NANYANG TECHNOLOGICAL UNIVERSITY

BIOPHYSICAL STUDY OF G-QUADRUPLEX STRUCTURES IN SOLUTION AND G-WIRE SUPERAMOLECULAR ASSEMBLY ON GRAPHENE

HU LANYING

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2012

BIOPHYSICAL STUDY OF G-QUADRUPLEX STRUCTURES IN SOLUTION AND G-WIRE SUPERAMOLECULAR ASSEMBLY ON GRAPHENE

HU LANYING

2012

BIOPHYSICAL STUDY OF G-QUADRUPLEX STRUCTURES IN SOLUTION AND G-WIRE SUPERAMOLECULAR ASSEMBLY ON GRAPHENE

HU LANYING

School of Physical and Mathematical Sciences

A thesis submitted to the Nanyang Technological University

in fulfillment of the requirement for the degree of

Doctor of Philosophy

2012

ABSTRACT

Guanine-rich nucleic acid sequences have a high propensity to adopt non-B DNA secondary structures like G-quadruplexes. Sequences that could potentially form G-quadruplexes are widespread throughout the genome and found to be more prominent in biologically critical regions. They may play biological significant roles in telomere maintenance, as well as regulation of gene transcription, replication and recombination.

G-quadruplex DNA is highly polymorphic. The study of G-quadruplex structures has attracted intense interests in the field of potential therapeutic targeting in human cancers. Structural uniqueness of G-quadruplex DNA can serve as specific recognition site for G-quadruplex interactive compounds. In Chapter 3, the structure of four-repeat *Giardia* telomeric sequence d[TAGGG(TAGGG)$_3$], which differs from the human counterpart d[TAGGG(TTAGGG)$_3$] only by one T deletion within the non-G linker in each repeat, was solved by NMR to explore the effect of loop length and sequence on the folding topology of G-quadruplexes. Two different intramolecular G-quadruplexes were found to coexist and interconvert in K^+ solution. Recurrence of several structural elements in the observed structures suggests a "cut and paste" principle for the design and prediction of G-quadruplex topologies, for which different elements could be extracted from one G-quadruplex and inserted into another.

The unique properties of G-quadruplex DNA compared to duplex DNA make them amenable for development of nanomaterials. In Chapter 4, Self-assembly of supramolecular G-wires, grown from the oligonucleotide d(GGGGTTGGGG), on graphene sheets were investigated. Atomic force microscope (AFM) and micro-Raman mappin results demonstrate for the first time that G-wires are well-ordered and preferentially oriented along the armchair direction of graphene. Such assembly could be further exploited for the development of graphene-based molecular device and biosensor. It can be envisioned that the development of the understanding of the thermodynamic properties and structural features of G-quadruplex DNA will endow new capabilities for structural nucleic acid nanotechnology.

ACKNOWLEDGEMENTS

The approval for my enrollment to the Division of Physics and Applied Physics in 2008 opened the door to a new world for me. During the long journey in the graduate program, I'm truly blessed to meet many great people along the way.

I would like to express my appreciation and gratitude to my advisor PHAN Anh Tuân for taking me under his supervision and for his guidance throughout the course of my Ph.D. research. His enthusiasm to science, academic sincerity and demand for perfection, have set a perfect role model for me to look up to. His efforts to get everyone involved in the laboratory group discussions and take responsibility in lab helped me understand the dynamics of group work. In spite of busy schedules and fluid timelines, he shows great patience for his students. His office door is always open and he is always welcome to develop, discuss, and debate ideas. It has always been his priority that I understand the principles behind experiments rather than just data-producing. Through his encouragement and tutoring, I found myself improving in numerous aspects, including, scientific writing and research aptitude. I truly value the opportunity to work on challenging projects and the freedom to explore my own ideas. The distinguished lessons and experiences have been invaluable and will benefit me lifelong.

I am thankful to the current and former group members of the Biophysics Lab, who have helped and enlightened me in different ways. Special thanks are deserved for Kah Wai Lim for his great help in structure computations of the *Giadia* project. My devout appreciation goes to Serge Bouaziz for coaching me structure calculation and NMR spectroscopy when he was the visiting scientist in our group. I also appreciate the efforts of Chun Li and Amrane Samir for their participation in the early stage of the project, Ngoc Quang Do for his assistance with the gel electrophoresis experiments. Heartfelt thanks to Herry Martadinata and Brahim Heddi, who were always there to lend a hand and being patient with me in spite of my annoying petty questions, thank you for your support. It has also been a pleasure to work with the undergraduates, Ci Ji Lim, Keyi Fang and Huiheng Siew. I wish the best luck to everyone in the Biophysics Lab.

Words could never contrive an accurate description to express my love and appreciation for my parents. I am eternally indebted to them for their unconditional love, endless support and wholehearted dedication over the years. I cannot forget they fight to hold back tears to be supportive every time I leave home. I miss you so much!

Living in a different country away from home is hard, but I'm very blissful to have my husband Bin Yan stood by me through thick and thin. He has always been there for me to cheer my spirits in any disparaging moment and make much more gratifying the good ones. He was also very helpful for my research. Many work-related

discussions we had on our walk-back home deepen my understanding of many physics topics. Our collaborative G-wire project is the best proof of this. I owe you immense gratitude for the time and effort you put forth on my behalf.

For all the unintentionally forgotten ones that should be mentioned here, thank you for your support!

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1 Introduction to DNA Structure and Nanotechnology

1.1 Deoxyribonucleic Acid (DNA)

Nucleic acids are biopolymers that contain genetic information[1]. Structurally speaking, nucleic acids are polymers made up of monomer units known as nucleotides. Each monomer unit, or nucleotide, has three components: nucleobase, pentose sugar and phosphate group. The nucleobases can be classified into two groups, purine (adenine or guanine) and pyrimidine (thymine or cytosine). The nucleobase is attached to the 1'- carbon of a sugar, which joined at the hydroxyl groups on the 5'- and 3'- carbons to the phosphate groups to form the backbone (**Figure 1.1**).

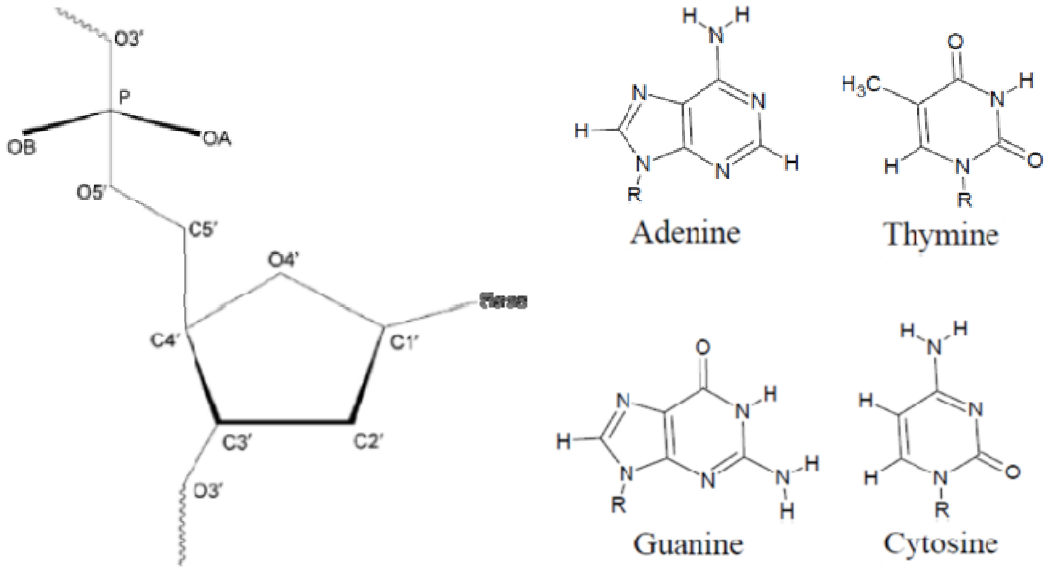

Figure 1.1 Structure of a repeating unit nucleotide in nucleic acids and the four nuleobases of DNA. This figure is adapted from Neidle's book[2].

In 1953, James Watson and Francis Crick proposed the double helix structure of DNA[3]. The Watson-Crick structure of so-called B-form DNA is a right-handed double helix, which is built up of Watson-Crick base pairs between two complementary strands in an anti-parallel orientation. The two strands are held together by hydrogen bonds in a specific pattern where adenine binds to thymine and guanine to cytosine through Watson-Crick edges (**Figure 1.2**). In addition to A·T and G·C Watson-Crick base pairs, different possible base pairing arrangements, namely non-canonical base pairs, play a prominent role towards the structural diversity of higher order DNA[4]. Typical examples of non-canonical structures are triplex[5] and G-quadruplex[6] DNA motifs with Hoogsteen base pairing.

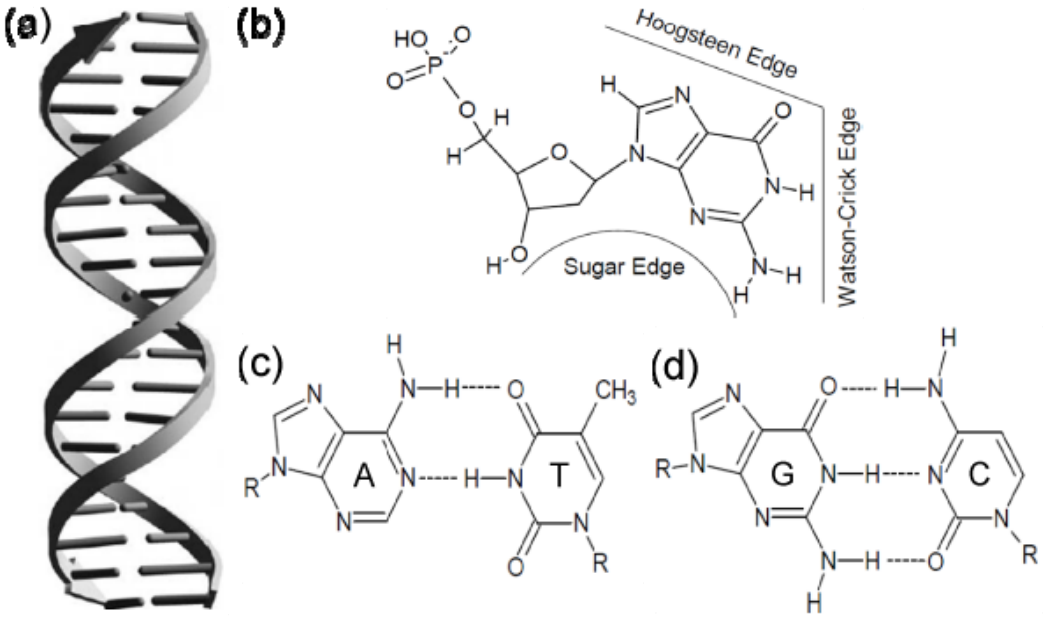

Figure 1.2 (a) Schematic view of canonical B-DNA. The backbone is represented by ribbon. (b) Three edges of deoxyribonucleotide for hydrogen bonding interactions. (c) A•T and (d) G•C Watson-Crick base pairs. The deoxyribose is represented with a letter R. This figure is adapted from Neidle's book[2] and Leontis's paper[7].

1.2 Structural Diversity of Biologically Relevant G-quadruplex DNA

Guanine-rich nucleic acid sequences have a high propensity to adopt non-B DNA secondary structures like G-quadruplex[6, 8], which is composed of multiple stacked guanine tetrads[9]. Guanine tetrad (G-quartet) is a co-planar arrangement of four guanines where hydrogen bonds are formed between the Watson-Crick edge and the Hoogsteen edge of two adjacent guanines. It involves a total of eight hydrogen bonds, which follow the same directions from hydrogen bond donors (N1 and N2) to hydrogen bond acceptors (N7 and O6) (**Figure 1.3**).

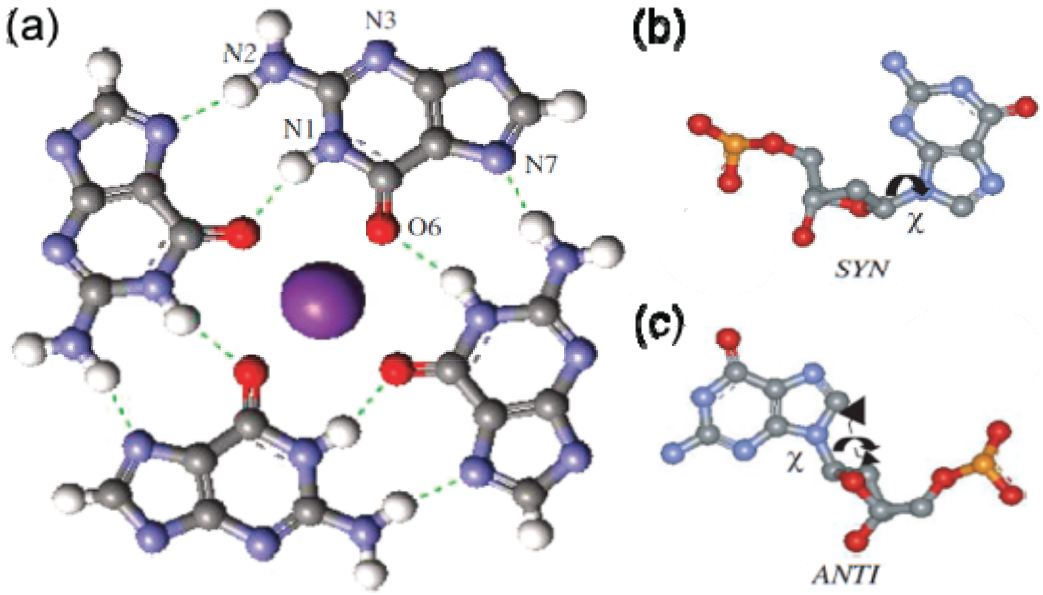

Figure 1.3 (a) Schematic drawing of a G-quartet with cyclic array of four guanines linked by Hoogsteen hydrogen bonds, including a cation at its core. Guanine nucleobase oriented in (b) syn and (c) anti conformation. This figure is adapted from Neidle's book[10].

Sequences that could potentially form G-quadruplexes are widespread throughout the genome. Genome-wide bioinformatic analyses suggest that over 376000 Putative Quadruplex Sequences (PQS) distribute throughout the human genome, although probably not all of these form in vivo[11]. Specially, these sequences have been found to be more prominent in biologically critical regions, such as telomeres[12], promoters[13], immunoglobulin switch regions[14] and recombination hot spots[15]. If G-quadruplexes form in vivo, even in a transient manner, they may play a role in telomere maintenance, as well as regulation of gene transcription, replication and recombination[16] (**Figure 1.4**). Some of the salient aspects of formation of G-quadruplex structures in these biological significant regions are illustrated below.

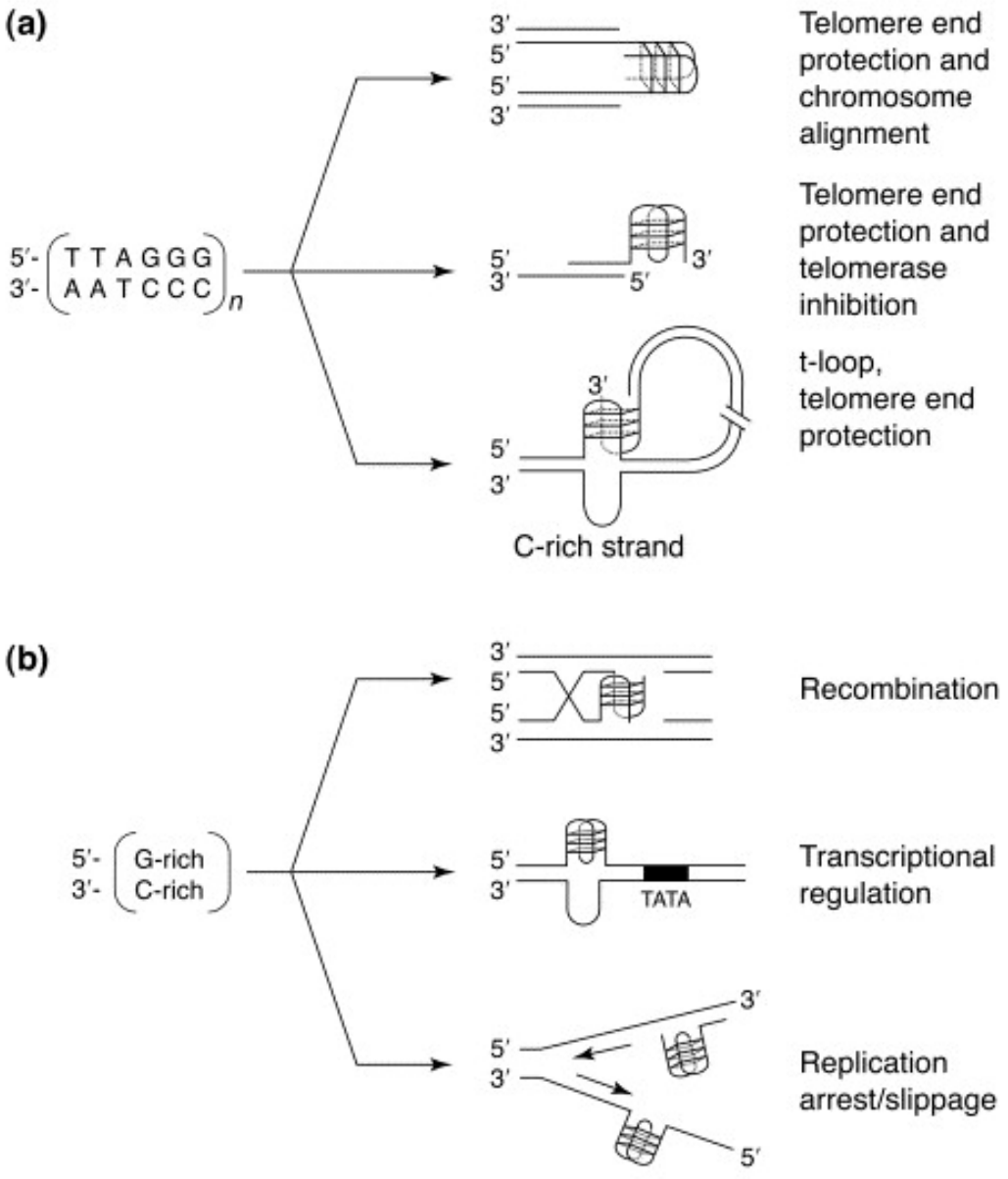

Figure 1.4 Schematic diagrams of G-quadruplex structures in various cellular events. This figure is adapted from Han's paper[16].

G-quadruplex forming motifs have been found in the promoter regions of many genes that play important roles in carcinogenesis, including myelocytomatosis viral oncogene homolog (c-MYC) [17], vascular endothelial growth factor A (VEGF-A)[18], B-cell lymphoma 2 (BCL-2)[19], kirsten rat sarcoma viral oncogene homologue (KRAS)[20], platelet-derived growth factor A (PDGF-A)[21], human telomerase reverse transcriptase (hTERT)[22], and retinoblastoma protein (pRb)[23] (**Figure 1.5**). For example, in the case of the c-MYC proto-oncogene, a single nucleotide mutation that disrupts the formation of G-quadruplex, results in three-fold increase in c-MYC expression[17a]. Over-expression of c-MYC is a hallmark of cancers[24]. In contrast, G-quadruplex-interacting agents that bind to and stabilize the G-quadruplex structure can suppress c-MYC expression[17a]. It has been suggested that G-quadruplex acts as a silencer element for the c-MYC oncogene expression[17].

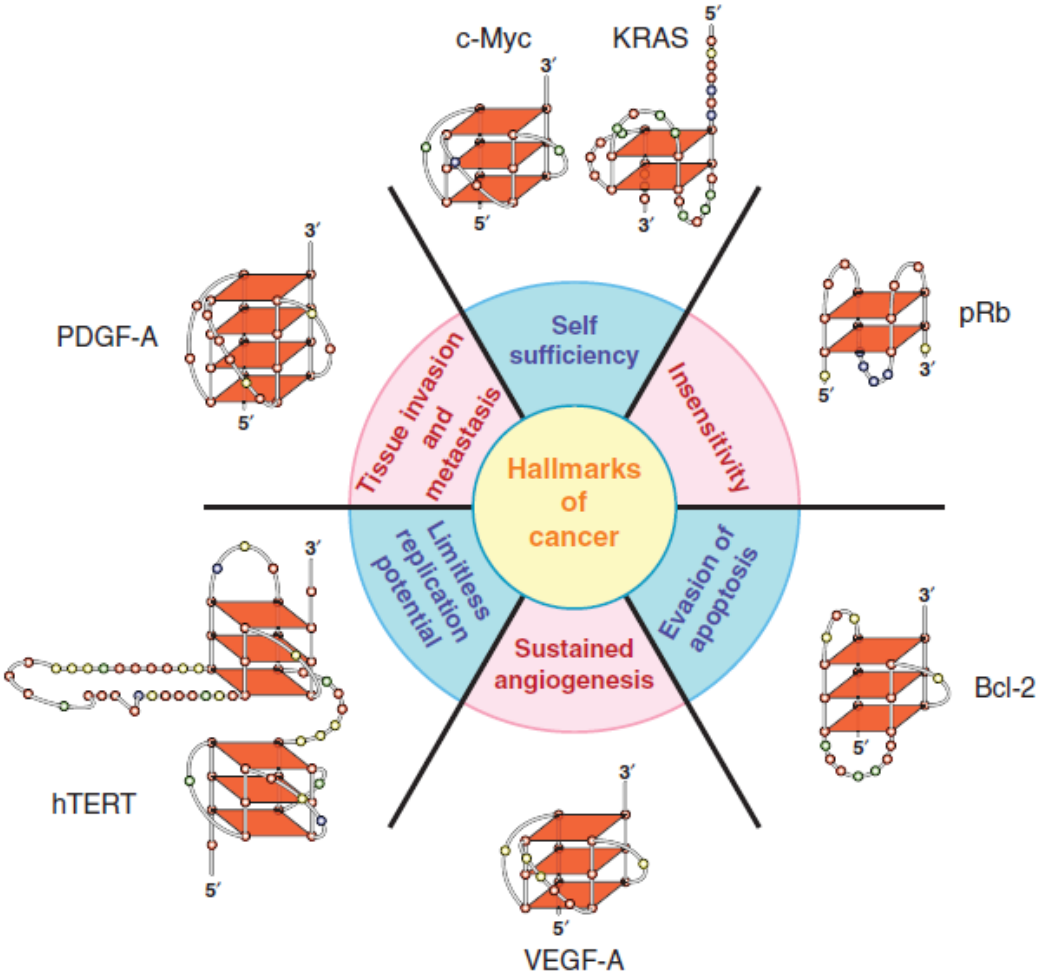

Figure 1.5 G-quadruplex DNAs formed in the promoter regions of oncogenes, shown with the associated hallmarks of cancer. This figure is adapted from Han's paper [17d].

Telomeres are protein-DNA complexes that cap the termini of linear eukaryotic chromosomes. They are essential for protecting chromosomes from end-to-end fusion and nucleolytic degradation, and they serve as substrates for elongating the G-rich stand of telomeric DNA by ribonucleoprotein telomerase[25]. In somatic cells, telomerase is inactive, and telomeres shorten progressively with each cell division and may act as a mitotic clock that eventually triggers cell senescence[26]. In contrast, telomerase is highly expressed in most tumor cells, resulting in stabilized telomere length[27]. Telomere maintenance has been shown to be crucial for immortality and proliferation of most cancer cells[27].

Telomeric DNAs are composed of non-coding tandem repeats of short sequences that have complementary guanine- and cytosine-rich strands. The guanine-rich strand orients in a 5' to 3' direction toward the chromosome end, which protrudes from the telomeric DNA duplex to form a single-stranded G-overhang. The 3' G-rich overhang could potentially fold into G-quadruplex, which acts as a structural inhibitor of telomerase by blocking its access to telomere[28]. In human cells, the telomeric repeat sequence of the guanine-rich strand is $(TTAGGG)_n$[29], which is conserved among vertebrates[30]. Human telomeric DNA sequences $d(TTAGGG)_n$ have been reported to form various G-quadruplex structures in vitro[6c, 8a, 31]. The first description of the structural information on human telomeric sequence was observed for the tetrameric parallel-stranded G-quadruplex formed by single-repeat d(TTAGGG) in K^+ solution[32] (**Figure 1.6a**). For the two-repeat human telomeric sequence d(TAGGGTTAGGGT),

a dimeric parallel-stranded G-quadruplex was observed in K^+-stabilized crystal[33](**Figure 1.6b**). A similar dimeric parallel and another dimeric antiparallel G-quadruplexes coexist and interconvert in K^+-solution[34] (**Figure 1.6c**). An asymmetric dimeric so-called (3+1) G-quadruplex was formed by the three-repeat human telomeric sequence d(GGGTTAGGGTTAGGGT) in Na^+ solution. In this G-quadruplex assembly, all three G-tracts of one strand associates with one G-tract of the other strand[35] (**Figure 1.6d**). The solution structure of four-repeat human telomeric sequence d[AGGG(TTAGGG)$_3$] in Na^+-containing solution was characterized to be an intramolecular antiparallel basket-type quadruplex[36] (**Figure 1.6e**). In contrast, the same sequence folds intramolecularly into a completely different parallel propeller-type G-quadruplex structure in K^+- containing crystal[33] (**Figure 1.6f**). Other G-quadruplex conformations have been deciphered or proposed for human telomeric DNA[31].

The conformational diversity available to human telomeric sequences confirms that G-quadruplex DNA is highly polymorphic in nature. The folding topology of G-quadruplex DNA is affected by many factors, including the strand concentration and the cation identity, and variations in nucleotide sequence and composition.

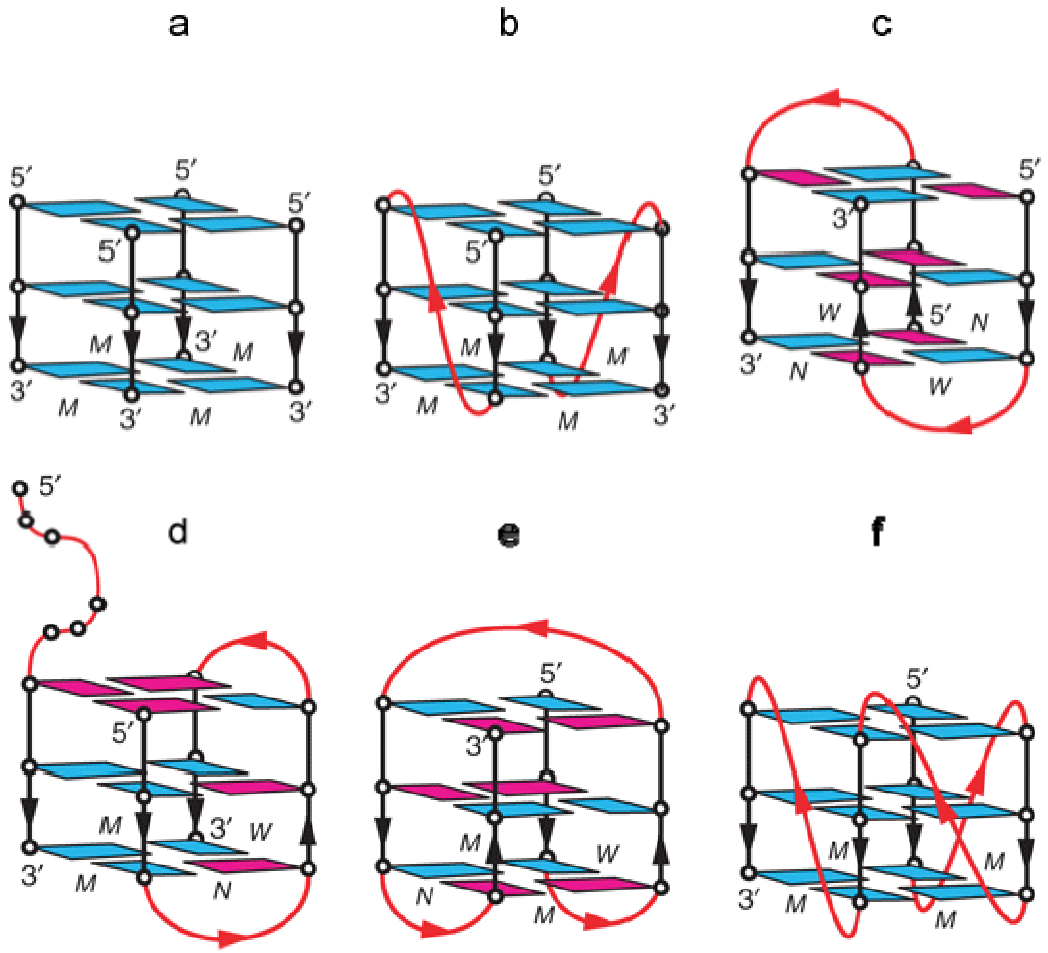

Figure 1.6 Schematic structures of G-quadruplexes formed by human telomeric sequences. Single-repeat human telomeric sequence d(TTAGGGG) forms (a) tetrameric parallel-stranded G-quadruplex structure in K^+ solution[32]. Two-repeat human telomeric sequence d(TAGGGTTAGGGT) adopts (b) dimeric parallel-stranded G-quadruplex structure in K^+-stabilized crystal[33] and (c) dimeric anti-parallel G-quadruplex structure in K^+ solution[34]. Three-repeat human telomeric sequence d(GGGTTAGGGTTAGGGT) exhibits (d) asymmetric dimeric (3+1) G-quadruplex structure in Na^+ solution[35]. Four-repeat human telomere sequence d[AGGG(TTAGGG)$_3$] folds into (e) intramolecular antiparallel basket-type G-quadruplex structure in Na^+ solution[36] and (f) parallel propeller-type G-quadruplex structure in K^+- containing crystal[33]. The schematic structures are color-coded as follows: *anti* guanines, cyan; *syn* guanines, magenta; loops, red. Groove types are indicated as narrow (*N*), medium (*M*) and wide (*W*). This figure is adapted from Phan's paper[31].

G-quadruplex DNAs vary in several key characteristics, namely strand stoichiometry, strand polarity, and loop connectivity.

Based on the organization of G-rich oligodeoxyribonucleic acid strands, G-quadruplex structure can be either intramolecular or intermolecular G-quadruplex[6a, 8a]. Intramolecular G-quadruplex DNA is comprised of a single strand of DNA, which is folded on itself to provide the four strands of the G-quadruplex scaffold (**Figure 1.6e and f**). Intermolecular G-quadruplex DNA is assembled by two (dimeric) (**Figure 1.6b, c, and d**) or four (tetrameric) (**Figure 1.6a**) separate strands.

Not only the strand stoichiometry presents important in characterizing a G-quadruplex, but the relative arrangement of adjacent backbones can also give rise to structural polymorphism. Irrespective of the number of strands that constitute a G-quadruplex, there are four possible variations in strand orientation. They can be all parallel (**Figure 1.6a, b, and f)**, three parallel and one antiparallel (3+1) (**Figure 1.6d)**, adjacent parallel (**Figure 1.6e)**, or alternating antiparallel (**Figure 1.6c)**.

Variations in strand polarity also affect the location of the loops connecting the strands. All loops that have been identified fall into one of these three main types: edgewise, diagonal, or double chain reversal loop[6c, d]. For antiparallel strands, edgewise loops joint two adjacent strands (**Figure 1.7a)**, diagonal loops connect the opposite G-strands (**Figure 1.7b)**. While adjacent parallel strands require a loop to connect the

bottom G-tetrad with the top one, leading to the formation of a double chain reversal loop (**Figure 1.7c)**.

G-quadruplex DNAs display diversity in their strand stoichiometry, strand polarity, and loop connectivity. It gives uniqueness to the structure and can serve as specific recognition site for the G-quadruplex interactive compound[37].

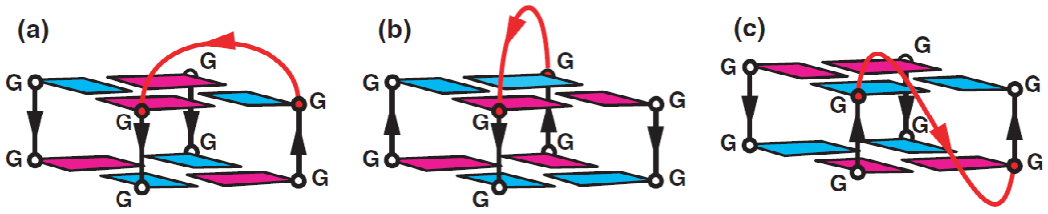

Figure 1.7 Schematic illustrating of (a) edge-wise, (b) diagonal, (c) double chain reversal loop, which connects individual strands of G-quadruplexes containing two stacked G-tetrads. The schematic structures are color-coded as follows: anti guanines, cyan; syn guanines, magenta; loops, red. This figure is adapted from Patel's paper[6d].

1.3 DNA Nanotechnology

DNA first was considered as the "blueprint of life"[1], now DNA has proven to be a powerful "construction brick" and vastly expanded its potential uses in nanotechnology and biotechnology applications[38].

DNA nanotechnology was pioneered by Nadrian C. Seeman in 1980s[39], and became a rapidly evolving research field. Over the past 20 years, DNA has been exploited in the fields of construction of nanostructures[40], mechanical[41] and electronic nanodevices[42], molecular detection[43] and computation[44], and DNA templated synthesis in chemistry[45] and biology[46].

