Imperial College London

Unravelling the role of multimolecular G-quadruplexes in promoting aggregation in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia

Federica Raguseo

A thesis presented for the degree of

Doctor of Philosophy

Imperial College London

Department of Chemistry

October 2023

To my parents, without whom I would have never made it this far.

Statement of Originality

I, *Federica Raguseo*, declare that this thesis is the result of my own work reporting original results conducted by myself. Any work performed in collaboration with others has been highlighted in the text. Sources obtained from others have been referenced accordingly.

Federica Raguseo

October 2023

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are just two of the neurodegenerative diseases characterised by the presence of pathological aggregates composed of proteins and RNA. The mechanism of aggregate formation has historically defined proteins as nucleation triggers, but recent studies have highlighted the fundamental role of nucleic acids in the aggregation events. In particular, 40% of familiar cases in ALS/FTD have been correlated to mutations in the hexanucleotide repeat $(GGGGCC)_n$, which in its RNA form has been identified within the aggregates. This work aimed to investigate the aggregation properties of $(GGGGCC)_n$ and unveil the underpinning molecular mechanism in the context of the ALS/FTD diseases. In these studies, I firstly showed that $(GGGGCC)_n$ DNA oligonucleotides can in the absence of proteins, and that aggregation is facilitated by higher RNA concentrations and number of repeats. I then demonstrated that aggregation is mediated by the formation of multimolecular G-quadruplexes (mG4s) and proceeded to disassemble the aggregates by employing G4-specific tools or altering the molecularity of the system *via* mutation analysis and antisense treatment. I validated the biological relevance of my findings by demonstrating the ability of RNA $(GGGGCC)_n$ to form mG4-rich aggregates similar to its DNA equivalents and by examining the interactions of these strands with TDP-43, a protein relevant to ALS/FTD. Furthermore, I confirmed the potential therapeutical relevance of the studies by demonstrating the ability of a G4-selective fluorescent probe to penetrate *C9orf72* mutant human motor neurons derived from ALS patients, which revealed clear fluorescent signal in putative condensates. My findings strongly suggest that $(GGGGCC)_n$ can form protein-free condensates sustained by multimolecular G-quadruplexes, highlighting their potential relevance as therapeutic targets for *C9orf72* mutation-related ALS and FTD.

Journal publications

- **Raguseo F.**, *et al.* The ALS/FTD-related C9orf72 hexanucleotide repeat expansion forms RNA condensates through multimolecular G-quadruplexes. *Nature Communications.* **14**, 1, 8272 (2023).

- Walczak M., Mancini L., Xu J., **Raguseo F.**, *et al*. A Synthetic Signalling Network Imitating the Action of Immune Cells in Response to Bacterial Metabolism. *Advanced Materials*. **35**, 33 (2023). DOI: 10.1002/adma.202301562.

- Lim Kee Chang W., Chan T.G., **Raguseo F.**, *et al*. Rapid short-pulses of focused ultrasound and microbubbles deliver a range of agent sizes to the brain. *Scientific Reports.* **13**, 1, 6963 (2023). DOI: 10.1038/s41598-023-33671-5.

- Williams S.L., Casas-Delucchi C.S., **Raguseo F.**, *et al*. Replication-induced DNA secondary structures drive fork uncoupling and breakage. *The EMBO Journal.* (2023). DOI: 10.15252/embj.2023114334.

- Liano D., Monti L., Chowdhury S., **Raguseo F.**, *et al*. Long-range DNA interactions: inter-molecular G-quadruplexes and their potential biological relevance. *Chemical Communications*. (2022). DOI: 10.1039/D2CC04872H.

- Robinson J., **Raguseo F.**, *et al*. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Research*. **49,** 15, 8419-8431 (2021).

- Minard A., Morgan D., **Raguseo F.**, *et al.* A short peptide that preferentially binds c-MYC G-quadruplex DNA. *Chemical Communications*. **56**, 63, 8940-8943 (2020).

- **Raguseo F.**, *et al.* Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. Chemical Communications, **56,** 9, 1317-1324 (2020).

