The C9Orf72 hexanucleotide repeat expansion aggregates by means of multimolecular G-quadruplex formation.

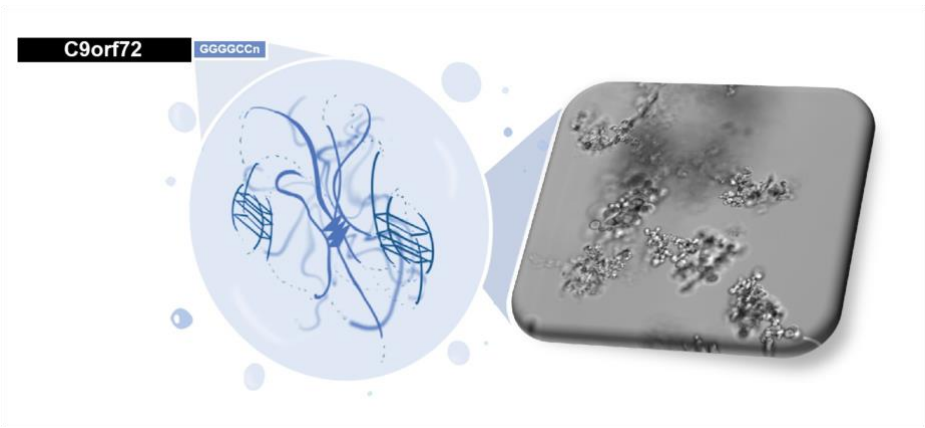
Federica Raguseo[1 2], Diana A. Tanase[1 2], Thomas E. Maher[1], Layla Malouf[1 2], Roger Rubio-Sánchez[2], Yuval Elani[3], Lorenzo Di Michele[1 2]*, Marco Di Antonio.[1 4] *

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. Tel: +44 (0)20 7594 5866; Email: m.di-antonio@imperial.ac.uk (MDA) and ld389@cam.ac.uk (LDM).

Abstract

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are two neurodegenerative diseases genetically characterised by the expansion of the hexanucleotide repeat $(GGGGCC)_n$. Here we investigated the formation of nucleic-acid secondary structures in the expansion repeats, and their role in generating aggregates that are characteristic of ALS/FDT. We observed significant aggregation of DNA $(GGGGCC)_n$, which we ascribe to the formation of multimolecular G-quadruplexes (G4s), as confirmed with G4-selective fluorescent staining. Exposing the aggregates to G4-unfolding conditions, such as photo-oxidation of guanines, led to prompt disassembly, suggesting an essential role of G4s in the aggregation process. We then characterised aggregation in RNA $(GGGGCC)_n$ repeats, which are directly relevant to ALS/FTD pathology. We observed that the RNA repeats aggregate at significantly lower concentrations compared to DNA, strengthening the hypothesis that nucleic acids may play a structural role in pathological aggregation at physiological concentrations. Our findings strongly suggest that RNA G-rich repetitive sequences can form protein-free aggregates sustained by multimolecular G-quadruplexes, highlighting their potential relevance as therapeutic targets for ALS and FTD.



Introduction

ALS and FTD are two of the most common types of fatal dementia in adults above > 65 y.o[1] with the first being found in approximately 2.1 cases per 100,000 people[2], and the second having an incidence of 4 cases per 100,000[3]. Although they appear to have different symptoms in their initial stages (ALS impairing movement[4] and FTD a patient's mental abilities[5]), both diseases are considered as a broad neurodegenerative continuum[6]. The link between the two is not only recognised from clinical observations[7], but also from overlapping in their genetics[1]. ALS and FTD, much like other neurodegenerative diseases, are characterised by the presence of pathological aggregates at the neuron level, containing a mix of proteins and nucleic acids[6,8]. Many of the existing studies have focused on the protein component as the leading aggregation trigger[9], often dismissing the role of nucleic acids as an element of additional stabilisation in the aggregation process[10]. 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[11,12,13,14]. In turn, recruitment of RNA in the aggregates has been regarded as a secondary step. However, despite the mainstream hypothesis proposing that proteins are 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% of familial cases (where the patient had a regressed family history of ALS) and 7% of the sporadic (no previous indication of family history of ALS) in Europe[15]. In fact, the mutation in the gene itself has been previously proposed to lead to the formation of aggregates *via* various mechanisms[16,17], including ones in which the RNA transcribed from the expansion repeat plays a structural role. However, a clear mechanism to support this hypothesis is yet to be described[18,19].

The hypothesis that the RNA expansion $(GGGGCC)_n$ could trigger aggregation in its own right in absence of additional proteins has become increasingly supported in recent years[8,20,21], given that nucleic acids do not need proteins to form biomolecular condensates[22]. Although it is true that endogenous nucleic acid sequences do not typically lead to aggregation, they have already shown the potential to form multimeric networks by both canonical and non-canonical base pairing interactions[23]. 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[24].

$r(GGGGCC)_n$, in particular, has previously been shown to arrange into hairpin and G-quadruplex (G4) structures, both of which have exhibited direct involvement in the disease progression by reducing aggregation upon treatment with small-molecule probes that can bind to, and stabilise these secondary structures[25,26,27,28,29]. In particular, the use of ligands to bind G4s has been shown to ameliorate ALS phenotypes in neuronal cells[28], whilst targeting of the hairpin affected the correct splicing of the C9Orf72 gene and consequently the disease development[25].

However, whilst multiple studies have shown that targeting RNA secondary structures formed by $r(GGGGCC)_n$ can have a significant effect on the biology of the disease, none have fully clarified the role of the nucleic acid structures in the regulation of pathological aggregation, which would be essential to devise therapeutic agents. Furthermore, there seems to be an apparent contradiction upon which secondary structure, G-quadruplex or hairpin, is relevant to the disease progression, and how.

We reasoned that if pathological aggregation in ALS and FTD is driven by RNA, given the G-rich nature of this repeat and the previously reported effect of G4-binding probes, multimolecular G-quadruplexes (mG4s) are likely candidates to provide the multi-molecular linkages required to stabilise the aggregates. Indeed, the non-canonical G-G base pairing underpinning G4-formation can readily occur under physiological conditions[30,31,32] (Figure 1A).

G4-formation has already been associated with cancer, neurodegenerative diseases (ALS and FTD) and rare-genetic diseases (Fragile X syndrome and Cockayne Syndrome)[33,34]. The hexanucleotide repeat $(GGGGCC)_n$ in C9orf72 has been shown to form G4s both in its RNA and DNA form[35,36,37] and the formation of these structures has been detected in cellular ALS aggregates[26]. 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[38,39,40]. Furthermore, the hypothesis that mG4s are key to pathological aggregation is consistent with the concept that targeting the hairpin conformation would slow down the progression of the disease, as locking the RNA in the hairpin conformation would prevent formation of mG4s. Additionally, treatment with G4-ligands could promote formation of unimolecular G4s, also preventing formation of mG4s and, thus, potentially justifying the reduction of aggregate size upon treatment with G4-ligands.

