

Understanding the Interactions of Guanine Quadruplexes with Peptides as Novel Strategies for Diagnosis or Tuning Biological Functions

Tom Miclot,^[a, b] Aurane Froux,^[a, c] Luisa D'Anna,^[a] Emmanuelle Bignon,^[b] Stéphanie Grandemange,^[c] Giampaolo Barone,^{*[a]} Antonio Monari,^{*[d]} and Alessio Terenzi^{*[a]}

Guanine quadruplexes (G4s) are nucleic acid structures exhibiting a complex structural behavior and exerting crucial biological functions in both cells and viruses. The specific interactions of peptides with G4s, as well as an understanding of the factors driving the specific recognition are important for the rational design of both therapeutic and diagnostic agents. In this review, we examine the most important studies dealing with

the interactions between G4s and peptides, highlighting the strengths and limitations of current analytic approaches. We also show how the combined use of high-level molecular simulation techniques and experimental spectroscopy is the best avenue to design specifically tuned and selective peptides, thus leading to the control of important biological functions.

Introduction

The structural organization of nucleic acids is highly complex and rather dynamic, and their polymorphism is also related to the exerted crucial biological functions. Indeed, in addition to the well-celebrated double-helical pairing of B-DNA, as firstly determined by Franklin, Watson, and Crick, other DNA structures have been highlighted, *in vitro* and *in vivo* and have been related to peculiar functions, including cellular regulation.^[1,2] If the structural complexity of DNA is nowadays assumed, the panorama of RNA, which usually lacks a strict double helical arrangement, is even more complex, involving motifs such as stem loops, knots, bulges, and multiloops. A structural organization which is common to both nucleic acids, and is present in guanine rich regions of the genome, is the so-called guanine quadruplex (G4).^[3] G4s can form in single- or multi-stranded nucleic acid sequences, and are basically constituted by stacking of quasi parallel plans each constituted by four guanines, named hereafter tetrads. A minimal number of three

tetrads is usually necessary to assure the stability of the G4, however longer arrangements are possible. The stability of the tetrad is assured by the fact that guanines are locked together through four Hoogsteen-type hydrogen bonds leading to a very rigid arrangement. The cooperative enhancement of the hydrogen bond network, and its role on the structural stability, have also been assessed computationally by *ab initio* methods and energy decomposition analysis.^[4,5] The global stability of the G4 also requires the presence of metal cations in the central channel defined by the nucleobases carbonyl oxygen atoms.^[6,7] The differential stabilizing role of the central cations^[8] and the interplay between their size and their chemical-physical properties has also been deeply analyzed, leading to the occurrence of well-defined preference orders.^[6] The nucleobases constituting a specific tetrad do not need to be contiguous in the primary sequence of the nucleic acid. Indeed, they can be separated by relatively long strands, which will result in the presence of flexible loops decorating the rigid tetrad core, adding conformational complexity.^[9,10] Finally, and despite the apparent simplicity of the G4 arrangement, an inherent polymorphism, derived from the mutual orientation of the strands and the glycoside angle of the guanine should be pointed out, which is translated on the presence of parallel, antiparallel, and hybrid arrangements, each one giving rise to characteristic features in their electronic circular dichroism (ECD) shape, used as fingerprints (Figure 1).^[11]

The equilibrium between the different arrangements remains highly complicated and, as a matter of fact, the dominant conformation is also dependent on the subtle interplay with environmental factors, such as crowding.^[12,13] Spectroscopic techniques, and in particular ECD, remain the methods of choice for a fast and usually accurate determination of the intrinsic G4 structural features.^[14–16] This is due to the high sensitivity of ECD to chiral secondary structure elements, and to the presence of specific spectroscopic signatures for the different G4 conformations (Figure 1). However, ECD by itself lacks

[a] Dr. T. Miclot, A. Froux, L. D'Anna, Prof. Dr. G. Barone, Prof. Dr. A. Terenzi
Università di Palermo, Department of Biological, Chemical and
Pharmaceutical Sciences and Technologies
Viale delle Scienze, 90128 Palermo (Italy)
E-mail: Giampaolo.barone@unipa.it
Alessio.terenzi@unipa.it

[b] Dr. T. Miclot, Dr. E. Bignon
Université de Lorraine and CNRS, UMR 7019 LPCT
54000 Nancy (France)

[c] A. Froux, Prof. Dr. S. Grandemange
Université de Lorraine and CNRS, UMR 7039 CRAN
54000 Nancy (France)

[d] Prof. Dr. A. Monari
Université Paris Cité and CNRS, ITODYS
75006 Paris (France)
E-mail: Antonio.monari@u-paris.fr

© 2023 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

the capacity to provide a fully resolved atomistic picture of the G4s arrangements, and in particular of the behavior of the flexible loops. In this respect, multiscale molecular modeling and simulation, involving both the sampling of the conformational space by classic molecular dynamics and the reproduc-

tion of the ECD signal by quantum chemistry approaches, represent a suitable strategy to achieve fully resolved and validated G4 structures.^[17,18] The role of molecular modeling in the field of G4 interactions will also be highlighted and further discussed in this minireview. On the other hand, bioinformatic



Tom Miclot obtained his Ph.D. in chemistry in 2022 jointly from University of Palermo and University of Lorraine. In 2023 he moved to pursue a postdoc fellowship at Heyrovský Institute, Prague in professor Timr's group. He has been working on modeling the structural stability of G-quadruplex nucleic acids and their interactions with proteins. Now, his main scientific interests are focused on structural organization of glycolytic enzymes assemblies.



Aurane Froux obtained her master's degree in biology from the French University of Lorraine in 2021. She then started a Ph.D. in bioinorganic chemistry and cellular and molecular biology in cotutelle between the Barone's group from the University of Palermo (Italy) and the Grandemange's one from the University of Lorraine (France). She is interested in the stabilization of G-quadruplexes by new transition metal complexes, and their applications as anticancer agents.



Luisa D'Anna obtained her master's degree in chemistry and pharmaceutical technologies in 2021 at the University of Palermo, where she started her Ph.D. in bioinorganic chemistry. She is working on Salphen metal complexes and their interaction with human and viral G-quadruplexes. Recently, she spent a research period at the University of Lorraine (France) in Prof. Grandemange's group.



Emmanuelle Bignon obtained her Ph.D. in Computational Chemistry in 2017 from the University of Lyon I (France). After five years as a postdoc at the Danish Cancer Society, Université Côte d'Azur and Université de Lorraine, she very recently obtained a CNRS associate professor position to start her own group at Université de Lorraine. Her research interests revolve around deciphering the molecular mechanisms of DNA compaction, DNA damage and DNA interaction with potential therapeutics.



Stéphanie Grandemange obtained her Ph.D. in cellular and molecular biology in 2005 from the University of Montpellier II. After three years as a postdoc at the University of Sciences III (Geneva) in Martinou's lab, she was recruited in 2008 to the University of Lorraine as Associate Professor; she has been a full professor there since 2020. Her interests are focused on finding new anticancer agents based on their DNA binding activities and their ability to modulate DNA structure through their interference with epigenetic events or their ability to induce DNA damages.



Giampaolo Barone obtained his Ph.D. in chemistry in 1998 from the University of Palermo, where he is currently Professor of Inorganic Chemistry. He has been a postdoc at the University of Bath, UK (1999), University of Salerno, Italy (1999–2001) and at the EMBL-Hamburg outstation, Germany (2001). He has been a visiting researcher in several European universities. His research interests are mainly focused on the synthesis of DNA-binding metal complexes and on the experimental and computational study of the chemical and spectroscopic properties of metal compounds.



Antonio Monari obtained his Ph.D. in chemistry at the University of Bologna in 2007. After several post-doctoral fellowships at the Universities of Bologna and Toulouse, he was recruited as associate professor at Université de Lorraine in Nancy (France). In 2021 he moved to the Université Paris Cité, in Paris, as full professor. He develops and uses modeling and simulation methods to study complex biological systems, including photobiological phenomena. He is an expert in classical and hybrid dynamical approaches, including enhanced sampling methodologies and non-adiabatic approaches. In particular he focuses on the study of DNA lesion and repair and, more recently, on the molecular bases behind viral replication.



Alessio Terenzi obtained his Ph.D. in chemistry from the University of Palermo (Italy) in 2011, during which he spent time at the Universities of Birmingham (UK) and La Coruña (Spain). After a postdoc in Palermo (2014), he moved to the University of Vienna (Austria) on a Marie Curie Cofund Fellowship where he worked on the in-cell interaction of metal complexes with G-quadruplex motifs. In 2018, he moved to the Donostia International Physics Center (San Sebastian, Spain) on a Marie Skłodowska-Curie Individual fellowship to develop photoactivatable Au^{III} anticancer prodrugs that incorporate nuclear imaging capability. Since 2020, he has been Associate Professor of Inorganic Chemistry at the University of Palermo.

