quadruplexes[28-30], which are helical complexes constructed through the multiple stacking of G•G•G•G tetrads[31]. The interest in G-quadruplex DNA targeting stems from the seminal finding that formation of G-quadruplex structures by telomeric substrates could inhibit the enzyme telomerase[159], which is responsible for the maintenance of telomere length in most cancer cells[43,154,155]. Induction/stabilization of these structures by specific ligands would thus make for an attractive anticancer strategy[30,61]. Targeting of genomic quadruplexes was first demonstrated against a G-rich fragment within nuclease hypersensitivity element III$_1$ (NHE III$_1$) of the c-myc promoter by the porphyrin TMPyP4[45], which led to downregulation of c-myc transcription. G-rich sequences have since been identified in other promoters including c-kit[237,239], KRAS[249], BCL-2[244], and hTERT[233,274,291], prompting the view that these motifs could play a role in gene regulation at the transcriptional level[46,47].

Prediction of quadruplex-forming motifs in the human genome. The G-quadruplex could exist in a diverse range of folding topologies, brought about by the relative alignments of the four core strands and the manner in which they are connected by linkers (known as loops). In addition, the number of component strand(s) making up the complex can differ. A typical intramolecular G-quadruplex–forming sequence would consist of four G-tracts interspersed with three loops (Figure 9.1a), while exceptions have also been observed[225,239]. Thermodynamic studies have concurred on the notion that shorter loops (preferably ≤7 nt) lead to more stable G-quadruplex structures[180,228-230,232]. Based on these observations, various algorithms have been implemented in the identification of putative quadruplex sequences (PQS) in the human genome[41,42,261,292,293], all variants of the generic expression $G_{X1}N_{L1}G_{X2}N_{L2}G_{X3}N_{L3}G_{X4}$, in which the G-tract (X) and loop length (L) were largely restricted to 2–5 and 7 nt, respectively. For instance, the *quadparser* algorithm (X≥3, L=1–7) identified >350,000 PQS in the human genome[42].

On the other hand, stem-loop (or hairpin) motifs (Figure 9.1b) have also been extensively studied, as they are intricately involved in nucleic acid secondary structure formations. Recent demonstrations of the stable formation of quadruplex–duplex constructs[265,273] called for the inclusion of duplex stem-loops as loop elements in the evaluation of PQS (Figure 9.1c), which would necessitate a consideration of longer

146
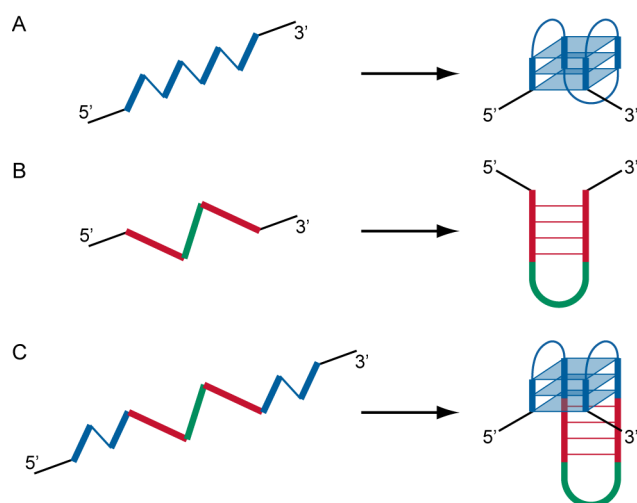
Figure 9.1 | Formation of quadruplex, duplex stem-loop, and stem-loop–containing quadruplex. (**a**–**c**) Schematic diagrams illustrating the formation from a single DNA strand of (**a**) a G-quadruplex, (**b**) a duplex stem-loop, and (**c**) a G-quadruplex containing a duplex stem-loop. G-tracts (thick lines) and quadruplex loops (thin lines) are colored in blue, self-complementary tracts are colored in red, while hairpin loops are colored in green.

loops (L>7). Here we identified stem-loop–containing PQS (SLC-PQS) in the human genome, some of which are situated at biologically interesting loci. We followed up with NMR characterization on a selected list of these sequences and validation of the folding topology adopted by one such motif. Structural features peculiar to these quadruplex–duplex hybrids could lead to novel drug design strategies.

9.2 Putative Quadruplex Motifs Containing Duplex Stem Elements in the Human Genome

Identification of stem-loop–containing PQS in the human genome. A previous survey of PQS (based on the definition $G_{\geq3}N_{1-7}G_{\geq3}N_{1-7}G_{\geq3}N_{1-7}G_{\geq3}$) in the human genome revealed more than 350,000 candidates[42]. It has further been shown that these motifs are highly overrepresented in the region immediately upstream of transcription start sites[47] (TSS) and are strongly associated with specific functional classes of genes[46]. These observations corroborated the proposition of a direct regulatory role of PQS in gene transcription. Here we sought to identify SLC-PQS in the human genome, with a particular focus on promoter (defined as the region ≤2 kb upstream of TSS) and genic (22,437 gene symbols in the human RefSeq database[294]) regions. To begin with, parameters for the general search algorithm $G_{X1}N_{L1}G_{X2}N_{L2}G_{X3}N_{L3}G_{X4}$ were relaxed (X=3–6, L=1–20; denoted as PQS_{L20}) to accommodate longer loops that might harbor stem-loop elements. Based on this algorithm, a total of 2,933,131 PQS_{L20} were

identified. Note that these sequences are overlapping, meaning different PQS_{L20} can share the same G-tract. PQS_{L20} in which L≤7 for all three loops were sub-classified as PQS_{L7}. For all other PQS_{L20} (i.e. any of L1, L2, or L3 is >7 nt in length), the long loops (L>7) were subsequently screened for the existence of stable stem-loops using the UNAFold package[295]. Only in cases where these loops form stable stem-loops with a base pair composition of >50% (with respect to the loop length) were the sequence classified as SLC-PQS.

Based on the selection criteria, 80,307 SLC-PQS (out of 60,172 unique stretches of DNA), as compared to 605,491 PQS_{L7} (out of 360,438 unique stretches of DNA) were identified in human genome. Subsequent analyses were performed on the 48,585 SLC-PQS located in the promoter and genic regions, which are spread across 12,315 genes. A detailed breakdown of their distribution across different gene segments (as referenced against the RefGene annotation, UCSC Genome Browser[296]) in comparison with PQS_{L7} is presented in Table 9.1. Out of these SLC-PQS in the promoter and genic regions, 33,247 are non-overlapping with PQS_{L7}. These would constitute an additional pool of genomic quadruplexes, distinct from the PQS_{L7} that have already been identified, that could potentially be targeted through ligand design[93]. We note that 397 SLC-PQS were identified in 285 genes for which PQS_{L7} were otherwise not detected. On the other hand, SLC-PQS that coincide with PQS_{L7} would extend the range of the putative quadruplex-forming regions.

Mapping of stem-loop–containing PQS to genomic loci of interest. Multiple annotation sets were overlaid across the sequence hits in order to isolate SLC-PQS with potential functional and therapeutic implications. First, the sequence hits were

Table 9.1 | Distribution of PQS$_{L7}$ and SLC-PQS across various gene segments.

Gene Segment	PQS$_{L7}$			SLC-PQS			Regulatory SLC-PQS ^{a}		
	Non-Temp ^{b}	Template	Symbols ^{c}	Non-Temp ^{b}	Template	Symbols ^{c}	Non-Temp ^{b}	Template	Symbols ^{c}
Promoter or Genic	166,964	173,924	19,099	27,365	23,184	12,315	7,478	5,967	6,172
Promoter	25,781	25,924	12,542	4,507	4,487	5,159	2,568	2,595	3,212
Genic	146,529	152,560	18,310	23,976	19,478	11,361	5,587	3,817	5,022
Exon	12,276	18,857	10,059	3,090	4,356	4,560	1,252	1,320	1,937
CDS	2,731	7,026	4,294	1,061	2,091	2,081	286	404	565
Intron	136,399	136,129	17,339	21,609	15,881	10,143	4,598	2,695	4,000
5'-UTR	3,238	4,283	3,774	990	984	1,417	640	600	948
3'-UTR	4,658	5,639	4,264	650	881	1,119	109	126	195

^{a} Number of SLC-PQS mapped to gene regulatory regions (chromatin accessibility and transcription factor-binding sites). ^{b} Sequence located on non-template strand. ^{c} Number of gene symbols for which the sequences are located.

mapped against chromatin accessibility regions and transcription factor-binding sites (annotations from UCSC Genome Browser[296]) to reveal 3,212 SLC-PQS within putative gene regulatory regions (Table 9.1). Cross-mapping and manual inspection of these sequences across various gene databases yielded numerous SLC-PQS that are located within transcriptional active sites of cancer-associated genes (oncogenes, key regulatory genes, or genes involved in cell proliferation and apoptosis), which could serve as prospective anticancer targets[61,235]. Examples of these genes, among which include the pim-1, cyclin Y, and midkine genes, are summarized in Table 9.2. In addition, sequence hits in the promoter and genic regions were also cross-checked against mutation (The Cancer Genome Atlas[297] (TCGA)) and SNP (dbSNP, NCBI[298]) databases to further identify SLC-PQS that are situated across these genomic loci of potential interest/relevance (Table 9.2).

Table 9.2 | Selective genes for which SLC-PQS have been located within.

Gene Symbol	Description	RefSeq ID(s)	SLC-PQS ID(s)	SLC-PQS Placement	Gene Segment(s)
PDK1	pyruvate dehydrogenase kinase, isozyme 1	NM_002610	G4ST02173128695	non-template	promoter
IGF2	insulin-like growth factor 2 (somatomedin A)	NM_006548, NM_001007225	G4ST03187025840	non-template	promoter
RICTOR	RPTOR independent companion of MTOR, complex 2	NM_152756	G4ST05039110287	non-template	promoter
PIM1	pim-1 oncogene	NM_002648	G4ST06037246104	template	5'UTR
MYB	v-myb myeloblastosis viral oncogene homolog (avian)	NM_001080416, NM_001144755	G4ST08067688172	template	promoter
CCNY	cyclin Y	NM_145012	G4ST10035665408	non-template	promoter
RET	ret proto-oncogene	NM_020630, NM_020975	G4ST10042891015, G4ST10042891022, G4ST10042892416	template	promoter
PDCD4	programmed cell death 4 (neoplastic transformation inhibitor)	NM_014456	G4ST10112621932	template	intron1
MDK	midkine (neurite growth-promoting factor 2)	NM_001012333	G4ST11046359281	template	promoter/ 5'UTR
PRKRIR	protein-kinase, interferon-inducible double stranded RNA dependent inhibitor, repressor of (P58 repressor)	NM_004705	G4ST11075769784	non-template	promoter
NDRG2	NDRG family member 2	NM_201535, NM_201540	G4ST14020563933, G4ST14020563945	template	promoter
HIC1	hypermethylated in cancer 1	NM_006497, NM_001098202	G4ST17001906024, G4ST17001906030, G4ST17001906034	template	promoter/ intron
RAB12	member RAS oncogene family	NM_001025300	G4ST18008599361	template	promoter
RAB3D	member RAS oncogene family	NM_004283	G4ST19011311393, G4ST19011311397	template	promoter
CD24	CD24 molecule	NM_013230	G4ST24019614192	non-template	promoter
BMP8A	bone morphogenetic protein 8a	NM_181809	G4ST01039729840	non-template	promoter
MYO9B	myosin IXB	NM_004145, NM_001130065	G4ST19017166594	template	exon
MYT1L	myelin transcription factor 1-like	NM_015025.2	G4ST02001786748	non-template	intron
ARHGAP31	Rho GTPase activating protein 31	NM_020754	G4ST03120547683	template	intron
ELFN1	extracellular leucine-rich repeat and fibronectin type III domain containing 1	NM_001128636.2	G4ST07001717478	template	intron

NMR characterization of stem-loop–containing PQS. To investigate the potential formation of quadruplex–duplex hybrids[273] by SLC-PQS, we proceeded with the NMR characterization on a selected list of sequence hits (Table 9.3). Imino protons (from guanine and thymine bases) of Watson–Crick base pairs typically resonate around 12.5–14.5 ppm while guanine imino protons of G-tetrads mostly resonate at 10.5–12.5 ppm. For the series of SLC-PQS inspected, most of them displayed imino proton peaks in both Watson–Crick and G-tetrad regions (Figure 9.2), pointing to the coexistence of duplex and quadruplex elements and the likely formation of quadruplex–duplex hybrids. Many of these sequences exhibited multiple conformations, as shown by the number and intensity of duplex and tetrad imino proton peaks (e.g. *CAG01*, *CAG03*, and *CAG07*; Figure 9.2a,c,g, respectively). In a few cases, the spectra showed the presence of a single major conformation. For instance, *OTH01* and *OTH03* displayed distinct sharp peaks amenable for detailed structural elucidation (Figure 9.2h,j, respectively; see validation of the former below). In other cases, weak/broadened (*CAG06*; Figure 9.2f) or fewer-than-expected (*CAG04*; Figure 9.2d) duplex imino proton peaks suggested the absence of a stable stem-loop or the adoption of an alternative structure altogether.

Effects of nucleotide changes (arising from mutation or SNP) on the formation of quadruplex–duplex hybrids by SLC-PQS were also examined. The MYO9B gene sequence *MUT01* and its mutated counterpart *MUT01del* with a GAGAGT deletion

Table 9.3 | Representative SLC-PQS oligonucleotides used for NMR study.

Name	Sequence[a,b,c]	SLC-PQS ID
CAG01	5'-GC **GGG** CGCGGCGCGCG **GGG** A **GGGG** AA **GGGG** T-3'	G4ST05039110287
CAG02	5'-GC **GGG** A **GGG** CGCGCCAGCG **GGG** TC **GGG** C-3'	G4ST06037246104
CAG03	5'-GA **GGG** C **GGGG** CT **GGGG** AGCTGGAAGCT **GGG** A-3'	G4ST08067688172
CAG04	5'-GA **GGG** C **GGG** CGCCGCTGGCGA **GGG** A **GGG** C-3'	G4ST10035665408
CAG05	5'-CA **GGG** AA **GGG** ACCTGATAGGTA **GGG** A **GGGG** C-3'	G4ST10042891022
CAG06	5'-CT **GGG** GCGGTTTCCGC **GGG** T **GGG** CA **GGG** A-3'	G4ST11046359281
CAG07	5'-GT **GGG** G **GGG** A **GGG** GGGAGCCACGCAGCTCCCA **GGGG** A-3'	G4ST17001906030
OTH01	5'-A **GGG** AGAGGAGAGCTCT **GGG** TT **GGG** T **GGG**-3'	G4ST02001786748
OTH02	5'-GT **GGG** TGCAGTCAGAGCTGCT **GGG** AGT **GGG** TAGCCTCAGGCTA **GGG**-3'	G4ST03120547683
OTH03	5'-CA **GGG** T **GGG** TCTGCTGTGCAG **GGG** T **GGG** T-3'	G4ST07001717478
MUT01	5'-CT **GGGG** T GG(G AGAGT)CTTCTCT **GGG** CTT **GGGG** C-3'	G4ST19017166594
MUT01del	5'-CT **GGGG** T GG CTTCTCT **GGG** CTT **GGGG** C-3'	
SNP01A	5'-CC **GGGGG** T **GGG** C **GGG** C(A)GCACAGACGGCTGAGCCG **GGG** C-3'	G4ST01039729840
SNP01G	5'-CC **GGGGG** T **GGG** C **GGG** C(G)GCACAGACGGCTGAGCCG **GGG** C-3'	

[a] G-tracts are shown in bold. *[b]* Self-complementary tracts are underlined. *[c]* Nucleotide deletion/polymorphisms are bracketed.
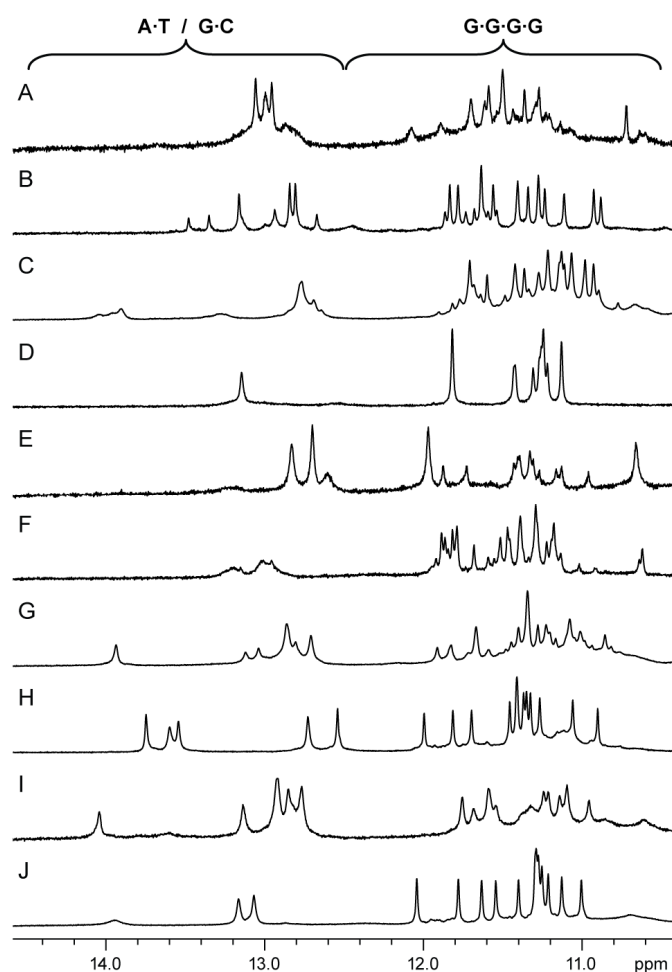
Figure 9.2 | 1D NMR imino proton spectra of SLC-PQS identified across various genes in the human genome. (**a**) *CAG01*, (**b**) *CAG02*, (**c**) *CAG03*, (**d**) *CAG04*, (**e**) *CAG05*, (**f**) *CAG06*, (**g**) *CAG07*, (**h**) *OTH01*, (**i**) *OTH02*, and (**j**) *OTH03*. Chemical shifts of typical Watson–Crick base-paired (A•T/G•C) and G-tetrad (G•G•G•G) imino protons are demarcated.

