



Review article

Review: Plant G-quadruplex (G4) motifs in DNA and RNA; abundant, intriguing sequences of unknown function

Brianna D. Griffin, Hank W. Bass*

Department of Biological Science, 319 Stadium Drive, Florida State University, Tallahassee, FL, 32306-4295, USA



ARTICLE INFO

Keywords:

G-quadruplex

G4

Gene regulation

cis-regulatory element

ABSTRACT

DNA sequences capable of forming G-quadruplex (G4) structures can be predicted and mapped in plant genomes using computerized pattern search programs. Non-telomeric G4 motifs have recently been found to number in the thousands across many plant species and enriched around gene promoters, prompting speculation that they may represent a newly uncovered and ubiquitous family of *cis*-acting elements. Comparative analysis shows that monocots exhibit five to ten times higher G4 motif density than eudicots, but the significance of this difference has not been determined. The vast scale and complexity of G4 functions, actual or theoretical, are reviewed in relation to the multiple modes of action and myriad genetic functions for which G4s have been implicated in DNA and RNA. Future experimental strategies and opportunities include identifying plant G4-interactomes, resolving the structures of G4s with and without their binding partners, and defining molecular mechanisms through reporter gene, genetic, or genome editing approaches. Given the global importance of plants for food, clothing, medicine, and energy, together with the potential role of G4 motifs as a widely conserved set of DNA sequences that could coordinate gene regulation, future plant G4 research holds great potential for use in plant improvement strategies.

1. Introduction

Genetic *cis*-regulatory elements control gene expression. These elements exist in DNA and RNA in a variety of structural forms. Among these, the non-B-form DNA structures have fascinated biologists for decades (reviewed in [1]), classified into categories such as Z-DNA, hairpin and cruciform DNA, slip-stranded DNA, triplex DNA, and G-quadruplex (G4) DNA. Here we review the relatively young field of plant G4 research. The G-quadruplex is a 4-stranded nucleic acid structure formed by G-G Hoogsteen base pair stacking in the presence of monovalent cations (such as potassium or sodium) as diagrammed in Fig. 1. The common and defining feature of a G4 structure is the planar stacking of guanine quartets (reviewed in [2,3]). The intrinsic properties of guanines to self-assemble and gel into non-random arrangements was first reported by Bang in 1910 [4]. The helical arrangement of stacked planes of guanine quartets was later described by Gellert et al. [5], providing a structural framework for our modern concept of G-quadruplex structures that form not only in solutions of guanylic acid but also in G-rich polymers of DNA and RNA. There also exists a remarkable amount of structural heterogeneity in both the number of structures a single sequence can form [6], and the variations in the G4s themselves resulting from strand polarities, and the size and

composition of loops and bulges, as shown in part in Fig. 2 (reviewed in [7]).

G4s were initially recognized as a hallmark of telomere repeat sequences for most eukaryotic species [8,9]. More recent studies implicate G4s in multiple genetic functions that include gene expression, DNA replication, recombination, and DNA repair [10,11]. It is now widely accepted that G4s can impact many genetic processes as summarized in Fig. 3. The genomic distribution of G4 motifs in and around genes has led to research investigating their roles as *cis*-acting regulatory elements [7,10,12–14]. One of the most active areas of G4 research focuses on the role of mammalian G4s in cell-cycle control and cancer [7,15–21].

In contrast, relatively little is known about G4s in plants, despite the widespread occurrence of these motifs in the genomes of eudicots and monocots [22–25]. This review summarizes recent studies on plant G4s, including species comparisons of genome-wide surveys, computer programs used to identify motifs, their locations relative to genetic elements, and their potential roles in both DNA and RNA processes. Finally, we discuss future experimental approaches needed to further identify and characterize the functional roles for DNA and RNA G4s in plant species.

* Corresponding author.

E-mail addresses: bdg13@my.fsu.edu (B.D. Griffin), bass@bio.fsu.edu (H.W. Bass).

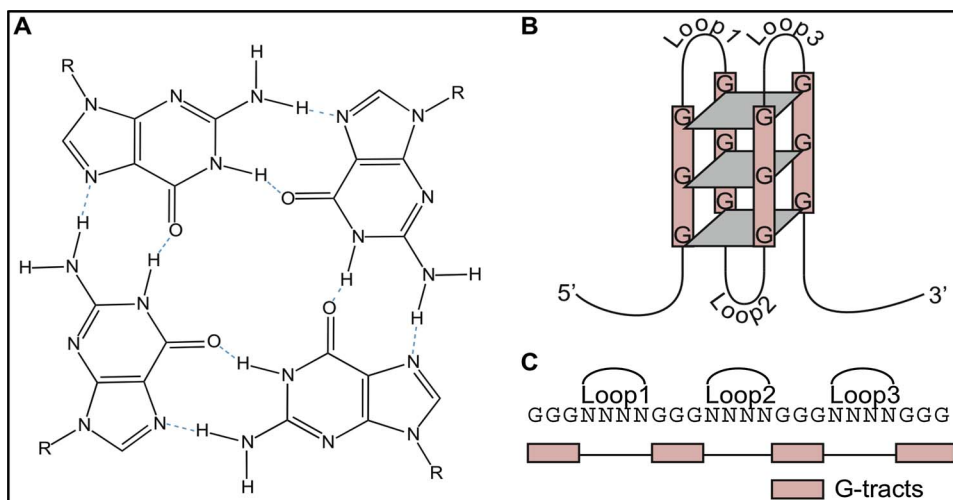


Fig. 1. Representations of G-quadruplex structures (A) G-quartet showing the four guanine bases that are hydrogen bonded to form one plane of a G-quadruplex. R represents the location of the sugar-phosphate backbone (not depicted). (B) Intramolecular G-quadruplex consisting of three G-quartets (grey planar squares). Typical intramolecular G4s have four stems (G-tracts) and three loops (Loop1-3). Not depicted here are the monovalent cations (typically K⁺ positioned in the center of the G4 and intercalated between each pair of adjacent, stacked planes). The structure depicted is but one of many possible topological variants of G4s. (C) Linear G-quadruplex motif sequence marking the locations of the G-tracts and the loop nucleotides. Figure reproduced from Fig. 1 in [56].