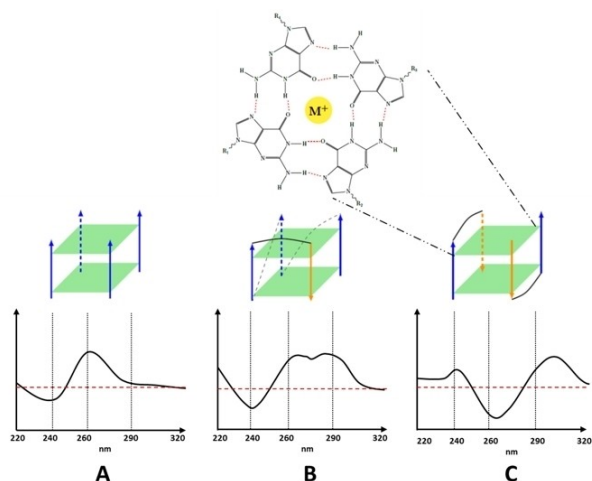


Figure 1. General structure of a guanine tetrad and schematic representation of A) parallel, B) hybrid, and C) antiparallel G-quadruplexes with the corresponding ECD spectra.

approaches leading to the accurate prediction of putative G4-forming spots in nucleic acid sequences have been developed and are publicly available.^[19,20]

From a biological point of view, G4s have been recognized in both RNA and DNA structures,^[21] and have been identified in either cellular or viral systems.^[22] They are involved in a wide range of biological functions through their role in telomere function, genome stability and structure, DNA replication, transcription, mRNA processing and protein translation. Based on their roles in many aspects of nucleic acid biology, studies have also reported their implications in the regulation of mitochondrial functions.^[23] G4s were particularly observed and described in telomeric sequences.^[24–26] In 1991, Zhaler et al. reported that the G4 formation inhibits telomerase activity, that is, the reverse transcriptase extending telomeric sequences at chromosome ends.^[27] Based on this, stabilizing G4 ligands have been considered as potent interesting anticancer agent that could inhibit telomere elongation resulting in inhibition of cancer progression.^[28–33]

From genome-wide sequences analyses, G4s have also frequently been identified in the promoting sequence, that is, 1 kb upstream the transcription start site of human genes.^[34–38] The majority of these studies were largely focused on cancer related genes such as the oncogenes *KRAS*, *c-MYC*, *c-RET*, *c-KIT*, *hTERT*, *BCL2* and *VEGF*. The formation of G4 in gene promoters has been largely associated with an inhibition of the transcription machinery. Indeed, G4 were initially characterized as obstacles for DNA transcription; later on, they have been associated with an increase of DNA instability, which also leads to inhibition of gene expression. Thus, G4 ligands and stabilizers have been designed to favor a down-regulation of oncogenes frequently overexpressed in cancers.

However, some results have also pointed out a positive regulation of G4 on gene expression bringing new evidence on the complex regulation of these structures depending on their localization and depending on the nature of G4-associated

proteins. Indeed, it has been postulated that G4s can be a high-affinity binding site for some transcription factor such as Sp1 and MAZ (Myc-associated zinc-finger protein)^[39] thus favoring transcription. Moreover, the localization of G4 structure in promoting sequences in front or behind the transcription start site or in the sense or antisense strand can influence their biological role as already discussed by Rigo et al.^[40] Besides telomeres and gene promoters regions, ribosomal DNA,^[41] 5'- and 3'-untranslated regions (UTR) of mRNA,^[42,43] TERRA,^[44] as well as some tRNA,^[45] rRNA,^[46,47] and other noncoding RNA (lncRNA, pre-miRNA, miRNA, piRNA)^[48] also contain G4s.^[49] RNA G4s, which mainly adopt compact parallel arrangements, are implicated in the regulation of translation, RNA metabolism and RNA splicing. For example, G4 in 5'-UTR mRNA mechanically impede the translational machinery and particularly translocation in ribosomes, hence slowing down translational efficiency.^[50] In 3'-UTR, G4s modulate RNA transport as well as its polyadenylation thus affecting RNA stability. Interestingly, RNA G4s have also been identified in retroviral RNAs where they favor RNA synthesis.

The stabilization of G4s with specific ligands represents an interesting therapeutic strategy.^[22,51] G4 ligands are typically planar and aromatic small molecules, including metal complexes, which often lack selectivity for a specific G4 or even for G4s versus B-DNA.^[32,33,52]

On the other hands, G4s develop intricate and very specific interaction patterns with other biological partners, including proteins and peptides. The specificity and the strength of these interactions, will in turn depend on a complex interplay between the recognition of either sequence, structural, or molecular shape.^[53,54] Indeed, a large part of the G4 biological activity results by their modulation of the interaction pattern between DNA and proteins, including telomerase, transcription factors and helicases. Furthermore, a valid therapeutic approach could consist in targeting G4 with protein-based compounds, which are surely more selective than small molecules. Such strategy, however suffers from different drawbacks, including the cost related to the production of intricate structures which also suffer of scarce cellular uptake and/or degradation by proteases, omnipresent in every cell compartment. Still, the interaction between G4s and proteins is an important area of research which deserves attention as witnessed from a recent excellent review.^[55] Because of these intrinsic limitations, understanding the recognition of particular G4s by specific peptides could be invaluable from a pharmacological or biotechnological point of view. Indeed, such specificity might lead on the one side to the development of highly sensitive G4 ligands, which could potentially be used as therapeutics but also for non-invasive diagnostic purposes as biomarkers for specific diseases.^[56] Specific peptides inhibiting G4s action by developing stable and selective interaction with the noncanonical nucleic acids could have a great therapeutic potential, especially in epigenetic-driven anticancer or antiviral therapies. In the following, we will first briefly review some of the most important studies dealing with DNA and RNA G4s interacting with small peptides, highlighting both the current limitations and the potentialities of the current approaches. Thus, we will

then propose some perspective on how to enhance the field, and in particular on how to combine high-level computational and experimental techniques, to go towards the rational design of peptides to finely tune and control the functions of G4 in cells and viruses, and their complex interplay.

Artificial and natural peptides interacting specifically with the *c*-MYC promoter

TH3 (Figure 2) is a thiazole peptide, structurally close to distamycin A, designed and synthesized by Dutta et al.^[57] This peptide binds preferentially to the G4 of the *c*-MYC promoter: two residues bind with the G4, one at the 5'-end and the other at the 3'-end. In addition to the FRET and NMR experiments that led to these conclusions, the authors further investigated the influence of the peptide on gene expression. To do so, they used three techniques: qRT-PCR, western blot and double luciferase assay. The result is that the TH3 peptide acts as a negative regulator of *c*-MYC expression, stabilizing the G4 of its promoter. Then, the use of flow cytometry and cell imaging allowed not only the localization of the peptide in the nucleus of the cells to be highlighted, but also its capacity to stop the proliferation phase of the cells and cell death.^[57]

Another artificial peptide, α,ϵ -poly-L-lysine (α,ϵ -PLL; Figure 3) which is a derivative of the natural ϵ -PLL peptide produced by the bacterium *Bacillus subtilis*, has been studied by Oliviero and collaborators.^[58] The authors have shown its interaction with both telomeric G4 DNA (*h*-Telo) and the *c*-MYC promoter, using surface plasmon resonance (SPR). While SPR shows no interaction specificity, ECD, UV-vis absorption, and fluorescence spectroscopy confirm that α,ϵ -PLL binds preferentially with the *c*-MYC G4. The binding kinetics inferred by ECD show that the interaction of α,ϵ -PLL with *c*-MYC evolve during 96 hours after mixing before reaching a stabilization. UV spectrometry shows a bathochromic effect, suggesting that the

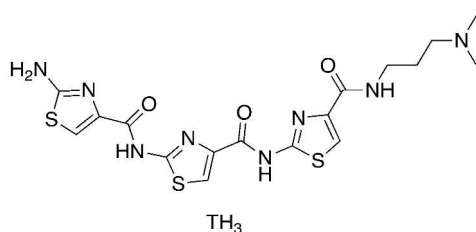


Figure 2. Chemical structure of the TH3 peptide.^[57]

