The C9Orf72 hexanucleotide repeat expansion transcript forms insoluble aggregates facilitated by multimolecular G-quadruplex structures

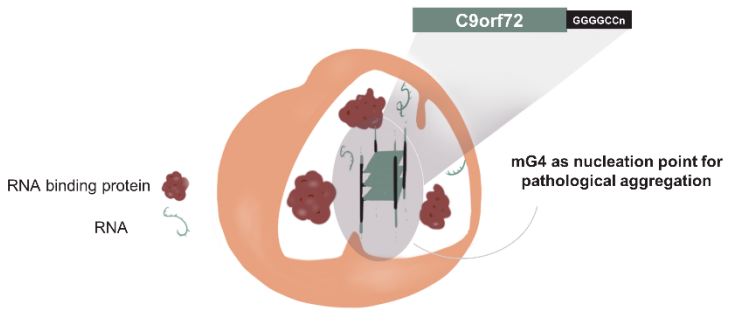
Federica Raguseo[1,2], Diana Tanase[1,2], Layla Malouf[1 2], Roger M. Rubio Sanchez[2], Yuval Elani[3], Lorenzo Di Michele*[1,2], Marco Di Antonio*[1,4]

1 Imperial College London, Chemistry Department, Molecular Sciences Research Hub, 82 Wood Lane, London W12 0BZ, UK
2 University of Cambridge, Department of Chemical Engineering and Biotechnology, West Cambridge Site, Cambridge, CB3 0AS, UK
3 Imperial College London, Department of Chemical Engineering, South Kensington, London SW7 2AZ, UK
4 The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

* To whom correspondence should be addressed. Email: m.di-antonio@imperial.ac.uk; ld389@cam.ac.uk.

Abstract

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are just two of the neurodegenerative diseases characterised by the presence of pathological aggregates[1]. 40% of familiar cases in ALS and FTD have been correlated to expansion of the hexanucleotide repeat $(GGGGCC)_n$, which has been previously shown to assemble into G-quadruplexes (G4s) structures[2]. In this study we investigated the role of nucleic acids and secondary structure formation in the generation of ALS/FTD aggregates, something that has often been relegated as a peripheral effect in the protein-led aggregation[3]. We showed a correlation between the emergence of multimolecular G4s (mG4s) formed by the DNA $(GGGGCC)_n$ repeats and the formation of protein free insoluble aggregates. We revealed that the aggregation is dependent on K^+ concentration and repeat-length, suggesting that G4-formation plays a role in the the formation of aggregates. G4-structures were detected in the aggregates by staining with the G4-specific fluorescent dye NMM. G4-unfolding promoted by NMM-mediated guanine photo-oxidation led to prompt disassembly of the insoluble aggregates, further confirming a G4-based mechanism. To reinforce the physiological relevance of our observations, we characterised the aggregation of RNA $(GGGGCC)_n$, which is thought to contribute to pathological aggregation in ALS/FTD. We observed that RNA repeats can aggregate at significant lower concentrations compared to DNA, suggesting that under physiological conditions RNA repeats can aggregate in the absence of any protein. Our findings constitute the first evidence supporting the formation of mG4-structures to drive protein-free aggregation, highlighting the potential of mG4s as therapeutic target to for the treatment of ALS and FTD.



Introduction

ALS and FTD are the of two most common types of fatal dementia in adults above > 65 y.o[4] with the first being found in approximately 2.1 cases per 100,000 people[5], whilst the second having an incidence of 4 cases per 100,000[6]. Although they appear to have different symptoms in the first stages of the diseases (ALS impairing movement[7] and FTD a patient's mental abilities[8]), both diseases are considered as a broad neurodegenerative continuum[1]. The link between the two is not only recognised from clinical observations[9], but also from overlapping in their genetics. In fact, mutations at TRDBP (TAR DNA-binding protein 43), SQSTM1 (Sequestrome-1), FUS (Fused in Sarcoma), VCP (Valosin Containing Protein), CHCHD10 (Coiled-Coil-Helix-Coiled-Coiled-Helix Domain Containing 10), TBK1 (TANK-Binding Kinase 1) and C9orf72 (Chromosome 9 Open Reading Frame 72) have been identified as hallmarks of both conditions[4].

ALS and FTD are, much like other neurodegenerative diseases, characterised by the presence of pathological aggregates at the neuron level, containing a mix of proteins and nucleic acids [1 10]. Many of the existing studies have focused on the protein component as the leading aggregation trigger[3], dismissing the role of nucleic acids as an element of additional stabilisation in the aggregation process[11]. Indeed, it has been repeatedly shown in ALS and FTD that RNA-binding proteins such as TDP-43/FUS present low-complexity domains which are prone to aggregation, due to low-affinity interactions[12 13 14 15]. In turn, recruitment of RNA in the aggregates has been regarded as a secondary step. However, despite the mainstream hypothesis proposing the mutated proteins as the primary drivers for aggregation, the hexanucleotide expansion $(GGGGCC)_n$ in the intronic region of C9orf72 is actually the most common mutation in ALS, accounting for about 40% or familiar cases (where the patient had a regressed family history of ALS) and 7% of the sporadic (no previous indication of family history of ALS)[16] in Europe. From statistical studies it has emerged that most people present a repeat number of 2, while patients affected by ALS the *n* value raises from hundreds to thousands of repeats[17]. It has been proposed that an expansion of 24 or more should be considered pathogenic[16]. The mutation in the gene has been previously proposed to lead to the formation of aggregates *via* various mechanisms[18 19], including ones in which the RNA plays a structural role, but conclusive evidence is yet to be provided[20].

In this work we bring forward the idea that the nucleic acid component could be the trigger of pathological aggregates, an hypothesis which has become increasingly accepted in recent years[21 22 10], based on the evidence that nucleic acids do not need the proteins to form biomolecular condensates[23]. Endogenous nucleic acid sequences do not typically lead to aggregation, but they do have the potential to form multimeric networks by both canonical and non-canonical base pairing interactions[24]. For example, $r(CAG)_n$ and $r(GGGGCC)_n$, both natural sequences related to neurodegenerative pathologies have been shown to phase-separate in vitro at a critical number of repeats, suggesting that they can form multimolecular structures[25]. However, there is no evidence to date that confirm which structures are indeed formed in the context of these aggregates, which would be essential to devise therapeutic agents. We reasoned that multimolecular G-quadruplex structures (mG4s) are ideal candidate to form such intermolecular networks, given their fast-folding kinetics and thermodynamic stability.

G-quadruplexes (G4) are stable secondary structures generated by non-canonical G-G base paring, which can form under physiological conditions[26]. G4-formation has already been associated with cancer, neurodegenerative diseases (ALS and FTD) and rare-genetic diseases (Fragile X syndrome and Cockayne Syndrome)[27 28]. From a structural standpoint, the G4 scaffold is composed of a G-tetrad, a planar tetrameric assembly of guanine (G) bases linked together by Hoogsten H-bonding (Figure 1A). The G-tetrad is stabilised by the coordination with a metal monovalent cation in its central cavity, the relative stability depending on the size of the metal (in order of increasing stability: $Li^+ << Na^+ < K^+$)[29], and more tetrads are held together by π-stacking interactions (Figure 1A). G4-formation stoichiometry could either be unimolecular or multimolecular (mG4s), depending on the number of strands involved in the structure formation[30].