Inspired by naturally occurring branched DNA molecules, Seeman and co-workers developed an artificial branched junction[39, 40b] (**Figure 1.8**). It is topologically equivalent to Holliday junction structure[47] (**Figure 1.9**) and a basic building block for complex structures with nanoscale features. Unlike natural Holliday junction, synthetic DNA junctions lack the sequence symmetry such that branch migration can be circumvented and thus make the junctions immobile and stable. In order to construct desired structures, programmed complementary sticky ends at the periphery of junctions are widely used to link DNA junctions.

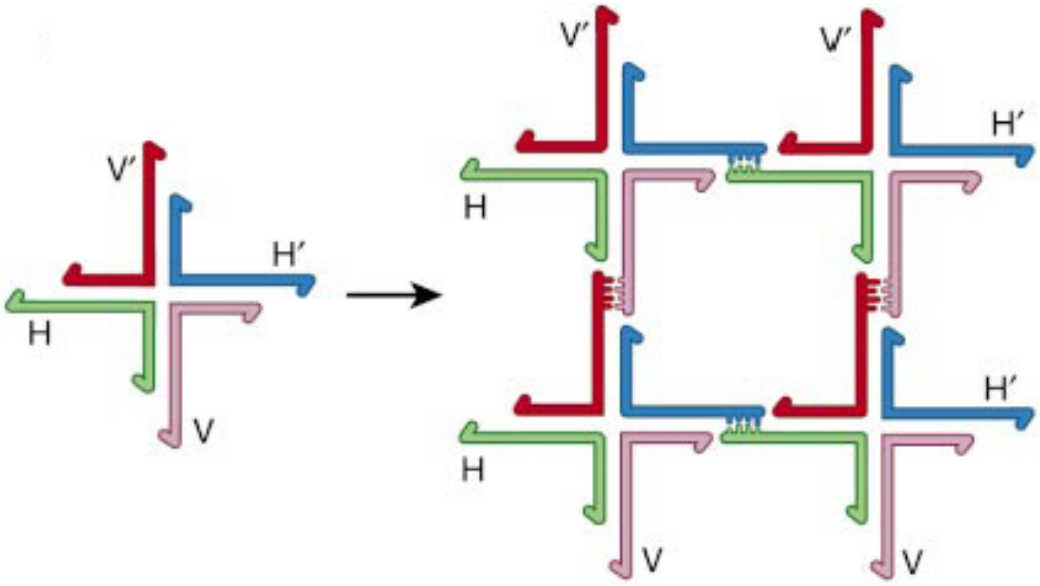

Figure 1.8 Schematic drawing of a four-arm junction with sticky ends. Four of the monomeric junctions can assemble into square-like unit with sticky ends on the outside, further allowing the addition of more monomers to produce two-dimensional structures. This figure is adapted from Seeman's paper[40b].

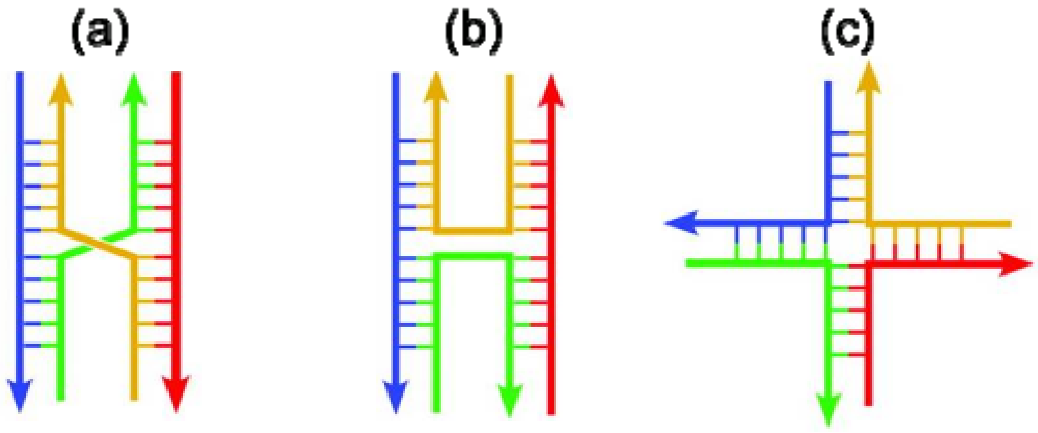

Figure 1.9 Schematic structures of Holliday Junction. Possible configurations for the Holliday junction: parallel stacked-X (a), antiparallel stacked-X (b), and open planar configuration (c). This figure is adapted from Hays's paper[47].

Although individual four-arm junction is stable, it is too flexible to hold the helical domains in the same plane in solution. Such conformational flexibility will prevent junction from further assembling into larger arrangements. The rigid double crossover (DX) DNA motif was first proposed by Seeman's group[40a, 48] (**Figure 1.10a**). DX molecules are made by joining two four-arm junctions together by two double helical arms. They can be used to produce two-dimensional arrays by sticky ends cohesion (**Figure 1.10b**). Two-dimesional DNA arrays can serve as templates for the organization of other functional materials with nanoscale precison. For example, double crossover molecules are assembled into two-dimensional DNA arrays, modified with alternating human R-thrombin and platelet derived growth factor (PDGF) binding aptamer to direct thrombin and PDGF into periodic nanoarrays[46c] (**Figure 1.11**). Following this direction, a variety of other rigid DNA motif have been designed and constructed into two-dimensional array, such as paranemic crossover[49], helix bundle[50], and branched DNA motif[46a].

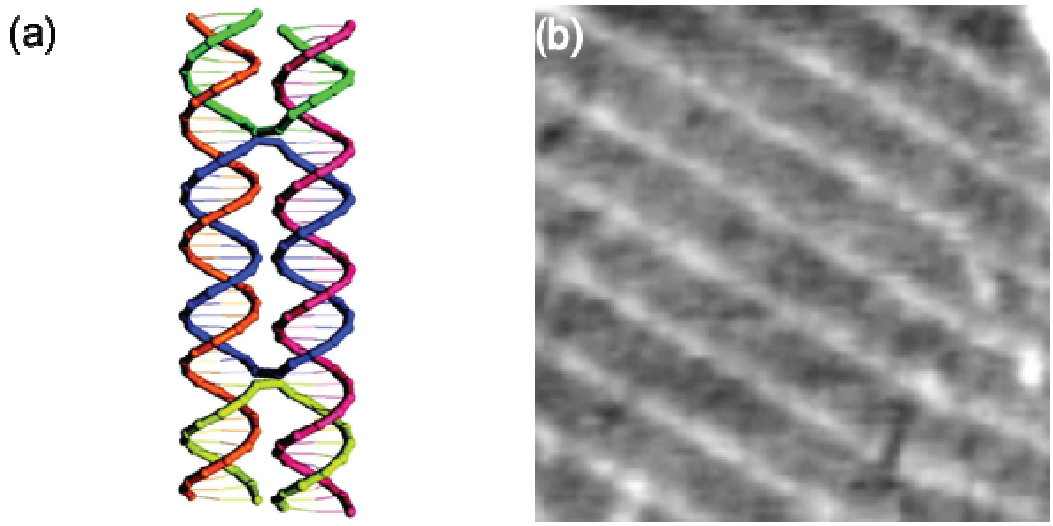

Figure 1.10 (a) Diagram of the DNA double crossover complex formed through joining two four-arm junctions together by two double helical arms. (b) AFM image of two-dimensional periodic stripe array constructed by double crossover DNA molecules. This figure is adapted from Winfree's paper[40a].

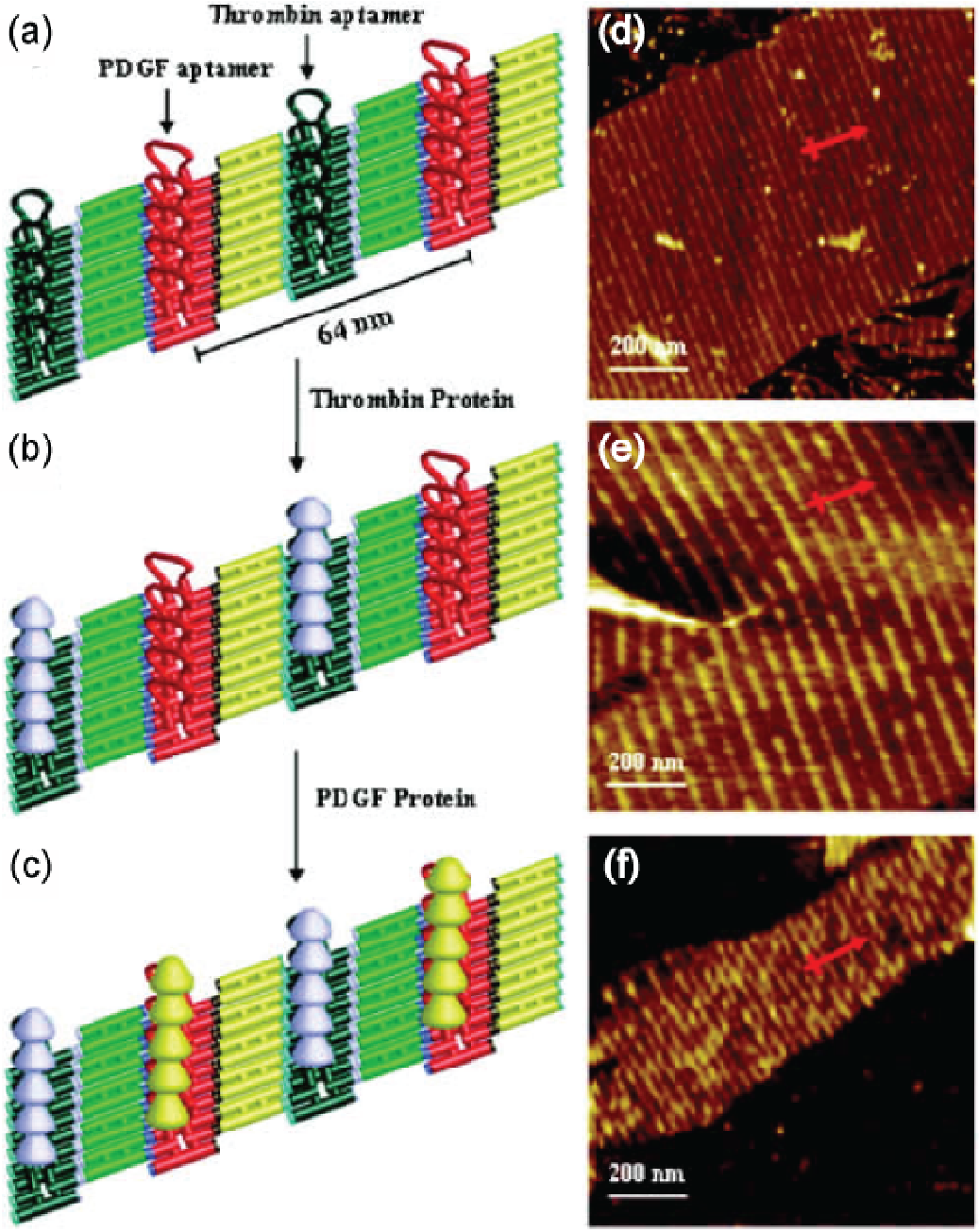

Figure 1.11 DNA templated assembly of multiprotein. Schematic drawing of the formation process of periodic two-dimensional alternating multiprotein nanoarrays templated by DNA nanoarrays (a-c). The corresponding AFM image of the nanoarray is shown on the right (d-f). This figure is adapted from Chhabra's paper[46c].

Rothemund ingeniously developed an alternative method, called "scaffolded DNA origami", to generate arbitrary DNA shapes and patterns[51] (**Figure 1.12**). The "scaffolded DNA origami" strategy is an approach for folding a long linear strand DNA with known sequence into an addressable shape which can display desired patterns. Many short-stranded DNA, complementary to multiple parts of the long scaffolding strand, are served as "staple strands".

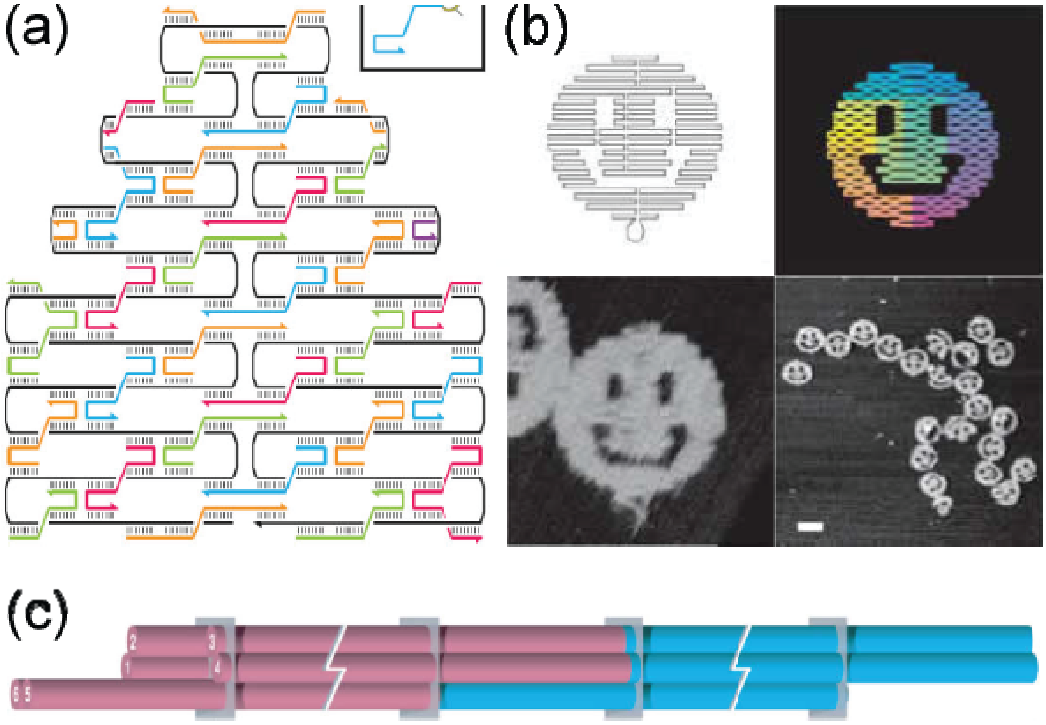

Figure 1.12 Design scheme and examples of scaffolded DNA origami. (a) Design layout of a two-dimensional origami. The black strand represents the long linear DNA strand that runs through the whole area of the structure, which is held together by the staple strands through binding to the multiple parts in the scaffold. (b) Selected one example of two-dimensional origami. The upper panel illustrates the design of a smiley face. AFM images corresponding to the design are shown in the lower panel. (c) Three-dimensional cartoon view of DNA nanotube synthesized by rolling a DNA origami sheet along its helical axis. This figure is adapted from Rothemund's paper[51] and Douglas's paper[52].

The construction of well-defined three-dimensional DNA objects could also be achieved by rationally designed self-assembly, which has been heavily pursued in recent years. Seeman and coworkers did the pioneer work on the construction of three-dimensional DNA objects by stepwise synthesis via ligation to create catenated polyhedra, such as a cube[53] or a truncated octahedron[54] (**Figure 1.13a, b**). Although the synthesis process is laborious, the yields of the resulting DNA polyhedra were low due to the multiple steps of ligation. Turberfield and coworkers invented a facile methodology to assemble three-dimensional DNA nanostructures. An elegant example is the one-pot assembly of the cage-like DNA tetrahedron[55] (**Figure 1.13c**), which is assembled within few seconds and has near quantitative yield. Another different design strategy of engineering three-dimensional DNA nanostructure is based on DNA origami folding technique. Shih et al. reported the design and synthesis of tube-like DNA six-helix bundles created by "rolling" a DNA origami array (**Figure 1.12c**). These nanotubes can form liquid crystals, which were applied to be a stable medium and induce the alignment of membrane proteins in order to facilitate their solution Nuclear magnetic resonance (NMR) structural determination[52]. It is notable that this study provides a powerful way to use DNA array as an organization framework for protein structure determination. It can be anticipated that more examples of such pratical application of DNA origami in both research and nanotechnology filed, will come out in the near future.

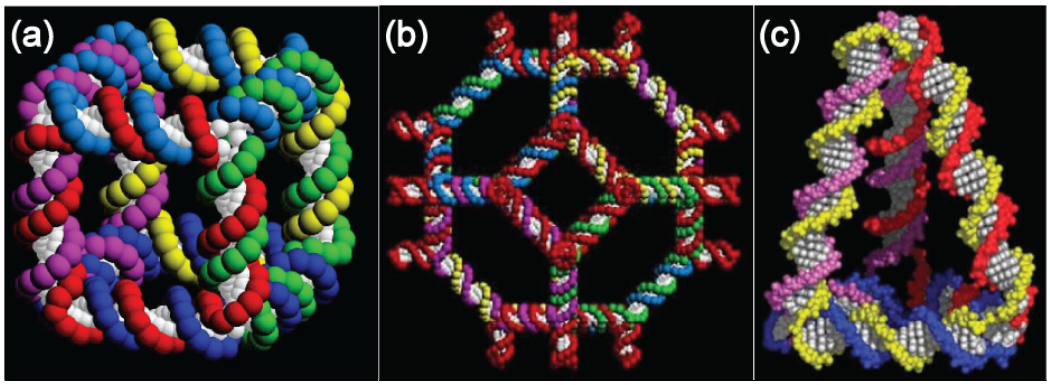

Figure 1.13 Model of self-assembled three-dimensional DNA architectures: (a) cube; (b) truncated octahedron; (c) tetrahedron. This figure is adapted from Chen's[53], Zhang's[54] and Goodman's[55] papers.

1.4 G-quadruplex as Building Blocks for Nanomaterials

By and large, structural DNA nanotechnology utilizes the bottom-up assembly of Watson-Crick base pairing to construct well-defined architectures[56]. A developing understanding of the non-Watson-Crick base-paired interactions should ultimately endow structural DNA nanotechnology with new capabilities.

Compared to Watson-Crick base paired duplex DNA, Hoogsteen base paired G-quadruplex DNA has higher resistance to degradation by nuclease, enhanced rigidity and thermal stability.[57] In addition, the thermodynamic property and structure of G-quadruplex DNA can be leveraged through the control of DNA sequence and external conditions. All of these features make G-quadruplex DNA amenable for development of nanomaterials[58].

Research groups of Mergny[59] and Tan[60] first conceptualize and validate the controlled transitions between compact G-quadruplex structure and its linear duplex form as the basis of nanoswitch. This finding inspired the development of a number of G-quadruplex DNA based molecular devices. For example, Simmel and co-workers constructed a molecular machine emulated the physiological function of the thrombin-binding aptamer in the cell[61]. In this artificial system, thrombin-binding aptamer (AP) can bind and release the thrombin (TB) protein in a controlled way. The

single-stranded overhang, concatenated at the 5' end of the G-quadurplex forming region of thrombin-binding aptamer, was devised to facilitate the release of thrombin. The thrombin release was triggered by the introduction of an "opening" strand Q, which is complementary to the overhang and a portion of the aptamer sequence. The resulting duplex (Q-AP) contains a toehold so that once a "removal" strand R, fully complementary to Q, is introduced the thrombin can be bound once again (**Figure 1.14**). In principle, such molecular devices can be constructed to bind to a variety of proteins and ligands only if they have a specific aptamer.

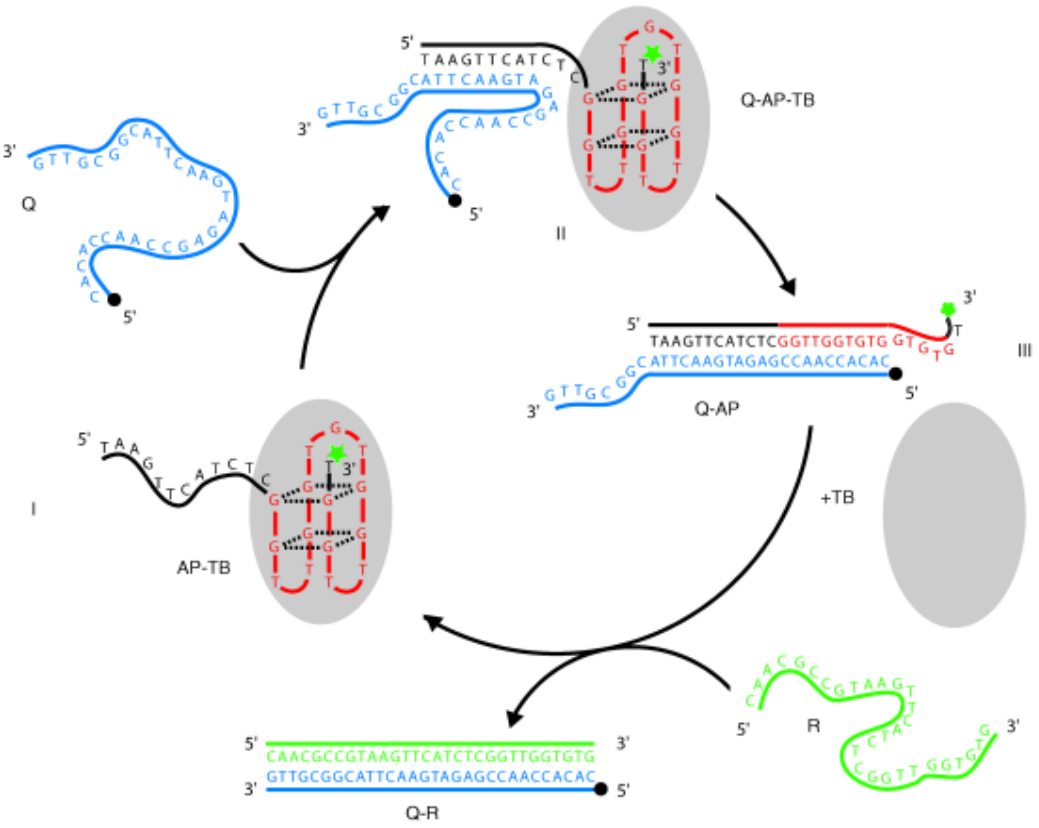

Figure 1.14 Schematic representation of the operation cycle of the aptamer-based molecular device that binds and releases thrombin. This figure is adapted from Dittmer's paper[61].

Since G-quadruplex formation is greatly affected by the external conditions, several G-quadruplex nanodevices have been developed by exploiting the conformational transition of DNA. Fahlman et al. have designed a "pinched duplex" nanodevice[62], which is constructed by a duplex DNA with intervening consecutive G-G mismatches. Upon addition of potassium or strontium ions, which are potent G-quadruplex conformation stabilizer, the G-G mismatches could fold into an intramolecular G-quadruplex and bend the overall duplex into a "pinched" conformation. This conformational transition can be reversed by sequestering the metal cations with effective chelator such as ethylenediaminetetraacetic acid (EDTA) (**Figure 1.15**). This unusual "pinched duplex" conformational transition could be used to introduce contractile elements into DNA architecture[62a] as well as construct a DNA-based electronic nanoswitch[63]. Another ion-sensitive G-quadruplex based nanodevice was developed by Sugimoto and co-workers[64]. In this work, a coordination unit for divalent cations, 2,2′-bipyridine, replaced thymines in the $d(G_4T_4G_4)$ to link two G-tracts. Upon binding to Ni^{2+} ion, bipyrimidine-modified antiparallel G-quadruplexes would transform into parallel G-quadruplexes, which can further associate intermolecularly into one-dimensional "G-wires". Reversible reaction could be activated with the addition of divalent ion chelator EDTA (**Figure 1.16**). The term "G-wires" was coined by Marsh et al. to describe the linear supermolecular structures formed by self-assembly of the telomeric oligonucleotide d(GGGGTTGGGG) as a consequence of slipped tetraplex association[65].

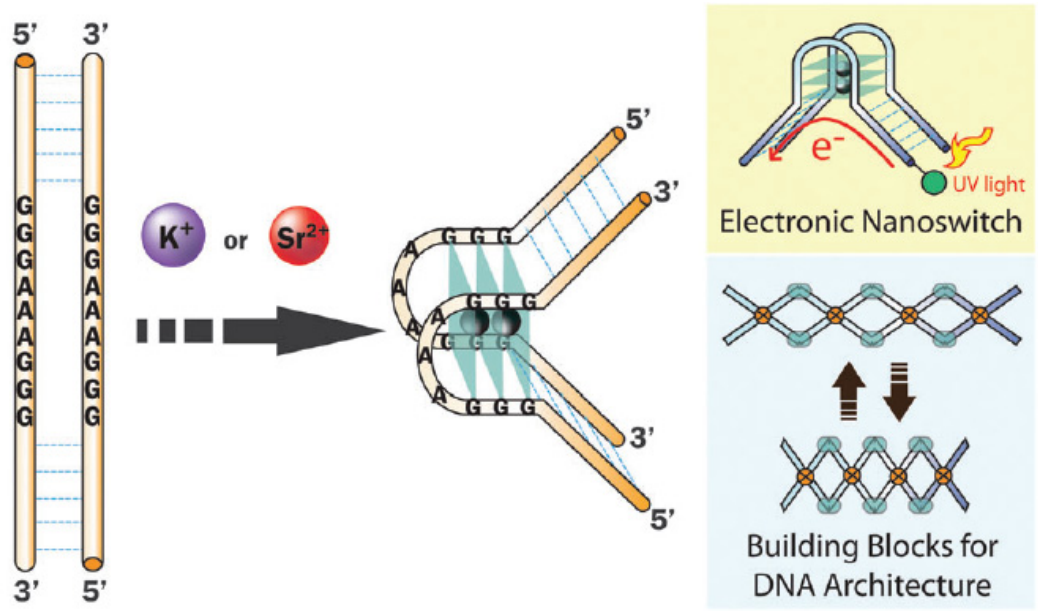

Figure 1.15 Schematic illustration of "pinched duplex" formation triggered by addition of the potassium or strontium ions. Two potential applications exploiting the conformational transition of "pinched duplex" are shown on the right. This figure is adapted from Choi's paper[62b].

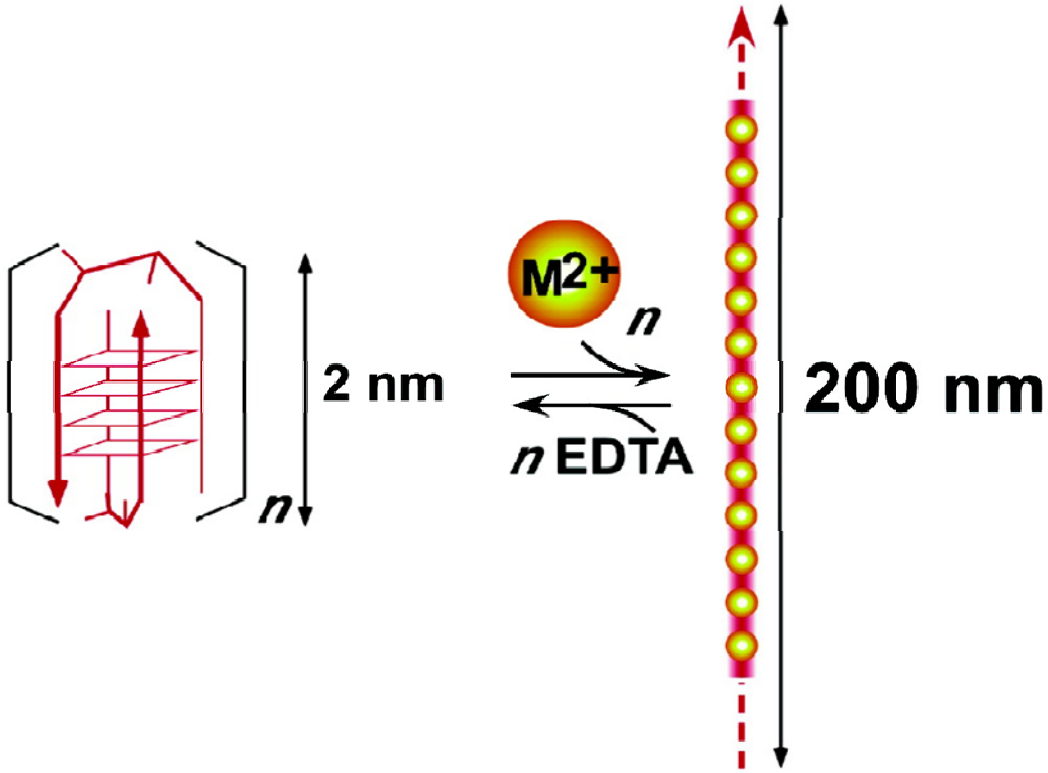

Figure 1.16 Schematic illustration of bipyridine-modified antiparallel G-quadruplex can be switched into a linear "G-wire" triggered by divalent cations and the system can be reset with EDTA. This figure is adapted from Miyoshi's paper[64]

It can be envisioned that understanding of G-quadruplexes will lead to a better exploitation of these structures in many fields, such as biomedicine, superamolecular chemistry and nanotechnology.

Chapter 2 Overview of Characterization Methods

A myriad of experimental techniques have been employed to provide insight into the formation and folding topology of G-quadruplexes. For example, molecular weight of G-quadruplexes can be estimated by gel electrophoresis. Circular dichroism (CD), on the other hand, is very useful in identifying the strand topology of a G-quadruplex. It is important to note, however, that such less precise and less definite techniques are insufficient to determine the G-quadruplex structure. Nuclear Magnetic Resonance (NMR) and X-ray crystallography are capable of giving a wealth of information for detailed structural analysis of G-quadruplexes in atomic-resolution.

2.1 Nuclear Magnetic Resonance

The underlying principle of Nuclear Magnetic Resonance (NMR) is the detection of the response of NMR active nuclei (such as ^{1}H or ^{13}C) in a constant magnetic field and under an electromagnetic (EM) pulse. The magnetic nuclear spins will align with the constant magnetic field. This alignment of the nuclear spins will be perturbed by employing an electro-magnetic, usually radio frequency pulse, which cause the nuclei to absorb energy at a specific frequency (**Figure 2.1a**). The resonant frequency of a nucleus, termed chemical shift, is dependent on the shielding effect of the electronic

environment next to that nucleus. Specific nuclear magnetic resonances signatures (i.e. imino proton chemical shifts) can be used to differentiate between G-quadruplex and Watson-Crick double helix[57a, 66]. The Watson-Crick base-paired imino protons are observed between 12-14 ppm, and the imino protons involved in the G-quadruplex Hoogsteen base pairing are observed in the region of 10-12 ppm (**Figure 2.1b**). The nuclear spins can interact with each other via cross-relaxation either coupling over bonds or through space. The through space interactions between protons result in magnetization transfer called Nuclear Overhauser Effect (NOE). The strength of the NOE is inversely dependent on inter-spin distance, thus it forms the basis for elucidation of three-dimensional molecular structure (**Figure 2.1c**). Compared with protons at a fixed distance (reference NOE), the distance between other protons can be calculated. As protons are abundant in nucleic acids, we can probe their three-dimensional structure based on the above principle.

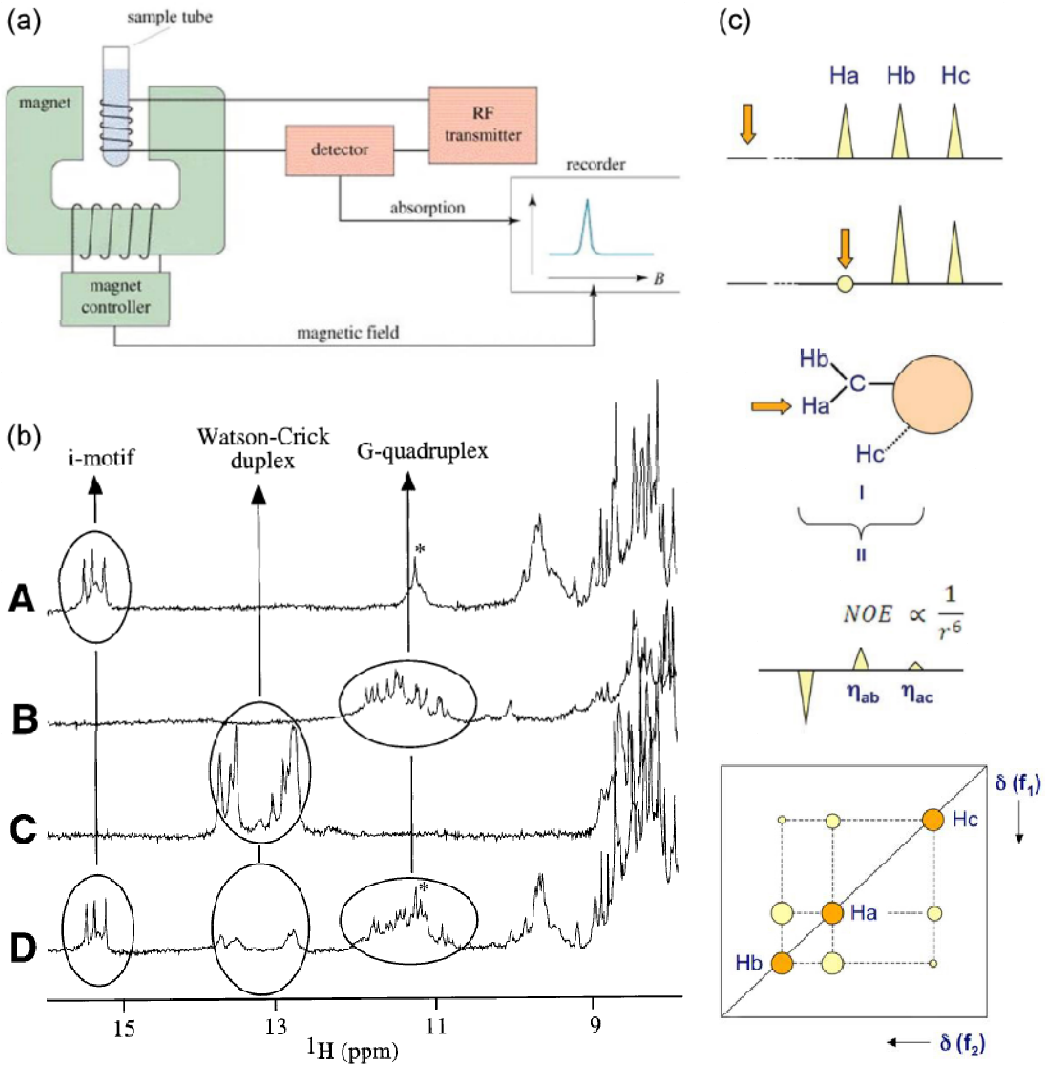

Figure 2.1 Schematic principle of (a) NMR spectrometer. (b) One-dimensional NMR spectra of d[AGGG(TTAGGG)$_3$] (22AG), d[(CCCTAA)$_3$CCCT] (22CT) and their 1:1 mixture at different experimental conditions, showing the imino protons for Watson-Crick base pairing (12–14 ppm) and for Hoogsteen base pairing (10–12 ppm). This figure is adapted from Phan's paper[57a]. (c) Nuclear Overhauser effect. If one moleculer has three protons (H_a, H_b, H_c), two of them (H_a and H_b) are at a fixed distance. When proton H_a is irradiated, its spin polarization will transfer to its neighboring protons, which is distance-dependent. In 2D spectra, the size of the cross-peak will depend on the internuclear distance.

