G-quadruplex structures: from basic understanding to biological application

Chun Kit Kwok[1,*] & Catherine J. Merrick[2,*]

[1]Department of Biology and Chemistry, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China

[2]Centre for Applied Entomology and Parasitology, School of Life Sciences, Keele University, Keele, Staffordshire, United Kingdom

*Corresponding authors:

Kwok C.K. (ckkwok42@cityu.edu.hk); Merrick C.J. (c.merrick@keele.ac.uk)

Keywords

Nucleic acid structure, G-quadruplex, cell imaging, next-generation sequencing, aptamer, biosensing

Abstract

Guanine-rich sequences in nucleic acids can assemble into G-quadruplex structures that involve G-quartets linked by loop nucleotides. The structural and topological diversity of G-quadruplexes have attracted great attention for decades. Recent methodological advances have enabled us to identify and characterize G-quadruplexes *in vivo* as well as *in vitro*, and at a much higher resolution and throughput, which has greatly expanded our current understanding of G-quadruplex structure and function. Accumulating knowledge about the structural properties of G-quadruplexes has facilitated the design and development of a repertoire of molecular and chemical tools for biological applications. This review highlights how these exciting methods and findings have opened new doors to investigate the potential functions and applications of G-quadruplexes in basic and applied biosciences.

Importance of G-quadruplex structures in basic and applied biosciences

Nucleic acid structures are fundamental to the cellular function and regulation of diverse biological events [1], including but not limited to DNA replication, transcription, and translation. DNA and RNA sequences can fold into myriad structural motifs such as duplexes, hairpins, triplexes, pseudoknots and G-quadruplexes, to assemble the functional structural conformation for their precise biological roles in specific cellular environments. Interestingly, G-rich sequences in DNA and RNA can self-associate into stacks of G-quartets (**Figure 1A**) to form complex structural motifs known as G-quadruplexes [2] (**Figure 1B**). G-quadruplexes are of growing interest in chemistry and biology, largely due to their peculiar and diverse molecular structures, which include parallel and antiparallel topologies (see glossary) (**Figure 1B**), their important regulatory roles in cellular processes (**Figure 1C**), and their versatile structural scaffolds for biological applications such as in biosensing and imaging (**Figure 1D**).

In this review, we first present the classical methods used to predict and identify G-quadruplex structures. Next, we highlight new molecular and chemical tools that enable detection of G-quadruplexes in cell imaging. Then, we focus on innovative next-generation sequencing techniques that map G-quadruplex structures on a genomic or transcriptomic scale. Novel biological insights that have resulted from these studies and limitations of these methods are discussed. After that, we describe approaches that utilize G-quadruplexes to address important problems. Lastly, we present our perspectives on future advances and challenges toward a more complete understanding of G-quadruplex structure-function relationships *in vivo*, which will help in the potential development of a new set of G-quadruplex-based biological applications.

Predicting G-quadruplexes using bioinformatics methods

Classically, a nucleic acid sequence containing four runs of at least three guanines, separated by short stretches of other bases, can potentially fold into an intramolecular G-quadruplex, so the potential to form these motifs can be predicted from primary sequence. In an intramolecular G-quadruplex the guanine runs all occur on the same strand of DNA, whereas in an intermolecular G-quadruplex they occur on both the sense and antisense strands (**Figure 1B**). For the former class, many predictive algorithms have been published over the past decade, including QuadParser [3], QGRS mapper [4], G4P Calculator [5], QuadBase [6] and most recently, G4 Hunter [7]: their relative features are reviewed in [8]. Predicting intermolecular G-quadruplexes is somewhat more complex because it requires

consideration of both DNA strands [9, 10], and algorithms have also been developed specifically for RNA [11] which, being single-stranded, can adopt myriad competing conformations besides G-quadruplexes [12].

Using such algorithms, it is of theoretical interest to predict how many putative G-quadruplex sequences (PQSs) a genome would be expected to contain at random – and therefore whether these sequences are over- or under-represented in real genomes. However, this remains a non-trivial problem because variable genome composition and biased base dyad frequencies render any simple model highly inadequate [3, 13].

The consensus sequence for PQSs has traditionally been $G_3 N_{1-7} G_3 N_{1-7} G_3 N_{1-7} G_3$ [3] but it is increasingly recognised that this does not accurately predict all the PQSs in a genome [14]: motifs with larger loops, non-guanine bulges, etc. may also form G-quadruplexes. Loops as large as N=30 can support G-quadruplex formation *in vitro* [15] and most predictive algorithms permit a user-defined loop length. Short loops are, however, a major factor in G-quadruplex stability [16, 17] and some algorithms incorporate this, together with other factors, into a sliding score for G-quadruplex propensity and stability, rather than a binary prediction [5, 7, 18]. The picture is further complicated by recent evidence that 'bulged' G-quadruplexes can occur *in vitro* (with a non-guanine base interrupting a 3-guanine track sequence) [14, 19], and also that two instead of three guanines quartets can suffice, particularly in RNA, giving rise to $G_2 N_x$ quadruplexes [20]. Since no predictive algorithm is perfect for all purposes, users must define their parameters appropriately and balance the chances of false positives against false negatives. It is then important to confirm *in silico* predictions via the *in vitro* and/or *in vivo* methods discussed below.

Identifying and characterizing G-quadruplexes using biophysical and biochemical methods

A number of experimental methods have been developed to provide support to the computational prediction of G-quadruplexes. These methods can be broadly defined into 2 classes: biophysical methods and biochemical methods (**Table 1**).

Thanks to the unusual structure and folding of the G-quadruplex, one can experimentally identify G-quadruplex formation and investigate their structural properties using biophysical techniques (**Table 1**). For example, the topology of the G-quadruplex structure can be determined by monitoring the positive or negative circular dichroism (CD) signals at specific wavelengths [21]. In general, G-quadruplexes with parallel topology

(**Figure 1B**) have negative and positive CD signals at 240 nm and 262 nm respectively, whereas G-quadruplexes with anti-parallel topology (**Figure 1B**) have these signals at 262 nm and 295nm respectively. Likewise, the thermostability of the G-quadruplex structure can be identified by observing the ultraviolet (UV) signal at 295nm [22]. Upon G-quadruplex melting, the UV absorbance at 295nm decreases, leading to a hypochromic shift that is a distinctive feature of G-quadruplex structure. These biophysical techniques (**Table 1**) are widely used under different *in vitro* conditions to verify G-quadruplex formation; however, these methods are limited to studying short oligonucleotides and thus do not account for the effect of flanking sequences on G-quadruplex formation.

To address this issue, biochemical techniques were employed to interrogate G-quadruplex formation in a longer sequence context (**Table 1**). In the DNA polymerase stop assay, the formation of a G-quadruplex in a DNA template can act as a roadblock and cause polymerase stalling, which halts the primer extension. Han *et al*. previously applied this to study the DNA G-quadruplex structure formed by telomeric DNA sequences, $d(T_2G_4)_4$ or $d(T_2AG_4)_4$, in the template strand [23]. The dimethyl sulfate (DMS) followed by piperidine cleavage assay is based on the fact that the formation of a G-quadruplex will prohibit the N7 guanine methylation caused by DMS, leading to a protection pattern observed at the DNA G-quadruplex region after piperidine cleavage. For example, Williamson *et al.* used this technique to interrogate telomeric DNA sequences, and observed such DMS protection pattern in the G-quadruplex site [24]. In-line probing (ILP) is a slow, spontaneous RNA cleavage reaction that measures the flexibility of each RNA nucleotide: this method was first developed to study the structure of riboswitches, and later applied to RNA G-quadruplexes [25]. Several recent studies have reported the use of ILP to probe the formation of G-quadruplexes in messenger RNAs [12, 26].

Recently, several new biochemical methods were developed to study RNA G-quadruplexes (**Table 1**). Reverse transcriptase can be stalled by RNA G-quadruplex structures during reverse transcription. Kwok *et al*. developed a reverse transcriptase stalling (RTS) assay and coupled this with ligation-mediated PCR to identify the *in vitro* G-quadruplex formation in low-abundance human telomerase RNA [27]. RNA structure can be probed by chemical probes such as DMS and SHAPE reagents [28, 29], and analysed by primer extension. Kwok *et al*., reported the novel use of lithium-based primer extension (LiPE) buffer in reverse transcription that alleviates RTS, and coupled it with SHAPE reagents and DMS to develop SHALiPE and DMSLiPE [30]. Application of *in vitro* SHALiPE and DMSLiPE has revealed G-quadruplex formation in precursor microRNA

149 [30]. Weldon *et al.* have developed a method called FOLDeR (footprinting of long 7-deazaguanine-substituted RNAs), which compared the RNase footprinting results between wildtype and 7-deazaguanine-substituted RNA [31]. Results from FOLDeR have revealed the *in vitro* formation and location of RNA G-quadruplexes in a 681-nucleotide fragment of Bcl-x RNA [31].

Most of the biophysical and biochemical assays described here (**Table 1**) are limited to *in vitro* studies; however, several methods can be adapted for *in vivo* applications, such as the DMS and piperdine cleavage assay, SHALiPE and DMSLiPE. Complementing these exciting biochemical methods with functional assays (e.g. reporter genes, western blotting) and cell imaging experiments (as discussed below) will enable us to uncover the structural and functional role of G-quadruplexes in cells.

Visualizing G-quadruplexes using cell imaging methods

Methods for detecting G-quadruplexes in the DNA or RNA of whole cells have advanced significantly in recent years (**Table 2**). Two structure-specific antibodies are now available to facilitate G-quadruplex immunofluorescence in a range of eukaryotic cells [32-34] and in cells infected with G-quadruplexes-rich viruses [35]. In parallel, a range of 'light-up' chemical probes is rapidly being developed, with the potential advantage over antibodies that they could be deployable in living cells.

The first quadruplex-specific antibody, Sty49, was used over a decade ago to visualise G-quadruplexes in ciliate macronuclei [36], where large amounts of telomeric DNA offer a super-abundance of G-quadruplexes. A long delay then ensued before successful detection was reported in mammalian cells (where telomere repeats are much less abundant). Whole-cell immunofluorescence assays may be particularly challenging because chromatin can obscure G-quadruplex epitopes, and because G-quadruplexes may be dynamic and fold only transiently *in vivo*. Several different protein probes were engineered, including a zinc-finger protein GQ1 [37], a range of designed ankyrin repeat binding proteins or 'DARPins' [38] and a single-chain antibody hf2 [39]: these could all detect G-quadruplexes *in vitro* but proved unsuitable for whole-cell immunofluorescence [38]. Since hf2 could be used to pull-down G-quadruplex DNA fragments from purified genomic DNA [40], the primary problem was probably not the sensitivity of these tools, but the *in cellulo* chromatin context.

