Transcriptional control by G-quadruplexes: *in vivo* roles and perspectives for specific intervention

Pablo Armas, Aldana David and Nora Calcaterra

Instituto de Biología Molecular y Celular de Rosario (IBR), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) - Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario (UNR), Ocampo y Esmeralda, (S2000EZP) Rosario, Argentina

Contact: calcaterra@ibr-conicet.gov.ar

Abstract

G-quadruplexes are non-canonical DNA secondary structures involved in several genomic and molecular processes. Here we summarize the main G-quadruplex features and evidences proving the *in vivo* role on the transcriptional regulation of genes required for zebrafish embryonic development. We also discuss alternative strategies for specifically interfering G-quadruplex *in vivo*.

According to the model proposed by James D. Watson and Francis Crick in 1953, the DNA molecule consists of two polynucleotide chains wound around each other to form a clockwise double helix. This structure is known as B form of DNA (B-DNA), and is the predominant conformation adopted by DNA under relaxed conditions. However, non-canonical DNA secondary structures are widespread in all living organisms where they have profound effects on replication, transcription and genome stability. Guanine quadruplex (G-quadruplex), i-motifs, triplexes, cruciforms and hairpins secondary structures can lead to double strand breaks; induction or inhibition of transcription; and initiation or stalling of replication (Figure 1A). The formation of secondary structures may contribute to the generation of genetic diversity, polymorphism, and genome evolution. Conversely, they may result in a variety of genetic disorders, hereditary diseases and cancer through chromosomal rearrangements, mutagenesis, dysregulation of gene expression, and changes in the DNA replication process.

Among various non-canonical DNA structures, G-quadruplexes have attracted enormous research attention as prospective targets for chemical intervention of biological functions. Guanine rich single-stranded DNA has a strong propensity to fold into G-quadruplex *in vitro*. The consensus sequence [$G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$] allows four sets of G triplets to form into three layers of G tetrads stabilized by Hoogsteen hydrogen bonding and K^+ chelation [1, 2]. There are 376,000 putative quadruplex sequences (PQS) in the human genome that have been identified through genome-wide surveys based on quadruplex folding rules [3, 4]. PQS are highly frequent in proto-oncogene promoters [5, 6]; however, not all of them may exist *in vivo*. A prerequisite for the formation of G-quadruplex is the destabilization of the B-DNA double-strand helix, which is highly stable. The current hypothesis states that transcription bubble generates regions of positive and negative supercoiling, which can be propagated along the DNA-helix until reaching sites susceptible of structural transitions. In this condition, transiently exposed single-strand segments become able to fold as G-quadruplexes [7] (Figure 1B). G-quadruplexes may affect gene transcriptional activity either by

upregulation or downregulation, a function that can be evidenced by stabilizing or disrupting G-quadruplex formation through both interacting small ligands (drugs) and/or specific nucleic acid binding proteins [8, 9].

The validation of drug-targeted G-quadruplex DNA and the modulation of oncogenes expression intensely increased in the recent past. Despite a few of new anticancer drugs have entered preclinical or clinical trials [10, 11], the selectivity of these compounds has yet to be improved. Up to now there are no drugs able to discriminate between G-quadruplexes affecting genome stability from those ones controlling gene expression. Even more, drugs cannot discern between G-quadruplexes controlling the transcription of oncogenes from those ones controlling the transcription of other essential genes. In this context, a novel strategy consisting of the use of short antisense DNA sequences or oligonucleotides (ASO) blocking the formation of a specific G-quadruplex has been recently reported [12-14].

Numerous studies performed *in cellulo* have demonstrated the influence of different PQS patterns and loop lengths [15, 16], the effect of ions [17, 18] and the action of specific G-quadruplex ligands [8, 9, 19] on G-quadruplex stability and the transcriptional process. Mostly, these studies consisted in assessing the effect of specific G-quadruplex ligands on the transcriptional expression of reporter genes governed by promoter elements containing PQS. Several reports suggest that G-quadruplexes act as transcriptional repressors by impeding transcription factor binding to duplex-DNA or stalling the progression of RNA polymerase, mostly when they are located downstream the transcription start site in the template strand [2, 20-22]. Conversely, other reports showed that G-quadruplex may enhance the transcription of particular genes by favoring the binding of specific transcription factors [2, 12, 22] or by holding the DNA molecule open thus facilitating the re-initiation of transcription [2, 7, 14, 22] (Figure 1B). Therefore, the hypothesis about a common behavior of G-quadruplexes on transcriptional control would be erroneous.