2.2 Atomic Force Microscopy

Atomic Force Microscopy (AFM), a descendent of the scanning probe microscopy family, was invented by Binning et al. in 1986 in answer to a call for greater versatility in atomic scale imaging.[67] Falling on the heels of the development of the scanning tunneling microscope (STM) that relied on electron tunneling between the surface and the probe, the AFM has become an established surface analytical technique, capable of high resolution imaging of non-conductive surfaces under a variety of environmental conditions (i.e., ambient air or even a liquid environment) without special sample treatments. These outstanding properties enable AFM to probe biological systems in aqueous solution approximating physiological conditions, without damaging them or changing their physical properties.

AFM images are created by scanning a sharp tip, integrated to the apex of a cantilever, over the surface of a sample and by monitoring cantilever deflections resulted from the interaction force between a few atoms at a tip on a cantilever and the sample to probe the surface topography,[68] as shown in **Figure 2.2**. AFM is able to provide atomic resolution under certain conditions; however, the lateral resolution is typically on nanometer scale (predominantly limited by the size of the tip) and angstrom scale vertically.[69] Scientists at IBM have shown to give true atomic resolution of the pentacene molecule in ultra-high vacuum.[70]

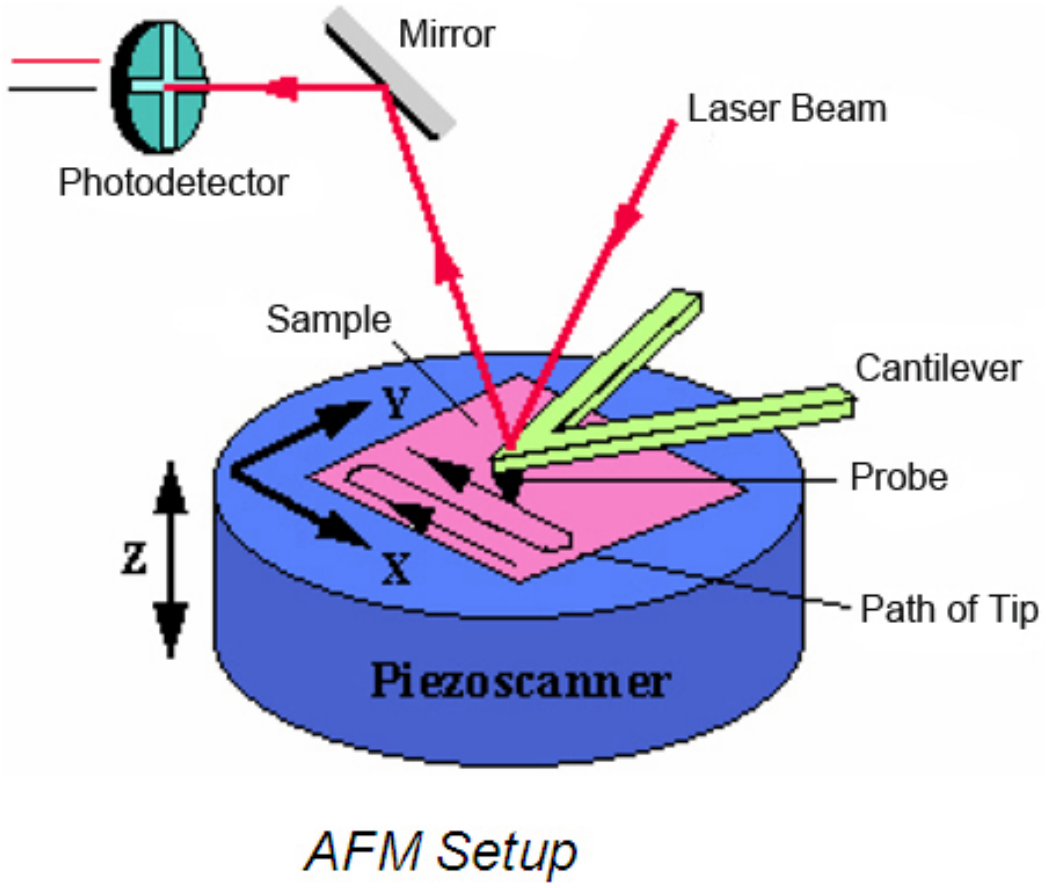

Figure 2.2 Schematic illustration of AFM as an imaging technique.

In general, imaging modes operated by AFM can be divided into static (contact) modes and dynamic (non-contact and tapping) modes relied on cantilever oscillation.

In the *static (contact)* mode, the cantilever is brought into continuous physical contact with the surface of the sample. Repulsive force, especially Pauli-repulsion, arises from overlapping of the electronic orbital will dominate. The force between the probe and the sample is kept constant by maintaining a constant cantilever deflection, while the topographic information of the sample is generated from the vertical displacement

of the cantilever. Contact mode is excellent for investigations of hard solid surfaces but the large repulsive force (in the order of 10^{-9} N) experienced between the tip and sample surface is often sufficiently high to cause inelastic deformation of soft materials. Dynamic AFM imaging modes were thus subsequently developed to avoid such problems.

In *dynamic non-contact* AFM the cantilever is oscillated at its resonant frequency with the probe situated in close proximity to, but not come in contact with the sample surface. The sample is not damaged during scanning, due to the sustained distance between tip and surface. However, its disadvantage lies in a significant decrease in resolution as a consequence of the lower sensitivity of attractive probe-surface interactions to subtle changes in surface topology. Furthermore, a liquid meniscus layer may develop on the sample and cantilever surfaces under ambient conditions. It would cause the tip to stick to the sample through capillary force while approaching the tip to the sample for set point minimization, which adversely affects resolution.

In *tapping mode*, the cantilever is sinusoidally vibrated driven by a piezo transducer and oscillates at or close to its resonance frequency. The oscillation amplitude, however, is bigger than in non-contact mode, so that the tip gently taps the sample each time it reaches the lowest point. When the sample intermittently interacts with the vibrating tip, the oscillation amplitude, phase and resonance frequency of the cantilever will be modified by interaction force between tip and sample. These

modifications with respect to the external reference provide information about the features on the sample surface. In contrast to contact mode, a constant cantilever oscillation amplitude rather than deflection is maintained. Topographic images of the surface are produced by changes in amplitude. Tapping mode also gives the possibility to image the phase shift recorded via a lock-in amplifier between the driven and the actual oscillations of the cantilever as a function of the X, Y coordinates analogue to the topographic image. Phase imaging can reveal variation in material properties, such as friction, adhesion and viscoelasticity, which is helpful in discriminating between different materials on surface.[71] Tapping mode imaging combines the high-resolution and rapid image acquisition ability of contact AFM with the non-destructive approach of non-contact AFM. The resolution is considerably improved compared to the non-contact mode; whereas smaller lateral force and shorter tip-sample contact time allow minimizing the deformation of the sample by tapping instead of constant contact.

2.3 Raman spectroscopy

Raman spectroscopy is a technique to visualize the inelastic scattering from the sample. When light is incident on the sample, the outgoing radiation can have the same or different energy compared to the incident light depending on the occurrence of elastic scattering (Rayleigh scattering) or inelastic scattering (Raman scattering). The energy difference between the scattered and incident photon in Raman scattering is caused by energy exchange between the incident photons and the molecules of the sample through phonon emission or absorption. The energy difference gives rise to a frequency shift in the scattered photon from the incident photon. Such frequency shift is termed as Raman shift and typically expressed in wavenumber (cm^{-1}). The positive wavenumber corresponds to energy loss (Stokes) and phonon emission while the negative wavenumber corresponds to energy gain (anti-Stokes) and phonon absorption (**Figure 2.3**). Raman spectra refer to the number of inelastically scattered photons as a function of Raman shift.

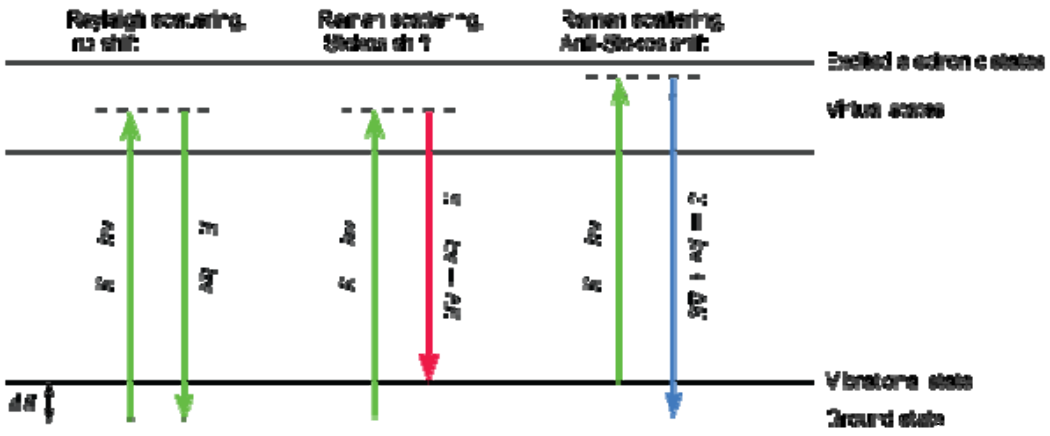

Figure 2.3 Schematic diagrams for Rayleigh and Raman scattering process.

2.4 Ultraviolet Absorption

Ultraviolet (UV) absorption at different temperatures is commonly used to measure the thermodynamic stability and analyze the kinetics of folding of secondary structures. The melting temperature relates to the stability of the folded G-quadruplex, and the nature of the melting and cooling spectra can also provide insights into the folding/unfolding kinetics. A reversible absorbance change at 295 nm, which is different from classical recording at 260 nm, allows to precisely monitoring guanine (G)-quatet formation and dissociation. This technique relies on the specific hyperchromicity of the UV absorption band at 295 nm for quadruplex structure that was first described by Mergny et al.[72]. The absorbance behavior was investigated with a guanine-rich oligonucleotide d(AGGGTTAGGGTTAGGGTTAGGG) at 1 and 90 ºC. The difference spectrum exhibits the difference between high and low temperature absorbance spectra, a maximal hyperchromism is observed upon formation of G-quartet at 295 nm[72] (**Figure 2.4a**). UV-melting experiments were conducted at 260 nm and 295 nm (**Figure 2.4b**). The absorbance profile at 295 nm shows a nice sigmoidal curve with melting temperature of 56 ºC. The heating and cooling curves are super-imposable indicating equilibrium processes. However, the absorption profile produced by monitoring at 260nm displays a smaller variation of absorbance.

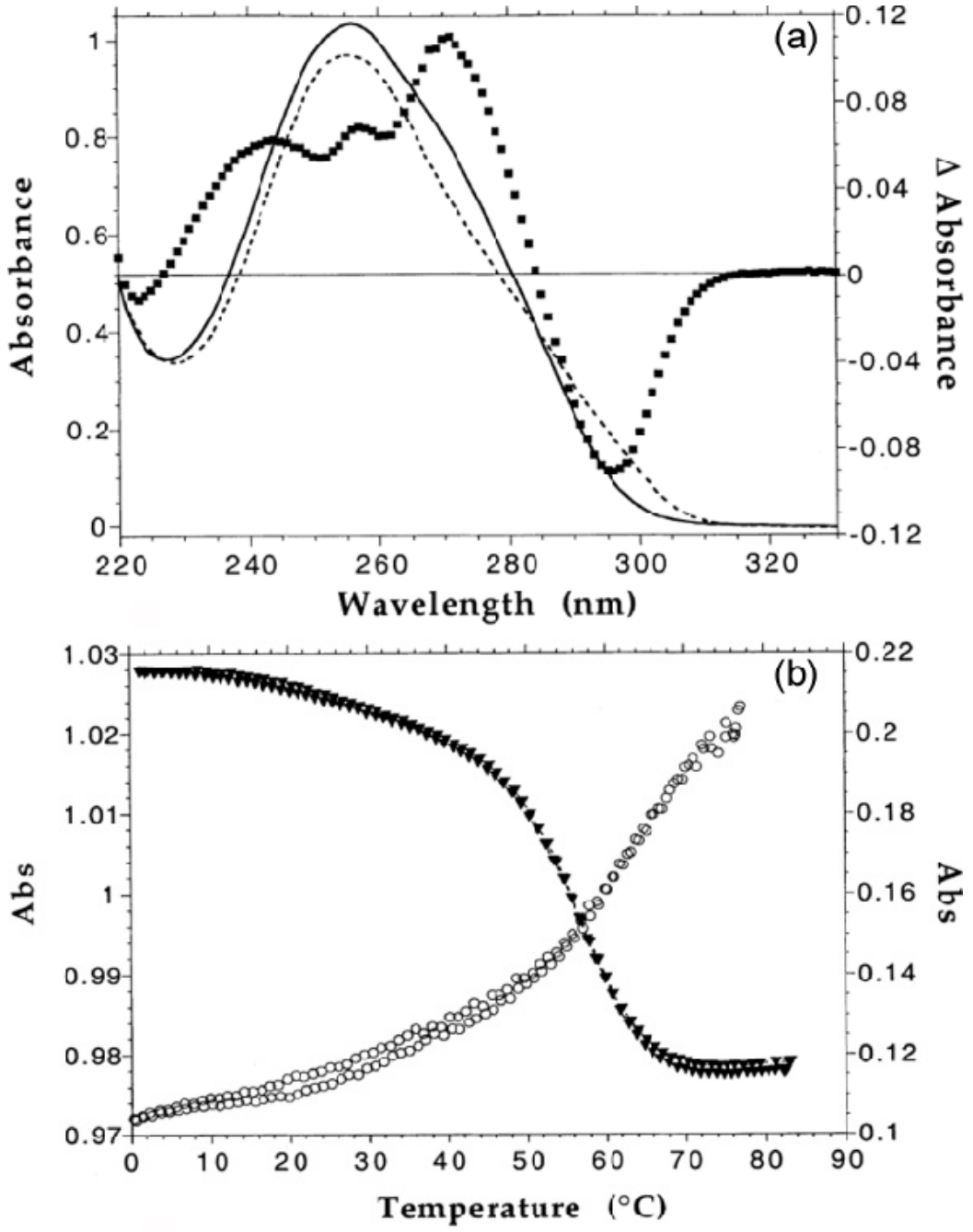

Figure 2.4 (a) Absorbance spectra at 90 ºC (full line) and 1ºC (dotted line). Differences in absorbance between the high and low temperature spectra (black squares, right scale). (b) Denaturation curves obtained at pH 7.0 at two different wavelengths: 260 nm (open circles, left vertical axis) and 295 nm (black triangles, right vertical axis). These data were derived from the oligonuleotide d(AGGGTTAGGGTTAGGGTTAGGG). This figure is adapted from Mergny's paper[72].

2.5 Circular Dichroism

Circular dichroism is a spectroscopic method which depends on the fact that chiral samples (including the vast majority of biomolecules) interact differently with left-handed and right-handed circularly polarized lights, which are mirror image of one another. Briefly, circular dichroism spectrum is defined as wavelength-dependent differential absorption of left-handed and right-handed circularly polarized light (**Figure 2.5**). A method which can discern such subtle differences between non-superimposable mirror image must be highly sensitive to the three-dimensional features of molecules, that is, to conformation. This is the basis of application of circular dichroism spectroscopy in studying conformations and structural change of proteins and nucleic acids[73]. G-quadruplex structures can be characterized by distinct circular dichroism signatures with a 295 nm positive and 265 nm negative peaks of anti-parallel G-quadruplex structures, while a parallel G-quadruplex structure exhibits a strong positive band at 260 nm and a negative band at 240 nm[74] (**Figure 3.22**). Gray et al.[75] proposed that the characteristic CD bands results from different stacking interactions of neighboring G-quartets with same or opposite polarities for hydrogen bonding.

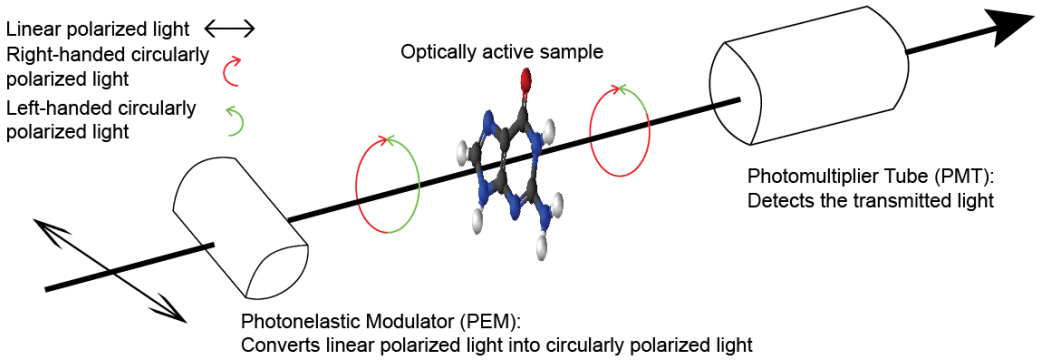

Figure 2.5 Schematic illustration of the principle behind CD spectroscopy.

2.6 Gel Electrophoresis

Gel electrophoresis facilitates the separation of charged molecules using an electric field applied to a gel matrix. Eletrophoretic separation of biomolecules is based on their difference in molecular size, relative hydrophobicity, net charge, and conformation. When exposed to an electric field, negatively charged DNA will travel through the matrix towards the anode. Small molecules migrate more freely and travel farther than large molecules when this molecular sieving works on the basis of their molecular size. Samples of unknown size are compared against standard markers with same migration speed to interpolate the actual size of the molecules. After the electrophoresis is complete, the molecules in the gel can either be visualized directly if the analyzed molecules glow under UV light or by using fluorescent dye/radioactive isotope to stain the sample to make them visible (**Figure 2.6**).

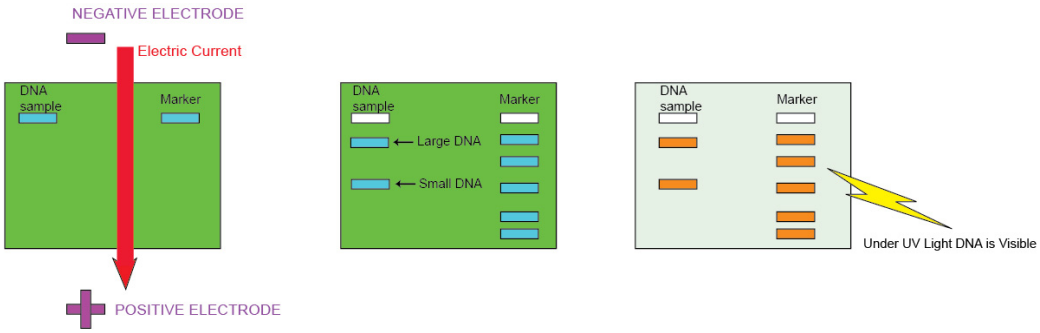

Figure 2.6 Schematic of gel electrophoresis used for the separation of DNA based on molecular size.

Chapter 3 *Giardia* Telomeric Sequence $d(TAGGG)_4$ Forms Two Intramolecular G-quadruplexes in K^+ Solution: Effect of Loop Length and Sequence on the Folding Topology

3.1 Introduction

It is important to be able to predict the G-quadruplex structure based on sequence, as well as to design a sequence that will adopt a desired G-quadruplex topology. Systematic studies on the effect of loop length/sequence and G-tract length on the topology and stability of G-quadruplexes have been conducted by several groups.[76] However, structural interpretations of G-quadruplexes in these works were limited by using mainly CD spectra.

As for high-resolution structural studies, five different intramolecular G-quadruplexes have been solved for DNA sequences containing human telomeric TTAGGG repeats under different experimental conditions.[33, 36, 77] In particular, the four-repeat human telomeric sequence $d[TAGGG(TTAGGG)_3]$ has been shown to form a (3 + 1) intramolecular G-quadruplex structure in K^+ solution, in which the G-tetrad core contains three strands oriented in one direction and the fourth in the opposite

direction.[77c]

Sequence variants of the human telomere are expressed in many other organisms. Variation on the loop length/sequence are found in telomeric repeats such as *Giardia lamblia* with TAGGG as the telomere repeat[74b], TTTAGGG from the flowering plant *Arabidopsis thaliana*[78] and *Chlamydomonas reinhardtii* telomere repeat TTTTAGGG[79]. Telomeric sequence bearing different G-tract length is found in the silkworm *Bombyx mori* consisted of pentanucleotide repeats TTAGG[80].

In K^+ solution, the *Bombyx mori* telomeric sequence d[TAGG(TTAGG)$_3$], which differs from the human counterpart only by one G deletion in each repeat, forms a chair-type intramolecular G-quadruplex, indicating an effect of G-tract length on the folding topology of G-quadruplexes.[80] A small variation within the non-G linkers might result in a dramatic change in G-quadruplex topology: the four-repeat variant human telomeric sequence d[AGGG(<u>C</u>TAGGG)$_3$] (variation is underlined) forms a chair-type intramolecular G-quadruplex involving two G-tetrads and a G•C•G•C tetrad in K^+ solution.[81]

Giardia has the features expected of eukaryotic cells, including linear chromosomes capped with telomeres composed of tandem arrays of TAGGG repeats.[82] To further explore the effect of loop length and sequence on the folding topology of G-quadruplexes, here we examine the structure of the four-repeat *Giardia* telomeric

sequence d[TAGGG(TAGGG)$_3$], i.e., d(TAGGG)$_4$, in which the non-G linkers are shorter than the human counterpart only by one T deletion in each repeat. The TAGGG repeats have also been detected as a potential variant that are interspersed within the human telomeres,[83] among the canonical TTAGGG repeats. We found that this sequence forms two different intramolecular G-quadruplexes in K^+ solution. The first one is a novel basket-type antiparallel-stranded G-quadruplex containing two G-tetrads, a G•(A-G) triad, and two A•T base pairs; the second one is a propeller-type parallel-stranded G-quadruplex involving three G-tetrads. Recurrence of several structural elements in the observed structures suggests a "cut and paste" principle for the design and prediction of G-quadruplex topologies.

3.2 Experimental Section

DNA Sample Preparation Unlabeled and site-specific labeled DNA oligonucleotides (**Table 3-1** and **Table 3-2**) were chemically prepared using an ABI 394 DNA/RNA synthesizer, as previously described.[80] DNA concentration was expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species.[84]

Table 3-1 Natural and Modified *Giardia* Telomeric DNA Sequences[a].

name	sequence							
natural sequence	TA	GGG	TA	GGG	TA	GGG	TA	GGG
I18-Form 1	TA	GGG	TA	GGG	TA	GGG	TA	**I**GG
ΔA12-Form 2	TA	GGG	TA	GGG	T-	GGG	TA	GGG

[a]Modifications from the natural sequence are shown in boldface.

Table 3-2 DNA sequences used in this study [(a)]

Name		Sequence							
Natural sequence		TA	GGG	TA	GGG	TA	GGG	TA	GGG
^{15}N-labeled natural sequences		TA	*GGG	TA	GGG	TA	GGG	TA	GGG
		TA	G*GG	TA	GGG	TA	GGG	TA	GGG
		TA	GG*G	TA	GGG	TA	GGG	TA	GGG
		TA	GGG	TA	*GGG	TA	GGG	TA	GGG
		TA	GGG	TA	G*GG	TA	GGG	TA	GGG
		TA	GGG	TA	GG*G	TA	GGG	TA	GGG
		TA	GGG	TA	GGG	TA	*GGG	TA	GGG
		TA	GGG	TA	GGG	TA	G*GG	TA	GGG
		TA	GGG	TA	GGG	TA	GG*G	TA	GGG
		TA	GGG	TA	GGG	TA	GGG	TA	*GGG
		TA	GGG	TA	GGG	TA	GGG	TA	G*GG
Inosine-substituted sequences	[I5]	TA	GGI	TA	GGG	TA	GGG	TA	GGG
	[I10]	TA	GGG	TA	GGI	TA	GGG	TA	GGG
	[I13]	TA	GGG	TA	GGG	TA	IGG	TA	GGG
	[I18]	TA	GGG	TA	GGG	TA	GGG	TA	IGG
^{15}N-labeled *I18-Form 1*		TA	GGG	TA	GGG	TA	G*GG	TA	IGG
		TA	GGG	TA	GGG	TA	GG*G	TA	IGG
		TA	GGG	TA	GGG	*TA	GGG	TA	IGG
Uracil-substituted *I18-Form 1*	U1	UA	GGG	TA	GGG	TA	GGG	TA	IGG
	U6	TA	GGG	UA	GGG	TA	GGG	TA	IGG
	U11	TA	GGG	TA	GGG	UA	GGG	TA	IGG
	U16	TA	GGG	TA	GGG	TA	GGG	UA	IGG
Δ*A12-Form 2*		TA	GGG	TA	GGG	T	GGG	TA	GGG
^{15}N -labeled Δ*A12-Form 2*		TA	*GGG	TA	GGG	T	GGG	TA	GGG
		TA	G*GG	TA	GGG	T	GGG	TA	GGG
		TA	GG*G	TA	GGG	T	GGG	TA	GGG
		TA	GGG	TA	*GGG	T	GGG	TA	GGG
		TA	GGG	TA	G*GG	T	GGG	TA	GGG
		TA	GGG	TA	GG*G	T	GGG	TA	GGG
		TA	GGG	TA	GGG	T	*GGG	TA	GGG
		TA	GGG	TA	GGG	T	G*GG	TA	GGG
		TA	GGG	TA	GGG	T	GG*G	TA	GGG
		TA	GGG	TA	GGG	T	GGG	TA	*GGG
		TA	GGG	TA	GGG	T	GGG	TA	G*GG
8-deuterodeoxy-guanine-substituted Δ*A12-Form 2*	D4	TA	G$^{\#}$GG	TA	GGG	T	GGG	TA	GGG
	D8	TA	GGG	TA	$^{\#}$GGG	T	GGG	TA	GGG
	D13	TA	GGG	TA	GGG	T	G$^{\#}$GG	TA	GGG
	D17	TA	GGG	TA	GGG	T	GGG	TA	$^{\#}$GGG

[(a)]The oligonucleotides were 2% ^{15}N-labeled at the position marked by an asterisk and 8-deutero-deoxyguanine substituted by a hash.

Gel Electrophoresis The molecular size of the structures formed by DNA oligonucleotides was probed by nondenaturing polyacrylamide gel electrophoresis (PAGE). The electrophoresis experiment was performed with 10 × 7 cm native gel containing 20% acrylamide (Acrylamide:Bis-acrylamide = 37.5:1) supplemented with 10 mM KCl in TBE buffer pH 8.3 at 120 mV, in 140 minutes. Gel was viewed by UV shadowing.

Circular Dichroism (CD) CD spectra were recorded on a Jasco J-810 spectropolarimeter using a 1-cm path-length quartz cuvette as previously described.[80] The buffer contained 10 mM KCl, 40 mM LiCl and 20 mM lithium phosphate (pH 7). For each spectrum, an average of 3 scans was taken. DNA concentration was typically around 5 μM.

UV Spectroscopy The UV melting experiments were performed on a Varian CARY-300 spectrophotometer by monitoring the UV absorption at 295 nm as a function of temperature.[72] The concentration of DNA varied from 3 to 300 μM. Samples were covered with approximately 100 μL of mineral oil to prevent evaporation. They were equilibrated at 90 °C for 10 min, then cooled to 20 °C and heated to 90 °C twice consecutively at a rate of 0.15 °C per minute. Data were collected every 1 °C during both cooling and heating processes.

NMR Spectroscopy Samples for NMR study were dialyzed successively against ~50 mM KCl solution and against water. Unless otherwise stated, the strand concentration of the NMR samples was typically 0.5-2.0 mM; the solutions contained 70 mM KCl and 20 mM potassium phosphate (pH 7). NMR experiments were performed on 600 and 700 MHz Bruker spectrometers at 25 °C, unless otherwise specified. Resonances for guanine residues were assigned unambiguously by using site-specific low-enrichment ^{15}N labeling,[85] site-specific ^{2}H labeling,[86] and JRHMBC through-bond correlations at natural abundance.[66a, 87] Resonances for thymine residues were assigned following systematic T-to-U substitutions. Spectral assignments were completed by NOESY, COSY, TOCSY, and $^{13}C-^{1}H$-HSQC, as described previously.[66a] Interproton distances were deduced from NOESY experiments at various mixing times. All spectral analyses were performed using the FELIX (Felix NMR, Inc.) program.

Structure Calculation Interproton distances for the *118-Form 1* G-quadruplex were deduced from NOESY experiments performed in H_2O (mixing time, 200 ms) and D_2O (mixing times, 100, 150, 200, and 350 ms). Structure computations were performed using the XPLOR-NIH program[88] in three general steps essentially as previously described:[77i] (i) distance geometry simulated annealing, (ii) distance-restrained molecular dynamics refinement, and (iii) relaxation matrix intensity refinement. Hydrogen bond restraints, interproton distance restraints, dihedral restraints, and planarity restraints were imposed during structure calculations. Structures were

displayed using the PyMOL program.[89]

Data Deposition The coordinates for the *I18-Form 1* G-quadruplex have been deposited in the Protein Data Bank (accession code 2KOW).

3.3 Results and Discussion

3.3.1 Four-Repeat *Giardia* Telomeric Sequence d(TAGGG)$_4$ Forms Two G-quadruplex Structures in K^+ Solution

Imino proton spectra of the four-repeat *Giardia* telomeric sequence d(TAGGG)$_4$ in K^+ solution (**Figure 3.1** and **Figure 3.2**) show two sets of peaks (differentiated by circles and asterisks, respectively) at 10.8−12.0 ppm, which could be distinguished by their respective intensities at different temperatures, indicating the formation of two different G-quadruplexes. At 25 °C (**Figure 3.1a**), the major conformation, designated Form 1, represented about 80% of the population (peaks with circles), while the minor conformation, designated Form 2, represented about 20% of the population (peaks with asterisks). When the temperature increased, the relative population of Form 2 with respect to Form 1 increased (**Figure 3.1b**). At temperatures higher than 40 °C, Form 2 was favored over Form 1 (**Figure 3.2**). Note that Form 1 was also characterized by an imino proton at 13.6 ppm.

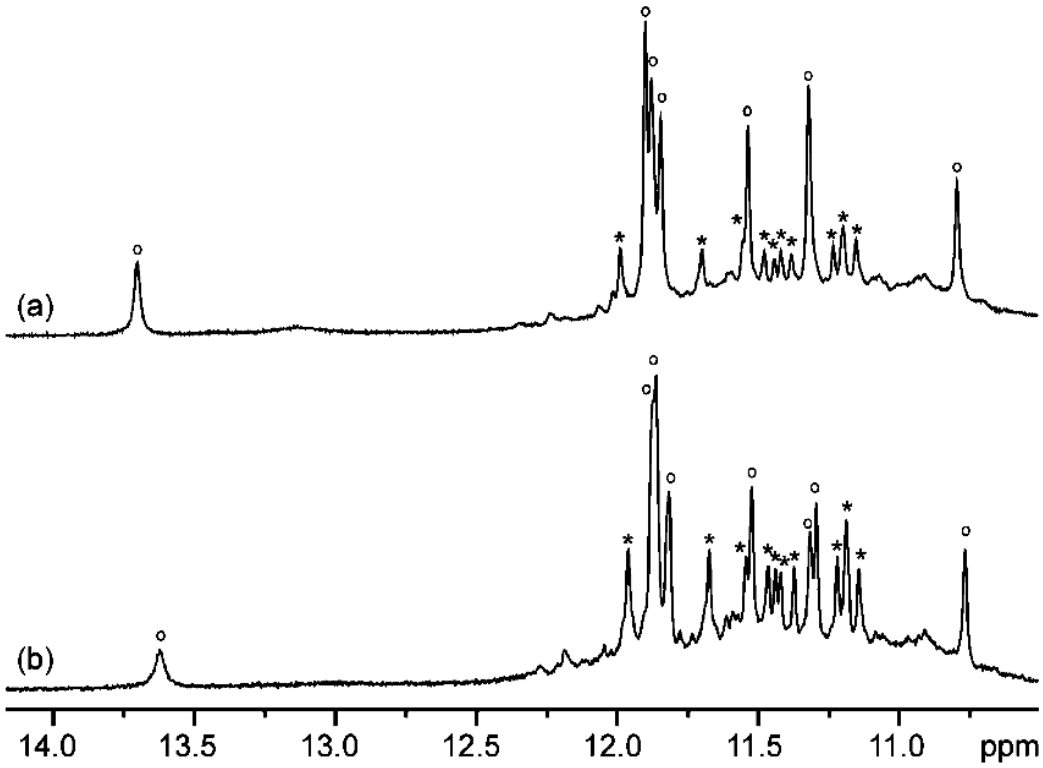

Figure 3.1 Imino proton spectra of the 20-nt *Giardia* telomeric $d(TAGGG)_4$ sequence in K^+ solution at (a) 25 °C and (b) 35 °C. Two sets of peaks, corresponding to two different conformations of G-quadruplexes, are labeled with circles and asterisks, respectively.