We hypothesised that mG4s could be of relevance in the mechanism of formation of the ALS and FTD pathological aggregates, as a series of recent studies have brough forward indirect evidence in favour

of this hypothesis. Firstly, the hexanucleotide repeat $(GGGGCC)_n$ in C9orf72 has shown the ability to form G4s both in its RNA and DNA form[2 31 32] and the formation of these structures has been detected in cellular ALS aggregates. In addition, TDP-43 and FUS proteins – two of the commonly known RNA-binding proteins present in the ALS/FTD pathological aggregates – have both shown G4-binding abilities, reinforcing the idea that G4s might play a key role in aggregation[33 34 35]. Furthermore, it has been shown that the addition of G4-binding ligands to the pathological aggregates affects the progression of the disease in cells[36 37], highlighting again how G4-manipulation is strongly correlated to the formation of aggregates. Despite this, direct evidence of G4-involvement is yet to be provided.

In this study, we assessed whether formation of multimolecular G4s can be key to the emergence of aggregates arising in ALS and FTD $(GGGGCC)_n$ repeats. We extensively characterised the biophysical behaviour of $(GGGGCC)_n$ repeats to assess their ability to form mG4s and aggregates in the absence of proteins. We observed substantial aggregation of the $(GGGGCC)_n$ repeats, which displayed a clear dependency on both repeat length (n) and oligonucleotide concentration. Next, we demonstrated that formation of mG4s is essential to observe the formation of such the aggregates, as exposure to G4-unfolding conditions led to the prompt disassembly of the insoluble aggregates. These results suggest an alternative mechanism that could be potentially responsible for the formation of pathological aggregates in ALS and FTD, where the C9orf72 transcript aggregates on its own right by forming multimolecular G4s (Figure 1B).

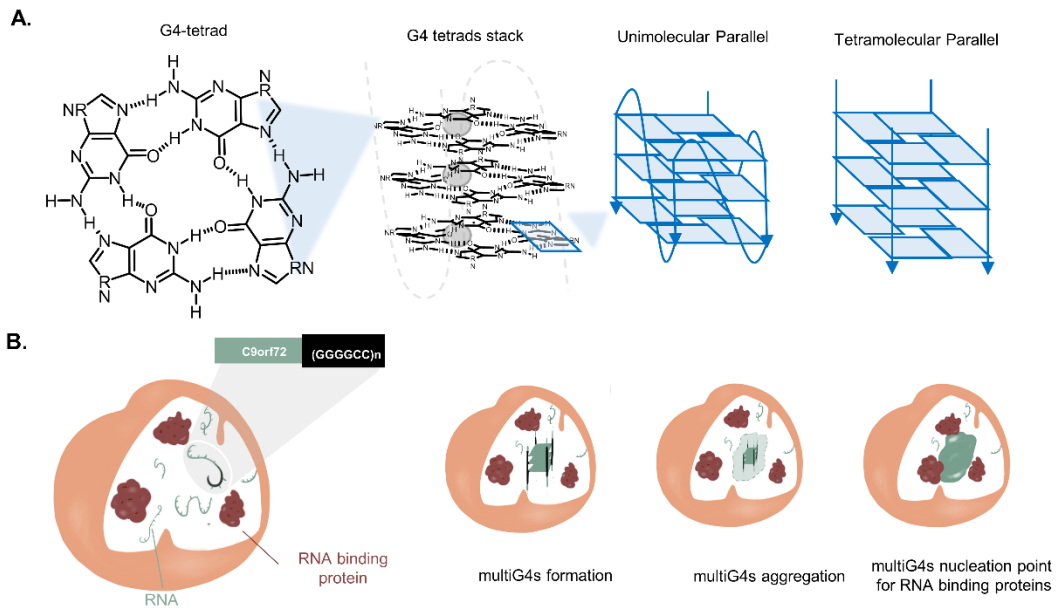


Figure 1: A. G4-tetrad is constituted by 4 Guanines linked together via non-canonical H-bonding interactions. G4-tetrads stack to form a typical G4-structure, which is stabilised by positive metal cations sitting at the centre of the tetrad. G4s can have different molecularities depending on how many strands constitute them. Here we show two molecularity examples: monomolecular and tetramolecular. **B.** Proposed mechanism of aggregation in ALS and FTD pathological aggregates. The $(GGGGCC)_n$ repeat sequence can aggregate in absence of proteins due to the formation of mG4s, creating a nucleation point for the aggregation to take place.

Results

To assess whether mG4s could be the aggregation trigger of the hexanucleotide repeat sequence $(GGGGCC)_n$ in the C9orf72 gene, we set out to investigate two key aspects: i) the ability of the $(GGGGCC)_n$ repeat to form mG4s under physiological conditions; and b) the ability of the $(GGGGCC)_n$ repeat to aggregate in the absence of proteins under conditions that favour mG4-formation, with these structures being the prevalent species detected in the aggregates.

$(GGGGCC)_n$ repeat forms mG4s

Firstly, we asked if the $(GGGGCC)_n$ repeat could form multimolecular G4-structures under physiological conditions. Although RNA is considered the toxic species in ALS and FTD pathological aggregates, DNA equivalents were initially tested for biophysical studies, as they were more easily synthesised and handled[38]. To achieve this, we have annealed the DNA $(GGGGCC)_n$ repeat of variable lengths (n=1 to 12) in a K^+ containing buffer to promote G4-formation. As multimolecular structures are favoured by crowding conditions and longer annealing times, the mG4s annealing protocol (Methods 1) was designed by including 30% PEG as crowding agent and a 0.02°C/min cooling rate (from 95°C to 25°C).

Circular dichroism (CD) spectroscopy was initially performed to globally assess the formation of G4-structures (Methods 2). CD spectra of the $(GGGGCC)_n$ repeats annealed in K^+ presented the characteristic negative peak at ~240 nm and the positive peak at ~263 nm that are typical of a parallel G4-topology (Figure 2A). Agarose gel was then employed to investigate if multimolecular G4s were formed (Methods 3, Figure 2B). Although the gel resolution in Figure 2B does not allow to visualise distinct molecular species, it is evident that samples n>6 present significant smearing, due to the formation higher molecularity species.

The combination of CD and gel electrophoresis results demonstrates the ability of $(GGGGCC)_n$ sequences to fold into parallel G4s that can adopt a multimolecular stoichiometry. To further demonstrate that the multimolecular species observed in the agarose gel could be ascribed to mG4s, we performed a KCl titration experiment, as G4-formation is highly dependent on the concentration of K^+. Agarose gel analysis of samples annealed at increasing KCl concentration revealed more pronounced formation of higher molecularity bands in a K^+ dependent manner, suggesting that their formation is likely mediated by G4s (Figure 2D). This result also revealed that the higher molecularity species cannot form under G4-unfavouring conditions (i.e. low K^+ concentrations).

To further confirm that the multimeric bands observed by agarose gels were mG4s, we also stained the samples with N-methyl meso porphyrin IX (NMM), a G4-specific fluorescent dye[39,40,41]. NMM is expected to fluoresce upon binding to the bands containing G4s but not in the non-G4 DNA control. Indeed, as displayed in Figure 2C, NMM stains selectively all the $(GGGGCC)_n$ bands, indicating that the higher molecularity species formed by the repeat sequence are composed of mG4s.

