Novel secondary structures of DNA; and development of a sensitive methodology for capturing DNA/RNA G-quadruplexes from living *Drosophila* salivary glands

by

Prince Kumar Lat

B.Tech, Indian Institute of Technology, Kanpur, India, 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY
Fall 2021

Declaration of Committee

Name:	**Prince Kumar Lat**
Degree:	**Doctor of Philosophy**
Thesis title:	**Novel secondary structures of DNA; and development of a sensitive methodology for capturing DNA/RNA G-quadruplexes from living *Drosophila* salivary glands**
Committee:	**Chair:** **Stephanie Vlachos** Lecturer, Molecular Biology and Biochemistry
	Dipankar Sen Supervisor Professor, Molecular Biology and Biochemistry
	Peter Unrau Committee Member Professor, Molecular Biology and Biochemistry
	Nicholas Harden Committee Member Professor, Molecular Biology and Biochemistry
	Mani Larijani Examiner Associate Professor, Molecular Biology and Biochemistry
	Sergei Mirkin External Examiner Professor, Biology Tufts University

Abstract

Work reported in this thesis, from three independent projects, highlights: first, a novel DNA secondary structure fold from a neurodegenerative disease-linked repeat sequence; second, a new approach for assembling and reversing a long and 1-dimensional DNA nanostructure. The third and most substantial project reports the development of and biological results from a highly selective and sensitive approach for *in vitro* and *in vivo* tagging of DNA and RNA G-quadruplexes. In the first project, a wholly novel higher-order fold of DNA, named as "iCD-DNA", was discovered and characterized. iCD-DNA was found to be formed uniquely by a hexanucleotide repeat expansion sequence, $d(C_2G_4)_n$, located at the 5' UTR of the *C9orf72* gene, causally linked to multiple neurological disorders such as Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD). It was found that incubating $d(C_2G_4)_n$ under mildly acidic conditions and in the presence of non-quadruplex supporting cations (e.g. Li^+, Mg^{2+}) gave rise to a distinctive higher order structure whose most striking feature was an inverted circular dichroism (CD) spectrum, distinguishable from the inverted CD spectra of either a left-handed duplex ("Z-DNA") or a left-handed G-Quadruplex ("Z-GQ"). On the basis of CD spectroscopy, gel mobility and chemical footprinting, structural models were proposed for iCD-DNA. In the second project, a new strategy for creating a long (~200-300 nm) and reversible 1-Dimensional DNA nanostructure/ nanowire (1DDN), named "$(TQs)_n$", was designed and carried out. "$(TQs)_n$" incorporates a hybrid of DNA triple and quadruplex helices. In this design, a novel approach for joining together DNA helices (called guanine-rich "glue junctions") was proposed and demonstrated. In the third project, a highly specific and sensitive methodology for uniquely biotin-tagging DNA/RNA G-quadruplexes (by way of their intrinsic peroxidase activity while complexed with heme) was deeply characterised, first, *in vitro,* and then applied to tag and pull down G-quadruplex forming RNAs and DNAs from living *Drosophila* larval salivary glands. Preliminary-sequencing data, so obtained, provided initial insights for the potential occurrence of G-quadruplexes in living cells but needs detailed future investigation.

Keywords: G-Quadruplexes; iCD-DNA; "$(TQs)_n$" 1-Dimensional DNA Nanostructure/Nanowire (1DDN); Heme-DNAzyme; Amyotrophic Lateral Sclerosis (ALS); DNA Nanotechnology

Dedication

I dedicate this thesis to my Brother Late ***Sumit Lat****.*

Acknowledgements

I would like to thank my Supervisor **Dr. Dipankar Sen** (Professor, Dept of MBB and Dept of Chemistry, SFU), for his continued support, guidance and suggestions during my research.

I would also like to extend my gratitude to my committee: **Dr. Peter J Unrau** (Professor, Dept of MBB, SFU) and **Dr. Nicholas Harden** (Professor, Dept of MBB, SFU) for their constructive criticism and supportive feedback.

I am extremely thankful to **Dr. Hua-Zhong Yu** (Professor, Department of Chemistry, SFU), **Dr. Ryan D Morin** (Associate Professor, Department of MBB, SFU), and **Dr. Lis Guimaraes de Azerado Melo** (4D Labs, SFU) for excellent and successful collaborations.

My sincere thanks to the funding agencies: ***NSERC***, Canada and SFU scholarship committee for supporting my financial needs with the following scholarships: (a) CD Nelson Entrance Scholarship, (b) Provost Prize of Distinction, (c) Special Grad Entrance Scholarship, (d) Weyerhaeuser Grad Scholarship, (e) David L. Baillie Grad Scholarship, (f) Graduate Fellowship (g) Thesis Completion Fellowship and (h) Dr. Bruce Brandhorst.Prize: Best publication award.

Finally, I would like to place on record ‘million thanks’, to my parents - ***Mr. Pradeep Lat & Mrs. Sudha Lat*** for their endless sacrifices; to my siblings - ***Ankit Lat, Dr. Smit Kumar*** & ***Nisha More*** for their immense love and care; to my Guru - ***Mr. Uma Shankar*** for imbibing within me the urge to live a life of simplicity; to my teachers - **Rajeev Ranjan, Kumar Saurabh, Prof. MMR Akhtar, Prof. S Ganesh, Prof. Ashwani Kumar Thakur, Prof Balaji Prakash & Prof S. Matheshwaran** for their constant motivation to do scientific research; to my friends whom I met in Vancouver - **Priyanka Kashyap** and **Akansha Vaish** for their immense support & cooperation; and to my colleagues from Dr. Sen lab – **Kun**, **Adam**, **Nisreen**, **Owen**, **Hanadi**, **Hamza, Carys** and **Fiona** for making such a touch journey reach its destination with endless beautiful memories.

Table of Contents

List of Tables

List of Figures

List of Acronyms

1DDN	1-Dimensional DNA Nanostructure
$4EM^+$	4-Ethyl Morpholinium$^+$
AFM	Atomic Force Microscopy
ALS/FTD	Amyotrophic Lateral Sclerosis / Frontotemporal Dementia
BT	Biotin Tyramide
CD	Circular Dichroism
CHE	Chelerythrine
COSY	Correlated Spectroscopy
DMS	Dimethyl Sulphate
EDTA	Ethylenediaminetetraacetic acid
GQ/GQs	G-Quadruplex / G-Quadruplexes
LC-ESI-MS	Liquid Chromatography-Electrospray Ionization-Mass Spectroscopy
NMM	N-methylmesoporphyrin IX
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
StAv/MAv	Streptavidin / Monovidin
TEM	Transmission Electron Microscopy
T_m	Melting Point
UTR	Untranslated Region

Glossary

Abzymes	Catalytic Antibodies, also known as catmab (“Catalytic Monoclonal Antibodies”)
Helicase	Protein enzymes that unpack DNA/RNA secondary structures.
iCD-DNA	A newly discovered secondary structure of $d(C_2G_4)_n$ sequence marked by an inversion in CD Spectra .
Nanowire/1DDN	A purely DNA-composed 1-Dimensional long nanostructure of substantial length (>100 nm).
Nuclease	Protein enzymes that cleave DNA/RNA strand into smaller/individual nucleotides/nucleosides pieces.
Rhau56	GQ binding domain of the RHAU helicase..
SELEX	A combinatory technique of finding ssDNA or RNA sequence to either recognise a particular target or catalyze a reaction.
TQ/TQ*s*	A non-sticky (TQ) and a purely G-quartet based sticky ended (TQ*s*) Triplex-Quadruplex Hybrid Tile

Preface

This thesis is an expanded version of the three independent research projects targeted at exploring different structural and functional aspects of nucleic acids originally published in (a) **Lat *et al***, $(C_2G_4)_n$ repeat expansion sequences from the *C9orf72* gene forms an unusual DNA higher order structure in the pH range of 5-6, *PLoS ONE,* (2018), 13(6): e0198418 (Chapter 2); (b) **Lat *et al***, A long and reversibly self-assembling 1D DNA nanostructure built from Triplex and Quadruplex hybrid tiles, *Angew Chemie Int Ed*., (2021), 60, 16, 8722-8727 (Chapter 3); and (c) **Lat *et al***, High specificity and tight spatial restriction of self-biotinylation by DNA and RNA G-Quadruplexes complexed *in vitro* and *in vivo* with Heme, Nucleic Acids Research, (2020), 48, 10, 5254-5267 (Chapter 4). Under the supervision of Dr. Dipankar Sen, I led all the three projects. Furthermore, I was also involved in training undergraduate and master students as well as in general management of the Sen lab.

Chapter 3, in particular, which talks about the rational designing of a hybrid DNA nanostructure, gave me a wonderful experience. It helped me grow as an independent researcher. I did some literature survey and came up with an idea of combining the structural properties of DNA triplexes and G-Quadruplexes to conceive of a novel way to build easily reversible 1-Dimensional DNA nanostructure/nanowire. I am thankful to Dr. Sen who after listening to my idea believed in me and allowed me to carry this project.

The work related to *Drosophila larvae* in the section 4.3.9 of Chapter 4 was done in collaboration with the Verheyen laboratory (Dr. Esther Verheyen, Department of MBB, SFU) and all the tissue samples were obtained from them. The RNA seq library preparation and preliminary sequencing data reported in section 4.3.10 of Chapter 4 were obtained from Sequencing and Bioinformatics Consortium (SBC), UBC, Canada.

Chapter 1.

Introduction

1.1. Overview

The research reported in this thesis can be broadly divided into two categories: First, exploration of novel secondary structures of DNA-- either pertaining to biological significance or for nanotechnological applications. And, second, development and utilization of a highly specific and sensitive methodology for tagging and capture, both *in vitro* and *in vivo*, of G-Quadruplex-folded DNAs and RNAs. The novel structures explored in the first part of this thesis (Chapter 2 and 3) have been formed by DNA sequences obtained synthetically. For the second part of the thesis (Chapter 4), standard quadruplex-forming sequences and oligonucleotides were used for method development and characterization, followed by the successful adoption of the method for use *in vivo*. Chapter 2 of this thesis describes the discovery and characterization of a completely new secondary fold of DNA, named as 'iCD-DNA', which is formed by the hexanucleotide repeat expansion sequence of $d(C_2G_4)_n$, from the gene *C9orf72* and linked with neurodegenerative disorders such as Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD). Chapter 3 describes a new strategy for using rationally designed triplex-quadruplex hybrid DNA "tiles" "TQ*s*", which can self-assemble to form a long and reversible 1-Dimensional DNA Nanostructure/Nanowire ("TQs")$_n$. I anticipate that such a nanostructure will find a variety of biochemical and nanotechnological applications. Chapter 4 focus to describe a reaction involving heme and G-quadruplex interaction for tagging both synthetically prepared and naturally occurring G-quadruplex sequences in genome.

The present chapter (Chapter 1) provides a comprehensive and relevant background information related to diverse structural and functional aspects of DNA. Specifically, it highlights non-canonical secondary folds of DNA and RNA, their biological significance, and their application in the field of nanotechnology. There is also a discussion of a specific catalytic (intrinsic peroxidase) property of DNA and RNA G-Quadruplexes (GQs) upon their complexation with ferric heme; a property that will form

the basis of a novel experimental approach to covalently tag and capture G-quadruplex-folded DNAs and RNAs, both *in vitro* and *in vivo* (described in detail in chapter 4).

1.2. DNA, RNA and Proteins

DNA, RNA and Proteins are the three most important biopolymers that regulate cellular machineries. As per the **Central Dogma** proposed by Francis Crick in 1958, DNA stores the genetic information. The Information encoded in DNA is then transcribed into RNA, which is finally translated into proteins. The monomeric unit of DNA/RNA polymer, a "nucleotide", is composed of a (i) negatively charged phosphate backbone, (ii) a 5-ring, 5-carbon sugar (furanose) and (iii) one out of several possible nitrogenous nucleobases (**Figure 1-1.a**). In the case of RNA, the pentose sugar is ribose whereas in the case of DNA, it is 2'-deoxyribose. There are essentially four naturally occurring nitrogenous bases both for DNA and RNA: cytosine (C), guanine (G) and adenine (A) (**Figure 1-1.b**) being common to both DNA & RNA, while the fourth base differs between DNA and RNA. While DNA has thymine (T), RNA has uracil (U) (**Figure 1-1.b**). These 4 nucleobases are either substituted purines or pyrimidines. Adenine and guanine each contain fused 5- and 6-membered purine heterocycles, while cytosine and thymine/uracil are substituted 6-membered pyrimidine heterocycles. While these are the only four bases strictly coding genetic information, DNA/RNA with chemically modified bases are found for in specialised niches. A few widely known "modified" bases are 8-oxoguanine, isoguanine, 5-methyl cytosine, xanthine, hypoxanthine, etc. (**Figure 1-1.c**). Chemical DNA/RNA synthesis or the replication/transcription of DNA/RNA, respectively, from a DNA template, results in a linear string of nucleotides that links consecutive sugars via phosphodiester linkages, while the nucleobases are linked by nitrogen to a sugar carbon via N-Glycosidic bonds (**Figure 1-1.d**).

Proteins, on the other hand, are very different from DNA/RNA in their chemical composition. The fundamental unit of a protein/peptide is the amino acid, whose structure consists of a central (alpha) carbon linked to a basic amino group ($-NH_2$), an acidic carboxylic acid group (-COOH) and an organic R group, which is referred to as side chain. The remaining one bond of the alpha carbon is satisfied by a hydrogen (H)

atom (**Figure 1-2.a**). Unlike DNA/RNA, which has only four different naturally occurring nitrogenous bases, proteins have 20 different side chains (**Figure 1-2.b**). Adjacent amino acids are linked through amide/peptide bonds, formed when the carboxylic acid of one amino acid condenses the amino group of another, releasing a water molecule (**Figure 1-3**).

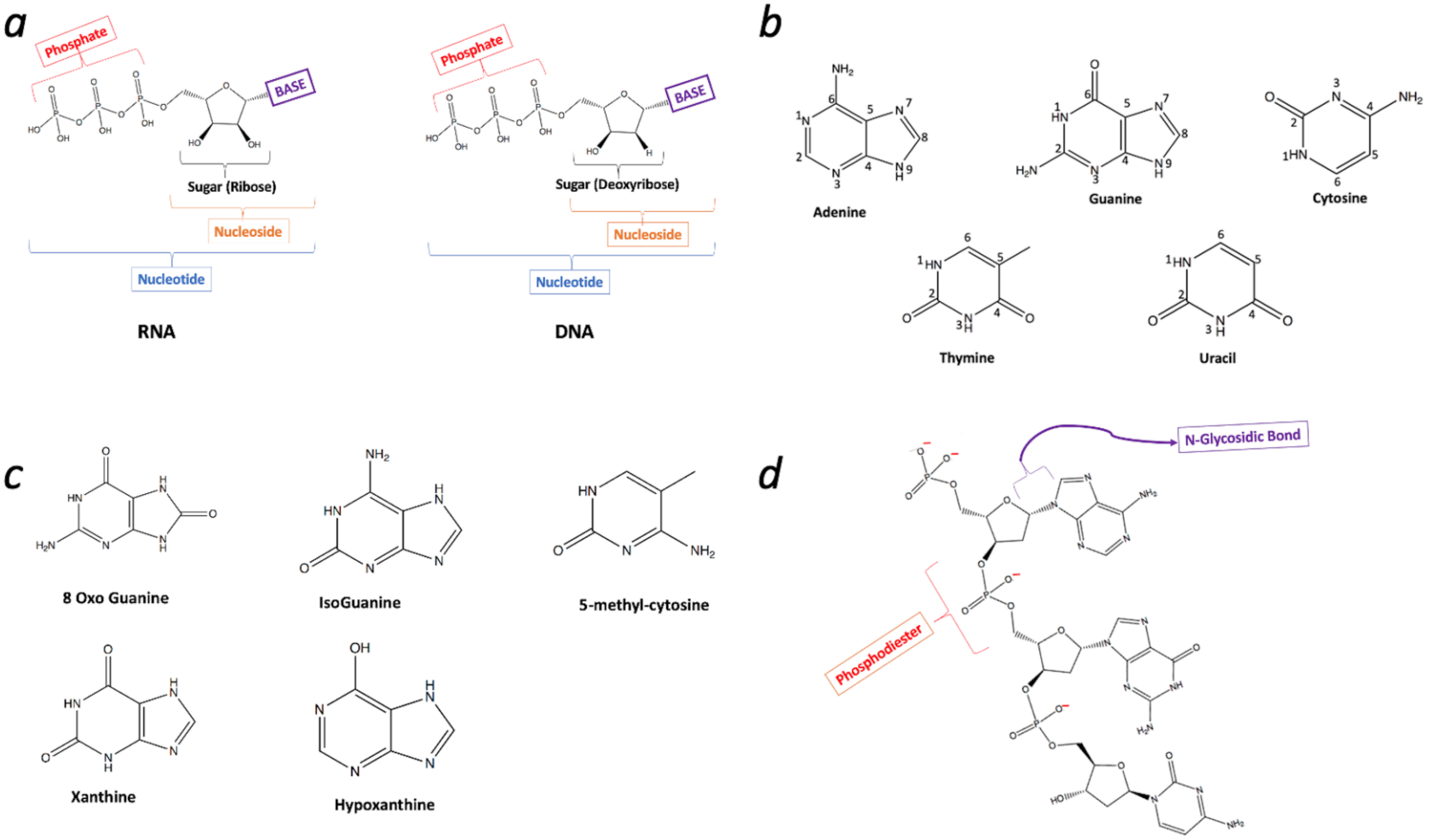


Figure 1-1. **Structure of DNA/RNA.** ***(a)*** Individual components of nucleic acids. **(b)** Naturally occurring nucleobases in DNA & RNA. **(c)** Other chemically modified nucleobase. **(d)** Full-length **DNA** polymer.

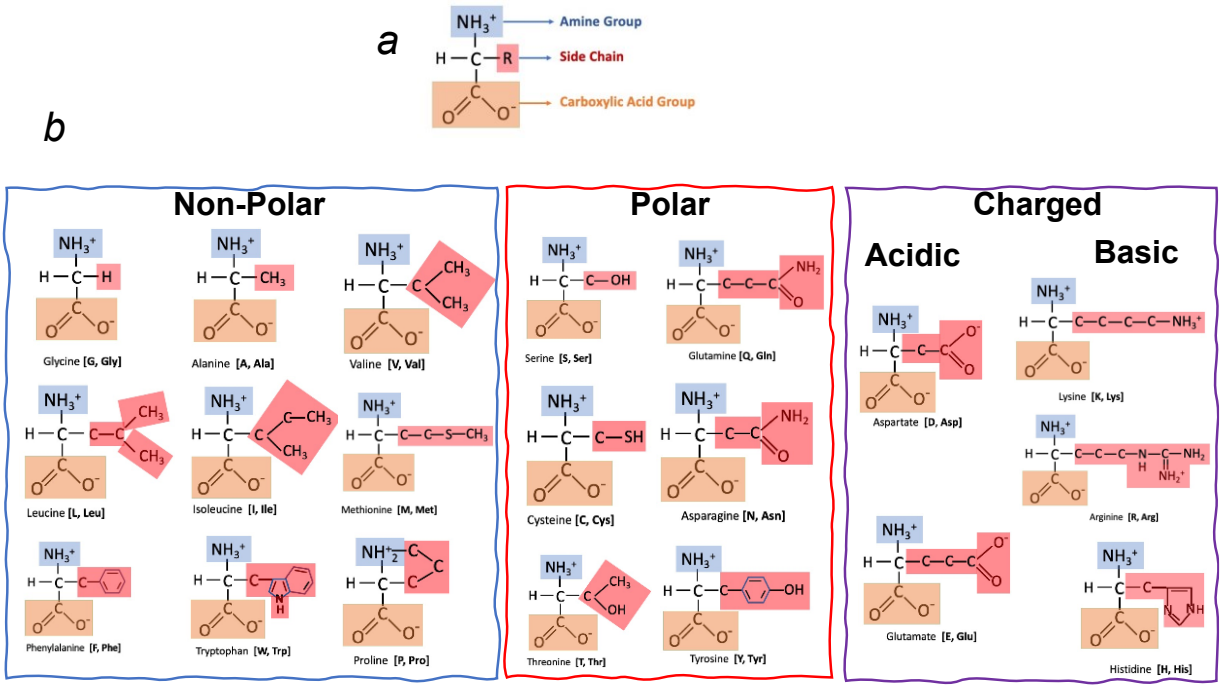


Figure 1-2. Amino Acid composition. **(a)** Generic structure of an amino acid. **(b)** the 20 naturally occurring side chains of amino acids found in proteins.

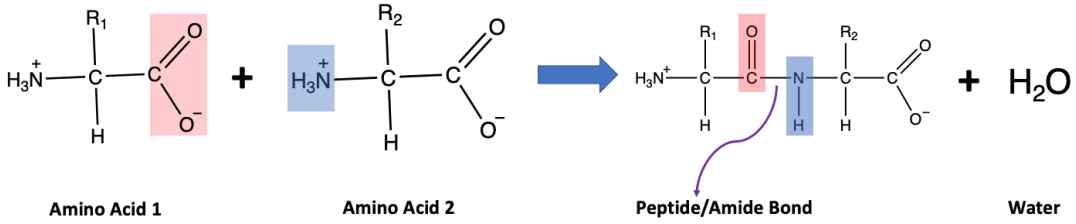


Figure 1-3. Peptide linkage to link subsequent amino acids resulting in peptide/protein formation.

The availability of diverse side chain makes protein an ideal candidate for the enzymatic activities required within living cells. This is a compelling reason why nature has evolved proteins for catalytic/enzymatic activities. DNA/RNA, on the other hand have less functional group diversity, containing only of 5 nucleobases between DNA and RNA. With increasing understanding of DNA and RNA, many non-canonical roles of nucleic acids within cells and beyond have been discovered. Non-genomic DNA/RNA sequences have been selected, through a well-established protocol called “SELEX” or

"*in vitro* selection" (see section 1.9), to act either as an "aptamers" or as "DNAzymes and ribozymes"[1]. An aptamer is a short fragment of DNA/RNA that, in its folded form, binds to a given specific molecular target with high selectivity and specificity. A DNAzyme or ribozyme, on the other hand, not only binds to a target but is also able to catalyze chemical reactions, in a manner paralleling the activity of protein enzymes.

Owing to its low cost of synthesis, a high level of chemical stability, and ease of self-assembly into complex structures, DNA has also been widely exploited *in vitro* for the creation of electric switches[2], logic gates[3], nanowires[4], reversible hydrogels[5], DNA origami[6] and other self-assembled structures suitable for nanotechnological applications.

1.3. Canonical secondary structures of Nucleic Acids

Secondary structures of DNA/RNA have an immense importance for the realization of their functional attributes. Inside the cell, DNA mainly exists as a double helix whereas RNA is transcribed as single stranded. However, RNA can also form double helices, internally/intramolecularly, or with a complementary RNA or DNA strand. There are specific rules based on which two strands of DNA/RNA come together to give rise to duplex "double helix" formation. Given that DNA/RNA strands consist of negatively charged phosphate backbones (with one full negative charge at neutral pH per nucleotide unit), individual DNA/RNA strands in aqueous solution without salt electrostatically repel each other. The self-assembly required for the formation of a double helix (or any other secondary fold) requires salts to counter (effectively partially neutralize) the negative charge of the phosphate backbone (**Figure 1-4**.**a**). Other crucial forces that enable the interaction between two strands of DNA are Van der Waals force (stacking- interactions between consecutive bases), hydrophobic interaction (between the hydrophobic nitrogenous nucleobases and surrounding sheath of water) and, importantly, hydrogen bonding between purines and pyrimidine bases located in two different (inter-) or same (intra-) strands. When two DNA strands come together to form a duplex, they orient themselves in antiparallel fashion, i.e., the sugar-phosphate linkage proceeds in mutually opposite direction on two complementary strands forming the duplex. In the Watson-Crick scheme of base-pairing, found in cellular duplex DNA and to a large extent in folded RNAs, 'A' pairs with 'T'/'U' and 'C' pairs with 'G' via hydrogen

bonding from their Watson-crick faces, as shown in **Figure 1-4**.**b**. The anti-parallel arrangement of the two continuous phosphate strands give rise to two unequal grooves in the duplex helix: the major groove and the minor groove **Figure 1-4.c**. The major and minor groove runs throughout the length of DNA, with the major groove being approximately 50 % wider than the minor groove. The specific features and functional groups exposed in the minor and major grooves of DNA play an essential role in determining specific interactions with proteins, small molecules, or even other DNA/RNA sequences.

RNA helices (of the 'A' family) are, on the whole, more structurally conserved relative to DNA helices (of the 'B' family under physiological conditions); the latter are structurally more polymorphic. The location of the bases with respect to the sugar (linked by the N-glycosidic bond) enables two alternative base orientations: *syn* or *anti* (**Figure 1-4.d**). The *syn* conformation, in general, involves more steric hindrance between the base and its appended sugar than the *anti*-conformation. This is a primary reason why nucleosides in both DNA and RNA duplexes exist overwhelmingly in the *anti*-conformation, with a few exceptions[7].

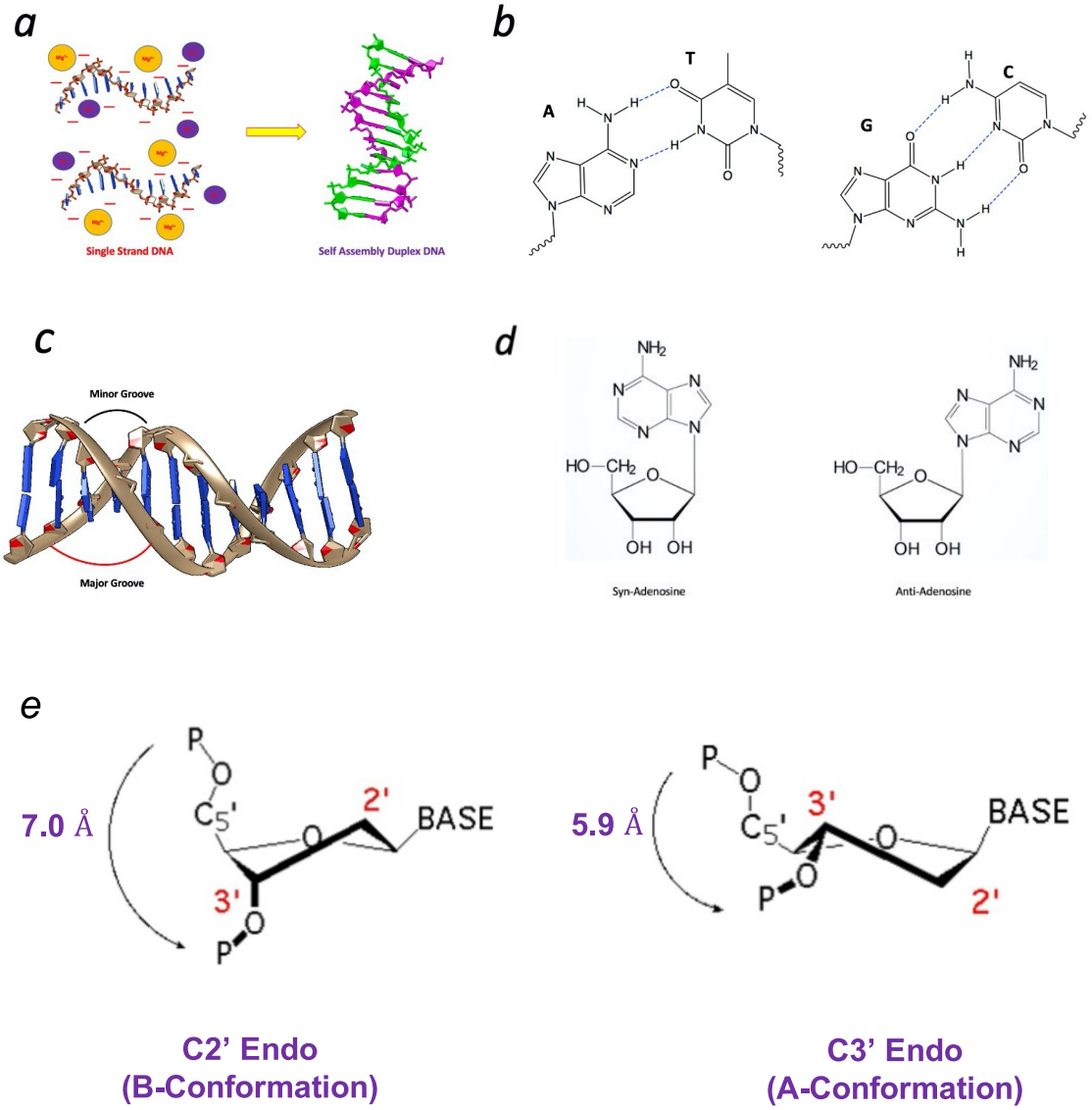


Figure 1-4. Watson Crick base pairing and characteristic features of duplex DNA. **(a)** Monovalent and divalent salts reduce the negative (-) phosphate backbone and propagates duplex formation. **(b)** Canonical Watson-Crick base pairing rules, A pairing with T/U and G pairing with C. **(c)** Standard B-type duplex showing major and minor groove. **(d)** Two different base orientations 'Syn' and 'Anti' with respect to glycosidic bond. **(e)** Two main conformation of sugar pucker found in nucleic acids.

With respect to the class of double-helix DNA and RNA adopt, helices have been divided into three broad categories as follow:

1.) the A Family: A, A'

2.) the B family: B, B', C, C', C'', D, E

3.) the Z family.

Of all these classes and subclasses, the three most important and physiologically occurring forms of helix are: the A type, the B type and the Z type (**Figure 1-5**). While RNA strictly forms A type helix, DNA is more flexible and can adopt all of the helix types listed above (although mostly favoring the B type helix under physiological conditions). The key difference driving the formation of an A type versus a B type helix is the sugar pucker found in individual nucleotides within those helices. To minimize non-bonded interaction between the substituents, the furanose ring (sugar) is twisted out of the plane. So, puckering occurs mainly when carbons C2' and C3' of the sugar ring show major displacement from the median plane defined by C1'-O4-C4'. In an A helix, the *endo* displacement of C-3' is greater than the *exo* displacement of C-2'; hence, a sugar residue in an A helix is labeled as C-3' *endo* whereas for a sugar in a B helix, the *endo* displacement of C-2' exceeds that of the *exo* displacement of C-3' and the sugar pucker is therefore referred to as C-2' *endo* (**Figure 1-4.e**). A summary of the structural similarities and differences between A type and B type helices is shown in **Table 1-1**.

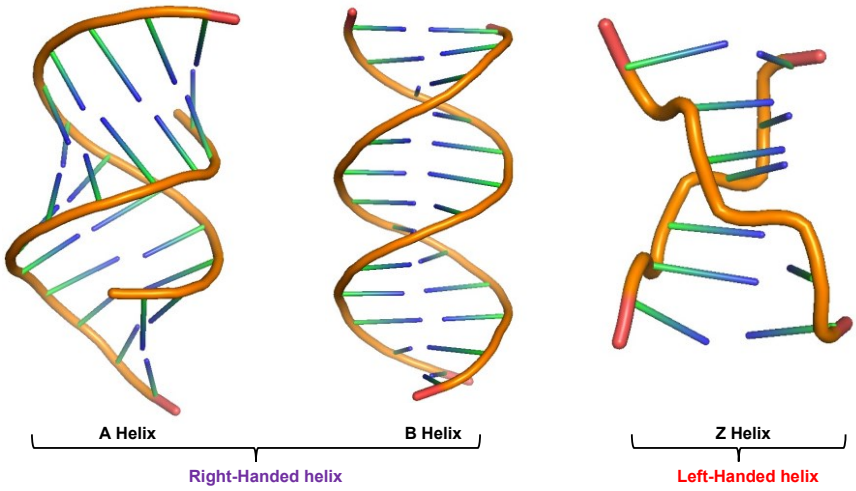


Figure 1-5. Three naturally occurring forms of double helix. (Left) 'A' Helix (PDB ID: 4jrt); (Middle) B Helix (PDB ID: 1lai); (Right) Z Helix (PDB ID: 4e4o). Both A and B helix are right-handed whereas Z-Helix is left-handed.

Table 1-1 Similarities and differences in A and B type Helix. A summary of A and B type helix, depicting their similarities and differences based on H-bond, strand orientation, sugar conformation and major and minor groove features.

Similarities		**Differences**	
B Type	**A type**	**B Type**	**A Type**
1. Hydrogen bond via Watson-Crick face 2. Strands run anti-parallel 3. anti-glycosidic conformation	1. Hydrogen bond via Watson-Crick face 2. Strands run anti-parallel 3. anti-Glycosidic conformation	1. C2'-endo sugar pucker 2. Deep, wide, major groove 3. Wider, deeper minor groove 4. Greater helix structural heterogeneity	1. C3'-*endo* sugar pucker 2. Narrower, shallower major groove 3. Smaller, shallower minor groove 4. Helices structurally homogenous

A third, prominent family of DNA duplexes is the **Z** helix. Z-DNA is a left-handed double helix (**Figure 1-6**) as opposed to the other helix families, which are right-handed. The Z-DNA sugar phosphate backbone runs in a zig-zag pattern because of the alternating syn-anti glycosidic bond conformation of the adjacent base pairs as opposed to the smooth trajectories found in A and B helices where all the nucleobases have an anti-glycosidic bond conformation. Z-DNA formation requires a very specific DNA sequence composition. Studies have shown that continuous strands of repetitive $d(CG)_n$ sequence switches from the B to the Z helix under very high salt concentration e.g. 4M NaCl[8,9,10]. Alternatively, divalent cations like Mg^{2+} (at low concentrations ~ 0.5 mM – 4 mM) have been found to shift equilibrium from B to Z helix on either a covalently modified DNA (m^5 dC-dG)[11] or on a non-modified DNA subjected to topologically constrained conditions[12]. The formation of Z-DNA *in vivo* has been demonstrated via various indirect methods; first in prokaryotes and then in eukaryotes[9]. The most persuasive evidence of Z-DNA came from the study on human genomes themselves, mainly in the context of auto-immune disorders namely lupus erythematosus. Patients with lupus erythematosus were found to produce antibodies which were highly specific

for Z helix DNA[13]. Recently, Shin *et al* used a Z-DNA specfic antibody coupled with chip-seq approach to identify Z-DNA forming sites (ZFS) in human genome and found 391 ZFS sites, majority of which occurred in the actively transcribed region[14].

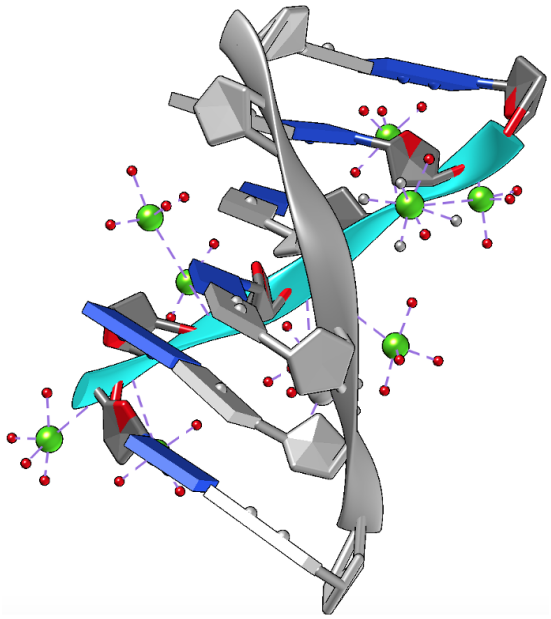


Figure 1-6. Left-handed DNA duplex Z-DNA (PDB ID: 4FS5). A 3D depiction of ZDNA, showing left-handed propagation of helix.

Although in all the standard double helix (A, B and Z), the pairing between the complementary strands of DNA/RNA happens via the mutual Watson-Crick edge interaction of the bases, however, the canonical Watson-Crick interaction is not the only possible edge-to-edge interaction that can occur between the DNA/RNA nucleobases (**Figure 1-7**). Advancement in the field of crystallography and NMR has shown evidence for non-Watson-Crick based hydrogen bond interactions, mainly Hoogsteen-hydrogen bonding, that yields several non-canonical secondary structures of DNA/RNA as discussed in the next section.

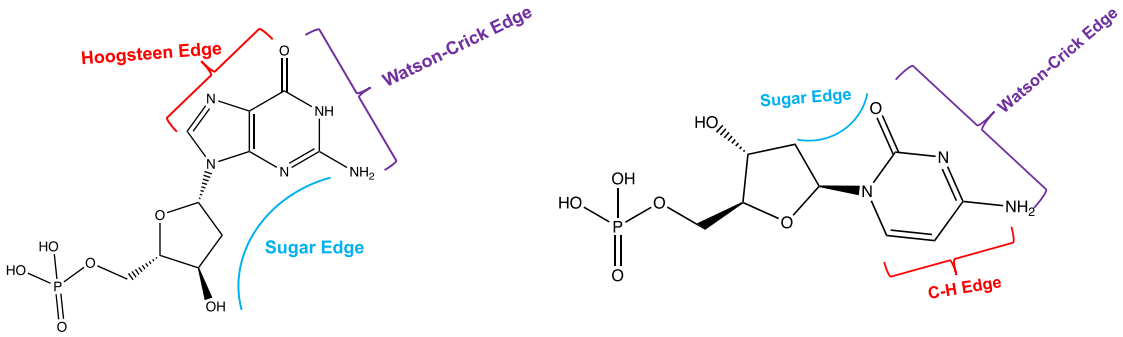


Figure 1-7. Different edges containing sites of H-bond acceptor and donor. Different edges (capable of forming H-bond in various combinations) of purine (left) and pyrimidine (right) have been shown. These can interact in different base pairing geometries to give rise to non-canonical secondary structures. Some of the most prominent ones have been discussed in Figure 1-8 of section 1.4.

1.4. Other secondary structures of DNA/RNA

DNA is often considered synonymous with the double-helix, a conformation governed by base pairing via specific hydrogen bonding between adenine (A) and thymine (T), and between guanine (G) and cytosine (C). The succession of these base pairs, giving rise to the overall double helical structure, was proposed by Watson and Crick in 1953[15]. However, other helical folds of DNA and RNA have also been recognized, in which DNA/RNA takes advantage of non-canonical (i.e., non-Watson-Crick) base pairing schemes. Three most important non-canonical secondary structures formed by DNA and RNA are triplex, i-motif and G-Quadruplex (GQ).

1.4.1. Triplexes

Early observations that the nucleic acids can form triple helices (triplexes) were reported by Felsenfeld almost 4 years after the discovery of double helix[16]. A strand of polyriboadenylic acid “poly(rA)” was found to form a three-stranded structure with two separate strands of polyribouridylic acid “poly(rU)” in the presence of Mg^{2+}. However, it took approximately 30 years since the initial observation of triplex formation for a fuller realization of its potential applications. It is now well established that DNA/RNA can form two classes of triplexes, namely: YR*Y and YR*R[17], with the third strand (bound in the major groove of the host duplex) being either all-pyrimidine (in YR*Y) or all-purine/ mixed pyrimidine-purine (in YR*R). The formation of triplexes requires the host duplex

to have its Watson and Crick strands to be wholly poly-purine and poly-pyrimidine, respectively. Such a duplex can then accommodate a third strand, known as 'Triplex Forming Unit (TFO)', in its major groove via non-Watson Crick hydrogen bonding known as 'Hoogsteen hydrogen bonding' in the case of YR*Y triplexes (**Figure 1-8**.**a**) (in the case of YR*R triplexes, the base-pairing involves Reverse Hoogsteen hydrogen bonding). The TFO strand can either be wholly poly-purine (for the YR*R triplex; involving canonical base triples that are CG*G and TA*A) or wholly poly-pyrimidine (for the YR*Y triplex; where the canonical base triples are CG^*C^+ and TA*T). YR*R class of triplex are more flexible and versatile. Although, initially it was felt that TFO of the YR*R class has to be completely poly-purine[18,19], later work showed it can be a mixture of both poly-purine and poly-pyrimidine. It was observed that TA*T triad may also be incorporated into YR*R triplex[20]. The formation of both, YR*R and YR*Y triplexes, requires the presence of divalent/multivalent cations (e.g., Mg^{2+}, Ca^{2+}, Mn^{2+}, Co^{2+}, Zn^{2+}, Cd^{2+} as well as polyamines like spermine (+4) and spermidine (+3)) in order to reduce the electrostatic repulsion of the phosphate backbone[17]. YR*Y triplexes (containing CG^*C^+ base triples), in addition to the divalent/multivalent cations, also require a low pH for its stabilization[17].

Depending upon the strand molecularity, triplexes are divided into two groups. These group include intermolecular triplex and intramolecular triplex (also known as H-DNA/*H-DNA). In intermolecular triplex, the TFO comes as a separate stand and is not a part of the duplex forming DNA strands whereas in case of intramolecular triplex, TFO is provided by one of the strands of the same duplex DNA molecule at a mirror repeat sequence. If the TFO of the intramolecular triplex is pyrimidine-rich, it is called H-DNA and if it is purine-rich, it is called *H-DNA. Given the strict requirement of specific sequence composition of a duplex DNA in triplex formation, Gaddis *et al* have designed computational algorithm to map TFO binding sites in mammalian genome and found ~2 million such sites both in human and mouse genomes, mostly enriched in promoter/transcribed gene region[21]. A high abundance of naturally occurring sequence capable of adopting H-DNA structure in human genome[22] (~1 in 50,000 bp) led to the investigation of such sites in an *in vivo* condition. A combination of "Hybridization probing"[23] (using fluorescent probes complementary to the single strand region exposed in an H-DNA) and "triplex-specific antibody"[23,24] approach provided indirect evidences for the occurrence of H-DNA, *in-vivo*. Since these experiments were not performed under a

completely physiological condition, it is hard to say whether H-DNA exist in a natural cellular milieu. Nonetheless, triplexes have found very attractive potential applications. Triplexes have been shown to be effective in modulating gene expression[25], directing site-specific DNA damage[26], targeting genes for molecular-targeted therapeutics[27] as well as for several nanotechnological applications[28]. One key pattern common in both the classes of triplexes (YR*Y and YR*R) is that the triplex forming third strand's (TFO's) base composition is same as one of the Watson/Crick strand of the host duplex; with the two strands of the same base composition oriented anti-parallel to each other. For e.g., in YR*Y, both the 'Y' strands have the same nucleobase composition but are oriented antiparallel with respect to each other. Similarly, in the YR*R, both the R strands have the same base composition and reverse strand polarity of the DNA sugar-phosphate backbone. This strict rule for the formation of triplexes rules out the possibility for an R-loop to either form or switch to a canonical triplex. R-loop is also a three-stranded structure of nucleic acid which is formed when an RNA strand invades a double-stranded DNA. It produces a DNA:RNA hybrid duplex and an associated non-template single strand DNA. The non-template single-stranded DNA in the R-loop happens to have the same strand polarity (together with the same base composition) as of the invading RNA strand. Thus, unlike canonical triplexes, DNA:RNA hybrid duplex, in an R-loop, doesn't form any hydrogen bonding with the associated single stranded non-template DNA. In contrast, the associated single stranded DNA in an R loop, when rich in guanine, folds into an intramolecular four-stranded secondary structure of nucleic acids called G-Quadruplexes (section 1.4.3). The formation of quadruplex further stabilizes the R-loop and has been shown to regulate the cellular machinery of transcription[29].

Other than the above two groups of canonical triplexes (YR*Y and YR*R), there is a third class of non-canonical triplex called a 'braided triplex'. Braided triplex formation occurs as a result of strand swapping between the Watson-crick and Hoogsteen hydrogen bond forming partner strands in a triplex as the triplex propagates. The concept of braided triplex has not been explored extensively. Initial observations about the formation of a braided triplex was reported in λ-phage DNA, as investigated by Atomic Force Microscopy by Linjing et al. in 1998[30]. The rules for the formation of braided triplex have not been fully defined; however, one factor anticipated to play a key

role in braided triplex formation is a high symmetricity and repetitive nature of specific DNA/RNA sequences.

1.4.2. Cytosine tetraplexes (the i-motif)

The cytosine tetraplexes or i-motif is a four-stranded secondary structure formed by DNA sequences rich in cytosines, at slightly acidic pH. The first description of an i-motif came from an NMR structure solved by Gehring K. et. al in 1993[31]. Cytosines at low pH undergo protonation and are then competent to form Hoogsteen hydrogen bonding with an unprotonated cytosine partner. Two pairs of strands, each consisting of C^+-C bonding, interdigitate with each other in anti-parallel fashion, giving rise to the final tetrameric structure (**Figure 1-8.b**). The intercalation of two duplexes to form i-motif gives rise to two minor and two major grooves. In terms of possible *in vivo* significance, the requirement for a pH < 7 was initially seen as a limitation for the i-motif to exist in a cellular environment. However, very recently in 2018, an antibody mediated approach was used to show the occurrence of i-motif *in vivo* by Zeraati et al[32].

The formation of i-motif at neutral pH, in particular, requires special environmental conditions[33]. These include low temperature coupled with either negative super helicity[34], molecular crowding [35], silico nanochannels[36], presence of silver[37] or presence of copper (I) cations[38]. Also, chemical modifications like 2'-deoxy-2'fluoro-arabinocytidine have been found to induce the formation of DNA i-motif at neutral pH[39,40].

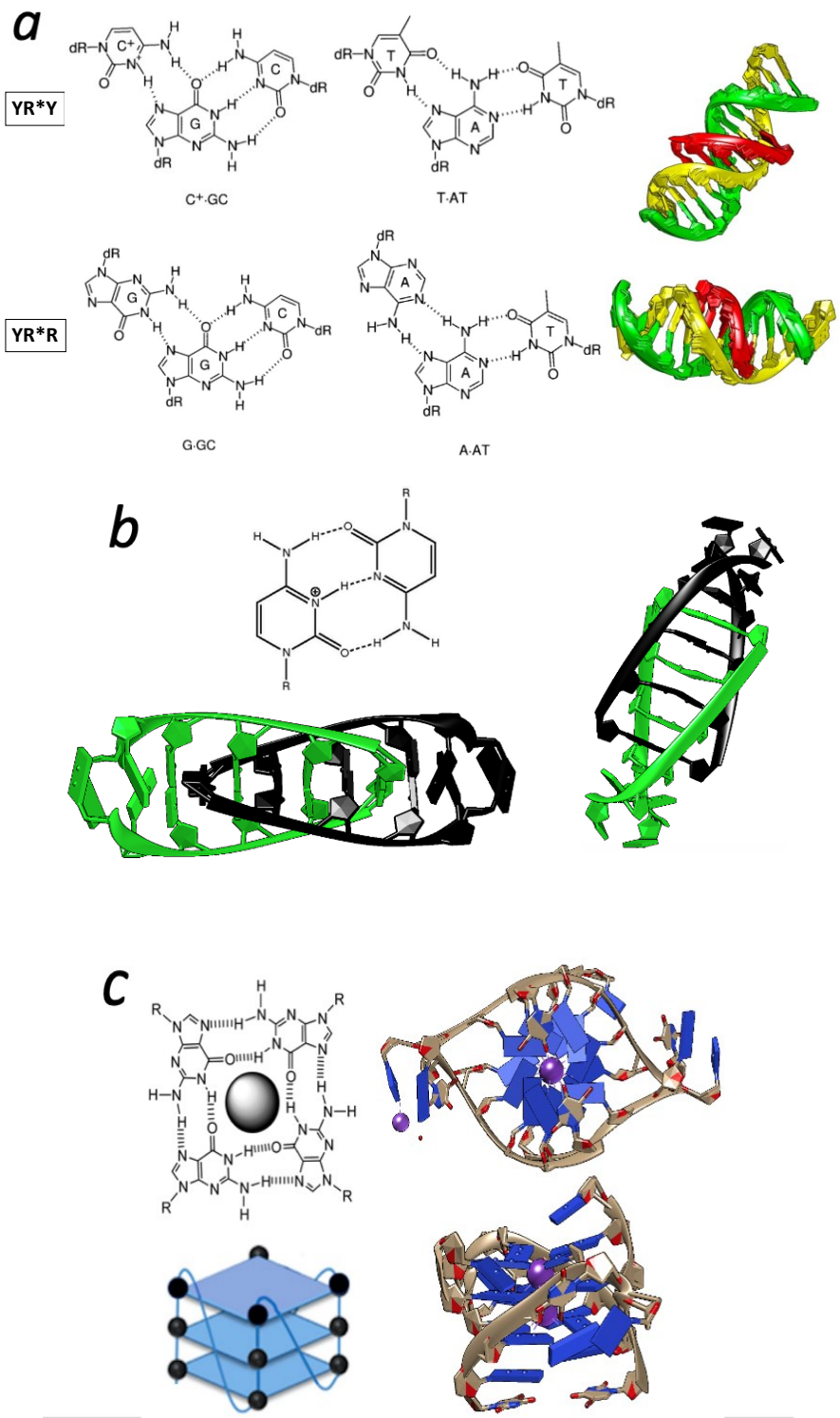


Figure 1-8. Non-canonical secondary structures of DNA/RNA. **(a) Triplex, Left**: Base pairing pattern of YR*Y {CG*C$^+$, TA*T} and YR*R {G*GC, A*AT}; **Right**: two different 3D views of triplex (PDB ID: 1BWG). **(b) i-motif, Left-Top**: Base pairing pattern of C$^+$- C; **Left-Bottom and Right:** two different 3D views of i-motif (PDB ID: 1YBI). **(c) Left-Top:** Base pairing pattern of G-Quartet; **Left-Bottom:** Schematic of 3-layer quartet; **Right:** two different 3D views of 3-layered G-Quartet (PDB ID: 1K8P).

1.4.3. G-Quadruplexes (GQs)

G-quadruplexes (GQs) are also a four-stranded structure of DNA and RNA, formed by G-rich oligonucleotides. The G-quadruplex nucleates via formation of a quartet by 4 guanines involving Hoogsteen hydrogen bonding. 2-4 layers of quartet then stack on top of each other by the virtue of π-stacking and in the presence of specific cations such as K^+ and Na^+, to form G-quadruplexes (**Figure 1-8.c**). G-quadruplexes are highly stable at physiological ion and pH conditions. The discovery of G-quartets dates to 1962[41]. Guanylic acid was found to form a viscous gel in the presence of Na^+. Using X-ray diffraction, a base-quartet model was proposed for these viscous gels. In 1988, Sen and Gilbert reported the formation of DNA G-quadruplexes forming *in vitro* from synthetic oligonucleotides derived from the G-rich DNA strand of mammalian immunoglobulin switch region responsible for class-switching recombination in B cells[42]. The next section of this chapter will discuss in detail the different conformations, topologies, biological and other significance of G-quadruplexes.

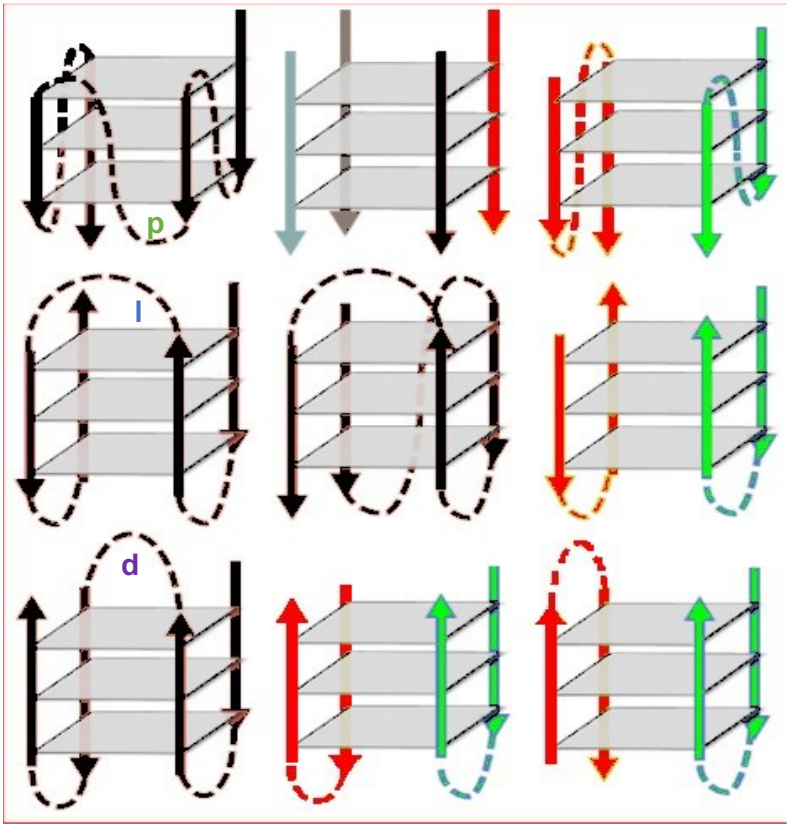


Figure 1-9. Multiple conformations (Polymorphism) of G-Quadruplexes. Parallel, Anti-parallel and Hybrid GQs. Strand Orientation: Unimolecular, Tetramolecular and Bi-molecular from left to right respectively. Arrows in the diagram represent the G-Quadruplexes strand polarity from 5' to 3'. Three different kinds of loops have been shown in left panel; **p**: propeller, **l**: lateral and **d**: diagonal.

1.5. G-quadruplexes: structural and functional relevance *in vitro* and *in vivo*

Pertaining to its high stability at physiological conditions, GQs have remained an area of extensive research since their discovery. Both DNA and RNA rich in guanines have been shown to form GQs *in vitro*[43,44]. In addition to every stabilizing factor required by DNA duplexes including base stacking, hydrogen bonding and electrostatic interaction, GQs have a specialised requirement for selective cations for its formation and stability. The formation of each layer of G-tetrad leads to a planar arrangement of O6 atoms (from guanine O6 carbonyl group) with a twist of 30^{o} and a rise of 3.3 A^{o} between each adjacent tetrad step within a quadruplex. This leads to the accumulation of a negative dipole cavity at the core of quartet and needs stabilization by coordination of metal cations. Thus, the size and charge of the cations has a large impact on the formation and stability of GQs. The general trend for the alkali ions from most stabilising to least is as follows: $K^{+} > Na^{+} > Rb^{+} > NH4^{+} > Cs^{+} > Li^{+}$. For divalent cations, low concentration initially stabilises quadruplex but starts destabilising at higher concentration. Venczel and Sen have reported the following trend of GQ stabilization by divalent cation[45] from most stabilizing to least as follows: $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$.

GQs, as a class, are highly polymorphic and can adopt a multitude of conformations and strand topologies, as shown in **Figure 1-9**. With regard to the directionality of the adjacent strands, GQs are classified as either parallel (all GQ forming strands point the same direction), anti-parallel (two strands run in the opposite direction than the other two) or hybrid (three strands are parallel to each other while the fourth one runs in opposite direction). *In vitro* studies have shown that RNA GQs are almost all parallel[46] whereas DNA GQs can exist in all the three conformations[47,48]. Further, the formation of GQs can involve either a single strand of DNA/RNA ('intramolecular' GQs) or involve 2-4 separate DNA/RNA strands ('intermolecular' GQs). This combination of different relative strand orientation leads to the formation of different loop arrangements. Loops in the GQs are classified in three groups which are diagonal ('d'), lateral ('l') and propeller ('p'), depending upon the way they connect the stems of the GQs as shown in **Figure 1-9**. Stems of the G-Quadruplex consist of a system of stacked guanosines which may obtain either an anti or a syn glycosidic bond conformation. GQs can again be divided into three categories depending upon the glycosidic bond orientations of the guanosines along the quadruplex stem. Group I,

consisting of the parallel-stranded quadruplex, are the ones in which all the guanosines adopts the same type of glycosidic bond orientation (e.g. all anti)[49]. Group II, consisting of hybrid quadruplexes, have adjacent stacking guanosines in combinations of both glycosidic bond orientations pairs i.e. of the same type (anti-anti, syn-syn) and of different type (syn-anti and anti-syn steps)[49]. Group III, consisting of the anti-parallel quadruplex, have their consecutive guanosine stacking with alternating glycosidic bond orientation[49] as shown in **Figure 1-10**.

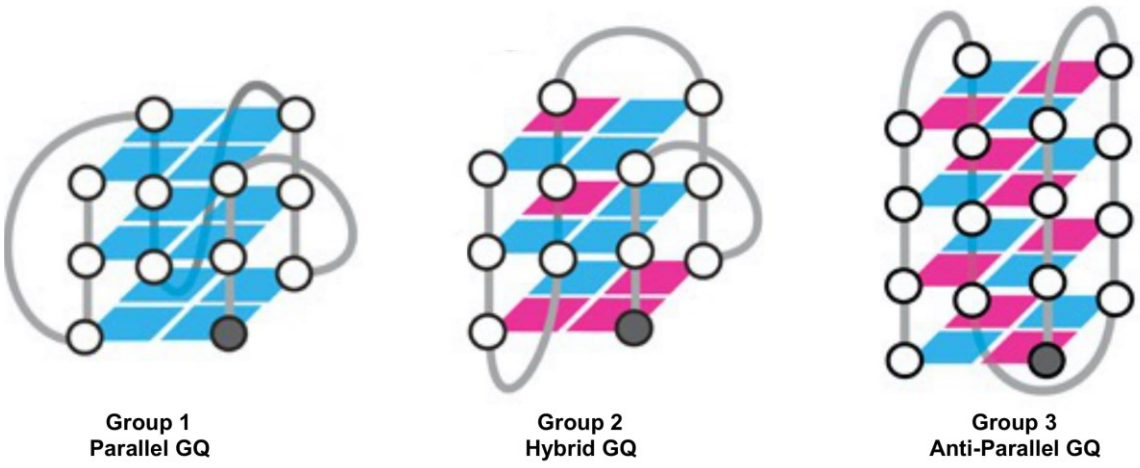


Figure 1-10. **G-Quadruplex classification based on adjacent-stacking guanosine glycosidic bond orientation**. Anti (Cyan) and Syn (Magenta) guanosine making quadruplex in different combination. The grey circle represents the 5' end of the quadruplex. Adapted, with permission, from ref. [49].

Whether GQs are formed *in vivo* or not remains an area of active research in the scientific community. Since the initial discovery of GQs formed *in vitro* by sequences linked with telomeric DNA, several studies have been carried out to probe for the presence of GQs in cells (both live and fixed cells). Following are the major classes of experimental approaches (**Figure 1-11**):

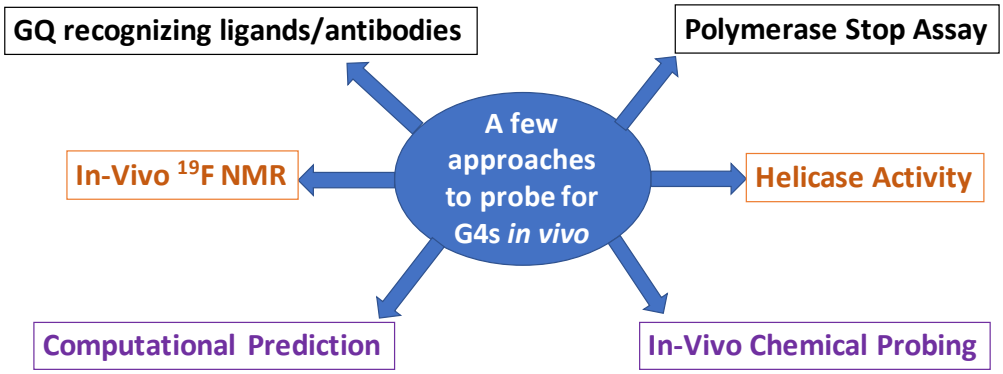


Figure 1-11. Major existing techniques of probing *in-cellular* GQs. Classification of the major existing techniques, in 6 broad categories, to explore the formation of GQs within living/fixed cells.