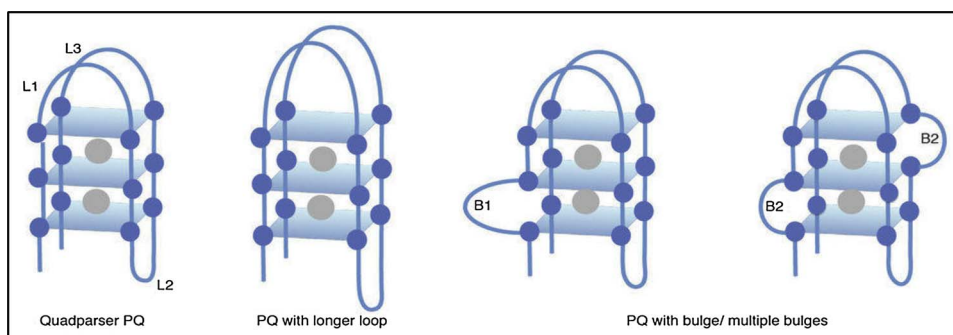


Fig. 2. Various types of G-quadruplex structures. Sample schematics of predicted quadruplexes (PQ) with normal (L1, L2, L3, 1st structure) loops or longer (2nd structure) loops. The last two structures depict G-quadruplexes with bulges in which stacked guanines are interrupted by extruded bases (bulges, B1, B2). Not depicted here are additional conformations that G4s can adopt depending on the direction of the strands in each stem of the quadruplex. Figure reproduced from Fig. S8 in [57].

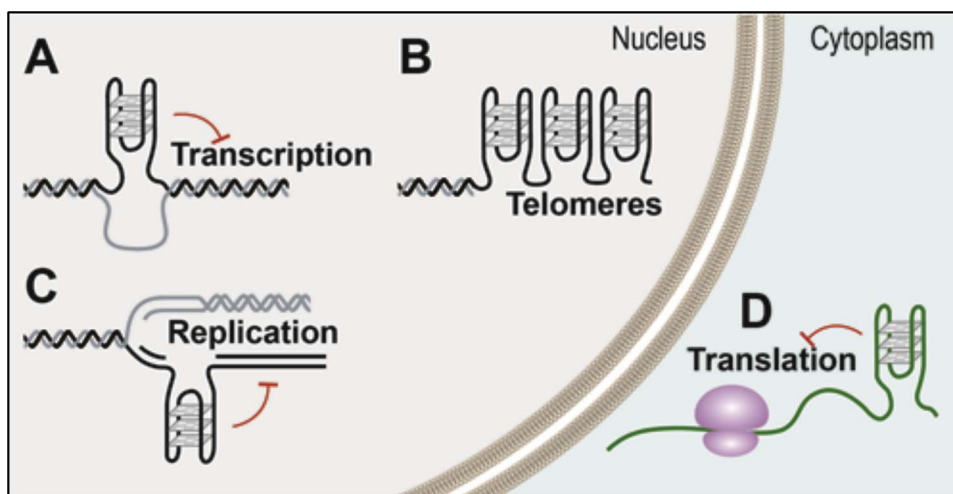


Fig. 3. Major processes involving G4s. Summary diagram of genetic processes (transcription, replication, translation) and structures (telomeres) associated with G4s. (A) In transcription, G4s can form upstream or downstream of the transcription start site and impede or increase (not depicted) transcription. (B) Telomeres can have G4s in the 3' G-rich strand, depicted here in the single-stranded region. (C) During replication, G4s can impede fork progression. (D) In translation, G4s can impede protein synthesis or form structures recognized by various ligands (not depicted). Figure reproduced from Fig. 2 in [12].

2. Genome-wide plant G4 motifs and informatics

Most plant G4 studies begin with a bioinformatic screen of genomic sequence data to find sequence patterns, or motifs, using one or more algorithms [26–28]. Genome-wide surveys show that G4 DNA sequence motifs with the potential to form G-quadruplexes are pervasive in plant genomes [22–25,29,30] and are notably abundant in monocots, as summarized in Fig. 4. For direct comparison, using the so-called canonical G4 search pattern (G₃L₁₋₇), Arabidopsis has only 9 G4 motifs/Mbp whereas rice has 92 G4 motifs/Mbp, a 10-fold difference. Gene ontology analysis has yet to point to any universally common plant function in conserved G4-containing genes, but the reason for this remains unclear. The assignment of G4-containing genes to multiple

different gene ontologies is not limited to plants but is also true for animals and bacteria [31–34]. In plants, it is possible that G4 elements evolve relatively quickly compared to other sequences, allowing them to be redeployed or acquire new functions in divergent taxa. Alternatively, it is possible that they have a shared function that is not easily recognized by examination of the genes that they reside in.

3. Numerous conceivable modes of action for all major genetic processes

Among the more intriguing and challenging aspects of G4 research are their vast array of possible roles and biochemical functions (reviewed in [13]). Indeed, they comprise a diverse collection of

Download English Version:

<https://daneshyari.com/en/article/8356679>

Download Persian Version:

<https://daneshyari.com/article/8356679>

[Daneshyari.com](https://daneshyari.com)