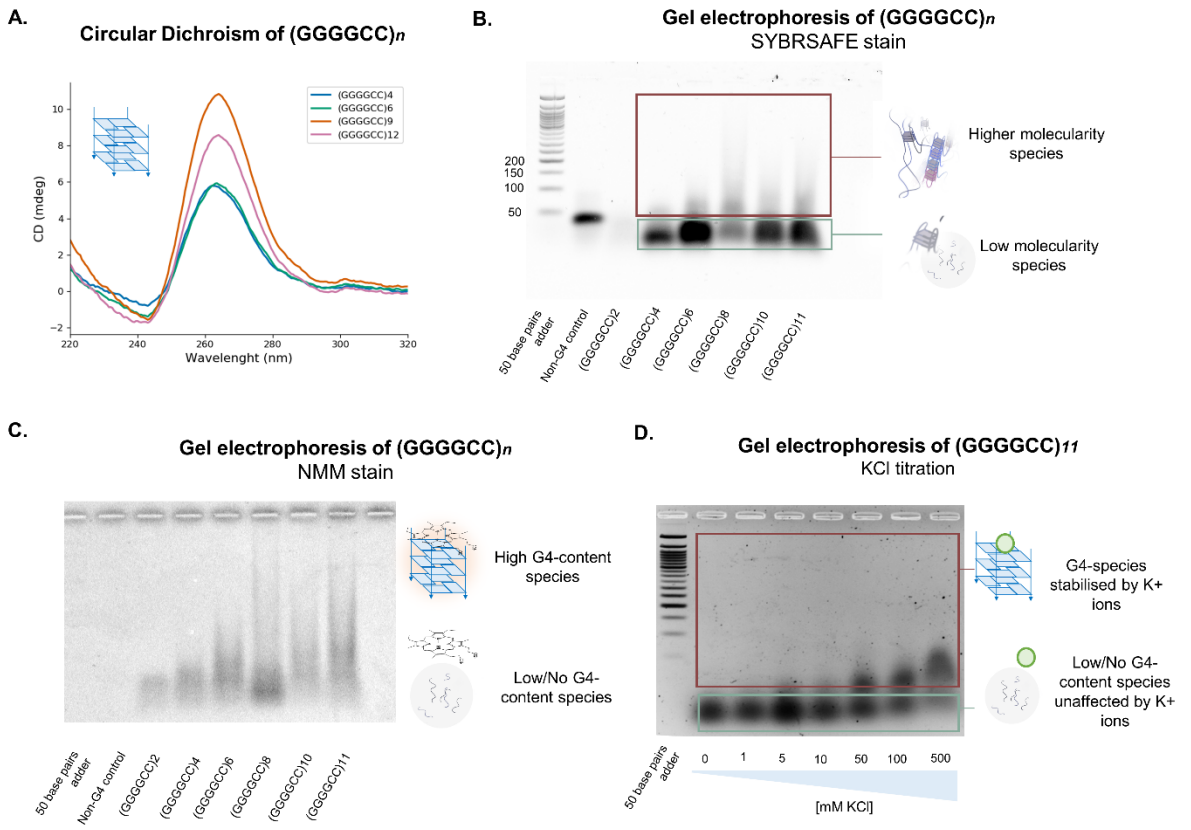


Figure 2: DNA (GGGGCC)$_n$ forms mG4s at different repeat lengths. A. (GGGGCC)$_n$ CD spectra. (GGGGCC)$_n$ (n=4,6,9,12) were annealed at relatively 500 µM, 250 µM, 100 µM and 50 µM in mG4-forming conditions. The annealing concentrations were chosen as to reflect a concentration range were the sequence of that repeat length was shown to aggregate in vitro. All the samples present a positive peak at about 264 nm and a negative peak at 240 nm, both associated to a parallel G4 conformation. **B. (GGGGCC)$_n$ agarose gel electrophoresis – SYBRSAFE stain.** (GGGGCC)$_n$ (n = 2-11) were annealed at 250 µM in mG4-forming conditions. **C. (GGGGCC)$_n$ agarose gel electrophoresis – NMM stain.** (GGGGCC)$_n$ (n = 2-11) were annealed at 250µM in mG4-forming conditions. The lanes containing any of the (GGGGCC)$_n$ were fluorescent upon staining with NMM, implying the presence of G4s in the higher molecularity species formed under these conditions. Both ladder and the non-G4 ssDNA control do not appear, confirming the specificity of the dye for G4-containing species. **D. (GGGGCC)$_{11}$ agarose gel electrophoresis.** (GGGGCC)$_{11}$ was annealed in presence of different [KCl]. As [KCl] increases, so does the molecularity of the species involved, indicating that the intermolecular structures forming have a strong KCl dependence.

mG4s lead to DNA-aggregation in absence of proteins

Having demonstrated the formation of multimolecular constructs in samples of DNA (GGGGCC)$_n$, whose size increases with [K^+] and repeat length, we proceeded to systematically screen for the formation of macroscopic aggregates. For this purpose, we used confocal microscopy (Methods 4) and assessed the formation of aggregates as a function of repeat length and annealing concentration. Since from clinical observations higher repeat lengths of (GGGGCC)$_n$ are associated with the presence of pathological aggregates[16], we hypothesized that a nucleic-acid driven aggregation pattern could be dependent on the repeat length under mG4-forming conditions. Indeed, confocal microscopy revealed significant formation of DNA-aggregates, which was more pronounced for longer repeats lengths (Figure 3A). From a physical point of view, the aggregates presented an amorphous morphology (Figure 3A, 3B,3D, S2A) and showed no recovery after photobleaching (FRAP), typical of a solid or gel-like state (Figure S2B).

To systematically investigate the relationship between aggregate formation and repeat length and concentration, we have generated a phase diagram (Figure 3B), highlighting a clear length and annealing concentration dependence, showing that lower nucleic acid concentrations are required for macroscopic aggregation with increasing length. We next sought to confirm whether the aggregates were constituted primarily by mG4s. To achieve this, we performed NMM staining as on the aggregate formed by the $(GGGGCC)_6$ repeat. As predicted, the aggregate displayed a detectable NMM-fluorescence signal, indicating the presence of G4s (Figure 3D). The selectivity of NMM-fluorescence for G4s aggregates over canonical dsDNA was assessed by staining a dsDNA condensate (formed by 4-pointed DNA nanostars – S3) under the same conditions. As expected, no fluorescence is detected in nucleic acids condensates that are not constituted of G4s (Figure S3B).

We also performed NMM photo-oxidation to confirm that G4s are the key structural species forming the aggregate (Method 4). NMM binds to the G4s by end-stacking interactions and, upon irradiation with a high-power laser, it allows the guanines to photo-oxidise, leading to the disassembling of the G4-structure. This phenomenon has been previously observed to not occur in the case of dsDNA, making it a very G4-specific disassembly experiment[42] and has been reported to lead to disassembly of mG4s in synthetic condensates formed from engineered DNA nanostructures[43]. As expected, the $(GGGGCC)_6$ NMM-stained aggregates disassemble immediately upon irradiation, proving that G4s are the prevalent structure holding the aggregate together (Figure 3E). Once again, we tested the selectivity on the assay on stained DNA-nanostars, which did not show any significant change upon irradiation (Figure S3C).