The *in cellulo* existence and potential impact of G-quadruplex on pathological processes is now accepted, but questions regarding their functions and mechanisms of action *in vivo* remain to be fully addressed. It was reported that microinjection in zebrafish embryos of small G-quadruplex ligands caused G-quadruplex stabilization along with downregulation of *Cdh5* transcription and the generation of embryonic phenotypes mimicking *Cdh5*-morphants [23]. However, such study was unable to conclusively demonstrate that phenotypes were due to the tested ligand on the *Cdh5* G-quadruplex. In view of the high number of PQS present in zebrafish genome [14], the possibility of nonspecific or pleiotropic effects of such ligands could not be ruled out.

Although in past years the knowledge about the biochemical features and cellular roles of the G-quadruplex has made significant progress, the challenge was to demonstrate the direct role of these structures on a specific biological process carried out by complex multicellular organisms. The embryonic development resulted as an appropriate scenario to accomplish this goal. During embryonic development, gene expression is orchestrated by specific and highly evolutionarily conserved mechanisms that take place accurately, both at spatial and temporal levels [24]. An intricate array of *cis*-regulatory sequences controlling individual genes leads to a fine-tuning of gene expression. The last decades have provided compelling evidence that not only protein-mediated transcriptional control but also chromatin state play essential roles in orchestrating stages of embryonic development [25, 26]. Although the enrichment of G-quadruplex in promoter regions of developmentally regulated genes had been found in *in silico* analyses several years ago [27], no experimental evidences proving the role of G-quadruplex in vertebrate embryonic development had been reported. In a recent work, evolutionarily conserved G-quadruplexes located within the proximal promoter region of genes required for proper craniofacial cartilage formation, notochord elongation and eye development were identified. The disruption of these G-quadruplexes *in vivo* by microinjection of specific ASOs in developing zebrafish resulted in lower transcription of the targeted genes, as well as in the recapitulation of the embryonic and larvae phenotypes reported for the respective morphants or mutants [14] (Figure 2). This pioneer work demonstrated the role *in vivo*

of G-quadruplexes as *cis*-acting elements contributing to the transcriptional regulation during the embryonic development, one of the most regulated processes of vertebrate´s biology.

PQS in gene promoters may contain more than four G-stretches or more than three guanines in each stretch, thus resulting in a mixture of conformational isomers in a multiple dynamic equilibrium. Additionally, PQS genetic polymorphisms in G-stretches or even loop-regions may modify G-quadruplex formation and/or stability, thus significantly affecting gene expression among individuals [28]. In a particular cellular context, specific proteins could associate with a subpopulation of conformational isomers shifting the equilibrium toward a particular structure or bind differentially to polymorphic PQS. Consequently, both G-quadruplex alternative conformation and PQS polymorphisms might function as molecular switches enabling gene expression modulation by transitions in DNA structure. Of note, several proteins were identified associated with PQS; most of them are highly conserved zinc-finger DNA binding proteins. Because the versatility of zinc-finger binding pocket is remarkable, it is interesting to consider the implications of G-quadruplex–zinc-finger interactions as a pair [29]. In this context, variations in G-quadrupex conformational isomers, PQS genetic polymorphisms and protein domains may contribute to a fine tuning of transcriptional control.

Conclusions and perspectives

Knowledge about G-quadruplex structure and biochemical features opens up a new field for exploring their biological functions, which might be relevant for understanding the regulation of several cellular and biological processes. G-quadruplexes existence had already been shown both *in cellulo* and *in vivo* and their role in trascriptional regulation is now evident. Profound insights into the mechanisms of G-quadruplex folding, polymorphism affecting G-stretches or even loop-regions, drugs and proteins binding affecting G-quadruplex stability should now be the focus of the scientific community.

