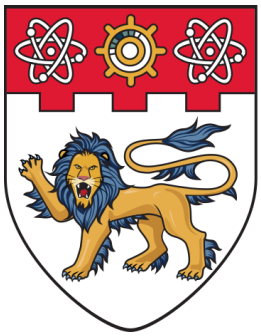


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

DEVELOPMENT AND APPLICATION OF NMR SPECTROSCOPY IN G-QUADRUPLEXES: STRUCTURE, DYNAMICS AND FORMATION IN BIOFILM

FERNALDO RICHTIA WINNERDY

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2019

DEVELOPMENT AND APPLICATION OF NMR SPECTROSCOPY IN G-QUADRUPLEXES: STRUCTURE, DYNAMICS AND FORMATION IN BIOFILM

FERNALDO RICHTIA WINNERDY

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A Thesis Submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Doctor of Philosophy

2019

Statement of Originality

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

14 February 2019

Date

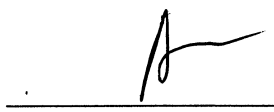
Fernaldo Richtia Winnerdy

Supervisor Declaration Statement

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

14 February 2019

Date



Anh Tuân Phan

Authorship Attribution Statement

This thesis contains material from 5 papers which are published (1), under review (1), or prepared for submission (3) in the following peer-reviewed journal where I was the first/co-first author:

1. Material from section 2.3 is published as Adrian M., **Winnerdy F.R.**, Heddi B. and Phan A.T. Rotation of guanine amino groups in G-quadruplexes: A probe for local structure and ligand binding. *Biophys. J.*, **2017**, 113 (4), 757-758. DOI: 10.1016/j.bpj.2017.05.053. **(M.A. and F.R.W. contributed equally)**.

The contributions of the co-authors are as follows:

- Dr. Adrian initiated the project, performed some NMR experiments and wrote the earlier version of the manuscript.
- I continued the NMR experiments, compiled the data, constructed the set of figures, and continued developing the manuscript.
- Dr. Heddi and Prof. Phan revised the manuscript, suggested some additional experimets and finalized the submitted version.

2. Material from section 2.4 is prepared as **Winnerdy F.R.**, Do, T.H.H, Heddi B., Pratviel, G. and Phan A.T. Determination of residence time of G-quadruplex-binding ligands with NMR exchange analysis. Manuscript is prepared for submission to *Chemical Science*.

The contributions of the co-authors are as follows:

- Ms. Do and Dr. Heddi initiated the project and performed some preliminary experiments.
- I performed all the experiments based on the preliminary data, prepared the set of figures and the manuscript.
- Dr. Pratviel provided the G4-binding ligands
- Prof. Phan designed and suggested some experimets as well as discussed the results together with me.

3. Materials from section 3.2 is prepared as **Winnerdy F.R.**, Heddi B. and Phan A.T. Solution structures of linear- and cyclic-dinucleotide (cGAMP) bound to a G-quadruplex. Manuscript is prepared for submission to *Journal of the American Chemical Society*.

The contributions of the co-authors are as follows:

- I begin the project, performed all experiments and structural calculations as well as prepared the set of figures and manuscript.
- Dr. Heddi helped in structure calculation of cyclic-dinucleotide complex
- Prof. Phan provided some direction and revised the manuscript.

4. Materials from section 3.3 is under review as Truong T.H.A, **Winnerdy F.R.** and Phan A.T. Discovery of an unprecedented G-quadruplex peripheral knot. Manuscript is under review in *Angewandte Chemie* (2019). **(T.T.H.A. and F.R.W. contributed equally)**.

The contributions of the co-authors are as follows:

- Dr. Truong initiated the project and performed the NMR experiments.
- I designed some NMR experiments, conducted the structure calculation and prepared the set of figures and manuscript.
- Prof. Phan provided some direction and revised the manuscript.

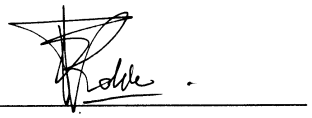
5. Materials from section 3.4 is prepared as **Winnerdy F.R.**, How J.Q. and Phan A.T. Left- and right-handed G-quadruplex hybrid structure of GGT_8 trinucleotide repeats. Manuscript is prepared for submission to *Nucleic Acids Research.*

The contributions of the co-authors are as follows:

- I designed the NMR experiments, conducted the structure calculation and prepared the set of figures and manuscript.
- Mr. How performed the DNA synthesis and NMR experiments.
- Prof. Phan provided some direction and revised the manuscript.

14 February 2019

Date



Fernaldo Richtia Winnerdy

Abstract

G-quadruplex is a secondary non-canonical structure of nucleic acid which gained popularity due to increasing evidence of its biological functions, potential therapeutic prospects and nanotechnology applications. Encouraged for a better comprehension of G-quadruplex as an alternative DNA structures, we address multiple angles of G-quadruplex properties including its dynamics, ligand binding kinetics, numerous novel topologies and potential role in biofilm networking. This research was accomplished through a combination of biophysical techniques, primarily Nuclear Magnetic Resonance (NMR) spectroscopy.

In this dissertation, G-quadruplex conformational and ligand-interaction dynamics are investigated with NMR spectroscopy. In particular, we developed two specific NMR methodologies for the measurements of the rotation rates of amino protons in guanines of G-quadruplexes via lineshape analysis and the determination of the residence time of ligands bound to G-quadruplex receptors via NOESY exchange analysis.

Applications of NMR spectroscopy in the investigation of novel G-quadruplex folding topologies and the analysis of G-quadruplex formation in biofilm are demonstrated. Multiple high-resolution structures are resolved using NMR spectroscopy: (1) Complex structures of 11-guanine G-quadruplex (*D4*) bound to cell-signaling metabolite cyclic guanosine-adenosine monophosphate (cGAMP), (2) a fold-back G-quadruplex structure featuring a triad and base-pair formation on an anti-parallel G-quadruplex core (*AT21S*), and (3) a left and right-handed G-quadruplex hybrid formation from multiple trinucleotide repeats $(GGT)_8$. Additionally, we determined that extracellular nucleic acid is the main contributor towards polymer networking property of P. *aeruginosa* biofilm (extracellular matrix). We proved the existence of non-canonical RNA-DNA complexes and potential formation of G-quadruplexes are responsible for the gel-networking characteristic of biofilms.

The methodologies presented in this thesis — lineshape analysis and NOESY exchange analysis — contribute as effective options in measuring G-quadruplex dynamics. Furthermore, the applications of NMR spectroscopy explored here provide insights on: (1) Vacancy-bearing G-quadruplex as metabolite binder, (2) fundamentals of G-quadruplex structural elements of fold-back diagonal loops, (3) expansion of G-quadruplex diversity based on the existence of left-handed right-handed junction, and (4) potential biological importance of G-quadruplexes in biofilm network.

Acknowledgements

I would like to express my sincere gratitude towards my supervisor **Prof. Anh Tuân Phan**, for giving me the opportunity to work in his group. I am grateful for his constant guidance, motivation, and patience throughout the course of my research.

I thank all collaborators involved in the projects presented in this thesis, **Dr. Michael Adrian**, **Dr. Truong Thi Hong Anh**, **Mr. Blaz Bakalar**, **Dr. Thomas Seviour** and **Dr. Shi Xiangyan** for their invaluable contributions towards the success of the projects.

I would like to express my gratefulness towards **Dr. Brahim Heddi** for introducing the world of scientific research in an amazing manner, **Dr. Kah Wai Lim** for his advices during difficulty in various problems, and **Dr. Vee Vee Cheong** for her enourmous aid inside and outside of the laboratory.

I am thankful to the lab manager **Ms Nguyen Thi Anh Kim** and my undergraduate student **Mr. How Jia Qi** for helping me in the best possible way. My appreciation extends to the past and present members of biophysics lab. I thank

them for all the help and valuable feedback during the 4-year period.

I want to convey my thanks to all friends for their pleasant company and the memories, especially my girlfriend **Ms. Poulomi Das**. I am grateful to **Dr. Anjali Sengar, Mr. Arijit Maity, Ms. Jeya Vandana** and **Mr. Nicholas Chan**, as well as **Mr. David Giovanni** and **Ms. Devina Wongso**, without whom I will not be able to greatly relish my time as a graduate student.

I would like to avail this opportunity to express my heartiest sense of reverence and love to all my family members, especially my Dad, Mom, and Sister, for their everlasting encouragement, blessings and love.

Abbreviation

1D	One dimensional
2D	Two dimensional
A	Adenine
AMP	Adenine monophosphate
C	Cytosine
CD	Circular dichroism
CMP	Cytosine monophosphate
COSY	Correlation spectroscopy
DMAc	Dimethyl acetamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSS	Dimethyl silapentane sulfonic acid
DTT	Dithiolthreitol
eDNA	Extracellular DNA
EDTA	Ethylene diamine tetra-acetic Acid
EM	Electron microscopy

EMIM-Ac	Ethyl methylimidazolium acetate
G	Guanine
G4	G-quadruplex
GMP	Guanine monophosphate
GTP	Guanine triphosphate
HETCOR	Heteronuclear correlation
HMBC	Heteronuclear multiple bond correlation
HSQC	Hetronuclear single quantum correlation
IDT	Integrated DNA Technologies
LB	Luria-Bertani
MAS	Magic-angle spinning
MWCO	Molecular weight cut-off
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
nt	nucleotide
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
pG4	Putative G-quadruplex
ppm	Part per million
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
SS-NMR	Solid state NMR
T	Thymine
TOCSY	Total correlation spectroscopy
U	Uracil
UMP	Uracil monophosphate
UV	Ultraviolet
WC	Watson-Crick

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 G-quadruplex nucleic acid

Deoxyribonucleic acid (DNA) serves as the fundamental molecule which holds all hereditary information in most living cells [1]. It is a polymer constituted by nucleotides as its building blocks. A nucleotide unit contains three main components: a phosphate group, a pentose sugar ring, and four types of nitrogenous bases or nucleobases (Figure 1.1A). The singular primary structure of DNA is a linear chain of nucleotides starting from the 5' hydroxy group of the first nucleotide (5'-end) and ending with the 3' hydroxy group of the last nucleotide (3'-end). This chain forms the sugar-phosphate backbone, with the nucleobases branching out of it (Figure 1.1B). In contrast, there exist several DNA secondary structures, of which the double helix (duplex) DNA (Figure 1.1C) is most commonly observed [2]. In this structure, two strands of DNA align in an anti-parallel orientation, stabilized by Watson-Crick base-pairing between nucleobases. There are two kinds of nucleobases pairings, an adenine-thymine pair and a guanine-cytosine pair, with two and three hydrogen bonds stabilizing the respective pairs.

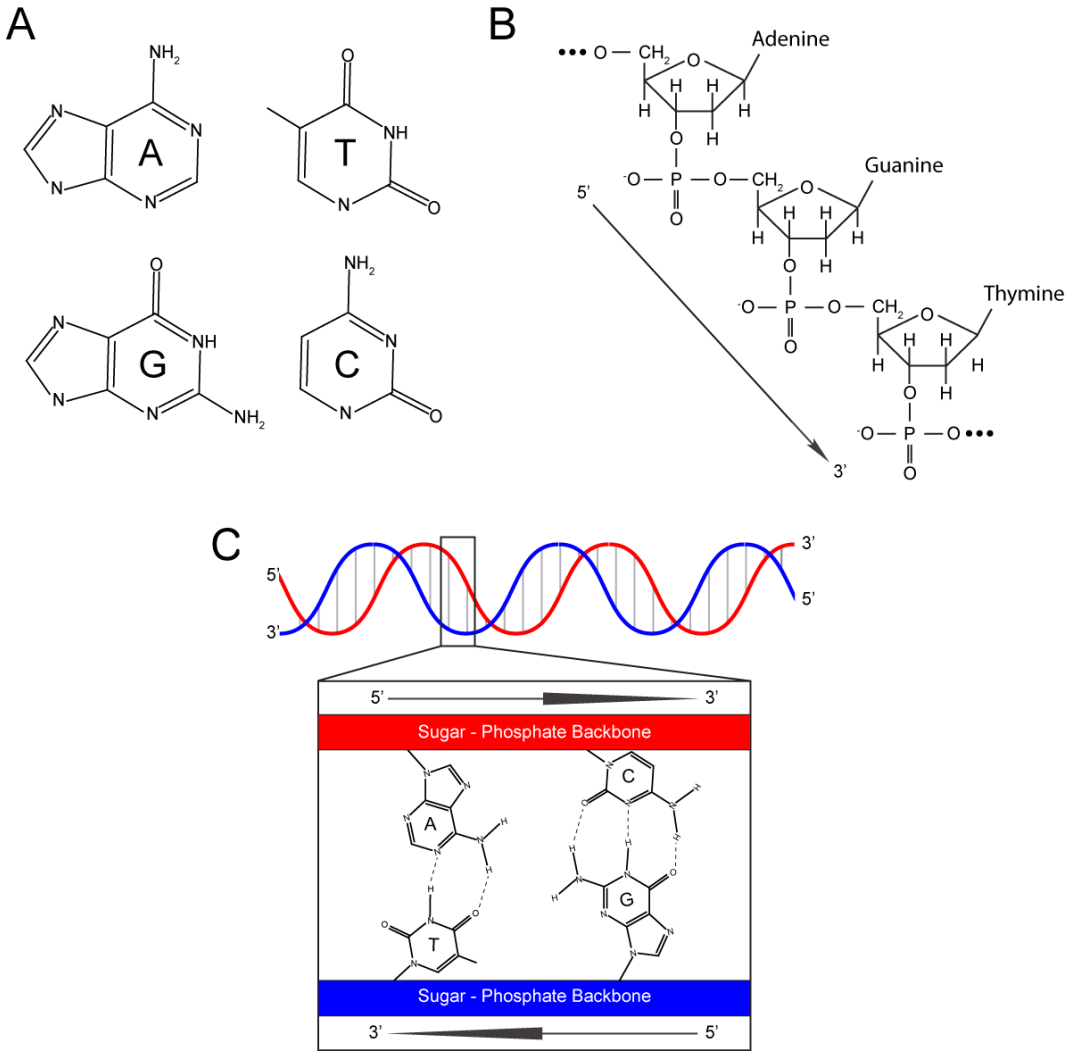


Figure 1.1: Morphology of deoxyribonucleic acid (DNA). A) Four types of nucleobases, adenine, thymine, guanine, and cytosine. B) DNA sugar-phosphate backbone progression from 5' to 3'. C) Schematic structure of DNA double helix, with inset containing the illustration of adenine-thymine and guanine-cytosine base pairs.

In guanine-rich oligonucleotides, four guanine bases can adopt a stable planar conformation known as a G-tetrad. Within the G-tetrad, each guanine base forms two hydrogen bonds to each of its neighboring guanines (Figure 1.2A) [3]. Specifically, Hoogsteen hydrogen bonds are formed between H1 and H21 of one guanine residue to O6 and N7 of the neighboring guanine respectively, with a total of eight hydrogen bonds stabilizing the G-tetrad configuration [4, 5]. These G-tetrads are further stabilized by a certain type of cation (such as Na^+, K^+, or $NH_4{}^+$) located

in the middle of the G-tetrad (Figure 1.2A), depending on their relative hydration free energies [6]. Consecutive layers of G-tetrads are stacked on top of each other, forming an alternative DNA secondary structure known as a G-quadruplex or G4 (Figure 1.2B) [7, 8].

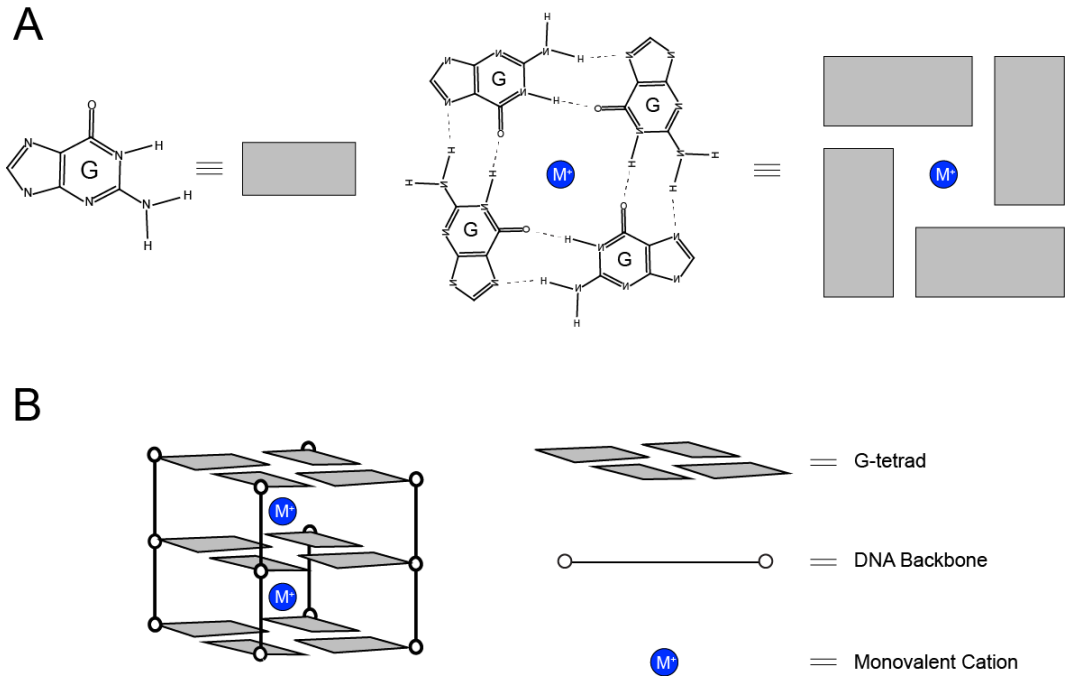


Figure 1.2: Building blocks of G-quadruplex. A) The representative of a guanine residue (left) and a G-tetrad (right). Single G-tetrad consists of four guanine residues with cyclic Hoogsteen hydrogen bonds between them, stabilized by one monovalent cation in the middle. B) Schematic of a three-layered G-quadruplex structure (left) and the symbols legend (right).

The G-quadruplex structure is diverse, and its topologies are influenced by structural elements such as strand orientations and loop arrangements as well as the differences in glycosidic bond angles and the number of molecules that constitute the structure (Figure 1.3) [9]. Based on its relative strand orientations, a parallel G-quadruplex exhibits unidirectional strand orientation, while anti-parallel or a hybrid (3+1) G-quadruplex possesses one or more strands with opposing strand orientations. Nucleotides connecting guanine tracts are termed as loops, and the loops can connect adjacent parallel strands (propeller loop), adjacent anti-parallel

strands (edgewise loop), diagonal anti-parallel strands (diagonal loop), or even forming additional structures such as a hairpin. As illustrated in Figure 1.3 (top), the nucleobase can rotate about its glycosidic bond to face towards or away from the ribose 5' end, resulting in a *syn* or *anti* configuration respectively. G-quadruplex can also be formed by a single molecule (monomeric), two molecules (dimeric), or multiple molecules (multimeric) [10, 11]. Recently, it was also found that G-quadruplex can adopt a left-handed strand progression as opposed to its usual right-handed progression [12].

G-quadruplex is polymorphic in nature, whereby the same oligonucleotide sequence can adopt different G-quadruplex structures under different sample environment. External factors affecting the G-quadruplex structure includes the type of cations, concentration of salts and the presence of crowding agents [13]. The effect of G-quadruplex polymorphism is illustrated in a well-studied G-quadruplex forming sequence, the human telomeric repeat $d(TTAGGG)_n$, where multiple folding topologies are adopted by these G-quadruplex forming sequences in different cations (Na^+ and K^+), and in crowded condition [14–17]. Most structural studies are performed only in the conditions where there is a single major conformation of G-quadruplex structure. In order to exert control over the type of G-quadruplex assembled, factors influencing the folding topology such as the buffer condition and the strand concentration need to be well-studied.

1.2 Biological importance of G-quadruplexes

Bioinformatic studies have shown that there are more than 350,000 potential G-quadruplex (pG4) forming sequences in human genome [18]. However, it was recently suggested that this number is an underestimation due to the exclusion of long-repetitive DNA sequences in the database [19, 20]. Moreover, the discovery

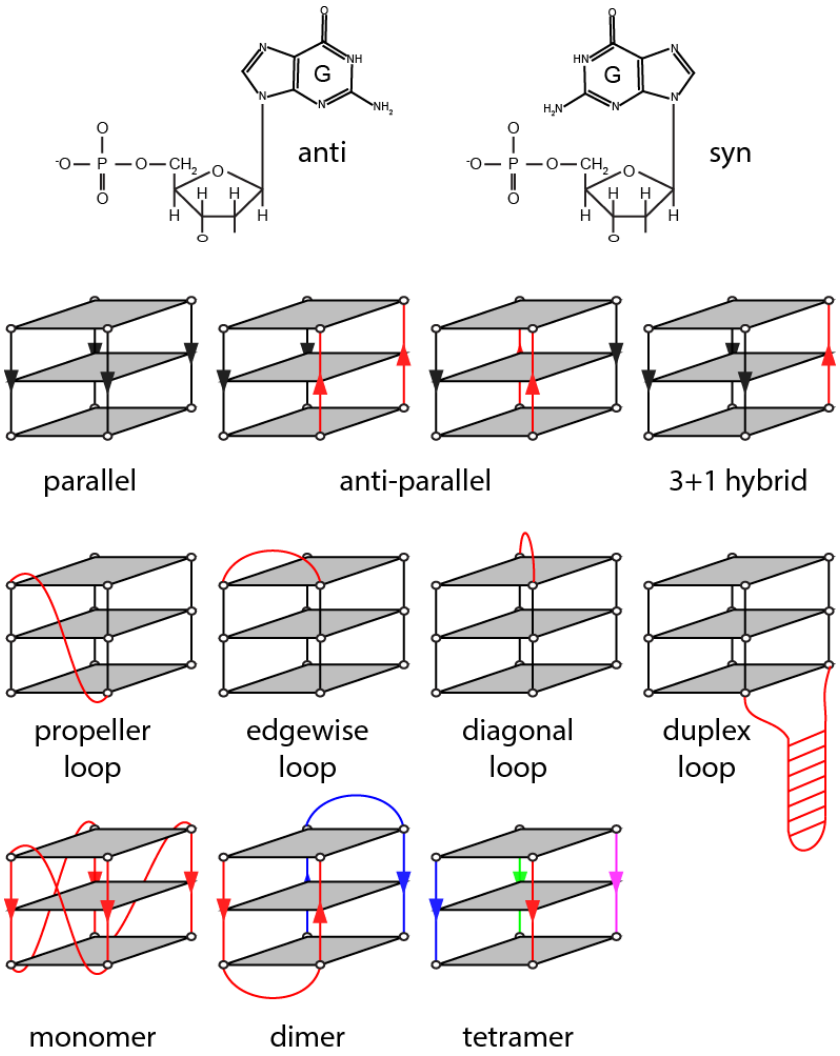


Figure 1.3: Diversity of G-quadruplex topology. (Top) Two different possible glycosidic bond orientations of G-tetrad guanine, *anti* and *syn*. (Middle top) Schematics of parallel, anti-parallel, and (3+1) hybrid G-quadruplexes, black and red arrows represent backbone progression in different orientations (Middle bottom) Schematics of different loops namely propeller, edgewise, diagonal and hairpin loops. (Bottom) Schematics of monomeric, dimeric, and tetrameric G-quadruplexes, different DNA strands are colored differently.

of more G-quadruplex structures forming *in vitro* from unprecedented DNA sequences, such as bulges and long loops also adds to the number of pG4 [21–24]. pG4 is evolutionarily conserved in different species [25–30], and localized with higher density in functional places such as the telomeric regions, gene promoters [31], DNA replication origins [32–34] and 5′-untranslated region of RNA [35, 36]. These

facts implicate pG4s to have some regulatory roles *in vivo*, most probably through G-quadruplex structure formation [20] (Figure 1.4).

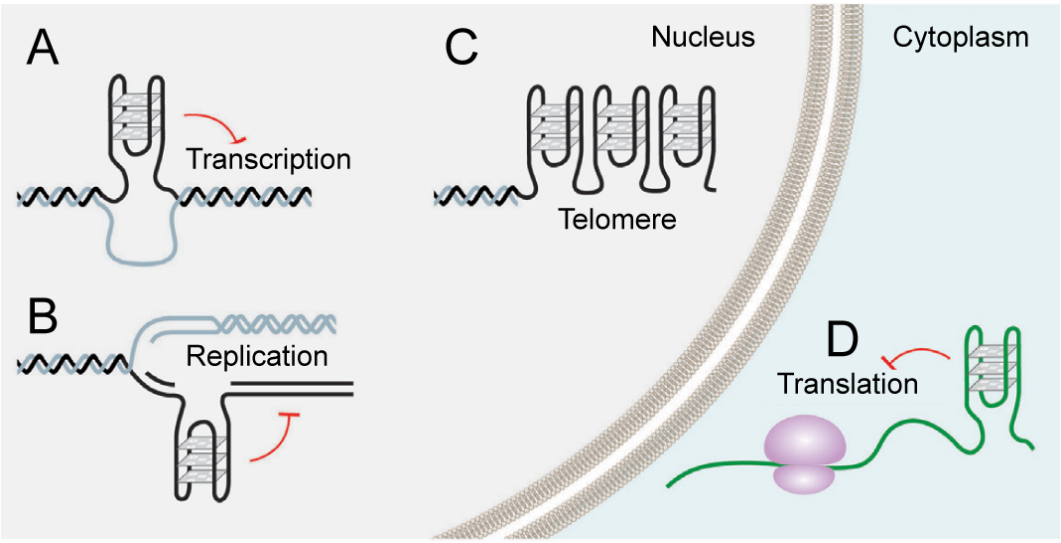


Figure 1.4: Known potential functions of G-quadruplexes. pG4 exists in A) promoter, B) origin of replication, C) telomere, and D) 5′-UTR of RNA regions. Figure is adapted from [20].

The most prominent and frequently studied pG4 is the human telomeric sequence. It consists of long repeats of d(TTAGGG) at the end of chromosomes, including the overhang regions [37]. This sequence is highly conserved in eukaryote cells [38] and its ability to form G-quadruplex in vitro is well known [14–17]. It was shown that the presence of G-quadruplex stabilizing ligand shortened the length of telomeric region [39–41], possibly due to the blockage of telomerase progression along the template sequence when encountered with the physical hindrance formed by the G-quadruplex structure [42]. Recently, it was shown explicitly that the G-quadruplex structure is formed in the telomeric region *in vivo* [43, 44]. Other approaches used to identify the presence of G-quadruplex in the telomeric region includes the use of radio-labeled G-quadruplex binding ligand [45] and structure-specific antibody [46, 47].

Another region within the human genome that is widely studied is the starting

point of DNA transcription, known as the promoter region. Around 50% of the human gene have been identified to possess pG4 sequence nearby their promoter regions, which strongly suggests the involvement of G-quadruplex in regulating the transcription process [20]. Many studies [48–50] including the mutation of pG4 sequence in c-myc promoter [51,52] and the addition of G-quadruplex binding ligand which affects DNA transcription *in vivo* [53,54] have proven the formation of G-quadruplex in the promoter regions.

The formations of G4 structures *in vivo* by pG4s are critical. To date, there are various evidences that G4 structures are formed in cells. The direct evidences mainly came from visualization of cells with fluorescence techniques [55,56], including quantification of the G4 structure formation with and without stabilizing small molecules added to the cells [46]. However, there were also studies which support the contrary, where the formation of RNA G4 was suggested to never happen *in vivo* due to a specific machinery in eukaryotic cells which can unfold any RNA G4 structure [57]. Overall, G-quadruplex has been the focus of many researches due to its enormous potential function in cells [58,59] as well as its unique robust structure that can be used for drug targeting [60–62] and in nanotechnology [63,64].

1.3 Spectroscopy of G-quadruplexes

This section gives an introduction for the three main spectroscopic techniques used throughout the dissertation: ultraviolet spectroscopy (UV), circular dichroism spectroscopy (CD), and nuclear magnetic resonance spectroscopy (NMR).

UV spectroscopy has been commonly used to detect and quantify the presence of biomolecules such as nucleic acids and amino acids [65,66]. Specifically, wavelengths ranging from 220 nm to 320 nm are frequently used to quantify the amount of DNA strands. DNA-containing solution fully transmits light at 320 nm, making

this wavelength the preferred zero-correction point [67]. Furthermore, the nucleobases in DNA chains include aromatic rings which absorb light at around 260 *nm* wavelength, with unique absorption values for different types of nucleobases [68]. In this study, the UV absorbance at 260 *nm* (A_{260}) is used to determine the DNA chain concentration based on its extinction coefficient (ϵ) and the Beer-Lambert law [69].

CD spectroscopy uses the principle of different absorbance of chiral molecules when irradiated with left circularly-polarized light and right circularly-polarized light [70]. The technique has been used widely in biomolecule studies including nucleic acids [71–73]. Foundation of G-quadruplex CD signal has been laid out both empirically [74,75] and analytically [76,77]. As illustrated in Figure 1.5A, parallel, anti-parallel, and hybrid (3+1) G-quadruplexes each exhibit different CD signature, characteristic of its folding topology [78]. Nevertheless, CD measurements always results in a combination of signals from all the structures present in the solution. Hence, the composition of multiple structures should be well understood prior to the use of CD technique as a deterministic experiment.

NMR spectroscopy is the main technique that is used in this study. It is used for a wide range of applications such as observation of G-quadruplex dynamics, determination of G-quadruplex structure formation, evaluation of G-quadruplex polymorphism in an assemble and elucidation of the high-resolution structure of G-quadruplex [15,16,79–82]. Guanine residues in G-quadruplex form cyclic-Hoogsteen hydrogen bonds, which is a unique feature not found in guanines in duplex structures or non-bonded guanines. The cyclic hydrogen bond connectivity for guanines participating in G-tetrad leads to imino proton precession with a unique frequency range of 10-12 ppm (Figure 1.5B) [83], making NMR spectroscopy a perfect tool to detect the existence and polymorphism of G-quadruplex structure.

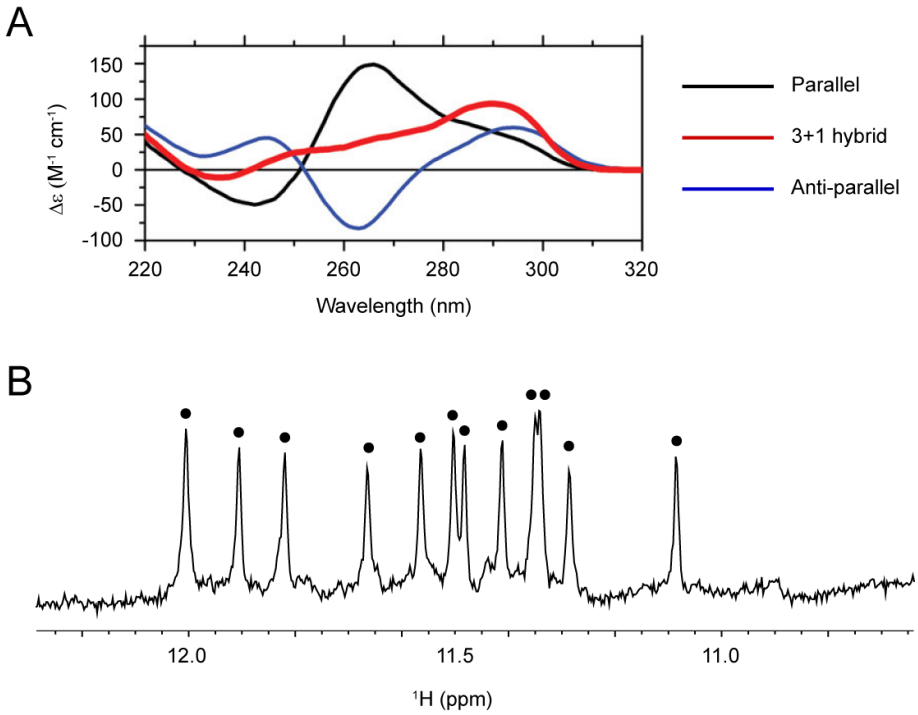


Figure 1.5: CD and NMR spectroscopy data of G-quadruplex. A) Characteristics CD profile of different type of G-quadruplexes. Black, red, and blue lines represent telomeric G-quadruplex in parallel, 3+1 hybrid, and anti-parallel conformations respectively. Figure is adapted from [78]. B) Characteristics NMR signal of G-quadruplex imino protons. Twelve sharp peaks (marked by black dots) in 10-12 ppm range represent twelve guanine imino protons involved in a 3-layer G-tetrad core.

1.4 Research scope and objectives

The scope of the study is divided into two main sections: the first section focuses on the development of NMR methodologies in G-quadruplex dynamics study which will be elaborated in Chapter 2 and the second section is on the applications of NMR spectroscopy in solving biological problems related to G-quadruplex which will be discussed in Chapter 3 and 4. The detailed objectives are as follows:

1. *Development of NMR methodologies in the study of G-quadruplex dynamics*

    - *Chapter 2.1:* Dynamics or molecular movement in biomolecules gives a notion of time in chemical processes, such as binding and unbinding of

ligands as well as changes between conformations [84, 85]. Previously, amino protons rotation for adenine residues in duplex DNA has been elucidated [86]. However, there exist little to no research on the behavior of amino proton in G-quadruplex guanines. In the first part of Chapter 2, the dynamics of amino proton in guanines of G-quadruplexes is investigated using lineshape analysis technique [87].

- *Chapter 2.2:* Stability of receptor-ligand complex is a crucial factor for the study of drug design. Generally, the parameters used for measurements of complex stability include dissociation constant (K_D) and change in melting temperature (ΔT_m). Nevertheless, the two parameters lack the temporal information of the receptor-ligand interaction, which is equally important in determining the functionality of the complex. In the second part of Chapter 2, NOESY exchange technique [88] is used to detect the residence time (t_{res}): the average time of the ligand to stay on the same receptor molecule before detaching itself of reattaching to another receptor molecule.

2. *Application of NMR in solving high resolution G-quadruplex structure and determining the component of biological sample*

- *Chapter 3.1:* The formation of a vacancy bearing G-quadruplex with $4n$-1 (n is integer representing number of G-tetrad layer) guanine residues is recently discovered [89, 90]. This discovery implicates a potential binding site for guanine derivatives and other small molecules on the G-quadruplex scaffold. In the first section of Chapter 3, the study of binding of metabolites to $4n$-1 G-quadruplex is investigated by solving two high-resolution structures of 11-guanine G-quadruplexes bound by a dinucleotide d(AG) and cyclic dinucleotide (cGAMP).

- *Chapter 3.2: AGRO100* [91, 92] is a G-rich aptamer that could potentially form multiple conformations of G-quadruplex in solution. Two of its derivatives *AT11* and *AT27*, both of which are parallel G-quadruplexes were previously studied. On the other hand, anti-parallel *AGRO100* derivatives — *AT21* — was also discovered. Diversity in the G-quadruplex conformations could prove essential for the binding of the aptamer towards certain ligands. Therefore, we aim to do a structural study on *AT21* sequence to observe its properties, which will be elaborated in the middle section of Chapter 3.

- *Chapter 3.3:* The existence of right-handed and left-handed G-quadruplex has been shown in previous studies [10–12]. However, unlike its duplex counterpart [93], left-handed and right-handed G-quadruplex structures were observed to always come separately. Hybrid G-quadruplex stacking structure has never been reported before, making its discovery highly important in the field of G-quadruplex structural polymorphism. In the last section of Chapter 3, the elucidation of a solution structure of the first left-handed right-handed hybrid G-quadruplex — $(GGT)_8$ — is described.

- *Chapter 4:* Biofilm formation results from extracellular polymer networking, whereby each polymer could contribute to the overall elastic property of the biofilm [94, 95]. However, the fundamental components of biofilm that contribute towards the polymer networking have not been studied in detail. Our objective is to identify the main components in this system responsible for forming the polymer network, as well as to evaluate the potential roles of nucleic acid including G-quadruplex in biofilm. In Chapter 4, biofilm formation in *Pseudomonas* is examined using NMR spectroscopy.

Chapter 2

NMR Methodologies in the Study of G-quadruplex Dynamics

Conformational fluctuation of biomolecules which plays a vital role in their biological functions can occur on a wide range of timescales (Figure 2.1). For instance, microsecond-to-millisecond dynamics are essential for the functions of nucleic acids [96, 97]. Adaptive response to molecular stimulus such as ligand and protein often causes dramatic change in RNA conformation involving global domain twisting and bending [98, 99]. The shape and flexibility of the DNA structure influence the readout of its genetic sequence and are thought to facilitate protein recognition [100, 101].

Part of this chapter is taken from the author's publication:

Adrian, M.*, **Winnerdy, F.R.***, Heddi, B. and Phan, A.T. Rotation in Guanine Amino Groups in G-Quadruplexes: A Probe for Local Structure and Ligand Binding. **2017**. *Biophysical Journal* 113, 775-784. with permission from Biophysical Journal. * denotes equal contribution.

Part of this chapter is taken from the author's unpublished work under the title of:

Winnerdy, F.R., Do, T.H.H., Heddi, B., Pratviel, G., Phan, A.T. Determination of residence time of G-quadruplex-binding ligands with NMR exchange analysis.

The experiments in these specific sections were performed by author's colleague Dr. Michael Adrian:

Subsection 2.3.1-2.3.5c — NMR experiments on *Platform G4*

NMR spectroscopy is an excellent technique in studying the dynamics of biomolecules over a broad range of timescales. With the recent developments of NMR techniques, observations of molecular movements can be achieved with high degree of details [102]. The study of molecular dynamics (with NMR spectroscopy) can be combined with rigid structures determination (with X-ray crystallography) to provide complete information required to understand the mechanism of biomolecules interactions such as binding and catalysis.

2.1 Dynamics in nucleic acids

Local motions including base-pair flipping, sugar re-puckering and phosphodiester backbone variation may allow enzymes to access functional groups of DNA molecules which are normally inaccessible. One of the important dynamic mechanisms in the cells is the base-pair opening, which is a prerequisite for DNA replication and transcription, and is also associated with DNA damage and repair processes [103, 104]. In duplex DNA, NMR studies on the imino proton exchange between the DNA and the surrounding water have indicated the opening rate of a base pair to be dependent on both its conformation and its nucleotide sequence [105–109]. For G-quadruplex, importance of dynamics include both base-pair opening and the folding of the G-quadruplex structure. Although experimental data from several thermodynamic and kinetic studies have revealed possible mechanisms of the G-quadruplex folding pathway [110, 111], the dynamics of G-quadruplexes are conventionally studied using computational methods [112–114]. This includes the *in silico* study by Islam and coworkers, where microsecond timescale simulation of a parallel-stranded G-quadruplex has determined the propeller loops to exhibit dynamic interactions with the G-tetrads [115].

Another type of internal motion occurring within the DNA base is the rotation

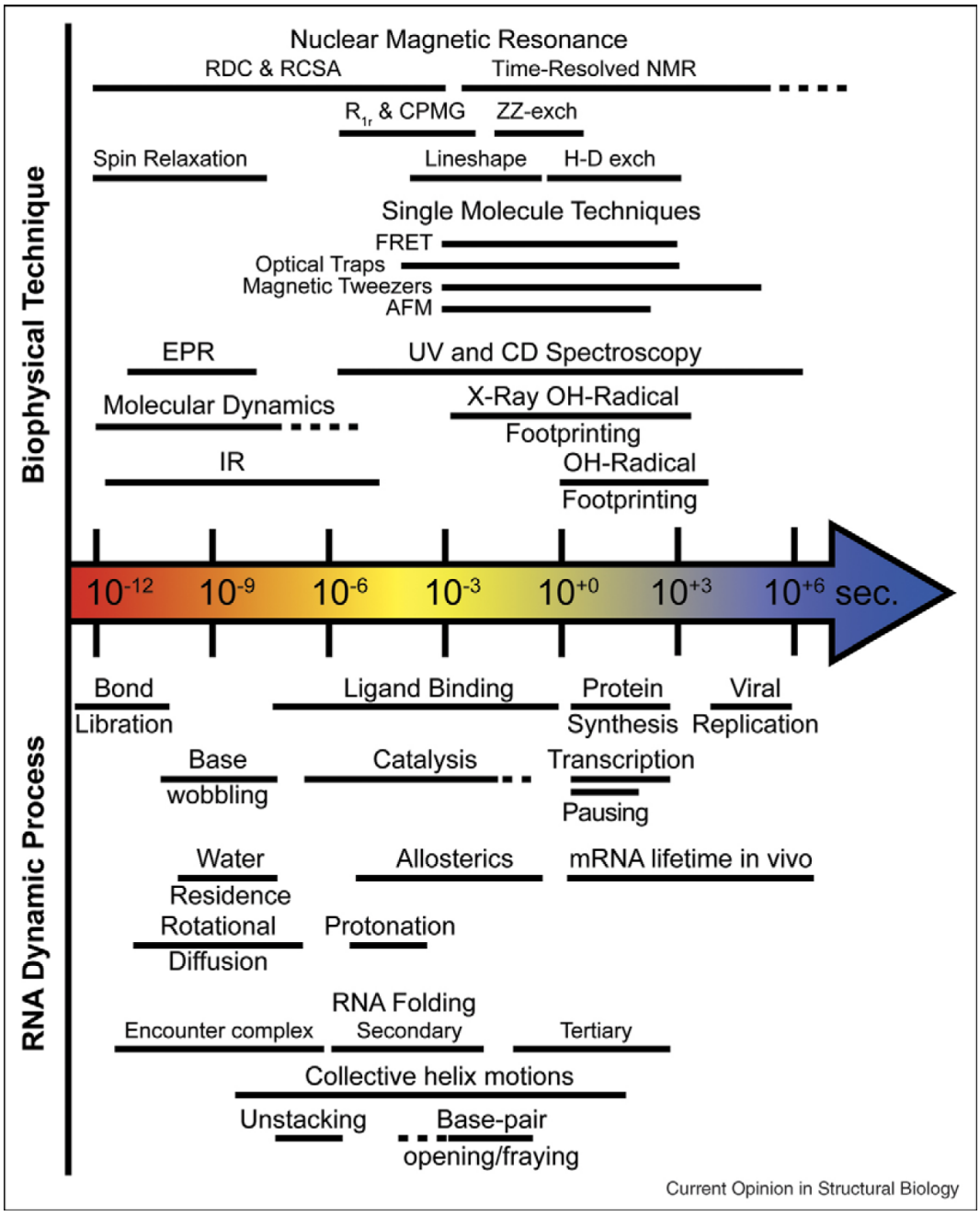


Figure 2.1: Timescales of biological techniques and biological processes. Figure taken from [116].

of its amino groups. The rotation of the amino group of adenine in the context of DNA duplex has been previously investigated using NMR spectroscopy, which suggest that apart from hydrogen-bonding, base stacking and DNA hydration also affect the rate of rotation [86]. The amino group of guanine is a key component of

the G-tetrads, providing half of the total hydrogen bonds that interconnect guanine bases in a planar configuration. Moreover, these amino groups are situated along G-quadruplex grooves and may act as sites for molecular recognition. Herein, we report the investigation of the rotational dynamics of guanine amino groups in G-quadruplexes using NMR spectroscopy, with the rotation rates of amino groups at distinct locations within various G-quadruplexes derived from their NMR lineshapes.

In addition to the study of the dynamics of within G-quadruplexes, a similar dynamic process of G-quadruplex-ligand interaction was also investigated. Interactions between a ligand to a receptor are generally assessed by the affinity of the ligand to the receptor (dissociation constant, K_D), and the stability of the receptor-ligand complex (change of melting temperature ΔT_m). However, in many cases, both the dissociation constant and the change of melting temperature are not representative in the biological function of the G-quadruplex-ligand complex, as they lack crucial information on the temporal lifetime of the complex [116]. Therefore, the concept of residence time was coined as a more direct measurement of temporal lifetime of a receptor-ligand complex [117], and is widely applied in pharmaceutical studies of the drug activity [118–121]. In this study, we examine the residence time of the G-quadruplex-ligand complex, with the dissociation rates of the complex derived from the NOESY exchange method at various experimental conditions.

2.2 Two-site chemical exchange methodologies

In this section, the basic theory of lineshape analysis method and NOESY exchange method is explored. The first method is used for determining the rate of guanine amino rotation of G-quadruplex, while the second method is applied for determining the residence time of ligand in complex with G-quadruplex.

The lineshape analysis is chosen due to its simple nature, whereby the observation of peak linewidth directly correlated to the exchange rate between two protons. This method might not work in certain order of exchange rate, as well as in the case of relatively larger molecules with crowded ^{1}H NMR signals. The advantages of the method include the cost-effectiveness as it does not require any isotope labelling, as well as the direct nature of the observation, leading to precise measurements.

The NOESY exchange method is a relatively simple and cost-effective method in determining the residence time of G4 ligands, especially for the purpose of screening, where only ligands with relatively high residence time would be of interest. Other methods, for instance ZZ exchange method would require labelling and only useful in detecting exchange rate that is faster than the relaxation rate of the magnetization or coherences present during mixing time. Relaxation and CEST methods are both useful in observing exchange process where the higher energy population is too low, which is not the case in our experiment, as the G4-ligand complexes population is comparable with the free-G4 states [102].

2.2.1 Lineshape analysis

NMR spectroscopy detects resonance precession frequencies of nuclei and converts them into readable peaks. Raw NMR data consists of free induction decay (FID) intensities, where the value of magnetization to a certain transverse direction is collected over a period of time in a designated interval. This FID, which is in time domain, is then converted into a frequency domain via Fourier transform. Therefore, the shape of the peak relies heavily on the amplitude, frequency, as well as decaying behavior of the FID (Figure 2.2). For a nucleus without any exchange process, the shape of an individual peak will follow the Lorentzian curve:

$$L(x) = A\frac{\frac{1}{2}\Gamma}{(x - x_0)^2 + (\frac{1}{2}\Gamma)^2} \tag{2.1}$$

where A is an intensity normalization constant, Γ is the full width at half maximum, and x_0 is the center of the peak. In the context of NMR, Γ is proportional to the decay rate of the transverse magnetization (k), which is the reciprocal of transverse relaxation time (T_2), $\Gamma \sim \frac{1}{T_2}$, and x_0 is the central precession frequency of the observed nucleus.

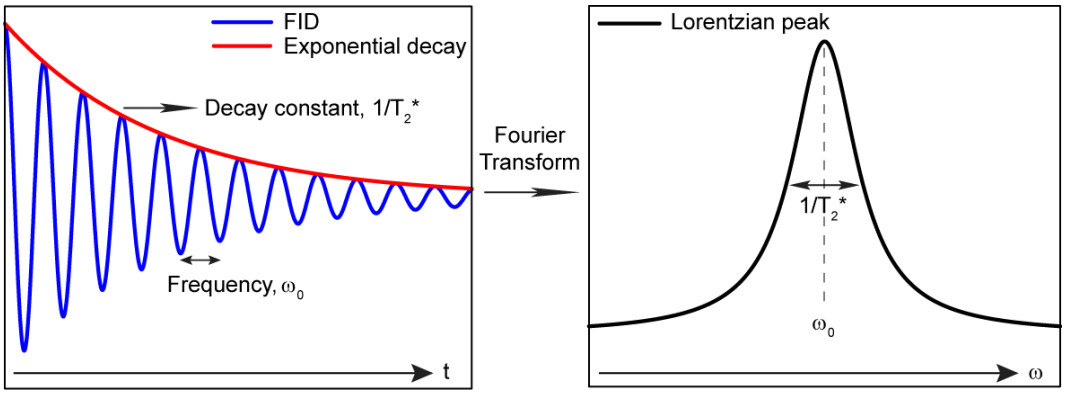


Figure 2.2: Schematic of NMR Fourier transformation process. (Left) Free induction decay (FID) signal as a direct output from NMR measurement. (Right) The Fourier transform of the FID results in Lorentzian peak with central frequency and linewidth related to the FID.

The two-site exchange lineshape analysis method is based on the above principle, with an additional interaction factor between the two individual peaks coming from the exchange process. When the chemical exchange between two conformations occurs at a much lower rate (k) than the initial frequency separation (2Ω) between two associated resonances, the resulting spectrum displays two separate peaks with relatively narrow linewidth (Figure 2.3). For $0 \leq k \leq \Omega$, the modified Lorentzian lineshape function of two exchanging resonance peaks is given by the following equation [87]:

$$f(x) = A\left\{ \frac{\left(\frac{1}{T_2^*} + 2k\right) + \left(\frac{k}{\epsilon}\right)x}{\left(\frac{1}{T_2^*} + k\right)^2 + \left(x + \epsilon\right)^2} + \frac{\left(\frac{1}{T_2^*} + 2k\right) - \left(\frac{k}{\epsilon}\right)x}{\left(\frac{1}{T_2^*} + k\right)^2 + \left(x - \epsilon\right)^2} \right\} \tag{2.2}$$

with

$$\epsilon = \sqrt{\Omega^2 - k^2} \tag{2.3}$$

A is an intensity normalization constant and T_2^* is intrinsic transverse relaxation lifetime of the nuclei. Increase in exchange rate broadens the linewidth and decreases apparent separation frequency until the two associated peaks coalesce at $k = \Omega$. Further increase in exchange rate reduces coalesced linewidth and strengthens the peak intensity. For $k > \Omega$, the modified Lorentzian lineshape function of a coalesced resonance peak is given by the following equation [87]:

$$f(x) = A\left\{ \left(1 + \frac{k}{\alpha}\right)\left[\frac{\frac{1}{T_2^*} + k - \alpha}{\left(\frac{1}{T_2^*} + k - \alpha\right)^2 + x^2} + \right] \left(1 - \frac{k}{\alpha}\right)\left[\frac{\frac{1}{T_2^*} + k + \alpha}{\left(\frac{1}{T_2^*} + k + \alpha\right)^2 + x^2} \right] \right\} \tag{2.4}$$

with

$$\alpha = \sqrt{k^2 - \Omega^2} \tag{2.5}$$

Here, we used the above described method to observe the exchange rate between two amino protons from the same amino group; the amino proton rotation rate in G-quadruplex. The intrinsic transverse relaxation rate $R_2^* = 1/T_2^*$ of the amino protons is a major contributing component to the linewidth at low temperatures. This rate is estimated using N2-modified G-quadruplex molecules, whereby one amino proton was substituted by a methyl group in one of G-tetrad participating guanine located at various positions (Figure 2.7A-D). The modification froze the

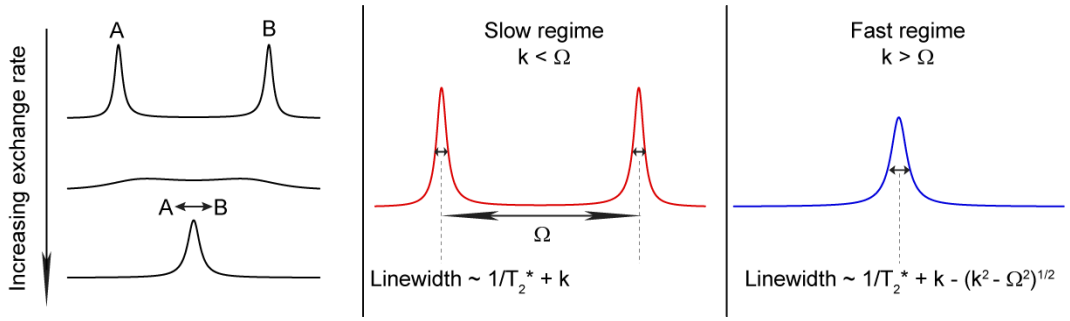


Figure 2.3: Schematics of Lineshape analysis. (Left) NMR spectra of two-site exchange process ranging from slow, intermediate, to fast exchange rate. (Middle) Slow exchanging protons give rise to two sharp and separated peaks, while (Right) fast exchanging protons results in one coalescent peak in the middle.

amino group rotation and consequently generated sharper resonance peak from the hydrogen-bonded amino proton, and the linewidth of the peak could be used to estimate R_2^* at various temperatures (Figure 2.7E and Table 2.1). The initial separation frequency 2Ω between two exchanging resonance peaks is dependent on the temperature and can only be precisely measured in the absence of exchange process. Consequently, Ω (and thus ϵ) is treated as a variable during fitting.

Amino proton peaks generally located at crowded regions in the NMR spectra, which compelled us to observe the NOESY trace imino to amino proton twin peaks instead of the direct 1D amino proton twin peaks. However, when using both of the NOESY trace imino to amino twin peaks to fit the lineshape function, several factors such as the chemical exchange between amino and water protons and the difference in spatial distance between the imino proton and the two neighboring amino protons might affect the quantitative determination of the rotation rates.

Exchange processes between amino and water protons have been qualitatively observed by looking at slightly asymmetric linewidths between the two amino proton peaks, one of which has a higher tendency to be involved in an exchange process with water protons. These exchange rates were measured to be low ($\sim 2Hz$ in AT

base pair of B-DNA at 70^oC, $\sim 0.02Hz$ in GC base pair of Z-DNA at 50^oC) [86,107], even for amino protons of free mononucleosides without any base pairing [122], and thus can be safely neglected in the lineshape analysis. Furthermore, there is negligible influence from small chemical shift anisotropy of protons [123].

Cross-peak intensity in NOESY spectra has a reciprocal dependency with the spatial distance between the two interacting protons ($I \sim 1/r^6$). Due to the difference in spatial distances between an imino proton and the two neighboring amino protons, the two peaks of amino protons would not have the same intensity in the NOESY trace spectrum. As a result, a specific intensity modulation is required to normalize this effect, which would add another unknown to the equation.

To avoid these problems, one of the two twin peaks which corresponds to the internal amino proton is used as the fitting peak, where it is assumed that the proton corresponding to this peak does not undergo any exchange with water due to its more protected position. The other peak corresponding to the water-exchanging proton is used only as a chemical shift reference to determine the initial estimate of Ω. The fitting of modified Lorentzian function to relevant data points around protected amino proton resonance peaks (excluding nearby resonance peaks originated from other protons) on NMR lineshapes of associated amino protons was done using the Origin software. Fitting iterations on coalesced lineshape were done by estimating the initial separation frequency between amino protons to be $\sim$ 3 ppm.

2.2.2 NOESY exchange analysis

NOESY exchange peak intensity correlate with the amount of exchange processes that happened during the mixing time period between the two exchanging states (Figure 2.4). In the mixing period, three distinct processes happen: the first is the

intrinsic spin-lattice relaxation of the nuclei, governed by the value T_1; the second is the exchange process between the two states, controlled by the rate constant k; and the third is the dipolar interaction between the two nuclei expressed by the rate constant R_D.

In our context, the kinetics of the two states is assumed to obey a simple first order chemical equation given by:

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B \tag{2.6}$$

with A and B as the two states, and k_1 and k_2 as the reaction rate constants in a given condition from A to B and B to A respectively.

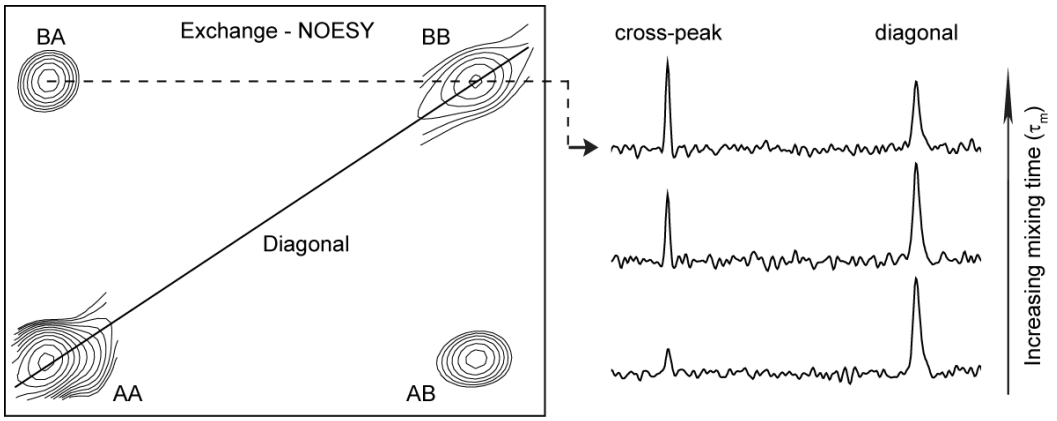


Figure 2.4: NOESY exchange analysis method. (Left) Example NOESY spectrum of two states of molecules (A state and B state). Diagonal peaks are labeled AA and BB, while cross peaks AB and BA. (Right) 1D slice projection of the NOESY spectrum, showing the change in intensity of the cross-peak and diagonal peak with changing mixing time.

The NOESY exchange matrix is given by [88]:

$$K = \begin{pmatrix} -\frac{1}{T_{AA}} & -\chi_A R_C \\ -\chi_B R_C & -\frac{1}{T_{BB}} \end{pmatrix} \tag{2.7}$$

with the following details,

$$\begin{aligned}
\frac{1}{T_{AA}} &= \frac{1}{T_{1A}} + 2\chi_B R_D + \chi_B k \\
\frac{1}{T_{BB}} &= \frac{1}{T_{1B}} + 2\chi_A R_D + \chi_A k \\
R_C &= R_D - k
\end{aligned}$$

T_{1A}, T_{1B}, χ_A and χ_B are the longitudinal relaxation rates and mole fractions of the states respectively, k is the exchange rate, and R_D is the distance-based dipolar interaction rate.

Here, we use two approximations suitable to our free and ligand-bound G-quadruplex states for simplification of the fitting formula. The first assumption is that both free and bound states have similar longitudinal imino proton relaxation rate, $T_{1A} \approx T_{1B} = T_1$, and the second assumption regards the dipolar interaction to be much lower than the exchange rate, $R_D \ll k$.