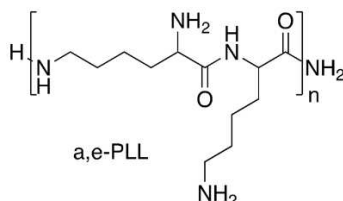


Figure 3. Chemical structure of the peptide α,ϵ -PLL.^[45]

peptide stacks between the quartets of the G4. Interestingly, HPLC size exclusion chromatography profile shows a high population of monomeric G4 and a low population of dimeric G4 in *c*-MYC solution, while the addition of the peptide does not significantly alter this equilibrium.^[45]

The KR12C peptide, of sequence Nter-KRIVKLIKKKWLR-Cter, has been designed by Sengupta et al.^[59] from the binding part of the human cathelicidin peptide LL37. The use of a multidisciplinary approach combining experimental measurements and molecular dynamics simulations has highlighted the specificity of the interaction between the peptide and the *c*-MYC promoter. The presence of the A6 and A15 nucleotides stacked above the first quartet forms a groove at the 5' terminus, which is specifically recognized by the peptide, developing a network of electrostatic interactions. Furthermore, the authors show that the peptide is able to stabilize the G4 of *c*-MYC in cellulose, further inducing a cascade involving E2F-1 and VEGF-A, which culminates in the reduction of *BCL*-2 expression accompanied by apoptosis.^[59]

Interestingly, Sengupta et al.,^[59] coherently with Minard et al.,^[60] also showed that the helicity of the peptide is important for the interaction with G4. Indeed, in solution the DM039 peptide (Ac-PGHLKGREIGLWYAKKQGQKNK-NH₂) is not structured but forms an α -helix when interacting with the *c*-MYC promoter. Based on this assumption, the authors chemically modified the peptide to obtain two variants whose α -helix structure is forced by hydrocarbon constrain. Unfortunately, both variants have shown a reduced affinity for the G4 of *c*-MYC, suggesting that a too sharp increase of the rigidity is detrimental to the selectivity of the interaction.^[60]

Finally, Scholz et al. demonstrated that DARPin, very small artificial proteins, are able to interact with G4s. al.^[61] The interaction can be non-specific, as in the case of DARPin 2G10, or specific, as in the case of DARPin 2E4, which recognizes the G4 of the *c*-MYC promoter. While ribosome display, ELISA tests, ECD spectroscopy and SPR have confirmed the interaction potentiality of DARPins the specific mode of interaction is not elucidated.^[61] Therefore, to establish whether the recognition is dictated by sequence or structure of G-quadruplexes, Miclot et al. used molecular dynamics simulations.^[53] By analyzing the contact frequency of the residues, they showed that DARPin 2E4 recognizes a specific motif formed by the G-quadruplex loops of the *c*-MYC promoter. This motif is composed by 1) a U shaped-loop with an extruded nucleotide, 2) a linear extension and 3) an appendage (Figure 4).

Artificially and naturally derived peptides interacting specifically with telomeric G4

Small polycyclic aromatic chromophores as acridine molecules are known for their G4 stabilization capacity, due to their high propensity of making π -interaction with guanines. However, those features characterize them also as B-DNA intercalators. Balasubramanian and collaborators aimed to increase the G4 specificity, targeting specifically loops and grooves over double-helical DNA.^[62] They synthesized several symmetrical 3,6-bis-

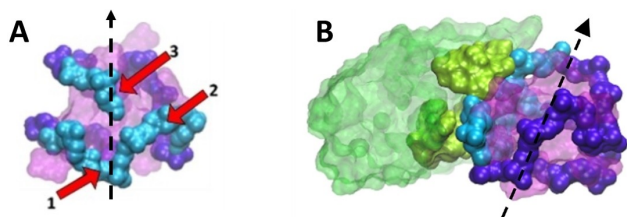


Figure 4. A) The structural motif recognized by the DARPin 2E4 is composed of 1) a U shaped-loop with an extruded nucleotide, 2) a linear extension, and 3) an appendage. B) Positioning of the G4 structure, in cyan, recognized by DARPin. The black dotted arrow is used to give the orientation of the G4 on the two images. Reproduced with permission from Ref. [41]. Copyright: 2022, Wiley-VCH.

peptide acridine and acridone conjugates (Figure 5). The capacity of those new hybrid molecules to selectively target the telomeric G4 versus duplex DNA was then assessed. The binding studies in solution demonstrated that all the new molecules bind to the nucleic acids stronger than the peptide-free analogues. The presence of the peptide also improved the specificity towards telomeric G4 of about tenfold, the selectivity of the binding being, however, peptide's sequence dependent. Thus, the addition of the two hanging peptides resulted in an effective way to convert a B-DNA intercalator in to a G4-binder, through the interaction of the peptide with the negatively charged pockets formed by the G4 grooves.^[62]

Using the same principle, the same research group synthesized trisubstituted acridine-peptide conjugates to provide opportunities for binding to DNA G4-forming sequences in the telomere but also in proto-oncogenes with low nanomolar affinities.^[63] Combining organic synthesis, biophysical techniques and molecular modeling, the authors investigated the binding affinities and the complex structures of acridine-based compounds exhibiting various peptide sequences. They showed that while the acridine moiety could be involved in π -stacking with the G4 tetrads, the peptide side chains could interact within the G4 grooves and loops, providing both quadruplex- and sequence specificities. Interestingly, they reveal the possibility to discriminate between different G4 types by tuning the peptide sequence and the site of modification on the acridine moiety of the binding compound. These results are of great importance for improving the selectivity of G4-targeting molecules.

With the aim to design G4 ligands with binding specificity for quadruplex over duplex DNA, the research group of

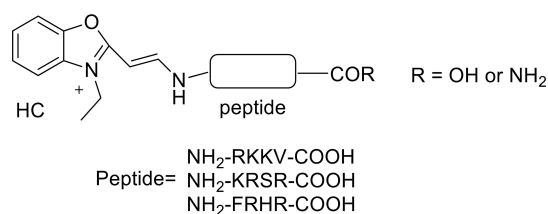


Figure 5. Chemical structure of the peptide-hemicyanine conjugates synthesized by Balasubramanian and collaborators.^[64]

Balasubramanian synthesized a peptide-hemicyanine conjugate, in which the presence of the small heterocyclic molecule hemicyanine (HC, Figure 5) should confer the ligand high binding affinity and high quadruplex/duplex discrimination.^[64] Three tetrapeptide moieties were selected and linked to HC: HC-RKKV, HC-KRSR and HC-FRHR, with both COOH and CONH₂ terminal groups. By using SPR and immobilized DNA, the authors showed that, although the *h-Telo* G4 binding was quite weak, on the micromolar range, nevertheless it was sensibly higher than duplex DNA binding, of more than 40 times in the case of HC-FRHR-CONH₂ conjugate.

The molecular details of quadruplex recognition by these heterocycle-peptide conjugates are still not fully understood. However, a loop-binding mode and groove interactions were proposed for the peptide-hemicyanine conjugates, without significant π -stacking with the DNA bases.^[64]

By employing a series of analytic techniques, including gel mobility shift assay, isothermal titration calorimetry, fluorescence spectroscopy and DNA-thermal denaturation experiments followed by ECD, Biswas et al.^[65] showed that a synthetic dendritic peptide, Cd2-(YEE)-E (Figure 6), specifically targets *h-Telo* G4. Interestingly, the same dendritic peptide inhibits the growth of HeLa and U2OS human cancer cells lines, inducing apoptosis in vitro. No significant inhibitory effect was noticed on the growth of normal human embryonic kidney cells. These experiments led to the conclusion that this peptide, upon targeting the telomeric G4 DNA, is an inhibitor of tumor cell growth by inducing apoptosis.^[65]

In order to improve the selectivity and specificity of the peptide ligands for *h-Telo* G4, Tyagi et al. designed the QW10 peptide, with sequence QQWQQQWQQ. QW10 was chosen to incorporate glutamine amino acid for its capacity as hydrogen bond donor and acceptor with guanine, and tryptophan to provide π -stacking with the quartet in an intercalative mode.^[66] No positively charged residues were incorporated to prevent non-specific binding to duplex DNA. Molecular docking studies demonstrated that this peptide is able to effectively intercalate between the G4 tetrads through tryptophan residues and develops hydrogen bonds through glutamine residues. Moreover, QW10 has been shown to bind and specifically stabilize *h-Telo* G4 only in its hybrid conformation, that is, in the presence of K⁺, and not in antiparallel organization (obtained in the

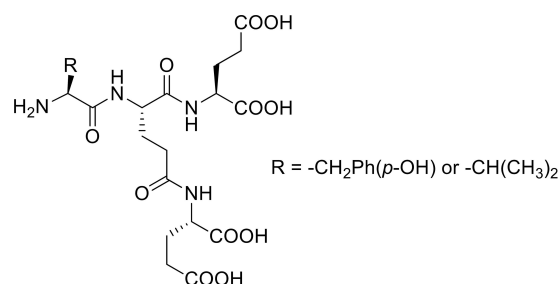


Figure 6. Chemical structure of the dendritic peptide synthesized by Biswas et al.^[65]

Table 1. Sequences of peptides derived from RNA helicase associated with AU-rich element (RHAU) protein. The residues used for cyclization or stapling are in red.