1.5.1. Ligands and antibodies for GQ recognition and binding

Several ligands as well as antibodies have been developed that uniquely recognise GQs as opposed to DNA/RNA duplexes (dsRNA and dsDNA) or single strands (ssDNA and ssRNA). Some widely used synthetic ligands are BioTASQ[50], Pyridostatin[51], Braco-19 HCL[52], NMM[53] and PhenDC3[54]. One of the most important, and intracellularly occurring, ligand for binding GQs is heme[53,55]. Heme is a naturally occurring cofactor present ubiquitously inside the cells. Other than small molecule ligands, such as the ones above, a widely used antibody, BG4, for GQ binding was developed by Biffi et al. BG4 binds to intramolecular and intermolecular DNA GQs with nanomolar affinity[56,57]. Other antibodies that have been reported to bind GQs are Sty3 (specific for parallel GQs[58]) and Sty 49 (antiparallel GQs[58]). In addition, certain synthetic ligands (e.g. BioTASQ[50]) as well as antibodies such as BG4[56,57], have been used for GQ's pull-downs from cellular chromatin and the DNA so obtained are sequenced by next-generation sequencing (ChIP-Seq). Also, fluorophore-attached antibodies and ligands have been used to visualize GQ both in fixed and live cells[57,59,60].

1.5.2. Computational algorithms

Looking *in vitro* at the patterns of sequences capable of forming GQ folds, several algorithms have been developed (e.g. $\mathbf{G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}}$) to predict potential quadruplex forming sequences in genomes[61,62]. G4 hunter is one such program that measures the G-richness and G-skewedness of a given query sequence and comes up with a quadruplex propensity score[63]. An another program called QGRS

mapper makes quadruplex prediction using the algorithm $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$, where x (number of tetrad) ≥ 2 and y1,y2,y3 (loop lengths) can be flexible and are user specified[64]. However, none of these existing algorithms are wholly satisfactory for making robust predictions of GQ formation, particularly where quadruplex formation involves unconventionally long loops and bulged out structures. Very recently, improved algorithms involving the use of machine learning have been incorporated to predict the GQ's formation, taking into account the role of unusually long loops and other subtle deviations seen in quadruplex formation[65]. A total of 400,000 potential GQs have been estimated to exist in human genome using computational algorithms. Further improvement in these algorithms are expected to be very helpful in advancing the mapping of GQs, both at the level of the genome as well as transcriptome, in association with experimental data.

1.5.3. Polymerase Stop Assay

In 2015, Chambers *et al*, came up with an alternate way of finding GQs within genomes[66]. The procedure involved extraction of total genomic DNA followed by primer extension under two different experimental conditions (**Figure 1-12**). First, a non-GQ-stabilizing condition (such as in the presence of Li^+) and second, a GQ stabilizing condition (e.g., in presence of K^+ or GQ stabilizing ligand Pyridostatin). The sites where a polymerase halted during primer extension under quadruplex-stabilizing conditions were mapped back to the genome to find the sites for potential GQs. While not necessarily giving a true representation of the in-cellular situation, this method nevertheless helps to identify genomic hotspots that have a high potential to fold into GQs.

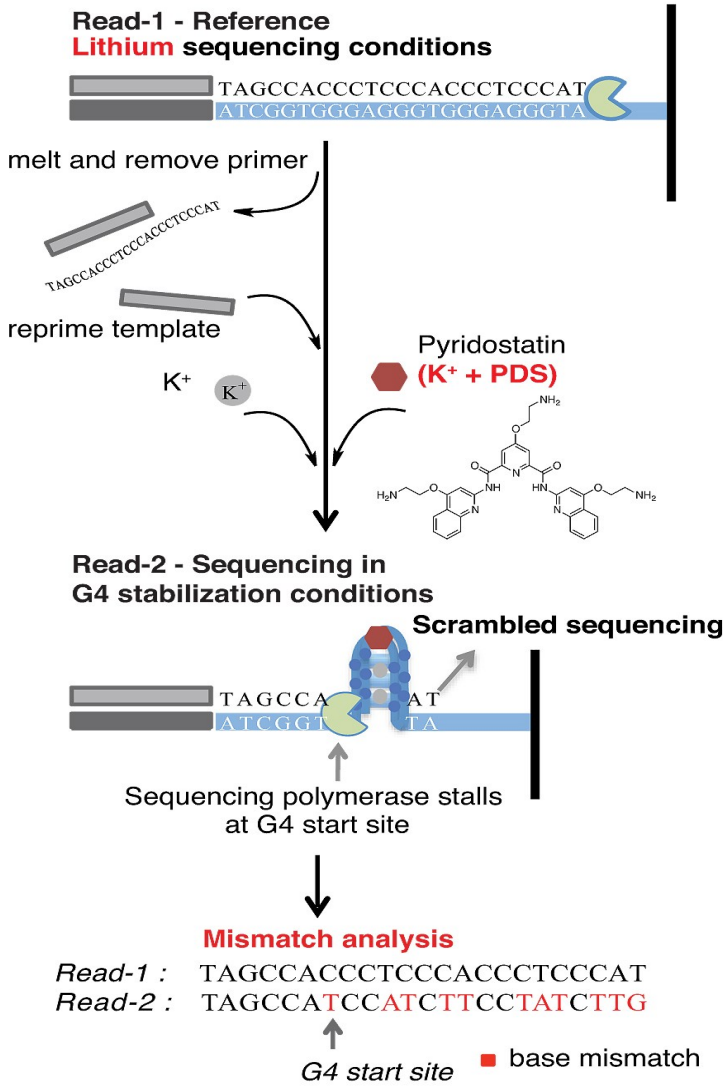


Figure 1-12. Polymerase stop assay. Extracted genomic DNA are sequenced under two conditions. Read 1: under Li^+ sequencing condition (which doesn't favor G-quadruplex) and Read 2: under G-Quadruplex stabilizing condition (K^+ and PDS). Sites of mismatch and difference in the sequencing quality between read 1 and read 2 is analyzed to find GQ forming loci. Adopted, with permission, from ref ([67]).

1.5.4. Helicase Activity

Indirect evidence for the formation of transient GQs come from "loss & gain-of-function" studies carried out with intracellular GQ-specific helicases. Some major GQ specific helicases are FANCJ[68] and the Bloom helicase[69]. Cell lines from patients with depleted FANCJ have been shown to accumulate large genomic DNA deletions that map to sites of potential GQs within the genome[70,71]. Similarly, depletion of Pif1 DNA

helicase in Yeast is shown to correlate with genetic instability of the alleles of G-rich human microsatellite CEB1 (shown to form stable G-quadruplexes *in vitro*) when inserted in the yeast *S. cerevisiae* genome[72].

1.5.5. *in vivo* ^{19}F NMR

Aside from mapping genomic G-Quadruplexes, a different approach has been used to test the capability of an *intracellular milieu* for the successful folding of a pre-denatured and synthetic, potentially GQ-forming sequence into GQ intracellularly. To validate this methodology, in-cell NMR was done after synthetic GQ-forming RNA sequences chemically derivatized with ^{19}F at 5' terminus , was transfected into *X.laevis* Oocyte[73]. The use of ^{19}F ensured a very low background signal and thus helped in minimizing noise in the experiments. *in-cell* NMR experiments were consistent with the proper folding of the injected synthetic sequences into GQs, confirming that intracellular milieus are competent for supporting and stabilizing G-quadruplex folds of RNA.

1.5.6. *In vivo* chemical probing

Chemical footprinting is often used as a classic approach, *in vitro*, for the characterization of various DNA secondary structures[74] (See Section 1.10.3). One of the most common modifications that is used to study GQs is alkylation by dimethyl sulphate (DMS). DMS modifies the N7 position of guanine when it is not involved in hydrogen bonding, such as in forming G-quartets and YR*Y base triples. Guanines that undergo DMS modification at the N7 position become prone to cleavage upon heating in the presence of aqueous piperidine (an organic base). DMS also methylates N1 position of adenine and N3 position of cytosine. DMS Modifications in the case of adenine and cytosine are usually detectable not by piperidine mediated cleavage but by primer extension blockage one nucleotide prior to modification site. To probe for GQs *in vivo*, however, potassium permanganate ($KMnO_4$), another common footprinting reagent, has been used preferentially. $KMnO_4$ easily penetrates the cells and oxidises unpaired/looped out thymines which often accompany GQ formation[75]. The DNA backbone at such oxidised bases can then be cleaved by specific nucleases such as nuclease S1, and used to map potential GQ formation *in vivo*.

Using a combination of the different approaches discussed above, a consensus of ~700,000 potential GQs have been estimated to be capable of forming in the human genome. A recent, separate, study done by the Shankar Balasubramanian group has reported a quantitative and comparative assessment of potential GQ sites in various model organisms[67]. The distribution of predicted GQs is very diverse on different chromosomes. Initial discoveries identified telomeric regions to be hotspots for GQ formation but later, upon detailed investigation, remarkably different observations were made. Incubating GQ-specific antibody (BG4) with individual chromosomes during metaphase revealed that ~75 % of the identified potential GQ-folding foci co-localized in non-telomeric regions[57]. A large proportion of these GQ sites, in the non-telomeric regions, lie in the promoter regions of either proto-oncogenes or genes related to neurological disorders. Some of the proto-oncogenes known to carry putative sites of GQ formation in their promoters are *c-MYC*[76], *BCL-2*[77], *KRAS*[78], *NRAS*[79], *KIT*[80], *VEGF*[81], *Her2*[82], *PIM1*[83] *etc*. Similarly, genes related to neurological disorders with genes linked of GQ formation include *CSTB* [84] (progressive myoclonus epilepsy type1), *FMR1*[85] (Fragile X Syndrome), *NOP56*[86] (spinocerebellar ataxia), *PRNP*[87] (familial Creutzfeldt-Jakob Disease), and *C9orf72*[88] (Amyotrophic Lateral Sclerosis/Frontotemporal dementia) etc.

1.6. Roles of GQs in Neurological Disorders: the $(CCGGGG)_n$ repeat expansion in ALS

GQs have been found to be key regulators in the etiology and symptoms of neurological disorders. Three major mechanisms regarding how GQs lead to pathology in these disorders are considered to be: *a)* at the level of genome; *b*) at the level of transcriptome; and *c)* at the level of the proteome. One of the classic cases that appears to involve in all the three levels of the above is the hexanucleotide repeat expansion sequence [$d(CCGGGG)_n$] linked to familial Amyotrophic Lateral Sclerosis (ALS) and frontotemporal dementia (FTD).

Hexanucleotide repeat expansion of the repeat sequence, $d(CCGGGG)_n$, is found in the 5' UTR of the *C9orf72* gene[89,90]. In normal population, this sequence is repeated 20-30 times or less; however, in the pathology of the disease, this repeat

expansion magnifies to hundreds- to thousands-fold[91,92]. At the level of DNA, $d(C_2G_4)_n$ single strand and its complementary strands have been shown, *in vitro*, to form unusual secondary structures, such as R-loops, i-motif, hairpins, and most prominently, GQs. These alternate structures, if present in repeat-expansion afflicted neurons, can potentially cause down-regulation in gene expression leading to reduced levels of the coded protein[93–95]. Indeed, the repeat expansion has been shown to decrease *c9orf72* expression[89]. At the level of RNA, $r(C_2G_4)_n$ repeats form tangled GQ foci, both in the nucleus and cytoplasm. These foci likely serve to sequester many RNA binding proteins including splicing factors[96–99] as well as cellular heme[100]. Such a sequestration leads to defective pre-mRNA splicing. Finally, at the level of proteins, the RNA transcripts that are produced, undergo non-AUG initiated translation (RAN) to generate dipeptide repeats and are hypothesised to cause toxicity[101,102]. Following on from these above identified roles of alternate secondary structures of $d(CCGGGG)_n$ in causing pathology, Chapter 2 of this thesis lays out the discovery of a unique secondary fold by this very DNA sequence under mildly acidic conditions, named as “iCD-DNA”.

1.7. DNA application in Nano-Technology

Owing to its easy replication, low cost of synthesis, precise and reversible self-assembly properties, robust chemical stability, and feasible charge conductivity through it, DNA has been considered as an excellent material for diverse use in the field of nanotechnology[2–6]. Different combinations of the secondary structures of DNA: in particular, GQs, i-motifs and DNA triplexes, have played pioneering roles in DNA’s utility for nanotechnological constructions and applications.

The Sen lab has done notable work in designing electric nano-switches using DNA. In this paradigm, electrical switching of a DNA construct involves a conformational change between two possible structures of the same DNA sequence construct. One of the two conformers has a higher electrical conductivity than the other. In 2014, Huang et al., designed a ‘P-duplex’, which incorporated a stretch of mismatched guanines in opposite strands of a duplex[103]. The P-duplex, in the absence of K^+, is a poorly conductive conformer (E-form), which in the presence of K^+ folds to form an intramolecular G-quadruplex with significantly higher conductivity (C-form). Non-

canonical structures of DNA have also been used to make logic gates[104,105]. A logic gate is an electronic circuit which forms a fundamental unit of any digital system. It consists of one or greater than one input portals and only one output. The relations between the input and output possibilities is represented in a “truth table”.

DNA nanowires or long 1-Dimensional DNA nanostructures (1DDN) have been an another burgeoning area of exploration in the field of DNA nanotechnology. GQs and i-motif have proven particularly useful in the designing of long DNA nanowires because of their increased stiffness, conductivity, thermal stability and resistance to nucleases compared to canonical duplexes. The wires, so generated, in the case of GQs is called ‘G-Wire’ and in the case of i-motif , is referred to as ‘C-Wire’[106,107]. DNA Duplexes, on the other hand, have been a very poor candidate for wire formation because of their low persistence length (~50 nm)[108].

Chapter 3 of this thesis discusses a novel approach for designing a long and reversible 1-Dimensional DNA nanowire/nanostructure using a hybrid of Triplex-Quadruplex DNA tile.

1.8. Heme-G-quadruplex interaction

Heme is a naturally occurring, ubiquitous cofactor present in every cell. Heme is a metalloporphyrin consisting of protoporphyrin IX complexed with iron (**Figure 1-13**). The chemical structure of heme (iron protoporphyrin IX) consists of four pyrrole rings with four methyl groups, two vinyl groups and two propionate groups attached. This makes heme chemically multifaceted. It has both hydrophobic groups (porphyrin ring, methyl and vinyl groups) and hydrophilic groups (iron and propionate). Such a chemical composition makes heme fit for both hydrophilic and hydrophobic environments. For example, heme can make stacking interactions through fitting suitably in hydrophobic pockets of certain enzymes in addition to interactions through salt bridges. The most stable and known oxidation state of iron in heme are Fe^{+2} (ferrous) and Fe^{+3}(ferric). Although ‘heme’ is used interchangeably as a generic term to identify both ferrous and ferric forms of iron protoporphyrin IX; in its correct form, heme refers to ferrous

protoporphyrin IX. Hemin is used for ferric protoporphyrin IX. The net charge of hemin is +1 and normally exits with counterion like Cl^-.

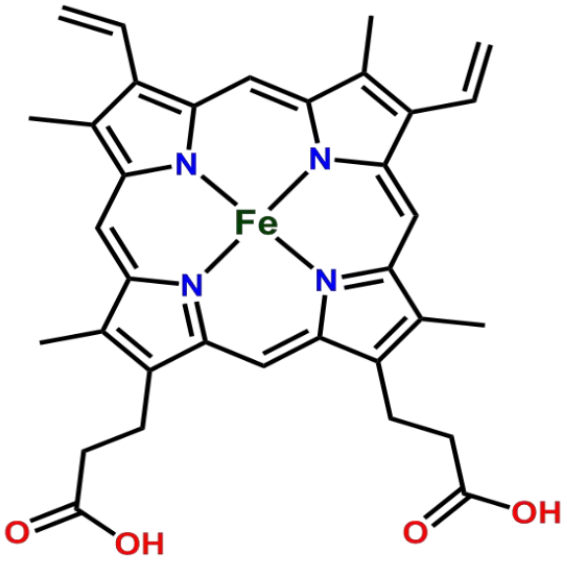


Figure 1-13. Chemical structure of iron protoporphyrin IX. Heme consists of four pyrrole group with four methyl groups, two vinyl groups and two propionate groups.

Heme level is tightly regulated inside cells. In human, around 80% of the heme is made and present in red blood cells; ~15 % is synthesized and is present in liver and the rest is distributed to other tissues[109]. Heme is used as a cofactor by a class of proteins called hemoproteins which perform a diversity of functions. These include: haemoglobin and myoglobin that transport and store blood oxygen, respectively; cytochromes and oxidoreductases that support cellular energy generation and biosynthesis; peroxidases that perform oxidative catalyses; cytochrome P450 monooxygenases that are involved in drug metabolism and synthesis of endogenous substances such as steroids and lipids; and heme oxygenases that control heme degradation processes as well as synthesis of important neuromodulators. Other than acting as a cofactor for many proteins, heme has been found to serve as a signalling molecule inside cells[110]. It has also been found to control basic molecular and cellular processes such as gene transcription, protein localization and assemblies[109].

The catalytic property of hemoprotein - peroxidases and monooxygenases - are highly related to a particular secondary fold of guanine-rich ssDNA/RNA called ‘G-Quadruplex’ that binds heme and possess similar activities. Peroxidases are the

enzymes that catalyze one electron oxidation of various organic and inorganic substrate using peroxide (usually H_2O_2). The most studied peroxidase are horseradish peroxidase (HRP), cytochrome c peroxidase (CcP), ascorbate peroxidase (APX), lignin peroxidase, myeloperoxidase (MPO) and lactoperoxidase (LPO). Monooxygenases, on the other hand, are the enzymes that catalyse the insertion of a single oxygen atom from O_2 into an organic substrate (two electron oxidation), e.g., cytochrome P450 monooxygenases.

Other than naturally occurring peroxidases, an interesting class of non-natural peptide sequences are also capable of performing peroxidase like functions. These peptides are referred to as catalytic antibodies or abzymes. Lerner**,** Schultz and co-workers have generated numerous abzymes that catalyze a wide variety of substrate utilizing the transition state mimics[111,112]. Discovery of such non-natural abzymes became an inspiration for the search of DNA/RNA that could be capable of binding and activating heme.

In late 1990s, Sen and coworkers found, for the first time, that GQs are excellent binders of heme (K_d in range of 10-500 nM)[53,113–119]. Comparison of the UV-visible absorption spectra of a heme-protein enzyme (e.g. metmyoglobin, the Fe(III)-oxidised form of myoglobin) and the heme-GQ complex resembled each other greatly[119]. Notably, the heme-GQ complex and the heme-protein enzyme[120] showed a similar and extensive hyperchromism in the Soret region (~400 nm)[119]. Such Soret band hyperchromism is an indicator of enhanced hydrophobicity in the environment of the bound heme relative to water.

It is well-known now that heme binds to GQs by end-stacking on the terminal quartets. And that heme, preferentially binds strongly to parallel quadruplex than the anti-parallel ones[121]. The mechanism of heme activation both by protein enzyme and by GQs has been found to be very similar. In the oxidized form of myoglobin, the Fe (III) has a 6-co-ordination where the proximal position coordinates to a nitrogen atom from a histidine residue and the distal site is occupied by water. In GQ-heme complex, the 6th distal position is coordinated with water which can be exchanged with peroxide under suitable conditions[122], and the proximal ligand is hypothesized to be guanine or water[55].

Phenolic substrate like Tyramine, Tyrosine and substituted Tyramide are very good substrate for Heme-GQ complex (also referred to as Heme-DNAzyme)[123]. Heme-

DNAzyme system can oxidize a broad range of substrate with around 70-fold enhanced rate as against heme alone[123]. Also, for several phenolic substrate Heme-DNAzyme system has been reported to exhibits a faster rate of oxidation than the Horseradish Peroxidase[123].

While GQs are an excellent binder of heme, no other naturally occurring fold of DNA/RNA has been found to bind and activate heme. In a recent study conducted by Shumayrikh et al., it was found that planar iso-guanine (iG) pentaplexes (but not the non-planar iG quadruplexes) bind and exhibit peroxidase activity comparably to heme-GQs[124]. iG-pentaplexes and G-Quadruplexes have a planar structure whereas iG-quadruplexes are non-planar. Given this observation, a planarity and distally coordinated nucleobase seems sufficient for heme activation.

Chapter 4 of this thesis will discuss in detail the relevance of this heme-GQ interaction, and will describe the development of a new and low-perturbative approach at capturing both DNA and RNA GQs, both *in vitro* and within living cells.

1.9. SELEX

Selex (**S**ystematic **E**volution of **L**igands by **EX**ponential enrichment), also known as "*in vitro* selection", is a combinatorial chemistry technique that allows simultaneous screening of a large and diverse pool (library) of single stranded DNA/RNA molecules against a given target with an aim to isolate either an aptamer (unique DNA/RNA sequence that binds to a given target with high affinity) or DNAzyme/Ribozyme (DNA/RNA sequence that catalyzes a chemical reaction upon binding to the target, respectively). Selex technology was first developed in 1990 by Ellington and Szostak[125] and Tuerk and Gold[126], separately. A general workflow of SELEX involves a random library of single stranded DNA sequence (10^{14}-10^{15}) which is incubated with a target molecule (**Figure 1-14**). A very small fraction of DNA sequence that successfully binds to the target in the first round of Selex is separated from the unbound fraction and is amplified using PCR. This whole cycle is repeated for multiple rounds (usually 8-15) till a sufficient enrichment in the functional sequence is obtained. There are few additional steps involved in the process of *in vitro* selection to find RNA aptamers. The RNA of

interest in each round is converted into complementary DNA by reverse transcription, followed by PCR amplification and in-vitro transcription to re-generate the RNA library for next round of selection. Since the initial conceptualization of this technique, there has been a lot of modifications and it has resulted in the generation of thousands of aptamers targeting amino acids, organic molecules, proteins, small metal ions, virus, bacteria and whole cells[1].

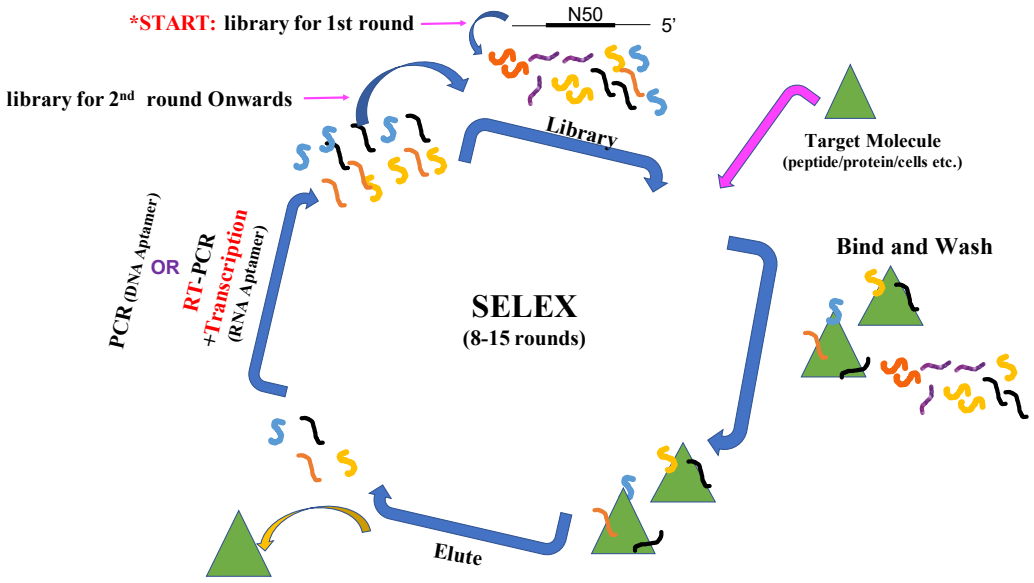


Figure 1-14. Schematic representation of Selex. DNA/RNA aptamer Selex protocol. The starting library consists of DNA/RNA sequence with 50 variable nucleotides (N50) enclosed by two constant flanking regions used for PCR amplification.

1.10. Biophysical and Biochemical tools to study DNA/RNA secondary structures

This section gives a detail on the tools and techniques that are used to study the secondary structures of DNA and RNA. It will form the basis to understand all the works that have been explained in the subsequent chapters.

1.10.1. CD-Spectroscopy

CD, or Circular Dichroism, Spectroscopy is an absorption spectroscopy that provides structural information of chiral molecules and assemblies. CD Spectra is the measure of the differential absorption of the right-handed Vs left-handed circularly

polarized light as a function of wavelength by an asymmetric molecule. Light is an electromagnetic wave where electric and magnetic field component oscillate perpendicularly in the direction of propagation of light. Light can either be un-polarized or polarized. Also, light can be polarized either linearly or circularly. Most of the light emitting sources emits un-polarized light. In un-polarized light the two components (electric and magnetic) oscillate randomly in many different directions. In linearly polarized light, the electromagnetic waves oscillated along a single plane while in circularly polarized light, two electromagnetic waves oscillated in planes with 90^{o} phase difference (**Figure 1-15**). Light can be circularly polarized in two directions: Left-handed circular polarization, when the vector rotates counter-clockwise looking down the axis of propagation and Right-handed circular polarization, when the vector rotates clockwise looking down the axis of propagation.

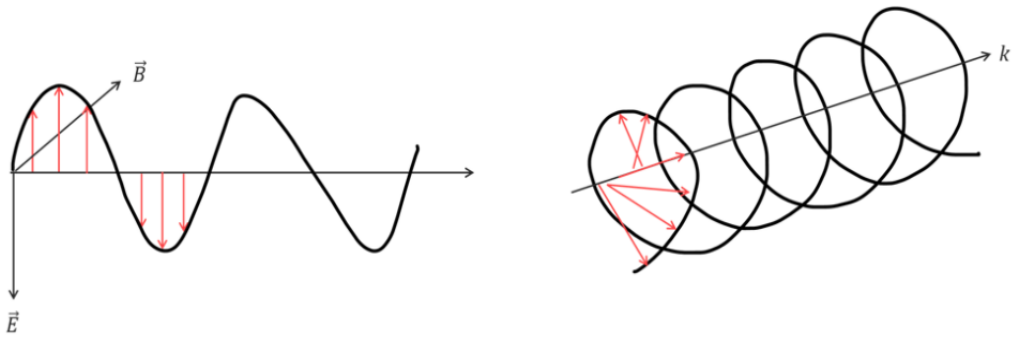


Figure 1-15. Linear and Circular Polarization of Light. (Left): Linear polarized light: Electromagnetic waves oscillating along a single plane parallel to the plane of polarization. (right) Circular polarized light: Electromagnetic waves oscillating in planes with 90^{o} phase difference.

Nucleic Acids, Proteins and many other biomolecules are chiral, i.e., they do not have a plane of symmetry. These chiral molecules are optically active as they can rotate the plane of polarization when they come in the path of plane/linear polarized light. Another interesting phenomenon observed with a chiral molecule is that it shows a differential absorption to right-handed vs left-handed circularly polarized light. As a result of which chiral molecule produces a unique CD spectrum.

The basic principle behind CD spectroscopy is summarized in **Figure 1-16**. A linear polarized light is converted to an alternating right and left-handed circularly polarized light which is then passed through an optically active (chiral) molecule. The

differential absorption of the two polarizations by an optically active molecule is then detected by a Photo Multiplier Tube (PMT) and is reported either in the units of absorbance or historically in unit of Molar Ellipticity (θ) in $deg.cm^2.dmol^{-1}$.

In theory, quantum mechanics approach can be used to calculate the CD spectra of nucleic acids, however, given the complexity of DNA/RNA, CD spectra of nucleic acids have mostly been interpreted empirically. CD spectra of the quadruplex arises primarily due the stacking arrangements of guanine base steps within a G-Quartet stack and the accompanying glycosidic bond orientation of guanosine. As a general rule-of-thumb, 'parallel GQ' with a positive CD peak at 260 nm and a negative one at 240 nm; and 'antiparallel-GQ' with a positive CD peak at 290 nm and a negative one at 260 nm are widely accepted for GQ characterization. Kypr J. et al. has summarized the signatory CD spectra of different secondary structures of DNA namely A, B and Z form of duplex, triplex, cytosine i-motif and G-Quadruplexes[127]. **Table 1-2** below summarizes the key CD spectra features for different secondary structures of DNA. Similarly, signatory CD spectra of different secondary structures of proteins have been documented by Norma J. Greenfield[128]. Briefly, alpha helix has a negative CD peak at 222 nm and 208 nm and a positive one at 190 nm; beta-sheet has a negative band at 218 nm and a positive at 196 nm; and the random coil shows a negative band at 195 nm and a positive band at 212 nm.

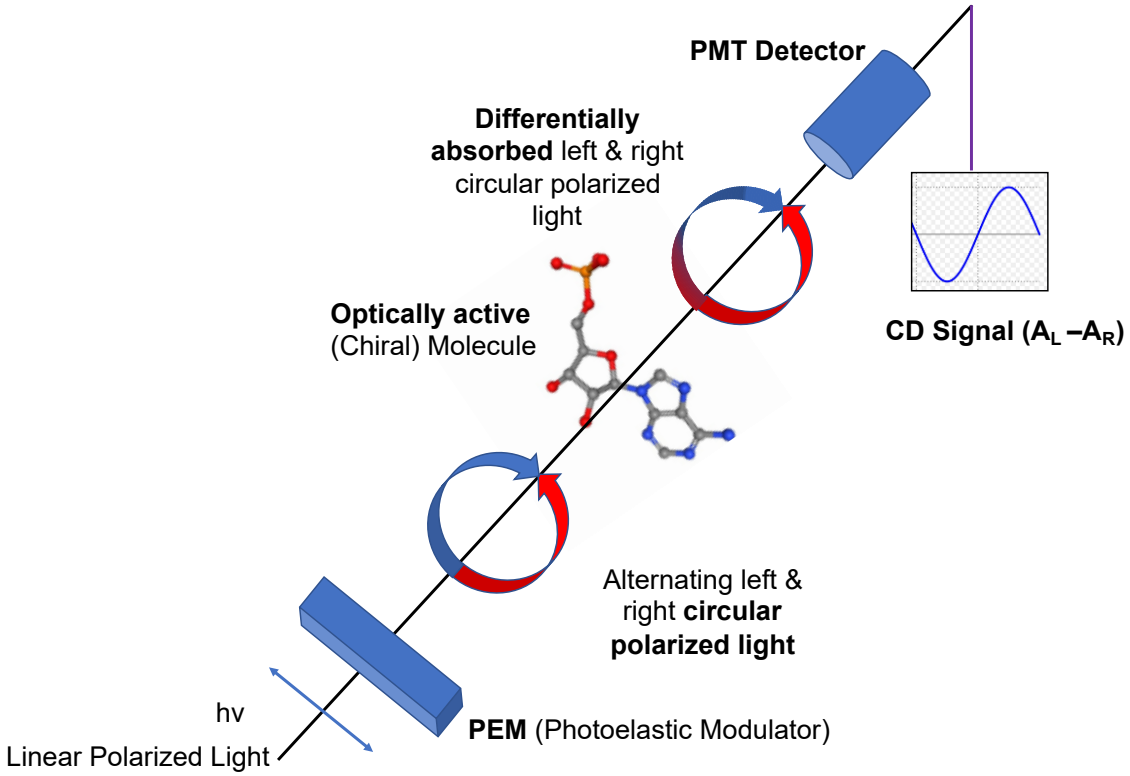


Figure 1-16. Basic principle behind CD Spectroscopy. A linear polarized light is passed through photoelastic modulator which generates alternating left and right-handed circular polarized light. The alternating left and right-handed circular polarized light are differentially absorbed by an optically active (chiral) molecule which is detected by a Photo Multiplier Tube (PMT).

Table 1-2. CD Signature for different secondary folds of nucleic acid. An empirical observation of the signatory CD features for different secondary structures of Nucleic acids. #: Anti-parallel and parallel directions are with respect to the purine strand participating in the Watson-Crick duplex strand of the triplex structure. Note that triplexes don't have characteristic CD spectrum like duplexes or quadruplexes, different triplex sequence produces different CD spectra[127].

Nucleic Acids secondary Structure	Handedness	Types	Ellipticity Maxima (in nm)	Ellipticity Minima (in nm)	Other features	Ref
Duplex	Right-Handed	A-type	~260	~210		[127]
		B-type	260-280	~245		
	Left-Handed	Z-type	~260	~290	Extreme negative band at ~210 nm	
Triplexes	Right-Handed	YR*R (Anti-Parallel)[#]	~260	~230		[129]
		YR*Y (Parallel)[#]	259 & 284	~248		[130]
G-Quadruplexes	Right-Handed	Parallel	~264	~245		[127]
		Anti-parallel	~295	~260		
		Mixed/Hybrid	~295	~245	Additional shoulder at ~260	
	Left-Handed	Parallel (ZG4)	~245	~280		[131]
i-motif (cytosine tetraplexes)	Right-Handed	Anti-parallel	290	~260		[127]

1.10.2. NMR

Although CD spectroscopy offers a rapid, easy and reliable method for the study of different secondary structures of DNA/RNA, it fails to give residue specific i.e., high-resolution information (e.g., Hydrogen bonding pattern) as well as overall 3D structural information. Nuclear magnetic resonance (NMR), in such a scenario, is a very helpful alternative approach, with the benefit of solving DNA/RNA 3D structure at atomic or near-atomic resolution in solutions. NMR is capable of providing a large variety of information on the structure and dynamics of DNA and RNA.

Nuclear magnetic resonance measures the chemical shift i.e., the resonant frequency of a spin/NMR active nuclei (e.g., ^{1}H, ^{13}C, ^{15}N, ^{31}P) with respect to a standard reference in a magnetic field. Multidimensional NMR spectra gives a lot of information about the structure, dynamics and functions of DNA/RNA. Using **1D proton NMR**, one can identify the nature of hydrogen bonding between the bases (e.g. Watson Crick H-bond show an imino proton signatory chemical shift between 12-15 ppm whereas Hoogsteen H-bond show a signatory chemical shifts between 10-12 ppm[132,133]). **2D NMR** on the other hand gives two general classes of distance information: A.) COSY (**CO**rrelated **S**pectroscop**Y**): It gives the distance information between neighbouring atoms using the bond data. Correlation involves a spin-spin coupling in neighbouring bonded atoms. B.) NOESY (**N**uclear **O**verhauser **E**ffect **S**pectroscop**Y**): It gives the distance information between two atoms through space data i.e., two atoms which are close in space even if they are not bonded. **3D NMR** experiments are constructed from the combination of two 2D NMR. 3D NMR provides some added advantage from 2D NMR in the sense that it introduces a third dimension and helps in overcoming the peak overlap problem of 2D NMR. However, it comes with a cost of increase in acquire time for capturing the spectra.

1.10.3. Gel Electrophoresis and chemical footprinting

CD and NMR provide a great degree of insight onto the structural aspect of the 3D structure of DNA/RNA. However, one of the major limitations for these technique is the necessity to have a pure/homogenous mixture. A heterogeneous mixture of DNA/RNA makes it very hard to explain the CD or the NMR data. One of the simplest and cost-effective way of determining the different higher order structures of DNA/RNA even in a heterogeneous mixture is gel electrophoresis.

Gel electrophoresis is used to separate DNA/RNA molecules as per their size. Gel electrophoresis are run mainly under two conditions: 1.) Native and 2.) Denaturing. Native gel electrophoresis is used when DNA/RNA are to be separated keeping their folded conformation intact whereas denaturing gel is used to separate DNA/RNA after they have been denatured as a single strand.

Gel electrophoresis is often coupled with different chemical footprinting experiments to inquire more about the hydrogen bonding patterns involved in a

secondary structure of DNA/RNA. The general procedure of chemical footprinting involves covalent modification of the DNA/RNA at specified sites with unique chemicals. These chemically modified sites then become amenable to cleavage when treated with mild bases like piperidine. Denaturing gel is then used to see the modified sites. Some of the most widely used chemical modifications are DMS (for 'N7' position of G and 'N3' position of A[74]), Hydrazine (for C and T mismatch[74]), Hydroxylamine (for C mismatch[134]), Osmium Tetroxide (for T and C mismatch[134]) and Bromide (for C-Bulge[135]).

1.10.4. Atomic Force Microscopy (AFM)

Atomic force microscopy was invented in 1985 by Binnig, Quate and Gerber as a high resolution version of scanning probe microscopy[136]. It enabled imaging compatibility with almost any type of surface, including glass, various polymers, ceramics, and, most importantly, with biological samples. AFM works on "surface sensing" mechanism. The basic principle of AFM involves a sharp tip (radius ~ 5 -10 nm) with a cantilever micro-fabricated with Si or Si_3N_4, which can be used to raster-scan over a surface pre-deposited with a target material of interest (in terms of what is presented in this thesis, it is DNA). The cantilever/tip assembly that interacts with the target material is also called a probe. In a tapping mode, the tip taps the surface in a systematic scanning process. The trajectory of the tip i.e., its up and down as well as side to side motion is monitored using a laser beam deflection system as shown in **Figure 1-17**. A laser beam continuously hits the probe during the scan and the reflected beam (from the probe) is tracked by a position sensitive photo-detector (PSPD). The reflected laser beam falling on the photo-detector captures the vertical and lateral motion of the probe. The motion of the probe, in turn, is influenced by surface topography which varies in the presence or absence of the sample being deposited over the surface. In tapping mode, the tip doesn't touch the surface physically rather the cantilever tip oscillates just above the surface while it scans. A very precise high speed feedback loop is used to prevent the cantilever tip from crashing into the surface. As the tip approaches the surface, tip-surface interaction causes the oscillation amplitude of the cantilever to decrease. The feedback loop corrects for the deviations in these amplitude and generates the image of the surface topography.

1.10.5. Transmission Electron Microscopy (TEM)

The earliest demonstration of TEM was given by Max Knoll and Ernst Ruska in 1931, and a commercialized version of an instrument came in 1939[137]. Since then, TEM has become a powerful tool for visualizing miniaturization in the order of 0.2 nm. TEM works on the principle of a light microscope; however, it uses electrons rather than the light. Electrons have a wavelength of $\lambda_{electron} = 10^{-10}$ m which is relatively very smaller compared to visible light's wavelength spectrum ($\lambda_{visible\text{-}light}$ = 380-700 nm). This enables TEM to achieve a very high magnification. The working principle of TEM involves a high energy beam of electron, focused into thin, small, and coherent beam using a condenser lens, that is shone through the specimen deposited on formvar/carbon coated copper mesh TEM grids. Detailed informative features of the sample are then reconstructed using the interaction between the electron and the atoms of the sample. Modern TEM instruments are further capable of an X-ray technique called EDX (Energy Dispersive Analysis, also known as EDS or EDAX), which allows identification of the elemental composition of target materials. The basic principle of EDX involves an electron beam hitting the inner shell of an atom. It knocks off an electron from this shell leaving a positively charged electron hole. This electron displacement attracts an electron from the outer shell to come and fulfill its vacancy. The displacement of an electron from an outer (high energy) to the inner (low energy) shell is released as X-ray. The energy of this X-ray is unique to specific elements. TEM instruments are equipped with silicon drift detector (SSD), an X-ray detector, to detect this energy and identify the individual elements in the specimen.

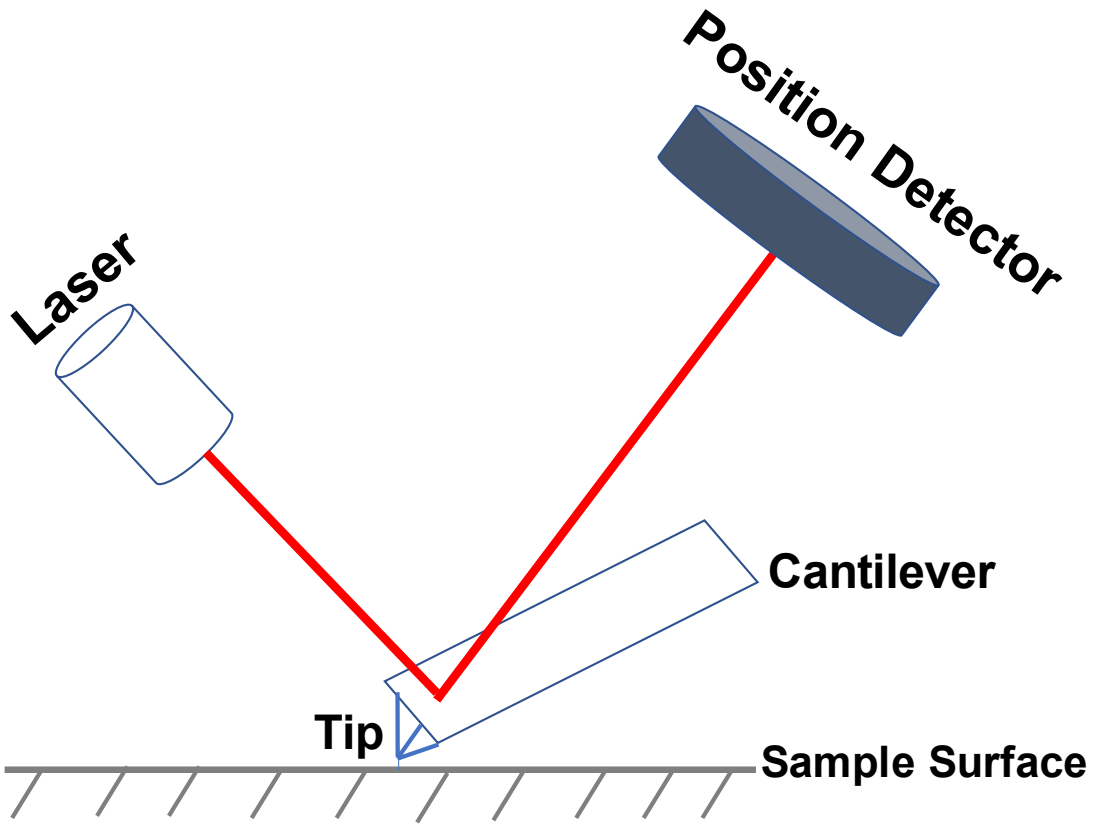


Figure 1-17. Basic principle of AFM. Schematic showing the 'surface sensing' mechanism of AFM. A sharp tip (radius ~5-10 nm) raster-scans a surface pre-deposited with a sample of interest e.g., DNA. The trajectory of the tip is monitored by a laser beam deflection system. In a tapping mode the cantilever tip just oscillates above the surface without touching it. As the proximity between the tip and the surface decreases, tip-surface interaction causes the oscillation amplitude to decrease. A very high precise feedback loop, which prevents the tip from crashing into the surface, corrects for these amplitudes and generates the image of surface topology.

Chapter 2.

$(C_2G_4)_n$ repeat expansion sequences from the *C9orf72* gene form an unusual DNA higher order structure in the pH range of 5-6

2.1. Introduction

Human genome roughly consists of around 30% loci with repetitive sequence scattered in the microsatellite regions[138]. These regions have a variable number of repeat segments in a given population which provides them the benefit of evolutionary plasticity. However, when these repeat regions expand beyond their normal repeat length, it leads to pathology. There are a range of neurodegenerative disease associated with repeat expansion disorders **(Table 2-1)**. One common observation across these repeat segments is that they are highly "GC" rich. Given the importance of 'G' and 'C' nucleobases in the formation of non-canonical secondary structures of DNA and RNA, it is highly conceivable that the presence of 'G' and 'C' in these repeats, beyond a particular threshold, may trigger pathology through the formation of non-canonical secondary structures.

In particular, a hexanucleotide sequence of $d(C_2G_4)_n$, present at the 5' UTR of *C9orf72* gene, when goes in the range of thousands of repeats (n~1000-1500), causes multiple disorders like Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD). In normal individuals, this hexanucleotide expansion lies in the range of 30-40 and normally remains double helical. However, upon expansion, it is expected to switch from duplex to other non-canonical secondary structures. In fact, using synthetic oligos of $d(G_4C_2)_n$, it has been shown *in vitro* that this sequence forms a mixture of parallel and antiparallel GQs at physiological pH in the presence of K^+[139].

In studying potential secondary structures formed by repeats of $d(C_2G_4)_n$ single stranded DNA sequences, I found the formation of an unexpected higher-order structure (referred to as "iCD-DNA"), in response to incubation at moderate to high DNA concentrations at mildly acidic condition. Described below is a study that uses circular

dichroism (CD), native gel mobility and footprinting analysis to investigate this unusual higher order structure.

Table 2-1. List of Repeat Expansion Causing Neurological Disorder. Table showing different repeat expansion related disorders with their repeat length in normal and pathological individuals.

Repeat Seq	**Diseases**	**Normal**	**Diseased**	**Ref**
C_2G_4	• Amyotrophic Lateral Sclerosis (ALS) • Frontotemporal Dementia (FTD)	~2-23	1000-1600	89
$(C_4G)_2CG$	• EPM1 (Myoclonic Epilepsy)	2-3	Upto 75	140
G_2C_2TG	• Spinocerebellar Ataxia Type 36	3-14	650-2500	141
AT_2CT	• Spinocerebellar Ataxia Type 10	10-29	800-4500	142
CCTG	• Myotonic Dystrophy Type 2	30	~5000	143
GCC	• *FRAXE* mental retardation	6-25	>200	144
GCG	• Oculopharyngeal muscular dystrophy	6	8-13	145
GAA	• Friedreich Ataxia	6-10	66-1800	146
CTG	• Myotonic Dystrophy type 1 • Huntington disease-like 2 • Spinocerebellar ataxia type 8 • Fuchs Corneal Dystrophy	5-30	~700->1000	147
CAG	• Huntington disease • Spinal and bulbar muscular Dystrophy • Dentatorubral-pallidoluysian atrophy and seven SCAs	<26	>65	138,148

2.2. Material and Methods

2.2.1. DNA preparation and incubation

All DNA oligonucleotides were purchased from the Core DNA Services Inc. (Calgary, Canada). Oligonucleotides were dissolved in TE buffer (10 mM Tris, 0.1 mM EDTA, pH 7.4), purified once by ethanol precipitation from TE containing 400 mM LiCl. DNA pellets so obtained were redissolved in TE buffer. Oligonucleotides used for native gel mobility analysis and for DMS footprinting experiments were 5' labelled with ^{32}P using γ-^{32}P ATP and a standard kinasing protocol, and then PAGE-purified following a pre-treatment with 10% (v/v) freshly prepared piperidine (v/v) at 90° C for 30 minutes prior to lyophilization.

For incubations, the DNA was heat denatured at 100° C in a water bath for 4 minutes, followed by immediate cooling in ice. Incubations were generally carried out with 700 μM DNA in the appropriate buffer solution, at 37° C. The DNA solution was then diluted with the same or another buffer to give 20 μM DNA, suitable for CD spectroscopy and other experiments. However, in many instances, incubations were carried out directly with 20 μM DNA, with end results indistinguishable from the higher concentration DNA incubations.

2.2.2. Native gel electrophoresis and DMS Footprinting

Native gel electrophoresis of $d(C_2G_4)_7$ was carried out in 7.5% bis/polyacrylamide (29:1) gels and run in TAE-Li Buffer (20 mM Tris, 1mM EDTA, 45 mM acetic acid and 20 mM lithium citrate, pH 5.2). Gels were run at 22° C for 4 hrs at 9 W with efficient cooling. Incubated DNA solutions were mixed with native gel loading buffer (50 mM Tris acetate, pH 5.2, 30% glycerol and loading dyes) prior to loading on gels.

Following incubation of ^{32}P-5'-labeled $d(C_2G_4)_7$ at a concentration of 700 μM in 150 mM Lithium Citrate, pH 5.2, at 37°C for 14 hrs, partial DNA modification with dimethyl sulfate (DMS) was carried out by addition of 0.2% DMS (freshly prepared in 10 mM Tris, pH 5.2). The mixture was allowed to incubate for 30 min at 22° C, and the reaction stopped using ß-mercaptoethanol. Treated solutions were then run on native gels. The observed slow (s) and fast (f) moving bands were excised from the gel and their DNA eluted into TE buffer. DNA was recovered by ethanol precipitation, washed

with cold 70% ethanol, dried, dissolved in freshly prepared 10% v/v piperidine, and heated at 90° C for 30 min. Following lyophilization, the DNA was dissolved in denaturing gel loading buffer (95% formamide, 1 mM EDTA, and loading dyes) and run in 20% denaturing/ sequencing gels.

2.2.3. Gel Data Analysis

Imaging and densitometry of native and sequencing gels running ^{32}P-labeled DNA were carried out on a Typhoon 9410 Phosphorimager (Amersham Biosciences). Quantitation was carried out using the ImageQuant 5.2 software (Amersham).

2.2.4. Circular Dichroism (CD) Spectroscopy

Following incubation and dilution of the DNA, as above, CD spectra was recorded in a Jasco-810 Spectropolarimeter (Jasco, Easton, MD) using a quartz cell of 0.5 mm optical path length. The scanning speed was fixed at 500 nm/min, with a response time of 1 s, and scans were carried out over a wavelength range of 220–320 nm. The spectra in the figures represent averages of 5 sequential scans, all measured at 22° C with baseline correction.

2.3. Results

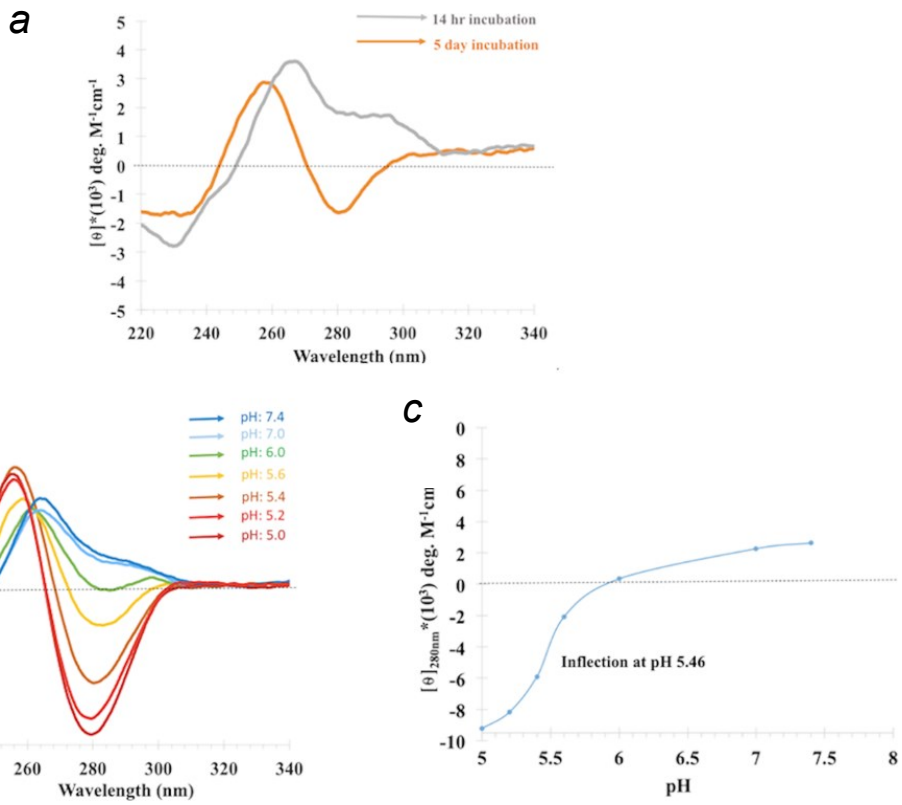
2.3.1. Inverted CD spectrum of d(C_2G_4)$_7$ in the absence of G-quadruplex stabilizing cations

While preparing a negative control for a CD spectroscopic study of G-quadruplex formation by d(C_2G_4)$_7$, I observed that this oligomer, dissolved at 700 μM concentration in TE-LiCl buffer (10 mM Tris, pH 7.4, 0.1 mM EDTA, 150 mM LiCl) and incubated at 37° C for up to 5 days, showed an unusual circular dichroism spectrum. **Figure 2-1.a**, shows spectra for 14-hour and 5-day incubations, with the latter being a smooth, inverted CD spectrum (with a maximum at ~255 nm and a minimum, with net negative ellipticity, at ~280 nm). Such a CD spectrum represents an “inversion” of CD spectra typically observed for A-DNA, B-DNA, as well as for classic DNA triplexes and G-quadruplexes[127]. The relative lack of shoulders in the appearance of the inverted

spectrum suggested either a unitary DNA species or a series of structurally related species rather than a complexly heterogeneous mixture. The long incubation at relatively high DNA concentration (at least at this pH) that gave rise to this CD signal, also suggested that these were slow-forming, thermodynamically rather than kinetically favoured DNA product or products (from this point referred to as "iCD-DNA").

Given that $d(C_2G_4)_7$ contains only two of the four nucleobases, G and C, and the known important role of protonated cytosines in the formation of non-canonical secondary DNA structures like triplex and i-motif, I investigated whether pH values of < 7.0 impacted on the inverted CD spectrum. **Figure 2-1.b** shows that low pH values do indeed accentuate the ellipticity inversion, with amplitudes intensifying even in the 5.2–5.0 pH range and at shorter incubation times than at neutral pH. **Figure 2-1.c** plots molar ellipticity at 280 nm as a function of pH in the 5.0–7.4 range, obtained from the data in **Figure 2-1.b**. The dependence was fitted with a sigmoidal function, with an inflection observed at pH 5.46. Roughly, such an inflection pH is consistent with cytosine protonation within the iCD-DNA structure or structures. To check that equilibrium was reached both at high (700 μM) and moderate (20 μM) DNA concentration incubations (analogous to the data shown in **Figure 2-1.b**), progressively longer incubations under these conditions were carried out. These latter experiments (**Figure 2-1.d & e**) also yielded similar computed inflection pH values.

I investigated whether the initial incubation at 700 μM DNA, such as described above, was strictly necessary for the formation of iCD-DNA. **Figure 2-1.f** shows that $d(C_2G_4)_7$ incubated at 700 μM in 150 mM lithium citrate, pH 5.2, already shows close to the maximal CD amplitude (observed at 16 hours of incubation) by 30 mins; while, a 20 μM DNA incubation does indeed show the characteristic shape (if not the full CD amplitude) after 16 hours of incubation. These data emphasize that iCD-DNA is a thermodynamically favored structure that is optimally but not exclusively generated by incubations at relatively high DNA concentrations. However, are long incubations needed at pH 5.2? **Figure 2-2. a shows** that while 20 μM $d(C_2G_4)_7$ incubated for 2 hrs at 37° C in buffers of various ionic strengths at pH 7.4, generates species with long-lived low CD ellipticities (corresponding presumably to the unfolded or partially base-paired DNA), incubations in increasing strengths of the Li buffer, all at pH 5.2, yields CD spectra characteristic of iCD-DNA in as little as 2 hrs, as can be seen by comparing with the 14 hr incubation (**Figure 2-2**. **b & c**).


a
14 hr incubation
5 day incubation
[θ]*(10³) deg. M⁻¹cm⁻¹
Wavelength (nm)
b
pH: 7.4
pH: 7.0
pH: 6.0
pH: 5.6
pH: 5.4
pH: 5.2
pH: 5.0
[θ]*(10³) deg. M⁻¹cm⁻¹
Wavelength (nm)
c
[θ]280nm*(10³) deg. M⁻¹cm
Inflection at pH 5.46
pH

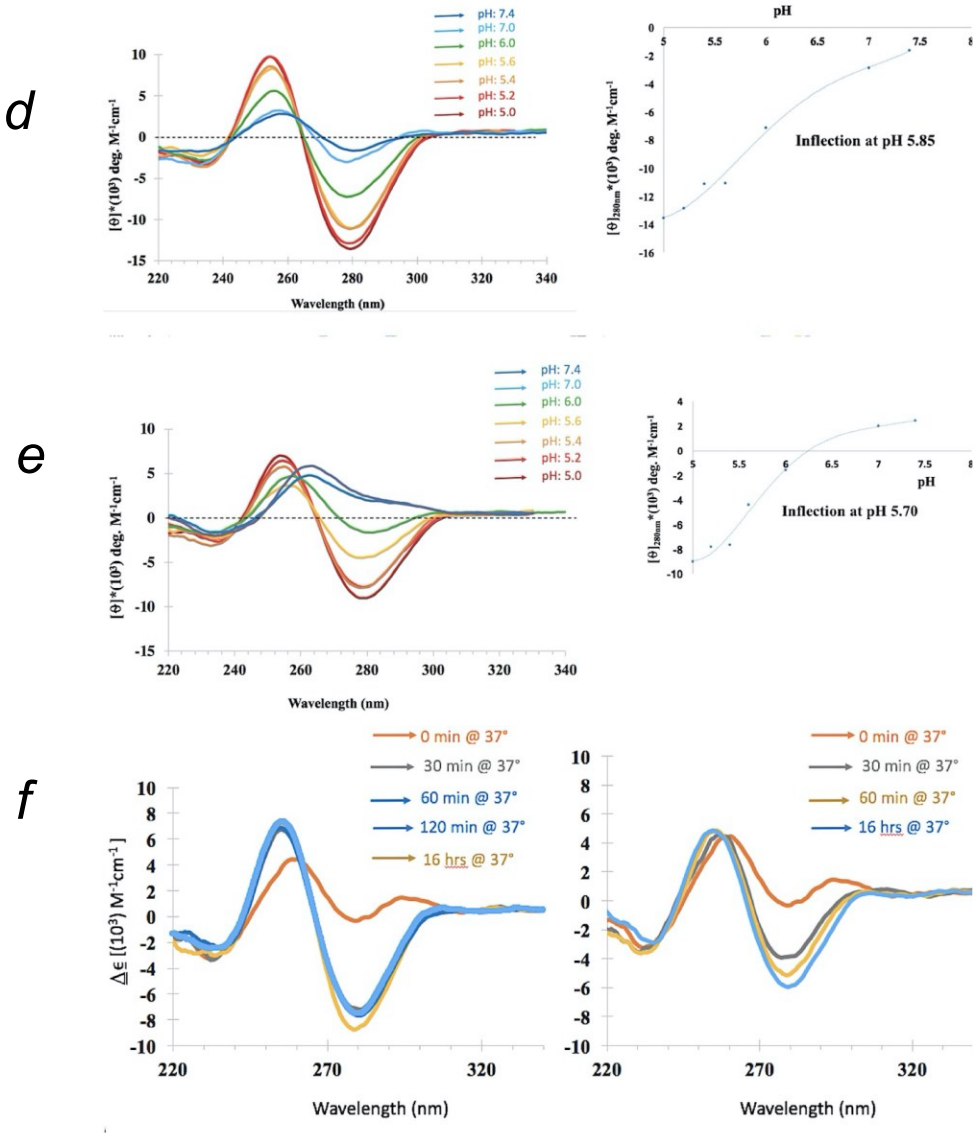


Figure 2-1. Formation of iCD-DNA: Related Kinetics and pH dependence. (a) CD spectra of 20 μM $d(C_2G_4)_7$ in TE Buffer plus 150 mM LiCl, pH 7.4. 700 μM DNA in this Buffer, at 37°C was incubated for 14 hrs and 5 days. CD spectra were taken immediately following dilution to 20 μM DNA in the same buffer, and measured at 22°C. **(b)** CD spectra of 20 μM $d(C_2G_4)_7$ in 150 mM lithium citrate buffer at different pH (5, 5.2, 5.4, 5.6, 6) as well as in TE Buffer plus 150 mM LiCl (at pH 7.0 and 7.4). Incubation was done at 700 μM concentration in the respective buffers for 14 hrs at 37°C and CD spectra was taken shortly after dilution to 20 μM DNA in appropriate buffer. **(c)** θ_{280} from **Figure 2-1.b** plotted as a function of pH. **(d) left:** CD spectra of 700 μM $d(C_2G_4)_7$ incubated in 150 mM lithium citrate, at different pH values, for 5 days at 37˚C, then diluted to 20 μM DNA in the buffer of the respective pH. Following dilution, the CD spectra were measured immediately. **Right: θ_{280}** values from figure at left, plotted as a function of pH. **(e) Left**: CD spectra of 700 μM $d(C_2G_4)_7$ incubated in 150 mM lithium citrate, at different pH values, for 5 days at 37˚C, then diluted to 20 μM DNA in the buffer of the same pH, followed by incubation at 37˚C for a further 14 hours prior to CD measurement. **Right**: **θ_{280}** values from figure at left, plotted as a function of pH. **(f)** CD spectra of $d(C_2G_4)_7$ incubated at 700 μM (*left*) and 20 μM (*right*) in 150 mM lithium citrate, pH 5.2. Left: 700 μM $d(C_2G_4)_7$ incubated in 150 mM lithium citrate, pH 5.2, at 37˚C, for the time indicated, followed by dilution to 20 μM of $d(C_2G_4)_7$ in 150 mM lithium citrate and immediate CD measurement. Right: 20 μM $d(C_2G_4)_7$ incubated in 150 mM lithium citrate, pH 5.2, at 37˚C for the times indicated. Incubations carried out under the two conditions, above, for 3 days, gave superimposable CD spectra.