Two new antibodies, BG4 [32] and 1H6 [33], have recently proved suitable for whole-cell immunofluorescence, although questions remain about whether they can be sensitive to the level of a single G-quadruplex, or can only detect high local densities of

multiple G-quadruplex motifs. The number of BG4 foci detected in fixed human cells is orders of magnitude lower than the number of G-quadruplex-forming sequences predicted throughout the genome, but dynamic and transient folding might be expected to limit detection *in vivo*, and most antibodies will not detect every possible structural variant of G-quadruplex. Concerning RNA G-quadruplexes, 1H6 apparently does not detect these [33] whereas BG4 does [41].

Turning to the development of small-molecule probes for G-quadruplexes, the status of this field has been recently reviewed [42, 43] and several representative probes are presented in **Table 2**. In brief, probes for use in living cells should be membrane-permeable and minimally cytotoxic, as well as highly selective for G-quadruplex motifs, with strong and specific 'light-up' (or 'light-off') emission versus minimal background. Furthermore, they should not actually induce G-quadruplex formation, but simply detect pre-existing G-quadruplexes – a difficult distinction to assess. This combination of criteria is very demanding and no perfect probe has yet been reported. However, there are many promising candidates, including sensors for DNA [44, 45], RNA [46, 47] and both DNA and RNA G-quadruplexes [48-50]. These are based on a wide variety of chemistries and may be targeted to a particular G-quadruplex-encoding sequence by conjugation with a gene-specific oligonucleotide [47]. Another approach is to attach a fluorophore to a known G-quadruplex-binding protein, thus circumventing the inherent tendency of guanines to quench fluorescence from small molecules, and potentially also mitigating cytotoxicity. A G-quadruplex-binding peptide from the RHAU helicase has recently showed promise as a sensor *in vitro* [51]. Importantly, all such probes will be subject to the same unanswered question as antibodies concerning their sensitivity: can a single G-quadruplex motif be ever detected *in cellulo*?

Genome/Transcriptome-wide mapping of G-quadruplexes using sequencing methods

In contrast to the above-described challenge of detecting single G-quadruplex motifs via whole-cell imaging, the advent of next-generation sequencing (NGS) has provided an excellent opportunity to explore the prevalence, location, diversity, dynamics and biology of G-quadruplexes in both the genome and transcriptome. G-quadruplexes can now be isolated from either chromatin [52] or purified genomic DNA [40] using some of the antibodies listed in **Table 2**, and they can also be detected in purified DNA and RNA by specialized sequencing methods called G4-seq [14] and rG4-seq [20].

Successful chromatin immunoprecipitation (ChIP) of G-quadruplexes was reported only very recently – possibly because suitable antibodies proved elusive, because the native

chromatin context tends to mask the majority of G-quadruplex epitopes, and/or because the PQSs are folded into G-quadruplexes only in specific cellular conditions. Circumventing some of these issues, an antibody pull-down was conducted from naked genomic DNA using the hf2 antibody (**Figure 2A**) but this detected only ~700 G-quadruplexes [40]; subsequently the BG4 antibody (which was also used in a chromatin context for whole-cell IFAs) was used to isolate G-quadruplex-containing fragments from human chromatin, yielding ~10,000 or 1,000 motifs from two different human cell lines (**Figure 2B**) [52]. The isolated regions tended to be non-coding regulatory regions of highly-transcribed genes, suggesting that nucleosome depletion and active transcription probably favour the folding of G-quadruplexes [52].

A powerful alternative to antibody-mediated pull-down is G4-seq: a method designed for the comprehensive identification of sequences that can form G-quadruplexes *in vitro* (**Figure 2C**) [14]. Here, sheared DNA is subjected to NGS in the presence or absence of conditions that favour quadruplex folding (potassium ions and/or the G-quadruplex-stabilising ligand pyridostatin). Under stabilising conditions, G-quadruplexes impede the polymerase, causing a characteristic increased mutation rate in sequence data at the G-quadruplex folded region. The G4-seq technique identified ~700,000 G-quadruplexes in the human genome: orders of magnitude more than ChIP and twice the number predicted *in silico* by standard algorithms. This is because many of the sequenced motifs were bulged or long-looped – and indeed, the majority of G-quadruplexes found by ChIP (79%) were not of the canonical $G_3\ N_{1\text{-}7}$ type either [52]. However, those that *were* of this type tended to represent the strongest ChIP peaks, and there was also a reassuring degree of consonance between the ChIP and G4-seq experiments: 87% of the DNA fragments from ChIP contained sequences identified in G4-seq [14].

The tremendous difference in size between the datasets obtained *in vitro* and *in vivo* (i.e. G4-seq versus G4-ChIP) suggests that most PQSs are dynamic and highly regulated *in vivo*, although more technical explanations such as limited sensitivity of ChIP and condition-dependence for G-quadruplex formation in native chromatin cannot be excluded. Specifically, the effect of cellular factors on these assays, such as G-quadruplex binding proteins and helicases (**Box 1**), is largely unknown and thus warrants detailed examination. Results are also likely to be influenced by sensitivity/specificity of the antibody used, choice of cell line, and variation in experimental protocols and bioinformatics pipelines. Future development in G-quadruplex-antibody-based sequencing methods should aim to address

these issues in order to establish a gold standard for robust mapping of DNA G-quadruplexes in native chromatin for different species.

A similar approach to structure-specific sequencing of RNA was recently reported (**Figure 2D**) [20], in which RNA G-quadruplexes impede the reverse transcriptase enzyme used to generate an NGS library from polyA-enriched RNAs. Like G4-seq, this 'rG4-seq' technique identified a preponderance of unconventional G-quadruplexes such as long loops, bulged, and 2-quartet structures. Notably, significant correlations were reported between RNA G-quadruplex and microRNA target sites and polyadenylation signals, which will likely motivate further in-depth studies to decipher their working mechanisms in cells. As described above and in **Table 1**, some RNA-structure-probing reagents can be used *in vivo*, allowing us to deduce *in vivo* transcriptome-wide RNA secondary structure [28, 29]. A recent report from Guo *et al*. which used transcriptome-wide RNA structure profiling techniques suggested that RNA G-quadruplexes are tightly controlled in human, mouse and yeast, and are under-represented in bacteria [53]. Future experiments may investigate whether these features are generally applicable to other cell types, cellular conditions and species by using multiple structurome methods [28, 29], and orthogonal antibody-based approaches as similarly performed for DNA G-quadruplexes. One future challenge is to identify and characterize the effect of known and unknown RNA G-quadruplex binding proteins (**Box 1**), and investigate whether they regulate the formation and dynamics of G-quadruplexes in cells under different cellular conditions. Overall, these genome and transcriptome-wide studies generate new testable hypotheses and offer future directions to explore the G-quadruplex-mediated biological processes across the tree of life.

G-quadruplexes as molecular tools for biological applications

The adaptability of a structural element under various conditions is an important parameter to consider in molecular tool design and development. Being versatile in nature, G-quadruplexes have been identified on many occasions, binding to diverse classes of targets by using combinatorial nucleic acid library screening or an *in vitro* selection process called SELEX (**Figure 3A**). Among the G-quadruplex-containing aptamers that were reported to recognize proteins and enzymes (**Table 3**), one of the well-studied examples is thrombin-binding aptamer (TBA) [54], which binds to the exosite I of human thrombin with high affinity and selectivity. The crystal structure of TBA revealed that the DNA G-quadruplex is in anti-parallel topology with 2-quartet planes [55], and biochemical studies showed that the presence of thrombin induces the TBA to fold into a G-quadruplex conformation for

binding [56]. Recent studies on TBA and variants suggest that they can be used for biosensing, with nanomolar affinity to thrombin [57, 58].

Besides proteins and enzymes, specific G-quadruplex-containing aptamers have also been found to recognize small molecules (**Table 3**). Using SELEX, Paige *et al*. obtained an aptamer called spinach that recognizes 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), and produces fluorescence upon ligand binding [59]. Recent crystal structural studies showed that the ligand stacked with the unique G-quadruplex [60] (**Figure 3B**). Notably, the G-quadruplex was found to be in a special structural scaffold in the spinach aptamer that is essential for the ligand binding and fluorescence [60]. Recently, the sequence requirement for the fluorescence of spinach RNA was extensively studied [61], and other fluorescent RNAs are suggested to contain the G-quadruplex motif, such as mango and other variants of spinach. These G-quadruplex containing fluorescent RNAs have been modified further to detect different molecules, including metabolite sensing and protein recognition, and have been applied for live cell imaging and sensing of biomolecules [62].

Another notable G-quadruplex application is quadruplex priming amplification (QPA). Taylor *et al.* developed an amplification strategy that involved the spontaneous dissociation of DNA duplex and formation of DNA G-quadruplex upon primer extension reaction, and the formation of G-quadruplex were detected via fluorescence signal from 2-aminopurine (2AP) (**Figure 3C**) [63]. The 2AP was designed to substitute the T at the first loop of dG_3T G-quadruplex, and the 2AP fluorescence signal increases upon G-quadruplex formation. The QPA can be employed for both linear [63] and exponential [64] signal amplification modes. Recently, QPA has been coupled with linear nicking amplification (LNA) [65] (**Figure 3D**), which allows sensitive detection of target molecules as low as the 10 fM range (~6000 molecules in 1 µl solution) [66]. This approach paves the way to detect low abundance target molecules such as pathogenic DNA in the near future. In QPA, the readout is fluorescence signal from 2AP or 3-methylisoxanthopterin (3MI); however, it might be interesting to see if this QPA can be performed without the need of exogenous fluorophores such as 2AP or 3MI, as several recent studies have reported the intrinsic fluorescence of the G-quadruplex alone [67-70]. Besides biosensing, RNA imaging, and QPA, other interesting G-quadruplex applications such as therapeutics and diagnostics (**Figure 1D**) have also been reported and reviewed elsewhere [71-73].

Future perspectives and challenges

Amongst the technologies discussed in this review, some are quite advanced while others remain in their infancy. Biophysical – and to some extent, biochemical – methods for identifying G-quadruplexes have been developed and applied for many years and a wealth of *in silico* algorithms is available for predicting G-quadruplex formation from nucleic acid sequences. However, it is increasingly clear that the rules for G-quadruplex folding are complex, subtle and context-dependent. Few of the algorithms comprehensively incorporate empirical experimental data, few biophysical techniques incorporate the wider sequence context, and few biochemical techniques are applicable *in vivo* to date (**Table 1**). As such, the field calls for new *in vivo* methods with superior resolution, throughput, and sensitivity to investigate the spatial-temporal formation of G-quadruplexes, their structure folding and dynamics, and the effect of cellular factors upon G-quadruplexes that prevail in cellular milieu.

Cell imaging methods have advanced tremendously in the past few years, but challenges remain around the feasibility of resolving single G-quadruplex motifs, the relative accessibility of G-quadruplexes in chromatin contexts, and the potentially transient nature of many G-quadruplexes *in vivo*. Also, it is yet to be seen if the G-quadruplex-specific probes (**Table 2**) can be easily applied to other biological systems. Notably, the cellular localization of G-quadruplexes in particular genes is largely untested: this would require sequence-specific G-quadruplex antibodies, oligonucleotide-conjugated G-quadruplex probes or a combination of IFA and FISH (both of which are highly demanding in terms of sensitivity). In addition, the production of the perfect light-up G-quadruplex probes for use in living cells remains elusive, although some recent progress has been made towards this goal [42, 43, 74].

