


 Cite this: *Chem. Commun.*, 2022, 58, 12753

Long-range DNA interactions: inter-molecular G-quadruplexes and their potential biological relevance

 Denise Liano,^{†a} Ludovica Monti,^{†ab} Souroprobho Chowdhury,^{ac} Federica Raguseo^{ac} and Marco Di Antonio^{id*abc}

Guanine-rich DNA sequences are known to fold into secondary structures called G-quadruplexes (G4s), which can form from either individual DNA strands (intra-molecular) or multiple DNA strands (inter-molecular, iG4s). Intra-molecular G4s have been the object of extensive biological investigation due to their enrichment in gene-promoters and telomers. On the other hand, iG4s have never been considered in biological contexts, as the interaction between distal sequences of DNA to form an iG4 in cells was always deemed as highly unlikely. In this feature article, we challenge this dogma by presenting our recent discovery of the first human protein (CSB) displaying astonishing picomolar affinity and binding selectivity for iG4s. These findings suggest potential for iG4 structures to form in cells and highlight the need of further studies to unravel the fundamental biological roles of these inter-molecular DNA structures. Furthermore, we discuss how the potential for formation of iG4s in neuronal cells, triggered by repeat expansions in the *C9orf72* gene, can lead to the formation of nucleic-acids based pathological aggregates in neurodegenerative diseases like ALS and FTD. Finally, based on our recent work on short LNA-modified probes, we provide a perspective on how the rational design of G4-selective chemical tools can be leveraged to further elucidate the biological relevance of iG4 structures in the context of ageing-related diseases.

 Received 2nd September 2022,
 Accepted 19th October 2022

DOI: 10.1039/d2cc04872h

rsc.li/chemcomm

1 Introduction

G-quadruplexes (G4s) are non-canonical structures originating from DNA (or RNA) sequences that are rich in guanines (Gs). Under physiological conditions, four Gs can form a planar tetrameric structure known as G-tetrad (Fig. 1), which is stabilised by Hoogsteen hydrogen bonds and coordinated by a monovalent alkali cation, typically K^+ or Na^+ . Depending on the stoichiometry of formation, G4s can be either intra-molecular (forming from a single strand of DNA), or inter-molecular (forming from multiple DNA strands, iG4s) (Fig. 1). These two G4-structures have shown to have different structural and physio-chemical properties, including kinetics of folding and thermodynamic stability (Fig. 1, table).^{1,2} In addition, the orientation of the nucleic acid strand(s) can produce G4-structures with different topologies, such as parallel, antiparallel, hybrid, and looped.³ Bioinformatic and genome-wide sequencing approaches have demonstrated that intra-molecular G4s are highly prevalent in the human

genome, with over 700 000 detected structures that are particularly enriched in functional regions, such as gene promoters and telomeres.⁴ It has been therefore proposed that G4s can potentially play a key role in physiological processes in eukaryotes, especially gene-expression regulation.⁵ Furthermore, the presence of G4s in promoters (e.g., *cMYC*) and/or telomers of cancer-related genes has prompted the study of these structures in the context of cancer, leading to the development of G4-ligands as novel potential chemotherapeutic agents.^{6,7}

More recent studies on G4s focused on investigating their formation in the genomic context of living cells. In fact, the abundant presence of G4-structures in the human genome, does not necessarily imply their formation *in cellulo*, which is something that required further validation. Immuno-precipitation experiments using a G4-specific antibody allowed the detection of G4s in human chromatin, revealing that these structures act as markers of active transcription and associate with accessible chromatin regions.⁸ These findings encouraged the development of G4-targeting approaches, as well as biological characterization of these unique genomic structures.⁹ Moreover, live cell imaging using small molecule probes have confirmed the formation of G4-structures in living cells under non-perturbative conditions, highlighting the potential of these structures to act as important biological regulators.^{10,11}

^a Imperial College London, Chemistry Department, Molecular Sciences Research Hub, 82 Wood Lane, W12 0BZ, London, UK. E-mail: m.di-antonio@imperial.ac.uk

^b The Francis Crick Institute, 1 Midland Road, NW1 1AT, London, UK

^c The Institute of Chemical Biology, Molecular Science Research Hub, 82 Wood Lane, W12 0BZ, London, UK

[†] Authors contributed equally to this work.


It is becoming increasingly apparent that G4s can represent a significant structural feature that impacts the overall chromatin architecture and its three-dimensional structure. The ability of chromatin to establish long-range interactions is crucial to regulate gene-expression.¹² It is therefore conceivable that G4s may play a role in this process through the formation of inter-molecular structures (iG4s) which allows connection between distal DNA sequences in the space. However, when

investigating the biological function of these secondary structures, iG4s have been often overlooked and deemed unlikely to be physiologically relevant, due to the low probability of multiple DNA strands to form such multi-molecular architectures in the context of chromatinized DNA. In this context, our laboratory has recently reported the first example of an endogenous human protein that is able to selectively bind to iG4s over equivalent intra-molecular G4-structures, thus supporting the



Denise Liano

Denise Liano obtained a PhD from Imperial College London, in March 2022. Her work focused on understanding the role of non-canonical G4s in promoting premature aging in Cockayne Syndrome patients. There, she challenged the scepticism considering inter-molecular G4s (iG4s) to be biologically irrelevant by reporting the first example of the human endogenous protein CSB that specifically recognizes and interacts with iG4s. As a

Post-doc Associate, Denise moved to US to join the Müschen group at the Center of Molecular and Cellular Oncology at the Yale Cancer Center where she employs molecular and cellular biology tools, including CRISPR-Cas9 and human-induced pluripotent stem cells to understand the expression of key genes involved in lymphoma and leukemia development.



