A computational study on the nature of DNA G-quadruplex structure

Gholamjani Moghaddam, Kiana

DOI:
10.33612/diss.159767021

1

DNA G-quadruplexes in Human Genome and Nanotechnology

DNA G-quadruplexes are higher-order structures self-assembled from guanine-rich oligonucleotides. These unique structures can be utilized as interesting building blocks in a wide range of applications from cancer therapeutics to nanodevices. This chapter provides a brief introduction to the G-quadruplex structure, focusing on the potential involvement of these structures in a range of different applications.

1.1. G-quadruplex structure

DNA stores and transfers hereditary information codes which play a critical role in various biological processes[1,2]. Generally, the most prevalent image of DNA is the double-helical structure discovered by Watson and Crick[3]. Apart from this structure, also called B-DNA, single-stranded guanine-rich DNA sequences can self-associate into non-canonical structures termed G-quadruplexes[4,5]. These DNA structures are composed of stacked planar G-quartets, each of which is held together by a network of Hoogsteen hydrogen bonds between four guanine bases[6,7] as shown in Figure 1.1. In fact, N1-H and N2-H of one guanine are hydrogen-bonded with O6 and N7 atoms of the neighboring guanine, resulting in eight hydrogen bonds per G-quartet. The G-quartets are stabilized by monovalent cations (K^+ or Na^+) coordinated within the plane of G-quartets, neutralizing the electrostatic repulsion of the four oxygen atoms of the guanines. Despite the simplicity of G-quartets,

guanine-rich DNA sequences can form different G-quadruplex topologies depending on the number of DNA strands, DNA strand orientation (antiparallel, parallel, or hybrid), size and type of loop structures (diagonal loops, lateral loops and double chain reversal loops) and glycosidic bond orientation (syn and anti conformations) [4,8] (Figure 1.2). Furthermore, they can be formed by the association of one strand (intramolecular) or multiple strands (intermolecular) [9]. In a parallel G-quadruplex all four strands are oriented in the same direction adopting the same glycosidic conformation (anti-anti-anti-anti or syn-syn-syn-syn). In the first type of antiparallel G-quadruplex, two strands have the same orientation and the other two are oriented in the opposite direction adopting syn-syn-anti-anti and syn-anti-syn-anti glycosidic conformations. The second type of antiparallel G-quadruplex, also called a hybrid-type (3+1), three strands have the same strand orientation while the fourth strand is oriented in the opposite direction. In this case, the G-quartets comprise three guanosines in the syn or anti conformation and the fourth one in an opposite conformation (syn-anti-anti-anti and anti-syn-syn-syn). G-quadruplex formation has been observed in telomeres at the end of eukaryotic chromosomes [10] and is particularly widespread in oncogene promoters such as k-RAS [11], BCL-2 [12], c-MYC [13], c-KIT [14,15], etc. These findings provide particular interest in the G-quadruplex structures as potential targets in cellular processes and cancer therapeutics [16,17]. Beyond the biological application, G-quadruplexes can be utilized as interesting building blocks in nanodevices [18].

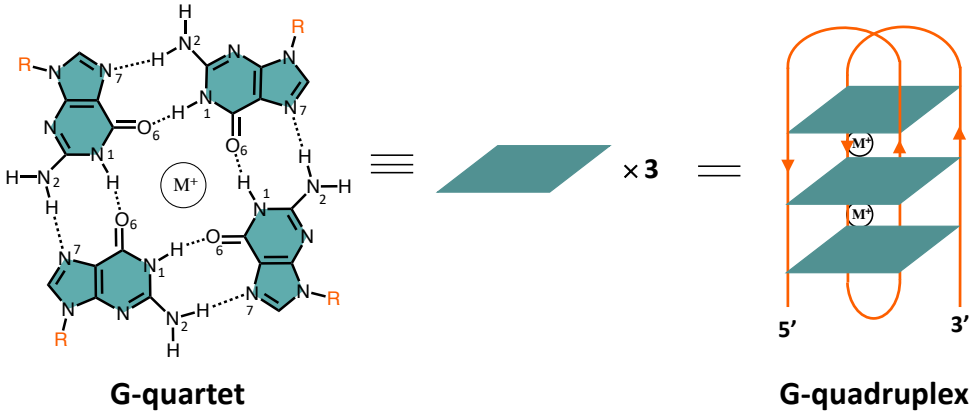

Figure 1.1 | Schematic representation of a G-quartet and G-quadruplex structure adopted from Ref. [19]. Four Guanine residues form a planar structure through Hoogsteen hydrogen bonding called G-quartet and then a G-quadruplex structure can be folded from G-quartets. M^+ is monovalent cations.