Following the two approximations, the mixing coefficients that describe the peak intensities in the NOESY spectrum can be expressed as [88]:

$$\begin{aligned}
a_{AA}(\tau_m) &= \chi_A \exp\left(-\left(\frac{1}{T_1} + \frac{k}{2}\right)\tau_m\right)\left[\cosh\left(\frac{k}{2}\tau_m\right) - (\chi_B - \chi_A)\sinh\left(\frac{k}{2}\tau_m\right)\right] \\
a_{BB}(\tau_m) &= \chi_B \exp\left(-\left(\frac{1}{T_1} + \frac{k}{2}\right)\tau_m\right)\left[\cosh\left(\frac{k}{2}\tau_m\right) + (\chi_B - \chi_A)\sinh\left(\frac{k}{2}\tau_m\right)\right] \\
a_{AB}(\tau_m) &= a_{BA}(\tau_m) = 2\chi_A\chi_B \exp\left(-\left(\frac{1}{T_1} + \frac{k}{2}\right)\tau_m\right)\sinh\left(\frac{k}{2}\tau_m\right)
\end{aligned}$$

with a_{AA} and a_{BB} are the intensity coefficients of the two diagonal peaks, a_{AB} and a_{BA} are the intensity coefficients of the two cross peaks, and τ_m is the NOESY mixing time.

Although there are several unknown parameters in the above equations, we can mathematically remove the exponential decay factor by taking the ratio of any two

of the intensity coefficients. The fraction of the free or bound state that remains after the mixing time period is defined by the ratio between the diagonal peak intensity at free or bound state (a_{AA} or a_{BB}) to the summation of intensity coming from the corresponding free or bound states ($a_{AA}+a_{AB}$ or $a_{BB}+a_{BA}$) respectively. In practice, one of the two ratios (either $a_{AA}/(a_{AA}+a_{AB}$) or $a_{BB}/(a_{BB}+a_{BA})$) is calculated from the above equation, with the following boundary conditions: ratio of 1 at zero mixing time; and ratio of equilibrium (χ_A or χ_B) at extremely long mixing time. The mathematical expression of the ratio (taking the free states as an example):

$$\frac{I_{diag}}{I_{total}} = \frac{a_{AA}}{a_{AA}+a_{AB}} = \frac{\cosh\left(\frac{k}{2}\tau_m\right) + (\chi_A - \chi_B)\sinh\left(\frac{k}{2}\tau_m\right)}{\left[\cosh\left(\frac{k}{2}\tau_m\right) + (\chi_A - \chi_B)\sinh\left(\frac{k}{2}\tau_m\right)\right] + 2\chi_B \sinh\left(\frac{k}{2}\tau_m\right)} \tag{2.8}$$

which can be simplified to:

$$\frac{I_{diag}}{I_{total}} = \chi_A + \chi_B \exp(-k\tau_m) \tag{2.9}$$

Measuring the above intensity ratio in various mixing time and fitting the experimental result to the equation would give the exchange rate constant k, where the reciprocal of k represents the time constant for the molecules initially in extreme state (either fully bound or free) to return to equilibrium. During this mixing time, when the exchange process from an extreme condition to equilibrium condition happen, the association and the dissociation processes occur simultaneously. For a long-lived bound state, the latter process will dominate the time required, and this time parameter is termed as the residence time of the ligand towards the receptor. Therefore, in context with our G-quadruplex-ligand complex, we take the reciprocal of k as the residence time of the ligand (t_{res}).

2.3 Rotation of amino group in G-quadruplexes

This work is published in *Biophysical Journal* with the title: "Rotation in Guanine Amino Groups in G-Quadruplexes: A Probe for Local Structure and Ligand Binding". Part of the section was presented earlier in the thesis of **Dr. Michael Adrian** (Chapter 5, 2013).

As illustrated in Figure 2.5A, an amino group can rotate about the C-N bond by flipping its twin protons. The rotation period of an amino group generally falls in the range of millisecond timescale, which largely affects the amino proton linewidth. In this study, the rotation rates of amino groups in different structural contexts of G-quadruplexes are investigated by NMR lineshape analysis.

2.3.1 NMR spectra of a model G-quadruplex

In K^+ solution, a tetrameric G-quadruplex was formed from the association of four d(TTAGGGGT) strands in an all-parallel arrangement (designated *Platform G4* thereafter) (Figure 2.5C) [110, 124, 125]. The imino proton spectrum of *Platform G4* displays four well-resolved peaks at 10-12 ppm (Figure 2.6B, top) corresponding to four G-tetrad layers [124, 125]. This symmetric G-quadruplex comprises four G-tetrads and yet displays a relatively simple NMR spectrum, allowing the demonstration of our NMR method for a systematic characterization of amino group rotation at different G-tetrad layers.

2.3.2 Blocking the rotation of amino groups by H-to-CH3 substitutions

The rotation of amino groups in *Platform G4* leads to broadened amino proton peaks with reduced peak intensity (Figure 2.6B, top). In order to freeze the rotation

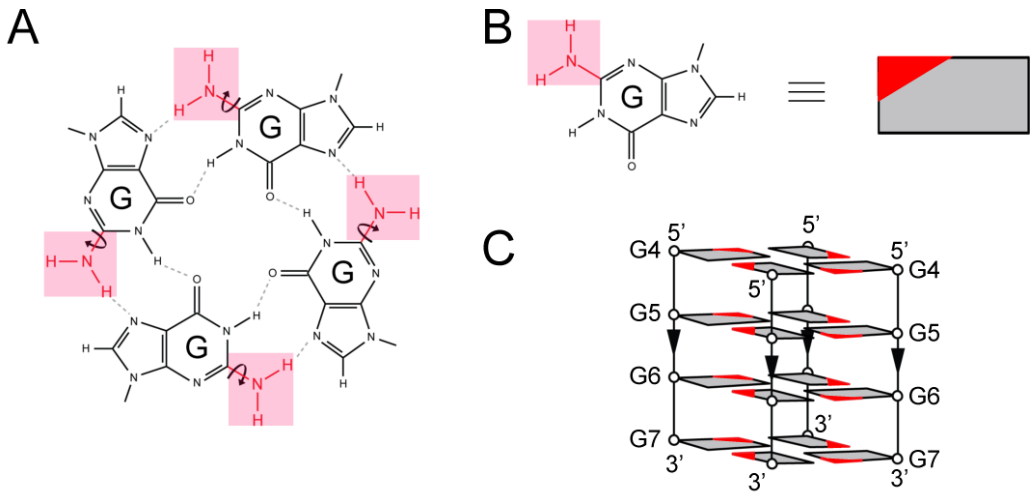


Figure 2.5: Guanine amino group rotation in G-quadruplexes. A) Chemical drawing of a G-tetrad, with the amino groups shaded in red. The amino group rotation about the C-N bond is represented by an arrow. B) Guanine base is drawn as a rectangle with one red-colored edge indicating the location of the amino group. C) Schematic structure of the tetrameric G-quadruplex $[d(TTAGGGGT)]_4$ (*Platform G4*). Only guanines are drawn for clarity.

of an amino group in a guanine, an amino proton was substituted by a methyl group (Figure 2.6A). For example, an N2-methyl guanine substitution at the G5 position of *Platform G4* resulted in the appearance of a sharp amino proton peak at ~ 9.1 ppm (Figure 2.6B, bottom) and a sharp methyl proton at $0.5 - 2.5$ ppm (not shown in the figure), consistent with the amino group rotation being blocked. Substitution of guanine by N2-methyl-guanine in the *Platform G4* sequence did not disrupt the G-quadruplex fold (data not shown), with the NMR spectra displaying four imino peaks characteristic of a symmetrical four-layer G-quadruplex (Figure 2.6B, bottom; Figure 2.7A-D).

The linewidth of the observed amino proton peak varied with temperature, providing a good estimate of the intrinsic transverse relaxation time of amino protons as a function of temperature within the examined G-quadruplex. Analyzing the lineshapes of these peaks facilitates the determination of the transverse relaxation

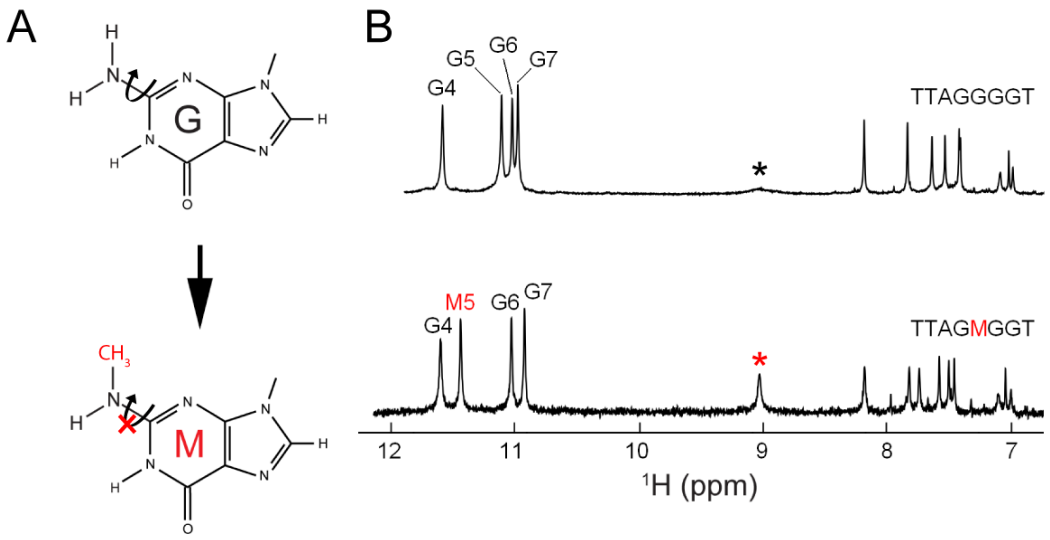


Figure 2.6: NMR spectra of *Platform G4* and a modified sequence. A) Chemical structure of a guanine (G) and an N2-methyl modified guanine (M). B) NMR spectrum of Platform G4 (top) showing a broadened amino proton peak (marked by an asterisk) and NMR spectrum of a modified sequence with N2-methyl guanine substitution at position 5 (bottom) showing a sharp amino proton peak (marked by a red asterisk). Imino proton assignments are indicated.

rate R_2^* as a function of temperature (Table2.1 and Figure 2.7E).

2.3.3 Observation of guanine amino protons spectra via NOE with imino protons

Guanine amino protons are highly overlapped and span the chemical shift range from 5 to 10 ppm, a spectral region crowded with other protons, such as aromatic and H1' protons (Figure 2.8A). Besides the problem of spectral overlap, baseline distortion by the signal of water protons further hampers the detection of individual amino protons. Consequently, a previous study have to use site-specific 15N-labeled samples to observe pairs of individual amino protons in 15N-edited experiments for their spectral lineshape analysis [86].

Here, we propose a simple and cost-effective method to extract residue-specific

Table 2.1: Estimated transverse relaxation rates (R_2^*) from N2-Methyl modified G-quadruplexes. Values indicated by * could not be determined due to superposition with another amino peak.

R_2^* (Hz)			
T (K)	**M5**	**M6**	**M7**
278	*	11.49	11.68
283	*	10.25	10.63
288	*	9.25	9.75
293	*	8.60	9.25
298	9.73	8.03	8.80
303	8.31	7.67	8.56
308	7.26	7.34	8.33
313	6.77	7.09	7.95
315	6.60	7.00	7.80
318	6.35	6.86	7.60

amino proton lineshape profiles of selected guanines without the use of rather expensive site-specific isotope labeling. This was based on the close proximity between imino and amino protons within a guanine, which would generate strong NOE cross-peaks in NOESY spectrum (Figure 2.8B). The identification of amino protons is straightforward, provided imino protons are well-resolved and assigned: they give the strongest cross-peaks in the NOE-trace of each imino proton (Figure 2.8C).

2.3.4 The rotation rates of amino groups estimated from NMR lineshape analysis

Qualitative assessment on the lineshape profiles (Figure 2.9) allowed an estimation of amino group rotation rates within *Platform G4*, which follow the order (from

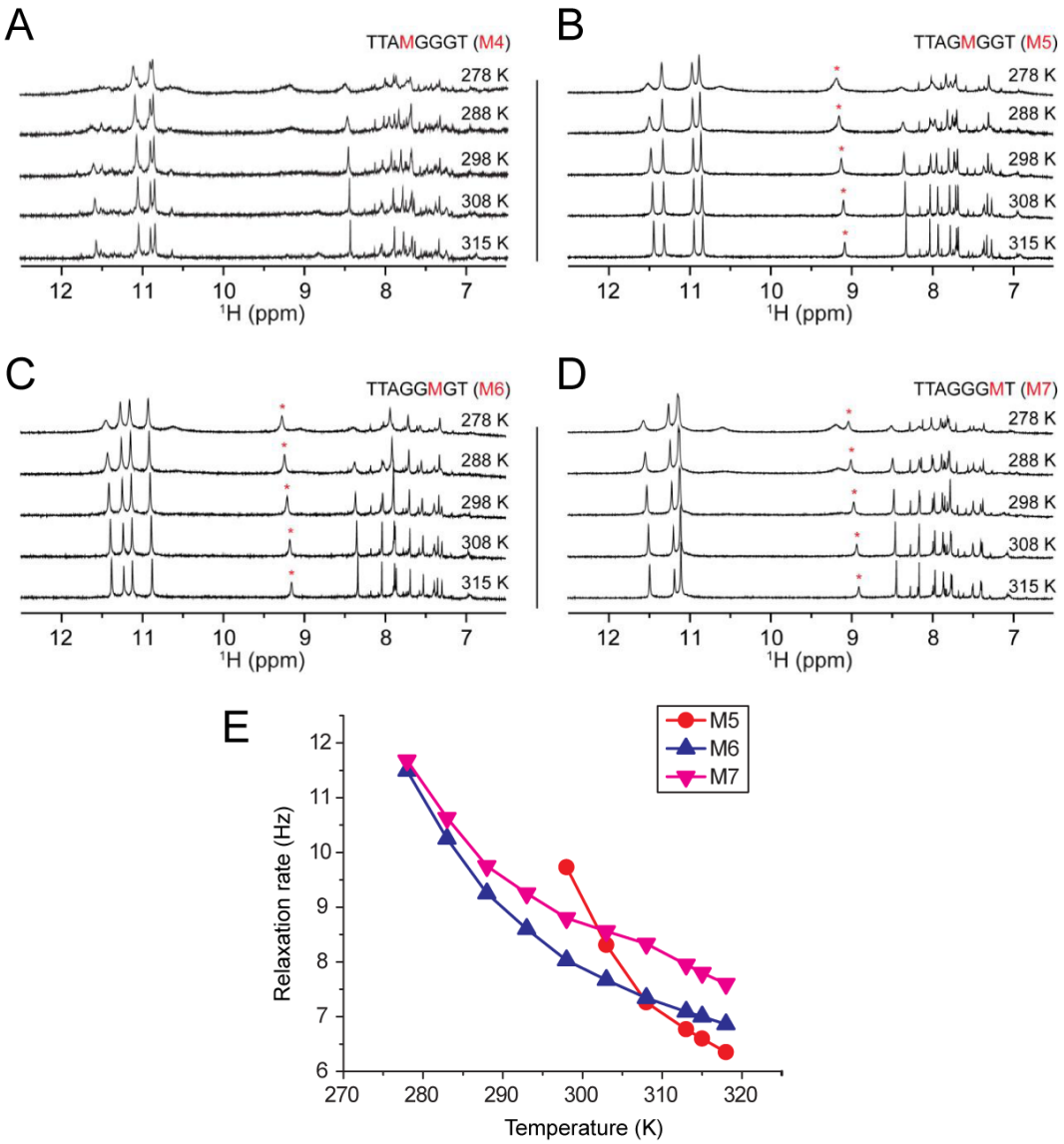


Figure 2.7: 1D NMR spectra of N2-modified G-quadruplexes at different indicated temperatures: A) TTAMGGGT (M4), B) TTAGMGGT (M5), C) TTAGGMGT (M6) and D) TTAGGGMT (M7). Amino proton peak is marked by asterisk. E) Temperature-dependent transverse relaxation rate plot from N2-Methyl G-quadruplexes (M6 and M7). Data points are connected by linear lines. The average values of the R_2^* are used in any further fitting method that requires T_2^* values.

higher to lowest), G7 > G4 > G5 > G6. At 298 K, *Platform G4* exhibited amino proton lineshape profiles with two distinctive peaks for each of the two guanines G5 and G6, and lineshape profiles in the coalescent regime for G4 and G7. The

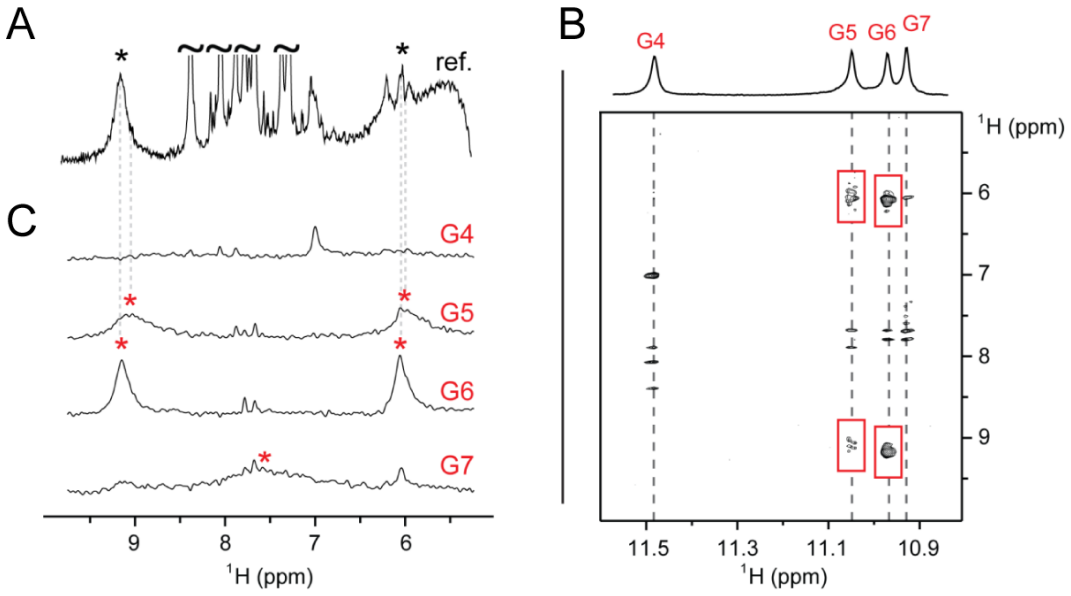


Figure 2.8: NOE-based approach: amino proton spectra of *Platform G4*. A) 1D NMR spectrum showing overlapping of proton peaks. Black asterisk indicates degenerated amino proton chemical shift. B) Expanded NOESY spectrum (mixing time, 300 ms) displaying imino/amino cross peaks at 298 K (marked by red rectangles). Fully assigned imino proton spectrum is shown at the top. C) Residue-specific amino proton spectra plotted from NOE-traces at different imino proton chemical shifts. Individual amino proton peak is marked by red asterisk.

lineshape profiles were observed to vary with the temperature: for example, amino protons of G5 showed distinctive peaks at 278 K, which merged to a very broad peak at 315 K (Figure 2.9B).

The values of amino group rotation rates were derived from fitting of modified Lorentzian function to the lineshape profiles (Figure 2.9). Depending on the G-tetrad position within *Platform G4*, the rate of guanine amino group rotation varies from ~20 Hz to ~5,000 Hz over the temperature range from 278 to 315 K (Figure 2.10; Table 2.2). At 298 K, the G6 amino group rotates with the lowest rate (~40 Hz) relative to those from other G-tetrads, followed by the G5 amino group (~140 Hz). The rotation rate of the G4 amino group could only be roughly estimated from their coalesced lineshape profiles (within intermediate regime), while the G7 amino group rotation (~3,000 Hz) is in the fast rotation regime. At elevated temperature

of 308 K, the G6 amino group rotation is still the slowest (∼75 Hz), while the lineshape profile of the G5 amino group coalesced to a low-intensity plateau shape corresponding to an intermediate rotation rate. The G4 amino protons barely formed one peak in the middle with estimated rotation rate of ∼2000 Hz, while those of G7 rose to a single observable peak corresponding to a rotation rate of ∼3,750 Hz. Over the range of experiment temperatures, the rotation rates of the G4 amino groups varies from low to high, while those of G6 amino groups are restricted within the slow regime (∼20 to ∼300 Hz; Figure 2.9).

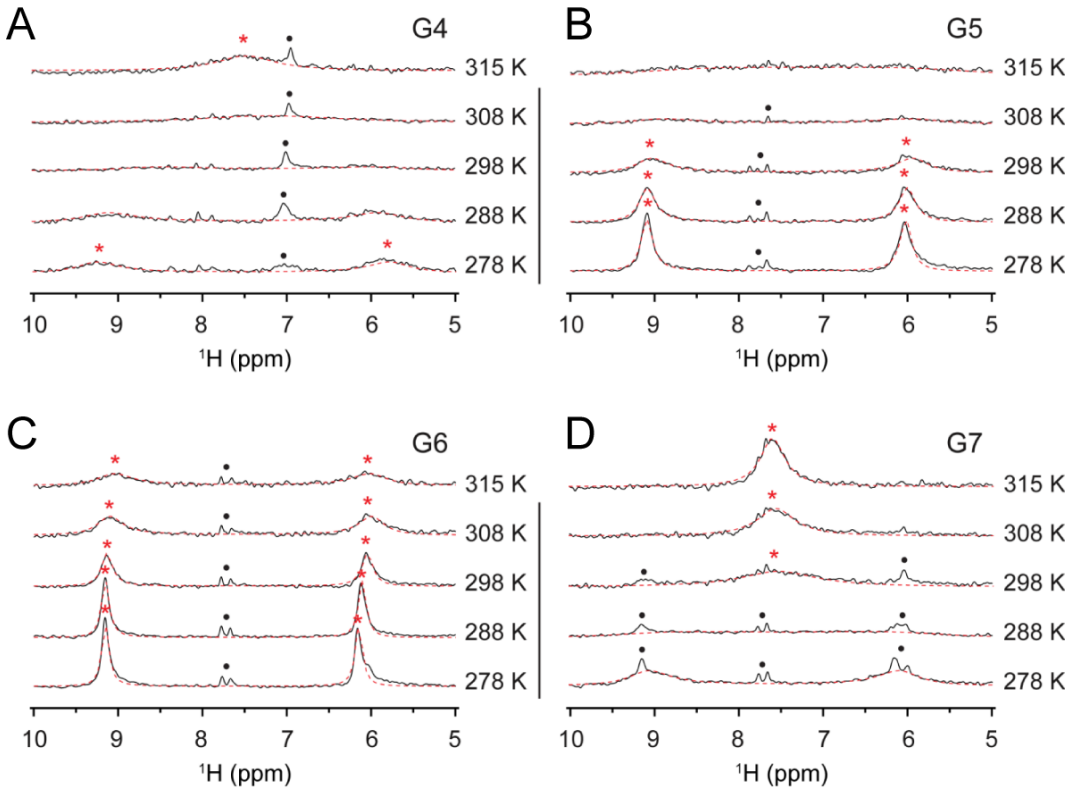


Figure 2.9: Chemical exchange of amino protons. Amino proton spectra of: A, B, C, D) G4, G5, G6 and G7 of the Platform G4 taken at different indicated temperatures showing sets of two coalesced peaks and broadened peaks. Amino proton peak is marked by asterisk. Broken red line represents fitting curve of two-site exchange lineshape function. Proton peak marked by dot was excluded in during fitting iteration.

Within the G-tetrad core, the rotation rates of amino groups of the internal tetrads G5 and G6 (∼140 Hz and ∼40 Hz at 298 K, respectively) were observed to be

significantly lower than those of the external tetrads G4 and G7 (the latter being ~3,000 Hz at 298 K). The difference in the rotation rates of amino groups may be related to the local structural contexts of different G-tetrads, with the central tetrads (G5 and G6) being sandwiched between two G-tetrads, while the external tetrads (G4 and G7) being positioned between a G-tetrad and an A-tetrad or possibly a T-tetrad [124].

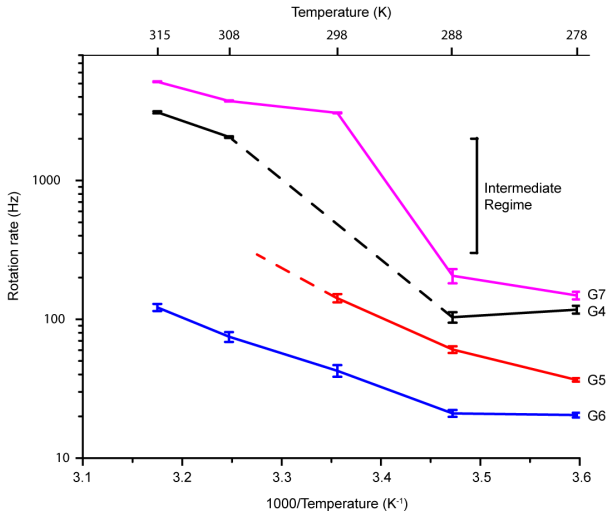


Figure 2.10: Plot of amino group rotation rate of *Platform G4* as a function of temperature. Rotation rate values from ~300 to ~2000 Hz lie within intermediate regime (indicated), in which the lineshape analysis method is unable to determine the exact values.

Table 2.2: Guanine amino proton rotation rates calculated from *Platform G4*.

Amino groups rotation rate (Hz)				
T (K)	**G4**	**G5**	**G6**	**G7**
278	117.5 ± 7.6	36.6 ± 1.0	20.4 ± 0.8	148.7 ± 9.3
288	103.6 ± 9.1	36.6 ± 1.0	20.4 ± 0.8	148.7 ± 9.3
298	N/A	142.0 ± 9.7	42.6 ± 4.1	3072 ± 24
308	2058 ± 28	N/A	74.8 ± 6.2	3746 ± 39
315	3103 ± 52	N/A	121.7 ± 7.2	5151 ± 42

2.3.5 Factors affecting amino group rotation rate

a. Stacking of G-tetrad

It was shown previously that removal of the 3'-end thymine residue from the Platform G4 sequence will result in the d(TTAGGGG) sequence, which favors the 3'-3' stacking of two G-quadruplex blocks (designated *Platform G4 Dimer*) (Figure 2.11A) [124, 125]. Remarkably, the amino groups of G7 exhibited substantially different rotation rates between *Platform G4* and *Platform G4 Dimer* (Figure 2.11B). At 298 K, two broadened but noticeable peaks could be seen in the G7 amino proton spectrum of *Platform G4 Dimer* (corresponding to a rotation rate of ∼130 Hz), while a coalesced peak was observed in the G7 amino proton spectrum of *Platform G4* (rotation rate ∼3000 Hz). Structurally, G7 tetrads in *Platform G4 Dimer* are mounted onto each other at 3'-3' interface through π-stacking interaction (Figure 2.11A), consequently adding restraints to the rotation of the G7 amino group. This observation at the 3'-3' stacking interface in *Platform G4 Dimer* is in line with the reduced amino group rotation of the inner G-tetrads due to stacking of multiple layers within *Platform G4*. Nonetheless, it should be noted that there are major differences between stacking of G-tetrads within *Platform G4* and of G-tetrads at 3'-3' interface in *Platform G4 Dimer*: (i) the G-tetrad stacking orientation is of similar polarity within *Platform G4*, while the G-tetrad stacking orientation in 3'-3' interface of *Platform G4 Dimer* is of opposite polarity, and (ii) there are no backbone constraints between 3'-3' stacked G-tetrads in *Platform G4 Dimer*.

b. Local conformation

Our methodology is applied to an intramolecular parallel G-quadruplex, which has different structural features from the previous G-quadruplexes. The presence of structural elements such as base pair [79, 126], triad [127], mixed-residue tetrad [128, 129], pentad [130], hexad [131] and heptad [132] have been observed

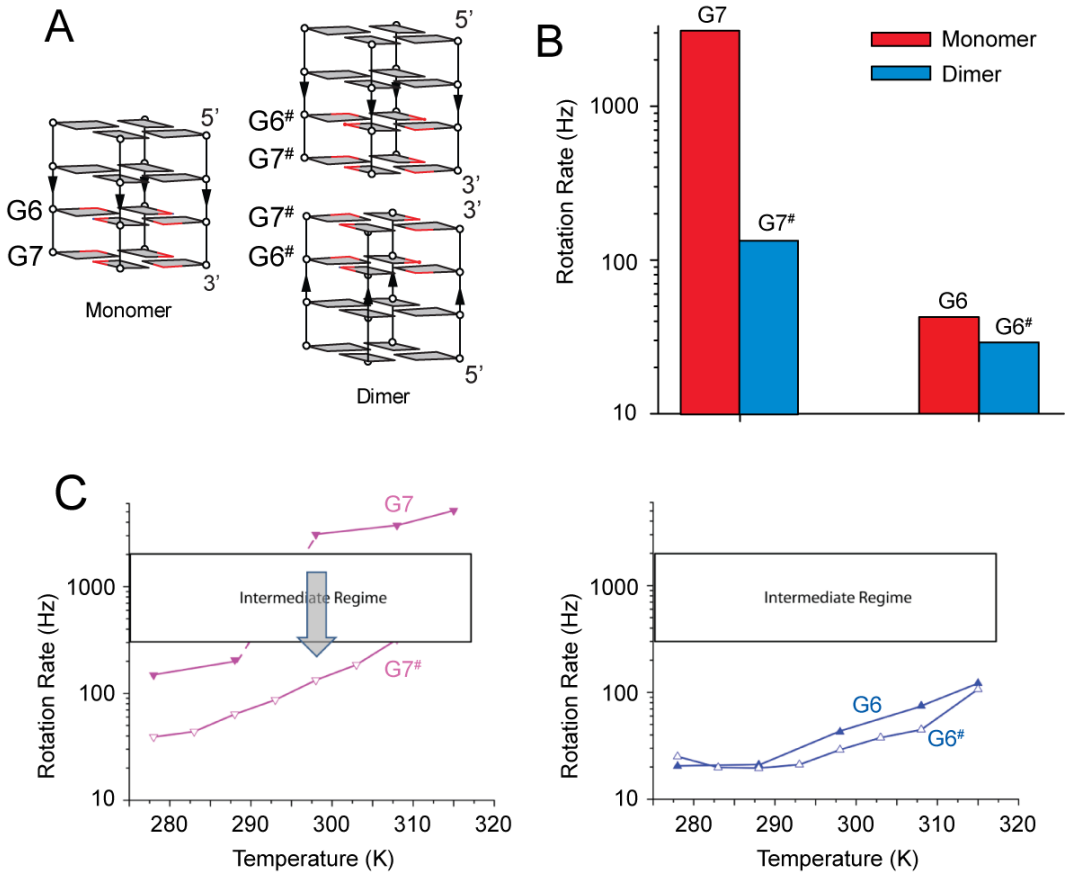


Figure 2.11: Stacking of G-quadruplexes affects amino group rotation. A) The schematic structures of *Platform G4* (left) and *Platform G4 Dimer* (right) are shown. The G-tetrad at the 3'-3' interface is highlighted (G7#), as well as one of the inner G-tetrad next to it (G6#). B) The amino group rotation rates of G6 and G6#, as well as G7 and G7# at 298 K are showing the bar chart. The data show a significant change for the amino group rotation rates at the stacking interface G7, as compared to that of the inner G-tetrad G6. C) The change in G7 and G6 amino proton rotation rates to its dimer counterparts G7# and G6# respectively.

within G-quadruplex molecules. An intramolecular parallel G-quadruplex formed by a minisatellite sequence (designated *CEB25*) possesses a set of base-pairs which anchors a 9-nt loop to the flanking residues. In contrast to a G-quadruplex containing a long loop, *CEB25* is thermodynamically more stable due to the formation of the base-pairs (Figure 2.12A) [126]. The rotation rate of an amino group in *CEB25* is generally dependent on its G-tetrad location. Nonetheless, two upper-tetrad guanines (G7 and G23) were observed to have low amino rotation rates (each exhibited

two separated amino peaks) relative to those of other outer-tetrad guanines (Figure 2.12C). The solution structure of *CEB25* reveals an A2-T18 base pair to stack onto the G3-G7-G19-G23 tetrad in such arrangement, where the base pair hovers on top of the amino groups of G7 and G23 (Figure 2.12B). This indicates the rotation of amino groups is influenced by the local structure arrangement, specifically by base pair stacking in this case.

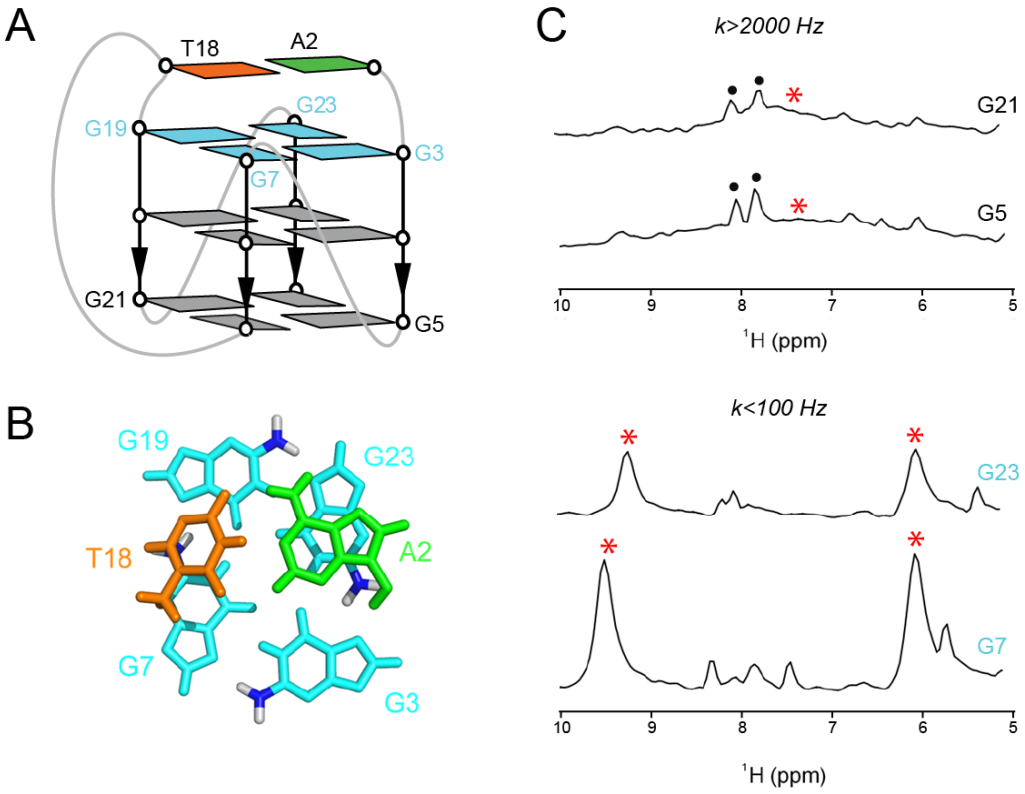


Figure 2.12: Local arrangement of a base pair affects amino group rotation. A) The schematic of the *CEB25* G-quadruplex. G-tetrad of interest (consisting G3, G7, G19 and G23) is indicated and shown in cyan. Paired adenine (A2) and thymine (T18) bases are drawn in green and orange respectively. Other bases are not drawn for clarity. B) Solution structure of the *CEB25* G-quadruplex showing A2-T18 base pair stacking onto G3-G7-G19-G23 tetrad (PDB ID:2LPW) [126]. C, D) Amino proton spectra and estimated rotation rates of (C) G21 and G5, as well as (D) G7 and G23. Observable amino proton peaks are marked by asterisks.

c. Ligand binding

Our methodology could also be applied to G-quadruplex-ligand complexes such as the G-quadruplex-L2H2 complex between a human telomeric G-quadruplex and a macrocyclic hexaoxazole (L2H2). The solution structure of the complex was recently reported and the complex was used as a platform for drug design [133]. In this complex, the ligand (L2H2) binding is facilitated by π-stacking interaction onto G3-G9-G17-G21 tetrad (Figure 2.13A) [133]. Amino proton spectra of the outer tetrad guanines (G3, G9, G17, G5, and G15) revealed dramatic differences between the amino rotation rates of ligand-associated guanines and the opposite site guanines (Figure 2.13C). In the free DNA, guanines are characterized with high rotation rates. In the complex, all guanines of the top tetrad are positioned directly under the ligand aromatic rings in the ligand-bound complex (Figure 2.13B), apparently decreasing their amino group rotation down to comparable rates as of those in the middle tetrad, whereby each amino group show two strong separated peaks (Figure 2.13C). Therefore, amino group rotation may also be correlated with the dynamic interaction between G-tetrad and the binding ligand.

d. Additional hydrogen bonds

The effect of hydrogen bond formation on the amino group rotation rate is examined on two different DNA structures, the first structure is a recently discovered left-handed G-quadruplex structure (designated *Z-G4*), and the second structure is a unique septad-forming quadruplex structure (designated $(GGA)_8$) [12, 132].

In Z-G4 (Figure 2.14A), the formation of hydrogen bonds between the O4' atom of a thymine sugar and one amino proton of an outer tetrad-bound guanine (e.g. G5) was reported (Figure 2.14B) [12]. The peaks of the two amino protons in this tetrad are well separated and are relatively sharp, indicating a lower exchange rate

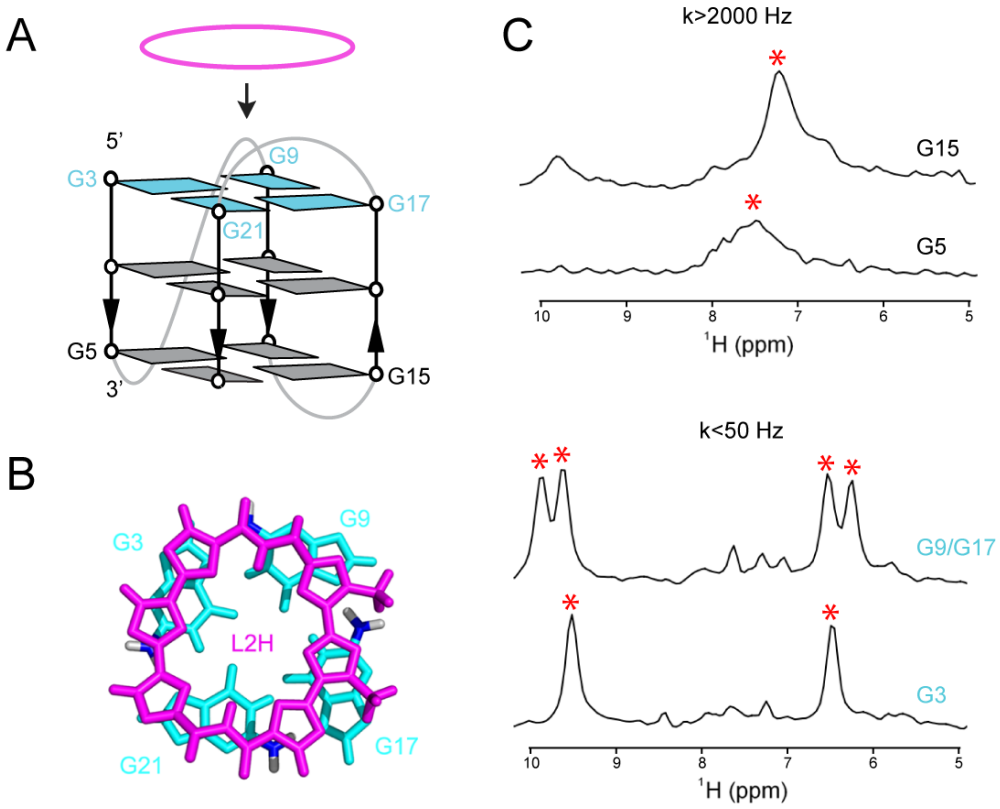


Figure 2.13: Ligand binding affects amino group rotation. A) The schematic of a telomeric G-quadruplex being targeted by the ligand L2H. Guanine residues at binding interface (G3, G9, G17 and G21) are indicated and drawn in cyan. L2H is represented by purple-colored ring. Other residues are not drawn for clarity. B) Solution structure of the complex showing L2H binding on G3-G9-G17-G21 tetrad (PDB ID:2MB3) [133]. C) Amino proton spectra and estimated rotation rates of G3, G9 and G17 in bound site and of G5 and G15 in free site are displayed. Observable amino proton peaks are marked by asterisks.

between them, in contrast to the low intensity peaks observed for amino proton in the inner tetrad of the same structure (Figure 2.14C). This effect is hypothesized to be due to the overall increase in energy required to break the additional hydrogen bond to facilitate a rotation event. We performed qualitative comparisons for different guanines of this structure at 298 K and found the amino groups of the outer-tetrad guanines (∼12 Hz) to have significantly lower rotation rates compared to that of those in the inner tetrads (∼2,400 Hz), in contrasts with the result of *Platform G4*. These results indicate that the hydrogen bonding effect on the amino

group rotation for the outer tetrad guanines in *Z-G4* is stronger than the tetrad stacking effect previously observed in the *Platform G4*. Furthermore, we note that the amino group rotation rate in the inner tetrads of the left-handed G-quadruplex *Z-G4* is higher than their counterparts in right-handed G-quadruplexes studied in this work.

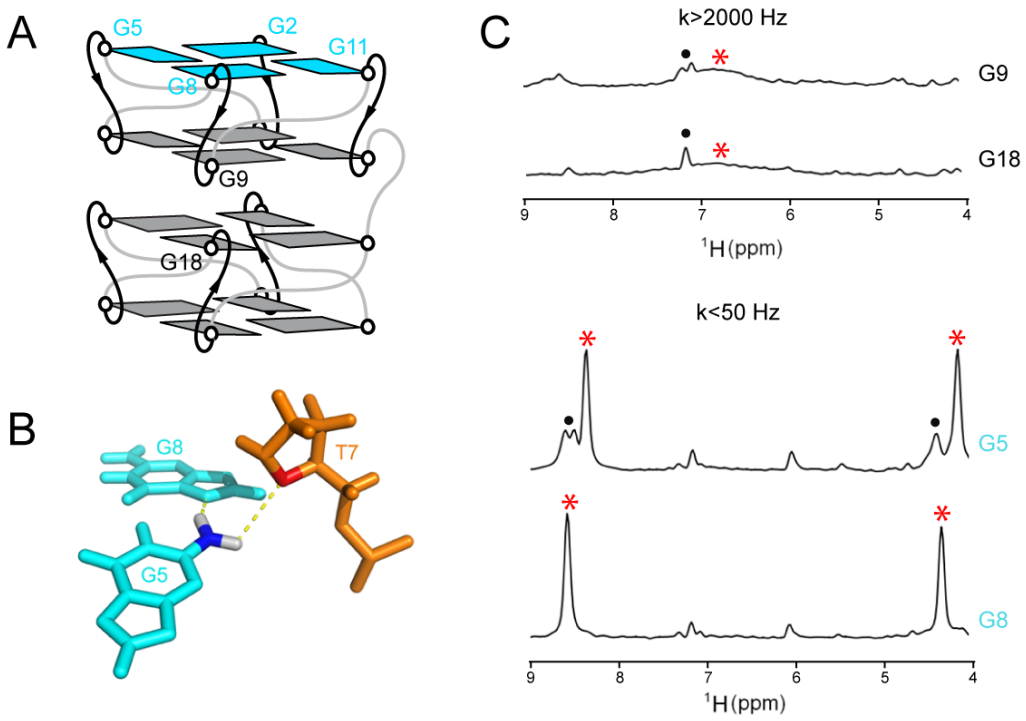


Figure 2.14: Additional hydrogen bond affects amino group rotation in *Z-G4*. A) The schematic of a left-handed G-Quadruplex, *Z-G4*. Guanine residues of interest (G2, G5, G8, G11) are indicated and drawn in cyan. B) The detail on the additional hydrogen bond that formed between G5 and T7 represented in stick figure [12]. C) Amino proton spectra of G5 and G8 (with additional hydrogen bond) are compared to those of G9 and G18 (without additional hydrogen bond). Observable amino proton peaks are marked with asterisks.

In $(GGA)_8$ (Figure 2.15A), there are two hydrogen bonds on the amino group of the inner-tetrad guanines. One of the protons is bonded to the nitrogen N7 of the neighboring guanine in the tetrad, while the other proton is bonded to another nitrogen atom of the side adenine. Based on the structure of the inner plane

(Figure 2.15B) [132], three guanines have their amino group situated next to the adenine aromatic protons (G1, G4, and G7), while one of the guanine has its amino group free (G10). A comparison of the amino rotation rate of the hydrogen bonded guanines (G7 and G1) to the free guanines from the outer tetrad (G11 and G23) has found the lower rotation rate on the hydrogen-bonded amino groups in G11 and G23 to fall in the intermediate regime, while G1 and G7 are in the slow regime (Figure 2.15C). This data supports our hypothesis that the hydrogen bond has a significant role in reducing the rate of amino groups.

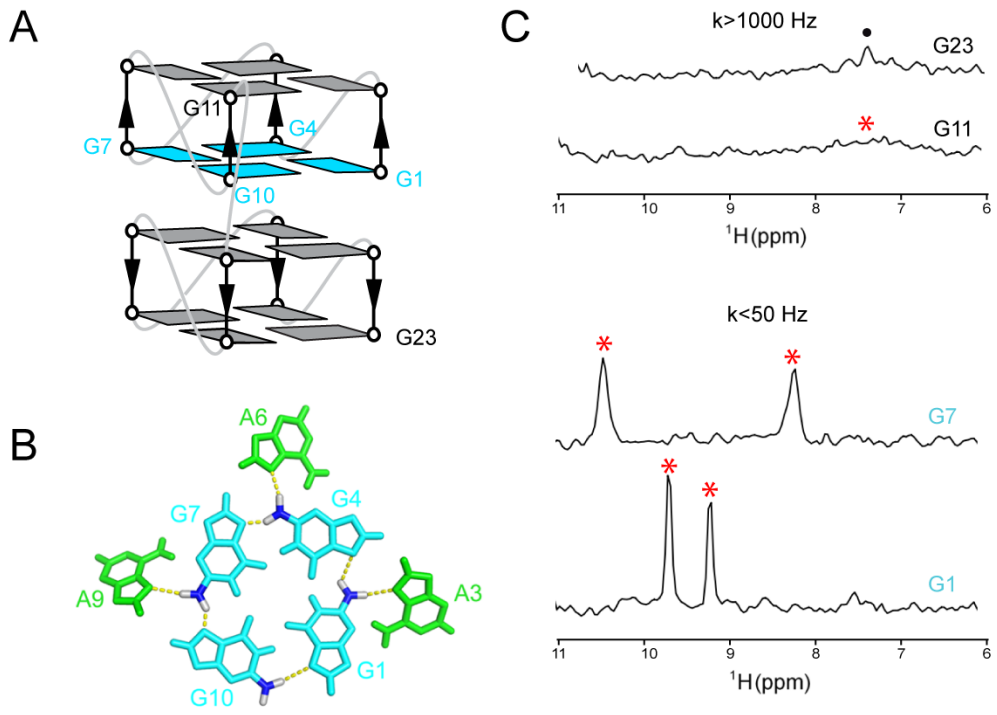


Figure 2.15: Additional hydrogen bond affects amino group rotation in *(GGA)₈*. A) The schematic of the *(GGA)₈* G-quadruplex, which form two septads in the middle. B) Stick figure of the middle septad of a *(GGA)₈* G-quadruplex solution structure solved by NMR spectroscopy [132]. Guanine and adenine are colored in cyan and green respectively. It shows an additional hydrogen bond formed between amino proton of a guanine to a nitrogen of the neighboring adenine. C) The comparison of NMR spectra between two guanines each possessing an additional hydrogen bond, located at the middle tetrad G7 and G1, and two guanines located at the outer tetrad G11 and G23. Observable amino proton peaks are marked by asterisks.

2.3.6 Potential application: detection of ligand-binding site

G-tetrad surfaces and grooves are the two main ligand binding sides reported for G-quadruplexes. We studied the effect of G-quadruplex-ligand stacking on the rate of amino group rotation, where the ligand binding would slow down the rotation of amino groups lying close to the ligand. In principle, groove-binding ligands act in the same way as any other residues that come to the groove, such as the adenines in $(GGA)_8$. Thus, steric hindrance or a hydrogen bond formed between the ligand and the guanine amino protons would slow down the rate of amino group rotation. Therefore, by simply observing the rotation rates of amino groups before and after the addition of the ligand and identifying the specific amino protons that undergo significant changes following the ligand addition, the determination of ligand binding site/s is possible.

2.4 Residence time of ligand in G-quadruplexes

This work is prepared for publication under the title: "Determination of residence time of G-quadruplex-binding ligands with NMR exchange analysis".

Residence time of G-quadruplex-ligand complex determines the effectiveness of G4 stabilization or any other functionality of the ligand. Since most biochemical activities are dependent on the interactions between ligands and their specific receptors, the lifetime of this binary complexes is thus highly related to their activity. The average enzyme-receptor lifetime in cells was determined to be in millisecond range ($\sim$50 ms) [116], alleging it suitable to be studied with NMR spectroscopy. Hence, we investigate the residence time of a ligand in G-quadruplex-ligand binary complex using the NOESY exchange technique.

2.4.1 NMR spectroscopy of a model G-quadruplex-ligand complex

A parallel monomeric G-quadruplex *Pu24T* (Figure 2.16C), d(5'-TGA GGG T GG TGA GGG T GGG GAAG G-3'), is used as the standard G-quadruplex receptor for the tetrad-binding ligand *TMPyP4* (Figure 2.16B). In K^+ solution, *Pu24T* forms a three-layered G-quadruplex, stabilized by a G-G-A triad in its 3' tetrad, resulting in 13 guanine imino protons observable by NMR spectroscopy [81]. On the other hand, the ligand *TMPyP4* is a small molecule in porphyrin family, which binds tightly to the tetrad planes of G-quadruplexes. However, due to the 3'-end triad capping of *Pu24T*, only 5'-end tetrad would be available as a binding site for the ligand. The structure of the *Pu24T-TMPyP4* complex has been studied before [81], giving us the imino proton assignment of both the free and bound G-quadruplexes (Figure 2.16A).

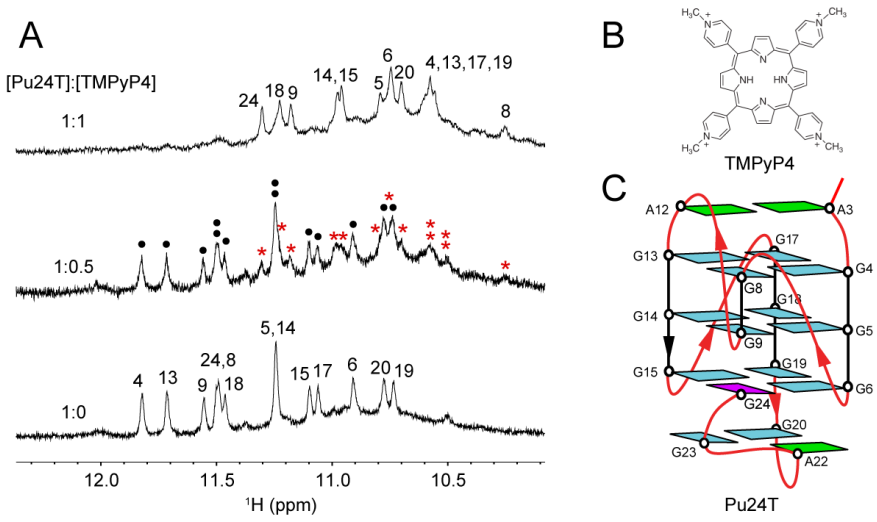


Figure 2.16: Titration of ligand to G-quadruplex DNA. A) 1D NMR spectra of *Pu24T* titrated with *TMPyP4*. All peaks are assigned based on Phan et al. [81], black dots and red stars indicate the peaks of free and bound *Pu24T* respectively. DNA strand concentration was 0.5 mM in K^+ solution. B) The chemical structure of *TMPyP4*, and C) the schematic of free *Pu24T* showing three-layered G-quadruplex with a stabilizing triad.

A titration study of *TMPyP4* into *Pu24T* was performed (Figure 2.16A), demonstrating the shift of all the imino proton peaks due to the binding of *TMPyP4* to *Pu24T* on its addition. At 0.5 mM concentration, a 1:1 ratio between G-quadruplex and its ligand brings most of the G-quadruplex to the bound state. The clear and highly resolved set of peaks in both the free (black dots in Figure 2.16A) and bound (red asterisks in Figure 2.16A) states of G-quadruplex allow us to demonstrate meticulously the use of NOESY exchange peak methodology.

2.4.2 Observation of chemical exchange cross-peak

The NOESY spectrum of *Pu24T-TMPyP4* complex shows clearly a set of exchange peaks in the imino proton region (Figure 2.17). Theoretically, there should be 13 cross-peaks corresponding to the 13 hydrogen-bonded imino proton from guanine residues. However, one of the cross-peaks corresponding to the exchange between the free and bound states of G20(H1) is too close to the diagonal, resulting in reduced visibility. Out of all the cross-peaks, the most shifted peaks are coming from G4, G8, G13, and G17, agreeable with the fact that these four guanines are located on the interaction site of *TMPyP4* [81]. For clarity and resolution purposes, we chose the cross peak of G13(H1) as the representative exchange peak between the two states of the G-quadruplex. As a remark, the same end results (residence time) were obtained when other resolved cross-peaks were chosen.

2.4.3 Estimation of dissociation rate using NOESY exchange analysis

To estimate the dissociation and association rates of the mixture, we focus our attention towards one slice of the NOESY spectrum. As explained previously, due to clarity and peak-resolution, we chose the peaks of G13(H1). In a NOESY

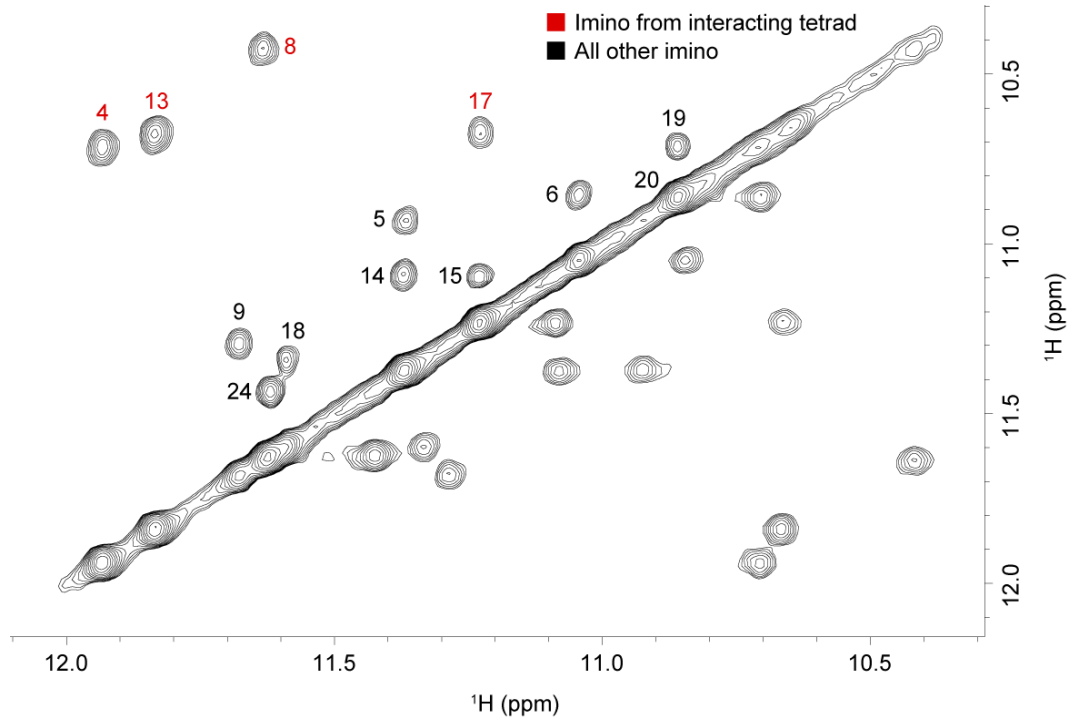


Figure 2.17: NOESY spectrum of *Pu24T-TMPyP4* complex. Exchange cross-peaks are labeled with guanine residue number, red indicates the guanine imino proton peaks of interacting tetrad, and black for all other guanine imino proton. Concentration of *Pu24T* was 1 mM and *TMPyP4* was 0.5 mM in K^+ buffer.

experiment, we isolated imino protons that precess in a certain frequency (in this case the frequency of free G13(H1) proton) by setting the evolution time t_1 to a specific value. Hence, before the mixing period, all G13(H1) protons that were in the free state were 'marked'. During the mixing time, part of the marked protons underwent exchange process and ended up as bound state protons at the end of the mixing time, giving rise to the cross-peak (AB). The other part of the marked protons remained in the free state and dictated the intensity of the diagonal peak (AA). In principle, the longer the mixing time, the more amount of marked protons will reach the bound state until equilibrium is established (detailed explanation in section 2.4.6).