Ludovica Monti

Ludovica Monti is a Research Fellow at Imperial College London supported by the Marie Skłodowska-Curie Fellowship. She obtained a PharmD degree (2014) and a PhD in Life Sciences (2018) from Sapienza University of Rome (Italy). She conducted half of her graduate studies training in world-leading universities in the USA. Her PhD work was recognised by Sapienza with the “Minerva award” for academic excellence. As a

postdoctoral researcher at the University of California San Diego (UCSD, 2018–2021), her research focused on drug discovery and optimisation for neglected parasitic diseases of global health significance, namely schistosomiasis and human African Trypanosomiasis. As a member of the Center for Discovery and Innovation in Parasitic Diseases at UCSD, Ludovica’s research showed clinical potential through the development of new chemical entities as therapeutic antiparasitic agents. Her current research applies interdisciplinary sciences, including chemistry, biology, and genomics, to establish novel molecular approaches targeting the DNA of Trypanosoma brucei parasites.



Souroprobho Chowdhury

Souro Chowdhury is currently a final year PhD student in Dr Marco Di Antonio’s group at Imperial College London. Starting his PhD in 2019 on a prestigious President’s PhD Scholarship, Souro’s work has mainly focused on achieving kinetic disruption of DNA G4 structures by using chemically modified oligonucleotide probes such as LNA.



Federica Raguseo

Federica Raguseo is a PhD student in the Chemistry Department at Imperial between the Di Antonio, Di Michele and Elani groups. Her work focuses on Amyotrophic Lateral Sclerosis and Frontotemporal Dementia, both characterized by the presence of pathological aggregates which affect neurological functions. Her aim is to understand the mechanism of formation such aggregates, aiding in the development of a new drug target. Her interest in biophysics

sparked in upon joining the Sensus2018 international competition for biosensing development and continued during her MSc in 2020 where she worked on antibodies binding modes to G4s, for which she was awarded the Alfred Bader Prize.





Fig. 1 Schematic representation of intra- and inter-molecular G-quadruplex (G4) structures and their properties. A planar G-tetrad structure is formed by four guanine residues stabilised by Hoogsteen hydrogen bonds and coordinated by a central monovalent cation. Intra-molecular G4s are formed within a single DNA strand, while inter-molecular G4s generate from two (or more) strands.

idea that iG4s can be biologically functional in the context of living cells.¹³ In this feature article, we provide an overview of the chemical-biology approaches we have developed in our laboratory to address unanswered questions on G4-biology, with particular interest on iG4s and their potential biological role(s) in the context of ageing-related disease models, such as cancer and accelerating ageing disorders.



Marco Di Antonio

Marco Di Antonio studied Chemistry in Pavia (Italy) where he got a MChem degree in 2007, before moving to Padua University for his PhD in Molecular Sciences, awarded in 2011. Marco then joined The Chemistry Department of Cambridge University (UK), as a Research associate first and in 2015 as a Senior Research Associate. Marco started his own group at Imperial College Chemistry in 2018 supported by a BBSRC David Phillips Fellowship, before being promoted to Lecturer in 2022. He also holds a satellite group at the Francis Crick Institute and a Lister Research Prize Fellowship (2022). His current research is focused on combining chemical biology tools with genetics to investigate the role of nucleic acids structures and modifications in ageing and cancer biology.

2 CSB as a selective iG4 protein binder

The possibility of high-ordered G4 topologies to form within multiple single-stranded G-rich sequences initially emerged from the study of telomeric DNA.¹⁴ In fact, telomeric sequences are constituted of $(G_3T_2A)_n$ repeats in humans, and of $(G_4T_2)_n$ and $(G_4T_4)_n$ repeats in the ciliates *Tetrahymena* and *Oxytricha*, respectively.¹⁵ *Tetrahymena* repeats are very polymorphic and most G-rich repeats fold into topologically heterogeneous conformations in solution, with iG4s being the preferred conformation under physiological conditions.^{15–17} Interestingly, Moye *et al.* reported that the human telomerase is able to recognise and extend telomeric DNA when folded into an iG4 conformation, suggesting that telomeric DNA can potentially assemble into iG4-structures to be recognised as a substrate of telomerase.¹⁸

Following these observations, our laboratory has recently demonstrated that the human Cockayne Syndrome B (CSB) protein is a selective iG4s interactor.¹³ CSB is a 168 kDa protein which belongs to SWI/SNF family of chromatin remodelling factors. CSB is codified by the *ercc6* gene on chromosome 10q11.¹⁹ Mutations of the *ercc6* gene are associated with approximately 70% of Cockayne Syndrome (CS) cases.^{20,21} CS patients exhibit a variety of clinical features which cause accelerating ageing and, ultimately, premature death.²² The pleiotropic phenotype of this disease derives from the diverse roles played by CSB, including chromatin remodelling and transcription-coupled DNA repair (TCR).¹⁹ Scheibye-Knudsen *et al.* demonstrated that CSB is able to resolve ribosomal DNA (rDNA) G4-structures in an ATP-independent fashion, albeit this protein has never been classified as a canonical helicase.^{23–25} The hypothesis that CSB is essential for the resolution of rDNA G4-structures derives from the observation that G-rich rDNA sequences can form G4s during rDNA transcription,^{26,27} as increased transcriptional pausing has been found at these sites in the absence of functional CSB.²³

Our group has confirmed that CSB is able to resolve G4 structures formed by a series of rDNA sequences (rDNA-1:5'-GGGGCCGGGGGTGGGGTCGGCCGGGAAA-3', rDNA-2:5'-GGGTCGGGGGGTGGGGCCCGGGCCGGGG-3', rDNA-3:5'-AGGGAGGGAGACGGGGGGG-3'), and, interestingly, this effect was also associated to negligible resolution of other types of G4s structures, such as c-KIT1, c-MYC, HRAS, and hTELO.¹³ These findings suggested that, in the presence of a short single-stranded tail (5'-ATAATTATAAATAAATAAT-3'), rDNA G4s might fold into peculiar G4-structures that are specifically recognised and resolved by CSB. Accordingly, we investigated the topology of different rDNA G4s annealed in conditions that either promote (K^+) or not (Li^+) G4-formation. For the majority of the rDNA G4s tested, Circular Dichroism (CD) revealed typical spectra of a parallel G4-topology, with a maximum peak at ~263 nm and minimum peak at ~240 nm.^{13,28,29} Given that parallel G4s are highly common topologies,^{1,9} we reasoned that it was unlikely that a topology-based selectivity was responsible for the observed selective resolvase activity of CSB.¹³ Indeed, on closer inspection, analysis of the rDNA substrates by native gel