DNA secondary structures known to show inverted CD spectra include the left handed Z-DNA duplex formed by $d(CG)_n$ in 4.0 M Na^+[149] as well as one reported instance of a left-handed G-quadruplex ("Z-G4") formed in the presence of ~100 mM K^+ at pH 7.0 by the DNA oligomer $d[T(GGT)_4TG(TGG)_3TGTT]$[131]. I measured the CD spectrum of $d(CG)_{25}$ in 4.0 M Na^+ (pH 7.0), as well as that of the Z-G4 G-quadruplex in 100 mM K^+ (pH 7.0), and compared them with the spectrum of $d(C_2G_4)_7$ in 150 mM lithium citrate, pH 5.2.

Figure 2-2.d shows these spectra, as well as the spectrum of the K^+-generated G-quadruplex products formed by $d(C_2G_4)_7$. It is clear that the "iCD-DNA" spectrum is utterly distinct from that of Z-DNA. With regard to the Z-G4, while its negative molar ellipticity region (270–290 nm) is roughly similar to that of iCD-DNA, the two spectra diverge significantly in the 230–270 region. It is therefore clear that iCD-DNA is not the left-handed Z-DNA duplex, though it may potentially have structural affinities with the one described left-handed G-quadruplex.

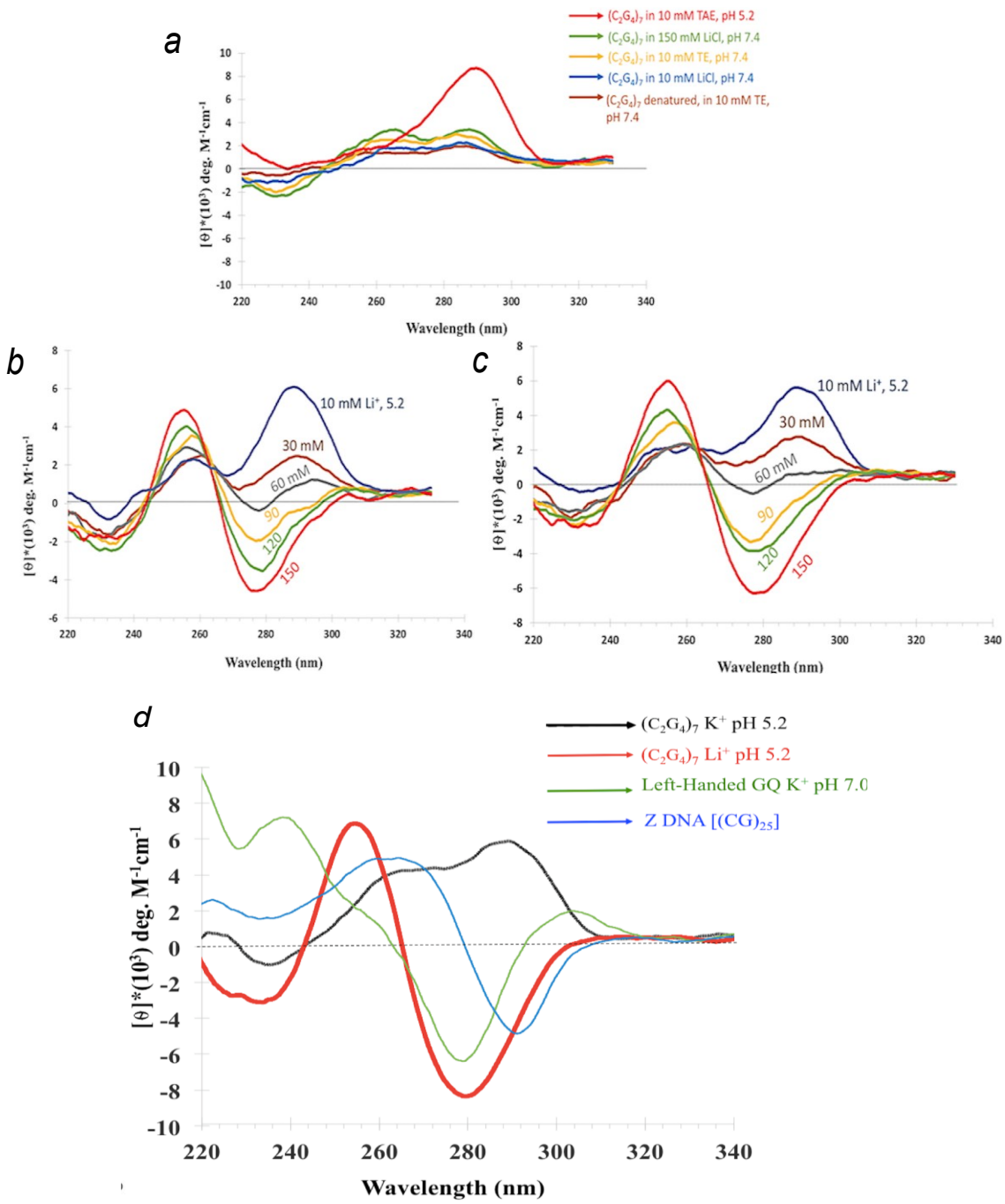


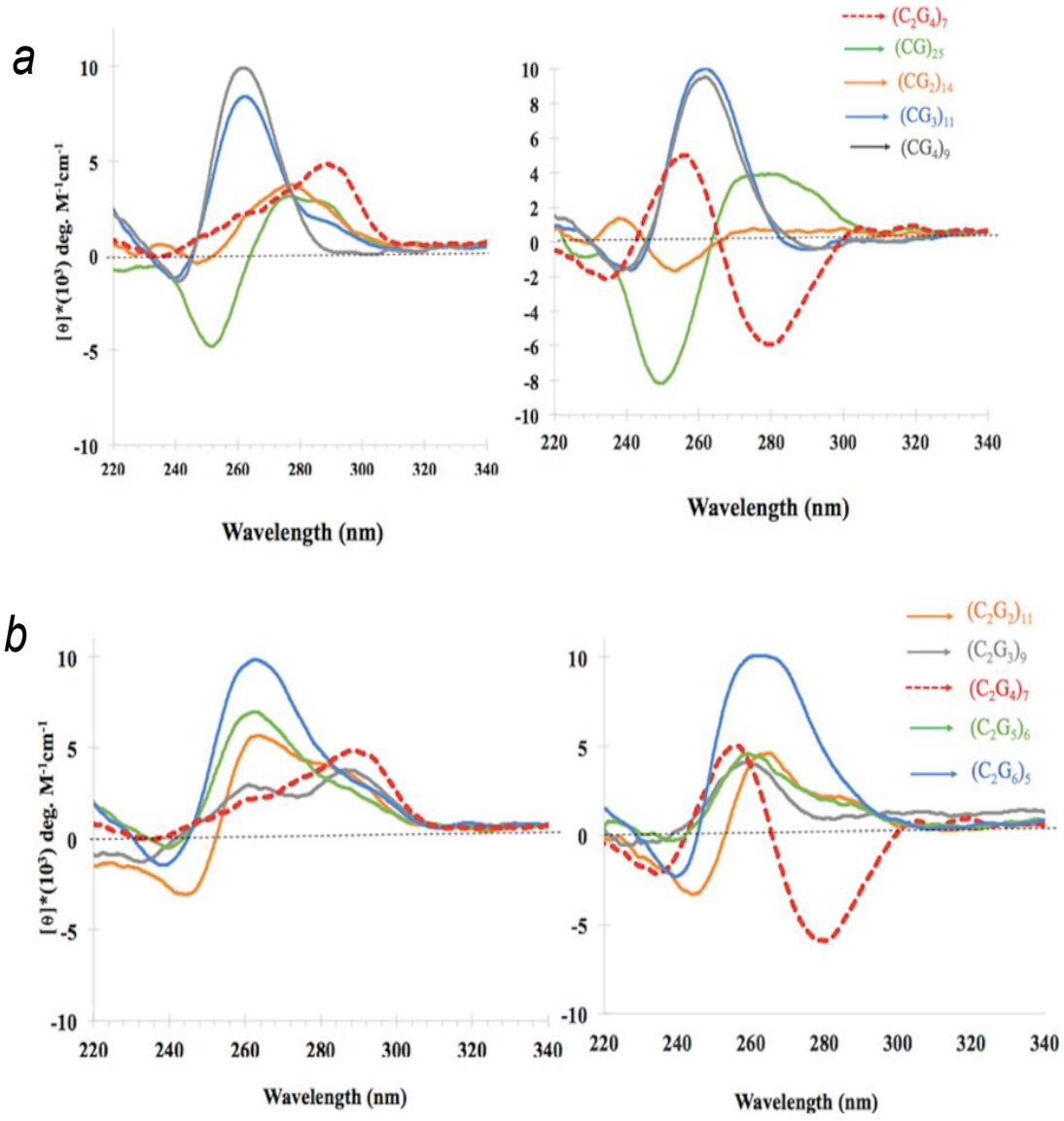
Figure 2-2. iCD-DNA formation as a function of pH and Li^+ as well as its comparison with Z-DNA and Z-G4. **(a)** CD spectra of 20 μM d$(C_2G_4)_7$, incubated for 2 hrs at 37˚C in buffers of various ionic strengths, all at pH 7.4. **(b)** and **(c)** CD spectra of 20 μM d$(C_2G_4)_7$, incubated at 37˚C in different concentrations of Li buffer, pH 5.2, for 2 hrs (b); and for 14 hrs (c). **(d)** Circular dichroism spectra of $(C_2G_4)_7$ in 150 mM potassium citrate, pH 5.2; $(C_2G_4)_7$ in 150 mM lithium citrate pH 5.2; a left-handed G-quadruplex (GQ) [ZG4: (T (GGT)4TG(TGG)3TGTT)] in TE (10mM Tris, 0.1mM EDTA, pH 7.0) plus 150 mM KCl; and a duplex Z-DNA [$(CG)_{25}$] in TE plus 4.0 M NaCl.

2.3.2. Do other GC repeats show inverted CD spectra?

Is the $(C_2G_4)_n$ sequence unique among GC-rich repeating sequences in forming iCD-DNA? I measured the CD spectra of a number of different GC-rich repeat sequences after the oligomers were incubated in either 150 mM $4EM^+$, pH 5.2 ("4EM buffer"); or in 150 mM lithium citrate, pH 5.2 ("lithium buffer"). **Figure 2-3.a & b** show the CD spectra of a variety of such repeating G/C-rich DNA oligomers. Both figures show that in 4EM buffer (*left*) none of the DNA oligomers shows a spectrum with the inversion features of iCD-DNA; in lithium buffer (*right*) the $d(CG_3)_{11}$ and $d(CG_4)_9$ oligomers show minor negative molar ellipticities in the 280–300 nm region, though not resembling the iCD-DNA spectrum to any great extent.

Figure 2-3.c compares spectra for $d(C_2G_4)_7$ with those of its complementary sequence, $d(C_4G_2)_7$. The $d(C_4G_2)_7$ sequence, which forms either i-motifs[150] or unusual quadruplexes proposed to contain C-G-C-G quartets[151] in the absence of potassium, does not generate the iCD-DNA spectrum in either incubation solution. **Figure 2-3.d** shows the spectra of two oligomers, $d(C_3G_4)_6$ and $d(C_3G_6)_5$. Like $d(CG_3)_{11}$ and $d(CG_4)_9$, $d(C_3G_4)_7$ shows a modest negative molar ellipticity in the 270–300 nm region, but again, its spectrum does not feature the intense negative ellipticity in this region characteristic of $d(C_2G_4)_7$.

I examined the ability of guanine-rich repeat sequences lacking cytosine to form iCD-DNA. **Figure 2-4** show that neither $d(T_2G_4)_7$ nor $d(A_2G_4)_7$ show the iCD-DNA spectrum over a pH range of 4.0-7.4


a
b
[θ]*(10³) deg. M⁻¹cm⁻¹
Wavelength (nm)
(C2G4)7
(CG)25
(CG2)14
(CG3)11
(CG4)9
(C2G2)11
(C2G3)9
(C2G4)7
(C2G5)6
(C2G6)5

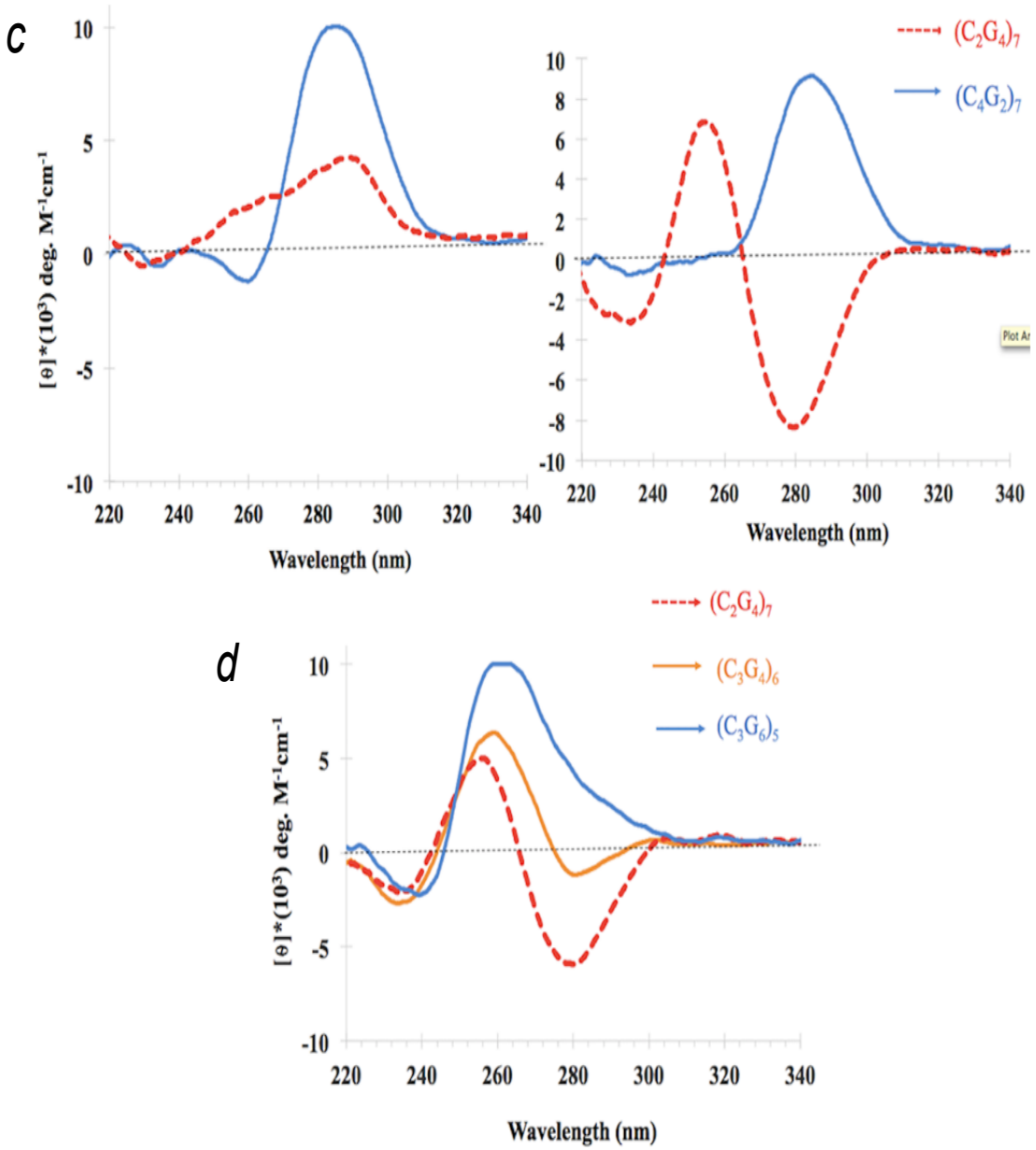


Figure 2-3. Effect of varying G:C ratios away from C_2G_4 on the formation of iCD-DNA. (a-c) Circular dichroism spectra of DNA repeats containing varying C:G ratios, independently incubated and diluted either in 150 mM 4-ethylmorpholine, pH 5.2 (left) or 150 mM lithium citrate, pH 5.2 (right); in all cases, given the different molecular weights of different oligonucleotides, the DNA mass was kept equal in each solution. **(d)** data, as above, but incubated and diluted only in 150 mM Lithium Citrate, pH 5.2.

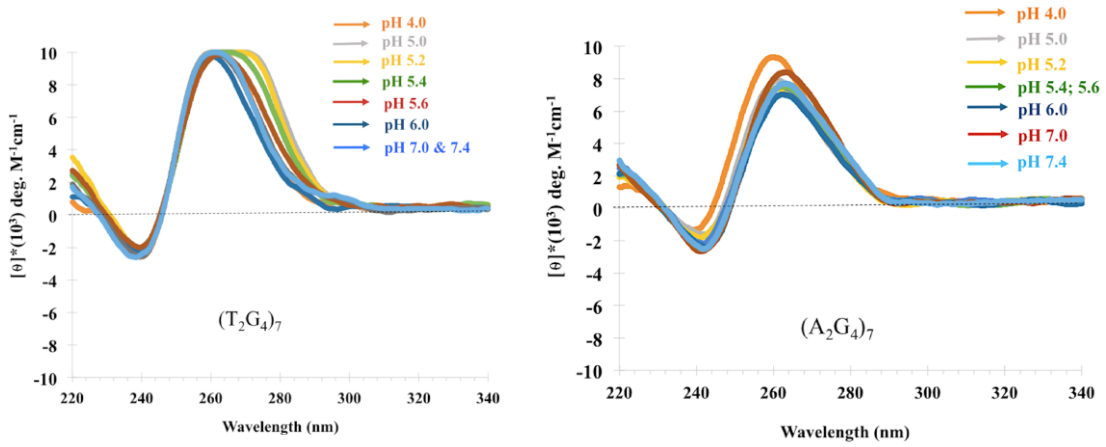


Figure 2-4. CD Spectra of $d(T_2G_4)_7$ and $d(A_2G_4)_7$ at different pH. None of sequence $d(T_2G_4)_7$ or $d(A_2G_4)_7$ show the negative CD spectra shown by $d(C_2G_4)_7$.

2.3.3. The role of counter-cations in iCD-DNA formation

Figure 2-5 shows the influence, variously, of 150 mM K^+; 150 mM Li^+; 10 mM Mg^{2+}; 10 mM Ca^{2+}; 150 mM 4-ethylmorpholinium$^+$ (4EM$^+$: pK_a = 7.67); and of 50 μM spermine^{4+} plus 150 mM 4EM$^+$. To generate the iCD-DNA conformer the following standard protocol was followed: DNA was incubated at 700 μM concentration independently in the above buffers, all at pH 5.2, at 37°C for 14 hrs, following which the solutions were diluted to 20 μM DNA in the same buffers. CD spectra were taken both immediately following dilution as well as after 14 hours of further incubation at 37°C following dilution. **Figure 2-5 .a**, *left* and *right*, show the CD data taken immediately following dilution and 14 hours after dilution, respectively. It can be seen that there is not a large difference in the two sets of spectra. Therefore, once formed, iCD-DNA doesn't change substantively over time. With regard to the individual incubations, both Li^+ and Mg^{2+} strongly support iCD-DNA formation; Ca^{2+} does so less efficiently; while the organic cation, 4-ethylmorpholinium, with or without added spermine, does not support it. The K^+ spectrum refers to G-quadruplex structures formed by $d(C_2G_4)_7$.

I wished to test for the stability/persistence of iCD-DNA in the presence of K^+, a cation known specifically to stabilize G-quadruplexes. 700 μM $d(C_2G_4)_7$ was incubated in 150 mM lithium citrate for 14 hrs at 37°C, followed by dilution to 20 μM of $d(C_2G_4)_7$ in different buffers. **Figure 2-5.b** shows CD spectra taken in 150 mM lithium citrate, pH 5.2 ("Li^+"), immediately following dilution; "Li^+/Mg^{2+}": spectra taken immediately following

dilution into 4 mM lithium citrate plus 10 mM magnesium acetate, pH 5.2. "$Li^+/Mg^{2+}/K^+(1)$" shows spectra taken 15 mins after dilution into a Li-Mg-K buffer (4 mM lithium citrate, 10 mM magnesium acetate and 25 mM potassium citrate, pH 5.2); and "$Li^+/Mg^{2+}/K^+(2)$" shows spectra taken 14 hrs after dilution into the Li-Mg-K buffer. It can be seen that even short incubations at 37° C after addition of K^+ lead to a disruption of the iCD-DNA spectra, and after 14 hrs in the presence of K^+, the CD spectra essentially resemble those of G-quadruplex structures formed in K^+ alone. To determine how much K^+ could be tolerated in this system, I carried out experiments exactly as above, except with potassium citrate, pH 5.2, added to 10 mM; 1 mM; and 0.1 mM (**Figure 2-5.b**). In 1 mM K^+, the inverted iCD-DNA spectrum persisted, although with lower amplitude, even after 14 hours of incubation at 37°; in 0.1 mM K^+, the iCD-DNA spectrum was stable even after 14 hours of incubation.

To investigate whether a minimum number of repeats of (C_2G_4) are necessary for iCD-DNA formation, I examined oligomers of the $d(C_2G_4)_n$ series, where n = 2–7. **Figure 2-6** shows spectra corrected to ensure a constant DNA mass (rather than molar concentration of oligomer), and it can be seen that under these experimental conditions $d(C_2G_4)_2$ does not form iCD-DNA; the larger oligomers do so progressively, until no further spectral difference can be seen between $d(C_2G_4)_6$ and $d(C_2G_4)_7$.

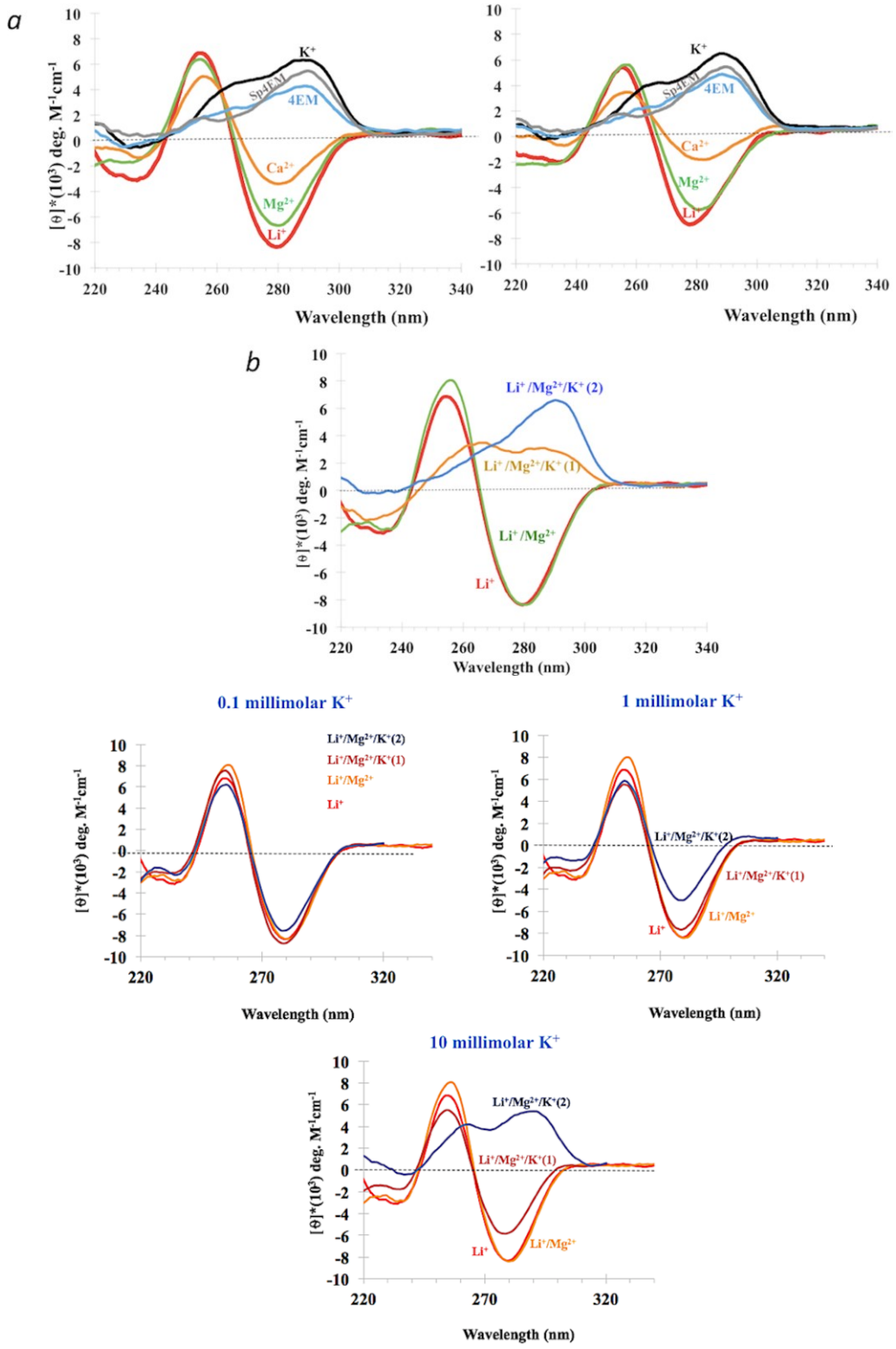


Figure 2-5. Effect of different counter-cations on iCD-DNA formation and stability. (a) Circular dichroism spectra of 20 μM $d(C_2G_4)_7$ in various buffered salt solutions at pH 5.2. The spectra on the left were taken immediately following dilution; the spectra on the right were taken 14 hrs following dilution and further incubation in the various buffers at 37˚C. **(b)** CD spectra of 20 μM $d(C_2G_4)_7$ diluted into different buffers at pH 5.2. All CD measurements were taken at 22℃.

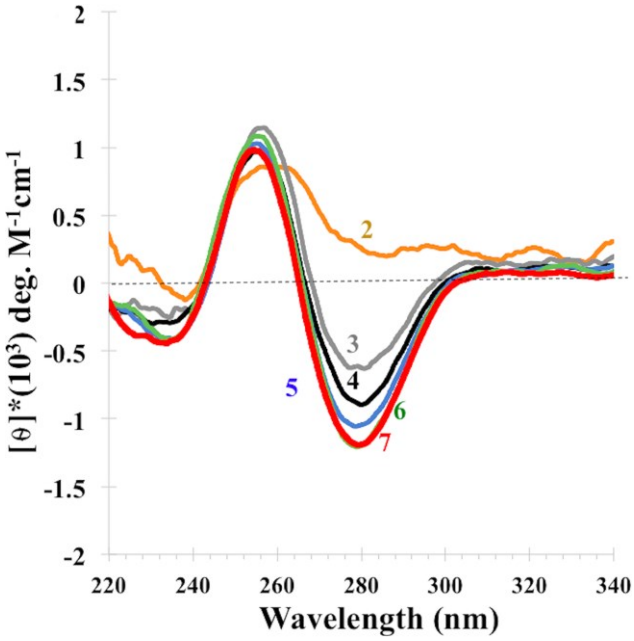


Figure 2-6. C_2G_4 repeat-length dependence on the formation of iCD-DNA. CD spectra of $d(C_2G_4)_n$ (where n = 2–7), in 150 mM lithium citrate, pH 5.2, at DNA concentrations adjusted to ensure unvarying DNA mass from solution to solution. Each DNA was incubated at 700 μM concentration in 150 mM lithium citrate, pH 5.2, at 37℃ for 14 hrs, following which it was diluted to ensure constant DNA mass into the same buffer as follows—"7": 2.85 μM $d(C_2G_4)_7$; "6": 3.33 μM $d(C_2G_4)_6$; "5": 4.0 μM $d(C_2G_4)_5$; "4": 5.0 μM $d(C_2G_4)_4$; "3": 6.66 μM $d(C_2G_4)_3$; and "2": 10.0 μM $d(C_2G_4)_2$.

2.3.4. The melting behavior of iCD-DNA

Figure 2-7.a shows the CD spectra of pre-formed iCD-DNA as a function of solution temperature, measured in buffered 150 mM lithium citrate, pH 5.2 ("lithium buffer"). The monotonic decomposition of the inverted CD spectrum (i.e. the lack of appearance of any other classic spectrum corresponding to either A- or B-family DNA duplexes, or canonical triplexes or right-handed G-quadruplexes) indicates that iCD-DNA has a homogenous structure that melts directly to unstructured, single-stranded DNA. **Figure 2-7.b** plots melting curves obtained by plotting θ_{280} values of $d(C_2G_4)_7$ in its iCD-DNA form, with data shown both for iCD-DNA in buffered 10 mM magnesium acetate, pH 5.2 ("magnesium buffer"); and in lithium buffer. Smooth two-state melting behaviour is observed in both cases, with T_m values calculated at 63°C in magnesium buffer and 60°C in lithium buffer.

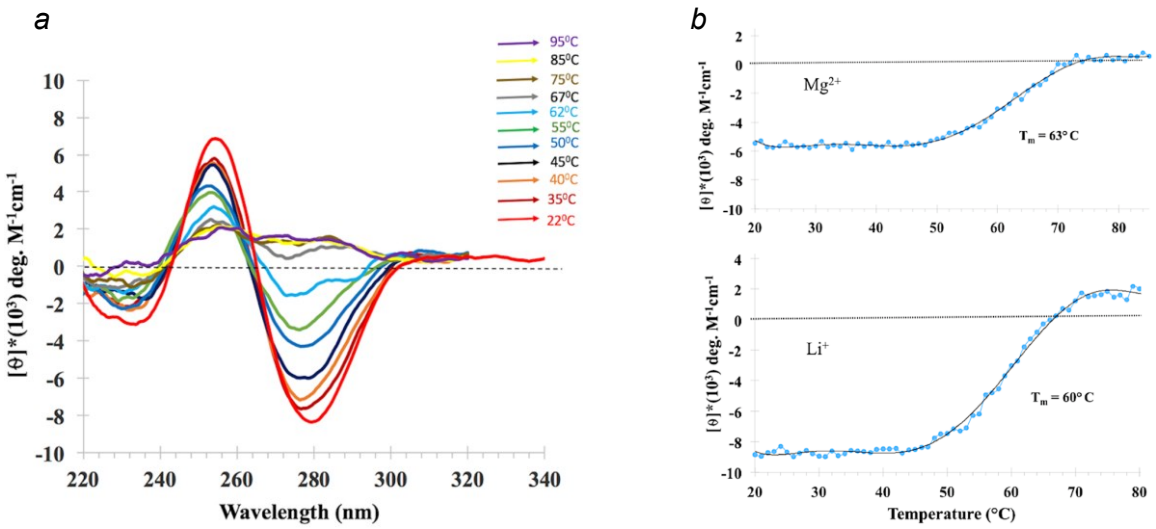


Figure 2-7. CD Melting profile of iCD-DNA. **(a)** Full CD melting profile of iCD-DNA incubated in Lithium Citrate, pH 5.2. **(b)** Melting Curves (Molar Ellipticity at 280 nm as a function of temperature) for iCD-DNA generated from incubation in either 10 mM Magnesium Acetate pH 5.2 (Top) or 150 mM Lithium Citrate pH 5.2 (bottom). Heating rate was 5° C/min.

2.3.5. Gel mobility and chemical protection data on iCD-DNA

Whether iCD-DNA consists of a single or multiple molecular species was examined by native gel electrophoresis. $d(C_2G_4)_7$ was first incubated, at different DNA concentrations (30 and 700 μM), for 1 or 14 hours at 37°C in lithium buffer. The resulting incubations were run in a 7.5% polyacrylamide non-denaturing gel run in TAE-Li buffer, pH 5.2.

Figure 2-8.a shows the data. Curiously, both sets of incubations gave rise to two distinct electrophoretic bands ("s": slower, and "f": faster). The same result was found with a 700 μM incubation of $d(C_2G_4)_4$, though the "s" band was overwhelmingly abundant for this oligomer. How robust were these "f" and "s" complexes—did their relative distribution in the native gel reflect their relative abundance in solution? To test this, "f" and "s" complexes from the lithium incubations of $d(C_2G_4)_7$ were excised and eluted from the native gel into lithium buffer, concentrated to ~5 μM without resorting to ethanol precipitation, and re-run into the native gel (TAE-Li buffer, pH 5.2).

Figure 2-8.d shows that ≥ 90% of each purified complex re-ran with its characteristic gel mobility. This suggests that the two complexes are generally stable and not in a rapid dynamic equilibrium under our incubation and dilution conditions.

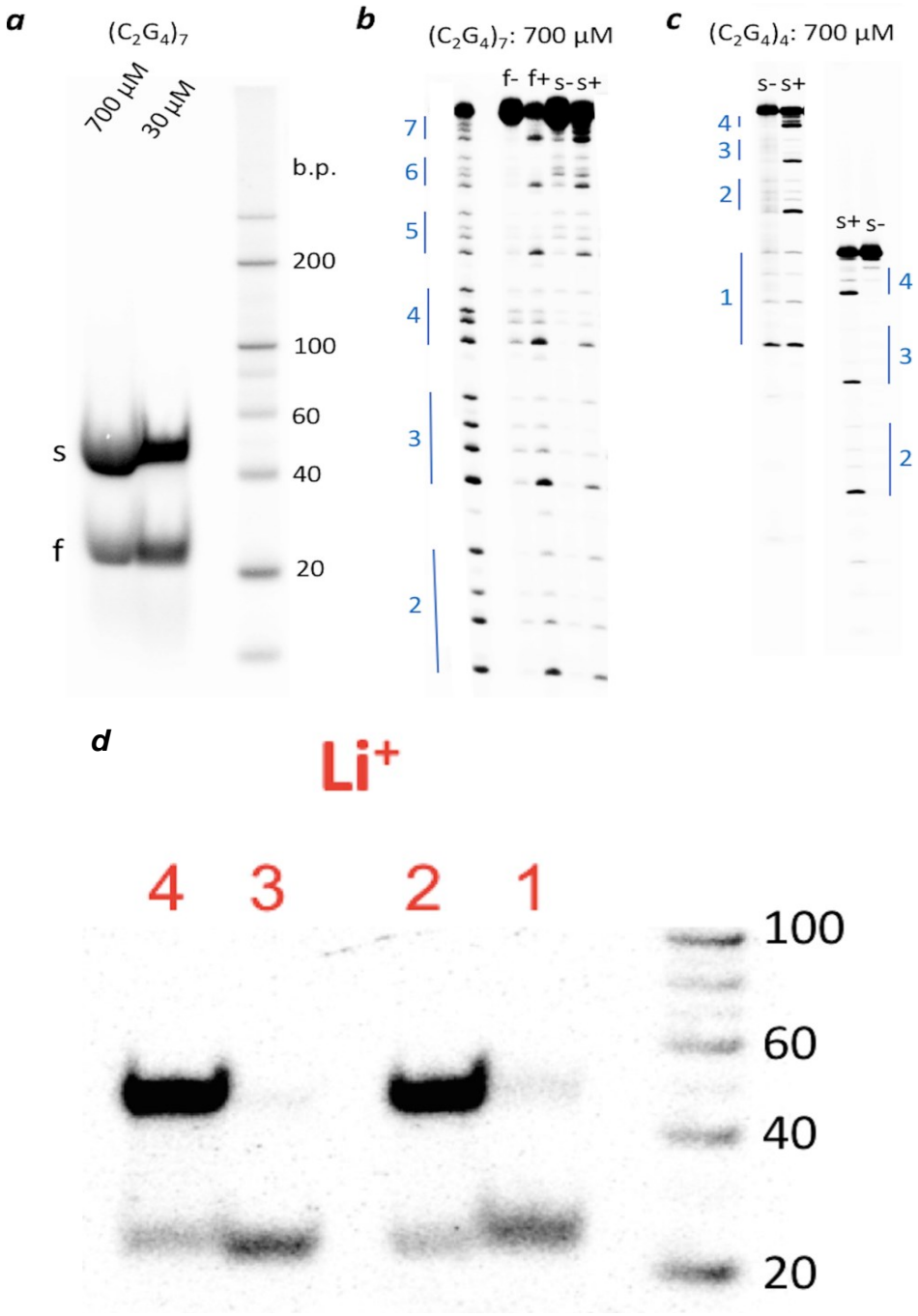


Figure 2-8. Native Gel and DMS footprinting of iCD-DNA. **(a)** Native gel analysis of $(C_2G_4)_7$ at two different concentrations. *s and f refer to slow- and fast-moving bands.* **(b-c)** DMS analysis of *s* and *f* DNA bands from $(C_2G_4)_7$ and $(C_2G_4)_4$. **(d)** Re-run of *s* and *f* species after elution from gel shown in **(a)** show a lack of interconversion of species. Lanes 1 and 2 refers to *f* and *s* species from 30 µM DNA incubation respectively and Lanes 3 and 4 refers to *f* and *s* species from 700 µM DNA incubation respectively.

Dimethyl sulfate (DMS) was used to try and define the base-pairing within the “s” and “f” complexes from the lithium buffer incubations. DMS selectively methylates guanines at their N7 position, which can be involved in Hoogsteen/Reverse Hoogsteen but not in Watson-Crick base pairing. **Figure 2-8.b** shows a 20% denaturing gel with the protection data for “f” and “s” complexes formed by $d(C_2G_4)_7$, and **Figure 2-8.c** shows the data for the predominant “s” band formed by $d(C_2G_4)_4$. A striking observation is that in all cases, the same distinctive methylation pattern can be seen, in which only the 5’-most guanine in a given GGGG stretch reacts strongly with DMS, while the other three are only modestly reactive or unreactive. Since DMS-methylation was carried out in the 30 μM or 700 μM DNA solution prior to loading on the native gel, it is therefore reasonable to deduce, since interconversion of the “f” and “s” complexes does not appear to be facile (**Figure 2-8.d**), that the “f” and “s” products represent fundamentally the same iCD-DNA structure, varying only in their strand stoichiometries.

Figure 2-9 shows a mixing experiment designed to investigate the strand stoichiometries of the “f” and “s” products seen in **Figure 2-8.a**. A slightly larger oligonucleotide than $(C_2G_4)_7$ was synthesized by adding a T_6 stretch to the 3’ end, to give a $(C_2G_4)_7T_6$ oligonucleotide. $(C_2G_4)_7$ and $(C_2G_4)_7T_6$, were now allowed to form iCD-DNA either individually (lanes 1, 3, 4, and 6), or as a mixture [equimolar $(C_2G_4)_7$ and $(C_2G_4)_7T_6$] (lanes 2 and 5). **Figure 2-9**.shows that from the mixtures, two distinct “f” bands formed while three distinct “s” bands formed (lanes 2 and 5). This is consistent with the “s” complex being a strand dimer and the “f” complex being a strand monomer (thus, the three “s” products seen from the mixture corresponding to $[(C_2G_4)_7]_2$; $(C_2G_4)_7•(C_2G_4)_7T_6$; and $[(C_2G_4)_7T_6]_2$).

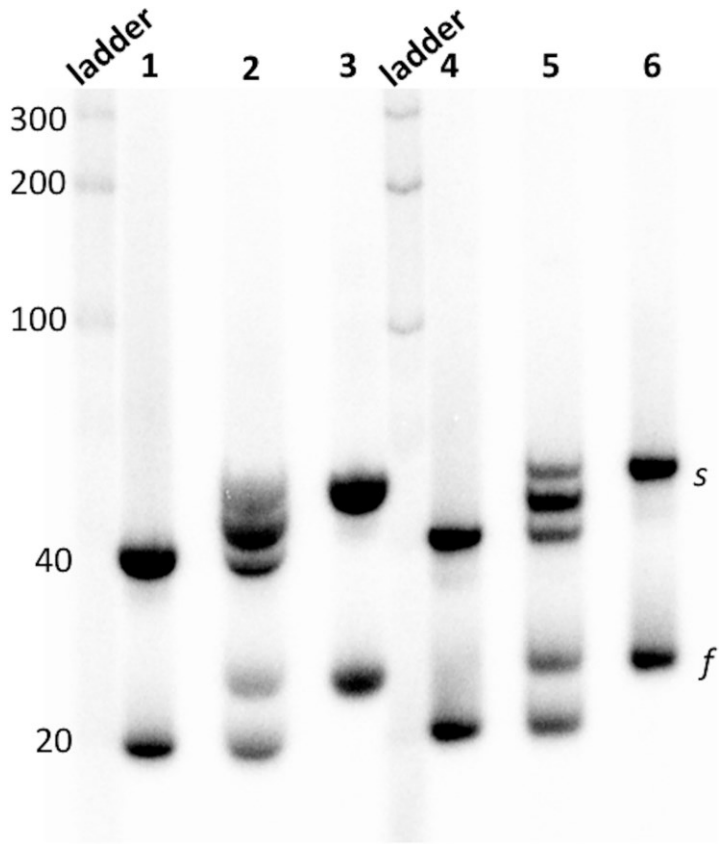


Figure 2-9. Investigation of the strand stoichiometries of the 'f' and 's' products. Lane 1-3 show incubation with [DNA] = 700 μM and lanes 4-6 show incubations with [DNA] = 30 μM. Lanes 1 and 4 show incubations of $d(C_2G_4)_7$; lanes 3 and 6 show incubations of $d(C_2G_4)_7T_6$; and, lanes 2 and 5 show incubations of equimolar mixes of $d(C_2G_4)_7$ and $d(C_2G_4)_7T_6$. The numbers shown to the left of the double-stranded ladder show the values of individual bands as base pairs.

To test whether the distinctive methylation pattern seen for $(C_2G_4)_7$ incubated in pH 5.2 lithium buffer (only the 5'-most G out of a GGGG stretch reacting strongly with DMS) is uniquely associated with iCD-DNA, I carried out DMS-methylation experiments on $(C_2G_4)_7$ incubations in pH 5.2 magnesium buffer (which also supports iCD-DNA formation, as defined by CD spectroscopy) and in pH 5.2 4EM and pH 5.2 potassium buffers (neither of which supports iCD-DNA formation). **Figure 2-10.c** shows that the methylation data in pH 5.2 magnesium buffer closely resembles the pattern found in pH 5.2 lithium buffer. **Figure 2-10.a & b**, however, shows that in pH 5.2 4EM buffer, neither the "f" nor the "s" complexes show methylation patterns characteristic of iCD-DNA (**Figure 2-8.b & c** and **Figure 2-10.b**). Notably, both the 5'-most and 3'-most guanines of a given GGGG tract are reactive to DMS. In pH 5.2 potassium buffer, expected to form G-quadruplexes, the methylation patterns are generally faint but resemble the pH

5.2 4EM buffer patterns more closely than those obtained from the two iCD-DNA supporting buffers. Most interestingly, pH 7.0 lithium buffer (**Figure 2-10.d**) gives methylation patterns for "f" and "s" that are distinct from each other, and both are very distinct from the iCD-DNA methylation signature. The "f" pattern resembles the G-ladder; whereas, the "s" pattern closely the pH 5.2 4EM pattern (**Figure 2-10.b**). It is clear therefore that both Li^+/Mg^{2+} and low pH are required for the distinctive methylation pattern (as well as CD signature) of iCD-DNA.

I investigated the methylation pattern of two other G/C-rich repeat sequences, $d(CG_4)_9$ and $d(CG_3)_{11}$, which have roughly the same molecular weight as $d(C_2G_4)_7$. Neither of these two new repeats shows the inverted CD signature characteristic of iCD-DNA (see **Figure 2-3.a**). **Figure 2-11.a** shows that in native gels run at pH 5.2, 700 μM oligonucleotide concentrations of $d(CG_4)_9$ and $d(C_2G_4)_7$ both run as two bands each, fast (f) and slow (s). Methylation data of these various products are also shown in **Figure 2-11.b**. It can be seen that the $d(CG_4)_9$-s and $d(CG_4)_9$-f complexes show methylation patterns distinct from each other as well as from $d(C_2G_4)_7$-s and $d(C_2G_4)_7$-f [the two $d(C_2G_4)_7$ complexes, of course, show similar patterns, with the 5'-most guanine in any **G**GGG stretch strongly methylated and the remaining three poorly/not methylated]. Strikingly, the $(CG_4)_9$-s complex shows the second guanine of each of its G**G**GG strongly methylated.

Figure 2-11.c-d shows the analogous native gel and methylation patterns for the "f" band formed from $(CG_3)_{11}$. Again, this methylation pattern is utterly different from those of the $(C_2G_4)_7$-s and $(C_2G_4)_7$-f complexes.

That distinctive **G**GGG iCD-DNA methylation pattern of iCD-DNA formed by $d(C_2G_4)_7$, however, does not immediately suggest a specific higher-order structure; most likely, there are a number of possible higher order folds of DNA can give rise to this methylation pattern. A methylation pattern alone is often insufficient to predict a detailed structure, given uncertainties about what kind of base-pairing may or may not occur particularly in various G-G base pairings. Nevertheless, this DMS protection pattern is useful to take into account for the building of one or more structural models for iCD-DNA, which are discussed, in the next section.

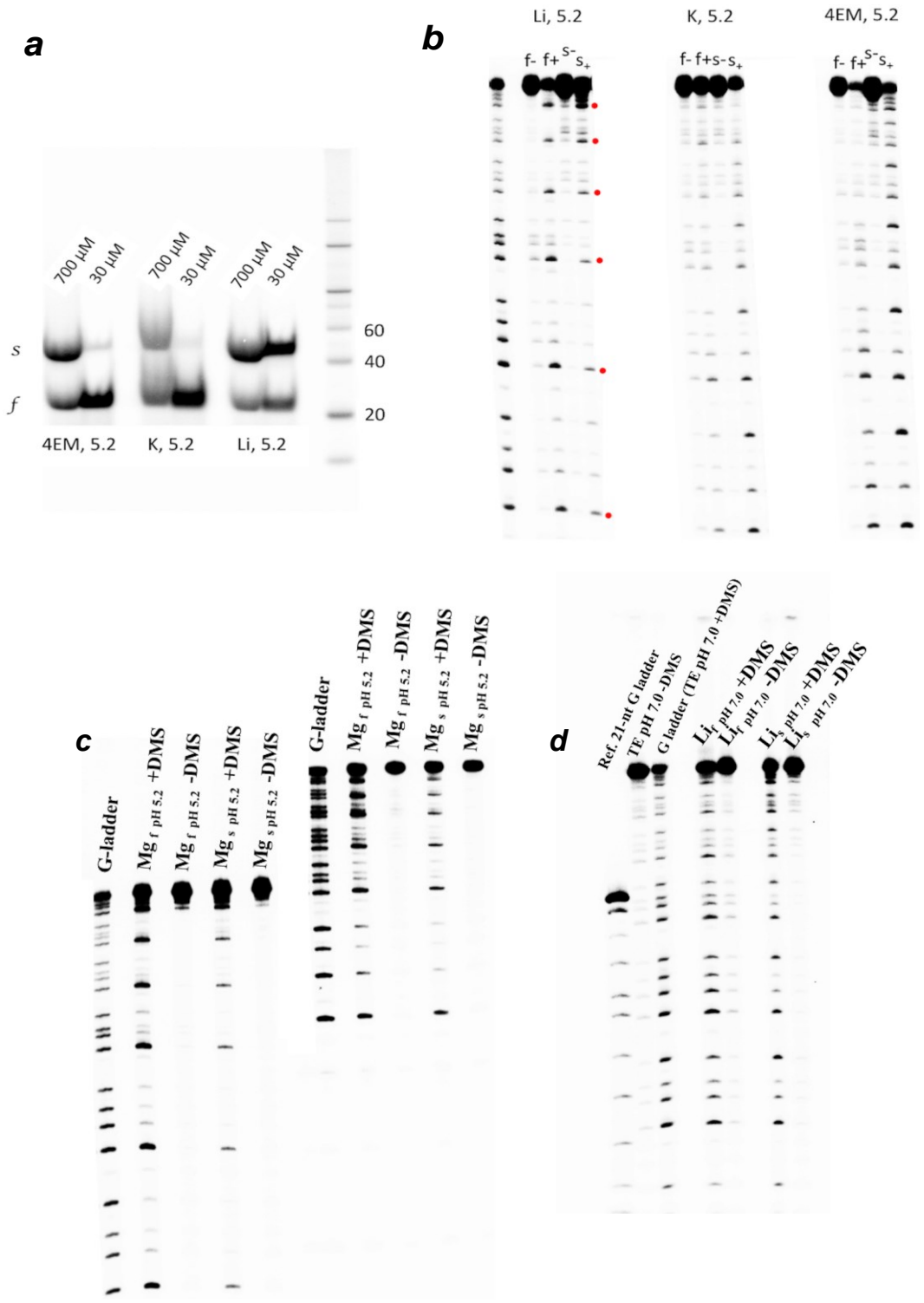


Figure 2-10. DMS footprinting of $(C_2G_4)_7$ under different salt and pH conditions. **(a)** Native gel run in TAE buffer pH 5.2 to show the product of incubation in Li buffer, K buffer and 4EM buffer. **(b)** Denaturing gel showing the DMS footprinting of *s* and *f*, product under these buffer conditions. **(c)** DMS footprinting of *s* and *f* product of $(C_2G_4)_7$ incubated in Magnesium Acetate, pH 5.2. The bands on the left and right side of the gel represent loadings at different times on the gel, to enable visualization of all seven repeats of (C_2G_4) in the $d(C_2G_4)_7$ oligonucleotide forming the iCD-DNA. **(d)** DMS pattern of *s* and *f* band from the incubations product of $d(C_2G_4)_7$ in Li buffer pH 7.0.

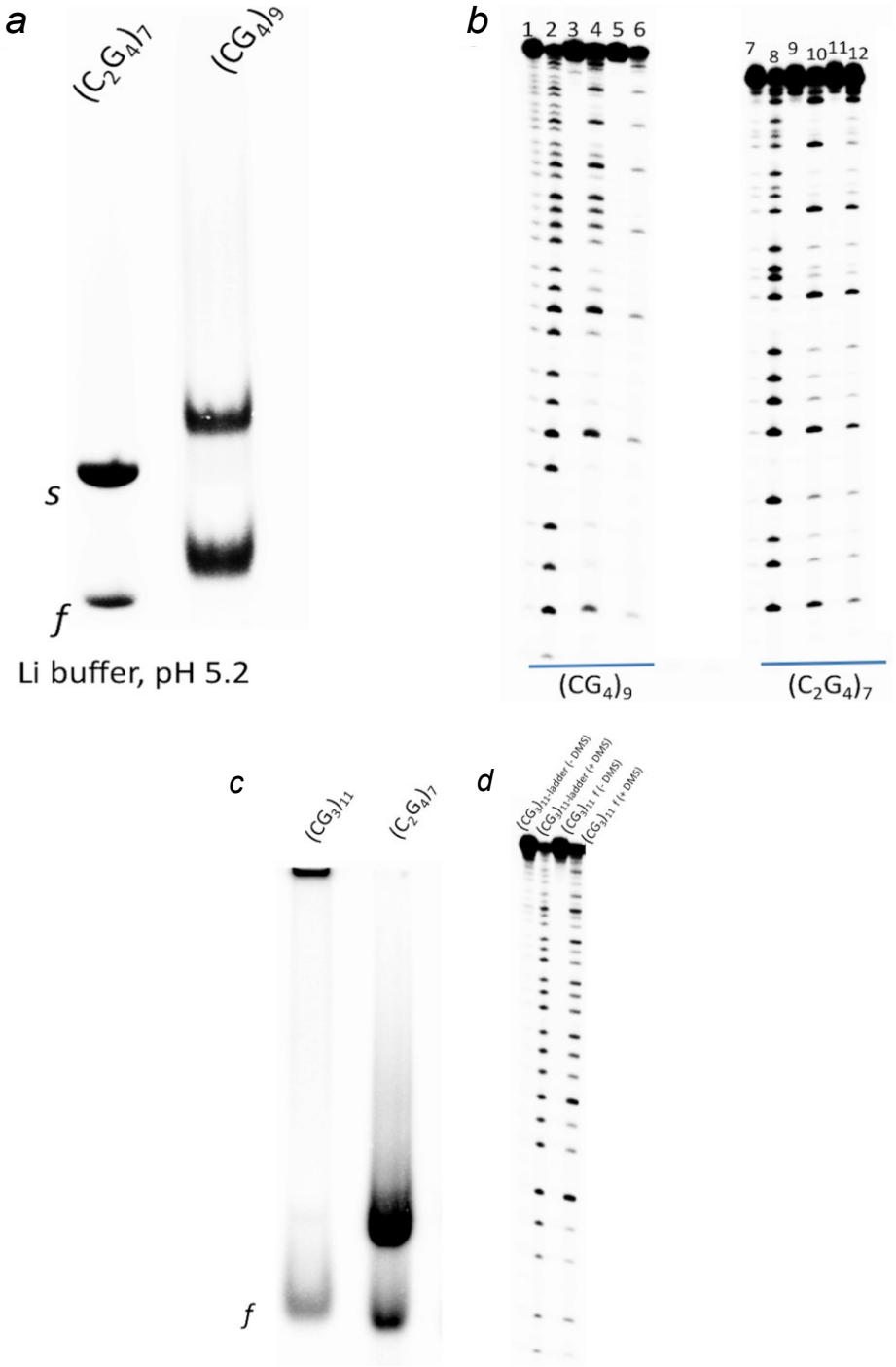


Figure 2-11. DMS footprinting of different DNA constructs as compared to d(C_2G_4)$_7$. (a,c) Native gel run in TAE buffer at pH 5.2 showing product of incubations of d(C_2G_4)$_7$, d(CG_4)$_9$ and d(CG_3)$_{11}$. **(b)** DMS footprinting for the *s* and *f* product of (C_2G_4)$_7$ and (CG_4)$_9$. Lanes 1, 7: G-ladder no DMS; Lanes 2, 8: G-ladder yes DMS; Lanes 3,9: *f* band no DMS; Lanes 4, 10: *f* band yes DMS; Lanes 5, 11: *s* band no DMS; Lanes 6, 12: *s* band yes DMS. **(d)** DMS footprinting of the *f* band of (CG_3)$_{11}$.

2.3.6. Structural models for iCD-DNA

To list what the above experiments reveal about iCD-DNA, I have the following: (a) an acidic pH is required for iCD-DNA formation; suggesting that the protonation of one or both cytosines in each C_2G_4 repeat is likely an important contributor; (b) the DMS methylation data show distinctive and consistent pattern, with the 5'-most G of each GGGG stretch reactive to DMS, and the others substantially protected; this holds true for both the "f" and "s" bands of iCD-DNA seen in acidic native gel (suggesting that "f" and "s" are effectively the same complex albeit with different strand molecularity; (c) Li^+ and Mg^{2+} cations are required for iCD-DNA formation; Ca^{2+} is only marginally effective, and a bulky organic monovalent cation, $4EM^+$, is ineffective. Spermine^{4+} is also ineffective. (d) The inverted CD signature of iCD-DNA suggests it is a structure not yet recorded in the literature[127] ; the partial similarity of this CD spectrum to that of one reported instance of a left-handed G-quadruplex[131] indicate that iCD-DNA may be an unusual variant of the classic G-quadruplex (which normally requires Na^+, K^+, or Sr^{2+} cations to form); indeed, I found that iCD-DNA converts relatively efficiently to classic G-quadruplexes when K^+ is added to iCD-DNA in lithium buffer. (e) The two-state melting curve suggests the formation of a homogenous structure, which directly melts to unstructured single stranded DNA. Certain classes of DNA helical structures, such as triple helices, generally show more complex melting behavior, with the Hoogsteen/Reverse-Hoogsteen bonded third strand melting away from the duplex at the lower temperature than the duplex itself, though there have been reports of the two-melting transition (i.e. between three states) located close to each other[152–158]. Thus, the observation of a single melting event between two states for iCD-DNA is not in itself sufficient to rule out the possibility of a conventional triplex, although the strand composition of $d(C_2G_4)_7$ is not formally suitable for forming a canonical YRR or YRY triplex.

It is possible to eliminate certain classes of higher-order DNA structure for iCD-DNA. First, the uniquely inverted CD spectrum of iCD-DNA rules out the possibility of B- or A- family double helices[127] ; right-handed Hoogsteen duplexes[159] , as well as conventional, right-handed G-quadruplexes[139] and classic i-motif structures[150].

So, what could iCD-DNA's structure be? Protonated cytosines are known to participate in Hoogsteen/Reverse Hoogsteen bonding[159,160] as well as in forming i-motifs[161]. Most simply, iCD-DNA could be left-handed Hoogsteen-bonded duplexes, "f"

being an intramolecular folded form, and “s” an intermolecular form involving two distinct strands. However, two further classes of structure that I propose here (below) do involve protonated cytosines in a more complex structures. While $d(C_2G_4)_n$ repeats contain bases that normally Watson-Crick base-pairs with each other, the requirement for acid pH to form iCD-DNA suggest that i-motifs may still be forming, even given the 2:1 excess of guanines over cytosine in the $d(C_2G_4)_n$ repeats. I propose that iCD-DNA may consist of short i-motifs stretches separating loose (i.e. not stabilized by K^+) G-quadruplexes, which could well be left-handed and so contribute to the inverted CD spectrum of iCD-DNA (**Figure 2-12**). Two alternative structures can be contemplated, which differ in the specifics of base-pairing. **Figure 2-12.a** shows a structure that contains only the i-motifs and loose G-quadruplexes, the interdigitated structure of i-motifs helping to hold together the Li^+ (or Mg^{2+} but not $4EM^+$)-stabilized, relatively loose G-quadruplex, whose outermost guanines (typically, only the 5’ G of a given run) could be susceptible to DMS-methylation. **Figure 2-12.b** shows a possible variant of the above structure, this one incorporating GCGC base quartets in addition G-quartets. Classic GCGC quartets have been observed in high-resolution structures of certain G-quadruplexes[162]. The “f” and “s” bands seen in the native gel of iCD-DNA refer to monomeric and dimeric complexes, respectively (*vide infra*).