Like whole-cell imaging, the ‘omic’-level sequencing techniques now available have exploded recently with a series of seminal papers that raise the exciting potential for future refinement and further study. In this field, defining the formation, structure dynamics, and interaction partners of G-quadruplexes *in cellulo* remains the key question. G-quadruplex formation can be influenced by a variety of factors, including metal ions, flanking sequence context, and protein interaction. Recent studies showed that G-quadruplexes could interconvert with stem-loop structures to regulate cellular processes [30, 75-77], suggesting that alternative structures such as duplexes or hairpins may compete with G-quadruplex formation *in vivo*. It will be interesting to structurally probe them *in vivo* and see how prevalent these G-quadruplex structure-switches are in the human and other genomes and transcriptomes. As mentioned above and detailed in **Box 1**, some known G-quadruplex binding proteins have been identified; however, their global effects on G-quadruplex

structures in cells are not characterized. In the future, further studies performed under knock-down/knock-out of G-quadruplex binding proteins, and under different cellular stress conditions, could provide clues about the formation, structure dynamics, and interactions of G-quadruplexes *in vivo.*

Most G-quadruplex studies consider only intramolecular G-quadruplex folding; however, bioinformatics searches have shown the prevalence of intermolecular DNA:RNA G-quadruplexes in human [78]. Given the vast number of predicted intermolecular G-quadruplexes (**Figure 1**), great experimental effort and robust analysis platforms are needed to reveal their pervasiveness, their structural conformational exchange with intramolecular G-quadruplexes or other structural motifs, and their potential functions in cells, such as in transcription. Innovative strategies are thus urgently needed to be able to detect and map these intermolecular G-quadruplex motifs *in vivo* at both the gene/transcript-specific and at the genome/transcriptome-wide levels.

Concluding remarks

Remarkable progress has been made in G-quadruplex research in the past 5 years. We are at an exciting time to explore the *in vivo* G-quadruplex structure at unprecedented resolution, throughput, and sensitivity. We are cautiously positive that the development of a suite of novel methodologies will enable us to reveal the *in vivo* structures and functions of G-quadruplexes in diverse organisms. As emerging evidence suggests a connection between G-quadruplexes, gene regulation, and development of diseases [79-81], these upcoming advancements in G-quadruplex research will likely help to illuminate the underlying biochemical mechanism and the molecular basis of diseases, and also facilitate the rational design and development of G-quadruplex-related tools for various biological applications including sensing, imaging, gene control, therapeutics and diagnostics (**Figure 1D and Table 3**). We look forward with great optimism to the next set of groundbreaking discoveries and applications to be unveiled in the next 5 years.

Acknowledgements

This work is supported by grants from City University of Hong Kong (project no. 9610363, 7200520) and Croucher Foundation (project no. 9500030) to C.K.K, and grants from the UK Medical Research Council (MR/L008823/1 and MR/P010873/1) to C.J.M. We acknowledge S.M. Leung for helping with Figures 2 and 3. We apologize to colleagues whose works are not cited due to space limitation.

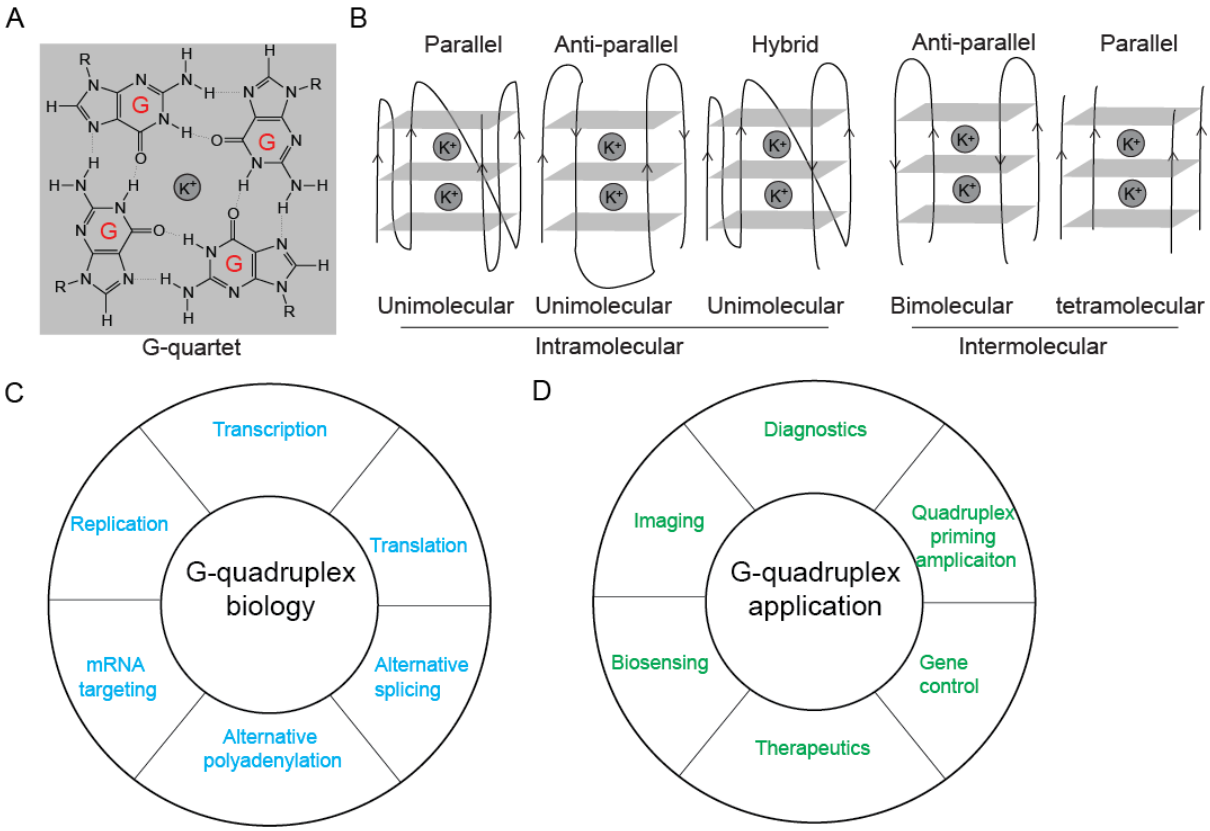


Figure 1. G-quadruplex structure, biology, and applications. A) Chemical structure of a G-quartet. Potassium ion (K^+) sits within the G-quartet for stabilization. G-quartets stack on each other to form G-quadruplex. B) Representative topologies of G-quadruplex structures. C) Representative G-quadruplex-associated biology. D) Representative biological applications that utilize G-quadruplexes.

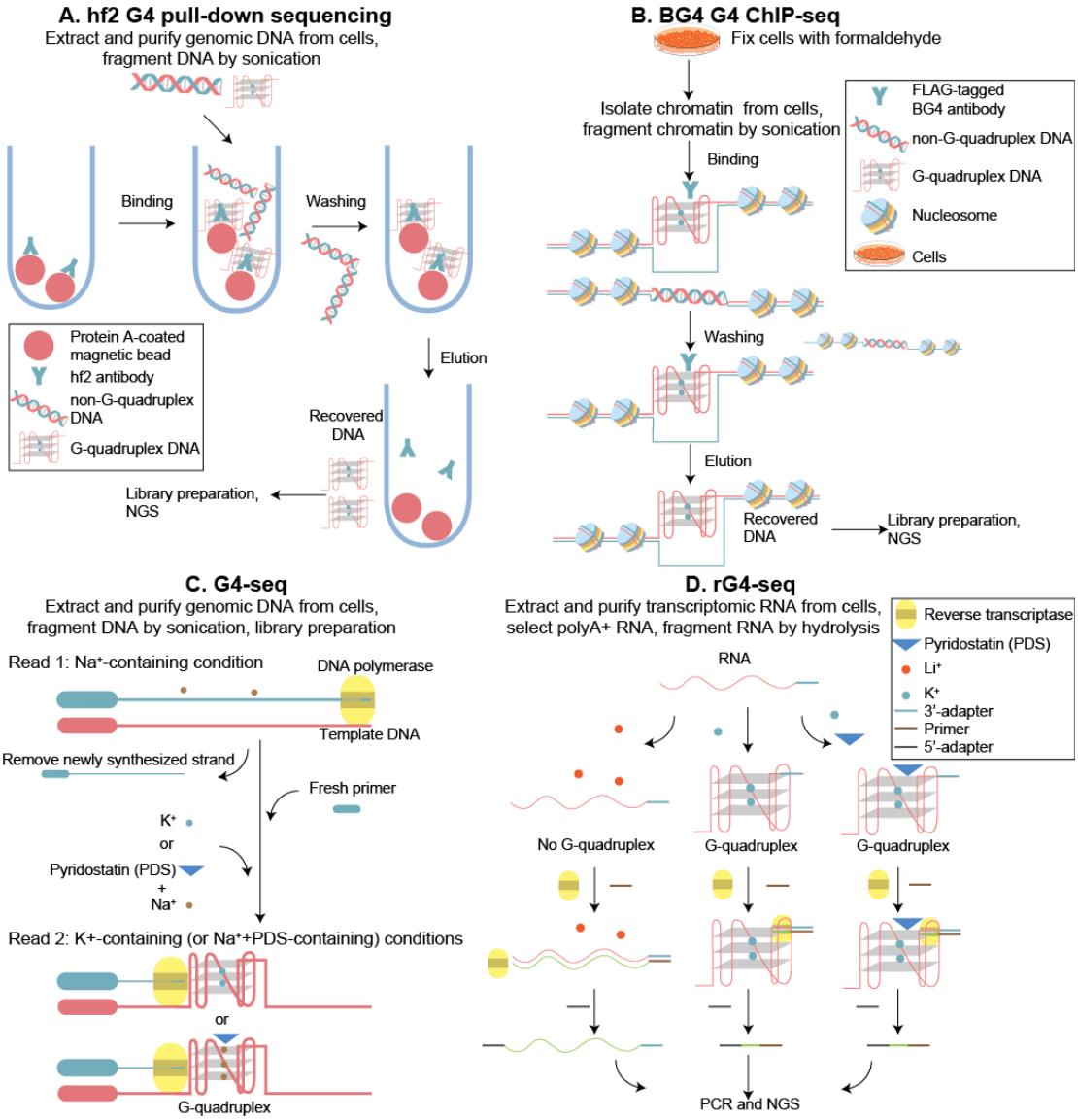


Figure 2. Schematics of G-quadruplex-specific NGS methods. A) hf2 G4 pull-down sequencing [40]: Genomic DNA was isolated from cultured cells and fragmented by sonication. The fragmented DNA was then incubated with G4-specific hf2 antibody. The mixture was washed to remove hf2-unbound DNA, and the hf2-bound DNA was subsequently eluted. The recovered DNA was then library prepared for NGS. B) BG4 ChIP-seq [82]: Cells were fixed with formaldehyde, then chromatin was isolated and fragmented by sonication. The fragmented chromatin was incubated with G4-specific BG4 antibody with a FLAG-tag. The mixture was washed to remove BG4-unbound chromatin, and the BG4-bound chromatin was subsequently eluted. The recovered DNA was library prepared for NGS. C) G4-seq [14]: Genomic DNA was isolated and fragmented by sonication. The fragmented DNA was then library prepared for NGS. The template DNA was first sequenced under Na^+-containing condition to yield read 1. The newly synthesized strand was denatured and washed away. The template DNA was renatured with fresh primer and sequenced again under K^+-containing (or Na^++PDS-containing) conditions to yield read 2. D) rG4-seq [20]: Transcriptomic RNA was isolated from cultured cells, polyA-RNA selected, and fragmented by hydrolysis. 3'-adapter ligation was performed to provide a handle for reverse transcription in the subsequent step. The ligated RNA was divided into three reactions (Li^+-containing, K^+-containing, K^++PDS- containing) and renatured, followed by reverse transcription. The cDNA fragments were ligated to a 5'-adaptor, followed by PCR and NGS.