Fig. 2 rDNA sequences form iG4s that are selectively bound by CSB. (A) Native gel electrophoresis showing that untailed rDNA is not able to form iG4s in the presence of Li⁺ or K⁺ (lanes 1 and 2, respectively). A tail of at least 3–5 bp is required for iG4s to form under K⁺ conditions (upper bands displayed in lanes 4 and 6, respectively), but not under Li⁺ conditions (lanes 3 and 5). A longer tail of 10–20 bp allows the formation of more stable iG4s which can mostly form in presence of K⁺ (lanes 8 and 10, respectively) compared to Li⁺ conditions (lanes 7 and 9). (B) NMM staining of agarose gel confirmed formation of folded G4 structures from rDNA sequences (lane 2: untailed rDNA in K⁺; lane 3 to 5: tailed rDNAs under K⁺ conditions; lane 6: untailed rDNA in Li⁺). ssDNA remained unstained (lane 1). iG4 bands are indicated with a blue arrow. (C) Agarose EMSA gel on tailed rDNA G4 using increasing concentrations of CSB in K⁺ conditions (lanes 2–10). Lane 1 is a control with the oligonucleotide annealed in Li⁺. iG4 bands are indicated with a blue arrow while the interaction between the iG4 and CSB is indicated within a yellow box. (D) Schematic representation of the selective binding of CSB towards iG4s.¹³

electrophoresis revealed that rDNA G4s sequences (rDNA-1, rDNA-2 and rDNA-3) can potentially assemble into iG4-structures in the presence of K⁺ and when a minimum 3-bp overhang (5'-ATA-3') is added at the 5'- or 3'-end (Fig. 2(A)). To confirm that all the G4-species observed under these conditions could be ascribed to folded G4s, we stained the native gel with the G4-specific probe, *N*-methylmesoporphyrin IX (NMM),³⁰ which is a selective fluorescent probe for G4-structures (Fig. 2(B)).¹³ Notably, gel-based binding analysis revealed a selective interaction of CSB with the iG4s formed by the rDNA templates (Fig. 2(C)). The binding affinity of CSB towards the multimeric rDNA conformations was surprisingly high, with a calculated dissociation constant (K_d) in the picomolar range ($K_d = 300\text{--}500$ pM). This K_d is comparable to other very high affinity DNA and RNA G4-binding proteins, such as DHX36, which also displays K_d values in the picomolar range.³¹ In

contrast, the binding of CSB towards intra-molecular rDNA G4s required higher concentrations of the protein (>50 nM) to trigger any binding event. Similarly, high concentrations of CSB were required to allow its interaction with established intra-molecular G4s: *c*-KIT1, *c*-MYC, HRAS, and hTELO, or unstructured ssDNA, indicating a significantly higher (~100 fold) binding affinity of CSB towards rDNA iG4s compared to intra-molecular G4s and ssDNA.¹³ Moreover, our data show that CSB is generally able to interact with iG4s and its binding towards these particular G4-structures is not limited to rDNA substrates. Indeed, we have observed remarkable low picomolar affinity ($K_d = 10$ pM) against iG4 formed from the *Oxytricha* telomeric sequence d(G₄T₄G₄),^{17,32} confirming the binding selectivity of CSB for iG4s (Fig. 2(D)).

The ability of CSB to bind G4s is not surprising *per se* since, recently, another chromatin remodelling factor, SMARCA4, has been observed to bind intra-molecular G4s with high affinity, suggesting that proteins belonging to SWI/SNF family might contain a conserved domain that can interact with G4 structures.³³ Moreover, CSB contains a glycine/arginine region at its N-terminal domain, which makes the protein reminiscent of other G4-binding proteins like nucleolin.^{34,35} However, what is surprising is the selective binding towards iG4s that CSB displays and that has never been reported for any other G4-binding protein to date. Although a full structural characterization is still required to fully understand this selectivity, our findings suggest that iG4s might form in cells, opening up to the idea that such structures can also be biologically relevant, something that has always been neglected to date. Importantly, given the very diverse structural architecture of iG4s compared to the more canonical unimolecular G4s, we anticipate that the biological functions of iG4s might be significantly different to what has been reported so far for the “canonical” intra-molecular G4s.

3 Formation of cellular iG4s in ribosomal DNA

The promising results obtained with CSB *in vitro* prompted us to investigate whether the specific interaction of this protein with iG4s could be observed in cells. Cellular localisation studies revealed that CSB is mainly distributed in the nucleoli of eukaryotic cells (Fig. 3(A)),³⁶ where primary ribosome biogenesis occurs.³⁷ In addition, recent studies demonstrated that CSB is required for rDNA synthesis as a key component of RNA PolI transcription machinery,³⁸ with CSB-deficient cells showing increased activity of poly-ADP ribose-polymerase 1 (PARP1).^{23,39,40} Persistent activation of PARP1 leads to premature ageing in *C. elegans* models treated with G4-stabilising ligands, suggesting that the stabilisation of G4s causes stalling of the RNA PolI transcription and activates DNA-damage response pathways.²³ rDNA genes are arranged in arrays of tandem repeats, resulting in a local concentration of effectors which contributes to generate a dense, G-rich, and transcriptionally active environment that is ideal for iG4-formation.¹²





Fig. 3 G4-ligands displace CSB from the nucleoli of human cells, while BG4-staining is reduced in the presence of the protein. (A) Localisation pattern of CSB in HeLa cells without treatment (Untreated) or upon 24 hours treatment with either (B) PDS or (C) CX-5461. CSB localisation is followed by fusing the protein with an EGFP tag (green). CSB signal is recorded at 470 nm. White bars indicate 20 μm . (D) Localisation pattern of EGFP in HeLa cells is not influenced by treatment with CX-5461. (E) BG4 immuno-staining in CS1AN cells indicating homogeneous BG4 localisation. BG4 signal is recorded at 590 nm. White bar indicates 5 μm . (F) BG4 immuno-staining in CS1AN cells re-expressing CSB revealed lack of BG4 staining in nucleoli occupied by CSB-EGFP (white circles). CSB signal is recorded at 470 nm, while BG4 signal is recorded at 590 nm. White bar indicates 10 μm . (G) Lack of nucleolar BG4 staining observed in HeLa cells (white circles). BG4 signal is recorded at 590 nm. White bar indicates 5 μm .¹³