The intensities of both the diagonal (AA) and cross-peaks (AB) in various mixing

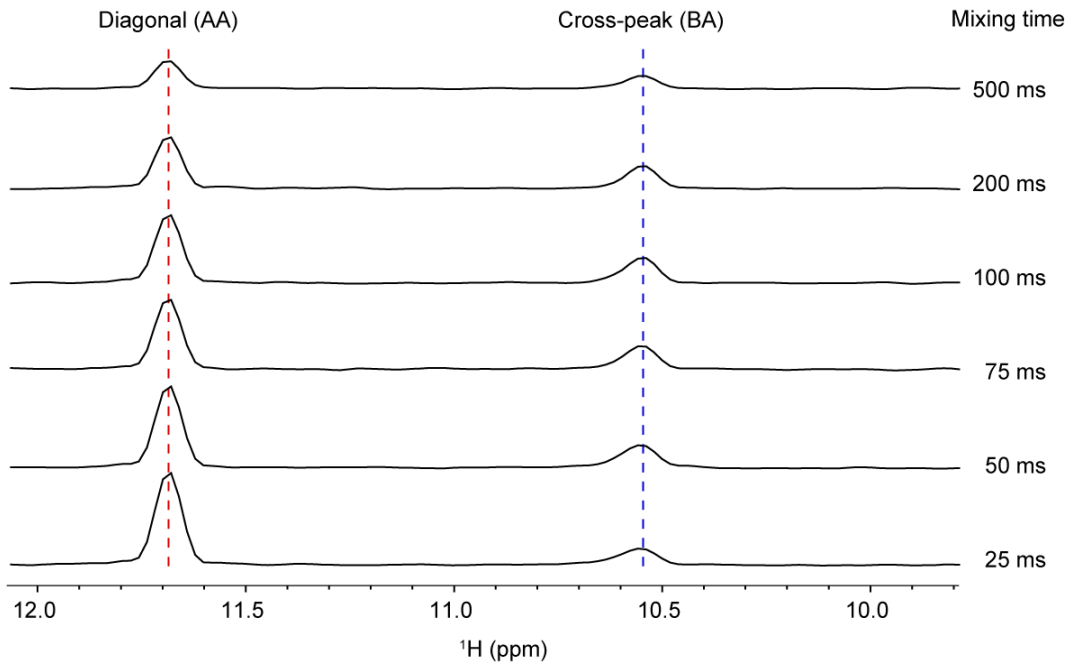


Figure 2.18: 1D projection of exchange-cross-peak and diagonal peak of NOESY exchange spectrum. *Pu24T* G13 imino proton diagonal and exchange peaks plotted along different NOESY mixing time. Red and blue dashed lines indicate the peak progression of the diagonal peak and cross-peak respectively.

times provide the information on the exchange process. For the diagonal peak, two decay processes are happening at the same time, the first is an intrinsic longitudinal relaxation process characterized by a time constant (T_1), and the second is an exchange process towards the bound state. On the other hand, for the cross-peak, in addition to the intrinsic relaxation process, a reverse exponential growth is also happening at the same time due to the exchange process from the free state.

Following the above reasons, we observed the diagonal peak intensity (AA) decayed with increasing mixing time due to relaxation (Figure 2.18 red), while the cross-peak intensity (AB) initially grew due to exchange process and ultimately decayed due to relaxation as well (Figure 2.18 blue). The plots of both diagonal and cross-peak intensities against mixing time can be found in Figure 2.19A-B.

In order to determine the rate of the exchange process, and the following residence

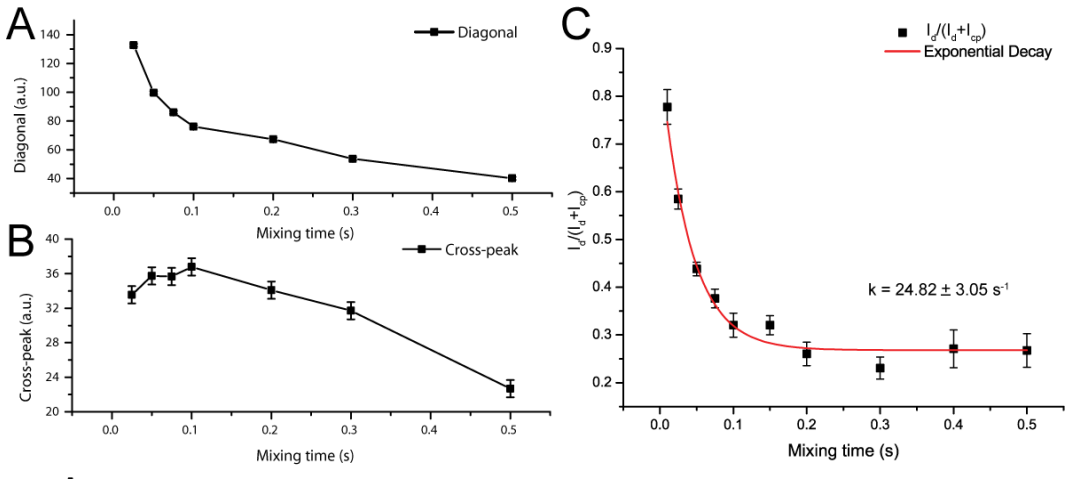


Figure 2.19: Estimation of residence time from NOESY exchange method. A) The diagonal (G13(H1)) peak intensity plotted against mixing time. Two exponential decay processes are expected corresponding to intrinsic relaxation time as well as exchange rate. B) The cross-peak intensity plotted against mixing time. A combination between exponential decay and growth are expected due to intrinsic relaxation and exchange rate respectively. C) Ratio of diagonal intensity to the sum of diagonal and cross-peak intensity plotted against mixing time. Red curve indicates exponential fitting of the data, giving the value of exchange rate to be $k = 24.82 \pm 3.05 s^{-1}$, corresponding to residence time of $t_{res} = 40.29 \pm 4.95 ms$.

time, we have to eliminate the effect of the intrinsic relaxation. Thus, a ratio must be taken between these two intensity values to cancel the relaxation term (assuming the relaxation of free and bound state to be the same). Moreover, sans the relaxation term, the sum of diagonal and cross-peak intensity is a constant expressing the amount of free-state in equilibrium condition. Therefore, by taking the ratio between the diagonal and the summation of both intensities, the value of the plateau in the fit (Figure 2.19C) carries a physical meaning of equilibrium mole fraction of the free state. The residence time of *TMPyP4* in *Pu24T* is determined to be ∼40 ms in the condition of 1 mM DNA and 0.5 mM ligand.

2.4.4 Factors affecting the residence time

Residence time is an important parameter that determine the effectiveness of drugs or ligands as it limits the running time of the biological function/s. In this section, we describe the possible factors that might affect the residence time of *TMPyP4* towards *Pu24T*.

a. G-quadruplex and ligand absolute concentration

In an equilibrium state of a solution, there always exist a mixture between the possible states of molecules. In our systems, we have three types of molecules i.e. the free G-quadruplex DNA (state A), the free ligand (L), and the bound DNA-ligand complex (state B). Without knowing or assuming the dissociation constant (K_D), we could only control the initial amount of the molecules. The absolute concentrations of both free DNA and ligand that we initially mixed were varied during the experiment, while keeping the concentration ratio between them constant. The results showed that with less amount of both DNA and ligand in the solution, the higher the residence time (Figure 2.20A, Table 2.3). The rate of dissociation was also plotted against the absolute DNA concentration (Figure 2.20C).

The variation range of absolute concentration of DNA was from 100-1200 μM while keeping the [DNA]:[ligand] ratio a constant 1:0.5. The variation of residence time in this range of concentration was found to be from ∼12 ms for the most concentrated sample (1200 μM DNA) to ∼66 ms for the least concentrated sample (100 μM). Extrapolating the results into nanomolar concentration would give us a residence time value of ∼170 ms, which is in the same order with biological processes such as base pair opening, secondary structure folding, and catalytic processes [96].

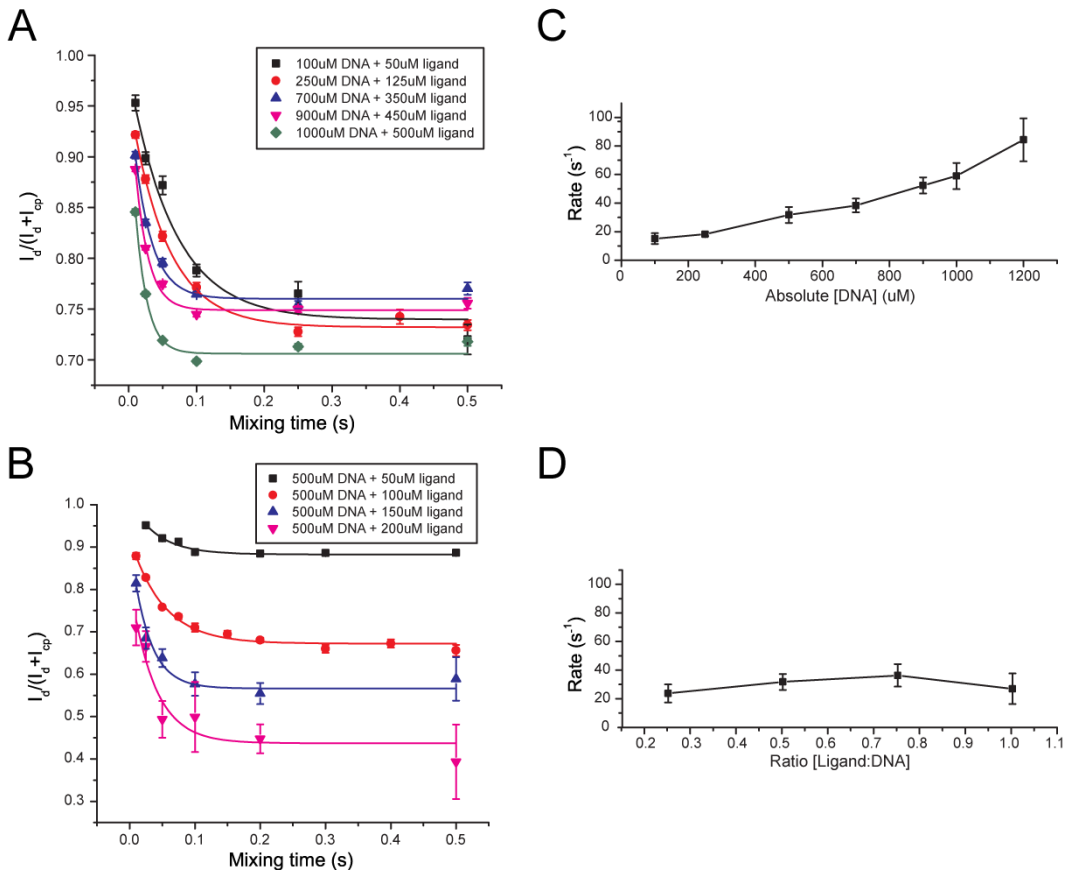


Figure 2.20: NOESY exchange analysis method for determination of the residence time of *TMPyP4* on *Pu24T*. A) Variation of absolute concentration of both the DNA and ligand while keeping the ratio between them constant lead to variation in exchange rate. B) Variation of ratio between DNA and ligand while keeping the absolute concentration of the DNA constant lead to similar exchange rate. Exchange rate plotted against C) absolute concentration and D) ratio between DNA and ligand.

b. Ratio of ligand to DNA

The second factor that could determine the residence time is the ratio between the DNA and the ligand inside the solution. As described earlier, the parameters that we can control are the initial concentrations of the DNA and ligands. To observe the effect of [DNA]:[ligand] ratio, the experiments were performed with constant DNA concentration, while varying ligand concentrations. The results showed negligible effect in the residence time with varying ratio (Figure 2.20B, Table 2.3). The rate

of dissociation was plotted against the concentration ratio (Figure 2.20D).

The variation range of [DNA]:[ligand] ratio was from 1:0.25 to 1:1. The DNA concentration was kept constant at 500 μM. The variation of residence time in these ratio range was from ~37 ms (1:1 ratio) to ~42 ms (1:0.25 ratio). The results suggest that the limiting factor for the residence time is the initial DNA concentration and not the ligand concentration.

c. Temperature

Temperature dependent measurements were done for verification purpose, since it is known that lower temperature would most likely increase the residence time of the ligand towards the DNA. Therefore, we measured the residence time of *Pu24T-TMPyP4* complex at two temperature points, 10°C and 25°C, and compared the decay curves produced (Figure 2.21). As expected, the residence time in lower temperature is ~2.5 times higher than the high temperature counterpart (~43 ms versus ~17 ms).

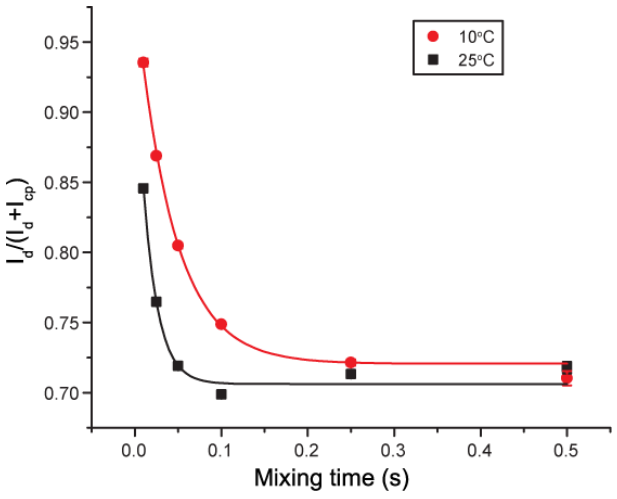


Figure 2.21: NOESY exchange analysis performed in two different temperatures. Intensity ratios in different temperatures show increasing exchange rate with temperature.

Table 2.3: Residence time of *TMPyP4* on *Pu24T* in different conditions determined by NOESY exchange method.

Varying [DNA]					
[DNA] (μM)	**[Ligand] (μM)**	**Ratio [DNA]:[Ligand]**	**T (°C)**	k (s^{-1})	t_{res} **(ms)**
100	50	1:0.50	25	15.16 ± 3.78	65.97 ± 16.43
250	125	1:0.50	25	18.22 ± 1.32	54.89 ± 3.98
500	250	1:0.50	25	31.74 ± 5.56	31.50 ± 5.52
700	350	1:0.50	25	38.34 ± 4.81	26.09 ± 3.27
900	450	1:0.50	25	52.37 ± 5.58	19.10 ± 2.03
1000	500	1:0.50	25	59.01 ± 9.15	16.95 ± 2.63
1200	600	1:0.50	25	84.35 ± 15.04	11.86 ± 2.11
Varying [DNA]:[Ligand] ratio					
[DNA] (μM)	**[Ligand] (μM)**	**Ratio [DNA]:[Ligand]**	**T (°C)**	k (s^{-1})	t_{res} **(ms)**
500	125	1:0.25	25	23.63 ± 6.40	42.32 ± 11.46
500	250	1:0.50	25	31.74 ± 5.56	31.50 ± 5.52
500	375	1:0.75	25	36.30 ± 7.89	27.55 ± 5.99
500	500	1:1.00	25	26.93 ± 10.69	37.13 ± 14.73
Varying temperature					
[DNA] (μM)	**[Ligand] (μM)**	**Ratio [DNA]:[Ligand]**	**T (°C)**	k (s^{-1})	t_{res} **(ms)**
1000	500	1:0.50	10	22.92 ± 1.13	43.64 ± 2.15
1000	500	1:0.50	25	59.01 ± 9.15	16.95 ± 2.63

d. Ligands

The same sets of NOESY exchange experiments were performed for two other compounds, Non-metalled 'medium arm' porphyrin (H_2MA) and Cobalt 'long-arm'

porphyrin (CoH_2LA) [134] by varying the initial concentration conditions. The chemical structure, the summary of results, and the residence time obtained are summarized below (Figure 2.22 and Table 2.4). The results show that increasing DNA concentration correspond to increasing dissociation rate for both compounds. Interestingly, there are enormous difference between the residence times of the two compounds, with t_{res} of H_2MA ($\sim$34 ms) being significantly lower than that of CoH_2LA ($\sim$140 ms) in 200 μM *Pu24T* concentration. This observation is hypothesized to be due to the long arm feature of CoH_2LA, which stabilizes the interaction and increases its residence time towards G-quadruplex.

Comparing the observed dissociation rates (Figure 2.23), the residence time of the three ligands were determined to be different from each other. At very low concentration, the residence time of *TMPyP4*, H_2MA, and CoH_2LA were determined to be $\sim$170 ms, $\sim$100 ms and $\sim$270 ms respectively, which come from the intersects of the linear extrapolations of the curves (detail in section 2.4.6). In addition, different slopes in Figure 2.23 states the different behavior of each ligand with differing concentrations, where H_2MA is observed to be most affected by additional DNA, followed by *TMPyP4* and CoH_2LA respectively.

2.4.5 Probing the binding strength of ligand

To fully characterize binding strength and/or affinity of a ligand towards a receptor, multiple parameters were previously formulated, including dissociation constant (K_D), increase in melting temperature (ΔT_m), and the residence time (t_{res}). Dissociation constant describes the binding affinity, it defines the number of complexes formed in an equilibrium condition based on the initial amount of receptor and ligand. The change in melting temperature expresses the stability of the binding, it illustrates the higher energy required to break the folded structure due to the

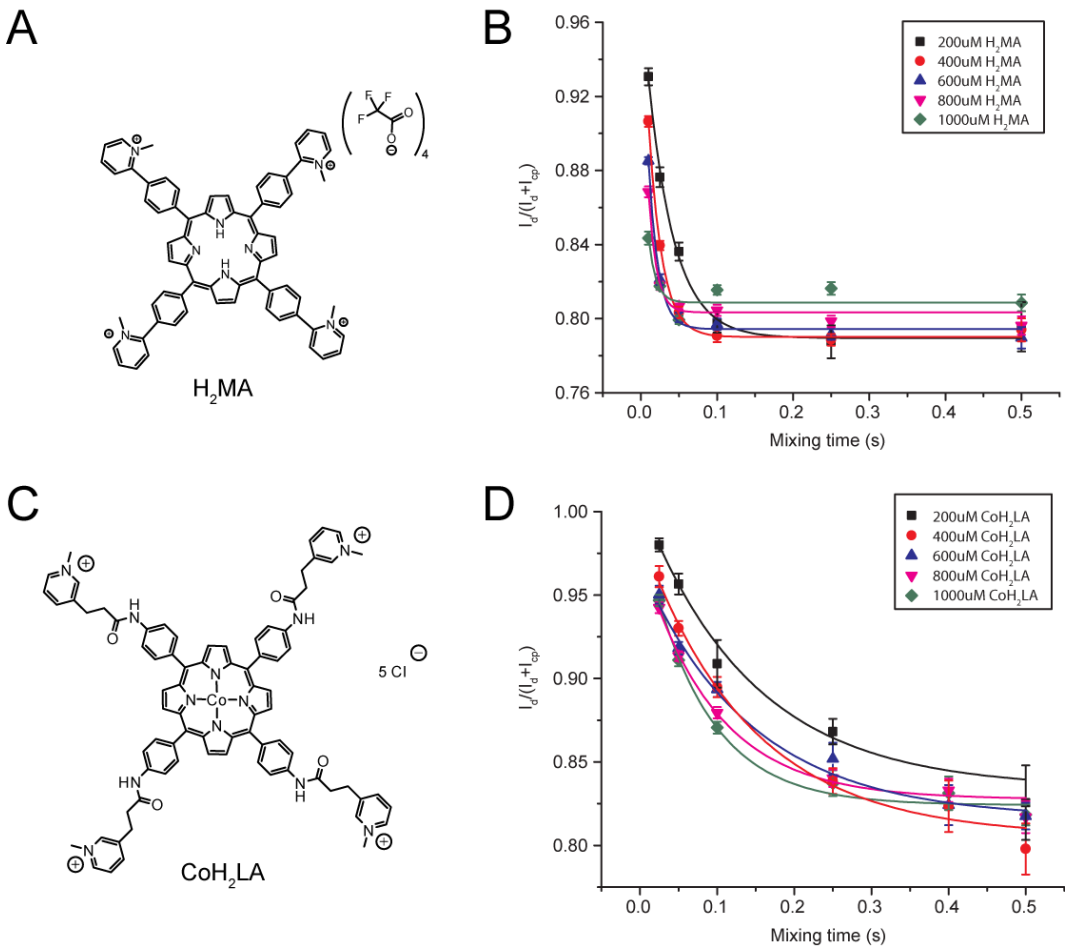


Figure 2.22: NOESY-exchange analysis method on two porphyrin variations H_2MA and CoH_2LA. Chemical structure of A) H_2MA and C) CoH_2LA compounds. The intensity ratios of B) H_2MA and D) CoH_2LA plotted against mixing time in different concentration of *Pu24T*, with constant ratio between the ligand to the receptor.

complex formation. Lastly, the residence time construes the time where the ligand and the receptor stayed in complex form, within which the biological function/s of the complex can be performed.

Having demonstrated the differences in the residence times for three different compounds, the NOESY exchange method is proposed to be an efficient and accurate way to measure residence times in the millisecond to second range. The method can detect accurately the residence times of the G4-ligand complexes given their

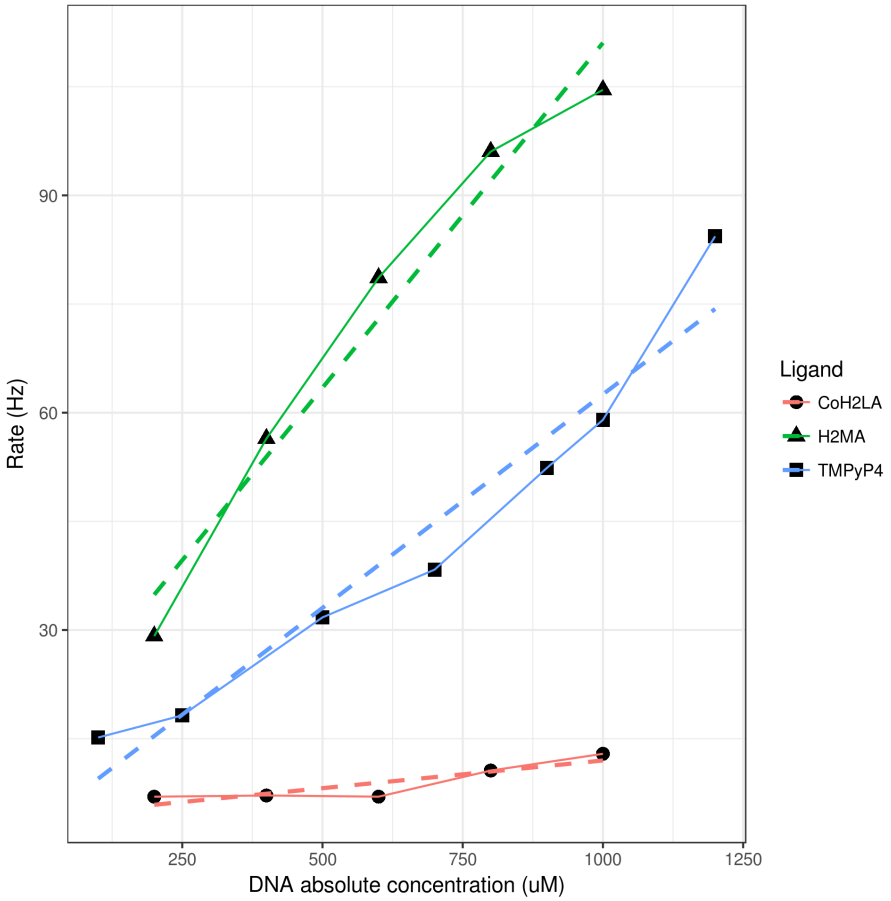


Figure 2.23: Linear approximation of association and dissociation rates of different ligands in different concentration conditions calculated with NOESY exchange analysis. Varying slopes are observed for each ligand indicating different behavior. Dashed lines indicate linear regression lines with blue, green and red indicate *TMPyP4*, H_2MA and CoH_2LA respectively.

concentration conditions under the limitation of similar relaxation time of the free and the bound receptors.

2.4.6 Preliminary model on receptor-ligand dynamics

In the context of ligand to receptor binding, the first order reaction approximation is the most commonly used model. Here, we used the one-step binding and one-step dissociation process [116] to build the mathematical model of the association and dissociation rate:

Table 2.4: Residence time of H_2MA and CoH_2LA on *Pu24T* in different conditions determined by NOESY exchange method.

H_2MA — Varying [DNA]					
Ligand	**[DNA] (μM)**	**[Ligand] (μM)**	**T (°C)**	k (s^{-1})	t_{res} **(ms)**
H_2MA	200	100	25	29.16 ± 2.35	34.30 ± 2.77
H_2MA	300	150	25	44.73 ± 3.36	22.36 ± 1.68
H_2MA	400	200	25	56.41 ± 2.06	17.73 ± 0.65
H_2MA	500	250	25	61.99 ± 1.75	16.13 ± 0.46
H_2MA	600	300	25	78.62 ± 8.21	12.72 ± 1.33
H_2MA	800	400	25	96.05 ± 17.40	10.41 ± 1.89
H_2MA	1000	500	25	104.57 ± 74.27	9.56 ± 6.79
CoH_2LA — Varying [DNA]					
Ligand	**[DNA] (μM)**	**[Ligand] (μM)**	**T (°C)**	k (s^{-1})	t_{res} **(ms)**
CoH_2LA	200	100	25	6.99 ± 2.02	143.08 ± 41.27
CoH_2LA	400	200	25	8.42 ± 0.70	118.71 ± 9.87
CoH_2LA	600	300	25	6.99 ± 1.62	143.00 ± 33.14
CoH_2LA	800	400	25	10.60 ± 1.10	94.33 ± 9.75
CoH_2LA	1000	500	25	12.91 ± 1.18	77.46 ± 7.06

$$A + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} B \tag{2.10}$$

with A, L, B, k_{on} and k_{off} represent free-receptor, free-ligand, receptor-ligand complex, association rate constant and dissociation rate constant respectively.

In our experimental system, the equilibrium is never disturbed, that is to say that the absolute as well as relative concentrations between A, L and B stay constant

throughout the experiment. Hence, we can safely assume that the rate constant k_{on} and k_{off} were invariable for each sample in a certain condition. However, these two 'constants' might not necessarily be unaffected by the change of experimental conditions such as absolute concentration, ratio between ligand to receptor, and temperature. Therefore, we separate the analysis into two parts, the first part analyzes a model of the spectra observables (intensities of diagonal and cross-peaks) in one specific experimental condition, and the second part investigates the model of the fitting observables (gained from exponential decay fitting analysis) throughout different conditions.

a. Model of spectra observables in one experimental condition

Although the equilibrium was not disturbed, observation can still be made by 'marking' all receptor molecules in a certain state (for simplification of explanation, let this state be the free state) just before the NOESY mixing period. The observation that follows, in the form of 1D projection of the 2D data, would only detect the marked molecules: whether they remain in free state (represented by the diagonal intensity (I_{AA})) or they have moved to the bound state (represented by the cross-peak intensity (I_{AB})). The kinetics of the marked molecule follow the first order rate equation:

$$\frac{d[A](t)}{dt} = -k_{on}[A][L] + k_{off}[B] \tag{2.11}$$

with $[A]$ represents the concentration of marked molecule remaining in the free state, $[B]$ represents the concentration of marked molecule in the bound state, and $[L]$ represents the concentration of free ligand which is a constant since equilibrium was not disturbed.

Conservation of the number of marked molecules gives the equation:

$$[B] = \chi_A A_0 - [A] \tag{2.12}$$

where A_0 is the total initial receptor concentration on the sample and χ_A is the mole fraction of the free receptor at equilibrium. The product of χ_A and A_0 indicate the initial free-receptor concentration, which is equal to the initial concentration of the marked receptors.

Solving equation 2.11 by substituting equation 2.12 yields:

$$[A](t) = \frac{\chi_A A_0}{k_{on}[L] + k_{off}} \left(k_{off} + k_{on}[L] e^{\left(-(k_{on}[L]+k_{off})t\right)} \right) \tag{2.13}$$

which can be further simplified with the following equation denoting the same forward and backward reaction rates in equilibrium:

$$k_{on}[L]\chi_A = k_{off}\chi_B \tag{2.14}$$

with χ_B denoting the mole fraction of the bound complex in equilibrium. Simplified version of equation 2.13 using equation 2.14 is a simple exponential decay equation:

$$[A](t) = \chi_A A_0 \left(\chi_A + \chi_B e^{\left(-(k_{on}[L]+k_{off})t\right)} \right). \tag{2.15}$$

Equation 2.15 denotes the amount of marked molecule in the free state as a function of mixing time, represented by the diagonal intensity (I_{AA}) in the spectra. However, the spectral intensities also contain the information of molecular relaxation which is not depicted in equation 2.15, and removal of this relaxation term would be necessary for the exchange analysis. Therefore, we took the ratio between the diagonal intensity of free state (I_{AA}) and the sum of diagonal and cross-peak intensities ($I_{AA} + I_{AB}$) for exchange rate fitting. The intensity summation carries a physical meaning of the total amount of marked molecules (in free and bound

states), which is conserved throughout the reaction, with a value of the product of χ_A and A_0. Hence, the intensity ratio:

$$\frac{I_{AA}}{I_{AA}+I_{AB}}(t) = \chi_A + \chi_B e^{\left(-(k_{on}[L]+k_{off})t\right)} \tag{2.16}$$

was used as a data point for each mixing time in a certain experimental condition. From each data point, an exponential curve was produced, and then an exponential decay fit was generated with the expression $k_{on}[L] + k_{off}$ as the decay rate.

b. Model of fitting observables across different conditions

The fitting results of part (a) were obtained in various experimental conditions. The variation of experimental conditions in this study came from three mutually exclusive parameters: (i) the overall concentration of receptor and ligand, (ii) the ligand to receptor concentration ratio and (iii) the temperature.

The subject of chemical reaction kinetics commonly regarded k_{on} and k_{off} as constants. However, our experimental data shows that the measurable decay rate k ($k = k_{on}[L] + k_{off}$), is increased upon increasing the overall concentration of receptor and ligand. Thus, we suspect that the two rate 'constants' are not really constant when exposed to different experimental conditions.

We propose a model where k_{on} is dependent on the concentration of both the free receptor and the ligand. The rationale behind the proposition is that the rate of association would be highly dependent on the amount of available ingredients for the association process to happen. Hence, we propose:

$$k_{on} = C_1 A^{\alpha} n^{\beta} \tag{2.17}$$

where α and β are numbers determining the dependency of k_{on} on both receptor and

ligand concentration, n is the ligand to receptor ratio and C_1 is a proportionality constant.

On the other hand, the dissociation rate is proposed to be independent on the concentration of receptor or ligand, as it only describe the process of detachment between two molecules in a complex. Hence, we propose:

$$k_{off} = C_2 \tag{2.18}$$

where C_2 is a constant.

Next, we aim to write the overall decay rate k in terms of overall receptor concentration (ligand and bound, A_0) and the ligand to receptor ratio. Assuming the dissociation constant be much smaller than the receptor concentration ($K_D \ll A_0$), both A and L can be written as:

$$A = (1 - n)A_0 \tag{2.19}$$

$$L = \frac{n}{1 - n}K_D \tag{2.20}$$

Substituting equation 2.19 and 2.20 into equation 2.17 and 2.18, the decay rate can be written as:

$$k = C_1(1 - n)^{\alpha-1}A_0^{\alpha}n^{\beta+1}K_D + C_2 \tag{2.21}$$

Experimental results show that k in the ∼100-500 μM concentration regime follows a linear dependency on A_0 and negligible dependency on n. Hence, the appropriate values of α and β would be 1 and -1 respectively, resulting in:

$$k = C_1 A_0 K_D + C_2 \tag{2.22}$$

The linear regression of the curves k versus A_0 generate slopes and y-intercepts for each ligand type. Applying equation 2.22, the slopes would be values of $C_1 K_D$ and the intercepts would be values of C_2. Ultimately, the value of residence time is the reciprocal of the y-intercept which is the time required for ligand-receptor complexes to dissociate in an infinitely dilute environment.

2.5 Conclusions

This chapter presents the development of two NMR-based techniques for detection of amino rotation rates and ligand dissociation rates or residence times in G-quadruplexes.

2.5.1 Summary

Lineshape analysis method is proposed to characterize the rotational dynamics of guanine amino groups in G4 nucleic acids. Residue-specific amino proton spectrum can be extracted from NOESY spectrum by taking advantage of strong intra-residue imino-amino proton NOEs. Using this cost-efficient method, stacking of multiple G-tetrads, local conformation, ligand binding, and hydrogen bond formation have been shown to affect amino group rotation. Additionally, the arrangement of G4 core stacking was also found to exert stronger restraints to amino group rotation as compared to that of double helix staggering. Based on the knowledge we have on the rotation rate behavior of amino groups, probing of localized dynamic processes is therefore possible.

NOESY exchange analysis method is proposed to quantify the residence time of

specific G4 ligands. The ratio between the diagonal peak to the summation of diagonal and cross-peak in NOESY spectrum is used to gauge the percentage of free G4 that stayed unbound after various mixing times. The results suggest that the absolute concentration of the ligand and G4 is affecting the ligand residence time in a linear fashion, while the effect of ratio between ligand and G4 is negligible. The method is proposed as a G4 ligand filter, with the ability to determine the compounds that fulfill a required threshold of residence time, in millisecond to second range, for their chemical or biological purposes.

2.5.2 Future works

The author has applied amino rotation determination with lineshape analysis to multiple G4 sequences to help determine their folding topology, including the fold-back anti-parallel G4 and trinucleotide repeat $(GGT)_8$ G4 explored in section 3.3 and 3.4 respectively.

The NOESY exchange method for determination of residence time has the following possible prospects:

- Intensive study on the effect of temperature towards residence time
- Ligand variations that goes to extreme residence times to probe the range of the technique
- Single-molecule experiments as alternative technique for residence time measurement
- Theoretical exploration of ligand kinetics, rigorous explanation on the physical meaning of the intersect and slope in the dissociation rate versus receptor concentration curves

Chapter 3

Structures of Novel G-quadruplex Systems

G-quadruplexes adopt diverse topologies [10], with the discovery of several novel topologies widening the possibilities of efficient drug targeting via site modification [135–137] and facilitating the development of nano-scaffolds via G4 engineering [138]. Some G4 ligands or helicases bind or unwind a certain G4 fold specifically [139], which raise the importance of G4 fold. The high-resolution structures provide precise measurements such as length and bond angles, which are useful in ligand or drug design. The structures also aid the field of G4 engineering, such as building G-wires [140], or performing chemical modifications on the sugar group and/or the loops to fulfil certain purposes [141].

Part of this chapter is taken from the author's unpublished work under the title of:

1. **Winnerdy, F.R.**, Heddi, B. and Phan, A.T. Solution structures of linear- and cyclic- dinucleotide (cGAMP) bound to a G-quadruplex.

2. Truong, T.H.A*, **Winnerdy, F.R.*** and Phan A.T. Discovery of an unprecedented G-quadruplex peripheral knot. * denotes equal contribution

3. **Winnerdy, F.R.**, How, J.Q and Phan, A.T. Left- and right-handed G-quadruplex hybrid structure of *(GGT)$_8$* trinucleotide repeats.

The experiments in these specific sections were performed by the author's colleague Dr. Brahim Heddi:

Subsection 3.2.2 — Titration of other guanine derivatives with *T4*

The experiments in these specific sections were performed by the author's colleague Dr. Truong Thi Hong Anh:

Subsection 3.3.2 — Resonance assignment of imino and aromatic protons in *AT21*

This chapter comprises of three novel G4 systems: the first is a unique G4-ligand system; second is a fold-back anti-parallel G4 with a stabilizing triad and base pair; and the third is a left and right-handed four-layered hybrid G4. NMR spectroscopy is the main technique used to probe all three systems, and as such, we will first delve into our standard procedure of NMR spectroscopy for solving high-resolution structure of G-quadruplexes.

3.1 Practical method: NMR spectroscopy of G-quadruplexes

The general workflow for determining G-quadruplex structures with NMR spectroscopy is presented in this section. In every subsection, one known quadruplex spectrum from the author's work is presented as an example. The detailed methodologies on sample preparation, preliminary UV and CD measurements, as well as detailed the procedures involved in structure calculation are presented in appendix A.

3.1.1 1D proton NMR

Imino protons (H1) of guanine residues forming a G-tetrad (Figure 3.1A) have a unique precession frequency range of 10-12 ppm when exposed to magnetic field [142]. This frequency range is unique due to hydrogen bond formation between the aforementioned protons and the oxygen of the carbonyl group in the neighboring guanines, as opposed to the formation of hydrogen bonds with the nitrogen of cytosines in the case of Watson-Crick base pairs, where the imino protons involved precess at 13-14 ppm [143] (Figure 3.1B).

The sequence *T95-2T* (d(5'-TT GGG T GGG T GGG T GGG T-3')), forms a

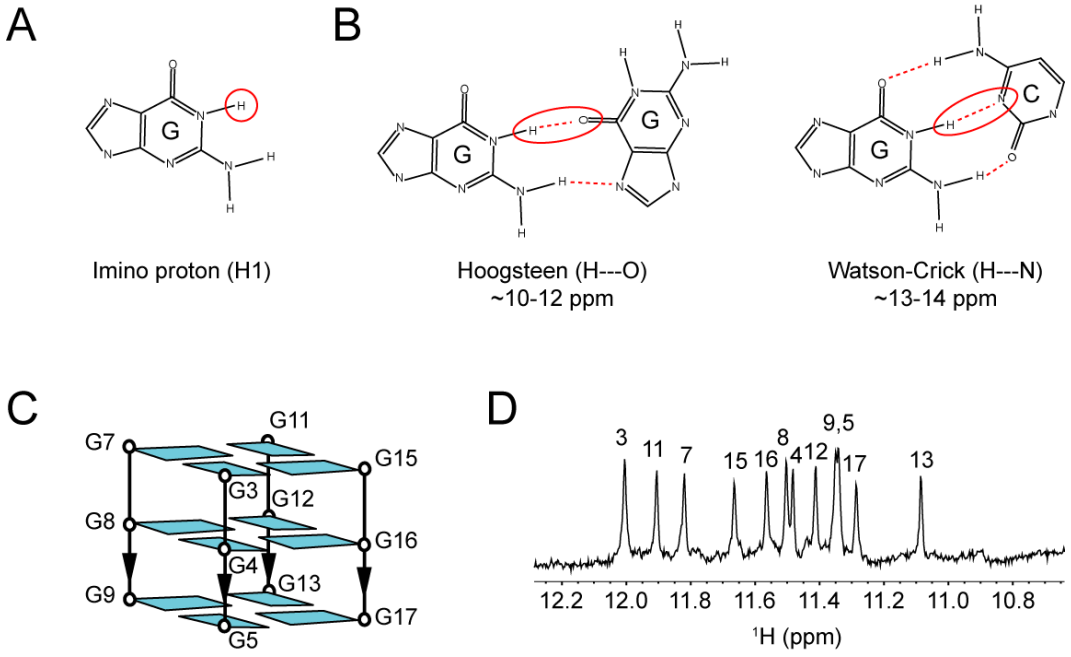


Figure 3.1: G-quadruplex 1D imino proton NMR. A) Guanine base, with imino proton indicated by red circle. B) G-G (left) and G-C (right) base pairs, guanine imino protons from G-tetrad and Watson-Crick base pair forming hydrogen bonds to oxygen and nitrogen atom respectively (red ellipses). C) Schematic of *T95-2T*, showing three-layered parallel G4 structure. D) 1D proton spectrum of *T95-2T* showing 12 distinct imino proton peaks, imino proton assignment is taken from [144].

monomeric parallel G-quadruplex in solution (Figure 3.1C) under specific conditions [144], where there are 12 guanine residues per molecule involved in G-tetrads formation. In the 1D ^{1}H-NMR spectrum, we observed 12 well-resolved peaks at ∼10-12 ppm, confirming the formation of a G-quadruplex (Figure 3.1D). The absence of humps or other smaller peaks also suggest the presence of a singular major conformation in solution. The parallel folding topology of the sequence is also confirmed by CD spectroscopy.

3.1.2 2D NOESY in H_2O solvent

Nuclear Overhauser Effect Spectroscopy (NOESY) is an NMR experiment that maps spatial range between two nuclei via their cross peak intensity [145]. For unlabeled samples, only ^{1}H-^{1}H NOESY is easily observable. The cross-peak intensity is correlated with the distance between two nuclei by, $I \sim 1/r^6$, where I is the cross-peak intensity and r is the distance between the two nuclei. Experimentally, cross-peaks are observable only if the two corresponding nuclei are close to each other ($<\sim 6.5\mathring{A}$) [146–150].

G-tetrads display a characteristic NMR spectral pattern in the imino (H1) to aromatic (H8) region of the NOESY spectra. Each guanine from a G-tetrad has its imino proton within the proximity of an amino proton, which is close to an aromatic proton of its neighboring guanine. NOESY intensities of the H1-H8 cross peaks are enhanced by diffusion pathways through amino protons [83]. Due to its cyclic nature, NOESY cross peaks coming from a single G-tetrad form a distinguishable pattern that we can observe in the 2D NOESY spectra (Figure 3.2A).

In theory, a three-layered G-quadruplex will have three H1-H8 cyclic patterns corresponding to each G-tetrad. Here, we use the *BS33* sequence (d(5'-TA GAGG C GGG A GTGG A GGG CG-3')) as an example. In Figure 3.2B, the cross peaks marked with the same color comes from the same tetrad. Note that in this example, we have assumed that we have obtained all the assignments for H1 and H8 of the guanines. 2D ^{1}H-^{1}H NOESY spectrum provides information on the groups of imino protons involved in the same tetrad, as well as the orientation of the guanines in the G-tetrad. H1 peaks coming from the same cyclic pattern belong to guanines of the same tetrad. Additionally, the cyclic direction of H1-H8 cross peaks determines the orientation of guanine residues within the tetrad (Figure 3.2C). Thus, the folding topology of G-quadruplex could be determined with this spectrum alone, provided

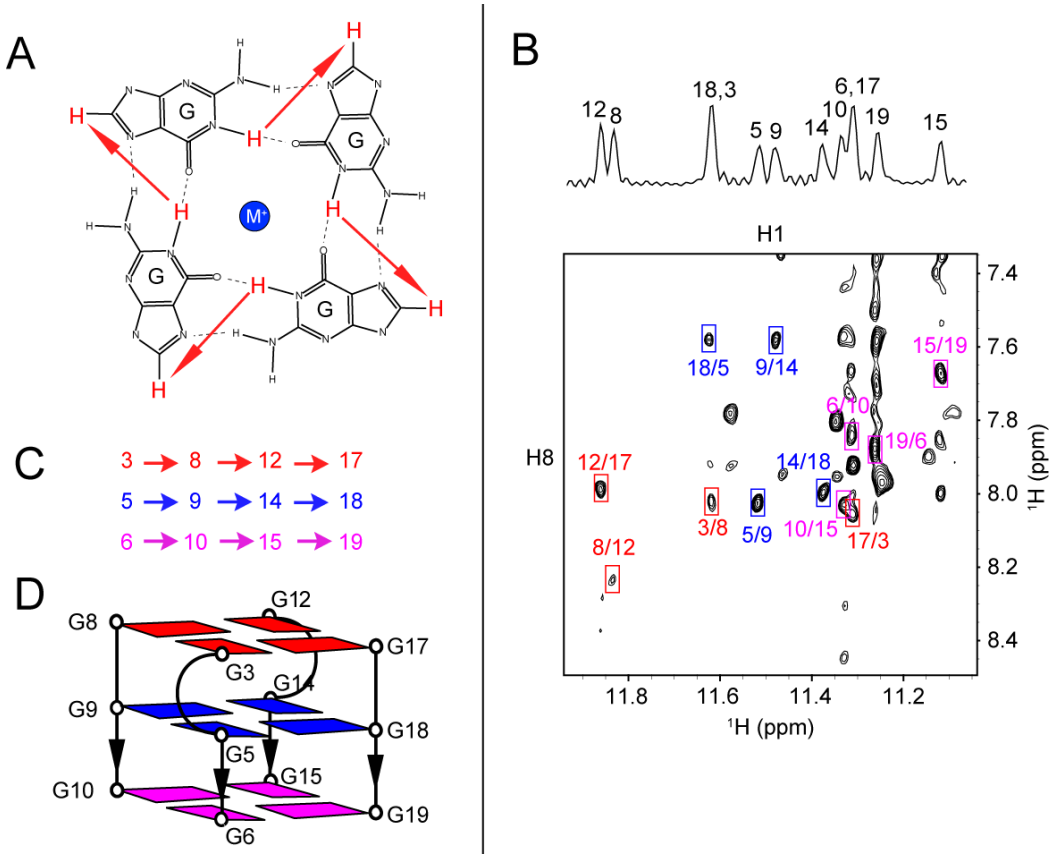


Figure 3.2: G-tetrad imino-aromatic cyclic connectivity pattern. A) G-tetrad schematic showing the proximity of imino protons to neighboring aromatic protons (red arrows) in a cyclic manner. B) 2D ^{1}H-^{1}H NOESY spectrum with the corresponding 1D spectrum on top. Cross-peaks between imino and aromatic protons from distinct G-tetrads are indicated by boxes of different colors. C) Cyclic orientation of the three G-tetrads of *BS33* derived from the cyclic pattern. D) Schematic of *BS33*. The colors red, blue, and magenta refer to the top, middle, and bottom tetrads of the G-quadruplex respectively.

that we have the assignments of H1 and H8 protons of all the participating guanines (Figure 3.2D). The rest of the cross peaks in the spectrum are used to generate distance constraints between the H1 protons and other protons.

3.1.3 Assignment of imino protons

Before the examination of 2D NOESY cross-peaks, it is ideal that the identity of the H1 peaks are known. However, since isotope labeling is required, it is prudent

that the assignment of the imino protons is carried out only when G4 formation is confirmed (via NOESY connectivity patterns). Assignments of imino protons are done by the low enrichment (1-2%) site-specific ^{15}N labeling technique [13]. Labeled guanines have ∼3-6 times the amount of ^{15}N as compared with that of unlabeled guanines (0.37% ^{15}N natural abundance). Therefore, ^{1}H-^{15}N HMQC experiment on each sample gives rise to a spectrum of a single imino proton peak, which corresponds to the labeled guanine in that sample (Figure 3.3). In this example, we use the sequence *BS33*.

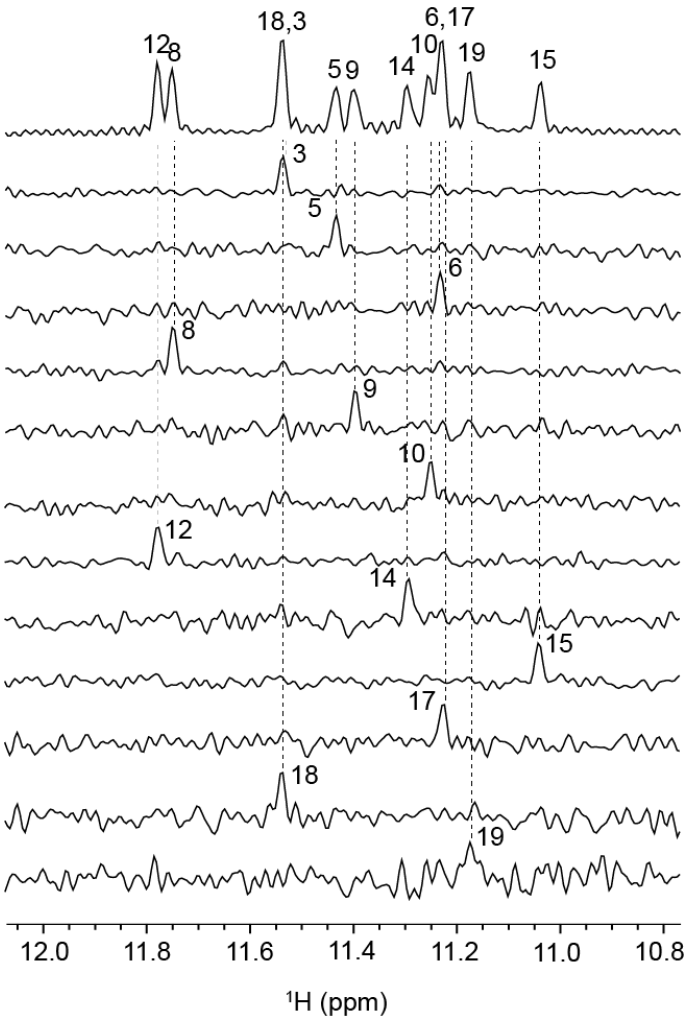


Figure 3.3: Imino proton assignment via ^{1}H-^{15}N HMQC spectra using 2% ^{15}N-labeled samples. Imino proton assignment of *BS33*.

3.1.4 Assignment of aromatic protons

Assignment of aromatic (H8) protons of guanines in the sample is done by performing ^{1}H-^{13}C HMBC experiments using parameters optimized for the magnetization transfer between intra-residue H1 to C5 (∼5 Hz J-coupling constant) and H8 to C5 (∼10 Hz J-coupling constant). By correlating two peaks with the same C5 chemical shift, and given the assignments of H1, assignment of H8 could be achieved [151] (Figure 3.4). In this example, we use the simple tetrameric G-quadruplex sequence *Platform G4* d(5'-TTAGGGGT-3')$_4$. Here, we can assign the two peaks from each group (H1 and H8) that correspond to the same carbon chemical shift value to be imino and aromatic protons from the same guanine residue.

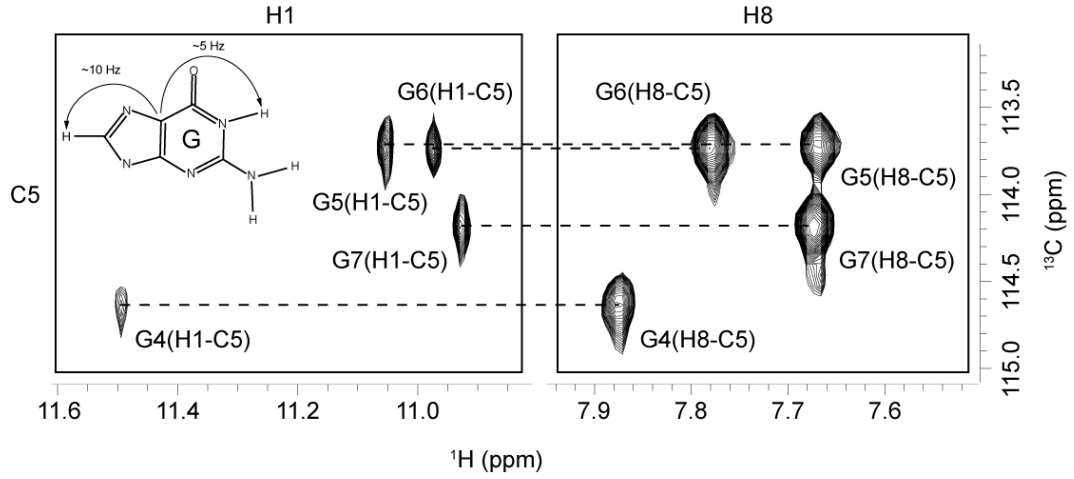


Figure 3.4: Aromatic proton assignment via ^{1}H-^{13}C HMBC spectrum. ^{1}H-^{13}C HMBC spectrum of *Platform G4*, showing through-bond correlation between intra-residue imino and aromatic protons via coupling with C5.

3.1.5 2D NOESY in D_2O solvent

Due to the water suppression scheme needed in NMR experiments for samples with H_2O solvent, the sugar proton peaks precessing around water proton frequency are diminished or unobservable [152]. The usage of D_2O solvent eliminates the need for water suppression, therefore resulting in clearer cross peaks at the aromatic and

sugar region. The only downside in utilizing D_2O solvent is that the H1 proton peaks are not observable since they can easily exchange with the deuterium of the solvent. Thus, both H_2O NOESY and D_2O NOESY complement each other to give all the distance information between protons.

One of the crucial regions to observe in D_2O NOESY is the aromatic (H8)-anomeric (H1') cross-peaks region. Inherently, every guanine has an H8-H1' intra-residue cross peak that appears in this region, as marked by the boxes in Figure 3.5A. Complementing that, based on the sequence of guanines as well as their glycosidic bond conformations, an H8 proton of one guanine could also be in close proximity with the H1' proton of the next and/or previous guanine (Figure 3.5B) [83]. Hence, using the various known sequential NOE patterns (e.g. H8(n)-H1'(n)-H8(n+1)-... in *anti-anti* guanines), the assignments of anomeric protons could be obtained. In this example, the *TP3-T6* (d(5'-TG GGG TCC GAGG CG GGG CTT GGG-3')) sequence is used.

Another vital and helpful information to be extracted from this NOESY region is the glycosidic bond angle of G-tetrad guanine residues. These angles are determined by looking at the intensity of H8-H1' intra-residue cross-peak: strong cross peaks suggest a *syn* conformation ($\chi = 60^o \pm 70^o$), while weaker cross peaks indicate an *anti* conformation ($\chi = 240^o \pm 70^o$) (Figure 3.5A,C and Figure 1.3 top).

3.1.6 Mapping of cross peak intensities to distance constraints

Having obtained the assignments of H1', the rest of the proton peaks could be assigned with the help of 2D TOCSY and COSY experiments, resulting in a complete identification of all the guanine peaks. Next, each NOE cross peak intensity

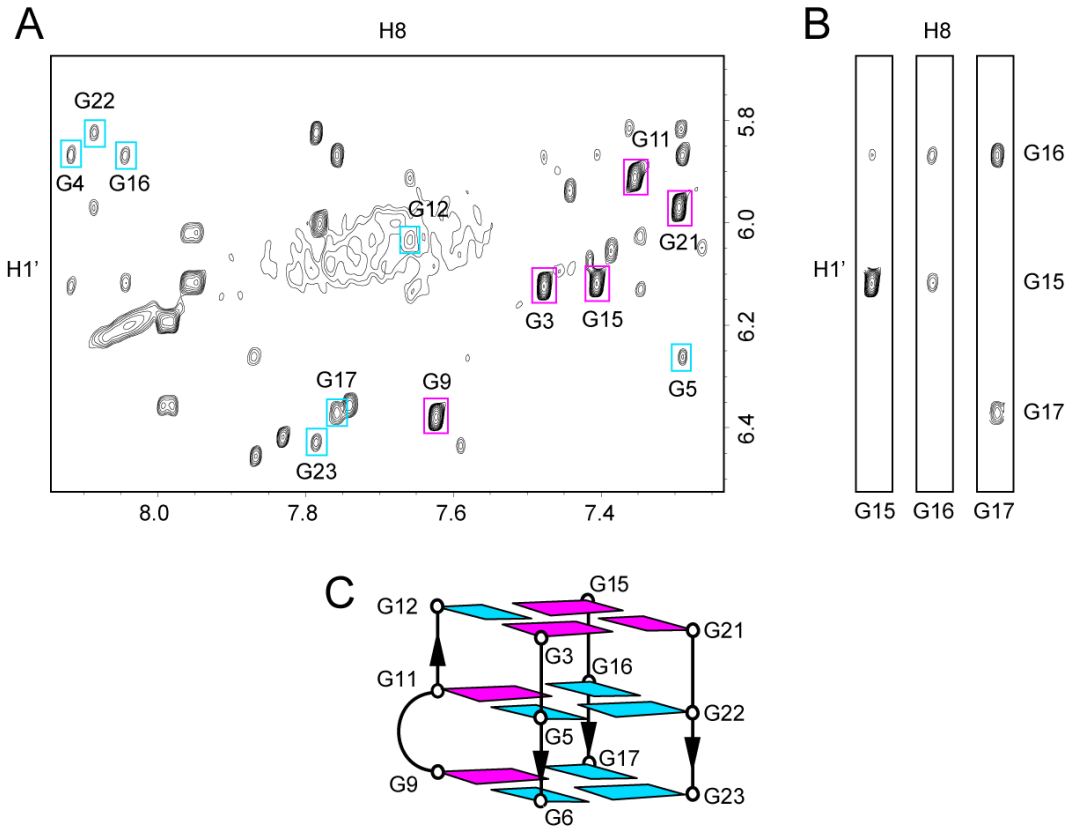


Figure 3.5: D_2O NOESY analysis of *TP3-T6* in the aromatic-anomeric region. A) 2D ^{1}H-^{1}H NOESY spectrum in D_2O solvent. Intra-residue H8-H1' peaks are marked with boxes, with cyan indicating an *anti* conformation, and magenta indicating a *syn* conformation. B) Strip view of G15, G16, and G17 aromatic protons, where we see the correlation between H8 and H1' of consecutive guanine residues. C) Schematic of *TP3-T6*; cyan indicates *anti*-guanines; and magenta indicates *syn*-guanines.

is mapped into distance constraints. Practically, all observed cross-peaks are categorized into several classes, from strong to weak, with each class corresponding to a distance range to be used as a parameter in structure calculation. Unless otherwise stated, the classification used for H_2O NOESY and D_2O NOESY in this study are stated in Table 3.1.

Table 3.1: Classification of NOESY peak intensities of G-quadruplex in H_2O and D_2O solvent.

H_2O NOESY		**D_2O NOESY**	
Classes	**Distance range ($\AA$)**	**Classes**	**Distance range ($\AA$)**
Strong	2.8-4.4	Strong	1.9-3.5
Medium	3.6-6.0	Medium	2.9-4.7
-	-	Intermediate	3.4-5.8
Weak	3.8-7.2	Weak	3.8-7.2

3.1.7 Structure calculation *in vacuo*

Amalgamation of all the NMR data provide information on the G-tetrad core composition and topology, glycosidic bond angles, as well as distance between protons. Furthermore, a combination of experimental data, oligonucleotide sequence as well as standard parameters and topology files (also known as force fields) [153] are used in annealing simulations and refinement. Out of 100 refined structures generated from the simulation, 10 of the lowest energy configurations are chosen.

Each resolved structure reported in this chapter is characterized by a structure table. The table consists of: the number of constraints used, grouped by the type of constraints; the number of violations, expressing the level of disagreement between the constraint inputs and the structure outputs; the deviation of the structures from ideal geometry; and the convergence levels between the 10 lowest energy structures.

3.2 Complex structures of incomplete G-quadruplex and dinucleotides

This work is prepared for publication under the title: "Interaction of (4n-1)-guanine G-quadruplex with metabolites: NMR structure of a quadruplex-dinucleotide complex".