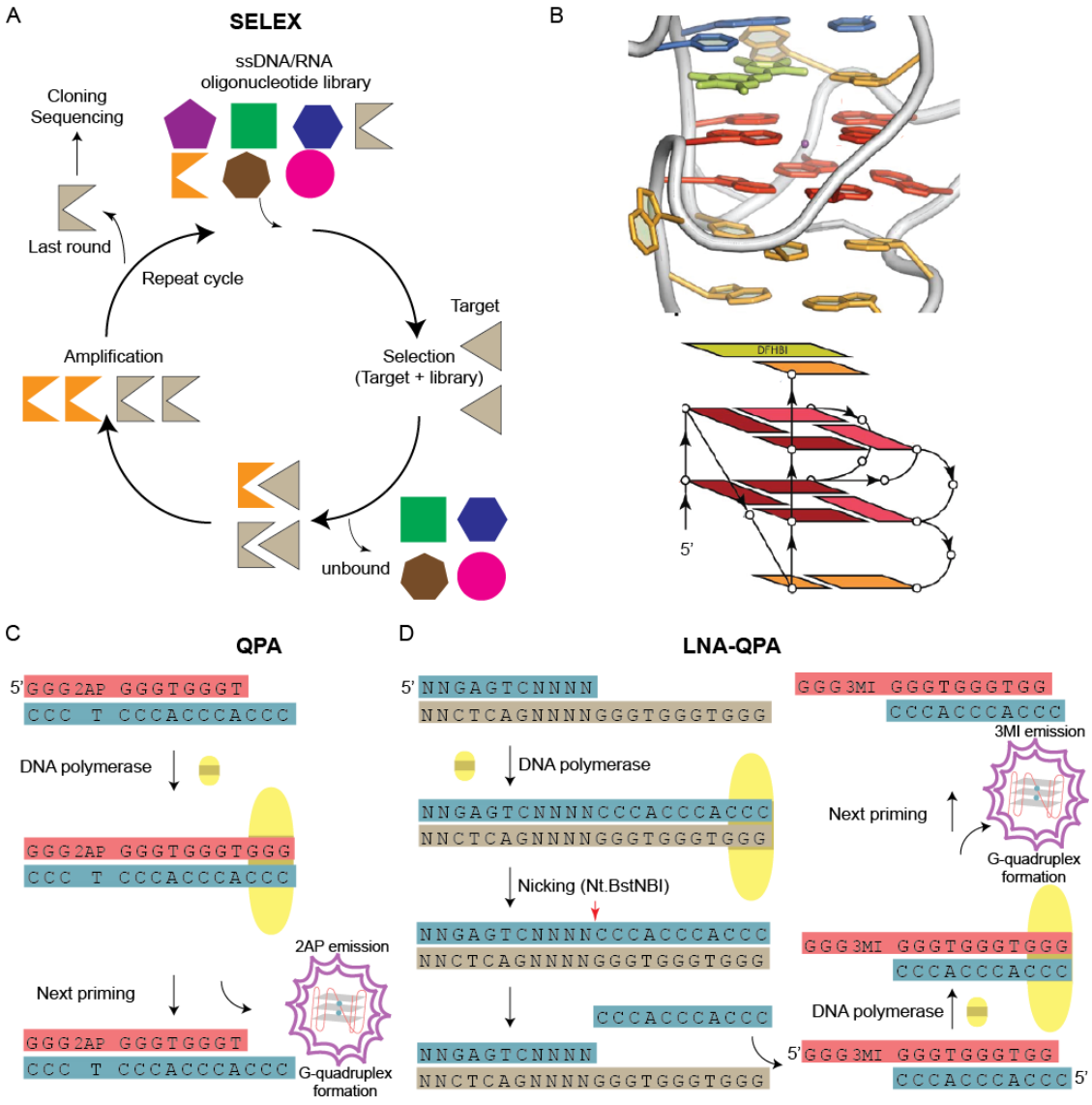


Figure 3. G-quadruplexes as molecular tools for biological applications. A) Schematic representation of SELEX. Random single-stranded DNA or RNA oligonucleotides are synthesized, and are subjected to interaction with the target of interest in the selection step. After that, the washing step removes the unbound oligonucleotides and retains the tightly bound ones. These bound oligonucleotides are then eluted out and are amplified by PCR (for ssDNA) or RT-PCR (for ssRNA). After several cycles, the final candidates are cloned and sequenced to identify the DNA or RNA sequence. B) The G-quadruplex domains of Spinach RNA aptamer (PDB ID:4KZD). The figures were adapted and modified from [60]. C) Schematic representation of QPA. The 2AP containing G-rich primer strand (red) that is missing one track of Gs required to form a G-quadruplex binds with the C-rich template strand (blue). Addition of DNA polymerase allows the missing Gs to be filled in, causes duplex dissociation and induces G-quadruplex formation. The formation of the G-quadruplex leaves the 2AP to be unstacked at the loop of the G-quadruplex, thus producing fluorescence signal for detection. D) Schematic representation of LNA-QPA. A triggering oligo (brown) is hybridized to the target strand (blue). Addition of DNA polymerase allows the target strand to be extended to produce the C-rich region. After nicking with the N.BstNBI endonuclease, the C-rich region can serve as the template strand for the QPA, similar to Fig. 3C. The formation of the G-quadruplex leaves the 3MI to be unstacked, thus producing fluorescence signal for detection.

Table 1. Representative biophysical and biochemical methods to study G-quadruplexes

Method[a]	G-quadruplex information	Features and limitations	References
CD	Topology	DNA and RNA, short oligonucleotide, not applicable in vivo	[21]
UV melting	Thermostability	DNA and RNA, short oligonucleotide, not applicable in vivo	[22]
FRET melting	Thermostabiliity	DNA and RNA, short oligonucleotide, not applicable in vivo	[83]
NMR	3D structure	DNA and RNA, short oligonucleotide, not directly applicable in vivo	[84]
X-ray crystallography	3D structure	DNA and RNA, short oligonucleotide, not applicable in vivo	[85]
Polymerase stop assay	Starting location at nucleotide resolution	DNA, no information on loops and other G-tracks, not applicable in vivo	[23]
DMS and piperdine cleavage assay	Location and structural reactivity at guanine nucleotide resolution	DNA, applicable in vivo, no information on loops	[24]
RNase T1	Location and structural reactivity at guanine nucleotide resolution	RNA, no information on loops, not applicable in vivo	[86]
In line probing	Location and structural reactivity at single nucleotide resolution	RNA, react with 2'OH, not applicable in vivo	[25]
RTS	Starting location at nucleotide resolution	RNA, no information on loops and other G-tracks, not applicable in vivo	[27]
SHALiPE (and DMSLiPE)	Location and structural reactivity at single nucleotide resolution	RNA, applicable in vivo, react with 2'OH	[30]
FOLDeR	Location and structural reactivity at single nucleotide resolution	RNA, requires multiple RNases and 7-deazaguanine substitution, not applicable in vivo,	[31]

[a]CD, circular dichroism; UV, ultraviolet; FRET, fluorescence resonance energy transfer; NMR, nuclear magnetic resonance; DMS, dimethyl sulphate; RTS, reverse transcriptase stalling; SHALiPE, selective 2'hydroxyl acylation analysed by lithium-based primer extension; FOLDeR, footprinting of long 7-deazaguanine-substituted RNAs

Table 2. Representative G-quadruplex-specific antibodies and chemical probes for cell imaging.

Antibodies/probes	Application to date	Features and limitations	References
GQ1 zinc-finger protein	*In vitro* detection of DNA G4s	Not applied in whole-cell IFA	[37]
G4 DARPins	*In vitro* detection of DNA G4s	Not successful in whole-cell IFA	[38]
hf2 single-chain antibody	*In vitro* detection of DNA G4s, pull-down of G4s from genomic DNA	Not applied in whole-cell IFA	[39, 40]
Sty49 single-chain antibody	*In vitro* detection of DNA G4s, IFA on fixed ciliate cells	Detects only high-abundance telomeric G4 DNA in macronuclei	[36]
BG4 single-chain antibody	IFA on human cells, DNA and RNA G4s	Requires 3-step antibody staining protocol. Sensitivity to single-G4 level unproven	[32, 41]
1H6 mouse monoclonal antibody	IFA on human cells, DNA G4s only	2-step antibody staining protocol. Detect DNA not RNA G4s. Sensitivity to single-G4 level unproven	[33, 35]
3,6-bis(1-methyl-4-vinylpyridinium) carbazole diiodide (BMVC)	DNA G4s in human cells	Light-up, cell permeable. Can induce G4s folding *in vivo*	[44]
squarylium dye TSQ1	DNA G4s in human cells	Light-up, cell permeable. Does not induce G4s folding *in vivo*	[45]
cyanine dye CyT	RNA G4s in human cells	Light-up, cell permeable.	[46]
G-quadruplex-triggered fluorogenic hybridization (GTFH) probe, ISCH-nras1	RNA G4 in 5' UTR of NRAS mRNA	Light-up, Gene-specific. Not cell-permeable, not sensitive to natural low RNA levels	[47]
Anthrathiophenedione dye (ATPD)	DNA and RNA G4s in human cells	Light-up, cell permeable.	[48]
NaphthoTASQ (N-TASQ)	DNA and RNA G4s in human cells	Light-up, affinity-triggered by contact with G4s ('smart probe'). Wavelength incompatible with standard light microscopy	[49]
triangulenium derivative DAOTA-M2	DNA and RNA G4s	Light-up, cell permeable, minimal toxicity	[50]
Fluorophore-conjugated RHAU helicase peptide	DNA G4s in vitro	Peptide-based. Not yet tested *in vivo*.	[51]

IFA, immunofluorescence assay

Table 3. Representative list of G-quadruplex-containing aptamers.