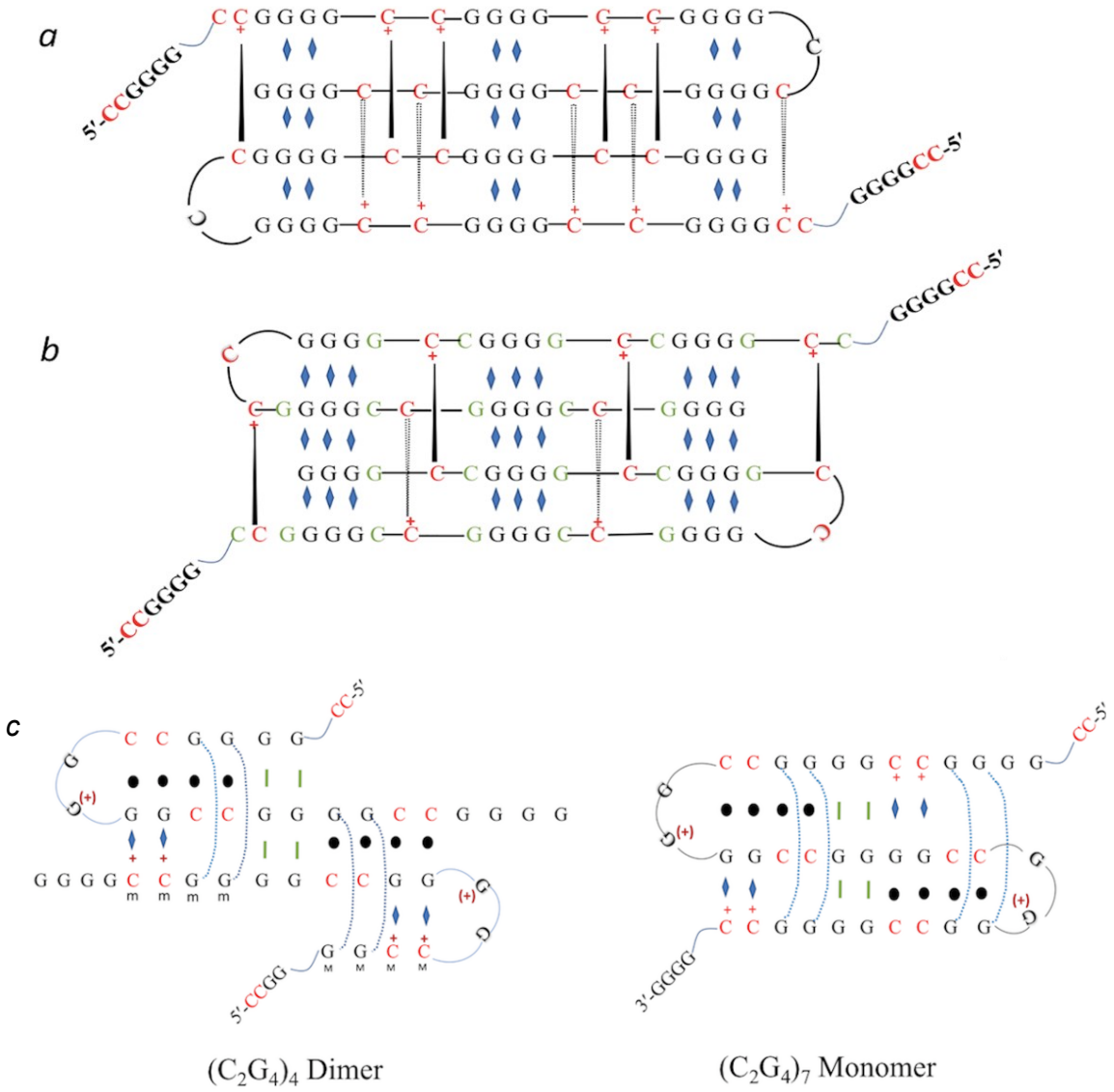


Figure 2-12. Proposed Structural Models for iCD-DNA. **(a)** Model consisting of purely i-motif and G-Quartets. **(b)** A related model but containing GCGC quartet in addition with i-motif and G-Quartet. **(c)** Braided Triplex model shown for two different repeats of d(C_2G_4) with different strand molecularity.

Alternatively, given the highly symmetric nature of the repetitive sequence $(C_2G_4)_n$, the potential exists for the formation of a non-canonical, braided or entangled structures, founded on Watson crick base-pairing between guanines and cytosines. Braiding occurs via 'partner swapping' of strands (or stretches of a given strand) participating in Watson-Crick base-pairing.

Figure 2-12.c shows two versions of such a 'braided' complex. The alternation of strands participating in Watson-crick base-pairing could be facilitated by conformationally fluid "buffer zones" made up of two consecutive G-triples. Such braided structures have been proposed by Bai and Colleagues to form from λ phage DNA; these authors carried out a computational simulation that featured alternating left-handed and right-handed helical elements[30,163]. Superficially, such braided structures would resemble DNA triplexes; although canonical triplexes (YRY and YRR, stabilized by Mg^{2+}, polyamines, and/or low pH) typically do not show inverted CD spectra, modest inversion (or close to zero ellipticity) has been observed at ~280 nm from certain "anti-parallel" triplexes where the third strand has a mixed purine and pyrimidine content[156]. Of course, the $d(C_2G_4)_n$ sequence is not suitable for forming canonical YRY and YRR triple helices (in which very little deviation is tolerated to the strict requirement for one all-purine and one all-pyrimidine strand forming a Watson-Crick duplex to which a third strand (all purine / all pyrimidine /purine-pyrimidine mixture) binds[17].

One prediction about such braided structures is that topological entanglement of the strands should override the strict canonical rules that hold for conventional triplexes (such as the requirement for the third strand to be anti-parallel to the duplex's purine strand in YRR triplexes and parallel to that strand in YRY triplexes). Each entrapped GGCC "third strand" stretch in a braided complex would therefore base-pair either conventionally (i.e., via Hoogsteen or reverse Hoogsteen base pairing) or unconventionally with the Watson-Crick base-paired tract adjacent to it. Another prediction is that in order to remain conventionally right-handed, the 'third strand' would need necessarily to alternate between lying in the major and minor grooves of the duplex. Precedent for minor-groove-bound third strands exist in RNA triplexes[164]. Alternatively, if the third-strand disposition within each triplex tract of iCD-DNA were required to be uniformly in the duplex's major groove, the triplex tracts would need to alternate from being left-handed and right-handed helices. Such dramatic changes in helical direction from tract to tract could, again, be enabled by conformationally fluid G-

triple “buffer zones”. In such braided structures, since non-canonical base triples would be expected to form, they could confound the ability to interpret the DMS methylation that I report here.

2.4. Discussion and Conclusion

I have reported here an unusual DNA structure—iCD-DNA—characterized by an inverted circular dichroism (CD) spectrum in the 220–310 nm wavelength range. iCD-DNA formation shows a pH dependence and optimizes in the pH range of 5.0–5.2. The inflection pH at equilibrium, for high DNA concentration (700 μM) incubations, is ~5.85, consistent with the pK_a for cytosine protonation (given that the inflection pH, even at equilibrium, may not precisely equate a pK_a value[165,166]). With my data, I am not able yet to propose a definitive structural model for iCD-DNA. Under my experimental conditions, native gel analysis has shown that two distinct species, albeit most likely of similar or identical structure, are obtained. This militates against immediate high-resolution structure determination using NMR spectroscopy or X-Ray crystallography. I have therefore proposed three general categories of structure that would likely be consistent with all the experimental data that I have obtained. These three proposed categories of structures (**Figure 2-12**) do have some typical distinctive features within them which can be probed, in future, using chemical footprinting experiments to distinguish them. i-motif / G-Quadruplex hybrid model, in particular, is expected to have more looped-out cytosines as compared to the braided-triplex model or the left-handed Hoogsteen duplex. Such looped-out cytosines can be distinguished from properly stacked and base-paired cytosines using chemical modifications of DNA bases such as by elemental bromine[135] and chloroacetaldehyde[17]. Elemental bromine leads to the bromination of unpaired cytosines whereas chloroacetaldehyde forms ethanoderivatives on the base-pairing position of cytosines, adenines and less prominently to guanines. All of these modified DNA bases can be detected at nucleotide resolution by piperidine cleavage followed by sequencing gel electrophoresis. Chloroacetaldehyde has been used successfully in past to confirm the presence of H-DNA both *in vivo* and *in-vitro*[17,167]. Further, antibody-mediated approach, using i-motif specific antibody iMab[32] and triplex-specific antibodies JEL 318 and JEL 466[168], can also be used to distinguish between i-motif / G-Quadruplex hybrid model and the braided triplex model. The repeating $(C_2G_4)_n$

DNA sequence from the human *C9orf72* gene is causally linked to the development of a number of neurodegenerative diseases (*vide infra*). This particular, very guanine-rich, repeat is known to favour the formation of G-quadruplexes in the presence of the potassium ion, the dominant monovalent cation in the intracellular environment. By contrast, the iCD-DNA structure reported here is destabilized by the presence of potassium ions. However, iCD-DNA formation is promoted by the Mg^{2+} ion, which is present in millimolar concentrations in the cell; so, the physiological relevance of iCD-DNA under specialized intracellular conditions cannot be ruled out. It is known that significant fluctuations of potassium ion can occur in the cell[169,170]. Although the exact quantification in the K^+ fluctuation is tough to measure given the transient and rapid dynamics of the K^+ concentration change, imbalance in the K^+ homeostasis has been shown to cause multiple diseases and abnormalities including hypertension, hyperkalemia, hypokalemia and chronic kidney disease[171]. Further, abnormalities in the functioning of potassium channel have been reported in the mouse model of Lou Gehrig's (ALS) disease[171,172]. In considering a broader picture of repeat expansion DNA sequences[173,174], many of which do not form G-quadruplexes but do form foldback hairpin structures, it is conceivable that with sufficiently long repeats, Watson-Crick base-paired tracts may switch strands, as in iCD-DNA, to give rise to braided structures; or, indeed, form non-conventional left-handed i-motif / G-quadruplex hybrids. Models proposed here for the iCD-DNA can have a potential involvement in the etiology of ALS/FTD. Formation of a topologically entangled iCD-DNA, in principle, can disturb the normal cellular machineries by either acting as a road-block to transcriptions, translations and replications or by forming DNA/RNA aggregates within the cell. Such aggregates can then sequester multiple proteins and thus, can further exaggerate the nature of pathology. Proposals somewhat akin to this have been made recently[175].

Chapter 3.

A long and reversibly self-assembling 1-D DNA nanostructure/nanowire built from Triplex and Quadruplex hybrid tiles

3.1. Introduction

DNA has long been considered as a promising material for the bottom-up construction of nanoscale objects and device[176,177,186,178–185], and this includes its utility for the construction of 1-dimensional "nanowires" (although electrical conductivity is not an obligate property of such structures)[187–190]. Much interest has focused on methods for assembling 1-D DNA nanostructures ("1DDN"), taking advantage of the well-known base-pairing and high-precision self-assembling properties of DNA[191–194]. Such 1DDN, whatever their inner architecture, are of potential interest for a variety of applications, including but not exclusive to their putative charge flow properties[195–198]. At a basic level, 1DDN can be understood as purely DNA-composed entities of substantial length (> 100 nm), formed by the programmed folding, base-pairing, and self-assembly of multiple DNA strands and/or oligonucleotides. DNA quadruple helices, such as G-quadruplexes and i-motifs, have proved to be particularly successful in the design of long and regular 1DDN (termed "G-wires" in the case of G-Quadruplexes")[199]. Another large class of DNA helices, triplexes (triple-helices)[17,28,200], however, have not been comparably explored for their promise to form such nanostructures. Making a purely triplex 1DDN (~100 nm) poses certain challenges. Using 'third strands' with long stretches of purely pyrimidine or purely purine bases to bridge non-continuous duplex units would likely generate issues of both proper phasing within and overall stability of the nanostructure so generated. Canonical DNA triple helices fall into two broad classes, YR*Y and YR*R (where the asterisk indicates the Hoogsteen/reverse Hoogsteen interaction); in these, a Watson-Crick duplex, in which one component is wholly purine ("R") and the other wholly pyrimidine ("Y"), is bound in the major groove by a wholly purine or wholly pyrimidine single strand to generate the triplex[17,28,200]. In the YR*Y triplex, the third (Y) strand binds via Hoogsteen hydrogen-bonding to the R strand of the duplex, with those two strands oriented parallel to each other. The canonical base-triples formed in the YR*Y triplex are

TA*T and CG*C^+; owing to the need to protonate cytosine in CG*C^+ triples, pH is an important factor in determining the stability of YR*Y triplexes containing CG*C^+[17,28,200].

Here I report the conception of as well as experimental data on a triplex-based 1DDN that is both assembled and precisely phased by the placement of short stretches of guanine base-quartets (G-quartets).

3.2. Materials and Methods

3.2.1. Chemicals and DNA oligonucleotides

All DNA oligonucleotides were purchased from the Core DNA Service Inc. (Calgary, Canada). The oligonucleotides were size-purified using 8% (containing 7 M urea) denaturing polyacrylamide gels. The DNA bands from such denaturing gel were excised, and the DNA eluted into TE buffer (10 mM Tris, 0.1 mM EDTA, pH 7.4), and then recovered via ethanol precipitation. The 70 % ethanol washed DNA pellet was then suspended in TE buffer to make up the desired stock concentration.

Oligonucleotides used for native gel analysis and DMS footprinting analysis were 5' radiolabeled with ^{32}P using γ-^{32}P ATP and following a standard kinasing protocol. DNAs to be so labeled were pre-treated with freshly diluted 10 % v/v piperidine at 90^oC for 30 min to remove any DNA strands with synthesis-related lesions. The kinased DNA was re-purified by size using a denaturing polyacrylamide gel.

3.2.2. Self-Assembly of the DNA TQ and TQ*s* tiles

For the self-assembly of DNA TQ and TQ*s* (as well as the intermediate DNA structures D, D*s*, T1, T1*s*, T2, or T2*s*), the respective DNA constituent strands (each at 1 μM concentration) were denatured at 100 ^oC for 4 min and diluted either in "low pH buffer" (50 mM Li^+ Acetate; 50 mM K^+ Acetate; 10 mM Mg^{2+} Acetate, pH 5.5) or in "neutral pH buffer" (10 mM Tris, 50 mM LiCl, 50 mM KCl and 10 mM $MgCl_2$, pH 7.4). The DNA mixtures were heat denatured at 100° C for 4 min and then slowly cooled from 100° C to 20° C in a thermocycler at a constant cooling rate of 7.5^oC/min. The DNA solution was then allowed to reset at 22^oC for 12 hrs.

3.2.3. Native Gel Electrophoresis, Gel Data Analysis, and DMS protection assay

The formation of different DNA species was investigated using either a “low pH” 7.5% (29:1) native polyacrylamide gel (containing and run in 50 mM Tris acetate buffer containing 5 mM K^+ Acetate, pH 5.2) or in a “neutral pH” 7.5 % (29:1) native polyacrylamide gel (containing and run in 50 mM Tris borate buffer with 5 mM KCl, pH 7.4). The gels were run at 9 W at 22° C with efficient cooling. The running buffer was made to circulate continuously from the bottom to the top reservoir and back, to maintain stable pH and ion concentration. The folded DNAs were mixed either with native low pH loading buffer (50 mM Tris Acetate, pH 5.2, 30 % glycerol and loading dyes) or with native neutral pH loading buffer (50 mM Tris HCl, pH 7.4, 30 % glycerol and loading dyes) prior to loading in the two kinds of gels. Imaging of the gels was carried out using a Typhoon 9410 Phosphorimager (Amersham Biosciences).

For the DMS treatment, pre-assembled DNA structures were treated with 0.2% DMS (freshly prepared in 10 mM Tris, pH 5.2). The solutions were allowed to incubate for 30 min at 22° C, following which the methylation reaction was quenched by the addition of ß-mercaptoethanol. Such DMS-treated DNA solutions was mixed with low pH native gel loading buffer and loaded into a 7.5 % low pH native gel. Bands corresponding to the TQ, T1 and D were then excised from the gel, their DNA eluted and recovered, and concentrated via ethanol precipitation. The recovered DNA was then treated with 10% v/v freshly prepared aqueous piperidine at 90^{o}C for 30 min followed by lyophilisation. The lyophilised DNA was dissolved in denaturing gel loading buffer (95 % formamide, 1 mM EDTA, and loading dyes) prior to loading on 20% denaturing/sequencing gels.

For a reference G-ladder preparation, ^{32}P-5’-labeled 5’-$Y_{24}G_3R’_{24}$ single-strand was heat denatured at 100° C for 3 min, following which DMS and piperidine treatments were applied, as described above.

3.2.4. Atomic Force microscopy (AFM)

Atomic force microscopy was used to directly image the triplex-quadruplex hybrid DNA nanostructures. Images were taken with a Dimension Icon AFM (Bruker; Billerica, MA) using a ScanAsyst-Air silicon nitride tip (2 nm radius) with a resonance frequency of

70 kHz in tapping mode. Images were constructed from 512 line scans stitched together, with the scan rate during imaging being between 0.8 and 1 Hz. Samples of DNA solution were pipetted onto a freshly peeled mica surface (prewashed with incubation buffer), allowed to dry, and then washed with double-deionized water to remove any traces of the buffer and to isolate the DNA prior to imaging.

3.2.5. Transmission Emission Microscopy (TEM) sample preparation, Imaging and EDX Analysis

TEM imaging was done to verify the formation of long DNA nano-structures. 4-5 μl samples were deposited on Formvar/Carbon Coated - Copper 300 mesh TEM grids (TED PELLA INC.) and incubated for 3 min. The excess solutions were removed from grid by absorbing on a filter paper. The grids were washed once with incubation buffer followed by washing with milli-Q water to remove any non-adhered DNA and excess salts. Without drying the samples out, the deposited DNA structure was stained with 4 % Neodymium Acetate[201] directly on the grid by pipetting 3-4 ul of staining solution on it and a further incubation of 2-3 min. After incubation, the excess solution was again removed using a filter paper. Finally, grids were washed twice with milli-Q water to remove all the excess chemicals. After drying the grids, TEM images were taken using an FEI Tecnai Osiris TEM microscope, equipped with Super-X EDX detection system based on Silicon Drift Detector (SDD) technology, and operated at 200 kV.

3.3. Results

3.3.1. Formation of Monomeric Triplex-Quadruplex Hybrid Tile "TQ"

Figure 3-1.a shows a schematic for the assembly of a triplex-quadruplex hybrid "tile", TQ, a key structural unit of this conception. The starter duplex here is "D", a 51-bp duplex comprising two 24 bp stretches of Watson-Crick duplex sandwiching a central stretch of three G•G mismatches: the two component single strands have sequences of 5'-$Y_{24}G_3R'_{24}$ and 5'-$Y'_{24}G_3R_{24}$, respectively, shown in **Table 3-1**.

When two distinct, 27-nt, pyrimidine-rich, non-mutually complementary single-strands, 5'-G_3Y_{24} and 5'-$G_3Y'_{24}$ (**Table 3-1**) are added to the duplex D at pH 5.5, that

leads to the formation of two alternative triplexes, T1 and T2, respectively. Addition of both single strands, sequentially or together, at pH 5.5 and in the presence of the K^+ cation (the last necessary for the specific stabilization of the three central G-quartets) gives rise to the Triplex/G-Quadruplex tile, TQ, which is stable under these solution conditions.

Table 3-1. List of DNA sequences used to make different constructs. DNA sequences for constructing non-sticky TQ tile and sticky TQ*s* tile.

TQ Tile DNA	
Name	Sequence (5′-3′)
$Y_{24}G_3R'_{24}$	**TCC TCT CTC TCT CTC TCT CTC TCT GGG AAG AAG AAG AAG AAG AAG AAG AAG**
$Y'_{24}G_3R_{24}$	**CTT CTT CTT CTT CTT CTT CTT CTT GGG AGA GAG AGA GAG AGA GAG AGA GGA**
G_3Y_{24}	**GGG TTC TTC TTC TTC TTC TTC TTC TTC**
$G_3Y'_{24}$	**GGG TCT CTC TCT CTC TCT CTC TCT CCT**
TQ*s* Wire DNA	
Name	Sequence (5′-3′)
$G2Y_{24}G_3R'_{24}G2$	**GG TCC TCT CTC TCT CTC TCT CTC TCT GGG AAG AAG AAG AAG AAG AAG AAG AAG GG**
$G2Y'_{24}G_3R_{24}G2$	**GG CTT CTT CTT CTT CTT CTT CTT CTT GGG AGA GAG AGA GAG AGA GAG AGA GGA GG**
$G_3Y_{24}G_4$	**GGG TTC TTC TTC TTC TTC TTC TTC TTC GGGG**
$G_3Y'_{24}G_4$	**GGG TCT CTC TCT CTC TCT CTC TCT CCT GGGG**

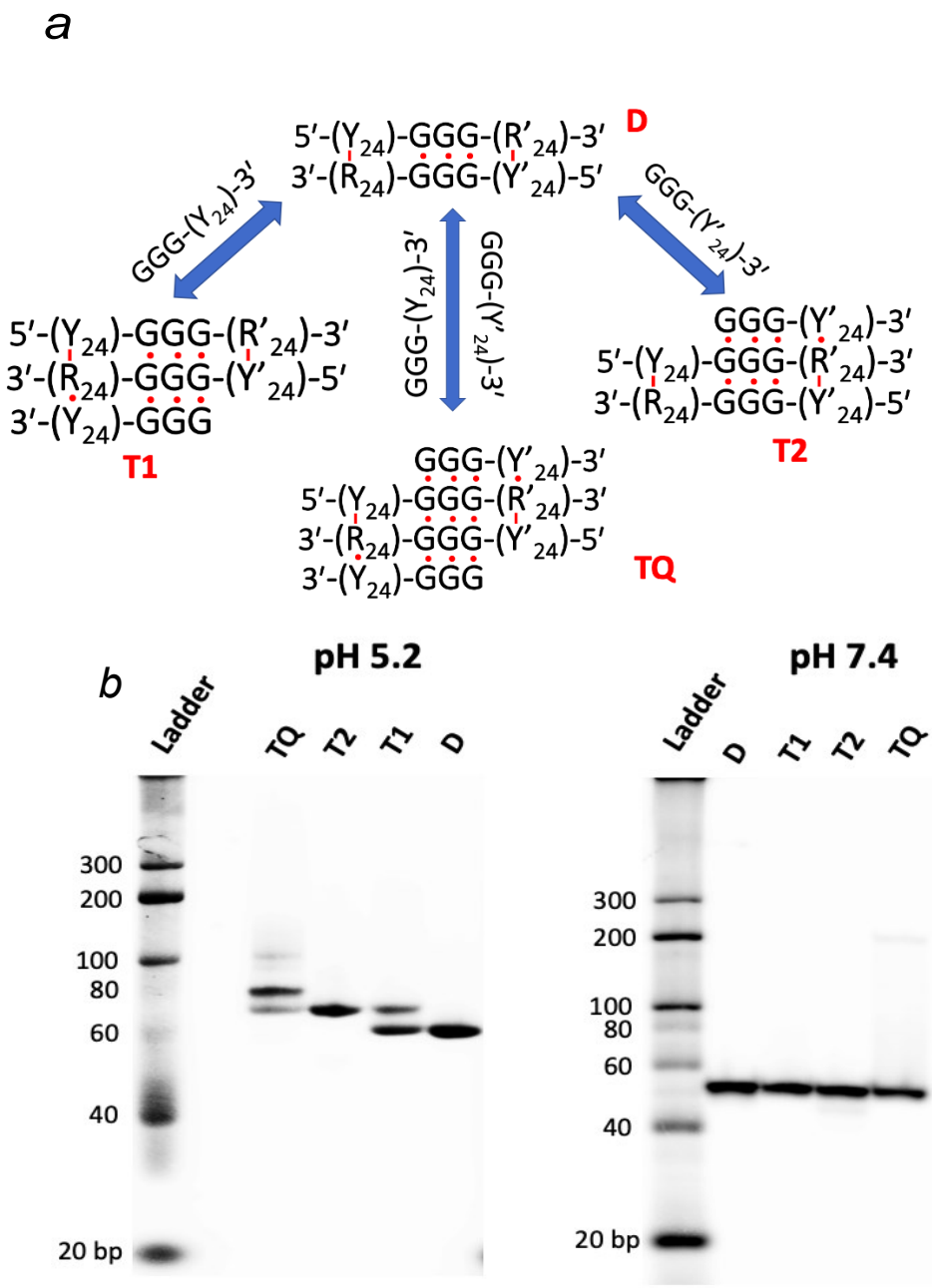


Figure 3-1. Characterization of Triplex-Quadruplex Hybrid Tile TQ. **(a)** Schematic of TQ design; D: duplex, T1/T2: triplexes; TQ: triplex-quadruplex hybrid. **(b)** Low pH and neutral pH native gels to analyze different products. (left) Native gel run in 50 mM TAE buffer plus 5 mM potassium acetate pH 5.2; (right) Native gel run in 50 mM TBE buffer plus 5 mM potassium chloride, pH 7.4. Samples D, T1, T2 and TQ were incubated either in low pH buffer (consisting of 50 mM Li^+ acetate, 50 mM K^+ acetate and 10 mM Mg^{2+} acetate, pH 5.5, and run in lower pH gel) or in neutral pH buffer (consisting of 10 mM Tris, 50 mM LiCl, 50 mM KCl and 10 mM $MgCl_2$ pH 7.4, and run in a neutral pH gel).

Figure 3-1.b shows native gels, run at pH 5.2 (left) and pH 7.4 (right), respectively of the various DNA constructs shown in **Figure 3-1.a** (the 5'-$Y_{24}G_3R'_{24}$ strand of the D duplex is ^{32}P-5'- end-labelled). It can be seen that in the pH 5.2 gel, the different complexes have characteristic native gel mobilities: the duplex D has the fastest mobility, the three-stranded and structurally equivalent complexes T1 and T2 share a mobility slightly slower than that of D, and that the four-stranded tile complex, TQ, has the slowest mobility of the four complexes. At pH 7.4, however, all four loadings show the same gel mobility, indicating that neither the purely triplex complexes, T1 and T2, nor the G-quartet containing tile, TQ, persists. At this pH, the CG*C^+ base triple-containing triplexes are not stable, despite the presence of 5 mM K^+ in the gel to stabilize the central 3-quartet G-Quadruplex within TQ.

To investigate in greater detail whether the expected base-pairing patterns within the D, T1 and TQ complexes were indeed occurring, I carried out footprinting experiments with dimethyl sulphate (DMS), a reagent that methylates the N7 position of guanines within ssDNA and DNA helices in which the N7 is not involved in hydrogen bonding, such as in a Watson-Crick duplex but not in the R strand of YR*Y triplexes or in G-quadruplexes: in the last two structures, this N7 position participates in Hoogsteen hydrogen-bonding and does not react with DMS. The various DNA complexes were treated with 0.2% DMS, and then loaded in a pH 5.5 native gel, shown in **Figure 3-2**, ***left***. DNA from the asterisked bands, representing the D, T1, and TQ complexes, were excised and purified, and treated with 10 % v/v hot aqueous piperidine (to break the DNA phosphodiester backbone at such modification sites) prior to loading on a sequencing gel**, Figure 3-2**, ***right***. As expected, all guanines within the ^{32}P-5'-labeled 5'-$Y_{24}G_3R'_{24}$ single-strand within the D duplex are reactive with DMS. In T1, the G3 stretch is reactive; however, the guanines participating in the triplex are not DMS-reactive, as expected; in TQ, all guanines within this strand are DMS-protected, consistent with their participation in triplex as well as G-quadruplex (for the G_3 stretch) formation. By contrast to the above, in the ^{32}P-5'-labeled 5'-$Y_{24}G_3R'_{24}$ single-strand alone ("+ DMS control"), all guanines are strongly reactive.

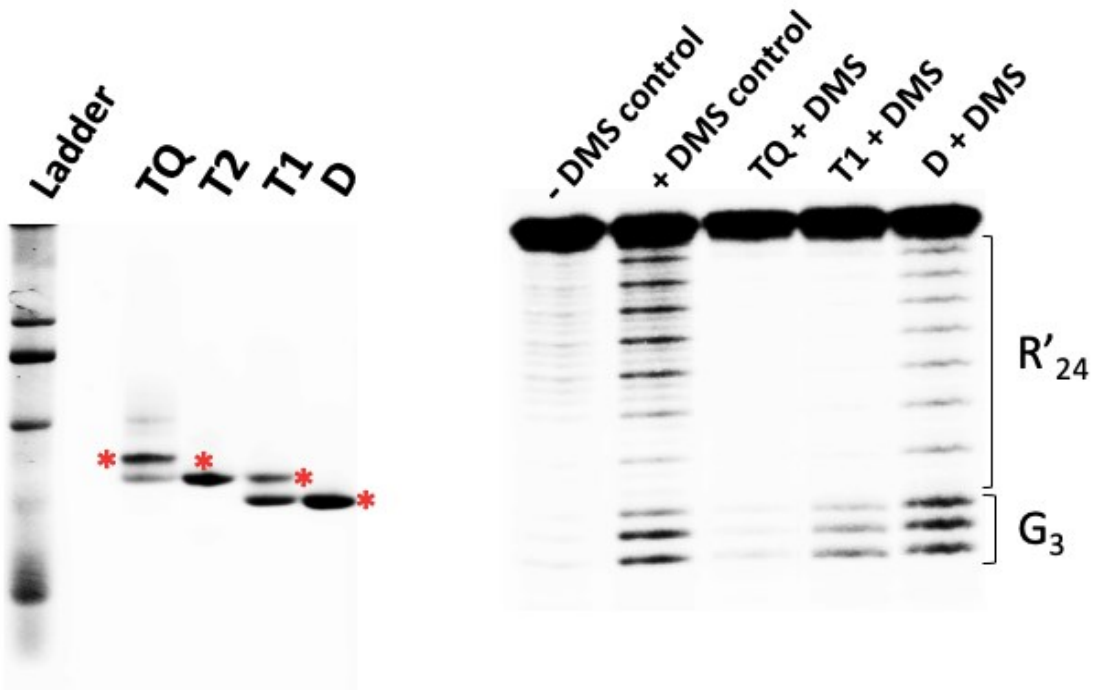


Figure 3-2. DMS footprinting of TQ tile. **(Left**) Native gel run at pH 5.5. showing (with red asterisks) DMS-methylated D, T1, T2 and TQ complexes. **(Right)** Denaturing gel showing guanine bands from each of the asterisked and DMS-treated complexes purified out of the native gel (left), followed by treatment with freshly prepared 10 % v/v hot aqueous piperidine and lyophilization to dryness prior to running on the denaturing gel. The "-DMS control" and "+DMS control" were obtained from the ^{32}P-5'-labeled 5'-$Y_{24}G_3R'_{24}$ single strand.

The formation of TQ tile was independently investigated using an alternate, fluorogenic dye-based approach. Chelerythrine (CHE) has been shown to bind specifically and undergo a large fluorescence enhancement from stacking upon the terminating ends of triplexes[202]. I incubated CHE with the TQ tile in a 2:1 ratio. **Figure 3-3** shows that CHE alone, or in the presence of Duplex D, exhibits only a low ("background") fluorescence. However, in the presence of triplex T1 or the tile TQ, its fluorescence is sharply enhanced. Construct T1 has a flush triplex terminus at one of its ends and a flush duplex terminus on the other; whereas, the tile TQ has flush terminating triplexes on both its ends. Incubation of TQ and T1 with CHE reveals that CHE's fluorescence in the presence of TQ is ~ 2-fold higher than its fluorescence in the presence of T1 (once the background-level fluorescence contribution from the duplex end of T1 has been subtracted). The above number is in agreement with the structure as proposed. Furthermore, the fluorescence decreases notably if the pH is raised from 5.5 to 8.0, confirming both the requirement of low pH for the stability of the TQ tile, and the facile, pH-dependent reversibility of TQ tile formation.

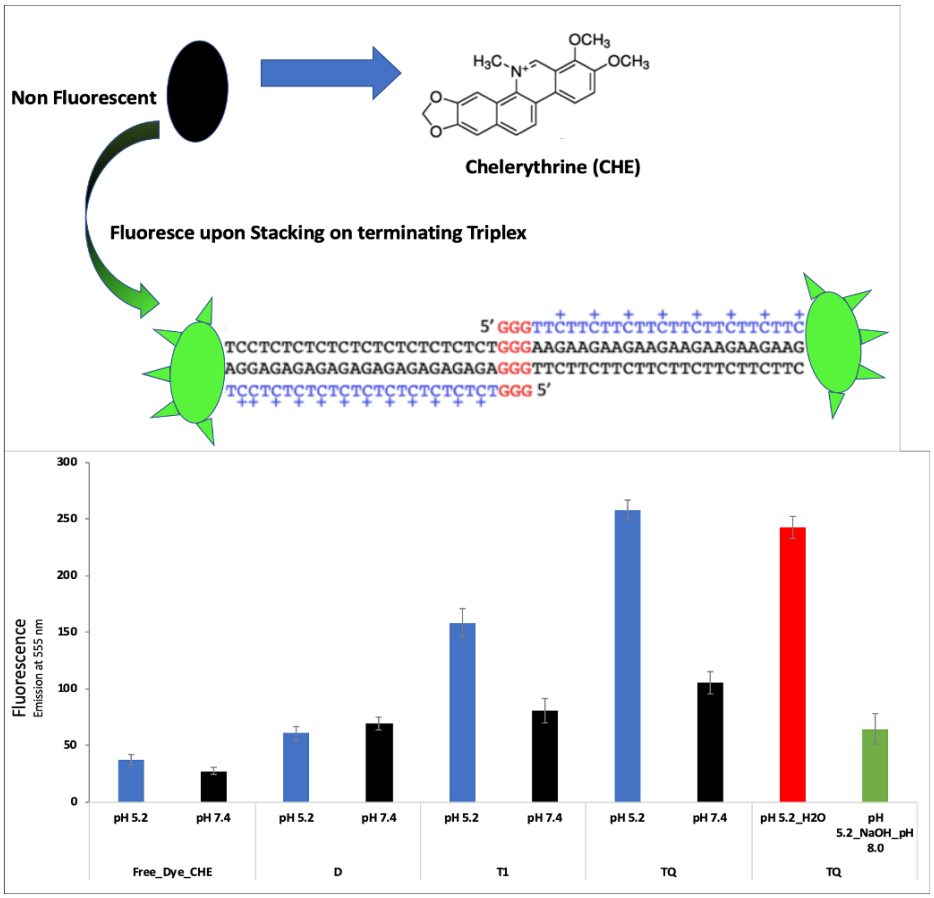


Figure 3-3. Dye based approach of investigating TQ formation. (Top) Schematic showing how CHE fluoresce after stacking to the triplex. **(Bottom)** Fluorescent measurement (Excitation/Emission: 333/555 nm) for different structures at two different pH. D: duplex, T1: Triplex, TQ: Triplex-Quadruplex hybrid tile. Last column: Effect of changing pH from pH 5.2 to 8.0. **"pH5.2_H_2O":** addition of equivalent amount of H_2O as NaOH to TQ tile to see effect of dilution. **"pH5.2_NaOH_pH8.0":** addition of NaOH to raise pH of TQ tile.

3.3.2. Sticky Version of TQ tile "TQ*s*" to Form long 1-D DNA Nanostructure (1DDN-"$(TQs)_n$")

Variants of the duplex D can be constructed by the inclusion of sticky "GG" additions at both 3' and 5' ends of each component single strand. This generates the D*s* ("*s*" for "sticky") duplex (**Figure 3-4.a**). This can be converted by the addition of either single strand, 5'-$G_3Y_{24}G_4$ or 5'-$G_3Y'_{24}G_4$, to form triplex complexes T1*s* or T2*s*, at pH 5.5. Addition of both single-strands would be expected to generate a triplex-quadruplex tile, TQ*s*, that has uniquely guanine-only sticky ends on both ends. Such sticky tiles should in principle assemble together (**Figure 3-4.b,** "1") in the presence of K^+ to form a triplex-quadruplex hybrid nanostructure/nanowire, $(TQs)_n$ (**Figure 3-4.b,** "2" and "3"). Such a final 1DDN would be expected to contain two distinct sets of short G-quadruplexes breaking up long stretches of YR*Y triplex. The first of these, $(G_3)_4$, would serve to correctly phase the triplex components of the final 1DDN; the second, $(G_4)_4$, would serve as a "glue junction", guanine-rich sticky ends capable of binding together to form a complete and well-stacked composite G-quadruplex involving two distinct TQ*s* tiles.

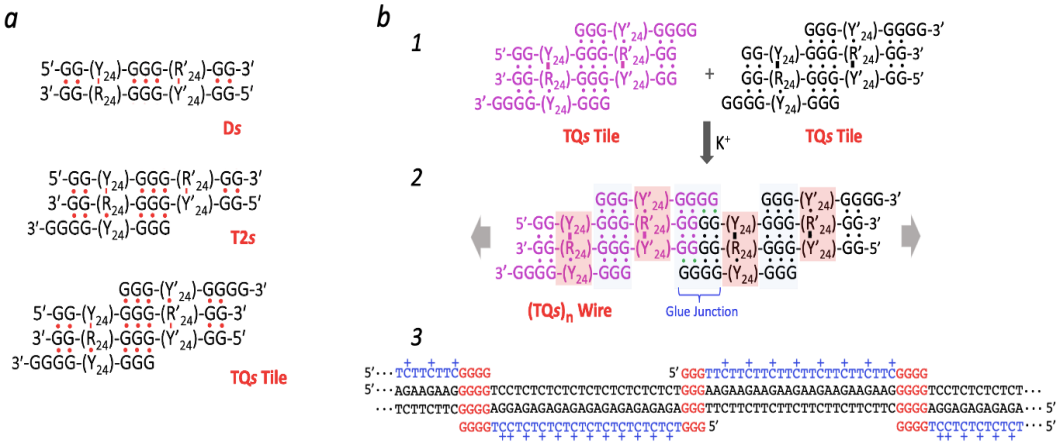


Figure 3-4. Design of sticky TQ*s* tiles. (a) Shown are a D*s* duplex to which addition of a single-strand, 5'-$G_3Y_{24}G_4$ at pH 5.5 creates a single-sided sticky-ended triplex tile, T2*s*; and the addition of a second single-strand, 5'-$G_3Y'_{24}G_4$, generates a double-sided sticky-ended triplex tile (TQ*s*). **(b)** TQ*s* tiles ("1") in the presence of K^+ should assemble to (TQ*s*)n 1DDN ("2"; full nucleotide sequence shown in "3").

Figure 3-5 shows the complexes Ds, T1*s*, T2*s*, and TQ*s* run in native gels run at pH 5.2 (***a***) and at pH 7.4 (***b***), respectively, with both gel running buffers containing 5 mM K^+. It can be seen that neither the triplexes, T1*s* and T2*s*, nor the triplex-quadruplex tile, TQ*s*, persist at pH 7.4; whereas at 5.2, they run as distinct complexes with defined gel mobilities. Interestingly, at pH 5.2, TQ*s* does not enter the gel at all and is restricted to the gel well, consistent with the TQ*s* tile self-assembling into a high molecular weight species, likely the triplex-quadruplex hybrid 1DDN, $(TQs)_n$. The requirement for *all four* DNA component strands (5'-$G_2Y_{24}G_3R'_{24}G_2$; 5'-$G_2Y'_{24}G_3R_{24}G_2$; 5'-$G_3Y_{24}G_4$ and 5'-$G_3Y'_{24}G_4$) for this to happen is indeed consistent with the formation of such a high molecular weight species/1DDN, $(TQs)_n$

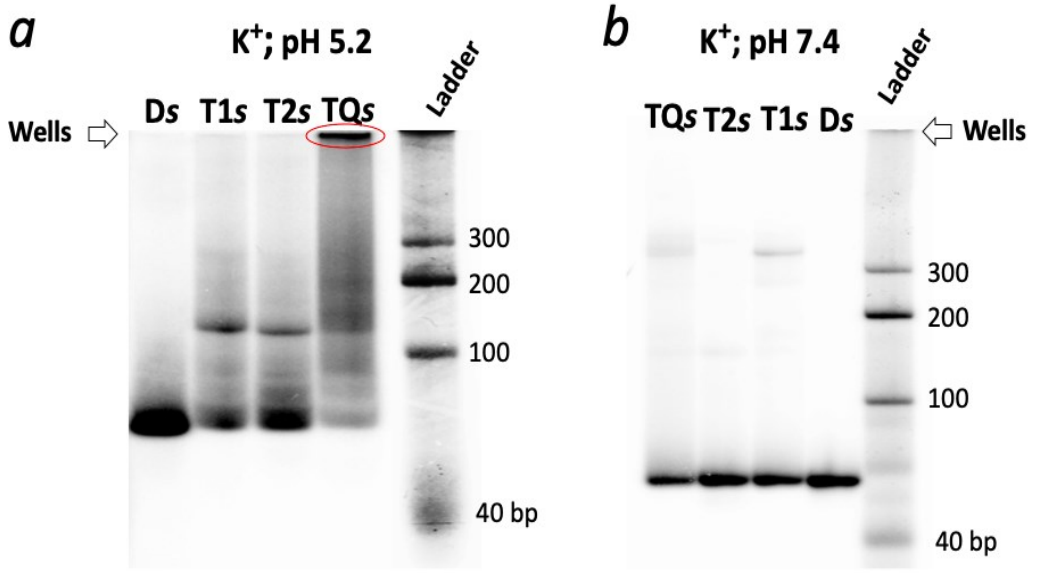


Figure 3-5. Native gels, at pH 5.2 and 7.4, showing the sticky constructs Ds, T1*s*, T2*s* and TQ*s*. (Left) Native gel run in 50 mM TAE buffer plus 5 mM potassium acetate, pH 5.2; **(Right)** Native gel run in 50 mM TBE buffer plus 5 mM potassium chloride, pH 7.4. D*s,* T1*s,* T2*s,* TQ*s* (incubated independently either in low pH or neutral pH buffer). The postulated $(TQs)_n$ 1DDN is shown circled in red.

The requirement for K^+ in the formation of $(TQs)_n$, was investigated by separately co-incubating its four constituent DNA single strands in a purely Li^+ containing buffer (pH 5.5); in Li^+ and Mg^{2+} (pH 5.5), and in a third buffer containing Li^+, Mg^{2+}, and K^+ (pH 5.5). **Figure 3-6.a,** shows that the incubations in the presence of either Li^+ alone or Li^+ and Mg^{2+} at pH 5.5 are not sufficient to hold $(TQs)_n$ together. The $(TQs)_n$ species is observed at the gel well only when K^+ is present during its formation. To investigate, further, the role of the mildly acidic pH in $(TQs)_n$'s formation and stability, the four constituent strands

were incubated with all three cations, Li^+, Mg^{2+} and K^+, at pH 5.5, but were then run in a pH 7.4 native gel **(Figure 3-6.b)**. Here, the high molecular weight species is no longer evident. Also, **Figure 3-6.c** shows that a significant proportion (~30 %) of the 1DDN dissociates if 18-Crown-6, a crown ether compound known to tightly and specifically sequester K^+, is added in excess of K^+. The combined result from **Figure 3-6** confirms that for both formation and stability of $(TQs)_n$, an acidic pH (to stabilize the DNA triplex stretches containing $CG*C^+$ base triples) *as well as* K^+ (to stabilize the short G-quadruplexes formed between successive TQ*s* tiles to form $(TQs)_n$ are required.

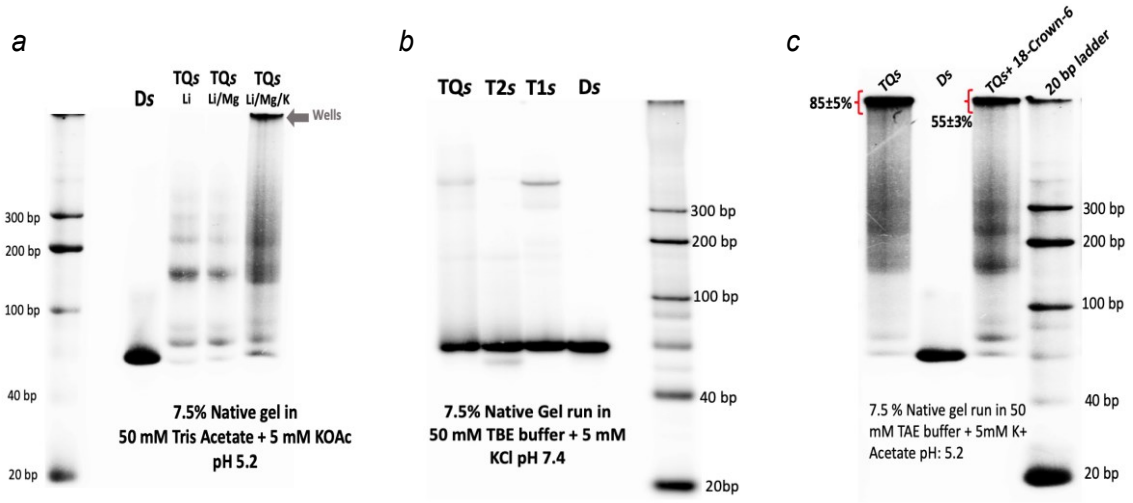


Figure 3-6. Analysis of $(TQs)_n$ stability in response to encountering a change in pH and K^+ depletion. **(a)** Incubation of TQ*s* in the presence of either a purely Li^+ containing buffer "Li" (pH 5.5); in Li^+ and Mg^{2+} containing buffer "Li/Mg" (pH 5.5), or in Li^+, Mg^{2+} and K^+ containing buffer "Li/Mg/K" (pH 5.5). The High molecular weight species, (TQ*s*)n seen in the well of the Li/Mg/K lane, is not found in lanes showing incubation lacking K^+. **(b)** TQ*s* was incubated at low pH buffer but was run in a neutral pH gel. The high molecular weight species, seen in (a), and corresponding to the hypothesized $(TQs)_n$, is no longer visible. **(c)** TQ*s* incubation was carried at pH 5.2 in "Li/Mg/K" buffer. Upon incubation one half of TQ*s* was mixed with K^+ sequester 18-crown-6 and other half with just equivalent volume of H_2O. The high molecular weight species corresponding to the hypothesized $(TQs)_n$ breaks down substantively in the lane where TQ*s* have been mixed with 18-crown-6.

3.3.3. Visualization of 1DDN "$(TQs)_n$" using Atomic Force Microscopy (AFM) and Transmission Electron Microscopy (TEM)

I used Atomic Force Microscopy (AFM) to investigate the physical appearance of $(TQs)_n$. Following its formation, as above, the product was deposited onto mica surfaces pre-washed with the incubation buffer. **Figure 3-7** shows three independent AFM images of *bona fide* $(TQs)_n$ assembly, of 150-300 nm length. **Figure 3-8** shows an average cross-section (depth) of $(TQs)_n$ nanostructure from multiple AFM images. The

cross section of (TQ*s*)$_n$ Wire (diameter ~2.2 nm) is comparable to the AFM cross section of G-wires as has been reported previously[107,203,204]. By contrast, incubation in the Li^+/ Mg^{2+}/K^+ buffer, at pH 5.5, of the duplex, D*s*; T1*s*/T2*s*, or the monomeric tile TQ, did not show any level of long nanostructure formation (**Figure 3-9**). D*s*, T1*s*/T2*s* as well as monomeric TQ tile are individually very small dimensionally (< 20 nm) and thus they appear as small dots on AFM image (the scanning window scale is 1 μm). DNA duplexes have been reported to have very non-defined/irregular AFM features with uncharacterised dimensions[107,203,204]. Because of its high relative flexibility, duplexes can get pinned or flattened upon mica surface. In our case, D*s* appear as floating island in the AFM image (suggesting limited aggregated and flattening) on mica surface whereas T1*s*/T2*s* as well as monomeric TQ appear as discrete dots.

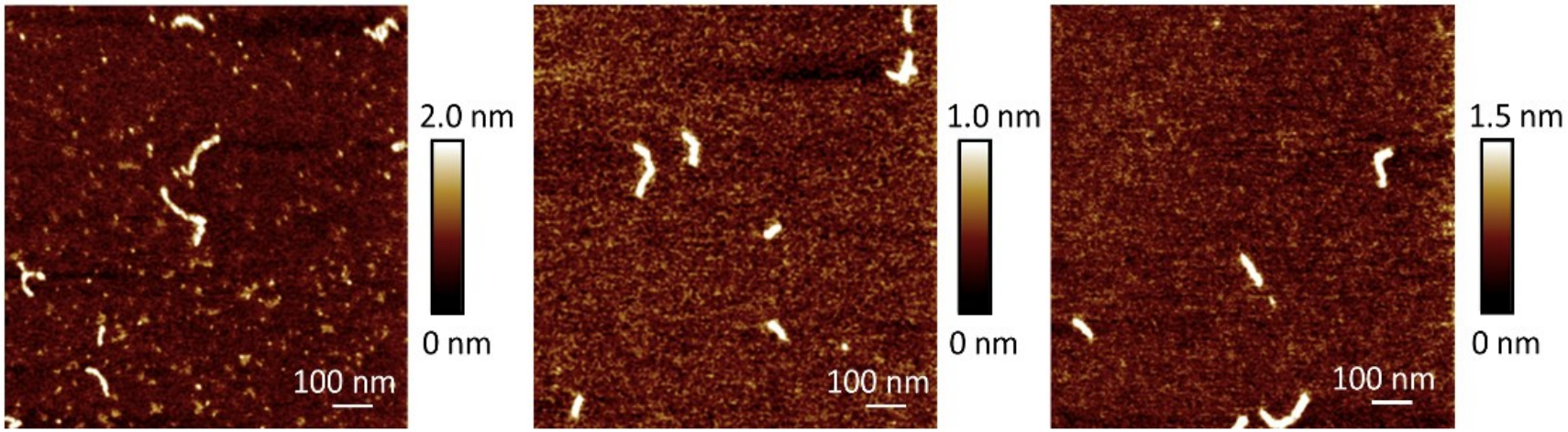


Figure 3-7. AFM Images of TQ*s* Hybrid DNA 1DDN. Incubated 1DDN were deposited onto mica and then imaged in tapping mode. Scale bar (color gradient) to the right of every AFM image denotes the height information of TQ*s* 1DDN.

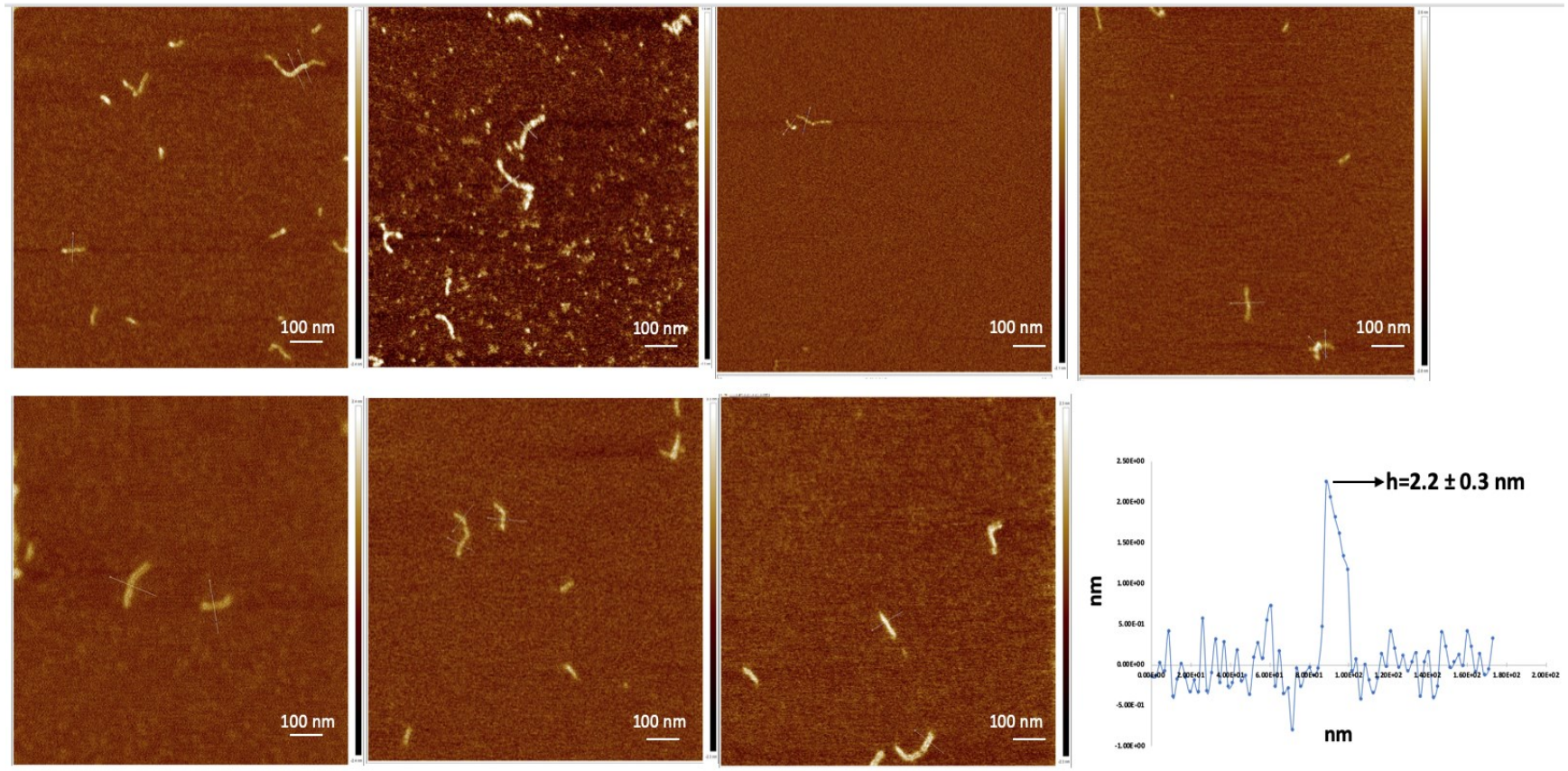


Figure 3-8. Cross-Section from multiple AFM images of (TQ*s*)$_n$ 1DDN. Actual sites of cross-section has been shown by a dashed line on the AFM image itself. An average cross-section plot has been shown at bottom right.

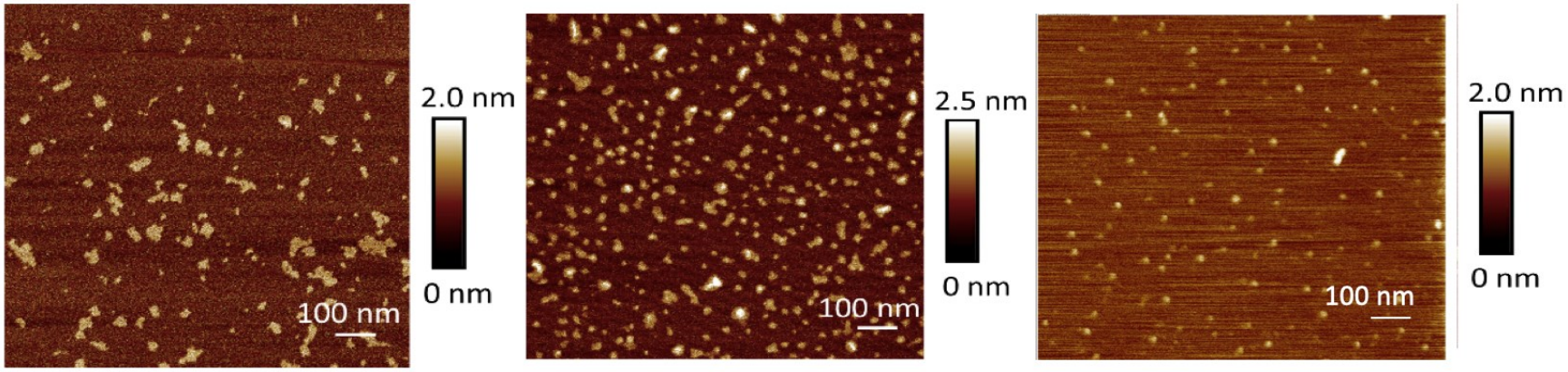


Figure 3-9. AFM images of duplex D*s*; triplex T1*s* and the monomeric tile TQ. None of these monomeric units show any sign of elongated Wire/1DDN formation.

I used TEM (Transmission Electron Microscopy) as an independent alternative method to investigate the appearance of TQ*s* nanostructure. **Figure 3-10** shows that TQ*s* tile, following incubation in Li^{+}/ Mg^{2+}/K^{+} buffer, at pH 5.5, forms a long DNA nanostructure (200-300 nm). The forks/branches seen in the STEM images reflect the overlap of two separate 1DDN/nanowire as a result of improper/incomplete segregation during adsorption of the deposited samples (nanowires) onto the TEM grids.

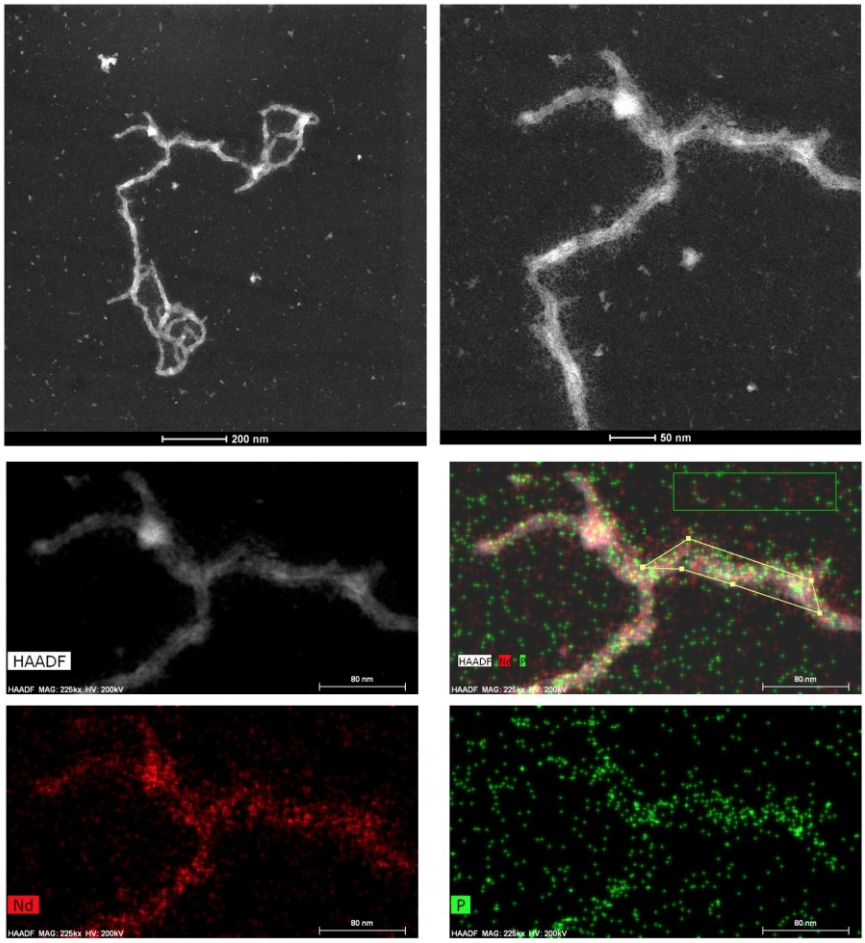


Figure 3-10. STEM-HAADF images of $(TQs)_n$ 1DDN. **(Top: left & right):** Images of 1DDN at two different scales**. (Medium and Bottom):** EDX analysis of 1DDN. Bottom-left: Signal (in red) of neodymium (which was used to stain the DNA for imaging). Bottom-right: Signal (in green) of Phosphorous (coming uniquely from DNA and not from any buffer). Medium-right: Overlapping signal from red and green channel.