(that removes a guanine residue constituting one of the four G-tracts, together with one half of the duplex stem), which is observed in tumor cell lines (TCGA mutation database[297]), showed strikingly different imino proton spectra (Figure 9.3a,a'),
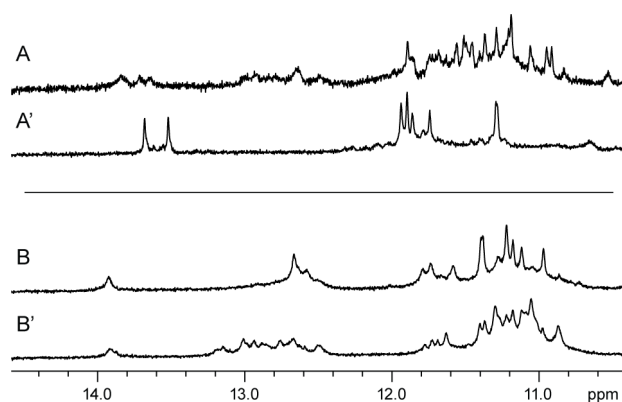
Figure 9.3 | Nucleotide changes affecting the populations of different forms of SLC-PQS. Shown here are the 1D NMR imino proton spectra of (**a**) *MUT01* and (**a'**) its mutated counterpart *MUT01del* from the MYO9B gene, as well as two single-nucleotide polymorphs of the BMP8A gene sequence (**b**) *SNP01A* and (**b'**) *SNP01G*.

151

indicating their adoption of completely different folding topologies. Single-nucleotide changes, in the form of SNP, could also affect the proportions of different SLC-PQS populations, as evidenced by two polymorphs of the BMP8A gene segment *SNP01* (*SNP01A* and *SNP01G*; Figure 9.3b,b').

Structural validation of a stem-loop–containing PQS. Detailed NMR structural characterization of the MYT1L gene sequence *OTH01* (d[**AGGG**AGAGGAGAG CTCT**GGG**TT**GGG**T**GGG**]; tetrad guanine residues in bold, self-complementary tracts underlined) was performed. Imino proton spectrum of *OTH01* indicated the adoption of a single predominant quadruplex–duplex hybrid structure; five major peaks at 12.5–13.8 ppm corresponded to the formation of five base pairs while twelve major peaks from 10.9–12.0 ppm corresponded to the establishment of three G-tetrads (Figure 9.4a). Concentration-independent melting profile of the CD spectra (melting temperature of around 54 °C at 40–50 mM K$^+$), as monitored at 295 nm, indicated the formation of an intramolecular structure by *OTH01*. Unambiguous assignments for H1 (Figure 9.4a) and H8 (Figure 9.4c) protons of selective guanines were achieved based on site-specific 2% ^{15}N-enrichment[192] and ^{2}H labeling[207] of samples, respectively. Through-bond correlation experiments (^{13}C–^{1}H-HSQC, TOCSY, and COSY) facilitated the determination of the H6/H8–H1' sequential connectivity of *OTH01* (Figure 9.4e). Guanine residues G2, G18, G19, G23, and G27 assume *syn* glycosidic conformation (as shown by their intense intraresidue H8–H1' NOE correlations), whereas the other tetrad guanine residues adopt *anti* glycosidic conformation. Cyclic H1–H8 NOE connectivity within individual tetrads (Figure 9.4d,h) indicated formation of a (3+1) G-tetrad core consisting of three G-tetrad layers, G20•G2•G27•G23, G19•G24•G28•G3, and G18•G25•G29•G4, with the first tetrad aligned in the opposite hydrogen-bond directionality as compared to the latter two (Figure 9.4i). Guanine glycosidic conformations are *anti•syn•syn•syn* around the first tetrad and *syn•anti•anti•anti* around the two other tetrads. The slower solvent exchange rate shown by H1 protons of G3, G19, G24, and G28 (Figure 9.4b) is consistent with the central placement of G19•G24•G28•G3 (between the G20•G2•G27•G23 and G18•G25•G29•G4 tetrads) in the tetrad core. Signature imino–H2 (A•T base pair; Figure 9.4f) and imino–amino (G•C base pair; Figure 9.4g) NOE cross-peaks (Figure 9.4d) verified the establishment of the four Watson–Crick base pairs (A5•T17, G6•C16, A7•T15, and G8•C14) constituting the stem-loop, which extends directly from the
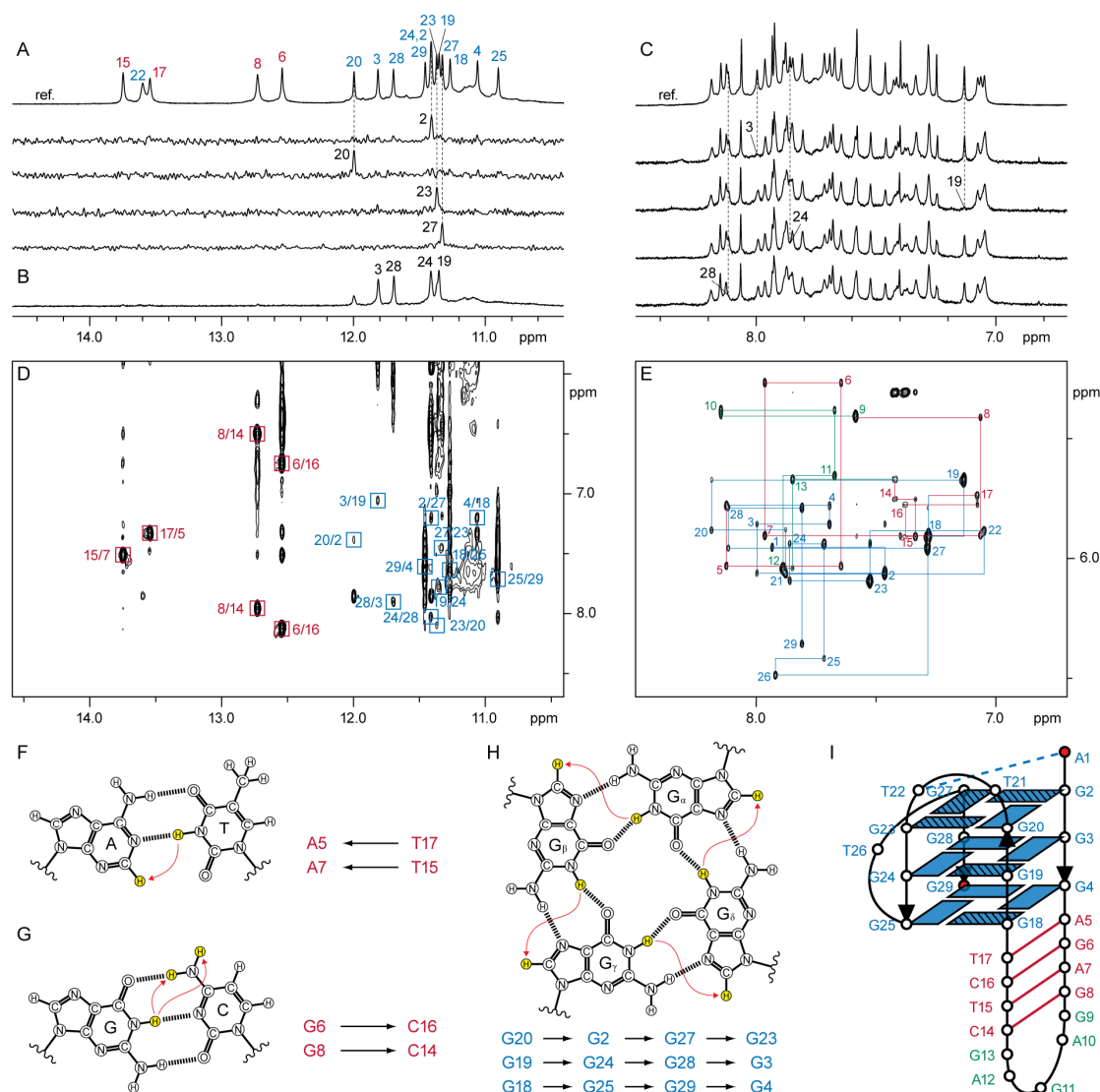
Figure 9.4 | NMR structural characterization of the MYT1L gene sequence *OTH01*. (a) Assignment of guanine H1 protons for the reference spectrum (ref.) is achieved through site-specific 2% ^{15}N-enrichment of samples at the indicated positions, one residue at a time, and their specific detection by ^{15}N-filtered experiments. **(b)** H1 proton spectra in ^{2}H$_2$O for 2 h at 25°C. **(c)** Assignment of guanine H8 protons for the reference spectrum (ref.) is achieved through site-specific ^{2}H substitution at the marked positions. **(d)** NOESY plot highlighting the characteristic cross-peaks corresponding to T(H3)–A(H2) for A•T Watson–Crick base pair **(f)** and G(H1)–C(H41)/G(H1)–C(H42) for G•C Watson–Crick base pair **(g)**, as well as guanine H1–H8 NOE connectivity **(h)** around each tetrad. **(e)** NOESY spectrum highlighting the H6/H8–H1' NOE connectivity of *OTH01* from A1 through G29. Residue numbers are marked on the corresponding intraresidue cross-peaks. **(i)** Schematic quadruplex–duplex hybrid structure of *OTH01*.

wide groove of the tetrad core (Figure 9.4i). Continuous stacking between the quadruplex and duplex segments was supported by NOE correlations between sugar and base protons of tetrad residues (G4 and G18) and the adjacent base pair step (A5 and T17). The duplex stem is closed off at the distal end by a five-nucleotide hairpin loop (GAGAG). The second, edgewise, loop (TT) of the quadruplex bends over a

153

narrow groove, while the last loop (T) traverses across a medium groove through the double-chain-reversal configuration. A non-canonical A1•T22 base pair capping the top of the tetrad core accounted for the H1 proton peak around 13.6 ppm. Overall, the quadruplex segment of *OTH01* bears identical loop arrangement as a number of (3+1) G-quadruplexes previously observed under various sequence contexts[37,167,233,244]. In particular, we note that *OTH01* closely resembles the structure of Form 1 hTERT promoter G-quadruplex[233], with the GAG loop of the latter replaced by the stem-loop in the present case.

SLC-PQS and PQS in the human genome. In this study, we explored the existence of SLC-PQS in the human genome, which would otherwise have been overlooked by conventional PQS algorithms exercising an arbitrary loop-length cut-off of 7-nt[41,42]. These motifs were found across various genomic loci of interest, and many of them do not overlap with PQS identified to date. Note that for practical considerations, here we limited the G-tract and initial loop length of SLC-PQS to 3–6 nt and 20 nt, respectively, while stable quadruplex–duplex hybrid complexes comprising only two G-tetrad layers (X=2), or containing longer stem-loops (L>20), may well exist. All in all, these SLC-PQS motifs complement the list of PQS[41,42] towards the compilation of a working set of genomic targets that could potentially be subjected to selective chemical interventions for biotechnological and therapeutic applications.

Biological and therapeutic implications of SLC-PQS. Out of 8,598 SLC-PQS identified in putative promoter, more than 56.4% (4,852) were mapped onto chromatin accessibility regions and transcription factor-binding sites. Some of these could exert a similar regulatory function in transcription as that posited for promoter PQS[46,47]. In addition, several SLC-PQS were identified at mutation and SNP loci, suggesting their possible associations with biological processes. Further experimental studies will be required to shed light on the biological significance, if any, of these motifs. Regardless, the localization of promoter SLC-PQS at open chromatin regions renders them attractive targets for drug design, especially those linked to cancer-associated/critical genes.

Targeting genomic quadruplex–duplex hybrids. Duplex-binding ligands have been extensively studied, and they range from intercalators to groove-binders to cross-linking agents. For instance, pyrrole–imidazole polyamides have been rationally

designed to recognize the minor grooves of DNA sequence-specifically[83]. On the other hand, quadruplex-binding ligands investigated thus far almost exclusively bind to G-quadruplexes through G-tetrad end-stacking[30,92], with a handful of groove-binders being discovered only recently[96,97]. These ligands typically exhibit excellent selectivity against duplexes, but they often show a lack of discrimination between different quadruplexes. Targeting of genomic quadruplex–duplex hybrid structures (adopted by SLC-PQS) would provide an opportunity to combine the best of both worlds; by coupling the specificity of duplex-binding ligands with the high affinity of quadruplex-binding ligands. Moreover, novel compounds could be designed to address structural features unique to these hybrid complexes, specifically at the quadruplex–duplex junctions. These could conceivably utilize similar recognition principle as that underlying the specific interactions between an RNA quadruplex–duplex hybrid and the RGG domain of the human fragile X mental retardation protein (FMRP)[266].

RNA quadruplex–duplex hybrids. Apart from their postulated role in transcriptional regulation, PQS have also been implicated in translational control through the formation of RNA G-quadruplexes[52,53]. Recently, it has been shown that non-coding RNAs are produced in a pervasive manner[299], and their biological functions remain poorly understood. It could be worthwhile to explore if RNA quadruplex–duplex hybrids reside within these transcripts. As illustrated by the complex folding topology adopted by the aptameric RNA quadruplex–duplex hybrid targeting the FMRP protein[266], folding principles governing their formations could deviate considerably from those that apply to the DNA counterparts. More of such structure will be needed before an algorithm can be developed to identify these motifs with any certainty.

Conclusion

Four-stranded G-quadruplex DNA have emerged as promising therapeutic targets in recent years, as their formations have been implicated in important genomic locations including telomeres and oncogenic promoters. Various algorithms have been implemented in the search for putative quadruplex sequences (PQS) in the human genome based on variants of the generic expression $G_{X1}N_{L1}G_{X2}N_{L2}G_{X3}N_{L3}G_{X4}$, wherein the loop length (L) was generally restricted to seven nucleotides (nt).

In this study, stem-loop–containing PQS (SLC-PQS) in the human genome were identified, with a particular focus on those located within promoter and genic regions. Many of these motifs do not overlap with PQS identified to date, representing an untapped pool of potential drug targets. Some of these were mapped to transcriptional active sites, cancer-associated genes, and mutation/single-nucleotide polymorphism (SNP) loci, suggesting possible biological involvements. A sampling of the sequence hits exhibited both NMR signatures of duplex and quadruplex elements, supporting the formation of quadruplex–duplex hybrids. The folding topology of one such motif, which adopts a (3+1) G-tetrad core topology with a stem-loop nested within the wide groove of the quadruplex, was validated by NMR. These quadruplex–duplex hybrids could open up new avenues for drug design efforts targeting non-canonical DNA structures.

Chapter 10

General Summary and Future Perspectives

The G-tetrad was discovered back in 1962, based upon the observation of gel formation by guanylic acid. It was not until the late 1980s that research on G-quadruplex structures started to gain traction, following a series of papers pointing to quadruplex formation under physiologically relevant conditions and the immediate implications of their potential involvement in biological processes. Quadruplex-forming motifs were subsequently identified in many genomic relevant loci, including telomeres, oncogenic promoters, introns, immunoglobulin class switch regions, rDNAs, minisatellites, and UTRs of mRNA. More recently, G-quadruplex formation was implicated in cellular processes including recombination and replication, further substantiating their potential role in biology. As such, it is vital to understand the possible structures that can be adopted by these guanine-rich motifs.

G-quadruplex exhibits a diverse range of structural forms. This arises from the multiple combinations of core and loop arrangement. The four strands constituting the G-tetrad core can be aligned in at least four distinct relative orientations. This is compounded by the many possible manners in which the corners of these strands can be joined by the loops, for which there are three general types. These result in an array of possible structures with widely varying shapes and dimensions. This structural polymorphism of G-quadruplex DNA perhaps is most aptly demonstrated by G-quadruplex structures of human telomeric repeats.

The human telomere is a protein–DNA complex that protects the telomere termini from being erroneously recognized as broken chromosome ends. Human telomeric DNA comprises many copies of the TTAGGG repeats. At the very end, it exists as a single-stranded TTAGGG overhang at the 3' terminal, ~100 to 200 nt long. Being G-rich, it has a high propensity to fold into G-quadruplexes. It was postulated that

formation of such telomeric G-quadruplexes could prohibit the recognition of the telomeric end by the enzyme telomerase, thereby leading to the down-regulation of its activity. This could serve as a promising anticancer strategy, as telomerase is found to be overactive in most cancer cells. Subsequent structural characterization of these repeats revealed their polymorphic nature, with at least four distinct folds discovered under various environments. The two human telomeric G-quadruplex structures presented in this study would build on this ensemble of structures which could serve as potential targets for ligand design. The stable structure of Form 3 revealed a new folding principle for G-rich sequences. By sacrificing the potential formation of a third tetrad layer, base pairing and stacking interactions in the loops are maximized, contributing to its superior stability with respect to the three-layered counterparts discovered in earlier studies. Hence the folding of G-rich sequences is driven not simply through the stacking of G-tetrad layers, but most probably involves the maximization of all possible interactions across the whole structure. On the other hand, the alternative (2+2) antiparallel folding topology discovered in Na^+ solution confirmed the polymorphic nature of these repeats beyond the influence of K^+ ions. This structure would serve as an additional reference for interpretation of studies conducted in Na^+ solution, for which only a single high-resolution structure was previously available.

Other areas of research on telomeric G-quadruplexes include variant sequence repeats and long telomeric repeats. An unusual G-quadruplex conformation stabilized through the establishment of up to two G•C•G•C tetrads was found to be favored by CTAGGG sequence variants, which could offer a possible explanation to the extreme instability associated with the occurrence of these repeat types in the telomere. On the other hand, a (3+1) G-quadruplex comprising a long propeller loop, which was shown to accommodate up to three TTAGGG repeats, was isolated from five- to seven-repeat human telomeric sequences. These long loops could potentially serve as recognition elements, either within the natural telomere context, or to be exploited for ligand design.