Peptide	Sequence
<i>Rhau18</i>	SMHPGHLKGREIGMWYAKKQ
<i>Rhau23</i>	GLHPGHLKGREIGMWYAKKQGQKNKNGL
<i>Rhau23c</i>	GLHPGHLKGREIGMWYAKKQGQKNKN
<i>Rhau</i> -LSAH1	HPGHLKGE ^E IG ^K WYAKKQGQKNK
<i>Rhau</i> -LSAH2	HPGHLKGR ^E IGM ^K YAKKQGQKNK

presence of Na⁺ based buffer), thus implying a structure-dependent binding mode.^[66]

Bloom syndrome (BLM) protein is a helicase able to unwind human telomeric G4. As a matter of fact, BLM possesses a helicase-and-RNase-D C-Terminal domain, named HRDC, able to bind single stranded DNA. Sharma and collaborators performed a molecular docking screening between several small peptides, derived from the HDRC domain of BLM, and polymorphic *h-Telo* G4.^[67] They identified the peptide HDRC (N-VAQKW-C) as biologically active and able to recognize the telomeric G4. Molecular simulations confirmed the ability of HDRC to recognize the G4 loops and intercalate through the tryptophane residues between the G-quartet, leading to a partial loss of stacking interactions between the quartets. Thus, the authors anticipated that the binding of HRDC to G4 might provoke the partial unfolding of the nucleic acid, as confirmed by all the further biophysical and biochemical studies. Finally, the treatment with HDRC appeared to inhibit the growth of MDA-MD-231 human breast cancer cells confirming the good potentiality of this kind of systems for therapeutic applications.^[67]

In 2013 Jana et al. considered the LL37 cathelicidin peptide (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLPRTES), naturally produced by humans against external and pathological attacks, and its capacity as a potent G4 binder. They showed that LL37 stably and selectively binds to G4, leading to an increase stability of the latter.^[68] LL37 has been shown to possess a α -helical motif, which is lost upon interaction with G4. Molecular dynamic studies confirmed those experimental results showing conformational fluctuation of LL37 all along the simulation, which ultimately leads to the loss of its helical conformation to adopt, instead, a circular organization over the G4 quartet.^[68] Later, the same group focused on the smallest FK13 region of LL37 identified as having antimicrobial activity to demonstrate its ability to inhibit telomerase. Indeed, FK13 (FKRIVQRIKDFLR) has been shown to bind to *h-Telo* G4, and thus provide additional structural stability. A high specificity towards G4 against double-strand DNA has also been observed.^[69] NMR spectroscopy has highlighted the interaction of the R7, R3, and R13 with the phosphate backbone of the G4, while MD simulations, coherently with the previous observation on the entire LL37 peptide, have proven that FK13 loses its helicity while interacting with the G4. On the other hand, FK13 treatment on MCF7 breast cancer cells leads to a significant inhibition of their growth, in a dose-dependent manner. Further experiments have characterized the growth inhibition as linked with the blockage of cell cycle in the subG0 phase, and early and late apoptosis. Finally, Jana et al. also found that the

treatment with FK13 is linked to a decrease in the telomerase activity, in particular caused by a lower mRNA level of hTERT, that is, the catalytic subunit of the telomerase enzyme. In conclusion, it was proposed that this short peptide can provoke cancer cell death through the inhibition of the telomerase activity as well as the binding and stabilization of the telomeric G4.^[69]

Peptides derived from RNA helicase associated with AU-rich element (RHAU) protein for specific G4 recognition

RNA helicase associated with AU-rich element (also known as RHAU) is a protein of the DEAH box family, which has been shown to resolve G4 structures with high specificity.^[70] In a recent review, Nguyen and Dang highlight the potential applications of different RHAU peptides which specifically recognize G4s.^[71]

Phan and collaborators reported the NMR solution structure of a G4 interacting with a peptide from the N-terminal region of RHAU,^[72] The portion of RHAU protein binding to G4s is a 53-residue fragment (from amino acid 53 to 105) with a 13-residue RHAU-specific motif (RSM; from amino acid 54 to 66) essential for the binding.^[73]

Through gel electrophoresis and NMR experiments, the authors demonstrated that the long peptide comprising the afore-mentioned protein portion of 53-residue fragment (named *Rhau53*) is selective for parallel G4s, with no binding to other DNA or RNA motifs, including other G4s and duplexes. In the attempt to find the shortest peptide able to still recognize G4s, binding assays were performed between a generic parallel G4 model (T95-2T) and peptides of different lengths. The 16- and 18-amino acid peptides termed *Rhau16* and *Rhau18*, respectively, containing the RSM motif, were the ones giving the best results (Table 1). The NMR solution structure (Figure 7) confirms that binding of *Rhau18* to parallel G4s occurs through stacking interactions on the planar G-tetrad surface and electrostatic interactions with the phosphate backbone of the peptide positively charged side chains. In particular, *Rhau18* forms an α -helix between residues R10 and A17, with a pronounced kink of the peptide loop before residue G9. The amino acids G7, I10, G11 and A15 are the ones involved in direct interaction with the G-tetrad face.^[72]

Even if *Rhau18* represents an excellent candidate to be used as a G4-specific ligand, since it is a linear peptide is, in principle, susceptible to peptidase digestion and hence not suitable for use in cells or in vivo. Trying to increase the stability of *Rhau18*,

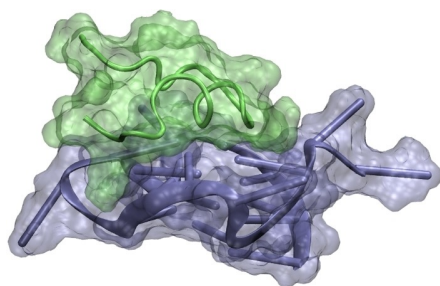


Figure 7. Representation of the NMR solution structure of the *Rhau18*-T95-2T complex (PDB ID: 2N21).^[72] The G4 is represented in blue and the peptide in green.

Phan et al. used asparaginyl endopeptidase from *Oldenlandia affinis* (OaAEP1) to induce an end-to-end joining of their peptide and obtain a cyclic variant of *Rhau18*, named *Rhau23c* (Figure 8).^[74] This was possible adding GL residues at the N terminus and NGL at the C terminus of *Rhau18* (to produce linear *Rhau23*) to make OaAEP1 work. ECD analysis revealed that the α -helix, essential for G4 binding, was not lost upon cyclization of the peptide. Remarkably, binding assays showed that *Rhau23c* binds to T95-2T at approximately 5 times higher affinity than the linear counterpart. Furthermore, *Rhau23c* was way more resistant to digestion by carboxypeptidase A, an exopeptidase that cleaves peptides from the C terminus, making this cyclic peptide an ideal candidate for therapeutic applications.^[74]

Another attempt to stabilize *Rhau18* against protease digestion was tried by the same authors using a stapling strategy, that is, connecting helical twists with covalent linkages.^[75] In particular, two stapled *Rhau* peptides, *Rhau*-LSAH1 and *Rhau*-LSAH2 involving different stapling sites, were produced and analyzed (Table 1). In both cases, the α -helix was preserved as confirmed by ECD experiments and, as for *Rhau23c*, the *Rhau*-LSAH1 and *Rhau*-LSAH2 showed enhanced stability against digestion by proteases. Tan and collaborators took advantage of the excellent G4 binding ability of *Rhau23* to formulate a peptide-PNA (peptide nucleic acid) conjugate with the aim to bind selectively G-triplex (G3), DNA structures which form in guanine-rich sequences but are generally much less stable than G4s.^[76] The peptide-PNA conjugate consists of a PNA G3-tract connected to *Rhau23* peptide. The authors demonstrated that the 3G PNA moiety (Figure 9) can bind to a

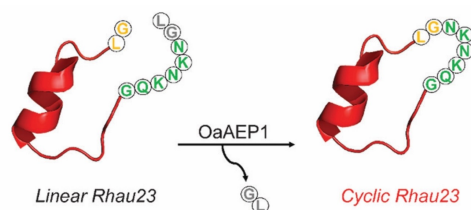


Figure 8. Schematic representation of *Rhau23* cyclization. Reproduced with permission from Ref. [74]. Copyright: 2020, The Royal Society of Chemistry.