Presentations

- Speaker at G4interact, organised by the University of Pavia, Italy (held virtually), 2021;
- Poster presenter at G4-thering, an international conference on G4s and non-canonical DNA secondary structures, Marienbad, Czech Republic, 2022;
- Poster presenter at Syncell, an international conference on synthetic cells, The Hague, Netherlands, 2022;
- Poster presenter at KeyStone Symphosia: Biomolecular Condensates, international conference on the role of condensates in biology, Vancouver, Canada, 2023.
- **Best Speaker** at G4interact, organised by the University of Pavia, Italy (held virtually), 2023;
- **Second Best Speaker** at Chemical Biology for Healthcare Postgraduate Symposium, Imperial College London, 2023;
- **Best Poster Presenter** at PostGraduate Showcase, Imperial College London, 2023;

Table of Contents

Statement of Originality **3**
Copyright Declaration **3**
Abstract **4**
Journal publications **5**
Presentations **6**
Table of Contents **7**
List of Abbreviations **10**
List of Figures **13**
List of Tables **15**
Chapter 1 ..... **16**
Introduction **16**
1.1 Neurodegenerative diseases 16
1.1.1 Socio-economic impact 16
1.1.2 Current treatments 21
1.1.3 Classification of NDs from clinical observations 24
1.1.4 Molecular hallmarks of neurodegeneration 27
1.1.5 Mechanism behind ALS/FTD onset 37
1.2 G-quadruplexes 45
1.2.1 Nucleic Acids: from helices to G4s 45
1.2.2 Biological relevance of G4s 48
1.2.3 Potential biological relevance of mG4s 51
1.2.4 mG4s role in phase-separation events 53
1.2.5 G4s as therapeutic targets 54
Chapter 2 ..... **56**
Project Aims **56**
Chapter 3 ..... **59**
Multimolecular G4s can trigger aggregation of (GGGGCC)$_n$ in absence of proteins **59**
3.1 Background and Aims 59
3.2 Experimental Techniques 62
3.2.1 Thermal Annealing 62
3.2.2 Microscopy 62
3.2.3 Agarose Gel Electrophoresis 63
3.2.3 Circular Dichroism 63
3.3 Results and Discussion 65

3.3.1 The C9orf72 repeat expansion $(GGGGCC)_n$ can arrange into mG4s 65
3.3.2 $(GGGGCC)_n$ can aggregate in vitro in absence of proteins 82
3.3.3 $(GGGGCC)_n$ aggregates are formed via mG4 self-assembly 89
3.4 Conclusions 97
3.5 Materials and Methods 99
3.5.1 Optimisation of the annealing conditions 99
3.5.2 Optical Microscopy 100
3.5.3 Agarose gel electrophoresis (AGE) 101
3.5.4 Circular Dichroism of DNA samples 101
3.5.5 Confocal microscopy 102
3.5.6 dsDNA condensates controls 103
Chapter 4 104
The molecularity of the $(GGGGCC)_n$ system affects its aggregation abilities 104
4.1 Background and Aims 104
4.2 Experimental Techniques 105
4.2.1 Mutation Analysis 105
4.2.2 PDS treatment 105
4.2.3 Antisense treatment 105
4.3 Results and Discussion 106
4.3.1 $(GGGGCC)_n$ mutation analysis 106
4.3.2 $(GGGGCC)_n$ aggregation is prevented by G4-ligand addition 112
4.3.3 $(GGGGCC)_n$ antisense treatment 119
4.4 Conclusions 123
4.5 Materials and methods 124
4.5.1 Annealing of $(GGGGCC)_n$ and its mutant sequences 124
4.5.2 PDS solubility assay 124
4.5.3 Antisense treatment assay 124
Chapter 5 125
Biological relevance of $(GGGGCC)_n$ aggregation 125
5.1 Background and Aims 125
5.2 Experimental Techniques and reagents 126
5.2.1 Polyacrylamide gel electrophoresis (PAGE) 126
5.2.2 Electrospray ionization mass spectrometry (ESI-MS) 126
5.2.3 Proteostat™ molecular rotor 126
5.2.4 Transmission Electron Microscopy (TEM) 127
5.2.5 IPSC-derived neuronal cells (iPSCs) 127