Given the immense conformational heterogeneity expressed by human telomeric repeats, it is becoming apparent that none of these structures should be regarded as the representative form. Rather, the diverse structural elements associated with them have provided ample opportunity for the design of chemical ligands that would selectively

158

recognize these motifs. The single-stranded overhang consists of up to 200 nt, and in principle multiple quadruplex units could simultaneously be formed and potentially cross-interact within the telomere context. Future research efforts will have to be directed towards the investigation of their potential interplay, which would be vital towards the understanding of telomere biology.

The targeting of promoter G-quadruplexes would represent another line of research that has been hotly pursued. Stabilization of c-myc promoter was shown to lead to the suppression of c-myc expression, opening up the possibility for the specific targeting of other oncogenic promoter G-quadruplexes for anticancer treatment. As opposed to telomeric G-quadruplexes, putative promoter quadruplexes identified to date largely comprise short loops of one or two nucleotides, which have been shown to display a tendency to adopt the propeller configuration. Consequently, many of the promoter quadruplexes investigated exhibit a propeller-type architecture. In this study, the interconversion between two quadruplex conformations that coexist in a G-rich hTERT promoter fragment was extensively investigated. Structural elucidation of the two forms, a (3+1) and a propeller-type G-quadruplexes, provided insights on how disparate quadruplex elements come together to stabilize distinct quadruplex conformations. To date, chemical ligands targeting quadruplexes mainly involve G-tetrad end stacking, and they typically exhibit poor discrimination between different quadruplex conformations. Perhaps the targeting of antiparallel G-quadruplexes would provide more asymmetric surfaces for their specific recognition.

In this study, the connectivity of duplex and quadruplex DNA was also established through the design and validation of five distinct quadruplex–duplex constructs. Each of these constructs demonstrates a different approach towards the attachment of a duplex stem onto a quadruplex core. The modularity of duplex–quadruplex connectivity was verified with the simultaneous incorporation of multiple duplex stems onto a single quadruplex scaffold to generate a G-junction motif. The thermal stability of these complexes was further investigated. It was found that incorporation of long duplex stems within the loops of G-quadruplexes leads to a general increase in stability, up to a point. Context at the junction is important towards the stabilization of these complexes, as well as the adoption of any desired construct. This understanding should facilitate the future design and prediction of such quadruplex–duplex hybrids. Akin to the double incorporation of the stacked X Holliday junction (double cross-

over motif) to generate immobile junction, G-junction, which has a stable quadruplex core, can potentially be incorporated in the design of DNA nanomaterials. For instance, G-junction motifs could be positioned at vertices of nanoconstructs, or employed as a universal "glue" for the connection of different structural elements.

The stable incorporation of long duplex stems in quadruplex loops pointed to the possibility of quadruplex-forming motifs comprising long intervening loop elements, which have not been explored in depth. An investigation on the existence of such motifs in the human genome was presented. These additional motifs would add onto the available pool of putative quadruplex sequences that could be useful towards biotechnological and therapeutic applications. Knowledge on duplex–quadruplex compatibility could also lead to alternative targeting strategies steered towards genomic G-rich sequences.

This study has presented a diverse range of G-quadruplex structures that can be adopted by guanine-rich sequences under various contexts, ranging from biologically relevant telomeric and promoter G-quadruplexes, to quadruplex–duplex hybrids that could provide nanotechnological and therapeutic applications. There still remains a need for the structural elucidation of more G-quadruplexes, in order to achieve a greater understanding on the folding principles governing these complex DNA structural forms, and hopefully the eventual control over these structures.

Statistics of Computed Structures

Statistics of the computed structures for 22-nt [Br]G7-Form 3 human telomeric sequence d[GGGTTA([Br]G)GGTTAGGGTTAGGGT] in K[+] solution.[a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	269	3
sequential (i, $i + 1$) distance restraints	199	20
long-range (i, $\geq i + 2$) distance restraints	44	45
other restraints		
hydrogen-bond restraints	42	
dihedral restraints	20	
planarity restraints	4	
repulsive restraints [b]	5	
intensity restraints		
non-exchangeable protons (each of four mixing times)	140	
B. Structure statistics for 10 molecules following intensity refinement		
NOE violations		
number [c] (>0.2 Å)	0.400 ± 0.490	
maximum violation (Å)	0.193 ± 0.029	
rmsd of violations (Å)	0.018 ± 0.002	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.005 ± 0.000	
bond angles (deg)	0.752 ± 0.028	
impropers (deg)	0.433 ± 0.015	
NMR R factor ($R_{1/6}$)	0.016 ± 0.001	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except T4, T5, T10, T11, A12, T16, T17, T22	1.06 ± 0.13	
all heavy atoms	1.54 ± 0.21	

[a] PDB code: 2KF7. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks; these restraints are removed during relaxation matrix intensity refinement. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for the 27-nt ^{Br}G-modified *4TTA16[Br22]* human telomeric sequence d[(TTAGGG)$_3$TTA(^{Br}G)GGTTA] in Na$^+$ solution.[a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	446	0
sequential (*i*, *i* + 1) distance restraints	197	26
long-range (*i*, ≥ *i* + 2) distance restraints	79	32
other restraints		
hydrogen-bond restraints	48	
dihedral restraints	12	
planarity restraints	3	
repulsive restraints [b]	2	
intensity restraints		
non-exchangeable protons (each of four mixing times)	140	
B. Structure statistics for 10 molecules following intensity refinement		
NOE violations		
number [c] (>0.2 Å)	0.800 ± 0.872	
maximum violation (Å)	0.195 ± 0.050	
rmsd of violations (Å)	0.022 ± 0.002	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.694 ± 0.007	
impropers (deg)	0.367 ± 0.004	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except T1, T2	1.04 ± 0.12	
all heavy atoms	1.49 ± 0.35	

[a] PDB code: 2MBJ. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks; these restraints are removed during relaxation matrix intensity refinement. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for the 22-nt *22CTA* variant human telomeric sequence d[AGGG(CTAGGG)$_3$] in K$^+$ solution.[a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	203	5
sequential (i, i + 1) distance restraints	179	29
long-range (i, $\geq i$ + 2) distance restraints	34	72
other restraints		
hydrogen-bond restraints	50	
dihedral restraints	34	
planarity restraints	5	
repulsive restraints [b]	6	
intensity restraints		
non-exchangeable protons (each of four mixing times)	108	
B. Structure statistics for 10 molecules following intensity refinement		
NOE violations		
number [c] (>0.2 Å)	0.300 ± 0.458	
maximum violation (Å)	0.168 ± 0.031	
rmsd of violations (Å)	0.019 ± 0.002	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.004 ± 0.000	
bond angles (deg)	0.746 ± 0.010	
impropers (deg)	0.413 ± 0.010	
NMR *R* factor (R$_{1/6}$)	0.016 ± 0.001	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except A1, G2	1.21 ± 0.14	
all heavy atoms	1.53 ± 0.25	

[a] PDB code: 2KM3. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks; these restraints are removed during relaxation matrix intensity refinement. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for the 20-nt hTERT promoter modified sequence d[AGGGIAGGGGCTGGGAGGGC] (Form 1) in K⁺ solution.[a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	329	0
sequential (i, $i + 1$) distance restraints	106	24
long-range (i, $\geq i + 2$) distance restraints	14	35
other restraints		
hydrogen-bond restraints	52	
dihedral restraints	32	
planarity restraints	13	
repulsive restraints [b]		
B. Structure statistics for 10 molecules following distance-restrained molecular dynamics refinement		
NOE violations		
number [c] (>0.2 Å)	0	
maximum violation (Å)	0.140 ± 0.026	
rmsd of violations (Å)	0.014 ± 0.000	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.692 ± 0.008	
impropers (deg)	0.362 ± 0.005	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except I5, A6, G7, A16, C20	0.64 ± 0.12	
all heavy atoms	1.24 ± 0.23	

[a] PDB code: 2KZD. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for the 20-nt hTERT promoter modified sequence d[AIGGGAGGGICTGGGAGGGC] (Form 2) in K$^+$ solution.[a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	310	0
sequential (i, i + 1) distance restraints	99	4
long-range (i, $\geq i$ + 2) distance restraints	10	47
other restraints		
hydrogen-bond restraints	48	
dihedral restraints	32	
planarity restraints	11	
repulsive restraints [b]		
B. Structure statistics for 10 molecules following distance-restrained molecular dynamics refinement		
NOE violations		
number [c] (>0.2 Å)	0	
maximum violation (Å)	0.127 ± 0.016	
rmsd of violations (Å)	0.015 ± 0.001	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.674 ± 0.008	
impropers (deg)	0.345 ± 0.007	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except A6, I10, C11, T12, A16, C20	0.95 ± 0.15	
all heavy atoms	1.54 ± 0.23	

[a] PDB code: 2KZE. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for the reference duplex hairpin.[a]

	D₂O	H₂O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	350	28
sequential ($i, i + 1$) distance restraints	213	29
long-range ($i, \geq i + 2$) distance restraints	15	46
other restraints		
hydrogen-bond restraints	32	
dihedral restraints	52	
planarity restraints	6	
repulsive restraints[b]	5	
B. Structure statistics for 10 molecules following distance-restrained molecular dynamics refinement		
NOE violations		
number[c] **(>0.2 Å)**	0.100 ± 0.300	
maximum violation (Å)	0.173 ± 0.035	
rmsd of violations (Å)	0.019 ± 0.001	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.705 ± 0.006	
impropers (deg)	0.354 ± 0.004	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except C1, G2, C14, G15	0.67 ± 0.13	
all heavy atoms	1.00 ± 0.39	

[a] PDB code: 2M8Y. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for quadruplex–duplex Construct I. [a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	544	34
sequential (*i*, *i* + 1) distance restraints	393	39
long-range (*i*, ≥ *i* + 2) distance restraints	59	51
other restraints		
hydrogen-bond restraints	64	
dihedral restraints	68	
planarity restraints	8	
repulsive restraints [b]	2	
B. Structure statistics for 10 molecules following distance-restrained molecular dynamics refinement		
NOE violations		
number [c] (>0.2 Å)	0.200 ± 0.400	
maximum violation (Å)	0.181 ± 0.026	
rmsd of violations (Å)	0.018 ± 0.001	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.690 ± 0.007	
impropers (deg)	0.354 ± 0.003	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms from quadruplex segment	0.77 ± 0.17	
all heavy atoms from hairpin segment	1.01 ± 0.29	
all heavy atoms	1.49 ± 0.54	

[a] PDB code: 2M8Z. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for quadruplex–duplex Construct II.[a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	600	29
sequential (*i, i* + 1) distance restraints	307	37
long-range (*i, ≥ i* + 2) distance restraints	49	106
other restraints		
hydrogen-bond restraints	82	
dihedral restraints	72	
planarity restraints	9	
repulsive restraints [b]	1	
B. Structure statistics for 10 molecules following distance-restrained molecular dynamics refinement		
NOE violations		
number [c] **(>0.2 Å)**	0	
maximum violation (Å)	0.168 ± 0.010	
rmsd of violations (Å)	0.020 ± 0.001	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.691 ± 0.005	
impropers (deg)	0.339 ± 0.004	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms from quadruplex segment except G27, A28, A29	1.12 ± 0.36	
all heavy atoms from hairpin segment	1.25 ± 0.48	
all heavy atoms	2.67 ± 0.72	

[a] PDB code: 2M90. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for quadruplex–duplex Construct III. [a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	583	34
sequential (i, i + 1) distance restraints	401	44
long-range (i, ≥ i + 2) distance restraints	44	71
other restraints		
hydrogen-bond restraints	72	
dihedral restraints	70	
planarity restraints	10	
repulsive restraints [b]	4	
B. Structure statistics for 10 molecules following distance-restrained molecular dynamics refinement		
NOE violations		
number [c] (>0.2 Å)	0.200 ± 0.400	
maximum violation (Å)	0.192 ± 0.048	
rmsd of violations (Å)	0.021 ± 0.001	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.688 ± 0.008	
impropers (deg)	0.347 ± 0.006	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms from quadruplex segment with G8, A24	0.89 ± 0.26	
all heavy atoms from hairpin segment	0.94 ± 0.18	
all heavy atoms	1.37 ± 0.26	

[a] PDB code: 2M91. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for quadruplex–duplex Construct IV.[a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	494	16
sequential ($i, i + 1$) distance restraints	278	31
long-range ($i, \geq i + 2$) distance restraints	38	97
other restraints		
hydrogen-bond restraints	88	
dihedral restraints	64	
planarity restraints	11	
repulsive restraints [b]	2	
B. Structure statistics for 10 molecules following distance-restrained molecular dynamics refinement		
NOE violations		
number [c] **(>0.2 Å)**	1.000 ± 0.632	
maximum violation (Å)	0.250 ± 0.057	
rmsd of violations (Å)	0.023 ± 0.002	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.688 ± 0.009	
impropers (deg)	0.348 ± 0.005	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms from quadruplex segment with G16, A32	1.28 ± 0.25	
all heavy atoms from hairpin segment	1.01 ± 0.21	
all heavy atoms	1.67 ± 0.33	

[a] PDB code: 2M92. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks. [c] The total number of violations divided by the number of structures.

Statistics of the computed structures for quadruplex–duplex Construct V. [a]

	D$_2$O	H$_2$O
A. NMR restraints		
distance restraints		
intraresidue distance restraints	574	15
sequential (*i, i* + 1) distance restraints	317	27
long-range (*i,* ≥ *i* + 2) distance restraints	21	89
other restraints		
hydrogen-bond restraints	50	
dihedral restraints	74	
planarity restraints	8	
repulsive restraints [b]	5	
B. Structure statistics for 10 molecules following distance-restrained molecular dynamics refinement		
NOE violations		
number [c] (>0.2 Å)	1.200 ± 1.166	
maximum violation (Å)	0.215 ± 0.031	
rmsd of violations (Å)	0.020 ± 0.001	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.684 ± 0.006	
impropers (deg)	0.348 ± 0.005	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms from quadruplex segment	1.08 ± 0.18	
all heavy atoms from hairpin segment	1.60 ± 0.45	
all heavy atoms	*2.49*	0.81

[a] PDB code: 2M93. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks. [c] The total number of violations divided by the number of structures.

173

Experimental Protocols

3.2.1   A Two-Tetrad Basket-Type G-Quadruplex Formed in Potassium Solution

Sample preparation. Unlabeled and site-specific low-enrichment (2% ^{15}N-labeled or ^{15}N,^{13}C-labeled) DNA oligonucleotides were chemically prepared using products from Glen Research, Spectra Gases and Cambridge Isotope Laboratories. Samples were dialyzed successively against ~50 mM KCl solution and against water. DNA concentration was expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species[300]. The same extinction coefficient was used for the natural and the ^{Br}G-substituted oligonucleotides. Unless otherwise stated, the strand concentration of the NMR samples was typically 0.5–2.5 mM; the solutions contained 70 mM KCl and 20mM potassium phosphate (pH 7).

UV-melting experiments. The thermal stability of different DNA oligonucleotides was characterized by recording the UV absorbance at 295 nm[127] as a function of temperature (20 to 90 °C) using a Varian Cary 300 Bio UV-Vis spectrophotometer. The heating and cooling rates were 0.25 ºC per minute. Two baselines corresponding to the completely folded (low temperature) and completely unfolded (high temperature) states were manually drawn in order to determine the fractions of folded and unfolded species during the melting/folding transition. The melting temperature (T_m) was defined as the temperature of the mid-transition point. Experiments were performed with quartz cuvettes, 1-cm path-length for low DNA concentrations and 0.2-cm path-length for high DNA concentrations.

Circular dichroism. CD spectra were recorded at 20 °C on a JASCO-810 spectropolarimeter using a 1-cm path length quartz cuvette in a reaction volume of 800 µl. The DNA oligonucleotides (8 µM) were prepared in 20 mM potassium phosphate buffer (pH 7.0) containing 70 mM KCl. The samples were heated at 95 °C for 5 minutes and cooled down to room temperature overnight. For each sample, an average of three scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm. The spectra were finally normalized to the concentration of the DNA samples.

NMR spectroscopy. Experiments were performed on 600 MHz and 700 MHz Varian and Bruker spectrometers at 25°C, unless otherwise specified. Resonances for guanine residues were assigned unambiguously by using site-specific low-enrichment

labeling[192] and through-bond correlations at natural abundance[193]. Imino protons of thymine residues were assigned unambiguously by using site-specific low-enrichment labeling[192]. Other resonances for thymine residues were assigned following systematic T-to-U substitutions[212]. Spectral assignments were completed by NOESY, COSY, TOCSY, and HSQC, as described previously[212]. All spectral analyses were performed using the FELIX (Felix NMR, Inc.) program.

Structure Calculations. Inter-proton distances for ^{Br}G7-Form 3 were deduced from NOESY experiments performed in H_2O (mixing time, 200 ms) and D_2O (mixing times, 100, 150, 200, and 300 ms). Structure computations were performed using the XPLOR-NIH program[128] in three general steps: (i) distance geometry simulated annealing, (ii) distance-restrained molecular dynamics refinement, and (iii) relaxation matrix intensity refinement. Hydrogen bond restraints, inter-proton distance restraints, dihedral restraints, as well as planarity restraints were imposed during structure calculations.

Hydrogen Bond Restraints. Each hydrogen bond was restrained by two distances (donor atom to acceptor atom and proton to acceptor atom) which correspond to ideal hydrogen bond geometry. The force constant for hydrogen bond restraints was kept at 100 $kcal.mol^{-1}.Å^{-2}$ throughout the computation.