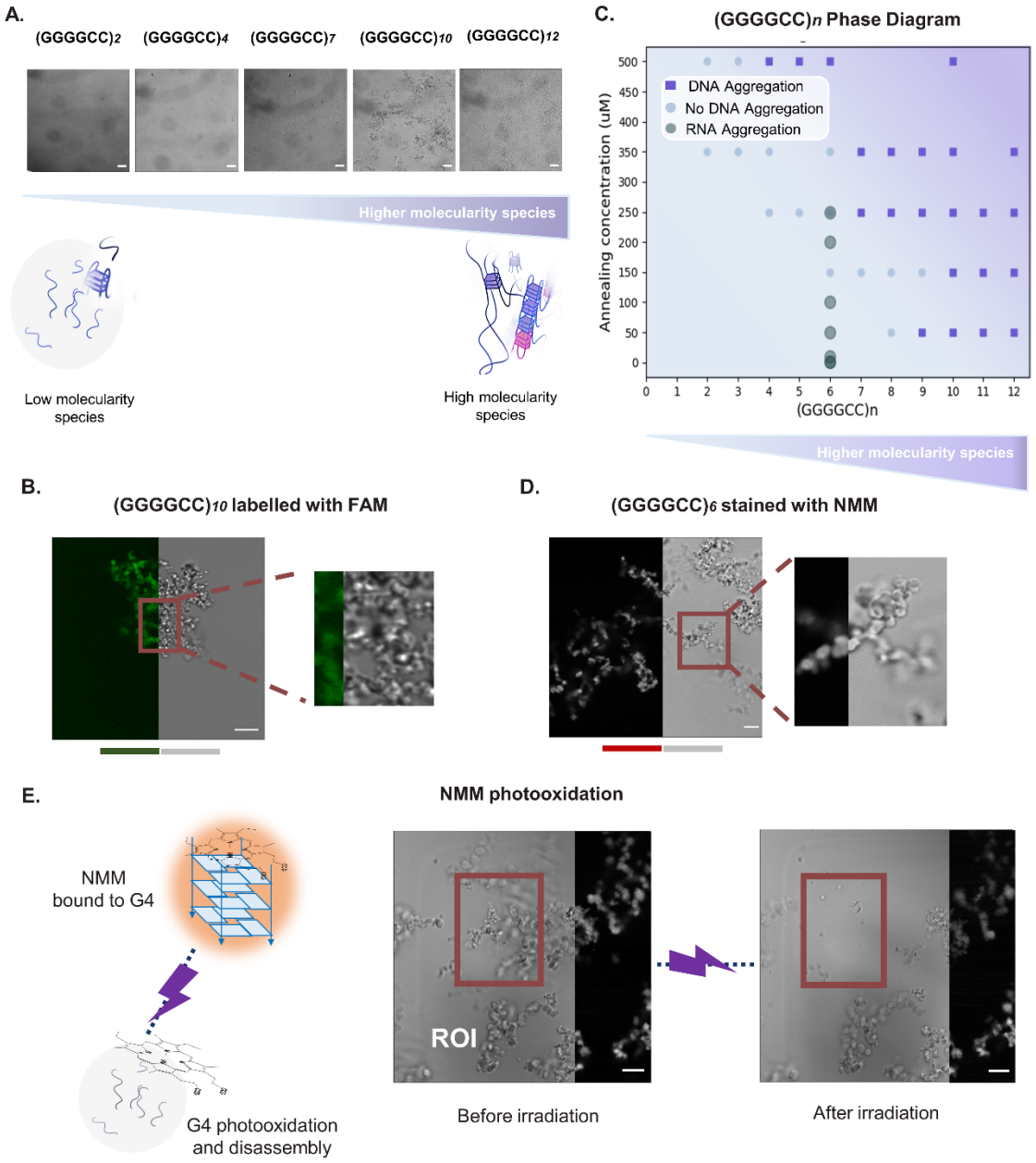


Figure 3: DNA $(GGGGCC)_n$ aggregates with a G4-dependant trend. A. (GGGGCC)n aggregates. Confocal imaging in the brightfield channel of (GGGGCC) (n = 2-11) annealed in multiG4s forming conditions at 250 µM. 10 µm Scalebar. **B. $(GGGGCC)_{10}$ FAM labelled–** $(GGGGCC)_{10}$ was annealed at 500 µM in mG4s forming and stained with NMM. The image is split to show the brightfield channel (on the right) and the FAM (on the left). **C. (GGGGCC)n phase diagram in KCl** – $(GGGGCC)_n$ was annealed in mG4s forming conditions. At lower annealing concentrations and lower repeat lengths, the sequence does not present the ability to aggregate. The higher the repeat length and the higher the concentration, the more significant the aggregation. **D. $(GGGGCC)_6$ NMM staining experiment –** $(GGGGCC)_6$ was annealed at mG4s forming conditions and stained with NMM. The image is split to show the brightfield channel (on the right) and the NMM fluorescence channel (on the left). **E. $(GGGGCC)_6$ NMM photooxidation experiment.** Upon laser excitation the NMM dye bound to the mG4s photo-oxidises the guanine in the quadruplex, leading to the disassembly of the structure. The region of interested highlighted in the figure (ROI) was irradiated for 1 minute and the second image was subsequently taken. The aggregate in the ROI disassembled due to photooxidation.10 µm scalebar.

To further validate that the mechanism of action employed by the $(GGGGCC)_n$ to aggregates relies on the formation of mG4s, we have characterised the aggregation potential of a series of mutated sequences (Figure 4A).

In fact, in addition to G4-formation it is possible that GC base pairing may play a role in stabilising the observed aggregates. To test this hypothesis, we characterised the $(CCCCGG)_6$ repeat, as it allows to retain the same number of canonical base pairing interactions to the ALS/FTD repeat but is unable to form G4s. CD was initially performed to confirm the absence G4s formation and was further validated by means of NMM staining of these sequences loaded on agarose gels (Figures 4B, C and D). Mutating the C to A in the $(GGGGAA)_6$ repeat preserves G4-forming ability but prevents formation of canonical base-pairing. CD of this repeat revealed a typical parallel-G4 trace and NMM staining confirmed G4 formation (Figure 4B and C). Similarly, $(GGGGTT)_6$ maintains the ability to form G4s preventing base-pairing, but due to G-T stacking interactions it is expected to favour a unimolecular G4s structure over multimolecular ones[44]. Indeed, CD and NMM staining revealed formation of stable G4s for this repeat (Figure 4C and 4D), but crucially only low molecular species were detected when running the $(GGGGTT)_6$ repeat in agarose gel.