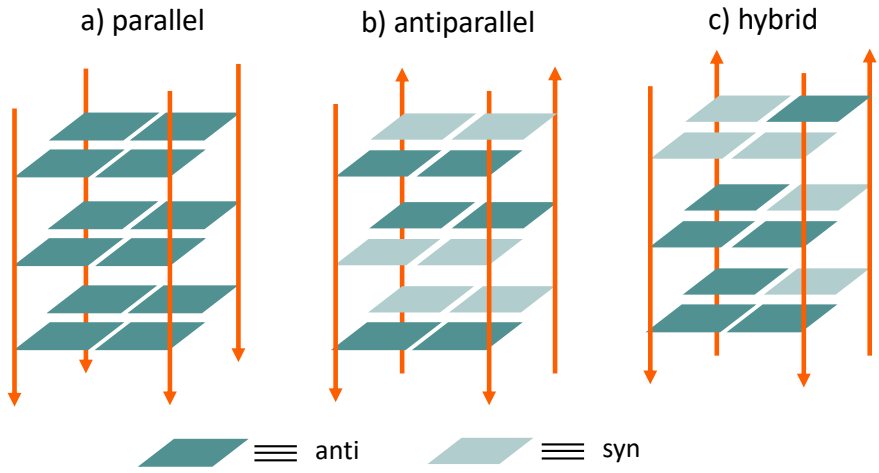

Figure 1.2 | Schematic representation of a) parallel, b) antiparallel and c) hybrid G-quadruplex structures, adopted from Ref. [20]

1.2. Application Domain

Cancer Therapy The concept of targeting G-quadruplex structures as an anticancer therapeutic strategy was first introduced for telomerase enzyme inhibition [21]. Human telomeres, which are located in the terminal part of chromosomes, are shortened during cell division through induction of apoptosis and senescence. However, the telomere length of >85% of cancer cells is maintained by telomerase activity that leads to telomere stabilization, cellular immortality and tumour progression. In most eukaryotes, telomeric DNA comprises repeated guanine-rich single-stranded sequences such as $(TTAGGG)_n$ which is predisposed to self-assemble into G-quadruplex structures with a variety of topologies [22,23]. In addition, G-quadruplex structures have been found in thousands of gene promoters. The expression of proto-oncogenes is necessary to control different normal cells growth, whereas overexpression or mutation of the proto-oncogenes has been noticed in various cancers including gastrointestinal stromal tumors (GIST), leukemias, melanoma, pancreatic cancers, etc [17,24]. After a decade of research about G-quadruplex formation *in vitro* and *in vivo*, numerous drug-like small molecules have been developed that target G-quadruplex structures in these two regions of the human genome inducing apoptosis and senescence in cancer cells or inhibition of oncogene expression [25]. Such small molecules provide a new direction for developing novel anticancer drugs. To date, a wide range of organic ligands such as anthraquinones, acridines, quinazolone, quindolines, quinacridines, porphyrins,

porphyrazines, berberines, metal complexes, etc have been synthesized and employed to stabilize G-quadruplex structures in the human genome[25–27]. In most cases, these ligands follow some principles: 1) planar aromatic surfaces interact with G-quartets through π-π stacking interactions, 2) cationic substituents interact with the backbone phosphates, grooves and loops, 3) the cationic core of ligands can bind to the negatively charged center of the G-quadruplex[28]. However, the design and development of G-quadruplex specific ligands is a challenging task. To discover G-quadruplex ligands, understanding of the nature of interactions and binding mechanism between the ligand and G-quadruplex is of paramount importance. These findings can pave the way to design ligands with high selectivity and binding affinity.