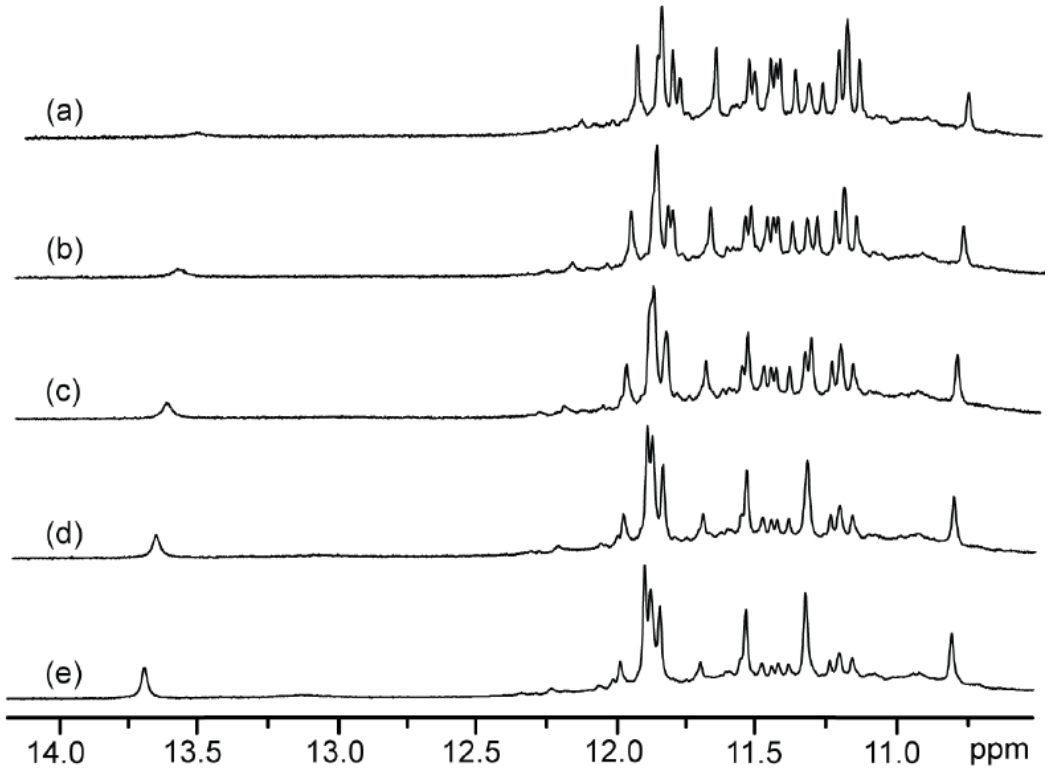

Figure 3.2 Imino proton spectra of the 20-nt *Giardia* telomeric $d(TAGGG)_4$ sequence in K^+ solution at (a) 45 ^oC, (b) 40 ^oC, (c) 35 ^oC, (d) 30 ^oC, and (e) 25 ^oC.

3.3.2 Favoring a Single Conformation by Sequence Modifications

We identified small sequence modifications (**Table 3-1**) that favored a single G-quadruplex conformation, thereby allowing detailed structural analysis of both conformations: substitution of G18 by an inosine favored Form 1 (**Figure 3.3b**), whereas the deletion of A12 favored Form 2 (**Figure 3.3c**). The rationale of these modifications will be discussed in view of the structures described below. Comparison of the NMR spectra of these modified sequences (designated *I18-Form 1* and *ΔA12-Form 2*, respectively) with those of the natural sequence indicated that the modified sequences adopted the same G-quadruplex folds as Form 1 and Form 2 of the natural sequence, respectively (**Figure 3.3** and data not shown).

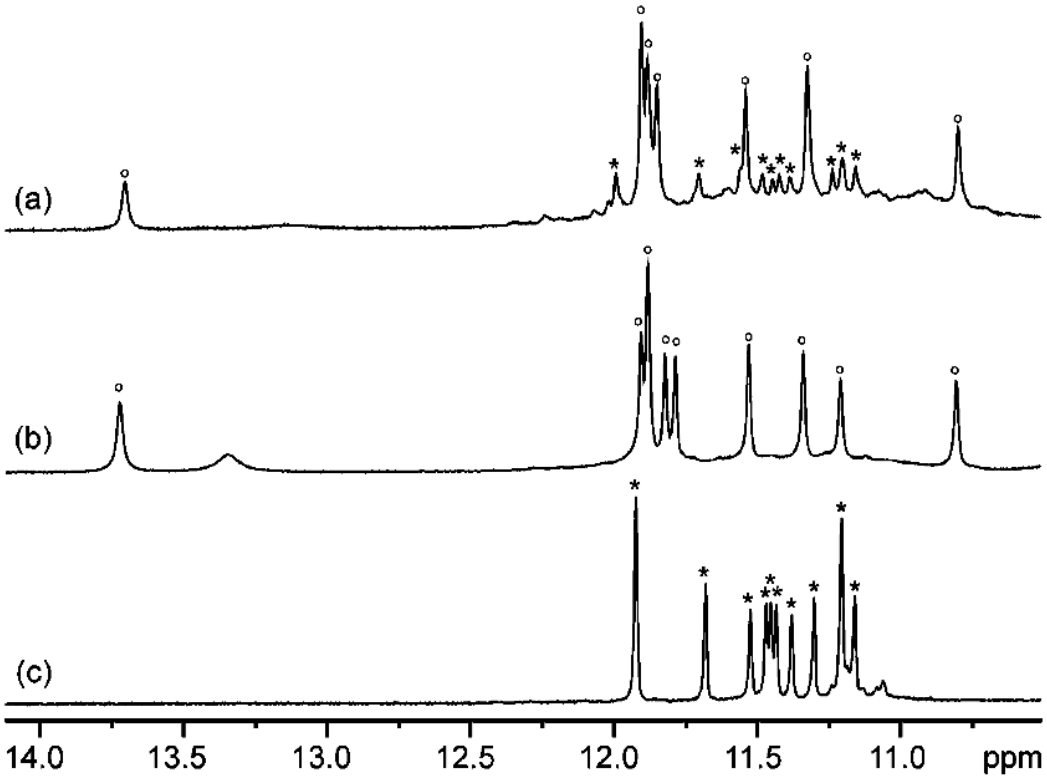

Figure 3.3 Imino proton spectra of natural and modified *Giardia* telomeric sequences in K^+ solution. (a) Natural $d(TAGGG)_4$ sequence; (b) *I18-Form 1*; and (c) *ΔA12-Form 2*. Peaks from Form 1 and Form 2 are labeled with circles and asterisks, respectively.

The molecular size of Form 1 and Form 2 G-quadruplexes formed by the natural and modified *Giardia* telomeric sequences (**Table 3-1**) was probed by native polyacrylamide gel electrophoresis (PAGE) (**Figure 3.4**). A single major band was observed for each sequence. The migration rate of the natural sequence, *I18-Form 1* and *ΔA12-Form 2* was comparable with that of a reference monomeric propeller-type parallel-stranded G-quadruplex containing three G-tetrads (unpublished data) and significantly faster than that of a dimeric interlocked G-quadruplex containing totally six G-tetrads,[90] arguing for an intramolecular structure for both Form 1 and Form 2.

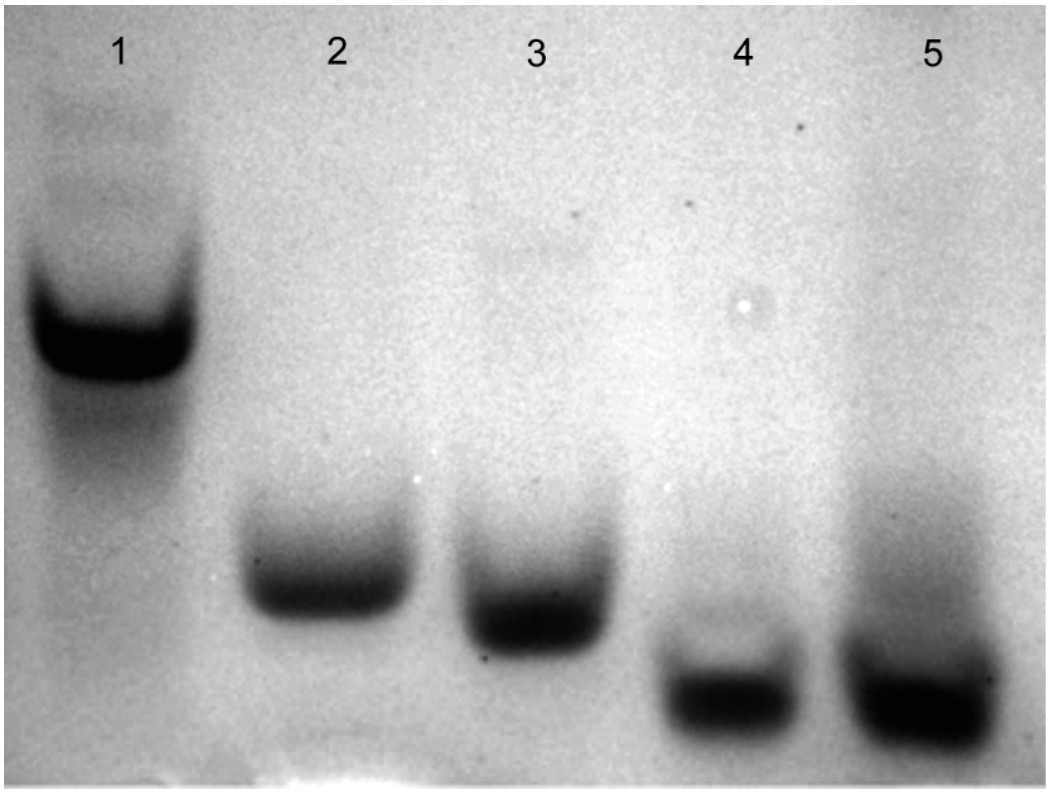

Figure 3.4 Nondenaturing PAGE analysis of the *Giardia* telomeric sequences. Migration markers are provided on the left. Lane 1: a migration marker of an interlocked dimeric G-quadruplex;[90] Lane 2: a migration marker of a monomeric propeller-type parallel-stranded G-quadruplex[91]; Lane 3: *ΔA12-Form 2*; Lane 4: *I18-Form 1*; Lane 5: natural *Giardia* telomeric sequence $d(TAGGG)_4$.

Melting experiments were conducted for *I18-Form 1* and *ΔA12-Form 2* by monitoring the UV absorbance at 295 nm. All transitions were reversible, indicating that the denaturation curves corresponded to a true equilibrium process (**Figure 3.5**). No significant difference in melting temperature (1 °C or less for *I18-Form 1*; 4 °C or less for *ΔA12-Form 2*) was observed upon 100-fold increase in concentration (from 3 to 300 μM), consistent with intramolecular G-quadruplex formation. A slightly larger concentration-dependent behavior observed for Form 2 could be explained by the stacking of two G-quadruplex blocks (see below) in the presence of high DNA and/or

K^+ concentration. Such a high-order structure was also observed in a gel electrophoresis experiment (**Figure 3.6**).

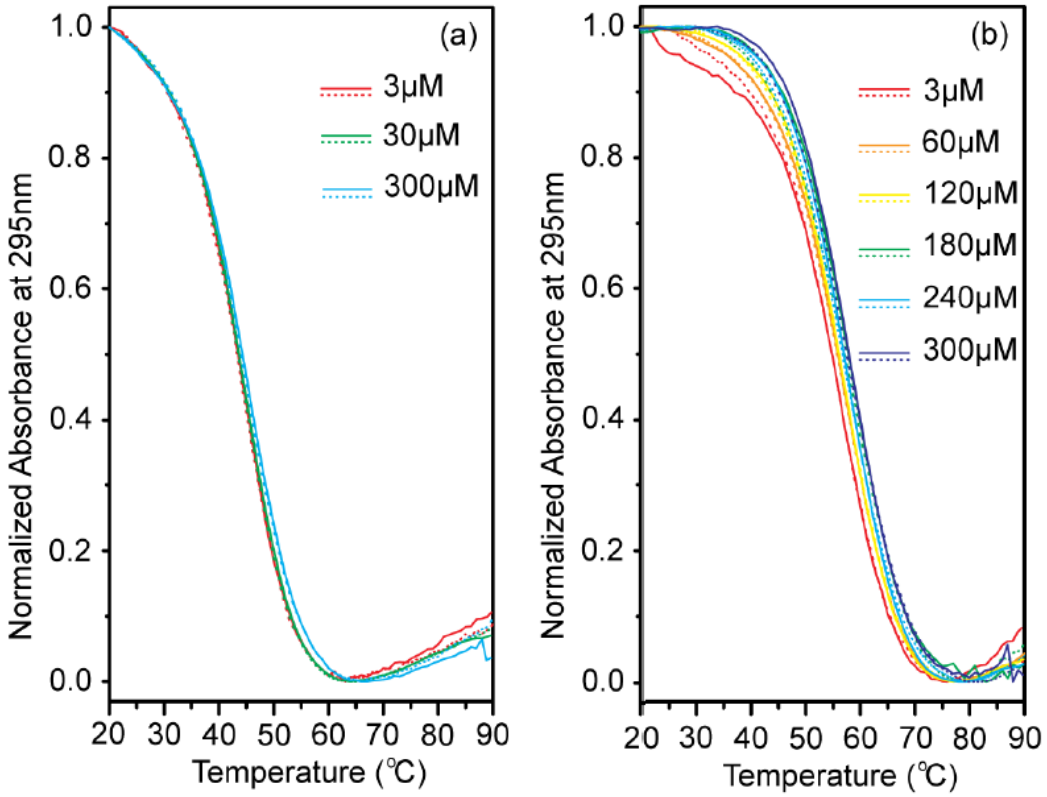

Figure 3.5 Normalized UV absorbance of different DNA concentrations at 295 nm as a function of temperature: (a) *I18-Form 1* and (b) *ΔA12-Form 2*. The heating (solid lines) and cooling (dashed lines) profiles are superimposable indicating equilibrium processes.

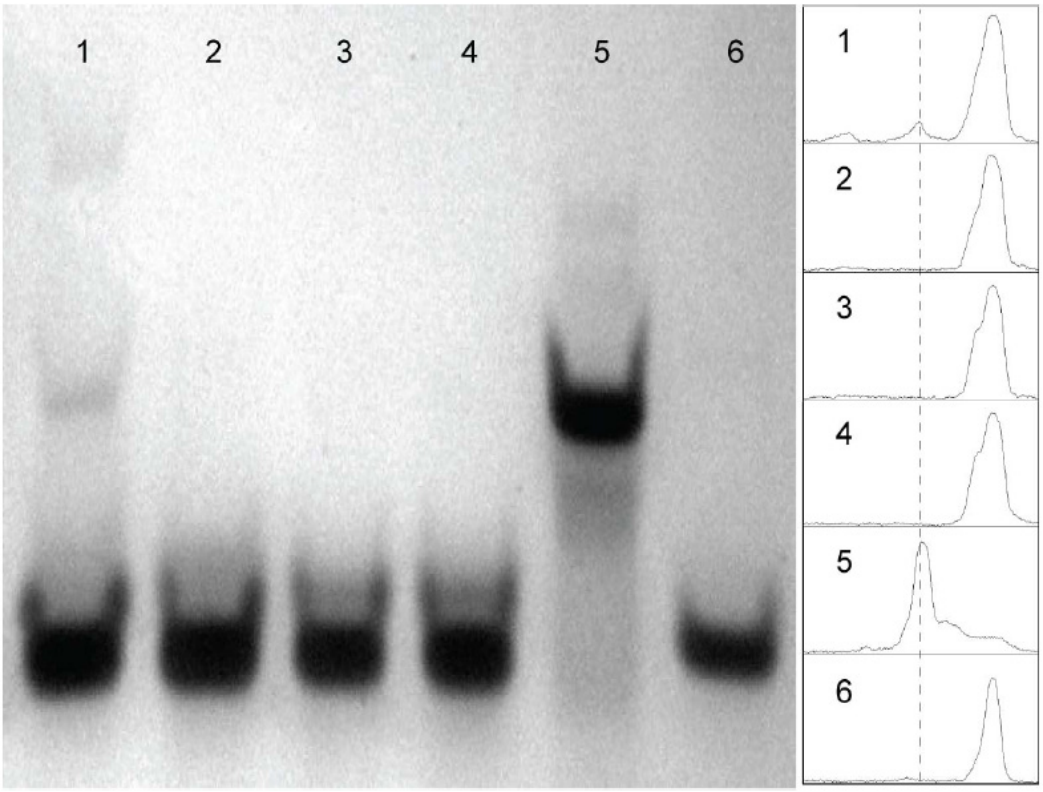

Figure 3.6 Non-denaturing PAGE analysis of the *ΔA12-Form 2 Giardia* telomeric sequences. Lane 1: *ΔA12-Form 2* (DNA concentration, 148.5 μM; solution contained 90 mM K^+); Lane 2: *ΔA12-Form 2* (DNA concentration, 148.5 μM; solution contained 90 mM K^+; sample was annealed before loading); Lane 3: *ΔA12-Form 2* (DNA concentration, 38.4 μM, solution contained 10 mM K^+); Lane 4: *ΔA12-Form 2* (DNA concentration, 38.4 μM, solution contained 10 mM K^+; sample was annealed before loading); Lane 5: a migration marker of an interlocked dimeric G-quadruplex;[90] Lane 6: a migration marker of a monomeric propeller-type parallel-stranded G-quadruplex[91]. Migration profiles from different lanes (indicated with the lanes' numbers) are shown on the right.

3.3.3 Structure of Form 1: Novel Basket-Type G-quadurplex with Two G-Tetrads and a G•(A-G) Triad

Rigorous approaches[66a, 85-87] were used for NMR spectral assignments. Guanine imino and H8 protons of *I18-Form 1* and *ΔA12-Form 2* sequences were unambiguously assigned (**Figure 3.7**−**Figure 3.11** and **Table 3-2**) using site-specific low-enrichment ^{15}N labeling[85] (**Figure 3.7 and Figure 3.9**), site-specific ^{2}H labeling[86] (**Figure 3.10**), and JRHMBC through-bond correlations between imino and H8 protons via ^{13}C5 at natural abundance[66a, 87] (**Figure 3.8** and **Figure 3.11**). Resonances for thymine residues were unambiguously assigned by T-to-U substitutions[66a] (**Table 3-2**). NMR spectral assignments were completed with other through-bond correlation experiments (COSY, TOCSY, and ^{13}C−^{1}H-HSQC) (data not shown) and through-space correlation NOESY experiments.[66a]

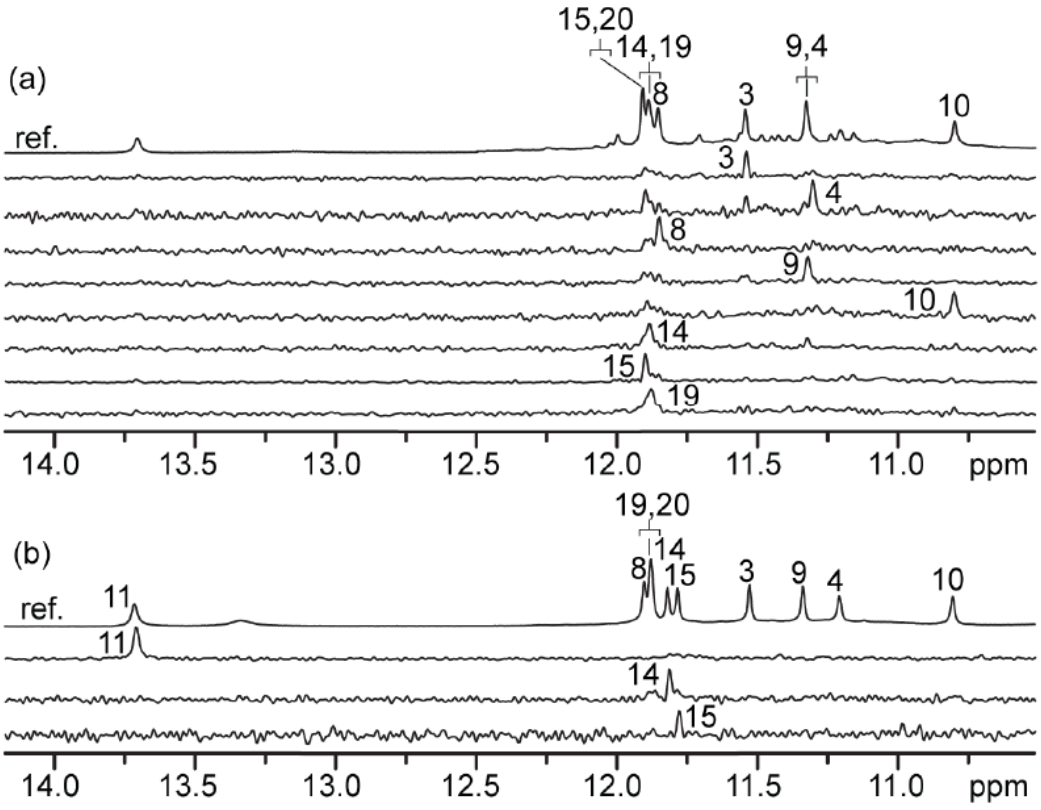

Figure 3.7 Imino proton spectra and assignments of (a) the 20-nt natural *Giardia* telomeric d(TAGGG)$_4$ sequence and (b) *I18-Form 1*. Imino protons were assigned in ^{15}N-filtered spectra of samples, 2% ^{15}N-labeled at the indicated positions. The reference spectra (ref.) are shown at the top.

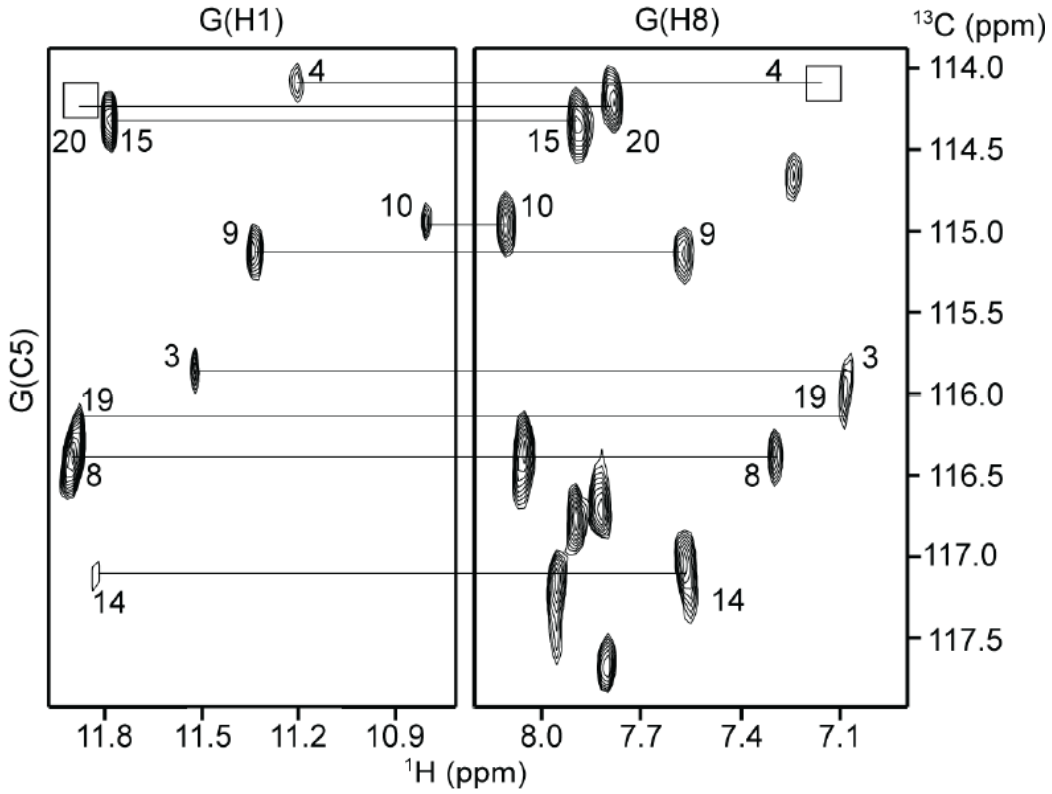

Figure 3.8 Guanine H8 proton assignments of *I18-Form 1* by through-bond correlations between guanine imino and H8 protons via ^{13}C5 at natural abundance, using long-range J-couplings. Boxes indicate the positions of peaks, observed at a lower threshold or in a separate experiment.

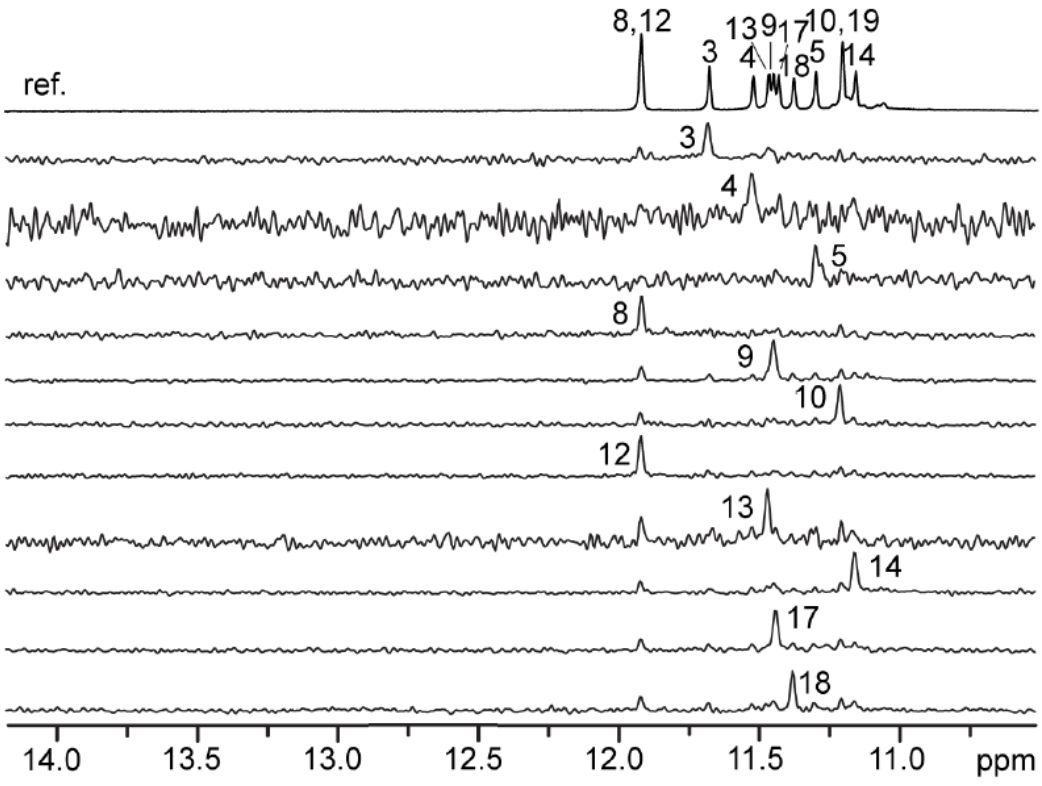

Figure 3.9 Guanine imino proton spectra and assignments of *ΔA12-Form 2*. Imino protons were assigned in ^{15}N-filtered spectra of samples, 2% ^{15}N-labeled at the indicated positions. The reference spectrum (ref.) is shown at the top.

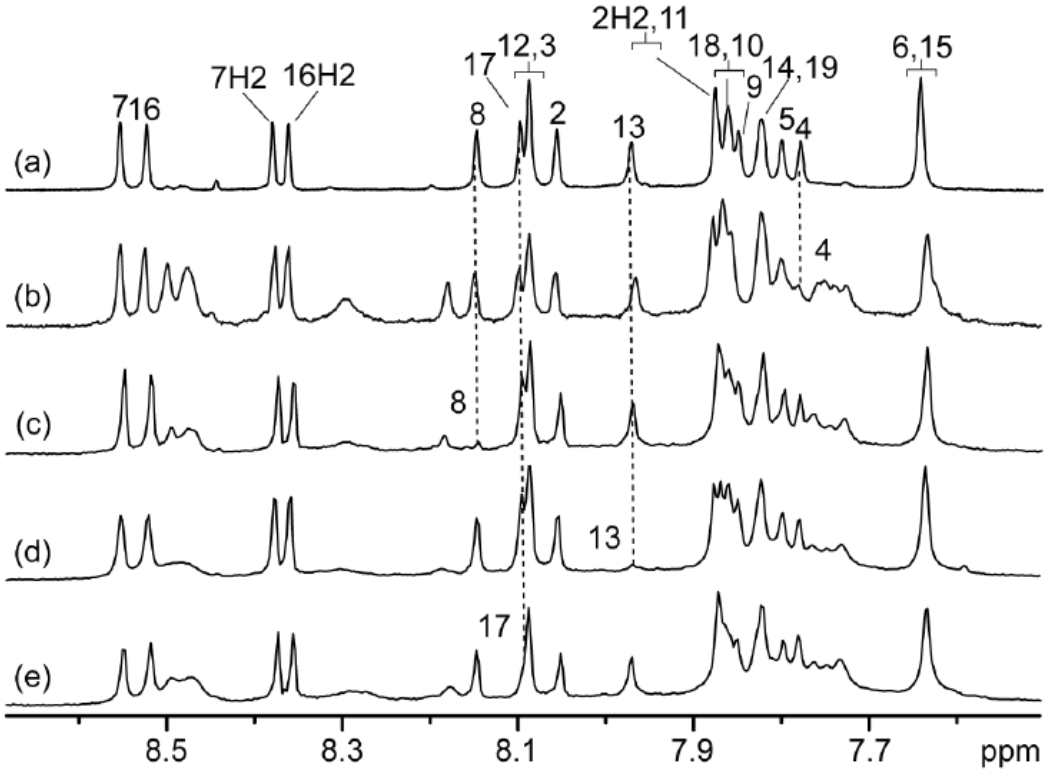

Figure 3.10 Guanine H8 proton assignments of *ΔA12-Form 2* by site-specific ^{2}H labeling at the indicated positions (b-e). The reference spectrum (a) is shown at the top. Spectrum (a) was recorded for a sample in D_2O; spectra (b-e) were recorded for samples in H_2O.

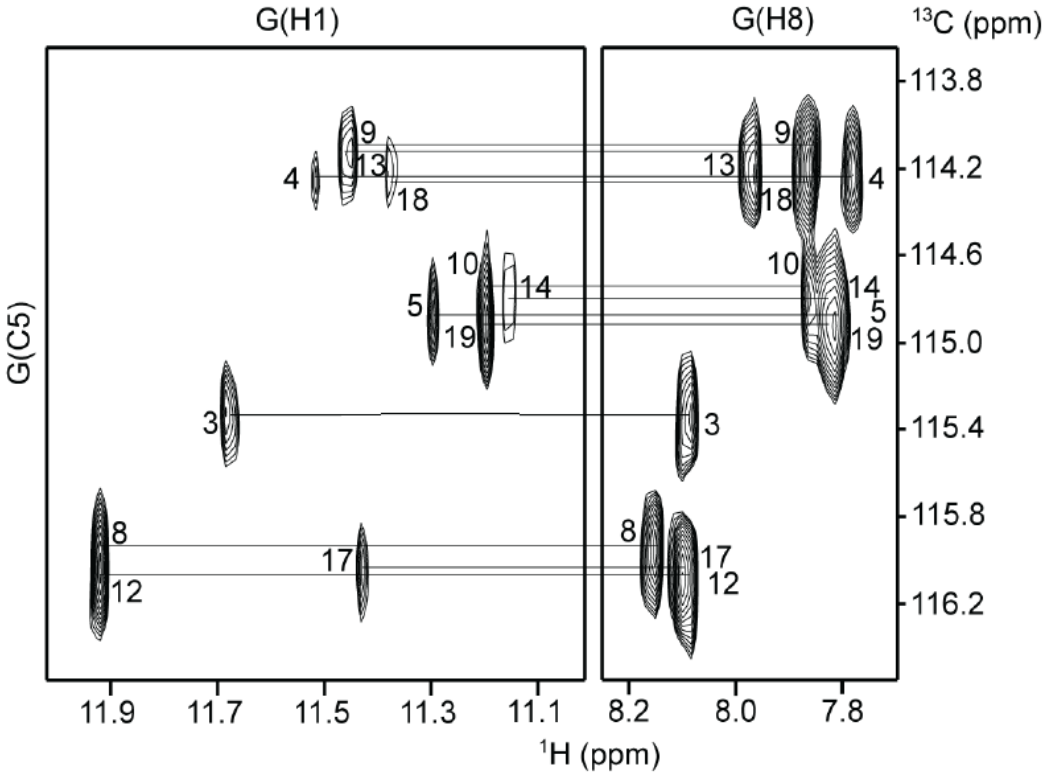

Figure 3.11 Guanine H8 proton assignments of *ΔA12-Form 2* by through-bond correlations between guanine imino and H8 protons via ^{13}C5 at natural abundance, using long-range J-couplings.