3.2.1 4*n*-1 G-quadruplexes

Basic G-quadruplex putative sequences have been described previously to be $G_{3+}N_{1\text{-}7}$-$G_{3+}N_{1\text{-}7}G_{3+}N_{1\text{-}7}G_{3+}$, where a minimum of four G-tracts as well as less than seven nucleotides in every loop are required for G-quadruplex formation [18, 154]. However, the findings of other G-quadruplex forming sequences *in vitro* [126, 155–159] including involvements of bulges [160] and duplex loops [23, 161] have compelled people to rethink the possibility of expanding the definition of putative G-quadruplex (pG4) sequences.

Motivated by this fact, we tried to discover other unconventional pG4 sequences, explicitly sequences with one less guanine. In general, pG4 sequences contain multiple of four ($4n$) number of guanine residues, simply because four guanines are required to form one G-tetrad. However, G4 with less guanines in the G-tetrad cores had been found computationally [162] and experimentally [163–168] to exist as an intermediate of G-quadruplex formation. Based on that fact, we investigated a G-quadruplex structure formed by 11 guanines *T4* [89] ($4n$-1, $n = 3$), generated by substituting a guanine residue at position 4 to a thymine from one of the most stable G-quadruplex forming sequence *T95-2T* (d(5'-TT GGG T GGG T GGG T GGG T-3')).

a. Structure of T4

Substitution of position four guanine to a thymine were performed on *T95-2T* sequence, with the new sequence designated as *T4* d(5'-TT G**T**G T GGG T GGG T GGG T-3')). Circular dichroism (CD) experiment of *T4* showed a parallel G-quadruplex signal, indicating the same type of structure as the original sequence. However, it has a lower melting temperature of 42.5°C, a drastic decrease from its 12-guanine counterpart (∼80°C), nevertheless still high enough to be relevant in

biological environment. Complete structure calculation of this sequence has been done by the author's mentor, summary of which is presented below [89] (Figure 3.6).

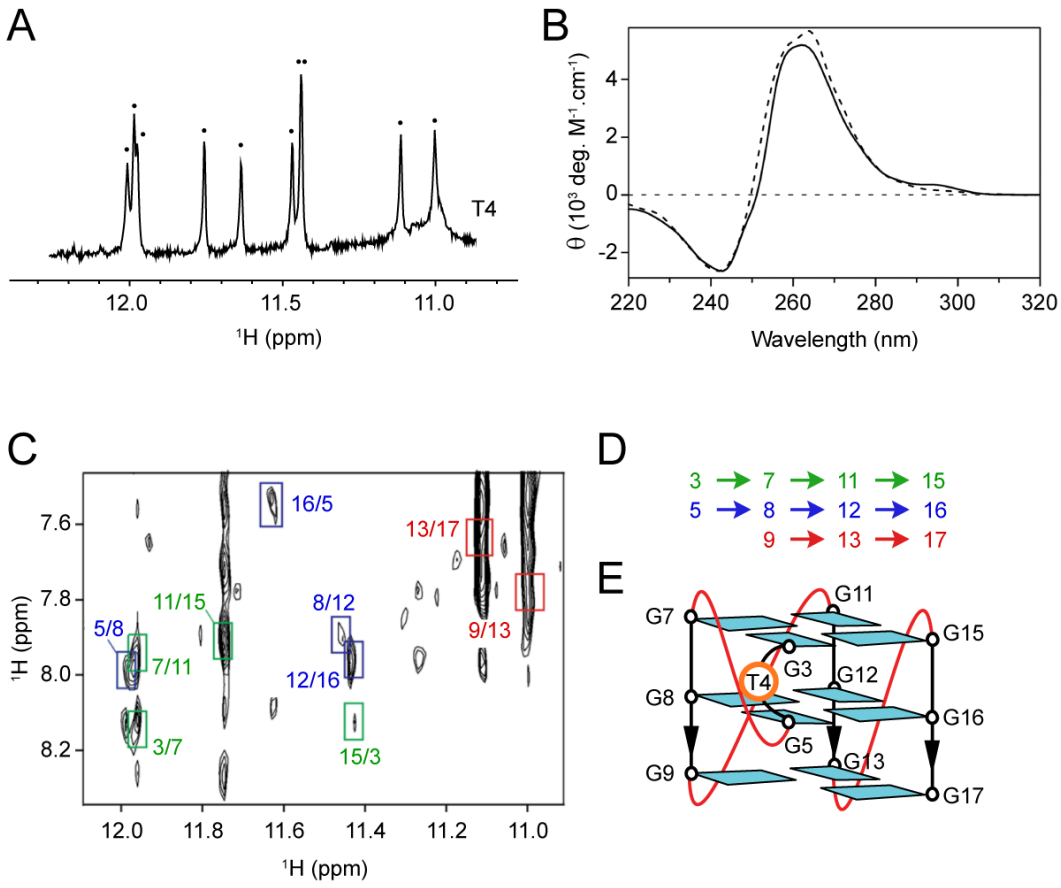


Figure 3.6: Summary of *T4* structural study. A) 1D NMR spectrum of *T4* 4, black dots indicate the 10 imino proton peaks. B) CD spectrum of *T4* (solid lines), showing parallel conformation. C) 2D NOESY showing cyclic-connectivity pattern of H1-H8 protons, different colors denote distinct G-tetrad. D) Orientation of guanines for each G-tetrad. E) Schematic of *T4* structures. All figures are adapted from [89].

The highlights of the structure are as follows: (1) it contains two layers of G-tetrad, (2) it has one thymine(T)-bulge between G3 and G5; and (3) it accommodates G-triad structure at the 3'-end with a vacant site. The last property opens the possibilities of metabolite or small molecule binding, leading us to a further study on the possible ligand interactions.

b. D4 sequence

In the solution structure of *T4*, we observed that the T-bulge has no importance in maintaining the stability nor plays any role in the loop arrangements. Hence, we decided to delete the T-bulge and coined a new sequence called *D4* (d(5'-TT GG T GGG T GGG T GGG T-3'). The melting temperature of *D4* is measured to be 48.1°C, rendering it to be slightly more stable than *T4* with $\Delta T_m = 5.6$°C. 1D proton NMR spectrum of *D4* was compared to that of *T4*, resulting in a very similar imino proton peak pattern (Figure 3.7). Hence, we assumed that the complete structure of *D4* is the same as *T4*, barring the T-bulge formation. The *D4* sequence is the selected sequence to be used in further study.

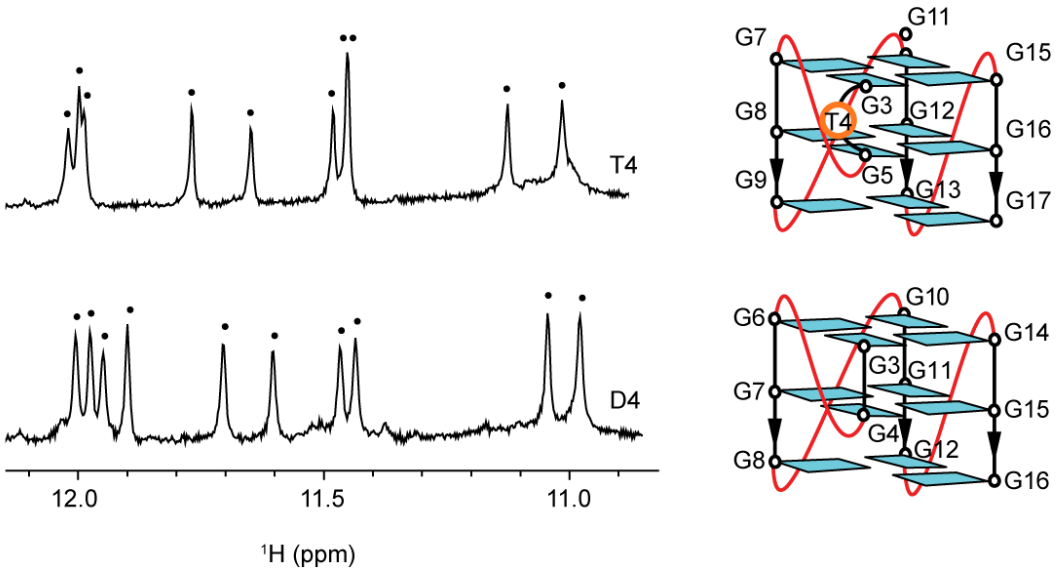


Figure 3.7: 1D NMR spectra of *T4* (top) and *D4* (bottom), with the respective structural schematics. Black dots in NMR spectra indicate the 10 imino proton peaks of *T4* and *D4*.

c. Other delta-sequences

To further examine the possibility and plausibility of the $4n$-1 G-quadruplex structure formation, deletions of one guanine in different tracts of *T95-2T* have been executed. The resulting sequences are named *D8*, *D12*, and *D16*, each with the

deletion of a single guanine residue in the corresponding position. The 1D proton NMR spectra for all three variants have at least two common features. Firstly, guanine imino proton peaks in G-quadruplex region were observed, suggesting the formation of G4. Secondly, more than 12 peaks occurred in the imino region, indicating multiple structure conformations in the solution (Figure 3.8A). Hypothetically, the two possible structures formed by *D8*, *D12*, and *D16* are parallel G-quadruplex structures, with vacant sites either in the 5'-end or 3'-end as illustrated in Figure 3.8B (*D16*). Regardless of which conformation they adopt, the formation of G-quadruplexes in these sequences indicate the possibility of many other 4*n*-1 G4 structures.

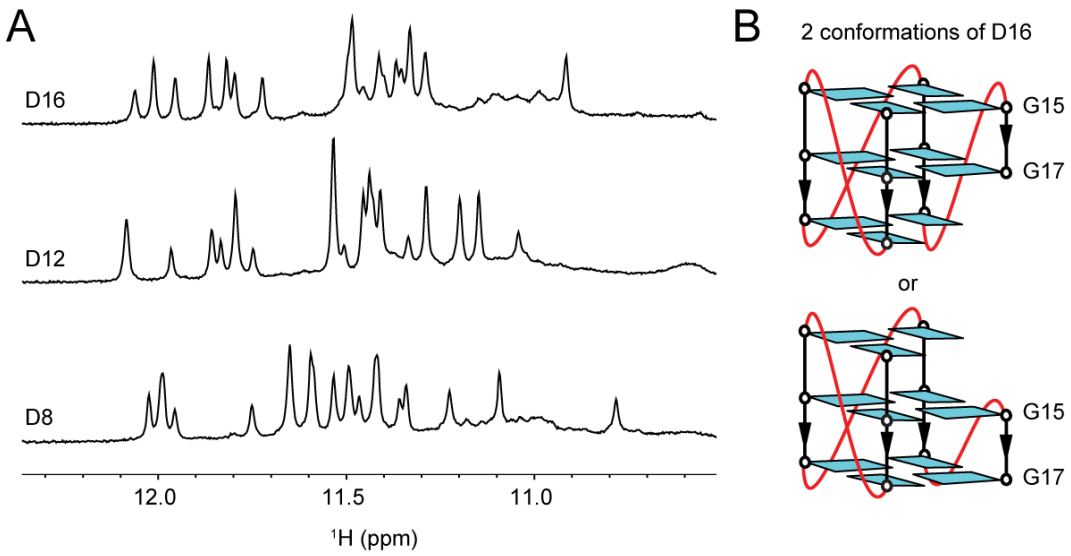


Figure 3.8: 1D NMR spectra and schematics of other 4*n*-1 sequences. A) 1D NMR spectra of *D8*, *D12*, and *D16*, each showing more than 12 unique peaks, indicating structure polymorphisms. B) Schematics of possible structure topologies of *D16* sequence.

In conclusion, it is affirmed that multiple 4*n*-1 pG4 sequences form G4 structures *in vitro*. Consequently, we might see further expansion on pG4 sequences in the future. Research in the vacant site of incomplete G-quadruplexes were done and presented in the next section.

3.2.2 Metabolites as G-quadruplex ligands

Ligand of G4 is a popular field of study ever since scientists realized that targeting G4s in telomeres might lead to a cancer treatment [62, 137, 169–176]. Throughout the years, ligands with higher affinity as well as specificity have been engineered in the field of drug design [39, 177, 177–183]. The most common site of G4 where small molecules, drugs, or peptides could bind onto is the planar interface of the G-tetrad [134]. The binding mode for these surface binders is usually $\pi - \pi$ stacking interaction between the aromatic rings of guanines and the ligands. Other than the interface, there has been a few other studies on G-quadruplex binding sites such as the grooves [184]. Herein, we present a novel ligand binding site, which is available only if the G-quadruplex has one less guanine residue. The specificity of the vacant site towards guanine base might prove to be a helpful feature for ligand design.

Nucleic acid derivatives are abundantly found inside cells as small molecule metabolites, the amount of which is comparable to protein as well as carbohydrate derivatives [185]. Cyclic dinucleotides, specifically, are found to have several important roles in cell signaling such as being secondary messengers in bacteria [186–190]. Due to their functions, it is important to maintain a certain concentration of these metabolites inside the cell at all time, as it may affect the signaling and ultimately influence the overall cell regulation. It is proposed that the presence of G-quadruplex as metabolite binders could have an effect in cell through riboswitch regulation [89]. Taken together, all mentioned facts motivated us to study the nature of G-quadruplex-metabolite binding.

a. Linear dinucleotides

Linear dinucleotides were chosen as the first potential G4 ligands due to their simplicity as guanine derivatives and their comparable size towards the aforementioned

cyclic-dinucleotide. Chemically, these molecules consist of two sugar-base groups, with connecting phosphate groups. We chose a few combinations of the two bases, with at least one guanine: d(AG), d(GA), d(GT), d(TG) and d(CG).

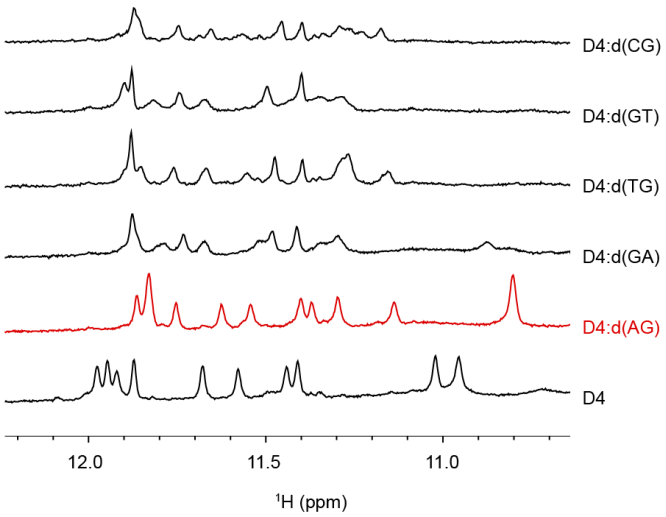


Figure 3.9: 1D NMR spectra of free *D4* and its mixtures with d(AG), d(GA), d(TG), d(GT), and d(CG) respectively (from bottom to top). All ratios of *D4*:dinucleotides are 1:4 equivalent. *D4*:d(AG) spectrum is drawn in red, showing 12 well-resolved imino proton peaks.

1D NMR spectra of the five complexes suggest that there certainly exists some interaction based on the imino proton peak shifts (Figure 3.9). Out of all the five combinations, *D4*-d(AG) complex had an especially well-resolved imino proton spectrum. Therefore, d(AG) was chosen to be the ligand of choice for structural study of *D4*-dinucleotide complex.

Titrations of d(AG) towards *D4* G-quadruplex in 100 μM concentration was performed (Figure 3.10), the results indicate clearly the formation of complete three-layer G-quadruplex. Initially, out of 11 guanine imino protons in *D4* sequence, only 10 are observable by NMR, due to the lack of hydrogen bond from one of the 3' triad guanines. However, with the introduction of the guanine-containing dinucleotide, a completion of G-triad into G-tetrad is expected, resulting in the appearances of two new imino proton peaks. Our results agreed with that conjecture

as 12 total peaks are observed in the imino region once excess d(AG) was added. In 100 μM concentration, a ratio of *D4*:d(AG) of 1:2 was needed for full transition from free *D4* to the complex.

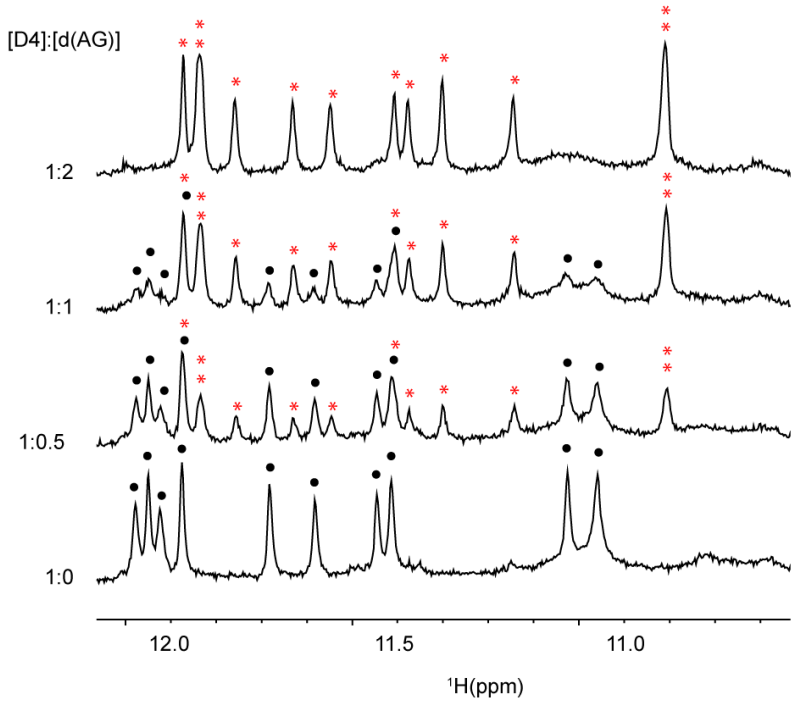


Figure 3.10: 1D NMR spectra from titration of d(AG) into *D4* oligonucleotide. Ratios used are 1:0, 1:0.5, 1:1 and 1:2. Black dots indicate the original 10 imino proton peaks from free DNA, while red star marked the rising imino proton peaks from the ligand-bound complex.

b. Cyclic-dinucleotides

Following the above results, we performed the titration experiment of *D4* with two cyclic-dinucleotides, cGAMP and cdiGMP. The two cyclic molecules were chosen based on their biological significance and availability. The results of the titrations were promising, albeit not as clean as d(AG) titration results, they still presented twelve resolved imino proton peaks on excess ligands (Figure 3.11).

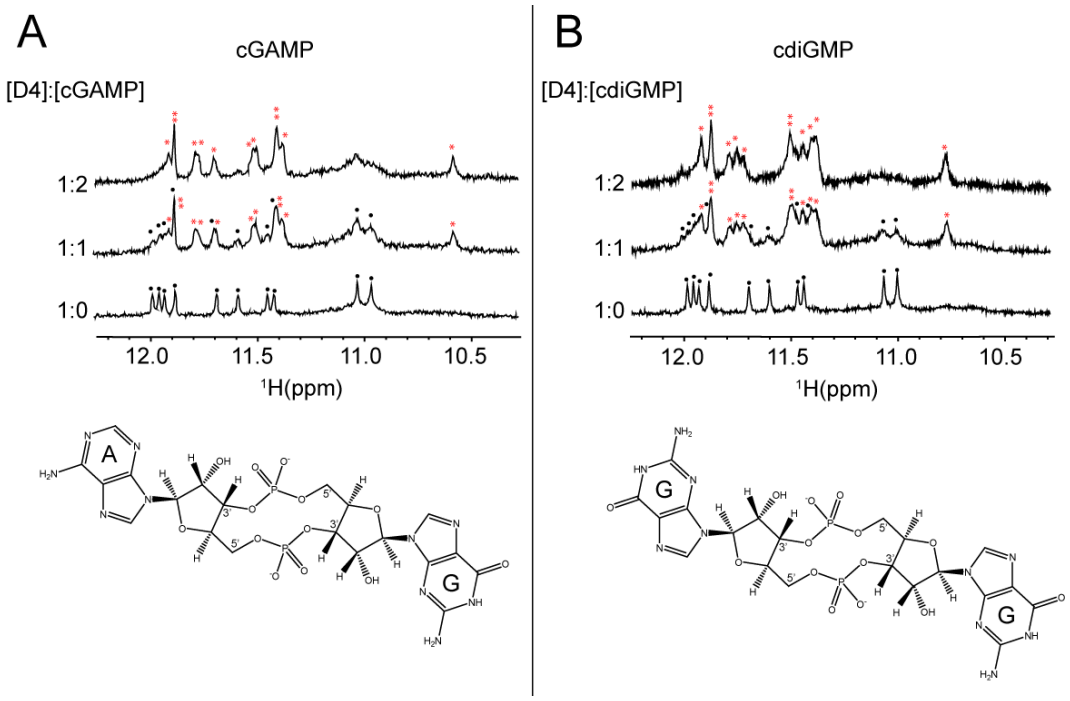


Figure 3.11: 1D NMR spectra (top) from titration of A) cGAMP and B) cdiGMP into *D4* oligonucleotide. Ratios used are 1:0, 1:1 and 1:2. Black dots indicate the original 10 imino proton peaks from free DNA, while red star marked the rising imino proton peaks from the *D4*-cGAMP and *D4*-cdiGMP complexes. The chemical structures of cGAMP and cdiGMP are shown (bottom).

c. Other guanine derivatives

Other guanine derivatives were also used as possible ligands that bind to the vacant sites of the 11-guanine G-quadruplex. The list of the small molecules titrated is the following: guanosine; guanine; inosine; guanosine tri-phosphate (GTP); isocanthopterin; and cytosine. The titrations of these molecules towards *T4* G-quadruplex were done by senior member in the group [89] (Figure 3.12).

The results show a wide range of binding affinity from these molecules towards the 11-guanine G-quadruplex, as well as some binding properties. To highlight a few, cytosine addition to *T4* requires 4-equivalent before really changing the *T4* original spectrum, while isocanthopterin destroys the triad structure of *T4* and keep the

rest intact in 1:1 ratio as can be seen in the disappearance of bottom tetrad peaks of *T4*.

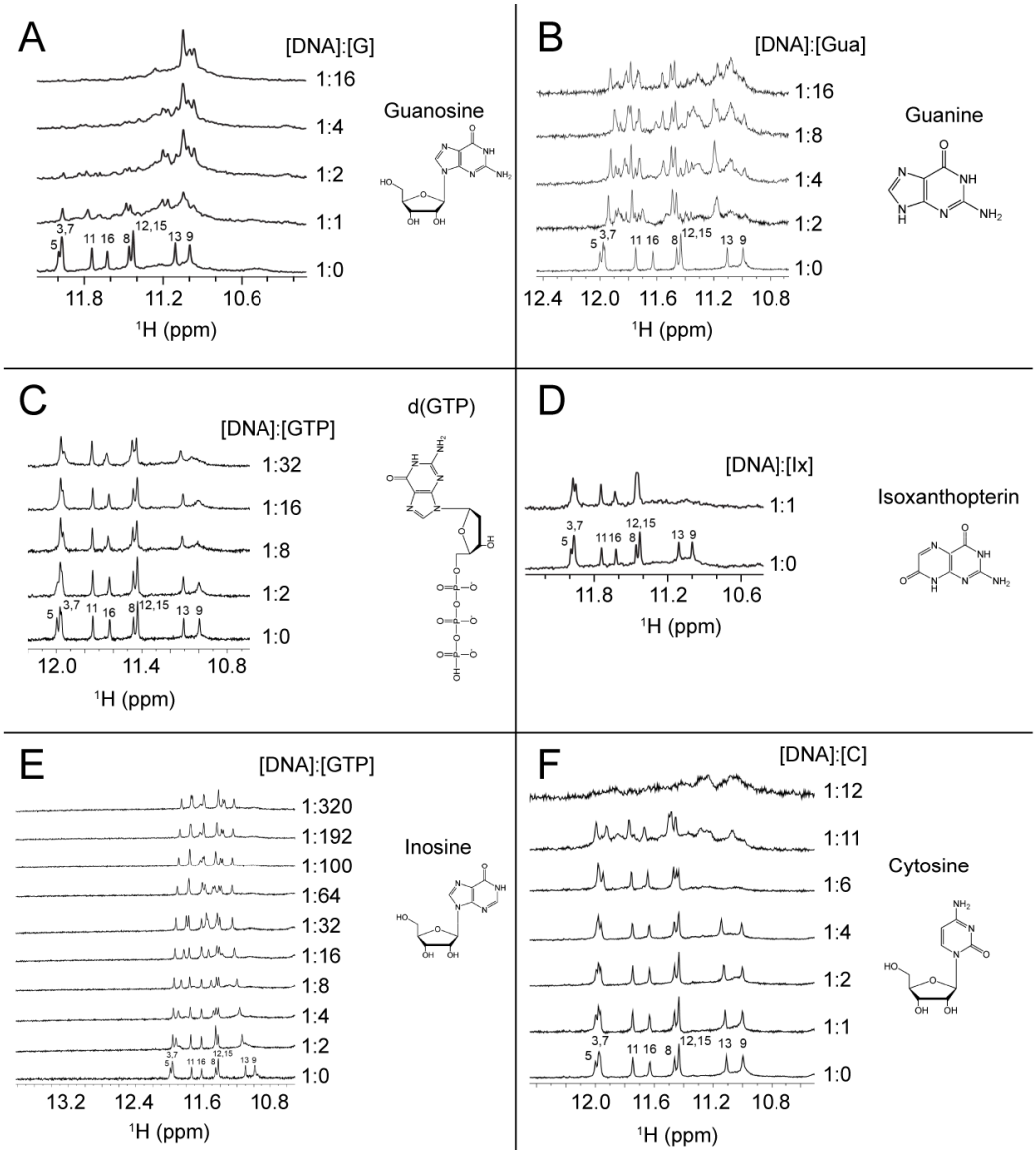


Figure 3.12: 1D NMR titration spectra of variety of small molecules into *T4*. The list of small molecules is, A) guanosine, B) guanine, C) guanosine tri-phosphate, D) isoxanthopterin, E) inosine, and F) cytosine. Chemical structure of each compound is drawn to the right of each spectrum.

Taken together, we deduced that incomplete G-quadruplex is a potential metabolite binder, especially in high metabolite concentration environment. The interaction

between cyclic-dinucleotides — important metabolites in cells — with vacant G-quadruplexes brings a possibility of gene expression regulation.

3.2.3 Solution structure of *D4*-d(AG) and *D4*-cGAMP

To comprehensively explore the interaction site of vacant G-quadruplex *D4* towards its ligands, we examined the high-resolution structures of *D4*-d(AG) complex and *D4*-cGAMP complex using NMR spectroscopy.

a. *D4*-d(AG) complex

The 1D imino proton spectra of the complex in 1:1.2 *D4*:d(AG) ratio are shown in room temperature (25°C) and lower temperature (10°C), with 1 mM *D4* strand concentration (Figure 3.13A). As previously shown, twelve imino proton peaks were observed from the complex structure, following the completion of G-triad into G-tetrad. Additionally, we observed one extra imino proton peak at ∼13 ppm indicating the existence of an additional base pair. In lower temperature, this peak is sharper, suggesting that the base pair formed is slightly loose and dynamic in nature.

The assignment of imino protons, aromatic protons, and all the other protons are done with methods described in section 3.1: ^{1}H-^{15}N HMQC, ^{1}H-^{13}C HMBC, and ^{1}H-^{1}H NOESY. Multiple 2% ^{15}N enriched samples were used for ^{1}H-^{15}N HMQC experiments (Figure 3.14), while an unlabeled sample was used for any other NMR experiments. The summary of the results is presented, together with the proposed schematic based on all NMR data (Figure 3.15A).

The followings are the structural information based on the data: (1) twelve imino proton peaks in 1D NMR show twelve guanine residues participating in the tetrads

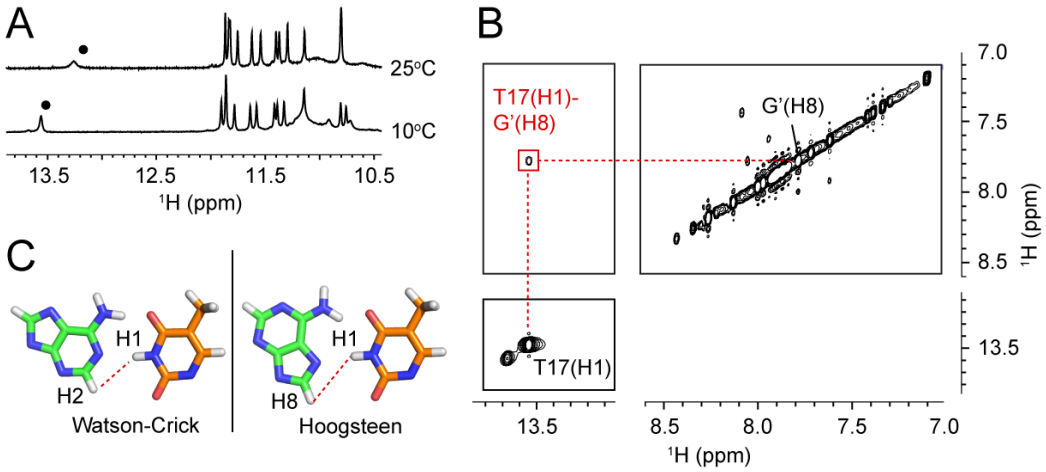


Figure 3.13: Additional hydrogen bond formed in *D4*-d(AG) complex in low temperature. A) 1D NMR spectra of *D4*-d(AG) complex in room (25°C) and low (10°C) temperature, black dots indicate the peak of T17 imino proton which is sharper in lower temperature. B) 2D NOESY of *D4*-d(AG) complex, the indicated cross-peak suggests that T17 imino proton is close to the aromatic proton of G' (guanine base from d(AG)). C) Comparison between Watson-Crick and Hoogsteen base pairing configurations, the dashed lines indicate the low spatial distance between T17(H1) to A'(H2) and A'(H8) respectively.

formation; (2) three H1-H8 NOESY cross-peak cyclic connectivity patterns indicate three G-tetrad formation, all in the same orientation; (3) strong H8-H1' cross-peak intensity of G' (guanine residue from the linear dinucleotide) stipulates *syn* conformation for G', while medium H8-H1' cross-peak intensities for all other guanines indicate *anti* conformations (Figure 3.15A); (4) low-temperature NOESY cross-peak between T17(H3)-A'(H8) signifies Hoogsteen base pairing between T17 and adenine residue of the ligand (Figure 3.13B, C).

Structure calculations *in vacuo* was done using distance, dihedral, hydrogen bond and planarity constraints. The 10 lowest energy structures and best ribbon representation are presented (Figure 3.16A). Table 3.2 shows structure statistics of *D4*-d(AG) complex.

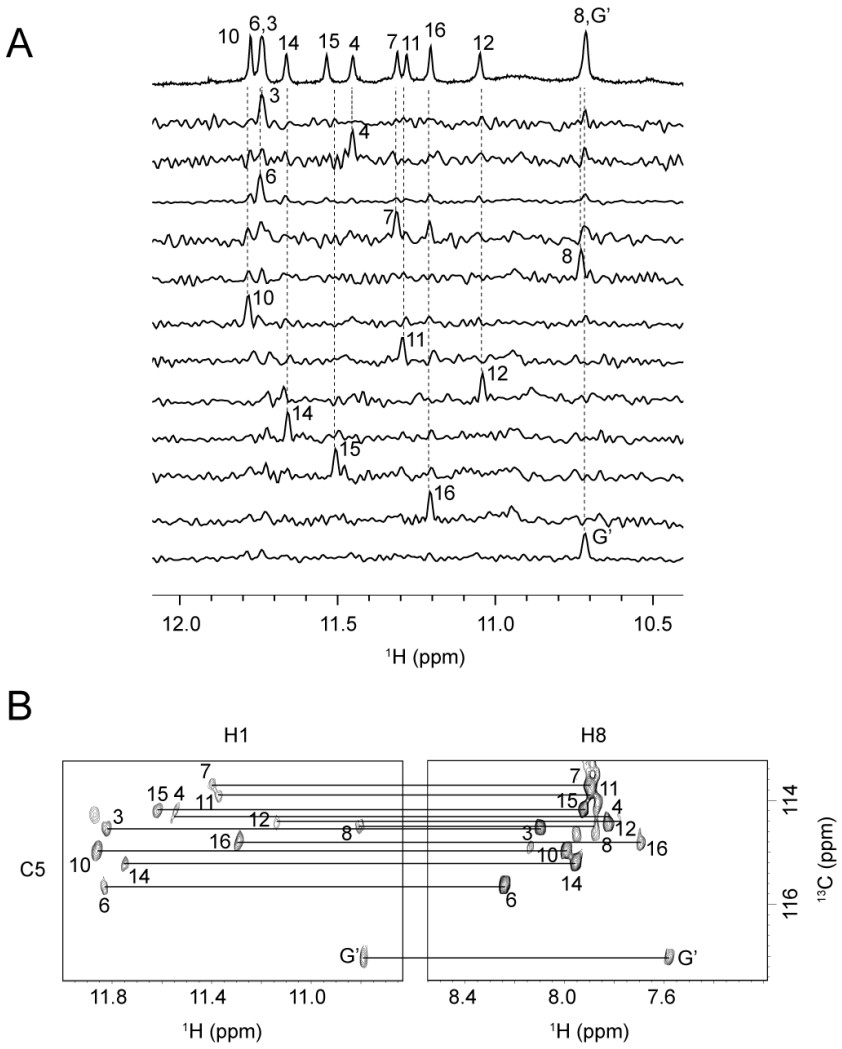


Figure 3.14: Spectral assignments of *D4*-d(AG) complex. A) Assignment of imino protons of *D4*-d(AG) complex using 2% ^{15}N site specific labeling technique. B) Assignment of the corresponding aromatic protons using long-range coupling HMBC method.

b. *D4*-cGAMP complex

For this complex, the ratio of 1:1.5 *D4*:cGAMP is required in 1 mM of *D4* strand concentration to generate clean spectra. As opposed to the linear counterpart, cGAMP addition did not result in additional appearance of an imino peak at ∼13-14 ppm.

The assignment of imino protons in this complex was inferred ambiguously from that of *D4*:d(AG) complex, as well as cross-reference work with the NOESY spectra

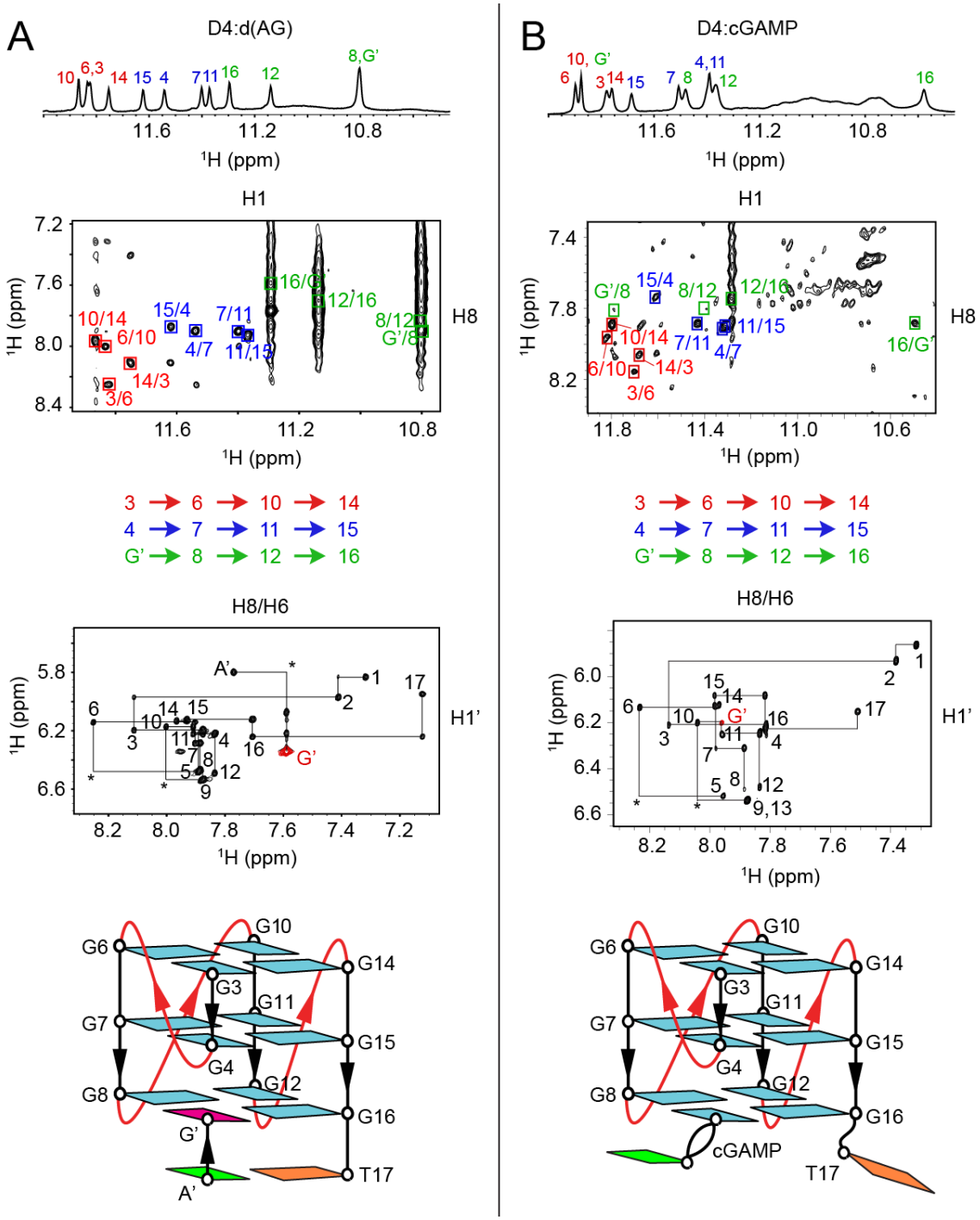


Figure 3.15: Summary of *D4*-d(AG) and *D4*-cGAMP complexes structures. Assigned 1D NMR spectra, 2D NOESY cyclic-connectivity patterns, G-tetrad orientations, H8/H6-H1' sequential walk, as well as structure schematics of A) *D4*-d(AG) complex and B) *D4*-cGAMP complex. Cyan, magenta, orange, and green represent *anti*-guanine, *syn*-guanine, thymine, and adenine residues respectively.

of both complexes. Similar set of NMR experiments, excluding the ^{1}H-^{15}N HMQC for imino proton assignment; were done for assignment and distance constraints generation purposes using an unlabeled sample. The notable differences are as follows: (1) ^{1}H-^{1}H NOESY shows no cross peak from both the aromatic protons (H2 and H8) of adenine residues, indicating no interaction; (2) all guanine intra-residue H8-H1' peaks, including that of cGAMP fell under medium intensity category indicating *anti* conformations (Figure 3.15B). Structure statistics are shown in Table 3.3.

Table 3.2: Statistics of the computed structure of *D4*-d(AG) complex.

A. NMR restraints		
Distance restraints	Exchangable	Non-exchangable
Intra-residue	0	339
Inter-residue	57	91
Other restraints		
Hydrogen bonds	52	
Dihedral angles	13	
B. Structure statistics		
NOE violations		
Number ($> 0.2\mathring{A}$)	0.7 ± 0.7	
Deviations from the ideal covalent geometry		
Bond lengths ($\mathring{A}$)	0.003 ± 0.000	
Bond angles (^{o})	0.671 ± 0.006	
Impropers (^{o})	0.357 ± 0.002	
Pairwise heavy atom rmsd value ($\mathring{A}$)		
All heavy atom	0.318 ± 0.055	
G-tetrad core	0.282 ± 0.080	

Table 3.3: Statistics of the computed structure of *D4*-cGAMP complex.

A. NMR restraints		
Distance restraints	Exchangable	Non-exchangable
Intra-residue	0	364
Inter-residue	39	79
Other restraints		
Hydrogen bonds	48	
Dihedral angles	12	
B. Structure statistics		
NOE violations		
Number ($> 0.2\mathring{A}$)	0.1 ± 0.3	
Deviations from the ideal covalent geometry		
Bond lengths ($\mathring{A}$)	0.003 ± 0.000	
Bond angles (^{o})	0.677 ± 0.011	
Impropers (^{o})	0.335 ± 0.006	
Pairwise heavy atom rmsd value ($\mathring{A}$)		
All heavy atom	0.529 ± 0.124	
G-tetrad core	0.449 ± 0.133	

Inferring from the two structures above, the guanine base is attached relatively tightly inside the previously vacant site in the G-quadruplex. In addition, the *D4*-cGAMP structure shows no interaction between the tail end of the cGAMP molecule and the *D4* molecule, suggesting that the binding of the guanine base alone is enough to stabilize the complex. Subsequently, it is possible to design various other chemicals to be attached to guanine base as a part of drug design.

In summary, we showed the existence of stable $4n$-1 G-quadruplex structures and their potential as a metabolite binder. In addition of increasing the number of

putative G-quadruplex forming sequences, it also provides the possibility of G-quadruplex-based structure to play regulatory roles inside cells. Two solution structures of *D4*-d(AG) and *D4*-cGAMP complexes are presented, showing the molecular detail of the binding site, potentially applicable to drug design.

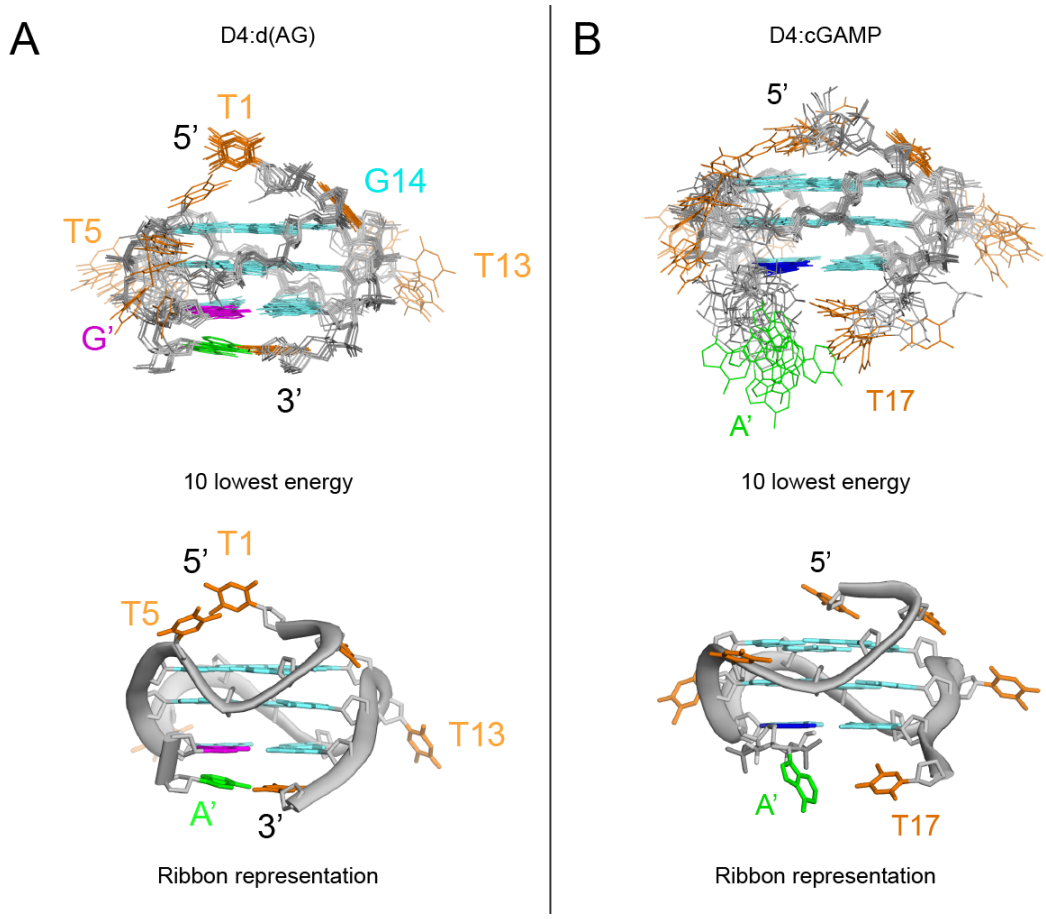


Figure 3.16: 10 superimpose lowest energy structures and the best ribbon representations of A) *D4*-d(AG) and B) *D4*-cGAMP complexes. The divergence of A' in cGAMP indicates that there is no interaction observed between the adenine base and the *D4* oligonucleotide. Cyan, magenta, orange, and green represent *anti*-guanine, *syn*-guanine, thymine, and adenine residues respectively.

3.3 Fold-back anti-parallel G-quadruplex structure with triad formation

This work is prepared for publication under the title: "Formation of an unprecedented G-quadruplex knot". Part of the section was presented earlier in the thesis of **Dr. Truong Thi Hong Anh** (Chapter 3, 2017).

Therapeutics studies of G-quadruplex has led to the development of many G4-based aptamers [149, 191–195]. The applications of these aptamers were stipulated to include anti-cancer (*AS1411, T40214*) [91,196], anti-virus (*93del, Zintevir*) [197] and anti-coagulation [198]. In this section, we would see further into the oligonucleotide aptamer *AS1411*, its behavior and derivatives.

3.3.1 G-quadruplex aptamer *AGRO100*

AS1411, or otherwise known as *AGRO100*, is the first oligonucleotide-based aptamer that reached Phase II clinical trials for treatment of acute myeloid leukemia and renal cell carcinoma [196]. Introduction of *AGRO100* and its variants to both normal and cancer cells showed selective anti-proliferation effect on cancer cells [199]. The proposed mechanism is binding of aptamer towards cytoplasmic nucleolin, which induces some signaling pathway deviation [200]. Nevertheless, the actual mechanism is still speculative.

Whilst the effect of the secondary structure towards the affinity and specificity of the aptamer towards cancer cell is uncertain, the fact that the aptamer contains G-rich oligonucleotide sequence piqued our interest. Therefore, we tried to explore the structural details of *AGRO100* in vitro using NMR spectroscopy. 1D NMR spectroscopy of *AGRO100* 26-mer sequence showed a highly overlapped spectrum,

which indicates its high polymorphism (Figure 3.17A). This result agreed with previous study using circular dichroism, where multiple different signals were observed from different fractions of size-exclusion chromatography [92].

Knowing this information, we tried sequence modifications of *AGRO100*, including a reduction of guanine residues from 17 to 16, as well as addition of a terminal thymine residue to increase stability. Previous studies explored two unique G-quadruplex structures from *AGRO100* variants: *AT27* [12] and *AT11* [201] (Figure 3.17A), whereby G26 and G10 of the original *AGRO100* sequence were substituted to thymine respectively, and one additional thymine at 5'-end were added for both sequences (hence the deviation in numbering). Both structures are constituted of two G-quadruplex blocks, where each block consists of two G-tetrad, and is connected via thymine linker/s to the other block. However, there is a fundamental difference between *AT27* and *AT11* stacking structures, with *AT11* adopting right-handed backbone progression, while *AT27* exhibiting a novel left-handed backbone progression (Figure 3.17B).

Motivated by striking differences between *AT27* and *AT11* structures, we continued to look for possible new G-quadruplex fold from the mutation of *AGRO100*. Herein, we present a unique structure coming from the substitution of guanine to thymine at position 20 of *AGRO100*, designated *AT21*. Oligonucleotide sequences mentioned in this subsection is listed in Table 3.4.

Table 3.4: DNA sequences of *AGRO100* and its variants used in this study.

Name	**5'-3' Oligonucleotide sequence**
AGRO100	GGTGGTGGTGGTTGTGGTGGTGGTGG
AT11	**T**GGTGGTGGT**T**GTTGTGGTGGTGGTGG**T**
AT27	**T**GGTGGTGGTGGTTGTGGTGGTGGTG**TT**
AT21	**T**GGTGGTGGTGGTTGTGGTG**T**TGGTGG**T**

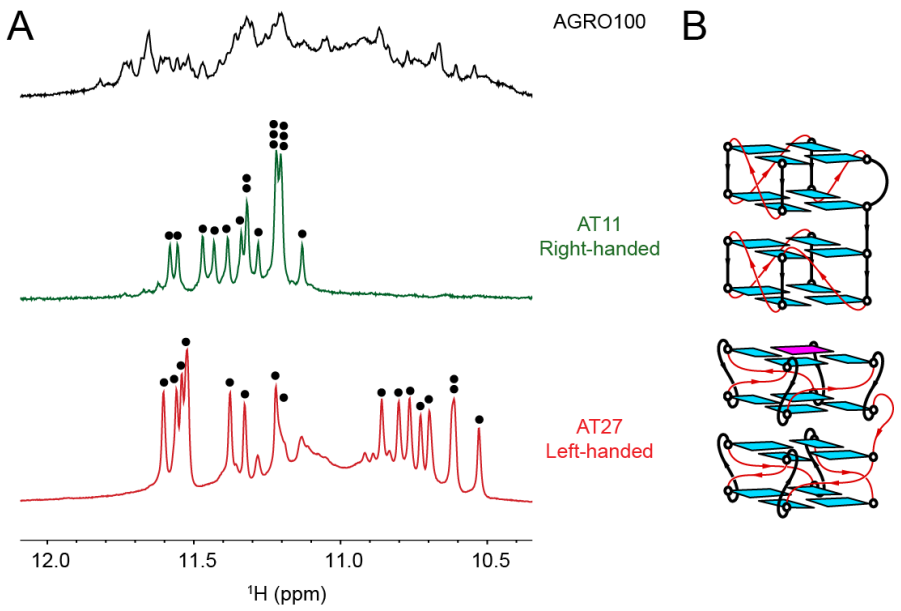


Figure 3.17: NMR analysis of *AGRO100* and its variants *AT11* and *AT27*. A) 1D NMR spectra of *AGRO100*, and its modification *AT11* [201] and *AT27* [12]. Black dots in the spectra indicate the imino proton peaks. B) Schematic of *AT11* and *AT27* showing right-handed and left-handed G-quadruplex respectively.

3.3.2 Structural interpretation of *AGRO100* derivative *AT21* and *AT21S*

Previous work by former colleague Dr. Truong Thi Hong Anh has elaborated the spectroscopic study of *AT21* (Figure 3.18), which shows an anti-parallel CD signature and unusual highly upshifted series of imino proton peaks around 10 ppm. Due to the dynamics of long loops, this sequence was not very suitable for structural determination, which led to sequence mutation from *AT21*, termed *AT21S*. The summary of structural interpretation of *AT21S* is shown (Figure 3.19), indicating formation of two G-tetrads as the core and additional stabilizing structure on the surface of the tetrad. Proton assignments and fold determination were performed by Dr. Truong.

The 9.2-10.4 ppm area of H_2O NOESY spectrum provide the information needed

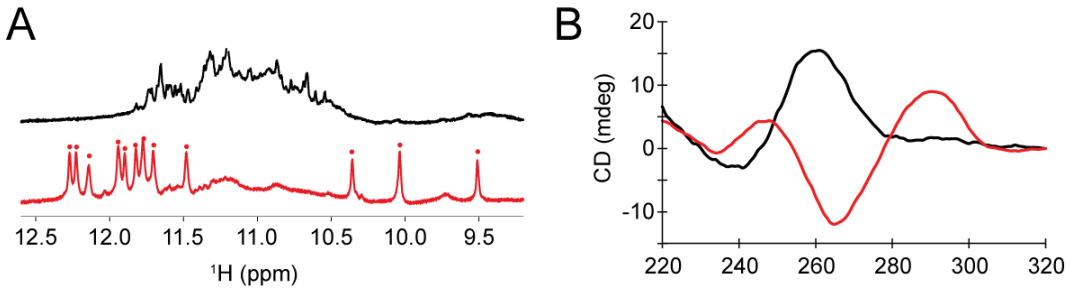


Figure 3.18: Spectroscopy of *AGRO100* versus *AT21*. A) 1D NMR spectra of *AGRO100* (black) and its modification *AT21* (red), red dots indicate the resolved imino proton peaks. B) CD spectra of *AGRO100* (black) and *AT21* (red).

for the determination of the stabilizing structure/s. The set of cross peaks between T15, T17, and G20 suggest the formation of G-T-T triad on top of the G-tetrad, and the set of cross peaks between G16 and T21 indicate the formation of a G-T base pair (Figure 3.20). The overall schematic of folding topology is presented (Figure 3.19B).

In detail, the structure of *AT21S* features the following: (1) two G-tetrad layers in opposite polarity; (2) four *syn*-guanines namely G2, G8, G13 and G19, as well as four *anti*-guanines G3, G6, G9 and G14; (3) three continuous tracts (G2-G3, G8-G9, G13-G14) and one discontinuous tract (G6-G19); (4) two edgewise loops (T4-T5 and T10-T12) and one V-shaped loop (T7); (5) one diagonal loop forming part of triad and base pair structures (T15-T18).

Structure calculation of *AT21S* resulted in 10 lowest energy structures and a best ribbon representation (Figure 3.21), with the structure statistics presented in Table 3.5.

3.3.3 Triad formation in G-quadruplex diagonal loops

Triad-containing G4 loops were previously studied [202–204], including specifically a diagonal fold-back loop [81]. The term 'fold-back' indicate the existence of one

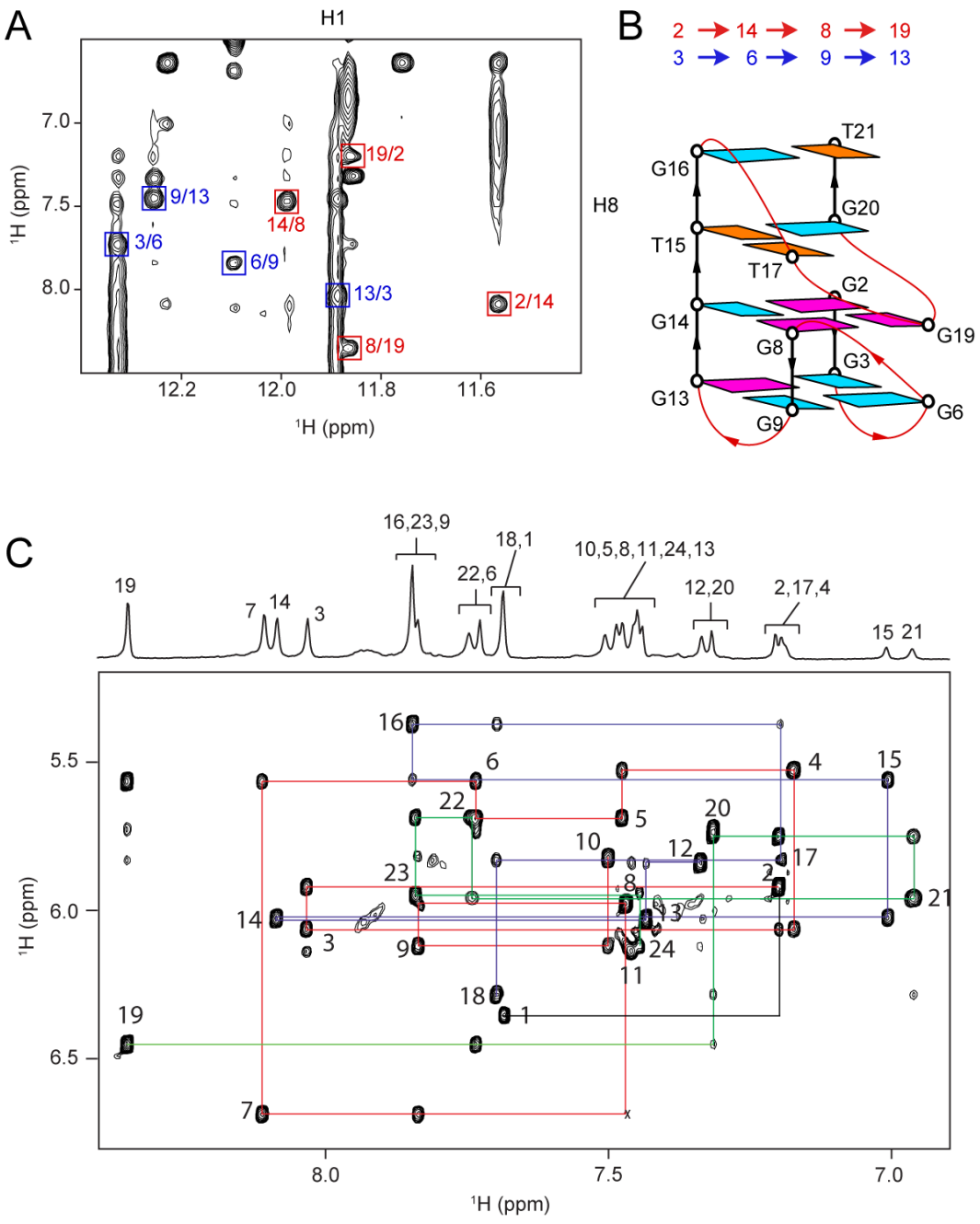


Figure 3.19: Summary of *AT21S* folding topology determination. A) NOESY spectrum of *AT21S* showing cyclic connectivity patterns of two G-tetrads. B) G-tetrad orientations and proposed schematic of *AT21S*. C) NOESY spectrum showing H8/H6-H1' sequential walk, different colors indicate broken connectivity.

guanine residue after the diagonal loop that completes the G-quadruplex structure, which we address as the "completing" guanine (Figure 3.22). In the previously

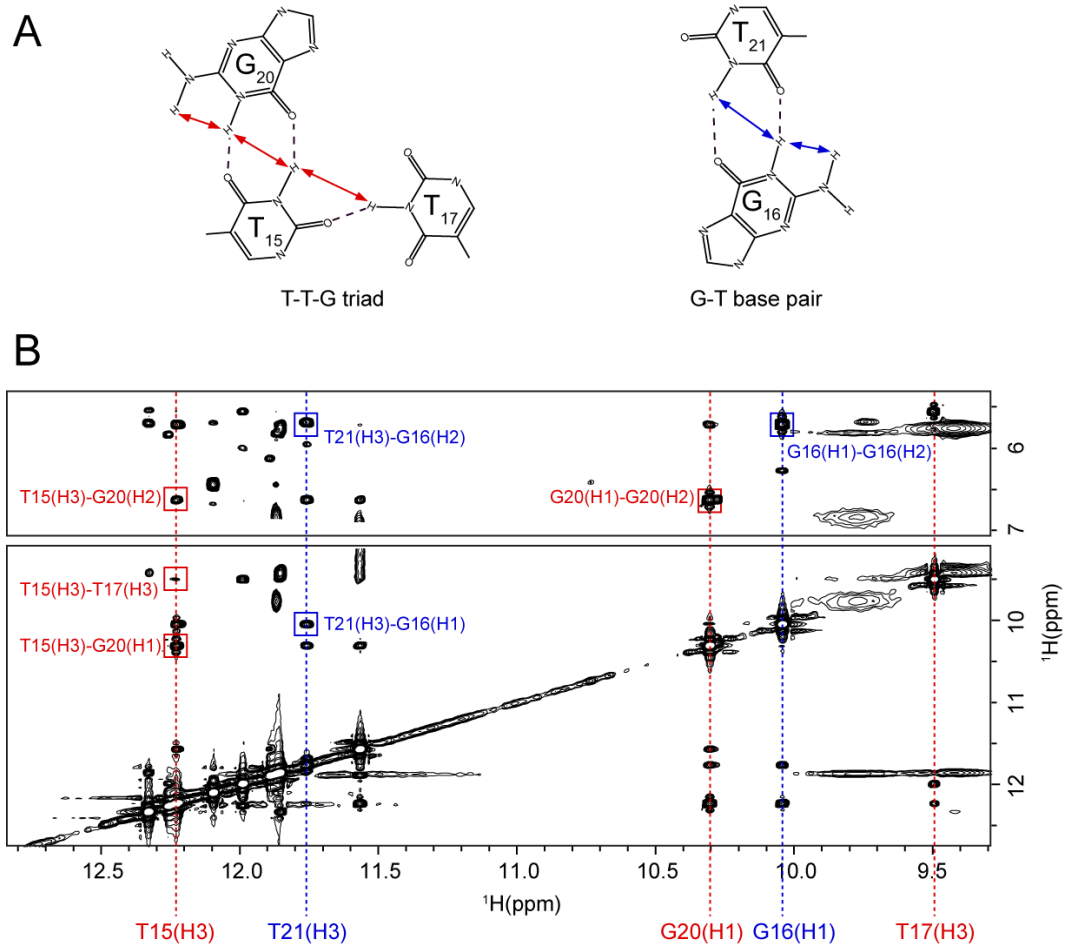


Figure 3.20: Summary of additional substructures formed in *AT21S*. A) T-T-G triad and G-T base pair schematics, red and blue arrows indicate protons that would affect each other in the NMR mixing time. B) ^{1}H-^{1}H NOESY spectrum of *AT21S* in the imino region, red and blue boxes indicate the relevant cross-peaks that constructed the structures of stabilizing triad and base pair respectively.

studied sequence *Pu24T*, a fold-back diagonal loop (**G**A**AG**, responsible residues in bold) was observed to form a G-A-G triad structure [81]. In AT21S sequence, the fold-back diagonal loop (**T**G**T**T) forms part of a T-T-G triad, completed with one G residue following the "completing" guanine.