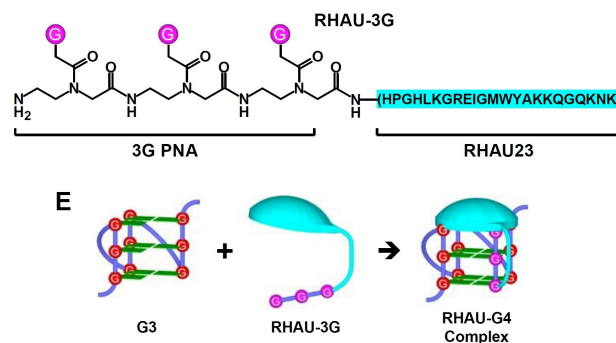


Figure 9. Structure of the 3G PNA conjugated to *Rhau23* peptide and schematic representation of its interaction with a G-triplex to give a G4 further stabilized by the peptide. Reproduced with permission from Ref. [76]. Copyright: 2019, The Royal Society of Chemistry.

G3 to form a bimolecular (3 + 1) G4 which is simultaneously stabilized by the action of *Rhau23*. Remarkably, affinities of low-nanomolar to sub-nanomolar towards G3 targets were achieved while no binding was observed for other DNA structures.^[76]

PNA oligomers have been developed in the early 90's^[77,78] yet their high interest in targeting G4s was revealed a decade later. They have shown that complementary PNA can invade G4 structures and provoke G4 to duplex conformational transitions in both DNA and RNA,^[79] offering interesting opportunities for designing new therapeutic agents. Interestingly, they also showed how tuning the PNA sequence from cysteine-rich to G-rich can lead to hybridization to heteroquadruplexes rather than heteroduplexes with DNA.^[80–84]

Cyclic peptides

As already mentioned, another way to avoid digestion by peptidases is to use synthetic unnatural cyclic peptides. Balasubramanian and co-workers developed a series of chiral cyclopeptides (Figure 10) able to bind G4s by top stacking thanks to their almost planar structures.^[85] The synthesized macrocycles are conveniently functionalized with alkylamines to increase water solubility and provide extra interaction with the DNA motif loops.

The authors, using SPR and FRET melting analysis evidenced that peptides **a–d** are selective for G4s motifs when compared

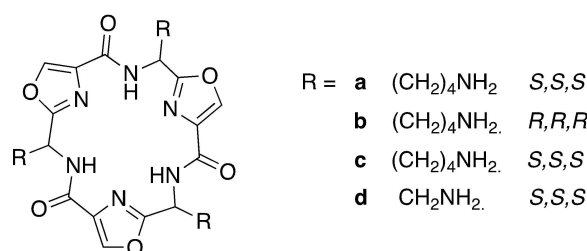


Figure 10. Structure of cyclopeptides developed by Balasubramanian and collaborators.^[85]

to double stranded DNA sequence and that, among G4s, they prefer to bind the one of the c-KIT sequence (parallel) rather than the telomeric counterpart *h-Telo* (hybrid).

Interestingly, enantiomers **a** and **b** were found to bind G4s with comparable strength, while a drop in binding affinity was observed for diastereoisomer **c**. The authors explained the results considering that **a** and **b** cyclopeptides bear three butylamine side chains just on one of their faces, leaving the other free for π - π stacking with the top tetrad of the quadruplex, whereas **c**, where there is the inversion of one stereocenter, presents side chains on both faces of the macrocycle, in this way hampering the top stacking interaction mechanism.^[85]

Proto-oncogenes have also been used as targets to develop (bi)cyclic-peptides showing selectivity to G4-forming sequences or the capacity to stabilize G4 structures. Inspired from the mode of recognition of G4 by the DHX36 helicase, Liu et al. characterized G4-specific bicyclic peptides selected by phage display.^[86] Their compounds showed selectivity towards G4 versus duplex DNA, with sub micromolar affinities, fostered by important hydrophobic interactions as revealed by molecular modeling. Selected from a 10^8 – 10^9 compounds library, bicyclic peptides of ACX_nCX_nCG sequence (X = any natural amino acids and $n = 3$ or 4) exhibited the capacity to mimic the protein structural preorganization driving G4 recognition allowing target specificity and high affinity to be achieved.

Cyclic peptides can also be used to form G4-bioconjugates, providing information on the quadruplex structure and a stable scaffold for ligand binding assays. Bonnat et al. used this technique to isolate and stabilize biologically relevant G4s, such as in the human immunodeficiency virus-1 (HIV-1) long terminal repeat promoter region.^[87] Using a cyclic peptide scaffold to attach specific oligomers, they could monitor the assembly of G4 in G-rich DNA oligonucleotides from HIV-1 and characterize its structure for further testing of different G4-binding ligands. G4 structures were validated by biophysical techniques (UV thermal difference studies, ECD and NMR), and binding analysis were performed by SPR, highlighting the high affinity of three different ligands (acridine and naphthalene derivatives) for the HIV-G4 structure.

Arginine rich peptides

The analysis above shows that peptides have some advantages over proteins and antibodies, due to their small size, ease of synthesis and purification, but also cell permeability, tumor-penetrating ability and improved biocompatibility.^[65]

Several G4 binding proteins, containing an arginine-glycine-rich sequence, known as RGG peptide, specifically control the biological functions of this kind of DNA motif. For this reason, the interaction between synthetic oligopeptides rich in RGG motifs with G4s is a very attracting strategy to reveal atomistic details of the G4-peptide binding.

In a recent investigation, Patra et al.^[88] performed single molecule experiments at micromolar concentrations, by using a zero-mode waveguide confocal microscope, combined with

dual color fluorescence cross-correlation spectroscopy (FCCS) and FRET. This approach allowed them to follow the binding dynamics, and to determine the association and dissociation rate constants of a 20-mer RGG peptide and six G4 forming sequences. The results obtained show that the binding dynamics occurs at sub-millisecond timescale and that high micromolar dissociation constants govern the G4-RGG complex, implying the occurrence of a medium to low DNA-peptide affinity, as expected. A large part of the interaction is driven by electrostatics. However, a specific affinity between G4s and the RGG-rich peptide is present. The latter might be the consequence of H-bonds established between arginine and the guanine bases of the G4s.^[88]

As the RGG domain contributes greatly to the G4 binding affinity, it is often present in G-quadruplex-binding proteins. To clarify the role of the RGG motif in G-quadruplex-protein interactions, the binding affinity and mode between RGG-motif peptides and G4s was systematically investigated.^[89] This study allowed the identification of the cold-inducible RNA-binding protein (CIRBP) as a new G4-binding protein both in vitro and in cellulo. It was found that the peptide RGSAGGRGFFRGGGRGRGRGFSRGG, containing seven RGG units, played a key role in the G-quadruplex binding of CIRBP.

Peptides binding to RNA G4s

Peptides can also bind RNA quadruplexes that recently emerged as excellent target both in human and non-human genomes.^[55]

C9orf72 (C9) is the gene responsible for the familial form of amyotrophic lateral sclerosis (ALS) and related frontotemporal dementia (FTD). This familial form is due to a hexanucleotide repeat expansion (HRE) of (GGGGCC)_n, inclined to form a G4 structure.^[90] It was shown that increased Pur-alpha (Pura) protein levels in animals alleviate certain cellular symptoms of ALS/FTD. It might be possible that Pura, by binding to the C9 repeat sequence of RNA, could negatively influence translation, thereby decreasing DPR production and helping to ameliorate neurodegeneration. A study conducted by Wortman et al. showed activities of Pur-like domain protein motifs against ALS or FTD and helps to optimize the design of peptide-based protective agents against this pathology.^[91] In particular, the authors synthesized and characterized a novel peptide (TZIP) of 89 amino acid residues, analogous to a Pur domain, with the aim to mimic and improve not only the binding of Pur proteins with the C9 G-rich hexanucleotide repeat sequence but also to favor cellular internalization. As a matter of fact, TZIP binds the C9 RNA and DNA hexanucleotide repeats with high affinities. Furthermore, it may bind both the G4 and the linear form of CC(GGGGCC)₄ RNA sequence with similar affinities. It also mediates quadruplex unwinding and secondary structural transitions among the RNA forms.^[91]

The loss of fragile X mental retardation protein (FMRP) function causes the fragile X mental retardation syndrome.^[92] FMRP harbors three RNA binding domains and is thought to regulate mRNA translation and/or localization; however, its

specific RNA targets are unknown. With the aim of better understanding the recognition patterns between the binding domain of TMPRSS and G-rich RNAs, Phan et al. determined the structure of the complex between an arginine-glycine-rich RGG peptide derived from FMRP and an in vitro-selected guanine-rich (G-rich) RNA (Figure 11). They found that arginine residue is critical in the recognition of the RNA duplex-quadruplex junction. The RGG peptide is positioned along the major groove of the RNA duplex, with the G4 forcing a sharp turn at the duplex-quadruplex junction.^[93]

