BBA - Proteins and Proteomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

G-quadruplex dynamics☆

Robert W. Harkness V, Anthony K. Mittermaier*

McGill University Department of Chemistry, 801 Sherbrooke St. W., Montreal, QC H3A 0B8, Canada

ARTICLE INFO

Keywords:
G-quadruplex
Dynamics
Conformational exchange
Folding
Ensemble
Nuclear magnetic resonance
Thermodynamics
Kinetics

ABSTRACT

G-quadruplexes (GQs) are four-stranded nucleic acid secondary structures formed by guanosine (G)-rich DNA and RNA sequences. It is becoming increasingly clear that cellular processes including gene expression and mRNA translation are regulated by GQs. GQ structures have been extensively characterized, however little attention to date has been paid to their conformational dynamics, despite the fact that many biological GQ sequences populate multiple structures of similar free energies, leading to an ensemble of exchanging conformations. The impact of these dynamics on biological function is currently not well understood. Recently, structural dynamics have been demonstrated to entropically stabilize GQ ensembles, potentially modulating gene expression. Transient, low-populated states in GQ ensembles may additionally regulate nucleic acid interactions and function. This review will underscore the interplay of GQ dynamics and biological function, focusing on several dynamic processes for biological GQs and the characterization of GQ dynamics by nuclear magnetic resonance (NMR) spectroscopy in conjunction with other biophysical techniques. This article is part of a Special Issue entitled: Biophysics in Canada, edited by Lewis Kay, John Baenziger, Albert Berghuis and Peter Tieleman.

1. Introduction

G-quadruplexes (GQ) are four-stranded, helical nucleic acid structures formed by guanosine (G)-rich DNA and RNA sequences [1–3] (Fig. 1a). They are typically formed by four tracts of three Gs (G_3) with intervening loops sequences of variable composition and length. Sequences with longer (G_{4-7}), shorter (G_2), or uneven G-tracts also form GQ structures [4–6]. GQs are found throughout the genome and in mRNA, notably in gene promoters, telomeres, and telomeric or virus RNA [3,7–9]. GQ structures have been implicated in regulation of gene expression, protein translation, and proteolysis [10–14], pointing to an overarching role in biological function. GQs are also important in biotechnology, forming a key structural component of aptamers [15,16] and catalytic DNA [17,18], among other applications. Intriguingly, many GQ sequences can exchange among multiple folded conformations of similar energies, producing a highly dynamic folded structural ensemble. The potentially large number of energetically similar conformational states for an individual GQ sequence creates an interesting landscape for biological and biophysical study.

Despite the existence of a rich literature on GQ structure and stability, their dynamics are not well understood. GQ dynamics are challenging to study by standard biophysical techniques. For example, solution nuclear magnetic resonance (NMR) spectra of dynamic GQs can be broadened beyond identification of any clear resonances [6,19,20], inhibiting structural characterization and application of standard NMR dynamics experiments. In addition, studying GQ folding with thermal melting experiments often yields multiple or broad transitions [21–23] which require application of sophisticated statistical thermodynamic approaches to appropriately extract thermodynamic information [24–27]. For these reasons, it is experimentally attractive to suppress GQ dynamics with nucleotide substitutions [19,28–30], or to study GQs that naturally populate a single ground conformation. In many cases, the properties of lesser-populated states in GQ ensembles are entirely ignored, meaning that their roles in folding, stability, and biological function are unknown. Conformational exchange processes in proteins and RNA can feature transient excursions to high energy, low-populated structural states that can play important roles in a variety of cellular processes [31–35], and GQ dynamics are likely similarly implicated. The importance of GQ dynamics is starting to be recognized and evidence is accumulating for its contribution to GQ function. This review will discuss important examples of conformationally heterogeneous GQs, focusing on the characterization of GQ dynamics by nuclear magnetic resonance (NMR) spectroscopy and emphasizing how these dynamics could contribute to biological processes.