Non-exchangeable Proton Distance Restraints. Distances between non-exchangeable protons were deduced from NOESY spectra in D_2O (mixing times, 100, 150, 200, and 300 ms) and implemented as distance restraints during structure calculations. For well-resolved cross peaks that are strong or medium at 150-ms mixing time, the initial slopes of the buildup curves were obtained by performing a second-order fit on the volumes measured at 100- and 150-ms mixing times. The slopes were then converted to inter-proton distances using the thymine H6-methyl distance of 2.99 Å as reference, and given bounds of ±30%. All other well-resolved cross peaks were converted to inter-proton distance constraints with the lower limit of 4.0 Å and the upper limit of 8.0 Å. Cross peaks that are not well-resolved were manually classified as strong or weak, and assigned to inter-proton distances of 2.0–4.5 Å or 3.8–7.5Å, respectively. Heavily overlapped cross peaks were given the distances of 2.0 to 7.5 Å. Cross peaks involving a methyl group were interpreted as distances to the methyl carbon. Necessary corrections were made to the bounds for these peaks to account for the radius of the methyl group (i.e. the bounds for well-

resolved peaks were increased to ±40%; the upper bounds for manually-treated peaks were increased by 0.5 Å).

Exchangeable Proton Distance Restraints. Cross peaks involving exchangeable protons were obtained from the NOESY spectrum in H_2O at 200-ms mixing time. The peaks were manually classified as strong (2.2–4.0 Å), medium (2.5–5.0 Å) or weak (3.5–6.5 Å).

NMR Dihedral Restraints. The glycosidic χ torsion angle for experimentally-determined *syn* guanine residues G1, G7, G14, and G19 were restrained to (60±30)°, while that for guanine residues G2, G8, G15, and G20 were fixed at (240±70)°, characteristic of *anti* conformation. The ε dihedral angles between residue pairs G1-G2, G7-G8, G14-G15, and G19-G20 were set to be (225±75)°, consistent with the stereochemically allowed range for this angle.

Planarity Restraints. Planarity restraints were enforced on the base atoms for G1•G14•G20•G8 and G2•G7•G19•G15 tetrads, G21•G9•G13 and A6•G3•A18 triples, and T11•T22 and T5•T17. Formation of base triples and base pairs was suggested by prior NOE-restrained computations performed without base triple/pair planarity constraints.

Distance Geometry Simulated Annealing. An extended conformation of the $^{Br}G7$-Form 3 sequence with ideal geometry was generated randomly using XPLOR-NIH. The extended DNA was then subjected to distance geometry simulated annealing (DGSA) by supplying the full set of hydrogen-bond, distance, dihedral, and planarity restraints. 100 structures were generated and subjected to further refinements.

Distance-restrained Molecular Dynamics Refinement. All 100 DGSA structures were refined with distance-restrained molecular dynamics. The system was heated from 300 K to 1000 K in 5 ps and allowed to equilibrate for 1 ps, wherein the force constants for the distance restraints were kept at 2 kcal.mol^{-1}.Å^{-2}. The force constants for non-exchangeable and exchangeable proton distance restraints were then scaled to final values of 50 and 40 kcal.mol^{-1}.Å^{-2}, respectively, in a 26-ps interval. Subsequently, the system was slowly cooled down to 300 K in 14 ps, after which equilibration was performed for another 10 ps. Coordinates of the molecule were saved every 0.5 ps during the last 4.0 ps and averaged. In the final step, the average structure was subjected to minimization until the gradient of energy was less than 0.1 kcal.mol^{-1}.

176

Dihedral (200 kcal.mol^{-1}.rad^{-2}) and planarity (2 kcal.mol^{-1}.Å^{-2} for tetrads, 1 kcal.mol^{-1}.Å^{-2} for triples and T•T base pairs) restraints were maintained throughout the course of refinement. 10 best structures were selected, based on the overall energy term, to undergo relaxation matrix intensity refinement.

Relaxation Matrix Intensity Refinement. Relaxation matrix intensity refinement was next performed to account for spin diffusion effects. The relaxation matrix was set up for the non-exchangeable protons. NOE intensity volumes of 191 cross peaks for each of four mixing times (100, 150, 200, and 300 ms) were employed. The exchangeable imino and amino protons were all replaced by deuterons. Heating was done from 5 K to 300 K. The force constant for NOE intensities was increased from 2 to 300 kcal.mol^{-1}.Å^{-2}, while the force constant for non-exchangeable proton distance restraints were decreased from 50 to 10 kcal.mol^{-1}.Å^{-2}. The force constant for exchangeable proton distance restraints was maintained at 40 kcal.mol^{-1}.Å^{-2}. After that, the system was equilibrated at 300 K for 3 ps and energy-minimized. The planarity restraints were set to very low values of 1 kcal.mol^{-1}.Å^{-2} for the tetrads and 0.5 kcal.mol^{-1}.Å^{-2} for the triples and T•T base pairs.

3.2.2 An Antiparallel (2+2) G-quadruplex Scaffold Formed in Sodium Solution

Sample preparation. Unlabeled and site-specific labeled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer using products from Glen Research, Spectra Gases, and Cambridge Isotope Laboratories. The oligonucleotides were de-protected following the manufacturer's protocols and purified using Poly-Pak^{TM} cartridges. Samples were dialyzed successively against water, ~25 mM NaCl solution, and water again. They were subsequently frozen, lyophilized, and suspended in a buffer containing 20 mM sodium phosphate (pH 7.0) and 70 mM NaCl. DNA concentration is expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species[300]. The same extinction coefficient was used for the natural and ^{Br}G-substituted oligonucleotides.

Ultraviolet spectroscopy. Thermal stability of different DNA oligonucleotides was characterized by recording the ultraviolet (UV) absorbance at 295 nm[127] as a function of temperature (20–90 °C) using a JASCO V-650 spectrophotometer. The heating and

cooling rates were 0.2 ºC per minute. Two baselines corresponding to the completely folded (low temperature) and completely unfolded (high temperature) states were manually drawn in order to determine the fractions of folded and unfolded species during the melting/folding transition. The melting temperature (T_m) is defined as the temperature for which there are equal fractions of folded and unfolded species. Experiments were performed with quartz cuvettes (1-cm path length for low DNA concentrations and 0.2-cm path length for high DNA concentrations).

Circular dichroism spectroscopy. Circular dichroism (CD) spectra were recorded at 20 °C on a JASCO-810 spectropolarimeter over the range of 220–320 nm using a 1-cm path length quartz cuvette with a reaction volume of 600 µl. For each sample, an average of 3 scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm.

NMR spectroscopy. Strand concentration of the NMR samples was typically 0.5–2.0 mM. NMR experiments were performed on Bruker AVANCE 600- and 700-MHz spectrometers at 25 °C, unless otherwise specified. Resonances for guanine residues were assigned unambiguously using site-specific low-enrichment ^{15}N labeling[192], site-specific ^{2}H labeling[207], and through-bond correlations at natural abundance[193,212]. Spectral assignments were assisted by NOESY, COSY, TOCSY, and ^{13}C–^{1}H-HSQC experiments. The spectra were processed with the software TopSpin[TM] and analyzed using the program FELIX (Felix NMR, Inc.).

Structure calculation. Inter-proton distances for the d[(TTAGGG)$_3$ TTA([Br]G)GGTTA] (*htel27[Br22]*) G-quadruplex were deduced from NOESY experiments performed in H_2O (mixing time, 200 ms) and D_2O (mixing times, 100 and 350 ms). Structures were calculated based on distance-restrained molecular dynamics refinement following distance geometry simulated annealing using the program XPLOR-NIH[128]. Hydrogen bond restraints, inter-proton distance restraints, dihedral restraints, planarity restraints, and repulsive restraints were imposed during structure calculations. Structures were displayed using the program PyMOL[301].

Hydrogen-Bond Restraints. Each hydrogen bond was restrained by two distances (donor atom to acceptor atom and proton to acceptor atom) which correspond to ideal hydrogen bond geometry. The force constant for hydrogen bond restraints was kept at 32 kcal.mol^{-1}.Å^{-2} throughout the computation.

Non-Exchangeable Proton Distance Restraints. Distances between non-exchangeable protons were deduced from NOESY spectra in D_2O (mixing times, 100 and 350 ms) and implemented as distance restraints during structure calculations. These interproton distance restraints were manually classified as very strong (1.80–3.00 Å), strong (2.25–3.75 Å), medium (2.85–4.75 Å), weak (3.60–6.00 Å), or very weak (4.50–7.50 Å). Heavily overlapped cross-peaks were given more permissible bounds of 1.80–4.60 Å or 3.50–7.50 Å. Cross-peaks involving a methyl group were interpreted as distances to the methyl carbon. The upper bounds for these peaks were increased by 0.5 Å to account for the radius of the methyl group. The upper bounds for cross-peaks involving ambiguous atom selection were also increased by 0.5 Å.

Exchangeable Proton Distance Restraints. Cross-peaks involving exchangeable protons were obtained from NOESY spectrum in H_2O at 200-ms mixing time. The peaks were manually classified as very strong (1.8–3.2 Å), strong (2.4–4.0 Å), medium (3.0–5.4 Å), weak (3.6–6.4 Å), or very weak (4.2–7.8 Å). Heavily overlapped cross-peaks were given more permissible bounds of 1.80–4.50 Å or 3.80–7.80 Å. Cross-peaks involving a methyl group were interpreted as distances to the methyl carbon. The upper bounds for these peaks were increased by 0.5 Å to account for the radius of the methyl group. The upper bounds for cross-peaks involving ambiguous atom selection were also increased by 0.5 Å.

Dihedral Restraints. The glycosidic χ torsion angle for experimentally-determined *syn* guanine residues were restrained to 60°, while that for *anti* guanine residues were fixed at 240°. The bounds for these dihedrals were set at ±40° (central tetrad) or ±70° (outer tetrads).

Planarity Restraints. Planarity restraints were enforced on the base atoms for the tetrads.

Repulsive Restraints. Repulsive restraints (4.2–80 Å) were applied on pairs of protons that do not exhibit cross-peaks in NOESY.

Distance Geometry Simulated Annealing. An extended conformation of the sequence d[(TTAGGG)₃TTA(ᴮʳG)GGTTA] with ideal geometry was generated randomly using XPLOR-NIH[128]. The extended DNA was then subjected to distance geometry simulated annealing (DGSA) by supplying the full set of hydrogen-bond,

distance, dihedral, planarity and repulsive restraints. 100 structures were generated and subjected to further refinements.

Distance-Restrained Molecular Dynamics Refinement. All 100 DGSA structures were refined with distance-restrained molecular dynamics. The system was heated from 300 K to 1000 K in 14 ps and allowed to equilibrate for 6 ps, wherein the force constants for the distance restraints were kept at 2 kcal.mol^{-1}.Å^{-2}. The force constants for non-exchangeable proton, exchangeable proton, and repulsive distance restraints were then scaled to final values of 16, 8, and 8 kcal.mol^{-1}.Å^{-2}, respectively, in a 20-ps interval and allowed to equilibrate for a further 200 ps. Subsequently, the system was slowly cooled down to 300 K in 42 ps, after which equilibration was performed for another 18 ps. Coordinates of the molecule were saved every 0.5 ps during the last 4 ps and averaged. In the final step, the average structure was subjected to minimization until the gradient of energy was less than 0.1 kcal.mol-1. Dihedral (50 kcal.mol^{-1}.rad^{-2}) and planarity (2 kcal.mol^{-1}.Å^{-2} for the tetrads) restraints were maintained throughout the course of refinement. 10 best structures with the lowest overall energy were selected. Structures were displayed using the program PyMOL[301].

4.2 Sequence Variant d[CTAGGG]$_n$] Prefers a G-Quadruplex Structure Comprising a G•C•G•C Tetrad

DNA sample preparation. Unlabeled and site-specific labeled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer or purchased from Eurogentec (Belgium). DNA concentration was expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species.

Gel electrophoresis. The molecular size of the structures formed by DNA oligonucleotides was probed by non-denaturing polyacrylamide gel electrophoresis (PAGE). DNA samples (^{32}P-labeled) of three different concentrations (0.05, 4, and 80 μM) were incubated in a 10 mM Tris-HCl pH 7.5 buffer supplemented with 100 mM NaCl or KCl. They were equilibrated at 90°C for 5 minutes and slowly cooled down (2 h) to 20°C. The samples were loaded on a 15% polyacrylamide gel supplemented with 20 mM of the corresponding salt and run at 26°C; 10% sucrose was added just before loading.

UV-melting experiments. The thermal stability of different oligonucleotides was characterized in heating/cooling experiments by recording the UV absorbance at 295 nm as a function of temperature using a Kontron-Uvikon 940 UV/Vis spectrophotometer. UV-melting experiments were conducted in a 10 mM lithium cacodylate pH 7.2 buffer containing 100 mM NaCl or KCl. The heating and cooling rates were 0.2°C per minute. Experiments were performed with 1-cm path length quartz cuvettes.

Circular dichroism. CD spectra were recorded on a JASCO-810 spectropolarimeter using a 1-cm path length quartz cuvette with a reaction volume of 600 μl. The DNA oligonucleotides (4 μM) were prepared in a 10 mM lithium cacodylate pH 7.2 buffer containing 100 mM NaCl or KCl. The samples were heated at 90°C for 5 minutes and cooled down to room temperature overnight. For each sample, an average of three scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm.

NMR spectroscopy. Samples for NMR study were dialyzed successively against ~50 mM KCl solution and against water. Unless otherwise stated, the strand concentration of the NMR samples was typically 0.5-2.0 mM; the solutions contained 70 mM KCl and 20 mM potassium phosphate (pH 7). NMR experiments were performed on 600 MHz and 700 MHz Bruker spectrometers at 25°C, unless otherwise specified. Resonances for guanine residues were assigned unambiguously by using site-specific low-enrichment ^{15}N labeling, site-specific ^{2}H labeling, and through-bond correlations at natural abundance. Resonances for thymine residues were assigned following systematic T-to-U substitutions. Spectral assignments were completed by analysis of NOESY, COSY, TOCSY, and ^{13}C–^{1}H-HSQC. Inter-proton distances were deduced from NOESY experiments at various mixing times. All spectral analyses were performed using the FELIX (Felix NMR, Inc.) program.

Structure calculation. Inter-proton distances for the d[AGGG(CTAGGG)$_3$] quadruplex were deduced from NOESY experiments performed in H$_2$O (mixing time, 200 ms) and D$_2$O (mixing times, 100, 200, and 350 ms). Structure computations were performed using the XPLOR-NIH program in three general steps: (i) distance geometry simulated annealing, (ii) distance-restrained molecular dynamics refinement, and (iii) relaxation matrix intensity refinement. Hydrogen bond restraints, inter-proton distance restraints, dihedral restraints, planarity restraints, and repulsive

restraints were imposed during structure calculations. Structures were displayed using the PyMOL program.

Hydrogen Bond Restraints. Each hydrogen bond was restrained by two distances (donor atom to acceptor atom and proton to acceptor atom) which correspond to ideal hydrogen bond geometry. The force constant for hydrogen bond restraints was kept at 100 kcal.mol^{-1}.Å^{-2} throughout the computation.

Non-exchangeable Proton Distance Restraints. Distances between non-exchangeable protons were deduced from NOESY spectra in D_2O (mixing times, 100, 200, and 350 ms) and implemented as distance restraints during structure calculations. For well-resolved cross-peaks, the initial slopes of the buildup curves were obtained by performing a second-order fit on the volumes measured at 100- and 200-ms mixing times. The slopes were then converted to inter-proton distances using the cytosine H5-H6 distance of 2.4 Å as reference, and given bounds of 20% or 30%, depending on the resolution of the peak. Cross-peaks that are not well-resolved were manually classified as strong or weak, and assigned to inter-proton distances of 1.8–4.2 Å or 3.6–7.8Å, respectively. Heavily overlapped cross-peaks were given the distances of 1.8 to 7.8 Å. Cross-peaks involving a methyl group were interpreted as distances to the methyl carbon. The upper bounds for these peaks were increased by 0.5 Å to account for the radius of the methyl group.

Exchangeable Proton Distance Restraints. Cross-peaks involving exchangeable protons were obtained from the NOESY spectrum in H_2O at 200-ms mixing time and classified manually.

NMR Dihedral Restraints. The glycosidic χ torsion angle for experimentally-determined *syn* guanine residues G3, G9, G15, and G21 were restrained to (60±30)°, while that for residues G4, C5, G8, G10, G16, C17, G20, and G22 were fixed at (240±40)°, characteristic of *anti* conformation. In addition, the H1'-C1'-N9/N1-C4/C2 torsion angle for G3, G9, G15, and G21 were set to (180±30)°, whereas that for G4, C5, G8, G10, G16, C17, G20, and G22 were set to (0±40)°. The ε dihedral angles between residue pairs G3-G4, G4-C5, G8-G9, G9-G10, G15-G16, G16-C17, G20-G21, and G21-G22 were set to be (225±75)°, consistent with the stereochemically allowed range for this angle. Prior NOE-restrained computations produced structures in which G2 adopted a near-*syn* conformation, in disagreement with the intensity of the

182

intraresidue H8-H1' cross-peak for this residue. As such, the glycosidic χ torsion angle and the H1'-C1'-N9-C4 angle of G2 were restrained to $(240\pm70)°$ and $(0\pm70)°$, respectively, for all subsequent calculations.

Planarity Restraints. Planarity restraints were enforced on the base atoms for the G3•G10•G15•G22 and G4•G21•G16•G9 tetrads, and the G8•C17, G20•C5, and G14•C11 base-pairs.

Repulsive Restraints. Repulsive restraints (4.5 - 80 Å) were applied on pairs of protons that do not exhibit cross-peaks in NOESY.

Distance Geometry Simulated Annealing. An extended conformation of the d[AGGG(CTAGGG)$_3$] sequence with ideal geometry was generated randomly using XPLOR-NIH. The extended DNA was then subjected to distance geometry simulated annealing (DGSA) by supplying the full set of hydrogen bond, distance, dihedral, planarity, and repulsive restraints. 100 structures were generated and subjected to further refinements.