Recently, the role of G4 in RNA viruses, including SARS-CoV-2, has attracted a lot of interest. For instance, it has been demonstrated that the formation of G4 structure in the RG-1 sequence of SARS-CoV-2 genome reduces the levels of the structural N-protein by inhibiting its translation.^[94] Furthermore, the precise structural details of the RG-1 arrangement have also been obtained by molecular modeling and simulations.^[95] In a recent study, Mukherjee et al. investigated the conformational dynamics of the RG-1 G4 affected by the various external conditions such as salt concentration, co-solutes, crowders and intrinsically disordered peptides (IDPs).^[96] Particularly they focused on α -synuclein and the human islet amyloid polypeptide, which are involved in Parkinson's disease (PD) and type-II diabetes mellitus, respectively. These IDPs could interact with RG-1 motifs through noncovalent interactions and might affect the formation of RG1 G4, thus affecting both translation and the virus cycle in the host cell. The authors have shown that hIAPP stabilizes a partially folded state of the RG-1 RNA-G4, whereas α -Syn strongly stabilizes the fully folded state of the quadruplex. The different stabilization could be explained by the different amino acid composition of the two peptides. α -Syn contains mainly lysine and arginine residues that can interact with the negatively charged phosphate groups of the nucleic acid backbone. On the contrary, hIAPP contains asparagine and glutamine residues, which can form hydrogen bonds with adenine, decreasing the stability of the G4^[93,97] and overall interaction strength.

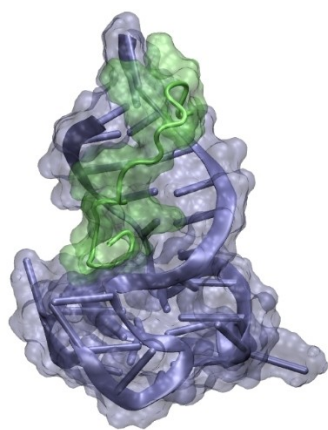


Figure 11. RNA duplex-quadruplex junction complex with FMRP RGG peptide. PDB ID: 2LA5.^[88]

To date, no antiviral drug or vaccine is available for infection caused by West Nile virus. Interestingly, potential guanine quadruplex forming sequences have been identified.^[98]

In 2016, the Armitage group reported γ -modified peptide nucleic acid (γ PNA) oligomers exhibiting potency to invade RNA G4 and inhibit translation of a luciferase reporter transcript harboring G4 structures.^[99] Indeed, functionalizing the γ -position of the N-(2-aminoethyl)glycine unit onto the PNA backbone allows the preorganization of the PNA fold, as reported by Ly et al.^[100,101] Taking advantage of this interesting feature, Armitage's group most recently designed γ PNA with low femtomolar affinities towards quadruplex-forming sequences from the West Nile virus genome.^[102] Using biophysical techniques, they studied how the PNA length, salt concentration, and γ -modifications can impact the selectivity towards G4s, also highlighting the higher potency of γ PNA against variants (e.g., deletions and mismatches in the target sequence) compared to unmodified PNA. Considering the crucial importance of selectivity for avoiding off-target effects, such these γ PNA molecules open incredible opportunities towards the design of high potential antivirals targeting viral replication and transcription.

Conclusions

With the examples examined above, we have highlighted the chemical and biological significance of G4s and their widespread diffusion. Furthermore, we have also shown how complex, yet specific, interaction patterns allow the development of precise interactions between G4 and peptides. Importantly, such interactions can be observed for different G4 types, including both DNA and RNA structures, and are in some instances correlated with the G4 recognition corresponding with an influence on its biological role. Although our review is clearly not exhaustive and it is not intended to provide a comprehensive state-of-the-art of present-day research, we have nonetheless presented crucial aspects of the interaction between G4 and peptides, underscoring the crucial factors leading to their selective recognition. If a full identification of the structural features of the interacting peptides appears difficult, also due to the conformational variability of G4s, we can nonetheless identify some susceptible hot-spots to enhance the interaction. Indeed, as expected, positively charged peptides will be suitable for developing favorable interactions with the negatively charged backbone of the nucleic acid. However, the presence of hydrophobic, and aromatic moieties, might favor π -stacking involving the exposed tetrads and might be most favorable to induce the stabilization of the G4. In addition, the presence of the lateral loops, which may form specific pockets might lead to interesting druggable spots; this could increase the selectivity of the peptide. In particular, a shape-based recognition, mimicking the epitope/paratope interactions in antibodies/antigens, as experienced by DARPins,^[53] seems a most promising opportunity.

A vast panel of biophysical and spectroscopic techniques exists to unravel the peculiar structures of G4/peptide adducts,

as well as the influence played by environmental factors on modulating their persistence and stability. The advances in biological structural methods, including not only X-ray crystallography or NMR, but also cryo-electron microscopy, are instrumental to providing high-resolution structures of peptide/G4 adducts. Yet, the former methods are usually time-consuming and expensive and can hardly be used routinely. On the other hand, while more common spectroscopic techniques, such as ECD, are readily available and highly sensitive to secondary structure arrangements, they cannot provide a full description at the atomistic level. Lacking such resolution, and more importantly, the perturbation brought by the interaction with the peptides or the recognition hot spots, is clearly a burden to attempts towards the rational design of specific peptide binders. Yet, such systems could be most welcome both for therapeutic reasons or as constituents of sensors allowing a rapid, and noninvasive, recognition of G4s. A selective interaction with specific G4 arrangements or sequences is clearly the most desirable feature for both approaches.

Supported by the examples chosen in this review, we believe that the combination of experimental techniques with multiscale molecular modeling and simulation could represent a valuable alternative that provides an atomistic structural resolution, one-to-one mapping with experiments, and finally the possibility to design specific binders. Of course, computational modeling should also rely on the use of sophisticated force fields, which provide a good description of noncanonical DNA and RNA structures and of their interactions with flexible peptides. While a large amount of work has been realized in the past year on this aspect, and further improvements are supposed to be observed thanks to machine learning approaches, the reliability, reproducibility, and validation of force fields is still an issue. Furthermore, the complex and rugged conformational landscape of G4s and peptides also calls for the use of appropriate enhanced sampling techniques and free-energy methods. In this respect, the reduction of dimensionality to access reasonable collective variables should also be taken into proper account. Finally, the use of quantum or hybrid methods to describe electronic excited states in complex environments, while giving access to ECD spectra (when properly combined with MD sampling) is still challenging, especially for multichromophoric systems like G4s. Constant dialogue between experiment and theory, and complementary methodological developments will push the accuracy of such protocols even further and will strengthen their accuracy and predictive capability.

In addition to the development of combined experimental and modeling protocols, we also believe that one of the most important issues at the forefront of G4 studies, and more specifically noncanonical nucleic acid structures, will be the precise exploration and understanding of the coupling between multiple secondary structures. In this sense, elucidating how the chemical and biological properties of multiple G4 aggregates change compared to their monomeric counterpart will be a fundamental question that needs to be answered too. In the same spirit, understanding how the presence of multiple G4s shapes interactions with peptides and proteins will allow not

only a specific G4, but also functional assemblies of secondary structures to be targeted to finely control a precise biological function. Once again, the role of molecular modeling and simulation when coupled with the proper chemical and biological assays will be crucial.

G4s are fascinating per se and can develop an intricate interaction network with other nucleic acids and with peptides and proteins. Chemical and structural biology, with the assistance of simulation, are nowadays ripe to unravel these complicated patterns, bringing together the promise of the possibility of a more efficient control of their function. Considering the signaling role exerted by G4s, this will clearly translate to the possibility of fine epigenetic control of the gene expression. The challenges are many, but the promises brought by the comprehension of the intricate G4 world already let us savor the premises of a bright future.