In this study we aimed to shed light on the nature of the multimolecular structures formed by repeat $(GGGGCC)_n$ and assessed their ability to form aggregates in absence of proteins. Specifically, we explored whether the formation of mG4s can lead to the emergence of aggregates in ALS and FTD $(GGGGCC)_n$ repeats. We extensively characterised the biophysical behaviour of $(GGGGCC)_n$ repeats to determine their ability to form mG4s and aggregates. We observed substantial aggregation, which depends on both repeat length *n* and oligonucleotide concentration. Next, we demonstrated that mG4s are structurally essential to the aggregates, as exposure to G4-unfolding conditions led to prompt aggregate disassembly. Consistently, we observed that aggregation is hindered by G4-ligand Pyridostatin (PDS), known to favour formation of unimolecular over multimolecular G4s, and by mutation of the repeat sequence. 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 in its own right by means of mG4s formation (Figure 1B).

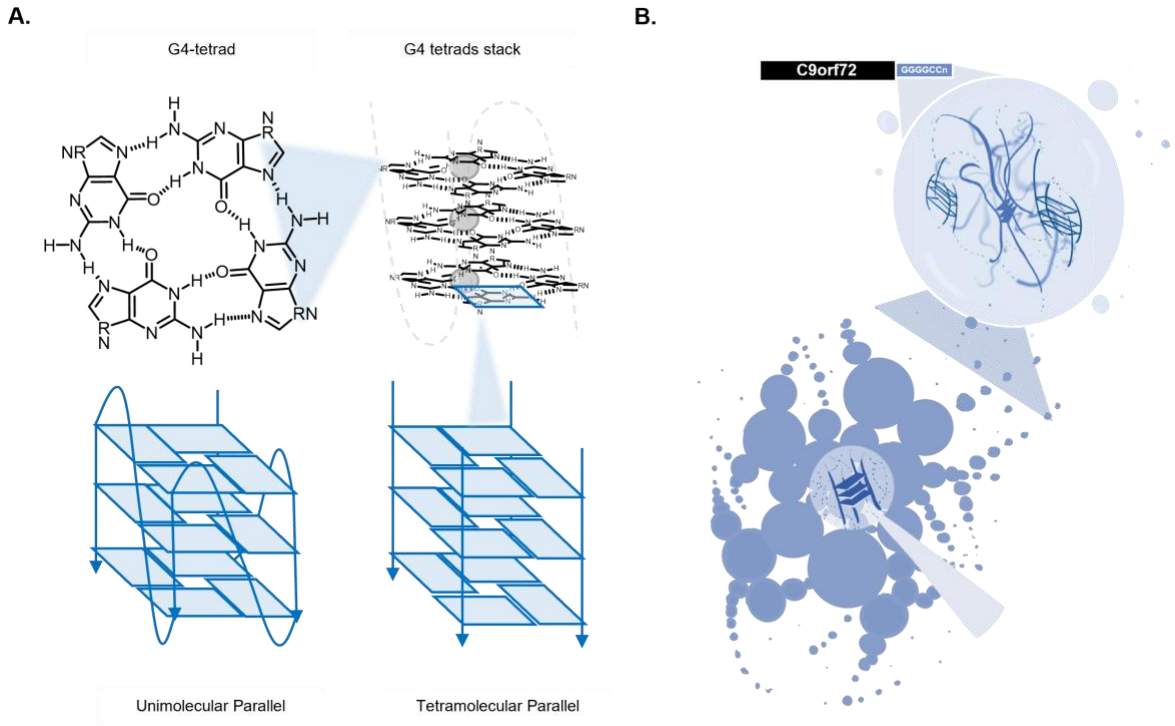


Figure 1: A. Characteristics of G-quadruplexes - The G4 scaffold is composed of a G-tetrad, a planar tetrameric assembly of guanine (G) bases linked together by Hoogsten H-bonding. G4-tetrads stack to form a typical G4-structure, which is stabilised by monovalent cations sitting at the centre of the tetrad, the relative stability depending on the size of the cation (in order of increasing stability: Li^+, Na^+, K^+), and more tetrads are held together by π-stacking interactions. **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.

Results

To confirm that multimolecular G4s (mG4s) promote aggregation of the hexanucleotide repeat sequence $(GGGGCC)_n$, two key conditions need to be met: i) the $(GGGGCC)_n$ must be able to form mG4s and ii) mG4s need to provide sufficiently stable cross links between strands to prompt their aggregation in the absence of proteins.

$(GGGGCC)_n$ forms mG4s

We first investigated the possibility that $(GGGGCC)_n$ forms multimolecular G4-structures. Although RNA is considered the toxic species in ALS and FTD pathological aggregates, we performed the bulk of our biophysical assays with DNA equivalents, as they are more easily sourced and handled[41]. To promote G4-formation, we annealed DNA $(GGGGCC)_n$ repeats of variable lengths (n=1 to 12) in a K^+-containing buffer (500 mM KCl)**.** To mimic intra-cellular crowding, 30% PEG was included in the buffer prior to slowly annealing the samples from 95 ℃ to 25 ℃ at a rate of -0.02 ℃/min (Methods 1).

Circular dichroism (CD) spectroscopy was initially performed to assess the formation of G4-structures (Methods 2). CD spectra of the $(GGGGCC)_n$ repeats annealed in mG4s-forming conditions presented the characteristic negative peak at ~240 nm and the positive peak at ~263 nm, which are typical of a parallel G4-topology (Figure 2A)[42]. Although CD confirmed G4-formation, it could not distinguish unimolecular from multimolecular topologies. Agarose gel electrophoresis (AGE) was thus employed to assess mG4s-formation (Methods 3, Figure 2B). Although the gel resolution in Figure 2B does not allow to visualise distinct bands, it is evident that samples with n>6 present significant smearing, attributed to the presence of higher molecularity species, such as mG4s.

The combination of CD and gel electrophoresis results suggested 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 (0-500 mM) experiment, as G4-formation is highly dependent on the concentration of K^+. AGE experiments revealed more pronounced slowly-migrating bands at higher K^+ concentrations, with no high-molecularity species being observed at sufficiently low K^+, thus confirming the role of G4s in mediating multimerization (Figure 2C).