Distance-restrained Molecular Dynamics Refinement. All 100 DGSA structures were refined with distance-restrained molecular dynamics. The system was heated from 300 K to 1000 K in 5 ps and allowed to equilibrate for 1 ps, wherein the force constants for the distance restraints were kept at 2 kcal.mol^{-1}.Å^{-2}. The force constants for non-exchangeable and exchangeable proton distance restraints were then scaled to final values of 50 and 40 kcal.mol^{-1}.Å^{-2}, respectively, in a 26-ps interval. Subsequently, the system was slowly cooled down to 300 K in 14 ps, after which equilibration was performed for another 10 ps. Coordinates of the molecule were saved every 0.5 ps during the last 4.0 ps and averaged. In the final step, the average structure was subjected to minimization until the gradient of energy was less than 0.1 kcal.mol^{-1}. Dihedral (200 kcal.mol^{-1}.rad^{-2}) and planarity (5 kcal.mol^{-1}.Å^{-2} for both tetrads and G•C base-pairs) restraints were maintained throughout the course of refinement. 10 best structures were selected, based on the overall energy term, to undergo relaxation matrix intensity refinement.

Relaxation Matrix Intensity Refinement. Relaxation matrix intensity refinement was next performed to account for spin diffusion effects. The relaxation matrix was set up for the non-exchangeable protons. NOE intensity volumes of 108 cross-peaks for each of three mixing times (100, 200, and 350 ms) were employed. The exchangeable

imino and amino protons were all replaced by deuterons. Heating was done from 5 K to 300 K. The force constant for NOE intensities was increased from 2 to 300 $kcal.mol^{-1}.Å^{-2}$, while the force constant for non-exchangeable proton distance restraints were decreased from 50 to 30 $kcal.mol^{-1}.Å^{-2}$. The force constant for exchangeable proton distance restraints was maintained at 40 $kcal.mol^{-1}.Å^{-2}$. After that, the system was equilibrated at 300 K for 3 ps and energy-minimized. The planarity restraints were set to very low values of 1 $kcal.mol^{-1}.Å^{-2}$ for the tetrads and 0.5 $kcal.mol^{-1}.Å^{-2}$ for the G•C base-pairs.

5.2   Human Telomeric G-Quadruplexes Harboring a Long Loop

DNA sample preparation. Unlabeled and site-specific modified DNA oligonucleotides were chemically synthesized using an automated ABI 394 DNA/RNA synthesizer. Upon the completion of the syntheses, the samples were de-protected and purified using Poly-Pak cartridge (Glen research) with respect to the manufacturer's protocol. The DNA oligonucleotides were then dialyzed against water, ~25 mM of KCl solution and water successively, followed by lyophilization to obtain pure DNA in powder form.

UV Experiments. The concentrations of DNA were quantified by measuring the UV absorbance at λ_{max} (~256 nm) at 90 °C on a Jasco V-650 spectrophotometer (scan speed: 400 nm/min. data pitch: 1.0 nm, bandwidth: 2.0 nm). For each sample, an average of three scans was taken with the buffer spectra subtracted, and the data were zero-corrected at 320 nm. The absolute concentration was then calculated by means of the nearest-neighbor approximation for the absorption coefficients of the unfolded species. The thermal stability of the DNA oligonucleotides was probed on a Jasco V-650 spectrophotometer. The heating and cooling experiments were carried out by monitoring the UV absorbance at 295 nm as a function of temperature. Unless otherwise stated, the DNA strand concentrations are 3–4 μM, suspending in a buffer containing 70 mM KCl and 20 mM potassium phosphate (pH 7.0). Samples were covered with paraffin oil to prevent evaporation when subjected to high temperature. The heating and cooling rates were 0.2 °C/min and the data was collected at every 1 °C interval for both processes. The spectra were then normalized manually to determine the fractions of folded and unfolded species during the melting and folding

transitions. The melting temperature was defined as the temperature of the middle transition point.

Circular Dichroism. The CD spectra were recorded at 25 °C on a JASCO-815 spectropolarimeter using a Hellma quartz cell of 1 cm path length, scanning from 220 nm to 320 nm (scan speed: 200 nm/min. data pitch: 1.0 nm, bandwidth: 2.0 nm). The oligonucleotides were suspended in a buffer containing 70 mM KCl and 20 mM potassium phosphate (pH 7.0). For each sample, an average of three or five scans was taken with the buffer spectra subtracted, and the data were normalized and zero-corrected at 320 nm.

NMR Spectroscopy. Unless otherwise stated, the NMR experiments were recorded at 25 °C on 600 and 700 MHz Bruker spectrometers. The strand concentration of the NMR samples was typically 0.5–2.0 mM which were suspended in 420–500 μL buffer solution containing 70 mM KCl and 20 mM potassium phosphate (pH 7.0). Resonances for guanine residues were unambiguously assigned using site-specific low-enrichment ^{15}N labeling.[3] The folding topology was analyzed by NOESY. All spectral analyses were performed using SpinWorks and FELIX (Felix NMR, Inc.) program.

6.2 G-Quadruplex Structures in the hTERT Promoter

DNA sample preparation. Unlabeled and site-specific labeled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer or purchased from Eurogentec (Belgium). DNA concentration was expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species.

Gel electrophoresis. In-gel migration properties (molecular size, shape and charge) of the structures formed by DNA oligonucleotides were probed by non-denaturing polyacrylamide gel electrophoresis (PAGE). Non-radiolabeled oligonucleotides (30 μM) were prepared in a 20 mM Tris-HCl pH 7.2 buffer supplemented with 100 mM KCl. Samples were heated at 90°C for 10 min and subsequently quenched in ice (fast-cooling) or cooled down to 30°C over 1 h. The samples were loaded on a 15% polyacrylamide gel supplemented with 20 mM KCl and run at 30°C; 10% sucrose was added just before loading. The gels were imaged by UV-shadowing.

UV-melting experiments and thermal difference spectra. The thermal stability of different oligonucleotides was characterized in heating/cooling experiments by recording the UV absorbance at 295 nm as a function of temperature using a Secomam-Uvikon XL UV-Vis spectrophotometer. UV-melting experiments were conducted in a 20 mM lithium cacodylate pH 7.2 buffer containing 100 mM monovalent cation (KCl, LiCl or a mixture of both). The heating and cooling rates were 0.2°C per minute. Experiments were performed with 0.2- or 1-cm path length quartz cuvettes. DNA concentration ranged from 0.5 to 100 μM. Two baselines corresponding to the completely folded (low temperature) and completely unfolded (high temperature) states were manually drawn in order to determine the fractions of folded and unfolded species during the melting process.

Circular dichroism. Unless otherwise stated, circular dichroism (CD) spectra were recorded at 20 or 25°C on a JASCO-815 spectropolarimeter using a 1-cm path length quartz cuvette with a reaction volume of 500 μl. DNA concentration was typically 4 μM. The DNA oligonucleotides were prepared in a pH 7.0 buffer containing 70 mM KCl and 20 mM potassium phosphate or in the same buffer as for the UV-melting experiments. For each sample, an average of three scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm. For equilibrium CD measurements, the 260-nm and 295-nm peaks were monitored until there were minimal fluctuations before the spectra were recorded. For unfolding and refolding kinetics study, a single scan was taken every 21 s. To create molecular crowding condition, PEG 200 (10%, 20%, 30% and 40% v/v) was added to a final DNA concentration of ~4 μM in 70 mM KCl and 20 mM potassium phosphate. The samples were incubated at 25 °C for 5 days.

Population estimation based on CD. The composite CD spectra of *GTERT-060* were computed based on the formula $C = \alpha A + (1-\alpha)B$, where α and $(1-\alpha)$ represent the proportions of the two species; A and B represent the respective component spectra; and C represents the composite spectrum, from 0% Form 1 (100% Form 2) to 100% Form 1 (0% Form 2). These computed fits were then compared against the original spectrum of the natural sequence and the respective RMSDs were calculated: for each composite spectrum (a total of 101 spectra), the difference in molar ellipticity at each wavelength (from that of natural *GTERT-060*) was squared, averaged across the range from 220 to 320 nm, and the root was subsequently obtained. The composite spectrum

displaying the lowest RMSD would produce the best fit to the original CD spectrum of *GTERT-060* and give an estimation of the proportions of the two coexisting conformations, assuming that the proportion of *GTERT-060* adopting other conformation(s) is negligible. "Pure" spectra of Form 1 and Form 2, calculated based on the average of the three spectra from both forms (*GTERT-060[I5]*, *GTERT-060[I7]* and *GTERT-060[I5/I7]* for Form 1; *GTERT-060[I2]*, *GTERT-060[I10]* and *GTERT-060[I2/I10]* for Form 2), were employed for fitting. The percentage populations of Form 1 and Form 2 thus derived were 64% and 36%, respectively. If other spectra (e.g. *GTERT-060[I5]*, *GTERT-060[I10]*, average of *GTERT-060[I5]* and *GTERT-060[I7]*, etc.) were used as the component spectra, the estimates for percentage populations of Form 1 and Form 2 were computed to be 50–80% and 20–50%, respectively, which still fairly agree with the NMR data.

NMR spectroscopy. Samples for NMR study were dialyzed successively against ~30 mM KCl solution and against water. Unless otherwise stated, the strand concentration of the NMR samples was typically 0.5–2.0 mM; the solutions contained 70 mM KCl and 20 mM potassium phosphate (pH 7.0). NMR experiments were performed on 600 MHz and 700 MHz Bruker spectrometers at 25°C, unless otherwise specified. Resonances for guanine residues were assigned unambiguously by using site-specific low-enrichment ^{15}N labeling, site-specific ^{2}H labeling and through-bond correlations at natural abundance. Spectral assignments were completed by NOESY, COSY, TOCSY and ^{13}C–^{1}H-HSQC. Inter-proton distances were deduced from NOESY experiments at various mixing times. All spectral analyses were performed using the FELIX (Felix NMR, Inc.) program.

Structure calculation. Inter-proton distances for the d[AGGGIAGGGGCTGGGA GGGC] (Form 1) and d[AIGGGAGGGICTGGGAGGGC] (Form 2) G-quadruplexes were classified based on NOESY experiments performed in H_2O (mixing time, 200 ms) and D_2O (mixing times, 100, 200 and 300 or 350 ms). Structure computations were performed using the XPLOR-NIH program in two general steps: (i) distance geometry simulated annealing and (ii) distance-restrained molecular dynamics refinement. Hydrogen bond restraints, inter-proton distance restraints, dihedral restraints, planarity restraints and repulsive restraints were imposed during structure calculations. Structures were displayed using the PyMOL program.

Hydrogen Bond Restraints. Each hydrogen bond was restrained by two distances

(donor atom to acceptor atom and proton to acceptor atom) which correspond to ideal hydrogen bond geometry. The force constant for hydrogen bond restraints was kept at 100 kcal.mol^{-1}.Å^{-2} throughout the computation.

Non-exchangeable Proton Distance Restraints. Distances between non-exchangeable protons were deduced from NOESY spectra in D_2O (mixing times, 100, 200 and 300/350 ms) and implemented as distance restraints during structure calculations. Cross-peaks that are strong at 100-ms mixing time were given bounds of 1.8–3.6 Å, cross-peaks that are strong at 200-ms mixing time were assigned distances of 3.0–4.6 Å, whereas cross-peaks that are observed only at 300- or 350-ms mixing time were converted to inter-proton distance constraints with the lower limit of 3.8 Å and the upper limit of 7.2 Å. Heavily overlapped cross-peaks were given the distances of 1.8 to 7.2 Å. Cross-peaks involving a methyl group were interpreted as distances to the methyl carbon. The upper bounds for these peaks were increased by 0.5 Å to account for the radius of the methyl group. The upper bounds for cross-peaks involving ambiguous atom selection were also increased by 0.5 Å.

Exchangeable Proton Distance Restraints. Cross-peaks involving exchangeable protons were obtained from the NOESY spectrum in H_2O at 200-ms mixing time. The peaks were manually classified as strong (1.8–3.8 Å or 1.8–4.5 Å), medium (2.8–5.2 Å), weak (3.6–6.0 Å or 3.5–7.5 Å) or very weak (4.5–7.5 Å). Cross-peaks for which the distance was hard to be determined were given a distance of 1.8–7.5 Å.

NMR Dihedral Restraints. The glycosidic χ torsion angle for experimentally-determined *syn* guanine residues were restrained to 60°, while that for *anti* guanine residues were fixed at 240°. In addition, the H1'–C1'–N9–C4 torsion angle for the former were set to 180°, whereas that for the latter were set to 0°. The bounds for these dihedrals were set at ±40° (central tetrad) or ±70° (outer tetrads). The ε dihedral angles between adjacent guanine residues constituting the core were set to be (225±75)°, consistent with the stereochemically allowed range for this angle.

Planarity Restraints. Planarity restraints were enforced on the base atoms for the tetrads, as well as additional base pairs, if they were determined to be present.

Repulsive Restraints. Repulsive restraints (4.5–80 Å) were applied on pairs of protons that do not exhibit cross-peaks in NOESY. A force constant of 40 kcal.mol^{-1}.Å^{-2} was applied.

188

Distance Geometry Simulated Annealing. An extended conformation of the d[AGGGIAGGGGCTGGGAGGGC] (Form 1) and d[AIGGGAGGGICTGGGAGGGC] (Form 2) sequences with ideal geometry was generated randomly using XPLOR-NIH. The extended DNA was then subjected to distance geometry simulated annealing (DGSA) by supplying the full set of hydrogen-bond, distance, dihedral, planarity and repulsive restraints. 100 structures were generated and subjected to further refinements.

Distance-Restrained Molecular Dynamics Refinement. All 100 DGSA structures were refined with distance-restrained molecular dynamics. The system was heated from 300 K to 1000 K in 5 ps and allowed to equilibrate for 1 ps, wherein the force constants for the distance restraints were kept at 2 kcal.mol^{-1}.Å^{-2}. The force constants for non-exchangeable proton, exchangeable proton, and repulsive distance restraints were then scaled to final values of 50, 40 and 40 kcal.mol^{-1}.Å^{-2}, respectively, in a 26-ps interval. Subsequently, the system was slowly cooled down to 300 K in 14 ps, after which equilibration was performed for another 10 ps. Coordinates of the molecule were saved every 0.5 ps during the last 4.0 ps and averaged. In the final step, the average structure was subjected to minimization until the gradient of energy was less than 0.1 kcal.mol^{-1}. Dihedral (200 kcal.mol^{-1}.rad^{-2}) and planarity (1 kcal.mol^{-1}.Å^{-2} for tetrads; 0.5 kcal.mol^{-1}.Å^{-2} for base pairs) restraints were maintained throughout the course of refinement. 10 best structures with the lowest overall energy were eventually selected.

7.2 Structural Basis of DNA Quadruplex–Duplex Junction Formation

DNA Sample Preparation. Unlabeled and site-specific labeled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer using products from Glen Research and Cambridge Isotope Laboratories. The oligonucleotides were de-protected following the manufacturer's protocols and purified using Poly-Pak^{TM} cartridges. Samples were dialyzed successively against ~20 mM KCl solution and against water, lyophilized, and prepared in a buffer containing 20 mM potassium phosphate (pH 7.0) and 20 mM KCl.

NMR Spectroscopy. Strand concentration of NMR samples was typically 0.2–2.0 mM. NMR experiments were performed on 400-, 600- and 700-MHz Bruker spectrometers at 25 °C, unless otherwise specified. Resonances for guanine residues

were assigned unambiguously using site-specific low-enrichment ^{15}N labeling, site-specific ^{2}H labeling, and through-bond correlations at natural abundance. Spectral assignments were assisted by NOESY, COSY, TOCSY, ^{13}C–^{1}H-HSQC and ^{31}P–^{1}H-HSQC experiments. The spectra were processed with the software TopSpin^{TM} and analyzed using the program FELIX (Felix NMR, Inc.).

Structure Calculation. Structures of the reference duplex hairpin and duplex–quadruplex constructs were calculated using the program XPLOR-NIH. Hydrogen-bond restraints, interproton distance restraints, dihedral restraints, planarity restraints and repulsive restraints were imposed during structure calculations. Interproton distance restraints, deduced from NOESY experiments at various mixing times, were classified as very strong, strong, medium, weak and very weak. Initially, an extended conformation of each sequence was generated randomly. The hairpin component was folded first by subjecting the extended chain to distance geometry simulated annealing and distance-restrained molecular dynamics refinement with the set of restraints involving the hairpin residues. Out of 100 conformers, the partially-folded structure with the lowest energy was then used as the template for the complete folding of the structure by supplying the full set of restraints for a successive round of calculations. 10 best structures with the lowest overall energy were selected. Structures were displayed using the program PyMOL.

8.2 Thermal Stability of DNA Quadruplex–Duplex Hybrids

DNA sample preparation. DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer using products from Glen Research. The oligonucleotides were de-protected following the manufacturer's protocols and purified using Poly-Pak^{TM} cartridges. Samples were dialyzed successively against water, 10 mM KCl solution, and water again. They were subsequently frozen, lyophilized, and suspended in a buffer containing 10 mM lithium phosphate (pH 7.0) and 10 mM KCl. DNA concentration is expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species.[300] Sample concentrations were determined from UV absorbance at 260 nm, based on molar extinction coefficients provided by the online application 'UV Spectrum of DNA Calculator' (http://biophysics.idtdna.com/UVSpectrum.html).

NMR spectroscopy. 1D NMR experiments were performed on Bruker AVANCE 600-MHz spectrometer at 25 °C, using JR-type pulse sequences[212] for water signal suppression.