With the help of the unambiguous assignments (described above), the classical H8/H6−H1′ NOE sequential connectivity of *I18-Form 1* could be traced (**Figure 3.12**). The intensity of intraresidue H8−H1′ NOE cross-peaks (**Figure 3.12** and **Figure 3.13**) indicated *syn* glycosidic conformation for G3, G8, G14, and G19, and *anti* conformation for other residues. Analysis of imino-H8 connectivity patterns revealed the formation of a novel intramolecular basket-type G-quadruplex with two G-tetrad layers, G3•G9•G20•G14 and G4•G15•G19•G8 (**Figure 3.14**). The structure of the *I18-Form 1* G-quadruplex in K^+ solution (**Figure 3.15**) was calculated on the basis of NMR restraints (**Table 3-3**). Two edgewise loops, formed by the G5-T6-A7 and

T16-A17-I18 segments, bridge across a narrow and a wide groove, respectively, at the bottom of the G-tetrad core. The diagonal loop, formed by the G10-T11-A12-G13 segment, connects two opposite corners of the top G-tetrad. In this diagonal loop, formation of the G10•(A12-G13) triad (**Figure 3.16c**) was supported by the observation of the imino proton of G10 at 10.8 ppm and the NOEs between A12(H8) and G13(H8). Formation of the Hoogsteen A2•T11 (**Figure 3.16b**) and Watson−Crick A7•T16 base pairs (**Figure 3.16d**) at the top and bottom of the structure, respectively, was supported by the observation of the imino protons of T11 at 13.6 ppm and T16 at 13.3 ppm (**Figure 3.3** and **Figure 3.7**), as well as a number of NOEs (data not shown). A continuous stacking between the G-tetrad core and the base triad/pairs at the top and the bottom was observed (**Figure 3.16a**).

G-quadruplex structures with only two G-tetrad layers have been observed when they are further stabilized by other base pairing and stacking.[80, 92] Recently, it has been observed that sequences containing four tracts of three consecutive Gs, can fold into G-quadruplex structures involving only two G-tetrad layers.[77i, 81] To compensate for the loss of a potential additional G-tetrad, again, base pairing and stacking in the loops contribute to the stability of the structure. The observation of Form 1 *Giardia* telomeric G-quadruplex with only two G-tetrad layer reinforces the view that the overall folding topology of a G-quadruplex is defined not only by maximizing the number of G-tetrads, but also by maximizing all possible base pairing and stacking in the loops.[77i, 81]

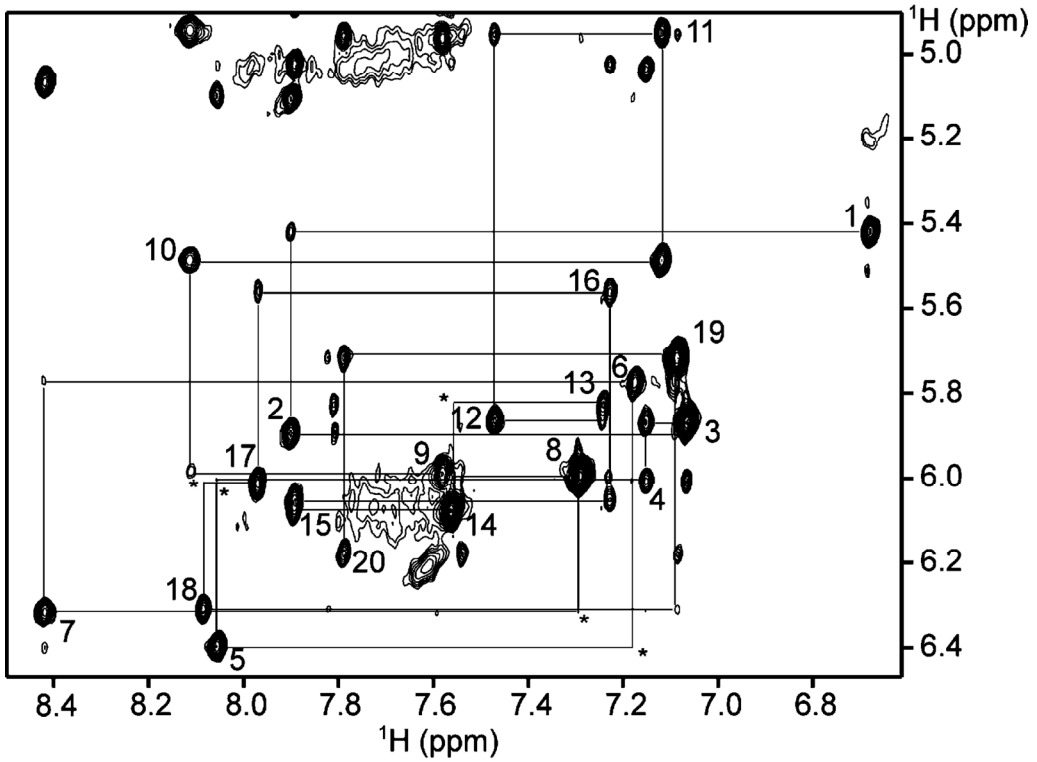

Figure 3.12 NOESY spectrum (mixing time, 350 ms) showing the H8/H6−H1′ connectivity of *I18-Form 1* in K^+ solution. Intraresidue H8/H6−H1′ NOE cross-peaks are labeled with residue numbers. Weak or missing sequential connectivities are marked with asterisks.

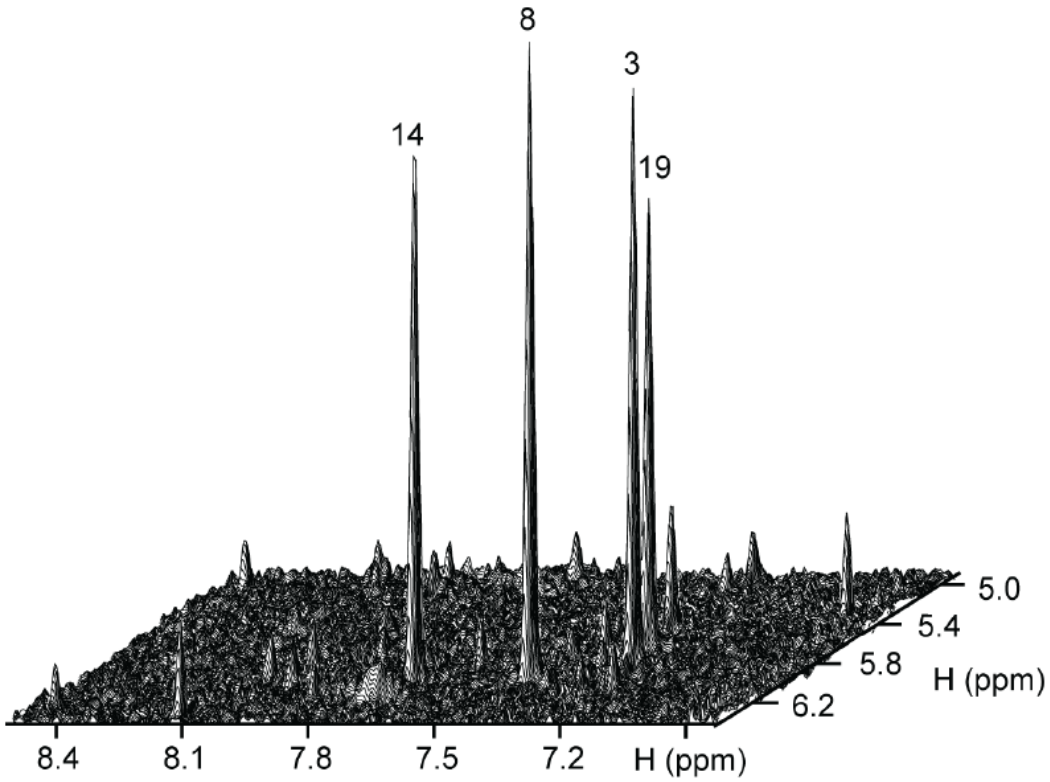

Figure 3.13 Stacked plot of NOESY spectrum (mixing time, 100 ms) of *I18-Form 1* showing four strong intraresidue H8-H1' cross-peaks for G3, G8, G14, and G19, indicating syn glycosidic conformation for these residues.

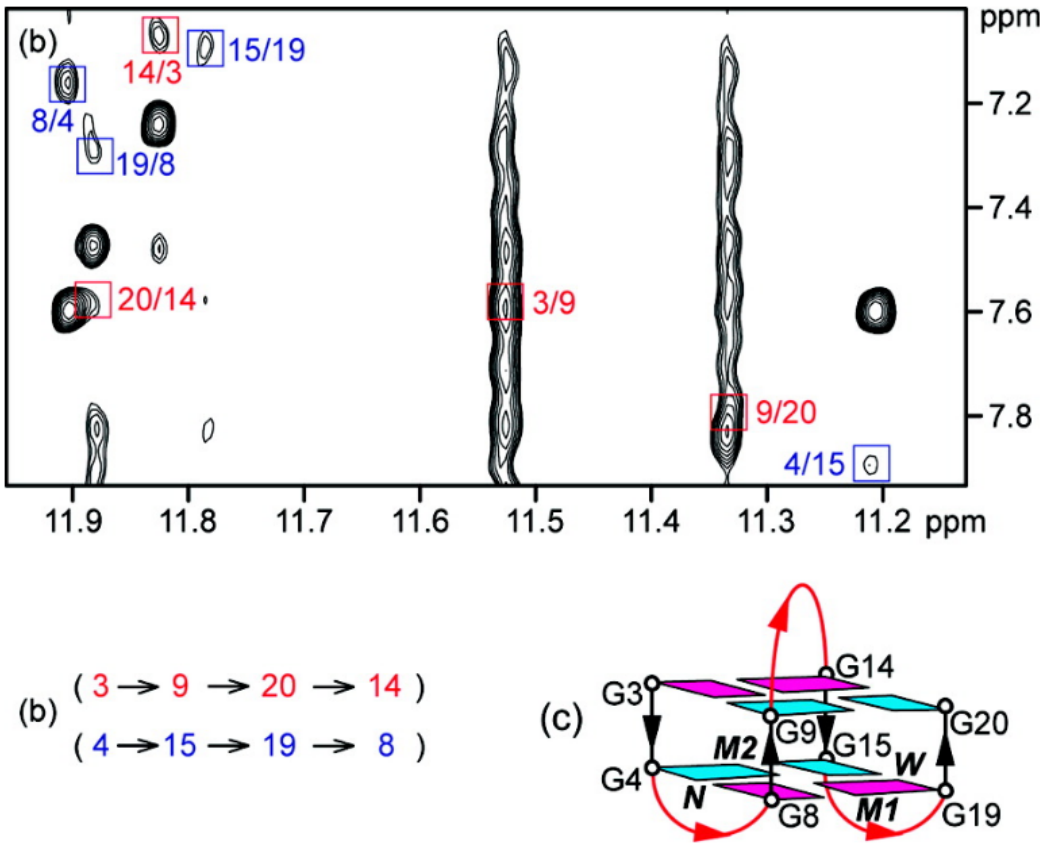

Figure 3.14 Determination of G-quadruplex folding topology. (a) NOESY spectrum (mixing time, 200 ms) showing imino-H8 connectivity of *I18-Form 1*. Cross-peaks that identify the two G-tetrads are framed and labeled with the residue number of imino protons in the first position and that of H8 protons in the second position. (b) Guanine imino-H8 NOE connectivities observed for *I18-Form 1* with G3•G9•G20•G14 and G4•G15•G19•G8 tetrads. (c) Schematic view of the *I18-Form 1* G-quadruplex. *anti* guanines are colored cyan; *syn* guanines are colored magenta. *W*, *M1*, *M2*, and *N* represent wide, medium 1, medium 2, and narrow grooves, respectively. The backbone of the core and loops is colored black and red, respectively.

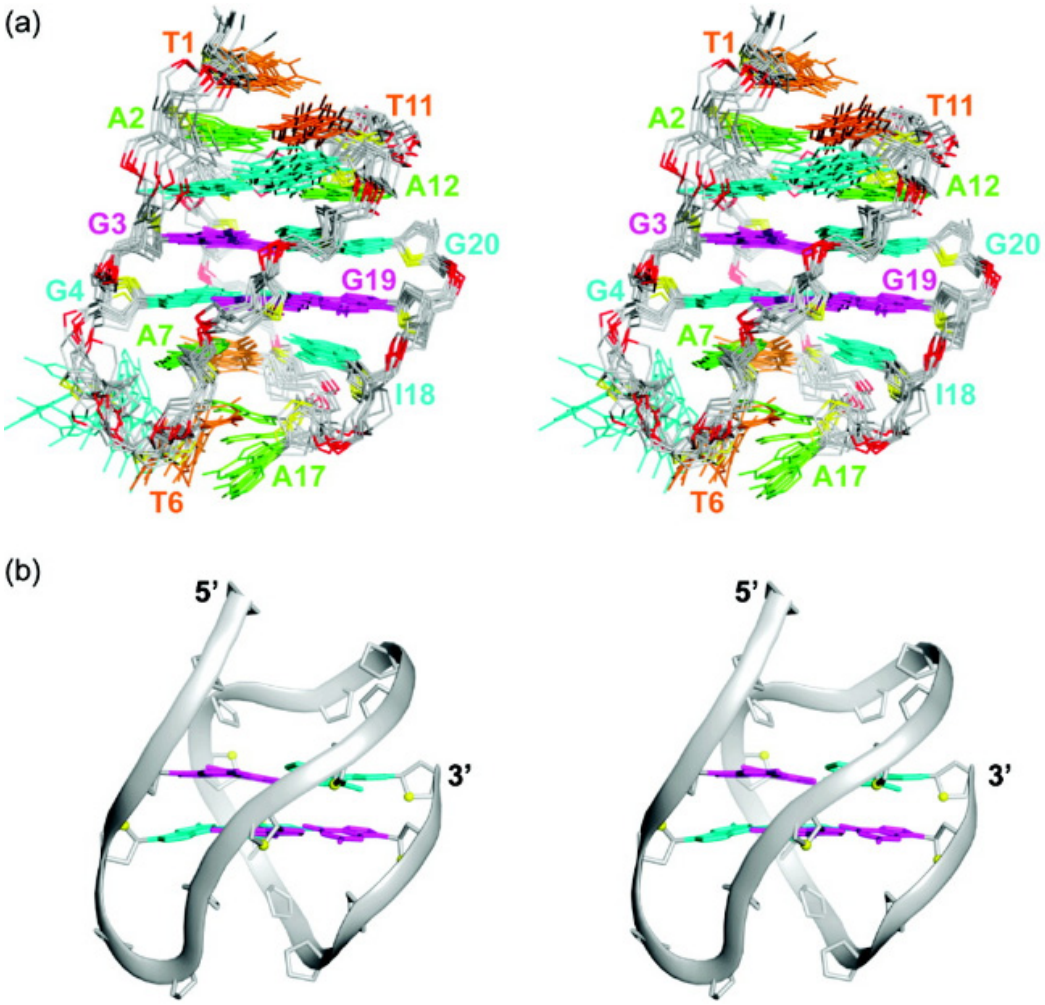

Figure 3.15 Stereo views of the *I18-Form 1* G-quadruplex structure in K^+ solution. (a) Ten superimposed refined structures of *I18-Form 1*. (b) Ribbon view of a representative structure. The *anti* and *syn* guanines are colored cyan and magenta, respectively; adenines are colored green; thymines, orange; inosines, cyan; backbone and sugar, gray; O4' atoms, yellow; phosphorus atoms, red.

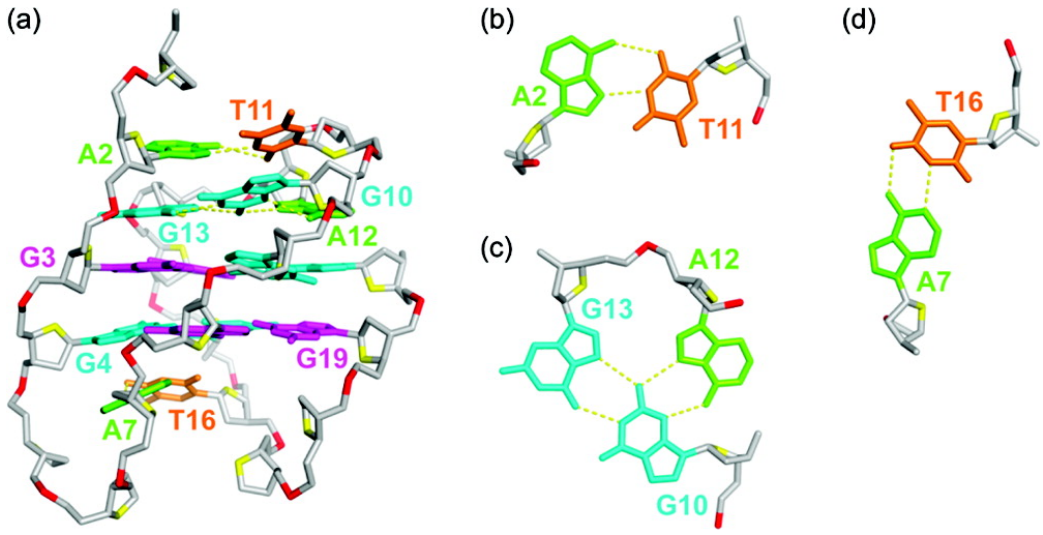

Figure 3.16 Base pairing and stacking in the *I18-Form 1* G-quadruplex structure. (a) Side view of the structure. (b) A2•T11 Hoogsteen base pair. (c) G10•(A12-G13) triad. (d) A7•T16 Watson−Crick base pair. Color coded as in Figure 3.18a. Hydrogen bonds between the base triad/pairs are shown by yellow dotted lines.

Table 3-3 Statistics of the Computed Structures of the 20-nt d[(TAGGG)$_3$TAIGG] Sequence

A. NMR restraints		
distance restraints	D_2O	H_2O
intraresidue distance restraints	252	0
sequential (i, $i + 1$) distance restraints	144	21
long-range (i, $\geq i + 2$) distance restraints	45	37
other restraints		
hydrogen bond restraints	48	
torsion angle restraints	20	
intensity restraints		
nonexchangeable protons (each of four mixing times)	196	
B. Structure statistics for 10 molecules following intensity refinement.		
NOE violations		
number (>0.2 Å)	0.200 ± 0.600	
maximum violation (Å)	0.159 ± 0.035	
rmsd of violations (Å)	0.020 ± 0.002	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.005 ± 0.000	
bond angles (deg)	0.796 ± 0.015	
impropers (deg)	0.452 ± 0.017	
NMR R-factor ($R_{1/6}$)	0.014 ± 0.001	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except G5, T6, and A17	0.96 ± 0.16	
all heavy atoms	1.37 ± 0.21	

Different basket-type intramolecular G-quadruplexes have been reported previously, containing two G-tetrads (formed by human telomeric sequence in K^+ solution[77i] (**Figure 3.17b**)), three G-tetrads (formed by human telomeric sequence in Na^+ solution[36] (**Figure 3.17c**)) or four G-tetrads (formed by *Oxytricha* telomeric sequence in Na^+ solution[93] (**Figure 3.17d**)). These three basket-type G-quadruplexes have similar loop arrangements, which are different from the novel basket-type of the Form 1 *Giardia* telomeric G-quadruplex (**Figure 3.17a**). While the G-tetrad cores of these four G-quadruplexes are oriented similarly according to the groove widths, the 5′-end of these two different basket types starts from different corners (**Figure 3.17**). An alternative way to distinguish these basket-type G-quadruplexes is to consider consecutive loop arrangements. In all basket-type G-qadruplexes, the loops are consecutively edgewise-diagonal-edgewise, but in Form 1 *Giardia* telomeric G-quadruplex the first (edgewise) loop spans across a narrow groove and the third (edgewise) loop spans across a wide groove, while in the other basket-type G-quadruplexes the reverse was observed: the first and third loops span across a wide and narrow groove, respectively.

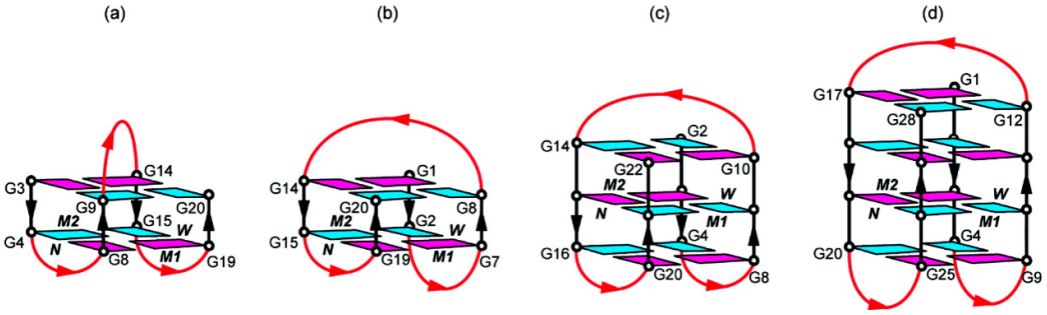

Figure 3.17 Schematic structures of intramolecular antiparallel basket-type G-quadruplexes formed by (a) the sequence d[$(TAGGG)_3TAIGG$] in K^+ solution (this work), (b) the sequence d[$(GGGTTA)_3GGGT$] in K^+ solution,[77i] (c) the sequence d[$AGGG(TTAGGG)_3$] in Na^+ solution,[36] and (d) the sequence d[$(GGGGTTTT)_3GGGG$] in Na^+ solution.[93] *anti* guanines are colored cyan; *syn* guanines are colored magenta. The backbone of the core and loops are colored black and red, respectively.

3.3.4 Structure of Form 2: Propeller-Type G-quadurplex

The H8/H6−H1′ NOE sequential connectivity of *ΔA12-Form 2* is displayed in **Figure 3.18**. All residues were observed to adopt *anti* glycosidic conformation (**Figure 3.18** and **Figure 3.19**). Analysis of imino-H8 connectivity patterns revealed the formation of an intramolecular propeller-type parallel-stranded G-quadruplex with three G-tetrad layers, G3•G8•G12•G17, G4•G9•G13•G18, and G5•G10•G14•G19 (**Figure 3.20**). There are three double-chain-reversal loops formed by single residue or two residues. At high concentration of K^+ and/or DNA, end stacking can occur between two such G-quadruplex blocks to form a higher-order structure[94] consistent with our gel-shift and spectroscopic data (see above).

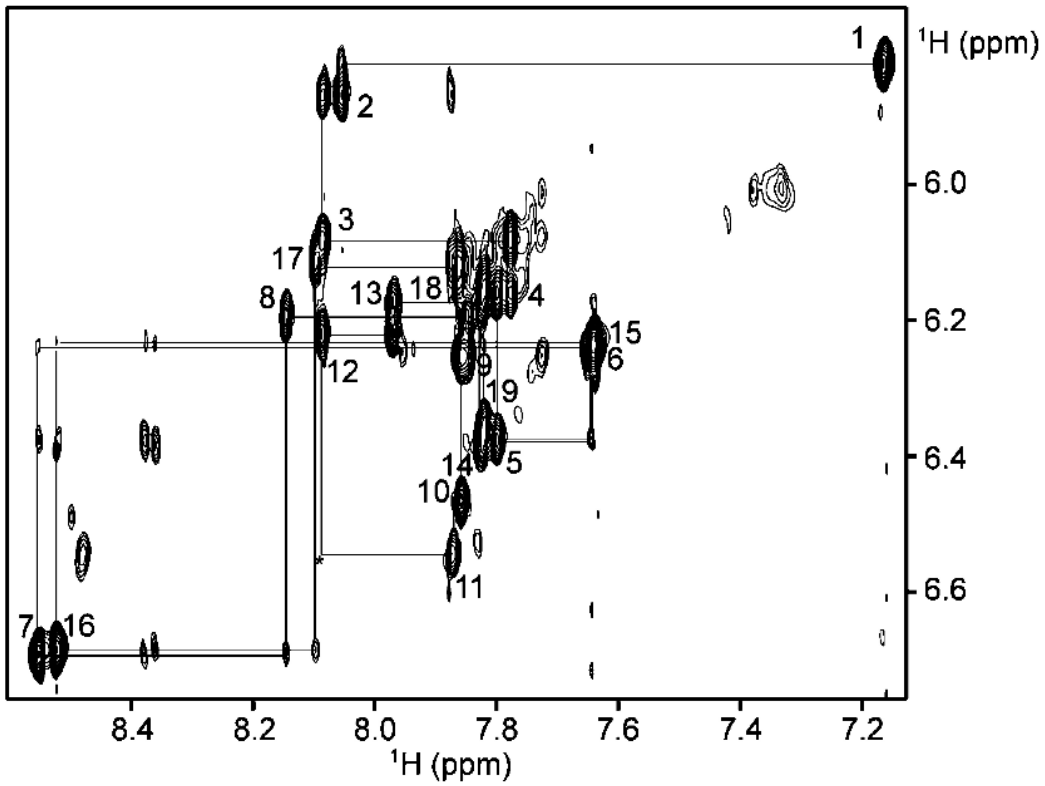

Figure 3.18 NOESY spectrum (mixing time, 350 ms) showing the H8/H6−H1′ connectivity of *ΔA12-Form 2* in K^+ solution. Intraresidue H8/H6−H1′ NOE cross-peaks are labeled with residue numbers. Weak or missing sequential connectivities are marked with asterisks.

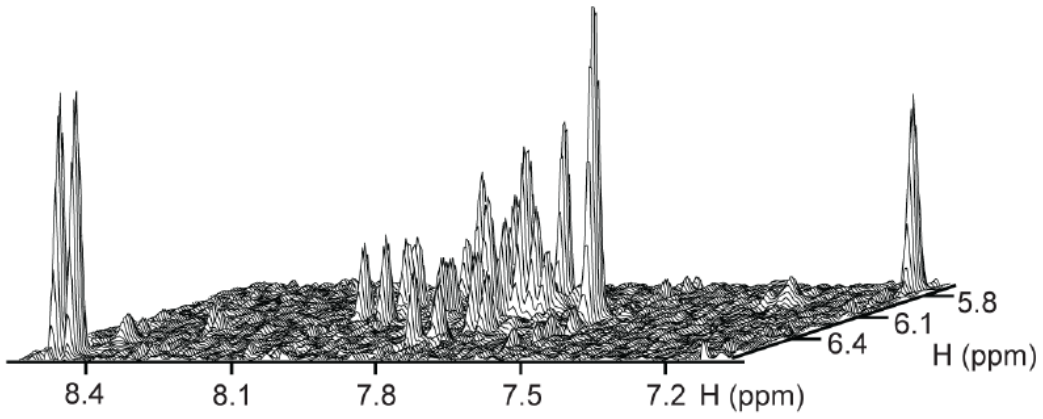

Figure 3.19 Stacked plot of NOESY spectrum (mixing time, 200 ms) of *ΔA12-Form 2*. No strong intraresidue H8/H1' cross-peaks were observed for any guanine, indicating *anti* glycosidic conformation for all guanines.

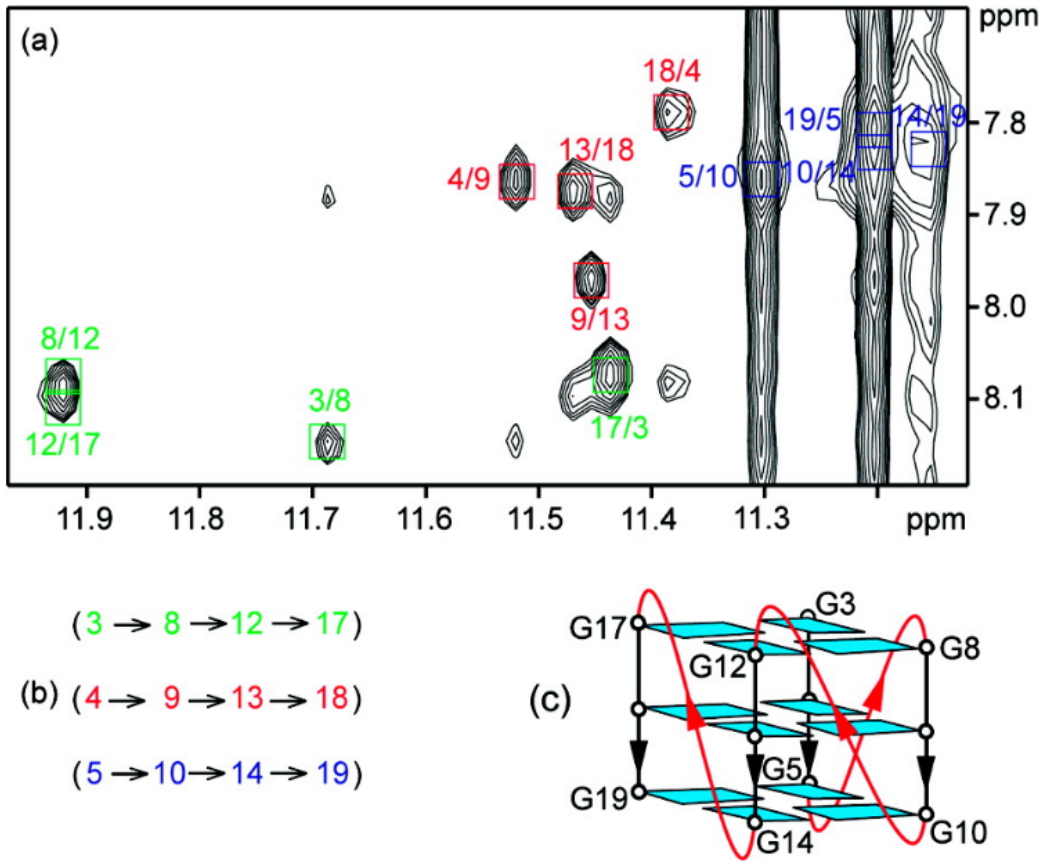

Figure 3.20 Determination of G-quadruplex folding topology. (a) NOESY spectrum (mixing time, 200 ms) showing imino-H8 connectivity of *ΔA12-Form 2*. Cross-peaks that identify the three G-tetrads are framed and labeled with the residue number of imino protons in the first position and that of H8 protons in the second position. (b) Guanine imino-H8 NOE connectivities observed for *ΔA12-Form 2* with G3•G8•G12•G17, G4•G9•G13•G18, and G5•G10•G14•G19 tetrads. (c) Schematic view of the *ΔA12-Form 2* G-quadruplex. The *anti* guanines are colored cyan. The backbone of the core and loops is colored black and red, respectively.

Intramolecular propeller-type parallel-stranded G-quadruplexes have been observed for human telomeric sequences in a K^+-containing crystal[33] but not in dilute solution.[36, 77] In contrast, the propeller-type G-quadruplex is observed for a *Giardia* telomeric sequence, in which each linker is shorter by one T. This finding is consistent with previous observations that shorter double-chain-reversal loops are more stable than the longer ones.[17c, 76, 95]

For the *Giardia* telomeric sequence $d(TAGGG)_4$, high temperatures favor the propeller-type parallel-stranded G-quadruplex over the basket-type antiparallel-stranded G-quadruplex. Similar temperature-dependent behavior was observed for the equilibrium between dimeric parallel-stranded and antiparallel-stranded G-quadruplexes formed by a two-repeat human telomeric sequence in K^+ solution.[34]

Our experimental data on the *Giardia* telomeric sequence $d(TAGGG)_4$ in K^+ solution show that the propeller-type parallel-stranded G-quadruplex form, in which the G-tetrad core's termini are exposed, is more favorable for end-stacking than the antiparallel-stranded G-quadruplex form, in which the G-tetrad core is capped at two ends with loops.

Imino proton spectra of sequences modified from the natural sequence, in which a guanine located in the loops of Form 1 (namely, G5, G10, G13 and G18) was

substituted by an inosine, are plotted in **Figure 3.21**. Except for the G10-substituted sequence, the three other sequences favored Form 1. In view of the structures of Form 1 and Form 2, these modifications do not affect Form 1 but destabilize Form 2. In contrast, the G-to-I substitution at position 10 removes a critical amino group for the G10•(A12-G13) in Form 1, and destabilizes this structure. This observation emphasized the importance of the G10•(A12-G13) in the stabilization of Form 1.

Deletion of A12 from the natural sequence (resulting in *ΔA12-Form 2*) favored Form 2 because the removal of this residue (i) abolishes the G10•(A12-G13) triad, destabilizing Form 1, and (ii) stabilizes Form 2 (a double-chain-reversal loop with one residue is more stable than that with two residues[17c, 76, 95]).