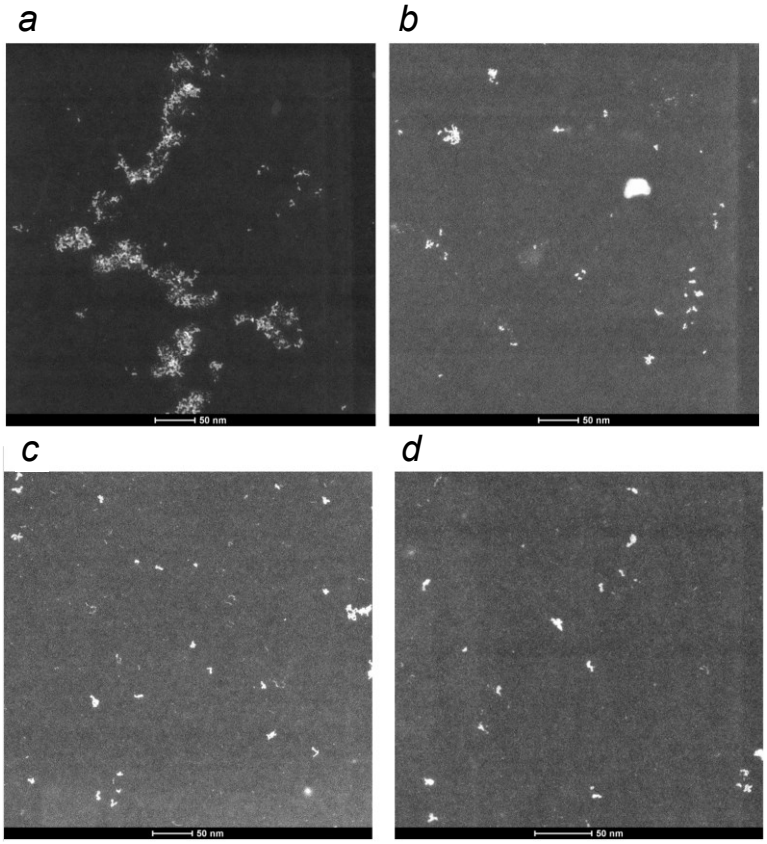


Figure 3-11. STEM-HAADF Images of 1DDN in response to increasing pH and K^+ depletion. Dissociation of $(TQs)_n$, incubated in "Li/Mg/K" buffer at pH 5.5, **(a)** 2 hrs and **(b)** 4hrs after raising the pH to 8.0 by addition of NaOH. **(c-d)** No 1DDN was seen when $(TQs)_n$, were incubated in "Li/Mg" buffer (pH 5.5), i.e., in absence of K^+.

Energy Dispersive X-Ray Analysis (EDX) was done to further confirm the presence of DNA in the TEM images as shown in **Figure 3-10.** Once again, the $(TQs)_n$ nanostructure were seen to disassemble from either an increase in the pH or the absence of K^+ from the incubation as shown in **Figure 3-11**.

3.4. Discussion and Conclusion

What applications might such TQ Hybrid nanostructure find? First, this is an easily reversed nanostructure, whose stability relies on two distinct enablers—the K^+ ion and pH < 5.5. These two requirements constitute an "AND" logic for the formation of the nanostructure, inasmuch as removal of either reagent/condition disassembles the

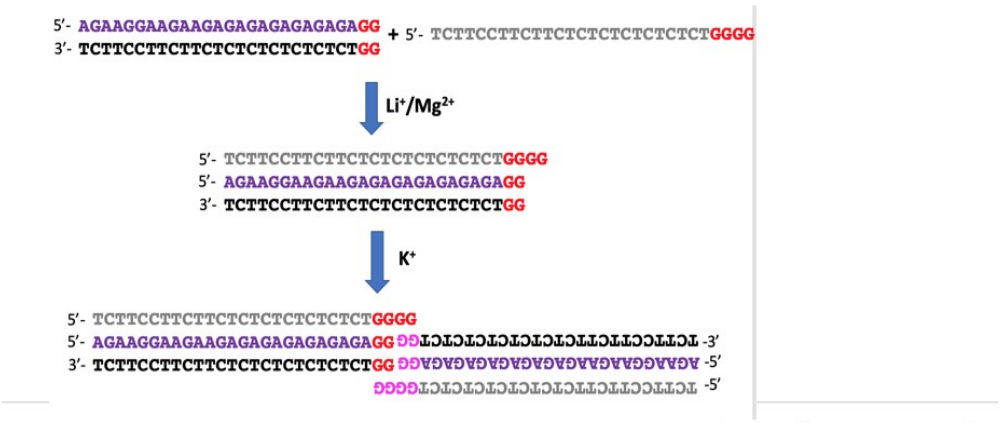
nanostructure. Conventional “G-Wires”, formed from canonical G-quartets, or other reported 1DDN, have, by contrast, very limited possibilities in terms of reversibility[205,206]. Potentially, then, such TQ Hybrid nanostructures could be used to link together proteins, protein complexes, DNA nanostructures or indeed non-biological nano-objects, whose linkage could conveniently be formed or broken by either the addition/sequestration of K^+ or by shunting the ambient pH between 5.5 and 7.5. Although in this work I report the self-assembly of a 1DDN of indefinite length, based on the reagent-limited self-assembly of the sticky TQ*s* tile (which contains G-rich sticky ends on both sides), nanostructures of discrete lengths could be assembled by the inclusion of low molar ratios of 1DDN-terminating tiles with only one-side sticky. The nanostructure reported here is built from YR*Y triplexes; however, it should also be possible to construct analogous 1DDN which harness, instead, “YR*R” triplexes, which would require only K^+ and Mg^{2+} ions and neutral pH but not pH < 5.5 for formation, stability and reversibility. A particular virtue of TQ Hybrid nanostructure as a class is their ease of synthesis; Other 1DDN, that are composed of multiple DNA strands, particularly different G-wires made up exclusively or largely out of guanine or other quartets, require high (mM range) concentrations of DNA to enable their efficient intermolecular self-assembly[199]. The TQ Hybrid 1DDN reported here, however, requires only ~1 μM total DNA to form. Further, abasic sites can be rationally positioned in these TQ Hybrid nanostructures, where specific fluorophores could bind and fluoresce, providing a novel approach to developing sensitive and reversible (pH- and K^+-dependent) sensors[207]. Among other anticipated uses for these nanostructures could be electric charge conduction (based on their relatively high guanine content), the construction of logic gates, DNA origami, and pH-responsive reversible hydrogels.

One of the most important innovations of this study is that, it introduces the concept of a new class of purely guanine-based “sticky ends”, which can be exploited to join together DNA entities not via Watson-Crick base pairing. To show this application, I took a small duplex with the sense and anti-sense strand consisting of mutually complementary poly-purine and poly-pyrimidine bases, respectively, as well as a set of ‘’GG’’ overhangs on each strand (on the 3’ end of the purine strand, and 5’ end of the pyrimidine strand- **Figure 3-12**.a). The triplex forming strand used here is poly-pyrimidine with a “GGGG” stretch at its 3’ end.

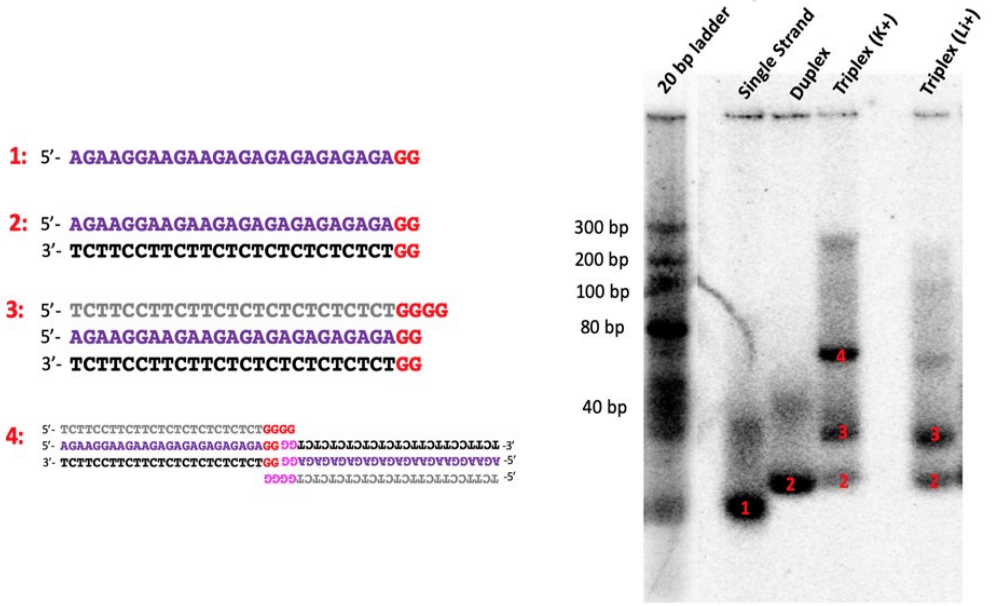
Figure 3-12 shows that in the presence of K^+ and low pH, this triplex can homo-dimerize to give rise to a triplex dimer -as shown both by low pH Native gel electrophoresis data and the CHE (fluorogenic dye) based data. The underlying importance of formation of canonical YR*Y triplexes to enable these nanostructures is shown in **Figure 3-13***;* if the rules for forming such triplexes are broken by way of G to C and compensating C to G mutations in the D duplex, the T1, T2 and TQ complexes are no longer able to form.

In conclusion, I report here a novel class of 1-D DNA nanostructure (1DDN), composed primarily of DNA triplexes but assembled and correctly phased by very short stretches of G-quadruplex. Future work will focus on fine-tuning the reversibility of these nanostructures; investigating the properties and utility of tiles and 1DDN that incorporate YR*R triplexes rather than the YR*Y triplexes used here; and of devising means, based on utilizing unique triplex sequences, for the assembly of 1DDN of precise length.

a


5′- AGAAGGAAGAAGAGAGAGAGAGAGG + 5′- TCTTCCTTCTTCTCTCTCTCTCTCTGGGG
3′- TCTTCCTTCTTCTCTCTCTCTCTCTGG
Li+/Mg2+
5′- TCTTCCTTCTTCTCTCTCTCTCTCTGGGG
5′- AGAAGGAAGAAGAGAGAGAGAGAGAGG
3′- TCTTCCTTCTTCTCTCTCTCTCTCTGG
K+
5′- TCTTCCTTCTTCTCTCTCTCTCTCTGGGG
5′- AGAAGGAAGAAGAGAGAGAGAGAGAGG
3′- TCTTCCTTCTTCTCTCTCTCTCTCTGG

b


1: 5′- AGAAGGAAGAAGAGAGAGAGAGAGAGG
2: 5′- AGAAGGAAGAAGAGAGAGAGAGAGAGG
3′- TCTTCCTTCTTCTCTCTCTCTCTCTGG
3: 5′- TCTTCCTTCTTCTCTCTCTCTCTCTGGGG
5′- AGAAGGAAGAAGAGAGAGAGAGAGAGG
3′- TCTTCCTTCTTCTCTCTCTCTCTCTGG
4:
20 bp ladder
Single Strand
Duplex
Triplex (K+)
Triplex (Li+)
300 bp
200 bp
100 bp
80 bp
40 bp
1
2
3
4

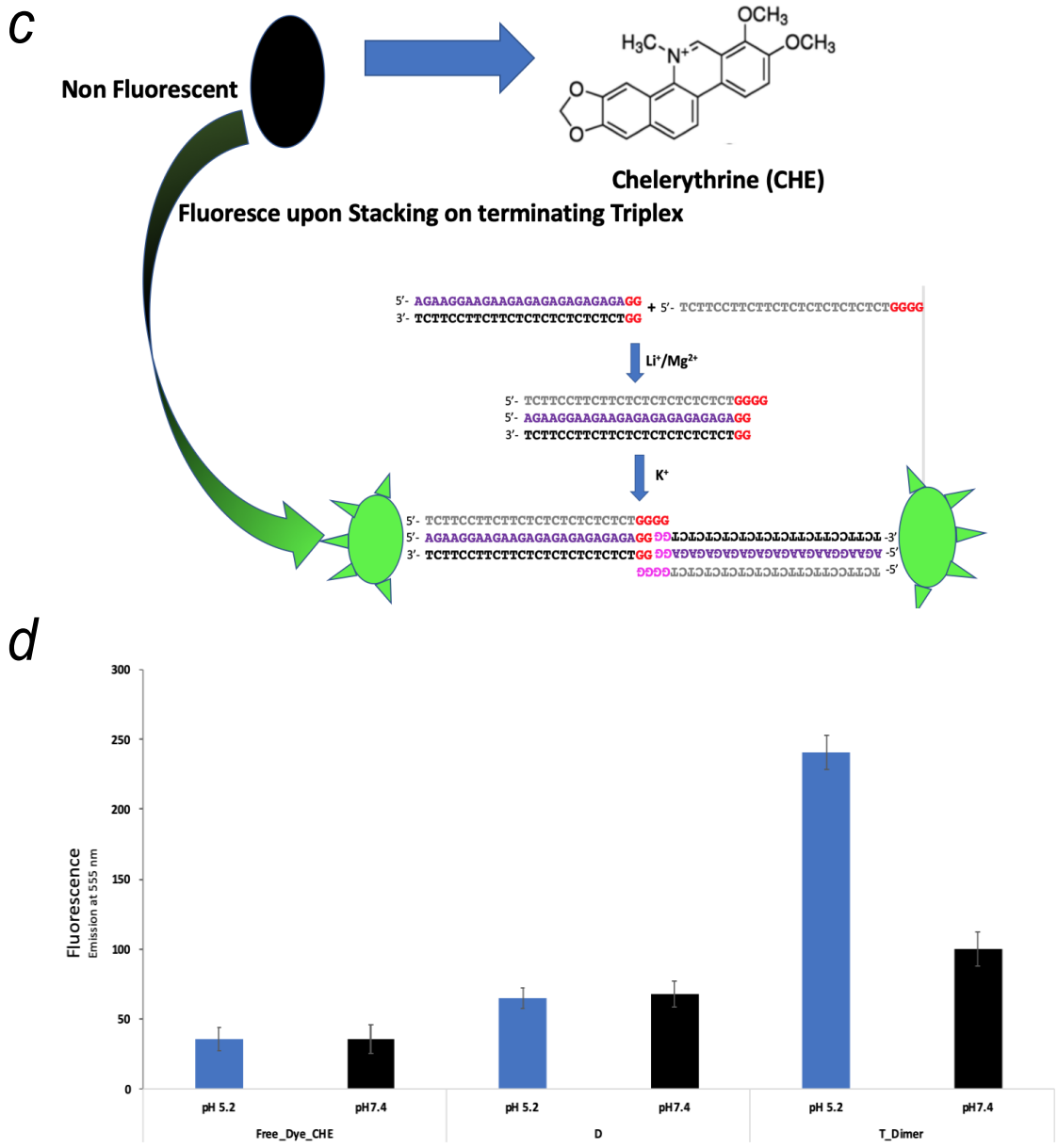


Figure 3-12. Dimerization of a short duplex sequence using purely G-Quartet-based sticky ends. (a) Schematic showing the principle of dimerization using quartet based sticky ends. **(b)** Low pH native gel showing K^+ dependent self-dimerization of a small duplex (See "species 4" formed only in the presence of K^+ (Lane: Triplex (K^+)) and not in the presence of Li^+ (Lane: Triplex (Li^+)). **(c-d)** Schematic and data to show the dimerization using an alternate dye-based approach.

TQ

```
                                  5′-GGG TTC TTC TTC TTC TTC TTC TTC TTC
5′-TCC TCT CTC TCT CTC TCT CTC TCT GGG AAG AAG AAG AAG AAG AAG AAG AAG
3′-AGG AGA GAG AGA GAG AGA GAG AGA GGG TTC TTC TTC TTC TTC TTC TTC TTC
   TCC TCT CTC TCT CTC TCT CTC TCT GGG-5′
```

T1-mut (Sense and Antisense G→C mutations)

```
                                  5′-GGG TTC TTC TTC TTC TTC TTC TTC TTC
5′-TGG TGT GTG TGT GTG TGT GTG TGT GGG AAC AAC AAC AAC AAC AAC AAC AAC
3′-ACC ACA CAC ACA CAC ACA CAC ACA GGG TTG TTG TTG TTG TTG TTG TTG TTG
```

T2-mut (Sense and Antisense G→C mutations)

```
5′-TGG TGT GTG TGT GTG TGT GTG TGT GGG AAC AAC AAC AAC AAC AAC AAC AAC
3′-ACC ACA CAC ACA CAC ACA CAC ACA GGG TTG TTG TTG TTG TTG TTG TTG TTG
   TCC TCT CTC TCT CTC TCT CTC TCT GGG-5′
```

TQ-mut (Sense and Antisense G→C mutations)

```
                                  5′-GGG TTC TTC TTC TTC TTC TTC TTC TTC
5′-TGG TGT GTG TGT GTG TGT GTG TGT GGG AAC AAC AAC AAC AAC AAC AAC AAC
3′-ACC ACA CAC ACA CAC ACA CAC ACA GGG TTG TTG TTG TTG TTG TTG TTG TTG
   TCC TCT CTC TCT CTC TCT CTC TCT GGG-5′
```

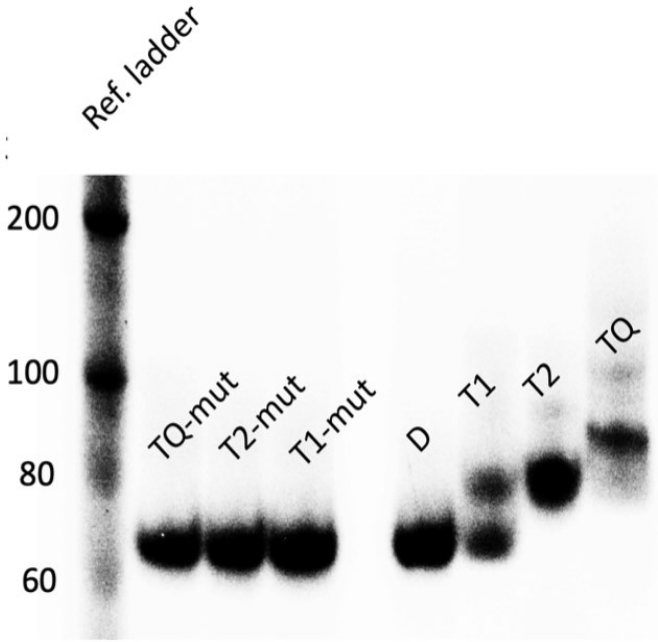


Figure 3-13. **Demonstration of the requirement for canonical YR*Y triplex formation for the assembly of the triplexes T1 and T2 and the tile TQ**. Introduction of G to C and compensating C to G mutations into the Watson-Crick duplex (shown in black, *left*) prevents the formation of triplexes using the original triplex-forming single-stranded DNA strands (shown in blue, *left)*.

Chapter 4.

A novel approach of capturing DNA/RNA G-Quadruplexes (*in vitro* and *in vivo*) using intrinsic Heme-DNAzyme peroxidase activity

4.1. Introduction

Single-stranded DNAs (ssDNAs) and RNAs are capable of forming a number of non-canonical secondary structures, including G-quadruplexes (GQs)[42,208,209], i -motif[210], R-loops[211], triplexes[164] and other triplex-like hybrids[212]. There structures have been discussed in detail in Section 1.4 of Chapter 1. In particular, Guanine-rich nucleic acids have a strong propensity to fold, under physiological temperature and solution conditions, into GQs, structurally polymorphic quadruple helices that can form either in intramolecular or intermolecular fashion, with a diversity of topologies and strand orientations in the presence of specific cations like K^+ and Na^+.

To date, a large number of distinct experimental approaches, explained in Section 1.5, have been used to search for GQs within living cells[46,47,213,214]. Extensive use has been made of strongly GQ-binding small molecules ('GQ-ligands') to stabilize and/or pull-down intracellular DNA and RNA GQs[214–217,50]. One caveat with regard to the use of extrinsic GQ-ligands to probe living cells is the possibility that such ligands may perturb the equilibrium of non-GQ but GQ-capable DNA or RNAs toward forming GQs. Yet GQ-ligands do not have to be either extrinsic or synthetic compounds. Ubiquitous cellular cofactor, hemin [ferric heme or Fe(III)-protoporphyrin IX], present in all cells, was found to be a GQ-ligand, complexing tightly with both RNA and DNA GQs (K_d values in the 10–500 nM range[53,113–119]. As explained in section 1.8, Hemin binds by end-stacking on the terminal G-quartets of GQs; it has been shown that parallel-stranded GQs bind hemin more strongly than antiparallel-stranded ones[121]. *In vitro*, such GQ–hemin complexes, in the presence of hydrogen peroxide (or gaseous oxygen in the presence of a reductant, such as ascorbate), show robust, DNA/RNA-enhanced oxidative catalysis of both 1-electron (peroxidase) and 2-electron (peroxygenase/oxygenase) oxidation reactions[53,113–119,121]. Since the discovery of the complexation of hemin by GQs in 1998,

this interaction has been intensively investigated, and no other fold of DNA or RNA (with the exception of the non-physiological isoguanine pentaplexes[124] has been reported either to complex hemin or to activate it toward oxidative activity[121,199]. In a most interesting development, experimental evidence has recently been reported to validate earlier conjectures[116,115] that cellular RNA/DNA GQs naturally sequester intracellular hemin[218].

Recently, Owen et. al. has reported that *in vitro*, treating GQ–hemin complexes with hydrogen peroxide and the commercially available reagent biotin tyramide (BT), leads to efficient covalent biotinylation of the GQs themselves[219]. This can be conveniently monitored by complexing the biotinylated GQs with streptavidin (StAv) and observation of concomitant electrophoretic mobility retardation of these GQ•biotin-StAV complexes in polyacrylamide gels. By contrast, initial experiments with DNA *duplexes* co-dissolved with GQ–hemin complexes showed no trace of duplex biotinylation and, correspondingly, no StAv-retarded gel mobility of the duplexes. That the StAv-shifted DNAs were indeed biotinylated was confirmed using MALDI-TOF mass spectrometry analysis[219].

Herein, I report a thorough investigation of the extent, specificity and spatial limits of the GQ self-biotinylation reaction as it occurs *in vitro*. I then go on to provide convincing results using this methodology for biotinylation of nucleic acids within freshly dissected, live *Drosophila* tissue.

4.2. Materials and Methods

4.2.1. Chemicals and DNA oligonucleotides

SAvPhire Monomeric StAv (‘monoavidin’) was purchased from Sigma-Aldrich. Trolox ((±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid) was from Sigma-Aldrich. Luminol (Pierce ECL Western Blotting Substrate) and StAv-HRP were from ThermoFisher Scientific.

All DNA and RNA oligonucleotides were purchased from the Core DNA Services Inc. (Calgary, Canada). The oligonucleotides were first treated for 30 min with 10% v/v aqueous piperidine at 90°C to cleave DNA strands containing synthesis-related chemical lesions, lyophilized, and size purified in 8–12% denaturing gels. The DNA was then ethanol precipitated and dissolved in TE buffer (10 mM Tris, pH 7.4, 0.1 mM ethylenediaminetetraacetic acid) to make stock solutions of desired concentrations. All DNA oligonucleotides used in this study were gel-purified by size. As required, oligonucleotides were 5′-labeled with ^{32}P using γ-^{32}P adenosine triphosphate (ATP) and a standard kinasing protocol, then polyacrylamide gel electrophoresis-purified. The DNA and RNA oligonucleotides used in this study are shown in **Table 4-1**

Table 4-1. DNA and RNA sequences used in this chapter. Various GQ and non-GQ forming DNA/RNA sequences used in this study.

CatG4	5′-TGG GTA GGG CGG GTT GGG AAA (21 nt)
CatG4-ext	5′-ACA TAG CTG ACT GGC TTG ATT TTG GGT AGG GCG GGT TGG GAA ATA TCG AAT TCT CAG CCT ACA CTG CAG TAC TAG TAC ATA TCA (84 nt)
dsDNA Watson strand	5′-CCC ATT CTA TCA TCA ACG GGT ACA AAC GAG TCC TGG CCT TGT CTG TGG AGA CGG ATT ACA CCT TCC CAC TTG CTG (75 nt)
dsDNA Crick strand	5′-CAG CAA GTG GGA AGG TGT AAT CCG TCT CCA CAG ACA AGG CCA GGA CTC GTT TGT ACC CGT TGA TGA TAG AAT GGG (75 nt)
CatG4-T7	5′-TGG GTA GGG CGG GTT GGG AAA TAT CCT ATA GTG AGT CGT ATT AAA G (46 nt)
ssDNA/BLD:	5′-CTT TAA TAC GAC TCA CTA TAG G (22 nt)
CatG4-ext2	5′-TGG GTA GGG CGG GTT GGG AAA TAT TTT AGC TCA CGA GAC GCT CCC ATA GTG ACC TAT AGT GAG TCG TAT TAA AGT CCG GAG GAC TGT CCT CCG GC (95 nt)
Comp 1	5′-TCA CTA TGG GAG CGT CTC GTG AGC TAA A (28 nt)
Comp 3	5′-GCC GGA GGA CAG TCC TCC GGA (21 nt)
MYC	5′- TGA GGG TGG GTA GGG TGG GTA A (22 nt)
KIT1	5′- AGG GAG GGC GCT GGG AGG AGG G (22 nt)
KIT2	5′- CGG GCG GGC GCG AGG GAG GGG (21 nt)
SPB1	5′- GGC GAG GAG GGG CGT GGC CGG C (22 nt)
TBA	5′- GGT TGG TGT GGT TGG (15 nt)
hTELO	5′- GGT TAG GGT TAG GGT TAG GGT TAG GGT TAG (30 nt)
rBCL2 (RNA)	5′- r(AGG GGG CCG UGG GGU GGG AGC UGG GG) (26 nt)
rNRAS (RNA)	5′- r(AGG GAG GGG CGG GUC UGG G) (19 nt)
r(ssRNA)	5′- r(UGA UUA GGA UCU GCC AAC CGU G) (22nt)

4.2.2. Standard GQ biotinylation reaction

GQ DNA stock solution was diluted appropriately into Q Buffer (40 mM HEPES, pH 8.0, 20 mM potassium chloride, 1% dimethylformamide and 0.05% Triton X-100), and then heat-denatured at 100°C for 3 min, followed by cooling to 22°C by directly transferring the reaction tubes from boiling water bath to that temperature. Hemin was added to the folded DNA, and the solution rested for 10 min to assist heme–GQ complex formation. BT and H_2O_2 were then added to initiate the biotinylation reaction for various time points. The reactions were quenched with the addition of 10 U bovine liver catalase (Sigma-Aldrich).

4.2.3. Native gel electrophoresis and gel data analysis

DNA biotinylation was verified using a StAv gel shift assay in 7.5% non-denaturing/native polyacrylamide gels (acrylamide:bis = 29:1) run in 50 mM Tris borate (TBE) buffer. Biotinylated DNA was mixed with StAv in aqueous solution prior to loading in native gels run at 22°C with efficient cooling. Imaging and densitometry analysis of ^{32}P-labeled gel bands were done using a Typhoon 9410 Phosphorimager (Amersham Biosciences). Quantitation was carried out using ImageQuant 5.2 software (Amersham).

4.2.4. Determination of biotinylation stoichiometry using 'monoavidin'

For determining biotinylation stoichiometries, 1 μM 'CatG4' DNA was denatured at 100°C and refolded in Q buffer for 30 min at 22°C. Hemin was added to 5 μM, and the mixture allowed to incubate for a further 10 min. A total of 500 μM BT and 1 mM H_2O_2 were now added to initiate the biotinylation reaction, which proceeded for 30 min. The reaction was quenched by the addition of 10 U of catalase. The treated DNA was recovered by ethanol precipitation and mixed with aqueous solutions of either StAv or Monomeric Avidin ('Monoavidin' or MAv). The biotinyl-DNA/protein complexes were then run in a native polyacrylamide gel run at 4°C.

4.2.5. Biotinylation competition experiments

For dilute solution competition experiments, 'CatG4-ext' (0.01 μM, either 5′-^{32}P-labeled or not, depending on the experiment) was heat-denatured for 3 min at 100°C. Following 5 min of incubation in Q buffer at 22°C, heme was added to 5 μM and rested 10 min to assist complexation. Duplex DNA (dsDNA, made from the annealing of 100 μM each of 'dsDNA Watson Strand' and 'dsDNA Crick Strand', 5′-^{32}P-labeled or not, depending on the experiment), or ssDNA (100 μM, 5′-^{32}P-labeled or not) was now added. The final reaction buffer (QD Buffer: 40 mM HEPES, 40 mM Tris, pH 8.0, 20 mM KCl, 26 mM $MgCl_2$, 2.5 mM spermidine-Cl_3, 1% dimethylformamide and 0.05% Triton X-100) contained both potassium to stabilize the GQ and magnesium and spermidine to stabilize the duplex. The total DNA/hemin mixture was equilibrated at 22°C for 30 min, following which, BT (to 500 μM) and H_2O_2 (to 1 mM) were added and the peroxidase reaction allowed to proceed for 30 min. The reaction was quenched using 10 U of bovine liver catalase (Sigma). The treated DNA mixture was recovered by ethanol

precipitation, and the 70% ethanol-washed DNA pellet was dissolved in an aqueous solution of 25 µg (45 µM) StAv prior to running in a 7.5% native polyacrylamide gel run in 50 mM TBE buffer.

For the labelling competition experiments carried out at ultra-high overall DNA concentrations, the procedure was the same as above, with the exception that sheared duplex salmon sperm DNA (Sigma) was added to 17.5 mg/ml to the dsDNA mixture with hemin/GQ prior to 30-min equilibration at 22°C and initiation of the peroxidase/biotinylation reaction.

4.2.6. Determination of distribution of biotinylation within a GQ

CatG4 (1 µM) and other singly riboside-modified oligonucleotides, 'CatG4_R_x' (*x* = 1, 2, 3) (1 µM), were denatured for 3 min at 100°C and refolded in Q Buffer for 30 min at RT. A total of 5 µM heme was then added and the solution equilibrated for 10 min. Following this, 500 µM BT and 1 mM H_2O_2 were added to initiate the reaction, which proceeded for 30 min at 22°C prior to quenching by addition of catalase. The DNA was then ethanol precipitated, dissolved in TE buffer, and the solution was divided into two halves. To cleave a given 'CatG4_R_x' at its internal ribonucleotide one half of the DNA solution was treated with 0.25 M NaOH at 90°C for 5 min; following which, the solution was neutralized with equimolar HCl. The NaOH-cleaved strands were resolved and purified from a 10% denaturing gel. Biotinylated DNA species were identified via treatment with StAv and subsequent analysis in 7.5% native gels.

4.2.7. Determination of the radius of active biotinylation

A total of 1 µM 'CatG4-T7' and 1 µM 'ssDNA' (both 5′-^{32}P-labeled with γ-^{32}P ATP) were pre-denatured separately for 3 min at 100°C in TE buffer. They were mixed together in QD Buffer and allowed to anneal by slow cooling (from 100°C to 20°C at a rate of 7.5°C/min) in a Thermocycler. The solution was now made up to 5 µM heme and allowed to equilibrate for 10 min. A total of 500 µM BT and 1 mM H_2O_2 were added and the biotinylation reaction allowed to proceed for 30 min. The reaction was quenched with catalase and the two component DNA strands ('CatG4-T7' and 'ssDNA') were separated and purified in an initial 8% denaturing gel. The recovered DNA, ethanol precipitated, was redissolved and treated with StAv prior to running in a 7.5% native gel.

For investigation of longer-range biotinylation along the length of GQ-duplex composites, three short oligonucleotides, ‘Comp-1’, ‘ssDNA’ and ‘Comp-3’—each complementary to a different stretch of the oligonucleotide ‘CatG4-ext2’—were annealed together at 1 μM concentration each to ‘CatG4-ext2’ (with only one of the shorter oligonucleotide ^{32}P-labeled at a time). The subsequent procedure was carried out as described for the complex formed between ‘CatG4-T7’ and ‘ssDNA’, above.

4.2.8. Competition with GQ-binding ligands

A total of 1 μM ‘CatG4-ext’ was denatured for 3 min at 100°C and refolded in Q buffer for 30 min at 22°C. It was then made up to 5 μM hemin and rested for 10 min. Different concentrations (0–200 μM) of a given GQ-binding ligand was added, and the solution equilibrated further for 10 min. Following this, the solution was made up to 500 μM BT and 1 mM H_2O_2 to initiate the biotinylation reaction, which proceeded for 30 min at 22°C prior to quenching by the addition of catalase. The DNA was then recovered by ethanol precipitation and co-dissolved with StAv prior to running on native gels for analysis.

4.2.9. In vivo biotinylation reaction and biotin blotting

Drosophila melanogaster salivary glands were dissected into PBS buffer (phosphate buffered saline, pH 7.4) and incubated with 50 μM hemin, 3 mM BT for 15 min, following which it was given a pulse of 10 mM H_2O_2 for 2–3 min (H_2O_2 was added and the solution containing the tissue shaken gently on a shaker for 2–3 min). Following this treatment, the glands were washed twice with PBS buffer containing radical quenchers and peroxidase inhibitors (5 mM Trolox, 10 mM sodium azide, 10 mM sodium ascorbate). Genomic DNA as well as total RNA were extracted from treated salivary glands using a ‘DNeasy Blood & Tissue Kit’ (Qiagen) and an ‘RNeasy Mini Kit (Qiagen), respectively, as per the manufacturer's protocols. To take extra precautions to ensure the purity of the DNA extractions, DNA as extracted using the kit was given further treatments of RNase and proteinase as follows: first, the DNA was incubated with RNase A at 37°C for 1 h, after which it was incubated with Proteinase K at 50°C for 12–14 h. Following the above treatment, the DNA solution was extracted twice with phenol: chloroform: isoamyl alcohol (25:24:1), and then once with chloroform alone. The treated DNA was then recovered by way of two successive ethanol precipitations, followed by a

70% ethanol wash. The purified DNA pellet was then dissolved in TE buffer for further downstream analysis.

A total of 2–5 µL of purified total RNA or genomic DNA was then spotted onto a Hybond N+ membrane using a standard dot blot transfer apparatus and were crosslinked to the membrane using a UV Crosslinker (a UV Stratalinker 2400) using 254 nm light source, at a flux of 120 000 µJ for 30 s. The membrane was then treated with Blocking Buffer [1 × Tris Buffered Saline (TBS) containing 5% bovine serum albumin, 0.5% sodium dodecyl sulphate (SDS), 0.1% Ficoll, 50 µg/ml salmon Sperm DNA and 0.5 ug/ml poly d-IC DNA] for 90 min at 22°C. The blocking buffer was discarded and the membrane swirled overnight at 4°C with fresh blocking buffer of the same composition but containing, additionally, 1:10 000 diluted StAv-HRP. The next day, the StAv-HRP containing buffer was discarded and the membrane thoroughly washed, with swirling in Washing Buffer (1 × TBS containing 0.5% SDS) twice for 45 min each at 22°C; twice for 20–25 min each at 45–50°C; and once more for 10 min at 22°C. The thoroughly washed membrane was then developed with the ECL System (ThermoFisher Scientific) using the manufacturer's protocol. Chemiluminescence from biotinylated RNA/genomic DNA was captured using a ChemiDoc™ Imaging System.

4.2.10. LC-MS protocol

A Bruker maXis Impact Quadrupole Time-of-Flight LC/MS System was used for the analysis. The system consists of an Agilent 1200 HPLC and a Bruker maXis Impact Ultra-High Resolution tandem TOF (UHR-Qq-TOF) mass spectrometer. The Software used was Compass 1.5. For the MS, the ionization mode was Negative Electrospray Ionization (-ESI). Gas Temp: 180°C. Gas Flow: 8 l/min. Nebulizer: 2 bar. Capillary Voltage: 3000 V. Mass Range: 50–1500 Da. Calibrant: Sodium Formate. For the HPLC, a Spursil C18 column with 3 micron particle size, 30 mm length × 3.0 mm diameter (Dikma Technologies) was used. The column temperature was: 30°C. For the solvent gradients, Solvent A was water with 0.1% formic acid; and Solvent B was acetonitrile with 0.1% formic acid.

4.2.11. Biotinylated RNA enrichment

Dynabeads M270 Streptavidin magnetic beads (Thermofisher Scientific, Catalog: 65305) were washed twice with solution A (0.1 M NaOH, 0.05 M NaCl). Beads were washed once more in 1X Solution B (Binding and Wash Buffer(2X): 10 mM Tris pH 7.4, 1mM EDTA, 2 M NaCl, 0.15 mg/ml Heparin) and were finally suspended in solution B. Total RNA, extracted from Drosophila salivary glands, were mixed with prewashed M270 beads and were allowed to bind on a rotor for 30 min at 22^{o}C. Post-binding, beads were washed thrice in solution B at room temperature and were then suspended in RNase free water. Beads (bound with the biotinylated RNA) were used either directly for cDNA synthesis using superscript III (reverse transcriptase) or the biotinylated RNA bound to the beads were eluted as per the protocol reported by Alice Ting and co-workers[220] for library preparation and sequencing.

4.2.12. High-Throughput Sequencing

Sequencing library was prepared using NEBNext Ultra II Directional RNA Library Prep Kit according to the standard protocol described in manufacturer's instructions. The sample library was sequenced with 2x75 bp reads on MiSeq nano flow cell, which produces upto 1 million raw reads.

4.2.13. Bioinformatics Analysis

RNA was either a) fragmented or b) left unfragmented prior to 75 bp paired-end sequencing on a MiSeq nano flow cell, resulting in 629,857 and 633,319 raw reads, respectively. Libraries were adaptor trimmed and quality filtered using Trimmomatic 0.36[221]. Reads were further aligned to PhiX and ribosomal RNA sequences using bbmap 38.87 (https://sourceforge.net/projects/bbmap/), resulting in the removal of 1.2 % and 89.0 % of the total reads, respectively. The final set of filtered reads contained 74,092 and 65,028 reads in a) fragmented or b) left unfragmented libraries, respectively.

Filtered reads were mapped to the *Drosophila melanogaster* genome (assembly Release 6 plus ISO1 MT; downloaded from RefSeq on 2020-10-28), using STAR 2.7.6a[222] with settings: -- genomeSAindexNbases 12 --sjdbOverhang 74. Filtered reads were mapped to the *D. melanogaster* transcriptome (Release 6 as above) using bowtie

1.0.0[223]. Further, filtered reads were assembled using rnaSPAdes 3.13.1[224], and mapped to the assembly with bowtie 1.0.0. Coverage BED files were generated for all alignments with bedtools 2.17.0[225] using settings: - bg -split. Feature alignments were counted using featureCounts 2.0.1[226].

4.3. Results

4.3.1. Biotinylation of diverse DNA and RNA GQs

As described in the Introduction, no DNA or RNA folded structure other than GQs have been shown either to bind hemin or to activate such bound hemin toward oxidative activity. GQs, however, are a highly polymorphic class of structures, with a variety of strand orientations and topologies[48]. I wished, first, to investigate, whether DNA GQs of different strand orientations (all-parallel, all-antiparallel and mixed orientation[47,48] as well as RNA GQs (which invariably form all-parallel stranded GQs[46]) could all self-biotinylate by way of the peroxidase activity of hemin complexed to them (sequences given in **Table 4-1**)[46,47,213]. 5′-^{32}P-labeled all-parallel DNA GQs (‘CatG4’, ‘MYC’, ‘KIT1’ and KIT2), all-antiparallel DNA GQs (‘SPB1’ and ‘TBA’), a mixed strand-orientation DNA GQ (‘hTELO’) as well as parallel RNA GQs (‘rNRAS’ and ‘rBCL2’) were subjected to standard biotinylation reactions (see ‘Materials and Methods’ section); together with a DNA oligonucleotide (‘BLD’) and a RNA oligonucleotide (‘r(ssRNA)), that are not capable of forming GQs[46,47,213]. Following quenching of the biotinylation reactions, the DNAs and RNAs were treated with StAv and analyzed in non-denaturing gels. **Figure 4-1** shows that the parallel GQs, whether DNA or RNA, were efficiently biotinylated (~38–72 %); whereas the antiparallel GQs were also biotinylated at lower but detectable levels. By contrast, no trace of biotinylation was seen with the negative control oligonucleotides, BLD and r(ssRNA).

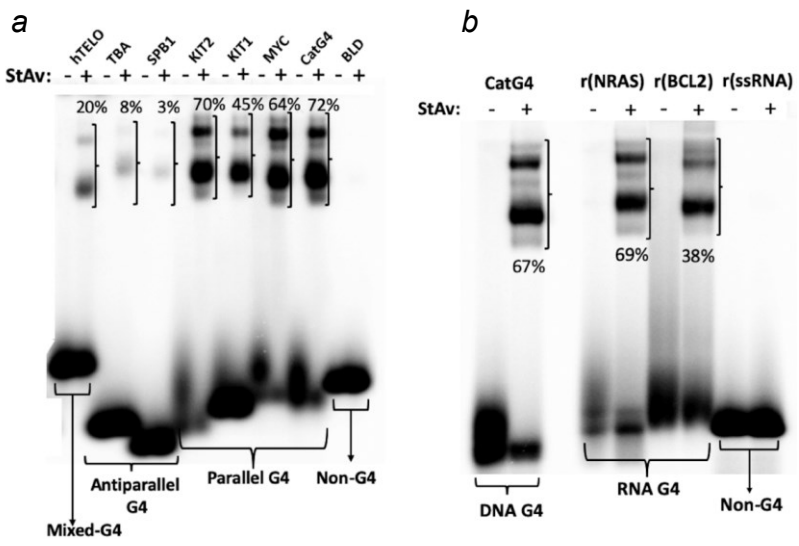


Figure 4-1. Verification of various GQ-forming DNAs (a) and RNAs (b). Biotinylation reaction performed on various orientations of DNA GQ as well as on RNA GQ with proper controls

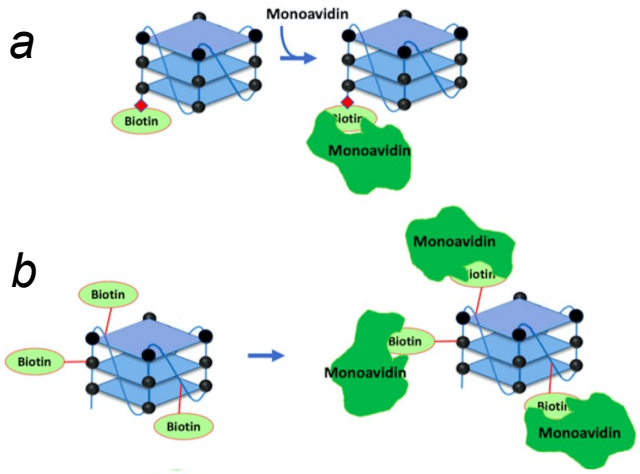
The biotinylation of all the GQs examined here is significant for the anticipated biotinylation experiments to be carried out *in vivo*; even low levels of biotinylation (as opposed to *no* biotinylation) of the pertinent DNA or RNA sequences should enable their pulldown using StAv, followed by NextGen sequencing. Recently, indeed, RNA self-biotinylation has been characterized *in vitro* in a separate study[227].

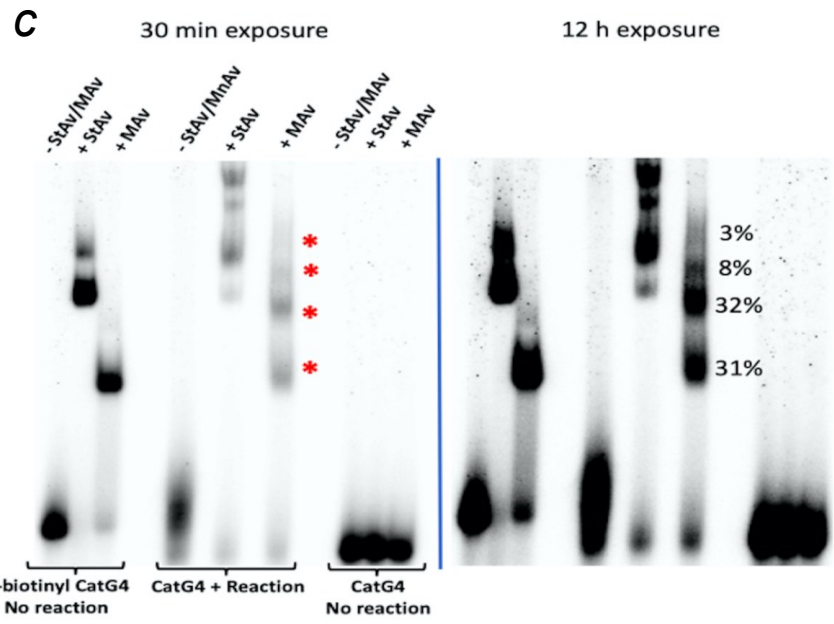
4.3.2. How many biotins covalently attach to each CatG4?

The tight and highly specific biotin-StAv interaction provides, in principle, a convenient approach for quantitating the number of biotins that covalently attach to each GQ-forming oligonucleotide under our reaction conditions. However, the use of tetrameric StAv (capable of binding up to four biotins/biotinylated DNAs) complicates such an estimation. Indeed, in a standard experiment involving GQ-forming oligonucleotides, multiple StAv-retarded gel bands of biotinylated GQ are seen (such as in **Figure 4-1**), which are difficult to interpret in terms of binding stoichiometries.

To address this problem, I decided to use a MAv (SAvPhire Monomeric StAv or 'monoavidin'), with its property of forming an exclusively 1:1 complex with biotin. To generate a reference complex consisting of the GQ forming 'CatG4' DNA oligonucleotide with a single appended biotin, in turn complexed to a single monoavidin molecule, I used

a ‘CatG4’ oligonucleotide chemically synthesized with a 3′-appended biotin moiety (‘3′-biotinyl CatG4’). **Figure 4-2.a** schematically shows such a DNA and its expected 1:1 complex with monoavidin. **Figure 4-2.b** shows the expected binding scenario of a multiply biotinylated ‘CatG4’ with monoavidins.





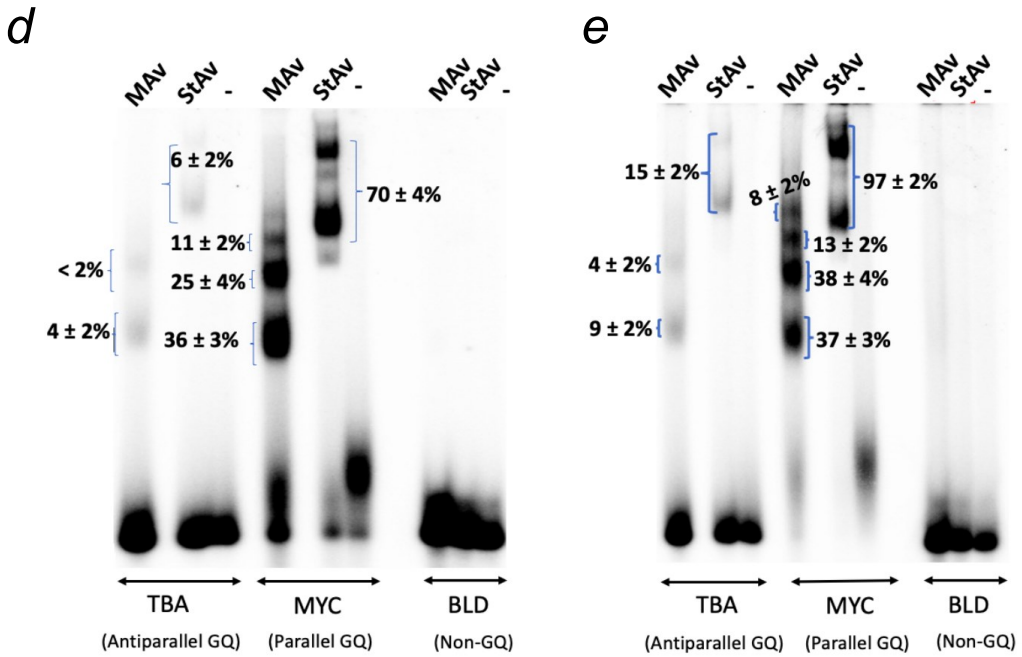


Figure 4-2. Quantitation of Self-Biotinylation of various GQs. ***(a, b)*** Schematic showing 1:1 complex formed by Biotin and MAv. **(c)** Quantitation of biotinylation for CatG4. Same gel has been shown at two different exposures. **(d, e)** Verification of GQ self-biotinylation efficiency at two different Heme concentrations.

Figure 4-2.c shows that while 3′-biotinyl 'CatG4' mixed with StAv generates two retarded bands of uncertain DNA:protein stoichiometry ('3′-biotinyl DNA; no reaction; + StAv'), when it is mixed instead with monoavidin ('3′-biotinyl DNA; no reaction; + MAv'), a single, clearly defined, monoavidin-retarded band results. This result provides us with a positive control for investigating the extent of biotinylation of 'CatG4' under our reaction conditions, based on the presumption that each biotin covalently linked to a given CatG4 molecule would bind a single monoavidin (shown conceptually in **Figure 4-2.b**). **Figure 4-2.c** shows two different exposures of the same native gel. Two strong retarded bands (corresponding to two biotins appended to one CatG4) can be seen but up to a total of four retarded bands can be discerned (shown by red asterisks in the lane marked 'CatG4 + reaction + MAv'). The percentages of these retarded bands are given in **Figure 4-2.c**. In summary, with folded 'CatG4' (a 21-nt, three G-quartet GQ), 2–4 biotins typically attach under our reaction conditions.

I asked the following further questions: (i) how reproducible are the levels of biotinylation observed for individual GQs under a given reaction condition (such as

shown in **Figure 4-1**)? Furthermore, (ii) is there a higher efficiency of biotinylation with a higher (50 μM rather than 5 μM) heme concentration? **Figure 4-2**.**d** shows a gel reporting the standard biotinylation of GQs with 5 μM heme. Two independent reactions were carried out, and the numbers shown in the gel reflect that. The percentage values of monoavidin- versus StAv-retarded DNA for both the TBA (antiparallel) and the MYC (parallel) GQs are mutually consistent. Further, the error bars (representing deviations from the mean) are small, showing a high reproducibility of the biotinylation reaction under a given reaction condition. **Figure 4-2.e** shows that notable biotinylation enhancements (97% for MYC and 15% for TBA) are achieved with 50 μM as opposed to 5 μM heme. Again, the reproducibility of biotinylation levels seen with two independent experiments using 50 μM heme is high.

4.3.3. Biotinylation competition experiments with large excesses of ssDNA or dsDNA in dilute and concentrated (gelated) solutions

Of the variety of secondary and tertiary structures formed by natural RNA and DNA, only GQs have been shown to bind hemin. Hemin neither binds to ssDNA or dsDNA nor is it activated toward oxidative catalysis by those DNA folds[115,121]. If self-biotinylation is intended to be a reliable method for tagging GQs *in vivo*, it would be necessary to estimate the degree to which, in complex mixtures of ss/dsDNA and GQs, promiscuous labelling of the ds or ss DNA might occur.

I investigated this first in dilute DNA solution, with 10 nM 'CatG4-ext' DNA co-dissolved with very large (10^4-fold) excesses of either 'ssDNA' or 'dsDNA'. This particular molar ratio was chosen based on rough calculations on the ratio of the size of the human genome ($\sim 3 \times 10^9$ bp) to the reported number of GQ-capable sites within the genome (7×10^5 bp)[66].

Figure 4-3.a illustrates the concept of the experiment. **Figure 4-3.b** shows the results of 'CatG4-ext' co-dissolved (10 nM) with 'ssDNA' (100 μM), the two DNAs being reciprocally ^{32}P-labeled. The DNA mixture was treated with hemin, BT and H_2O_2, purified by ethanol purification, mixed with StAv, and run in a native polyacrylamide gel. **Figure 4-3.b** shows that StAv-shifted gel bands (~50% of the total DNA in that lane, indicated with a red bracket) are visible only in the lane containing a mixture of ^{32}P-labeled 'CatG4-ext' DNA co-dissolved with unlabelled 'ssDNA'; and, that no trace of

such shifted bands can be seen from the lane containing ^{32}P-labeled 'ssDNA' mixed with unlabelled 'CatG4-ext' DNA. The biotinylation of 'CatG4-ext' is therefore completely specific, even in the background of a 10^4-fold excess of 'ssDNA'. **Figure 4-3.c** shows results almost indistinguishable from those seen in **Figure 4-3.b**, except, in this experiment, 10^4-fold excess 'dsDNA' replaces 10^4-fold excess 'ssDNA'.

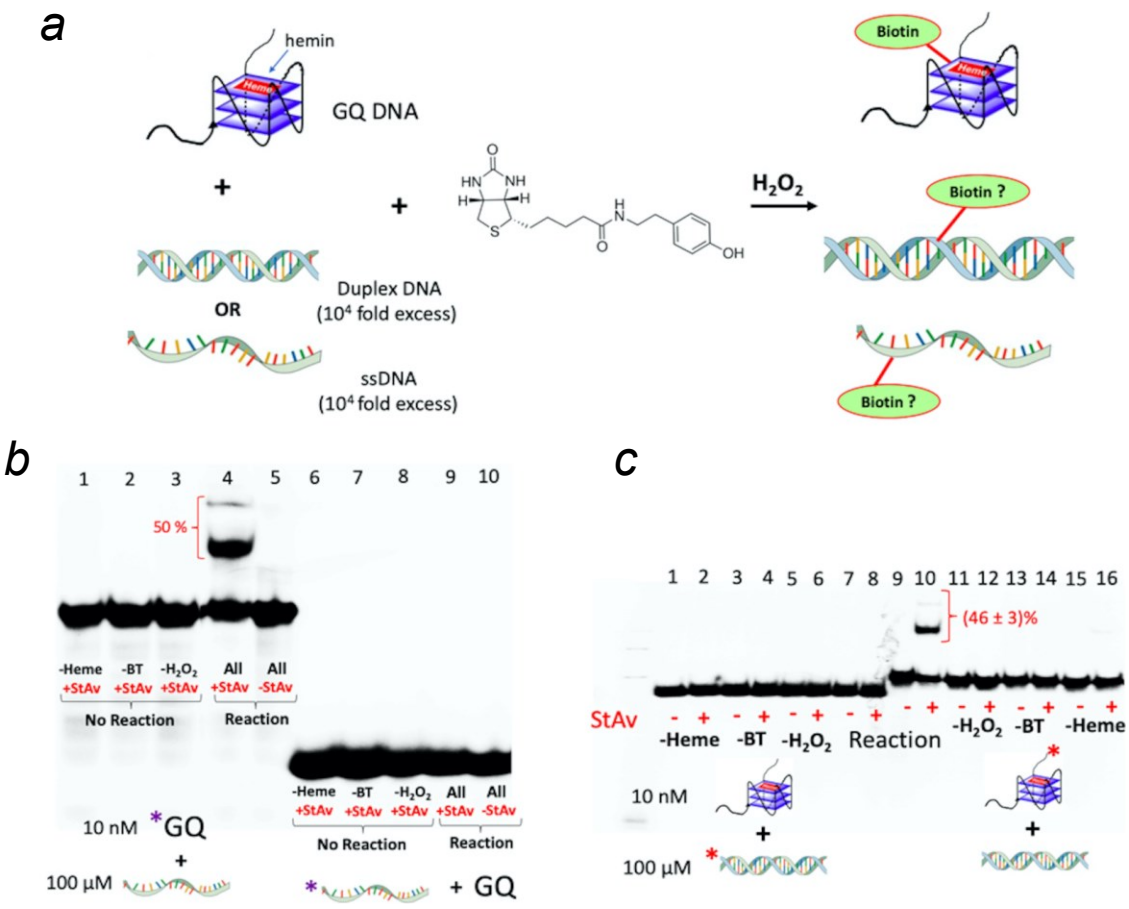


Figure 4-3. Specificity of GQ biotinylation over ss or dsDNA in dilute co-solutions. (a) Schematic showing design of the experiment. **(b and c)** Native gel showing reciprocally ^{32}P-end-labelled 'CatG4_ext' (10 nM) co-dissolved 100 µM 'ssDNA' (b) or 100 'dsDNA' (c) and treated with hemin, H_2O_2, and BT. Upon treatment with StAv, retarded mobility (biotinylated) bands (indicated by red brackets) were observed and then quantitated relative to the unbiotinylated DNA in those same lanes.

4.3.4. Does GQ biotinylation occur even in a highly concentrated DNA solution?

What happens when the biotinylation experiment, as above, is carried out not in dilute solution but in a milieu of highly concentrated, gelated DNA, such as might be

found in a cellular nucleus? To explore this question, we performed a new competition experiment using ^{32}P-labeled 10 nM 'CatG4-ext' and 100 μM 'dsDNA', either with no further added DNA (dilute solution) or within a highly concentrated (gelated) DNA environment containing a very high concentration of sheared salmon sperm genomic DNA (17.5 mg/ml), in QD Buffer. Under these DNA and salt conditions, such a solution has been shown rigorously to form a viscous and isotropic gel[228]; this was also our own observation. **Figure 4-4** shows that under these two very different solution conditions, biotinylation of the GQ formed by the 'CatG4-ext' remains comparably efficient in the dilute solution as well as in the highly concentrated DNA solution.