To further confirm that the multimeric bands observed by agarose gels contained mG4s, we also stained the samples with N-methyl mesoporphyrin IX (NMM), a G4-specific fluorescent dye[43,44,45]. NMM is expected to fluoresce upon binding to the bands containing G4s but not the non-G4 DNA control. Indeed, as displayed in Figure 2D, NMM selectively stains the slowly-migrating portions of the $(GGGGCC)_n$ bands, indicating that the higher molecularity species contain G4s.

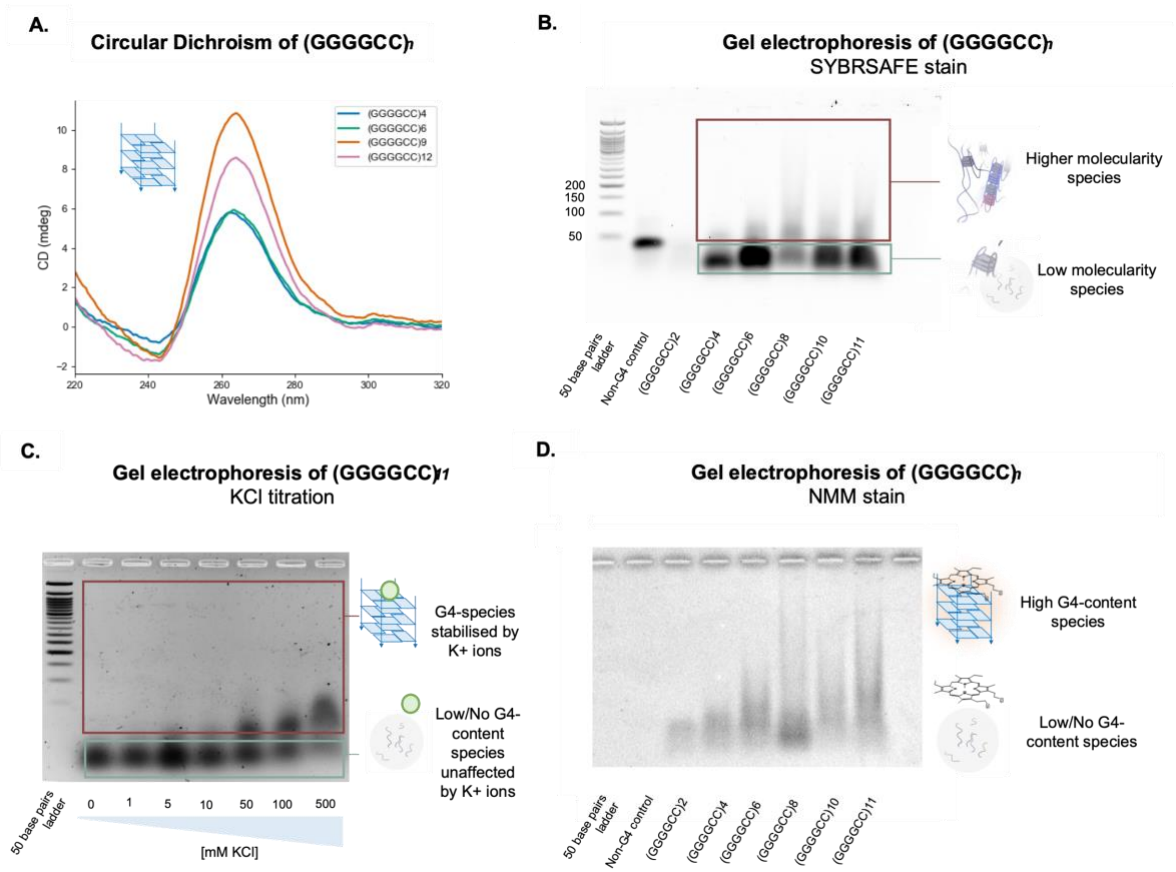


Figure 2: DNA $(GGGGCC)_n$ forms mG4s at different repeat lengths. A. $(GGGGCC)_n$ CD spectra - $(GGGGCC)_n$ (n = 4, 6, 9, 12) were annealed at 500 µM, 250 µM, 100 µM and 50 µM under mG4-forming conditions. All the samples present a positive peak at about 263 nm and a negative peak at 240 nm, both associated 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 and the gel was stained with SYBRSAFE. **C. $(GGGGCC)_{11}$ agarose gel electrophoresis -** $(GGGGCC)_{11}$ was annealed in the presence of different KCl concentrations. As KCl increases, so does the molecularity of the species involved, indicating that the multimolecular structures forming have a strong KCl dependence, relating them to G4s. In the schematics, the green dot represents the K^+ ion. **D. $(GGGGCC)_n$ agarose gel electrophoresis – NMM stain.** $(GGGGCC)_n$ (n = 2-11) were annealed at 250 µM 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.

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 the concentration of K^+ and the length of the repeat, we proceeded to systematically screen for the emergence of macroscopic aggregates using confocal microscopy. Figure 3A shows representative images of the aggregates for all tested repeat lengths, at a fixed concentration of 250 µM. Consistent with the observed clinical correlation between repeat length and pathological aggregation in ALS and STD[15], we note that *in vitro* aggregation is more prominent for higher n. Visually, the aggregates presented an amorphous morphology (Figure 3A, 3B, 3C, S1A) and showed no fluorescence recovery after photobleaching (FRAP), indicating a solid or gel-like state (Figure S1 – Carboxyfluorescein labelled aggregates).

The phase diagram in Figure 3D (Figure S2 systematically maps the aggregation state of the system as a function of repeat length and oligonucleotide concentration. Strands with repeat lengths of 2 and 3 showed no aggregation even at the highest tested concentration of 500 µM. The shortest repeat-length producing aggregates at 500 µM is n = 4, while constructs with $n \geq 9$ aggregated at oligonucleotide concentrations as low as 50 µM. The strong dependence of the phase-boundary location on repeat length confirms the stabilising effect that an increased repeat length has on the formation of multimeric structures, consistent with AGE data in Figures 2B and D.

Microscopy results thus confirm that $(GGGGCC)_n$ are capable of aggregation in the absence of additional proteins. Staining aggregates formed by $(GGGGCC)_6$ with NMM, expectedly produced a detectable 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 – Figure S3) under the same conditions. As expected, no fluorescence was detected in DNA condensates that do not have G4s (Figure S3B).