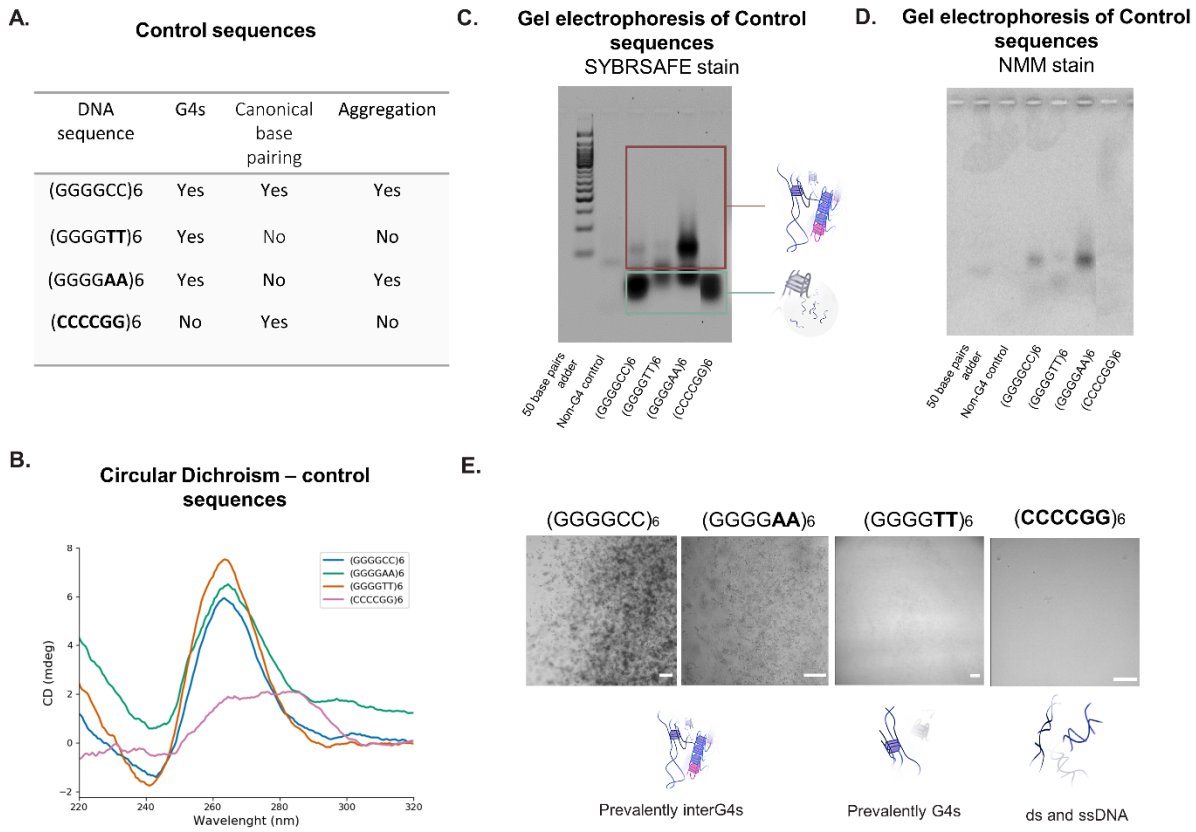


Figure 4: Role of base pairing and monomolecular G4s in the aggregation mechanism. A. Table of controls. B. Circular Dichroism spectra of Controls. Controls of n = 6 were annealed at in mG4-forming conditions. Only the predicted G4 forming sequences present the characteristic peaks at >260 nm and negative 240 nm. **C. Controls agarose gel electrophoresis – SYBRSAFE stain.** Controls of n = 6 were annealed in mG4-forming conditions. Sequences able to form G4s ($(GGGGAA)_6$, $(GGGGTT)_6$, $(GGGGCC)_6$) present a band at higher molecularity due to mG4s. **D. $(GGGGCC)_n$ agarose gel electrophoresis – NMM stain.** Controls of n = 6 were annealed in mG4-forming conditions. All G4-forming sequences ($(GGGGAA)_6$, $(GGGGTT)_6$, $(GGGGCC)_6$) presented fluorescence upon staining, confirming the presence of the G4. Both ladder, non-G4 ssDNA control and $CCCCGG_6$ do not appear, confirming the specificity of the dye for G4-containing species. **E. Controls imaging.** Controls of n = 6 were annealed in mG4-forming conditions. Aggregates were only clearly distinguishable for $(GGGGAA)_6$ and $(GGGGCC)_6$ – the only sequences forming mG4s. 100 µm scalebar.

We then assessed the aggregation properties of these mutant sequences. Aggregation was not observed for $(CCCCGG)_6$ sequence, suggesting that canonical base pairing interactions were not driving the aggregation phenomenon. Similarly, no significant aggregation was detected for $(GGGGTT)_6$, which

arranges preferentially in unimolecular G4 species, suggesting that it is not G4-formation *per se* that lead to aggregation, but the multimolecular network generated by G-G base pairing in mG4-structures. Indeed, $(GGGGAA)_6$ and $(GGGGCC)_6$ were the only two sequences able to form preferentially mG4s and for which aggregation was visibly observed, demonstrating that such structures are a necessary feature to form complex multimeric aggregates (Figure 4E).

RNA repeats aggregate at lower concentrations compared to DNA

As previously mentioned, RNA is considered the toxic species in ALS and FTD pathological aggregates, but DNA equivalents were initially tested for biophysical studies. It is, however, important to highlight some key differences between G4s formed by the two types of nucleic acids. RNA G4s (rG4s) tend to arrange preferentially into a parallel conformation and to be more thermodynamically stable than their equivalent DNA G4s. They also possess a more flexible backbone due to the absence of the extra -OH group, which could have implications on their tendency to form multimolecular structures. For these reasons, aggregation due to mG4s in r(GGGGCC)n is expected to occur at lower RNA concentrations, but once formed the aggregates should possess similar physical properties.

Therefore, we sought to confirm the RNA ability to aggregate in a similar manner to its DNA equivalents by subjecting $r(GGGGCC)_6$ to the same mG4-forming conditions applied for the DNA repeats (Methods 1). As expected, aggregation of $r(GGGGCC)_6$ was observed in a range of concentration going from 250 µM to 500 nM, forming in all cases amorphous aggregates similar in morphology to DNA $(GGGGCC)_6$, which stained with NMM confirming mG4-formation (Figures 5A). As anticipated, RNA required a far lower concentration of annealing than DNA to observe aggregation, reflecting the known ability of RNA to form more stable G4s compared to DNA. To demonstrate that mG4s were responsible for the aggregation observed, the RNA aggregates stained with NMM were subjected to irradiation, leading to immediate disassembly of the aggregates (Figure 5B). These results confirmed that $(GGGGCC)_n$ aggregation is a mG4-dependant phenomenon and will occur in a similar manner for RNA and DNA equivalents.

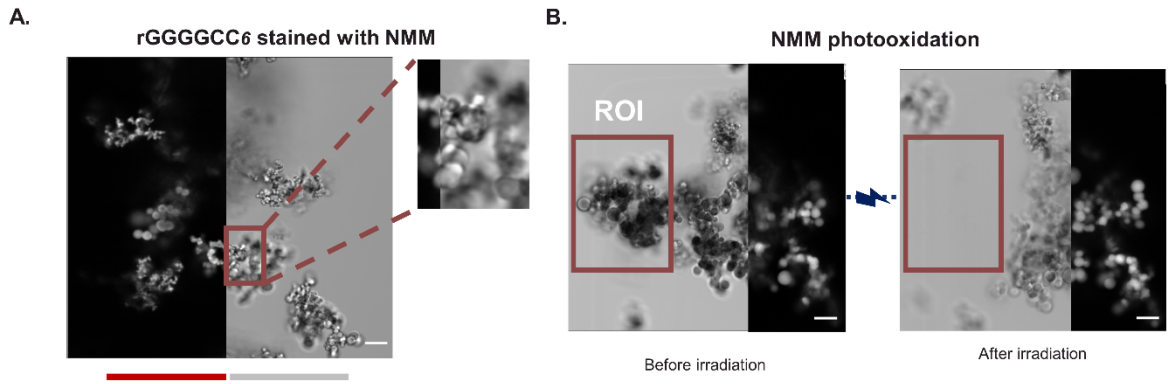


Figure 5: RNA $(GGGGCC)_n$ aggregates with a G4-dependant trend. A. $r(GGGGCC)_6$ NMM staining experiment – $r(GGGGCC)_6$ was annealed in mG4s forming conditions. It was then subjected to NMM staining prior to acquisition. The image is split to show the brightfield channel (on the right) and the NMM fluorescence channel (on the left). **B. $r(GGGGCC)_6$ NMM photooxidation experiment -** Upon laser excitation the NMM dye bound to the mG4s photo-oxidises the guanine in the quadruplex, leading to the disassembly of the structure. The region of interested highlighted in the figure (ROI) was irradiated for 1 minute and the second image was subsequently taken. The aggregate in the ROI disassembled due to photooxidation. 10 µm scalebar.