G4/rG4-containing aptamer	Targets[a]	Aptamer sequences	References
T40214	STAT3	d(GGGCGGGCGGGCGGGC)	[87]
HJ24	Shp2	d(AGCGTCGAATACCACACGGGGGTTTT GGTGGGGGGGGCTGGGTTGTCTTGGGG GTGGGCTAATGGAGCTCGTGGTCAT)	[88]
3R02	VEGF	d(TGTGGGGGTGGACTGGGTGGGTACC)	[89]
ISIS 5320	HIV gp120	d(T*T*G*G*G*G*T*T)	[90]
AS1411	Nucleolin	d(GGTGGTGGTGGTTGTGGTGGTGGTGG)	[91]
93del	HIV Integrase	d(GGGGTGGGAGGAGGGT)	[92]
RT6	HIV Reverse Transcriptase	d(ATCCGCCTGATTAGCGATACTCAGGCG TTAGGGAAGGGCGTCGAAAGCAGGGTG GGACTTGAGCAAAATCACCTGCAGGGG)	[93]
ODN 93	HIV RNase H	d(GGGGGTGGGAGGAGGGTAGGCCTTAGGTTTC TGA)	[94]
ODN 112	HIV Rnase H	d(CCAGTGGCGGGTGGGTGGGTGGTGGGGGGAC TTGG)	[94]
TBA	Thrombin	d(AGTCCGTGGTAGGGCAGGTTGGGGTGACT)	[54]
RA-36	Thrombin	d(GGTTGGTGTGGTTGGTGGTTGGTGTGGTTGG)	[95]
Scl 2	Sclerostin	d(TTGCGCGTTAATTGGGGGGGTGGGTGGGTT)	[96]
R12	PrP^C	r(GGAGGAGGAGGA)	[97]
PPK2 G9	PPK2	d(AACACATAGGTTTGGTTAGGTTGGTTGGTTGA ATTA)	[98]
Spinach	DFHBI	r(GACGCAACUGAAUGAAAUGGUGAAGGACGGG UCCAGGUGUGGCUGCUUCGGCAGUGCAGCUUG UUGAGUAGAGUGUGAGCUCCGUAACUAGUCGC GUC)	[59]
Mango	TO1	r(UACGAAGGGACGGUGCGGAGAGGAGAGUA)	[99]

* = phosphorothioate bond

[a] STAT3, Signal transducer and activator of transcription 3. VEGF, Vascular endothelial growth factor. HIV, human immunodeficiency virus. PrP^C, cellular prion protein, PPK2, polyphosphate kinase. DFHBI, 3,5-difluoro-4-hydroxybenzylidene imidazolinone. TO1, thiazole orange.

Box 1. Proteins that bind and/or metabolize G-quadruplexes

Many cellular proteins have been identified that interact with DNA and/or RNA quadruplexes. These proteins can be used as tools to probe the distribution and function of quadruplex motifs, as well as being subjects of intense study themselves.

DNA G-quadruplexes

Proteins partners of DNA G-quadruplexes include several groups of structure-specific helicases, such as PIF1, RECQ and FANCJ (recently reviewed in [100]), the transcriptional helicases XPD/XPB [101] and certain non-helicase proteins like nucleolin [102]. Deficiencies in RECQ, FANCJ, and XP helicases are linked to rare human disease syndromes: Fanconi's anaemia for FANCJ, Xeroderma Pigmentosum for XPD/XPB and Bloom's, Werner's and Rothmund-Thomson syndromes for three members of the five-member RECQ family. The diseases are generally characterized by chromosomal instability, telomere deficiency, cancer proneness, etc.: the expected phenotypes for cells that cannot resolve non-canonical DNA secondary structures like G-quadruplexes, and hence suffer high rates of DNA replication fork stalling. At a molecular level, ChIP has demonstrated that these helicases tend to associate with PQSs in the genome, particularly when cells are treated with G-quadruplex-stabilising drugs [103], while in helicase-deficient cells, genes whose expression is deregulated likewise tend to contain PQSs [104]. However, the correlation is not direct and exclusive because some of the helicases also target other structures such as hairpins and chicken-foot structures [105]. Accordingly, ChIP experiments for PIF1 and RECQs may yield many more targets than direct G4-ChIP.

RNA G-quadruplexes

Protein partners of RNA G-quadruplexes include helicases such as RHAU (DHX36) and DHX9, as well as non-helicase proteins like FMRP and Aven. For more details, please see recent excellent reviews [81, 106]. RHAU is one of the most studied helicases for RNA G-quadruplexes. RHAU is shown to be involved in the maturation of human telomerase RNA (hTERC) by unwinding the RNA G-quadruplex at the 5'end of hTERC [107]. FMRP is an important protein that is responsible for fragile X syndrome, and is crucial for cognitive development. Binding assays and bioinformatics analysis of NGS data suggested that it interacts with RNA G-quadruplexes [108, 109]. Recently, a crystal structure revealed that it requires an RNA duplex-quadruplex junction for recognition [110]. Similar to DNA G-quadruplex binding proteins, RNA G-quadruplex binding proteins also target other structures such as triple helices [109, 111, 112]. Thus, one should be cautious about the RNA-protein NGS data obtained on these RNA G-quadruplex binding proteins, as they likely also contain structural motifs that do not fold into G-quadruplexes.

Glossary box

Parallel & anti-parallel G-quadruplex topology: A parallel G-quadruplex has all the guanine-bearing strands in the same 5'/3' polarity, necessitating linking by 'propeller type' loops that run top-to-bottom of the folded motif. In an anti-parallel quadruplex, the strands do not all have the same polarity, and thus the linking loops can be at the top or bottom of the folded motif. See **Figure 1B**.

Intra- & inter-molecular G-quadruplex: An intramolecular quadruplex is formed from a single DNA strand, which bears four runs of guanine residues in close proximity. An intermolecular quadruplex is formed from runs of guanines on more than one DNA strand, or from a hybrid of DNA and RNA strands. See **Figure 1B**.

ChIP: 'Chromatin ImmunoPrecipitation'. A technique to locate proteins – and also DNA motifs such as G-quadruplexes – in native chromatin. Chromatin is formaldehyde-fixed, extracted from cells, fragmented and treated with an antibody to the entity of interest in order to isolate associated DNA fragments. These are then identified by sequencing (ChIP-seq) or by hybridization to a microarray (ChIP-on-chip).

NGS: 'Next-Generation Sequencing'. Modern, high-throughput sequencing techniques such as Illumina, Ion Torrent and 454, all of which produce sequence data concurrently on a genomic/transcriptomic scale in the form of millions of short sequence fragments (usually <1kb).

SHAPE: 'Selective 2′ -Hydroxyl Acylation analyzed by Primer Extension. SHAPE is used to determine RNA secondary structures by treating RNA with an acylation reagent that selectively acylates the flexible (unpaired) nucleotides of the RNA at the 2'hydroxyl (2'OH) group. These modifications can stall reverse transcriptase and thus provide an electrophoresis-based or NGS-based readout of nucleotide reactivity, which can then be used to infer RNA structure.

Light up or light off probe: A 'light up' probe displays enhanced fluorescence upon binding to its target whereas a 'light off' probe undergoes fluorescence quenching.

Aptamer: A biological molecule – usually a peptide or oligonucleotide – that binds to a specific target such as a protein or small molecule. Oligonucleotide aptamers (which may form G-quadruplexes) can be generated by SELEX experiment.

SELEX: 'Systematic Evolution of Ligands by EXponential Enrichment'. A technique for generating highly target-selective oligonucleotides with strong binding affinity from a library of random sequences via repeated rounds of binding to the target ligand, washing, elution, reverse transcription (for RNA aptamer), and PCR amplification. See **Figure 3A**.

References

1. Bloomfield, V.A., Crothers, D.M., and Tinoco, I. (2000) *Nucleic acids: structures, properties, and functions*. Sterling Publishing Company.
2. Neidle, S. and Balasubramanian, S. (2006) *Quadruplex nucleic acids*. Royal Society of Chemistry. Cambridge, UK.
3. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 33, 2908-2916.
4. Kikin, O., D'Antonio, L., and Bagga, P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 34, W676-682.
5. Eddy, J. and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* 34, 3887-3896.
6. Yadav, V.K., Abraham, J.K., Mani, P., Kulshrestha, R., and Chowdhury, S. (2008) QuadBase: genome-wide database of G4 DNA--occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.* 36, D381-385.
7. Bedrat, A., Lacroix, L., and Mergny, J.L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746-1759.
8. Wong, H.M., Stegle, O., Rodgers, S., and Huppert, J.L. (2010) A toolbox for predicting g-quadruplex formation and stability. *J Nucleic Acids* 2010, 564946.
9. Cao, K., Ryvkin, P., and Johnson, F.B. (2012) Computational detection and analysis of sequences with duplex-derived interstrand G-quadruplex forming potential. *Methods* 57, 3-10.
10. Kudlicki, A.S. (2016) G-quadruplexes involving both strands of genomic DNA are highly abundant and colocalize with functional sites in the human genome. *Plos One* 11.
11. Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011) ViennaRNA package 2.0. *Algorithms Mol Biol* 6, 26.
12. Beaudoin, J.D., Jodoin, R., and Perreault, J.P. (2014) New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res.* 42, 1209-1223.
13. Stanton, A., Harris, L.M., Graham, G., and Merrick, C.J. (2016) Recombination events among virulence genes in malaria parasites are associated with G-quadruplex-forming DNA motifs. *BMC Genomics* 17, 859.
14. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P., and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33, 877-881.
15. Guedin, A., Gros, J., Alberti, P., and Mergny, J.L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* 38, 7858-7868.
16. Piazza, A., Adrian, M., Samazan, F., Heddi, B., Hamon, F., Serero, A., Lopes, J., Teulade-Fichou, M.P., Phan, A.T., and Nicolas, A. (2015) Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.* 34, 1718-1734.
17. Kim, M., Kreig, A., Lee, C.Y., Rube, H.T., Calvert, J., Song, J.S., and Myong, S. (2016) Quantitative analysis and prediction of G-quadruplex forming sequences in double-stranded DNA. *Nucleic Acids Res.* 44, 4807-4817.
18. Yano, M. and Kato, Y. (2014) Using hidden Markov models to investigate G-quadruplex motifs in genomic sequences. *BMC Genomics* 15 Suppl 9, S15.
19. Mukundan, V.T. and Phan, A.T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.* 135, 5017-5028.
20. Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., and Balasubramanian, S. (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* 13, 841-844.