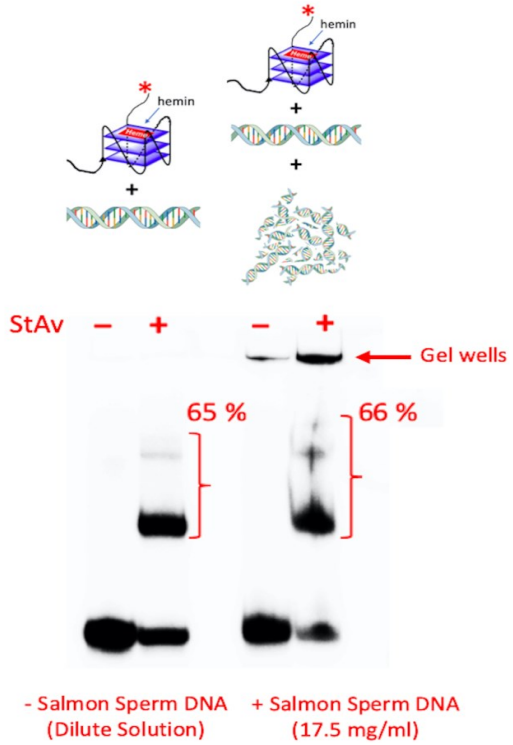


Figure 4-4. Specificity of GQ biotinylation under highly concentrated (gelated) solutions. Co-dissolved 10^4:1 molar mixtures of 'dsDNA':'CatG4-ext' were treated with hemin, H_2O_2 and biotin-tyramide (BT) either in dilute solution ('- Salmon Sperm DNA') or in a highly concentrated dsDNA solution ('Salmon Sperm DNA: 17.5 mg/ml').

It could be argued that within the gelated DNA regime generated by the salmon sperm DNA (which itself is not radiolabelled in our experiment) some promiscuous labelling of the concentrated salmon sperm DNA may occur. To address this issue, we describe experiments, below, to determine what the 'active zone' (or spatial restriction)

for biotinylation may be around a GQ-complexed hemin that oxidizes BT to its phenolic radical, in turn capable of labelling DNA.

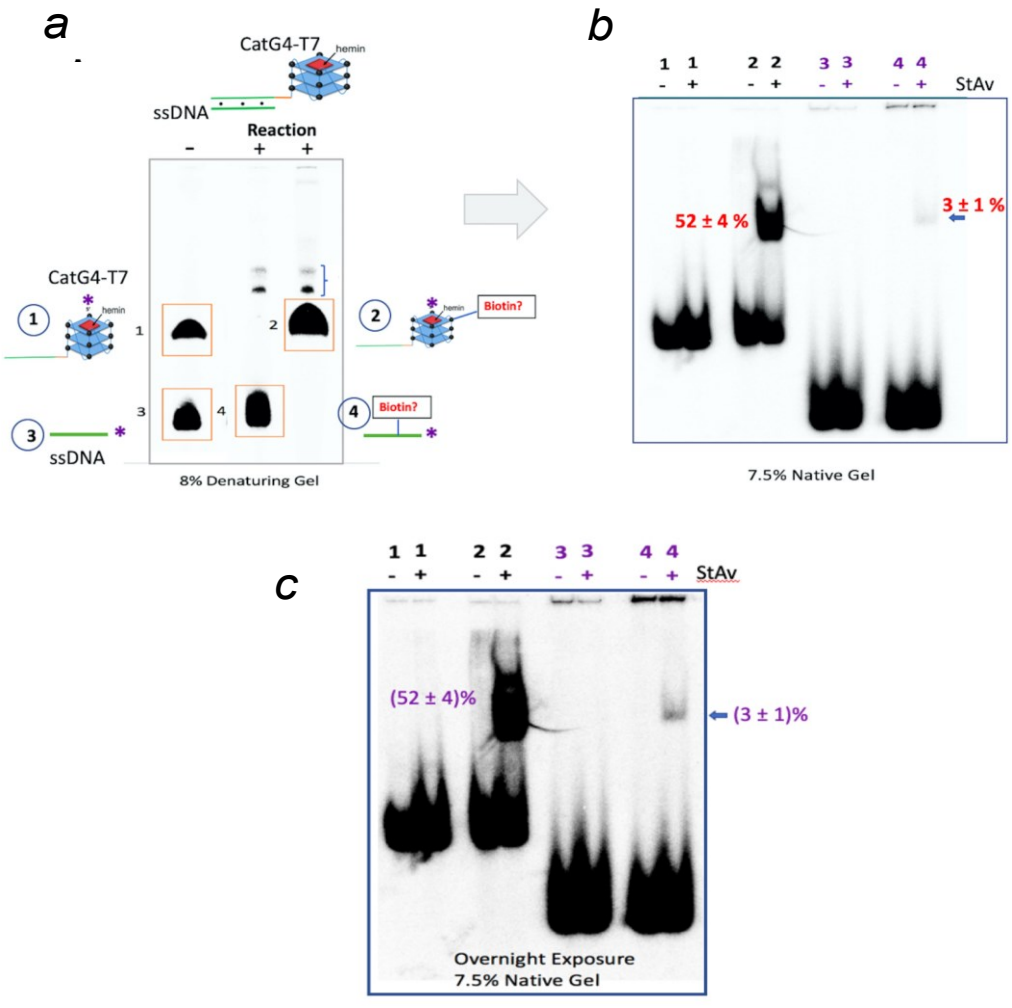
4.3.5. What is the spatial restriction of biotinylation in a GQ-duplex chimeric DNA?

To determine if non-GQ DNA duplex elements physically linked and therefore highly proximal to a GQ (such as might be found within a living cell) might also be targets for biotinylation, and to determine how far away from the CatG4-bound hemin such biotinylation might occur, I hybridized together the 46-nt 'CatG4-T7' and 22-nt 'ssDNA' to generate a GQ-duplex chimera consisting of a 22-bp duplex pendant from the 'CatG4' GQ by an AAA nucleotide linker (shown schematically in **Figure 4-5.a**). Biotinylation was performed with reciprocally 5′-^{32}P-labeled 'CatG4-T7' and 5′-^{32}P-labeled 22-nt 'ssDNA' (bound, to unlabelled 'ssDNA' and 'CatG4-T7', respectively), and complexed with hemin.

After the biotinylation reaction, a denaturing gel (**Figure 4-5.a**) was used, first, to separate and purify the radiolabelled 'ssDNA' and 'CatG4-T7' strands from their unlabelled, partially complementary strands. The purified DNA was then mixed with StAv and analyzed in a native gel (**Figure 4-5.b**) to quantitate the proportion of the purified, ^{32}P-labeled oligonucleotide that had undergone biotinylation. **Figure 4-5.b** shows that ~52% of the radiolabelled 'CatG4-T7' strand was StAv-shifted under our reaction conditions. Of course, this longer of the two oligonucleotides participates in both the GQ and duplex domains of the duplex-GQ chimera. The radiolabelled 'ssDNA', however, participates only in the duplex portion of the chimera, and **Figure 4-5.b** shows that only ~3% of this oligonucleotide is StAv-shifted (**Figure 4-5.c** shows a longer exposure of this gel).

To test the further reaches of a longer duplex linked to a GQ, I generated a different chimeric DNA complex (shown in **Figure 4-5.d**), in which the duplex component is 74 bp long. As shown in this figure, three short oligonucleotides 'Comp-1', 'ssDNA' and 'Comp-3' (with only ssDNA being 5′-^{32}P-labeled) were hybridized simultaneously to different stretches of the tailed GQ-forming oligonucleotide 'CatG4-ext2', to generate a quasi-continuous duplex appended to the GQ. The denaturing gel (**Figure 4-5.d**) shows the dissociated, ^{32}P-labeled 'CatG4-ext2' and 'ssDNA' out of the complete GQ-duplex

chimera, following either biotinylation under the specified conditions or not biotinylation. **Figure 4-5.e** shows a native gel that highlights StAv-shifted bands (shown with a red bracket) obtained from the DNA species 1–6 shown in **Figure 4-5.d** following their purification and mixing with StAv. Crucially, it can be seen that *no* StAv-shifted band is seen from bands 4 and 6 (representing 'ssDNA'). Thus, 'ssDNA', hybridized 32–53 bp away from the GQ in this construct, is not biotinylated at all.



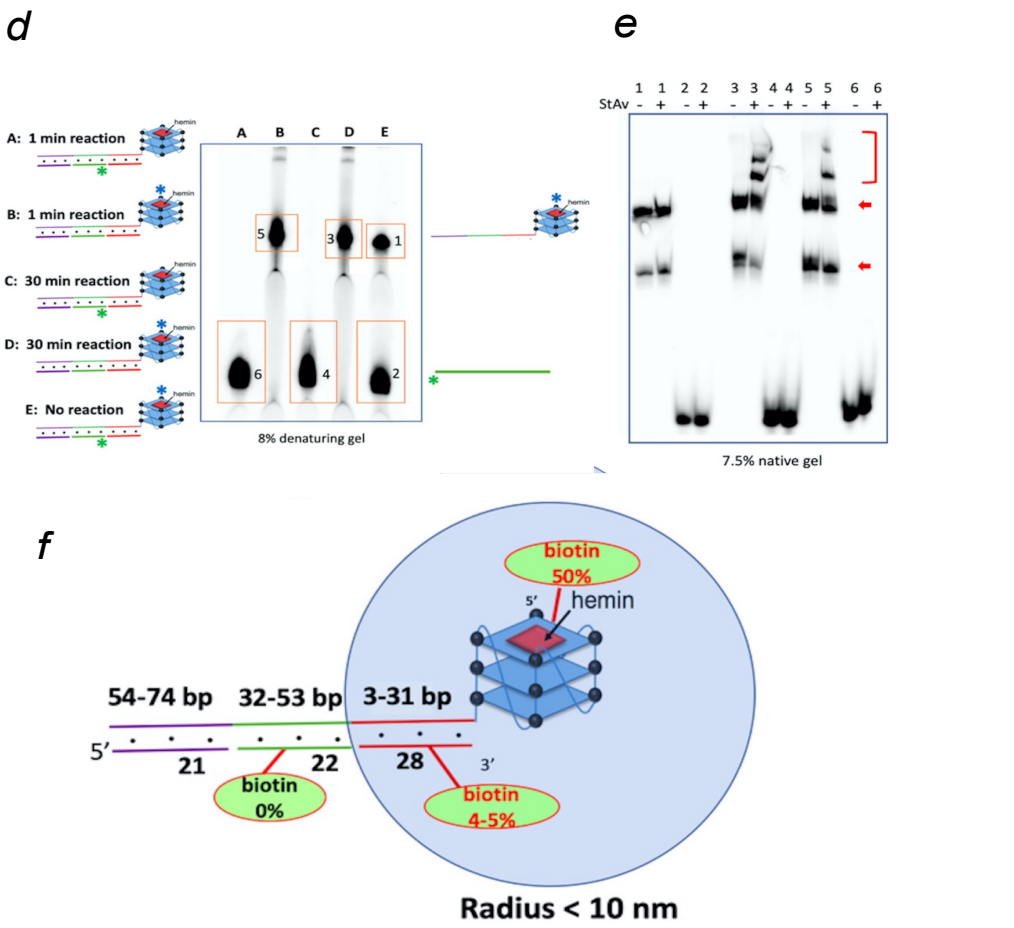


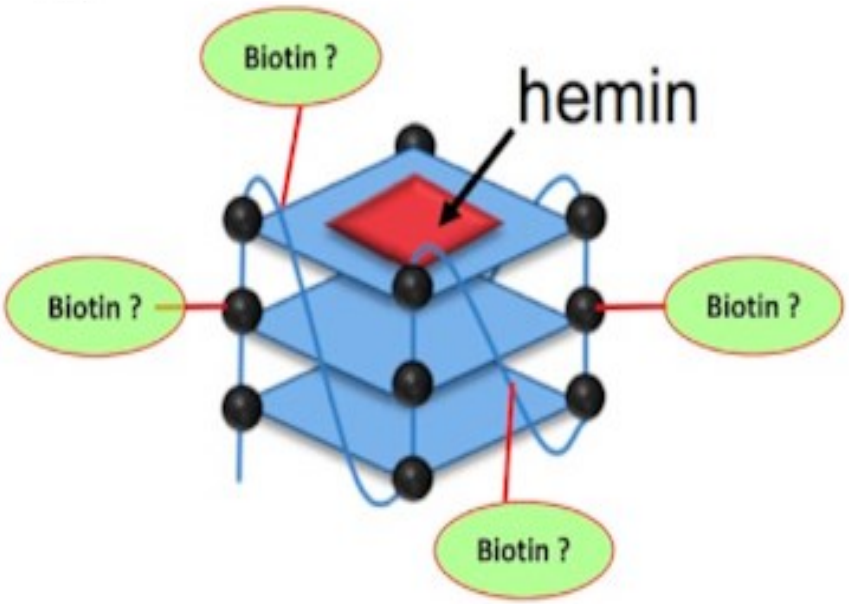
Figure 4-5. Radius of Biotinylation from GQ. **(a)** Top, the design of a duplex-GQ chimera, 46-nt 'CatG4-T7' hybridized to 22-nt 'ssDNA'. Bottom, an 8% denaturing gel showing the individually ^{32}P-labeled component oligonucleotides, 'CatG4-T7' (band '1') and 'ssDNA' (band '3') that make up the duplex-GQ chimera. Bands '2' and '4' represent, respectively, post-biotinylation ^{32}P-labeled 'CatG4-T7' (out of the duplex-GQ chimera containing non-radiolabeled 'ssDNA') and post-biotinylation ^{32}P-labeled 'ssDNA' (out of the duplex-GQ chimera containing non-radiolabeled 'CatG4-T7'). The minor bands shown with a bracket ({) represent inter-strand crosslinked minor products formed between 'ssDNA' and 'CatG4-T7'**.** **(b)** A native gel showing purified DNA isolated and purified from bands 1–4 shown in the denaturing gel (in a), run with either StAv added (+) or not added (−). The numbers shown in red indicate the percentage of total DNA StAv-shifted (and are therefore biotinylated) in the relevant lanes. **(c)** A longer exposure of the same gel shown in b. **(d)** Does biotinylation extend past 31 bp in a duplex linked to a GQ? To the left are shown a GQ-duplex chimera consisting of three short oligonucleotides, 'Comp-1', 'ssDNA' and 'Comp-3', hybridized simultaneously to different stretches of the tailed GQ-forming oligonucleotide 'CatG4-ext2'. Here, only either 'CatG4-ext2' or 'ssDNA' were 5′ ^{32}P-labeled (shown, respectively, as a blue asterisk and a green asterisk). The denaturing gel shows the radiolabeled 'CatG4-ext2' or 'ssDNA' from the complete GQ-duplex chimera, either biotinylated under the specified conditions or not. **(e)** Native gel showing StAv-shifted bands from DNA species 1–6 following purification from the denaturing gel shown in d. The bands shown with the red bracket are the StAv shifted bands. The multiple bands seen from 'CatG4-ext2' (shown with red arrows) represent different folded conformers formed in the native gel by this large oligomer. **(f)** The effective radius of biotinylation (~10 nm, representing a duplex of ~31 bp), estimated from the above experiments.

Figure 4-5.f summarizes the above data and proposes that the effective radius of biotinylation estimated from the above experiments is ~10 nm (representing a duplex of ~31 bp) away from the GQ.

4.3.6. Covalently appended biotins are distributed along the entire length of a GQ

I was interested to determine the distribution of biotin attachment along the 21-nt length of the 'CatG4' oligonucleotide (shown schematically in **Figure 4-6.a**). Does biotinylation occur, for instance, uniformly along the length of 'CatG4', or preferentially to its 3′ or 5′ ends? To address this question, we devise a number of variants of 'CatG4', all of which had the same base sequence as CatG4 itself but incorporated a single ribonucleotide at different locations within the deoxyribonucleotide. **Figure 4-6.b** shows the sequences of three CatG4 variants, 'CatG4_R1', 'CatG4_R2' and 'CatG4_R3'. The nucleotide marked in red in each case is the single embedded ribonucleotide. Asterisks show the site of ^{32}P-labeling (i.e., either at the 5′ or the 3′ end) in a given oligonucleotide. Following biotinylation under the standard conditions described above, the 5′- or 3′-labeled 'CatG4_R1' to 'CatG4_R3' oligonucleotides were treated, first, with hot alkali, to cleave their phosphodiester chains at the embedded ribonucleotides. Second, the two resulting pieces obtained from each of the oligonucleotides were separated and purified by denaturing gel electrophoresis. Third, the extent of gross biotinylation (i.e., the coupling of at least one biotin to a given ^{32}P-end-labeled DNA fragment, relative to that in the ^{32}P-end-labeled but uncleaved 'CatG4') was determined by StAv band-shift analysis. **Figure 4-6.c** shows a schematic for this experimental approach, and **Figure 4-6.d** shows the overall experimental data obtained from an analysis of the oligonucleotide 'CatG4_R1'. Quantitation of the DNA bands is displayed in **Figure 4-6.b**. The numbers shown below each sequence of 'CatG4_Rx' (where x = 1-3) indicate raw percentages of StAv-shifted bands relative to the total DNA in a given gel band. **Figure 4-6.e** tabulates the absolute percentages of biotin localization along the full length of the 'CatG4' oligonucleotide. My approach does not attempt to identify individual bases or nucleotides or sites within them as specific atomic loci for biotin attachment; nevertheless, it can be seen that biotins are more or less evenly distributed along the length of 'CatG4'. The stretch …GGGC… within 'CatG4' has somewhat higher levels of biotinylation relative to the other three quadrants; the reason for this is not immediately clear. It is conceivable that the single dC in this stretch or its neighbouring phosphates are preferred sites for biotin attachment.

a



b

CatG4_R1: *TGG GTrA GGG CGG GTT GGG AAA
10 ± 2%

CatG4_R2: *TGG GTA GGG rCGG GTT GGG AAA
10 ± 2% 22 ± 1 %
32 ± 3%

CatG4_R3: *TGG GTA GGG CGG GTrU GGG AAA
10 ± 2% 22 ± 1% 10 ± 1%
42 ± 3%

CatG4_R3: TGG GTA GGG CGG GTrU GGG AAA*
10 ± 2% 22 ± 1 10 ± 1% 7 ± 1%
50 ± 2%

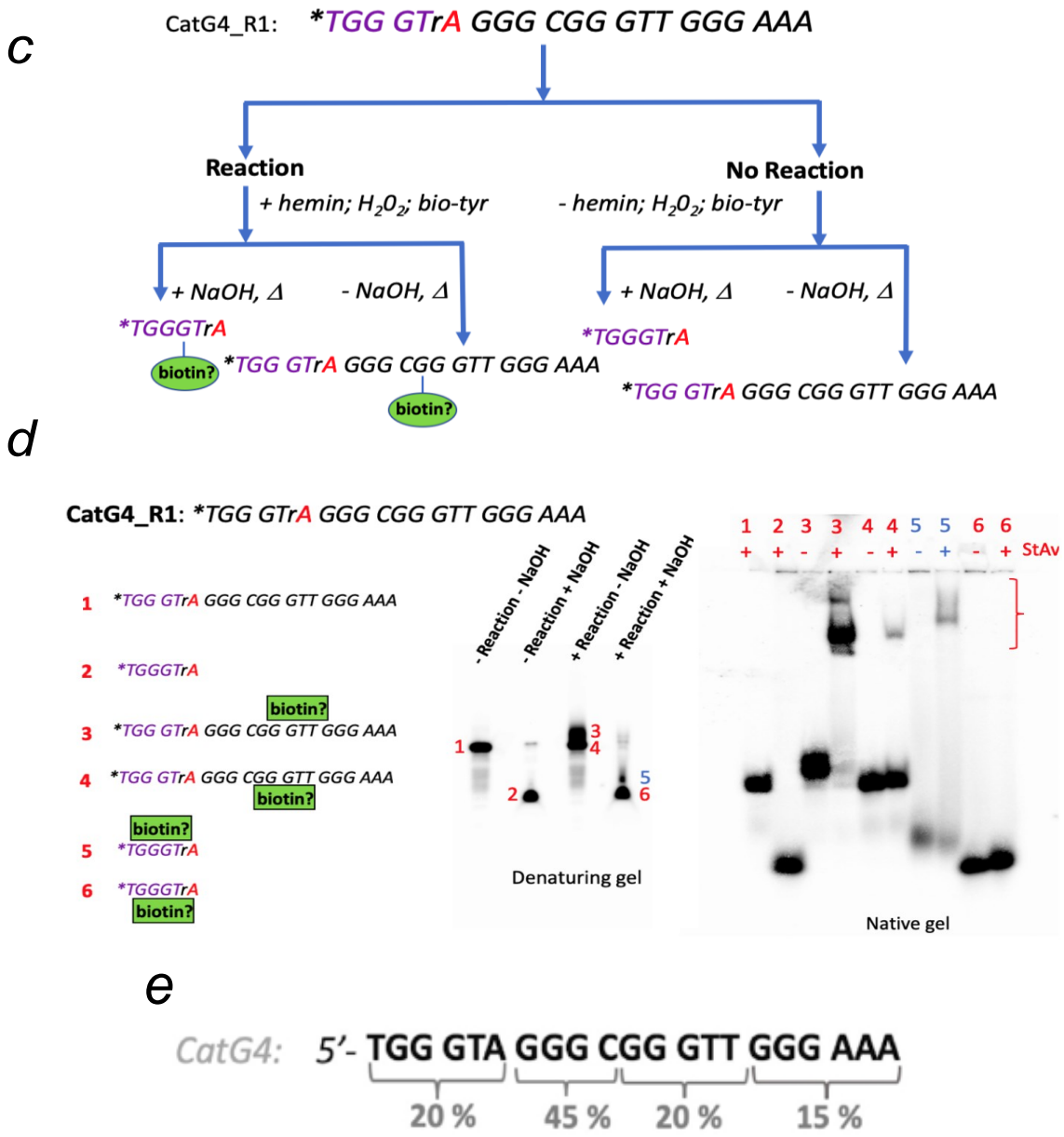
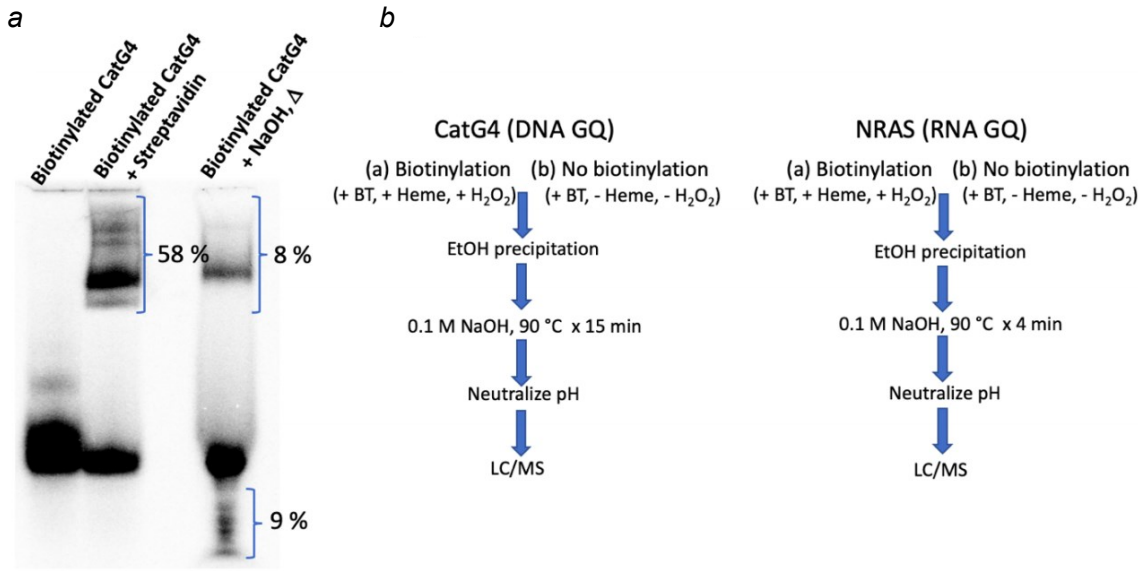
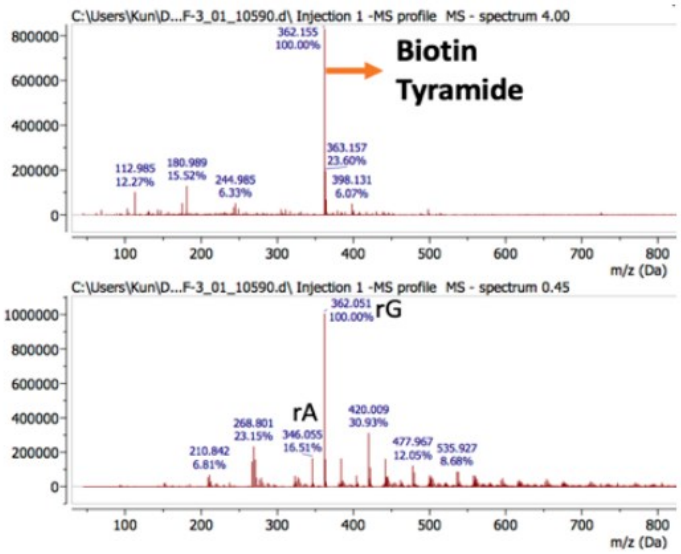


Figure 4-6. Preferential Sites of GQ biotinylation. **(a)** A schematic of the question being posed, about the distribution of biotin along the length of the CatG4 oligonucleotide**.** **(b)** The sequences of the four 'CatG4' variants, CatG4_Rx (where x = 1-3). The nucleotide marked in red is the single ribonucleotide within each of these oligonucleotides. The asterisk shows the site of ^{32}P-labeling (at the 5′ or the 3′ end). The numbers shown below each sequence indicate raw percentages of StAv-shifted bands relative to total DNA in a given gel band. The numbers obtained were from two independent sets of measurements. **(c)** Schematic showing the experimental approach to determining the distribution of appended biotins in the GQ formed by "CatG4". **(d)** Determination of the extent of biotinylation within two component pieces (sequences 1 and 2: shown in **Left**) of the oligonucleotide "G4_R1" (or CatG4_R1). Middle: A denaturing gel showing G4_R1 treated with hemin, BT, and H_2O_2 ("Reaction") or not treated, either cleaved in two by heating with NaOH (or not). The six ^{32}P-labeled fragments, 1-6, were purified, and then mixed with soluble streptavidin (StAv). These were then run on a native gel (Right). The red bracket on gel shows the streptavidin-retarded bands. **(e)** Absolute percentages of the likelihood of biotin localization along segments of the total sequence of the 'CatG4' oligonucleotide.

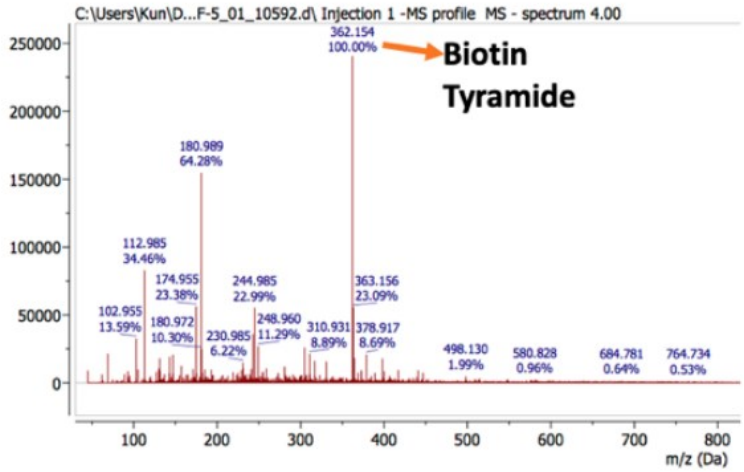
To obtain a level of deeper insight into the sites of GQ biotinylation, I carried out chemical and mass spectrometric experiments on biotinylated DNA (and RNA) GQs. First, I attempted to degrade biotinylated and unbiotinylated DNA (CatG4) GQs with the highly efficient Nuclease P1, to mononucleotides (5′-NMPs), such that the biotinylated mononucleotides could be analyzed by mass-spectrometry. However, I found that while the unmodified GQ was readily degraded by P1, biotinylated GQ was poorly and incompletely degraded. Attempting an alternative strategy, I tested whether the GQ biotinyl labels were hot base-labile (heating to 90° at neutral pH does not destroy the GQ biotinyl labels[219]). Accordingly, the DNA (CatG4) GQ was biotinylated in a standard reaction, and a portion of the resulting DNA was subjected to heating at 90°C in 0.1 M NaOH. **Figure 4-7.a** shows that upon such treatment, the level of DNA biotinylation (indicated by the proportion of DNA band-shifted by StAv binding in a native gel) decreased from 58% to only 8% of the DNA. Interestingly, ~9% of the starting DNA broke down into smaller fragments (such as might be expected if alkylation/biotinylation had occurred to the N7 position of guanines, for example).



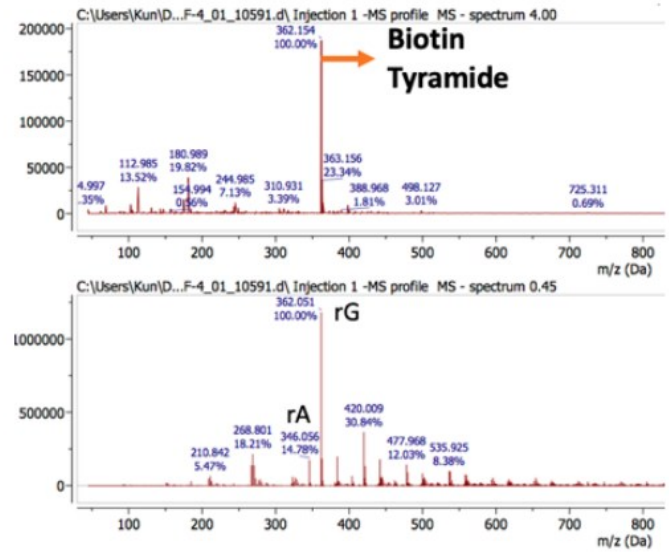
c


C:\Users\Kun\D...F-3_01_10590.d\ Injection 1 -MS profile MS - spectrum 4.00
362.155 100.00%
Biotin
Tyramide
363.157 23.60%
112.985 12.27%
180.989 15.52%
244.985 6.33%
398.131 6.07%
m/z (Da)
C:\Users\Kun\D...F-3_01_10590.d\ Injection 1 -MS profile MS - spectrum 0.45
362.051 100.00%
rG
rA
420.009 30.93%
268.801 23.15%
346.055 16.51%
210.842 6.81%
477.967 12.05%
535.927 8.68%
m/z (Da)

d


C:\Users\Kun\D...F-5_01_10592.d\ Injection 1 -MS profile MS - spectrum 4.00
362.154 100.00%
Biotin
Tyramide
180.989 64.28%
112.985 34.46%
174.955 23.38%
244.985 22.99%
363.156 23.09%
102.955 13.59%
180.972 10.30%
230.985 6.22%
248.960 11.29%
310.931 8.89%
378.917 8.69%
498.130 1.99%
580.828 0.96%
684.781 0.64%
764.734 0.53%
m/z (Da)

e


C:\Users\Kun\D...F-4_01_10591.d\ Injection 1 -MS profile MS - spectrum 4.00
362.154 100.00%
Biotin
Tyramide
180.989 19.82%
112.985 13.52%
363.156 23.34%
244.985 7.13%
310.931 3.39%
388.968 1.81%
498.127 3.01%
725.311 0.69%
m/z (Da)
C:\Users\Kun\D...F-4_01_10591.d\ Injection 1 -MS profile MS - spectrum 0.45
362.051 100.00%
rG
rA
420.009 30.84%
268.801 18.21%
346.056 14.78%
210.842 5.47%
477.968 12.03%
535.925 8.38%
m/z (Da)

f

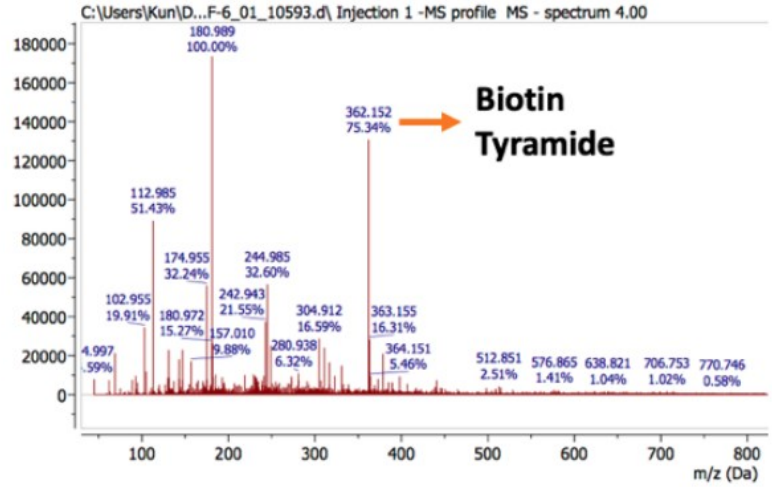

C:\Users\Kun\D...F-6_01_10593.d\ Injection 1 -MS profile MS - spectrum 4.00
180.989 100.00%
362.152 75.34%
Biotin
Tyramide
112.985 51.43%
174.955 32.24%
244.985 32.60%
242.943 21.55%
102.955 19.91%
180.972 15.27%
157.010 9.88%
304.912 16.59%
363.155 16.31%
280.938 6.32%
364.151 5.46%
4.997 .59%
512.851 2.51%
576.865 1.41%
638.821 1.04%
706.753 1.02%
770.746 0.58%
m/z (Da)

Figure 4-7. Deeper insight into site of biotinylation using gel electrophoresis and ESI-MS. (a) Effect of 0.1 M NaOH and heating upon the biotinylated GQ, CatG4. Shown is a 7.5% native (non-denaturing) gel, whose first two lanes on the left show, respectively, biotinylated CatG4 (generated by the standard heme/H_2O_2/biotin-tyramide reaction) without and with streptavidin added prior to loading in the gel. From the second lane, it can be seen that 58% of the DNA has been biotinylated (retarded by streptavidin binding). Treatment with 0.1 M NaOH at 90°C (lane at far right), however, shows that such a treatment leaves only 8% of the DNA biotinylated. A portion of the DNA (9%) is also shown to be broken down in size by this treatment. **(b):** A reaction scheme for an LC/MS-based scheme for looking at whether biotin tyramide (or a related product) is in fact released from biotinylated DNA (CatG4) and RNA (NRAS) GQs following the hot base treatment. ***(c-f)*** LC- ESI Mass spectrometry analysis of solutions (c) "NRAS RNA biotinylation"; (e) "NRAS RNA no biotinylation"; (d) "CatG4 DNA biotinylation"; (f); "CatG4 DNA no biotinylation". c and e show both biotin-tyramide mass and abundances, as well as the mass and abundances of rGMP and rAMP generated from alkaline hydrolysis of the NRAS RNA (rCMP and rUMP are very small peaks). d and f show biotin-tyramide mass and abundances.

The above data suggested that the phosphate groups of the GQ were important (though not necessarily exclusive) sites for biotinylation—DNA phosphotriesters are known to be notably base-labile[229]. To test whether BT or a derivative was indeed being released from base-treated GQs (both RNA and DNA), a mass-spectrometry based quantitation was carried out (experimental scheme shown in **Figure 4-7.b)**. Precisely equal amounts of a biotinylated and, separately, unbiotinylated RNA GQ (NRAS) and also biotinylated and unbiotinylated DNA GQ (CatG4) were purified first, in parallel, by two successive ethanol precipitations each, then hot base-treated, followed by analysis and quantitation of their contents by LC-ESI-MS.

Figure 4-7.c-f, shows, first, that in the hot-based treated RNA solutions, the two most abundant base-hydrolyzed NMP products, GMP ('rG') and AMP ('rA') were detected in equivalent quantities in the biotinylated and unbiotinylated samples, thus serving as internal standards for quantitation of any released biotin-tyramide in these same solutions. Indeed, an at least 4-fold excess of BT was measured in the base-treated biotinylated RNA solution (relative to the unbiotinylated RNA, **Figure 4-7**.**c & e**). The DNA analysis (**Figure 4-7.d & f**) showed an absence of hydrolyzed dNMP products, and found an at least two-fold excess of BT in the hot-base treated biotinylated sample, relative to the unbiotinylated sample.

4.3.7. Do GQ-binding ligands compete with heme with respect to GQ self-biotinylation?

As discussed above, a number of tight-binding, synthetic GQ ligands have been reported in the literature. These include N-methylmesoporphyrin IX (NMM; 'L1')[53], pyridostatin ('L2')[51]; and BRACO19 ('L3')[52], all of which bind to GQs by end-stacking upon terminal G-quartets, much as hemin does[55] **Figure 4-8.a**. It might therefore be expected that addition of excess GQ-ligands, that compete for binding with hemin, should reduce GQ–hemin concentration and hence, the overall levels of GQ self-biotinylation. **Figure 4-8.b** shows that this is indeed the case. Individual competition with four-fold excesses (over hemin) of the three ligands leads to significant decreases in the overall percentage of StAv-shifted gel bands (67% overall biotinylation seen in the absence of any competing ligand; 22–23% in the case of competition with NMM or BRACO19, and only 1% in the presence of pyridostatin). To get more precise and quantitative data on these inhibitions of hemin-mediated biotinylation by GQ binders,

levels of ‘CatG4-ext’ biotinylation were tested in the presence of 5 μM hemin and varying concentrations of the three GQ ligands. These data are shown in **Figure 4-8.c-f**. It can be seen that ~100 μM concentrations of either NMM or BRACO19 are required to fully abolish ‘CatG4-ext’ biotinylation; whereas, only ~10 μM Pyridostatin is required to do the same. The reported binding affinities of the three ligands to GQs are roughly in the same range as that of hemin[51–53,55]. The data shown here, however, suggest a superior binding affinity of Pyridostatin, relative to the other two ligands, to the specific GQ formed by ‘CatG4-ext’.

Importantly, the above data confirm, again, that it is GQ-bound hemin (as opposed to free-floating hemin) that is responsible for the observed GQ self-biotinylation. This result also builds confidence that when GQ self-biotinylation is sought for *in vivo* using this method, the identification of purely GQ DNA/RNA-generated self-biotinylation (as opposed to adventitious biotinylation caused by non-GQ peroxidases or other oxidases) can be confirmed by checking the competitive impact of GQ-ligands on intracellular DNA/RNA biotinylation. This is because these three chemically diverse GQ ligands are uniquely directed to binding GQs and not the active sites of proteinaceous oxidases within living cells.

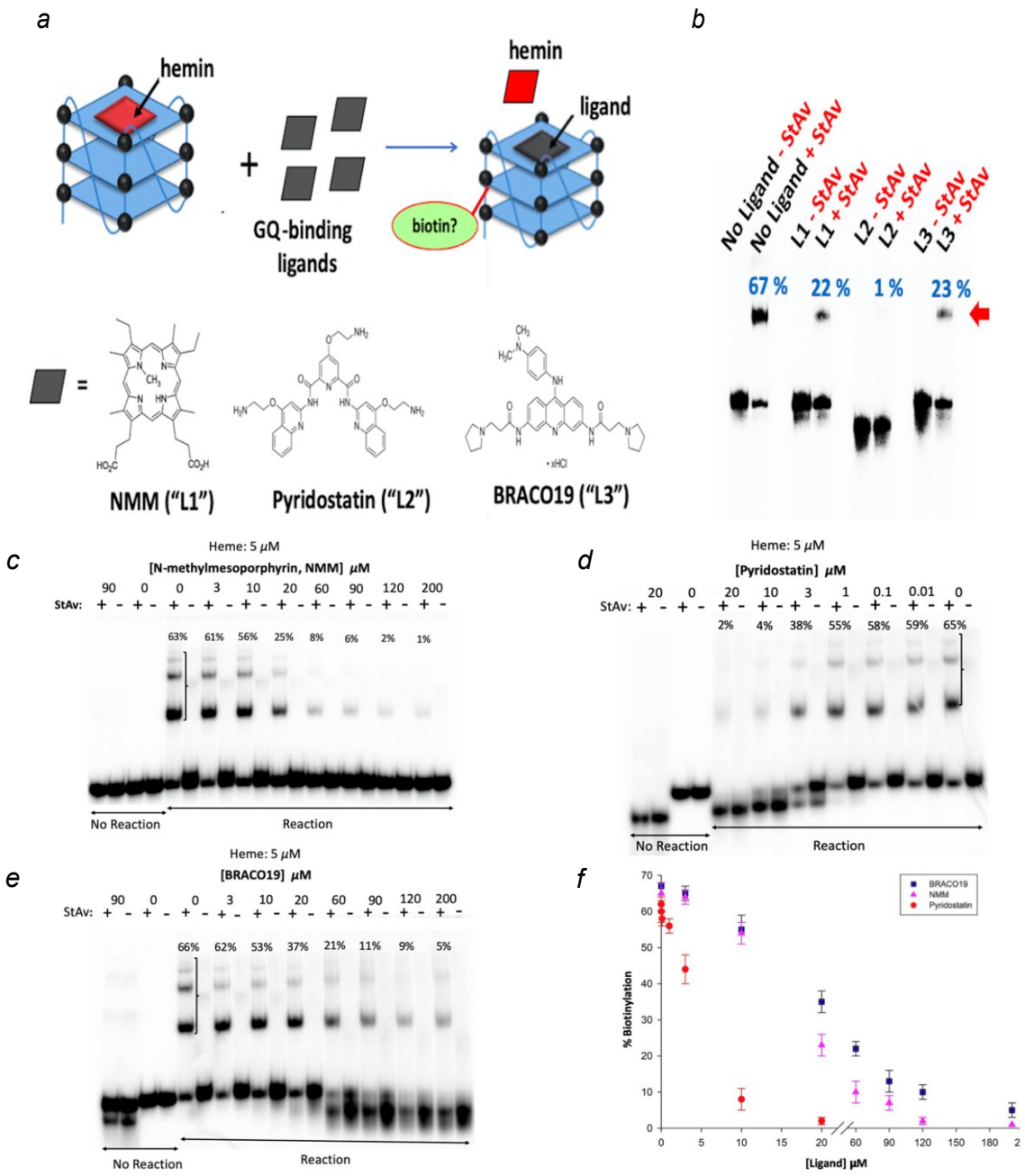


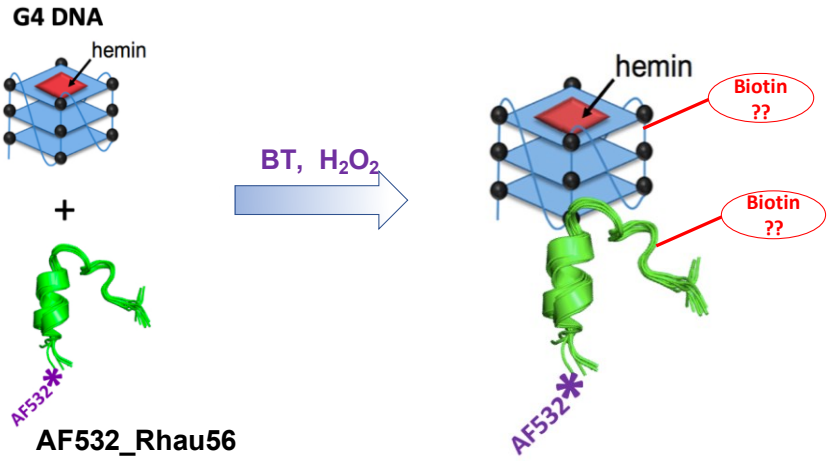
Figure 4-8. Competitive Effect of GQ binding ligand. **(a)** Schematic and Structure of three different GQ binding ligands. **(b)** A native gel showing reduction in biotinylation at a fixed concentration of ligand (20 μM) to CatG4-ext (1 μM). **(c-f)** A better quantitation of the competing effect of three different GQ ligand by titrating different concentration of the ligands.

4.3.8. Biotinylation of GQ binding protein/peptides

DNA/RNA-Protein interactions play a crucial role for a lot of biological processes in every living organism. Existing techniques for identifying GQs, with the use of antibodies or small molecules, fail to identify the GQs-interacting proteins. This current methodology using the Heme-GQ system provides, in principle, the added advantage of tagging (with biotin) proteins interacting with the GQs (given the radius of biotinylation from GQ ~10 nm as found in Section 4.3.5). I was interested in exploring two questions: a.) Can the GQs retain their own sites of biotinylation after binding to a peptide? b.) Can a GQ-binding peptide be biotinylated (Summarized in **Figure 4-9**)? A synthetic peptide of 56 amino acid (a part of RHAU helicase, "Rhau56"), known to bind parallel GQs at its terminating quartet[230], was taken as a peptide candidate to achieve the above said objectives. Rhau56 has an L-shaped structure containing an alpha-helix spanning from Gly6 to Ala18 (sequence shown in **Figure 4-9**) with a hydrophobic core comprising Leu8, Ile13 and Trp16. The peptide makes stacking interaction with the terminal quartets of a parallel GQ.

A titration was done to find an optimal concentration of GQ (DNA), Rhau56 (Peptide) and Heme, where both Heme and peptide could bind to the GQ concomitantly. GQ-Peptide mixture was incubated with 0, 5 or 50 μM heme and loaded on a native gel as shown in **Figure 4-10.a**. It can be seen that addition of 5 μM heme doesn't perturb GQ-peptide complex ("Bound DNA"), however, addition of 50 μM heme displaces a significant proportion of peptide from the GQ-peptide complex. Further, the bands labelled 1-7 in **Figure 4-10.a** were excised from the native gel and the excised gel pieces 1-6 were soaked in a solution containing BT and H_2O_2 but no heme for 2-3 min. Gel piece from band 7 was first soaked in a solution containing pre-folded (5'-P^{32} labelled) GQ for 20 min. following which it was soaked with a solution BT and H_2O_2 but no heme for 2-3 min. Following this, the reaction was quenched using DTT. DNA from the gel pieces were then recovered and mixed with Streptavidin before loading on an another native gel (**Figure 4-10.b**). It can be seen that DNA from gel bands 3, 5 and 6 does show biotinylation whereas bands 1,2, 4 and 7 doesn't. From this, two conclusions can be safely drawn: first, any Heme that is needed to carry out the reaction is coming from the gel pieces itself and second, the heme coming from the gel pieces are actually bound (co-migrating) with the peptide-GQ complex and are not there merely as a result of an undefined mobility of "heme smear" throughout the lane. This is evident more from

the gel Band 7, which when soaked with P^{32}-labelled pre-folded-GQ, BT and H_2O_2, doesn't show any biotinylation. Given that the reactions were carried out on the gel pieces itself (i.e., without pre-eluting the DNA out from the gel), an assumption can be made that the GQ-peptide complex didn't have too much of a freedom to dissociate and thus it is reasonable to say that GQ did retain its sites of biotinylation even when it was bound with a peptide.



RHAU56: CSM HPG HLK GRE IGM WAY KKQ GQK NKE AER QER AVV HMD ERR EEQ IVQ LLN SVQ AK

Figure 4-9. Investigation of G4-binding peptide biotinylation. Rhau56 binds to parallel GQ by making stacking interaction with the terminal quartets.

Second, to see if the peptide (Rhau56) itself was tagged with a biotin or not, a fluorescent version of Rhau56 (attached with Alexa-Fluor AF532 dye at N-terminal) was incubated with ^{32}P-labelled GQ and was supplied with 5 μM Heme, 500 μM BT and 1 mM H_2O_2 to initiate the reaction. The reaction was quenched at various time points (1 min, 5 min and 30 min) using DTT. Parallelly, peptide was incubated independently (without the GQ) and the reaction was initiated by adding 5 μM Heme, 500 μM BT and 1 mM H_2O_2 for the same time point as above. Post reaction, GQ and Rhau56 were separated by running them on SDS gel as shown in **Figure 4-11 (left)**. Rhau56 was then transferred on a Nitrocellulose membrane using a "Western Blot" transfer apparatus. Biotinylation on the peptide was verified using Streptavidin-HRP and development of chemiluminescence using ECL ('luminol') substrate. **Figure 4-11** (right) shows the result of western blot. Signal shown in blue comes from the dye "AF532" covalently attached to the peptide and signal shown in green is the chemiluminescence

showing the presence of biotin. Within the reaction time of 1-5 min, incubating peptide in the presence of GQ (as against peptide incubation without the GQ) leads to the peptide biotinylation as evident from the overlapping blue and green signal in lane "G4+P 1" and "G4+P 5" as compared to "P 1" and "P 5" respectively (**Figure 4-11**, **right**). However, once the reaction is allowed to proceed for longer duration (~30 min), peptide Rhau56 starts showing biotinylation even in the absence of GQ (Comparing lane "G4+P 30" Vs "P 30", **Figure 4-11**, **right**). Strikingly, it was found that multiple freezing and thawing of peptide makes it more efficient in its degree of self-biotinylation in absence of GQ whereas fresh stock solution of the peptide shows biotinylation only in the presence of GQ within short range of reaction time (<~5 min, **Figure 4-12**). We hypothesised that this peptides (Rhau56) upon longer incubation (or multiple freeze-thaw) may lead to the formation of alternate secondary structures which could bind and activate heme. Nonetheless, under optimized conditions and reaction time, this technique may be utilized to probe GQ-interacting proteomics profile in living cells as bolstered by the *in vitro* data. A similar observation was made very recently by Masuzawa T *et al*[231].

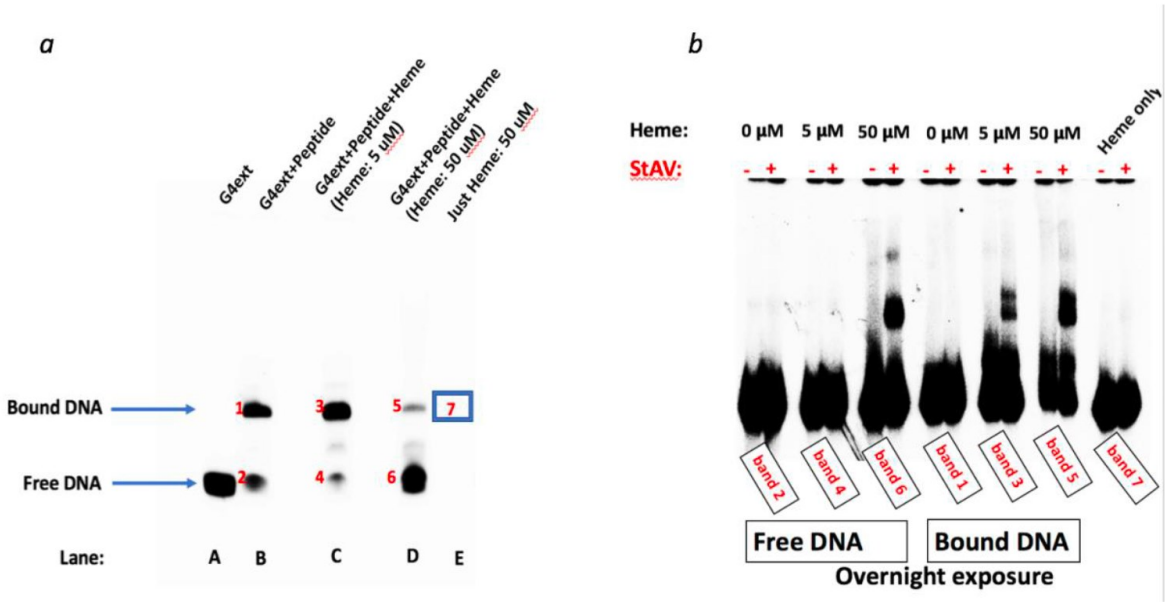


Figure 4-10. GQs retain self-biotinylation site upon binding to a peptide. (a) Native gel showing mobility shift of peptide bound G4 with or without different heme concentrations. **(b)** Native gel showing biotinylation of gel band species 1-7 (shown on gel **a**).

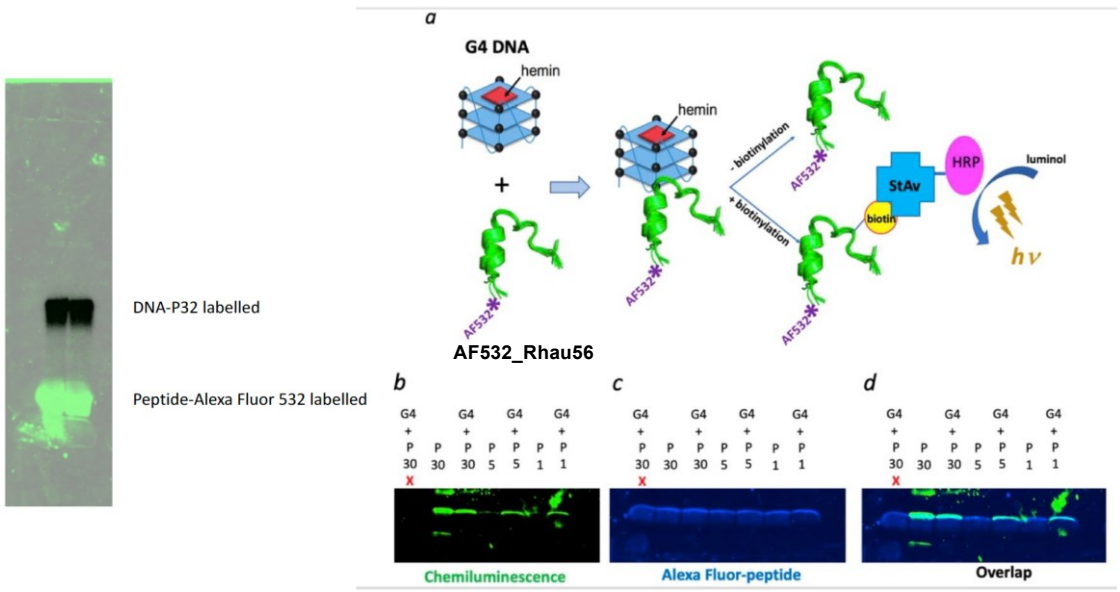


Figure 4-11. Biotinylation of GQ binding peptide. (Left) SDS gel showing separation of peptide and GQ post reaction. **(Right) (a)**Schematic of western blot assay to verify biotinylation of peptide; western blot assay showing signal from chemiluminescence (**panel b: (green)** presence of biotin); signal from Alexa-Fluor532 (**panel c: (blue)** presence of peptide); **panel d**: overlapping signal from both b and c.

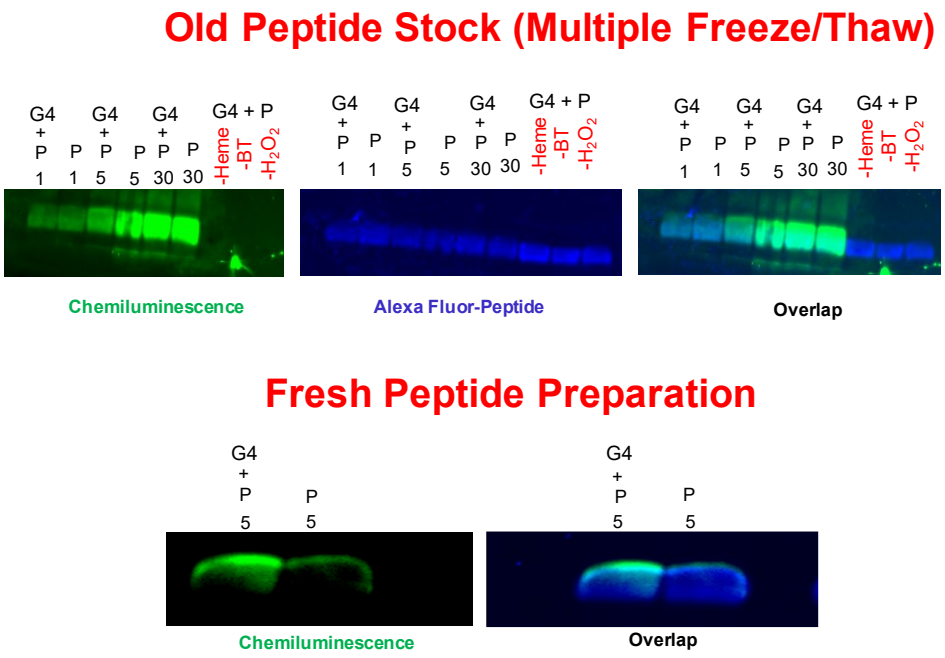


Figure 4-12. Peptide self-biotinylation triggered by multiple freeze-Thaw. (Top) old stocks of peptide which has undergone multiple freeze/thaw show more independent and less G4 dependent biotinylation at all time points of reaction. **(Bottom)** Fresh peptide solution show more GQ dependent biotinylation upto 5 min of reaction.

4.3.9. Evidence for GQ-mediated self-biotinylation for tagging RNA and DNA in living tissue

Clearly, GQ self-biotinylation can work efficiently *in vitro*. However, its application *in vivo* could, in principle, encounter unanticipated challenges including, potentially, a too-low oxidative activity of GQ–hemin in the intracellular milieu. I elected to try a few key experiments to test GQ–hemin peroxidase activity in living tissue. Initially, I chose to investigate salivary glands dissected from the larvae of *D. melanogaster* (**Figure 4-13.a**). The goal of these experiments was to test for detectable levels of biotinylation within total tissue RNA and genomic DNA extracted from the glands following treatment of the glands with hemin, BT and H_2O_2. It was also important to provide evidence that any observed biotinylation was not adventitious but a likely consequence of intracellular heme–GQ oxidative activity.

The feeding of low concentrations of heme to mammalian cells in tissue culture has been successfully exploited by Ting and co-workers to optimize the activity of heme-based recombinant peroxidases that they have expressed in such cells[232,233] . I used the concentrations of reagents used by those studies as starting points but adjusted them to suit my experimental system. To assay for covalent biotinylation within total RNA as well as genomic DNA extracted from treated whole salivary glands, a 'biotin blot' protocol was devised (**Figure 4-13.b**), whereby rigorously purified RNA, and separately, DNA from hemin, BT and H_2O_2-incubated salivary glands (as well as from negative control experiments in which incubation with one key reagent at a time, hemin, BT or H_2O_2, was omitted) were immobilized via UV crosslinking to a nylon membrane, and then probed for biotinylation status by the specific binding of commercially available StAv-HRP conjugates and subsequent oxidized luminol chemiluminescence (**Figure 4-13.c**). As a positive control for this protocol, I used pure 3′-biotinyl CatG4 (**Figure 4-1**; vide *infra*) in two different amounts (2 and 0.2 ng), which were spotted directly onto the membrane and crosslinked.

Salivary glands were dissected from live third instar *Drosophila* larvae using standard methods into phosphate-buffered saline (PBS) buffer, and were incubated with 5–50 μM heme and 3 mM BT. They were then briefly (2–3 min) pulsed with 10 mM H_2O_2. Total RNA and, separately, genomic DNA, were purified using commercial purification kits and 400 ng (for RNA) and 200 ng (for DNA) were spotted onto a nylon

membrane using a dot blot device, followed by UV-crosslinking. How do we know that the commercial kits used in this study indeed purified genomic DNA and RNA, respectively? And, that these preparations were not contaminated by proteins from the live tissue? Three kinds of evidence are shown in **Figure 4-14**. First, **Figure 4-14.a**, shows a 1% agarose gel running the purified total RNA aliquot and a purified genomic DNA aliquot. It can be seen that the two aliquots are indeed different from each other. **Figure 4-14.b** shows the UV-vis spectra of the purified genomic DNA and total RNA, respectively. The spectra report A_{260}/A_{280} values > 1.8, consonant with nucleic acid preparations uncontaminated by proteins. **Figure 4-14.c & d** shows that the chemiluminescence 'output' signals from both 3′-biotinyl 'CatG4' (a positive control) and genomic DNA biotinylated *in vivo* were comparably abolished by digestion of these DNAs with active DNAase I, and were correspondingly *not* abolished by digestion of these DNAs with inactivated DNAase I.