G-quadruplexes have been linked with diseases such as cancer, neurodegenerative and genetic disorders, thus providing new clues for the customized design of novel therapeutic strategies. The design of specific molecules capable to discriminate among G-quadruplexes could be helpful at the moment of planning efficient strategies to fight against diseases. The ASO approach allows blocking *in vivo* the formation of a specific G-quadruplex, thus controlling the expression of defined genes. This and similar approaches would emerge as alternative tools to selectively modulate the G-quadruplex-mediated transcription of precise genes allowing to inhibit either disease development or progression reducing undesired collateral effects.

References

1. Murat P, Balasubramanian S. Existence and consequences of G-quadruplex structures in DNA. Current opinion in genetics & development 2014; 25:22-9.

2. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet 2012; 13:770-80.

3. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. Nucleic Acids Res 2005; 33:2908-16.

4. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res 2005; 33:2901-7.

5. Maizels N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nat Struct Mol Biol 2006; 13:1055-9.

6. Verma A, Yadav VK, Basundra R, Kumar A, Chowdhury S. Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. Nucleic Acids Res 2009; 37:4194-204.

7. Du Z, Zhao Y, Li N. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. Genome Res 2008; 18:233-41.

8. Duchler M. G-quadruplexes: targets and tools in anticancer drug design. J Drug Target 2012; 20:389-400.

9. Balasubramanian S, Hurley LH, Neidle S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nature reviews Drug discovery 2011; 10:261-75.

10. Duan W, Rangan A, Vankayalapati H, Kim MY, Zeng Q, Sun D, et al. Design and synthesis of fluoroquinophenoxazines that interact with human telomeric G-quadruplexes and their biological effects. Mol Cancer Ther 2001; 1:103-20.

11. Bates PJ, Laber DA, Miller DM, Thomas SD, Trent JO. Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. Experimental and molecular pathology 2009; 86:151-64.

12. Kumar N, Patowary A, Sivasubbu S, Petersen M, Maiti S. Silencing c-MYC expression by targeting quadruplex in P1 promoter using locked nucleic acid trap. Biochemistry 2008; 47:13179-88.

13. Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature 2015; 525:56-61.

14. David AP, Margarit E, Domizi P, Banchio C, Armas P, Calcaterra NB. G-quadruplexes as novel cis-elements controlling transcription during embryonic development. Nucleic Acids Res 2016; 44:4163-73.

15. Arora A, Maiti S. Stability and molecular recognition of quadruplexes with different loop length in the absence and presence of molecular crowding agents. J Phys Chem B 2009; 113:8784-92.

16. Guedin A, Gros J, Alberti P, Mergny JL. How long is too long? Effects of loop size on G-quadruplex stability. Nucleic Acids Res 2010; 38:7858-68.

17. Marathias VM, Bolton PH. Determinants of DNA quadruplex structural type: sequence and potassium binding. Biochemistry 1999; 38:4355-64.

18. Schultze P, Hud NV, Smith FW, Feigon J. The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G(4)T(4)G(4)). Nucleic Acids Res 1999; 27:3018-28.

19. Monchaud D, Teulade-Fichou MP. A hitchhiker's guide to G-quadruplex ligands. Org Biomol Chem 2008; 6:627-36.

20. Agarwal T, Roy S, Kumar S, Chakraborty TK, Maiti S. In the sense of transcription regulation by G-quadruplexes: asymmetric effects in sense and antisense strands. Biochemistry 2014; 53:3711-8.

21. Maizels N. G4-associated human diseases. EMBO Rep 2015; 16:910-22.

22. Smestad JA, Maher LJ, 3rd. Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription. BMC medical genetics 2015; 16:91.

23. Agarwal T, Lalwani MK, Kumar S, Roy S, Chakraborty TK, Sivasubbu S, et al. Morphological effects of G-quadruplex stabilization using a small molecule in zebrafish. Biochemistry 2014; 53:1117-24.

24. Wolpert L. Do we understand development? Science 1994; 266:571-2.

25. Misteli T. Self-organization in the genome. Proc Natl Acad Sci U S A 2009; 106:6885-6.

26. Rajapakse I, Perlman MD, Scalzo D, Kooperberg C, Groudine M, Kosak ST. The emergence of lineage-specific chromosomal topologies from coordinate gene regulation. Proc Natl Acad Sci U S A 2009; 106:6679-84.

27. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res 2007; 35:406-13.