Acknowledgements

A.M. thanks ANR and CGI for their financial support of this work through Labex SEAM ANR 11 LABX 086, ANR 11 IDEX 05 02. The support of the IdEx "Université Paris 2019" ANR-18-IDEX-0001 and of the Platform P3MB is gratefully acknowledged.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: DNA · guanine quadruplexes · molecular modeling · peptides · RNA · spectroscopy

- [1] N. Saini, Y. Zhang, K. Usdin, K. S. Lobachev, *Biochimie* **2013**, *95*, 117–123.
- [2] H. Tateishi-Karimata, N. Sugimoto, *Chem. Commun.* **2020**, *56*, 2379–2390.
- [3] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474.
- [4] C. Fonseca Guerra, T. van der Wijst, J. Poater, M. Swart, F. M. Bickelhaupt, *Theor. Chem. Acc.* **2010**, *125*, 245–252.
- [5] J. Szolomájer, G. Paragi, G. Batta, C. F. Guerra, F. M. Bickelhaupt, Z. Kele, P. Pádár, Z. Kupihár, L. Kovács, *New J. Chem.* **2011**, *35*, 476–482.
- [6] C. Nieuwland, F. Zaccaria, C. Fonseca Guerra, *Phys. Chem. Chem. Phys.* **2020**, *22*, 21108–21118.
- [7] F. Zaccaria, G. Paragi, C. Fonseca Guerra, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20895–20904.
- [8] E. Lary, J. L. Mergny, V. Gabelica, *Met. Ions Life Sci.* **2016**, *16*, 203–258.
- [9] P. Hazel, J. Huppert, S. Balasubramanian, S. Neidle, *J. Am. Chem. Soc.* **2004**, *126*, 16405–16415.
- [10] J. Jana, Y. M. Vianney, N. Schröder, K. Weisz, *Nucleic Acids Res.* **2022**, *50*, 7161–7175.
- [11] J. Spiegel, S. Adhikari, S. Balasubramanian, *Trends Chem.* **2020**, *2*, 123–136.
- [12] A. Verdian Doghaei, M. R. Housaindokht, M. R. Bozorgmehr, *J. Theor. Biol.* **2015**, *364*, 103–112.
- [13] S. Matsumoto, H. Tateishi-Karimata, S. Takahashi, T. Ohyama, N. Sugimoto, *Biochemistry* **2020**, *59*, 2640–2649.
- [14] M. Vorlíčková, I. Kejnovská, K. Bednářová, D. Renčíuk, J. Kypr, *Chirality* **2012**, *24*, 691–698.
- [15] M. Vorlíčková, I. Kejnovská, J. Sagi, D. Renčíuk, K. Bednářová, J. Motlová, J. Kypr, *Methods* **2012**, *57*, 64–75.

- [16] J. Kyr, I. Kejnovská, D. Renčluk, M. Vorlič, *Nucleic Acids Res.* **2009**, *37*, 1713–1725.
- [17] H. Gattuso, A. Spinello, A. Terenzi, X. Assfeld, G. Barone, A. Monari, *J. Phys. Chem. B* **2016**, *120*, 3113–3121.
- [18] E. Bignon, T. Miclot, A. Terenzi, G. Barone, A. Monari, *Chem. Commun.* **2022**, *58*, 2176–2179.
- [19] E. Puig Lombardi, A. Londoño-Vallejo, *Nucleic Acids Res.* **2020**, *48*, 1–15.
- [20] H. B. Cagirici, H. Budak, T. Z. Sen, *BMC Bioinf.* **2022**, *23*, 10.1186/s12859-022-04782-z.
- [21] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.* **2020**, 10.1038/s41580-020-0236-x.
- [22] E. Ruggiero, S. N. Richter, *Annu. Rep. Med. Chem.* **2020**, *54*, 101–131.
- [23] M. Falabella, R. J. Fernandez, F. B. Johnson, B. A. Kaufman, *Curr. Med. Chem.* **2018**, *26*, 2918–2932.
- [24] T. M. Bryan, *Molecules* **2020**, *25*, 10.3390/molecules25163686.
- [25] A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
- [26] A. T. Phan, *FEBS J.* **2010**, *277*, 1107–1117.
- [27] A. M. Zahler, J. R. Williamson, T. R. Cech, D. M. Prescott, *Nature* **1991**, *350*, 718–720.
- [28] A. Terenzi, H. Gattuso, A. Spinello, B. K. Keppler, C. Chipot, F. Dehez, G. Barone, A. Monari, *Antioxidants* **2019**, *8*, 10.3390/antiox8100472.
- [29] T. M. Ou, Y. J. Lu, J. H. Tan, Z. S. Huang, K. Y. Wong, L. Q. Gu, *ChemMedChem* **2008**, *3*, 690–713.
- [30] A. De Cian, G. Cristofari, P. Reichenbach, E. De Lemos, D. Monchaud, M. P. Teulade-Fichou, K. Shin-ya, L. Lacroix, J. Lingner, J. L. Mergny, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17347–17352.
- [31] S. Neidle, *FEBS J.* **2010**, *277*, 1118–1125.
- [32] S. Neidle, *J. Med. Chem.* **2016**, *59*, 5987–6011.
- [33] S. Neidle, *Nat. Chem. Rev.* **2017**, *1*, 0041.
- [34] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
- [35] A. K. Todd, M. Johnston, S. Neidle, *Nucleic Acids Res.* **2005**, *33*, 2901–2907.
- [36] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2007**, *35*, 2105–2105.
- [37] J. H. Chariker, D. M. Miller, E. C. Rouchka, *PLoS One* **2016**, *11*, 10.1371/journal.pone.0165101.
- [38] S. Lago, M. Nadai, F. M. Cernilogar, M. Kazerani, H. Domínguez Moreno, G. Schotta, S. N. Richter, *Nat. Commun.* **2021**, *12*, 3885.
- [39] S. Cogoi, M. Paramasivam, A. Membrino, K. K. Yokoyama, L. E. Xodo, *J. Biol. Chem.* **2010**, *285*, 22003–22016.
- [40] R. Rigo, M. Palumbo, C. Sissi, *Biochim. Biophys. Acta Gen. Subj.* **2017**, *1867*, 1399–1413.
- [41] M. Wallgren, J. B. Mohammad, K. P. Yan, P. Pourbozorgi-Langroudi, M. Ebrahimi, N. Sabouri, *Nucleic Acids Res.* **2016**, *44*, 6213–6231.
- [42] S. Kumari, A. Bugaut, J. L. Huppert, S. Balasubramanian, *Nat. Chem. Biol.* **2007**, *3*, 218–221.
- [43] J. D. Beaudoin, J. P. Perreault, *Nucleic Acids Res.* **2013**, *41*, 5898–5911.
- [44] C. Da Xiao, T. Shibata, Y. Yamamoto, Y. Xu, *Chem. Commun.* **2018**, *54*, 3944–3946.
- [45] S. M. Lyons, D. Gudanis, S. M. Coyne, Z. Gdaniec, P. Ivanov, *Nat. Commun.* **2017**, *8*, 10.1038/s41467-017-01278-w.
- [46] K. Havlová, J. Fajkus, *Front. Plant Sci.* **2020**, *11*, 10.3389/fpls.2020.593692.
- [47] S. Mestre-Fos, C. Ito, C. M. Moore, A. R. Reddi, L. D. Williams, *J. Biol. Chem.* **2020**, *295*, 14855–14865.
- [48] S. Ghafouri-Fard, A. Abak, A. Baniahmad, B. M. Hussien, M. Taheri, E. Jamali, M. E. Dinger, *Cancer Cell Int.* **2022**, *22*, 10.1186/s12935-022-02601-2.
- [49] D. Varshney, S. M. Cuesta, B. Herdy, U. B. Abdullah, D. Tannahill, S. Balasubramanian, *Sci. Rep.* **2021**, *11*, 10.1038/s41598-021-01847-6.
- [50] T. Endoh, N. Sugimoto, *Molecules* **2019**, *24*, 10.3390/molecules24081613.
- [51] C. Jaubert, A. Bedrat, L. Bartolucci, C. Di Primo, M. Ventura, J. L. Mergny, S. Amrane, M. L. Andreola, *Sci. Rep.* **2018**, *8*, 8120.
- [52] T. Kench, R. Vilar, *Annu. Rep. Med. Chem.* **2020**, *54*, 485–515.
- [53] T. Miclot, E. Bignon, A. Terenzi, S. Grandemange, G. Barone, A. Monari, *Chem. Eur. J.* **2022**, 10.1002/chem.202201824.
- [54] C. C. Hognon, T. Miclot, C. G. Iriepa, A. Francés-Monerris, S. Grandemange, A. Terenzi, M. Marazzi, G. Barone, A. Monari, C. Garcia-iriepa, et al., *J. Phys. Chem. Lett.* **2020**, *11*, 5661–5667.
- [55] V. Meier-Stephenson, *Biophys. Rev. Lett.* **2022**, *14*, 635–654.
- [56] Y. Wu, R. M. Brosh, *FEBS J.* **2010**, *277*, 3470–3488.
- [57] D. Dutta, M. Debnath, D. Müller, R. Paul, T. Das, I. Bessi, H. Schwalbe, J. Dash, *Nucleic Acids Res.* **2018**, *46*, 5355–5365.
- [58] M. Marzano, A. P. Falanga, D. Marasco, N. Borbone, S. D'Errico, G. Piccialli, G. N. Roviello, G. Oliviero, *Mar. Drugs* **2020**, *18*, 10.3390/md18010049.
- [59] P. Sengupta, N. Banerjee, T. Roychowdhury, A. Dutta, S. Chattopadhyay, S. Chatterjee, *Nucleic Acids Res.* **2018**, *46*, 9932–9950.
- [60] A. Minard, D. Morgan, F. Raguseo, A. Di Porzio, D. Liano, A. G. Jamieson, M. Di Antonio, *Chem. Commun.* **2020**, *56*, 8940–8943.
- [61] O. Scholz, S. Hansen, A. Plückthun, *Nucleic Acids Res.* **2014**, *42*, 9182–9194.
- [62] S. Ladame, J. A. Schouten, J. Stuart, J. Roldan, S. Neidle, S. Balasubramanian, *Org. Biomol. Chem.* **2004**, *2*, 2925–2931.
- [63] J. E. Redman, J. M. Granadino-Roldán, J. A. Schouten, S. Ladame, A. P. Reszka, S. Neidle, S. Balasubramanian, *Org. Biomol. Chem.* **2009**, *7*, 76–84.
- [64] J. A. Schouten, S. Ladame, S. J. Mason, M. A. Cooper, S. Balasubramanian, *J. Am. Chem. Soc.* **2003**, *125*, 5594–5595.
- [65] S. Biswas, S. Samui, A. K. Das, S. Pasadi, K. Muniyappa, J. Naskar, *RSC Adv.* **2020**, *10*, 26388–26396.
- [66] S. Tyagi, S. Saxena, N. Kundu, T. Sharma, A. Chakraborty, S. Kaur, D. Miyoshi, J. Shankaraswamy, *RSC Adv.* **2019**, *9*, 40255–40262.
- [67] T. Sharma, N. Kundu, S. Kaur, A. Chakraborty, A. K. Mahto, R. P. Dewangan, J. Shankaraswamy, S. Saxena, *RSC Adv.* **2022**, *12*, 21760–21769.
- [68] J. Jana, R. K. Kar, A. Ghosh, A. Biswas, S. Ghosh, A. Bhunia, S. Chatterjee, *Mol. Biosyst.* **2013**, *9*, 1833–1836.
- [69] J. Jana, P. Sengupta, S. Mondal, S. Chatterjee, *RSC Adv.* **2017**, *7*, 20888–20899.
- [70] J. P. Vaughn, S. D. Creacy, E. D. Routh, C. Joyner-Butt, G. S. Jenkins, S. Pauli, Y. Nagamine, S. A. Akman, *J. Biol. Chem.* **2005**, *280*, 38117–38120.
- [71] L. T. A. Nguyen, D. T. Dang, *Mol. Biotechnol.* **2022**, 10.1007/s12033-022-00552-7.
- [72] B. Heddi, V. V. Cheong, H. Martadinata, A. T. Phan, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9608–9613.
- [73] S. Lattmann, B. Giri, J. P. Vaughn, S. A. Akman, Y. Nagamine, *Nucleic Acids Res.* **2010**, *38*, 6219–6233.
- [74] K. H. Ngo, R. Yang, P. Das, G. K. T. Nguyen, K. W. Lim, J. P. Tam, B. Wu, A. T. Phan, *Chem. Commun.* **2020**, *56*, 1082–1084.
- [75] M. Y. Yaneva, V. V. Cheong, J. K. Cheng, K. W. Lim, A. T. Phan, *Biochem. Biophys. Res. Commun.* **2020**, *531*, 62–66.
- [76] C. J. Wen, C. J. Wen, J. Y. Gong, J. Y. Gong, K. W. Zheng, Y. D. He, J. Y. Zhang, Y. H. Hao, Z. Tan, Z. Tan, *Chem. Commun.* **2020**, *56*, 6567–6570.
- [77] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden, P. E. Nielsen, *Nature* **1993**, *365*, 566–568.
- [78] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497–1500.
- [79] B. A. Armitage, *Methods Mol. Biol.* **2019**, *2035*, 333–345.
- [80] A. Gupta, L. L. Lee, S. Roy, F. A. Tanious, W. D. Wilson, D. H. Ly, B. A. Armitage, *ChemBioChem* **2013**, *14*, 1476–1484.
- [81] V. L. Marin, B. A. Armitage, *Biochemistry* **2006**, *45*, 1745–1754.
- [82] B. Datta, M. E. Bier, S. Roy, B. A. Armitage, *J. Am. Chem. Soc.* **2005**, *127*, 4199–4207.
- [83] K. A. Kormuth, J. L. Woolford, B. A. Armitage, *Biochemistry* **2016**, *55*, 1749–1757.
- [84] S. Roy, F. A. Tanious, W. D. Wilson, D. H. Ly, B. A. Armitage, *Biochemistry* **2007**, *46*, 10433–10443.
- [85] K. Jantos, R. Rodriguez, S. Ladame, P. S. Shirude, S. Balasubramanian, *J. Am. Chem. Soc.* **2006**, *128*, 13662–13663.
- [86] K. C. Liu, K. Röder, C. Mayer, S. Adhikari, D. J. Wales, S. Balasubramanian, *J. Am. Chem. Soc.* **2020**, *142*, 8367–8373.
- [87] L. Bonnat, L. Bar, B. Génaro, H. Bonnet, O. Jarjayes, F. Thomas, J. Dejeu, E. Defrancq, T. Lavergne, *Chem. Eur. J.* **2017**, *23*, 5602–5613.
- [88] S. Patra, J. B. Claude, J. V. Naubron, J. Wenger, *Nucleic Acids Res.* **2021**, *49*, 12348–12357.
- [89] Z. L. Huang, J. Dai, W. H. Luo, X. G. Wang, J. H. Tan, S. Bin Chen, Z. S. Huang, *J. Am. Chem. Soc.* **2018**, *140*, 17945–17955.
- [90] K. Reddy, B. Zamiri, S. Y. R. Stanley, R. B. Macgregor, C. E. Pearson, *J. Biol. Chem.* **2013**, *288*, 9860–9866.
- [91] M. J. Wortman, A. V. Dagdanova, A. M. Clark, E. W. Godfrey, S. M. Pascal, E. M. Johnson, D. C. Daniel, *Biochim. Biophys. Acta Mol. Cell Res.* **2020**, *1867*, 10.1016/j.bbamcr.2020.118674.
- [92] G. J. Bassell, S. T. Warren, *Neuron* **2008**, *60*, 201–214.