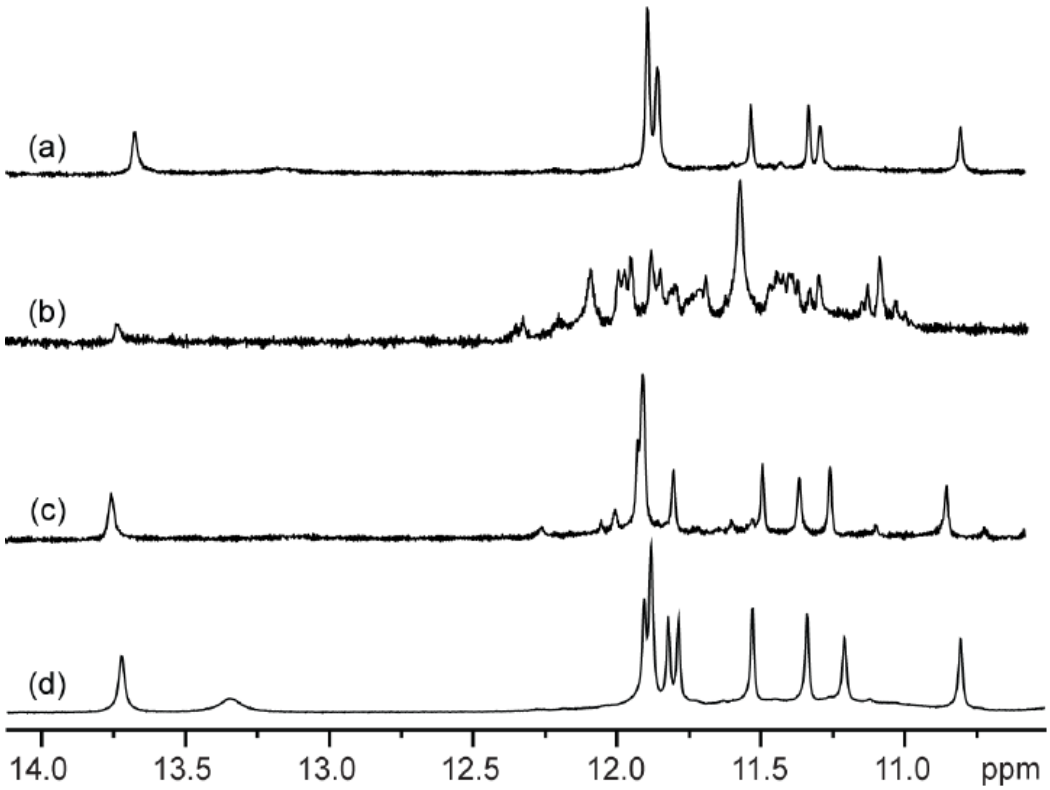

Figure 3.21 Imino proton spectra of (a) d(TAGGITAGGGTAGGGTAGGG), (b) d(TAGGGTAGG**I**TAGGGTAGGG), (c) d(TAGGGTAGGGTA**I**GGTAGGG), and (d) d(TAGGGTAGGGTAGGGTA**I**GG), where **I** represents inosine. Inosine substitution at position 5, 13, and 18 of the 20-nt natural *Giardia* telomeric sequence $d(TAGGG)_4$ all favor the formation of Form 1, whereas inosine substitution at position 10 of the sequence destabilizes Form 1.

We have shown that the four-repeat *Giardia* telomeric sequence $d(TAGGG)_4$ forms two different intramolecular G-quadruplex structures in K^+ solution: the first one is a basket-type antiparallel-stranded G-quadruplex containing two G-tetrads, a G•(A-G) triad, and two A•T base pairs, while the second one is a propeller-type parallel-stranded G-quadruplex containing three G-tetrads. This result shows a dramatic effect of linker length and sequence on the folding topology of G-quadruplexes. Human telomeric sequence, which has the same terminal residues and differs by an extra T in each repeat, forms a (3 + 1) G-quadruplex structure.[12] This is consistent with previous findings that a small change to a G-rich sequence might result in a dramatic change in G-quadruplex topology.[96]

Technically, this work demonstrates again that small sequence modifications can be used to manipulate the relative populations of different G-quadruplex conformations and to favor a single conformation for structural analysis.[34, 77, 87]

3.3.5 Interconversion between G-quadruplexes

At 25 °C, the CD profile of *I18-Form 1* was very different from that of *ΔA12-Form 2* (**Figure 3.22**). The former showed positive peaks at 245 and 295 nm and a negative peak at 265 nm, typical of an antiparallel-stranded G-quadruplex, while the latter showed a positive peak at 260 nm, characteristic of a parallel-stranded G-quadruplex.[74a] Note that the normalized intensity of the 260-nm peak of *ΔA12-Form 2* (red curve, putative parallel G-quadruplex) is roughly three times that of the 295-nm peak of *I18-Form 1* (black curve, putative antiparallel G-quadruplex).

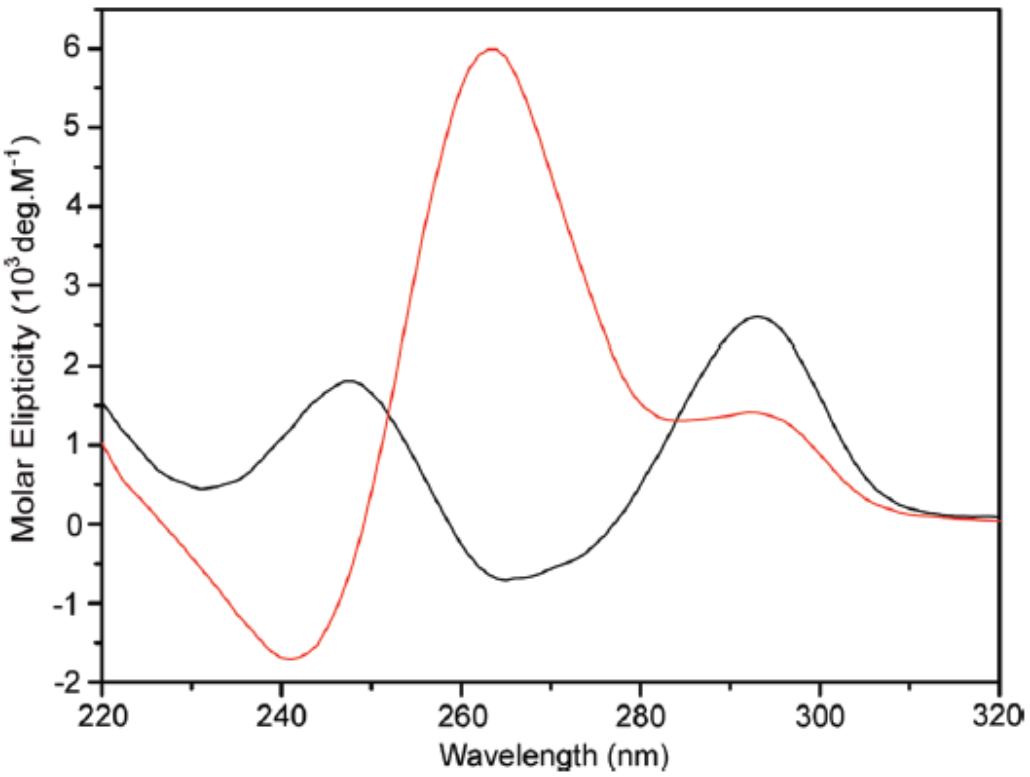

Figure 3.22 Normalized CD spectra of *I18-Form 1* (black line) and *ΔA12-Form 2* (red line) at 25 °C. The former displays the CD signature of antiparallel G-quadruplexes, whereas the latter shows the CD signature of parallel G-quadruplexes.

CD spectra of the natural sequence d(TAGGG)$_4$ recorded at various temperatures (**Figure 3.23**) were consistent with this sequence forming two G-quadruplex structures (as monitored by peaks at 295 and 260 nm, respectively) and that the parallel-stranded form was favored at high temperatures as observed by NMR (see above). We remark that these CD data show that an interpretation of G-quadruplex topology based only on a single CD spectrum might be misleading due to the coexistence of multiple G-quadruplex conformations. A combination of different experimental conditions (temperature, pH, etc.) and sequence modifications can be used to vary the relative populations of different forms, aiding structural interpretations.

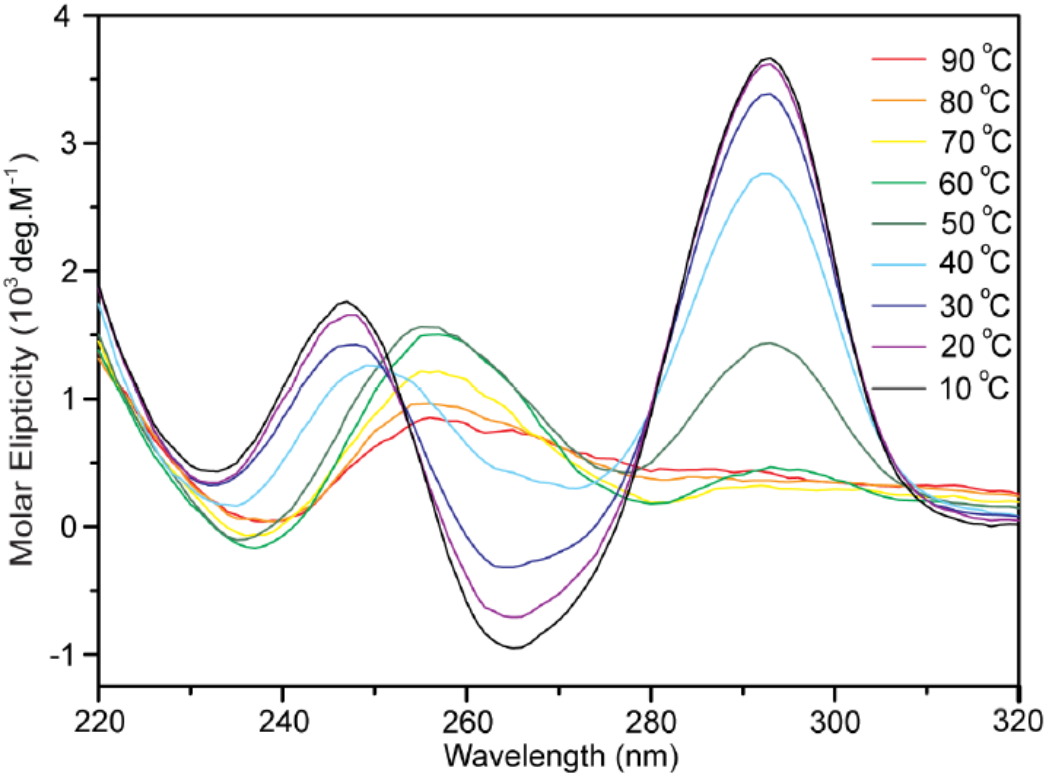

Figure 3.23 Normalized CD spectra of the 20-nt natural *Giardia* telomeric d(TAGGG)$_4$ sequence recorded at different temperatures (color coded on the right).

Formation of Form 1 and Form 2 G-quadruplexes by the natural sequence d(TAGGG)$_4$ and their interconversion could be monitored by CD (**Figure 3.24**). After a quick sample cooling from 90 to 10 °C, formation of the two conformations was observed (as monitored by peaks at 295 and 260 nm, respectively) at similar rates (in the order of minutes), which was followed by their interconversion toward the equilibrium with Form 1 being the major conformation.

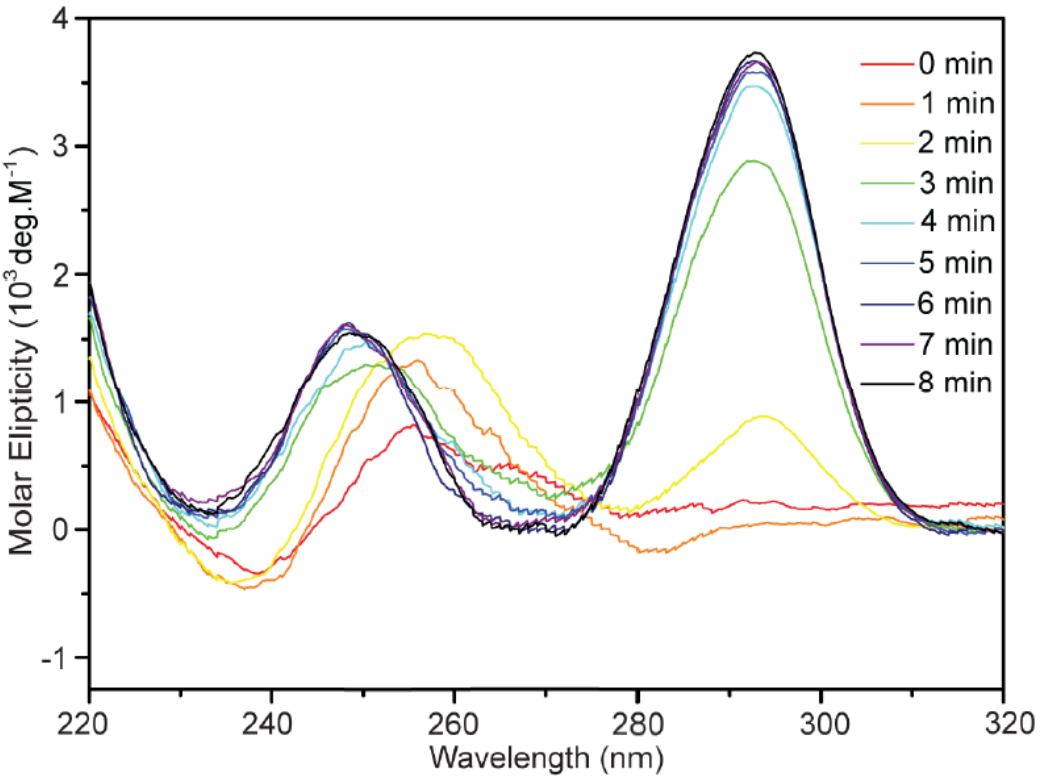

Figure 3.24 Normalized CD spectra of the 20-nt natural *Giardia* telomeric d(TAGGG)$_4$ sequence at different time intervals (color coded on the right) after the sample, initially equilibrated at 90 ^{o}C, was quickly cooled down to 10 ^{o}C.

In this work, we observed the interconversion between two distinct G-quadruplexes for the four-repeat *Giardia* telomeric sequence $d(TAGGG)_4$ in K^+ solution, indicating that these different G-quadruplex conformations are isoenergetic. The interconversion between different G-quadruplexes has been previously observed for *Tetrahymena* and human telomeric sequences.[34, 97] Formation of a mixture of various G-quadruplex forms might be a general property of telomeric sequences,[34, 77, 80-81, 97] as well as other G-rich genomic sequences.[17c, 98]

The kinetics of G-quadruplex formation and unfolding might also be an important factor to determine the relevance of different G-quadruplex forms in different biological processes. Different folding/unfolding kinetics could result in different G-quadruplex populations, when the system is not at equilibrium. Previously, distinct folding and unfolding rates have been observed for two human telomeric G-quadruplexes;[34] relative populations of these two forms could vary during the folding/unfolding processes.[34] For the four-repeat *Giardia* telomeric sequence $d(TAGGG)_4$ in K^+ solution, although Form 1 is the major form at equilibrium at low temperatures, the two forms might have comparable populations after a few minutes of folding time (**Figure 3.24**).

In nature, different proteins or other cellular factors can recognize a particular G-quadruplex form and selectively promote or abolish this conformation as a step within a regulation pathway. It could be interesting to specifically target this structure by small-molecule ligands with high affinity toward only one form. In these cases,

proteins or small-molecule ligands should recognize and distinguish between different structural elements of G-quadruplexes, such as G-tetrad cores and loops.

3.4 Conclusion

Recurrence of several structural elements was found in the structures of Form 1 and Form 2 *Giardia* telomeric G-quadruplexes. (i) Single nucleotide has been reported to form a very stable double-chain-reversal loop;[17c, 76, 95] such a loop was found to stabilize Form 2 in *ΔA12-Form 2*. (ii) Multiple base pairs and base triads have been observed in the loops of G-quadruplexes and stabilize these structures by stacking continuously on the G-tetrad cores;[77c-i, 81] a Hoogsteen A•T base pair, a G•(A-G) triad, and a Watson−Crick A•T base pair were observed to stack at the top and bottom of the G-tetrad core and stabilize Form 1. (iii) The G•(A-G) triad in Form 1, formed in the diagonal loop G10-T11-A12-G13 (single-nucleotide turn is underlined) (**Figure 3.25a**), matches well to a G•(A-G) triad previously observed in the *MYC* promoter,[98] formed in a diagonal loop G-A-A-G (**Figure 3.25b**). Although the G-tetrad cores of the two structures are quite different, the configurations of the triad-containing diagonal loops are very similar: the first (G), third (A) and fourth (G) bases in the loop participate in a G•(A-G) triad, while the second base (T or A) forms a single-base turn and stacks on top of the triad. The adjacent G-tetrads in the two structures are different (with glycosidic conformations being *syn•syn•anti•anti* and *syn•anti•anti•anti*, respectively), but the configurations of the connection points are

similar: both loops connect an *anti* guanine to a *syn* guanine and span the same diagonal distance across the corresponding G-tetrad. The recurrence of different structural elements in these structures suggests a "cut and paste" principle for the design and prediction of G-quadruplex topologies, for which different elements could be extracted from one G-quadruplex and inserted in another one. However, one should be cautious that the formation of a particular G-quadruplex "structural element" by a "sequence element" might also be context-dependent due to the presence of different competing conformations.

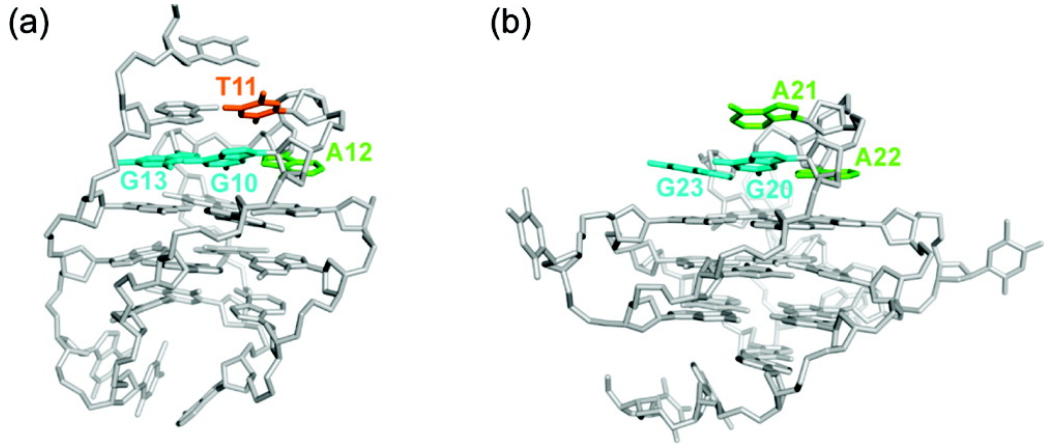

Figure 3.25 Recurrence of G-quadruplex structural motifs. G•(A-G) triad observed in (a) *I18-Form 1* (this work) and (b) *Pu24I* (pdb code: 2A5P). Only the bases from the diagonal loops are highlighted, all other atoms are colored gray.

Chapter 4 Self-assembly of DNA supramolecular wires on graphene

4.1 Introduction

Graphene, an isolated single atomic layer of hexagonally arranged carbon atoms, is a basic building block for some carbon allotropes including fullerence, carbon nanotubes and graphite. As a model system for condensed matter physics[99] and the formidable potential of future applications in nanoelectronics,[100] it has stimulated intense research interest. Considering the large surface area of graphene,[101] a high loading capacity for biomolecules could plausibly be expected. Graphene and graphene-based nanomaterials have been documented to be interfaced with DNA,[102] proteins[103] and even cells.[104] A thorough knowledge and understanding of the DNA-graphene interaction has important implications for the interpretation of DNA interactions with carbon-based substances. This is relevant to efforts dedicated over the past decade to DNA-assisted sorting of structure-specific carbon nanotubes,[105] seamlessly rolled up graphene sheets, DNA sequencing with graphene-based nanochannel device,[106] and DNA sensing with carbon nanotubes,[107] graphene[108] and graphene derivatives.[109]

Most previous investigations focused on the interaction between graphene and single-stranded/ duplex DNA.[110] Besides the canonical double helix, guanine-rich

nucleic acid sequences can adopt a four-stranded helical structure called the G-quadruplex through stacking of multiple G-tetrads, each of which arises from the planar association of four guanines in a cyclic Hoogsteen hydrogen-bonding arrangement.[6] Because of these peculiar structural features, G-quadruplex DNA exhibits higher structural stability,[57a] rigidity[57b] and conductivity[28,29] than duplex DNA.[63, 111 63, 111 63, 111 93,168 93,167 93,166 93,165 93,165 93,165 93,165 93,164 93,164 93,164] These advantages, along with the ability of forming extended wires of hundreds nanometers, G-wires, possibly by slippage in G-quadruplex DNA,[112] make this structure appealing for the development of biomolecular electronics.[64] Here we report on the formation of G-wires by the d(GGGGTTGGGG) oligonucleotide on graphene. Our experimental results demonstrate that G-wires are self-assembled into well-ordered arrays and preferentially oriented along the armchair direction.

4.2 Experimental Section

Sample preparation DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer as previously described.[74b] Reagents for syntheses including the dSpacer,[113] a deoxynucleotide lacking the nucleobase, were purchased from Glen Research, Sterling, VA. DNA concentration was expressed in strand molarity using a nearest-neighbour approximation for the absorption coefficients of the unfolded species.[84]

G-wires were grown with 20 μg/μl (6.29×10^{-3} μM) of the d(GGGGTTGGGG)

oligonucleotide dissolved in appropriate buffer, treated by heating at 95°C for 10 min, followed by gently cooling at a rate of 0.3°C min^{-1} and incubation at 37°C for 24 hrs. The buffer contained 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, and 1 μM spermidine, supplemented with 50 mM of KCl, or NaCl, or LiCl. Unless otherwise specified, incubation buffer containing 50 mM KCl was used for the growth of G-wires.

500-bp DNA duplex was obtained by isolating from "100-bp DNA ladder" (Promega) on a 3% agarose gel. The 500-bp DNA fragment was recovered from agarose using a QIAquick gel extraction kit (Qiagen). Duplex DNA was diluted with 50 mM KCl incubation buffer to a concentration of 4 ng/μl and heated ~10 min at 65°C.

Graphene samples were prepared by mechanical exfoliation of highly ordered pyrolytic graphite (HOPG) on top of an oxidized Si wafer using the established "scotch tape" technique[100].

Circular dichroism (CD) CD spectra were recorded on a Jasco J-815 spectropolarimeter using a 1-cm path-length quartz cuvette. Formation of G-quadruplexes in solution was followed as a function of incubation time using CD spectroscopy.

Raman spectroscopy Raman spectra and images were acquired by using a WITec CRM200 confocal microscopy Raman system in the backscattering configuration. The excitation laser (532 nm) was focused onto the sample using a 100× objective

lens (N.A. = 0.95). The piezostage movement and data acquisition were controlled using Scan Ctrl Spectroscopy Plus software from WITec GmbH, Germany.

Atomic force microscopy (AFM) 10 μl of the sample was deposited onto the substrate surface. After 3 min adsorption, the substrate was gently rinsed with deionized water and dried with nitrogen gas. Atomic force microscopy was performed under ambient conditions, using a Veeco Dimension V (Veeco Instruments, Santa Barbara, CA) system in tapping mode. Topographic images were taken at a scanning rate of 1 Hz and resolution of 768 × 768 lines. Height analysis and angle measurement were carried out by the accompanying Veeco Nanoscope software (version 7.20). Two-dimensional fast Fourier transform by means of Digital Micrograph software (Gatan, Inc., Plasanton, CA) was used for estimation of the preferred direction of G-wire orientation on graphene.

4.3 Results and Discussion

4.3.1 G-wire Formation on Graphene Imaged by AFM

Previously, the DNA sequence d(GGGGTTGGGG) has been shown to form higher-order structures in solution[65] and to self-assemble on the mica surface as G-wires.[112] One generally accepted explanation for the adsorption of DNA to the mica is due to the “salt bridge” effect. Divalent or higher valence cations mediated the interaction between the negative phosphate groups of the DNA backbone and the

negatively charged mica surface[114]. Here we investigate the self-assembly of this sequence on the graphene surface by using the tapping-mode AFM (Error! Reference ource not found.). Wire-like structures of ~3.5 nm high were observed only for samples incubated in KCl or NaCl solution, but not for samples in LiCl solution, for which only structures of ~1.0 nm high were observed (Error! Reference source not ound.). As it has been shown that G-quadruplexes are stable in K^+ and Na^+ solution, but not in Li^+ solution,[6a] it is likely that the 3.5-nm-high wires (called G-wires) observed in K^+ and Na^+ solution are based on G-quadruplexes, while the structures observed in Li^+ solution are single-stranded DNA.[115] Another control is the d(TTTTTTTTTT) sequence in K^+ solution, which form structures of ~1.0 nm high on the graphene surface (**Figure 4.2**). It should be noted that the height of G-wires on top of the mica surface is ~2 nm.[112] Formation of G-quadruplexes by d(GGGGTTGGGG) in solution was supported by the CD spectra (**Figure 4.3**), which display a positive peak at 260 nm, characteristic of parallel-stranded G-quadruplexes.[74a] A possible scenario for G-wire formation involving parallel-stranded slipped strands has be postulated previously.[112] A model of such a G-wire on graphene is shown in **Figure 4.4**.

Theoretic Molecular Dynamics modeling is our future work. The proposed structure model is based on the structural knowledge of G-quadruplex DNA. Some Molecular Dynamics modeling research have been done by other research groups to discuss the structures and interaction energies of stacked graphene–nucleobase complexes.[116, 117] The geometry optimization was performed. The relative orientation of different

nucleobases is different. We assume that the dT to dU substitution might affect the relative orientation of G-wire to graphene.

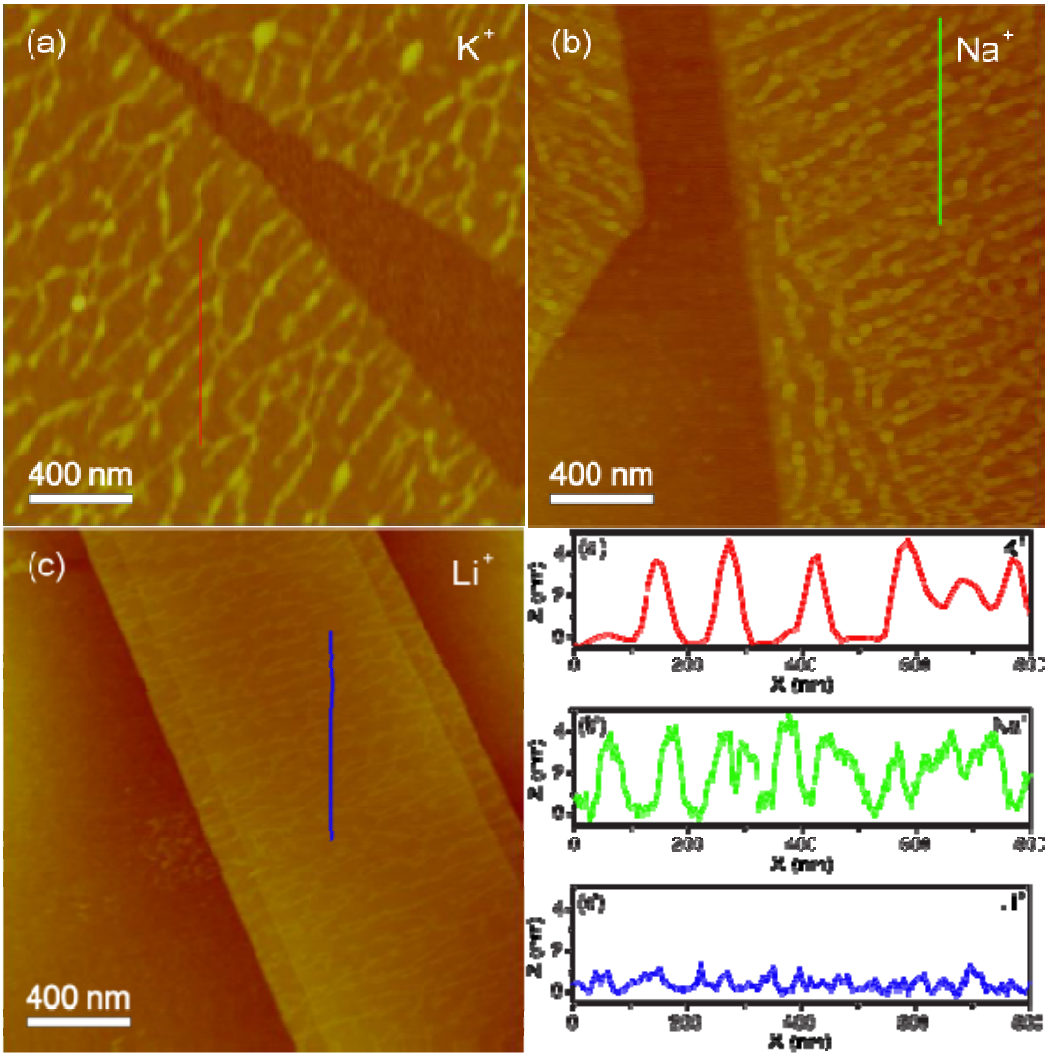

Figure 4.1 Tapping-mode AFM height images of G-wires grown with 20 μg/μl of the d(GGGGTTGGGG) oligonucleotide dissolved in buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$ and 1 μM spermidine, supplemented with different salt (a) 50 mM KCl, (b) 50 mM NaCl, and (c) 50 mM LiCl, (a'-c') The corresponding height profiles along the lines shown in the AFM images.

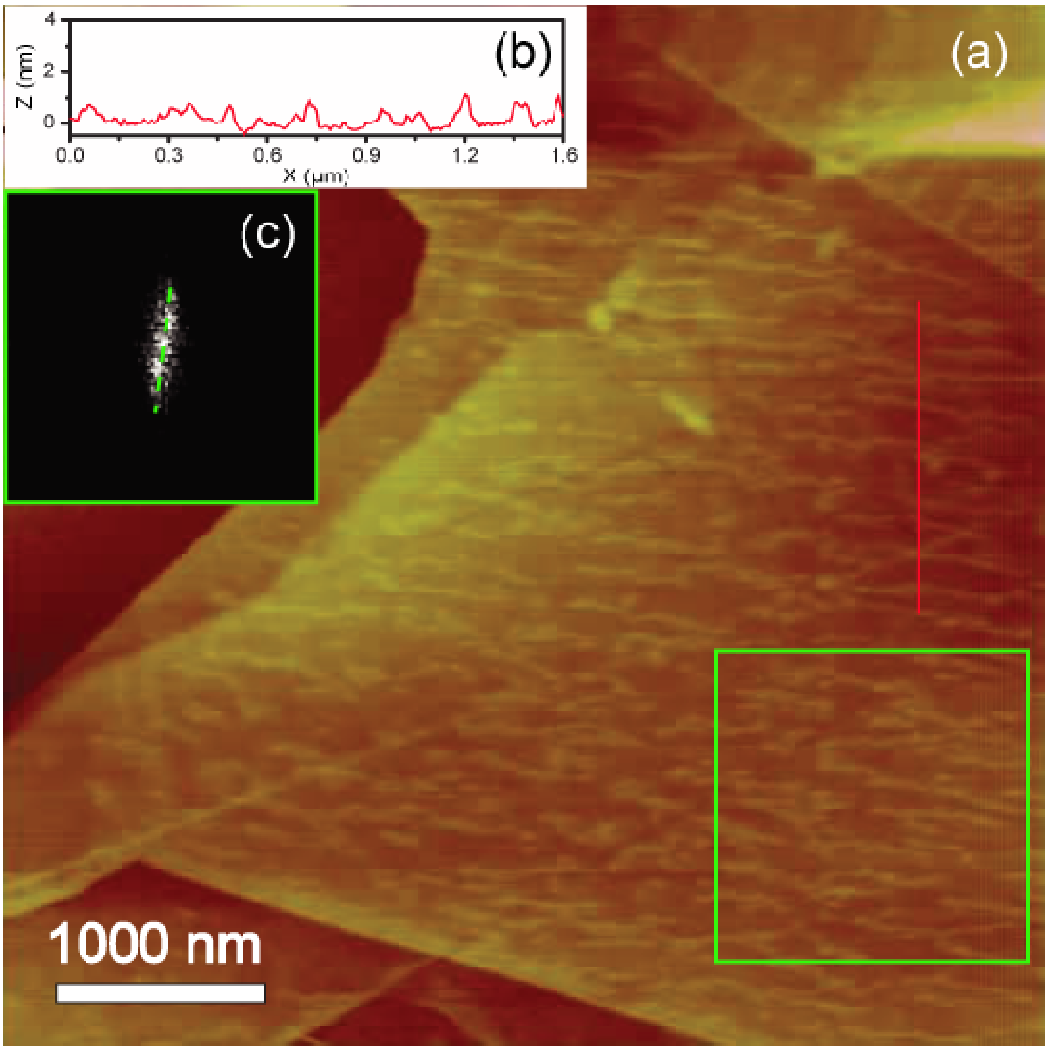

Figure 4.2 (a) Tapping-mode AFM height images formed by the oligonucleotide d(TTTTTTTTTT) on graphene. (b) The corresponding height profiles along the red line shown in the AFM image. (c) Two dimensional fast Fourier transform of the boxed regions indicated by the green square in the AFM image, showing the preferred orientation.