21. Paramasivan, S., Rujan, I., and Bolton, P.H. (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods* 43, 324-331.
22. Mergny, J.L., Phan, A.T., and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.* 435, 74-78.
23. Han, H., Hurley, L.H., and Salazar, M. (1999) A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.* 27, 537-542.
24. Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* 59, 871-880.
25. Beaudoin, J.D., Jodoin, R., and Perreault, J.P. (2013) In-line probing of RNA G-quadruplexes. *Methods* 64, 79-87.
26. Kwok, C.K., Ding, Y., Shahid, S., Assmann, S.M., and Bevilacqua, P.C. (2015) A stable RNA G-quadruplex within the 5'-UTR of Arabidopsis thaliana ATR mRNA inhibits translation. *Biochem J* 467, 91-102.
27. Kwok, C.K. and Balasubramanian, S. (2015) Targeted detection of G-quadruplexes in cellular RNAs. *Angew. Chem. Int. Ed.* 54, 6751-6754.
28. Kwok, C.K., Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2015) The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem. Sci.* 40, 221-232.
29. Kwok, C.K. (2016) Dawn of the *in vivo* RNA structurome and interactome. *Biochem. Soc. Trans.* 44, 1395-1410.
30. Kwok, C.K., Sahakyan, A.B., and Balasubramanian, S. (2016) Structural analysis using SHALiPE to reveal RNA G-quadruplex formation in human precursor microRNA. *Angew. Chem. Int. Ed.* 55, 8958-8961.
31. Weldon, C., Behm-Ansmant, I., Hurley, L.H., Burley, G.A., Branlant, C., Eperon, I.C., and Dominguez, C. (2017) Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA. *Nat. Chem. Biol.* 13, 18-20.
32. Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* 5, 182-186.
33. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Jr., Sen, D., and Lansdorp, P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* 42, 860-869.
34. Harris, L.M., Chappell, L., Stanton, A., Veneziano-Broccia, P., Smargiasso, N., Damblon, C., Claessens, A., Rayner, J.C., and Merrick, C.J. (2017) RecQ helicases and G-quadruplex-forming DNA motifs affect virulence gene expression in the malaria parasite *Plasmodium falciparum*. *submitted*.
35. Artusi, S., Perrone, R., Lago, S., Raffa, P., Di Iorio, E., Palu, G., and Richter, S.N. (2016) Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. *Nucleic Acids Res.* 44, 10343-10353.
36. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J., and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U S A* 98, 8572-8577.
37. Isalan, M., Patel, S.D., Balasubramanian, S., and Choo, Y. (2001) Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. *Biochemistry* 40, 830-836.
38. Scholz, O., Hansen, S., and Pluckthun, A. (2014) G-quadruplexes are specifically recognized and distinguished by selected designed ankyrin repeat proteins. *Nucleic Acids Res.* 42, 9182-9194.
39. Fernando, H., Rodriguez, R., and Balasubramanian, S. (2008) Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry* 47, 9365-9371.

40. Lam, E.Y., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat Commun* 4, 1796.
41. Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* 6, 75-80.
42. Vummidi, B.R., Alzeer, J., and Luedtke, N.W. (2013) Fluorescent Probes for G-Quadruplex Structures. *Chembiochem* 14, 540-558.
43. Ma, D.L., Zhang, Z., Wang, M., Lu, L., Zhong, H.J., and Leung, C.H. (2015) Recent developments in G-quadruplex probes. *Chem Biol* 22, 812-828.
44. Tseng, T.Y., Wang, Z.F., Chien, C.H., and Chang, T.C. (2013) In-cell optical imaging of exogenous G-quadruplex DNA by fluorogenic ligands. *Nucleic Acids Res.* 41, 10605-10618.
45. Chen, Y., Yan, S., Yuan, L., Zhou, Y., Song, Y., Xiao, H., Weng, X., and Zhou, X. (2014) Nonlinear optical dye TSQ1 as an efficiently selective fluorescent probe for G-quadruplex DNA. *Organic Chemistry Frontiers* 1, 267-270.
46. Xu, S., Li, Q., Xiang, J., Yang, Q., Sun, H., Guan, A., Wang, L., Liu, Y., Yu, L., Shi, Y., Chen, H., and Tang, Y. (2015) Directly lighting up RNA G-quadruplexes from test tubes to living human cells. *Nucleic Acids Res.* 43, 9575-9586.
47. Chen, S.B., Hu, M.H., Liu, G.C., Wang, J., Ou, T.M., Gu, L.Q., Huang, Z.S., and Tan, J.H. (2016) Visualization of NRAS RNA G-quadruplex structures in cells with an engineered fluorogenic hybridization probe. *J. Am. Chem. Soc.* 138, 10382-10385.
48. Cogoi, S., Shchekotikhin, A.E., Membrino, A., Sinkevich, Y.B., and Xodo, L.E. (2013) Guanidino anthrathiophenediones as G-quadruplex binders: uptake, intracellular localization, and anti-Harvey-Ras gene activity in bladder cancer cells. *J. Med. Chem.* 56, 2764-2778.
49. Shivalingam, A., Izquierdo, M.A., Le Marois, A., Vysniauskas, A., Suhling, K., Kuimova, M.K., and Vilar, R. (2015) The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat Commun* 6, 8178.
50. Laguerre, A., Wong, J.M., and Monchaud, D. (2016) Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic method. *Sci Rep* 6, 32141.
51. Dang, D.T. and Phan, A.T. (2016) Development of fluorescent protein probes specific for parallel DNA and RNA G-quadruplexes. *Chembiochem* 17, 42-45.
52. Hansel-Hertsch, R., Beraldi, D., Lensing, S.V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., Tannahill, D., and Balasubramanian, S. (2016) G-quadruplex structures mark human regulatory chromatin. *Nat Genet* 48, 1267-1272.
53. Guo, J.U. and Bartel, D.P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* 353.
54. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., and Toole, J.J. (1992) Selection of single-stranded-DNA molecules that bind and inhibit human thrombin. *Nature* 355, 564-566.
55. Russo Krauss, I., Merlino, A., Randazzo, A., Novellino, E., Mazzarella, L., and Sica, F. (2012) High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Res.* 40, 8119-8128.
56. Nagatoishi, S., Tanaka, Y., and Tsumoto, K. (2007) Circular dichroism spectra demonstrate formation of the thrombin-binding DNA aptamer G-quadruplex under stabilizing-cation-deficient conditions. *Biochem.* 352, 812-817.

57. Deng, B., Lin, Y., Wang, C., Li, F., Wang, Z., Zhang, H., Li, X.F., and Le, X.C. (2014) Aptamer binding assays for proteins: the thrombin example--a review. *Anal Chim Acta* 837, 1-15.
58. Wang, Q., Zhou, Z., Zhai, Y., Zhang, L., Hong, W., Zhang, Z., and Dong, S. (2015) Label-free aptamer biosensor for thrombin detection based on functionalized graphene nanocomposites. *Talanta* 141, 247-252.
59. Paige, J.S., Wu, K.Y., and Jaffrey, S.R. (2011) RNA mimics of green fluorescent protein. *Science* 333, 642-646.
60. Huang, H., Suslov, N.B., Li, N.S., Shelke, S.A., Evans, M.E., Koldobskaya, Y., Rice, P.A., and Piccirilli, J.A. (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* 10, 686-691.
61. Autour, A., Westhof, E., and Ryckelynck, M. (2016) iSpinach: a fluorogenic RNA aptamer optimized for in vitro applications. *Nucleic Acids Res.* 44, 2491-2500.
62. Strack, R.L., Song, W.J., and Jaffrey, S.R. (2014) Using Spinach-based sensors for fluorescence imaging of intracellular metabolites and proteins in living bacteria. *Nat. Protoc.* 9, 146-155.
63. Taylor, A., Joseph, A., Okyere, R., Gogichaishvili, S., Musier-Forsyth, K., and Kankia, B. (2013) Isothermal quadruplex priming amplification for DNA-based diagnostics. *Biophys. Chem.* 171, 1-8.
64. Kankia, B.I. (2011) Self-dissociative primers for nucleic acid amplification and detection based on DNA quadruplexes with intrinsic fluorescence. *Anal. Biochem.* 409, 59-65.
65. Van Ness, J., Van Ness, L.K., and Galas, D.J. (2003) Isothermal reactions for the amplification of oligonucleotides. *Proc. Natl. Acad. Sci. U S A* 100, 4504-4509.
66. Gogichaishvili, S., Lomidze, L., and Kankia, B. (2014) Quadruplex priming amplification combined with nicking enzyme for diagnostics. *Anal Biochem* 466, 44-48.
67. Dao, N.T., Haselsberger, R., Michel-Beyerle, M.-E., and Phan, A.T. (2011) Following G-quadruplex formation by its intrinsic fluorescence. *FEBS Lett.* 585, 3969-3977.
68. Kwok, C.K., Sherlock, M.E., and Bevilacqua, P.C. (2013) Effect of loop sequence and loop length on the intrinsic fluorescence of G-quadruplexes. *Biochemistry* 52, 3019-3021.
69. Dao, N.T., Haselsberger, R., Michel-Beyerle, M.E., and Phan, A.T. (2013) Excimer formation by stacking G-quadruplex blocks. *Chemphyschem* 14, 2667-2671.
70. Sherlock, M.E., Rumble, C.A., Kwok, C.K., Breffke, J., Maroncelli, M., and Bevilacqua, P.C. (2016) Steady-State and Time-Resolved Studies into the Origin of the Intrinsic Fluorescence of G-Quadruplexes. *Journal of Physical Chemistry B* 120, 5146-5158.
71. Collie, G.W. and Parkinson, G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem Soc Rev* 40, 5867-5892.
72. Tian, T., Xiao, H., and Zhou, X. (2015) G-quadruplex's applications in biological target detection and drug delivery. *Curr Top Med Chem* 15, 1988-2001.
73. Platella, C., Riccardi, C., Montesarchio, D., Roviello, G.N., and Musumeci, D. (2016) G-quadruplex-based aptamers against protein targets in therapy and diagnostics. *Biochim Biophys Acta*, http://doi.org/10.1016/j.bbagen.2016.1011.1027.
74. Ma, D.L., Wang, M., Lin, S., Han, Q.B., and Leung, C.H. (2015) Recent development of G-quadruplex probes for cellular imaging. *Curr Top Med Chem* 15, 1957-1963.
75. Pandey, S., Agarwala, P., Jayaraj, G.G., Gargallo, R., and Maiti, S. (2015) The RNA stem-loop to G-quadruplex equilibrium controls mature microRNA production inside the cell. *Biochemistry* 54, 7067-7078.