28. Baral A, Kumar P, Halder R, Mani P, Yadav VK, Singh A, et al. Quadruplex-single nucleotide polymorphisms (Quad-SNP) influence gene expression difference among individuals. Nucleic Acids Res 2012; 40:3800-11.

29. Kumar P, Yadav VK, Baral A, Saha D, Chowdhury S. Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. Nucleic Acids Res 2011; 39:8005-16.

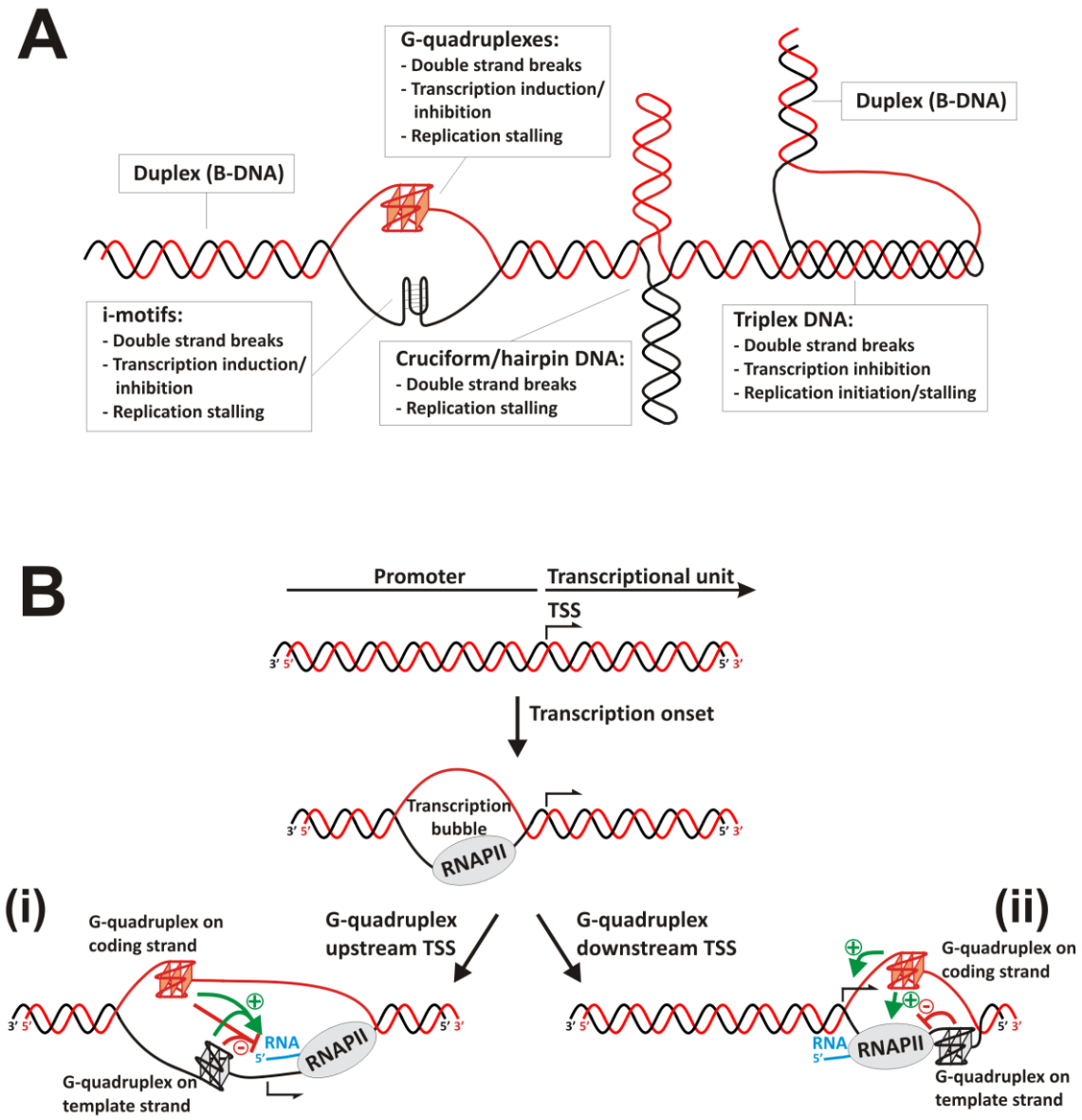

Figure 1. Non-canonical DNA secondary structures influencing transcription. A. Scheme representing the most relevant non-canonical DNA secondary structures and their effects on genome and gene expression. **B.** Transcriptional regulation by G-quadruplexes. After transcription onset, transcription bubble generates transiently exposed single-strand segments able to fold as G-quadruplexes. Two putative scenarios are represented: (i) G-quadruplexes may form upstream the transcription start site (TSS), causing positive or negative effects on transcription depending on