Considering the data in Figure 2, it is natural to hypothesise that macroscopic aggregation is mediated by mG4s cross-linking the oligonucleotides. However, NMM staining does not *per se* imply a structural role of quadruplexes but just demonstrates their presence. To assess whether G4s play a structural role within the aggregates, we performed NMM photo-oxidation experiments (Method 4). NMM binds to the G4s by end-stacking interactions and, upon irradiation with a high-power laser, it causes guanine photo-oxidation, leading to the formation of 8-oxo-guanine and consequent disassembly of the G4-structure[46]. This phenomenon has been previously observed to not occur for dsDNA[46], making it a highly G4-specific disassembly route, which we have previously leveraged to achieve light-controlled disassembly of mG4s in engineered DNA nanostructures[47]. As shown in Figure 3E, we irradiated a portion of the $(GGGGCC)_6$ NMM-stained aggregates to trigger photo-oxidation and, gratifyingly, disassembly of the aggregates was immediately observed. Given that no other relevant multimolecular interaction is known to be affected by NMM photo-oxidation, these results proved that mG4s were the prevalent structure holding the aggregate together, and that in the absence of guanine-guanine hydrogen bonding, necessary for G4-formation, no aggregation can be observed. Moreover, we tested the selectivity of the assay on NMM-stained dsDNA-nanostars, which did not show any significant change upon irradiation (Figure S3C).

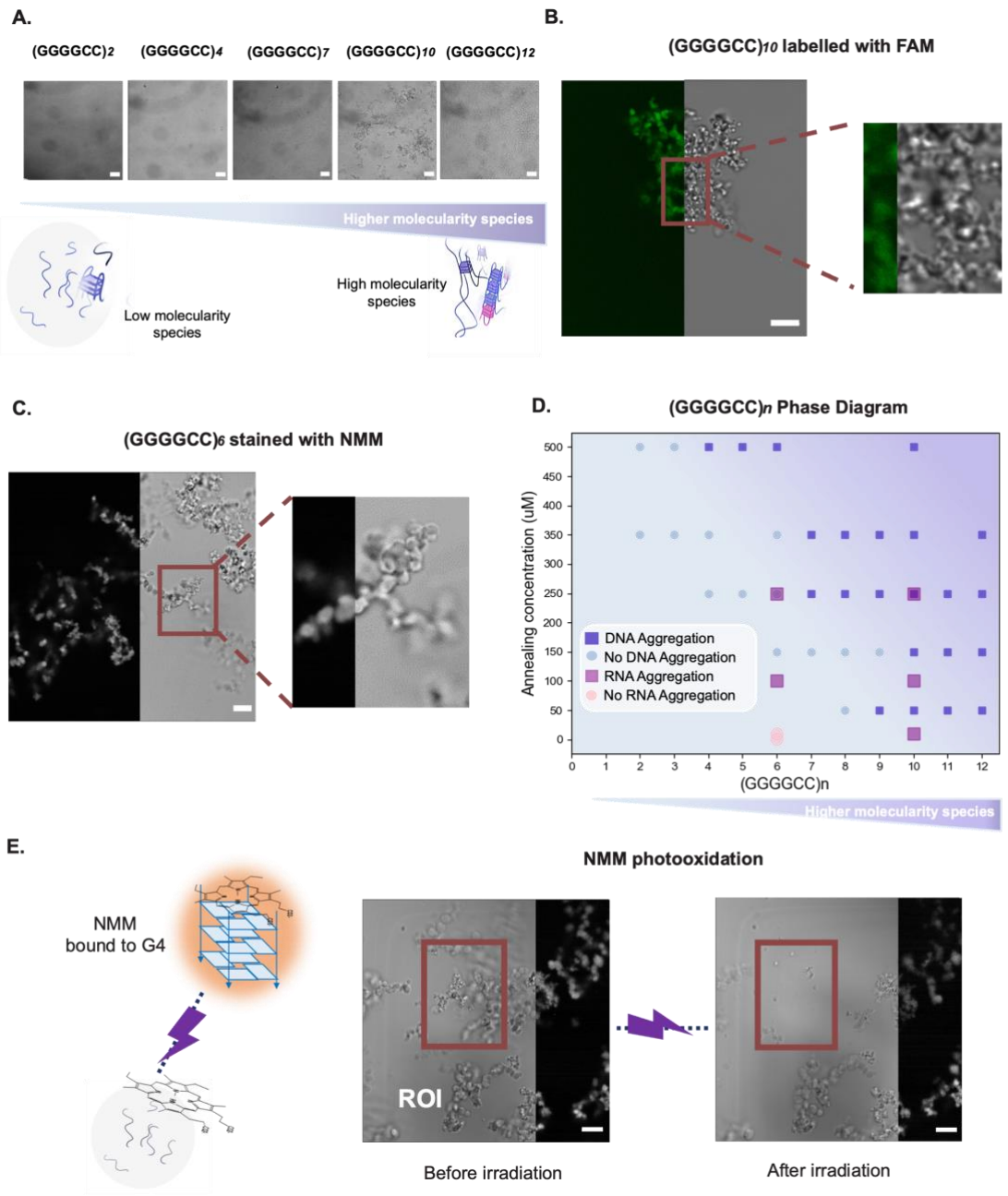


Figure 3: DNA $(GGGGCC)_n$ aggregates in a G4-dependent fashion. A. $(GGGGCC)_n$ aggregates - Confocal imaging in the brightfield channel of $(GGGGCC)_n$ (n = 2-12) annealed under mG4-forming conditions at 250 µM. 10 µm scalebar. **B. FAM labelled $(GGGGCC)_{10}$–** Carboxyfluorescein (FAM) labelled $(GGGGCC)_{10}$ was annealed in mG4-forming conditions. The image is split to show the brightfield channel (on the right) and the FAM (on the left). 10 µm scalebar. **C. $(GGGGCC)_6$ NMM staining experiment –** $(GGGGCC)_6$ was annealed under 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). **D. $(GGGGCC)_n$ phase diagram in KCl** – $(GGGGCC)_n$ was annealed under mG4-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. $r(GGGGCC)_{6,10}$ were also annealed under mG4s forming conditions and presented aggregation at lower concentrations than their DNA equivalents. **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 acquired. The aggregate in the ROI disassembled due to photooxidation. The image is split to show the brightfield channel (on the left) and the NMM (on the right). 10 µm scalebar.

Rational design of $(GGGGCC)_n$ mutants to prevent mG4-formation reduces aggregation

To further confirm that mG4s are mainly responsible for aggregation in $(GGGGCC)_n$ we carried out aggregation experiments on a series of mutated sequences (Figure 4A).

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 studied the $(CCCCGG)_6$ repeat, which retains the same number of potential canonical base pairing interactions as $(GGGGCC)_n$ but is unable to form G4s. CD was initially performed to confirm the absence of G4-formation, which was also validated *via* NMM staining of these sequences loaded on an agarose gel (Figures 4B, C and D). Replacing Cs with As 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 of the agarose gel confirmed G4 formation (Figures 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 mono-molecular G4-structures over multimolecular ones[48], allowing us to disentangle the effect of unimolecular vs multimolecular G4s in the aggregation process. CD and agarose gel-NMM staining of the T-to-C mutant indeed revealed formation of G4s (Figures 4C and D).