Discussion

Although G4s are a widely recognised target for many diseases ranging from cancer and Cockayne Syndrome to Fragile X syndrome[26], the biological relevance of mG4s has only recently been put into question[45]. Indeed, there is growing evidence of how of G4-formation could aid long-range interactions at enhancer-promoters regions in cells[46] and they could promote formation of phase-separated aggregates in membrane-less organelles[19 45]. Furthermore, the discovery of an endogenous human protein that selectively binds mG4s over unimolecular structures, strongly suggest that these multimeric species can be formed in cells[28].

Although unimolecular species form more readily and, therefore, have a kinetic advantage, mG4s are usually more thermodynamically stable in cell crowding conditions[47]. But despite the uprising biological role of mG4s is now part of a discussion, there is no specific tool to test for their presence in vitro or in vivo. Much like when unimolecular G4s where first discovered and there were no specific probes that would separate it from dsDNA in live cells, currently all G4-specific probes (such as PDS[48], PhenDC3[49], TMPyP4[50], NMM[41]) have not been thoroughly studied in the context of mG4s, making differentiation in live-cell experiment impossible. Even in simulations, it is more challenging to predict the formation of multimolecular species owing to the greater computational burden. In fact, prediction of mG4s is currently only possible based on pre-existing X-ray crystallography data[51].

In this study we show how $(GGGGCC)_n$ can arrange into mG4s using a combination of biophysical and biochemical strategies: CD proved the presence of G4s in the samples, agarose gel confirmed the formation of multimolecular species and both NMM staining and [KCl] titration performed on the gel gave further evidence that such species were due to the presence of mG4s over other higher molecularity interactions. We then observed a direct correlation between mG4s formation and the ability of $(GGGGCC)_n$ to aggregate in vitro. Since the pathological aggregate formation in ALS and FTD is strongly correlated with the presence of $(GGGGCC)_n$ at high repeat numbers, we sought to investigate the relationship between our nucleic acid-driven aggregation and repeat length itself. Importantly, we were able to show that higher repeat lengths require lower concentrations for aggregation to occur, which allowed us to draw a parallel with the disease itself. The repeat lengths used for this study are significantly below the recognised pathological threshold (n=24)[16], which is paraphs why we could only observe aggregation in the micromolar range. However, given the trend of the data obtained it is possible to infer that higher repeat lengths would require lower and lower concentrations to see aggregation, potentially reaching biological relevant ranges (albeit variable, a human cell contains about 6 pg of DNA[52]).

The aggregates obtained were amorphous and solid appearance. Although pathological aggregates in ALS and FTD are often referred to as "liquid-like"[53 54 55], it is reasonable to assume that while the nucleic-acid only aggregate might be solid, the additional protein-component in ALS/FTD-affected cells could modify the final physical state. In fact, two of the main proteins tied with the disorder (TDP-43 and FUS) have shown significant G4-binding abilities, suggesting that paraphs the protein could act in conjunction with G4-triggered aggregation in a live cell environment.

The aggregates responded to staining with the G4-specific dye NMM and were successfully disassembled via guanine photooxidation when NMM was irradiated, which could be further explored as an approach to disassemble the pathological aggregates in live cells. As a proof of concept, the biological relevance of our study was assessed by testing the equivalent $r(GGGGCC)_6$ sequence, successfully confirming the G4-driven aggregation behaviour.

In summary, this interpretation offers an alternative mechanism of action for ALS/FTD diseases which sees mG4s, a new biologically relevant target, as aggregation triggers. Currently, the only two FDA-approved drugs that have shown an increased survival rate in randomised ALS clinical trials are Riluzole and Edaravone[56 57], but ALS and FTD remain deadly diseases. So far, candidate drugs have been developed to target protein–lead aggregation but have shown only partial effectiveness[58],

suggesting the need to explore alternative approaches. A recent study has shown potential for ALS treatment in a small molecule with the ability to bind to r(GGGGCC)$_n$, stabilising the intermolecular hairpin conformation formed by the repeat[59]. This is in agreement with our observations and might suggest that depletion of mG4s by stabilisation of intermolecular hairpins could offer a starting point for a new therapeutical pathway for ALS and FTD[59].

Methods

1. (GGGGCC)n mG4s annealing protocol

(GGGGCC)$_n$ (n=1-12) was purchased from IDT in its lyophilised form and diluted in MilliQ water to reach a stock concentration of 1 mM. The stocks were heated to 40 °C for 5 minutes to ensure complete solvation. The DNA samples were then further diluted to the desired annealing concentration (500-0μM) in 10 mM TRIS-HCl buffer pH 7.4 (filtered with 0.2 μm sterile syringe filters) and 30% Poly(ethylene glycol) BioUltra 200 (PEG) was added to the clear solution. The samples were vortexed and subjected to a first denaturing step in a C1000 Touch™ Thermal Cycler (heating to 95 °C for 1h followed by cooling to 25 °C with a temperature ramp of 0.25 °C/min). 500 mM KCl was added to the denatured solution and a further heating cycle was applied (heating to 95 °C for 1h followed by cooling to 25 °C with a temperature ramp of 0.02 °C/min). To note is that after the long annealing procedure, significant evaporation is observed in the samples. The phase diagram was developed by reducing evaporation as much as possible to estimate an accurate concentration. For this purpose the author recommends using an annealing volume of > 30 μL. For other applications where exact concentration is not of interest smaller volumes can also be adopted.

2. Circular Dichroism spectroscopy protocol

Pre-annealed (GGGGCC)$_n$ (n=1-12) DNA samples were diluted to 5μM in 10 mM TRIS-HCl buffer pH 7.4 with a final volume of 150 μL. The samples were transferred in Quartz High Precision cell (1 mm light path) and analysed in the Jasco J-715 Spectropolarimeter. The parameters used for each run were the following: Sensitivity (100 mdeg); Wavelength start and end (320 nm – 220 nm); Data Pitch 0.5 nm with continuous scanning mode and scanning speed of 100 nm/min; Response (2 sec); Band Width (2.0 nm); Accumulation (3).

3. Agarose gel electrophoresis protocol

SYBRSAFE staining

Pre-annealed (GGGGCC)$_n$ (n=1-12) DNA samples were diluted to 10 μM in 10 mM TRIS-HCl buffer pH 7.4 with a final volume of 12 μL. 2 μL of Purple Dye were added to the solution to visualise the samples. The agarose gel matrix was prepared by mixing 1.5 g of agarose in 50 μL 1X TBE buffer and 3 μL SYBRSAFE stain, heated by microwaving and left to polymerise for 20 minutes prior to running. The samples were run at 65 V for 85 mins at room temperature and imaged with Typhoon FLA 9500 on the EtBr channel.

NMM staining

Pre-annealed (GGGGCC)$_n$ (n=1-12) DNA samples were diluted to 10μM in 10 mM TRIS-HCl buffer pH 7.4 with a final volume of 12 μL. 2 μL of 50% glycerol were added to the solution to visualise the samples. The agarose gel matrix was prepared by mixing 1.5 g of agarose in 50 μL 1X TBE buffer, heated by microwaving and left to polymerise for 20 minutes prior to running. The samples were run at 65 V for 85 mins at room temperature. The gel was then incubated in a solution of 50 μL 2mM NMM and 50 mL 1X TBE for 1h covered to avoid photobleaching of the dye. The gel was washed once with 1X TBE for 10 minutes and imaged with Typhoon FLA 9500 on the EtBr channel.

4. Imaging via confocal microscopy

Phase diagram

Leica SP8 Inverted Confocal microscope (HC PL APO CS2 10x/0.40 DRY | HC PL APO CS2 20x/0.75 DRY objectives) was used for the imaging. In order to define aggregation on the phase diagram, pre-annealed $(GGGGCC)_n$ (n=1-12) DNA samples were directly transferred in PDMS wells and imaged by irradiating the sample with the 458 nm excitation laser on the brightfield channel.