their capability of interfering with RNA Polymerase II or transcription factors binding, recruiting G-quadruplex binding proteins or maintaining an open DNA conformation that facilitates transcription re-initiation. (ii) G-quadruplexes may form downstream the TSS, usually causing positive effects on transcription when located in the coding strand due to favoring transcription re-initiation, or negative effects on transcription when located in the template strand due to stalling the progression of RNA polymerase.

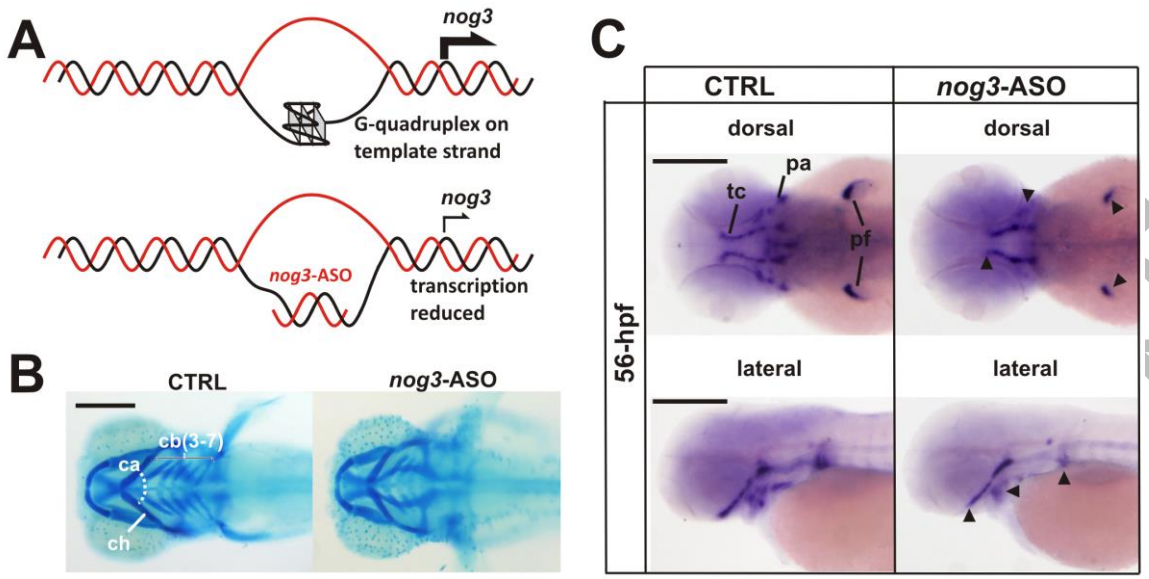

Figure 2. ***Noggin 3*** **(*nog3*), a gene required for proper craniofacial cartilages development, is regulated *in vivo* by G-quadruplex. A.** Strategy to specifically block G-quadruplex formation using an antisense oligonucleotide (*nog3*-ASO) microinjected in zebrafish embryos. **B.** Alcian blue staining showing craniofacial cartilages (ca: ceratohyal cartilages angle; cb (3-7): ceratobranchial cartilages 3 to 7; ch: ceratohyal cartilage) of 4 days post fertilization (4-dpf) larvae. Compared to controls (CTRL), *nog3*-ASO microinjected lavae display reduced head structures and abnormal craniofacial cartilage pattern. **C.** Lateral and dorsal views of whole-mount *in situ* hybridizations showing reduced expression of *nog3*-mRNA in 56 hours post-fertilization (56-hpf) larvae microinjected with *nog3*-ASO when compared with controls (CTRL). pa: pharyngeal arches; pf: pectoral fin; tc: trabeculae cranii. Scale bars = 200 μm.