Having gauged the G4-forming tendencies of the mutants, we proceeded to assess their aggregation properties. No significant aggregation was observed for $(CCCCGG)_6$, confirming that canonical base pairing interactions are not sufficient to drive aggregation in the absence of G4s. Similarly, no significant aggregation was detected for $(GGGGTT)_6$, which predominantly forms unimolecular G4s as opposed to mG4s (see AGE in Figure 4C). This evidence further supports our hypothesis that mG4s are the essential cross linking elements in the aggregates, while formation of G4s is *per se* also insufficient to prompt aggregation. Fully consistently, $(GGGGAA)_6$ and $(GGGGCC)_6$ were the only two sequences revealing significant mG4-formation in the AGE experiments, and also the only ones showing significant aggregation under confocal microscopy (Figure 4E).

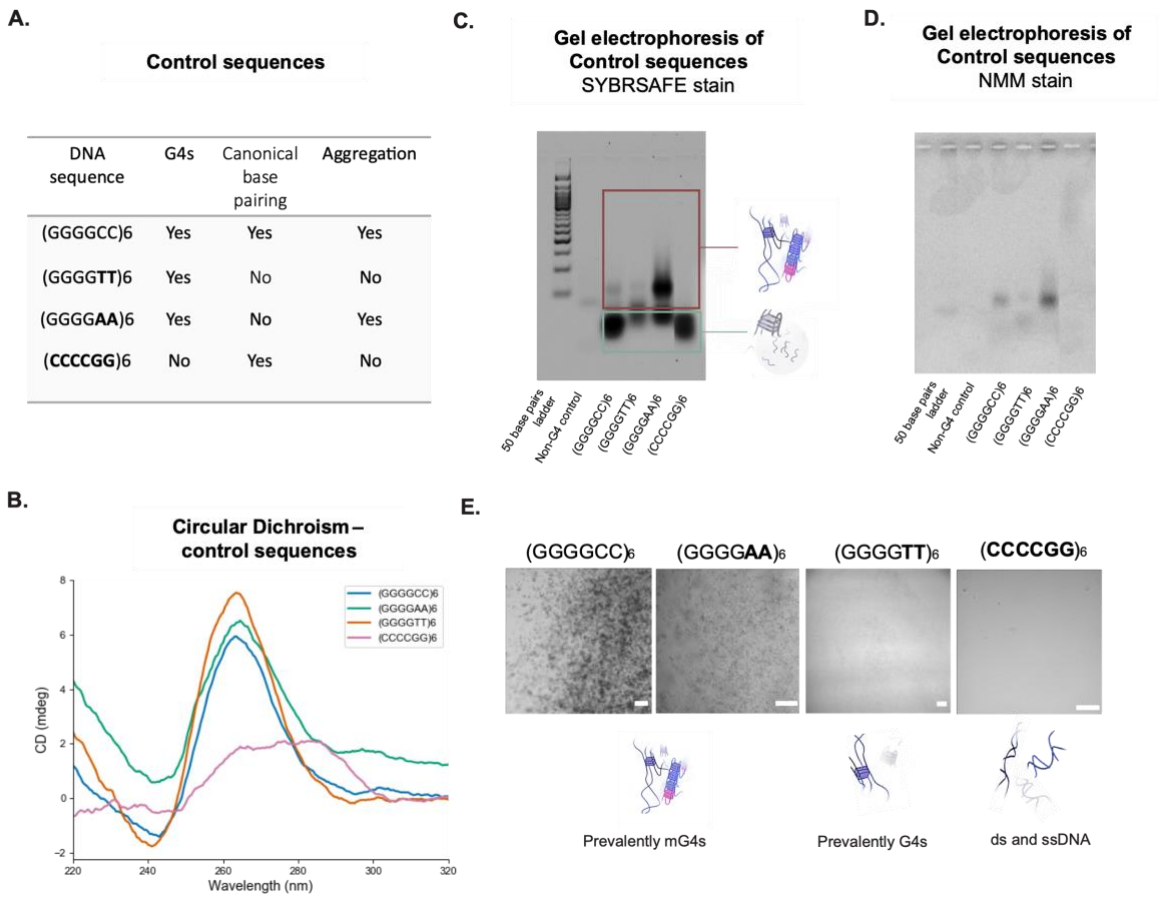


Figure 4: Role of base pairing and monomolecular G4s in the aggregation mechanism. A. Table of controls. B. Circular Dichroism spectra of controls - Controls of *n* = 6 were annealed in mG4-forming conditions. Only the predicted G4-forming sequences present the characteristic peaks at >260 nm and 240 nm. **C. Controls agarose gel electrophoresis – SYBRSAFE stain -** Controls of n = 6 were annealed in mG4-forming conditions. **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. Ladder, non-G4 ssDNA control, and $CCCCGG_6$ do not appear, confirming the specificity of the dye for G4-containing species. **E. Aggregation of the control sequences -** Controls of n = 6 were annealed under mG4-forming conditions. Aggregates were only clearly distinguishable for $(GGGGAA)_6$ and $(GGGGCC)_6$ – the only sequences forming mG4s. 100 µm scalebar.

G4-binding ligand PDS prevents $(GGGGCC)_n$ aggregation

As mentioned, G4 ligands have been observed to reduce the size of pathological ALS aggregates in cells. The ligands have been speculated to act by disrupting the interactions between G4-binding proteins and unimolecular G4s, thought to underpin aggregation[49]. Having demonstrated that mG4s alone are able to prompt aggregation in protein-free samples, it is natural to ask whether G4 could destabilise the aggregates through a different route, namely by promoting the formation of unimolecular as opposed to multimolecular quadruplexes. For this purpose, we identified pyridostatin (PDS), a well-characterised G4-ligand[50], known to promote kinetically the formation unimolecular G4s[50] and compatible with our experimental conditions.

Figure S4 shows the results of NMM-stain AGE experiments of $(GGGGCC)_{10}$ annealed in the presence of 10 µM PDS (Methods 6), revealing a substantial reduction in high molecularity species compared to the PDS-free sample. The ability of PDS to suppress large multimers is echoed by substantial reduction in the number and size of aggregates visible with confocal

microscopy (Methods 4), as displayed in Figure 5. It is worth noting that the small percentage of DMSO introduced to solubilise the ligand does not directly impact aggregation.

Besides further confirming the critical importance of G4 molecularity in stabilising $(GGGGCC)_n$ aggregates, these data outline a new mechanism for the reported ability of G4 ligands to ameliorate reported ALS phenotype in cells.