Therefore, we hypothesised that (i) iG4s could be formed in cells and be bound by CSB, and (ii) treatment with G4-ligands could cause displacement of CSB and induce the observed rDNA transcriptional arrest. To assess whether well-known G4-ligands, such as pyridostatin (PDS)⁴¹ and CX-5461,⁷ compete with CSB for binding to iG4s and cause CSB-displacement from the DNA structure, we conducted *in vitro* competition-binding assays. After incubation of a pre-bound iG4 rDNA-CSB complex with the G4-ligands, we observed a dose-dependent displacement of CSB from the iG4 substrate. We then asked if a similar displacement of CSB could also be observed in cells. To assess this, we incubated HeLa cells, which endogenously express CSB,

with either PDS or CX-5461. Strikingly, treatment of HeLa cells with G4-ligands caused a significant reduction of CSB in the nucleoli (Fig. 3(B) and (C)). Conversely, clear CSB staining was observed in the nucleoli of cells that were not treated with G4-ligands (Fig. 3(A)). These results suggest that G4-ligands can compete with the protein for binding to G4-structures formed within nucleolar rDNA sequences, confirming the potential for iG4s to be formed in cells at the rDNA level.¹³ Interestingly, Lyama *et al.* described the nucleolar displacement of CSB upon treatment of HeLa cells with CX-5461, however, this effect was attributed to a transcriptional inhibitory effect caused by the ligand.^{36,42} In addition, it has been demonstrated that CX-5461 acts as a potent G4-binder and -stabiliser,⁷ suggesting that the displacement of CSB from the nucleoli upon treatment with the ligand is more likely to be caused by binding of the molecule to iG4s rather than by transcriptional regulatory effects, as we also observed with PDS. To validate this, we used immuno-fluorescence to measure changes in expression levels (*i.e.*, transcriptional inhibition) of a green fluorescent protein (EGFP) upon 4 and 24 hours treatment with CX-5461. Notably, we failed to detect any transcriptional inhibition (Fig. 3(D)), which reinforced the hypothesis of CSB being displaced by rDNA iG4s formed in cells, as we also observed *in vitro* and upon treatment of HeLa cells with validated G4-ligands like PDS.

To confirm that CSB is bound to iG4s in the nucleoli of living cells, we performed immunostaining experiments using a synthetic G4-specific antibody, BG4,^{8,43} in Cockayne Syndrome cells (CS1AN) that are genetically impaired of the CSB gene. Immuno-staining of G4s with the BG4 antibody revealed a strong staining in the nucleoli of CS1AN cells, suggesting G4-formation at the rDNA level (Fig. 3(E)). Interestingly, we observed that the re-expression of CSB in CS1AN cells induced CSB-nucleolar localisation and caused a strong reduction of the BG4-staining in the nucleoli of these cells (Fig. 3(F)), suggesting that CSB is bound to iG4s in the nucleoli of living cells.¹³ In addition, we observed a similar lack of nucleolar BG4 staining in HeLa cells that express endogenous CSB. This effect was comparable to what was observed in CS1AN cells that re-express CSB (Fig. 3(F)), confirming the ability of the CSB protein to compete with BG4 antibody for binding to nucleolar iG4s.¹³

Overall, our findings strongly suggest that iG4s could be formed in the nucleoli of human cells where CSB acts as an endogenous binder of these structures. The dense nature of the nucleoli supports a model where multiple rDNA genes form long-range iG4 interactions (Fig. 4(A)), which might be key to sustain nucleolar cellular homeostasis and of particular importance to the premature ageing phenotypes observed in CS patients that lack functional CSB (Fig. 4(B)).¹³

4 The potential role of iG4s in neurodegenerative aggregates

4.1 ALS and FTD genetic traits

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are fatal neurodegenerative diseases that affect the





Fig. 4 iG4s can form within nucleoli of cells and their homeostasis is regulated by CSB. (A) Schematic representation of a cellular nucleus containing dense regions, referred to as nucleoli (in yellow). The nucleolar environment is enriched in rDNA sequences which are prone to interact and form iG4s that are recognised and handled by CSB. (B) Our model suggests that functional CSB binds iG4 structures promoting healthy ageing (green boxes). When CSB is not functional or after treatment with G4-ligands, iG4s are mis-regulated, promote stalling of PolI, and activate PARP1 response. All these effects are known to trigger premature ageing (red boxes).^{13,23}

motor system and cognitive behaviors with severe consequences on execution of functions, behaviors, and language abilities. ALS affects the motor neurons in the frontal cortex, brainstem, and spinal cord, while FTD is related to motor neuron disorder. The link between the two diseases can be appreciated both at the clinical level and by the overlapping genetic profiles. In fact, both diseases present mutations in a series of genes and a characteristic expansion of the Chromosome 9 Open Reading Frame 72 gene (C9orf72).⁴⁴ More specifically, C9orf72 presents a hexanucleotide repeat expansion (GGGGCC)_n that is the most common cause of ALS, accounting for approximately 40% of the familiar cases (*i.e.*, family history of ALS) and 7% of sporadic cases in Europe.⁴⁵ Statistical studies have shown that healthy people usually present a repeat number (*n*) of 2, while in ALS patients the *n* value typically raises from hundreds to thousands of repeats.⁴⁶ In addition, a smaller part of the population presents an intermediate number of

repeats, from 20 to 30. It has been reported that an expansion of 24 or more (*p* value = 2×10^{-4}) should be considered pathogenic.⁴⁵

Due to the significant overlap in genetic hallmarks, ALS and FTD often form a neurodegenerative continuum, where both diseases have been observed within same families, or even the same individual.⁴⁷