The similarity between these two fold-back diagonal loops from *Pu24T* and *AT21S* could be a basis for further study of stabilizing triad in G-quadruplexes. Basic cut-and-paste study was performed to see if the two fold-back diagonal loops could be

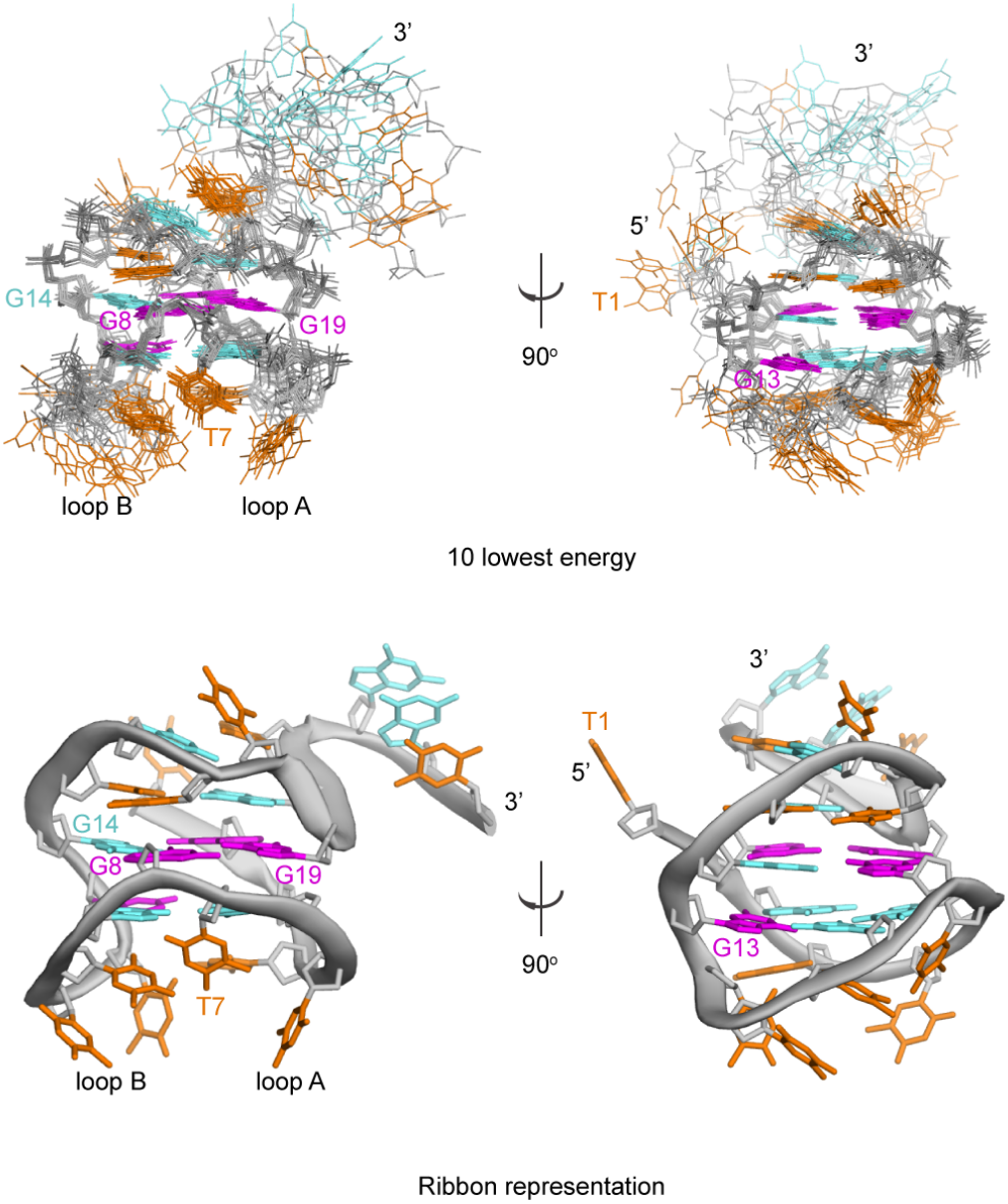


Figure 3.21: 10 superimposed lowest energy structures and the best ribbon representation of *AT21S*. Cyan, magenta, and orange represent *anti*-guanine, *syn*-guanine and thymine residues respectively.

interchangeable (Figure 3.23). The results show dynamic movements of the loops indicated by sharpening of peaks in lower temperature. The possible explanation of the unstable nature of the interchanged structures (*Pu24T* base with *AT21S* loop and vice versa) is the difference in the grooves lengths of the two core structures. *Pu24T* has four medium grooves due to its parallel conformation, while *AT21S* has

Table 3.5: Statistics of the computed structure of *AT21S*.

A. NMR restraints		
Distance restraints	Exchangable	Non-exchangable
Intra-residue	4	275
Inter-residue	82	115
Other restraints		
Hydrogen bonds	42	
Dihedral angles	10	
B. Structure statistics		
NOE violations		
Number ($> 0.2\mathring{A}$)	0.3 ± 0.5	
Deviations from the ideal covalent geometry		
Bond lengths ($\mathring{A}$)	0.003 ± 0.000	
Bond angles (^{o})	0.695 ± 0.006	
Impropers (^{o})	0.366 ± 0.003	
Pairwise heavy atom rmsd value ($\mathring{A}$)		
All heavy atom	0.726 ± 0.162	
G-tetrad core	0.483 ± 0.123	

two wide and two narrow grooves characteristic of anti-parallel conformation. In addition, as a side result, we observed that despite the broadened peaks, the two G-tetrad of *AT21S* are still formed when we pasted the diagonal loop of *Pu24T* (GAAG) in place of the original sequence (TGTT_GT) (Figure 3.23A, B). Importantly, modifying the diagonal loop into poly T destroyed the structure, indicating the necessity of the triad structure.

In summary, we showed a unique anti-parallel structure of *AGRO100* derivative *AT21S*. The structure contains two G-tetrad as the core, with stabilizing triad

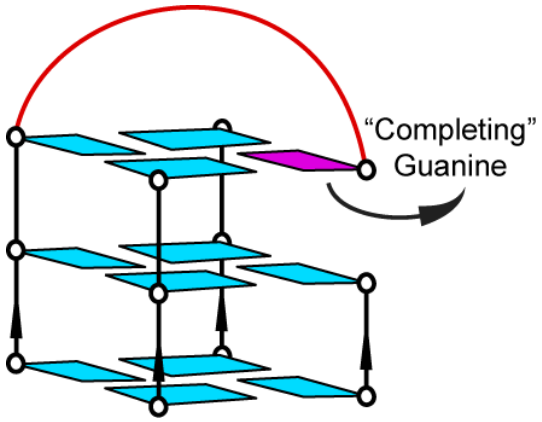


Figure 3.22: Schematic of G-quadruplex with fold-back loop (red curve). "Completing guanine" is indicated in magenta.

and base pair as a part of the fold-back diagonal loop. Together with previously solved fold-back structure (*Pu24T*), the solved structure of *AT21S* provide a good foundation for design of stabilizing triad loops.

3.4 G-quadruplex structure of trinucleotide repeats *(GGT)₈*

This work is prepared for publication under the title: "Left-handed right-handed G-quadruplex structure of *(GGT)₈* trinucleotide repeats".

GGT (or TGG/GTG) trinucleotide repeats were previously studied to have possible importance *in vivo*, the repeats were found in mRNA sequence in human lymphocytes [205] and were evolutionary conserved in *Citrus* genome [206]. It also acts as catalyst in genomic rearrangement including microdeletion [207]. Research in conformational properties of DNA containing GGT repeats using UV and CD spectroscopy revealed the polymorphism of the sequence, ranging from octameric complex into hairpin formation in different conditions [208]. Herein, we discuss the G-quadruplex structure formed by a sequence of eight GGT repeats.

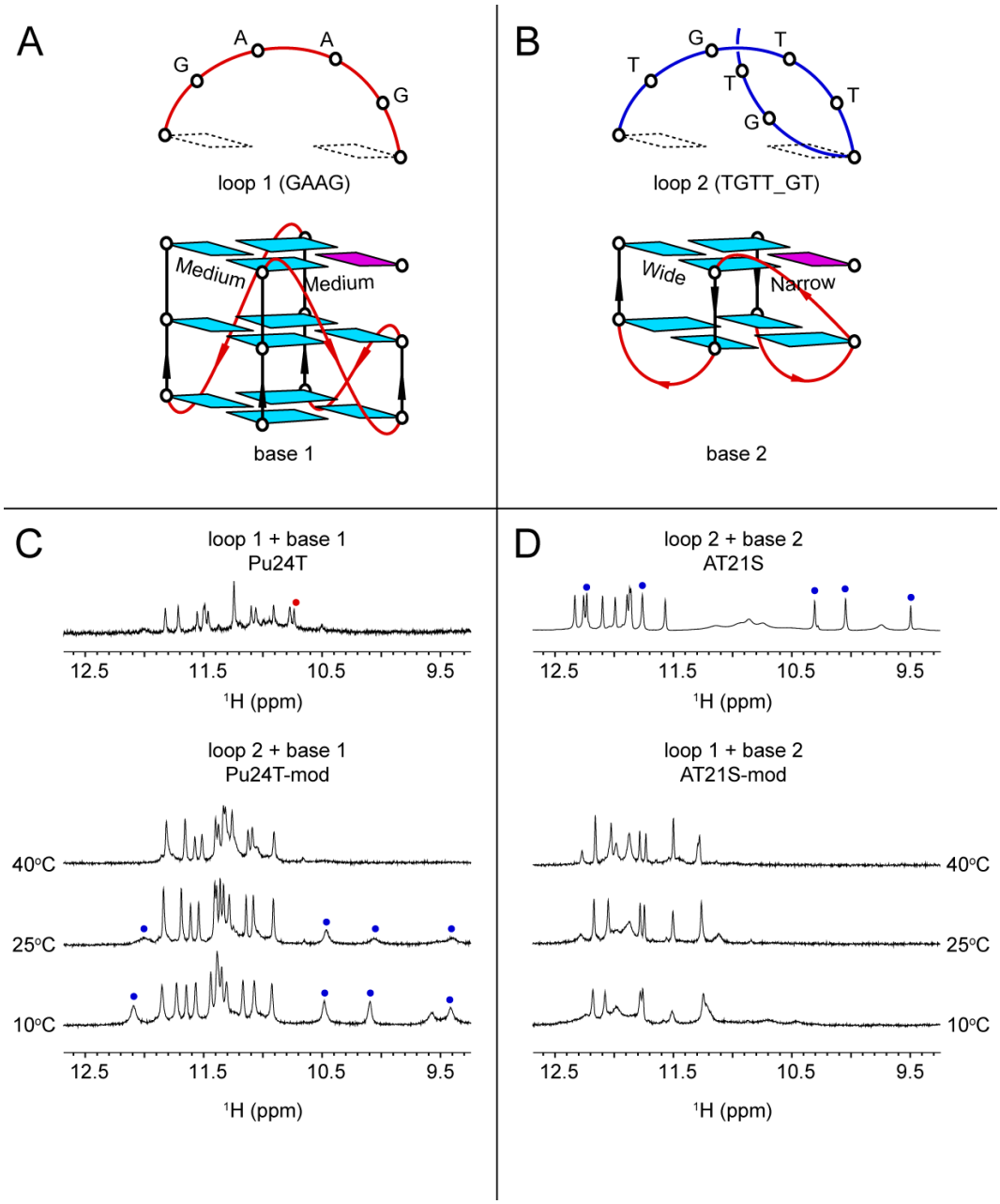


Figure 3.23: Cut-and-paste study of *Pu24T* and *AT21S*. Diagonal loops and bases of A) *Pu24T* and B) *AT21S*. Dashed rectangles represent the start and end positions of the diagonal loops. "Completing guanines" are shown in magenta. C, D) Resulting NMR spectra of cut-and-paste study between the loops and bases. Red and blue dots marked peaks from loop 1 (*Pu24T*) and loop 2 (*AT21S*) respectively.

3.4.1 Folding topology of G-quadruplex *(GGT)$_8$*

1D NMR spectrum of eight GGT trinucleotide repeat sequence *(GGT)$_8$*, d(5'-GGTGGTGGTGG T GGTGGTGGTGG T-3') showed 16 characteristic imino proton peaks for Hoogsteen hydrogen bonds, indicating a possible 4-layer G-quadruplex structure.

a. CD and 1D NMR spectroscopy

CD and 1D NMR spectroscopy experiments of *(GGT)$_8$* were performed as preliminary measurements. The resulting NMR spectrum showed a significant hump in the imino proton area suggesting multiple conformations (Figure 3.24C). Hence, native gel purification experiment was done to selectively acquire the monomeric species (Figure 3.24A). The left most lane in Figure 3.24A indicates a standard 24-mer 3-layer G-quadruplex sample *Pu24T* showing both monomeric and dimeric bands. The next 4 lanes were loaded with *(GGT)$_8$* sample showing some higher order structures (smeared bands) and the monomeric structure/s (thick bands at the bottom). The monomeric band of *(GGT)$_8$* was cut, incubated, desalted, and re-dissolved in water.

Gel-purified *(GGT)$_8$* sample was re-measured with both CD and NMR spectroscopy, generating a clean 1D NMR spectrum and an extremely different CD spectrum (Figure 3.24B). The gel purification assured monomeric species was chosen, however since the sequence is relatively long (24-mer), other minor conformations [208] or structures from shorter unsuccessful sequences could also present. Thus, the CD spectrum would not be reliable in determining the fold of the major G-quadruplex structure formed by the sequence. On the other hand, the NMR spectrum of gel-purified *(GGT)$_8$* showed 16 peaks indicating possible formation of a 4-layer monomeric G-quadruplex structure (Figure 3.24C).

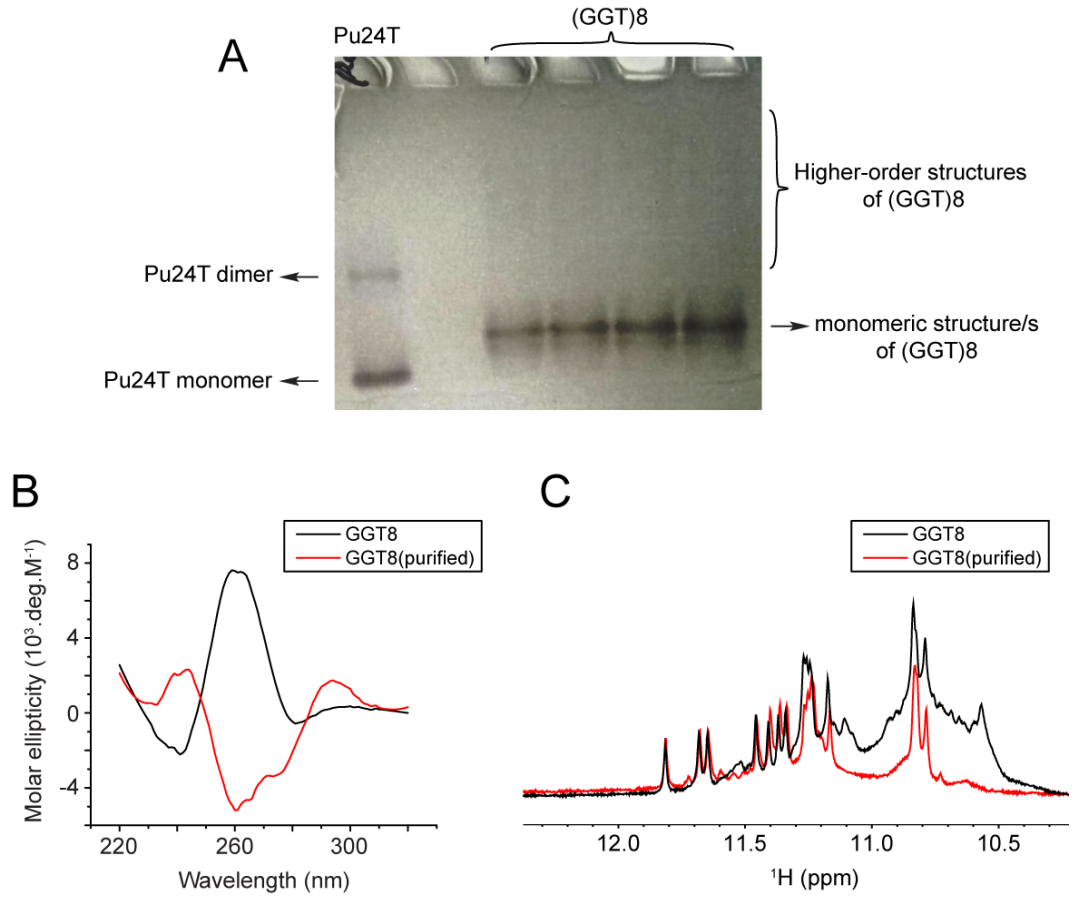


Figure 3.24: Gel purification of *(GGT)₈* and the corresponding CD and NMR spectra. A) Native gel snapshot showing DNA migration. The left-most lane was loaded with standard 3-layer G-quadruplex *Pu24T*, and the 4 lanes to the right were loaded with *(GGT)₈* sample. B) CD and C) NMR spectra of untreated *(GGT)₈* (black) superimposed with gel-purified *(GGT)₈* (red).

b. Imino and aromatic proton assignments

The assignments of imino protons were performed with 2% ^{15}N site-specific labeled samples [13]. ^{15}N-filtered 1D ^{1}H-^{15}N HMQC spectra for the labeled samples are compiled and presented (Figure 3.25A). The aromatic assignments were done by ^{1}H-^{13}C HMBC spectrum connecting the imino protons to their intra-residue aromatic protons via C5 [151] (Figure 3.25B).

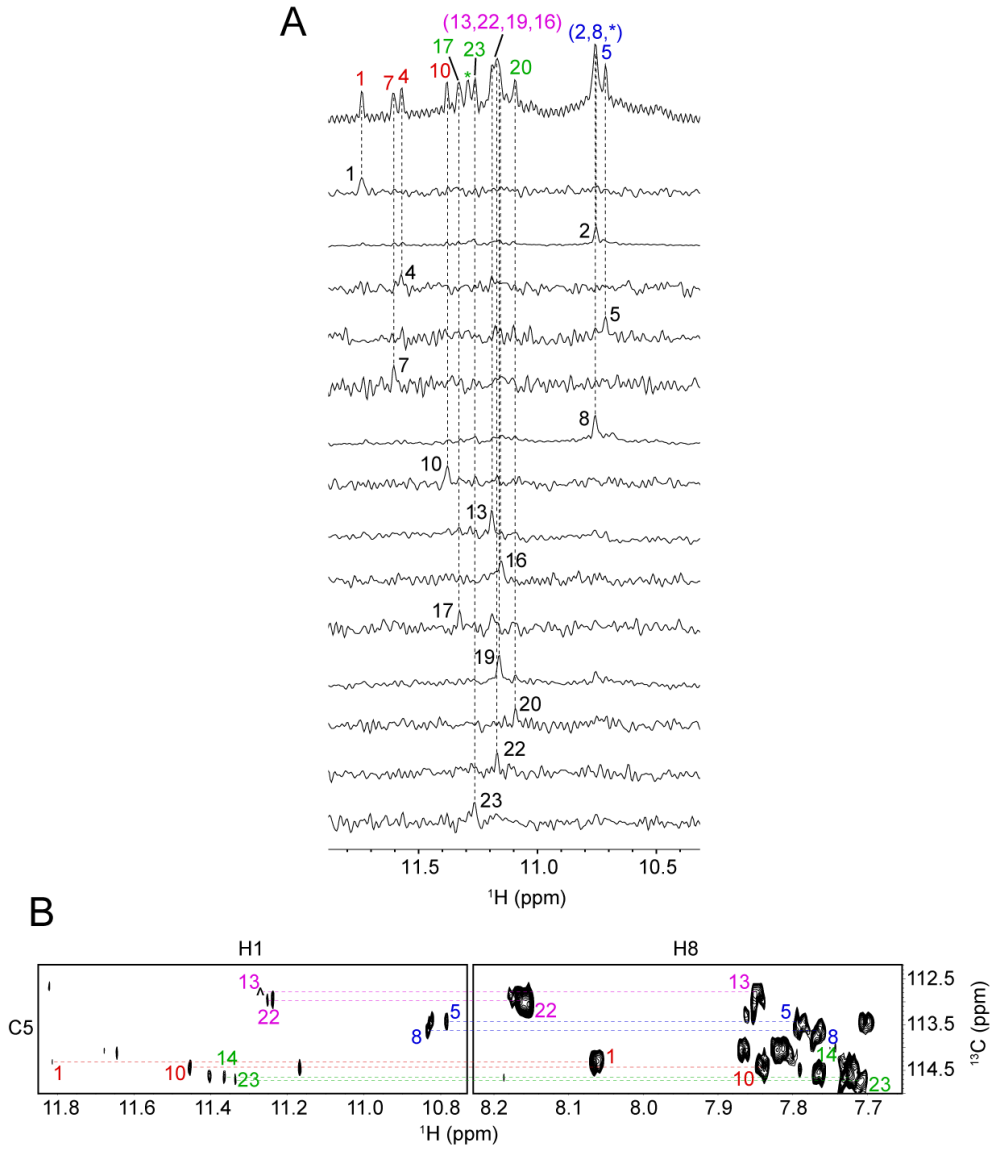


Figure 3.25: A) Imino and B) aromatic protons assignments of *(GGT)₈*. Site-specific ^{15}N labeling method and long-range through-bond coupling method were used as described previously (section 3.1). Only 8 out of 16 pairing of H1-H8 are shown for clarity reason. Same colored peaks belong to the same G-tetrad.

c. D_2O exchange

The time dependent D_2O exchange 1D NMR spectra are presented below (Figure 3.26). The results indicate guanine residues G1, G4, G7 and G10, as well as G14, G17, G20 and G23 as outer tetrad guanines, and guanine residues G2, G5, G8 and G11, as well as G13, G16, G19 and G22 as inner tetrad guanines.

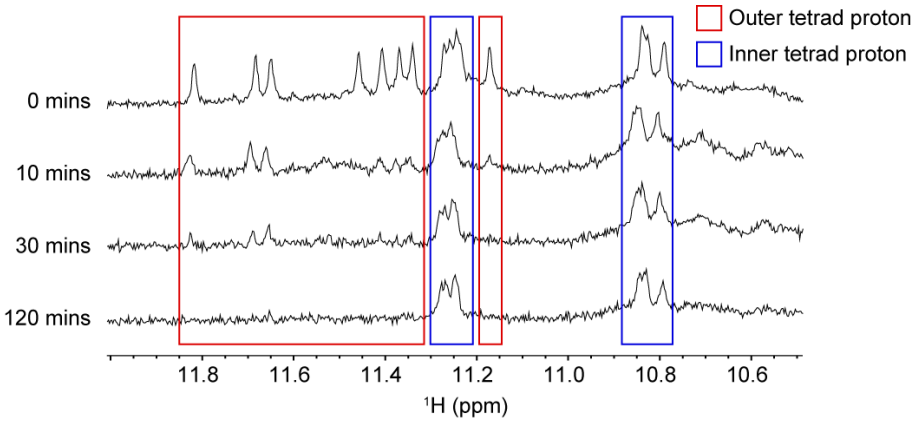


Figure 3.26: 1D proton NMR spectra of *(GGT)$_8$* in imino region, plotted in different times (0, 10, 30, and 120 mins) after D_2O addition. Peaks in red boxes were assigned to outer tetrad imino protons due to their fast disappearance following D_2O exchange. Peaks in blue boxes were assigned to inner tetrad imino protons due to their persistence following D_2O exchange.

d. Folding topology

Assignments of all the imino and aromatic protons of the guanine residues allowed the distinction between guanines from different G-tetrads. Using the cyclic-connectivity patterns described in the previous section (section 3.1), four G-tetrads were determined to be formed by the following residues, in order of H1-H8 cyclic connectivity: G1-G4-G7-G10; G2-G5-G8-G11; G13-G16-G19-G22; and G14-G17-G20-G23 (Figure 3.27B, C).

Out of all the NOESY cross-peaks, one particularly interesting region is the H1'-H1' region, whereby the cross peaks between G5(H1')-G19(H1'), G8(H1')-G16(H1'), as well as G11(H1')-G13(H1') were observed (Figure 3.28A). These cross peaks suggested that the Hoogsteen hydrogen bonds orientation of the 2nd and 3rd tetrad are opposites of each other, as shown in the folding schematic (Figure 3.27D).

Moreover, the relatively strong intensities of these H1'-H1' cross-peaks suggest that the stacking between the 2nd and 3rd tetrads (designated as middle-interface further) is a 5'-5' stacking (Observation 1), where the two H1' protons are close to

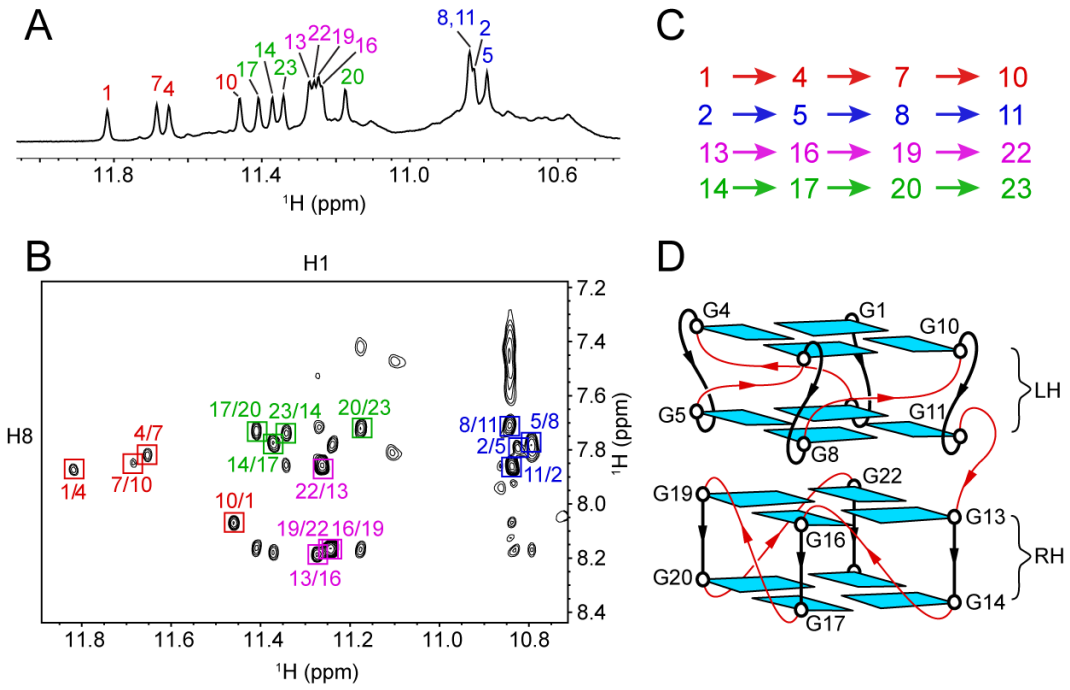


Figure 3.27: Folding topology of *(GGT)₈*. A) 1D proton NMR spectrum of *(GGT)₈* in imino region, showing assignments of all 16 guanines involved in G4 core. B) H_2O NOESY spectrum at 250 ms mixing time in H1-H8 region, showing four sets of cyclic-connectivity patterns as described previously (section 3.1). C) G-tetrad progression schemes. D) Proposed schematic of *(GGT)₈* folding showing left-handed block in the first-half and right-handed block in the second half. Red, blue, magenta, and green color codes are used for 1st, 2nd, 3rd, and 4th G-tetrad respectively.

each other (Figure 3.28B). However, data acquired from D_2O exchange experiment stated that the overall backbone direction is inward from 1st (G1, G4, G7 and G10) to 2nd (G2, G5, G8 and G11) tetrad, and it is outward from 3rd (G13, G16, G19 and G22) to 4th (G14, G17, G20 and G23) tetrad (Observation 2). The two observations led to a proposition that the inward G4 block (1st and 2nd tetrads) is in left-handed conformation, unconventionally exposing 5'-end towards the middle-interface, while the outward G4 blocks (3rd and 4th tetrads) is in right-handed conformation, conventionally exposing 5'-end towards the middle-interface.

To strengthen the proposition on the folding topology and determine the high-resolution structure, we executed an in-depth NMR pattern recognition study of

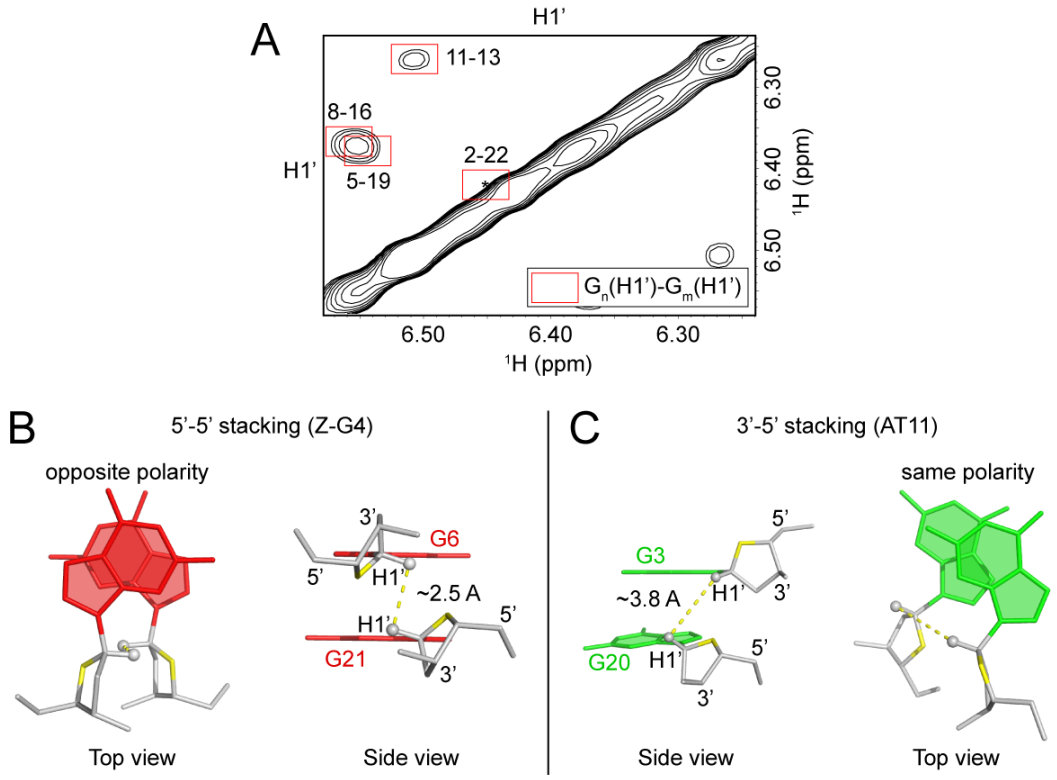


Figure 3.28: Inter-block stacking interface. A) NOESY spectrum of $(GGT)_8$ at 300 ms mixing time in H1'-H1' region. The cross-peaks shown in red boxes are strong H1'-H1' cross-peaks between the 2nd and 3rd G-tetrads, indicating opposite polarity between the two. Asterisk in the red box indicate possible cross-peak quenched by the diagonal. B) 5'-5' and C) 3'-5' stacking interface, showing opposite polarity and close-range proximity of H1's of the interacting guanines for 5'-5' interface, and same polarity with medium-range proximity of H1's between the stacking guanines for 3'-5' interface.

left-handed G4 (section 3.4.3). But first we introduce left-handed DNA and left-handed G-quadruplexes.

3.4.2 Left-handed G-quadruplexes

G-quadruplex structures and stacking have been known and studied for decades resulting in enormous numbers of G-quadruplex topologies. However, until recently [82, 156, 201, 209], all of discovered G-quadruplex structures were based on right-handed tetrad progression, characterized by clockwise rotation of the tetrads from

5'-end to 3'-end. In 2015, our group discovered the first left-handed G-quadruplex crystal structure [12] with opposite tetrad progression orientation. The structure is termed *Z-G4* since it can be considered as the analogue of Z-DNA in G-quadruplex field. Specifically, it comprises of two blocks of 2-layer G-quadruplex stacking with each other while connected by thymine linker. The work also elucidated the differences of stacking interface between two guanines in right-handed and left-handed tetrad progression (Figure 3.29).

In detail, the structure of *Z-G4* features two blocks of two-layered parallel G-quadruplexes, each oriented from outer tetrad to inner tetrad. Normally, this configuration would result in the exposal of 5'-faces of the sugar groups on both surfaces toward the solvent. However, the relative orientation of the backbone and the sugar ring for every guanine residue is the opposite of that observed in right-handed G-quadruplexes (further described in section 3.4.3). This structural feature caused the exposal of 3'-faces of the sugar groups, as well as 5'-5' inter-block stacking interface between the two inner tetrads of the G-quadruplex. Other structural features will be described in the next subsection.

Circular dichroism (CD) spectroscopy spectrum of left-handed G-quadruplex is almost the opposite the right-handed parallel G-quadruplex counterpart. The CD signal of left-handed G-quadruplex features a shallow peak at ∼245 nm and a sharp trough at ∼270 nm, while that of right-handed G-quadruplex shows a shallow trough at ∼240 nm and a sharp peak at ∼260 nm (Figure 3.30). This result somewhat agrees with the previously studied Z-DNA and B-DNA, where opposite CD signals were observed between the two [210].

The minimal oligonucleotide sequence motif that is required for the formation of a left-handed G-quadruplex (Bakalar, B. *et. al.*, unpublished (1)) as well as the effect of different cations in driving the formation of left-handed G-quadruplex structure

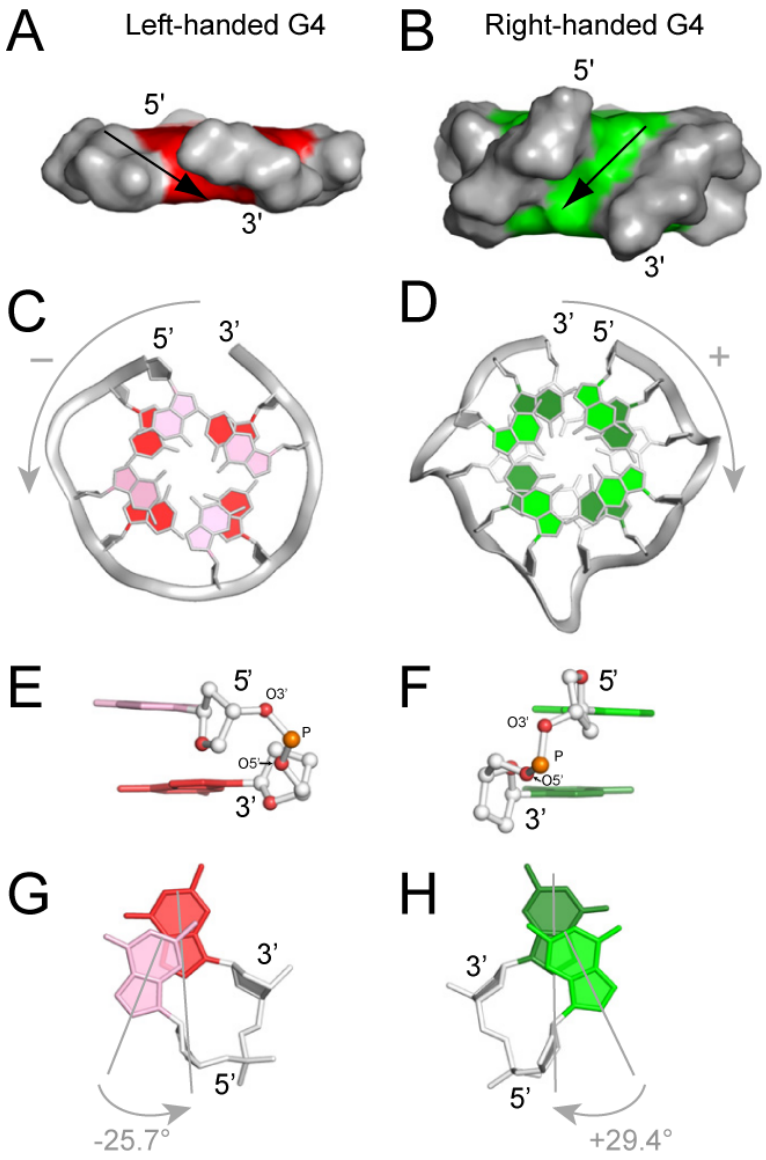


Figure 3.29: Summary of geometrical differences between left-handed (red) and right-handed (green) G4. A) Left and B) right-handed side view of surface representations. C) Left and D) right-handed top view of ribbon representations showing counter-clockwise and clockwise rotation orientation from 5' to 3' respectively. E) Left and F) right-handed detailed view on the sugar phosphate backbone, showing the relative orientations of the sugar rings to the base stacking. G) Left and H) right-handed base stacking rotations, showing different direction and magnitude of rotation with left-handed stacking slightly less shifted compared with right-handed arrangement. Figure adapted from [12].

(Bakalar, B. *et. al.*, unpublished (2)) have been actively studied in our group.

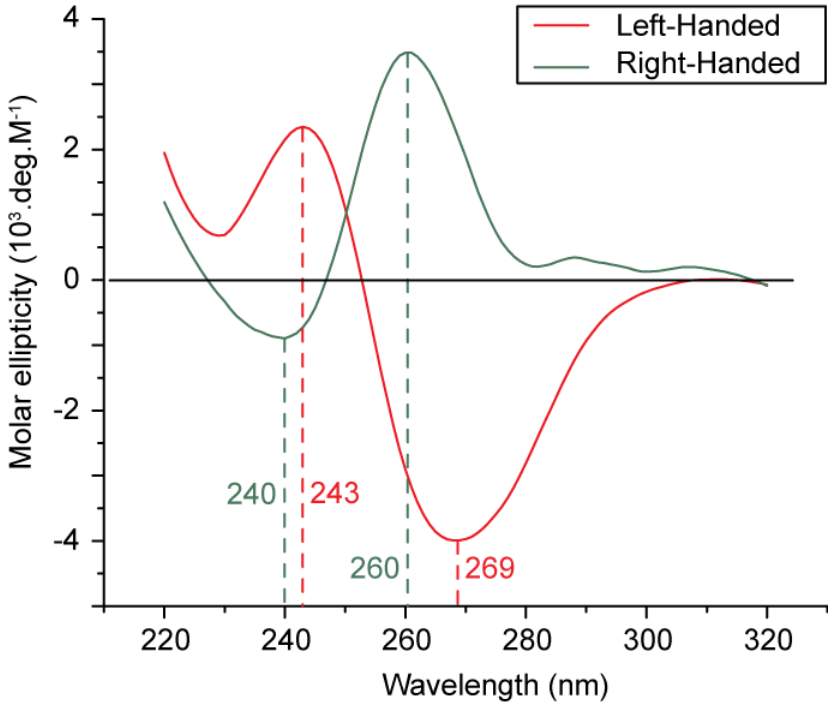


Figure 3.30: Typical circular dichroism (CD) signal of left-handed (red) and right-handed (green) G4. A relatively small peak at 243 nm and a relatively large trough at 269 nm was observed for left-handed fold, while a small trough at 240 nm and a large peak at 260 nm was observed for right-handed fold. The DNA sequences used to generate the CD signal were second-half of *Z-G4* (14 bases) and first-half of *Z-G4* (12 bases) for left and right-handed fold respectively.

3.4.3 NMR observables of left-handed G-quadruplexes

In this subsection, three unique structural features of left-handed G-quadruplex blocks are described based on the solved crystal structure of *Z-G4*. These unique features generate equally unique NMR observables that can effectively be used to probe the structure occurences without fully solving the high-resolution structure.

a. Thymine capping

Z-G4 (d(5'-T GGTGGTGGTGG TT GTGGTGGTGGTG TT-3')) sequence consists of multiple thymine residues, where except the two thymine residues in the middle (T13 and T14) as well as the two at the end (T27 and T28), all the other

thymine residues are propeller loops. Usually, a single-base propeller loop will have its nitrogenous base facing outward, away from the G-tetrad core. Nevertheless, in the case of *Z-G4*, the thymine propeller loops are directed inwards, projecting on top or bottom of the G-tetrad core. These thymine residues are forming, albeit not fully, additional stacks of bases (Figure 3.31).

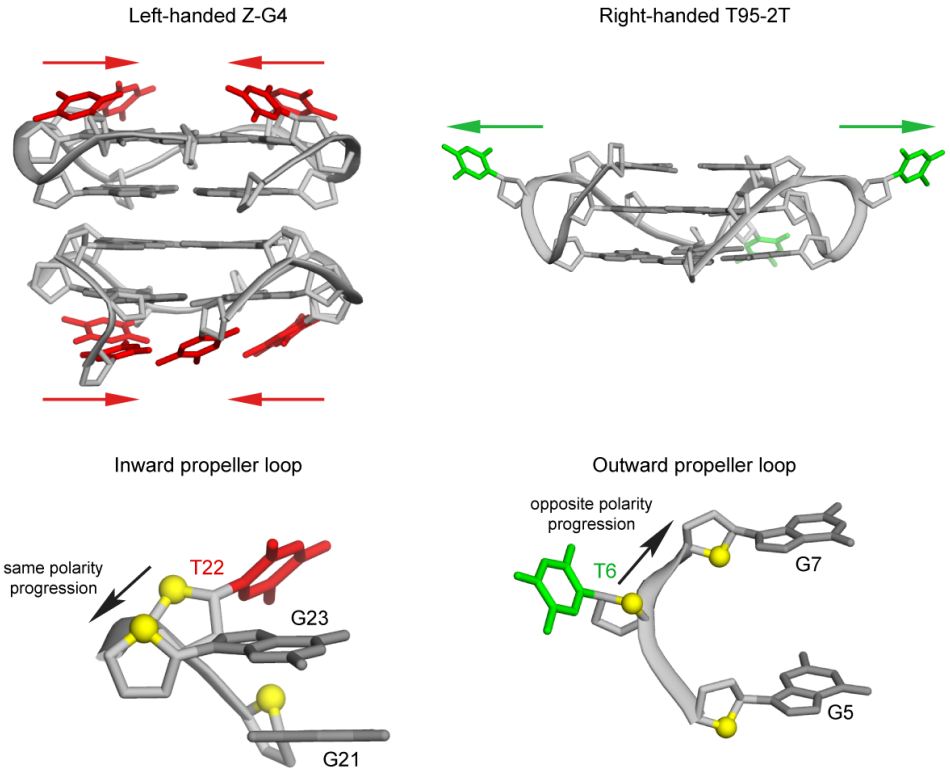


Figure 3.31: Thymine (T) loops relative orientation of left and right-handed G4 structures. (Top) Inward and outward orientation of the T-loops relative to the G4 core were shown in the left-handed and right-handed G4 structures respectively. (Bottom) Detailed view on the orientation of the thymine backbone relative to the stacking guanine bases. Left-handed T-loop has the same polarity progression as can be seen from the O4' atoms (indicated in yellow) of the sugar rings all pointing to the relatively similar direction. On the contrary, right-handed T-loop has the opposite polarity progression. Left and right-handed G4 structures used are *ZG-4* and *T95-2T* respectively.

NMR observables of this feature are detected in the aromatic (H8/H6)-anomeric (H1') region of NOESY experiment. The thymine residues that are directed inward

and stacked onto guanine residues in the G-tetrad core produced a characteristic H1'(n)-H8(n+1) cross-peaks (Figure 3.32A), which is not observed in any right-handed thymine propeller loops. In addition, cross-peaks between methyl groups of the thymines to the guanines were also identified (Figure 3.32B), further stating the inward orientation of the thymine loops.

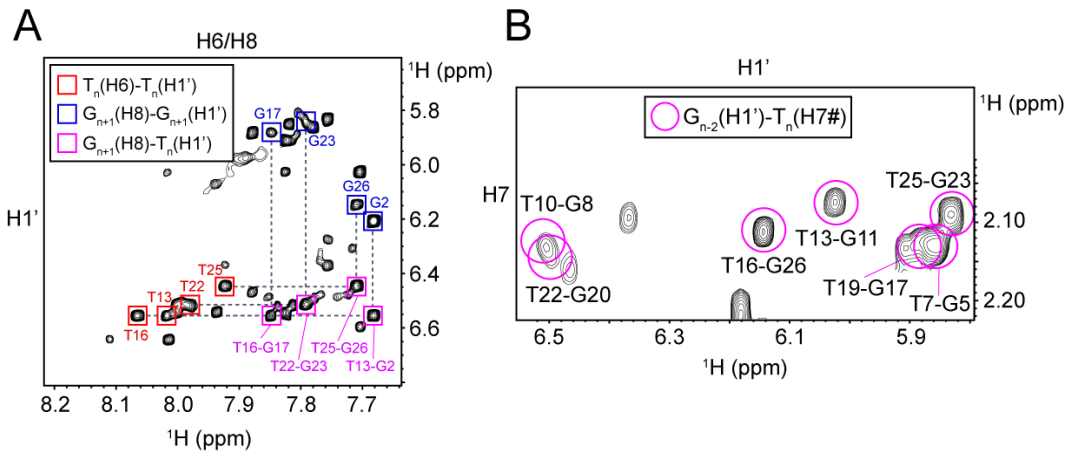


Figure 3.32: NOESY spectra of *Z-G4* at 300 ms mixing time. A) H6/H8-H1' region. Shown in red and blue boxes are the intra-residue H1'-H6/H8 cross-peaks of T-cap and G-tetrad guanine respectively. Shown in purple boxes are the inter-residue cross-peaks form the T-loops to the G-tetrad guanines T_n(H1')-G_{n+1}(H8), which uniquely appear only in presence of T-cap. B) H1'-H7# region. Shown in purple circles are inter-residue cross-peaks from the T-cap methyl groups to the sugar proton of G-tetrad guanines T_n(H7#)-G_{n-1}(H1').

The significance of the additional stacks of thymines (further: T-cap) in stabilizing the G-quadruplex structures as well as their necessity towards formation of left-handed G-quadruplexes are still to be explored. However, the ramification of the T-capping has been observed via amino proton NMR signal. Using the lineshape analysis method explored in section 2.3, we observed the amino rotation rate of the guanines in the outer tetrads of *Z-G4*. The results suggested that the amino proton might form additional hydrogen bonds towards the sugar oxygen of the T-cap residues (Figure 2.14). Based on this result, we used the amino rotation rate as a probe towards the existence of T-capping feature for further left-handed

G-quadruplex studies.

b. Twisted backbone progression

The sugar-phosphate backbone in right handed double-helix DNA as well as in G-quadruplex has a certain default orientation. Specifically, if we look at the DNA backbone strand from 5' to 3' as a forward direction, all the O4' atoms of the sugars will be oriented backwards. This arrangement is more "natural" since the sugar orientation is parallel with the overall backbone/strand orientation.

To explain further, consider a vector formed with origin at 5'C and end point at 3'C of the same residue, and a second vector formed from the first (5') guanine base to the second (3') guanine base (Figure 3.33 top). In right handed G-quadruplex, these two vectors are aligned in parallel (or close to parallel) with each other. However, in the case of *Z-G4*, the two vectors are at a high angle (almost 90°) with each other, resulting in a "twisted" backbone progression.

As a feature of a left-handed G-quadruplex, this twisted backbone is observable by NMR. The distance between protons H2'/H2"(n+1)-H1'(n) is significantly nearer than the "natural" backbone counterpart (Figure 3.33 middle). This feature is reflected clearly in the NOESY spectrum of *Z-G4* (Figure 3.34), where cross-peaks between H1' atoms of a guanine and H2'/H2" atoms of the subsequent guanine were observed. This NOESY pattern is not observed in any parallel right-handed G-quadruplex data.

c. Rotation angle between stacking guanines

As a consequence of the "twisted" backbone orientation of the left-handed G-quadruplex block, the stacking arrangement between two guanines is changed as oppose to the usual right-handed stacking. The angle of rotation between the

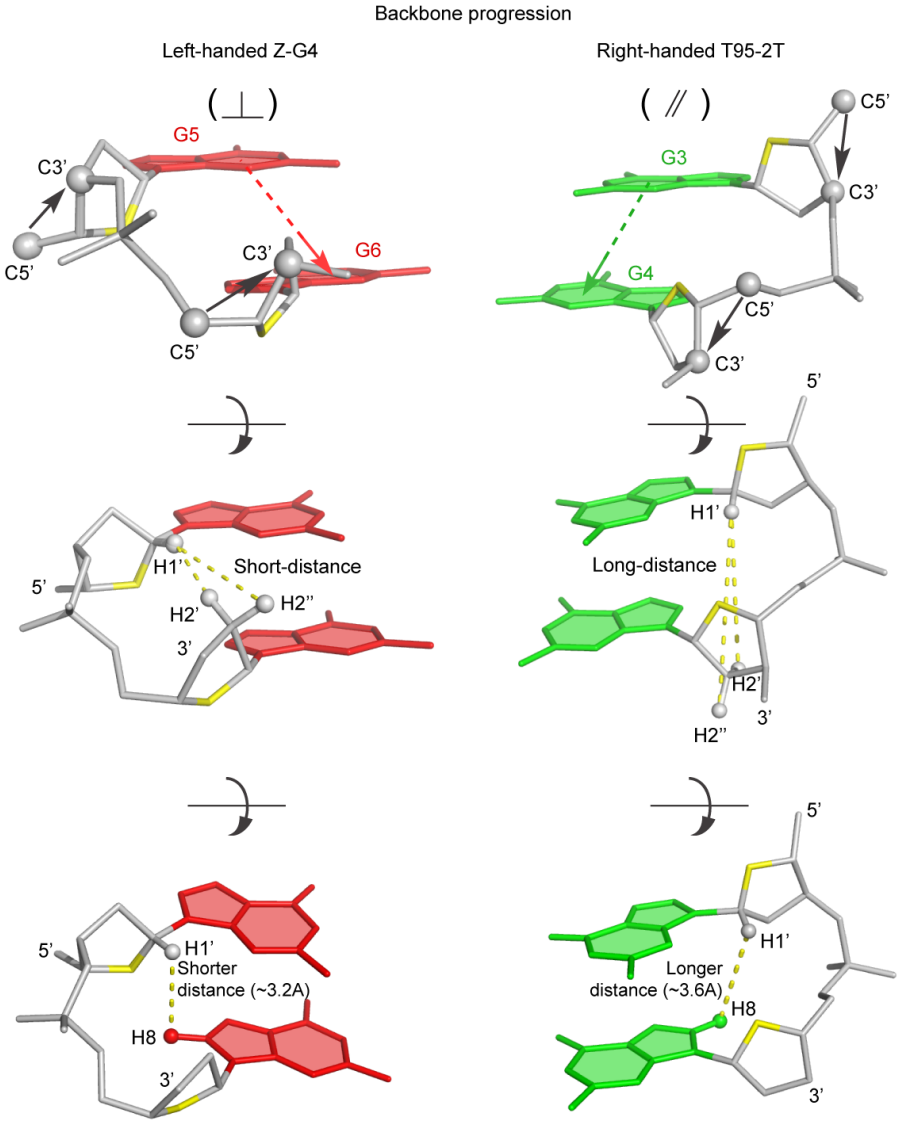


Figure 3.33: Detailed view on the sugar-phosphate backbone of left-handed and right-handed G4 (Top). Perpendicular relative orientation was observed between intra-residue C5'-C3' vectors (black arrows) and the guanine stacking orientation (red dashed arrow) for left-handed G4, while parallel relative orientation was recognized between the same intra-residue vectors (black arrows) with the guanine stacking (green dashed arrow) for right-handed G4. (Middle) Short distances between inter-residue protons H1'(n)-H2'/H2"(n+1) (yellow dashed lines) in left-handed stacking and long distances from the same set of protons in right-handed stacking. (Bottom) Shorter distance between inter-residue protons H1'(n)-H8(n+1) (yellow dashed line) of left-handed stacking guanines compared to those of right-handed stacking guanines.

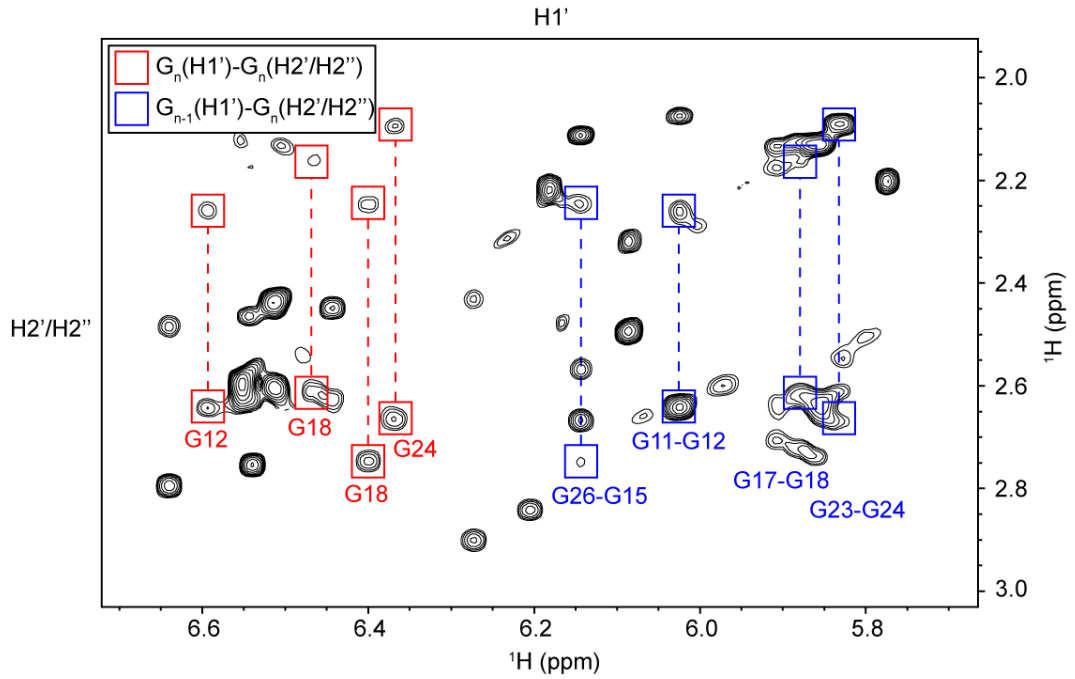


Figure 3.34: NOESY spectrum of *Z-G4* at 300 ms mixing time in H1'-H2'/H2" region. Red pairings (two boxes and connecting dashed lines) indicate the intra-residue cross-peaks of H1'-H2'/H2". Blue pairings indicate the inter-residue cross-peaks between left-handed stacking guanines G_{n-1}(H1')-G_n(H2'/H2"). The blue cross-peaks are direct consequences of observed distance in Figure 3.33 (middle) and can be used for probing left-handed stacking.

guanine base are different in direction as well as in magnitude. Previous study showed that left-handed stacking guanines rotate by ~25.7° while the right-handed counterparts rotate by ~29.4° in the opposite direction (Figure 3.29G, H).