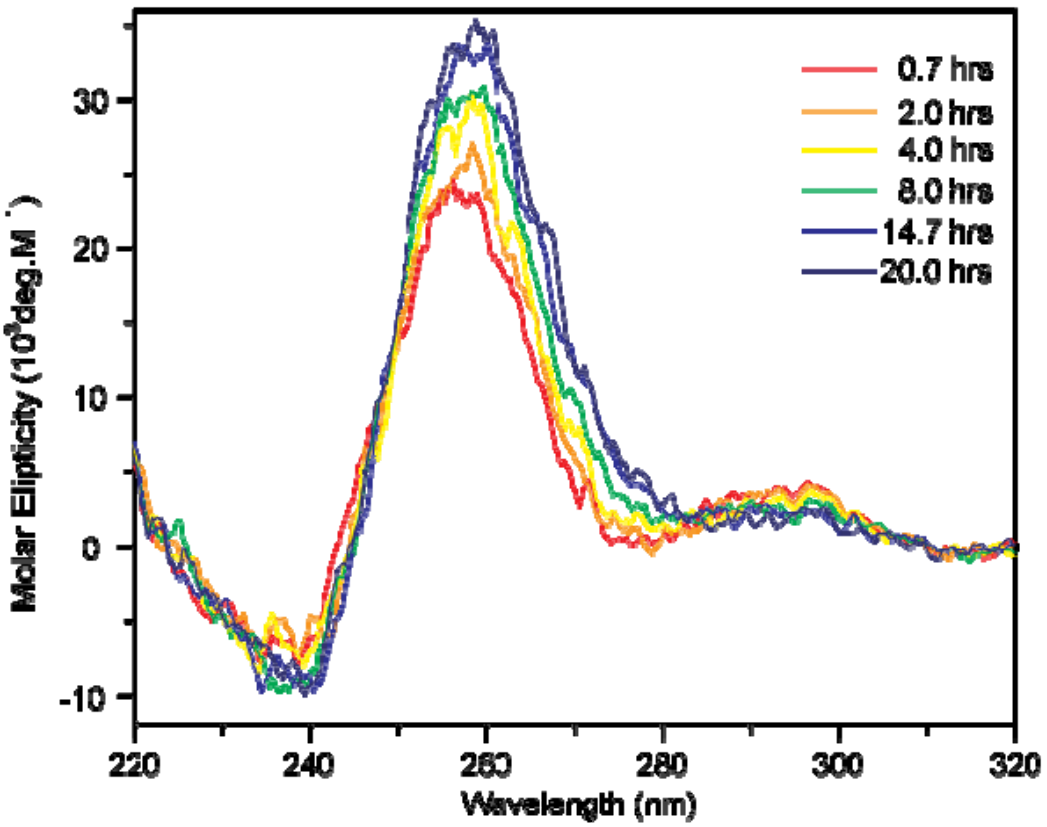

Figure 4.3 CD spectra of d(GGGGTTGGGG) in K^+ solution as a function of incubation time.

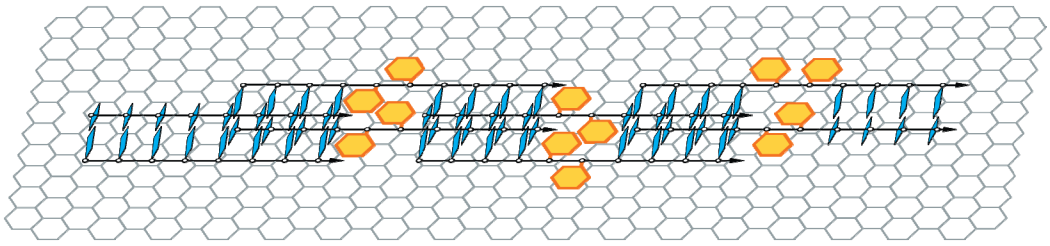

Figure 4.4 Model of a G-wire, formed by the oligonucleotide d(GGGGTTGGGG), self-assembled into a well-ordered structure on graphene with the aid of projected thymines. Graphene is colored gray; guanines, cyan; thymines, orange; and DNA backbone, black.

4.3.2 Alignment of G-wires on Graphene

AFM images reveal that the G-wires form well-ordered arrays once deposited onto graphene (**Figure 4.5**). Fourier transformation (FT) of the region in the red box shows a single preferred direction (**Figure 4.5b**), while FT of the region in the green square reveals a sharp azimuthal distribution (**Figure 4.5c**). The orientations of G-wires can thus be identified, each separated by 60 degrees with respect to the others, reflecting the lattice symmetry of graphene.

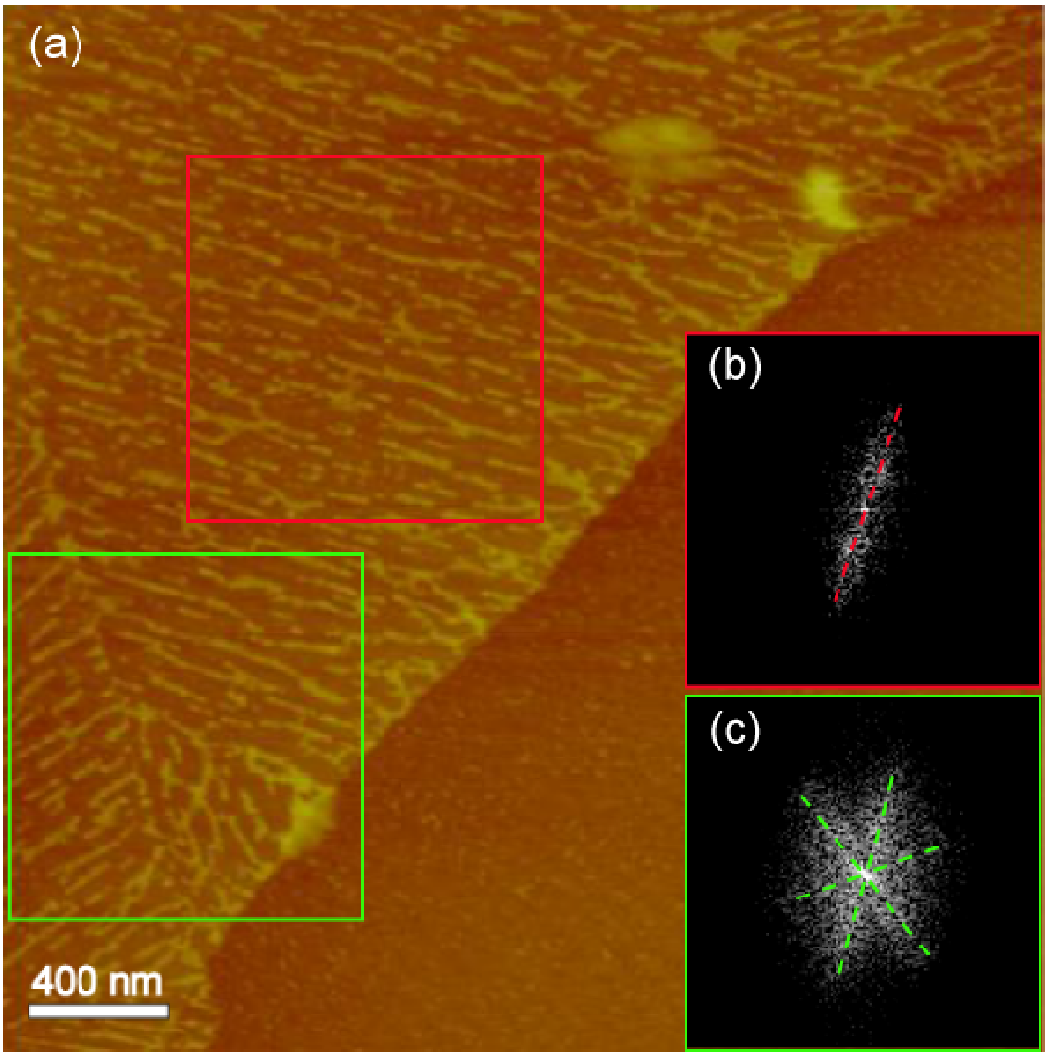

Figure 4.5 (a) AFM image showing G-wire alignments atop graphene. (b-c) Two-dimensional fast Fourier transform of the boxed regions indicated by the red and green square in the AFM image, showing well-defined orientation of the G-wire.

In a model previously proposed by Marsh et al.,[112] a G-wire formed by d(GGGGTTGGGG) can be viewed as an array G-quadruplex units connected by TT linkers. As guanine bases are involved in the formation of the core of the G-quadruplexes, it should be the thymine bases of the linkers that are mainly responsible for the interaction between the DNA and graphene via π–π stacking interactions between the nucleobases and the hexagonal cells of the graphene.[110, 118] The optimized stacking arrangement has been obtained by means of density functional theory (DFT) calculations.[116-117] In the optimized structure of graphene-nucleobase complex, nucleobase stacks parallel to the graphene with the carbonyl oxygen atoms and most of the nitrogen atoms located above the centers of the graphene hexagonal carbon rings. Consequently, we could have TT linkers acting as anchors to maintain the G-wires on graphene, leading to the organization of G-wires into highly ordered architectures atop graphene.

To test the role of the TT linkers for the development of the well-ordered assembly of G-wires, we studied the structure formed by a mutated sequence where two thymine bases were removed. AFM result of the mutated oligomer d(GGGGSSGGGG), where S is a dSpacer without a base, showed that the ordered arrangement of G-wires was disturbed and no preferred orientation was observed (**Figure 4.6a**). As a control, structures formed by a 500-bp duplex DNA deposited on graphene were also observed to be randomly distributed (**Figure 4.6b**). These results support the mechanism, in which the TT linkers, served as a fixation point for G-wires, are a

critical factor affecting the patterning behavior of G-wires on graphene.

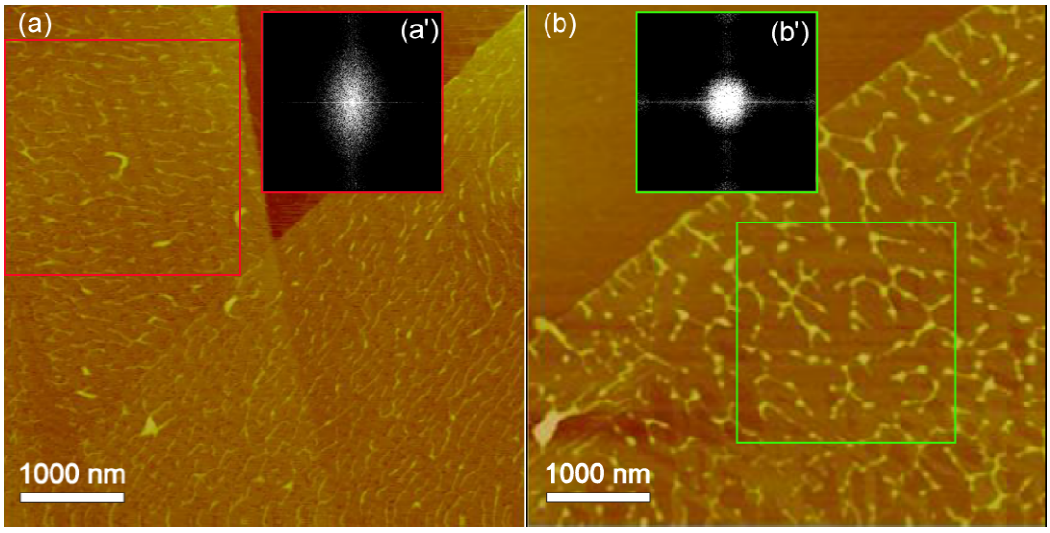

Figure 4.6 AFM topographic images of (a) structures formed by the modified oligonucleotide d(GGGGSSGGGG) incorporating dSpacer as the linker sequence and (b) 500-bp duplex DNA on graphene. Fast Fourier transform of the boxed regions indicated by the red square (a') and green square (b'), shown in the insets.

We also attempted to form supramolecular DNA structures on the surface of graphene oxide (GO). However, we did not observe the formation of G-wires on the surface of GO (for example, see **Figure 4.7**).

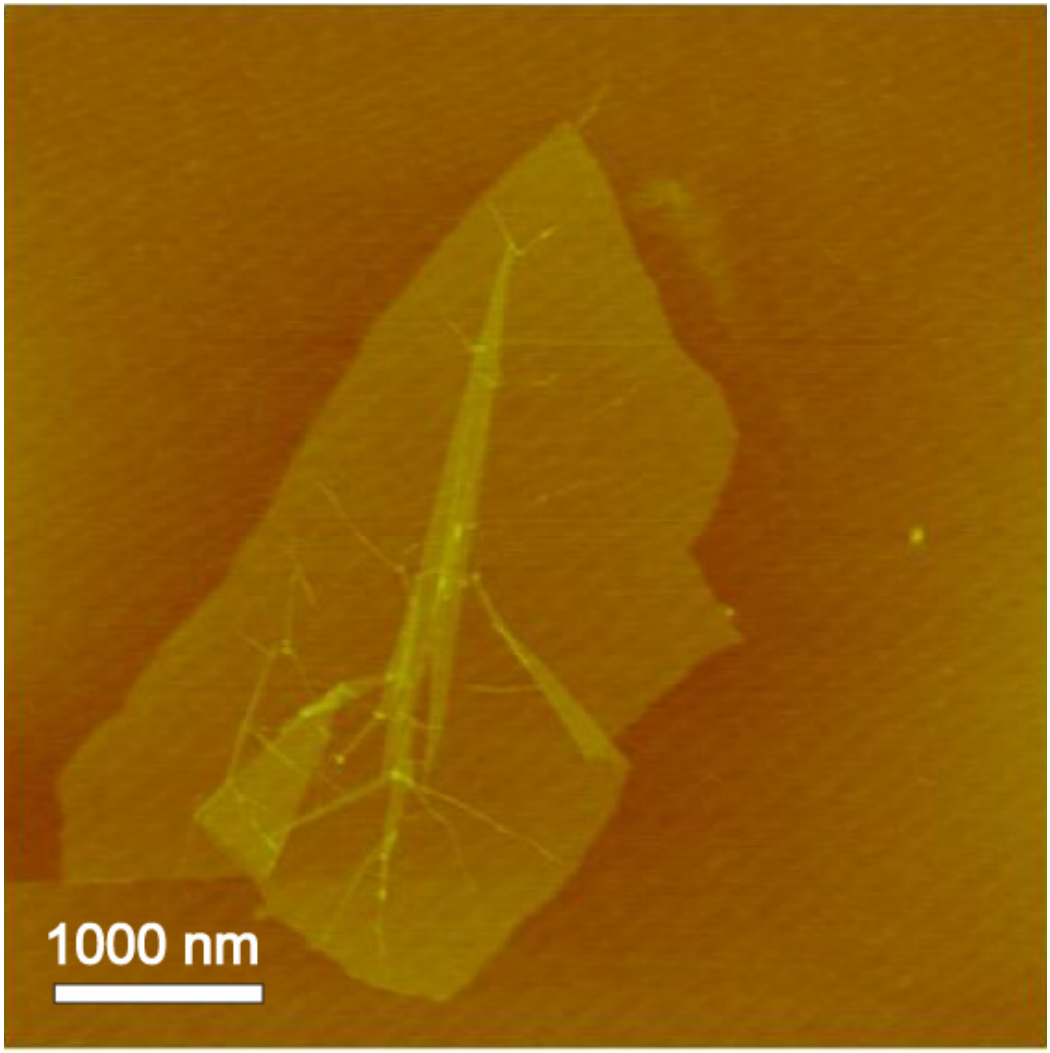

Figure 4.7 AFM image of graphene oxide after incubation with d(GGGGTTGGGG).

The ability to functionalize graphene with robust, well-ordered G-wire arrays presents opportunities for fields ranging from graphene-based biomolecular electronics to sensor design. Raman spectra measurements were carried out before and after patterning G-wires on top of graphene (**Figure 4.8**). It should be noted that G and 2D bands exhibited no obvious shift and no D band was developed, indicating that graphene was essentially unaffected by doping after functionalization.[119] Additionally, G-wire monolayer could also be expected to be utilized as a template for further chemistry or materials deposition.[120]

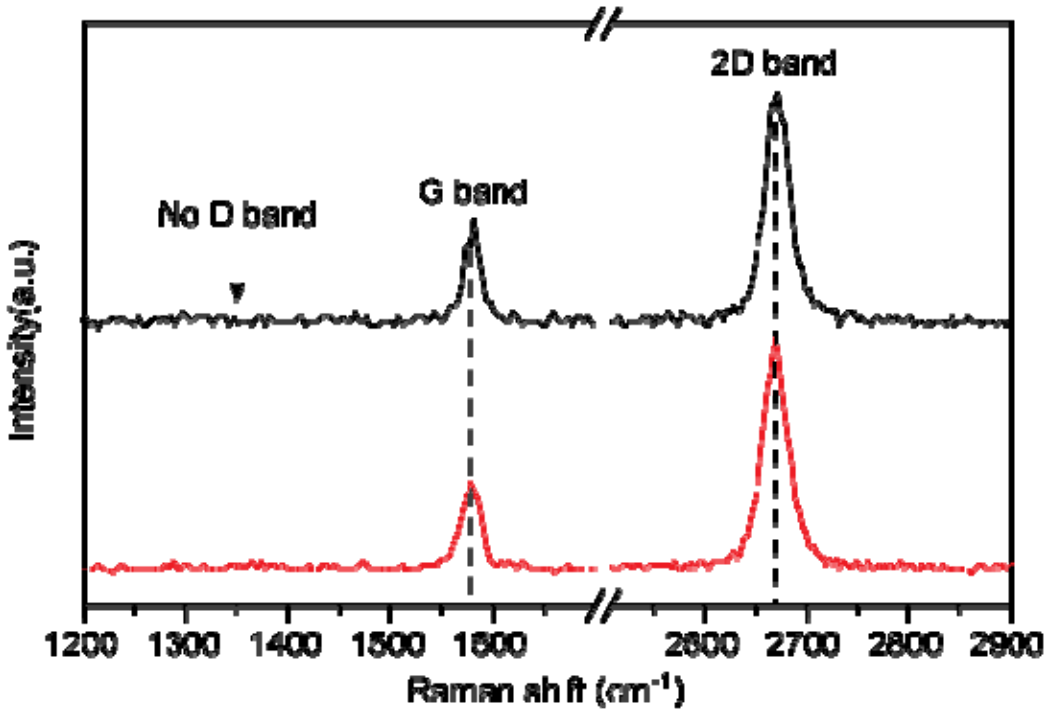

Figure 4.8 Raman spectrum collected before (black curve) and after (red curve) the deposition of G-wires on the graphene.

4.3.3 Orientation of G-wires with respect to Graphene Crystallographic Lattice

It is known that the crystallographic orientation of the edges plays an important role in determining the graphene electronic properties. As a versatile and nondestructive technique, Raman spectroscopy is widely used to characterize the microstructure of carbon-based materials[121]. For example, the D band, which is related to double resonance Raman scattering and requires scattering between two nonequivalent cones, will not be activated by a perfect zigzage edge[122]. So the edge chirality can be identified by using polarized Raman spectroscopy.

In order to determine the relative orientation of G-wires with respect to the graphene substrate lattice, the crystallographic orientation of graphene is identified using polarized micro-Raman image constructed by the intensity of D peak. It is commonly accepted that the D band at 1350 cm^{-1} is related to the intervalley double-resonance scattering[121] and the armchair edges are expected to give rise to a stronger D peak compared to that of the zigzag edges.[122-123]

Figure 4.9 shows a piece of graphene, in which two adjacent edges form an angle of 30 degree, indicating that they have different edge chiralities. Based on the D-peak intensity, the brighter edge is armchair while the other one is zigzag. The crystallographic arrangements of graphene thus can be identified. Thereafter, we determined the relative orientation of each G-wire fragment with respect to the atomic

lattice of graphene substrate using the following formula: $\varphi = n \times 60^{o} + \theta$, where n denotes an integer and θ is defined as the angle between each G-wire fragment and the zigzag direction. The various orientation angles φ were chosen within one symmetric period (60 degree). It was observed that the frequency distribution histogram of the orientation angles φ follows a Gaussian distribution, with the mean value around 30 degree (**Figure 4.9d**). This suggests that G-wires are oriented along the armchair direction on the graphene surface. We also performed polarized Raman scattering studies on oriented G-wires atop another piece of graphene, where the angle between two adjacent edges is 150 degree (**Figure 4.9e-h**). The statistical result is in accordance with the previous conclusion.

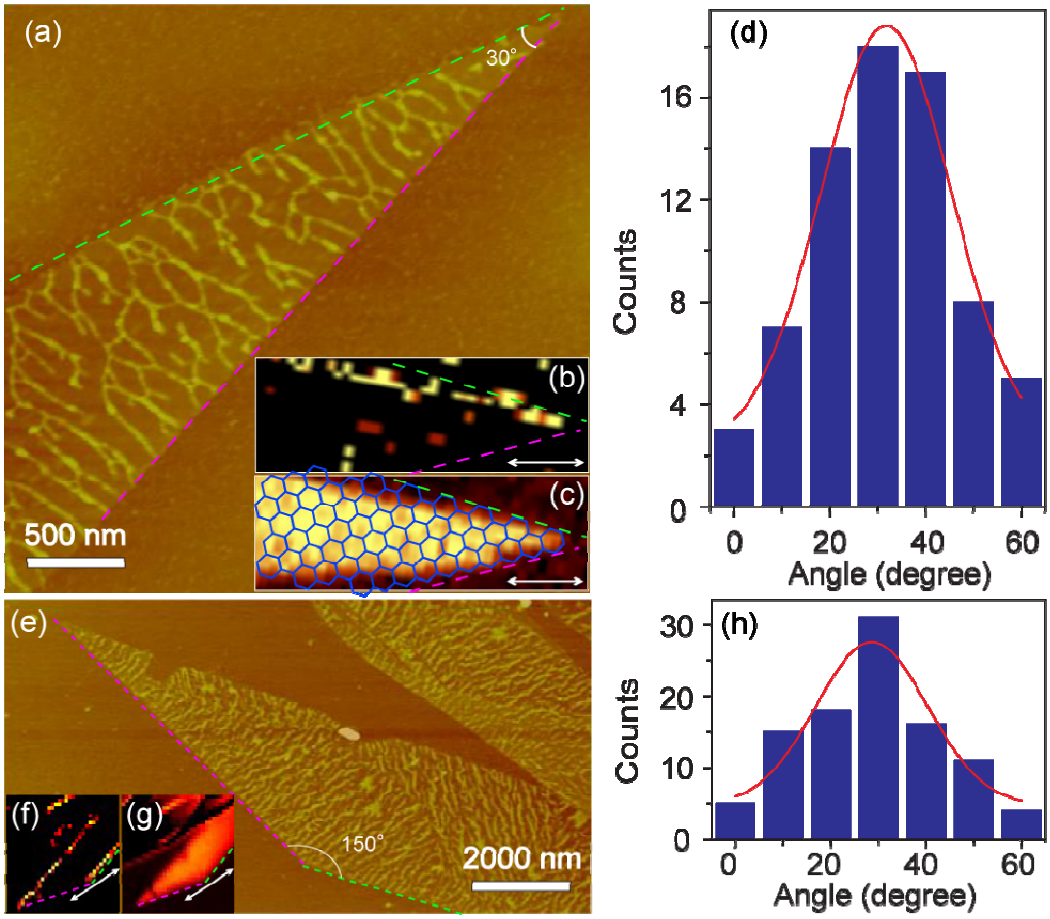

Figure 4.9 Two adjacent graphene edges form an angle of 30 degree (a-d) and 150 degree (e-h). (a, e) AFM image of graphene after immobilization with G-wires. Inset Raman images of the same piece of graphene constructed by the intensity of (b,f) D peak and (c,g) G peak with laser polarization direction indicated by the arrow. Dashed lines mark the general contour of the crystallographic orientation at the graphene edges as follows: armchair (green), zigzag (magenta). (d,h) Histogram showing the orientation distribution of the G-wires with respect to the graphene crystallographic lattices.

Thus, the sensing ability of G-wires can be used to determine the crystal orientation and hence the edge chirality of graphene. The physics of graphene orientation is the focus of extensive research,[124] which account for novel physical phenomena such as superconductivity,[125] ferromagnetism,[126] and quantum Hall effect.[127] Conventional methods for acquiring atomic resolution information,[128] such as scanning tunneling microscope[129] and transmission electron microscopy,[124b] require stringent experimental conditions like low temperature, high vacuum or sample conductivity. Recently, Raman spectroscopy has been explored as a diagnostic tool to characterize the graphene edge structure, but its sensitivity strongly depends on laser polarization, relative position of the laser spot with respect to the edge, and especially the amount of edge disorder.[121-123] Selectively oriented G-wires provide an easy method for probing the graphene orientation, which would accelerate the pace of applications of graphene-based devices, e.g. graphene nanoribbon semiconductors.[130]

4.4 Conclusion

In summary, supramolecular G-wires could be grown from the d(GGGGTTGGGG) oligonucleotide on graphene surface. We found that the G-wires self-assembled into well-ordered arrays on graphene, where thymine bases in the TT linkers served as anchors for the assembly. Using atomic force microscopy in combination with micro-Raman mapping we showed that G-wires preferentially oriented along the armchair direction of graphene. The formation of G-wires could open new avenues for graphene research and application.

Chapter 5 Summary and Perspectives

5.1 Summary of the Dissertation

G-quadruplex DNA is highly polymorphic, which might be associated with different cellular process. A thorough understanding of the G-quadruplex structure is essential for structure-based anti-cancer drug design[16]. In order to explore the effect of loop length and sequence on the folding topology of G-quadruplexes, high resolution structure of *Giardia* telomeric sequence $d(TAGGG)_4$ was solved by nuclear magnetic resonance spectroscopy. Compared to the human counterpart $d[TAGGG(TTAGGG)_3]$, the only difference lies in one T is deleted within the non-G-linker in each repeat. Four-repeat human telomeric sequence $d[TAGGG(TTAGGG)_3]$ has been shown to form (3+1) intramolecular G-quadruplex structure in solution containing K^+ ions[77c]. However, two different intramolecular G-quadruplex folding topologies are demonstrated to be adopted by *Giardia* telomeric sequence in K^+ solution. One is a novel antiparallel basket-type G-quadruplex containing only two layers of G-tetrads, another one is a parallel propeller-type G-quadruplex. Recurrence of several structural elements in the observed structures suggests a "cut and paste" principle for the design and prediction of G-quadruplex topologies.

Besides the important biological role of G-quadruplex DNA, it can also serve as a versatile building block for nanomaterials. G-wires, grown from oligonucleotide d(GGGGTTGGGG), are self-assembled into well-ordered arrays on graphene. The

preferred orientation of G-wires atop graphene, was investigated using a combination of AFM and micro-Raman spectroscopy and found to be along the armchair direction. Based on the structural knowledge of G-quadruplex DNA, the self–assembly properties can be controlled through sequence modification. Such assembly could open new avenues for graphene research and application.

5.2 Perspectives

Mechanical properties of DNA play an essential role in cellular process which binds or reads DNA for purposes of packaging, recognition and modification. Despite the wide interest in the topology of G-quadruplexes, their mechanical properties are not well-understood. The mechanical properties of G-quadruplex DNA can be measured through single-molecule mechanical studies. In contrast to ensemble-based techniques, single molecule measurements make the determination of low probability events or subpopulations and not synchronizable dynamics possible. As both single molecule studies and G-quadruplex are relatively new research fields, as far as we know, only two studies are reported on G-quadruplex single molecule mechanical studies[131]. The evidences provided by both studies are not persuasive as the structure of corresponding G-quadruplexes is not fully controlled. *Giardia* work demonstrates that structure of G-quadrupelx can be leveraged through sequence modification and experimental conditions. Coupled with our research group core strength in NMR

structural studies, we can verify the conformation of G-quadruplex and complement data from single molecule studies.

As graphene is a basic building block for some carbon allotropes, the DNA-graphene interaction study will help understanding the DNA interactions with carbon-based substances. For example, such interaction study could give indications on DNA-assisted sorting of carbon nanotube[105], DNA sequencing with graphene-based nanochannel device[106], and DNA sensing with carbon nanotubes[107], graphene[108] and graphene derivatives[109]. In addition, G-wires are found to preferentially orient along the armchair direction. Such alignment property of G-wires atop graphene could be used to identify the crystal orientation and hence the edge chirality of graphene. Graphene exhibits a variety of novel physical phenomena such as superconductivity[125], ferromagnetism[126], quantum Hall effect[127] and so on. These exceptional phenomena might be related to the physics of graphene orientation[124]. The sensing ability of G-wires provide an facile method for probing the graphene orientation, which would accelerate the pace of development of graphene-based devices, e.g. graphene nanoribbon semiconductor[130].

Bibliography

1. Lewin, B., *Genes VII*. Oxford University Press: 2000.

2. Neidle, S., *Principles of nucleic acid structure*. Academic Press: 2008.

3. Watson, J. D.; Crick, F. H. C., MOLECULAR STRUCTURE OF NUCLEIC ACIDS - A STRUCTURE FOR DEOXYRIBOSE NUCLEIC ACID. *Nature* **1953,** *171* (4356), 737-738.

4. (a) Wells, R. D., Non-B DNA conformations, mutagenesis and disease. *Trends Biochem. Sci.* **2007,** *32* (6), 271-278; (b) Choi, J.; Majima, T., Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011,** *40* (12).

5. (a) Vasquez, K. M.; Glazer, P. M., Triplex-forming oligonucleotides: principles and applications. *Q. Rev. Biophys.* **2002,** *35* (1), 89-107; (b) Duca, M.; Vekhoff, P.; Oussedik, K.; Halby, L.; Arimondo, P. B., The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* **2008,** *36* (16), 5123-5138.

6. (a) Simonsson, T., G-quadruplex DNA structures - Variations on a theme. *Biol. Chem.* **2001,** *382* (4), 621-628; (b) Davis, J. T., G-Quartets 40 Years Later: From 5'-GMP to Molecular Biology and Supramolecular Chemistry. *Angew. Chem., Int. Ed.* **2004,** *43* (6), 668-698; (c) Phan, A. T.; Kuryavyi, V.; Patel, D. J., DNA architecture: from G to Z. *Curr. Opin. struct. Biol.* **2006,** *16* (3), 288-298; (d) Patel, D. J.; Phan, A. T.; Kuryavyi, V., Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **2007,** *35* (22), 7429-7455.

7. Leontis, N. B.; Westhof, E., Geometric nomenclature and classification of RNA base pairs. *RNA* **2001,** *7*, 499–512.

8. (a) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006,** *34* (19), 5402-5415; (b) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J., In *Quadruplex Nucleic Acids*, Neidle, S.; Balasubramanian, S., Eds. Royal Society of Chemistry: Cambridge: UK, 2006; pp 81-99; (c) Gilbert, D. E.; Feigon, J., Multistranded DNA structures. *Curr. Opin. Struct. Biol.* **1999,** *9* (3), 305-314; (d) Williamson, J. R., G-Quartet Structures in Telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **1994,** *23*, 703-730

9. Gellert, M.; Lipsett, M. N.; Davies, D. R., Helix fomation by guanylic acid. *Proc. Natl Acad. Sci. USA* **1962,** *48* (12), 2013-2018.

10. Stephen Neidle (Editor), S. B. E., *Quadruplex Nucleic Acids*. The Royal Society of Chemistry: 2006.

11. (a) Huppert, J. L.; Balasubramanian, S., Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33* (9), 2908-2916; (b) Todd, A. K.; Johnston, M.; Neidle, S., Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005,** *33* (9), 2901-2907.

12. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J., Structure of the human telomere in K(+) solution: An intramolecular (3+1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **2006,** *128* (30), 9963-9970.

13. Huppert, J. L.; Balasubramanian, S., G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007,** *35* (2), 406-413.

14. Larson, E. D.; Duquette, M. L.; Cummings, W. J.; Streiff, R. J.; Maizels, N., MutSα Binds to and Promotes Synapsis of Transcriptionally Activated Immunoglobulin Switch Regions. *Curr. Biol.* **2005,** *15* (5), 470-474.

15. Mani, P.; Yadav, V. K.; Das, S. K.; Chowdhury, S., Genome-Wide Analyses of Recombination Prone Regions Predict Role of DNA Structural Motif in Recombination. *Plos One* **2009,** *4* (2).

16. Han, H. Y.; Hurley, L. H., G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol. Sci.* **2000,** *21* (4), 136-142.

17. (a) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H., Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA* **2002,** *99* (18), 11593-11598; (b) Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H., The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* **2004,** *126* (28), 8702-8709; (c) Phan, A. T.; Modi, Y. S.; Patel, D. J., Propeller-type parallel-stranded g-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **2004,** *126* (28), 8710-8716; (d) Brooks, T. A.; Hurley, L. H., The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer* **2009,** *9* (12), 849-861.

18. (a) Sun, D. Y.; Guo, K. X.; Rusche, J. J.; Hurley, L. H., Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive

agents. *Nucleic Acids Res.* **2005,** *33* (18), 6070-6080; (b) Guo, K.; Gokhale, V.; Hurley, L. H.; Sun, D., Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res.* **2008,** *36* (14), 4598-4608.

19. (a) Dexheimer, T. S.; Sun, D.; Hurley, L. H., Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* **2006,** *128* (16), 5404-5415; (b) Dai, J. X.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. Z., NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **2006,** *34* (18), 5133-5144; (c) Dai, J. X.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. Z., An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* **2006,** *128* (4), 1096-1098.

20. Cogoi, S.; Xodo, L. E., G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **2006,** *34* (9), 2536-2549.

21. Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H., Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.* **2007,** *35* (22), 7698-7713.

22. Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H., Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the hTERT Core Promoter with

Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands. *J. Am. Chem. Soc.* **2009,** *131* (31), 10878-10891.