4.2 Pathological aggregates in ALS and FTD

Much like most neurodegenerative diseases, ALS and FTD are characterised by the presence of pathological aggregates.⁴⁸ The mechanism of formation of such aggregates is still widely uncharacterised, but their composition comprises primarily proteins with low complexity domains and nucleic acids (principally RNA).⁴⁹

Protein aggregation has been considered the main potential trigger for G4^{50–52} and/or condensate formation;^{48,53,54} more recently, increasing evidence highlights the relevance of RNA in such aggregates, due to its ability to stabilise pre-existing protein aggregates through non-specific stacking and electrostatic interactions.⁵⁵ In addition, it is also thought that RNA may play a role in the mechanism of aggregate formation,^{56,57} but this hypothesis is yet to be proved. In this context, Fay *et al.* have shown cell-aggregation of mRNA transcript of C9orf72 gene in ALS affected neurons, proving that nucleic acids can form aggregates.⁵⁷ These results make the C9orf72 gene an incredibly appealing target to underpin the mechanism of action of nucleic acid in the context of pathological aggregates. Interestingly, the C9orf72 gene has shown *in vitro* polymorphism to form RNA hairpin and G4s secondary structures, implying that nucleic acid secondary structures might contribute to trigger aggregation. Formation of iG4s within such G-rich nucleic acid repeats would cluster together multiple nucleic strands, providing the potential to the system to aggregate on a microscopic level (nucleation point, Fig. 5). Further structural characterization of the gene (GGGGCC)_n was performed by Fay *et al.*, confirming the ability of C9orf72 to form G4s from both RNA and DNA, albeit further elucidation on if and how such G4s might contribute to the aggregation is still required. Moreover, a recent crystallography study has shown the ability of (GGGGCC)_n to form iG4 DNA:RNA tetrameric hybrids, supporting the hypothesis that this sequence has the potential to form intermolecular links.⁵⁸

Currently, there is no FDA-approved cure for ALS and FTD, although the drugs Riluzole (developed for Parkinson's disease) and Edaravone have shown to increase survival rates in randomized ALS clinical trials.^{59,60} So far, candidate drugs have been developed to target protein-lead aggregation; since recent studies highlight the potential of nucleic acids to potentially be key components of pathological aggregates, there is scope to pursue novel therapeutic pathways that target nucleic acid structures, including iG4s. In fact, it has been shown that G4-targeting ligands have potential to affect the disease progression, although it is not clear how they affect the aggregation pathway.⁶¹ Considering our recent work that identified an iG4s selective resolvase,¹³ we anticipate that increasing attention should be pointed at these G4-structures, which might be also





Fig. 5 Schematic illustration of iG4s as nucleation triggers for aggregation in ALS/FTD affected cells. (i) ALS and FTD are characterised by a mutation at the C9orf72 gene at the repeat expansion (GGGGCC)_n, a sequence with the potential to form iG4s. (ii) iG4s can aggregate at high concentrations, giving rise to a nucleation point. (iv) RNA binding proteins use the nucleation point created by the iG4s to form pathological aggregates.

relevant in the context of neurodegenerative diseases like ALS and FTD.

5 Use of LNA probes to disrupt G4s

Our work provides robust evidence to support the idea that G4s can have a range of biological functions that might not simply depend on their genomic location, but also on the stoichiometry of their formation (*i.e.*, iG4s).¹² G4s might indeed play a significant role in nucleolar homeostasis and could contribute to the formation of pathological aggregates. This highlights the need of chemical probes to selectively target specific G4s within the genome to further elucidate their cellular functions. Furthermore, selective G4-probes that bind to resolve/disrupt G4s (rather than to stabilise them) could represent an alternative therapeutic strategy to alleviate ageing-related diseases like ALS and FTD.⁶² Surprisingly, there is only a limited number of examples of G4-disrupting agents in the literature, highlighting the chemical challenges associated with developing such probes. Waller *et al.* reported one of the earliest examples of the disruption of a DNA G4s, which led to enhanced expression of the *c-KIT* gene in cells upon treatment with such small-molecule.⁶³ More recently, Monchaud and co-workers developed an innovative platform to facilitate the identification of small-molecules that can disrupt G4s, and they have shown that G4-disrupting ligands can cooperate with DNA helicases to accelerate G4-unfolding.⁶⁴

A significant drawback of small-molecule G4-ligands is their lack of sequence specificity. The inherent inability of G4-ligands to selectively target specific G4-sequences, together with the high abundance of G4-structures (>700 000 putative G4 sequences) in the human genome, leads to unwanted genome-wide perturbations. This is particularly relevant when considering the diverse biological functions that G4s located in different genomic locations might elicit. For example, a potential G4-targeting based treatment for ALS would require a molecule that selectively recognises G4s formed in the C9orf72 gene. To this end, a possible alternative to G4-disrupting small-molecules is the use of chemically-modified oligonucleotide probes,⁶⁵ which are ideal sequence-specific G4-disrupting agents as they contain the necessary sequence information *via* Watson-Crick base pairing. Accordingly, early approaches showed the use of Peptide Nucleic Acids (PNA) to disrupt G4s *in vitro*.⁶⁶ Another highly promising chemically-modified alternative are Locked Nucleic Acids (LNA), which have shown to thermodynamically drive greater duplex formation.⁶⁷ LNA-modified probes are particularly attractive for chemical-biology applications due to their enhanced base-pairing, reduced endonuclease degradation, and significantly improved mismatch discrimination, as compared to unmodified DNA.^{20,68}