Circular dichroism. Unless otherwise stated, circular dichroism (CD) spectra were recorded at 25 °C on a JASCO-815 spectropolarimeter over the range of 220–320 nm using a 1-cm path length quartz cuvette with a reaction volume of 500 μL. DNA concentration was typically 4 μM. For each sample, an average of three scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm. For CD-melting experiments, cooling and heating were successively performed across the temperature range of 15–95 °C over a total of 14 h. The full spectrum was recorded at intervals of 1 °C, after which molar ellipticity at the appropriate wavelength was extracted (295 nm for Construct I; 262 nm for Construct II). Two baselines corresponding to the completely folded (low temperature) and completely unfolded (high temperature) states were manually drawn in order to determine the fractions of folded and unfolded species during the melting process. Melting temperature (T_m) is defined as the temperature for which there are equal fractions of folded and unfolded species. For each sequence, the average T_m from the folding and unfolding experiments is presented. For clarity, only the folding curves are presented in the main text.

9.2 Putative Quadruplex Motifs Containing Duplex Stem Elements in the Human Genome

DNA Sample Preparation. Unlabeled and site-specific labeled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer. The oligonucleotides were de-protected, purified, dialyzed successively against ~20 mM KCl and against water, and prepared in a buffer containing 20 mM KCl and 20 mM potassium phosphate (pH 7.0).

NMR Spectroscopy. NMR experiments were performed on Bruker 600 and 700 MHz spectrometers at 25 °C, unless otherwise specified. Resonances for guanine residues were assigned unambiguously by using site-specific low-enrichment ^{15}N-labeling[192] and site-specific ^{2}H labeling[207]. Spectral assignments were assisted by NOESY, COSY, TOCSY, and ^{13}C–^{1}H-HSQC, as previously described[212]. All spectral analyses were performed using the program FELIX (Felix NMR, Inc.).

References

1.   Watson, J.D. and Crick, F.H.C. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.

2.   Wilkins, M.H.F., Stokes, A.R., and Wilson, H.R. (1953) Molecular structure of deoxypentose nucleic acids. *Nature* **171**, 738–740.

3.   Franklin, R.E. and Gosling, R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741.

4.   Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R.E. (1981) Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **78**, 2179–2183.

5.   Verdaguer, N., Aymami, J., Fernandezforner, D., Fita, I., Coll, M., Huynhdinh, T., Igolen, J., and Subirana, J.A. (1991) Molecular structure of a complete turn of *A*-DNA. *J. Mol. Biol.* **221**, 623–635.

6.   Franklin, R.E. and Gosling, R.G. (1953) Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **172**, 156–157.

7.   Arnott, S., Hukins, D.W.L., and Dover, S.D. (1972) Optimized parameters for RNA double-helices. *Biochem. Biophys. Res. Commun.* **48**, 1392–1399.

8.   Wang, A.H.J., Fujii, S., Vanboom, J.H., Vandermarel, G.A., Vanboeckel, S.A.A., and Rich, A. (1982) Molecular structure of r(GCG)d(TATACGC): A DNA–RNA hybrid helix joined to double helical DNA. *Nature* **299**, 601–604.

9.   Drozdzal, P., Michalska, K., Kierzek, R., Lomozik, L., and Jaskolski, M. (2012) Structure of an RNA/DNA dodecamer corresponding to the HIV-1 polypurine tract at 1.6 Å resolution. *Acta Crystallogr. D* **68**, 169–175.

10.   Wang, A.H.J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., Vanboom, J.H., Vandermarel, G., and Rich, A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686.

11.   Wang, A.H.J., Quigley, G.J., Kolpak, F.J., Vandermarel, G., Vanboom, J.H., and Rich, A. (1981) Left-handed double helical DNA: Variations in the backbone conformation. *Science* **211**, 171–176.

12.   Rich, A., Nordheim, A., and Wang, A.H.J. (1984) The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **53**, 791–846.

13.   Schroth, G.P., Chou, P.J., and Ho, P.S. (1992) Mapping Z-DNA in the human genome: Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J. Biol. Chem.* **267**, 11846–11855.

14.   Lilley, D.M.J. (2000) Structures of helical junctions in nucleic acids. *Q. Rev. Biophys.* **33**, 109–159.

15. Holliday, R. (1964) A mechanism for gene conversion in fungi. *Genet. Res.* **5**, 282–304.

16. Liu, Y.L. and West, S.C. (2004) Happy Hollidays: 40th anniversary of the Holliday junction. *Nat Rev Mol Cell Bio* **5**, 937–946.

17. Eichman, B.F., Vargason, J.M., Mooers, B.H.M., and Ho, P.S. (2000) The Holliday junction in an inverted repeat DNA sequence: Sequence effects on the structure of four-way junctions. *Proc. Natl. Acad. Sci. USA* **97**, 3971–3976.

18. Ha, S.C., Lowenhaupt, K., Rich, A., Kim, Y.G., and Kim, K.K. (2005) Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183–1186.

19. Hoogsteen, K. (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **16**, 907–916.

20. Thuong, N.T. and Hélène, C. (1993) Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed.* **32**, 666–690.

21. Felsenfeld, G., Davies, D.R., and Rich, A. (1957) Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **79**, 2023–2024.

22. Radhakrishnan, I. and Patel, D.J. (1994) Solution structure of a pyrimidine•purine•pyrimidine DNA triplex containing T•AT, C^+•GC and G•TA triples. *Structure* **2**, 17–32.

23. Moser, H.E. and Dervan, P.B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **238**, 645–650.

24. Beal, P.A. and Dervan, P.B. (1991) Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science* **251**, 1360–1363.

25. Duca, M., Vekhoff, P., Oussedik, K., Halby, L., and Arimondo, P.B. (2008) The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* **36**, 5123–5138.

26. Gehring, K., Leroy, J.L., and Guéron, M. (1993) A tetrameric DNA structure with protonated cytosine•cytosine base pairs. *Nature* **363**, 561–565.

27. Simonsson, T. (2001) G-quadruplex DNA structures – Variations on a theme. *Biol. Chem.* **382**, 621–628.

28. Davis, J.T. (2004) G-quartets 40 years later: From 5 '-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed.* **43**, 668–698.

29. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415.

30. Patel, D.J., Phan, A.T., 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 Res.* **35**, 7429–7455.

31. Gellert, M., Lipsett, M.N., and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. USA* **48**, 2013–2018.

32. Phan, A.T. (2010) Human telomeric G-quadruplex: Structures of DNA and RNA sequences. *FEBS J.* **277**, 1107–1117.

33. Parkinson, G.N., Lee, M.P.H., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–880.

34. 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. *Proc. Natl. Acad. Sci. USA* **90**, 3745–3749.

35. Schultze, P., Macaya, R.F., and Feigon, J. (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Biol.* **235**, 1532–1547.

36. Smith, F.W. and Feigon, J. (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature* **356**, 164–168.

37. Wang, Y. and Patel, D.J. (1994) Solution structure of the *Tetrahymena* telomeric repeat d(T$_2$G$_4$)$_4$ G-tetraplex. *Structure* **2**, 1141–1156.

38. 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**, 364–366.

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

40. Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* **59**, 871–880.

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

42. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916.

43. Blackburn, E.H. (1991) Structure and function of telomeres. *Nature* **350**, 569–573.

44. Simonsson, T., Pecinka, P., and Kubista, M. (1998) DNA tetraplex formation in the control region of c-*myc*. *Nucleic Acids Res.* **26**, 1167–1172.

45. 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. *Proc. Natl. Acad. Sci. USA* **99**, 11593–11598.

46. Eddy, J. and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887–3896.

47. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406–413.

48. Eddy, J. and Maizels, N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* **36**, 1321–1333.

49. Shimizu, A. and Honjo, T. (1984) Immunoglobulin class switching. *Cell* **36**, 801–803.

50. Hanakahi, L.A., Sun, H., and Maizels, N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.* **274**, 15908–15912.

51. Mitani, K., Takahashi, Y., and Kominami, R. (1990) A GGCAGG motif in minisatellites affecting their germline instability. *J. Biol. Chem.* **265**, 15203–15210.

52. 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. *Nat. Chem. Biol.* **3**, 218–221.

53. Huppert, J.L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008) G-quadruplexes: The beginning and end of UTRs. *Nucleic Acids Res.* **36**, 6260–6268.

54. Cahoon, L.A. and Seifert, H.S. (2009) An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science* **325**, 764–767.

55. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nat. Struct. Mol. Biol.* **12**, 847–854.

56. Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D., and Lipps, H.J. (2008) Telomerase recruitment by the telomere end binding protein-β facilitates G-quadruplex DNA unfolding in ciliates. *Nat. Struct. Mol. Biol.* **15**, 598–604.

57. Paeschke, K., Capra, J.A., and Zakian, V.A. (2011) DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* **145**, 678–691.

58. 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. *EMBO J.* **30**, 4033–4046.

59. Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.* **13**, 1055–1059.

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

61. Balasubramanian, S. and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **13**, 345–353.

62. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., and Toole, J.J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566.

63. Travascio, P., Li, Y.F., and Sen, D. (1998) DNA-enhanced peroxidase activity of a DNA aptamer-hemin complex. *Chem. Biol.* **5**, 505–517.

64. Ge, B.X., Huang, Y.C., Sen, D., and Yu, H.Z. (2010) A robust electronic switch made of immobilized duplex/quadruplex DNA. *Angew. Chem. Int. Ed.* **49**, 9965–9967.

65. Bates, P.J., Kahlon, J.B., Thomas, S.D., Trent, J.O., and Miller, D.M. (1999) Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. *J. Biol. Chem.* **274**, 26369–26377.

66. Simonsson, T. and Henriksson, M. (2002) c-myc suppression in Burkitt's lymphoma cells. *Biochem. Biophys. Res. Commun.* **290**, 11–15.

67. Jing, N.J., Li, Y.D., Xiong, W.J., Sha, W., Jing, L., and Tweardy, D.J. (2004) G-quartet oligonucleotides: A new class of signal transducer and activator of transcription 3 inhibitors that suppresses growth of prostate and breast tumors through induction of apoptosis. *Cancer Res.* **64**, 6603–6609.

68. Qi, H.Y., Lin, C.P., Fu, X., Wood, L.M., Liu, A.A., Tsai, Y.C., Chen, Y.J., Barbieri, C.M., Pilch, D.S., and Liu, L.F. (2006) G-quadruplexes induce apoptosis in tumor cells. *Cancer Res.* **66**, 11808–11816.

69. Choi, E.W., Nayak, L.V., and Bates, P.J. (2010) Cancer-selective antiproliferative activity is a general property of some G-rich oligodeoxynucleotides. *Nucleic Acids Res.* **38**, 1623–1635.

70. Wyatt, J.R., Vickers, T.A., Roberson, J.L., Buckheit, R.W., Klimkait, T., Debaets, E., Davis, P.W., Rayner, B., Imbach, J.L., and Ecker, D.J. (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proc. Natl. Acad. Sci. USA* **91**, 1356–1360.

71. Rando, R.F., Ojwang, J., Elbaggari, A., Reyes, G.R., Tinder, R., Mcgrath, M.S., and Hogan, M.E. (1995) Suppression of human immunodeficiency virus type 1 activity *in vitro* by oligonucleotides which form intramolecular tetrads. *J. Biol. Chem.* **270**, 1754–1760.

72. Jing, N.J., Gao, X.L., Rando, R.F., and Hogan, M.E. (1997) Potassium-induced loop conformational transition of a potent anti-HIV oligonucleotide. *J Biomol Struct Dyn* **15**, 573–585.

73. Jing, N.J., Marchand, C., Liu, J., Mitra, R., Hogan, M.E., and Pommier, Y. (2000) Mechanism of inhibition of HIV-1 integrase by G-tetrad-forming oligonucleotides *in vitro*. *J. Biol. Chem.* **275**, 21460–21467.

74. Jing, N.J., Xiong, W.J., Guan, Y.L., Pallansch, L., and Wang, S.M. (2002) Potassium-dependent folding: A key to intracellular delivery of G-quartet oligonucleotides as HIV inhibitors. *Biochemistry* **41**, 5397–5403.

75. 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. *Proc. Natl. Acad. Sci. USA* **102**, 634–639.

76. Marsh, T.C., Vesenka, J., and Henderson, E. (1995) A new DNA nanostructure, the G-wire, imaged by scanning probe microscopy. *Nucleic Acids Res.* **23**, 696–700.

77. Zhou, C.Q., Tan, Z.K., Wang, C., Wei, Z.Q., Wang, Z.G., Bai, C.L., Qin, J.F., and Cao, E.H. (2001) Branched nanowire based guanine rich oligonucleotides. *J Biomol Struct Dyn* **18**, 807–812.

78. Batalia, M.A., Protozanova, E., Macgregor, R.B., and Erie, D.A. (2002) Self-assembly of frayed wires and frayed-wire networks: Nanoconstruction with multistranded DNA. *Nano Lett.* **2**, 269–274.

79. Dutta, K., Fujimoto, T., Inoue, M., Miyoshi, D., and Sugimoto, N. (2010) Development of new functional nanostructures consisting of both DNA duplex and quadruplex. *Chem Commun* **46**, 7772–7774.

80. Seeman, N.C. (2010) Nanomaterials Based on DNA. *Annu. Rev. Biochem.* **79**, 65–87.

81. Neidle, S. (2001) DNA minor-groove recognition by small molecules. *Nat Prod Rep* **18**, 291–309.

82. Spink, N., Brown, D.G., Skelly, J.V., and Neidle, S. (1994) Sequence-dependent effects in drug-DNA interaction: the crystal structure of Hoechst 33258 bound to the d(CGCAAATTTGCG)$_2$ duplex. *Nucleic Acids Res.* **22**, 1607–1612.

83. Trauger, J.W., Baird, E.E., and Dervan, P.B. (1996) Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559–561.

84. Kapuscinski, J. (1995) DAPI: A DNA-specific fluorescent probe. *Biotech Histochem* **70**, 220-233.

85. Gottesfeld, J.M., Neely, L., Trauger, J.W., Baird, E.E., and Dervan, P.B. (1997) Regulation of gene expression by small molecules. *Nature* **387**, 202–205.

86. Mapp, A.K., Ansari, A.Z., Ptashne, M., and Dervan, P.B. (2000) Activation of gene expression by small molecule transcription factors. *Proc. Natl. Acad. Sci. USA* **97**, 3930–3935.

87. Hélène, C. and Toulme, J.J. (1990) Specific regulation of gene expression by antisense, sense and antigene nucleic acids. *Biochim. Biophys. Acta* **1049**, 99–125.

88. Nielsen, P.E., Egholm, M., Berg, R.H., and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254**, 1497–1500.

89. Obika, S., Nanbu, D., Hari, Y., Morio, K., In, Y., Ishida, T., and Imanishi, T. (1997) Synthesis of 2'-*O*,4'-*C*-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed $C_{3'}$-*endo* sugar puckering. *Tetrahedron Lett.* **38**, 8735–8738.

90. Koshkin, A.A., Singh, S.K., Nielsen, P., Rajwanshi, V.K., Kumar, R., Meldgaard, M., Olsen, C.E., and Wengel, J. (1998) LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* **54**, 3607–3630.

91. Sun, J.S., Francois, J.C., Montenaygarestier, T., Saisonbehmoaras, T., Roig, V., Thuong, N.T., and Hélène, C. (1989) Sequence-specific Intercalating agents: Intercalation at specific sequences on duplex DNA via major groove recognition by oligonucleotide–intercalator conjugates. *Proc. Natl. Acad. Sci. USA* **86**, 9198–9202.

92. Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* **19**, 239–250.

93. Read, M., Harrison, R.J., Romagnoli, B., Tanious, F.A., Gowan, S.H., Reszka, A.P., Wilson, W.D., Kelland, L.R., and Neidle, S. (2001) Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. USA* **98**, 4844–4849.

94. Haider, S.M., Parkinson, G.N., and Neidle, S. (2003) Structure of a G-quadruplex–ligand complex. *J. Mol. Biol.* **326**, 117–125.

95. Shin-ya, K., Wierzba, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Hayakawa, Y., and Seto, H. (2001) Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*. *J. Am. Chem. Soc.* **123**, 1262–1263.

96. Cosconati, S. *et al.* (2010) Structural and conformational requisites in DNA quadruplex groove binding: Another piece to the puzzle. *J. Am. Chem. Soc.* **132**, 6425–6433.

97. Hamon, F., Largy, E., Guédin-Beaurepaire, A., Rouchon-Dagois, M., Sidibe, A., Monchaud, D., Mergny, J.L., Riou, J.F., Nguyen, C.H., and Teulade-Fichou, M.P. (2011) An acyclic oligoheteroaryle that discriminates strongly between diverse G-quadruplex topologies. *Angew. Chem. Int. Ed.* **50**, 8745–8749.

98. Breaker, R.R. (1997) DNA aptamers and DNA enzymes. *Curr. Opin. Chem. Biol.* **1**, 26–31.

99. Silverman, S.K. (2008) Catalytic DNA (deoxyribozymes) for synthetic applications–current abilities and future prospects. *Chem Commun*, 3467–3485.

100. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510.

101. Pinheiro, A.V., Han, D.R., Shih, W.M., and Yan, H. (2011) Challenges and opportunities for structural DNA nanotechnology. *Nat Nanotechnol* **6**, 763–772.

102. Winfree, E., Liu, F.R., Wenzler, L.A., and Seeman, N.C. (1998) Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544.

103. Chen, J.H. and Seeman, N.C. (1991) Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633.

104. Goodman, R.P., Schaap, I.A.T., Tardin, C.F., Erben, C.M., Berry, R.M., Schmidt, C.F., and Turberfield, A.J. (2005) Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**, 1661–1665.

105. Rothemund, P.W.K. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302.

106. Douglas, S.M., Dietz, H., Liedl, T., Hogberg, B., Graf, F., and Shih, W.M. (2009) Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418.

107. Han, D.R., Pal, S., Nangreave, J., Deng, Z.T., Liu, Y., and Yan, H. (2011) DNA origami with complex curvatures in three-dimensional space. *Science* **332**, 342–346.