Cellular Processes G-quadruplexes are known to have biological roles in cellular processes including DNA replication, transcription, and translation[24,29]. However, these structures can be detrimental for replication and gene expression. G-quadruplex structures must be unfolded for completion of replication and transcription of the DNA employed by helicase enzymes, any unfolded G-quadruplex blocks transcription and/or replication and down-regulates gene expression leading to DNA damage[30]. Such G-quadruplex barrier can be counteracted by a set of DNA helicases, such as BLM[31], FANCJ[32], PIF1[33], WRN[34] and REV1. Furthermore, RNA helicase associated with AU-rich element (RHAU), a member of the ATP-dependent RNA helicases, was identified to bind and unwind the G-quadruplex structures[35,36]. To date, various specific functional roles have been assigned to RHAU. Different studies have shown the role of RHAU in transcriptional regulation, mRNA stability and controlling gene expression[37]. Furthermore, the role of RHAU in the recognition and remodeling of G-quadruplex structures is critical in a number of key cellular regulatory processes. RHAU includes a core DEAH-box helicase domain which is flanked by N-terminal and C-terminal extensions[38]. The conserved N-terminal domain known as the RHAU-specific motif (RSM) is required for interaction with G-quadruplexes, but it is insufficient for G-quadruplex unfolding. The full-length protein is necessary for the G-quadruplex unwinding process. In fact, understanding the detailed mechanism concerning how this protein can recognize G-quadruplex structures can help us to suggest a strategy for the recognition of G-quadruplex structures by proteins.

Nanotechnology The G-quadruplex structures not only are the main regulatory elements in the human genome, but also can play key roles in nanodevices and optomechanical molecular motor research[39–41]. The unique structure of the G-quadruplex makes it an interesting building block for the development of nanodevices. Indeed, G-quadruplex structures can undergo reversible conformational changes controlled by external triggers such

as light[42], pH[43], metal cations[44,45] and small molecules[41,46,47]. Among external stimuli, light is a promising external trigger which offers great advantages for controlling movement and conformation of the systems. For example, irradiation is a precise method with high selectivity and non-invasiveness features. Moreover, the timing, dosage and location of light can be easily regulated[48]. This process can be introduced by using photolabile groups or photoswitchable molecules[49,50]. Azobenzene derivatives are widely used as excellent photoswitchable molecules[51] because they possess intriguing photo-chemical characteristics. Introduction of azobenzenes into G-quadruplex structures is one of the most widely used methods to regulate G-quadruplex formation[52,53]. Indeed, the azobenzenes can be switched between trans and cis configurations under visible and UV light, respectively that plays a key role in the regulation of G-quadruplex formation. Therefore, understanding the photoisomerization reaction of azobenzene derivatives within G-quadruplex can assist the development of nanodevices with high efficiency.

1.3. Thesis Outline

The aim of this thesis is to investigate different aspects of G-quadruplex based systems for biological and nanotechnological applications such as cancer therapy, nanodevices and cellular processes with the aid of computational techniques.

Chapter 2 introduces briefly the electronic structure methods, atomistic molecular dynamics (MD) simulations, quantum mechanics/molecular mechanics (QM/MM) simulations and coarse-grain simulations which are the main computational methods used in this thesis.

Chapter 3 presents the G-quadruplex application in cancer therapy. Stabilization of G-quadruplex structures in the oncogenic promoter regions such as c-KIT with small molecules has attracted considerable attention as a promising target for cancer therapeutics. We investigate the binding interactions of some quinazolone derivatives with c-KIT G-quadruplex by MD simulations which can pave the way to rational ligand design.

Following the results obtained from the MD simulations, we provide detailed insight into the nature of interactions between other quinazolone derivatives and c-KIT G-quadruplex which is presented in **Chapter 4**.

Chapter 5 presents the G-quadruplex application in nanodevices. Introducing photoswitches into DNA G-quadruplex provides excellent opportunities to control folding and unfolding of these assemblies, demonstrating their potential in the development of novel nanodevices. We applied QM/MM simulations to identify the effect of the size and

substitution patterns of three azobenzene derivatives on the smallest photoswitchable G-quadruplex reported to date.

Chapter 6 describes the G-quadruplex application in the cellular process. The recognition process of G-quadruplex structures within cells by proteins is considered important for replication and gene expression. We investigate the binding mechanism between RNA helicase associated with AU-rich element (RHAU) and different G-quadruplex structures. These findings can help us to suggest a strategy for the recognition of G-quadruplex structures by proteins.