We recently explored the use of LNA-modified oligonucleotide probes to achieve the disruption of individual G4s. Using a similar approach to the antisense oligonucleotide (ASO) design, we developed a series of LNA-modified cDNA sequences that target model G4 sequences (Fig. 6).⁶⁹ Intriguingly, we found that the G4-disrupting potential of the LNA-modified probes varied significantly depending on the *placement* of the LNA-modification in the complementary sequence. Specifically, probes with the LNAs placed complementary to the Gs (guanines) showed significantly lower G4-disrupting half-lives compared to probes where the LNA modification was in loops or randomly distributed. These observations were consistent across two G4 model systems with different topologies that were tested: a parallel G4 (c-KIT1) and a mixed-type G4 (hTelo). These results suggest that our approach might be an efficient and versatile strategy to achieve disruption of individual G4s with high specificity. Notably, we found that the G4-disrupting potency of the LNA-modified probes was preserved (or even improved) when the length of the sequence was shortened from ~20 nucleotides to ~10 nucleotides. These results allowed us to propose the potential application of such short LNA-modified probes beyond their use in *in vitro* biophysical assays. For example, using a combination of single-molecule experiments and biochemical assays, we demonstrated that short LNA-modified probes can disrupt G4s responsible for stalling of DNA polymerases, thus suggesting that these probes could be used to modulate essential biological functions and rescue the progression of the stalled polymerases. Finally, we investigated whether the short G4-targeting LNA probes could be used to modulate gene-expression in cells through a targeted disruption of G4s in promoter regions. Accordingly, we used a previously validated dual-reporter plasmid system⁷⁰ consisting





Fig. 6 Schematic representation of LNA-mediated G4-disruption. LNA-modified probes contain the necessary sequence information *via* Watson-Crick base pairing to disrupt a specific G4. The position of the LNA-modification along the complementary sequence (in red) can be varied to optimise the G4-disrupting potential of the probe, depending on the requirement. LNA probes can be used to achieve efficient G4-disruption in both biophysical *in vitro* assays and in cells. Our studies revealed that the LNA–DNA complex induces an increase in gene expression without hampering biological processes such as transcription and replication, thus suggesting that this is a viable strategy to achieve G4-disruption in cells.⁶⁹

of *renilla* and *firefly* genes. These genes, when expressed, can generate luminescence signals that can be used to quantitatively measure levels of gene-expression. Within the plasmid construct, the promoter of the *renilla* gene is modified to contain the c-KIT promoter sequence, allowing the c-KIT1 G4 present in this promoter to control the expression of the *renilla* gene. On the other hand, the *firefly* gene serves as a control of transfection efficiency between individual experiments. Consistently with our biophysical and biochemical data, we found that short LNA-modified probes significantly enhanced c-KIT1 gene-expression, whereas this effect was not observed when the iso-sequential unmodified DNA probe was used (Fig. 6).⁶⁹

Taken together, our studies demonstrated that short LNA-modified oligonucleotides are viable chemical-biology tools

that can individually target and disrupt specific G4s both *in vitro* and in cells. This opens-up new possibilities of the use of these sequence-specific probes to study the biological effect of disrupting individual G4s (rather than using unspecific small-molecules). For example, these probes could be used to study biological phenotypes associated with diseases, such as G4-helicase deficiencies, characteristic of the Cockayne Syndrome, or neurodegenerative diseases such as ALS/FTD that have been linked to the presence of excessive G4-forming repeat-sequences.

6 Conclusions

Intra-molecular G4s have been the focus of G4-based research, with remarkable findings showing the importance of these atypical DNA secondary structures in regulating many essential physiological processes, including epigenetic regulation and transcriptional control. In the last decades, highly innovative chemical-biology techniques have been developed to map and visualise the formation of G4s in living cells. However, little attention has been dedicated to study and unravel the biological role of G4-structures formed by multiple DNA-strands (iG4s), which have been deemed not to be biologically relevant due to the low probability of long-range DNA-structures to form in chromatinized DNA.

Our laboratory has recently challenged this dogma and generated new evidence supporting the potential for formation of iG4s in human cells, as well as their possible implication in pathologies like the Cockayne Syndrome. Specifically, our recent work on the CSB protein contributes to demonstrating that iG4s could be physiological components of cellular nucleoli and, in the presence of functional CSB, these structures contribute to cellular homeostasis. In the absence of functional CSB, iG4s accumulate in the nucleoli unresolved, contributing to premature ageing-phenotypes.

The discovery of CSB being an iG4-selective binding protein motivated us to reconsider the role that G4-structures may play in certain diseases, accounting for the potential of also multi-molecular G4-structures (iG4s) to be formed in living cells and contributing to certain phenotypes. For example, in ALS and FTD cells, the accumulation of unresolved iG4-structures might contribute to the formation of the pathological aggregates that are typical hallmarks of these neurological disorders, acting as nucleation points for condensation. These observations lead to re-think the use of molecules to target G4s for therapeutic intervention, and to approach G4-targeting therapies from a fresh perspective. In this context, the use of G4-specific targeting agents, such as LNA-modified oligonucleotide probes, represent a viable alternative to selectively disrupt unresolved, accumulated iG4s and possibly maintain cellular homeostasis. Overall, this feature article aims to provide insights into new strategies and approaches to further investigate iG4s, aiming to better characterise these unique nucleic acid structures and their functional role in human biology, as well as their potential as therapeutic targets in ageing-derived diseases, including



neurodegenerative diseases. We anticipate that iG4s will emerge as an important structural feature that contributes towards maintaining chromatin and cellular homeostasis and, therefore, they should be taken into account when investigating G4-biology.

Author contributions

Conceptualization: M. D. A.; writing – original draft preparation M. D. A., L. M., D. L., F. R. and S. C.; writing – review and editing L. M., and M. D. A. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) David Phillips Fellowship to Marco Di Antonio [BB/R011605/1]. Marco Di Antonio is supported by a Lister Institute Research Prize awarded in 2022. Ludovica Monti is supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 101027645. Souroprobho Chowdhury is supported by a President Scholarship from Imperial College London. Federica Raguseo is supported by a Leverhulme Trust, Cellular Bionics scholarship [EP/S023518/1]. Open access funding were kindly provided by Imperial College London.