Following the spotting of DNA/RNA upon the Hybond Nylon+ membranes, above, the membranes were blocked, washed and treated with HRP-StAv; and, finally, after extensive washing, they were developed using bound HRP-mediated luminol oxidation and the recording of chemiluminescence.

Figure 4-13.c & d show the results. **Figure 4-13.c** shows that both total salivary gland RNA and genomic DNA show evidence of strong biotinylation following treatment with hemin and BT, followed by a pulse of H_2O_2, whereas omission of any one of these key reagents leads to a full loss of signal. At this point, however, these results do not guarantee that the observed biotinylation of salivary gland RNA and DNA results from oxidation of BT by intracellular GQs complexed with hemin; some other kind of adventitious oxidation of BT and concomitant DNA/RNA biotinylation could, in principle, be involved. To eliminate such a possibility, I investigated biotinylation of whole intracellular RNA in the presence not only of heme but of either one of the two strong GQ-ligands, Pyridostatin and BRACO19.

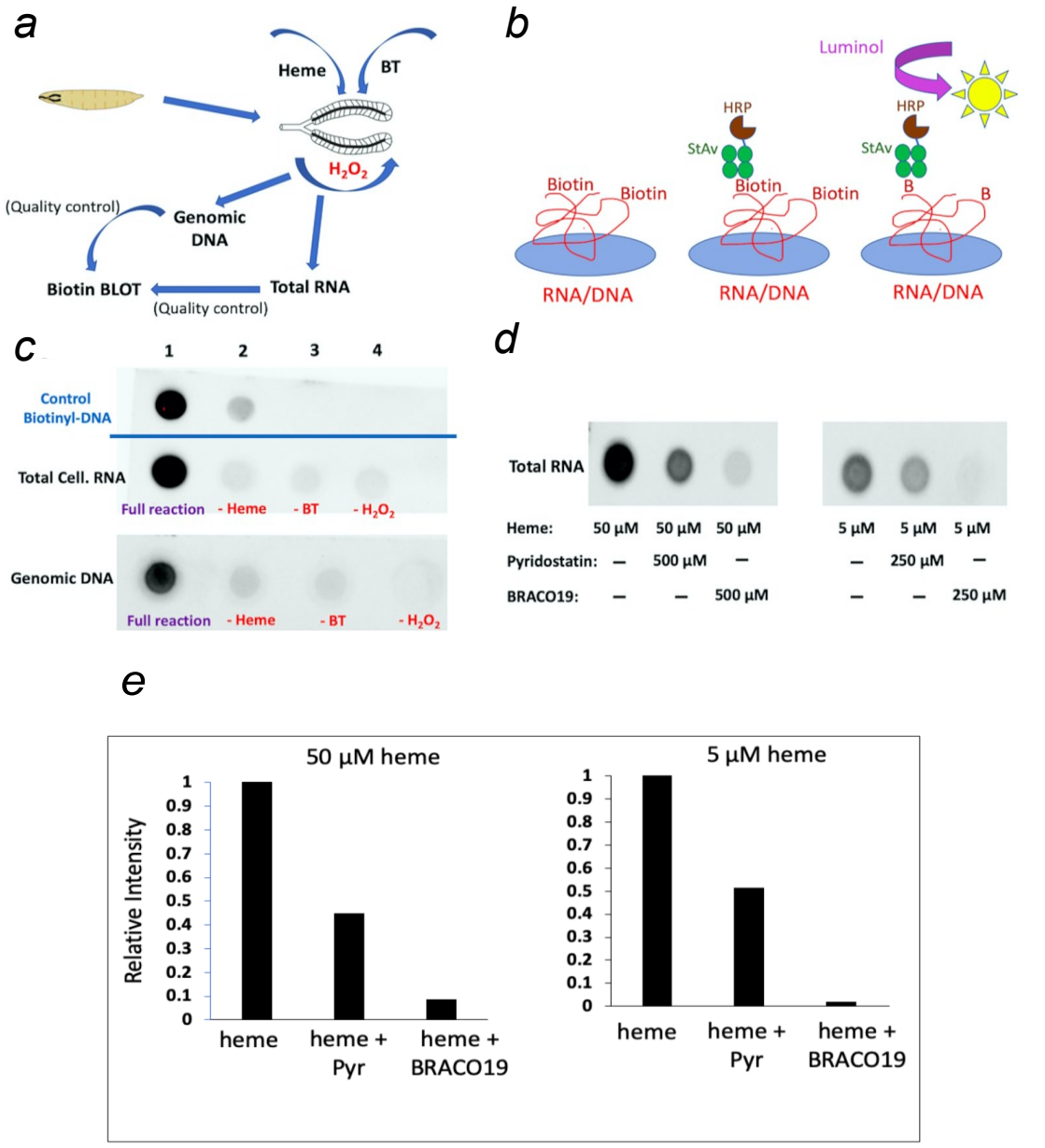


Figure 4-13. Biotinylation of RNA and DNA within live Drosophila Salivary glands. (**a**) A schematic diagram showing the design of the experiment. (**b**) Design of biotin dot-blots of total cellular RNA and genomic DNA. (**c**) Developed biotin dot blots from 2–5 µl (DNA = 200 ng; RNA = 400 ng for all dot blots) of total gland RNAs and genomic DNA extracted from living Drosophila larvae that have been treated with 50 µM heme, 3 mM BT and pulsed briefly (2–3 min) with 10 mM H_2O_2 followed by quenching. Blots '1' show chemiluminescence from the full reaction, where all the above reagents are present; '2', '3' and '4' show negative controls, where one of the participating reagents at a time is left out. The positive control spots, 'Control Biotinyl-DNA' show a standard 3′-biotinylated DNA spotted at two different concentrations: 2 ng (Spot 1) and 0.2 ng (Spot 2). (**d**) Developed biotin dot blots from 2–5 µl (DNA = 200 ng; RNA = 400 ng for all dot blots) of total cellular RNAs isolated from live salivary glands that have been treated with 50 µM heme plus 500 µM GQ-ligand (or 5 µM heme plus 250 µM GQ-ligand) and 3 mM BT, all pulsed briefly (2–3 min) with 10 mM H_2O_2 followed by quenching. **(e)** Quantitation of the chemiluminescence output data from the biotinylated total RNA blots shown in (d).

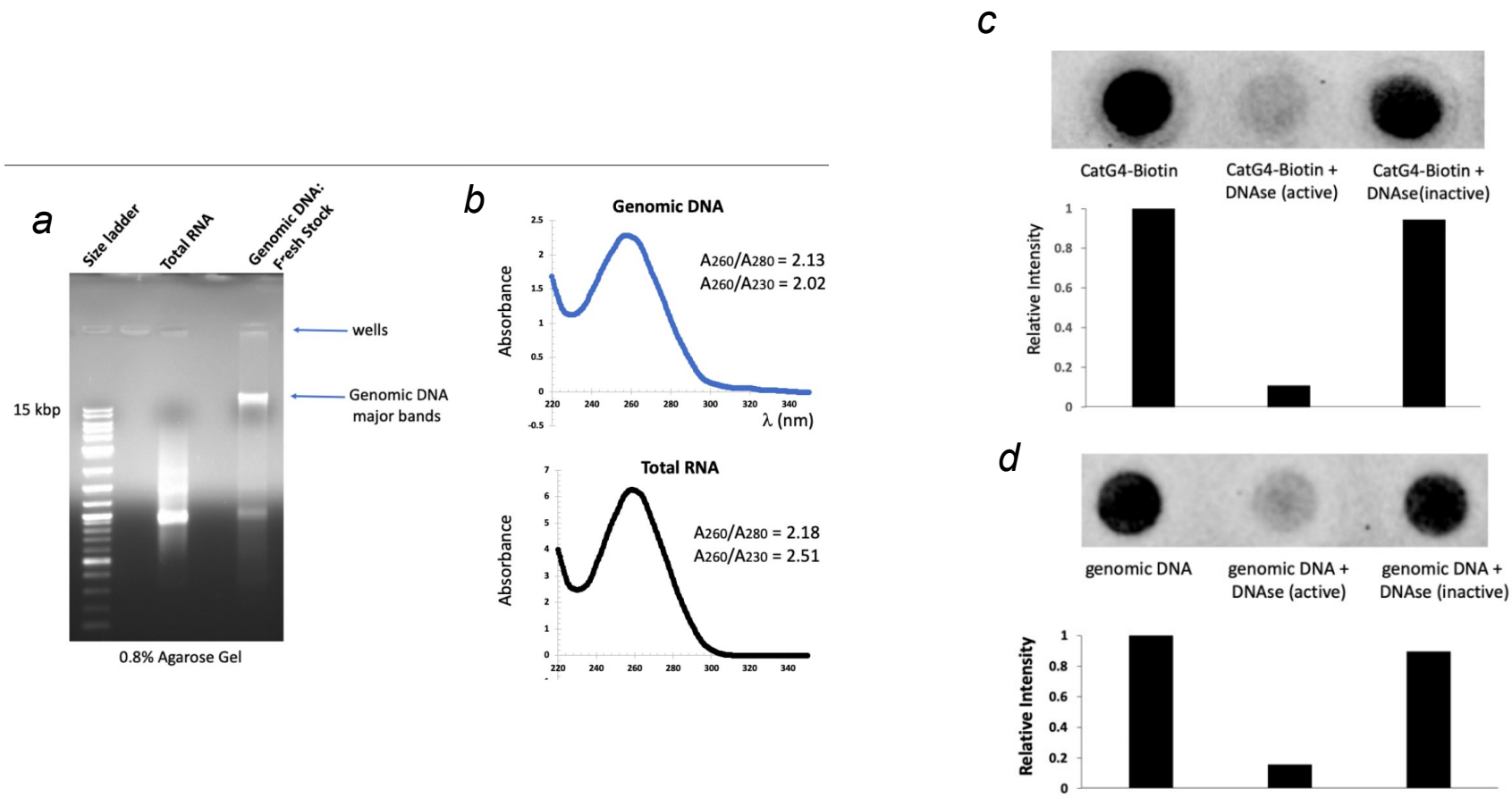


Figure 4-14. Examination of the purity of biotinylated DNA and RNA extracted using commercial kits from treated Drosophila salivary glands. **(a)**: a 0.8% agarose gel showing (left) total RNA and (right) genomic DNA purified from the glands. **(b)**: UV-Vis spectra of the purified total RNA and genomic DNA. The A260/A280 ratios, which are > 1.8, indicate high purities (free of proteins) for both DNA and RNA preparations. **(c-d)** Evidence that the biotin detected, using streptavidin-HRP-generated chemiluminescence output signal, is from biotinylated DNA (and not biotinylated proteins) extracted from treated Drosophila embryonic salivary glands. **c:** A chemically biotinylated oligonucleotide (3'-biotinyl CatG4) was used as a positive control, spotted by itself (left); treated with DNAase I (centre); and with heat-deactivated DNAase I (right). **d**: the same analysis as in panel c, except on total genomic DNA extracted from Drosophila 3rd instar larval salivary glands.

Figure 4-13.**d** shows that co-treatment (along with hemin) of either of these compounds leads to significant (with Pyridostatin) or almost complete (with BRACO19) loss of biotinylation signal. These results broadly mirror the data obtained in vitro (shown in **Figure 4-8.b**), though there are differences. **Figure 4-13.e** shows quantitation of the in vivo biotinylation suppression data. It is interesting that BRACO19 displays somewhat superior activity to Pyridostatin *in vivo* (*in vivo*, an ensemble of different GQs, each with its own binding preference to these ligands, is likely encountered); whereas, the opposite is true in vitro, where a specific GQ has been used (**Figure 4-8**). Therefore, the difference in our in vitro and in vivo results may reflect the use of a single, defined GQ in vitro, whereas the in vivo data represent the cumulative activity of many different GQs found within living Drosophila cells.

As mentioned above, these non-heme GQ ligands (structurally distinct from hemin as well as from each other) are expected *uniquely* to compete with heme for binding GQs, and not to have other inhibitory interactions with proteinaceous oxidases or peroxidases within living cells (even if one of these mutually structurally distinct ligands did, it is very unlikely that *both* would have the same effect). Furthermore, it is highly unlikely that these ligands are inhibiting the BT oxidation by comprehensively competing (at 250 μM concentration) with the outstanding phenolic substrate BT (at 3 mM) as *substrates* for oxidation by hemin–GQ or, indeed, other oxidases in the cell.

The experiments shown in **Figure 4-13** report semi-quantitative in vivo data; they do not yet supply direct sequence information on the Drosophila RNAs and DNA sequences biotinylated by the procedure. Nevertheless, the data presented in **Figure 4-13** provide compelling evidence that under the right experimental conditions, GQ–hemin-mediated nucleic acid self-biotinylation does work *in vivo*. As described earlier, there is now evidence of the natural sequestration of intracellular hemin by cellular RNA/DNA GQs[218]; the biotin blots reported here are not sensitive enough to detect that.

4.3.10. RNA Sequencing post biotinylation in *Drosophila Melanogaster*

To obtain the actual sequencing data, total RNA was extracted post-biotinylation from the 3rd stage larvae of *Drosophila Melanogaster*. Extracted RNA was mixed with streptavidin-conjugated magnetic beads to separate and enrich the biotinylated RNA

away from the non-biotinylated fraction (see section 4.2.11, for detailed protocol). Reverse Transcription, using a random hexamer primer, was done directly on the RNA bound to the beads to enable formation of cDNA. The gel in **Figure 4-15** shows that treatment of the dissected salivary gland (in PBS buffer and at room temperature) with heme, BT and H_2O_2 led to a significant enrichment of biotinylated RNA, as evident from the successful formation of cDNA in the lane: "+Heme Rxn", (**Figure 4-15**") as contrasted with a no reaction ("Untreated") negative control (lane: "-Rxn", **Figure 4-15**). I wanted to see if the intrinsic cellular heme concentration was sufficient to enable the biotinylation reaction inside cells. For this, salivary glands were treated only with BT and H_2O_2 but no extrinsically added heme ("-Heme Rxn"). Lane: "-Heme Rxn", at a relatively high gel exposure, does show a very faint level of cDNA formation.

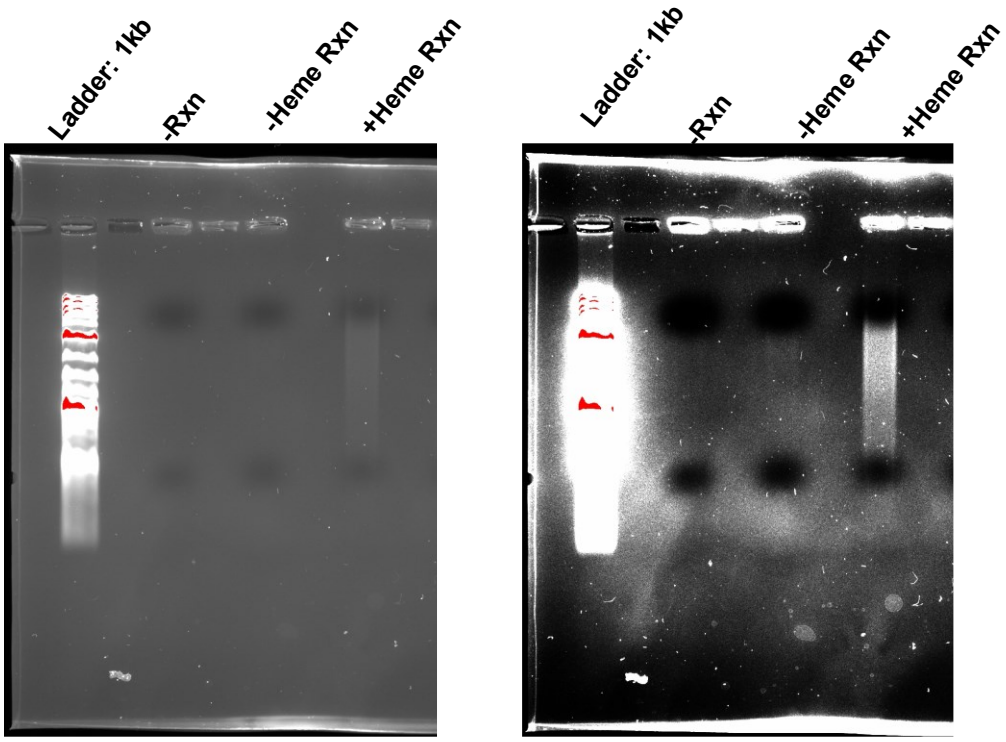


Figure 4-15. Reverse Transcription Directly on Streptavidin-coupled Magnetic beads. Two different exposure of a 0.8 % Agarose gel. "**–Rxn"** refers to salivary glands that were not treated with any of BT, Heme or H_2O_2. "**–Heme Rxn"** is one where BT, H_2O_2 were supplied but no extrinsic heme. "**+Heme Rxn"** is one where all three extrinsic Heme, BT and H_2O_2 were provided. Gels were stained with Sybr Gold.

After this initial verification of the cDNA formation, biotinylated RNA bound to the streptavidin-conjugated magnetic beads (prior to reverse transcription) was eluted and purified.

Notably, in the case of “-Heme Rxn” (Treatment with only BT & H_2O_2) and “-Rxn” (untreated), the eluted RNA concentration was below the limits of detection when quantified using a Qubit fluorometer; hence, we proceeded to use only the RNA eluted from the “+Heme Rxn” sample for the initial round of library preparation and sequencing.

Here, we report the preliminary sequencing data. The aim of this experiment was not to do the differential transcript enrichment analysis between the reaction (“+Heme Rxn” or “-Heme Rxn”) and the no reaction (“-Rxn”) control. Rather, we wished to make a qualitative measurement of the overall representation of the biotinylated transcripts that were showing up (as a result of streptavidin pull down) in the case of “+Heme Rxn” sample alone. Also, for the initial round of sequencing, total RNA eluted from streptavidin beads (following multiple rounds of wash, see section 4.2.11) were used for library preparation without any explicit rRNA depletion or Poly A enrichment. Since we were enriching the biotinylated transcript using streptavidin conjugated beads, we assumed that most of the non-biotinylated rRNA (or non-biotinylated non-rRNA) contaminants would get naturally depleted as they will not be captured by streptavidin beads.

Sequencing data of the “+Heme Rxn” sample alone revealed some interesting observations. Strikingly, we observed a very high abundance of **rRNA** in our sequencing result (~90 %) (See Appendix: “Drosophila_G4_RNAseq.xlsx”, Sheet name: “Read_distribution”). Two possible explanations can be provided for such a high bias of rRNA in our library. First, it may be a completely non-specific contamination, given the high overall abundance of rRNA inside the cells. A second and more intriguing possible explanation is that at least a proportion of these rRNA are the real biotinylated transcripts, highlighting potential GQ forming sites within them. Indeed, recently, there have been significant reports for the potential formation of GQs in rRNA. Particularly, studies from the Loren Dean Williams lab (Georgia Institute of Technology)[234,235] have shown evidences for (a) the occurrence of GQs,in expanded segments (ES) of human rRNA, i.e. those segments of rRNA on the “periphery” of the assembled ribosome – hence, not visible in the published crystal structures; and (b) sequestration of heme by

these rRNA loci both in *vitro* and *in vivo* . Future experiments with appropriate negative controls - (where the larvae will be treated with only one of the three components (heme or BT or H_2O_2) at a time or in a combinations of two i.e. (Heme+H_2O_2) or (Heme+BT)) - will be used to validate if there is a true biotinylation-based enrichment of the rRNA loci with a potential to fold in GQ. These experiments will help in ruling out one of the two hypotheses stated above i.e., whether these rRNAs enrich owing to *bona fide* GQ formation within them or they are enriched just as part of the "noise" of purification.

Next, we looked at the distribution and enrichment of **non-rRNA** transcripts. Although, the high abundance of rRNA made our library difficult to analyze for non-rRNA transcript enrichment, we were nevertheless able to gather the raw counts of the non-rRNA transcripts that showed up in our sequencing data. Excel file "Drosophila_G4_RNAseq.xlsx" (Sheet name: "non-rRNA (Heme-GQ)", See Appendix) summarizes the transcript distribution with their respective raw counts. To begin with, we compared our results with an existing dataset containing the reported sites of GQ formation in *Drosophila* (at the level of genomic DNA) using an *in-vitro* "*polymerase stop essay"* devised by Marsico *et al*[67]. Notably, out of the 8291 hits/transcripts that we obtained from our sequencing data, 4375 were in common with those reported by "polymerase stop assay" (See Appendix: "Drosophila_G4_RNAseq.xlsx", Sheet name: "common with stop pol"). Further, this overlapping list of genes between the two datasets contained some of the widely known quadruplex forming genes (as reported in humans) such as *MYC*[236], *FMR1*[85], *NOP56*[86], *VEGF*[81], *RET*[237], etc.

Although not conclusive, these preliminary data are promising and consistent with GQ forming both in the rRNA and in non-rRNA transcripts. However, multiple replicates with proper negative controls will be needed to be analyzed with proper statistical validation before firm conclusions can be drawn on whether transcripts enriched by our method strongly correlate with *in vivo* GQ localization.

4.4. Discussion and Conclusion

Heme-G-Quadruplex intrinsic peroxidase property can be utilised in order to develop a highly sensitive approach to tag both DNA and RNA G-Quadruplexes with biotin (Summarized in **Figure 4-16**). I have shown that *in vitro*, GQ self-biotinylation is (i) overwhelmingly specific for GQ over duplex or single-strands, even under conditions of very high, gelated, concentrations of DNA; (ii) distributed relatively evenly over the length of a DNA that folds to a GQ, with two–four biotinylation events occurring per 21-nt 'CatG4' GQ-forming oligonucleotide under our reaction conditions; (iii) that the effective range of biotinylation is <10 nm from the GQ itself. (iv) I carried out a preliminary investigation of whether the biotinylation protocol could be applied to live tissue and found compelling evidence that intracellular nucleic acid biotinylation did occur, as indicated by a biotin-blot procedure. (v) I used GQ-specific ligands, that compete for GQ binding with hemin, to show that application of such ligands in competition with hemin lowered or abolished self-biotinylation levels, both *in vitro* and *in vivo*. (vi) Finally, I have presented early sequencing data, which gives promising insights for the existence of GQs in living tissues/cells and for heme-GQ interaction *in vivo*.

On the cumulative strength of these above data, I believe that the self-biotinylation procedure can indeed be productively used to tag, identify and pull down DNAs and RNAs folded into GQs within living cells. I anticipate that at the level of ChIP-Seq with subsequent multiple replications of Next-Gen RNA Sequencing, GQs that are naturally interacting with heme within the cell (such as demonstrated in ref. ([218])) will be identified with certainty without the provision/feeding of extraneous heme to the cells/tissues being examined. Labile heme concentration has been estimated to be present in nanomolar range within different cellular compartments in normal non-erythroid human cell lines[238]. Subcellular localization of labile heme is heterogenous with the highest concentration present in cytoplasm (~25-300 nM). In contrast, mitochondria and nucleus retains a low level of labile heme < 2.5 nM[238–241]. In cancer cells (e.g. prostate cancer), labile heme concentration is higher in nucleus than in cytoplasm[242]. Given that the current estimated quadruplex sites lie in nanomolar range (~700,000 putative quadruplex[66]) in the human genome with nanomolar to micromolar affinity towards heme[53,113–119], it is highly likely that GQs may help in sequestering the excess labile heme which otherwise may be toxic to the cells. However, the feeding of

extraneous heme will help to identify additional DNA and RNA sequences that show a *capability* to form GQs within the cell, perhaps in response to specific cellular conditions or environmental stimuli. Where fed extraneously, heme is expected to functionally resemble GQ-stabilizing ligands, such as NMM, BRACO19, pyridostatin and others, in terms of their strengths and limitations for use in probing for GQs within cells.

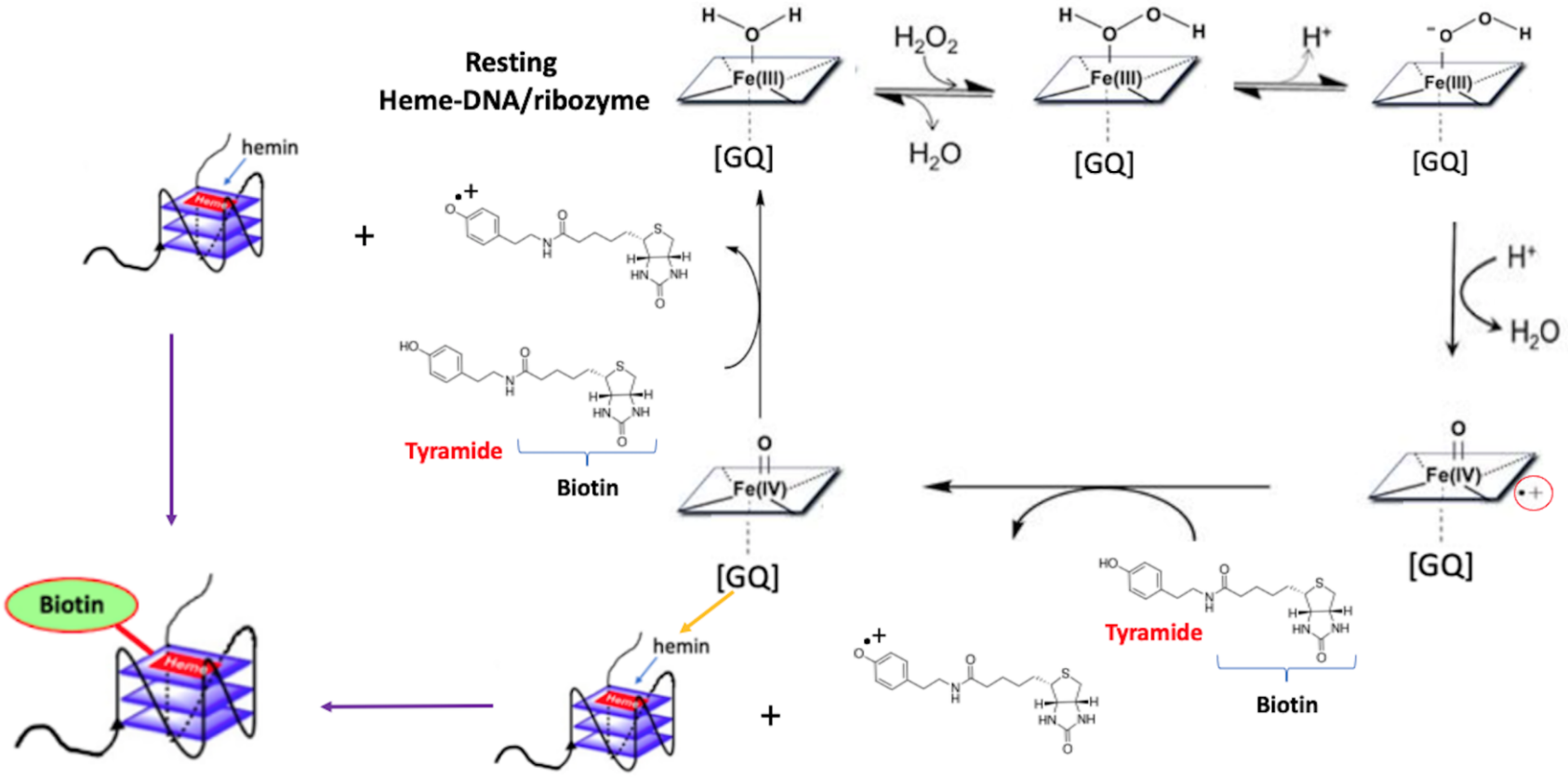


Figure 4-16. A proposed mechanism of Heme-Quadruplex peroxidase activity for self biotinylation. GQ's planar surface provides the electron-rich push effect required for the heterolytic cleavage of H_2O_2 and subsequent oxidation of Fe^{3+} to an oxidation equivalent of +5. Iron gets back to its ferric (Fe3+) state in two consecutive step by taking electrons from the biotin-tyramide substate which creates a highly reactive phenolic radical. This phenolic radical forms a covalent adduct with the GQs itself.

Chapter 5.

Future perspectives

This thesis highlights fundamental discoveries as well as potentially diverse practical applications of newly discovered and rationally designed DNA/RNA non-canonical secondary structures. The different project covered in this thesis opens the door for a series of interesting future investigations. Here, I summarize some interesting directions of work that can follow in future.

With regard to the **first** project**,** the discovery of iCD-DNA (Chapter 2), once again, highlights the robust polymorphism of DNA secondary/tertiary structures. iCD-DNA is clearly different from either a classical B type duplex or hairpin or any of the conventional right-handed G-Quadruplexes that have been reported in the literature this far. Any new secondary structure formed by DNA has significant intrinsic importance, even if it forms under conditions that deviate slightly from the "normal conditions" of physiology. A classic example of the above is the design and application of an i-motif (stable at pH 4.5- 5.2)-based DNA nanostructure, called the "I-switch" developed by Krishnan group[243]. I-switch is the first example of an i-motif-driven nanomachine capable of sensing and reporting pH changes as a function of endosomal maturation both inside living cells in culture[243] as well as within multicellular organisms[244]. Clearly, the full application of any novel structure requires a thorough understanding of its 3D folding. Although, I have proposed three plausible structural models for the iCD DNA, the exact crystal structure is yet unknown. Thus, **NMR and X-Ray crystallography experiments** can be carried out to solve the high-resolution structure of iCD-DNA. Knowing the detailed structure of iCD-DNA will both facilitate iCD-DNA based nanotechnological development as well as to test for its occurrence *in vivo*. Further, the availability of an atomic structure of iCD-DNA will contribute to deepening the theoretical modelling of CD spectra for different secondary folds of nucleic acids.

Given that the intracellular pH is maintained between 7.0 to 7.5[245] and that the iCD-DNA forms from a disease-significant repetitive DNA sequence, it will unquestionably be important to test if iCD-DNA formation can be triggered at neutral pH. Considering the role of protonated cytosines in all the three proposed models of iCD-

DNA (See Chapter 2), various structural and environmental factors known to influence the pK_a of cytosine, such as the presence of crowding agents[35], negative supercoiling[34] as well as covalent cytosine modifications (such as methylation or hydroxymethylation)[39,40] can be screened to examine if iCD-DNA can form or persist at neutral pH. Concomitantly, both small molecules and proteins can be screened for the property of binding and stabilizing iCD-DNA. This may indeed enable the persistence of iCD-DNA in the presence of physiological potassium concentrations (where, from our results to date, there is a strong tendency for iCD-DNA to convert to G-quadruplexes). All these will help to understand the exceptional situations in which an iCD-DNA like structure could exist under intracellular conditions.

With regard to my **second** project, demonstration of the rational design of triplex-quadruplex hybrid tiles (TQ) (Chapter 3) establishes a firm precedence for the structural robustness of a mixture of different DNA secondary structures to co-exist as one unit. Tile "TQ" represents a continuous structure, where smooth transition is facilitated from triplex (triple-helices) to quadruplex and then back to triplex, in a precise way. Notably, the "sticky-ended" version of "TQ" tile ("TQ*s*") facilitates a facile and reversible assembly of triplex-quadruplex hybrid wires of length comparable to or greater than the various reported forms of purely G-quadruplex-based G-wire[199]. With TQ tiles, the pH dependence and the presence of long stretches of triplexes in $(TQs)_n$ wire provide incremental advantages over a classic G-Wire, especially with regard to dye-binding based sensor development. Most of the GQ recognising dyes bind to GQs via end stacking to the terminal quartets[59,246]. Thus, planar terminal quartets in a G-Wire are an obligatory requirement (to a great extent) for its usability in terms of developing a fluorogenic methodology for diagnostic or sensor applications. By contrast, the presence of long triplex stretches in TQ*s* wire/1DDN offers sites for intercalating dyes that bind specifically to triplexes and thereby fluoresce. This property could be used in future to design highly sensitive pH responsive sensors. Another area of interest in the field of "DNA-wire" is the feasibility and the extent of charge transfer. Although initially perceived as promising, DNA duplexes turned out to be a poor substrate for electric charge transport over >100 nm length scale. G-wires, by contrast, have turned out to be very good candidates for charge transfer because of their high guanine content and high stiffness. Currents more than 100 pA have been measured over distance ranging to > 100 nm in G-wires[197]. The design of hybrid tiles like TQ*s* may prove useful in charge

conduction applications given the presence of two quartets in every repeating unit of the wire (one three-layered quartet for triplex phasing and another four-layered quartet at glue junction) and sufficient numbers of guanine content in the triplex part. Triplex stretch of the TQ*s* tile can further be modified with an increased guanine content for better charge transport. Assembly of $(TQs)_n$ wires of finite and pre-determined length will pose certain challenges, but may be optimized by carefully controlling the assembly of sticky ended TQs tiles with low molar ratios of wire-terminating tiles (i.e. TQ*s* tiles with sticky ends on only one side).

Finally, one of the most innovative lines of investigation to come out from the TQs design is the notion of a purely guanine-based sticky ends. Complementary Watson-crick base pairing is commonly used to bridge non-continuous and non-blunt ended DNA/RNA duplexes (via classic single-stranded "sticky ends"). Thus, complementary Watson-crick base pairing rules (A pairing with T/U and C pairing with G) define a conventional class of sticky ends **Figure 5-1**. Other than Watson-Crick base pairing, a second class of sticky ends involves ligation of the discontinuous strands through G-C-G-C quartet formation[107], held together by Hoogsteen hydrogen bonding. However, purely G-quartet based sticky ends have not been explored/investigated in detail. The formation of $(TQs)_n$ involves an entirely new category of "sticky ends" based purely on the formation of G-quartets. Future experiments will focus on establishing rules, in terms of orientations, directions, as well as the number of quartet layers that may be sufficient to bind and ligate together the blunt ends of duplexes (and triplexes) **(Figure 5-2)**. Defining such rules will help in designing efficient DNA couplers to reversibly associate/dissociate different TQ tiles of interest in a given mixture.

With regard to my **third** project, this thesis presents a very robust and minimally intrusive methodology for probing and tagging RNA/DNA GQs *in living cells* (Chapter 4). Using an indirect approach of dot blot assay and GQ ligand competition experiments, I have been able to show evidence for two things, first, sequestration of heme by G-quadruplexes within living cells and second, utilization of the intrinsic and well-studied peroxidase activity of heme-quadruplex complexes to tag genomic DNA and total RNA with biotin (**Figure 4-13**). The direct evidence for GQ biotinylation *in vivo*, however, will require deep sequencing of the biotinylated DNA/RNA strands. Initial RNA sequencing data enclosed in this thesis, although promising, are too preliminary for making any concrete conclusions. Thus, future experiments will focus on doing deep sequencing of

both biotinylated DNA and RNA under two conditions: first, following feeding of cells with extrinsic heme ('Extrinsic Heme Reaction') and second, relying upon the intrinsic cellular heme that may be bound to intracellular G-Quadruplexes ("No Extrinsic Heme Reaction"). Recently, studies from several other groups have bolstered the hypothesis that heme does indeed interact with and is activated by nucleic acids in living cells[218,234,247]. GQ biotinylation and pull down in the case of "No Extrinsic Heme Reaction" condition will be particularly interesting as this condition represents a scenario very close to the native and non-perturbed *in-cellular* milieu.

To date, we have used this method exclusively in *Drosophila* larvae. However, future experiments will focus on generalizing the applicability of this approach to different model organisms. This in turn will facilitate a direct comparison of GQ-profiles in the different model organisms. Another important aspect of this project will be to probe the GQ-dependent (based on proximity and/or binding) proteomics profile in different cells and cellular compartments. Our methodology is able to biotinylate DNA/proteins located within ~10 nm range of a GQ site *in vitro* (See Section 4.3.5 and Section 4.3.8). Given that interactions of proteins with DNA/RNA underlie almost many key biological processes[248,249] , it will be interesting to identify the proteins that recognise and bind to RNA and DNA G-quadruplex. This will help in understanding different regulatory mechanism at molecular/cellular level.

The above future experiments are expected to take this work to ambitious long-term goals *i.e.,* mapping the 3D organisation of genome in interphase nuclei as well as proteomics of different cellular compartments. Recently, three different approaches have been reported to do the proteomics mapping. The first one utilizes an engineered version of horseradish peroxidase called APEX2[233]. In the presence of a phenolic substrate like Biotin tyramide and H_2O_2, APEX2 creates a short-lived radicle and biotinylates nearby proteins. A second approach uses an another proximity based biotinylation technique call the "BioID"[250]. In BioID, a 'bait' polypeptide of interest is fused with a mutant *E.Coli* biotin ligase (BirA*) and the fusion protein is expressed in cultured cells. The abortive BirA* ligase releases biotinoyl AMP in the local environment and proteins residues within the 10 nm radius of the bait protein get covalently modified with biotin at lysine sites. The third and the most recent approach incorporates a CMPP (Co-binding mediated protein profiling) to investigate the interaction between GQs and associated proteins[251]. In this method, GQ stabilizing small ligand such as pyridostatin

(PDS) are covalently attached to photo-crosslinkers that facilitate the irreversible attachment to cellular protein targets in the vicinity of GQ. I anticipate that the future experiments targeted to do GQ and the associated proteomics mappings (using our "GQ-Heme" assay) will further complement these studies in making a more broad and holistic understanding of the internal cellular milieu.

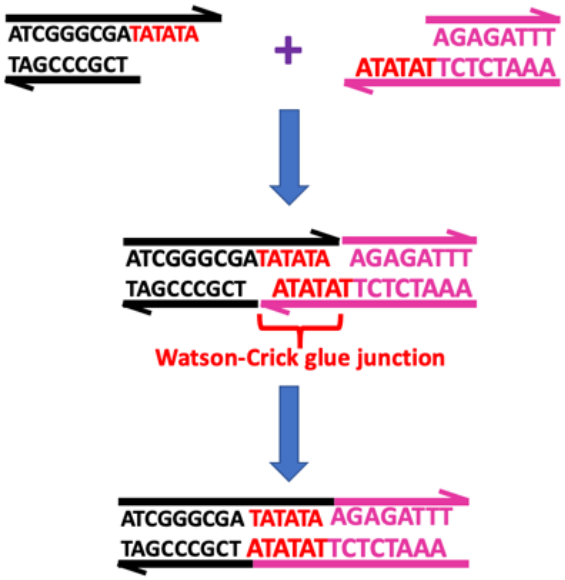


Figure 5-1. Watson Crick based Glue Junction (Sticky ends). The conventional rule of A pairing with T/U and G pairing up with C constitutes the canonical Watson-crick based sticky ends to ligate different fragments of DNA.

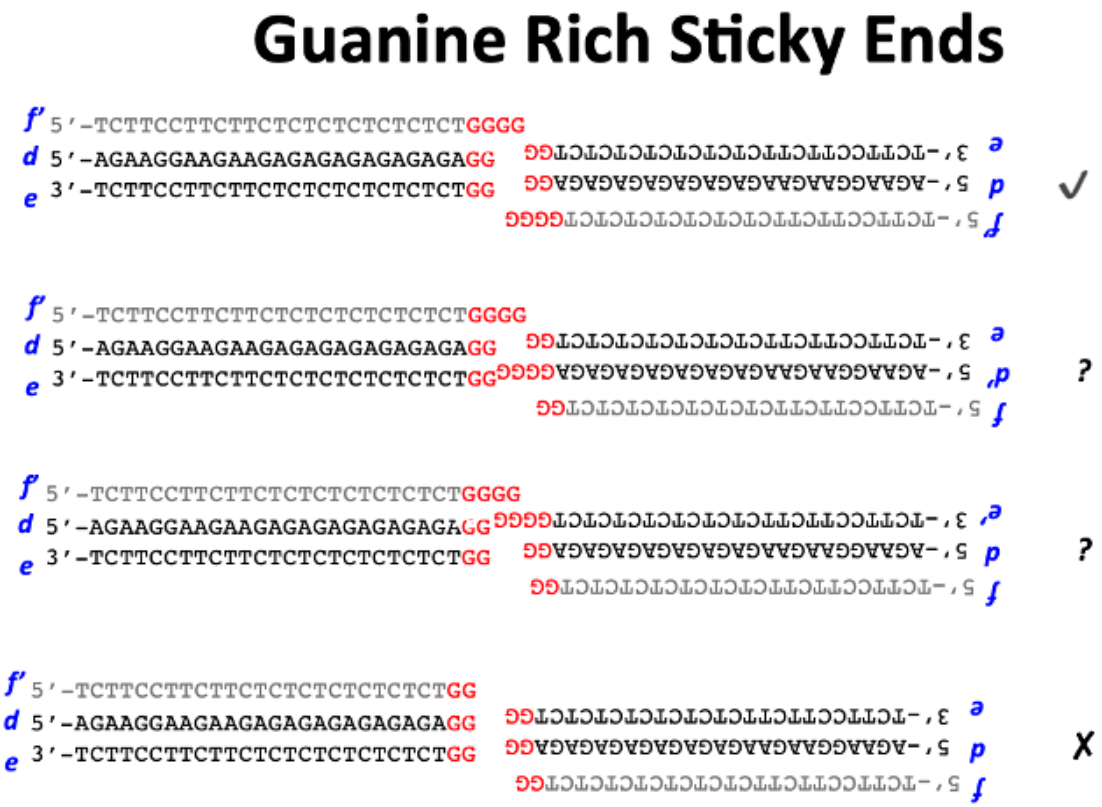


Figure 5-2. **Guanine Based Glue junction. (Top)** Glue junction pattern used in making $(TQs)_n$ with a continuous stretch of 'GGGG' in the third strand of the triplex. **(Bottom ones)** Other potential class of G-Quartet based glue junction, with the continuous stretch of 'GGGG' in Watson/Crick strand of the duplex,. that needs to be tested for their successful utilization.

References

(1) Zhuo, Z.; Yu, Y.; Wang, M.; Li, J.; Zhang, Z.; Liu, J.; Wu, X.; Lu, A.; Zhang, G.; Zhang, B. Recent Advances in SELEX Technology and Aptamer Applications in Biomedicine. *Int. J. Mol. Sci.* **2017**, *18* (10), 2140. https://doi.org/10.3390/ijms18102142.

(2) Huang, Y. C.; Sen, D. A Twisting Electronic Nanoswitch Made of DNA. *Angew. Chemie - Int. Ed.* **2014**, *53* (51), 14055–14059. https://doi.org/10.1002/anie.201407729.

(3) Ma, D. L.; He, H. Z.; Chan, D. S. H.; Leung, C. H. Simple DNA-Based Logic Gates Responding to Biomolecules and Metal Ions. *Chem. Sci.* **2013**, *4* (9), 3366–3380. https://doi.org/10.1039/c3sc50924a.

(4) Vecchioni, S.; Capece, M. C.; Toomey, E.; Nguyen, L.; Ray, A.; Greenberg, A.; Fujishima, K.; Urbina, J.; Paulino-Lima, I. G.; Pinheiro, V.; Shih, J.; Wessel, G.; Wind, S. J.; Rothschild, L. Construction and Characterization of Metal Ion-Containing DNA Nanowires for Synthetic Biology and Nanotechnology. *Sci. Rep.* **2019**, *9* (1), 1–15. https://doi.org/10.1038/s41598-019-43316-1.

(5) Gačanin, J.; Synatschke, C. V.; Weil, T. Biomedical Applications of DNA-Based Hydrogels. *Adv. Funct. Mater.* **2020**, *30* (4). https://doi.org/10.1002/adfm.201906253.

(6) Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440* (7082), 297–302. https://doi.org/10.1038/nature04586.

(7) Sokoloski, J. E.; Godfrey, S. A.; Dombrowski, S. E.; Bevilacqua, P. C. Prevalence of Syn Nucleobases in the Active Sites of Functional RNAs. *Rna* **2011**, *17* (10), 1775–1787. https://doi.org/10.1261/rna.2759911.

(8) Lafer, E. M.; Sousa, R.; Ali, R.; Rich, A.; Stollar, B. D. The Effect of Anti-Z-DNA Antibodies on the B-DNA-Z-DNA Equilibrium. *J. Biol. Chem.* **1986**, *261* (14), 6438–6443. https://doi.org/10.1016/s0021-9258(19)84581-7.

(9) Herbert, A.; Rich, A. The Biology of Left-Handed Z-DNA. *J. Biol. Chem.* **1996**, *271* (20), 11595–11598.

(10) Herbert, A. Z-DNA and Z-RNA in Human Disease. *Commun. Biol.* **2019**, *2* (1), 1–10. https://doi.org/10.1038/s42003-018-0237-x.

(11) Fujii, S.; Wang, A. H. J.; van der Marel, G.; van Boom, J. H.; Rich, A. Molecular Structure of $(m^5\ dC\text{-}dG)^3$: The Role of the Methyl Group on 5-Methyl Cytosine in Stabilizing Z-DNA. *Nucleic Acids Res.* **1982**, *10* (23), 7879–7892. https://doi.org/10.1093/nar/10.23.7879.

(12) Zhang, Y.; Cui, Y.; An, R.; Liang, X.; Li, Q.; Wang, H.; Wang, H.; Fan, Y.; Dong, P.; Li, J.; Cheng, K.; Wang, W.; Wang, S.; Wang, G.; Xue, C.; Komiyama, M. Topologically Constrained Formation of Stable Z-DNA from Normal Sequence under Physiological Conditions. *J. Am. Chem. Soc.* **2019**, *141* (19), 7758–7764. https://doi.org/10.1021/jacs.8b13855.

(13) Lafer, E. M.; Valle, R. P. C.; Moller, A.; Nordheim, A.; Schur, P. H.; Rich, A.; Stollar, B. D. Z-DNA-Specific Antibodies in Human Systemic Lupus Erythematosus. *J. Clin. Invest.* **1983**, *71*, 314–321. https://doi.org/10.1172/JCI110771.

(14) Shin, S. I.; Ham, S.; Park, J.; Seo, S. H.; Lim, C. H.; Jeon, H.; Huh, J.; Roh, T. Y. Z-DNA-Forming Sites Identified by ChIP-Seq are Associated with Actively Transcribed Regions in the Human Genome. *DNA Res.* **2016**, *23* (5), 477–486. https://doi.org/10.1093/dnares/dsw031.

(15) Crick, F.; Watson, J. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.

(16) Felsenfeld,G., Davies,D.R. and Rich, A. Formation of a Three-Stranded Polynucleotide Molecule. *JACS* **1957**, *79, 8, 2023-2024*.

(17) Frank-Kamenetskii, M. D.; Mirkin, S. M. Triplex DNA Structures. *Annu. Rev. Biochem.* **1995**, *64*, 65–95. https://doi.org/10.1146/annurev.bi.64.070195.000433.

(18) Broitman, S. L.; Im, D. D.; Fresco, J. R. Formation of the Triple-Stranded Polynucleotide Helix, Poly(A.A.U). *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 5120–5124. https://doi.org/10.1073/pnas.84.15.5120.

(19) Kohwi, Y.; Kohwi-shigematsu, T. Magnesium Ion-Dependent Triple-Helix Structure Formed by Homopurine-Homopyrimidine Sequences in Supercoiled Plasmid DNA. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 3781–3785.

(20) Beal, P.; Dervan, P. Second Structural Motif for Recognition of DNA by Oligonucleotide-Directed Triple-Helix Formation. **1991**, *251*, 1360–1363.

(21) Gaddis, S. S.; Wu, Q.; Thames, H. D.; Digiovanni, J.; Walborg, E. F.; Macleod, M. C.; Vasquez, K. M. A Web-Based Search Engine for Triplex-Forming Oligonucleotide Target Sequences. *Oligonucleotides* **2006**, *16*, 196–201. https://doi.org/10.1089/oli.2006.16.196.

(22) Schroth, G. P.; Ho, P. S. Occurrence of Potential Cruciform and H-DNA Forming Sequences in Genomic DNA. *Nucleic Acids Res.* **1995**, *23* (11), 1977–1983. https://doi.org/10.1093/nar/23.11.1977.

(23) Ohno, M.; Fukagawa, T.; Lee, J. S.; Ikemura, T. Triplex-Forming DNAs in the Human Interphase Nucleus Visualized in Situ by Polypurine/Polypyrimidine DNA Probes and Antitriplex Antibodies. *Chromosoma* **2002**, *111* (3), 201–213. https://doi.org/10.1007/s00412-002-0198-0.

(24) Agazie, Y. M.; Burkholder, G. D.; Lee, J. S. Triplex DNA in the Nucleus: Direct Binding of Triplex-Specific Antibodies and Their Effect on Transcription, Replication and Cell Growth. *Biochem. J.* **1996**, *316*, 461–466. https://doi.org/10.1042/bj3160461.

(25) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. Site-Specific Oligonucleotide Binding Represses Transcription of the Human c-Myc Gene in Vitro. *Science (80-.).* **1988**, *241*, 456–459.

(26) Christensen, L. A.; Finch, R. A.; Booker, A. J.; Vasquez, K. M. Targeting Oncogenes to Improve Breast Cancer Chemotherapy. *Cancer Res.* **2006**, *66* (8), 4089–4094. https://doi.org/10.1158/0008-5472.CAN-05-4288.

(27) Jain, A.; Wang, G.; Vasquez, K. M. DNA Triple Helices: Biological Consequences and Therapeutic Potential. *Biochimie* **2008**, *90* (8), 1117–1130. https://doi.org/10.1016/j.biochi.2008.02.011.

(28) Chandrasekaran, A. R.; Rusling, D. A. Triplex-Forming Oligonucleotides: A Third Strand for DNA Nanotechnology. *Nucleic Acids Res.* **2018**, *46* (3), 1021–1037. https://doi.org/10.1093/nar/gkx1230.

(29) Robinson, J.; Raguseo, F.; Nuccio, S. P.; Liano, D.; Di Antonio, M. DNA G-Quadruplex Structures: More than Simple Roadblocks to Transcription? *Nucleic Acids Res.* **2021**. https://doi.org/10.1093/nar/gkab609.

(30) Yang, L.; Bai, C.; Liu, C.; Shi, X.; Lee, I. Theoretical Studies on Conformation Comparison of Braid-like and Triplex DNA. *J. Mol. Struct.* **1999**, *477* (1–3), 7–13. https://doi.org/10.1016/S0022-2860(98)00510-9.

(31) Gehring, K; Gueron, J.-L. L M. A Tetrameric DNA Structure with Protonated Cytosine· Cytosine Base Pairs. *Nature* **1993**, *363*, 561–565. https://doi.org/10.1038/363561a0.

(32) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures are Formed in the Nuclei of Human Cells. *Nat. Chem.* **2018**, *10* (6), 631–637. https://doi.org/10.1038/s41557-018-0046-3.

(33) Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. Formation of i-motif Structure at Neutral and Slightly Alkaline pH. *Mol. Biosyst.* **2010**, *6* (3), 580–586. https://doi.org/10.1039/b919600e.

(34) Sun, D.; Hurley, L. H. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* **2009**, *52* (9), 2863–2874. https://doi.org/10.1021/jm900055s.

(35) Rajendran, A.; Nakano, S. I.; Sugimoto, N. Molecular Crowding of the Cosolutes Induces an Intramolecular i-motif Structure of Triplet Repeat DNA Oligomers at Neutral PH. *Chem. Commun.* **2010**, *46*, 1299–1301. https://doi.org/10.1039/b922050j.

(36) Satpathi, S.; Das, K.; Hazra, P. Silica Nano-Channel Induced i-motif Formation and Stabilization at Neutral and Alkaline pH. *Chem. Commun.* **2018**, *54*, 7054–7057. https://doi.org/10.1039/c8cc02811g.

(37) Day, H. A.; Huguin, C.; Waller, Z. A. E. Silver Cations Fold i-motif at Neutral pH. *Chem. Commun.* **2013**, *49*, 7696–7698. https://doi.org/10.1039/c3cc43495h.

(38) Abdelhamid, M. A. S.; Fábián, L.; Macdonald, C. J.; Cheesman, M. R.; Gates, A. J.; Waller, Z. A. E. Redox-Dependent Control of i-Motif DNA Structure Using Copper Cations. *Nucleic Acids Res.* **2018**, *46* (12), 5886–5893. https://doi.org/10.1093/nar/gky390.

(39) Assi, H. A.; Harkness, R. W.; Martin-Pintado, N.; Wilds, C. J.; Campos-Olivas, R.; Mittermaier, A. K.; Gonzalez, C.; Damha, M. J. Stabilization of i-motif Structures by 2'-β-Fluorination of DNA. *Nucleic Acids Res.* **2016**, *44* (11), 4998–5009. https://doi.org/10.1093/nar/gkw402.

(40) Assi, H. A.; El-Khoury, R.; González, C.; Damha, M. J. 2-Fluoroarabinonucleic Acid Modification Traps G-Quadruplex and i-motif Structures in Human Telomeric DNA. *Nucleic Acids Res.* **2017**, *45* (20), 11535–11546. https://doi.org/10.1093/nar/gkx838.

(41) GELLERT, M.; LIPSETT, M. N.; DAVIES, D. R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013–2018. https://doi.org/10.1073/pnas.48.12.2013.

(42) Sen, D.; Gilbert, W. Formation of Parallel Four-Stranded Complexes by Guanine-Rich Motifs in DNA and its Implications for Meiosis. *Nature* **1988**, *334* (6180), 364–366. https://doi.org/10.1038/334364a0.

(43) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415. https://doi.org/10.1093/nar/gkl655.

(44) Mergny, J. L., Lacroix, L. UV Melting of G- Quadruplexes. *Curr. Protoc. Nucleic Acid Chem.* **2009**, *37* (0), 17.1.1-17.1.15.

(45) Venczel, E. A.; Sen, D. Parallel and Antiparallel G-DNA Structures from a Complex Telomeric Sequence. *Biochemistry* **1993**, *32*, 6220–6228. https://doi.org/10.1021/bi00075a015.

(46) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, *429* (14), 2127–2147. https://doi.org/10.1016/j.jmb.2017.05.017.

(47) Lightfoot, H. L.; Hagen, T.; Tatum, N. J.; Hall, J. The Diverse Structural Landscape of Quadruplexes. *FEBS Lett.* **2019**, *593* (16), 2083–2102. https://doi.org/10.1002/1873-3468.13547.

(48) Webba Da Silva, M. Geometric Formalism for DNA Quadruplex Folding. *Chem. - A Eur. J.* **2007**, *13* (35), 9738–9745. https://doi.org/10.1002/chem.200701255.

(49) Dvorkin, S. A.; Karsisiotis, A. I.; da Silva, M. W. Encoding Canonical DNA Quadruplex Structure. *Sci. Adv.* **2018**, *4* (8). https://doi.org/10.1126/sciadv.aat3007.

(50) Yang, S. Y.; Lejault, P.; Chevrier, S.; Boidot, R.; Robertson, A. G.; Wong, J. M. Y.; Monchaud, D. Transcriptome-Wide Identification of Transient RNA G-Quadruplexes in Human Cells. *Nat. Commun.* **2018**, *9* (1) 4730. https://doi.org/10.1038/s41467-018-07224-8.

(51) Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J. F.; Balasubramanian, S. A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *J. Am. Chem. Soc.* **2008**, *130* (47), 15758–15759. https://doi.org/10.1021/ja805615w.

(52) Gowan, S. M.; Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. A G-Quadruplex-Interactive Potent Small-Molecule Inhibitor of Telomerase Exhibiting in Vitro and in Vivo Antitumor Activity. *Mol. Pharmacol.* **2002**, *61* (5), 1154–1162. https://doi.org/10.1124/mol.61.5.1154.

(53) Li, Y.; Geyer, C. R.; Sen, D. Recognition of Anionic Porphyrins by DNA Aptamers. *Biochemistry* **1996**, *35* (21), 6911–6922. https://doi.org/10.1021/bi960038h.

(54) Monchaud, D.; Allain, C.; Bertrand, H.; Smargiasso, N.; Rosu, F.; Gabelica, V.; De Cian, A.; Mergny, J. L.; Teulade-Fichou, M. P. Ligands Playing Musical Chairs with G-Quadruplex DNA: A Rapid and Simple Displacement Assay for Identifying Selective G-Quadruplex Binders. *Biochimie* **2008**, *90* (8), 1207–1223. https://doi.org/10.1016/j.biochi.2008.02.019.

(55) Yamamoto, Y.; Araki, H.; Shinomiya, R.; Hayasaka, K.; Nakayama, Y.; Ochi, K.; Shibata, T.; Momotake, A.; Ohyama, T.; Hagihara, M.; Hemmi, H. Structures and Catalytic Activities of Complexes between Heme and All Parallel-Stranded Monomeric G-Quadruplex DNAs. *Biochemistry* **2018**, *57* (41), 5938–5948. https://doi.org/10.1021/acs.biochem.8b00792.

(56) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and Selective Chemical Targeting of RNA G-Quadruplex Structures in the Cytoplasm of Human Cells. *Nat. Chem.* **2014**, *6* (1), 75–80. https://doi.org/10.1038/nchem.1805.

(57) Biffi, G; Tannahill, D; McCafferty, J; Balasubramanian S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5* (3), 182–186. https://doi.org/10.1038/nchem.1548.

(58) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Plückthun, A. *In Vitro* Generated Antibodies Specific for Telomeric Guanine-Quadruplex DNA React with *Stylonychia Lemnae* Macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (15), 8572–8577. https://doi.org/10.1073/pnas.141229498.

(59) Chilka, P.; Desai, N.; Datta, B. Small Molecule Fluorescent Probes for G-Quadruplex Visualization as Potential Cancer Theranostic Agents. *Molecules* **2019**, *24* (4). https://doi.org/10.3390/molecules24040752.

(60) Summers, P. A.; Lewis, B. W.; Gonzalez-Garcia, J.; Porreca, R. M.; Lim, A. H. M.; Cadinu, P.; Martin-Pintado, N.; Mann, D. J.; Edel, J. B.; Vannier, J. B.; Kuimova, M. K.; Vilar, R. Visualising G-Quadruplex DNA Dynamics in Live Cells by Fluorescence Lifetime Imaging Microscopy. *Nat. Commun.* **2021**, *12* (1), 1–11. https://doi.org/10.1038/s41467-020-20414-7.

(61) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* **2005**, *33* (9), 2908–2916. https://doi.org/10.1093/nar/gki609.

(62) Todd, A. K.; Johnston, M.; Neidle, S. Highly Prevalent Putative Quadruplex Sequence Motifs in Human DNA. *Nucleic Acids Res.* **2005**, *33* (9), 2901–2907. https://doi.org/10.1093/nar/gki553.

(63) Brázda, V.; Kolomazník, J.; Lýsek, J.; Bartas, M.; Fojta, M.; Šťastný, J.; Mergny, J. L. G4Hunter Web Application: A Web Server for G-Quadruplex Prediction. *Bioinformatics* **2019**, *35* (18), 3493–3495. https://doi.org/10.1093/bioinformatics/btz087.

(64) Kikin, O.; D'Antonio, L.; Bagga, P. S. QGRS Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Res.* **2006**, *34* (WEB. SERV. ISS.), 676–682. https://doi.org/10.1093/nar/gkl253.

(65) Garant, J. M.; Perreault, J. P.; Scott, M. S. Motif Independent Identification of Potential RNA G-Quadruplexes by G4RNA Screener. *Bioinformatics* **2017**, *33* (22), 3532–3537. https://doi.org/10.1093/bioinformatics/btx498.

(66) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* **2015**, *33* (8), 877–881. https://doi.org/10.1038/nbt.3295.