76. Kuo, M.H., Wang, Z.F., Tseng, T.Y., Li, M.H., Hsu, S.T., Lin, J.J., and Chang, T.C. (2015) Conformational Transition of a Hairpin Structure to G-Quadruplex within the WNT1 Gene Promoter. *J. Am. Chem. Soc.* 137, 210-218.
77. Onel, B., Carver, M., Wu, G., Timonina, D., Kalarn, S., Larriva, M., and Yang, D. (2016) A new G-quadruplex with hairpin loop immediately upstream of the human BCL2 P1 promoter modulates transcription. *J. Am. Chem. Soc.* 138, 2563-2570.
78. Zheng, K.W., Xiao, S., Liu, J.Q., Zhang, J.Y., Hao, Y.H., and Tan, Z. (2013) Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.* 41, 5533-5541.
79. Maizels, N. (2015) G4-associated human diseases. *Embo Reports* 16, 910-922.
80. Simone, R., Fratta, P., Neidle, S., Parkinson, G.N., and Isaacs, A.M. (2015) G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett.* 589, 1653-1668.
81. Cammas, A. and Millevoi, S. (2016) RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* 45, 1584-1595.
82. Hansel-Hertsch, R., Antonio, M.D., and Balasubramanian, S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell. Biol.*, doi: 10.1038/nrm.2017.1033.
83. Mergny, J.L., Lacroix, L., Teulade-Fichou, M.P., Hounsou, C., Guittat, L., Hoarau, M., Arimondo, P.B., Vigneron, J.P., Lehn, J.M., Riou, J.F., Garestier, T., and Helene, C. (2001) Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay. *Proc. Natl. Acad. Sci. U S A* 98, 3062-3067.
84. Webba da Silva, M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods* 43, 264-277.
85. Parkinson, G.N., Lee, M.P., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 417, 876-880.
86. Morris, M.J. and Basu, S. (2009) An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry* 48, 5313-5319.
87. Jing, N.J., Li, Y.D., Xiong, W.J., Sha, W., Jing, L., and Tweardy, D.J. (2004) G-quartet oligonucleotides: A new class of signal transducer and activator of transcription 3 inhibitors that suppresses growth of prostate and breast tumors through induction of apoptosis. *Cancer Research* 64, 6603-6609.
88. Hu, J., Wu, J., Li, C., Zhu, L., Zhang, W.Y., Kong, G.P., Lu, Z.X., and Yang, C.J. (2011) A G-quadruplex aptamer inhibits the phosphatase activity of oncogenic protein Shp2 in vitro. *Chembiochem* 12, 424-430.
89. Nonaka, Y., Yoshida, W., Abe, K., Ferri, S., Schulze, H., Bachmann, T.T., and Ikebukuro, K. (2013) Affinity improvement of a VEGF aptamer by in silico maturation for a sensitive VEGF-detection system. *Anal. Chem.* 85, 1132-1137.
90. Wyatt, J.R., Vickers, T.A., Roberson, J.L., Buckheit, R.W., Klimkait, T., Debaets, E., Davis, P.W., Rayner, B., Imbach, J.L., and Ecker, D.J. (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human-immunodeficiency-virus envelope-mediated cell-fusion. *Proc. Natl. Acad. Sci. U S A* 91, 1356-1360.
91. Bates, P.J., Laber, D.A., Miller, D.M., Thomas, S.D., and Trent, J.O. (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Experimental and Molecular Pathology* 86, 151-164.
92. de Soultrait, V.R., Lozach, P.Y., Altmeyer, R., Tarrago-Litvak, L., Litvak, S., and Andreola, M.L. (2002) DNA aptamers derived from HIV-1 RNase H inhibitors are strong anti-integrase agents. *J. Mol. Biol.* 324, 195-203.

93. Michalowski, D., Chitima-Matsiga, R., Held, D.M., and Burke, D.H. (2008) Novel bimodular DNA aptamers with guanosine quadruplexes inhibit phylogenetically diverse HIV-1 reverse transcriptases. *Nucleic Acids Res.* 36, 7124-7135.
94. Andreola, M.L., Pileur, F., Calmels, C., Ventura, M., Tarrago-Litvak, L., Toulme, J.J., and Litvak, S. (2001) DNA aptamers selected against the HIV-1 RNase H display in vitro antiviral activity. *Biochemistry* 40, 10087-10094.
95. Zavyalova, E., Golovin, A., Timoshenko, T., Babiy, A., Pavlova, G., and Kopylov, A. (2012) DNA aptamers for human thrombin with high anticoagulant activity demonstrate target- and species-specificity. *Current Medicinal Chemistry* 19, 5232-5237.
96. Shum, K.T., Chan, C., Leung, C.M., and Tanner, J.A. (2011) Identification of a DNA aptamer that inhibits sclerostin's antagonistic effect on Wnt signalling. *Biochem. J.* 434, 493-501.
97. Mashima, T., Nishikawa, F., Kamatari, Y.O., Fujiwara, H., Saimura, M., Nagata, T., Kodaki, T., Nishikawa, S., Kuwata, K., and Katahira, M. (2013) Anti-prion activity of an RNA aptamer and its structural basis. *Nucleic Acids Res.* 41, 1355-1362.
98. Shum, K.T., Lui, E.L., Wong, S.C., Yeung, P., Sam, L., Wang, Y., Watt, R.M., and Tanner, J.A. (2011) Aptamer-mediated inhibition of Mycobacterium tuberculosis polyphosphate kinase 2. *Biochemistry* 50, 3261-3271.
99. Dolgosheina, E.V., Jeng, S.C., Panchapakesan, S.S., Cojocaru, R., Chen, P.S., Wilson, P.D., Hawkins, N., Wiggins, P.A., and Unrau, P.J. (2014) RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. *ACS Chem Biol* 9, 2412-2420.
100. Mendoza, O., Bourdoncle, A., Boule, J.B., Brosh, R.M., Jr., and Mergny, J.L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.* 44, 1989-2006.
101. Gray, L.T., Vallur, A.C., Eddy, J., and Maizels, N. (2014) G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* 10, 313-318.
102. Gonzalez, V., Guo, K., Hurley, L., and Sun, D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* 284, 23622-23635.
103. Paeschke, K., Capra, John A., and Zakian, Virginia A. (2011) DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* 145, 678-691.
104. Johnson, J.E., Cao, K., Ryvkin, P., Wang, L.S., and Johnson, F.B. (2010) Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res.* 38, 1114-1122.
105. Mohaghegh, P., Karow, J.K., Brosh, R.M., Jr., Bohr, V.A., and Hickson, I.D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.* 29, 2843-2849.
106. Millevoi, S., Moine, H., and Vagner, S. (2012) G‐quadruplexes in RNA biology. *WIREs RNA* 3, 495-507.
107. Sexton, A.N. and Collins, K. (2011) The 5' guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. *Mol. Cell. Biol.* 31, 736-743.
108. Zhang, Y., Gaetano, C.M., Williams, K.R., Bassell, G.J., and Mihailescu, M.R. (2014) FMRP interacts with G-quadruplex structures in the 3'-UTR of its dendritic target Shank1 mRNA. *RNA Biol.* 11, 1364-1374.
109. Suhl, J.A., Chopra, P., Anderson, B.R., Bassell, G.J., and Warren, S.T. (2014) Analysis of FMRP mRNA target datasets reveals highly associated mRNAs mediated by

G-quadruplex structures formed via clustered WGGA sequences. *Hum Mol Genet* 23, 5479-5491.

110. Vasilyev, N., Polonskaia, A., Darnell, J.C., Darnell, R.B., Patel, D.J., and Serganov, A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a beta-turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. U S A* 112, E5391-5400.

111. Jain, A., Bacolla, A., Chakraborty, P., Grosse, F., and Vasquez, K.M. (2010) Human DHX9 helicase unwinds triple-helical DNA structures. *Biochemistry* 49, 6992-6999.

112. Booy, E.P., McRae, E.K., Howard, R., Deo, S.R., Ariyo, E.O., Dzananovic, E., Meier, M., Stetefeld, J., and McKenna, S.A. (2016) RNA helicase associated with AU-rich element (RHAU/DHX36) interacts with the 3'-tail of the long non-coding RNA BC200 (BCYRN1). *J. Biol. Chem.* 291, 5355-5372.

Outstanding questions box

1) Dynamic regulation of G-quadruplexes *in vivo*: Under what conditions and in which cellular compartments do they form? How are they regulated?

2) Cellular imaging: Can G-quadruplexes be imaged in living, as well as fixed, cells in different biological systems? Can they be imaged at single-motif resolution?

3) *In vivo* G-quadruplex 'omics': Can *in vivo* genome-wide and transcriptome-wide mapping of G-quadruplex structures and G-quadruplex interactions be achieved in diverse organisms? Can new sequencing methods be developed to complement the current approaches?

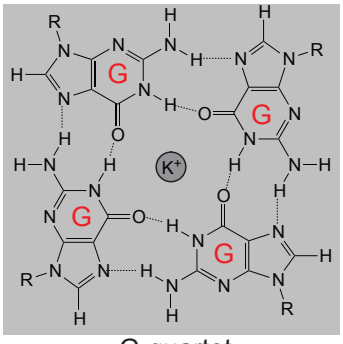
4) Intermolecular G-quadruplexes: Can intermolecular G-quadruplex motifs involving DNA, RNA or DNA:RNA hybrids be identified and characterized *in vivo*? Can new chemicals and molecular tools be developed to achieve this?

5) G-quadruplexes across the tree of life: how are G-quadruplexes used in the biology of diverse species of prokaryotes, eukaryotes, and viruses? Are they evolutionarily conserved?

6) The link between G-quadruplexes and diseases: What are the underlying biochemical mechanisms? Can new G-quadruplex-based diagnostic and therapeutic applications be developed?