Notes and references

- 1 I. Frasson, *et al.*, Multimeric G-quadruplexes: A review on their biological roles and targeting, *Int. J. Biol. Macromol.*, 2022, **204**, 89–102.
- 2 J. Jana and K. Weisz, Thermodynamic stability of G-quadruplexes: Impact of sequence and environment, *ChemBioChem*, 2021, **22**(19), 2848–2856.
- 3 J. Spiegel, S. Adhikari and S. Balasubramanian, The Structure and Function of DNA G-Quadruplexes, *Trends Chem.*, 2020, **2**(2), 123–136.
- 4 V. S. Chambers, *et al.*, High-throughput sequencing of DNA G-quadruplex structures in the human genome, *Nat. Biotechnol.*, 2015, **33**(8), 877–881.
- 5 D. Rhodes and H. J. Lipps, G-quadruplexes and their regulatory roles in biology, *Nucleic Acids Res.*, 2015, **43**(18), 8627–8637.
- 6 R. Hänsel-Hertsch, M. Di Antonio and S. Balasubramanian, DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**(5), 279–284.
- 7 H. Xu, *et al.*, CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours, *Nat. Commun.*, 2017, **8**, 14432.
- 8 R. Hänsel-Hertsch, *et al.*, G-quadruplex structures mark human regulatory chromatin, *Nat. Genet.*, 2016, **48**(10), 1267–1272.
- 9 F. Raguseo, *et al.*, Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome, *Chem. Commun.*, 2020, **56**(9), 1317–1324.
- 10 M. Di Antonio, *et al.*, Single-molecule visualization of DNA G-quadruplex formation in live cells, *Nat. Chem.*, 2020, **12**, 832–837.
- 11 P. A. Summers, *et al.*, Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy, *Nat. Commun.*, 2021, **12**(1), 162.
- 12 J. Robinson, *et al.*, DNA G-quadruplex structures: More than simple roadblocks to transcription?, *Nucleic Acids Res.*, 2021, **49**(15), 8419–8431.
- 13 D. Liano, S. Chowdhury and M. Di Antonio, Cockayne syndrome B protein selectively resolves and interact with intermolecular DNA G-quadruplex structures., *J. Am. Chem. Soc.*, 2021, **143**(49), 20988–21002.
- 14 S. Kolesnikova and E. A. Curtis, Structure and function of multimeric G-quadruplexes, *Molecules*, 2019, **24**, 17.
- 15 E. Demkovičová, *et al.*, Telomeric G-quadruplexes: From human to, *J. Nucleic Acids*, 2017, **2017**, 9170371.
- 16 Q. Guo, M. Lu and N. R. Kallenbach, Effect of thymine tract length on the structure and stability of model telomeric sequences, *Biochemistry*, 1993, **32**(14), 3596–3603.
- 17 S. Burge, *et al.*, Quadruplex DNA: Sequence, topology and structure, *Nucleic Acids Res.*, 2006, **34**(19), 5402–5415.
- 18 A. L. Moye, *et al.*, Telomeric G-quadruplexes are a substrate and site of localization for human telomerase, *Nat. Commun.*, 2015, **6**, 7643.
- 19 V. Tiwari, *et al.*, Current and emerging roles of Cockayne syndrome group B (CSB) protein, *Nucleic Acids Res.*, 2021, **49**(5), 2418–2434.
- 20 V. Laugel, *et al.*, Mutation update for the CSB/ERCC6 and CSA/ERCC8 genes involved in Cockayne syndrome, *Hum. Mutat.*, 2010, **31**(2), 113–126.
- 21 S. Wang, *et al.*, Rescue of premature aging defects in Cockayne syndrome stem cells by CRISPR/Cas9-mediated gene correction, *Protein Cell*, 2020, **11**(1), 1–22.
- 22 B. T. Wilson, *et al.*, The Cockayne Syndrome Natural History (CoSynH) study: Clinical findings in 102 individuals and recommendations for care, *Genet. Med.*, 2016, **18**(5), 483–493.
- 23 M. Scheibye-Knudsen, *et al.*, Cockayne syndrome group A and B proteins converge on transcription-linked resolution of non-B DNA, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**(44), 12502–12507.
- 24 C. P. Selby and A. Sancar, Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II, *J. Biol. Chem.*, 1997, **272**(3), 1885–1890.
- 25 N. Beerens, *et al.*, The CSB protein actively wraps DNA, *J. Biol. Chem.*, 2005, **280**(6), 4722–4729.
- 26 S. Chiarella, *et al.*, Nucleophosmin mutations alter its nucleolar localization by impairing G-quadruplex binding at ribosomal DNA, *Nucleic Acids Res.*, 2013, **41**(5), 3228–3239.
- 27 D. Drygin, *et al.*, Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis, *Cancer Res.*, 2009, **69**(19), 7653–7661.
- 28 J. L. Huppert, Four-stranded nucleic acids: Structure, function and targeting of G-quadruplexes, *Chem. Soc. Rev.*, 2008, **37**(7), 1375–1384.
- 29 R. Del Villar-Guerra, J. O. Trent and J. B. Chaires, G-quadruplex secondary structure obtained from circular dichroism spectroscopy, *Angew. Chem., Int. Ed.*, 2018, **57**(24), 7171–7175.
- 30 N. C. Sabharwal, *et al.*, N-Methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes, *FEBS J.*, 2014, **281**(7), 1726–1737.
- 31 P. Schult and K. Paeschke, The DEAH helicase, *Biol. Chem.*, 2021, **402**(5), 581–591.
- 32 X. Guo, S. Liu and Z. Yu, Bimolecular quadruplexes and their transitions to higher-order molecular structures detected by ESIF-TICR-MS, *J. Am. Soc. Mass Spectrom.*, 2007, **18**(8), 1467–1476.
- 33 X. Zhang, *et al.*, Chemical profiling of DNA G-quadruplex-interacting proteins in live cells, *Nat. Chem.*, 2021, **13**(7), 626–633.
- 34 N. L. Batenburg, *et al.*, Cockayne Syndrome group B protein interacts with TRF2 and regulates telomere length and stability, *Nucleic Acids Res.*, 2012, **40**(19), 9661–9674.
- 35 Z. L. Huang, *et al.*, Identification of G-quadruplex-binding protein from the exploration of RGG Motif/G-quadruplex Interactions, *J. Am. Chem. Soc.*, 2018, **140**(51), 17945–17955.
- 36 T. Iyama, *et al.*, Regulation of the intranuclear distribution of the cockayne syndrome proteins, *Sci. Rep.*, 2018, **8**(1), 17490.
- 37 M.-L. Dubois and F. O.-M. Boisvert, in *The Dynamic Functions of the Nucleolus 2016*, ed. D. J. Bazett-Jones and G. Dellaire, Springer International Publishing, Switzerland, 2016.
- 38 J. Bradsher, *et al.*, CSB is a component of RNA pol I transcription, *Mol. Cell*, 2002, **10**(4), 819–829.