☆ This article is part of a Special Issue entitled: Biophysics in Canada, edited by Lewis Kay, John Baenziger, Albert Berghuis and Peter Tieleman.

* Corresponding author.
E-mail address: anthony.mittermaier@mcgill.ca (A.K. Mittermaier).

http://dx.doi.org/10.1016/j.bbapap.2017.06.012
Received 24 April 2017; Accepted 16 June 2017

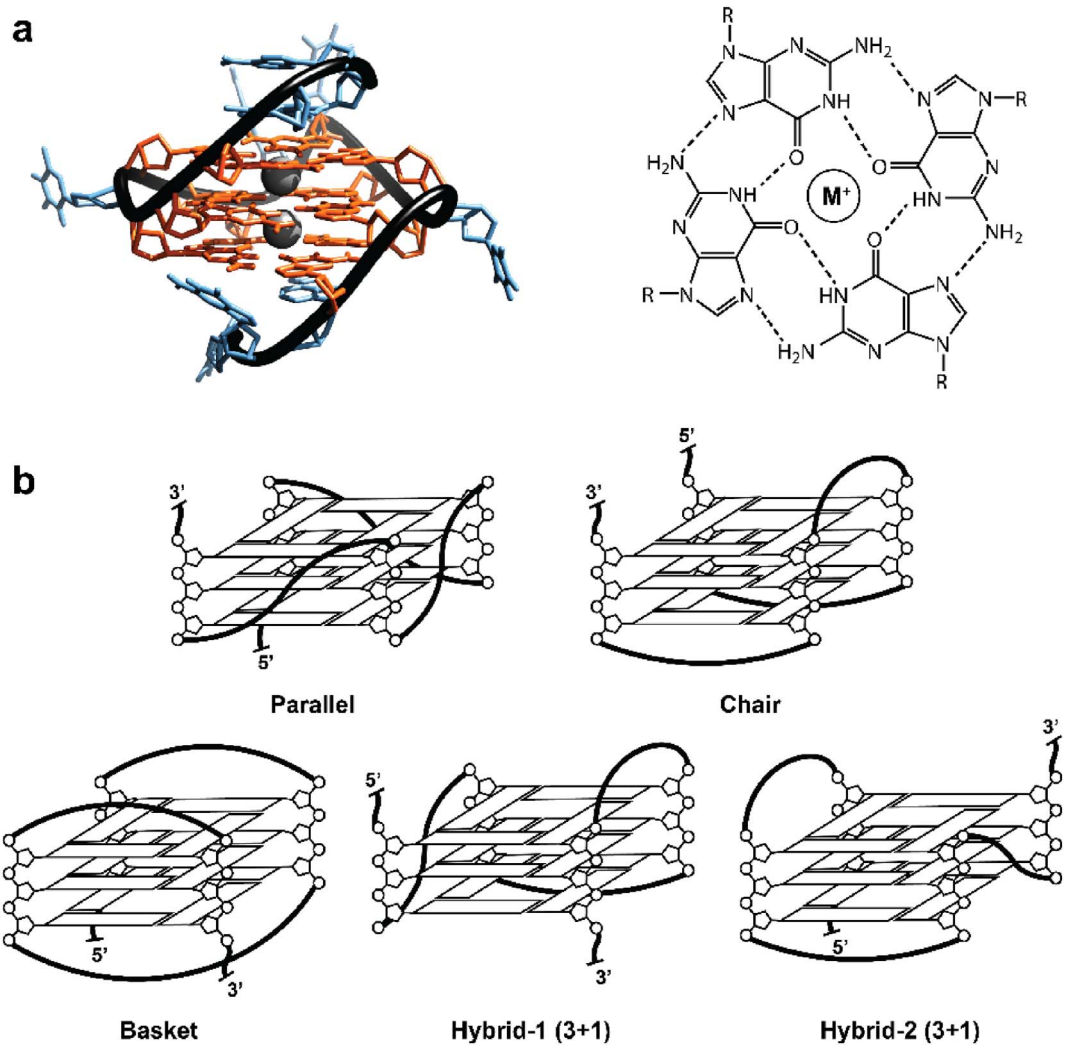

Fig. 1. GQ structure. (a) Solution NMR structure of the c-myc Pu22 sequence 5′-TGAGGGTGGGTAGGGTGGGTAA-3′ (PDB ID: 1XAV) with G-tetrad and coordinated metal depicted immediately to the right. In the 3D structure, core G residues are colored orange, loop and flanking residues are colored blue. Bound K^+ ions are colored grey and the backbone is colored black. (b) GQ topologies.

2. GQ structure

GQs typically form from sequences following 5′-$G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$-3′ [36] where is N is A, C, G, T, or U, although divergent sequences can also adopt a GQ fold [37,38]. The canonical four G_3-tract GQ is formed by assembling the G-tracts into three stacked G-tetrads while adopting energetically favorable loop interactions [29,39] (Fig. 1a, left). G-tetrads are planar arrangements of four Gs engaged in Hoogsteen hydrogen bonding interactions, where each G in the tetrad is rotated roughly 90° with respect to the adjacent ones (Fig. 1a, right). The second and third G tetrads are rotated relative to the first. This rotation gives rise to the helical nature of GQs, which are almost always right handed, with interesting exceptions [40,41]. Loop sequences can form intra- and inter-loop interactions [42,43], in addition to loop-core arrangements where longer loops form capping structures and stack onto outer G-tetrads [44,45]. GQs may also feature flanking sequences to the 5′- and 3′-ends that form stabilizing capping interactions with the GQ core [7,28] (Fig. 1a, left).