FAM-labelled samples

Leica SP8 Inverted Confocal microscope (HC PL APO CS2 10x/0.40 DRY | HC PL APO CS2 20x/0.75 DRY objectives) was used for the imaging. Pre-annealed samples labelled with 2-50% FAM were diluted to 10 µM in 10 mM TRIS-HCl buffer pH 7.4, imaged by irradiating with the 488 nm laser and with an emission filter from 510-550 nm.

NMM-stained samples

Leica SP8 Inverted Confocal microscope (HC PL APO CS2 10x/0.40 DRY | HC PL APO CS2 20x/0.75 DRY objectives) was used for the imaging. Pre-annealed samples were diluted to 10µM in 10 mM TRIS-HCL buffer pH 7.4 and irradiated with the 405 nm laser. 10µM NMM was then added to the sample and left to incubate for 10 minutes in a dark environment. The sample was then imaged on the emission filter 600-650 nm.

NMM-photooxidation

Leica SP8 Inverted Confocal microscope (HC PL APO CS2 10x/0.40 DRY | HC PL APO CS2 20x/0.75 DRY objectives) was used for the imaging. Pre-annealed $GGGGCC_6$ sample was diluted to approximately 10µM in 10 mM TRIS-HCL buffer pH 7.4 and irradiated with the 405 nm laser. 10 µM NMM was then added to the sample and left to incubate for 10 minutes in a dark environment. A portion of the sample was then irradiated at 405 nm at 100% for 1 minute and complete disassembly of that portion of the aggregate was observed both in brightfield or on the emission filter 600-650 nm.

FRAP (Fluorescence recovery after photobleaching)

Leica SP8 Inverted Confocal microscope (HC PL APO CS2 10x/0.40 DRY | HC PL APO CS2 20x/0.75 DRY objectives) was used for the imaging. Pre-annealed samples labelled with 10% FAM were imaged by irradiating with the 488 nm laser and with an emission filter from 510-550 nm. A region of the well was irradiated at 100% power intensity for 30 seconds until no fluorescence was observed. Recovery after photobleaching was monitored every five minutes for 30 minutes.

5. dsDNA nanostars annealing protocol

The samples were prepared by mixing 4 nanostars cores (N4_core1, N4_core2, N4_core3, N4_core4) and two linkers (L_AA1 and L_AA2) in a TrisHCl buffer containing 500 mM KCl (Table 2). The annealing protocol involved heating the samples at 95°C for 30 minutes to then cool down between 85C and 25°C (-0.1°C/min cooling rate). Samples were transferred from a glass capillary to a PDMS well for imaging.

NAME	VOLUME (µL)
TrisHCl pH 7.5	52.3
3M KCL	11.7
A4_CORE1	1.4
A4_CORE2	1.5
A4_CORE3	1.5

A4_CORE4	2.0
L_1	2.3
L_2	2.0

Table 2: Volumes used for the annealing protocol of dsDNA control.

Acknowledgements

Graphics were designed using Procreate design app and ChemDraw.

Funding

This work was supported by the following funding bodies:

M.D.A is supported by Biotechnology and Biological Sciences Research Council (BBSRC) David Phillips Fellowship [BB/R011605/1]; M.D.A. is a recipient of Lister Institute Research Prize that support his group. F.R. is supported by a Leverhulme Trust, Leverhulme Cellular Bionics scholarship [EP/S023518/1]; L.D.M. is supported by Royal Society University Research Fellowship [UF160152, URF\R\221009] and Royal Society Research Fellows Enhanced Research Expenses [RF/ERE/210029], also supporting R.R.S. L.D.M., R.R.S., D.T. and L.M. are supported by the European Research Council (ERC) under the Horizon 2020 Research and Innovation Programme [ERC-STG No 851667 NANOCELL]. Y.E. is supported by United Kingdom Research and Innovation Future Leaders Fellowship [MR/S031537/1].

Facility for Imaging by Light Microscopy (FILM) at Imperial College London is part-supported by funding from the Wellcome Trust [grant 104931/Z/14/Z] and BBSRC [grant BB/L015129/1].

Funding for open access charge: Imperial College London.

References

1. Ferrari, R., Kapogiannis, D., Huey, E. D. & Momeni, P. FTD and ALS: a tale of two diseases. *Curr Alzheimer Res* 8, 273–294 (2011).
2. Zhou, B. *et al.* Characterizations of distinct parallel and antiparallel G-quadruplexes formed by two-repeat ALS and FTD related GGGGCC sequence. *Scientific Reports* 8, 2366 (2018).
3. Gao, F.-B., Almeida, S. & Lopez-Gonzalez, R. Dysregulated molecular pathways in amyotrophic lateral sclerosis–frontotemporal dementia spectrum disorder. *The EMBO Journal* 36, 2931–2950 (2017).
4. Abramzon, Y. A., Fratta, P., Traynor, B. J. & Chia, R. The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Front. Neurosci.* 14, (2020).
5. Chiò, A. *et al.* Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. *Neuroepidemiology* 41, 118–130 (2013).
6. Ratnavalli, E., Brayne, C., Dawson, K. & Hodges, J. R. The prevalence of frontotemporal dementia. *Neurology* 58, 1615–1621 (2002).
7. Amyotrophic Lateral Sclerosis (ALS) Fact Sheet | National Institute of Neurological Disorders and Stroke. https://www.ninds.nih.gov/amyotrophic-lateral-sclerosis-als-fact-sheet.
8. Frontotemporal dementia. *nhs.uk* https://www.nhs.uk/conditions/frontotemporal-dementia/ (2017).
9. Burrell, J. R., Kiernan, M. C., Vucic, S. & Hodges, J. R. Motor neuron dysfunction in frontotemporal dementia. *Brain* 134, 2582–2594 (2011).
10. Louka, A., Zacco, E., Temussi, P. A., Tartaglia, G. G. & Pastore, A. RNA as the stone guest of protein aggregation. *Nucleic Acids Research* 48, 11880–11889 (2020).

11. Silva, J. L. & Cordeiro, Y. The 'Jekyll and Hyde' Actions of Nucleic Acids on the Prion-like Aggregation of Proteins. *J Biol Chem* 291, 15482–15490 (2016).
12. Ling, S.-C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron* 79, 416–438 (2013).
13. Jo, M. *et al.* The role of TDP-43 propagation in neurodegenerative diseases: integrating insights from clinical and experimental studies. *Exp Mol Med* 52, 1652–1662 (2020).
14. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162, 1066–1077 (2015).
15. Proteostatic imbalance and protein spreading in amyotrophic lateral sclerosis. https://www.embopress.org/doi/epdf/10.15252/embj.2020106389 doi:10.15252/embj.2020106389.
16. Iacoangeli, A. *et al.* C9orf72 intermediate expansions of 24–30 repeats are associated with ALS. *Acta Neuropathologica Communications* 7, 115 (2019).
17. Van Mossevelde, S., van der Zee, J., Cruts, M. & Van Broeckhoven, C. Relationship between C9orf72 repeat size and clinical phenotype. *Curr Opin Genet Dev* 44, 117–124 (2017).
18. Blokhuis, A. M., Groen, E. J. N., Koppers, M., van den Berg, L. H. & Pasterkamp, R. J. Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathol* 125, 777–794 (2013).
19. Fay, M. M., Anderson, P. J. & Ivanov, P. ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. *Cell Rep* 21, 3573–3584 (2017).
20. Balendra, R. & Isaacs, A. M. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol* 14, 544–558 (2018).
21. Garcia-Jove Navarro, M. *et al.* RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nature Communications* 10, 3230 (2019).