- 39 M. N. Okur, *et al.*, Cockayne syndrome group A and B proteins function in rRNA transcription through nucleolin regulation, *Nucleic Acids Res.*, 2020, **48**(5), 2473–2485.
- 40 A. D. Edwards, *et al.*, G-Quadruplex loops regulate PARP-1 enzymatic activation, *Nucleic Acids Res.*, 2021, **49**(1), 416–431.
- 41 R. Rodriguez, *et al.*, A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres, *J. Am. Chem. Soc.*, 2008, **130**(47), 15758–15759.
- 42 M. J. Bywater, *et al.*, Inhibition of RNA polymerase I as a therapeutic strategy to promote cancer-specific activation of p53, *Cancer Cell*, 2012, **22**(1), 51–65.
- 43 G. Biffi, *et al.*, Quantitative visualization of DNA G-quadruplex structures in human cells, *Nat. Chem.*, 2013, **5**(3), 182–186.
- 44 M. DeJesus-Hernandez, *et al.*, Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS, *Neuron*, 2011, **72**(2), 245–256.
- 45 A. Iacoangeli, *et al.*, C9orf72 intermediate expansions of 24–30 repeats are associated with ALS, *Acta Neuropathol. Commun.*, 2019, **7**(1), 115.
- 46 S. Van Mossevelde, *et al.*, Relationship between C9orf72 repeat size and clinical phenotype, *Curr. Opin. Genet. Dev.*, 2017, **44**, 117–124.
- 47 Y. A. Abramzon, *et al.*, The overlapping genetics of amyotrophic lateral sclerosis and frontotemporal dementia, *Front. Neurosci.*, 2020, **14**, 42.
- 48 C. A. Ross and M. A. Poirier, Protein aggregation and neurodegenerative disease, *Nat. Med.*, 2004, **10**(suppl), S10–S17.
- 49 S. F. Banani, *et al.*, Biomolecular condensates: Organizers of cellular biochemistry, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**(5), 285–298.
- 50 G. Fang and T. R. Cech, Characterization of a G-quartet formation reaction promoted by the beta-subunit of the Oxytricha telomere-binding protein, *Biochemistry*, 1993, **32**(43), 11646–11657.
- 51 G. Fang and T. R. Cech, The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA, *Cell*, 1993, **74**(5), 875–885.
- 52 H. Masai, *et al.*, Rif1 promotes association of G-quadruplex (G4) by its specific G4 binding and oligomerization activities, *Sci. Rep.*, 2019, **9**(1), 8618.
- 53 S. Alberti and A. A. Hyman, Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing, *Nat. Rev. Mol. Cell Biol.*, 2021, **22**(3), 196–213.
- 54 J. S. Fassler, *et al.*, Protein aggregation and disaggregation in cells and development, *J. Mol. Biol.*, 2021, **433**(21), 167215.
- 55 A. Louka, *et al.*, RNA as the stone guest of protein aggregation, *Nucleic Acids Res.*, 2020, **48**(21), 11880–11889.
- 56 Y. Zhang, *et al.*, G-quadruplex structures trigger RNA phase separation, *Nucleic Acids Res.*, 2019, **47**(22), 11746–11754.
- 57 M. M. Fay, P. J. Anderson and P. Ivanov, ALS/FTD-Associated C9ORF72 repeat RNA Promotes phase transitions *in vitro* and in cells, *Cell Rep.*, 2017, **21**(12), 3573–3584.
- 58 Y. Geng, *et al.*, Crystal structure of parallel G-quadruplex formed by the two-repeat ALS- and FTD-related GGGGCC sequence, *Nucleic Acids Res.*, 2021, **49**(10), 5881–5890.
- 59 A.-T. Af Degeneration.F., ALS and Frontotemporal Degeneration. 2018.
- 60 R. M. Liscic, *et al.*, From basic research to the clinic: Innovative therapies for ALS and FTD in the pipeline, *Mol. Neurodegener.*, 2020, **15**(1), 31.
- 61 R. Simone, *et al.*, G-quadruplex-binding small molecules ameliorate, *EMBO Mol. Med.*, 2018, **10**(1), 22–31.
- 62 P. Lejault, *et al.*, How to untie G-quadruplex knots and why? *Cell, Chem. Biol.*, 2021, **28**(4), 436–455.
- 63 Z. A. Waller, *et al.*, A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression, *J. Am. Chem. Soc.*, 2009, **131**(35), 12628–12633.
- 64 J. Mitteaux, *et al.*, Identifying G-quadruplex-DNA-disrupting small molecules, *J. Am. Chem. Soc.*, 2021, **143**(32), 12567–12577.
- 65 E. Cadoni, *et al.*, Beyond small molecules: Targeting G-quadruplex structures with oligonucleotides and their analogues, *Nucleic Acids Res.*, 2021, **49**(12), 6638–6659.
- 66 J. J. Green, *et al.*, Kinetics of unfolding the human telomeric DNA quadruplex using a PNA trap, *J. Am. Chem. Soc.*, 2003, **125**(13), 3763–3767.
- 67 N. Kumar and S. Maiti, Role of locked nucleic acid modified complementary strand in quadruplex/Watson-Crick duplex equilibrium, *J. Phys. Chem. B*, 2007, **111**(42), 12328–12337.
- 68 J. Wengel, *et al.*, Chemistry of locked nucleic acids (LNA): Design, synthesis, and bio-physical properties, *Lett. Pept. Sci.*, 2003, 237–253.
- 69 S. Chowdhury, *et al.*, Short LNA-modified oligonucleotide probes as efficient disruptors of DNA G-quadruplexes, *Nucleic Acids Res.*, 2022, **50**(13), 7247–7259.
- 70 C. Ducani, *et al.*, Interplay of three G-quadruplex units in the KIT promoter, *J. Am. Chem. Soc.*, 2019, **141**(26), 10205–10213.