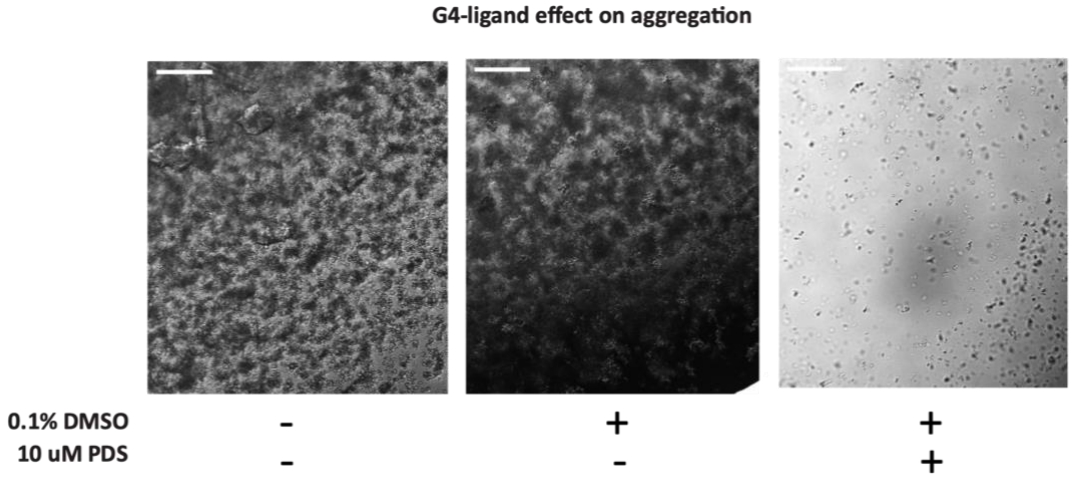


Figure 5: $(GGGGCC)_{10}$ aggregation in presence of 10 µM PDS - $(GGGGCC)_{10}$ was annealed at 150 µM under mG4-forming conditions. The sample containing the G4-ligand presents aggregates to a far lesser extent with respect to the non-PDS containing controls. 100 µm scalebar.

RNA repeats aggregate at lower concentrations compared to DNA

RNA is considered the toxic species in ALS and FTD pathological aggregates, but DNA equivalents were initially tested for biophysical studies as they are more easily sourced and handled[41]. 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 the equivalent DNA G4s[51,52]. For these reasons, we hypothesised that aggregation of $r(GGGGCC)_n$ could occur at lower concentrations compared to the aggregation observed with DNA but, overall, the aggregates would possess similar physical and chemical properties.

Therefore, we sought to confirm the ability of RNA to aggregate in a similar manner to its DNA equivalent by subjecting $r(GGGGCC)_{6/10}$ to the same mG4-forming conditions used for DNA (Methods 1). Aggregation was observed for concentrations as low as 100 µM for repeat length 6, and 10 µM for n=10 (Figure S5), forming in all cases amorphous aggregates similar in morphology to their DNA equivalents. For comparison, DNA $(GGGGCC)_6$ was seen to aggregate only above 350 µM in concentration, reflecting the known ability of RNA to form more stable G4s compared to DNA. To further validate mG4-formation in the RNA aggregates, the samples were stained with NMM, confirming the presence of G4s (Figure 6A). Furthermore, the RNA aggregates stained with NMM were subjected to irradiation at 405 nm to induce guanine photo-oxidation, which caused immediate disassembly of the irradiated aggregates (Figure 6B), as previously observed for DNA. These results confirmed that $(GGGGCC)_n$ aggregation is an mG4-dependant phenomenon and occurs in a similar manner for RNA and DNA repeats, but crucially requires lower nucleic acid concentrations in the case of RNA.

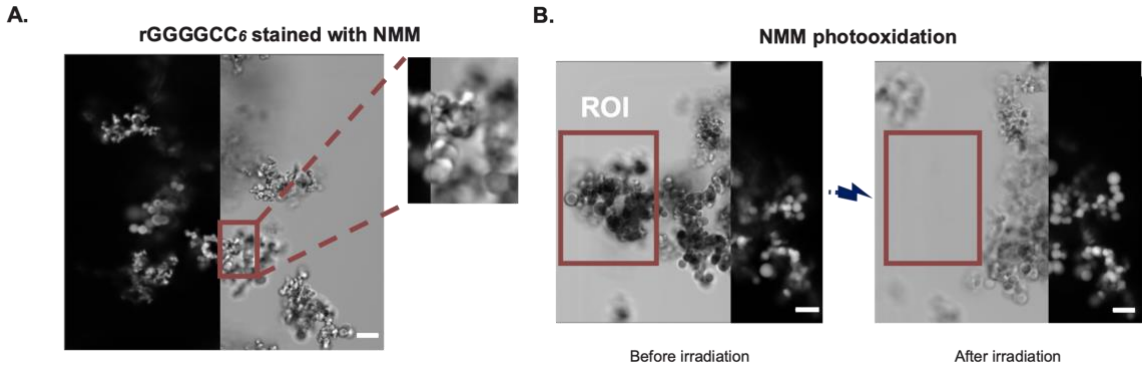


Figure 6: RNA (GGGGCC)$_n$ aggregates with a G4-dependent trend. A. r(GGGGCC)$_6$ NMM staining experiment - r(GGGGCC)$_6$ was annealed in mG4s forming conditions. It was then subjected to NMM staining prior to acquisition. The image is split to show the brightfield channel (on the right) and the NMM fluorescence channel (on the left). **B. r(GGGGCC)$_6$ NMM photo-oxidation 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 interest highlighted in the figure (ROI) was irradiated for 1 minute and the second image was subsequently acquired. The aggregate in the ROI disassembled due to photo-oxidation. 10 µm scalebar.

Discussion

Although unimolecular G4s are increasingly recognised drug targets for many diseases ranging from cancer and Cockayne Syndrome to Fragile X syndrome[30], the biological relevance of mG4s has only recently started to be investigated[53]. Indeed, there is growing evidence of how mG4-formation could aid long-range interactions at enhancer-promoters regions in cells[54] and how they could promote formation of phase-separated aggregates in membrane-less organelles[17,53]. Furthermore, the recent discovery of an endogenous human protein that selectively binds mG4s over unimolecular structures strongly suggests that these multimeric species could be formed in cells[34].

Although unimolecular species have a kinetic advantage and therefore form more readily, mG4s are usually thermodynamically favoured in the crowded conditions of the cell[55]. However, despite the uprising biological role of mG4s, there is no specific tool to test for their presence *in vitro* or *in vivo*. Much like when unimolecular G4s were first discovered and there were no specific probes that would distinguish them from dsDNA in live cells, currently all G4-specific probes (such as PDS[56], PhenDC3[57], BG4[58] and NMM[45]) have not been thoroughly studied in the context of mG4s, making differentiation in live-cell experiment impossible.