GQ assembly requires suitable cations that are bound between each pair of tetrads, coordinated by the carbonyl groups of the tetrad Gs [46] (Fig. 1a left). A variety of cations can play this structural role including Ca^{2+}, Pb^{2+}, Sr^{2+}, Na^+ and K^+ [47–51], although the last two are considered most biologically relevant as their intracellular concentrations (~10 mM and 140 mM in mammalian cells respectively [52]) are much larger than those of the others. Typically K^+ is bound with highest affinity [53,54]. Interestingly, NH_4^+ is readily coordinated in the GQ core and $^{15}NH_4^+$ has been used to probe cation dynamics within GQs by NMR spectroscopy [55,56]. Cations may also be coordinated by loop sequences and can play a role in loop dynamics [57,58].

GQs fold into different topologies that are characterized by the relative orientations of the four G-tracts and the types of loop motifs that connect them [1,59] (Fig. 1b). G-tracts can be aligned in the GQ core in the same or opposing directions, known as parallel and antiparallel respectively. When all four G-tracts are aligned in the same direction, the GQ is parallel or 4 + 0 topology (4 parallel + 0 antiparallel G-tracts, Fig. 1b parallel). GQs with G-tracts aligned in alternating opposing directions are antiparallel or 2 + 2 topology (2 parallel + 2 antiparallel G-tracts, Fig. 1b chair, basket). Topologies where 3 of 4 G-tracts are parallel while the 4th is antiparallel are called hybrid or 3 + 1 (3 parallel + 1 antiparallel G-tract, Fig. 1b hybrid-1 and hybrid-2). Loops that connect parallel G-tracts are termed double-chain-reversal or propeller type (because they appear similar to a propeller blade, Fig. 1b parallel) and are typically formed by short loop sequences of 1-2 nucleotides in length [19,29,60,61]. Edge-wise (Fig. 1b chair) and diagonal (Fig. 1b basket) loops connect antiparallel G-tracts. These types of loops run along the edge or bridge the diagonal of the top or bottom face of the GQ and are typically formed by longer loop sequences of 3 + nucleotides in length [62,63]. GQ topologies can also be