22. Kovachev, P. S. *et al.* RNA modulates aggregation of the recombinant mammalian prion protein by direct interaction. *Scientific Reports* 9, 12406 (2019).

23. Locatelli, E., Handle, P. H., Likos, C. N., Sciortino, F. & Rovigatti, L. Condensation and Demixing in Solutions of DNA Nanostars and Their Mixtures. *ACS Nano* 11, 2094–2102 (2017).

24. Morya, V., Walia, S., Mandal, B. B., Ghoroi, C. & Bhatia, D. Functional DNA Based Hydrogels: Development, Properties and Biological Applications. *ACS Biomater. Sci. Eng.* 6, 6021–6035 (2020).

25. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* 546, 243–247 (2017).

26. Raguseo, F., Chowdhury, S., Minard, A. & Antonio, M. D. Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem. Commun.* 56, 1317–1324 (2020).

27. Paulson, H. Repeat expansion diseases. *Handb Clin Neurol* 147, 105–123 (2018).

28. Liano, D., Chowdhury, S. & Di Antonio, M. Cockayne Syndrome B Protein Selectively Resolves and Interact with Intermolecular DNA G-Quadruplex Structures. *J. Am. Chem. Soc.* 143, 20988–21002 (2021).

29. Davis, J. T. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl* 43, 668–698 (2004).

30. Frasson, I., Pirota, V., Richter, S. N. & Doria, F. Multimeric G-quadruplexes: A review on their biological roles and targeting. *International Journal of Biological Macromolecules* 204, 89–102 (2022).

31. Reddy, K., Zamiri, B., Stanley, S. Y. R., Macgregor, R. B. & Pearson, C. E. The Disease-associated r(GGGGCC)n Repeat from the C9orf72 Gene Forms Tract Length-dependent

Uni- and Multimolecular RNA G-quadruplex Structures*. *Journal of Biological Chemistry* 288, 9860–9866 (2013).

32. Zhang, Y. *et al.* G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Research* 47, 11746–11754 (2019).
33. Frontiers | FUS Recognizes G Quadruplex Structures Within Neuronal mRNAs. https://www.frontiersin.org/articles/10.3389/fmolb.2020.00006/full.
34. Ishiguro, A., Kimura, N., Watanabe, Y., Watanabe, S. & Ishihama, A. TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. *Genes Cells* 21, 466–481 (2016).
35. Ishiguro, A., Katayama, A. & Ishihama, A. Different recognition modes of G-quadruplex RNA between two ALS/FTLD-linked proteins TDP-43 and FUS. *FEBS Lett* 595, 310–323 (2021).
36. Simone, R. *et al.* G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Molecular Medicine* 10, 22–31 (2018).
37. Zamiri, B., Reddy, K., Macgregor, R. B. & Pearson, C. E. TMPyP4 Porphyrin Distorts RNA G-quadruplex Structures of the Disease-associated r(GGGGCC)n Repeat of the C9orf72 Gene and Blocks Interaction of RNA-binding Proteins*. *Journal of Biological Chemistry* 289, 4653–4659 (2014).
38. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-quadruplexes in biology: principles and molecular mechanisms. *J Mol Biol* 429, 2127–2147 (2017).
39. Xu, S. *et al.* Fluorescent G-quadruplex–NMM DNA probe for the detection of silver nanoparticles in aqueous media. *Anal. Methods* 7, 1672–1675 (2015).
40. Yett, A., Lin, L. Y., Beseiso, D., Miao, J. & Yatsunyk, L. A. N-methyl mesoporphyrin IX as a highly selective light-up probe for G-quadruplex DNA. *J Porphyr Phthalocyanines* 23, 1195–1215 (2019).

41. Sabharwal, N. C. *et al.* N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J* 281, 1726–1737 (2014).
42. McBrayer, D., Schoonover, M., Long, K. J., Escobedo, R. & Kerwin, S. M. N-Methylmesoporphyrin IX Exhibits G-Quadruplex-Specific Photocleavage Activity. *ChemBioChem* 20, 1924–1927 (2019).
43. Fabrini, G., Minard, A., Brady, R. A., Di Antonio, M. & Di Michele, L. Cation-Responsive and Photocleavable Hydrogels from Noncanonical Amphiphilic DNA Nanostructures. *Nano Lett.* 22, 602–611 (2022).
44. Lago, S., Tosoni, E., Nadai, M., Palumbo, M. & Richter, S. N. The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1861, 1371–1381 (2017).
45. Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D. & Di Antonio, M. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Research* 49, 8419–8431 (2021).
46. Li, L. *et al.* YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* 17, 161–168 (2021).
47. Merkina, E. E. & Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophysical Journal* 89, 365–373 (2005).
48. Koirala, D. *et al.* A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nature Chemistry* 3, 782–787 (2011).
49. Molnár, O. R., Végh, A., Somkuti, J. & Smeller, L. Characterization of a G-quadruplex from hepatitis B virus and its stabilization by binding TMPyP4, BRACO19 and PhenDC3. *Sci Rep* 11, 23243 (2021).

50. Carvalho, J., Queiroz, J. A. & Cruz, C. Circular Dichroism of G-Quadruplex: a Laboratory Experiment for the Study of Topology and Ligand Binding. *J. Chem. Educ.* 94, 1547–1551 (2017).
51. Cagirici, H. B., Budak, H. & Sen, T. Z. G4Boost: a machine learning-based tool for quadruplex identification and stability prediction. *BMC Bioinformatics* 23, 240 (2022).
52. Piovesan, A. *et al.* On the length, weight and GC content of the human genome. *BMC Res Notes* 12, 106 (2019).
53. Carey, J. L. & Guo, L. Liquid-Liquid Phase Separation of TDP-43 and FUS in Physiology and Pathology of Neurodegenerative Diseases. *Frontiers in Molecular Biosciences* 9, (2022).
54. Farrawell, N. E. *et al.* Distinct partitioning of ALS associated TDP-43, FUS and SOD1 mutants into cellular inclusions. *Sci Rep* 5, 13416 (2015).
55. Koehler, L. C. *et al.* TDP-43 Oligomerization and Phase Separation Properties Are Necessary for Autoregulation. *Frontiers in Neuroscience* 16, (2022).
56. Fang, T. *et al.* Stage at which riluzole treatment prolongs survival in patients with amyotrophic lateral sclerosis: a retrospective analysis of data from a dose-ranging study. *Lancet Neurol* 17, 416–422 (2018).
57. Mitsubishi Tanabe Pharma Development America, Inc. *A Phase 3b, Multicenter, Randomized, Double-Blind Study to Evaluate Efficacy and Safety of Oral Edaravone Administered for a Period of 48 Weeks in Subjects With Amyotrophic Lateral Sclerosis (ALS)*. https://clinicaltrials.gov/ct2/show/NCT04569084 (2022).
58. Malik, R. & Wiedau, M. Therapeutic Approaches Targeting Protein Aggregation in Amyotrophic Lateral Sclerosis. *Front Mol Neurosci* 13, 98 (2020).

59. Bush, J. A. *et al.* A blood–brain penetrant RNA-targeted small molecule triggers elimination of r(G4C2)exp in c9ALS/FTD via the nuclear RNA exosome. *Proceedings of the National Academy of Sciences* 119, e2210532119 (2022).