- [93] A. T. Phan, V. Kuryavyi, J. C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, A. Polonskaia, C. Chen, D. Clain, et al., *Nat. Struct. Mol. Biol.* **2011**, *18*, 796–804.
- [94] C. Zhao, G. Qin, J. Niu, Z. Wang, C. Wang, J. Ren, X. Qu, *Angew. Chem. Int. Ed.* **2021**, *60*, 432–438; *Angew. Chem.* **2021**, *133*, 436–442.
- [95] T. Miclot, C. Hognon, E. Bignon, A. Terenzi, M. Marazzi, G. Barone, A. Monari, *J. Phys. Chem. Lett.* **2021**, *12*, 10277–10283.
- [96] S. K. Mukherjee, J. M. Knop, R. Winter, *Chem. Eur. J.* **2022**, *28*, 10.1002/chem.202104182.
- [97] N. M. Luscombe, R. A. Laskowski, J. M. Thornton, *Nucleic Acids Res.* **2001**, *29*, 2860–2874.
- [98] J. Xu, H. Huang, X. Zhou, *JACS Au* **2021**, *1*, 2146–2161.
- [99] S. N. Oyaghire, C. J. Cherubim, C. A. Telmer, J. A. Martinez, M. P. Bruchez, B. A. Armitage, *Biochemistry* **2016**, *55*, 1977–1988.
- [100] B. Sahu, I. Sacui, S. Rapireddy, K. J. Zanotti, R. Bahal, B. A. Armitage, D. H. Ly, *J. Org. Chem.* **2011**, *76*, 5614–5627.
- [101] A. Dragulescu-Andrasi, S. Rapireddy, B. M. Frezza, C. Gayathri, R. R. Gil, D. H. Ly, *J. Am. Chem. Soc.* **2006**, *128*, 10258–10267.
- [102] S. Sarkar, B. A. Armitage, *ACS Infect. Dis.* **2021**, *7*, 1445–1456.

Manuscript received: November 2, 2022
Revised manuscript received: January 3, 2023
Accepted manuscript online: January 4, 2023
Version of record online: February 2, 2023