In this study we showed how (GGGGCC)$_n$ can arrange into mG4s *via* a combination of biophysical tools: Circular dichroism demonstrating the formation of G4-structures, agarose gel electrophoresis confirming the formation of multimolecular species, and both NMM staining and KCl titrations providing further evidence to support the formation of these species in ALS/FTD repeats. We further observed a clear correlation between mG4-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 investigated 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)[15], which is possibly 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 aggregate at lower concentrations, potentially reaching biologically relevant ranges (a human cell contains ~ 10–30 pg total RNA and 6 pg of DNA[59]).

The aggregates obtained were amorphous and solid in appearance. Although pathological aggregates in ALS and FTD are often referred to as "liquid-like"[60,61,62], it is reasonable to assume that additional protein components in ALS/FTD-affected cells could modify the physical state of the aggregates. In fact, two of the main proteins tied with the disorder (TDP-43 and FUS) have shown significant G4-binding abilities, suggesting that perhaps 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 photo-oxidation 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 RNA $(GGGGCC)_6$ sequence, successfully confirming the G4-driven aggregation behaviour.

Furthermore, we demonstrated how the G4-ligand PDS can diminish mG4-induced aggregation of $(GGGGCC)_{10}$ by driving the system to a lower molecularity state. In previously reported studies, G4-ligands have been shown to prevent aggregation events in ALS/FTD phenotypes[63,28]. It was hypothesised that the ligands interfere with aggregation by disrupting the interaction between G4s and G4-binding proteins[49]. Although this pathway is still plausible, here we offer additional insights into the aggregation properties of such repeat expansion in the absence of proteins, demonstrating that PDS could also be affecting the ALS phenotype by shifting G4 molecularity towards lower molecularity species.

In summary, our data depict a protein-free pathway for the formation of aggregates that could be relevant for ALS and FTD, making mG4s a potential therapeutic target for these neurodegenerative diseases. Importantly, this does not exclude the relevance of proteins or other canonical and non-canonical nucleic acid structures as contributors to these aggregates.

Currently, the only two FDA-approved drugs that have shown an increased survival rate in randomised ALS clinical trials are Riluzole and Edaravone[64,65], but ALS and FTD remain deadly diseases. So far, candidate drugs have been developed to target protein–led aggregation but have shown only partial effectiveness[66], suggesting the need to explore alternative approaches. A recent study has shown potential in a small molecule with the ability to cross the blood-brain barrier and binding to $r(GGGGCC)_n$ promoting formation of a hairpin structure over G4s, which could offer the starting point for a new therapeutic pathway[29]. Altogether, these studies and our work strongly suggest that targeting RNA structures within the $(GGGGCC)_n$ repeat could be a valuable strategy for treating ALS and FTD, which should be further investigated in the future.

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 ℃ for 5 minutes to ensure complete solvation. The DNA samples were then further diluted to the desired annealing concentration (500-0 µM) in 10 mM TRIS-HCl buffer at 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 ℃ for 1h followed by cooling to 25 ℃ with a temperature ramp of 0.25 ℃/min). 500 mM KCl was added to the denatured solution and a further heating cycle was applied (heating to 95 ℃ for 1h followed by cooling to 25 ℃ with a

temperature ramp of 0.02 ℃/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. Running the gel on ice makes it easier to distinguish multimolecular species.

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 10 µM 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.

G4-ligands containing 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 $(GGGGCC)_{10}$ in presence of the 10 µM PDS was directly transferred in PDMS wells and imaged by irradiating the sample with the 458 nm excitation laser on the brightfield channel. As PDS is solubilised in presence of 0.1% DMSO, a DMSO control was also added to the experiment.

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 10 µM 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

Condensates made from dsDNA nanostars were prepared by mixing four nanostar core strands (N4_core1, N4_core2, N4_core3, N4_core4) and two linker strands (L_AA1 and L_AA2) in a TrisHCl buffer containing 500mM KCl at an final nanostar concentration of 1 µM and linkers concentration of 2 µM. A volume of 70µL was then loaded into borosilicate glass capillaries (inner dimensions: 0.4 mm × 4 mm × 50 mm, CM Scientific) with a micropipette. Prior to loading, capillaries were cleaned by sonication in 1% Hellmanex III (HellmaAnalytics) at 60°C for 15 minutes. The surfactant was removed through five rounds of rinsing with deionised (DI) water, followed by a round of sonication in ultrapure water (Milli-Q). After loading the sample, the capillary ends were capped with mineral oil and sealed with epoxy glue (Araldite) on a coverslip. The sample was annealed by incubating at 95℃ for 30 minutes and cooled down between 85℃ and 25℃ (-0.1℃/min cooling rate). Samples were transferred from the glass capillary to a PDMS well for imaging. All strands were purchased from Integrated DNA Technologies (IDT) and reconstituted in TE. Sequences are provided in the SI, Figure S3 (A).

6. $(GGGGCC)_n$ mG4s annealing protocol in presence of PDS

$(GGGGCC)_{10}$ was purchased from IDT in its lyophilised form and diluted in MilliQ water to reach a stock concentration of 1 mM. The stock was heated to 40 ℃ for 5 minutes to ensure

complete solvation. The DNA sample was then further diluted to the desired annealing concentration (150 µM) in 10 mM TRIS-HCl buffer at 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 ℃ for 1h followed by cooling to 25 ℃ with a temperature ramp of 0.25 ℃/min). 500 mM KCl and 10 µM PDS were added to the denatured solution and a further heating cycle was applied (heating to 95 ℃ for 1h followed by cooling to 25 ℃ with a temperature ramp of 0.02 ℃/min). To solubilise the PDS, 0.1% DMSO was also added to the solution, therefore a DMSO control has also been added to the results. 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.

References

1. Abramzon, Y. A., Fratta, P., Traynor, B. J. & Chia, R. The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Front. Neurosci.* 14, (2020).
2. Chiò, A. *et al.* Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. *Neuroepidemiology* 41, 118–130 (2013).
3. Ratnavalli, E., Brayne, C., Dawson, K. & Hodges, J. R. The prevalence of frontotemporal dementia. *Neurology* 58, 1615–1621 (2002).
4. Amyotrophic Lateral Sclerosis (ALS) Fact Sheet | National Institute of Neurological Disorders and Stroke. https://www.ninds.nih.gov/amyotrophic-lateral-sclerosis-als-fact-sheet.
5. Frontotemporal dementia. *nhs.uk* https://www.nhs.uk/conditions/frontotemporal-dementia/ (2017).
6. Ferrari, R., Kapogiannis, D., Huey, E. D. & Momeni, P. FTD and ALS: a tale of two diseases. *Curr Alzheimer Res* 8, 273–294 (2011).
7. Burrell, J. R., Kiernan, M. C., Vucic, S. & Hodges, J. R. Motor neuron dysfunction in frontotemporal dementia. *Brain* 134, 2582–2594 (2011).
8. 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).