Indirect NMR observables of this feature are observed via inter-residue H1'-H8 NOESY cross-peak intensities. The inter-residue cross-peaks H1'(n)-H8(n+1) are observed to be considerably stronger than the intra-residue cross peaks H1'(n)-H8(n) in the case of left-handed arrangement (*Z-G4*), while they are of the same intensities for a right-handed arrangement (*T95-2T* dimer, *J19*) (Figure 3.35). Since we know that all intra-guanine distances of H1'(n)-H8(n) for these structures are constant (due to all guanine residues adopting *anti* glycosidic bond conformation), we can use the intra-guanine cross-peaks as control and infer that the

inter-residue H1'(n)-H8(n+1) distances of left-handed stacking guanines are considerably closer compared with those of right-handed stacking guanines.

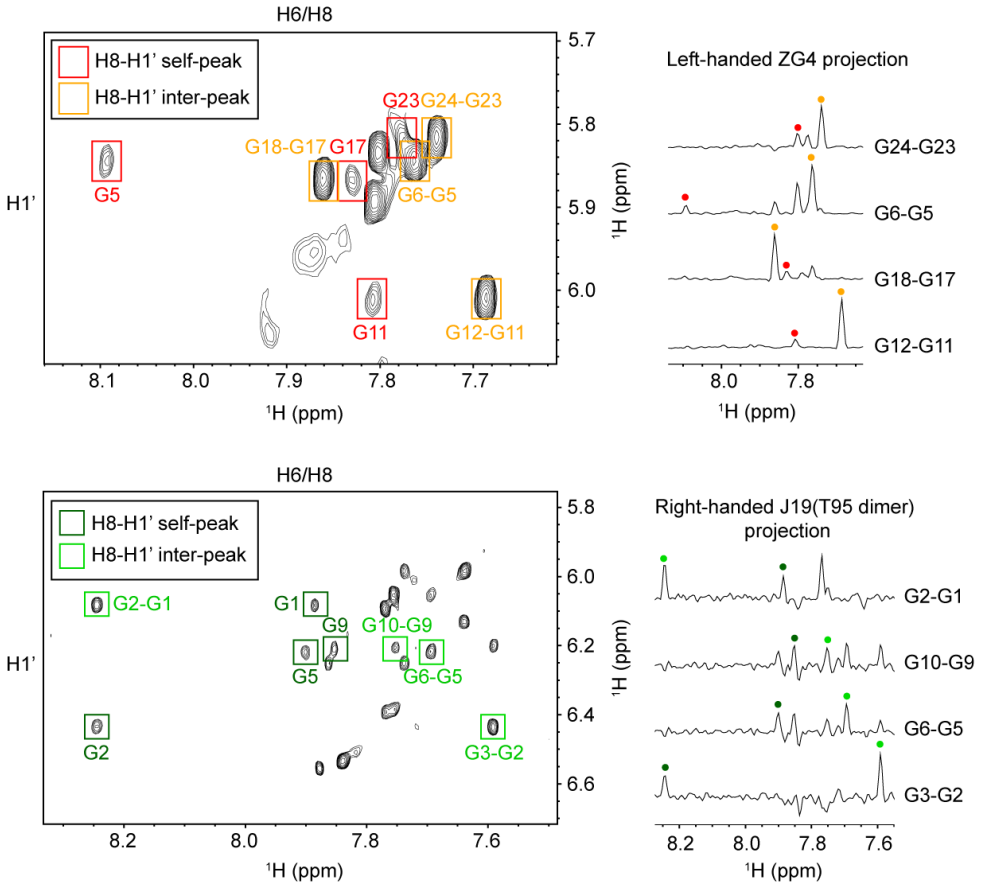


Figure 3.35: NOESY spectra of *Z-G4* and *J19* at 100 ms mixing time. NOESY spectra were marked with intra-residue H1'-H8 cross peaks (red and dark green boxes respectively) and inter-residue G_n(H1')-G_{n+1}(H8) cross-peaks between stacking guanines (orange and light green boxes respectively). Horizontal slice 1D projections were shown to the right, with the corresponding cross-peaks being color coded the same way. Significant differences between intra (red dots) and inter-residue (orange dots) cross peaks in *Z-G4* were observed. On the contrary, similar intensity peaks between intra (dark green dots) and inter-residue (light green dots) cross-peaks in *J19* were observed. The observations are direct consequence of observed distance in Figure 3.33 (bottom).

Intuitively, the change in the rotation magnitude and the inter-residue H1'(n)-H8(n+1) distance might be geometrically related. Thus, we could use the characteristic strong peaks in left-handed arrangement for probing left-handed stacking

in other sequences/structure. Based on previous studies [211, 212], we hypothesized that the different stacking layout of the two tetrads might affect the cation preference and stability inside the G-quadruplex ion channel. Further exploration of the idea is currently ongoing (Bakalar, B. *et. al.*, unpublished (2)).

3.4.4 High-resolution structure of $(GGT)_8$

Structure determination of $(GGT)_8$ by NMR spectroscopy was performed with the help of the left-handed G4 fold characteristics listed in section 3.4.3.

a. Observed left-handed features

To further prove the hypothesis of left-handed block formation in one-half of the structure, the NMR signal was divided into two halves and observed in detail. The results of (G1-T12) NMR signal matches the characteristic left-handed G4 feature previously studied. On the other hand, lack of left-handed features was also observed for the NMR signal of second-half (G13-T24) of the structure, indicating a G-quadruplex left to right junction in between the two halves. Presented below is the detailed explanation on the evidences:

First, sharp imino to amino NOE cross-peaks of G1, G4 and G7 residues were observed indicating hydrogen bonds formed from G1(H22) to T3(O4'), G4(H22) to T6(O4') and G7(H22) to T9(O4'). In contrast, imino to amino NOE cross-peaks for G14, G17, G20 or G23 were not observed (Figure 3.36A). These hydrogen bonds are supporting evidences of the T-capping formation from the single thymine propeller loop specifically T3, T6 and T9. The unobservable G10(H1)-G10(H22) cross peak — indicating lack of hydrogen bond between G10(H22) to T12(O4') — is expected because T12 has to act as a linker between the two halves of the structure and could not adopt the T-cap orientation.

Second, the inter-residue NOE cross-peaks between H1'(n)-H2'/H2"(n+1) were only observed for the first block, indicating its left-handed "twisted" strand progression (Figure 3.36B). These strong cross-peaks were observed for all four guanine stacks of the first block (G1-G2, G4-G5, G7-G8 and G10-G11), yet unobservable for all four guanine stacks of the second block (G14-G13, G17-G16, G20-G19 and G23-G22), clearly suggesting their differences.

Third, the relatively strong inter-residue cross-peak intensities of H1'(n)-H8(n+1) were observed only for the first block, while medium inter-residue cross-peak intensities were observed for the second block (Figure 3.36C, D).

Based on all the observations, the stable structure conformation of *(GGT)$_8$* is determined to be a parallel left-handed right-handed G-quadruplex hybrid.

b. High-resolution structure

The high-resolution structure calculated using NMR spectroscopy data is presented in the form of 10 superimposed lowest energy structures and the best ribbon representation (Figure 3.37). The structure statistics is presented in Table 3.6.

Similar to Z-to-B duplex DNA junction [93], there might exist a "junction" (or in G4 case, "stacking") in G4 structures, which connect a left-handed G-quadruplex block to a right-handed G-quadruplex block. The discovery of left-handed to right-handed stacking is analogous to the discovery of Z to B-DNA junction in DNA double helix, implying even more diversity in G-quadruplex structure topologies. In this section, we studied one DNA sequence *(GGT)$_8$* that displays the properties of left to right-handed G-quadruplex stacking and solved its high-resolution structure with NMR spectroscopy. In addition, we also showed that NMR spectroscopy can be used to determine the left-handed G4 progression via various NMR observables

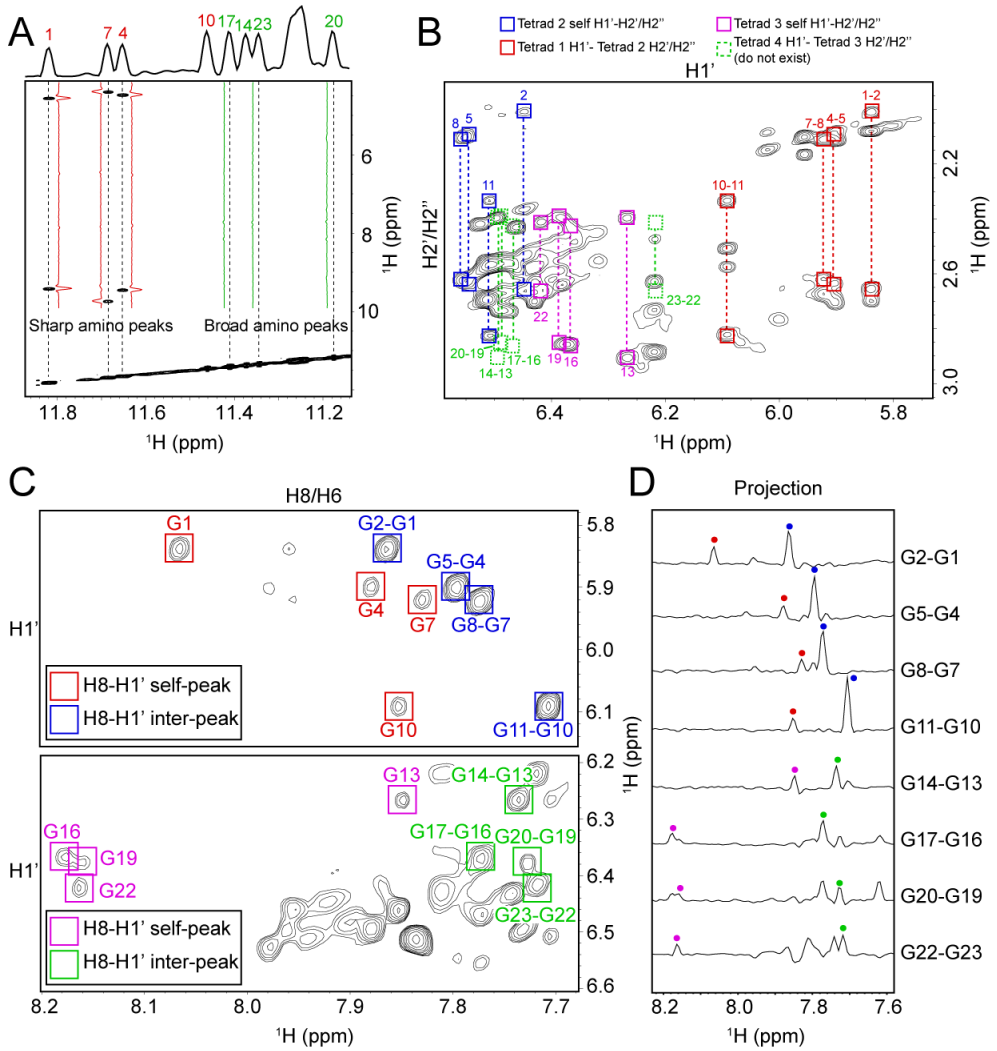


Figure 3.36: Left and right-handed evidences of *(GGT)₈* structure. A) NOESY spectrum of *(GGT)₈* at 250 ms mixing time in H1-H21/H22 region. Indicated by dashed lines are the outer tetrad imino proton slices: 1st tetrad (red projection) imino slices show strong twin peaks; 4th tetrad (green projection) imino slices do not show any considerable peak. B) NOESY spectrum of *(GGT)₈* at 300 ms mixing time in H1'-H2'/H2" region: red (1st tetrad) and blue (2nd tetrad) pairings shows left-handed H1'-H2'/H2" cross-peaks; magenta (3rd tetrad) and green (4th tetrad) pairings do not show the feature. C) NOESY spectra of *(GGT)₈* at 300 ms mixing time in H8/H6-H1' regions and D) their horizontal slice projections: red and blue boxes and dots represent first-half of the sequence, showing significantly higher intensity for the inter-residue H8-H1' cross peaks (blue); magenta and green boxes and dots represent second-half of the sequence, showing comparable intensity between the intra and inter-residue H8-H1' cross-peaks.

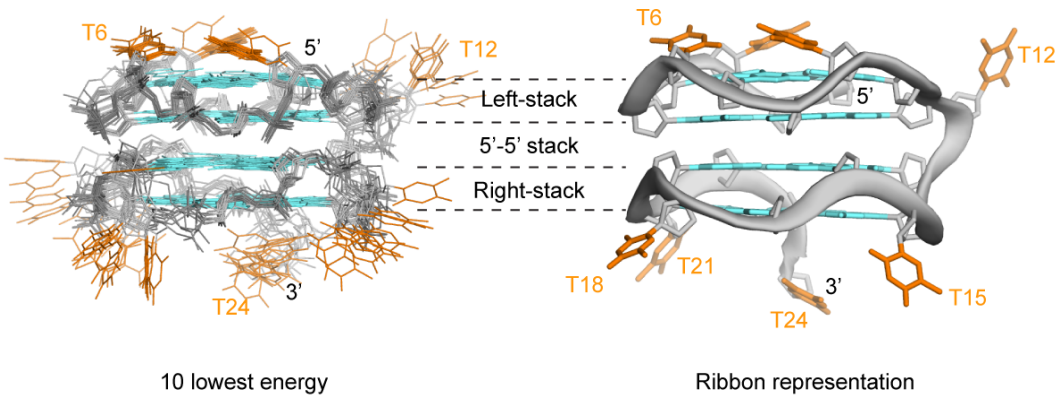


Figure 3.37: 10 superimposed lowest energy structures and the best ribbon representation of $(GGT)_8$. Left-handed stacking, 5'-5' interface, and right-handed stacking are indicated. Cyan and orange represent *anti*-guanine and thymine residues respectively.

produced by left-handed structural features.

3.5 Conclusions

3.5.1 Summary

In 4n-1 G-quadruplex-dinucleotide complexes, the following facts are discovered:

- Discovery of an empty pocket in 4n-1 G4 that can act as a binding site for G4 ligand, especially guanine derivatives
- Binding of cyclic-dinucleotides, which are known cell-signaling metabolites in bacteria cells, potentially leads to regulatory function

In *AT21S* study, novel G-quadruplex fold has been revealed, which highlights the following:

- Two-layer anti parallel G4 core with stabilizing G-T-T triad and a G-T base pair, completed with two edgewise loops, a V-shaped loop and a diagonal loop

Table 3.6: Statistics of the computed structure of *(GGT)$_8$*.

A. NMR restraints		
Distance restraints	Exchangable	Non-exchangable
Intra-residue	6	464
Inter-residue	60	178
Other restraints		
Hydrogen bonds	70	
Dihedral angles	16	
B. Structure statistics		
NOE violations		
Number ($> 0.2\AA$)	0.2 ± 0.4	
Deviations from the ideal covalent geometry		
Bond lengths ($\AA$)	0.003 ± 0.000	
Bond angles (^{o})	0.697 ± 0.004	
Impropers (^{o})	0.392 ± 0.009	
Pairwise heavy atom rmsd value ($\AA$)		
All heavy atom	0.528 ± 0.088	
G-tetrad core	0.404 ± 0.058	

- Conserved G-G-T triad structure upon changing G4 core with other sequence (*Pu24T*)

In *(GGT)$_8$* hybrid structure, we explored:

- Several defining characteristics of left-handed block observed by NMR: T-capping; twisted backbone progression; and difference in rotation angle of the stacking guanines
- Stacking conformation between a right-handed G4 block and a left-handed G4 block

3.5.2 Future works

The concept of 11-guanine G-quadruplexes could be extended towards native sequences in human genome. The evidence of ligand binding on the vacant spot provided here could be the basis for drug design that specifically target the 11-guanine G-quadruplex forming in vivo.

Cut and paste study of the diagonal loop is the main future perspective on the *AT21S* project. Further NMR examinations of the structure with G-quadruplex core extension, base mutation, etc. would provide the necessary information towards understanding the structural elements that affects stability of the fold-back loop as well as the G-quadruplex structure as a whole.

Minimal motif of left-handed G-quadruplexes and effect of ammonium ion in left-handed G4 stability are actively studied. Future perspective would be to explain the CD spectroscopy signal of the left-handed folding topology, further exploration on the left-handed stack (e.g. three-layer or anti-parallel possibility) as well as finding or improving the most efficient and deterministic way of probing the left-handed fold.

Chapter 4

G-quadruplex-based Network in Biofilm

Biofilms are ecosystems that include microbial communities and their environment. They are key factors that contribute to bacterial pathogenicity [213], disruption of water filtration systems [214], and assistance to wastewater treatment bioprocess [215] among others. Microbes forming biofilms have benefits of increased antibiotic tolerance [216], enhanced resource availability [217], and establishment of ecological microniches [218]. These properties are mediated by an extracellular polymeric matrix secreted by the cells [219], which are unique to biofilms and not to other planktonic bacteria. In this chapter, CD and NMR spectroscopy are used to study the gel-forming property of biofilms and identify the components in biofilms that contribute towards polymer networking.

Part of this chapter is taken from the author's unpublished work under the title of:

Seviour, T., **Winnerdy, F.R.**, Wong, L.L., Shi, X., Mugunthan, S., Kohli, G.S., Shewan, H.M., Stokes, J.R., Rice, S.A., Phan, A.T. and Kjelleberg, S. The biofilm matrix scaffold of *Pseudomonas* species consists of non-canonically base paired extracellular DNA and RNA.

The experiments in these sections were performed by the author's collaborators Dr. Thomas Seviour et. al.:

Subsection 4.1.1 — Rheology experiment

Subsection 4.2.1 — Isolation of nucleic acid using EMIM-Ac

Subsection 4.5.2 — Staining and visualization of DNA and RNA in isolated and crude biofilm

Subsection 4.5.3 — Biofilm growth in different cations

The experiment in this section was performed with help of the author's collaborator Dr. Shi Xiangyan:

Subsection 4.5.3 — Solid state NMR, ^{1}H-^{15}N HETCOR of isolated biofilm

4.1 Biofilm of *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is chosen as the object of study due to its extensive use in literature [220], as well as its contribution towards one of the five clinical infections caused by microbial biofilms [221]. Eight polymers have been found to exist in the extracellular matrix, therefore called exopolymers, that support key traits in *P. aeruginosa* biofilms. These exopolymers include three exopolysaccharides [222], four proteins [223–225], and an extracellular DNA (eDNA) [226]. Previously, the polysaccharides were described to have a certain structural role in the biofilm with functional redundancy with each other [227], while other exopolymers were also analyzed to have multiple roles [228, 229].

4.1.1 Measurement of biofilm elasticity

In order to track the main component in biofilm formation, a method to quantify the biofilm networking property has to be established. Since biofilm is a result of polymer networking [95] that produces certain viscoelasticity [94], detection of the elastic property of biofilms is a good way to determine biofilm formation. Nevertheless, the exact quantification of the elastic property might not be a simple task since each exopolymer in the biofilm can contribute to this property interdependently or independently of each other. Thus, we used the reported ability of the ionic liquid, 1-ethyl-3-methyl-imidazolium acetate or EMIM-Ac, to dissolve a range of recalcitrant biopolymers including DNA [230] and cellulose [231].

Although the gel-like property of crude biofilm is not conserved upon the dissolution of the biofilm with the ionic liquid, the elastic properties of the polymers are carried over. This elastic behavior can be measured with a rheology technique, where the normal pressure difference ($N_1 - N_2$) of the liquid is detected while spinning the sample with varying shear-rates. For polymer solutions, the secondary normal

stress is considered zero. The dependence of the normal pressure (N_1) towards the shear rate is regulated by the elasticity of the solutions, with higher dependency correlating to higher elasticity. Overall, the elastic property of the ionic liquid-dissolved biofilm can be quantified using rheological method.

4.1.2 DNA as a structural component in biofilm

For specific determination of the main polymer component that contributes to the elastic response in biofilm, various biological enzymatic processes were used to specifically eliminate certain polymers. For instance, genetic knockout mutants were used to remove the polysaccharides Pel and Psl, and enzymatic treatments with pronase, RNase A, and DNase I were done to discard the contribution of protein, RNA, and DNA respectively. The resulting graph of normal pressure versus the shear rate shows that only the solution treated with DNase I displays a complete removal of the dependency of normal pressure on shear rate, suggesting very low elasticity (Figure 4.1). Hence, it can be concluded that the elastic response of the biofilm is attributed to DNA or DNA interaction with other polymer/s.

Experiments on planktonic cells showed that treatment with the ionic liquid leaves the cells intact (without cell lysis), as indicated by the preservation of the cell morphology. Thus, we can further specify that the elastic response from the biofilm is due to extracellular DNA (eDNA) and not chromosomal DNA. A previous study showed that eDNA is indeed a component that contributes towards the extracellular matrix of cells [232,233]. However, the fact that eDNA is responsible for the elastic response suggests that eDNA is not an incidental matrix component but is instead an integral element contributing to the structure of the *P. aeruginosa* biofilm.

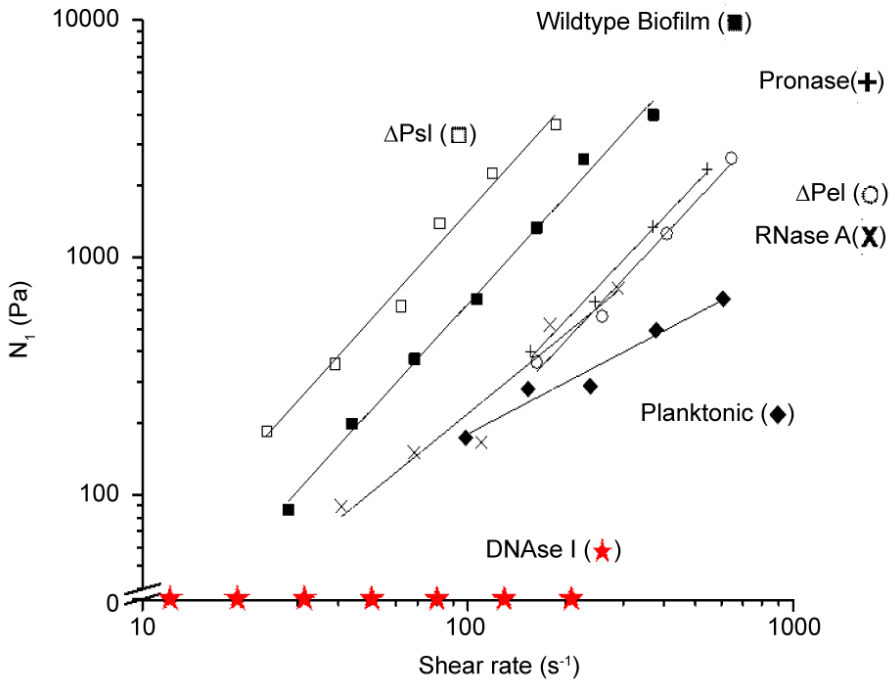


Figure 4.1: Primary normal stress differences (N_1) of *Pseudomonas aeruginosa* biofilm: wild type; ΔPsl;ΔPel; pronase digested; RNase A digested; DNase I digested; and planktonic cells dissolved in 1-ethyl-3-methylimidazolium acetate (40 mg/ml) at 25°C. Lines indicate power-law fits to the data, any data below 0.1 Pa were set to 0. Figure and data courtesy of Thomas Seviour *et. al.*

4.2 Isolation of nucleic acids from biofilms

The results of the rheology study suggest that nucleic acids are responsible for the structural property of crude biofilms. However, there is no evidence that nucleic acids are the sole biopolymers that contribute to this property of biofilm. In order to address this issue, the isolation of nucleic acids was performed. The process of isolation includes dissolving the biofilm in ionic liquid, recovering the nucleic acids through perchloric acid precipitation as well as re-dissolving the isolate with water. The isolation process is also helpful for spectroscopic studies as the resultant samples produce simpler CD and NMR spectra, which allow quantitative analysis

to be performed more efficiently.

4.2.1 Isolate mimics the gel-forming property of biofilms

The first step of isolation involves the dissolution of the biofilm in ionic liquid, which was done for the rheology experiments previously. The next step comprises of a selective precipitation process using perchloric acid. After this step, the resultant precipitate is mostly nucleic acid, while the supernatant contains proteins and other components. Following re-dissolution into water, the nucleic acid-containing precipitate reassembles into a gel (gel-isolate), mimicking the ability of native biofilms in forming polymeric networks (Figure 4.2). On the other hand, other fractions that did not get precipitated by the perchloric acid do not self-assemble into gels. These results suggest that nucleic acids in the extracellular matrix environment can indeed form a network similar to those in native biofilms independently without the existence of other exopolymers.

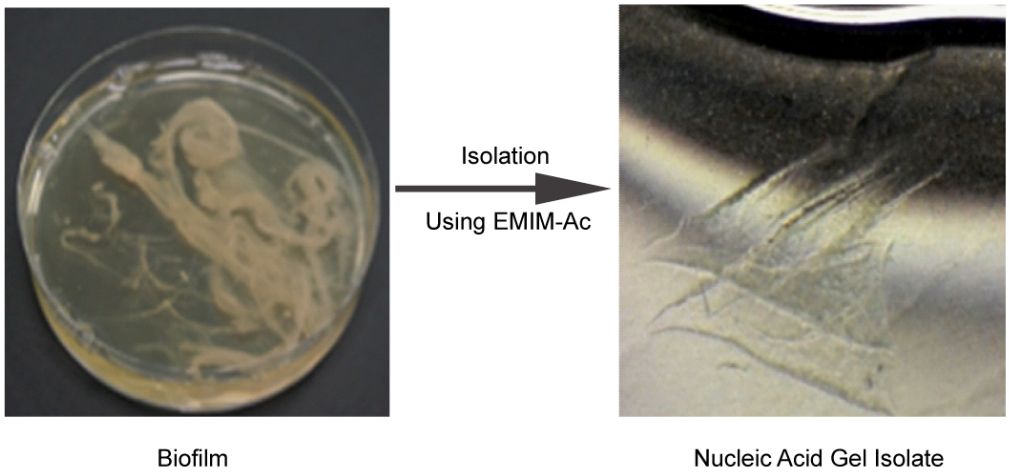


Figure 4.2: Extracellular nucleic acids isolation. Nucleic acid extracted from *P. aeruginosa* biofilm reconstituted into a gel upon transfer from 1-ethyl-3-methylimidazolium acetate into water. Figure and data courtesy of Thomas Seviour *et. al.*

To make sure that the gel-forming property is not a universal property of all nucleic

acids but rather specific to eDNA, the same process was carried out using another cell system. The corresponding isolate of calf-thymus DNA did not have the ability to form a gel, which positively indicates that the gel-forming ability is a unique characteristic of extracellular nucleic acids present in the biofilm isolate.

4.2.2 Nucleic acid conformation is preserved upon isolation

Further characterization of the gel-isolate is required to confirm the preservation of the nucleic acid when isolated from its native-biofilm gel counterpart. The two samples i.e. the native biofilm gel (further designated as crude or native) and the gel-isolate (further designated as isolate) were analyzed by circular dichroism spectroscopy (Figure 4.3).

For the crude sample, the resulting CD spectra at various temperatures showed a large trough at ∼220-230 nm and a relatively small peak at ∼260-285 nm. The trough was designated to protein signal [234], which is further confirmed by its absence in the spectrum for a sample that was subjected to proteolysis and fractional precipitation for removal of protein. The signal at ∼260-285 nm was identified as signal from nucleic acids [235]. This experiment also suggested that the amount of nucleic acid inside the crude biofilm is relatively low compared to the amount of protein. As for the isolate sample, the resulting CD spectra showed a major peak at ∼260-285 nm which corresponds to signal from nucleic acids. As expected of the isolation process, most of the material in the isolate is indeed nucleic acid. In addition, the fact that the peak is spread over the wavelength of ∼260-285 nm suggests that the nucleic acids are polymorphic.

By comparing the wavelength of the maximum peak from both the crude and the isolate samples, it is possible to infer that the nucleic acid components in both

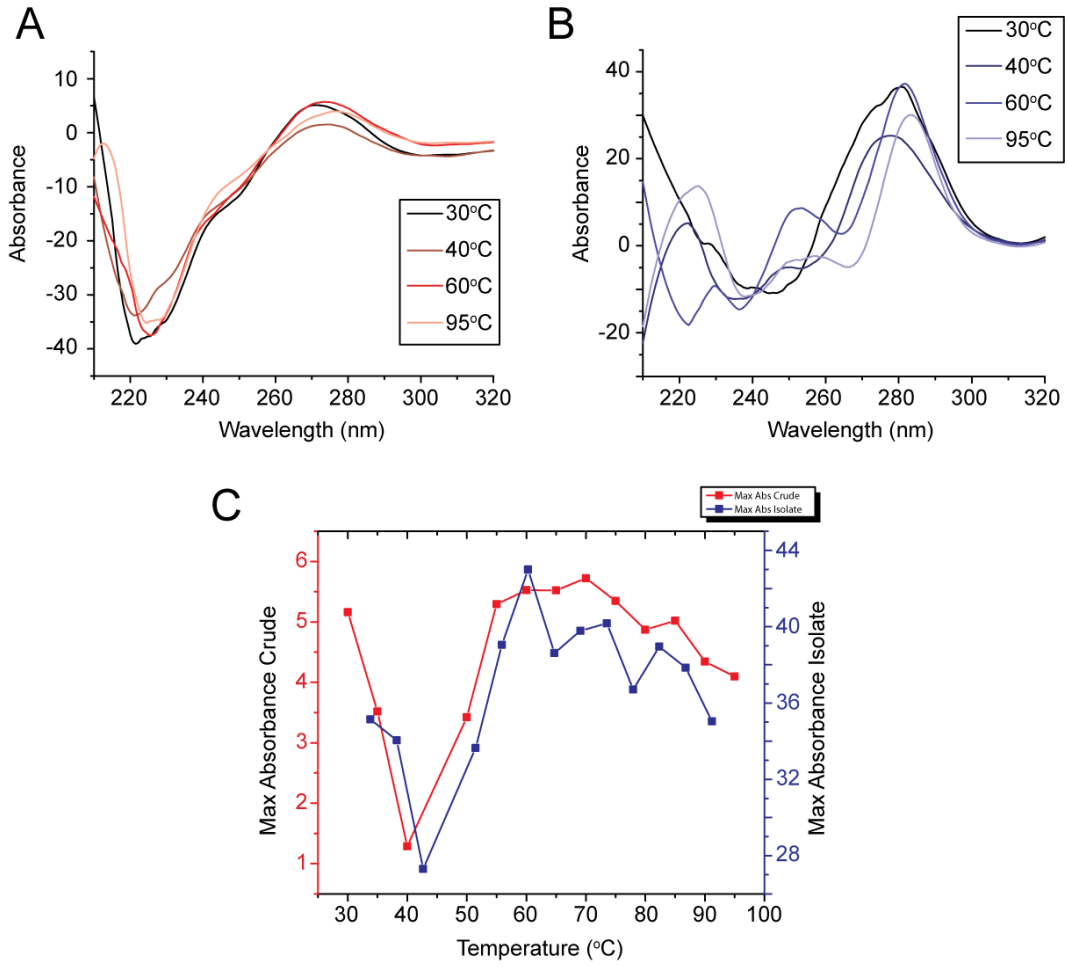


Figure 4.3: Circular dichroism spectra of A) crude *Pseudomonas aeruginosa* biofilm and B) extracellular nucleic acids gel isolate at temperatures ranging between 30 and 95°C. C) Amplitudes of dominant nucleic acid peak (260-285 nm) from both crude and isolate sample plotted against the temperature.

samples were similar in their polymorphic configuration. To confirm the hypothesis further, melting experiments were done for both the samples, and the maximum CD ellipticity signals were recorded at each temperature point (Figure 4.3C). The curves show a similar pattern between the crude and isolate samples, with both having the least maximum absorbance at ∼40°C and the highest maximum at ∼60°C, with a steady decrease until 90°C. Based on the experimental results, the nucleic acid conformations were known to be preserved through the isolation process. Furthermore, the existence of CD signal even at 90°C indicates more stable

conformations than typical double-helix DNA duplexes [236]. Following this result, the isolate sample was therefore determined to be an accurate substitute for the study of the networking behavior of native biofilms.

4.3 Nucleic acid conformation in biofilms

NMR spectroscopy was chosen to further study the conformation of the nucleic acids present in the isolate sample. However, a complete dissolution of the sample is required which could not be achieved by dissolving the sample in water at neutral pH. Thus, alkali treatment up to pH 12 [237] was performed to observe the composition of the isolate sample using ^{31}P NMR. The principles and results of ^{31}P NMR of nucleic acid phosphate groups are presented below:

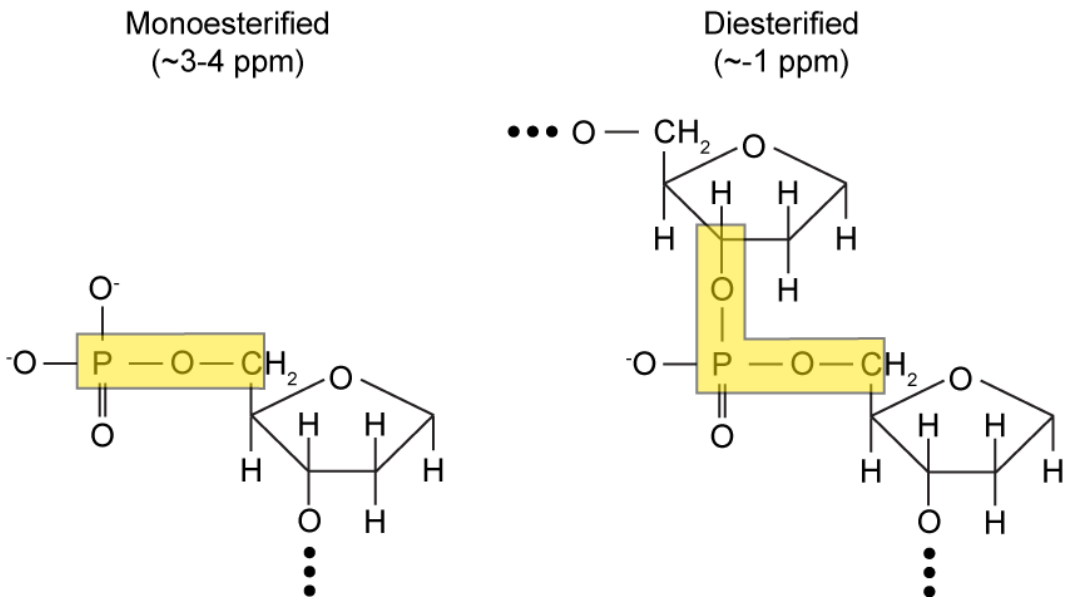


Figure 4.4: Schematic of monoesterified (left) and diesterified (right) phosphate groups the in nucleic acid backbone. Highlighted in yellow are the ester groups formed, a single ester group produces phosphorous peaks at ∼3-4 ppm and double ester groups result in peaks at ∼-1 ppm.

4.3.1 Identification of di-esterified and mono-esterified phosphate

In the 1D ^{31}P NMR spectra, phosphorous nuclei have distinctly different resonance frequencies depending on the local conformation of the atoms surrounding them. Herein, we focus on the phosphorous atom in a phosphate group, where it could be in a mono-esterified state or in a di-esterified state (Figure 4.4). In the former state, the phosphorous peak appears in 1D ^{31}P NMR at ∼3-4 ppm region for pH 7 to 12, whereas in the latter, the phosphorous nucleus produces a peak at ∼-1 ppm. This difference in chemical shift is big enough that there is no ambiguity in determining the esterification state of the phosphorous nuclei using this method.

In practice, a mono-esterified phosphorous atom indicates a single monomeric state of the nucleotide (mononucleotide). On the other hand, di-esterified phosphorous atoms are indicative of polynucleotide or chain structures, such as DNA or RNA. Due to the large size of the polynucleotide chains, the NMR peaks of the di-esterified phosphorous atoms are generally broad.

4.3.2 Relative abundance of polymeric and monomeric nucleic acids

At pH 7, the abundance of mono-esterified and di-esterified nucleotides in both the crude and isolate samples was observed using solid-state NMR spectroscopy (Figure 4.5A). As expected, all of the observed phosphorous peaks showed a polymeric nucleotide or chain conformation of the nucleic acid. However, at pH 12, the same measurements were made but the same observations were not detected. Part of the nucleic acids disintegrated into their monomeric counterpart, while some retained the chain structure (Figure 4.5B). This result could be translated into the possibility of two different initial conformations of the nucleic acid being present

in both the crude and isolate samples at neutral pH, with one being more stable than the other.

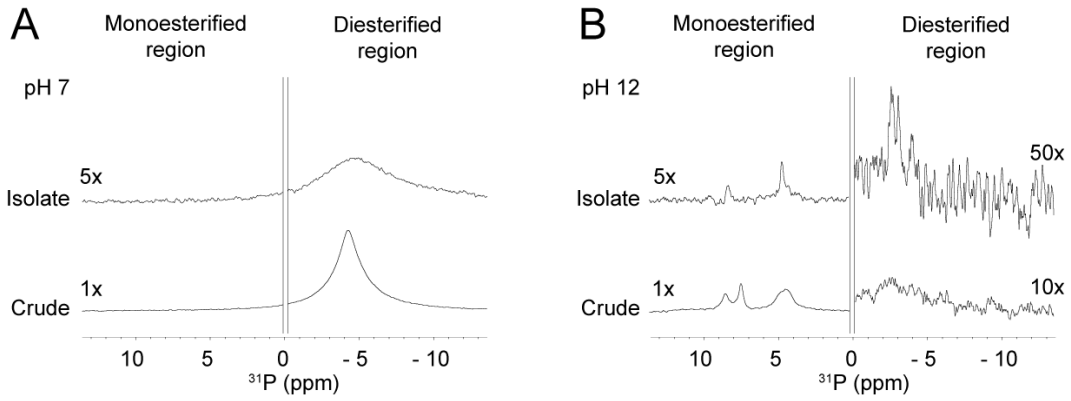


Figure 4.5: Solid state ^{31}P NMR spectra at T = 25°C of extracellular nucleic acid gel isolate and native crude Pseudomonas aeruginosa biofilm in A) neutral condition and in B) alkali condition post-lyophilization. The isolate spectra are amplified 5x from the crude spectra due to significantly lower isolate sample concentration. Only diesterified peaks appeared in the neutral condition, while both monoesterified and diesterified peaks coexisted in the alkali condition, indicating transesterification during alkali treatment.

At elevated pH, solution NMR spectra of crude and isolate samples allowed the observation of 8 sharp peaks pertaining to the mononucleotide conformation of the nucleic acids (Figure 4.13). The identities of these peaks would allow the determination of the exact composition of nucleic acids present within the isolate, specifically the part that was disintegrated during alkali treatment. Lastly, comparison between the crude and isolate ^{31}P spectra once again shows the similarity in the nucleic acid conformation between the two.

4.4 Solution NMR analysis of isolate *Pseudomonas* biofilm at elevated pH

Solution state NMR is a very useful tool to determine the composition and structure of relatively small molecules. However, it is not very good for studying large or insoluble molecules. Therefore, an elevated pH condition is needed to completely dissolve the gel-isolate and produce small molecules of interest that can be observed with solution NMR. Thus, although the presence of polynucleotide chains would not be dismissed, the focus of this subsection is to study mononucleotides that are produced upon alkali treatment.

4.4.1 ^{13}C-^{1}H HSQC

Heteronuclear Single Quantum Coherence (HSQC) NMR experiment correlates one type of nuclei to another type of nuclei via through-bond magnetization transfer (Figure 4.6A). In this context, ^{13}C-^{1}H HSQC experiment was used to correlate the carbon resonance frequencies in nucleic acids to those of the neighboring protons. The spectrum contains the information of every carbon-hydrogen pair stored in a 2D layout, with every cross-peak representing one pair.

Herein, two main regions were studied, namely the aromatic region (proton-carbon pair in the aromatic rings of nitrogenous bases) and the anomeric region (H1'-C1' pair in the sugar rings). The former region is located at δ_{1H} =∼7-9 ppm, where 8 sharp peaks and broad humps were detected. Similarly, in the anomeric region at δ_{1H} =∼5-6 ppm, the same set of peaks were found (Figure 4.7). These results indicate the existence of two families of nucleic acids at elevated pH, which most likely came from two different original states at neutral pH. Hence, it is assumed that the first original state of the nucleic acids is unaffected by an increase in pH and remains as part of chain structures — producing broad humps, while the second

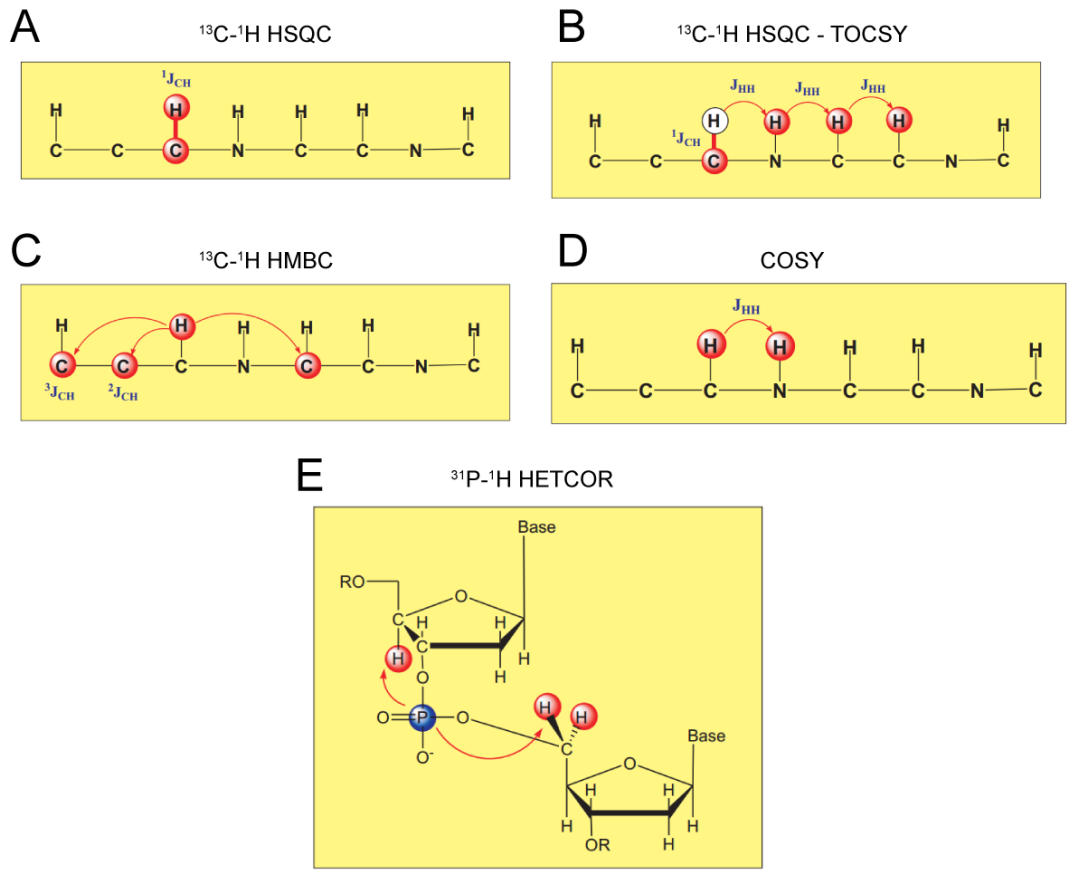


Figure 4.6: Schematics of magnetization transfer in solution NMR experiments with each coupling constant. A) ^{13}C-^{1}H Heteronuclear Single Quantum Coherence (HSQC), B) ^{13}C-^{1}H HSQC-Total Correlation Spectroscopy (TOCSY), C) ^{13}C-^{1}H Heteronuclear Multiple Bond Correlation (HMBC), D) ^{1}H-^{1}H Correlation Spectroscopy (COSY), and E) ^{31}P-^{1}H Heteronuclear Correlation (HETCOR). Figure is adapted from ([238]).

original state break down during alkali treatment and forms mononucleotides — generating a set of sharp peaks (note: the possibility that mononucleotides were originally there at the neutral pH can be dismissed based on the solid-state NMR result in the previous section (4.3.2)).

In order to study the importance of the two original states, the pH for the same sample was brought down to neutral conditions again. Unsurprisingly, gel property of the isolate was not restored, indicating the importance of the second original state as neutral pH would not ligate back the mononucleotides into a polynucleotide

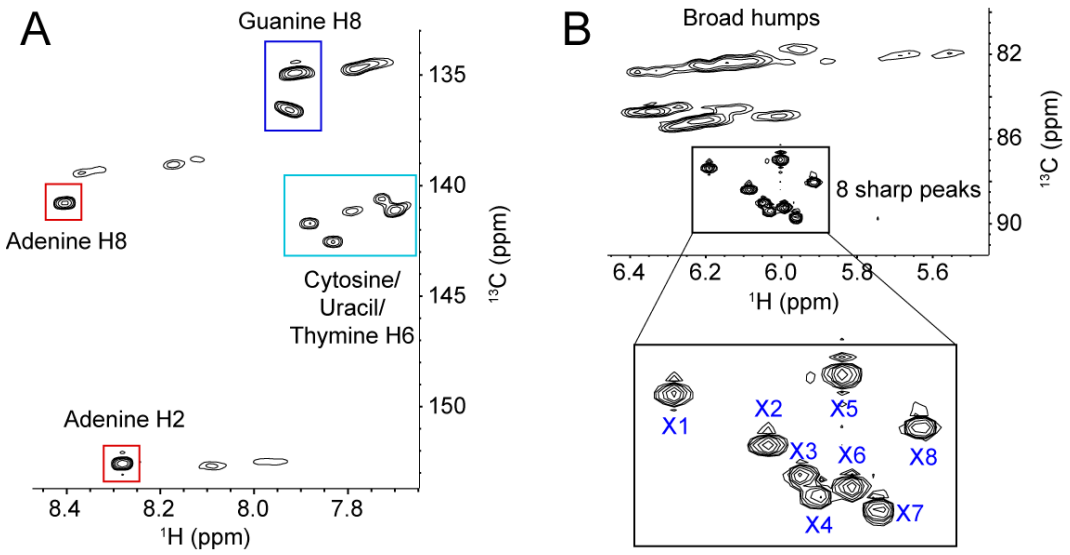


Figure 4.7: HSQC spectrum of isolate gel sample at T = 25°C and pH = 12. A) Aromatic region: sharp carbon-proton coupling peaks are marked with color coded boxes. Red indicates adenine C8-H8 and C2-H2 pairs, blue indicates guanine C8-H8 pairs and cyan indicates cytosine and thymine/uracil C6-H6 pairs. B) Anomeric region: 8 well-resolved C1'-H1' peaks (labeled X1-X8) corresponding to mononucleotides, and broad humps suggesting chain formation.

form. In conclusion, either the second state alone is responsible for the gel property of the isolate or, the more likely scenario whereby both the original nucleic acid states are responsible for the maintenance of the networking property in the gel isolate. Either way, the exact identities of the two states are very important for understanding the networking nature of biofilms.

4.4.2 HSQC-TOCSY

To further analyze the identities of the sharp peaks and the broad humps in HSQC, another NMR experiment was performed. HSQC-TOCSY (Total Correlation Spectroscopy) experiment could observe not only the direct carbon-hydrogen pairs, but also the carbon-hydrogen pairs 2-3 bonds away (indirect pairs) (Figure 4.6B). Theoretically, the spectrum of HSQC would be a subset of the spectrum of HSQC-TOCSY, with the latter experiment being especially useful since we already have

the direct C-H coupling determination from the HSQC spectrum. In other words, all the additional peaks in HSQC-TOCSY can be assigned to be indirect coupling cross-peaks.

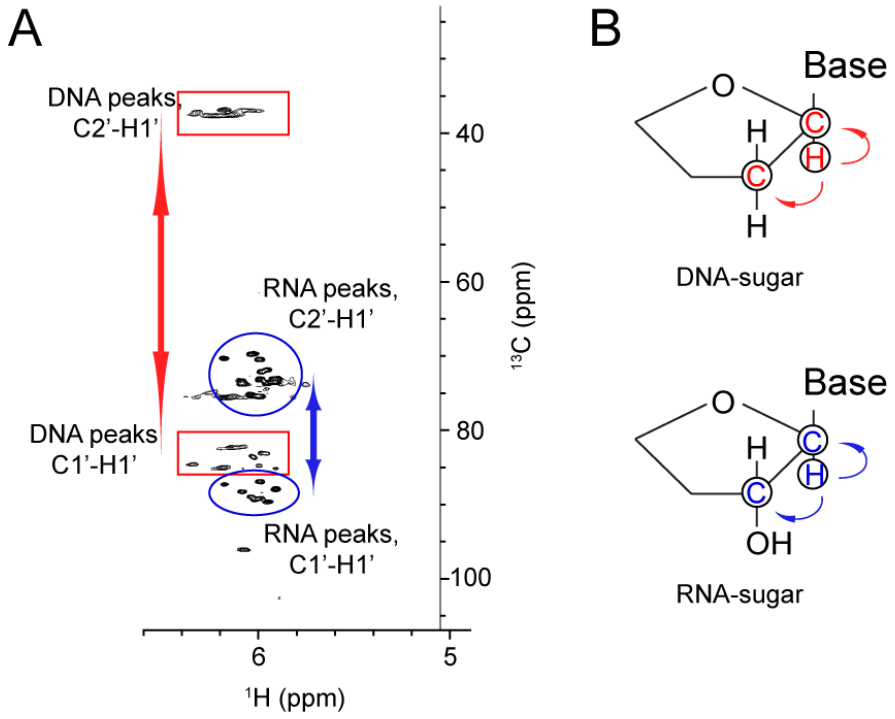


Figure 4.8: A) HSQC-TOCSY spectrum of isolate gel sample at T = 25°C and pH = 12. Red boxes show couplings of C1'-H1' and C2'-H1' of DNA chains in the form of broad peaks. Blue ellipses show similar couplings of C1'-H1' and C2'-H1' of ribo-mononucleotides in the form of 8 sharp peaks. B) Schematic of the observed couplings of DNA (top) and RNA (bottom) sugar rings.

Generally, nucleic acids contain a few carbon-hydrogen pairs in the nitrogenous bases per nucleotide; the exact numbers of which depend on the type of the nitrogenous base. However, the sugar component is uniform, as it includes five carbons in the sugar ring (C1'-C5'), with 7 carbon-coupled hydrogen atoms in the deoxyribose conformation (H1', H2', H2", H3', H4', H5', H5"), or 6 carbon-coupled hydrogen atoms in the ribose conformation (H1', H2', H3', H4', H5', H5"). The

main difference between the two conformations are the chemical shift of the C2'-H2' cross-peak, appearing at ∼40 ppm/2 ppm for deoxyribose sugar (constitutes DNA), in contrast with ∼70 ppm/4 ppm for ribose sugar (constitutes RNA).

Using this contrast, and knowledge of the C1'-H1' region from the HSQC spectrum, the indirect correlation cross-peak C2'-H1' can be determined. The broad humps at the C1'-H1' region (∼85 ppm/6 ppm) were shown to have correlations to another set of broad humps at ∼40 ppm/6 ppm region (Figure 4.8). This result indicated that these broad humps correspond to deoxyribose sugars (DNA), and since the peaks were broad, it is highly likely that the nucleotides were in the chain form with high molecular weight. On the other hand, the 8 sharp peaks at the C1'-H1' region were shown to have correlations to another 8 sharp peaks at ∼70 ppm/6 ppm, indicating the presence of ribose sugars (RNA). In contrast with the previous result, sharp peaks imply that the nucleotides were in the monomeric form with low molecular weight.

Taken together, the two families of nucleic acids at elevated pH were determined to be DNA in the polynucleotide chain form as well as RNA in the mononucleotide form. There are only 4 types of ribonucleotides i.e. adenosine, guanosine, cytosine, and uracil and yet there are 8 distinct sharp peaks observed. Therefore, determination of their identities was of further interest.

4.4.3 ^{13}C-^{1}H HMBC

To identify the 8 sharp peaks, ^{13}C-^{1}H HMBC (Heteronuclear Multiple Bond Correlation) experiments were performed. HMBC experiment provides information on C-H coupling multiple bonds away, even when the bond connection is separated by nitrogen atoms (Figure 4.6C). In other words, it allows observation of the aromatic carbon (e.g. C8 of guanine) coupled with sugar protons (e.g. H1' of guanine)

which were not observable in the HSQC-TOCSY spectrum. The disadvantages of this experiment include weak signals and long acquisition times for obtaining a decent spectrum. Moreover, it can get very crowded due to the appearance of all correlation cross-peaks.

Herein, HMBC spectrum for the isolate with an elevated pH is shown (Figure 4.9). The 8 spin-systems were identified, consisting of 4 pairings of 2 spin-systems. The pairings suggested that out of the 8 spin systems, each of the 4 different nitrogenous bases (A, G, C, U) were represented by 2 spin systems, with a slight difference in the overall chemical structure between them. It is to be noted that in the direct coupling (one-bond) e.g. C1'-H1', the effect of scalar coupling splitting was unavoidable, resulting in split peaks.

Standard samples of ribo-mononucleotides were used to assign each pair of spin systems with their base types. Based on the HMBC spectra of the standards, the types of nitrogenous bases of the 8 spin systems were determined together with the full assignments of aromatic and anomeric protons and their corresponding carbons (Figure 4.10). The relevant chemical shift data and base types are shown below (Table 4.1 column 1-3).

In summary, using HMBC experiments, the information of the type of bases for all the 8 spin-systems was acquired. However, the difference in the two components of a pair of spin systems were not yet understood.

4.4.4 COSY and ^{31}P-^{1}H HETCOR

COSY (Correlation Spectroscopy) experiment provides information of H-H couplings that is separated by 3 bonds (Figure 4.6D). The resulting spectrum of COSY is considered clean since it only shows correlation of direct proton neighbors (e.g.

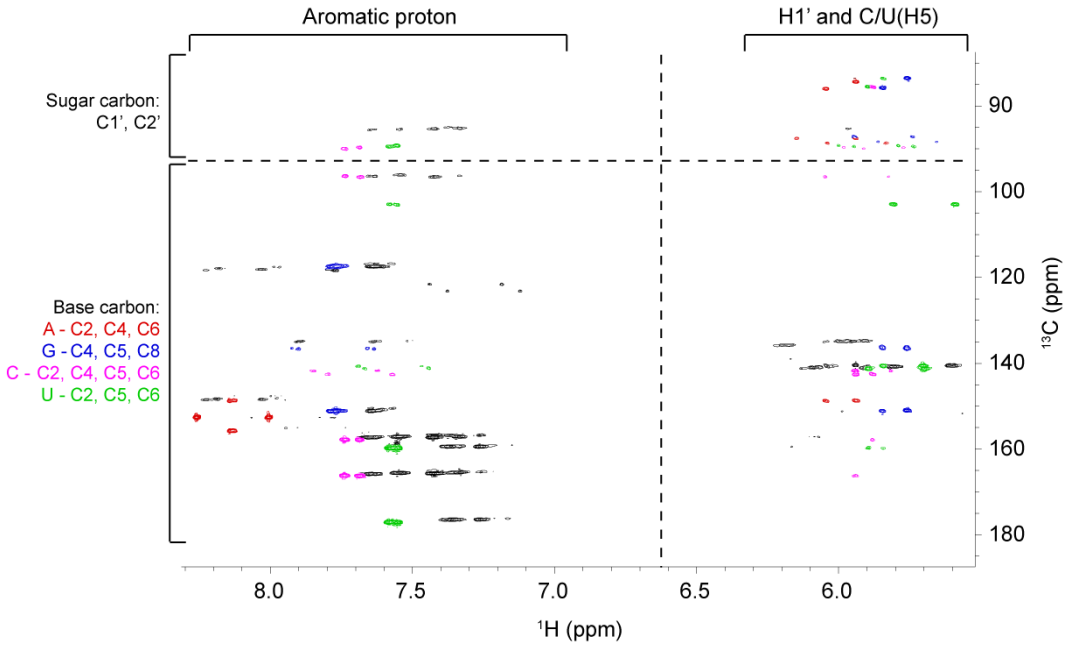


Figure 4.9: HMBC spectrum of the isolate gel sample at T = 25°C and pH = 12. The spectrum is divided into 4 regions, C1'/C2' to aromatic proton region (top-left), C1'/C2' to anomeric proton region (top-right), base carbon to aromatic proton region (bottom-left), and base carbon to anomeric proton region (bottom-right). The peaks that were determined to be ribo-mononucleotide peaks are color coded: red peaks are adenine, blue are guanine, magenta are cytosine, and green are uracil. The splitting of the peaks are a result of one-bond scalar coupling constants. The DNA chains were assumed to be responsible for all other peaks.

H1'-H2' or H2'-H3').