108. Wei, B., Dai, M.J., and Yin, P. (2012) Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623–626.

109. Zheng, J.P., Birktoft, J.J., Chen, Y., Wang, T., Sha, R.J., Constantinou, P.E., Ginell, S.L., Mao, C.D., and Seeman, N.C. (2009) From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* **461**, 74–77.

110. Simmel, F.C. and Dittmer, W.U. (2005) DNA nanodevices. *Small* **1**, 284–299.

111. Mao, C.D., Sun, W.Q., Shen, Z.Y., and Seeman, N.C. (1999) A nanomechanical device based on the B–Z transition of DNA. *Nature* **397**, 144–146.

112. Yan, H., Zhang, X.P., Shen, Z.Y., and Seeman, N.C. (2002) A robust DNA mechanical device controlled by hybridization topology. *Nature* **415**, 62–65.

113. Maune, H.T., Han, S.P., Barish, R.D., Bockrath, M., Goddard, W.A., Rothemund, P.W.K., and Winfree, E. (2010) Self-assembly of carbon nanotubes

into two-dimensional geometries using DNA origami templates. *Nat Nanotechnol* **5**, 61-66.

114. Endo, M., Katsuda, Y., Hidaka, K., and Sugiyama, H. (2010) Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure. *J. Am. Chem. Soc.* **132**, 1592–1597.

115. Sannohe, Y., Endo, M., Katsuda, Y., Hidaka, K., and Sugiyama, H. (2010) Visualization of dynamic conformational switching of the G-quadruplex in a DNA nanostructure. *J. Am. Chem. Soc.* **132**, 16311–16313.

116. Rajendran, A., Endo, M., and Sugiyama, H. (2012) Single-molecule analysis using DNA origami. *Angew. Chem. Int. Ed.* **51**, 874–890.

117. Liu, J.W., Cao, Z.H., and Lu, Y. (2009) Functional nucleic acid sensors. *Chem Rev* **109**, 1948–1998.

118. Aldaye, F.A. and Sleiman, H.F. (2007) Modular access to structurally switchable 3D discrete DNA assemblies. *J. Am. Chem. Soc.* **129**, 13376–13377.

119. Andersen, E.S. *et al.* (2009) Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73–76.

120. Douglas, S.M., Bachelet, I., and Church, G.M. (2012) A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834.

121. Saccà, B., Meyer, R., Erkelenz, M., Kiko, K., Arndt, A., Schroeder, H., Rabe, K.S., and Niemeyer, C.M. (2010) Orthogonal protein decoration of DNA origami. *Angew. Chem. Int. Ed.* **49**, 9378–9383.

122. Dutta, P.K., Varghese, R., Nangreave, J., Lin, S., Yan, H., and Liu, Y. (2011) DNA-directed artificial light-harvesting antenna. *J. Am. Chem. Soc.* **133**, 11985–11993.

123. Merrifield, R.B. (1963) Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154.

124. Matteucci, M.D. and Caruthers, M.H. (1981) Synthesis of deoxyoligonucleotides on a polymer support. *J. Am. Chem. Soc.* **103**, 3185–3191.

125. Beaucage, S.L. and Caruthers, M.H. (1981) Deoxynucleoside phosphoramidites–A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862.

126. Plateau, P. and Guéron, M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.* **104**, 7310–7311.

127. Mergny, J.L., Phan, A.T., and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.* **435**, 74–78.

128. Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73.

129. Zakian, V.A. (1989) Structure and function of telomeres. *Annu. Rev. Genet.* **23**, 579–604.

130. Zakian, V.A. (1995) Telomeres: beginning to understand the end. *Science* **270**, 1601–1607.

131. Hodes, R.J., Hathcock, K.S., and Weng, N.P. (2002) Telomeres in T and B cells. *Nat Rev Immunol* **2**, 699–706.

132. Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., and Wu, J.R. (1988) A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA* **85**, 6622–6626.

133. Allshire, R.C., Dempster, M., and Hastie, N.D. (1989) Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.* **17**, 4611–4627.

134. Makarov, V.L., Hirose, Y., and Langmore, J.P. (1997) Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88**, 657–666.

135. Wright, W.E., Tesmer, V.M., Huffman, K.E., Levene, S.D., and Shay, J.W. (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev.* **11**, 2801–2809.

136. McElligott, R. and Wellinger, R.J. (1997) The terminal DNA structure of mammalian chromosomes. *EMBO J.* **16**, 3705–3714.

137. de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110.

138. de Lange, T. (2009) How telomeres solve the end-protection problem. *Science* **326**, 948–952.

139. Sandell, L.L. and Zakian, V.A. (1993) Loss of a yeast telomere: arrest, recovery, and chromosome Loss. *Cell* **75**, 729–739.

140. Baumann, P. and Cech, T.R. (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171–1175.

141. Li, B.B., Oestreich, S., and de Lange, T. (2000) Identification of human Rap1: Implications for telomere evolution. *Cell* **101**, 471–483.

142. Kim, S.H., Kaminker, P., and Campisi, J. (1999) TIN2, a new regulator of telomere length in human cells. *Nat. Genet.* **23**, 405–412.

202

143. Houghtaling, B.R., Cuttonaro, L., Chang, W., and Smith, S. (2004) A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Curr. Biol.* **14**, 1621–1631.

144. Liu, D., Safari, A., O'Connor, M.S., Chan, D.W., Laegeler, A., Qin, J., and Zhou, S.Y. (2004) PTOP interacts with POT1 and regulates its localization to telomeres. *Nat. Cell Biol.* **6**, 673–680.

145. Ye, J.Z.S., Hockemeyer, D., Krutchinsky, A.N., Loayza, D., Hooper, S.M., Chait, B.T., and de Lange, T. (2004) POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes Dev.* **18**, 1649–1654.

146. Zhong, Z., Shiue, L., Kaplan, S., and Delange, T. (1992) A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Mol. Cell. Biol.* **12**, 4834–4843.

147. Chong, L., Vansteensel, B., Broccoli, D., Erdjumentbromage, H., Hanish, J., Tempst, P., and Delange, T. (1995) A human telomeric protein. *Science* **270**, 1663–1667.

148. Bilaud, T., Brun, C., Ancelin, K., Koering, C.E., Laroche, T., and Gilson, E. (1997) Telomeric localization of TRF2, a novel human telobox protein. *Nat. Genet.* **17**, 236–239.

149. Broccoli, D., Smogorzewska, A., Chong, L., and deLange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* **17**, 231–235.

150. Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., and de Lange, T. (1999) Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514.

151. Harley, C.B., Futcher, A.B., and Greider, C.W. (1990) Telomeres shorten during aging of human fibroblasts. *Nature* **345**, 458–460.

152. Greider, C.W. (1996) Telomere length regulation. *Annu. Rev. Biochem.* **65**, 337–365.

153. Hemann, M.T., Strong, M.A., Hao, L.Y., and Greider, C.W. (2001) The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107**, 67-77.

154. Counter, C.M., Avilion, A.A., Lefeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B., and Bacchetti, S. (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **11**, 1921–1929.

155. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015.

156. Greider, C.W. and Blackburn, E.H. (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405–413.

157. De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.F., and Mergny, J.L. (2008) Targeting telomeres and telomerase. *Biochimie* **90**, 131–155.

158. Zaug, A.J., Podell, E.R., and Cech, T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*. *Proc. Natl. Acad. Sci. USA* **102**, 10864–10869.

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

160. De Cian, A., Cristofari, G., Reichenbach, P., De Lemos, E., Monchaud, D., Teulade-Fichou, M.P., Shin-Ya, K., Lacroix, L., Lingner, J., and Mergny, J.L. (2007) Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proc. Natl. Acad. Sci. USA* **104**, 17347–17352.

161. Sun, D.Y., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S., and Hurley, L.H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.* **40**, 2113–2116.

162. He, Y.J., Neumann, R.D., and Panyutin, I.G. (2004) Intramolecular quadruplex conformation of human telomeric DNA assessed with ^{125}I-radioprobing. *Nucleic Acids Res.* **32**, 5359–5367.

163. Neidle, S. and Parkinson, G. (2002) Telomere maintenance as a target for anticancer drug discovery. *Nat. Rev. Drug Discov.* **1**, 383–393.

164. Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d[AG$_3$(T$_2$AG$_3$)$_3$] G-tetraplex. *Structure* **1**, 263–282.

165. 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. *J. Am. Chem. Soc.* **128**, 9963–9970.

166. Phan, A.T., Luu, K.N., and Patel, D.J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K$^+$ solution. *Nucleic Acids Res.* **34**, 5715–5719.

167. Phan, A.T., Kuryavyi, V., Luu, K.N., and Patel, D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K$^+$ solution. *Nucleic Acids Res.* **35**, 6517–6525.

168. Xu, Y. and Sugiyama, H. (2004) Highly efficient photochemical 2'-deoxyribonolactone formation at the diagonal loop of a 5-iodouracil-containing antiparallel G-quartet. *J. Am. Chem. Soc.* **126**, 6274–6279.

169. Xu, Y., Noguchi, Y., and Sugiyama, H. (2006) The new models of the human telomere d[AGGG(TTAGGG)$_3$] in K$^+$ solution. *Bioorgan Med Chem* **14**, 5584–5591.

170. Matsugami, A., Xu, Y., Noguchi, Y., Sugiyama, H., and Katahira, M. (2007) Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K$^+$ solution. *FEBS J.* **274**, 3545–3556.

171. Okamoto, K., Sannohe, Y., Mashimo, T., Sugiyama, H., and Terazima, M. (2008) G-quadruplex structures of human telomere DNA examined by single molecule FRET and BrG-substitution. *Bioorgan Med Chem* **16**, 6873–6879.

172. Ambrus, A., Chen, D., Dai, J.X., Bialis, T., Jones, R.A., and Yang, D.Z. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **34**, 2723–2735.

173. Dai, J.X., Punchihewa, C., Ambrus, A., Chen, D., Jones, R.A., and Yang, D.Z. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: A novel adenine triple formation. *Nucleic Acids Res.* **35**, 2440–2450.

174. Dai, J.X., Carver, M., Punchihewa, C., Jones, R.A., and Yang, D.Z. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K$^+$ solution: Insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* **35**, 4927–4940.

175. Ying, L.M., Green, J.J., Li, H.T., Klenerman, D., and Balasubramanian, S. (2003) Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA* **100**, 14629–14634.

176. Redon, S., Bombard, S., Elizondo-Riojas, M.A., and Chottard, J.C. (2003) Platinum cross-linking of adenines and guanines on the quadruplex structures of the AG$_3$(T$_2$AG$_3$)$_3$ and (T$_2$AG$_3$)$_4$ human telomere sequences in Na$^+$ and K$^+$ solutions. *Nucleic Acids Res.* **31**, 1605–1613.

177. Ourliac-Garnier, I., Elizondo-Riojas, M.A., Redon, S., Farrell, N.P., and Bombard, S. (2005) Cross-links of quadruplex structures from human telomeric DNA by dinuclear platinum complexes show the flexibility of both structures. *Biochemistry* **44**, 10620–10634.

178. 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 Res.* **36**, 4079–4087.

179. D'Isa, G., Galeone, A., Oliviero, G., Piccialli, G., Varra, M., and Mayol, L. (2004) Effect of γ-hydroxypropano deoxyguanosine, the major acrolein-derived adduct, on monomolecular quadruplex structure of telomeric repeat d(TTAGGG)$_4$. *Bioorg Med Chem Lett* **14**, 5417–5421.

180. Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **126**, 16405–16415.

181. Risitano, A. and Fox, K.R. (2005) Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium. *Bioorg Med Chem Lett* **15**, 2047–2050.

182. Rezler, E.M., Seenisamy, J., Bashyam, S., Kim, M.Y., White, E., Wilson, W.D., and Hurley, L.H. (2005) Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J. Am. Chem. Soc.* **127**, 9439–9447.

183. Rujan, I.N., Meleney, J.C., and Bolton, P.H. (2005) Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic Acids Res.* **33**, 2022–2031.

184. Wlodarczyk, A., Grzybowski, P., Patkowski, A., and Dobek, A. (2005) Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. *J Phys Chem B* **109**, 3594–3605.

185. Qi, J.Y. and Shafer, R.H. (2005) Covalent ligation studies on the human telomere quadruplex. *Nucleic Acids Res.* **33**, 3185–3192.

186. Vorlíčková, M., Chládková, J., Kejnovská, I., Fialová, M., and Kypr, J. (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res.* **33**, 5851–5860.

187. Li, J., Correia, J.J., Wang, L., Trent, J.O., and Chaires, J.B. (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res.* **33**, 4649–4659.

188. Lee, J.Y., Okumus, B., Kim, D.S., and Ha, T.J. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. USA* **102**, 18938–18943.

189. Yu, H.Q., Miyoshi, D., and Sugimoto, N. (2006) Characterization of structure and stability of long telomeric DNA G-quadruplexes. *J. Am. Chem. Soc.* **128**, 15461–15468.

190. Xue, Y., Kan, Z.Y., Wang, Q., Yao, Y., Liu, J., Hao, Y.H., and Tan, Z. (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K^+ solution under molecular crowding condition. *J. Am. Chem. Soc.* **129**, 11185–11191.

191. Balagurumoorthy, P., Brahmachari, S.K., Mohanty, D., Bansal, M., and Sasisekharan, V. (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* **20**, 4061–4067.

192. Phan, A.T. and Patel, D.J. (2002) A site-specific low-enrichment $^{15}N,^{13}C$ isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.* **124**, 1160–1161.

193. Phan, A.T. (2000) Long-range imino proton-^{13}C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR* **16**, 175–178.

194. Kuryavyi, V., Kettani, A., Wang, W.M., Jones, R., and Patel, D.J. (2000) A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads. *J. Mol. Biol.* **295**, 455–469.

195. Gray, D.M., Wen, J.D., Gray, C.W., Repges, R., Repges, C., Raabe, G., and Fleischhauer, J. (2008) Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* **20**, 431–440.

196. Lim, K.W., Amrane, S., Bouaziz, S., Xu, W.X., Mu, Y.G., Patel, D.J., Luu, K.N., and Phan, A.T. (2009) Structure of the human telomere in K$^+$ solution: A stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.* **131**, 4301–4309.

197. Smith, F.W., Lau, F.W., and Feigon, J. (1994) d(G$_3$T$_4$G$_3$) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-*syn-syn-anti* and 5'-*syn-anti-anti* N-glycosidic conformations. *Proc. Natl. Acad. Sci. USA* **91**, 10546–10550.

198. Strahan, G.D., Shafer, R.H., and Keniry, M.A. (1994) Structural properties of the [d(G$_3$T$_4$G$_3$)]$_2$ quadruplex: Evidence for sequential syn-syn deoxyguanosines. *Nucleic Acids Res.* **22**, 5447–5455.

199. Črnugelj, M., Hud, N.V., and Plavec, J. (2002) The solution structure of d(G$_4$T$_4$G$_3$)$_2$: A bimolecular G-quadruplex with a novel fold. *J. Mol. Biol.* **320**, 911–924.

200. Heddi, B. and Phan, A.T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.* **133**, 9824–9833.

201. Črnugelj, M., Sket, P., and Plavec, J. (2003) Small change in a G-rich sequence, a dramatic change in topology: New dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.* **125**, 7866–7871.

202. Webba da Silva, M., Trajkovski, M., Sannohe, Y., Hessari, N.M., Sugiyama, H., and Plavec, J. (2009) Design of a G-quadruplex topology through glycosidic bond angles. *Angew. Chem. Int. Ed.* **48**, 9167–9170.

203. Marušič, M., Šket, P., Bauer, L., Viglasky, V., and Plavec, J. (2012) Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Res.* **40**, 6946–6956.

204. Kuryavyi, V., Majumdar, A., Shallop, A., Chernichenko, N., Skripkin, E., Jones, R., and Patel, D.J. (2001) A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G(*syn*)•G(*syn*)•G(*anti*)•G(*anti*) tetrads flanked by a G•(T-T) triad and a T•T•T triple. *J. Mol. Biol.* **310**, 181–194.

205. Baird, D.M., Jeffreys, A.J., and Royle, N.J. (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *EMBO J.* **14**, 5433–5443.

206. Mendez-Bermudez, A., Hills, M., Pickett, H.A., Phan, A.T., Mergny, J.L., Riou, J.F., and Royle, N.J. (2009) Human telomeres that contain (CTAGGG)$_n$ repeats show replication dependent instability in somatic cells and the male germline. *Nucleic Acids Res.* **37**, 6225–6238.

207. Huang, X.N., Yu, P.L., LeProust, E., and Gao, X.L. (1997) An efficient and economic site-specific deuteration strategy for NMR studies of homologous oligonucleotide repeat sequences. *Nucleic Acids Res.* **25**, 4758–4763.

208. Kettani, A., Kumar, R.A., and Patel, D.J. (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.* **254**, 638–656.

209. Kettani, A., Bouaziz, S., Gorin, A., Zhao, H., Jones, R.A., and Patel, D.J. (1998) Solution structure of a Na cation stabilized DNA quadruplex containing G•G•G•G and G•C•G•C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.* **282**, 619–636.

210. Zhang, N., Gorin, A., Majumdar, A., Kettani, A., Chernichenko, N., Skripkin, E., and Patel, D.J. (2001) Dimeric DNA quadruplex containing major groove-aligned A•T•A•T and G•C•G•C tetrads stabilized by inter-subunit Watson-Crick A•T and G•C pairs. *J. Mol. Biol.* **312**, 1073–1088.

211. Bouaziz, S., Kettani, A., and Patel, D.J. (1998) A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats. *J. Mol. Biol.* **282**, 637–652.

212. Phan, A.T., Guéron, M., and Leroy, J.L. (2002) Investigation of unusual DNA motifs. *Methods Enzymol.* **338**, 341–371.

213. Phan, A.T., Guéron, M., and Leroy, J.L. (2000) The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.* **299**, 123–144.

214. Mergny, J.L., De Cian, A., Amrane, S., and da Silva, M.W. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res.* **34**, 2386–2397.

215. 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 Res.* **35**, 3064–3075.