9. 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).
10. 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).
11. Ling, S.-C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron* 79, 416–438 (2013).
12. 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).
13. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162, 1066–1077 (2015).
14. Proteostatic imbalance and protein spreading in amyotrophic lateral sclerosis. https://www.embopress.org/doi/epdf/10.15252/embj.2020106389 doi:10.15252/embj.2020106389.
15. Iacoangeli, A. *et al.* C9orf72 intermediate expansions of 24–30 repeats are associated with ALS. *Acta Neuropathologica Communications* 7, 115 (2019).
16. 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).
17. 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).
18. Balendra, R. & Isaacs, A. M. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol* 14, 544–558 (2018).

19. Wang, Z.-F. *et al.* The Hairpin Form of r(G4C2)exp in c9ALS/FTD is Repeat-associated non-ATG Translated and a Target for Bioactive Small Molecules. *Cell Chem Biol* 26, 179-190.e12 (2019).
20. 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).
21. Kovachev, P. S. *et al.* RNA modulates aggregation of the recombinant mammalian prion protein by direct interaction. *Scientific Reports* 9, 12406 (2019).
22. 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).
23. 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).
24. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* 546, 243–247 (2017).
25. Wang, Z.-F. *et al.* The Hairpin Form of r(G4C2)exp in c9ALS/FTD Is Repeat-Associated Non-ATG Translated and a Target for Bioactive Small Molecules. *Cell Chemical Biology* 26, 179-190.e12 (2019).
26. 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).
27. 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).

28. Cheng, A. *et al.* Selective C9orf72 G-Quadruplex-Binding Small Molecules Ameliorate Pathological Signatures of ALS/FTD Models. *J. Med. Chem.* 65, 12825–12837 (2022).
29. 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).
30. Raguseo, F., Chowdhury, S., Minard, A. & Di Antonio, M. Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem. Commun.* 56, 1317–1324 (2020).
31. 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).
32. 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).
33. Blice-Baum, A. C. & Mihailescu, M.-R. Biophysical characterization of G-quadruplex forming FMR1 mRNA and of its interactions with different fragile X mental retardation protein isoforms. *RNA* 20, 103–114 (2014).
34. 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).
35. 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).
36. 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).

37. Zhang, Y. *et al.* G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Research* 47, 11746–11754 (2019).
38. Imperatore, J. A., McAninch, D. S., Valdez-Sinon, A. N., Bassell, G. J. & Mihailescu, M. R. FUS Recognizes G Quadruplex Structures Within Neuronal mRNAs. *Front Mol Biosci* 7, 6 (2020).
39. 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).
40. 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).
41. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-quadruplexes in biology: principles and molecular mechanisms. *J Mol Biol* 429, 2127–2147 (2017).
42. del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-quadruplex secondary structure from circular dichroism spectroscopy. *Angew Chem Int Ed Engl* 57, 7171–7175 (2018).
43. 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).
44. 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).
45. Sabharwal, N. C. *et al.* N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J* 281, 1726–1737 (2014).

46. 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).

47. 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).

48. 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).

49. Ishiguro, A. & Ishihama, A. Essential Roles and Risks of G-Quadruplex Regulation: Recognition Targets of ALS-Linked TDP-43 and FUS. *Frontiers in Molecular Biosciences* 9, (2022).

50. Santos, T., Salgado, G. F., Cabrita, E. J. & Cruz, C. G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals (Basel)* 14, 769 (2021).

51. Lyu, K., Chow, E. Y.-C., Mou, X., Chan, T.-F. & Kwok, C. K. RNA G-quadruplexes (rG4s): genomics and biological functions. *Nucleic Acids Research* 49, 5426–5450 (2021).

52. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol* 21, 459–474 (2020).

53. 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).

54. Li, L. *et al.* YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* 17, 161–168 (2021).

55. Merkina, E. E. & Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophysical Journal* 89, 365–373 (2005).

56. 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).

57. 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).

58. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chem* 5, 182–186 (2013).

59. Piovesan, A. *et al.* On the length, weight and GC content of the human genome. *BMC Res Notes* 12, 106 (2019).

60. 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).

61. Farrawell, N. E. *et al.* Distinct partitioning of ALS associated TDP-43, FUS and SOD1 mutants into cellular inclusions. *Sci Rep* 5, 13416 (2015).

62. Koehler, L. C. *et al.* TDP-43 Oligomerization and Phase Separation Properties Are Necessary for Autoregulation. *Frontiers in Neuroscience* 16, (2022).

63. Simone, R. *et al.* G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol Med* 10, 22–31 (2018).

64. 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).

65. 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).

66. Malik, R. & Wiedau, M. Therapeutic Approaches Targeting Protein Aggregation in Amyotrophic Lateral Sclerosis. *Front Mol Neurosci* 13, 98 (2020).

Acknowledgements

Graphics were designed using Procreate design app and ChemDraw.

Authors contributions

FR developed the experimental protocols and conducted all experiments apart from design and annealing of dsDNA nanostars (DT). TM synthesised PDS and aided FR in the PDS solubility assay (not reported in the paper) for the G4-ligands addition experiment. LM helped FR in performing the RNA and DNA photooxidation experiments. RRS helped FR in obtaining the 3D reconstructed image of the fluorescence aggregates. FR analysed all the data. FR wrote the paper, aided by MDA and LDM. FR, MDA, LDM and YE designed the research. MDA, LDM and YE supervised the research. All authors discussed the results and edited the paper.

Funding

This work was supported by the following funding bodies:

MDA is supported by Biotechnology and Biological Sciences Research Council (BBSRC) David Phillips Fellowship [BB/R011605/1]; MDA is a Lister Institute Fellow and is supported by a Lister Institute Research Prize. FR is supported by a Leverhulme Trust, Leverhulme Cellular Bionics scholarship [EP/S023518/1]; LDM 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 RRS. LDM, RRS, DT and LM are supported by the European Research Council (ERC) under the Horizon 2020 Research and Innovation Programme [ERC-STG No 851667 NANOCELL]. YE is supported by United Kingdom Research and Innovation Future Leaders Fellowship [MR/S031537/1]. TM is supported by [EPSRC G98210].

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.

Materials and Correspondence

For further information refer to corresponding authors Dr Di Antonio and Dr Di Michele.

Tel: +44 (0)20 7594 5866; Email: m.di-antonio@imperial.ac.uk (MDA) and ld389@cam.ac.uk (LDM).