23. Xu, Y.; Sugiyama, H., Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* **2006,** *34* (3), 949-954.

24. Hanahan, D.; Weinberg, R. A., The hallmarks of cancer. *Cell* **2000,** *100* (1), 57-70.

25. (a) Zakian, V. A., TELOMERES - BEGINNING TO UNDERSTAND THE END. *Science* **1995,** *270* (5242), 1601-1607; (b) de Lange, T., Protection of mammalian telomeres. *Oncogene* **2002,** *21* (4), 532-540.

26. Harley, C. B.; Futcher, A. B.; Greider, C. W., TELOMERES SHORTEN DURING AGING OF HUMAN FIBROBLASTS. *Nature* **1990,** *345* (6274), 458-460.

27. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W., Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science* **1994,** *266* (5193), 2011-2015.

28. Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M., Inhibition of telomerase by G-quartet DNA structures. *Nature* **1991,** *350* (6320), 718-720.

29. Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R., A HIGHLY CONSERVED REPETITIVE DNA-SEQUENCE, (TTAGGG)N, PRESENT AT THE TELOMERES OF HUMAN-CHROMOSOMES. *Proc. Natl Acad. Sci. USA* **1988,** *85* (18), 6622-6626.

30. Meyne, J.; Ratliff, R. L.; Moyzis, R. K., CONSERVATION OF THE HUMAN

TELOMERE SEQUENCE (TTAGGG)N AMONG VERTEBRATES. *Proc. Natl Acad. Sci. USA* **1989,** *86* (18), 7049-7053.

31. Phan, A. T., Human telomeric G-quadruplex: structures of DNA and RNA sequences. *Febs Journal* **2010,** *277* (5), 1107-1117.

32. Wang, Y.; Patel, D. J., Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* **1992,** *31* (35), 8112-8119.

33. Parkinson, G. N.; Lee, M. P. H.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002,** *417* (6891), 876-880.

34. Phan, A. T.; Patel, D. J., Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.* **2003,** *125* (49), 15021-15027.

35. Zhang, N.; Phan, A. T.; Patel, D. J., (3+1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Am. Chem. Soc.* **2005,** *127* (49), 17277-17285.

36. Wang, Y.; Patel, D. J., Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1993,** *1* (4), 263-282.

37. Balasubramanian, S.; Hurley, L. H.; Neidle, S., Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **2011,** *10* (4), 261-275.

38. Condon, A., Designed DNA molecules: principles and applications of molecular nanotechnology. *Nat. Rev. Genet.* **2006,** *7* (7), 565-575.

39. (a) Seeman, N. C., Nucleic acid junctions and lattices. *J. Theor. Biol.* **1982,** *99* (2), 237-247; (b) Kallenbach, N. R.; Ma, R. I.; Seeman, N. C., An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **1983,** *305* (5937), 829-831.

40. (a) Winfree, E.; Liu, F. R.; Wenzler, L. A.; Seeman, N. C., Design and self-assembly of two-dimensional DNA crystals. *Nature* **1998,** *394* (6693), 539-544; (b) Seeman, N. C., DNA in a material world. *Nature* **2003,** *421* (6921), 427-431; (c) Feldkamp, U.; Niemeyer, C. M., Rational design of DNA nanoarchitectures. *Angew. Chem. Int. Ed.* **2006,** *45* (12), 1856-1876.

41. (a) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L., A DNA-fuelled molecular machine made of DNA. *Nature* **2000,** *406* (6796), 605-608; (b) Bath, J.; Turberfield, A. J., DNA nanomachines. *Nat. Nanotechnol.* **2007,** *2* (5), 275-284; (c) Seeman, N. C., Nanomaterials Based on DNA. In *Annu. Rev. Biochem.*, Kornberg, R. D.; Raetz, C. R. H.; Rothman, J. E.; Thorner, J. W., Eds. 2010; Vol. 79, pp 65-87.

42. Keren, K.; Krueger, M.; Gilad, R.; Ben-Yoseph, G.; Sivan, U.; Braun, E., Sequence-specific molecular lithography on single DNA molecules. *Science* **2002,** *297* (5578), 72-75.

43. (a) Tyagi, S.; Kramer, F. R., Molecular beacons: Probes that fluoresce upon hybridization. *Nature Biotechnol.* **1996,** *14* (3), 303-308; (b) Weizmann, Y.; Beissenhirtz, M. K.; Cheglakov, Z.; Nowarski, R.; Kotler, M.; Willner, I., A virus spotlighted by an autonomous DNA machine. *Angew. Chem. Int. Ed.* **2006,** *45* (44), 7384-7388.

44. (a) Adleman, L. M., Molecular Computation of Solutions to Combinatorial Problems. *Science* **1994,** *266* (5187), 1021-1024; (b) Mao, C. D.; LaBean, T. H.; Reif, J. H.; Seeman, N. C., Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature* **2000,** *407* (6803), 493-496.

45. (a) Alivisatos, A. P.; Johnsson, K. P.; Peng, X. G.; Wilson, T. E.; Loweth, C. J.; Bruchez, M. P.; Schultz, P. G., Organization of 'nanocrystal molecules' using DNA. *Nature* **1996,** *382* (6592), 609-611; (b) Li, X. Y.; Liu, D. R., DNA-Templated organic synthesis: Nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew. Chem. Int. Ed.* **2004,** *43* (37), 4848-4870; (c) Mirkin, C. A.; Letsinger, R. L.; Mucic, R. C.; Storhoff, J. J., A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **1996,** *382* (6592), 607-609; (d) Braun, E.; Eichen, Y.; Sivan, U.; Ben-Yoseph, G., DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature* **1998,** *391* (6669), 775-778.

46. (a) Yan, H.; Park, S. H.; Finkelstein, G.; Reif, J. H.; LaBean, T. H., DNA-templated self-assembly of protein arrays and highly conductive nanowires. *Science* **2003,** *301* (5641), 1882-1884; (b) Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F., Assembling materials with DNA as the guide. *Science* **2008,** *321* (5897), 1795-1799; (c) Chhabra, R.; Sharma, J.; Ke, Y.; Liu, Y.; Rinker, S.; Lindsay, S.; Yan, H., Spatially Addressable Multiprotein Nanoarrays Templated by Aptamer-Tagged DNA Nanoarchitectures. *J. Am. Chem. Soc.* **2007,** *129* (34), 10304-10305.

47. Hays, F. A.; Watson, J.; Ho, P. S., Caution! DNA Crossing: Crystal Structures of

Holliday Junctions. *J. Biol. Chem.* **2003,** *278* (50), 49663-49666.

48. Fu, T. J.; Seeman, N. C., DNA Double-Crossover Molecules. *Biochemistry* **1993,** *32* (13), 3211-3220.

49. Shen, Z. Y.; Yan, H.; Wang, T.; Seeman, N. C., Paranemic crossover DNA: A generalized Holliday structure with applications in nanotechnology. *J. Am. Chem. Soc.* **2004,** *126* (6), 1666-1674.

50. Mathieu, F.; Liao, S. P.; Kopatscht, J.; Wang, T.; Mao, C. D.; Seeman, N. C., Six-helix bundles designed from DNA. *Nano Lett.* **2005,** *5* (4), 661-665.

51. Rothemund, P. W. K., Folding DNA to create nanoscale shapes and patterns. *Nature* **2006,** *440* (7082), 297-302.

52. Douglas, S. M.; Chou, J. J.; Shih, W. M., DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc. Natl Acad. Sci. USA* **2007,** *104* (16), 6644-6648.

53. Chen, J. H.; Seeman, N. C., Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **1991,** *350* (6319), 631-633.

54. Zhang, Y. W.; Seeman, N. C., Construction of a DNA-Truncated Octahedron. *J. Am. Chem. Soc.* **1994,** *116* (5), 1661-1669.

55. Goodman, R. P.; Schaap, I. A. T.; Tardin, C. F.; Erben, C. M.; Berry, R. M.; Schmidt, C. F.; Turberfield, A. J., Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **2005,** *310* (5754), 1661-1665.

56. Seeman, N. C.; Lukeman, P. S., Nucleic acid nanostructures: bottom-up control of geometry on the nanoscale. *Rep. Prog. Phys.* **2005,** *68* (1), 237-270.

57. (a) Phan, A. T.; Mergny, J. L., Human telomeric DNA: G-quadruplex, i-motif and watson-crick double helix. *Nucleic Acids Res.* **2002,** *30* (21), 4618-4625; (b) Kotlyar, A. B.; Borovok, N.; Molotsky, T.; Cohen, H.; Shapir, E.; Porath, D., Long, Monomolecular Guanine-Based Nanowires. *Adv. Mater.* **2005,** *17* (15), 1901-1905.

58. Krishnan, Y.; Simmel, F. C., Nucleic Acid Based Molecular Devices. *Angew. Chem. Int. Ed.* **2011,** *50* (14), 3124-3156.

59. Alberti, P.; Mergny, J. L., DNA duplex-quadruplex exchange as the basis for a nanomolecular machine. *Proc. Natl Acad. Sci. USA* **2003,** *100* (4), 1569-1573.

60. Li, J. J.; Tan, W., A Single DNA Molecule Nanomotor. *Nano Lett.* **2002,** *2* (4), 315-318.

61. Dittmer, W. U.; Reuter, A.; Simmel, F. C., A DNA-Based Machine That Can Cyclically Bind and Release Thrombin. *Angew. Chem. Int. Ed.* **2004,** *43* (27), 3550-3553.

62. (a) Fahlman, R. P.; Hsing, M.; Sporer-Tuhten, C. S.; Sen, D., Duplex Pinching: A Structural Switch Suitable for Contractile DNA Nanoconstructions. *Nano Lett.* **2003,** *3* (8), 1073-1078; (b) Choi, J.; Majima, T., Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011,** *40* (12), 5893-5909.

63. Huang, Y. C.; Sen, D., A Contractile Electronic Switch Made of DNA. *J. Am. Chem. Soc.* **2010,** *132* (8), 2663-2671.

64. Miyoshi, D.; Karimata, H.; Wang, Z.-M.; Koumoto, K.; Sugimoto, N., Artificial G-Wire Switch with 2,2‘-Bipyridine Units Responsive to Divalent Metal Ions. *J. Am. Chem. Soc.* **2007,** *129* (18), 5919-5925.

65. Marsh, T. C.; Henderson, E., G-Wires: Self-Assembly of a Telomeric Oligonucleotide, d(GGGGTTGGGG), into Large Superstructures. *Biochemistry* **1994,** *33* (35), 10718-10724.

66. (a) Phan, A. T.; Guéron, M.; Leroy, J. L., Investigation of unusual DNA motifs. In *Nuclear Magnetic Resonance of Biologica Macromolecules, Pt A*, 2001; Vol. 338, pp 341-371; (b) Rangan, A.; Fedoroff, O. Y.; Hurley, L. H., Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule. *J. Biol. Chem.* **2001,** *276* (7), 4640-4646.

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

68. Seo, Y., Atomic force microscopy and spectroscopy. *Rep. Prog. Phys.* **2008,** *71* (1), 016101.

69. Engel, A.; Muller, D. J., Observing single biomolecules at work with the atomic force microscope. *Nat. Struct. Biol.* **2000,** *7* (9), 715-718.

70. Gross, L.; Mohn, F.; Moll, N.; Liljeroth, P.; Meyer, G., The Chemical Structure of a Molecule Resolved by Atomic Force Microscopy. *Science* **2009,** *325* (5944), 1110-1114.

71. Albrecht, T. R.; Grutter, P.; Horne, D.; Rugar, D., Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity. *J. Appl. Phys.* **1991,** *69* (2), 668-673.

72. Mergny, J. L.; Phan, A. T.; Lacroix, L., Following G-quartet formation by UV-spectroscopy. *Febs Letters* **1998,** *435* (1), 74-78.

73. (a) Beychok, S., Circular dichroism of biological macromolecules. *Science* **1966,** *154* (3754), 1288-1299; (b) Gray, D. M.; Ratliff, R. L.; Vaughan, M. R.; David, M. J. L.; James, E. D., Circular dichroism spectroscopy of DNA. *Methods Enzymol.* **1992,** *211*, 389-406.

74. (a) Balagurumoorthy, P.; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V., Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* **1992,** *20* (15), 4061-4067; (b) Hu, L.; Lim, K. W.; Bouaziz, S.; Phan, A. T., Giardia Telomeric Sequence d(TAGGG)4 Forms Two Intramolecular G-Quadruplexes in K+ Solution: Effect of Loop Length and Sequence on the Folding Topology. *J. Am. chem. Soc.* **2009,** *131* (46), 16824-16831.

75. Gray, D., M. ; Jin-Der, W.; Carla, W. G.; Rudolf, R.; Charlotte, R.; Gerhard, R.; Jörg, F., Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* **2008,** *20* (3-4), 431-440.

76. (a) Rachwal, P. A.; Brown, T.; Fox, K. R., Effect of G-Tract Length on the Topology and Stability of Intramolecular DNA Quadruplexes. *Biochemistry* **2007,** *46* (11), 3036-3044; (b) Rachwal, P. A.; Findlow, I. S.; Werner, J. M.; Brown, T.; Fox, K. R., Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res.* **2007,** *35* (12), 4214-4222; (c) Bugaut, A.; Balasubramanian, S., A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes. *Biochemistry* **2008,** *47* (2), 689-697; (d) Guédin, A.; De Cian, A.; Gros, J.; Lacroix, L.; Mergny, J.-L., Sequence effects in single-base loops for quadruplexes. *Biochimie* **2008,** *90* (5),

686-696.

77. (a) Xu, Y.; Noguchi, Y.; Sugiyama, H., The new models of the human telomere d[AGGG(TTAGGG)3] in K+ solution. *Bioorg. Med. Chem.* **2006,** *14* (16), 5584-5591; (b) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D., Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **2006,** *34* (9), 2723-2735; (c) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J., Structure of the human telomere in K+ solution: An intramolecular (3+1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **2006,** *128* (30), 9963-9970; (d) Phan, A. T.; Luu, K. N.; Patel, D. J., Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. *Nucleic Acids Res.* **2006,** *34* (19), 5715-5719; (e) Dai, J. X.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. Z., Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res.* **2007,** *35* (7), 2440-2450; (f) Matsugami, A.; Xu, Y.; Noguchi, Y.; Sugiyama, H.; Katahira, M., Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K+ solution. *Febs Journal* **2007,** *274* (14), 3545-3556; (g) Dai, J. X.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Z., Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* **2007,** *35* (15), 4927-4940; (h) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J., Structure of two intramolecular G-quadruplexes formed by natural human telomere

sequences in K+ solution. *Nucleic Acids Res.* **2007,** *35* (19), 6517-6525; (i) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T., Structure of the Human Telomere in K+ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **2009,** *131* (12), 4301-4309.

78. Richards, E. J.; Ausubel, F. M., Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. *Cell* **1988,** *53* (1), 127-136.

79. Petracek, M. E.; Lefebvre, P. A.; Silflow, C. D.; Berman, J., Chlamydomonas Telomere Sequences are A+T-rich but ContainThree Consecutive G-C Base-pairs. *Proc. Natl. Acad. Sci. U. S. A.* **1990,** *87* (21), 8222-8226.

80. Amrane, S.; Ang, R. W. L.; Tan, Z. M.; Li, C.; Lim, J. K. C.; Lim, J. M. W.; Lim, K. W.; Phan, A. T., A novel chair-type G-quadruplex formed by a Bombyx mori telomeric sequence. *Nucleic Acids Res.* **2009,** *37* (3), 931-938.

81. Lim, K. W.; Alberti, P.; Guédin, A.; Lacroix, L.; Riou, J.-F.; Royle, N. J.; Mergny, J.-L.; Phan, A. T., Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G·C·G·C tetrad. *Nucleic Acids Res.* **2009,** *37* (18), 6239-6248.

82. Leblancq, S. M.; Kase, R. S.; Vanderploeg, L. H. T., Analysis of a Giardia Lamblia rRNA Encoding Telomere with TAGGG as the Telomere Repeat. *Nucleic Acids Research* **1991,** *19* (20), 5790-5790.

83. Allshire, R. C.; Dempster, M.; Hastie, N. D., Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.* **1989,** *17*

(12), 4611-4627.

84. Cantor, C. R.; Warshaw, M. M.; Shapiro, H., Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleolides. *Biopolymers* **1970,** *9* (9), 1059-1077.

85. Phan, A. T.; Patel, D. J., A site-specific low-enrichment N-15,C-13 isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.* **2002,** *124* (7), 1160-1161.

86. Huang, X.; Yu, P.; LeProust, E.; Gao, X., An efficient and economic site-specific deuteration strategy for NMR studies of homologous oligonucleotide repeat sequences. *Nucleic Acids Res.* **1997,** *25* (23), 4758-4763.

87. Phan, A. T., Long-range imino proton-C-13 J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR* **2000,** *16* (2), 175-178.

88. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M., The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2003,** *160* (1), 65-73.

89. DeLano, W. L. The PyMOL User's Manual.

90. Phan, A. T.; Kuryavyi, V.; Ma, J. B.; Faure, A.; Andréola, M. L.; Patel, D. J., An interlocked dimeric parallel-stranded DNA quadruplex: A potent inhibitor of HIV-1 integrase. *Proc. Natl Acad. Sci. USA* **2005,** *102* (3), 634-639.

91. Do, N. Q.; Lim, K. W.; Teo, M. H.; Heddi, B.; Phan, A. T., Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV

and anticancer activity. *Nucleic Acids Res.* **2011,** *39* (21), 9448-9457.

92. Kelly, J. A.; Feigon, J.; Yeates, T. O., Reconciliation of the X-ray and NMR Structures of the Thrombin-Binding Aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Biol.* **1996,** *256* (3), 417-422.

93. (a) Wang, Y.; Patel, D. J., Solution Structure of theOxytrichaTelomeric Repeat d[G4(T4G4)3] G-tetraplex. *J. Mol. Biol.* **1995,** *251* (1), 76-94; (b) Smith, F. W.; Schultze, P.; Feigon, J., Solution structures of unimolecular quadruplexes formed by oligonucleotides containing Oxytricha telomere repeats. **1995,** *3* (10), 997-1008.

94. Martadinata, H.; Phan, A. T., Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric RNA Sequences in K+ Solution. *J. Am. Chem. Soc.* **2009,** *131* (7), 2570-2578.

95. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S., Loop-Length-Dependent Folding of G-Quadruplexes. *J. Am. Chem. Soc.* **2004,** *126* (50), 16405-16415.

96. (a) Crnugelj, M.; Sket, P.; Plavec, J., Small Change in a G-Rich Sequence, a Dramatic Change in Topology: New Dimeric G-Quadruplex Folding Motif with Unique Loop Orientations. *J. Am. Chem. Soc.* **2003,** *125* (26), 7866-7871; (b) Črnugelj, M.; Hud, N. V.; Plavec, J., The Solution Structure of d(G4T4G3)2: a Bimolecular G-quadruplex with a Novel Fold. *J. Mol. Biol.* **2002,** *320* (5), 911-924.

97. (a) Phan, A. T.; Modi, Y. S.; Patel, D. J., Two-repeat Tetrahymena telomeric d(TGGGGTTGGGGT) sequence interconverts between asymmetric dimeric G-quadruplexes in solution. *J. Mol. Biol.* **2004,** *338* (1), 93-102; (b) Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S., Studies on the structure and dynamics

of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA* **2003,** *100* (25), 14629-14634; (c) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T., Extreme conformational diversity in human telomeric DNA. *Proc. Natl Acad. Sci. USA* **2005,** *102* (52), 18938-18943.

98. Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J., Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* **2005,** *1* (3), 167-173.

99. Katsnelson, M. I.; Novoselov, K. S.; Geim, A. K., Chiral tunnelling and the Klein paradox in graphene. *Nature Phys.* **2006,** *2* (9), 620-625.

100. Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Zhang, Y.; Dubonos, S. V.; Grigorieva, I. V.; Firsov, A. A., Electric field effect in atomically thin carbon films. *Science* **2004,** *306* (5696), 666-669.

101. Rao, C. N. R.; Sood, A. K.; Subrahmanyam, K. S.; Govindaraj, A., Graphene: The New Two-Dimensional Nanomaterial. *Angew. Chem. Int. Ed.* **2009,** *48* (42), 7752-7777.

102. (a) Mohanty, N.; Berry, V., Graphene-Based Single-Bacterium Resolution Biodevice and DNA Transistor: Interfacing Graphene Derivatives with Nanoscale and Microscale Biocomponents. *Nano Lett.* **2008,** *8* (12), 4469-4476; (b) Patil, A. J.; Vickery, J. L.; Scott, T. B.; Mann, S., Aqueous Stabilization and Self-Assembly of Graphene Sheets into Layered Bio-Nanocomposites using DNA. *Adv. Mater.* **2009,** *21* (31), 3159-3164.

103. (a) Ohno, Y.; Maehashi, K.; Yamashiro, Y.; Matsumoto, K., Electrolyte-Gated

Graphene Field-Effect Transistors for Detecting pH and Protein Adsorption. *Nano Lett.* **2009,** *9* (9), 3318-3322; (b) Liu, J.; Fu, S.; Yuan, B.; Li, Y.; Deng, Z., Toward a Universal "Adhesive Nanosheet" for the Assembly of Multiple Nanoparticles Based on a Protein-Induced Reduction/Decoration of Graphene Oxide. *J. Am. chem. Soc.* **2010,** *132* (21), 7279-7281.

104. Wang, Y.; Li, Z. H.; Hu, D. H.; Lin, C. T.; Li, J. H.; Lin, Y. H., Aptamer/Graphene Oxide Nanocomplex for in Situ Molecular Probing in Living Cells. *J. Am. Chem. Soc.* **2010,** *132* (27), 9274-9276.

105. (a) Zheng, M.; Jagota, A.; Semke, E. D.; Diner, B. A.; McLean, R. S.; Lustig, S. R.; Richardson, R. E.; Tassi, N. G., DNA-assisted dispersion and separation of carbon nanotubes. *Nature Mater.* **2003,** *2* (5), 338-342; (b) Zheng, M.; Jagota, A.; Strano, M. S.; Santos, A. P.; Barone, P.; Chou, S. G.; Diner, B. A.; Dresselhaus, M. S.; McLean, R. S.; Onoa, G. B.; Samsonidze, G. G.; Semke, E. D.; Usrey, M.; Walls, D. J., Structure-Based Carbon Nanotube Sorting by Sequence-Dependent DNA Assembly. *Science* **2003,** *302* (5650), 1545-1548; (c) Tu, X.; Manohar, S.; Jagota, A.; Zheng, M., DNA sequence motifs for structure-specific recognition and separation of carbon nanotubes. *Nature* **2009,** *460* (7252), 250-253.

106. (a) Lagerqvist, J.; Zwolak, M.; Di Ventra, M., Fast DNA Sequencing via Transverse Electronic Transport. *Nano Lett.* **2006,** *6* (4), 779-782; (b) Postma, H. W. C., Rapid Sequencing of Individual DNA Molecules in Graphene Nanogaps. *Nano Lett.* **2010,** *10* (2), 420-425.

107. (a) Star, A.; Tu, E.; Niemann, J.; Gabriel, J. C. P.; Joiner, C. S.; Valcke, C.,

Label-free detection of DNA hybridization using carbon nanotube network field-effect transistors. *Proc. Natl Acad. Sci. USA* **2006,** *103* (4), 921-926; (b) Jeng, E. S.; Moll, A. E.; Roy, A. C.; Gastala, J. B.; Strano, M. S., Detection of DNA Hybridization Using the Near-Infrared Band-Gap Fluorescence of Single-Walled Carbon Nanotubes. *Nano Lett.* **2006,** *6* (3), 371-375; (c) Yang, R.; Jin, J.; Chen, Y.; Shao, N.; Kang, H.; Xiao, Z.; Tang, Z.; Wu, Y.; Zhu, Z.; Tan, W., Carbon Nanotube-Quenched Fluorescent Oligonucleotides: Probes that Fluoresce upon Hybridization. *J. Am. chem. Soc.* **2008,** *130* (26), 8351-8358; (d) Diao, P.; Liu, Z., Vertically Aligned Single-Walled Carbon Nanotubes by Chemical Assembly – Methodology, Properties, and Applications. *Adv. Mater.* **2010,** *22* (13), 1430-1449.

108. (a) Dong, X.; Shi, Y.; Huang, W.; Chen, P.; Li, L.-J., Electrical Detection of DNA Hybridization with Single-Base Specificity Using Transistors Based on CVD-Grown Graphene Sheets. *Adv. Mater.* **2010,** *22* (14), 1649-1653; (b) Tang, L. A. L.; Wang, J.; Loh, K. P., Graphene-Based SELDI Probe with Ultrahigh Extraction and Sensitivity for DNA Oligomer. *J. Am. Chem. Soc.* **2010,** *132* (32), 10976-10977.

109. Dreyer, D. R.; Park, S.; Bielawski, C. W.; Ruoff, R. S., The chemistry of graphene oxide. *Chem. Soc. Rev.* **2010,** *39* (1), 228-240.

110. Lu, C.-H.; Yang, H.-H.; Zhu, C.-L.; Chen, X.; Chen, G.-N., A Graphene Platform for Sensing Biomolecules. *Angew. Chem. Int. Ed.* **2009,** *48* (26), 4785-4787.

111. Guo, A.-M.; Xiong, S.-J., Effects of contact and efficient charge transport in G4-DNA molecules. *Phys. Rev. B* **2009,** *80* (3), 035115.

112. Marsh, T. C.; Vesenka, J.; Henderson, E., A new DNA nanostructure, the G-wire,

imaged by scanning probe microscopy. *Nucleic Acids Res.* **1995,** *23* (4), 696-700.

113.Takeshita, M.; Chang, C. N.; Johnson, F.; Will, S.; Grollman, A. P., Oligodeoxynucleotides containing synthetic abasic sites. Model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J. Biol. Chem.* **1987,** *262* (21), 10171-10179.

114.Pastré, D.; Piétrement, O.; Fusil, S.; Landousy, F.; Jeusset, J.; David, M.-O.; Hamon, L.; Le Cam, E.; Zozime, A., Adsorption of DNA to Mica Mediated by Divalent Counterions: A Theoretical and Experimental Study. *Biophys. J.* **2003,** *85* (4), 2507-2518.

115.Hansma, H. G.; Revenko, I.; Kim, K.; Laney, D. E., Atomic force microscopy of long and short double-stranded, single- stranded and triple-stranded nucleic acids. *Nucleic Acids Res.* **1996,** *24* (4), 713-720.

116.Antony, J.; Grimme, S., Structures and interaction energies of stacked graphene-nucleobase complexes. *Phys. Chem. Chem. Phys.* **2008,** *10* (19), 2722-2729.

117.Gowtham, S.; Scheicher, R. H.; Ahuja, R.; Pandey, R.; Karna, S. P., Physisorption of nucleobases on graphene: Density-functional calculations. *Phys. Rev. B* **2007,** *76* (3), 033401.

118.He, S.; Song, B.; Li, D.; Zhu, C.; Qi, W.; Wen, Y.; Wang, L.; Song, S.; Fang, H.; Fan, C., A Graphene Nanoprobe for Rapid, Sensitive, and Multicolor Fluorescent DNA Analysis. *Adv. Fun. Mat.* **2010,** *20* (3), 453-459.

119.(a) Das, A.; Pisana, S.; Chakraborty, B.; Piscanec, S.; Saha, S. K.; Waghmare, U. V.; Novoselov, K. S.; Krishnamurthy, H. R.; Geim, A. K.; Ferrari, A. C.; Sood, A. K.,

Monitoring dopants by Raman scattering in an electrochemically top-gated graphene transistor. *Nature Nanotechnol.* **2008,** *3* (4), 210-215; (b) Ni, Z. H.; Wang, Y. Y.; Yu, T.; Shen, Z. X., Raman Spectroscopy and Imaging of Graphene. *Nano Res.* **2008,** *1* (4), 273-291.

120. Lyonnais, S.; Gorelick, R. J.; Mergny, J. L.; Le Cam, E.; Mirambeau, G., G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA. *Nucleic Acids Res.* **2003,** *31* (19), 5754-5763.

121. Dresselhaus, M. S.; Jorio, A.; Saito, R., Characterizing Graphene, Graphite, and Carbon Nanotubes by Raman Spectroscopy. *Annu. Rev. Condens. Matter. Phys.* **2010,** *1* (1), 89-108.

122. You, Y. M.; Ni, Z. H.; Yu, T.; Shen, Z. X., Edge chirality determination of graphene by Raman spectroscopy. *Appl. Phys. Lett.* **2008,** *93* (16), 163112.

123. (a) Cancado, L. G.; Pimenta, M. A.; Neves, B. R. A.; Dantas, M. S. S.; Jorio, A., Influence of the atomic structure on the Raman spectra of graphite edges. *Phys. Rev. Lett.* **2004,** *93* (24), 247401; (b) Casiraghi, C.; Hartschuh, A.; Qian, H.; Piscanec, S.; Georgi, C.; Fasoli, A.; Novoselov, K. S.; Basko, D. M.; Ferrari, A. C., Raman Spectroscopy of Graphene Edges. *Nano Lett.* **2009,** *9* (4), 1433-1441; (c) Gupta, A. K.; Russin, T. J.; Gutiérrez, H. R.; Eklund, P. C., Probing Graphene Edges via Raman Scattering. *ACS Nano* **2008,** *3* (1), 45-52; (d) Cong, C. X.; Yu, T.; Wang, H. M., Raman Study on the G Mode of Graphene for Determination of Edge Orientation. *ACS Nano* **2010,** *4* (6), 3175-3180.

124. (a) Castro Neto, A. H.; Guinea, F.; Peres, N. M. R.; Novoselov, K. S.; Geim,

A. K., The electronic properties of graphene. *Rev. Mod. Phys.* **2009,** *81* (1), 109-162; (b) Girit, Ç. Ö.; Meyer, J. C.; Erni, R.; Rossell, M. D.; Kisielowski, C.; Yang, L.; Park, C.-H.; Crommie, M. F.; Cohen, M. L.; Louie, S. G.; Zettl, A., Graphene at the Edge: Stability and Dynamics. *Science* **2009,** *323* (5922), 1705-1708; (c) Yazyev, O. V.; Katsnelson, M. I., Magnetic Correlations at Graphene Edges: Basis for Novel Spintronics Devices. *Phys. Rev. Lett.* **2008,** *100* (4), 047209.

125. Kopnin, N. B.; Sonin, E. B., BCS Superconductivity of Dirac Electrons in Graphene Layers. *Phys. Rev. Lett.* **2008,** *100* (24), 246808.

126. Feldner, H.; eacute; egrave; ne; Meng, Z. Y.; Lang, T. C.; Assaad, F. F.; Wessel, S.; Honecker, A., Dynamical Signatures of Edge-State Magnetism on Graphene Nanoribbons. *Phys. Rev. Lett.* **2011,** *106* (22), 226401.

127. Ghahari, F.; Zhao, Y.; Cadden-Zimansky, P.; Bolotin, K.; Kim, P., Measurement of the nu =1/3 Fractional Quantum Hall Energy Gap in Suspended Graphene. *Phys. Rev. Lett.* **2011,** *106* (4), 046801.

128. Wu, Y. H.; Yu, T.; Shen, Z. X., Two-dimensional carbon nanostructures: Fundamental properties, synthesis, characterization, and potential applications. *J. Appl. Phys.* **2010,** *108* (7), 071301-38.

129. (a) Ritter, K. A.; Lyding, J. W., The influence of edge structure on the electronic properties of graphene quantum dots and nanoribbons. *Nature Mater.* **2009,** *8* (3), 235-242; (b) Tao, C.; Jiao, L.; Yazyev, O. V.; Chen, Y.-C.; Feng, J.; Zhang, X.; Capaz, R. B.; Tour, J. M.; Zettl, A.; Louie, S. G.; Dai, H.; Crommie, M. F., Spatially resolving edge states of chiral graphene nanoribbons. *Nature Phys.* **2011,** *7* (8),

616-620.

130. Li, X.; Wang, X.; Zhang, L.; Lee, S.; Dai, H., Chemically Derived, Ultrasmooth Graphene Nanoribbon Semiconductors. *Science* **2008,** *319* (5867), 1229-1232.

131. (a) Yu, Z. B.; Schonhoft, J. D.; Dhakal, S.; Bajracharya, R.; Hegde, R.; Basu, S.; Mao, H. B., ILPR G-Quadruplexes Formed in Seconds Demonstrate High Mechanical Stabilities. *J. Am. Chem. Soc.* **2009,** *131* (5), 1876-1882; (b) Susanna, L.; Heather, B.; Byker, S. G.; Dejian, Z.; Kumar, S., Single Molecule Force Spectroscopy on G-Quadruplex DNA. *Chem.—Eur. J.* **2009,** *15* (33), 8113-8116.

List of Publications

1. **Hu, L.**; Lim, K. W.; Bouaziz, S.; Phan, A. T., *Giardia* telomeric sequence d(TAGGG)$_4$ forms two intramolecular G-quadruplexes in K^+ solution: effect of loop length and sequence on the folding topology. *J. Am. Chem. Soc.* **2009,** *131* (46), 16824-16831.

2. **Hu, L.** Y.; Yan, B.; Yu, T.; Phan, A. T., Self-assembly of DNA supramolecular G-wires on graphene. (*2012, submitted*)

3. Swaminathan, V.; Liew, H. F.; Lew, W. S.; **Hu, L**.; Phan, A. T., Photoelectrochemical studies of DNA-tagged biomolecules on Au and Au/Ni/Au multilayer nanowires. *Nanoscale Res. Lett.* **2011,** *6* (1), 535.

4. Miao, Y.; Liu, Y.; **Hu, L.**; Helseth, L. E., Colloidal clustering of protein-coated microspheres in evaporating droplets. *Soft Matter* **2012**, 8 (7), 2267-2273.