216. Pedroso, I.A., Duarte, L.F., Yanez, G., Burkewitz, K., and Fletcher, T.M. (2007) Sequence specificity of inter- and intramolecular G-quadruplex formation by human telomeric DNA. *Biopolymers* **87**, 74–84.

217. Tang, J., Kan, Z.Y., Yao, Y., Wang, Q., Hao, Y.H., and Tan, Z. (2008) G-quadruplex preferentially forms at the very 3 ' end of vertebrate telomeric DNA. *Nucleic Acids Res.* **36**, 1200–1208.

218. Petraccone, L., Trent, J.O., and Chaires, J.B. (2008) The tail of the telomere. *J. Am. Chem. Soc.* **130**, 16530–16532.

219. Haider, S., Parkinson, G.N., and Neidle, S. (2008) Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys. J.* **95**, 296–311.

220. Xu, Y., Sato, H., Sannohe, Y., Shinohara, K., and Sugiyama, H. (2008) Stable lariat formation based on a G-quadruplex scaffold. *J. Am. Chem. Soc.* **130**, 16470–16471.

221. Xu, Y., Ishizuka, T., Kurabayashi, K., and Komiyama, M. (2009) Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: A protective capping structure for telomere ends. *Angew. Chem. Int. Ed.* **48**, 7833–7836.

222. Wang, H., Nora, G.J., Ghodke, H., and Opresko, P.L. (2011) Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. *J. Biol. Chem.* **286**, 7479–7489.

223. Petraccone, L., Spink, C., Trent, J.O., Garbett, N.C., Mekmaysy, C.S., Giancola, C., and Chaires, J.B. (2011) Structure and stability of higher-order human telomeric quadruplexes. *J. Am. Chem. Soc.* **133**, 20951–20961.

224. Zhang, N., Phan, A.T., and Patel, D.J. (2005) (3+1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Am. Chem. Soc.* **127**, 17277–17285.

225. Phan, A.T., Kuryavyi, V., Gaw, H.Y., and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human *MYC* promoter. *Nat. Chem. Biol.* **1**, 167–173.

226. Hu, L.Y., 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. *J. Am. Chem. Soc.* **131**, 16824–16831.

227. Jiang, F., Patel, D.J., Zhang, X.H., Zhao, H., and Jones, R.A. (1997) Specific labeling approaches to guanine and adenine imino and amino proton assignments in the AMP-RNA aptamer complex. *J. Biomol. NMR* **9**, 55–62.

228. Phan, A.T., Modi, Y.S., and Patel, D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human *c-myc* promoter. *J. Am. Chem. Soc.* **126**, 8710–8716.

229. Rachwal, P.A., Findlow, I.S., Werner, J.M., Brown, T., and Fox, K.R. (2007) Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res.* **35**, 4214–4222.

230. Bugaut, A. and Balasubramanian, S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **47**, 689–697.

231. Guédin, A., De Cian, A., Gros, J., Lacroix, L., and Mergny, J.L. (2008) Sequence effects in single-base loops for quadruplexes. *Biochimie* **90**, 686–696.

232. Guédin, 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 Res.* **38**, 7858–7868.

233. Lim, K.W., Lacroix, L., Yue, D.J.E., Lim, J.K.C., Lim, J.M.W., and Phan, A.T. (2010) Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J. Am. Chem. Soc.* **132**, 12331–12342.

234. Borman, S.T.U. (2009) Promoter quadruplexes. *Chem Eng News* **87**, 28–30.

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

236. Ambrus, A., Chen, D., Dai, J.X., Jones, R.A., and Yang, D.Z. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44**, 2048–2058.

237. Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S., and Neidle, S. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene. *J. Am. Chem. Soc.* **127**, 10584–10589.

238. Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S., and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry* **45**, 7854–7860.

239. Phan, A.T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Am. Chem. Soc.* **129**, 4386–4392.

240. Gunaratnam, M., Swank, S., Haider, S.M., Galesa, K., Reszka, A.P., Beltran, M., Cuenca, F., Fletcher, J.A., and Neidle, S. (2009) Targeting human gastrointestinal stromal tumor cells with a quadruplex-binding small molecule. *J. Med. Chem.* **52**, 3774–3783.

241. Hsu, S.T.D., Varnai, P., Bugaut, A., Reszka, A.P., Neidle, S., and Balasubramanian, S. (2009) A G-Rich sequence within the *c-kit* oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Am. Chem. Soc.* **131**, 13399–13409.

242. Kuryavyi, V., Phan, A.T., and Patel, D.J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human *c-kit2* promoter. *Nucleic Acids Res.* **38**, 6757–6773.

243. Wei, D.G., Parkinson, G.N., Reszka, A.P., and Neidle, S. (2012) Crystal structure of a *c-kit* promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Res.* **40**, 4691–4700.

244. Dai, J.X., Chen, D., Jones, R.A., Hurley, L.H., and Yang, D.Z. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **34**, 5133–5144.

245. Sun, D.Y., Guo, K.X., Rusche, J.J., and Hurley, L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **33**, 6070–6080.

246. De Armond, R., Wood, S., Sun, D.Y., Hurley, L.H., and Ebbinghaus, S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1α promoter. *Biochemistry* **44**, 16341–16350.

247. Qin, Y., Rezler, E.M., Gokhale, V., Sun, D., and Hurley, L.H. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the *PDGF-A* promoter and modulation of *PDGF-A* promoter activity by TMPyP4. *Nucleic Acids Res.* **35**, 7698–7713.

248. Xu, Y. and Sugiyama, H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* **34**, 949–954.

249. 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 Res.* **34**, 2536–2549.

250. Guo, K., Pourpak, A., Beetz-Rogers, K., Gokhale, V., Sun, D., and Hurley, L.H. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the *RET* oncogene. *J. Am. Chem. Soc.* **129**, 10220–10228.

251. Tong, X.T., Lan, W.X., Zhang, X., Wu, H.M., Liu, M.L., and Cao, C.Y. (2011) Solution structure of all parallel G-quadruplex formed by the oncogene *RET* promoter sequence. *Nucleic Acids Res.* **39**, 6753–6763.

252. Rachwal, P.A., Brown, T., and Fox, K.R. (2007) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.* **581**, 1657–1660.

253. Miyoshi, D., Nakao, A., and Sugimoto, N. (2002) Molecular crowding regulates the structural switch of the DNA G-quadruplex. *Biochemistry* **41**, 15017–15024.

254. Lane, A.N., Chaires, J.B., Gray, R.D., and Trent, J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36**, 5482–5515.

255. Phan, A.T. and Patel, D.J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.* **125**, 15021–15027.

256. Zimmerman, S.B. and Trach, S.O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **222**, 599–620.

257. Smith, F.W. and Feigon, J. (1993) Strand orientation in the DNA quadruplex formed from the *Oxytricha* telomere repeat oligonucleotide d($G_4T_4G_4$) in solution. *Biochemistry* **32**, 8682–8692.

258. Renčiuk, D., Kejnovská, I., Školáková, P., Bednářová, K., Motlová, J., and Vorlíčková, M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K^+ solutions. *Nucleic Acids Res.* **37**, 6625–6634.

259. Feldkamp, U. and Niemeyer, C.M. (2006) Rational design of DNA nanoarchitectures. *Angew. Chem. Int. Ed.* **45**, 1856–1876.

260. Krishnan, Y. and Simmel, F.C. (2011) Nucleic acid based molecular devices. *Angew. Chem. Int. Ed.* **50**, 3124–3156.

261. Huppert, J.L. (2008) Hunting G-quadruplexes. *Biochimie* **90**, 1140–1148.

262. Jaeger, L. and Chworos, A. (2006) The architectonics of programmable RNA and DNA nanostructures. *Curr. Opin. Struct. Biol.* **16**, 531–543.

263. Zhu, L.M., Chou, S.H., Xu, J.D., and Reid, B.R. (1995) Structure of a single-cytidine hairpin loop formed by the DNA triplet GCA. *Nat. Struct. Biol.* **2**, 1012–1017.

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

265. Balkwill, G.D., Garner, T.P., Williams, H.E.L., and Searle, M.S. (2009) Folding topology of a bimolecular DNA quadruplex containing a stable mini-hairpin motif within the diagonal loop. *J. Mol. Biol.* **385**, 1600–1615.

266. Phan, A.T. *et al.* (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* **18**, 796–804.

267. Do, N.Q. and Phan, A.T. (2012) Monomer–dimer equilibrium for the 5 '–5 ' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chem-Eur J* **18**, 14752–14759.

268. Bourdoncle, A., Torres, A.E., Gosse, C., Lacroix, L., Vekhoff, P., Le Saux, T., Jullien, L., and Mergny, J.L. (2006) Quadruplex-based molecular beacons as tunable DNA probes. *J. Am. Chem. Soc.* **128**, 11094–11105.

269. Venczel, E.A. and Sen, D. (1996) Synapsable DNA. *J. Mol. Biol.* **257**, 219–224.

270. Wu, B., Girard, F., van Buuren, B., Schleucher, J., Tessari, M., and Wijmenga, S. (2004) Global structure of a DNA three-way junction by solution NMR: Towards prediction of 3H fold. *Nucleic Acids Res.* **32**, 3228–3239.

271. Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S.M., Driver, D.A., Berg, R.H., Kim, S.K., Norden, B., and Nielsen, P.E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature* **365**, 566–568.

272. Yurke, B., Turberfield, A.J., Mills, A.P., Simmel, F.C., and Neumann, J.L. (2000) A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605–608.

273. Lim, K.W. and Phan, A.T. (2013) Structural basis of DNA quadruplex–duplex junction formation. *Angew. Chem. Int. Ed.* **52**, 8566–8569.

274. Palumbo, S.L., Ebbinghaus, S.W., and Hurley, L.H. (2009) Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc.* **131**, 10878–10891.

275. Huang, Y.C. and Sen, D. (2010) A contractile electronic switch made of DNA. *J. Am. Chem. Soc.* **132**, 2663–2671.

276. Yu, Z.B., Gaerig, V., Cui, Y.X., Kang, H.J., Gokhale, V., Zhao, Y., Hurley, L.H., and Mao, H.B. (2012) Tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. *J. Am. Chem. Soc.* **134**, 5157–5164.

277. Kumar, N. and Maiti, S. (2008) A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res.* **36**, 5610–5622.

278. Risitano, A. and Fox, K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.* **32**, 2598–2606.

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

280. Adrian, M., Heddi, B., and Phan, A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods* **57**, 11–24.

281. Vorlíčková, M., Kejnovská, I., Bednářová, K., Renčiuk, D., and Kypr, J. (2012) Circular dichroism spectroscopy of DNA: From duplexes to quadruplexes. *Chirality* **24**, 691–698.

282. Do, N.Q., Lim, K.W., Teo, M.H., Heddi, B., and Phan, A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.* **39**, 9448–9457.

283. Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Itakura, K., and Breslauer, K.J. (1982) Extra adenosine stacks into the self-complementary d(CGCAGAATTCGCG) duplex in solution. *Biochemistry* **21**, 445–451.

284. Joshua-Tor, L., Rabinovich, D., Hope, H., Frolow, F., Appella, E., and Sussman, J.L. (1988) The three-dimensional structure of a DNA duplex containing looped-out bases. *Nature* **334**, 82–84.

285. Summers, M.F., Byrd, R.A., Gallo, K.A., Samson, C.J., Zon, G., and Egan, W. (1985) Nuclear magnetic resonance and circular dichroism studies of a duplex single-stranded hairpin loop equilibrium for the oligodeoxyribonucleotide sequence d(CGCGATTCGCG). *Nucleic Acids Res.* **13**, 6375–6386.

286. Lim, K.W., Alberti, P., Guédin, A., Lacroix, L., Riou, J.F., Royle, N.J., Mergny, J.L., and Phan, A.T. (2009) Sequence variant (CTAGGG)$_n$ in the human telomere favors a G-quadruplex structure containing a G•C•G•C tetrad. *Nucleic Acids Res.* **37**, 6239–6248.

287. Yue, D.J.E., Lim, K.W., and Phan, A.T. (2011) Formation of (3+1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats. *J. Am. Chem. Soc.* **133**, 11462–11465.

288. Zhou, J., Bourdoncle, A., Rosu, F., Gabelica, V., and Mergny, J.L. (2012) Tri-G-quadruplex: Controlled assembly of a G-quadruplex structure from three G-rich strands. *Angew. Chem. Int. Ed.* **51**, 11002–11005.

289. Dai, J.X., Dexheimer, T.S., Chen, D., Carver, M., Ambrus, A., Jones, R.A., and Yang, D.Z. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* **128**, 1096–1098.

290. Amrane, S., Adrian, M., Heddi, B., Serero, A., Nicolas, A., Mergny, J.L., and Phan, A.T. (2012) Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite. *J. Am. Chem. Soc.* **134**, 5807–5816.

291. Emanuela, M., Martufi, M., Cacchione, S., De Santis, P., and Savino, M. (2010) Self-organization of G-quadruplex structures in the hTERT core promoter stabilized by polyaminic side chain perylene derivatives. *Biophys. Chem.* **153**, 43–53.

292. Kikin, O., D'Antonio, L., and Bagga, P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676–W682.

293. Rawal, P., Kummarasetti, V.B.R., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K., and Chowdhury, S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: Role in *Escherichia coli* global regulation. *Genome Res.* **16**, 644–655.

294. Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012) NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130–D135.

295. Markham, N.R. and Zuker, M. in *Bioinformatics: Data, sequence analysis and evolution* Vol. II (ed J. M. Keith) Ch. 1, 3–31 (Humana Press, Totowa, 2008).

296. Dreszer, T.R. *et al.* (2012) The UCSC Genome Browser database: Extensions and updates 2011. *Nucleic Acids Res.* **40**, D918–D923.

297. Chin, L. *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068.

298. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001) dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311.

299. Djebali, S. *et al.* (2012) Landscape of transcription in human cells. *Nature* **489**, 101–108.

300. Cantor, C.R., Warshaw, M.M., and Shapiro, H. (1970) Oligonucleotide interactions. III. Circular dichroism studies of conformation of deoxyoligonucleotides. *Biopolymers* **9**, 1059–1077.

301. DeLano, W.L. and Bromberg, S. *PyMOL User's Guide*. (DeLano Scientific LLC, 2004).

List of Publications

Amrane, S., Ang, R.W.L., Tan, Z.M., Li, C., Lim, J.K.C., Lim, J.M.W., <u>Lim, K.W.</u>, and Phan, A.T. (2009) A novel chair-type G-quadruplex formed by a *Bombyx mori* telomeric sequence. *Nucleic Acids Res.* **37**, 931–938.

<u>Lim, K.W.</u>, Amrane, S., Bouaziz, S., Xu, W.X., Mu, Y.G., Patel, D.J., Luu, K.N., and Phan, A.T. (2009) Structure of the human telomere in K^+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.* **131**, 4301–4309 (reviewed in *Faculty 1000*).

<u>Lim, K.W.</u>, Alberti, P., Guédin, A., Lacroix, L., Riou, J.F., Royle, N.J., Mergny, J.L., and Phan, A.T. (2009) Sequence variant (CTAGGG)$_n$ in the human telomere favors a G-quadruplex structure containing a G•C•G•C tetrad. *Nucleic Acids Res.* **37**, 6239–6248 (cover article).

Hu, L., <u>Lim, K.W.</u>, 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. *J. Am. Chem. Soc.,* **131**, 16824–16831 (reviewed in *Faculty 1000*).

<u>Lim, K.W.</u>, Lacroix, L., Yue, D.J.E., Lim, J.K.C., Lim, J.M.W., and Phan, A.T. (2010) Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J. Am. Chem. Soc.* **132**, 12331–12342.

Martadinata, H., Heddi, B., <u>Lim, K.W.</u>, and Phan, A.T. (2011) Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. *Biochemistry* **50**, 6455–6461.

Yue, D.J.E., <u>Lim, K.W.</u>, and Phan, A.T. (2011) Formation of (3+1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats. *J. Am. Chem. Soc.* **133**, 11462–11465 (reviewed in *Faculty 1000*).

Do, N.Q., <u>Lim, K.W.</u>, Teo, M.H., Heddi, B., and Phan, A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.* **39**, 9448–9457 (cover article).

<u>Lim, K.W.</u> and Phan, A.T. (2013) Structural basis of DNA quadruplex–duplex junction formation. *Angew. Chem. Int. Ed.* **52**, 8566–8569 (reviewed in *Faculty 1000*).

<u>Lim, K.W.</u>, Ng, V.C.M., Martin-Pintado, N., Heddi, B., and Phan, A.T. (2013) Structure of the human telomere in Na^+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.* (in press).

<u>Lim, K.W.</u>, Khong, Z.J., and Phan, A.T. (2013) Thermal stability of DNA quadruplex–duplex hybrids. *Biochemistry* (under revision).

<u>Lim, K.W.</u>, Jenjaroenpun, P., Low, Z.J., Ow, G.S., Ng, Y.S., Khong, Z.J., Kuznetsov, V.A., and Phan, A.T. (2013) Putative quadruplex motifs containing duplex stem elements in the human genome. (to be submitted).

Posters and Talks

Posters

Lim, K.W. and Phan, A.T. (2009) Structures of the telomeres. International Conference on Materials for Advanced Technology (ICMAT), Singapore, June 28–July 3, 2009.

Lim, K.W. and Phan, A.T. (2011) G-quadruplex, telomere and telomerase. Albany 2011: The 17th Conversation, Albany NY, U.S.A. June 14–18, 2011.

Talks

Lim, K.W. and Phan, A.T. (2011) G-quadruplex, telomere and telomerase. Albany 2011: The 17th Conversation, Albany NY, U.S.A. June 14–18, 2011 (Young Scientists Lecture Program).

Lim, K.W. (2013) Structural basis of DNA quadruplex–duplex junction formation. 4th International Meeting on G-Quadruplex Nucleic Acids: From Structure to Chemistry and Biology, Singapore. July 1–4, 2013.