(67) Marsico, G.; Chambers, V. S.; Sahakyan, A. B.; McCauley, P.; Boutell, J. M.; Antonio, M. Di; Balasubramanian, S. Whole Genome Experimental Maps of DNA G-Quadruplexes in Multiple Species. *Nucleic Acids Res.* **2019**, *47* (8), 3862–3874. https://doi.org/10.1093/nar/gkz179.

(68) Wu, C. G.; Spies, M. G-Quadruplex Recognition and Remodeling by the FANCJ Helicase. *Nucleic Acids Res.* **2016**, *44* (18), 8742–8753. https://doi.org/10.1093/nar/gkw574.

(69) Budhathoki, J. B.; Stafford, E. J.; Yodh, J. G.; Balci, H. ATP-Dependent G-Quadruplex Unfolding by Bloom Helicase Exhibits Low Processivity. *Nucleic Acids Res.* **2015**, *43* (12), 5961–5970. https://doi.org/10.1093/nar/gkv531.

(70) London, T. B. C.; Barber, L. J.; Mosedale, G.; Kelly, G. P.; Balasubramanian, S.; Hickson, I. D.; Boulton, S. J.; Hiom, K. FANCJ Is a Structure-Specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts. *J. Biol. Chem.* **2008**, *283* (52), 36132–36139. https://doi.org/10.1074/jbc.M808152200.

(71) Wu, Y.; Shin-ya, K.; Brosh, R. M. FANCJ Helicase Defective in Fanconia Anemia and Breast Cancer Unwinds G-Quadruplex DNA To Defend Genomic Stability. *Mol. Cell. Biol.* **2008**, *28* (12), 4116–4128. https://doi.org/10.1128/mcb.02210-07.

(72) Ribeyre, C.; Lopes, J.; Boulé, J. B.; Piazza, A.; Guédin, A.; Zakian, V. A.; Mergny, J. L.; Nicolas, A. The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences in Vivo. *PLoS Genet.* **2009**, *5* (5). https://doi.org/10.1371/journal.pgen.1000475.

(73) Bao, H. L.; Xu, Y. Investigation of Higher-Order RNA G-Quadruplex Structures i*n Vitro* and in Living Cells by ^{19}F NMR Spectroscopy. *Nat. Protoc.* **2018**, *13* (4), 652–665. https://doi.org/10.1038/nprot.2017.156.

(74) Maxam, A. M.; Gilbert, W. A New Method for Sequencing DNA. *Proc. Natl. Acad. Sci. USA* **1976**, *74* (2), 560–564.

(75) Kouzine, F.; Wojtowicz, D.; Yamane, A.; Casellas, R.; Przytycka, T. M.; Levens, D. L. In Vivo Chemical Probing for G-Quadruplex Formation. *Methods Mol. Biol.* **2019**, *2035*, 369–382. https://doi.org/10.1007/978-1-4939-9666-7_23.

(76) Brooks, T. A.; Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **2010**, *1* (6), 641–649. https://doi.org/10.1177/1947601910377493.

(77) Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D. The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-Nt Loop in K^+ Solution. *J. Am. Chem. Soc.* **2014**, *136* (5), 1750–1753. https://doi.org/10.1021/ja4118945.

(78) Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the *KRAS* Proto-Oncogene and its Effect on Transcription. *Nucleic Acids Res.* **2006**, *34* (9), 2536–2549. https://doi.org/10.1093/nar/gkl286.

(79) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. An RNA G-Quadruplex in the 5′ UTR of the *NRAS* Proto- Oncogene Modulates Translation. *Nat. Chem Biol.* **2007**, *3* (4), 218–221. https://doi.org/10.1038/nchembio864.

(80) Mazzini, S.; Gargallo, R.; Musso, L.; De Santis, F.; Aviñó, A.; Scaglioni, L.; Eritja, R.; Di Nicola, M.; Zunino, F.; Amatulli, A.; Dallavalle, S. Stabilization of C-KIT G-Quadruplex DNA Structures by the RNA Polymerase I Inhibitors BMH-21 and BA-41. *Int. J. Mol. Sci.* **2019**, *20* (19), 4927. https://doi.org/10.3390/ijms20194927.

(81) Agrawal, P.; Hatzakis, E.; Guo, K.; Carver, M.; Yang, D. Solution Structure of the Major G-Quadruplex Formed in the Human VEGF Promoter in K^+: Insights into Loop Interactions of the Parallel G-Quadruplexes. *Nucleic Acids Res.* **2013**, *41* (22), 10584–10592. https://doi.org/10.1093/nar/gkt784.

(82) Cui, X.; Chen, H.; Zhang, Q.; Xu, M.; Yuan, G.; Zhou, J. Exploration of the Structure and Recognition of a G-Quadruplex in the Her2 Proto-Oncogene Promoter and Its Transcriptional Regulation. *Sci. Rep.* **2019**, *9* (1), 1–12. https://doi.org/10.1038/s41598-019-39941-5.

(83) Arora, A.; Suess, B. An RNA G-Quadruplex in the 3′ UTR of the Proto-Oncogene PIM1 Represses Translation. *RNA Biol.* **2011**, *8* (5), 1–5. https://doi.org/10.4161/rna.8.5.16038.

(84) Saha, T.; Usdin, K. Tetraplex Formation by the Progressive Myoclonus Epilepsy Type-1 Repeat: Implications for Instability in the Repeat Expansion Diseases. *FEBS Lett.* **2001**, *491*, 184–187. https://doi.org/10.1016/S0014-5793(01)02190-1.

(85) Didiot, M. C.; Tian, Z.; Schaeffer, C.; Subramanian, M.; Mandel, J. L.; Moine, H. The G-Quartet Containing FMRP Binding Site in FMR1 mRNA is a Potent Exonic Splicing Enhancer. *Nucleic Acids Res.* **2008**, *36* (15), 4902–4912. https://doi.org/10.1093/nar/gkn472.

(86) Zhang, Y.; Roland, C.; Sagui, C. Structural and Dynamical Characterization of DNA and RNA Quadruplexes Obtained from the GGGGCC and GGGCCT Hexanucleotide Repeats Associated with C9FTD/ALS and SCA36 Diseases. *ACS Chem Neurosci.* **2018**, *9* (5), 1104–1117. https://doi.org/10.1021/acschemneuro.7b00476.

(87) Pradhan, P.; Srivastava, A.; Singh, J.; Biswas, B.; Saini, A.; Siddique, I.; Kumari, P.; Khan, M. A.; Mishra, A.; Yadav, P. K.; Kumar, S.; Bhavesh, N. S.; Venkatraman, P.; Vivekanandan, P.; Kundu, B. Prion Protein Transcription is Auto-Regulated through Dynamic Interactions with G-Quadruplex Motifs in its Own Promoter. *Biochim Biophys Acta Gene Regul Mech* **2020**, *1863* (3), 194479. https://doi.org/10.1016/j.bbagrm.2019.194479.

(88) Simone, R.; Fratta, P.; Neidle, S.; Parkinson, G. N.; Isaacs, A. M. G-Quadruplexes: Emerging Roles in Neurodegenerative Diseases and the Non-Coding Transcriptome. *FEBS Lett.* **2015**, *589* (14), 1653–1668. https://doi.org/10.1016/j.febslet.2015.05.003.

(89) Dejesus-hernandez, M.; Mackenzie, I. R.; Boeve, B. F.; Boxer, A. L.; Baker, M.; Rutherford, N. J.; Nicholson, A. M.; Finch, N. A.; Flynn, H.; Adamson, J.; Kouri, N.; Wojtas, A.; Sengdy, P.; Hsiung, G. R.; Karydas, A.; Seeley, W. W.; Josephs, K. A.; Coppola, G.; Geschwind, D. H.; Wszolek, Z. K.; Feldman, H.; Knopman, D. S.; Petersen, R. C.; Miller, B. L.; Dickson, D. W.; Boylan, K. B.; Graff-radford, N. R.; Rademakers, R. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of *C9ORF72* Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **2011**, *72*, 245–256. https://doi.org/10.1016/j.neuron.2011.09.011.

(90) Renton, A. E.; Majounie, E.; Waite, A.; Simo´n-Sa´nchez, J.; Rollinson, S.; Gibbs, J. R.; Schymick, J. C.; Laaksovirta, H.; Swieten, J. C. van; Myllykangas, L.; Kalimo, H.; Paetau, A.; Abramzon, Y.; Remes, A. M.; Kaganovich, A.; Scholz, S. W.; Duckworth, J.; Ding, J.; Harmer, D. W.; Hernandez, D. G.; Johnson, J. O.; Mok, K.; Ryten, M.; Trabzuni, D.; Guerreiro, R. J.; Orrell, R. W.; Neal, J.; Murray, A.; Pearson, J.; Jansen, I. E.; Sondervan, D.; Seelaar, H.; Blake, D.; Young, K.; Halliwell, N.; Callister, J. B.; Toulson, G.; Richardson, A.; Gerhard, A.; Snowden, J.; Mann, D.; Neary, D.; Nalls, M. A.; Peuralinna, T.; Jansson, L.; Isoviita, V.-M.; Kaivorinne, A.-L.; Ho¨ ltta¨-Vuori, M.; Ikonen, E.; Sulkava, R.; Benatar, M.; Wuu, J.; Chio`, A.; Restagno, G.; Borghero, G.; Sabatelli, M.; Consortium, T. I.; Heckerman, D.; Rogaeva, E.; Zinman, L.; Rothstein, J. D.; Sendtner, M.; Drepper, C.; Eichler, E. E.; Alkan, C.; Abdullaev, Z.; Pack, S. D.; Dutra, A.; Pak, E.; Hardy, J.; Singleton, A.; Williams, N. M.; Heutink, P.; Pickering-Brown, S.; Morris, H. R.; Tienari, P. J.; Traynor, B. J. A Hexanucleotide Repeat Expansion in *C9ORF72* is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron* **2011**, *72*, 257–268. https://doi.org/10.1016/j.neuron.2011.09.010.

(91) Gómez-Tortosa, E.; Gallego, J.; Guerrero-López, R.; Marcos, A.; Gil-Neciga, E.; Sainz, M. J.; Díaz, A.; Franco-Macías, E.; Trujillo-Tiebas, M. J.; Ayuso, C.; Pérez-Pérez, J. C9ORF72 Hexanucleotide Expansions of 20–22 Repeats Are Associated with Frontotemporal Deterioration. *Neurology* **2013**, *80*, 366–370.

(92) Beck, J.; Poulter, M.; Hensman, D.; Rohrer, J. D.; Mahoney, C. J.; Adamson, G.; Campbell, T.; Uphill, J.; Borg, A.; Fratta, P.; Orrell, R. W.; Malaspina, A.; Rowe, J.; Brown, J.; Hodges, J.; Sidle, K.; Polke, J. M.; Houlden, H.; Schott, J. M.; Fox, N. C.; Rossor, M. N.; Tabrizi, S. J.; Isaacs, A. M.; Hardy, J.; Warren, J. D.; Collinge, J.; Mead, S. Large *C9orf72* Hexanucleotide Repeat Expansions Are Seen in Multiple Neurodegenerative Syndromes and Are More Frequent Than Expected in the UK Population. *Am. J. Hum. Genet.* **2013**, *92*, 345–353.

(93) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-motif Functions in Oncogene Promoters. *FEBS J.* **2010**, *277* (17), 3459–3469. https://doi.org/10.1111/j.1742-4658.2010.07759.x.

(94) Kendrick, S.; Hurley, L. H. The Role of G-Quadruplex/i-motif Secondary Structures as Cis-Acting Regulatory Elements. *Pure Appl. Chem.* **2010**, *82* (8), 1609–1621. https://doi.org/10.1351/PAC-CON-09-09-29.

(95) Qin, Y.; Hurley, L. H. Structures, Folding Patterns, and Functions of Intramolecular DNA G-Quadruplexes Found in Eukaryotic Promoter Regions. *Biochimie* **2008**, *90* (8), 1149–1171. https://doi.org/10.1016/j.biochi.2008.02.020.

(96) Mizielinska, S.; Lashley, T.; Norona, F. E.; Clayton, E. L.; Ridler, C. E.; Fratta, P.; Isaacs, A. M. C9orf72 Frontotemporal Lobar Degeneration is Characterised by Frequent Neuronal Sense and Antisense RNA Foci. *Acta Neuropathol* **2013**, *126*, 845–857.

(97) Lee, Y.-B.; Chen, H.-J.; Peres, J. N.; Gomez-Deza, J.; Attig, J.; Stalekar, M.; Troakes, C.; Nishimura, A. L.; Scotter, E. L.; Vance, C.; Adachi, Y.; Sardone, V.; Miller, J. W.; Smith, B. N.; Gallo, J.-M.; Ule, J.; Hirth, F.; Rogelj, B.; Houart, C.; Shaw, C. E. Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and are Neurotoxic. *Cell Rep.* **2013**, *5*, 1178–1186.

(98) Gendron, T.F; Belzil, V.V; Zhang, Y-J; Petrucelli, L. Mechanisms of Toxicity in C9FTLD/ALS. *Acta Neuropathol* **2014**, *127*, 359–376.

(99) Gendron, T. F.; Bieniek, K. F.; Zhang, Y.-J.; Jansen-West, K.; Ash, P. E. A.; Caulfield, T.; Daughrity, L.; Dunmore, J. H.; Castanedes-Casey, M.; Chew, J.; Cosio, D. M.; Blitterswijk, M. van; Lee, W. C.; Rademakers, R.; Boylan, K. B.; Dickson, D. W.; Petrucelli, L. Antisense Transcripts of the Expanded *C9ORF72* Hexanucleotide Repeat Form Nuclear RNA Foci and Undergo Repeat-associated Non-ATG Translation in C9FTD / ALS. *Acta Neuropathol* **2013**, *126*, 829–844. https://doi.org/10.1007/s00401-013-1192-8.

(100) Grigg, J. C.; Shumayrikh, N.; Sen, D. G-Quadruplex Structures Formed by Expanded Hexanucleotide Repeat RNA and DNA from the Neurodegenerative Disease-Linked *C9orf72* Gene Efficiently Sequester and Activate Heme. *PLoS One* **2014**, *9* (9), e106449. doi:10.1371/journal.pone.0106449. https://doi.org/10.1371/journal.pone.0106449.

(101) Mori, K.; Weng, S.; Arzberger, T.; May, S.; Rentzsch, K.; Kremmer, E.; Schmid, B.; Kretzschmar, H. A.; Cruts, M.; Broeckhoven, C. Van; Haass, C.; Edbauer, D. The C9orf72 GGGGCC Repeat is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science (80-.).* **2013**, *339*, 1335–1338.

(102) Mori, K.; Arzberger, T.; Grässer, F. A.; Gijselinck, I.; May, S.; Rentzsch, K.; Weng, S.-M.; Schludi, M. H.; Zee, J. van der; Cruts, M.; Broeckhoven, C. Van; Kremmer, E.; Kretzschmar, H. A.; Haass, C.; Edbauer, D. Bidirectional Transcripts of the Expanded C9orf72 Hexanucleotide Repeat are Translated into Aggregating Dipeptide Repeat Proteins. *Acta Neuropathol* **2013**, *126*, 881–893. https://doi.org/10.1007/s00401-013-1189-3.

(103) Huang, Y. C.; Yu, H. Z.; Sen, D. DNA Mechatronic Devices Switched by K^+ and by Sr^{2+} are Structurally, Topologically, and Electronically Distinct. *Biopolymers* **2015**, *103* (8), 460–468. https://doi.org/10.1002/bip.22595.

(104) Ge, L.; Wang, W.; Sun, X.; Hou, T.; Li, F. Versatile and Programmable DNA Logic Gates on Universal and Label-Free Homogeneous Electrochemical Platform. *Anal. Chem.* **2016**, *88* (19), 9691–9698. https://doi.org/10.1021/acs.analchem.6b02584.

(105) Gao, R. R.; Yao, T. M.; Lv, X. Y.; Zhu, Y. Y.; Zhang, Y. W.; Shi, S. Integration of G-Quadruplex and DNA-Templated Ag NCs for Nonarithmetic Information Processing. *Chem. Sci.* **2017**, *8* (6), 4211–4222. https://doi.org/10.1039/c7sc00361g.

(106) Kotlyar, A. B.; Borovok, N.; Molotsky, T.; Cohen, H.; Shapir, E.; Porath, D. Long, Monomolecular Guanine-Based Nanowires. *Adv. Mater.* **2005**, *17* (15), 1901–1905. https://doi.org/10.1002/adma.200401997.

(107) Troha, T.; Drevenšek-Olenik, I.; Webba Da Silva, M.; Spindler, L. Surface-Adsorbed Long G-Quadruplex Nanowires Formed by G:C Linkages. *Langmuir* **2016**, *32* (28), 7056–7063. https://doi.org/10.1021/acs.langmuir.6b01222.

(108) Mitchell, J. S.; Glowacki, J.; Grandchamp, A. E.; Manning, R. S.; Maddocks, J. H. Sequence-Dependent Persistence Lengths of DNA. *J. Chem. Theory Comput.* **2017**, *13* (4), 1539–1555. https://doi.org/10.1021/acs.jctc.6b00904.

(109) Zhang, L. *Heme Biology, The Secret Life of Heme in Regulating Diverse Biological Processes*; Singapore: World Scientific Publishing Co. Pte. Ltd, 2011.

(110) Mense, S. M.; Zhang, L. Heme: A Versatile Signaling Molecule Controlling the Activities of Diverse Regulators Ranging from Transcription Factors to MAP Kinases. *Cell Res.* **2006**, *16* (8), 681–692. https://doi.org/10.1038/sj.cr.7310086.

(111) Shokat, K. .; Schultz, P. G. Catalytic Antibodies. *Annu. Rev. Immunol.* **1990**, *8*, 335–363. https://doi.org/10.1002/3527603662.

(112) Tramontano, A.; Janda, K. D.; Lerner, R. A. Catalytic Antibodies. *Science (80-.).* **1986**, *234*, 1566–1570.

(113) Golub, E.; Freeman, R.; Willner, I. A Hemin/G-Quadruplex Acts as an NADH Oxidase and NADH Peroxidase Mimicking DNAzyme. *Angew. Chemie - Int. Ed.* **2011**, *50* (49), 11710–11714. https://doi.org/10.1002/anie.201103853.

(114) Tao, L.; Shaojun, D.; Erkang, W. G-Quadruplex Aptamers with Peroxidase-like DNAzyme Functions: Which the Best and How Does It Work? *Chem. - An Asian J.* **2009**, *4* (6), 918–922. https://doi.org/10.1002/asia.200900019.

(115) Sen, D.; Poon, L. C. H. RNA and DNA Complexes with Hemin [Fe(III) Heme] are Efficient Peroxidases and Peroxygenases: How Do They Do It and What Does It Mean? *Crit. Rev. Biochem. Mol. Biol.* **2011**, *46* (6), 478–492. https://doi.org/10.3109/10409238.2011.618220.

(116) Poon, L. C. H.; Methot, S. P.; Morabi-Pazooki, W.; Pio, F.; Bennet, A. J.; Sen, D. Guanine-Rich RNAs and DNAs That Bind Heme Robustly Catalyze Oxygen Transfer Reactions. *J. Am. Chem. Soc.* **2011**, *133* (6), 1877–1884. https://doi.org/10.1021/ja108571a.

(117) Travascio, P.; Witting, P. K.; Mauk, A. G.; Sen, D. The Peroxidase Activity of a Hemin-DNA Oligonucleotide Complex: Free Radical Damage to Specific Guanine Bases of the DNA. *J. Am. Chem. Soc.* **2001**, *123* (7), 1337–1348. https://doi.org/10.1021/ja0023534.

(118) Travascio, P.; Bennet, A. J.; Wang, D. Y.; Sen, D. A Ribozyme and a Catalytic DNA with Peroxidase Activity: Active Sites versus Cofactor-Binding Sites. *Chem. Biol.* **1999**, *6* (11), 779–787. https://doi.org/10.1016/S1074-5521(99)80125-2.

(119) Travascio, P.; Li, Y.; Sen, D. DNA-Enhanced Peroxidase Activity of a DNA Aptamer-Hemin Complex. *Chem. Biol.* **1998**, *5* (9), 505–517. https://doi.org/10.1016/S1074-5521(98)90006-0.

(120) Franzen, S.; Boxer, S. G. On the Origin of Heme Absorption Band Shifts and Associated Protein Structural Relaxation in Myoglobin Following Flash Photolysis. *J. Biol. Chem.* **1997**, *272* (15), 9655–9660. https://doi.org/10.1074/jbc.272.15.9655.

(121) Kong, D. M.; Yang, W.; Wu, J.; Li, C. X.; Shen, H. X. Structure-Function Study of Peroxidase-like G-Quadruplex-Hemin Complexes. *Analyst* **2010**, *135* (2), 321–326. https://doi.org/10.1039/b920293e.

(122) Saito, K.; Tai, H.; Hemmi, H.; Kobayashi, N.; Yamamoto, Y. Interaction between the Heme and a G-Quartet in a Heme-DNA Complex. *Inorg. Chem.* **2012**, *51* (15), 8168–8176. https://doi.org/10.1021/ic3005739.

(123) Rojas, A. M.; Gonzalez, P. A.; Antipov, E.; Klibanov, A. M. Specificity of a DNA-Based (DNAzyme) Peroxidative Biocatalyst. *Biotechnol. Lett.* **2007**, *29* (2), 227–232. https://doi.org/10.1007/s10529-006-9228-y.

(124) Shumayrikh, N.; Huang, Y. C.; Sen, D. Heme Activation by DNA: Isoguanine Pentaplexes, but not Quadruplexes, Bind Heme and Enhance its Oxidative Activity. *Nucleic Acids Res.* **2015**, *43* (8), 4191–4201. https://doi.org/10.1093/nar/gkv266.

(125) Ellington, A. D.; Szostak, J. W. *In Vitro* Selection of RNA Molecules that Bind Specific Ligands. *Nature* **1990**, *346*, 818–822. https://doi.org/10.1038/346818a0.

(126) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science (80-.).* **1990**, *249*, 505–510. https://doi.org/10.1126/science.2200121.

(127) Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. Circular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37* (6), 1713–1725. https://doi.org/10.1093/nar/gkp026.

(128) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2006**, *1* (6), 2876–2890. https://doi.org/10.1038/nprot.2006.202.

(129) Johnson, K. H.; Durland, R. H.; Hogan, M. E. The Vacuum UV CD Spectra of GGC Triplexes. *Nucleic Acids Res.* **1992**, *20* (15), 3859–3864.

(130) Pilch, D. S.; Levensont, C.; Shafert, R. H. Structural Analysis of the (dA)10.2(dT)10 Triple Helix. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 1942–1946.

(131) Chung, W. J.; Heddi, B.; Schmitt, E.; Lim, K. W.; Mechulam, Y.; Phan, A. T. Structure of a Left-Handed DNA G-Quadruplex. *Proc. Natl. Acad. Sci.* **2015**, *112* (9), 2729–2733. https://doi.org/10.1073/pnas.1418718112.

(132) Silva, M. W. da. NMR Methods for Studying Quadruplex Nucleic Acids. *Methods* **2007**, *43*, 264–277. https://doi.org/10.1016/j.ymeth.2007.05.007.

(133) Varani, G.; Aboul-ela, F.; Allain, F. H.-T. NMR Investigation of RNA Structure. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 51–127.

(134) Cotton, R. G.; Rodrigues, N. R.; Campbell, R. D. Reactivity of Cytosine and Thymine in Single-Base-Pair Mismatches with Hydroxylamine and Osmium Tetroxide and its Application to the Study of Mutations. *Proc. Natl. Acad. Sci.* **1988**, *85* (12), 4397–4401. https://doi.org/10.1073/pnas.85.12.4397.

(135) Ross, S. A.; Burrows, C. J. Cytosine-Specific Chemical Probing of DNA Using Bromide and Monoperoxysulfate. *Nucleic Acids Res.* **1996**, *24* (24), 5062–5063. https://doi.org/10.1093/nar/24.24.5062.

(136) Binnig, G.; Quate, C. F. Atomic Force Microscope. *Phys. Rev. Lett.* **1986**, *56* (9), 930–933.

(137) Edward J. Big. A Short History of the Electron Microscope. *Bios* **1956**, *27*, 33–37.

(138) Paulson, H. *Repeat Expansion Diseases*; 2018; 147, 105-123. https://doi.org/10.1016/B978-0-444-63233-3.00009-9.

(139) Zhou, B.; Liu, C.; Geng, Y.; Zhu, G. Topology of a G-Quadruplex DNA Formed by *C9orf72* Hexanucleotide Repeats Associated with ALS and FTD. *Sci. Rep.* **2015**, *5*, 16673.

(140) Pataskar, S. S.; Dash, D.; Brahmachari, S. K. Progressive Myoclonus Epilepsy [EPM1] Repeat d(CCCCGCCCCGCG)$_n$ Forms Folded Hairpin Structures at Physiological pH. *J. Biomol. Struct. Dyn.* **2001**, *19* (2), 293–305. https://doi.org/10.1080/07391102.2001.10506740.

(141) Obayashi, M.; Stevanin, G.; Synofzik, M.; Monin, M. L.; Duyckaerts, C.; Sato, N.; Streichenberger, N.; Vighetto, A.; Desestret, V.; Tesson, C.; Wichmann, H. E.; Illig, T.; Huttenlocher, J.; Kita, Y.; Izumi, Y.; Mizusawa, H.; Schöls, L.; Klopstock, T.; Brice, A.; Ishikawa, K.; Dürr, A. Spinocerebellar Ataxia Type 36 Exists in Diverse Populations and can be Caused by a Short Hexanucleotide GGCCTG Repeat Expansion. *J. Neurol. Neurosurg. Psychiatry* **2015**, *86* (9), 986–995. https://doi.org/10.1136/jnnp-2014-309153.

(142) Goel, D.; Suroliya, V.; Shamim, U.; Mathur, A.; Faruq, M. Spinocerebellar Ataxia Type 10 (SCA10): Mutation Analysis and Common Haplotype Based Inference Suggest its Rarity in Indian Population. *eNeurologicalSci* **2019**, *17*, 100211. https://doi.org/10.1016/j.ensci.2019.100211.

(143) Liquori, C. L.; Ricker, K.; Moseley, M. L.; Jacobsen, J. F.; Kress, W.; Naylor, S. L.; Day, J. W.; Ranum, L. P. W. Myotonic Dystrophy Type 2 Caused by a CCTG Expansion in Intron I of *ZNF9*. *Science (80-.).* **2001**, *293* (5531), 864–867. https://doi.org/10.1126/science.1062125.

(144) Sharma, D; Gupta, M; Thelma, B K. Expansion Mutation Frequency and CGG/GCC Repeat Polymorphism in *FMR1* and *FMR2* Genes in an Indian Population. *Genet. Epidemiol.* **2001**, *20* (1), 129–144. https://doi.org/10.1002/1098-2272(200101)20:1<129::AID-GEPI11>3.0.CO;2-2.

(145) Mller, T.; Schrder, R.; Zierz, S. GCG Repeats and Phenotype in Oculopharyngeal Muscular Dystrophy. *Muscle and Nerve* **2001**, *24* (1), 120–122. https://doi.org/10.1002/1097-4598(200101)24:1<120::AID-MUS17>3.0.CO;2-0.

(146) Patel, P. I.; Isaya, G. Friedreich Ataxia: From GAA Triplet-Repeat Expansion to Frataxin Deficiency. *Am. J. Hum. Genet.* **2001**, *69* (1), 15–24. https://doi.org/10.1086/321283.

(147) Ohshima, K.; Kang, S.; Wells, R. D. CTG Triplet Repeats from Human Hereditary Diseases are Dominant Genetic Expansion Products in Escherichia Coli. *J. Biol. Chem.* **1996**, *271* (4), 1853–1856. https://doi.org/10.1074/jbc.271.4.1853.

(148) Carroll, L. S.; Massey, T. H.; Wardle, M.; Peall, K. J. Dentatorubral-Pallidoluysian Atrophy: An Update. *Tremor and Other Hyperkinetic Movements* **2018**, *8*: 577 https://doi.org/10.7916/D81N9HST.

(149) Jang, Y. J.; Lee, C.; Kim, S. K. Formation of Poly[d(A-T)$_2$] Specific Z-DNA by a Cationic Porphyrin. *Sci. Rep.* **2015**, *5*, 9943. https://doi.org/10.1038/srep09943.

(150) Kovanda, A.; Zalar, M.; Šket, P.; Plavec, J.; Rogelj, B. Anti-Sense DNA d(GGCCCC)$_n$ Expansions in C9ORF72 Form i-motifs and Protonated Hairpins. *Sci. Rep.* **2016**, *5*, 17944. https://doi.org/10.1038/srep17944.

(151) Zamiri, B.; Mirceta, M.; Bomsztyk, K.; Macgregor, R. B.; Pearson, C. E. Quadruplex Formation by both G-Rich and C-Rich DNA Strands of the *C9orf72* (GGGGCC)8•(GGCCCC)8 Repeat: Effect of CpG Methylation. *Nucleic Acids Res.* **2015**, *43* (20), 10055–10064. https://doi.org/10.1093/nar/gkv1008.

(152) Plum, G. E.; Park, Y. W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. Thermodynamic Characterization of the Stability and the Melting Behavior of a DNA Triplex: A Spectroscopic and Calorimetric Study. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (23), 9436–9440. https://doi.org/10.1073/pnas.87.23.9436.

(153) Pilch, D. S.; Levenson, C.; Shafer, R. H. Structure, Stability, and Thermodynamics of a Short Intermolecular Purine-Purine-Pyrimidine Triple Helix. *Biochemistry* **1991**, *30* (25), 6081–6087.

(154) Khomyakova, E. B.; Gousset, H.; Liquier, J.; Huynh-Dinh, T.; Gouyette, C.; Takahashi, M.; Florentiev, V. L.; Taillandier, E. Parallel Intramolecular DNA Triple Helix with G and T Bases in the Third Strand Stabilized by Zn^{2+} Ions. *Nucleic Acids Res.* **2000**, *28* (18), 3511–3516. https://doi.org/10.1093/nar/28.18.3511.

(155) Gondeau, C.; Maurizot, J.; Durand, M. Spectroscopic Investigation of an Intramolecular DNA Triplex Containing Both G . G : C and T . A : T Triads and its Complex with Netropsin. *J. Biomol. Struct. Dyn.* **1998**, *15* (6), 1134–1145. https://doi.org/10.1080/07391102.1998.10509007.

(156) Gondeau, C.; Maurizot, J. C.; Durand, M. Circular Dichroism and UV Melting Studies on Formation of an Intramolecular Triplex Containing Parallel T*A:T and G*G:C Triplets: Netropsin Complexation with the Triplex. *Nucleic Acids Res.* **1998**, *26* (21), 4996–5003. https://doi.org/10.1093/nar/26.21.4996.

(157) Chen, F. Intramolecular Triplex Formation of the Purine.Purine.Pyrimidine Type. *Biochemistry* **1991**, *30*, 4472–4479.

(158) Asensio, J. L.; Brown, T.; Lane, A. N. Solution Conformation of a Parallel DNA Triple Helix with 5 ′ and 3 ′ Triplex – Duplex Junctions. *Structure* **1999**, *7*, 1–11.

(159) Bolli, M.; Christopher Litten, J.; Schu¨tz, R.; Leumann, C. J. Bicyclo-DNA: A Hoogsteen-Selective Pairing System. *Chem. Biol.* **1996**, *3*, 197–206. https://doi.org/10.1016/S1074-5521(96)90263-X.

(160) Nikolova, E. N.; Goh, G. B.; Brooks, C. L.; Al-hashimi, H. M. Characterizing the Protonation State of Cytosine in Transient G.C Hoogsteen Base Pairs in Duplex DNA. *J. Am. Chem. Soc.* **2013**, *135*, 6766−6769. https://doi.org/10.1021/ja400994e

(161) Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. Fundamental Aspects of the Nucleic Acid i-motif Structures. *RSC Advances*. 2014, pp 26956–26980. https://doi.org/10.1039/c4ra02129k.

(162) Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R. A.; Patel, D. J. 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* **1998**, *282* (3), 619–636. https://doi.org/10.1006/jmbi.1998.2030.

(163) Liu, C. Q.; Bai, C. L.; Wang, Y.; Huang, J. F.; Wang, S. S.; Zhu, X. Q.; Cao, E. H.; Zhao, J. S. A Molecular Model of Braid-like DNA Structure. *J. Theor. Biol.* **1995**, *177* (4), 411–416. https://doi.org/10.1006/jtbi.1995.0257.

(164) Devi, G.; Zhou, Y.; Zhong, Z.; Toh, D. F. K.; Chen, G. RNA Triplexes: from Structural Principles to Biological and Biotech Applications. *WIREs RNA* **2015**, *6* (1), 111–128. https://doi.org/10.1002/wrna.1261.

(165) Lyamichev, V. I.; Mirkin, S. M.; Frank-Kamenetskii, M. D.; Cantor, C. R. A Stable Complex between Homopyrimidine Oligomers and the Homologous Regions of Duplex DNAs. *Nucleic Acids Res.* **1988**, *16* (5), 2165–2187. https://doi.org/10.1093/nar/16.5.2165.

(166) Reijenga, J.; van Hoof, A.; van Loon, A.; Teunissen, B. Development of Methods for the Determination of pK_a Values. *Anal. Chem. Insights* **2013**, *8*, 53–71. https://doi.org/10.4137/ACI.S12304.

(167) Mirkin, S. M.; Frank-kamenetskiil, M. D. H-DNA AND RELATED STRUCTURES. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 541–576.

(168) Zain, R.; Sun, J. S. Do Natural DNA Triple-Helical Structures Occur and Function *in vivo*? *Cell. Mol. Life Sci.* **2003**, *60* (5), 862–870. https://doi.org/10.1007/s00018-003-3046-3.

(169) Sterns, R. H.; Cox, M.; Feig, P. U.; Singer, I. Internal Potassium Balance and the Control of the Plasma Potassium Concentration. **1981**, *60*, 339–354.

(170) Kalkhoff, R. K.; Yorde, D. E.; Roman, R. J.; Siegesmund, K. A.; Dragen, R. F. Fluctuations of Alpha Cell Calcium, Potassium and Sodium during Amino Acid Perifusion of Rat Pancreatic Islets. *Endocrinology* **1987**, *121* (1), 429–431. https://doi.org/10.1210/endo-121-1-429.

(171) Udensi, U. K.; Tchounwou, P. B. Potassium Homeostasis, Oxidative Stress, and Human Disease. *Int. J. Clin. Exp. Physiol.* **2017**, *4* (3), 111–122.

(172) Maglemose, R.; Hedegaard, A.; Lehnhoff, J.; Dimintiyanova, K. P.; Moldovan, M.; Grøndahl, L.; Meehan, C. F. Potassium Channel Abnormalities are Consistent with Early Axon Degeneration of Motor Axons in the G127X SOD1 Mouse Model of Amyotrophic Lateral Sclerosis. *Exp. Neurol.* **2017**, *292*, 154–167. https://doi.org/10.1016/j.expneurol.2017.03.008.

(173) Mirkin, S. M. DNA Structures , Repeat Expansions and Human Hereditary Disorders. *Curr. Opin. Struct. Biol.* **2006**, *16*, 351–358. https://doi.org/10.1016/j.sbi.2006.05.004.

(174) La Spada, A. R.; Taylor, J. P. Repeat Expansion Disease: Progress and Puzzles in Disease Pathogenesis. *Nat. Rev. Genet.* **2010**, *11* (4), 247–258. https://doi.org/10.1038/nrg2748.

(175) Jain, A.; Vale, R. D. RNA Phase Transitions in Repeat Expansion Disorders. *Nature* **2017**, *546* (7657), 243–247. https://doi.org/10.1038/nature22386.

(176) Bath, J.; Turberfield, A. J. DNA Nanomachines. *Nat. Nanotechnol.* **2007**, *2*, 275–284.

(177) Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F. Assembling Materials with DNA as the Guide. *Science (80-.).* **2008**, *321* (5897), 1795–1799. https://doi.org/10.1126/science.1154533.

(178) Endo, M.; Sugiyama, H. DNA Origami Nanomachines. *Molecules* **2018**, *23* (1766). https://doi.org/10.3390/molecules23071766.

(179) Seeman, N. C. DNA Nanotechnology at 40. *Nano Lett.* **2020**, *20*, 1477–1478. https://doi.org/10.1021/acs.nanolett.0c00325.

(180) Yan, X.; Huang, S.; Wang, Y.; Tang, Y.; Tian, Y. Bottom-up Self-Assembly Based on DNA Nanotechnology. *Nanomaterials* **2020**, *10* (2047). https://doi.org/10.3390/nano10102047.

(181) Jaekel, A.; Lill, P.; Whitelam, S.; Saccà, B. Insights into the Structure and Energy of DNA Nanoassemblies. *Molecules* **2020**, *25* (5446). https://doi.org/10.3390/molecules25235466.

(182) Nicolson, F.; Ali, A.; Kircher, M. F.; Pal, S. DNA Nanostructures and DNA-Functionalized Nanoparticles for Cancer Theranostics. *Adv. Sci.* **2020**, *7* (2001669). https://doi.org/10.1002/advs.202001669.

(183) Zhou, Z.; Fan, D.; Wang, J.; Sohn, Y. S.; Nechushtai, R.; Willner, I. Triggered Dimerization and Trimerization of DNA Tetrahedra for Multiplexed miRNA Detection and Imaging of Cancer Cells. *Small* **2021**, *17* (2007355). https://doi.org/10.1002/smll.202007355.

(184) Bhatia, D.; Wunder, C.; Johannes, L. Self-Assembled, Programmable DNA Nanodevices for Biological and Biomedical Applications. *ChemBioChem* **2021**, *22*, 763–778. https://doi.org/10.1002/cbic.202000372.

(185) Gu, H.; Chao, J.; Xiao, S. J.; Seeman, N. C. A Proximity-Based Programmable DNA Nanoscale Assembly Line. *Nature* **2010**, *465*, 202–205. https://doi.org/10.1038/nature09026.

(186) Gerling, T.; Wagenbauer, K. F.; Neuner, A. M.; Dietz, H. Dynamic DNA Devices and Assemblies Formed by Shape-Complementary, Non–Base Pairing 3D Components. *Science (80-.).* **2015**, *347*, 1446–1452.

(187) Marsh, T. C.; Henderson, E. G-Wires: Self-Assembly of a Telomeric Oligonucleotide, d(GGGGTTGGGG), into Large Superstructures. *Biochemistry* **1994**, *33*, 10718–10724. https://doi.org/10.1021/bi00201a020.

(188) Hannestad, J. K.; Sandin, P.; Albinsson, B. Self-Assembled DNA Photonic Wire. *Nucleic Acids Symposium Series,* **2008**, *52*, 685. https://doi.org/10.1093/nass/nrn346.

(189) Kankia, B. Quadruplex-Based Reactions for Dynamic DNA Nanotechnology. *J. Phys. Chem. B* **2020**, *124*, 4263–4269. https://doi.org/10.1021/acs.jpcb.0c02540.

(190) Zhuravel, R.; Stern, A.; Fardian-Melamed, N.; Eidelshtein, G.; Katrivas, L.; Rotem, D.; Kotlyar, A. B.; Porath, D. Advances in Synthesis and Measurement of Charge Transport in DNA-Based Derivatives. *Adv. Mater.* **2018**, *30*, 1706984. https://doi.org/10.1002/adma.201706984.

(191) Green, L. N.; Subramanian, H. K. K.; Mardanlou, V.; Kim, J.; Hariadi, R. F.; Franco, E. Autonomous Dynamic Control of DNA Nanostructure Self-Assembly. *Nat. Chem.* **2019**, *11*, 510–520. https://doi.org/10.1038/s41557-019-0251-8.

(192) Gao, Y.; Li, Q.; Zhang, J.; Wu, C.; Shen, Z.; Xue, C.; Chang, H. T.; Wu, Z. S. Bead-String-Shaped DNA Nanowires with Intrinsic Structural Advantages and Their Potential for Biomedical Applications. *ACS Appl. Mater. Interfaces* **2020**, *12*, 3341–3353. https://doi.org/10.1021/acsami.9b16249.

(193) Ono, A.; Kanazawa, H.; Ito, H.; Goto, M.; Nakamura, K.; Saneyoshi, H.; Kondo, J. A Novel DNA Helical Wire Containing Hg^{II}-Mediated T:T and T:G Pairs. *Angew. Chemie - Int. Ed.* **2019**, *58*, 16835–16838. https://doi.org/10.1002/anie.201910029.

(194) Kondo, J.; Tada, Y.; Dairaku, T.; Hattori, Y.; Saneyoshi, H.; Ono, A.; Tanaka, Y. A Metallo-DNA Nanowire with Uninterrupted One-Dimensional Silver Array. *Nat. Chem.* **2017**, *9*, 956–960. https://doi.org/10.1038/nchem.2808.

(195) Aich, P.; Labiuk, S. L.; Tari, L. W.; Delbaere, L. J. T.; Roesler, W. J.; Falk, K. J.; Steer, R. P.; Lee, J. S. M-DNA: A Complex between Divalent Metal Ions and DNA Which Behaves as a Molecular Wire. *J. Mol. Biol.* **1999**, *294*, 477–485. https://doi.org/10.1006/jmbi.1999.3234.

(196) Houlton, A.; Pike, A. R.; Angel Galindo, M.; Horrocks, B. R. DNA-Based Routes to Semiconducting Nanomaterials. *Chem. Commun.* **2009**, No. 14, 1797–1806. https://doi.org/10.1039/b818456a.

(197) Livshits, G. I.; Stern, A.; Rotem, D.; Borovok, N.; Eidelshtein, G.; Migliore, A.; Penzo, E.; Wind, S. J.; Di Felice, R.; Skourtis, S. S.; Cuevas, J. C.; Gurevich, L.; Kotlyar, A. B.; Porath, D. Long-Range Charge Transport in Single G-Quadruplex DNA Molecules. *Nat. Nanotechnol.* **2014**, *9* (12), 1040–1046. https://doi.org/10.1038/nnano.2014.246.

(198) Nagashima, S.; Ha, H. D.; Kim, D. H.; Košmrlj, A.; Stone, H. A.; Moon, M. W. Spontaneous Formation of Aligned DNA Nanowires by Capillarity-Induced Skin Folding. *PNAS* **2017**, *114* (24), 6233–6237. https://doi.org/10.1073/pnas.1700003114.

(199) Mergny, J. L.; Sen, D. DNA Quadruple Helices in Nanotechnology. *Chem. Rev.* **2019**, *119* (10), 6290–6325. https://doi.org/10.1021/acs.chemrev.8b00629.

(200) Gowers, D. M.; Fox, K. R. Towards Mixed Sequence Recognition by Triple Helix Formation. *Nucleic Acids Res.* **1999**, *27* (7), 1569–1577. https://doi.org/10.1093/nar/27.7.1569.

(201) Kuipers, J.; Giepmans, B. N. G. Neodymium as an Alternative Contrast for Uranium in Electron Microscopy. *Histochem. Cell Biol.* **2020**, *153* (4), 271–277. https://doi.org/10.1007/s00418-020-01846-0.

(202) Hu, Y.; Lin, F.; Wu, T.; Wang, Y.; Zhou, X. S.; Shao, Y. Fluorescently Sensing of DNA Triplex Assembly Using an Isoquinoline Alkaloid as Selector, Stabilizer, Inducer, and Switch-On Emitter. *Chem. - An Asian J.* **2016**, *11* (14), 2041–2048. https://doi.org/10.1002/asia.201600459.

(203) Vesenka, J.; Marsh, T.; Henderson, E.; Vellandi, C. The Diameter of Duplex and Quadruplex DNA Measured by Scanning Probe Microscopy. *Scanning Microsc.* **1998**, *12* (2), 329–342.

(204) Bose, K.; Lech, C. J.; Heddi, B.; Phan, A. T. High-Resolution AFM Structure of DNA G-Wires in Aqueous Solution. *Nat. Commun.* **2018**, *9* (1). https://doi.org/10.1038/s41467-018-04016-y.

(205) Miyoshi, D.; Karimata, H.; Wang, Z. M.; Koumoto, K.; Sugimoto, N. Artificial G-Wire Switch with 2,2′-Bipyridine Units Responsive to Divalent Metal Ions. *J. Am. Chem. Soc.* **2007**, *129* (18), 5919–5925. https://doi.org/10.1021/ja068707u.

(206) Usui, K.; Okada, A.; Sakashita, S.; Shimooka, M.; Tsuruoka, T.; Nakano, S. I.; Miyoshi, D.; Mashima, T.; Katahira, M.; Hamada, Y. DNA G-Wire Formation Using an Artificial Peptide is Controlled by Protease Activity. *Molecules* **2017**, *22* (1991), 1–12. https://doi.org/10.3390/molecules22111991.

(207) Ni, Z.; Ye, T.; Yu, Y.; Gao, L.; Fei, Y.; Li, Q.; Zhou, Y.; Shao, Y.; Zeng, L. Triplex-Forming Oligonucleotide as a Lighting-up Switch for a DNA Abasic Site-Binding Fluorescent Ligand. *J. Lumin.* **2018**, *198* (2018), 193–197. https://doi.org/10.1016/j.jlumin.2018.02.011.

(208) Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. DNA G-Quadruplexes in the Human Genome: Detection, Functions and Therapeutic Potential. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (5), 279–284. https://doi.org/10.1038/nrm.2017.3.

(209) Rouleau, S.; Jodoin, R.; Garant, J.-M.; Perreault, J.-P. RNA G-Quadruplexes as Key Motifs of the Transcriptome. *Adv. Biochem. Eng. Biotechnol.* **2017**, 170, 1–20. https://doi.org/10.1007/10_2017_8.

(210) Assi, H. A.; Garavís, M.; González, C.; Damha, M. J. i-Motif DNA: Structural Features and Significance to Cell Biology. *Nucleic Acids Res.* **2018**, *46* (16), 8038–8056. https://doi.org/10.1093/nar/gky735.

(211) Crossley, M. P.; Boceka, M.; Cimprich, K. A. R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell* **2019**, *73* (3), 398–411. https://doi.org/10.1016/j.molcel.2019.01.024.

(212) Lat, P. K.; Sen, D. $(C_2G_4)_n$ Repeat Expansion Sequences from the *C9orf72* Gene Form an Unusual DNA Higher-Order Structure in the pH Range of 5-6. *PLoS One* **2018**, *13* (6), 1–20. https://doi.org/10.1371/journal.pone.0198418.

(213) Rhodes, D.; Lipps, H. J. Survey and Summary G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Res.* **2015**, *43* (18), 8627–8637. https://doi.org/10.1093/nar/gkv862.

(214) Hänsel-Hertsch, R.; Spiegel, J.; Marsico, G.; Tannahill, D.; Balasubramanian, S. Genome-Wide Mapping of Endogenous G-Quadruplex DNA Structures by Chromatin Immunoprecipitation and High-Throughput Sequencing. *Nat. Protoc.* **2018**, *13* (3), 551–564. https://doi.org/10.1038/nprot.2017.150.

(215) Müller, S.; Kumari, S.; Rodriguez, R.; Balasubramanian, S. Small-Molecule-Mediated G-Quadruplex Isolation from Human Cells. **2010**, *2* (12), 1095–1098. https://doi.org/10.1038/nchem.842.

(216) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. Small-Molecule-Induced DNA Damage Identifies Alternative DNA Structures in Human Genes. *Nat. Chem. Biol.* **2012**, *8*, 301–310. https://doi.org/10.1038/nchembio.780.

(217) Renard, I.; Grandmougin, M.; Roux, A.; Yang, S. Y.; Lejault, P.; Pirrotta, M.; Wong, J. M. Y.; Monchaud, D. Small-Molecule Affinity Capture of DNA/RNA Quadruplexes and their Identification *in vitro* and *in vivo* through the G4RP Protocol. *Nucleic Acids Res.* **2019**, *47* (11), 5502–5510. https://doi.org/10.1093/nar/gkz215.

(218) Gray, L. T.; Puig Lombardi, E.; Verga, D.; Nicolas, A.; Teulade-Fichou, M. P.; Londoño-Vallejo, A.; Maizels, N. G-Quadruplexes Sequester Free Heme in Living Cells. *Cell Chem. Biol.* **2019**, *26* (12), 1681–1691. https://doi.org/10.1016/j.chembiol.2019.10.003.

(219) Einarson, O. J.; Sen, D. Self-Biotinylation of DNA G-Quadruplexes via Intrinsic Peroxidase Activity. *Nucleic Acids Res.* **2017**, *45* (17), 9813–9822. https://doi.org/10.1093/nar/gkx765.

(220) Kaewsapsak, P.; Shechner, D. M.; Mallard, W.; Rinn, J. L.; Ting, A. Y. Live-Cell Mapping of Organelle-Associated RNAs via Proximity Biotinylation Combined with Protein-RNA Crosslinking. *Elife* **2017**, *6* (e29224), 1–31. https://doi.org/10.1101/153098.

(221) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

(222) Dobin, A.; Davis, C. A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T. R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29* (1), 15–21. https://doi.org/10.1093/bioinformatics/bts635.

(223) Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S. L. Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biol.* **2009**, *10* (3). https://doi.org/10.1186/gb-2009-10-3-r25.

(224) Bushmanova, E.; Antipov, D.; Lapidus, A.; Prjibelski, A. D. RnaSPAdes: A de Novo Transcriptome Assembler and its Application to RNA-Seq Data. *Gigascience* **2019**, *8* (9), 1–13. https://doi.org/10.1093/gigascience/giz100.

(225) Quinlan, A. R.; Hall, I. M. BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* **2010**, *26* (6), 841–842. https://doi.org/10.1093/bioinformatics/btq033.

(226) Liao, Y.; Smyth, G. K.; Shi, W. FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30* (7), 923–930. https://doi.org/10.1093/bioinformatics/btt656.

(227) Li, W.; Zeng, W.; Chen, Y.; Wang, F.; Wu, F.; Weng, X.; Zhou, X. Biotinylation and Isolation of an RNA G-Quadruplex Based on its Peroxidase-Mimicking Activity. *Analyst* **2019**, *144* (15), 4472–4476. https://doi.org/10.1039/c9an00353c.

(228) Fried, M. G.; Bloomfield, V. A. DNA Gelation in Concentrated Solutions. **1984**, *23* (11), 2141–2155.

(229) Siebenlist, U.; Gilbert, W. Contacts between Escherichia Coli RNA Polymerase and an Early Promoter of Phage T7. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77* (1), 122–126. https://doi.org/10.1073/pnas.77.1.122.

(230) Heddi, B.; Cheong, V. V.; Martadinata, H.; Phan, A. T. Insights into G-Quadruplex Specific Recognition by the DEAH-Box Helicase RHAU: Solution Structure of a Peptide-Quadruplex Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (31), 9608–9613. https://doi.org/10.1073/pnas.1422605112.

(231) Masuzawa, T.; Sato, S.; Niwa, T.; Taguchi, H.; Nakamura, H.; Oyoshi, T. G-Quadruplex-Proximity Protein Labeling Based on Peroxidase Activity. *Chem. Commun.* **2020**, *56* (78), 11641–11644. https://doi.org/10.1039/d0cc02571b.

(232) Rhee, H.-W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K.; Carr, S. A.; Ting, A. Y. Proteomic Mapping of Mitochondria in Living Cells via Spatially-Restricted Enzymatic Tagging. *Science (80-.).* **2013**, *339* (6125), 1328–1331. https://doi.org/10.1126/science.1230593.

(233) Hung, V.; Udeshi, N. D.; Lam, S. S.; Loh, K. H.; Cox, K. J.; Pedram, K.; Carr, S. A.; Ting, A. Y. Spatially Resolved Proteomic Mapping in Living Cells with the Engineered Peroxidase APEX2. *Nat. Protoc.* **2016**, *11* (3), 456–475. https://doi.org/10.1038/nprot.2016.018.

(234) Mestre-Fos, S.; Ito, C.; Moore, C. M.; Reddi, A. R.; Williams, L. D. Human Ribosomal G-Quadruplexes Regulate Heme Bioavailability. *J. Biol. Chem.* **2020**, *295* (44), 14855–14865. https://doi.org/10.1074/jbc.RA120.014332.

(235) Mestre-Fos, S.; Penev, P. I.; Suttapitugsakul, S.; Hu, M.; Ito, C.; Petrov, A. S.; Wartell, R. M.; Wu, R.; Williams, L. D. G-Quadruplexes in Human Ribosomal RNA. *J. Mol. Biol.* **2019**, *431* (10), 1940–1955. https://doi.org/10.1016/j.jmb.2019.03.010.

(236) Wang, W.; Hu, S.; Gu, Y.; Yan, Y.; Stovall, D. B.; Li, D.; Sui, G. Human MYC G-Quadruplex: From Discovery to a Cancer Therapeutic Target. *Biochim Biophys Acta Rev Cancer* **2020**, *1874* (2), 188410. https://doi.org/10.1016/j.bbcan.2020.188410.

(237) Kumarasamy, V. M.; Sun, D. Demonstration of a Potent RET Transcriptional Inhibitor for the Treatment of Medullary Thyroid Carcinoma Based on an Ellipticine Derivative. *Int. J. Oncol.* **2017**, *51* (1), 145–157. https://doi.org/10.3892/ijo.2017.3994.

(238) Donegan, R. K.; Moore, C. M.; Hanna, D. A.; Reddi, A. R. Handling Heme: The Mechanisms Underlying the Movement of Heme within and between Cells. *Free Radic. Biol. Med.* **2019**, *133*, 88–100. https://doi.org/10.1016/j.freeradbiomed.2018.08.005.

(239) Atamna, H.; Brahmbhatt, M.; Atamna, W.; Shanower, G. A.; Dhahbi, J. M. ApoHRP-Based Assay to Measure Intracellular Regulatory Heme. *Metallomics* **2015**, *7* (2), 309–321. https://doi.org/10.1039/c4mt00246f.

(240) Hanna, D. A.; Harvey, R. M.; Martinez-Guzman, O.; Yuan, X.; Chandrasekharan, B.; Raju, G.; Outten, F. W.; Hamza, I.; Reddi, A. R. Heme Dynamics and Trafficking Factors Revealed by Genetically Encoded Fluorescent Heme Sensors. *PNAS* **2016**, *113* (27), 7539–7544. https://doi.org/10.1073/pnas.1523802113.

(241) Song, Y.; Yang, M.; Wegner, S. V.; Zhao, J.; Zhu, R.; Wu, Y.; He, C.; Chen, P. R. A Genetically Encoded FRET Sensor for Intracellular Heme. *ACS Chem. Biol.* **2015**, *10*, 1610–1615. https://doi.org/10.1021/cb5009734.

(242) Canesin, G.; Janovicova, L.; Wegiel, B. Measurement of Labile and Protein-Bound Heme in Fixed Prostate Cancer Cells and in Cellular Fractions. *STAR Protoc.* **2021**, *2*, 100491. https://doi.org/10.1016/j.xpro.2021.100491.

(243) Modi, S.; Swetha, M. G.; Goswami, D.; Gupta, G. D.; Mayor, S.; Krishnan, Y. A DNA Nanomachine that Maps Spatial and Temporal pH Changes Inside Living Cells. *Nat. Nanotechnol.* **2009**, *4*, 325–330. https://doi.org/10.1038/nnano.2009.83.

(244) Surana, S.; Bhat, J. M.; Koushika, S. P.; Krishnan, Y. An Autonomous DNA Nanomachine Maps Spatiotemporal pH Changes in a Multicellular Living Organism. *Nat. Commun.* **2011**, *2* (340), 1–7. https://doi.org/10.1038/ncomms1340.

(245) Webb, B. A.; Chimenti, M.; Jacobson, M. P.; Barber, D. L. Dysregulated pH: a Perfect Storm for Cancer Progression. *Nat. Rev. Cancer* **2011**, *11*, 671–677. https://doi.org/10.1038/nrc3110.

(246) Ma, D. L.; Zhang, Z.; Wang, M.; Lu, L.; Zhong, H. J.; Leung, C. H. Recent Developments in G-Quadruplex Probes. *Chem. Biol.* **2015**, *22* (7), 812–828. https://doi.org/10.1016/j.chembiol.2015.06.016.

(247) Canesin, G.; Di Ruscio, A.; Li, M.; Ummarino, S.; Hedblom, A.; Choudhury, R.; Krzyzanowska, A.; Csizmadia, E.; Palominos, M.; Stiehm, A.; Ebralidze, A.; Chen, S. Y.; Bassal, M. A.; Zhao, P.; Tolosano, E.; Hurley, L.; Bjartell, A.; Tenen, D. G.; Wegiel, B. Scavenging of Labile Heme by Hemopexin is a Key Checkpoint in Cancer Growth and Metastases. *Cell Rep.* **2020**, *32* (12), 108181. https://doi.org/10.1016/j.celrep.2020.108181.

(248) Emamjomeh, A.; Choobineh, D.; Hajieghrari, B.; MahdiNezhad, N.; Khodavirdipour, A. DNA–Protein Interaction: Identification, Prediction and Data Analysis. *Mol. Biol. Rep.* **2019**, *46* (3), 3571–3596. https://doi.org/10.1007/s11033-019-04763-1.

(249) Liu, S.; Li, B.; Liang, Q.; Liu, A.; Qu, L.; Yang, J. Classification and Function of RNA–Protein Interactions. *Wiley Interdiscip. Rev. RNA* **2020**, *11* (6), 1–27. https://doi.org/10.1002/wrna.1601.

(250) Go, C. D.; Knight, J. D. R.; Rajasekharan, A.; Rathod, B.; Hesketh, G. G.; Abe, K. T.; Youn, J.-Y.; Samavarchi-Tehrani, P.; Zhang, H.; Zhu, L. Y.; Popiel, E.; Lambert, J.-P.; Coyaud, É.; Cheung, S. W. T.; Rajendran, D.; Wong, C. J.; Antonicka, H.; Pelletier, L.; Palazzo, A. F.; Shoubridge, E. A.; Raught, B.; Gingras, A.-C. A Proximity-Dependent Biotinylation Map of a Human Cell. *Nature* **2021**, 595, 120-124. https://doi.org/10.1038/s41586-021-03592-2.

(251) Zhang, X.; Spiegel, J.; Cuesta, S. M.; Adhikari, S.; Balasubramanian, S. Chemical Profiling of DNA G-Quadruplex-Interacting Proteins in Live Cells. *Nat. Chem.* **2021**, *13*, 626–633. https://doi.org/https://doi.org/10.1038/s41557-021-00736-9.

Appendix.

Sequencing Data File

Excel file: "Drosophila_G4_RNAseq.xlsx" (Attached)

This excel sheet consists of 4 sheets. Their details are as follows:

1.) Sheet name: **'Readme'** gives gene information.

2.) Sheet name: **'Read_distribution'** gives % read distribution of rRNA and non-rRNA transcripts in our sequencing data.

3.) Sheet name: **'non-rRNA (Heme-GQ)'** summarizes the total non-rRNA transcripts/hits that were enriched using our 'Heme-GQ' biotinylation assay.

4.) Sheet name: **'common with stop pol'** enlists a set of genes that were enriched in our sequencing data and were also in common with the ones reported to form GQ in *Drosophila* by Shankar Balasubramanian group[67] (using polymerase stop assay).