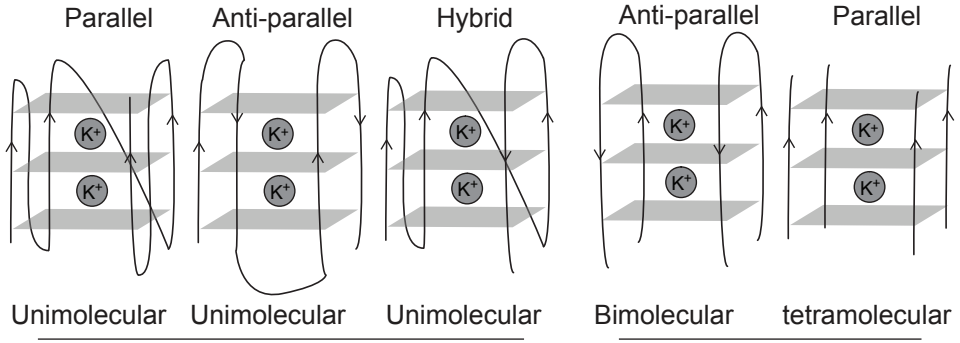
Figure1

A

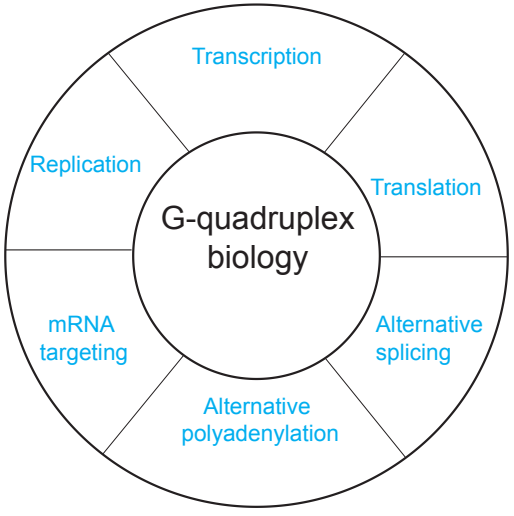


G-quartet

B



C



D

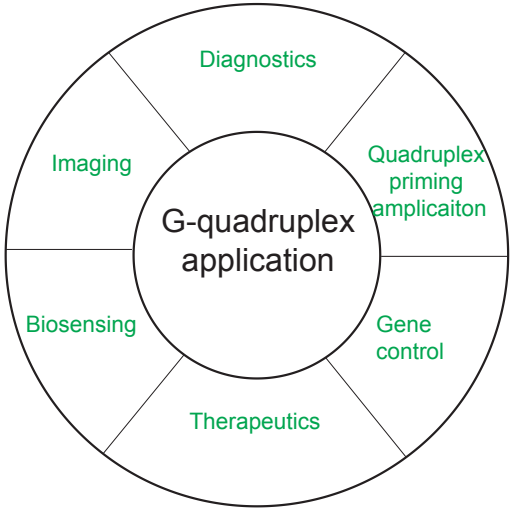


Figure 2

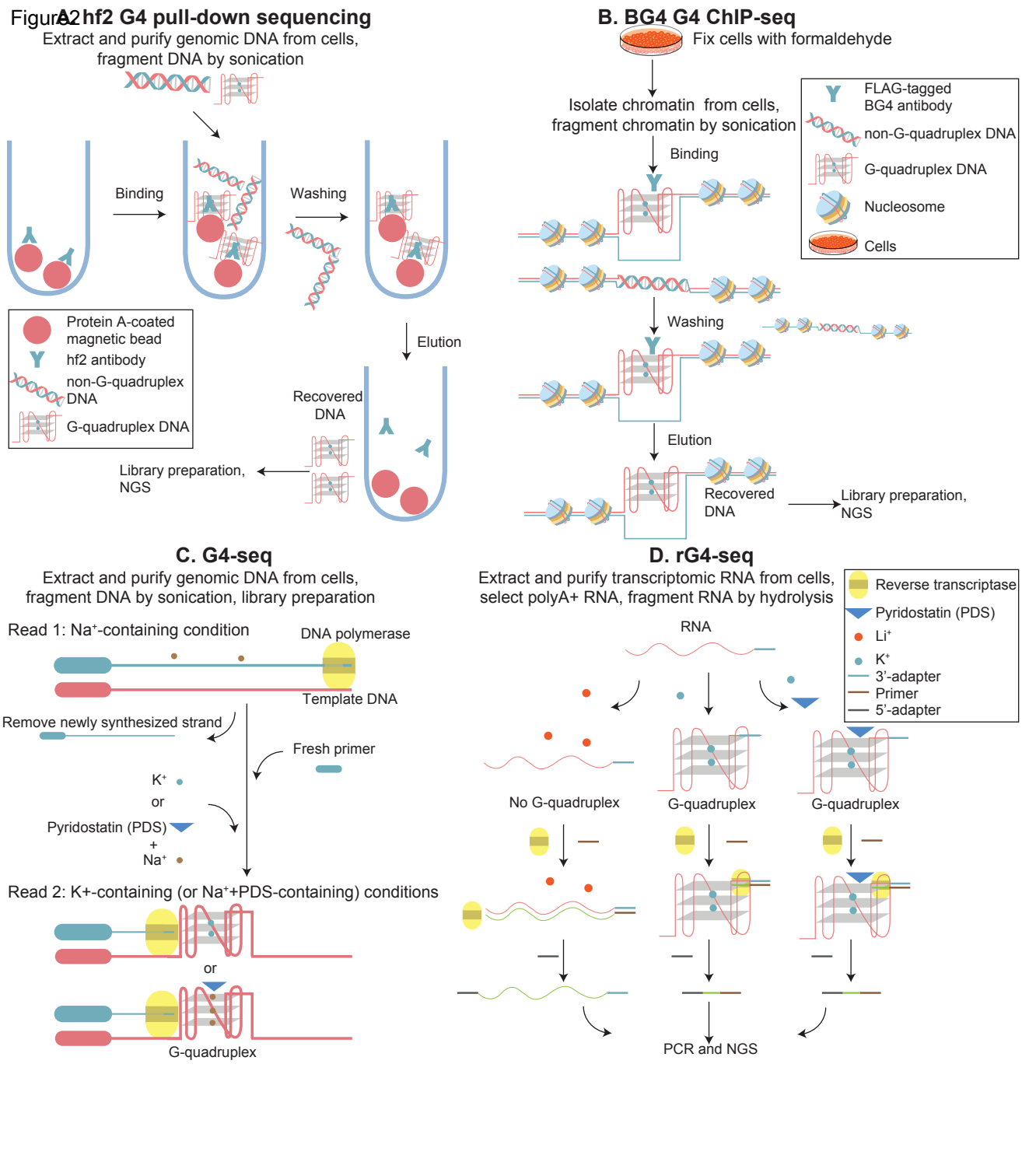
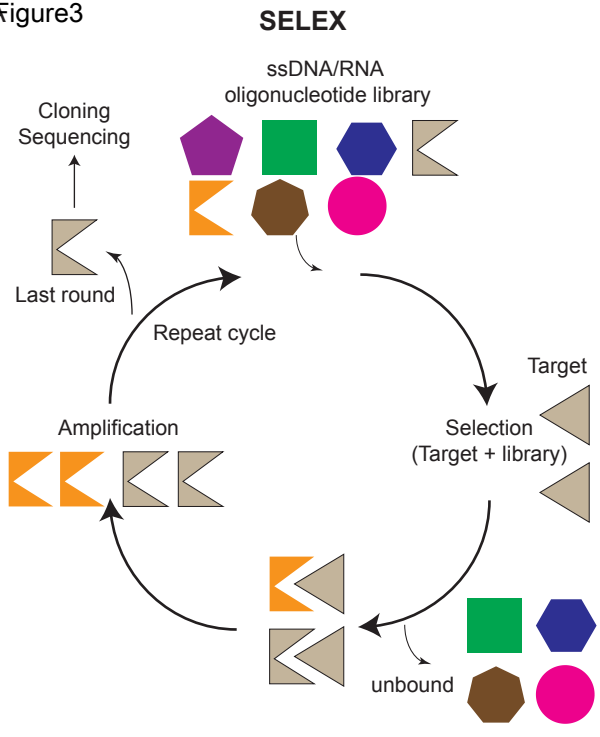
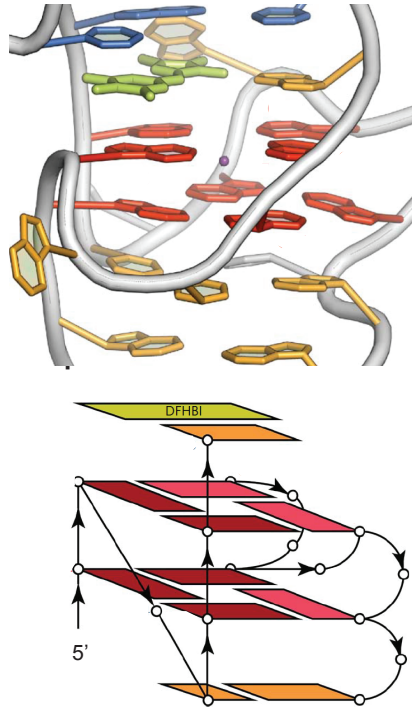

A. hf2 G4 pull-down sequencing
Extract and purify genomic DNA from cells, fragment DNA by sonication
Binding
Washing
Elution
Protein A-coated magnetic bead
hf2 antibody
non-G-quadruplex DNA
G-quadruplex DNA
Recovered DNA
Library preparation, NGS
B. BG4 G4 ChIP-seq
Fix cells with formaldehyde
Isolate chromatin from cells, fragment chromatin by sonication
Binding
Washing
Elution
FLAG-tagged BG4 antibody
non-G-quadruplex DNA
G-quadruplex DNA
Nucleosome
Cells
Recovered DNA
Library preparation, NGS
C. G4-seq
Extract and purify genomic DNA from cells, fragment DNA by sonication, library preparation
Read 1: Na+-containing condition
DNA polymerase
Template DNA
Remove newly synthesized strand
Fresh primer
K+
or
Pyridostatin (PDS)
+
Na+
Read 2: K+-containing (or Na+ +PDS-containing) conditions
or
G-quadruplex
D. rG4-seq
Extract and purify transcriptomic RNA from cells, select polyA+ RNA, fragment RNA by hydrolysis
Reverse transcriptase
Pyridostatin (PDS)
Li+
K+
3'-adapter
Primer
5'-adapter
RNA
No G-quadruplex
G-quadruplex
G-quadruplex
PCR and NGS

Figure3

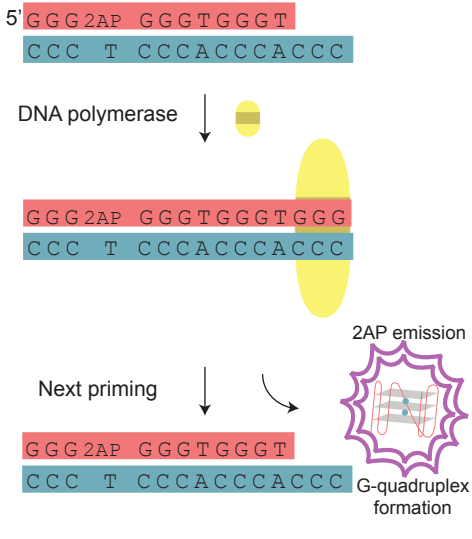
A



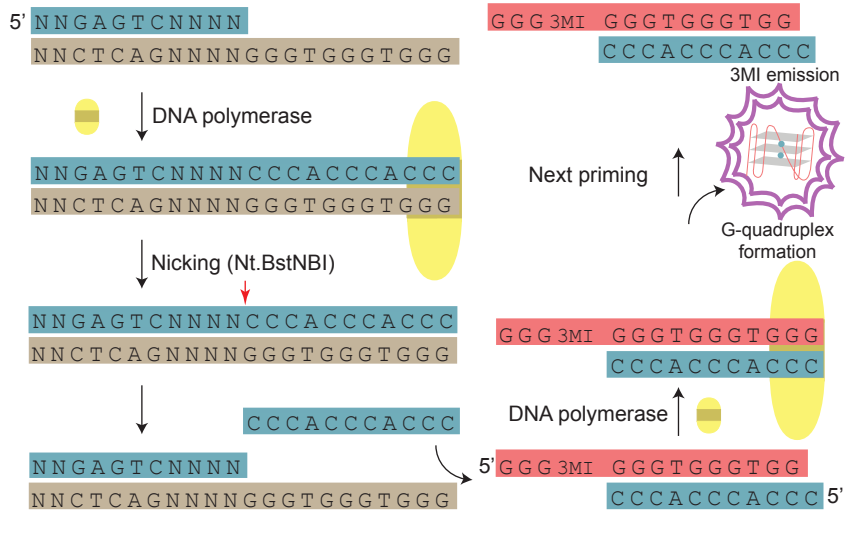
B



C



D



Permission to reprint