In our context, we used COSY spectrum to determine the chemical shift of H2' and H3' of our 8 spin systems. Starting from the H1' chemical shifts, correlation peaks were traced to H2' and then to H3' (Figure 4.11), resulting in identification of the proton chemical shifts values of H1', H2', and H3' of all 4 pairs of spin systems (Table 4.1 column 4-5).

^{31}P-^{1}H HETCOR (Heteronuclear Correlation) experiment provide information on P-H couplings, with decaying intensity depending on the number of bonds and angle between the two nuclei (Figure 4.6E). In a chain of B-DNA, the phosphate group is always bound to the C5' from one side, and to C3' from the other side.

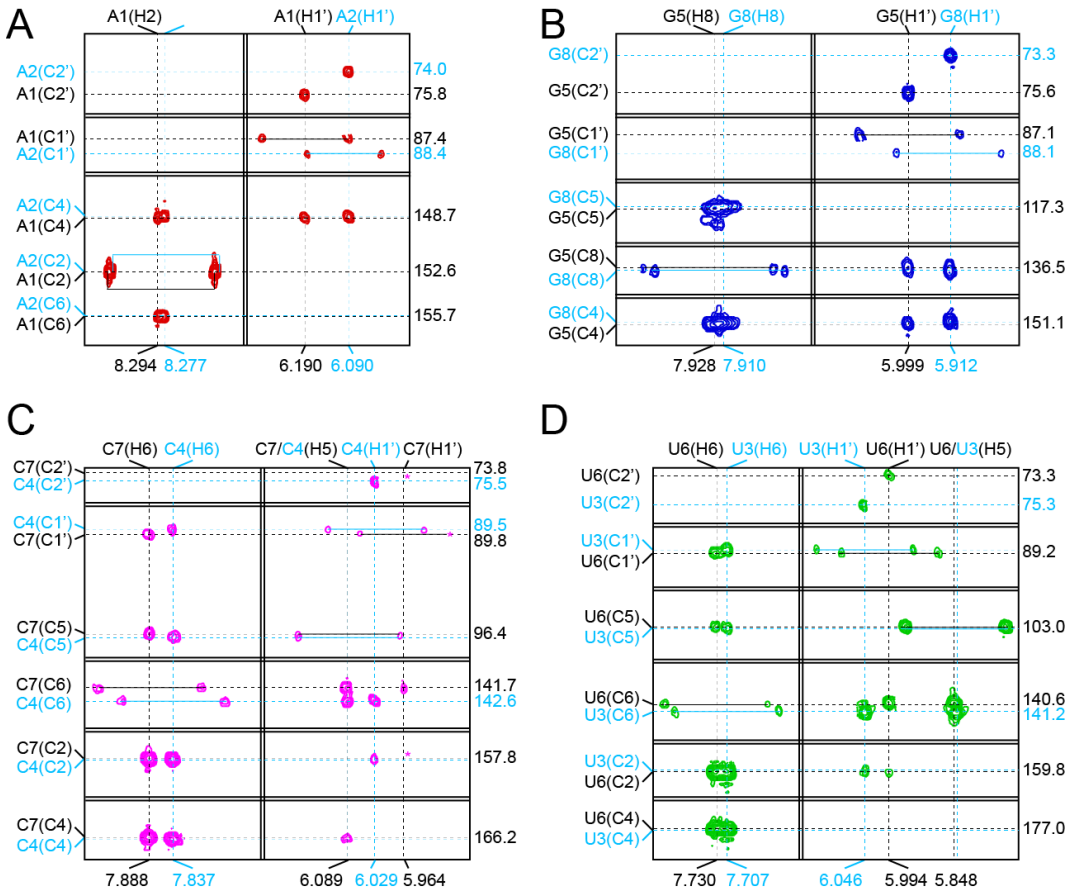


Figure 4.10: Complete assignments of relevant carbons and protons in the HMBC spectrum. Isolated A) adenine monophosphate (AMP) peaks, B) guanosine monophosphate (GMP) peaks, C) cytosine monophosphate peaks (CMP), as well as D) uracil monophosphate (UMP) peaks and their assignments. Two solid lines indicate cuts in the axis, dashed lines indicate the assignment (left and top side) and chemical shift values (right and bottom side), and solid lines indicate J-splitting of direct bonded carbon-proton pairs. Black and cyan color mark two different spin-system for each base type.

However, in the case of mononucleotides, the phosphate group is monoesterified, which means that there is only a single side that is connected towards a sugar group.

Herein, we use ^{31}P-^{1}H HETCOR experiment to determine the connection point between the sugar and phosphate group. We observed that the spectrum was indeed showing broad cross peaks between the diesterified phosphorous atoms and

Table 4.1: Assignments of 8 spin-systems of gel isolate sample in alkali condition.

Label	Type-specific label	H1'	H2'	H3'	P	Final assignment
X1	A1	6.190	**5.054**	4.603	3.509	2'-AMP
X2	A2	6.090	4.844	**4.698**	4.003/4.031	3'-AMP
X3	U3	6.046	**4.664**	4.371	3.661/3.683	2'-UMP
X4	C4	6.029	**4.676**	4.362	3.661/3.683	2'-CMP
X5	G5	5.999	**5.002**	4.578	3.431	2'-GMP
X6	U6	5.994	4.419	**4.526**	3.873/3.907	3'-UMP
X7	C7	5.964	4.397	**4.495**	3.873/3.907	3'-CMP
X8	G8	5.912	4.812	**4.689**	4.003/4.031	3'-GMP

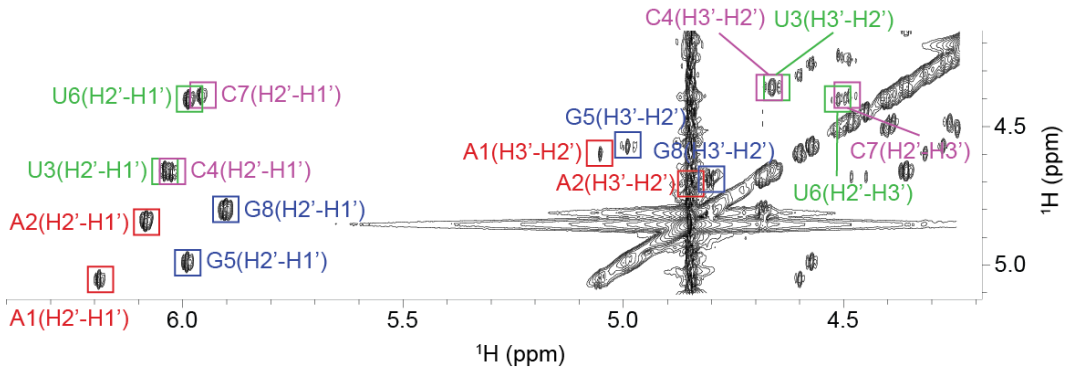


Figure 4.11: COSY spectrum of the isolate gel sample at T = 25°C and pH = 12. Indicated in boxes are intramolecular cross-peaks of each ribo-mononucleotides, i.e. H1'-H2' and H2'-H3'. The boxes are color coded with red mark adenine, blue mark guanine, magenta mark cytosine and green mark uracil.

both the H3' and H5'/H5" atoms (Figure 4.12). The interesting part however, is the cross peaks generated by the monoesterified phosphate group. As expected, 8 sharp cross peaks appeared in the monoesterified phosphorous chemical shift region. Matching these with the chemical shifts of H2' and H3' of the ribo-mononucleotides that were obtained (Table 4.1 column 4-5), 4 of 8 cross peaks were identified to

be a 3'-phosphorylated ribonuclease and the other 4 were determined to be a 2'-phosphorylated ribonuclease.

Summarizing the results, we obtained the exact identities of the 8 spin-systems of ribo-mononucleotides, i.e. 3'-AMP, 3'-GMP, 3'-CMP, 3'-UMP, 2'-AMP, 2'-GMP, 2'-CMP and 2'-UMP as well as the chemical shift data for the phosphorous atoms of each of them (Table 4.1 column 6-7).

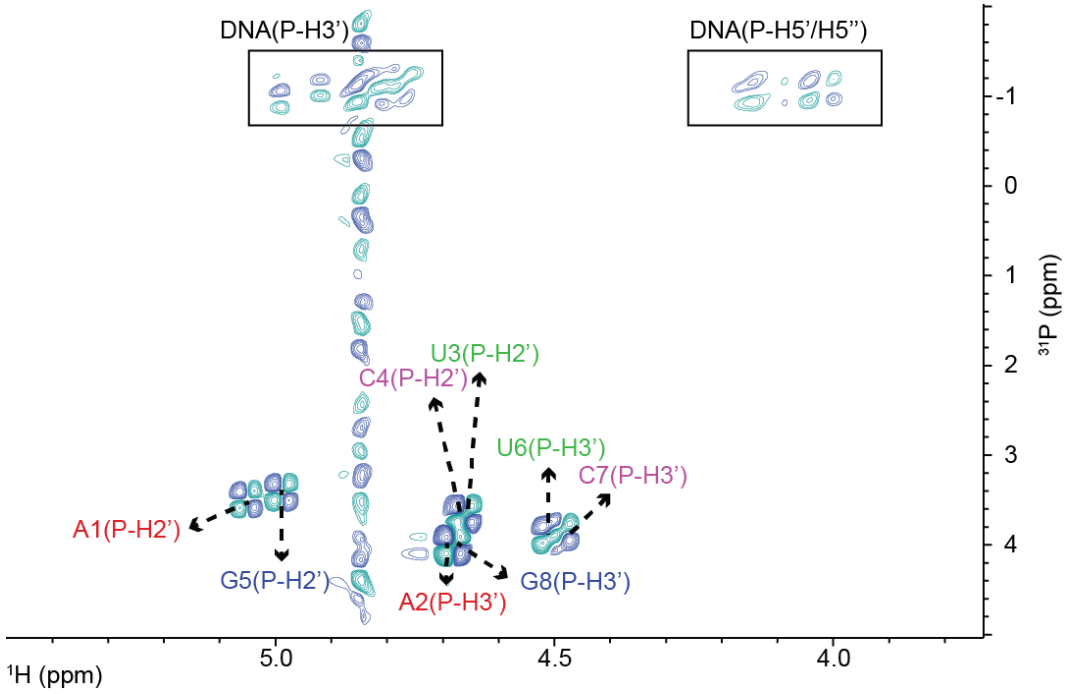


Figure 4.12: ^{31}P-^{1}H HETCOR spectrum of the isolate gel sample at T = 25°C and pH = 12. Dispersive-type cross peaks indicate a correlation between the phosphorous atom of each spin-system to its nearest neighboring proton. Four spin systems suggest a 2'- phosphate position, while the other four suggest a 3'- phosphate position. DNA chains' phosphate group correlations with both H3' and H5'/H5" are shown inside the black boxes.

4.4.5 1D ^{31}P spectrum allows quantitative composition determination

Having all the information on the identities of the spin systems, we were able to determine quantitatively the relative amounts of each mononucleotide. ^{31}P 1D spectrum was chosen as the method due to several reasons: first, each ribo-mononucleotide has exactly one phosphorous nucleus, which allows direct relative quantity determination without any normalization constant; second, the gel-isolate sample has well-resolved phosphorous peaks as compared to the proton counterparts; third, the use of direct detection for ^{31}P (not feasible in the case of ^{13}C due to long relaxation time) which ignores all the magnetization transfer efficiency values between nuclei results in accurate comparison between states.

Hence, the ^{31}P 1D spectrum of the gel-isolate (Figure 4.13) was indicative of the relative abundance of each form of ribo-mononucleotides. Assuming minimal peak shifts, the corresponding spectrum of native biofilm was also analyzed. Some of the peaks were overlapped, resulting in ambiguity for some of the forms.

4.5 RNA-DNA complex network

The existence of both RNA and DNA (section 4.4) in the gel-isolate as well as in the native biofilm (section 4.3) has brought the possibility of RNA-DNA networking. Herein, several direct and indirect evidences for the formation of higher order RNA-DNA structures are laid out.

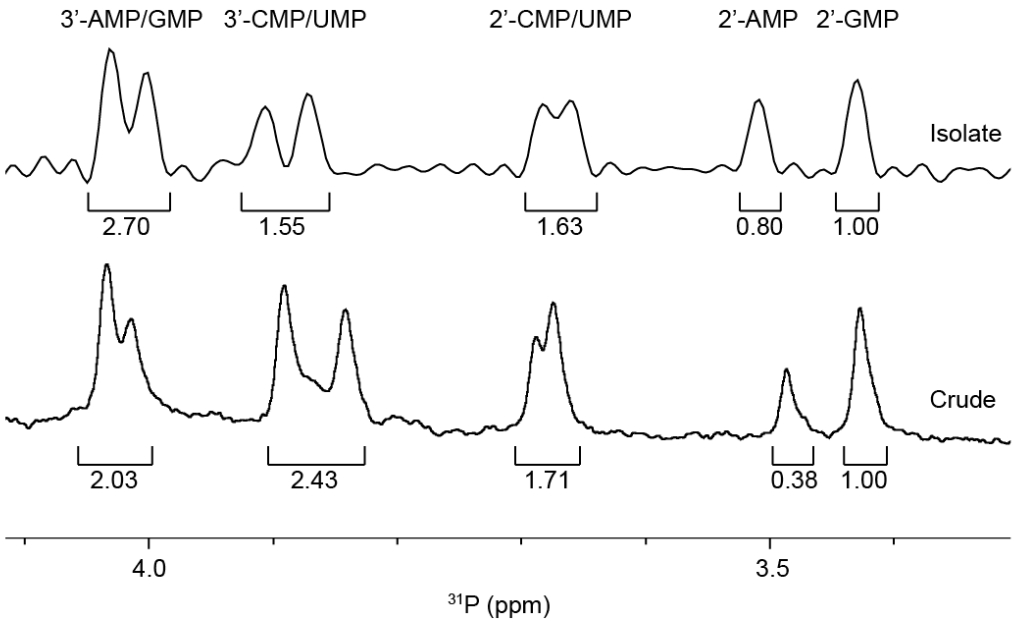


Figure 4.13: ^{31}P NMR spectrum of the isolate gel (top) and the crude biofilm (bottom) at T = 25°C and pH = 12. Both spectra are assigned using the chemical shift data of the phosphorous, with only two peaks were unambiguous, while the other six peaks were ambiguous between the two possibilities indicated above the peaks. The numbers below the peaks are the integration results relative to the right-most peak from each spectrum.

4.5.1 Purine-rich ribonucleotides

As previously described, the sharp ribo-mononucleotides peaks in the ^{31}P spectra only became prevalent at an elevated pH, in contrast with the broad peaks evident at pH 7. Solid state NMR suggests that these ribo-mononucleotides were derived from chain structures that exist at biological pH and assemble with DNA chain into higher-order network structures. To explore this possibility further, the compositions of RNA chain structures were determined by examining the relative abundance of their building blocks i.e. the ribo-mononucleotides.

Quantitative analysis has been performed via peak integration of the ^{31}P 1D NMR spectrum for both the native biofilm sample as well as the isolate sample at pH 12 (Figure 4.13). The relative abundance of each mononucleotide type is shown in Table 4.2. Interestingly, the isolate gel sample was purine-dominated (∼59%),

while the crude biofilm sample was slightly pyrimidine-dominated (∼55%).

Table 4.2: Relative abundance of each mononucleotide types in isolate gel and crude biofilm.

Type	**Name**	**Relative quantity in isolate**		**Relative quantity in crude**	
Purine	2'-AMP	10.42%	58.58%	5.01%	45.26%
	2'-GMP	13.01%		13.27%	
	3'-AMP + 3'-GMP	35.15%		26.98%	
Pyrimidine	2'-CMP + 2'-UMP	21.22%	41.58%	22.71%	54.74%
	3'-CMP + 3'-UMP	20.19%		32.03%	

The difference between the two compositions suggests that some of the RNA within the biofilm was removed in the process of extraction or isolation. One possible interpretation of our results would be that only those RNA chains that had some networking interaction were preserved in the isolation process. A previous study indicates that purine-rich RNA is commonly seen in RNA-DNA duplexes due to their increased thermodynamic stability [239] as RNA purines bind more strongly to their DNA pyrimidine counterpart than vice versa. These facts, when combined with our results, suggest the formation of higher order structures involving both RNA and DNA.

4.5.2 Extracellular DNA and RNA network

The solid state (SS)-NMR results described in section 4.3.2. showed that both RNA and DNA exist in chain forms at neutral pH. Alkali transesterification was also confirmed by the observation of the full conversion of diesterified phosphates to monoesterified phosphates in the RNA standard, when dissolved in 0.1 M NaOH (Figure 4.14). Conversely, RNA diesterified phosphate peaks were preserved in the

RNA standard spectrum following dissolution in EMIM-Ac (ionic liquid) and recovery by perchloric acid. Therefore, the conventional method for *P. aeruginosa* biofilm dissolution in alkali [237] may in fact work by trans-esterifying RNA as a primary structural component and illustrates the importance of the ionic liquid-based extraction protocol described here as a non-destructive method for interrogating biofilm structural polymers.

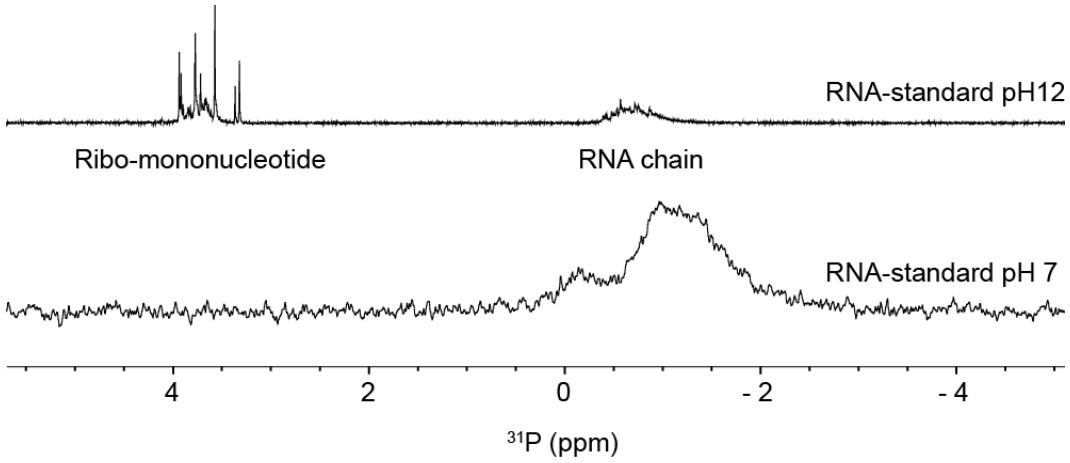


Figure 4.14: ^{31}P NMR spectra of RNA standard in pH 12 (top) and pH 7 (bottom) at T = 25°C. The results show the existence of monoesterified peaks in alkali condition and their absence in neutral condition, suggesting the trans-esterification effect of alkali treatment.

To visualize the organization of the DNA-RNA network, we used nucleic acid-specific stains, i.e. eDNA-specific TOTO1 dye (Figure 4.15A) and the RNA specific SYTO RNA Select dye (Figure 4.15C) [240]. Fibrous structures, which are typical of networked polymer gels [241] were observed in the isolate as well as in the crude biofilm (Figure 4.15B). This is consistent with our observation that the biofilm is readily degraded by DNase I (Figure 4.1) [242]. However, digestion with RNases III, A and H, did not degrade the crude biofilm. This may suggest that the RNase binding was shielded by hairpins in the network, or the presence of non-canonical DNA-RNA interactions [243, 244].

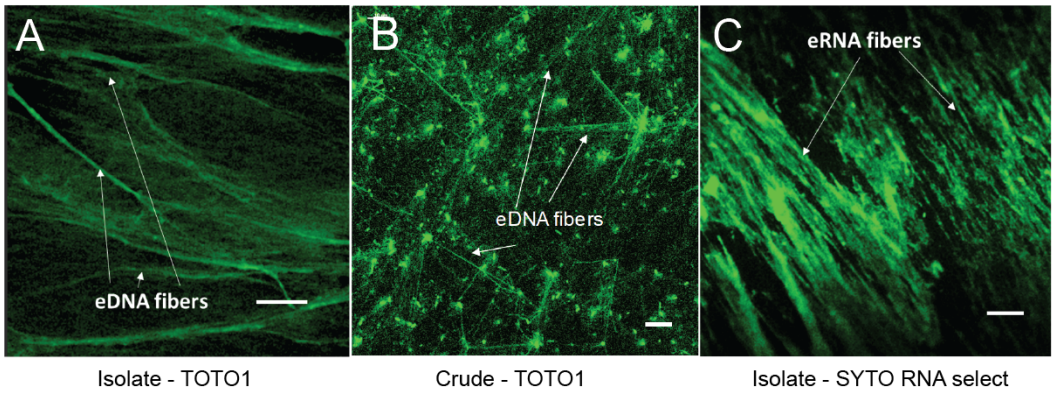


Figure 4.15: Staining results of A) isolated gel and B) crude biofilm with DNA specific dye TOTO1, and C) isolated gel with RNA specific dye SYTO RNA Select at neutral pH. The results indicate the coexistence of eDNA and RNA chains in the isolated sample. Figure and data courtesy of Thomas Seviour *et. al.*

4.5.3 Non-canonical and Watson-Crick base pairs

To elucidate the mechanism of RNA-DNA gelation, we generated a 2D, through-space, ^{15}N-^{1}H HETCOR spectrum of ^{15}N labeled gel sample isolated from the matrix of *P. aeruginosa* biofilms (Figure 4.16A). This experiment was performed with relatively long cross polarization contact times so that correlations between distant ^{1}H and ^{15}N atoms could be observed. This confirmed the complete absence of proteins and supported our conclusion that proteins do not contribute to the RNA-DNA interaction. The HETCOR spectrum showed four signal clusters at δ_H ∼10-14 ppm and δ_N ∼140-160 ppm (Figure 4.16A), which arose from direct N-H couplings of T/U and G nucleobase imino groups. Two of these clusters (δ_H 12-14 ppm) result from imino protons hydrogen-bonded to a nucleobase nitrogen (i.e. N-H — N), while the other two (δ_H 10 -12 ppm) resulted from imino protons hydrogen-bonded to a nucleobase carbonyl oxygen (i.e. N-H — O). Moreover, a strong intermolecular long-range correlation at δ_N 196 ppm arising from G-C Watson-Crick base pairing, i.e. C(N3)-G(H1) was ascertained, as well as a weak intermolecular correlation at δ_N 220 ppm resulting from A-T/U Watson-Crick base

pairing, i.e. A(N1) to U/T(H1).

The absence of long-range correlations between the clusters at δ_H 10 -12 ppm is consistent with non-Watson-Crick pairings for the G and T/U nucleobases. The observation of N-H to O interactions suggests the formations of G-G, T/U-T/U, or G-T/U Hoogsteen base pairs, or G-tetrads. G-quadruplexes are stabilized by the presence of a monovalent cation, with potassium providing the greatest stability. Herein, it was experimentally confirmed that potassium had a similar effect in a static biofilm growth assay. Specifically, *P. aeruginosa* biofilm growth was found to be greater in the presence of potassium than in either sodium or lithium media (Figure 4.16B). Additionally, it was also confirmed that this was a biofilm-specific effect as there was no effect of monovalent cation on planktonic growth. These results present strong evidence that the formation of G-quadruplexes contributes to extracellular DNA-RNA networking.

Non-canonical base pairing and hairpin formations are commonly seen with RNA and cause multivalent intermolecular interactions between RNA strands that lead to a sol-gel transition [245]. The greater structural versatility of RNA could be a key to the ability of the eDNA to form networks, with hairpins, as well as canonical and non-canonical base pair interactions all contributing to the assembly of higher order nucleic acid structures. Additionally, G-rich RNA sequences are more predisposed to forming G-quadruplexes because they do not have a complementary strand and form more stable quadruplexes than DNA [246]. It is therefore probable that the RNA instead of DNA enables RNA-DNA networking and gelation.

4.5.4 Biological relevance of RNA-DNA hybrids

RNA-DNA hybrids perform a number of biologically crucial functions. For example, they contribute to a range of medical conditions including amyotrophic lateral

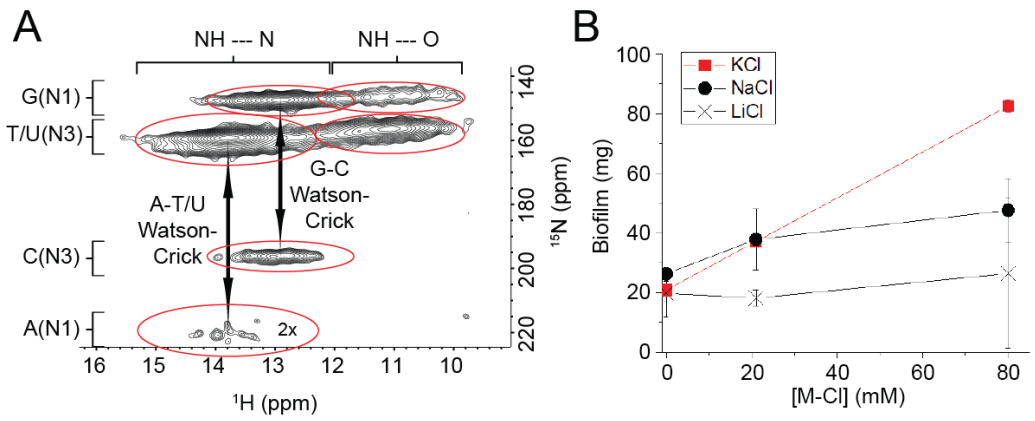


Figure 4.16: A) Solid state 2D ^{1}H-^{15}N through-space heteronuclear correlation (HETCOR) spectrum of gel isolate in H_2O at pH 7, T = 25°C. Red circles indicate distinct cross-peaks: The left pair connected with a double headed arrow indicates T/U-A Watson-Crick (WC) base pairings, where T/U(H3) correlates with T/U(N3) as well as A(N1) (2x decreased threshold due to weak signal); the middle pair connected with a double headed arrow indicates G-C WC base pairing, where G(H1) correlates with G(N1) as well as C(N3); the right sets of peaks indicate T/U and G imino protons hydrogen bonded to carbonyl groups which suggest non-canonical G-G, G-T/U, or T/U-T/U base pairings or tetrads. B) *Pseudomonas aeruginosa* biofilm growth yield at 37°C (5-days) as a function of $[K^+]$, $[Na^+]$ or $[Li^+]$ showing growth dependence on monovalent cation concentration ($K^+ > Na^+ > Li^+$). Solid state NMR experiments were performed with help from Shi Xiangyan; figure and data of biofilm growth courtesy of Thomas Seviour *et. al.*

sclerosis [242] and are thought to possess an increased sensitivity to chromosomal breakage [239]. Furthermore, RNA-DNA hybrids may contribute to pathogenicity during infection. RNA-DNA hybrids expressed by enterohemmorrhagic *Escherichia coli* were reported to activate the host inflammosome during infection and stimulate an innate immune stimulatory response [247]. It is interesting to speculate, therefore, that RNA-DNA hybrids formed by *P. aeruginosa* biofilms could have multiple roles, including being responsible for the gel-properties of biofilms as well as directly influencing infection and associated host responses.

4.6 Conclusions

4.6.1 Summary

The results presented here provide an unprecedented resolution of the biofilm exopolymeric matrix and its key foundational structural components. We have developed a methodology that enabled us to preserve the molecular organization of the foundational polymer in its native state upon extraction and isolation. This enabled us to use NMR spectroscopy to describe intermolecular associations at the atomic level. Solution NMR has shown that both the native biofilm and its isolated state contain RNA as well as extracellular DNA. Moreover, solid-state NMR has provided information on extracellular networking between RNA and DNA which prevents RNA enzymatic degradation.

Our findings that an RNA-DNA gel network can provide the foundation for biofilm matrices in *P. aeruginosa* mark a departure from the prevailing paradigm that biofilm gelation occurs only due to polysaccharides. In addition to *Pseudomonas spp.*, other organisms are known to have eDNA, including *Staphylococcus aureus* [248], *Staphylococcus epidermidis* [249], and *Mycobacterium abscessus* [250] and it is possible that the RNA-DNA foundational gel networks are a broad-scoping phenomenon in biofilms. This study has revealed how nucleic acids, including RNA, are integral to the biophysical and other emergent properties imparted on the biofilm via the extracellular matrix.

4.6.2 Future works

More detailed examination of the native biofilm sample as well as thorough explanation of the roles of nucleic acids and G-quadruplexes in polymer networking can be achieved with the following experiments:

- Solution NMR characterization (e.g. HSQC) of the fully labeled purified biofilm sample dissolved in ionic liquid at normal pH
- Solid-state NMR on the labeled native biofilm sample — G-quadruplex spectroscopic features are expected to be clearly observable upon using a labeled sample
- RNA sequencing in the native and isolate biofilm samples — might provide information on any preferential preservation upon isolation by accurately determining the RNA compositions in both samples.

Chapter 5

Summary and Future Perspectives

5.1 Summary

The dissertation explored nuclear magnetic resonance (NMR) spectroscopy as a biophysical technique to solve variative G-quadruplex-related biological problems. The thesis is divided into two main sections: the first section focuses on the development of NMR methodologies in the dynamic study of G-quadruplexes (Chapter 2); the second section delves into the applications of NMR in resolving multiple high-resolution G-quadruplex structures (Chapter 3) and exploring the role of nucleic acids in biofilm networking (Chapter 4). The followings are the summary for each section:

1. *Development of NMR methodologies in the study of G-quadruplex dynamics*
   Two methodologies of NMR spectroscopy were developed to answer specific problems in the field of molecular interaction and dynamics. The two techniques use the same basic principles of molecular relaxation and exchange rates to measure the kinetics of molecules.

   - *Chapter 2.1:* The first method is based on lineshape analysis [87]. The kinetics of rotation of amino protons were measured by examining the

linewidths of their NOESY cross-peaks with the resolved imino proton peaks. We discovered that the rotation rates of amino groups of guanine bases in G-quadruplexes are highly dependent on a few factors such as: temperature, stacking arrangement, local conformation, ligand binding and additional hydrogen bond formation. Potential applications of the technique include probing of ligand binding sites in G-quadruplex structures and determination of extra hydrogen bond(s) formed involving the G-tetrad amino proton/s.

- *Chapter 2.2:* The second method is based on two-site exchange analysis [88]. In this work, we used the NOESY exchange cross-peaks intensities to identify the behavior of G-quadruplex-ligand complexes in the condition of excess receptors. We showed that there is a measurable value for each G-quadruplex-ligand pair that states the temporal stability or the residence time of the ligand. This variable states an average time of residence of the ligands before they leave their corresponding receptors (i.e. G-quadruplexes). This technique could prove useful in designing chemicals or drugs, where it acts as a preliminary test to determine whether the ligands would have sufficient time being attached to the receptors to be chemically or biologically relevant.

2. *Application of NMR in solving high resolution G-quadruplex structures and determining the component of biological samples*

   - *Chapter 3.1: D4*-cGAMP complex formation implies that G-quadruplexes can be used as potential receptors in cell-signaling. The complex is stable with melting temperature $> 45^{o}$C, with its structural detail revolving around the filling of the vacancy pocket of *D4* by the guanine base of cGAMP.

   - *Chapter 3.2: AT21S* structure featured the formation of a stabilizing

triad and base pair as a part of a fold-back diagonal loop of a G-quadruplex. Preliminary results of a cut-and-paste study were shown using *AT21S* and a known G-quadruplex fold-back structure *Pu24T*. The result showed the possibility of diagonal loop design based on the fold-back property.

- *Chapter 3.3:* $(GGT)_8$ structure contain a new stacking arrangement in G-quadruplexes, where a left-handed G-tetrad is directly stacked on a right-handed G-tetrad. The structure includes 5'-5' stacking interface between the two blocks and a formation of T-capping on the left-handed block, as previously observed in *Z-G4* structure [12]. The discovery of left-right stacking in G-quadruplexes is analogous to the Z-to-B junction in DNA double helix.

- *Chapter 4:* eDNA and RNA were revealed to have significant roles on the networking property of *Pseudomonas aeruginosa* biofilm. NMR spectroscopy was used to determine the existence of eDNA and RNA in the *Pseudomonas aeruginosa* biofilm as well as to quantify the relative amount of mononucleotides that comprise the RNA structures in the isolated sample. We also proved the existence of non-canonical DNA/RNA base pairing in the biofilm by using solid-state NMR spectroscopy. Together with the fact that the mentioned biofilm grows at elevated rates in the presence of potassium, we hypothesized that G-quadruplexes are present among the detected non-canonical structures. The results suggest that nucleic acids in canonical or non-canonical forms, are necessary in maintaining the networking property of biofilms, which would lead to targeting possibility for biofilm dissolution.

Overall, we used NMR spectroscopy to study G-quadruplexes from different angles including dynamics, deterministic, and structural studies. Other biophysical

technique such as UV and CD spectroscopy, surface plasmon resonance (SPR), gel electrophoresis, and X-ray diffraction were also explored, with or without appreciable results.

5.2 Future perspectives

1. *Dynamics assessment*

   NMR spectroscopy is one of the many tools used for molecular structure determination for small to moderate size molecules. Importantly, one of the advantages that NMR possesses over other technique such as X-ray crystallography is the ability to observe the dynamics of molecules such as movements, interactions, and exchange processes. G-quadruplex dynamics was studied in a few different ways including simulation [112–114] and our own lineshape analysis [251]. The assessments of molecular movement in G-quadruplexes as well as nucleic acids in general might lead to the determination of DNA folding mechanism and molecular recognition [110].

2. *Size augmentation*

   One persistent drawback of solution NMR spectroscopy is the limitation of molecular size due to increased tumbling time required for large molecules. Biomolecules with molecular weight > 100 kDa would therefore be not ideal for structural or dynamic analyses using solution NMR spectroscopy [252]. Fortunately, the recent development in solid state NMR spectroscopy, including magic angle spinning (MAS) technique, has resolved not only the size limitation problem, but also the insolubility problem. Future application on solid state NMR would allow us to do a structural analysis on a larger and insoluble G-quadruplex systems such as G-wires [140].

3. *G-quadruplex structural elements*

Various novel G-quadruplex folds and conformations have been elucidated with NMR spectroscopy. These discoveries include unique structural elements of G-quadruplexes such as: bulges formation [160]; duplex loop in edgewise, propeller, and diagonal position [23,161]; fold-back G-quadruplexes [81]; stabilizing triad formations [203, 204]; left-handed backbone progression [12]; inward propeller loop orientation (T-capping); and left-handed right-handed junction in a G-quadruplex structure. The number of unidentified G-quadruplex folds and conformations has kept on increasing with the number of new discoveries. Comprehensive study on G-quadruplex structural elements would undoubtedly be important in the process of understanding the field of G-quadruplex structure topologies.

4. *In vivo NMR*

   The traditional NMR spectroscopy of extracted sample *in vitro* had given us enormous results in terms of structures and interactions of biomolecules. However, the lack of information on the molecule behavior inside cells was an obstacle in the field. Thus, direct examination of biomolecules *in vivo* has been developed with the use of in-cell NMR spectroscopy [253–255]. Progressive work on protein-ligand interaction [256, 257] as well as secondary structure determination of nucleic acid [258, 259] have been performed. We believe that this advancing technology could be used to study G-quadruplex formations and/or biological roles in living cells.

5. *Other biophysical techniques*

   The number of deposited high-resolution structures in the protein data bank (PDB) via NMR spectroscopy has been on a decline since 2007 [260], yet the overall number of structures solved from all techniques is steadily increasing. This statistics clearly tell us that X-ray crystallography is a more effective technique in terms of solving high-resolution structures, in addition

to having a better accuracy. Hence, the future of NMR spectroscopy might not be as a structure determination technique, but as an analytical tool in specific investigation such as dynamics, interactions, and compositional analysis. Looking forward, G-quadruplexes could certainly be explored with the combination of NMR spectroscopy with other techniques such as X-ray and electron microscopy (EM).

Appendix A

Materials and Methods

DNA sample preparation

Unlabeled and site-specific labeled DNA oligonucleotides were purchased form Integrated DNA Technologies (IDT) or were chemically synthesized in-house on an ABI 394 DNA/RNA synthesizer using products from Glen Research and Cambridge Isotope Laboratories. Synthesized oligonucleotides were deprotected with ammonium hydroxide in 55°C for 16 hours, after which they were purified using PolyPak™ cartridges. The purified products were then dialyzed successively against ∼25 mM potassium chloride (KCl) solution for ∼19h and distilled water for ∼1h. Samples were subsequently frozen, lyophilized, and generally dissolved in a buffer containing 20 mM potassium phosphate (KPi) and 70 mM KCl with pH 7 to desired volumes.

Gel purification

Native PAGE (Polyacrylamide gel electrophoresis) was used to separate monomeric species from higher-order structure species in DNA sample. Native mini-gel was casted with 10 lanes, with each lane filled with 20 μL DNA sample. Electrophoresis was performed for ∼40 minutes in constant 120 V potential. The resulting gel was viewed under UV light, followed by selective cutting of the desired location

corresponding to the monomeric purified species. The gel splices were dissolved in water and spun in 220 rpm, 27°C, for ∼20 hours. The supernatant of the sample was extracted, desalted, and dried. The result is dissolved in the desired buffer volume.

Ultraviolet Spectroscopy

Ultraviolet (UV) absorption spectroscopy was used for two main purposes of DNA concentration determination and thermal stability characterization. The measurements were done using JASCO V-650 UV/Vis spectrophotometer.

For determination of concentration, oligonucleotide samples were first diluted to a concentration of ∼2-5 μM in 500 μL, 1-cm path-length cuvettes. They were heated to 95°C for the unfolding of the secondary structures to take place, then the absorption spectra were measured with three accumulations in 220 nm - 320 nm wavelength range. The resulting absorption data were then corrected by subtracting the blank signal from the sample signal, generating the final spectra. The absorption values at 260 nm wavelength were used to determine the oligonucleotides concentration via Beer-Lambert law using nearest-neighbor approximation for the absorption coefficient of the unfolded species [72].

For determination of melting temperature, oligonucleotide samples were prepared similarly in 500 μL, 1-cm path-length cuvettes with ∼2-5 μM concentration. They were heated to 95°C as an unfolded starting position. The samples were then slowly cooled down to 20°C with a rate of 0.2°C/min, while the absorption values at 295 nm were recorded every 1°C. Optional reverse curve could be generated by re-heating the samples to 95°C and perform the same temperature-dependent absorption values. Two baselines were generated using the completely unfolded

(95°C) and completely folded (20°C) states, then they were extrapolated to determine the folded fractions of oligonucleotides during the melting process. Melting temperatures (T_m) were determined by looking at the temperature of the sample when exactly half of the samples are folded [261].

Circular Dichroism

Circular dichroism (CD) spectra were recorded on a JASCO-810 spectropolarimeter using 500 μL, 1-cm path-length cuvettes. Oligonucleotide samples were diluted to ~2-5 μM, and the spectra were recorded in 20°C temperature with a wavelength range of 220 nm - 320 nm. The resulting CD spectra were averaged between 3 accumulations, subtracted with the blank signal, as well as zero-corrected at 320 nm. The absorbance values at 260 nm were also used to determine the concentration of the oligonucleotides, and further to normalize the CD spectra in terms of molar ellipticity.

Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) spectroscopy experiments were performed on Bruker Avance II 600 and 700 MHz, as well as Bruker Avance III 800 MHz spectrometers. The oligonucleotide samples were typically in 0.1 - 2 mM concentrations with 400 μL volume, and 5 mm NMR tubes were used for all experiments. Unless otherwise specified, the temperature of measurements was set to be 25°C. All spectra were zero-corrected with 25 μM DSS.

In method development of amino rotations and residence time, NOESY Quantitative analysis were performed with different G-quadruplex samples as well as in different mixing time (from 10 ms to 500 ms). Spectral assignments and relative abundance prediction of the biofilm samples were done using HSQC, HSQC-TOCSY,

HMBC, and ^{31}P HETCOR experiments among others. The detailed methodology of chapter 2 and chapter 4 are embedded in the corresponding chapters.

For structure determination, the resonance assignments were done via site-specific low-enrichment ^{15}N labelling [13] and through-bond correlations at natural abundance [151]. Complete spectral assignments information was obtained from the following spectra: COSY, TOCSY, NOESY, ^{13}C-^{1}H HMBC, as well as ^{13}C-^{1}H HSQC [262]. Proton-proton distances were deduced from NOESY experiments performed in H_2O (200 ms mixing time) as well as D_2O (100 ms and 300 ms mixing times). All spectral analyses were done using TOPSPIN and NMRFAM-SPARKY [263] software.

Solid state NMR

Solid-state NMR experiments were performed on a 14.1 T Bruker Avance III instruments equipped with a 1.9 mm MAS probe operated in double mode. The typical ^{1}H, ^{15}N and ^{31}P $\pi/2$ pulse lengths were 2.3, 3.7, and 4.5 μs, respectively. 2D dipolar-based ^{15}N-^{1}H heteronulcear-correlation (HETCOR) experiments were conducted at 37 kHz MAS spinning frequency. Variable temperature was regulated at -20°C and the sample temperature was 12°C (calibrated using ethylene glycol). In the ^{15}N-^{1}H HETCOR experiments, the initially excited ^{1}H magnetization was transferred to ^{15}N through a cross polarization step followed by t_1 evolution. Then, the ^{15}N magnetization was flipped to longitudinal axis and 400 ms proton saturation pulses were applied for water suppression. Subsequently, the ^{15}N magnetization was flipped to the transverse plane and transferred to ^{1}H via a second CP step for signal acquisition. Two ^{15}N-^{1}H HETCOR experiments were collected, one with 400 μs and the other with 2 ms contact times applied for both of the CP steps. Low power XiX ^{1}H decoupling ($\sim$10 kHz) was employed during ^{15}N evolution and WALTZ-16 decoupling (10 kHz) was implemented on ^{15}N

channel during ^{1}H acquisition. 1D ^{31}P experiments were performed with 15 kHz MAS spinning frequency and a sample temperature of 27 °C. 75 kHz SPINAL64 ^{1}H decoupling was applied during ^{31}P acquisition time. All chemical shifts were indirectly referenced using adamantane as a secondary standard (downfield peak is at 40.48 ppm, DSS scale).

Structure Calculation

Structure computations were performed using the XPLOR-NIH program. The basic procedures are the following two steps: (i) distance geometry simulated annealing and (ii) distance-restrained molecular dynamics refinement as previously described [264]. The restraints imposed in the simulations include inter-proton distance restraints, dihedral angle restraints, hydrogen-bond restraints, planarity restraints, as well as repulsive restraints. The 10-lowest energy out of 100 generated structures in vacuo were used as the structure models, from which all the NMR structure calculation statistics were generated.

Inter-proton distances were deduced from NOESY spectra in H_2O as well as D_2O solvents in multiple mixing times. In details, H_2O NOE peaks were classified into 3 classes from strong to weak, with the distance restraints as follows: $3.6\pm1.0\mathring{A}, 4.8\pm 1.4\mathring{A}$, and $5.5 \pm 1.7\mathring{A}$, while D_2O NOE peaks divided into 4 classes as follows: $2.7\pm0.8\mathring{A}, 3.8\pm0.9\mathring{A}, 4.6\pm1.2\mathring{A}$, and $5.5\pm1.7\mathring{A}$. The dihedral restraints were based on the cross-peak intensity of guanine anomeric protons (H1') to their neighboring aromatic protons (H8), strong cross-peaks lead to $60^o \pm 70^o$, while weaker cross-peaks lead to $240^o\pm70^o$ dihedral angle (χ). Hydrogen bond restraints were imposed based on the knowledge of G-tetrad structure, with 4 hydrogen bond constraints were applied per guanine base inside G-tetrad. Specifically, inter-residue H21-N7 as well as H1-O6 were constrained in $2.0 \pm 0.2\mathring{A}$, and N2-N7 as well as N1-O6 at $2.9 \pm 0.3\mathring{A}$. The planarity restraints were used in agreement with previously

observed G-quadruplex X-ray structures. Finally, the repulsive restraints were introduced to a few pairs of well-resolved protons which did not give any NOE cross-peak. Calculated structures were aligned and displayed using PyMOL graphical program [265].

Appendix B

Collaborative work with SCELSE

Bacterial strains and media

Pseudomonas aeruginosa PDO300(Δpel) and PDO300(Δpsl) mutant strains were gratefully received from Professor Bernd H. A. Rehm, Institute of Molecular Biosciences, Massey University, New Zealand (10). *P. protogens Pf-5 and P. putida* ATCC BAA-477 and S12 strain were used.

P. aeruginosa strains PAO1 and PDO300, and their isogenic mutants, were grown on Luria-Bertani (LB) agar (5 g litre^{-1} NaCl, 5 g litre^{-1} yeast extract, 10 g litre^{-1} tryptone, 15 g litre^{-1} agar) or *Pseudomonas* isolation agar (PIA) (20 g litre^{-1} peptone, 10 g litre^{-1} K_2SO_4, 1.4 g litre^{-1} MgCl2, Irgasan^{TM} 25 mg litre^{-1}, agar 13.6 g litre^{-1}) at 37°C for 16 hours. *P. protogens* and *P. putida* were grown on LB agar at 30°C for 16 hours.

Overnight pre-cultures of *P. aeruginosa* strains were grown in either 10 mL Luria-Bertani broth (LB) (5 g litre^{-1} NaCl, 5 g litre^{-1} yeast extract, 10 g litre^{-1} tryptone) or 10 mL of M9 minimum media (MP Biomedicals) (according to requirements of experiment) at 37°C under shaking conditions (200 rpm). For solid-state NMR experiments performed on the NA gel isolate, ^{15}N labeled NH_4Cl-supplemented M9

was used for biofilm growth.

Briefly, M9 minimum media consisted of 9.552 g litre^{-1} $Na_2HPO_4.2H_2O$, 4.41 g litre^{-1} KH_2PO_4, 1.71 g litre^{-1} NaCl, 1 g litre^{-1} NH_4Cl, 0.24 g litre^{-1} $MgSO_4$, 0.011 g litre^{-1} $CaCl_2$, 2g litre^{-1} Casamino acid, and 0.4 g litre^{-1} glucose.

Effects of monovalent cations (K^+, Na^+ and Li^+) on eDNA secondary structure were examined by cultivating 5 days old PAO1 WT biofilm in modified M9 minimum media in the conditions described above. Modified M9 minimum media comprised of standard M9 media was used with NaCl salt replaced with either KCl, LiCl or NaCl at either of three different concentrations (0, 0.02 mM and 0.08 mM). $Na_2HPO_4.2H_2O$ and KH_2PO_4 concentration were reduced to 8.5 g litre^{-1} and 1 g litre^{-1} respectively with 2 g litre^{-1} of $NaH_2PO_4.H_2O$ salt added. $MgSO_4$ was raised from 0.002 mM to 0.004 mM.

Cultivation of planktonic cells and static polymicrobial biofilms

To encourage planktonic cell growth, 10 mL aliquots of *P. aeruginosa* wild-type PAO1 pre-culture (OD_{600} 2.40) were diluted 50 times with fresh LB broth in 1 L Erlenmeyer flask and incubated for 24 h at 37°C with shaking (200 rpm). Overnight cultured planktonic cells were concentrated by centrifugation at 10,000 x g for 10 minutes.

For biofilm cultures, 10 mL of respective overnight pre-cultures were incubated overnight, diluted 50 times with fresh LB broth in 2 L Erlenmeyer flasks and incubated for 5 days at 37°C under static conditions. Biofilms were concentrated by centrifugation at 10,000 x g for 15 minutes. Centrifugation stratified the biofilm mixture into three layers. The bottom two layers were collected and lyophilized.

Enzymatic digestions

20 mg 5-day old freeze-dried concentrated biofilm were resuspended in 1 mL of either i) RNase buffer (50 mM Tris HCl, 10 mM EDTA, pH 8) with 0.2 mg RNaseA from bovine pancreas (Sigma Aldrich R 6513), ii) storage buffer (10 mM NaCl, 10 mM Tris-HCl) with 0.1 mg Pronase E from *Streptomyces grisens* (Sigma Aldrich) with 0.5% (v/v), iii) 1X RNase H reaction buffer (20 mM Tris-HCl pH 7.8, 40 mM KCl, 8 mM $MgCl_2$, 1 mM DTT) with 0.4 mg RNase H (Thermofisher Scientific), iv) 1X RNase III reaction buffer (500 mM NaCl, 100 mM Tris pH 7.9, 100 mM $MgCl_2$, 10 mM DTT) with 0.6 mg RNase II (Thermofisher Scientific) and v) DNase I buffer (100 mM Tris (pH 7.5), 25mM $MgCl_2$ and 5mM $CaCl_2$) with 0.2 mg DNAse I from bovine pancreas (Sigma Aldrich). All digestions were performed with shaking at 200 rpm at 37°C for 16 hours.

Normal force measurement

40 mg/mL solutions of lyophilised samples (biofilm and planktonic cells) were prepared in 1-ethyl-3-methylimidazolium acetate and incubated in 55°C for 2 hours. A Haake Mars 3 (Thermo Scientific) stress-controlled rotational rheometer with Peltier controlled element at to 25°C, was used for rheological measurements. 35-mm diameter parallel plate geometry was used with smooth titanium plates. Prior to measurement the gap error was zeroed at 4 N and gap error calculated as previously described [266–268]. 100 μL of sample was deposited on the plates. The plates were closed to 100 μm, the sample trimmed and the sample allowed to sit for 5 minutes prior to measurement. All measurements with Normal force (F_N) less than the resolution of the rheometer (i.e. < 0.1 N) were set to 0 before calculation of N_1. Corrections were made to viscosity, F_N and $N_1 - N_2$ to account for non-Newtonian fluid behavior in parallel plates and for inertia as described by Davies and Stokes [267].

Oscillatory measurements were carried out on the ARG2 (TA Instruments) with Peltier controlled element at 25°C on smooth plates with a 200 μm gap. Controlled strain 0.1 across a frequency range of 0.6 to 60 rad/s.

Extracellular polymeric substances (EPS) extraction

Lyophilised biofilms were dissolved in ionic liquid mixture (40% (v/v) 1-ethyl-3-methylimidazolium acetate (EMIM Ac): 60% (v/v) dimethyl acetamide (DMAc)) at 55°C for 16 hours. The solution was centrifuged (10,000 x g) to remove any undissolved material. Perchloric acid (70%) was added (0.05% v/v) to the viscosified centrate (on ice). After 15 minutes incubation, the solution was centrifuged at 10,000 x g at 4°C for 15 minutes and the pellet recovered. This was repeated on the centrate 2 to 4 times until the solution was not viscous. The precipitate was dialysed against Milli-Q water for 2 days at 4°C (SnakeSkin^{TM} Dialysis Tubing, 3.5K MWCO, 22 mm) and the retentate was lyophilized (FreeZone Plus 4.5 Liter Cascade Benchtop Freeze Dry System).

eDNA purification

20 mg of lyophilized perchloric precipitate were dissolved in 1 mL of 40% (v/v) EMIM-Ac in 60% DMAc (v/v) (55°C, 16 h). Chromatographic separation was achieved in a Shimadzu system comprising DGU-20A 3r Prominence Degasser and LC-20AD solvent Delivery Unit, fitted with two Agilent PLgel 10 μm column of 105$\mathring{A}$ pore size for separation across the MW range 200 kDa to 2000 kDa. The eluent flow rate was 3.0 mL min^{-1} and the injection volume 1 mL. The fractions with molecular weight range of 2000-800 kDa and 800-200 kDa were pooled and dialysed for 48 hours against double distilled water to induce gelation. The gel was collected manually from the dialysis tubing.

DNA agarose gel

1% of agarose gel solution was prepared from Viviantis LE grade agarose using 1x TAE buffer (40 mM Tris, 20 mM Acetate and 1 mM EDTA, pH 8.6). Gels were run horizontally. After electrophoresis, the gel was stained for 0.5 hours with ethidium bromide and visualized under UV [269].

SDS page gel

SDS-PAGE was performed on 1-mm thick 12% Bis-Tris Protein gels (Thermo Fisher Scientific) in XCell SureLock^{TM} Mini-Cell Electrophoresis System. The NuPAGE MES SDS Running Buffer 20x was diluted to 1x when running the gel. The gel was run at 200 V for 0.5 hours and stained with Coomassie Blue G250.

Genomic DNA extraction

Genomic DNA was extracted from planktonic cells or biofilm cells using FastDNA SPIN Kit for soil (MP Biomedicals, USA) as per the standard protocol.

Sequencing

GPC fractionate material was resuspended in 500 μL of 1x Protease K solution (10x Protease K solution: 10mM Tris HCl, 1% SDS and 10 mM EDTA, pH 8 buffer (10x proteas K buffer containing 500 mM Tris-HCl, 10%SDS, 10mM $CaCl_2$) and 10 μL of Protease K (20 mg/mL, ThermoFisher Scientific) was added and the mixture incubated at 56°C for 2-16 hours, after which DNA was extracted as previously described in the Phenol-chloroform method [270].

Samples before sequencing were further purified to remove any remaining protein and RNA by RNase and Proteinase K treatment. The DNA was then isolated

using Phenol-Chloroform precipitation as described above. The DNA precipitate was dissolved in TE buffer and the purity confirmed by 260/280 value in Nanodrop (acceptable range value: 1.8-2.0) and Qubit 2.0 fluorometer. Three replicates were used for each DNA sequence analysis. Library was produced using Illumina DNA sample preparation kit. The libraries were sequenced using Illumina MiSeq platform (Illumina, San Diego, Ca) with paired-end protocol to read lengths of 600 nt generating a total of 1,614,106 and 1,848,846 paired end reads. Raw reads were quality filtered (reads remaining after trimming: PPG1-1549104, PBLC1-1666280) and aligned to the *Pseudomonas aeruginosa* PAO1 (AE004091) genome using CLC Genomics Workbench 9.0 (CLC bio, Cambridge, MA).

Staining and microscopy

Microscopic imaging was conducted on a confocal microscope Zeiss LSM 780 with a x63 objective. eDNA in crude biofilm and isolated gel were stained with BacLight Live/Dead viability stain (Molecular Probes) for 15 minutes and imaged on confocal microscope. Live/Dead stain was prepared by adding 6 μL of 1:1 SYTO 9 (3.34 mM in DMSO) and propidium iodide (20 mM in DMSO) mixture to 1 mL of 0.9% (w/v) filtered NaCl solution.

Extracellular RNA in the gel isolate were stained using SYTO RNA SELECT green fluorescent cell stain (5 mM solution in DMSO). 5 μM stain solution was prepared from 1 μL of stock in 1X PBS solution (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4). The gel isolate was labelled with 5 μM stain solution, kept at 37^{o}C for 20 minutes and then transferred to glass side for imaging.

eDNA staining was achieved by depositing PAO1 biofilm or NA gel isolate on a glass slide, air-drying overnight and incubating with 2 μM TOTO-1 iodide (1 mM solution in DMSO, Thermofischer) for 15 minutes.

Bibliography

[1] O. T. Avery, C. M. Macleod, and M. McCarty. Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *J Exp Med*, 79(2):137–58, 1944.

[2] J. D. Watson and F. H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–8, 1953.

[3] M. Gellert, M. N. Lipsett, and D. R. Davies. Helix formation by guanylic acid. *Proc Natl Acad Sci U S A*, 48:2013–8, 1962.

[4] D. Sen and W. Gilbert. Formation of parallel four-stranded complexes by guanine-rich motifs in dna and its implications for meiosis. *Nature*, 334(6180):364–6, 1988.

[5] W. I. Sundquist and A. Klug. Telomeric dna dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, 342(6251):825–9, 1989.

[6] Nicholas V Hud, Flint W Smith, Frank AL Anet, and Juli Feigon. The selectivity for k+ versus na+ in dna quadruplexes is dominated by relative free

energies of hydration: a thermodynamic analysis by 1h nmr. *Biochemistry*, 35(48):15383–15390, 1996.

[7] C. C. Hardin, E. Henderson, T. Watson, and J. K. Prosser. Monovalent cation induced structural transitions in telomeric dnas: G-dna folding intermediates. *Biochemistry*, 30(18):4460–72, 1991.

[8] C. C. Hardin, T. Watson, M. Corregan, and C. Bailey. Cation-dependent transition between the quadruplex and watson-crick hairpin forms of d(cgcg3gcg). *Biochemistry*, 31(3):833–41, 1992.

[9] J. T. Davis. G-quartets 40 years later: from 5'-gmp to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl*, 43(6):668–98, 2004.

[10] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, and S. Neidle. Quadruplex dna: sequence, topology and structure. *Nucleic Acids Res*, 34(19):5402–15, 2006.

[11] A. T. Phan, V. Kuryavyi, and D. J. Patel. Dna architecture: from g to z. *Curr Opin Struct Biol*, 16(3):288–98, 2006.

[12] W. J. Chung, B. Heddi, E. Schmitt, K. W. Lim, Y. Mechulam, and A. T. Phan. Structure of a left-handed dna g-quadruplex. *Proc Natl Acad Sci U S A*, 112(9):2729–33, 2015.

[13] A. T. Phan and J. L. Mergny. Human telomeric dna: G-quadruplex, i-motif and watson-crick double helix. *Nucleic Acids Res*, 30(21):4618–25, 2002.

[14] B. Heddi and A. T. Phan. Structure of human telomeric dna in crowded solution. *J Am Chem Soc*, 133(25):9824–33, 2011.

[15] K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, and A. T. Phan. Structure of the human telomere in k+ solution: a stable

basket-type g-quadruplex with only two g-tetrad layers. *J Am Chem Soc*, 131(12):4301–9, 2009.

[16] A. T. Phan, V. Kuryavyi, K. N. Luu, and D. J. Patel. Structure of two intramolecular g-quadruplexes formed by natural human telomere sequences in k+ solution. *Nucleic Acids Res*, 35(19):6517–25, 2007.

[17] A. T. Phan, K. N. Luu, and D. J. Patel. Different loop arrangements of intramolecular human telomeric (3+1) g-quadruplexes in k+ solution. *Nucleic Acids Res*, 34(19):5715–9, 2006.

[18] J. L. Huppert and S. Balasubramanian. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*, 33(9):2908–16, 2005.

[19] D. Schiavone, G. Guilbaud, P. Murat, C. Papadopoulou, P. Sarkies, M. N. Prioleau, S. Balasubramanian, and J. E. Sale. Determinants of g quadruplex-induced epigenetic instability in rev1-deficient cells. *EMBO J*, 33(21):2507–20, 2014.

[20] D. Rhodes and H. J. Lipps. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res*, 43(18):8627–37, 2015.

[21] D. J. Yue, K. W. Lim, and A. T. Phan. Formation of (3+1) g-quadruplexes with a long loop by human telomeric dna spanning five or more repeats. *J Am Chem Soc*, 133(30):11462–5, 2011.

[22] H. Martadinata and A. T. Phan. Formation of a stacked dimeric g-quadruplex containing bulges by the 5'-terminal region of human telomerase rna (hterc). *Biochemistry*, 53(10):1595–600, 2014.

[23] K. W. Lim, P. Jenjaroenpun, Z. J. Low, Z. J. Khong, Y. S. Ng, V. A. Kuznetsov, and A. T. Phan. Duplex stem-loop-containing quadruplex motifs

in the human genome: a combined genomic and structural study. *Nucleic Acids Res*, 43(11):5630–46, 2015.

[24] T. Q. N. Nguyen, K. W. Lim, and A. T. Phan. A dual-specific targeting approach based on the simultaneous recognition of duplex and quadruplex motifs. *Sci Rep*, 7(1):11969, 2017.

[25] S. L. Konig, A. C. Evans, and J. L. Huppert. Seven essential questions on g-quadruplexes. *Biomol Concepts*, 1(2):197–213, 2010.

[26] S. Frees, C. Menendez, M. Crum, and P. S. Bagga. Qgrs-conserve: a computational method for discovering evolutionarily conserved g-quadruplex motifs. *Hum Genomics*, 8:8, 2014.

[27] N. Beaume, R. Pathak, V. K. Yadav, S. Kota, H. S. Misra, H. K. Gautam, and S. Chowdhury. Genome-wide study predicts promoter-g4 dna motifs regulate selective functions in bacteria: radioresistance of d. radiodurans involves g4 dna-mediated regulation. *Nucleic Acids Res*, 41(1):76–89, 2013.

[28] W. I. Sundquist and S. Heaphy. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic rna. *Proc Natl Acad Sci U S A*, 90(8):3393–7, 1993.

[29] J. Norseen, F. B. Johnson, and P. M. Lieberman. Role for g-quadruplex rna binding by epstein-barr virus nuclear antigen 1 in dna replication and metaphase chromosome attachment. *J Virol*, 83(20):10336–46, 2009.

[30] B. Tuesuwan, J. T. Kern, P. W. Thomas, M. Rodriguez, J. Li, W. M. David, and S. M. Kerwin. Simian virus 40 large t-antigen g-quadruplex dna helicase inhibition by g-quadruplex dna-interactive agents. *Biochemistry*, 47(7):1896–909, 2008.

[31] N. Maizels and L. T. Gray. The g4 genome. *PLoS Genet*, 9(4):e1003468, 2013.

[32] C. Cayrou, P. Coulombe, A. Puy, S. Rialle, N. Kaplan, E. Segal, and M. Mechali. New insights into replication origin characteristics in metazoans. *Cell Cycle*, 11(4):658–67, 2012.

[33] C. Cayrou, P. Coulombe, A. Vigneron, S. Stanojcic, O. Ganier, I. Peiffer, E. Rivals, A. Puy, S. Laurent-Chabalier, R. Desprat, and M. Mechali. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res*, 21(9):1438–49, 2011.

[34] E. Besnard, A. Babled, L. Lapasset, O. Milhavet, H. Parrinello, C. Dantec, J. M. Marin, and J. M. Lemaitre. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with g-quadruplex consensus motifs. *Nat Struct Mol Biol*, 19(8):837–44, 2012.

[35] A. Bugaut and S. Balasubramanian. 5'-utr rna g-quadruplexes: translation regulation and targeting. *Nucleic Acids Res*, 40(11):4727–41, 2012.

[36] J. D. Beaudoin and J. P. Perreault. Exploring mrna 3'-utr g-quadruplexes: evidence of roles in both alternative polyadenylation and mrna shortening. *Nucleic Acids Res*, 41(11):5898–911, 2013.

[37] R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff, and J. R. Wu. A highly conserved repetitive dna sequence, (ttaggg)n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A*, 85(18):6622–6, 1988.

[38] J. Meyne, R. L. Ratliff, and R. K. Moyzis. Conservation of the human telomere sequence (ttaggg)n among vertebrates. *Proc Natl Acad Sci U S A*, 86(18):7049–53, 1989.

[39] D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S. Neidle, and L. H. Hurley. Inhibition of human telomerase by a g-quadruplex-interactive compound. *J Med Chem*, 40(14):2113–6, 1997.

[40] M. Read, R. J. Harrison, B. Romagnoli, F. A. Tanious, S. H. Gowan, A. P. Reszka, W. D. Wilson, L. R. Kelland, and S. Neidle. Structure-based design of selective and potent g quadruplex-mediated telomerase inhibitors. *Proc Natl Acad Sci U S A*, 98(9):4844–9, 2001.

[41] A. M. Burger, F. Dai, C. M. Schultes, A. P. Reszka, M. J. Moore, J. A. Double, and S. Neidle. The g-quadruplex-interactive molecule braco-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res*, 65(4):1489–96, 2005.

[42] A. M. Zahler, J. R. Williamson, T. R. Cech, and D. M. Prescott. Inhibition of telomerase by g-quartet dna structures. *Nature*, 350(6320):718–20, 1991.

[43] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps, and A. Pluckthun. In vitro generated antibodies specific for telomeric guanine-quadruplex dna react with stylonychia lemnae macronuclei. *Proc Natl Acad Sci U S A*, 98(15):8572–7, 2001.

[44] K. Paeschke, T. Simonsson, J. Postberg, D. Rhodes, and H. J. Lipps. Telomere end-binding proteins control the formation of g-quadruplex dna structures in vivo. *Nat Struct Mol Biol*, 12(10):847–54, 2005.

[45] C. Granotier, G. Pennarun, L. Riou, F. Hoffschir, L. R. Gauthier, A. De Cian, D. Gomez, E. Mandine, J. F. Riou, J. L. Mergny, P. Mailliet, B. Dutrillaux,

and F. D. Boussin. Preferential binding of a g-quadruplex ligand to human chromosome ends. *Nucleic Acids Res*, 33(13):4182–90, 2005.

[46] G. Biffi, D. Tannahill, J. McCafferty, and S. Balasubramanian. Quantitative visualization of dna g-quadruplex structures in human cells. *Nat Chem*, 5(3):182–6, 2013.

[47] A. Henderson, Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, Jr. Brosh, R. M., D. Sen, and P. M. Lansdorp. Detection of g-quadruplex dna in mammalian cells. *Nucleic Acids Res*, 42(2):860–9, 2014.

[48] J. L. Huppert, A. Bugaut, S. Kumari, and S. Balasubramanian. G-quadruplexes: the beginning and end of utrs. *Nucleic Acids Res*, 36(19):6260–8, 2008.

[49] H. Fernando, S. Sewitz, J. Darot, S. Tavare, J. L. Huppert, and S. Balasubramanian. Genome-wide analysis of a g-quadruplex-specific single-chain antibody that regulates gene expression. *Nucleic Acids Res*, 37(20):6716–22, 2009.

[50] L. T. Gray, A. C. Vallur, J. Eddy, and N. Maizels. G quadruplexes are genomewide targets of transcriptional helicases xpb and xpd. *Nat Chem Biol*, 10(4):313–8, 2014.

[51] T. Simonsson, P. Pecinka, and M. Kubista. Dna tetraplex formation in the control region of c-myc. *Nucleic Acids Res*, 26(5):1167–72, 1998.

[52] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, and L. H. Hurley. Direct evidence for a g-quadruplex in a promoter region and its targeting with a small molecule to repress c-myc transcription. *Proc Natl Acad Sci U S A*, 99(18):11593–8, 2002.

[53] A. Verma, K. Halder, R. Halder, V. K. Yadav, P. Rawal, R. K. Thakur, F. Mohd, A. Sharma, and S. Chowdhury. Genome-wide computational and expression analyses reveal g-quadruplex dna motifs as conserved cis-regulatory elements in human and related species. *J Med Chem*, 51(18):5641–9, 2008.

[54] S. G. Hershman, Q. Chen, J. Y. Lee, M. L. Kozak, P. Yue, L. S. Wang, and F. B. Johnson. Genomic distribution and functional analyses of potential g-quadruplex-forming sequences in saccharomyces cerevisiae. *Nucleic Acids Res*, 36(1):144–56, 2008.

[55] Wei-Chun Huang, Ting-Yuan Tseng, Ying-Ting Chen, Cheng-Chung Chang, Zi-Fu Wang, Chiung-Lin Wang, Tsu-Ning Hsu, Pei-Tzu Li, Chin-Tin Chen, Jing-Jer Lin, et al. Direct evidence of mitochondrial g-quadruplex dna by using fluorescent anti-cancer agents. *Nucleic acids research*, 43(21):10102–10113, 2015.

[56] Alexander Henderson, Yuliang Wu, Yu Chuan Huang, Elizabeth A Chavez, Jesse Platt, F Brad Johnson, Robert M Brosh, Dipankar Sen, and Peter M Lansdorp. Detection of g-quadruplex dna in mammalian cells. *Nucleic acids research*, 42(2):860–869, 2013.

[57] Junjie U Guo and David P Bartel. Rna g-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, 353(6306):aaf5371, 2016.

[58] M. Subramanian, F. Rage, R. Tabet, E. Flatter, J. L. Mandel, and H. Moine. G-quadruplex rna structure as a signal for neurite mrna targeting. *EMBO Rep*, 12(7):697–704, 2011.

[59] Y. Wu and Jr. Brosh, R. M. G-quadruplex nucleic acids and human disease. *FEBS J*, 277(17):3470–88, 2010.

[60] T. A. Brooks and L. H. Hurley. The role of supercoiling in transcriptional control of myc and its importance in molecular therapeutics. *Nat Rev Cancer*, 9(12):849–61, 2009.

[61] S. Balasubramanian, L. H. Hurley, and S. Neidle. Targeting g-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov*, 10(4):261–75, 2011.

[62] A. De Cian, L. Lacroix, C. Douarre, N. Temime-Smaali, C. Trentesaux, J. F. Riou, and J. L. Mergny. Targeting telomeres and telomerase. *Biochimie*, 90(1):131–55, 2008.

[63] C. Leiterer, A. Csaki, and W. Fritzsche. G-wire synthesis and modification with gold nanoparticle. *Methods Mol Biol*, 749:141–50, 2011.

[64] D. Miyoshi and N. Sugimoto. G-quartet, g-quadruplex, and g-wire regulated by chemical stimuli. *Methods Mol Biol*, 749:93–104, 2011.

[65] G. H. Beaven and E. R. Holiday. Ultraviolet absorption spectra of proteins and amino acids. *Adv Protein Chem*, 7:319–86, 1952.

[66] S. Ulitzur. Rapid determnation of dna base composition by ultraviolet spectroscopy. *Biochim Biophys Acta*, 272(1):1–11, 1972.

[67] J. C. Sutherland and K. P. Griffin. Absorption spectrum of dna for wavelengths greater than 300 nm. *Radiat Res*, 86(3):399–409, 1981.

[68] M. J. Cavaluzzi and P. N. Borer. Revised uv extinction coefficients for nucleoside-5'-monophosphates and unpaired dna and rna. *Nucleic Acids Res*, 32(1):e13, 2004.

[69] A. Fleck and H. N. Munro. The precision of ultraviolet absorption measurements in the schmidt-thannhauser procedure for nucleic acid estimation. *Biochim Biophys Acta*, 55:571–83, 1962.

[70] John K. Snyder. Circular dichroism: Principles and applications k. nakanishi and n. berova (columbia university) and r. w. woody (colorado state university). *Journal of Natural Products*, 59(12):1219–1219, 1996.

[71] D. M. Gray, R. L. Ratliff, and M. R. Vaughan. Circular dichroism spectroscopy of dna. *Methods Enzymol*, 211:389–406, 1992.

[72] C. R. Cantor, M. M. Warshaw, and H. Shapiro. Oligonucleotide interactions. 3. circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, 9(9):1059–77, 1970.

[73] P. N. Borer, L. S. Kan, and P. O. Ts'o. Conformation and interaction of short nucleic acid double-stranded helices. i. proton magnetic resonance studies on the nonexchangeable protons of ribosyl apapgpcpupu. *Biochemistry*, 14(22):4847–63, 1975.

[74] A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo, and M. Webba da Silva. Topological characterization of nucleic acid g-quadruplexes by uv absorption and circular dichroism. *Angew Chem Int Ed Engl*, 50(45):10645–8, 2011.

[75] M. Vorlickova, I. Kejnovska, J. Sagi, D. Renciuk, K. Bednarova, J. Motlova, and J. Kypr. Circular dichroism and guanine quadruplexes. *Methods*, 57(1):64–75, 2012.

[76] S. Masiero, R. Trotta, S. Pieraccini, S. De Tito, R. Perone, A. Randazzo, and G. P. Spada. A non-empirical chromophoric interpretation of cd spectra of dna g-quadruplex structures. *Org Biomol Chem*, 8(12):2683–92, 2010.

[77] A. Randazzo, G.P. Spada, and M.W. da Silva. Circular dichroism of quadruplex structures. *Topics in Current Chemistry*, 330, 2012.

[78] R. Del Villar-Guerra, R. D. Gray, and J. B. Chaires. Characterization of quadruplex dna structure by circular dichroism. *Curr Protoc Nucleic Acid Chem*, 68:17 8 1–17 8 16, 2017.

[79] V. Kuryavyi, A. T. Phan, and D. J. Patel. Solution structures of all parallel-stranded monomeric and dimeric g-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res*, 38(19):6757–73, 2010.

[80] K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, and D. J. Patel. Structure of the human telomere in k+ solution: an intramolecular (3 + 1) g-quadruplex scaffold. *J Am Chem Soc*, 128(30):9963–70, 2006.

[81] A. T. Phan, V. Kuryavyi, H. Y. Gaw, and D. J. Patel. Small-molecule interaction with a five-guanine-tract g-quadruplex structure from the human myc promoter. *Nat Chem Biol*, 1(3):167–73, 2005.

[82] A. Sengar, B. Heddi, and A. T. Phan. Formation of g-quadruplexes in poly-g sequences: structure of a propeller-type parallel-stranded g-quadruplex formed by a g(1)(5) stretch. *Biochemistry*, 53(49):7718–23, 2014.

[83] M. Adrian, B. Heddi, and A. T. Phan. Nmr spectroscopy of g-quadruplexes. *Methods*, 57(1):11–24, 2012.

[84] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus. Dynamics of ligand binding to myoglobin. *Biochemistry*, 14(24):5355–73, 1975.

[85] D. A. Case and M. Karplus. Dynamics of ligand binding to heme proteins. *J Mol Biol*, 132(3):343–68, 1979.

[86] R. Michalczyk and I. M. Russu. Rotational dynamics of adenine amino groups in a dna double helix. *Biophys J*, 76(5):2679–86, 1999.

[87] LW Reeves and KN Shaw. Nuclear magnetic resonance studies of multi-site chemical exchange. i. matrix formulation of the bloch equations. *Canadian Journal of Chemistry*, 48(23):3641–3653, 1970.

[88] J. Jeener, B. H. Meier, P. Bachmann, and R. R. Ernst. Investigation of exchange processes by two dimensional nmr spectroscopy. *The Journal of Chemical Physics*, 71(11):4546–4553, 1979.

[89] B. Heddi, N. Martin-Pintado, Z. Serimbetov, T. M. Kari, and A. T. Phan. G-quadruplexes with (4n - 1) guanines in the g-tetrad core: formation of a g-triad.water complex and implication for small-molecule binding. *Nucleic Acids Res*, 44(2):910–6, 2016.

[90] X. M. Li, K. W. Zheng, J. Y. Zhang, H. H. Liu, Y. D. He, B. F. Yuan, Y. H. Hao, and Z. Tan. Guanine-vacancy-bearing g-quadruplexes responsive to guanine derivatives. *Proc Natl Acad Sci U S A*, 112(47):14581–6, 2015.

[91] P. J. Bates, D. A. Laber, D. M. Miller, S. D. Thomas, and J. O. Trent. Discovery and development of the g-rich oligonucleotide as1411 as a novel treatment for cancer. *Exp Mol Pathol*, 86(3):151–64, 2009.

[92] M. M. Dailey, M. C. Miller, P. J. Bates, A. N. Lane, and J. O. Trent. Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. *Nucleic Acids Res*, 38(14):4877–88, 2010.

[93] S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim, and K. K. Kim. Crystal structure of a junction between b-dna and z-dna reveals two extruded bases. *Nature*, 437(7062):1183–6, 2005.

[94] S. C. Chew, B. Kundukad, T. Seviour, J. R. van der Maarel, L. Yang, S. A. Rice, P. Doyle, and S. Kjelleberg. Dynamic remodeling of microbial biofilms by functionally distinct exopolysaccharides. *MBio*, 5(4):e01536–14, 2014.

[95] X. Chen and P. S. Stewart. Role of electrostatic interactions in cohesion of bacterial biofilms. *Appl Microbiol Biotechnol*, 59(6):718–20, 2002.

[96] H. M. Al-Hashimi and N. G. Walter. Rna dynamics: it is about time. *Curr Opin Struct Biol*, 18(3):321–9, 2008.

[97] B. Furtig, J. Buck, V. Manoharan, W. Bermel, A. Jaschke, P. Wenter, S. Pitsch, and H. Schwalbe. Time-resolved nmr studies of rna folding. *Biopolymers*, 86(5-6):360–383, 2007.

[98] N. Leulliot and G. Varani. Current topics in rna-protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry*, 40(27):7947–7956, 2001.

[99] J. R. Williamson. Induced fit in rna-protein recognition. *Nature Structural Biology*, 7(10):834–837, 2000.

[100] R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, and B. Honig. The role of dna shape in protein-dna recognition. *Nature*, 461(7268):1248–U81, 2009.

[101] G. A. Meints and G. P. Drobny. Dynamic impact of methylation at the m. hhai target site: A solid-state deuterium nmr study. *Biochemistry*, 40(41):12436–12443, 2001.

[102] Anthony K Mittermaier and Lewis E Kay. Observing biological dynamics at atomic resolution using nmr. *Trends in biochemical sciences*, 34(12):601–611, 2009.

[103] R. J. Roberts and X. D. Cheng. Base flipping. *Annual Review of Biochemistry*, 67:181–+, 1998.

[104] J. J. P. Perry, E. Cotner-Gohara, T. Ellenberger, and J. A. Tainer. Structural dynamics in dna damage signaling and repair. *Current Opinion in Structural Biology*, 20(3):283–294, 2010.

[105] M. Gueron, M. Kochoyan, and J. L. Leroy. A single-mode of dna base-pair opening drives imino proton-exchange. *Nature*, 328(6125):89–92, 1987.

[106] M. Kochoyan, J. L. Leroy, and M. Gueron. Processes of base-pair opening and proton exchange in z-dna. *Biochemistry*, 29(20):4799–805, 1990.

[107] E. Folta-Stogniew and I. M. Russu. Sequence dependence of base-pair opening in a dna dodecamer containing the caca/gtgt sequence motif. *Biochemistry*, 33(36):11016–24, 1994.

[108] U. Dornberger, M. Leijon, and H. Fritzsche. High base pair opening rates in tracts of gc base pairs. *J Biol Chem*, 274(11):6957–62, 1999.

[109] C. Chen and I. M. Russu. Sequence-dependence of the energetics of opening of at basepairs in dna. *Biophys J*, 87(4):2545–51, 2004.

[110] C. Bardin and J. L. Leroy. The formation pathway of tetramolecular g-quadruplexes. *Nucleic Acids Res*, 36(2):477–88, 2008.

[111] A. N. Lane, J. B. Chaires, R. D. Gray, and J. O. Trent. Stability and kinetics of g-quadruplex structures. *Nucleic Acids Res*, 36(17):5482–515, 2008.

[112] J. Sponer and N. Spackova. Molecular dynamics simulations and their application to four-stranded dna. *Methods*, 43(4):278–90, 2007.

[113] S. Haider, G. N. Parkinson, and S. Neidle. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys J*, 95(1):296–311, 2008.

[114] L. Petraccone, N. C. Garbett, J. B. Chaires, and J. O. Trent. An integrated molecular dynamics (md) and experimental study of higher order human telomeric quadruplexes. *Biopolymers*, 93(6):533–48, 2010.

[115] B. Islam, M. Sgobba, C. Laughton, M. Orozco, J. Sponer, S. Neidle, and S. Haider. Conformational dynamics of the human propeller telomeric dna quadruplex on a microsecond time scale. *Nucleic Acids Res*, 41(4):2723–35, 2013.

[116] P. J. Tummino and R. A. Copeland. Residence time of receptor-ligand complexes and its effect on biological function. *Biochemistry*, 47(20):5481–92, 2008.

[117] R. A. Copeland, D. L. Pompliano, and T. D. Meek. Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov*, 5(9):730–9, 2006.

[118] Robert A Copeland. A practical introduction to structure, mechanism, and data analysis. *Enzymes, 2nd ed. John Wiley and Sons, New York, NY*, page 104, 2000.

[119] G. Vauquelin, F. Fierens, and I. Van Liefde. Long-lasting angiotensin type 1 receptor binding and protection by candesartan: comparison with other biphenyl-tetrazole sartans. *J Hypertens Suppl*, 24(1):S23–30, 2006.

[120] G. Vauquelin and I. Van Liefde. Slow antagonist dissociation and long-lasting in vivo receptor protection. *Trends Pharmacol Sci*, 27(7):356–9, 2006.

[121] David C Swinney. Can binding kinetics translate to a clinically differentiated drug? from theory to practice. *Letters in Drug Design and Discovery*, 3(8):569–574, 2006.

[122] A. Kettani, M. Gueron, and J. L. Leroy. Amino proton exchange processes in mononucleosides. *J Am Chem Soc*, 119(5):1108–1115, 1997.

[123] J. Lounila and J. Jokisaari. Anisotropies in spin spin coupling-constants and chemical-shifts as determined from the nmr-spectra of molecules oriented by liquid-crystal solvents. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 15:249–290, 1982.

[124] E. Gavathiotis and M. S. Searle. Structure of the parallel-stranded dna quadruplex d(ttagggt)4 containing the human telomeric repeat: evidence for a-tetrad formation from nmr and molecular dynamics simulations. *Org Biomol Chem*, 1(10):1650–6, 2003.

[125] Y. Kato, T. Ohyama, H. Mita, and Y. Yamamoto. Dynamics and thermodynamics of dimerization of parallel g-quadruplexed dna formed from d(ttagn) (n=3-5). *J Am Chem Soc*, 127(28):9980–1, 2005.

[126] S. Amrane, M. Adrian, B. Heddi, A. Serero, A. Nicolas, J. L. Mergny, and A. T. Phan. Formation of pearl-necklace monomorphic g-quadruplexes in the human ceb25 minisatellite. *J Am Chem Soc*, 134(13):5807–16, 2012.

[127] A. Kettani, G. Basu, A. Gorin, A. Majumdar, E. Skripkin, and D. J. Patel. A two-stranded template-based approach to g center dot(c-a) triad formation: Designing novel structural elements into an existing dna framework. *Journal of Molecular Biology*, 301(1):129–146, 2000.

[128] N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin, and D. J. Patel. Dimeric dna quadruplex containing major groove-aligned a-t-a-t and g-c-g-c tetrads stabilized by inter-subunit watson-crick a-t and g-c pairs. *J Mol Biol*, 312(5):1073–88, 2001.

[129] P. K. Patel, A. S. Koti, and R. V. Hosur. Nmr studies on truncated sequences of human telomeric dna: observation of a novel a-tetrad. *Nucleic Acids Res*, 27(19):3836–43, 1999.

[130] N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin, and D. J. Patel. V-shaped scaffold: A new architectural motif identified in an a center dot(g center dot g center dot g center dot g) pentad-containing dimeric dna quadruplex involving stacked g(anti)center dot g(anti)center dot g(anti)center dot g(syn) tetrads. *Journal of Molecular Biology*, 311(5):1063–1079, 2001.

[131] A. Kettani, A. Gorin, A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. Jones, and D. J. Patel. A dimeric dna interface stabilized by stacked a.(g.g.g.g).a hexads and coordinated monovalent cations. *J Mol Biol*, 297(3):627–44, 2000.

[132] A. Matsugami, T. Okuizumi, S. Uesugi, and M. Katahira. Intramolecular higher order packing of parallel quadruplexes comprising a g:g:g:g tetrad and a g(:a):g(:a):g(:a):g heptad of gga triplet repeat dna. *J Biol Chem*, 278(30):28147–53, 2003.

[133] W. J. Chung, B. Heddi, M. Tera, K. Iida, K. Nagasawa, and A. T. Phan. Solution structure of an intramolecular $(3 + 1)$ human telomeric g-quadruplex bound to a telomestatin derivative. *J Am Chem Soc*, 135(36):13495–501, 2013.

[134] Isabelle M Dixon, Frederic Lopez, Agueda M Tejera, Jean-Pierre Esteve, Maria A Blasco, Genevieve Pratviel, and Bernard Meunier. A g-quadruplex ligand with 10000-fold selectivity over duplex dna. *Journal of the American Chemical Society*, 129(6):1502–1503, 2007.

[135] S. Balasubramanian and S. Neidle. G-quadruplex nucleic acids as therapeutic targets. *Curr Opin Chem Biol*, 13(3):345–53, 2009.

[136] H. Han and L. H. Hurley. G-quadruplex dna: a potential target for anti-cancer drug design. *Trends Pharmacol Sci*, 21(4):136–42, 2000.

[137] L. Oganesian and T. M. Bryan. Physiological relevance of telomeric g-quadruplex formation: a potential drug target. *Bioessays*, 29(2):155–65, 2007.

[138] V. V. Cheong, B. Heddi, C. J. Lech, and A. T. Phan. Xanthine and 8-oxoguanine in g-quadruplexes: formation of a g.g.x.o tetrad. *Nucleic Acids Res*, 43(21):10506–14, 2015.

[139] Brahim Heddi, Vee Vee Cheong, Herry Martadinata, and Anh Tuân Phan. Insights into g-quadruplex specific recognition by the deah-box helicase rhau: Solution structure of a peptide–quadruplex complex. *Proceedings of the National Academy of Sciences*, 112(31):9608–9613, 2015.

[140] K. Bose, C. J. Lech, B. Heddi, and A. T. Phan. High-resolution afm structure of dna g-wires in aqueous solution. *Nat Commun*, 9(1):1959, 2018.

[141] Christopher Jacques Lech and Anh Tuân Phan. Ball with hair: modular functionalization of highly stable g-quadruplex dna nano-scaffolds through n2-guanine modification. *Nucleic acids research*, 45(11):6265–6274, 2017.

[142] J. Feigon, K. M. Koshlap, and F. W. Smith. 1h nmr spectroscopy of dna triplexes and quadruplexes. *Methods Enzymol*, 261:225–55, 1995.

[143] D. J. Patel and A. E. Tonelli. Assignment of the proton nmr chemical shifts of the t-n3h and g-n1h proton resonances in isolated at and gc watson-crick base pairs in double-stranded deoxy oligonucleotides in aqueous solution. *Biopolymers*, 13(10):1943–64, 1974.

[144] N. Q. Do and A. T. Phan. Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded g-quadruplexes: Nmr structural study. *Chemistry*, 18(46):14752–9, 2012.

[145] Albert W. Overhauser. Polarization of nuclei in metals. *Physical Review*, 92(2):411–415, 1953.

[146] F. W. Smith and J. Feigon. Quadruplex structure of oxytricha telomeric dna oligonucleotides. *Nature*, 356(6365):164–8, 1992.

[147] Y. Wang, C. de los Santos, X. O. Gao, K. Greene, D. Live, and D. J. Patel. Multinuclear nuclear magnetic resonance studies of na cation-stabilized complex formed by d(g-g-t-t-t-t-c-g-g) in solution. implications for g-tetrad structures. *J Mol Biol*, 222(3):819–32, 1991.

[148] Y. Wang and D. J. Patel. Solution structure of a parallel-stranded g-quadruplex dna. *J Mol Biol*, 234(4):1171–83, 1993.

[149] R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe, and J. Feigon. Thrombin-binding dna aptamer forms a unimolecular quadruplex structure in solution. *Proc Natl Acad Sci U S A*, 90(8):3745–9, 1993.

[150] F. Aboul-ela, A. I. Murchie, and D. M. Lilley. Nmr study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(tg4t). *Nature*, 360(6401):280–2, 1992.

[151] A. T. Phan. Long-range imino proton-13c j-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled dna. *J Biomol NMR*, 16(2):175–8, 2000.

[152] Pierre Plateau and Maurice Gueron. Exchangeable proton nmr without baseline distorsion, using new strong-pulse sequences. *Journal of the American Chemical Society*, 104(25):7310–7311, 1982.

[153] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[154] A. K. Todd, M. Johnston, and S. Neidle. Highly prevalent putative quadruplex sequence motifs in human dna. *Nucleic Acids Res*, 33(9):2901–7, 2005.

[155] A. T. Phan, V. Kuryavyi, J. B. Ma, A. Faure, M. L. Andreola, and D. J. Patel. An interlocked dimeric parallel-stranded dna quadruplex: a potent inhibitor of hiv-1 integrase. *Proc Natl Acad Sci U S A*, 102(3):634–9, 2005.

[156] M. Adrian, D. J. Ang, C. J. Lech, B. Heddi, A. Nicolas, and A. T. Phan. Structure and conformational dynamics of a stacked dimeric g-quadruplex formed by the human ceb1 minisatellite. *J Am Chem Soc*, 136(17):6297–305, 2014.

[157] V. Kocman and J. Plavec. A tetrahelical dna fold adopted by tandem repeats of alternating ggg and gcg tracts. *Nat Commun*, 5:5831, 2014.

[158] L. Cerofolini, J. Amato, A. Giachetti, V. Limongelli, E. Novellino, M. Parrinello, M. Fragai, A. Randazzo, and C. Luchinat. G-triplex structure and formation propensity. *Nucleic Acids Res*, 42(21):13393–404, 2014.

[159] V. Limongelli, S. De Tito, L. Cerofolini, M. Fragai, B. Pagano, R. Trotta, S. Cosconati, L. Marinelli, E. Novellino, I. Bertini, A. Randazzo, C. Luchinat, and M. Parrinello. The g-triplex dna. *Angew Chem Int Ed Engl*, 52(8):2269–73, 2013.

[160] V. T. Mukundan and A. T. Phan. Bulges in g-quadruplexes: broadening the definition of g-quadruplex-forming sequences. *J Am Chem Soc*, 135(13):5017–28, 2013.

[161] K. W. Lim, T. Q. Nguyen, and A. T. Phan. Joining of multiple duplex stems at a single quadruplex loop. *J Am Chem Soc*, 136(52):17969–73, 2014.

[162] P. Stadlbauer, L. Trantirek, 3rd Cheatham, T. E., J. Koca, and J. Sponer. Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie*, 105:22–35, 2014.

[163] T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran, and H. Sugiyama. Folding pathways of human telomeric type-1 and type-2 g-quadruplex structures. *J Am Chem Soc*, 132(42):14910–8, 2010.

[164] F. Rosu, V. Gabelica, H. Poncelet, and E. De Pauw. Tetramolecular g-quadruplex formation pathways studied by electrospray mass spectrometry. *Nucleic Acids Res*, 38(15):5217–25, 2010.

[165] M. Boncina, J. Lah, I. Prislan, and G. Vesnaver. Energetic basis of human telomeric dna folding into g-quadruplex structures. *J Am Chem Soc*, 134(23):9657–63, 2012.

[166] R. D. Gray, J. O. Trent, and J. B. Chaires. Folding and unfolding pathways of the human telomeric g-quadruplex. *J Mol Biol*, 426(8):1629–50, 2014.

[167] W. Li, X. M. Hou, P. Y. Wang, X. G. Xi, and M. Li. Direct measurement of sequential folding pathway and energy landscape of human telomeric g-quadruplex structures. *J Am Chem Soc*, 135(17):6423–6, 2013.

[168] A. Y. Zhang and S. Balasubramanian. The kinetics and folding pathways of intramolecular g-quadruplex nucleic acids. *J Am Chem Soc*, 134(46):19297–308, 2012.

[169] L. H. Hurley. Dna and its associated processes as targets for cancer therapy. *Nat Rev Cancer*, 2(3):188–200, 2002.

[170] J. L. Mergny, J. F. Riou, P. Mailliet, M. P. Teulade-Fichou, and E. Gilson. Natural and pharmacological regulation of telomerase. *Nucleic Acids Res*, 30(4):839–65, 2002.

[171] E. M. Rezler, D. J. Bearss, and L. H. Hurley. Telomere inhibition and telomere disruption as processes for drug targeting. *Annu Rev Pharmacol Toxicol*, 43:359–79, 2003.

[172] S. Neidle and D. E. Thurston. Chemical approaches to the discovery and development of cancer therapies. *Nat Rev Cancer*, 5(4):285–96, 2005.

[173] N. Maizels. Dynamic roles for g4 dna in the biology of eukaryotic cells. *Nat Struct Mol Biol*, 13(12):1055–9, 2006.

[174] M. Fry. Tetraplex dna and its interacting proteins. *Front Biosci*, 12:4336–51, 2007.

[175] L. Kelland. Targeting the limitless replicative potential of cancer: the telomerase/telomere pathway. *Clin Cancer Res*, 13(17):4960–3, 2007.

[176] B. Pagano and C. Giancola. Energetics of quadruplex-drug recognition in anticancer therapy. *Curr Cancer Drug Targets*, 7(6):520–40, 2007.

[177] P. J. Perry, A. P. Reszka, A. A. Wood, M. A. Read, S. M. Gowan, H. S. Dosanjh, J. O. Trent, T. C. Jenkins, L. R. Kelland, and S. Neidle. Human telomerase inhibition by regioisomeric disubstituted amidoanthracene-9,10-diones. *J Med Chem*, 41(24):4873–84, 1998.

[178] P. J. Perry, M. A. Read, R. T. Davies, S. M. Gowan, A. P. Reszka, A. A. Wood, L. R. Kelland, and S. Neidle. 2,7-disubstituted amidofluorenone derivatives as inhibitors of human telomerase. *J Med Chem*, 42(14):2679–84, 1999.

[179] R. J. Harrison, A. P. Reszka, S. M. Haider, B. Romagnoli, J. Morrell, M. A. Read, S. M. Gowan, C. M. Incles, L. R. Kelland, and S. Neidle. Evaluation of by disubstituted acridone derivatives as telomerase inhibitors: the importance of g-quadruplex binding. *Bioorg Med Chem Lett*, 14(23):5845–9, 2004.

[180] S. M. Haider, G. N. Parkinson, and S. Neidle. Structure of a g-quadruplex-ligand complex. *J Mol Biol*, 326(1):117–25, 2003.

[181] O. Y. Fedoroff, M. Salazar, H. Han, V. V. Chemeris, S. M. Kerwin, and L. H. Hurley. Nmr-based model of a telomerase-inhibiting compound bound to g-quadruplex dna. *Biochemistry*, 37(36):12367–74, 1998.

[182] M. J. Moore, C. M. Schultes, J. Cuesta, F. Cuenca, M. Gunaratnam, F. A. Tanious, W. D. Wilson, and S. Neidle. Trisubstituted acridines as g-quadruplex telomere targeting agents. effects of extensions of the 3,6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. *J Med Chem*, 49(2):582–99, 2006.

[183] C. M. Schultes, B. Guyen, J. Cuesta, and S. Neidle. Synthesis, biophysical and biological evaluation of 3,6-bis-amidoacridines with extended 9-anilino substituents as potent g-quadruplex-binding telomerase inhibitors. *Bioorg Med Chem Lett*, 14(16):4347–51, 2004.

[184] Qian Li, Junfeng Xiang, Xudong Li, Lirong Chen, Xiaojie Xu, Yalin Tang, Qiuju Zhou, Lin Li, Hong Zhang, and Hongxia Sun. Stabilizing parallel g-quadruplex dna by a new class of ligands: two non-planar alkaloids through interaction in lateral grooves. *Biochimie*, 91(7):811–819, 2009.

[185] Bryson D Bennett, Elizabeth H Kimball, Melissa Gao, Robin Osterhout, Stephen J Van Dien, and Joshua D Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nature chemical biology*, 5(8):593, 2009.

[186] A. Ablasser, M. Goldeck, T. Cavlar, T. Deimling, G. Witte, I. Rohl, K. P. Hopfner, J. Ludwig, and V. Hornung. cgas produces a 2'-5'-linked cyclic dinucleotide second messenger that activates sting. *Nature*, 498(7454):380–4, 2013.

[187] E. J. Diner, D. L. Burdette, S. C. Wilson, K. M. Monroe, C. A. Kellenberger, M. Hyodo, Y. Hayakawa, M. C. Hammond, and R. E. Vance. The innate immune dna sensor cgas produces a noncanonical cyclic dinucleotide that activates human sting. *Cell Rep*, 3(5):1355–61, 2013.

[188] F. Rao, R. Y. See, D. Zhang, D. C. Toh, Q. Ji, and Z. X. Liang. Yybt is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a ggdef domain with atpase activity. *J Biol Chem*, 285(1):473–82, 2010.

[189] C. Shu, G. Yi, T. Watts, C. C. Kao, and P. Li. Structure of sting bound to cyclic di-gmp reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat Struct Mol Biol*, 19(7):722–4, 2012.

[190] K. Sureka, P. H. Choi, M. Precit, M. Delince, D. A. Pensinger, T. N. Huynh, A. R. Jurado, Y. A. Goo, M. Sadilek, A. T. Iavarone, J. D. Sauer, L. Tong,

and J. J. Woodward. The cyclic dinucleotide c-di-amp is an allosteric regulator of metabolic enzyme function. *Cell*, 158(6):1389–1401, 2014.

[191] J. W. Shim, Q. Tan, and L. Q. Gu. Single-molecule detection of folding and unfolding of the g-quadruplex aptamer in a nanopore nanocavity. *Nucleic Acids Res*, 37(3):972–82, 2009.

[192] B. I. Kankia and L. A. Marky. Folding of the thrombin aptamer into a g-quadruplex with sr(2+): stability, heat, and hydration. *J Am Chem Soc*, 123(44):10799–804, 2001.

[193] T. Li, S. Dong, and E. Wang. Label-free colorimetric detection of aqueous mercury ion (hg2+) using hg2+-modulated g-quadruplex-based dnazymes. *Anal Chem*, 81(6):2144–9, 2009.

[194] T. Li, L. Shi, E. Wang, and S. Dong. Multifunctional g-quadruplex aptamers and their application to protein detection. *Chemistry*, 15(4):1036–42, 2009.

[195] B. Gatto, M. Palumbo, and C. Sissi. Nucleic acid aptamers based on the g-quadruplex structure: therapeutic and diagnostic potential. *Curr Med Chem*, 16(10):1248–65, 2009.

[196] J. E. Rosenberg, R. M. Bambury, E. M. Van Allen, H. A. Drabkin, Jr. Lara, P. N., A. L. Harzstark, N. Wagle, R. A. Figlin, G. W. Smith, L. A. Garraway, T. Choueiri, F. Erlandsson, and D. A. Laber. A phase ii trial of as1411 (a novel nucleolin-targeted dna aptamer) in metastatic renal cell carcinoma. *Invest New Drugs*, 32(1):178–87, 2014.

[197] S. H. Chou, K. H. Chin, and A. H. Wang. Dna aptamers as potential anti-hiv agents. *Trends Biochem Sci*, 30(5):231–4, 2005.

[198] C. Platella, C. Riccardi, D. Montesarchio, G. N. Roviello, and D. Musumeci. G-quadruplex-based aptamers against protein targets in therapy and diagnostics. *Biochim Biophys Acta*, 1861(5 Pt B):1429–1447, 2017.

[199] P. J. Bates, E. M. Reyes-Reyes, M. T. Malik, E. M. Murphy, M. G. O'Toole, and J. O. Trent. G-quadruplex oligonucleotide as1411 as a cancer-targeting agent: Uses and mechanisms. *Biochim Biophys Acta*, 1861(5 Pt B):1414–1428, 2017.

[200] A. C. Girvan, Y. Teng, L. K. Casson, S. D. Thomas, S. Juliger, M. W. Ball, J. B. Klein, Jr. Pierce, W. M., S. S. Barve, and P. J. Bates. Agro100 inhibits activation of nuclear factor-kappab (nf-kappab) by forming a complex with nf-kappab essential modulator (nemo) and nucleolin. *Mol Cancer Ther*, 5(7):1790–9, 2006.

[201] N. Q. Do, W. J. Chung, T. H. A. Truong, B. Heddi, and A. T. Phan. G-quadruplex structure of an anti-proliferative dna sequence. *Nucleic Acids Res*, 45(12):7487–7493, 2017.

[202] V. Kuryavyi, A. Kettani, W. Wang, R. Jones, and D. J. Patel. A diamond-shaped zipper-like dna architecture containing triads sandwiched between mismatches and tetrads. *J Mol Biol*, 295(3):455–69, 2000.

[203] V. Kuryavyi, A. Majumdar, A. Shallop, N. Chernichenko, E. Skripkin, R. Jones, and D. J. Patel. A double chain reversal loop and two diagonal loops define the architecture of a unimolecular dna quadruplex containing a pair of stacked g(syn)-g(syn)-g(anti)-g(anti) tetrads flanked by a g-(t-t) triad and a t-t-t triple. *J Mol Biol*, 310(1):181–94, 2001.

[204] A. Kettani, G. Basu, A. Gorin, A. Majumdar, E. Skripkin, and D. J. Patel. A two-stranded template-based approach to g.(c-a) triad formation: designing novel structural elements into an existing dna framework. *J Mol Biol*, 301(1):129–46, 2000.

[205] C. Epplen and J. T. Epplen. Expression of (cac)n/(gtg)n simple repetitive sequences in mrna of human lymphocytes. *Hum Genet*, 93(1):35–41, 1994.

[206] T. Matsuyama, T. Akihama, Y. Ito, M. Omura, and K. Fukui. Distribution of tgg repeat-related sequences in 'trovita' orange (citrus sinensis osbeck) chromosomes. *Genome*, 42(6):1251–4, 1999.

[207] F. Bena, S. Gimelli, E. Migliavacca, N. Brun-Druc, K. Buiting, S. E. Antonarakis, and A. J. Sharp. A recurrent 14q32.2 microdeletion mediated by expanded tgg repeats. *Hum Mol Genet*, 19(10):1967–73, 2010.

[208] M. Zemanek, J. Kypr, and M. Vorlickova. Conformational properties of dna containing (cca)n and (tgg)n trinucleotide repeats. *Int J Biol Macromol*, 36(1-2):23–32, 2005.

[209] B. De Nicola, C. J. Lech, B. Heddi, S. Regmi, I. Frasson, R. Perrone, S. N. Richter, and A. T. Phan. Structure and possible function of a g-quadruplex in the long terminal repeat of the proviral hiv-1 genome. *Nucleic Acids Res*, 44(13):6442–51, 2016.

[210] T. Miyahara, H. Nakatsuji, and H. Sugiyama. Helical structure and circular dichroism spectra of dna: a theoretical study. *J Phys Chem A*, 117(1):42–55, 2013.

[211] Peter Schultze, Nicholas V Hud, Flint W Smith, and Juli Feigon. The effect of sodium, potassium and ammonium ions on the conformation of the dimeric

quadruplex formed by the oxytricha nova telomere repeat oligonucleotide d (g 4 t 4 g 4). *Nucleic acids research*, 27(15):3018–3028, 1999.

[212] Peter Podbevšek, Nicholas V Hud, and Janez Plavec. Nmr evaluation of ammonium ion movement within a unimolecular g-quadruplex in solution. *Nucleic acids research*, 35(8):2554–2563, 2007.

[213] P. L. Phillips and G. S. Schultz. Molecular mechanisms of biofilm infection: Biofilm virulence factors. *Adv Wound Care (New Rochelle)*, 1(3):109–114, 2012.

[214] K. Drescher, Y. Shen, B. L. Bassler, and H. A. Stone. Biofilm streamers cause catastrophic disruption of flow with consequences for environmental and medical systems. *Proc Natl Acad Sci U S A*, 110(11):4345–50, 2013.

[215] T. W. Seviour, L. K. Lambert, M. Pijuan, and Z. Yuan. Selectively inducing the synthesis of a key structural exopolysaccharide in aerobic granules by enriching for candidatus "competibacter phosphatis". *Appl Microbiol Biotechnol*, 92(6):1297–305, 2011.

[216] N. Hoiby, O. Ciofu, and T. Bjarnsholt. Pseudomonas aeruginosa biofilms in cystic fibrosis. *Future Microbiol*, 5(11):1663–74, 2010.

[217] A. Kurniawan, T. Yamamoto, Y. Tsuchiya, and H. Morisaki. Analysis of the ion adsorption-desorption characteristics of biofilm matrices. *Microbes Environ*, 27(4):399–406, 2012.

[218] M. K. de Kreuk, C. Picioreanu, M. Hosseini, J. B. Xavier, and M. C. van Loosdrecht. Kinetic model of a granular sludge sbr: influences on nutrient removal. *Biotechnol Bioeng*, 97(4):801–15, 2007.

[219] H. C. Flemming and J. Wingender. The biofilm matrix. *Nat Rev Microbiol*, 8(9):623–33, 2010.

[220] T. Seviour, A. K. Malde, S. Kjelleberg, Z. Yuan, and A. E. Mark. Molecular dynamics unlocks atomic level self-assembly of the exopolysaccharide matrix of water-treatment granular biofilms. *Biomacromolecules*, 13(6):1965–72, 2012.

[221] G. P. Bodey, R. Bolivar, V. Fainstein, and L. Jadeja. Infections caused by pseudomonas aeruginosa. *Rev Infect Dis*, 5(2):279–313, 1983.

[222] A. Ghafoor, I. D. Hay, and B. H. Rehm. Role of exopolysaccharides in pseudomonas aeruginosa biofilm formation and architecture. *Appl Environ Microbiol*, 77(15):5238–46, 2011.

[223] T. Seviour, S. H. Hansen, L. Yang, Y. H. Yau, V. B. Wang, M. R. Stenvang, G. Christiansen, E. Marsili, M. Givskov, Y. Chen, D. E. Otzen, P. H. Nielsen, S. Geifman-Shochat, S. Kjelleberg, and M. S. Dueholm. Functional amyloids keep quorum-sensing molecules in check. *J Biol Chem*, 290(10):6457–69, 2015.

[224] B. R. Borlee, A. D. Goldman, K. Murakami, R. Samudrala, D. J. Wozniak, and M. R. Parsek. Pseudomonas aeruginosa uses a cyclic-di-gmp-regulated adhesin to reinforce the biofilm extracellular matrix. *Mol Microbiol*, 75(4):827–42, 2010.

[225] M. Allesen-Holm, K. B. Barken, L. Yang, M. Klausen, J. S. Webb, S. Kjelleberg, S. Molin, M. Givskov, and T. Tolker-Nielsen. A characterization of dna release in pseudomonas aeruginosa cultures and biofilms. *Mol Microbiol*, 59(4):1114–28, 2006.

[226] M. Okshevsky and R. L. Meyer. The role of extracellular dna in the establishment, maintenance and perpetuation of bacterial biofilms. *Crit Rev Microbiol*, 41(3):341–52, 2015.

[227] K. M. Colvin, Y. Irie, C. S. Tart, R. Urbano, J. C. Whitney, C. Ryder, P. L. Howell, D. J. Wozniak, and M. R. Parsek. The pel and psl polysaccharides provide pseudomonas aeruginosa structural redundancy within the biofilm matrix. *Environ Microbiol*, 14(8):1913–28, 2012.

[228] Y. Irie, B. R. Borlee, J. R. O'Connor, P. J. Hill, C. S. Harwood, D. J. Wozniak, and M. R. Parsek. Self-produced exopolysaccharide is a signal that stimulates biofilm formation in pseudomonas aeruginosa. *Proc Natl Acad Sci U S A*, 109(50):20632–6, 2012.

[229] F. A. Herbst, M. T. Sondergaard, H. Kjeldal, A. Stensballe, P. H. Nielsen, and M. S. Dueholm. Major proteomic changes associated with amyloid-induced biofilm formation in pseudomonas aeruginosa pao1. *J Proteome Res*, 14(1):72–81, 2015.

[230] Hua Zhao. Dna stability in ionic liquids and deep eutectic solvents. *Journal of Chemical Technology and Biotechnology*, 90(1):19–25, 2015.

[231] JÃ$\frac{1}{4}$rgen Vitz, Tina Erdmenger, Claudia Haensch, and Ulrich S Schubert. Extended dissolution studies of cellulose in imidazolium based ionic liquids. *Green chemistry*, 11(3):417–424, 2009.

[232] U. Bockelmann, A. Janke, R. Kuhn, T. R. Neu, J. Wecke, J. R. Lawrence, and U. Szewzyk. Bacterial extracellular dna forming a defined network-like structure. *FEMS Microbiol Lett*, 262(1):31–8, 2006.

[233] L. K. Jennings, K. M. Storek, H. E. Ledvina, C. Coulon, L. S. Marmont, I. Sadovskaya, P. R. Secor, B. S. Tseng, M. Scian, A. Filloux, D. J. Wozniak, P. L. Howell, and M. R. Parsek. Pel is a cationic exopolysaccharide that cross-links extracellular dna in the pseudomonas aeruginosa biofilm matrix. *Proc Natl Acad Sci U S A*, 112(36):11353–8, 2015.

[234] N. J. Greenfield. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*, 1(6):2876–90, 2006.

[235] Jaroslav Kypr, Iva Kejnovska, Daniel Renciuk, and Michaela Vorlickova. Circular dichroism and conformational polymorphism of dna. *Nucleic acids research*, 37(6):1713–1725, 2009.

[236] G. Khandelwal and J. Bhyravabhotla. A phenomenological model for predicting melting temperatures of dna sequences. *PLoS One*, 5(8):e12433, 2010.

[237] L. Friedman and R. Kolter. Two genetic loci produce distinct carbohydrate-rich structural components of the pseudomonas aeruginosa biofilm matrix. *J Bacteriol*, 186(14):4457–65, 2004.

[238] Mike Engelhardt *et. al.* Nmr guide. `http://triton.iqfr.csic.es/guide/eNMR/eNMR2Dinv/`, 2000.

[239] J. Nadel, R. Athanasiadou, C. Lemetre, N. A. Wijetunga, O. Broin P, H. Sato, Z. Zhang, J. Jeddeloh, C. Montagna, A. Golden, C. Seoighe, and J. M. Greally. Rna:dna hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin*, 8:46, 2015.

[240] M. Okshevsky and R. L. Meyer. Evaluation of fluorescent stains for visualizing extracellular dna in biofilms. *J Microbiol Methods*, 105:102–4, 2014.

[241] Daniel J Cornwell and David K Smith. Expanding the scope of gels — combining polymers with low-molecular-weight gelators to yield modified self-assembling smart materials with high-tech applications. *Materials Horizons*, 2(3):279–293, 2015.

[242] Y. Z. Chen, C. L. Bennett, H. M. Huynh, I. P. Blair, I. Puls, J. Irobi, I. Dierick, A. Abel, M. L. Kennerson, B. A. Rabin, G. A. Nicholson, M. Auer-Grumbach, K. Wagner, P. De Jonghe, J. W. Griffin, K. H. Fischbeck, V. Timmerman, D. R. Cornblath, and P. F. Chance. Dna/rna helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (als4). *Am J Hum Genet*, 74(6):1128–35, 2004.

[243] C. Geerts-Dimitriadou, Y. Y. Lu, C. Geertsema, R. Goldbach, and R. Kormelink. Analysis of the tomato spotted wilt virus ambisense s rna-encoded hairpin structure in translation. *PLoS One*, 7(2):e31013, 2012.

[244] H. Nakamura, Y. Oda, S. Iwai, H. Inoue, E. Ohtsuka, S. Kanaya, S. Kimura, C. Katsuda, K. Katayanagi, K. Morikawa, and et al. How does rnase h recognize a dna.rna hybrid? *Proc Natl Acad Sci U S A*, 88(24):11535–9, 1991.

[245] A. Jain and R. D. Vale. Rna phase transitions in repeat expansion disorders. *Nature*, 546(7657):243–247, 2017.

[246] S. Millevoi, H. Moine, and S. Vagner. G-quadruplexes in rna biology. *Wiley Interdiscip Rev RNA*, 3(4):495–507, 2012.

[247] S. Kailasan Vanaja, V. A. Rathinam, M. K. Atianand, P. Kalantari, B. Skehan, K. A. Fitzgerald, and J. M. Leong. Bacterial rna:dna hybrids are activators of the nlrp3 inflammasome. *Proc Natl Acad Sci U S A*, 111(21):7765–70, 2014.

[248] E. E. Mann, K. C. Rice, B. R. Boles, J. L. Endres, D. Ranjit, L. Chandramohan, L. H. Tsang, M. S. Smeltzer, A. R. Horswill, and K. W. Bayles. Modulation of edna release and degradation affects staphylococcus aureus biofilm maturation. *PLoS One*, 4(6):e5822, 2009.

[249] B. Adam, G. S. Baillie, and L. J. Douglas. Mixed species biofilms of candida albicans and staphylococcus epidermidis. *J Med Microbiol*, 51(4):344–9, 2002.

[250] S. J. Rose, L. M. Babrak, and L. E. Bermudez. Mycobacterium avium possesses extracellular dna that contributes to biofilm formation, structural integrity, and tolerance to antibiotics. *PLoS One*, 10(5):e0128772, 2015.

[251] M. Adrian, F. R. Winnerdy, B. Heddi, and A. T. Phan. Rotation of guanine amino groups in g-quadruplexes: A probe for local structure and ligand binding. *Biophys J*, 113(4):775–784, 2017.

[252] H. Yu. Extending the size limit of protein nuclear magnetic resonance. *Proc Natl Acad Sci U S A*, 96(2):332–4, 1999.

[253] Z. Serber and V. Dotsch. In-cell nmr spectroscopy. *Biochemistry*, 40(48):14317–23, 2001.

[254] Z. Serber, R. Ledwidge, S. M. Miller, and V. Dotsch. Evaluation of parameters critical to observing proteins inside living escherichia coli by in-cell nmr spectroscopy. *J Am Chem Soc*, 123(37):8895–901, 2001.

[255] A. Y. Maldonado, D. S. Burz, and A. Shekhtman. In-cell nmr spectroscopy. *Prog Nucl Magn Reson Spectrosc*, 59(3):197–212, 2011.

[256] P. Selenko and G. Wagner. Nmr mapping of protein interactions in living cells. *Nat Methods*, 3(2):80–1, 2006.

[257] J. Xie, R. Thapa, S. Reverdatto, D. S. Burz, and A. Shekhtman. Screening of small molecule interactor library by using in-cell nmr spectroscopy (smili-nmr). *J Med Chem*, 52(11):3516–22, 2009.

[258] R. Hansel, S. Foldynova-Trantirkova, F. Lohr, J. Buck, E. Bongartz, E. Bamberg, H. Schwalbe, V. Dotsch, and L. Trantirek. Evaluation of parameters

critical for observing nucleic acids inside living xenopus laevis oocytes by in-cell nmr spectroscopy. *J Am Chem Soc*, 131(43):15761–8, 2009.

[259] D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Walchli, B. O. Smith, M. Shirakawa, P. Guntert, and Y. Ito. Protein structure determination in living cells by in-cell nmr spectroscopy. *Nature*, 458(7234):102–5, 2009.

[260] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[261] J. L. Mergny, A. T. Phan, and L. Lacroix. Following g-quartet formation by uv-spectroscopy. *FEBS Lett*, 435(1):74–8, 1998.

[262] A. T. Phan, M. Gueron, and J. L. Leroy. Investigation of unusual dna motifs. *Methods Enzymol*, 338:341–71, 2001.

[263] W. Lee, M. Tonelli, and J. L. Markley. Nmrfam-sparky: enhanced software for biomolecular nmr spectroscopy. *Bioinformatics*, 31(8):1325–7, 2015.

[264] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. M. Clore. The xplor-nih nmr molecular structure determination package. *J Magn Reson*, 160(1):65–73, 2003.

[265] LLC Schrodinger. The pymol molecular graphics system, version 1.8. 2015.

[266] GA Davies and JR Stokes. Thin film and high shear rheology of multiphase complex fluids. *Journal of Non-newtonian fluid mechanics*, 148(1-3):73–87, 2008.

[267] Georgina A Davies and Jason R Stokes. On the gap error in parallel plate rheometry that arises from the presence of air when zeroing the gap. *Journal of Rheology*, 49(4):919–922, 2005.

[268] Olena Kravchuk and Jason R Stokes. Review of algorithms for estimating the gap error correction in narrow gap parallel plate rheology. *Journal of Rheology*, 57(2):365–375, 2013.

[269] Pei Yun Lee, John Costumbrado, Chih-Yuan Hsu, and Yong Hoon Kim. Agarose gel electrophoresis for the separation of dna fragments. *Journal of visualized experiments: JoVE*, (62), 2012.

[270] Frederick M Ausubel. *Short protocols in molecular biology: a compendium of methods from current protocols in molecular biology*. 2002.