5.2.6 Fluorescence-activated cell sorting (FACS) 127
5.2.7 Anti β-tubulin cells staining 127
5.3 Results and discussion 128
5.3.1 The biologically relevant *r*(GGGGCC)$_n$ behaves similarly to (GGGGCC)$_n$ 128
5.3.2 TDP-43 enhances mG4-driven aggregation of (GGGGCC)$_n$ 131
5.3.3 Evidence of G4-driven aggregation in iPSC derived neuronal cells 149
5.4 Conclusios 158
5.5 Methods 161
5.5.1 RRM1–2 expression and purification 161
5.5.2 SDS-PAGE 161
5.5.3 Electrospray ionization mass spectrometry (ESI-MS) 162
5.5.4 Proteostat Binding Assays 162
5.5.5 RRM1–2 labelling with ALEXA 633 162
5.5.6 DNA/protein aggregates formation 162
5.5.7 Fluorescein self-quenching tests 162
5.5.8 Aggregates quantification analysis from microscopy 162
5.5.9 Transmission electron microscopy (TEM) 163
5.5.10 IPSC patient derived neuronal cell culture 163
5.5.11 Immunocytochemistry (ICC) 163
5.5.12 Fluorescence-activated cell sorting (FACS) 164
6. General Conclusions and Future Outlooks 165
References 169
Appendix 186
Figures copyright 210
Acknowledgements 211

List of Abbreviations

A	Adenine
AD	Alzheimer's Disease
AGE	Agarose Gel Electrophoresis
ALP	Autophagy-Lysosome Pathway
ALS	Amyotrophic Lateral Sclerosis
APP	Amyloid Beta Precursor Protein
ATG	Autophagy Related Gene
ATM	Ataxia-Telangiectasia Mutated
ATXN2	Ataxin-2
C	Cytosine
C9orf72	Chromosome 9 Open Reading Frame 72
CD	Circular Dichroism
CHCHD10	Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 10
ChIP	Chromatin
CHMP2B	Charged Multivesicular Body Protein 2B
CNS	Central Nervous System
CSB	Cockayne syndrome group B protein
CTE	Chronic Traumatic Encephalopathy
DJ-1	Protein Deglycase
DMPK	Myotonic Dystrophy Protein Kinase
ds	Double Stranded
EC50	Half Maximal Effective Concentration
ESI-MS	Electrospray Ionization Mass Spectrometry
FACS	Fluorescence-Activated Cell Sorting
FDA	Food and Drug Administration
FRAP	Fluorescence Recovery After Photobleaching
FTD	Frontotemporal Dementia
FUS	FUsed in Sarcoma
G	Guanine
G3DP1	Stress Granule Assembly Factor 1
G4	G-quadruplex
GBA1	Glucosylceramidase Beta 1
GBP	Great British Pound
GFP	Green Fluorescent Protein
GNR	Granulin
H-bonding	Hydrogen bonding
HD	Huntington Disease
hnRNP	Heterogeneous Nuclear Ribonucleoproteins
HTT	Huntingtin
ICC	Immunocytochemistry
IDT	Integrated DNA Technology
IPSC	Induced Pluripotent Stem Cells
LCD	Low Complexity Domains
LDB	Lewy Body Dementia

LNA	Locked Nucleic Acids
LRRK2	Leucine-Rich Repeat Kinase 2
MAPT	Microtubule-Associated Protein Tau
mG4	multimolecular G-quadruplexes
MS	Multiple sclerosis
MSA	Multisystem atrophy
ND	Neurodegenerative Diseases
NHS	National Health Service
NMM	N-methyl mesoporphyrin IX
NUPACK	Nucleic Acid Package
OPMD	Oculopharyngeal Muscular Dystrophy
OT2	Opentrons Pipetting robot
OXPHOS	Oxidative Phosphorylation
PAGE	Polyacrylamide Gel Electrophoresis
PB	Phosphate Buffer
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PD	Parkinson's Disease
PDS	Pyridostatin
PEG	Polyethylene Glycol
polyQ	Polyglutamine
PS1	Presenilin-1
QGRS	Quadruplex Forming G-rich Sequence(s)
RBD	RNA-Binding Domain
RE	Repeat Expansion
RG4	RNA G-quadruplex
ROI	Region Of Interest
ROS	Oxygen Free Radicals
RRM1-2	RNA Binding Domain 1-2
SCA	Spinocerebellar Ataxias
SCI	Stroke and Spinal Cord Injury
SETX	Senataxin
SG	Stress Granule
SMN1	Survival Motor Neuron 1
SNX14	Sorting Nexin 14
SOD1	Superoxide Dismutase Type 1
SQSTM1	Sequestosome 1
ss	Single Stranded
T	Thymine
TBK1	TANK-binding kinase 1
TDP1/2	Tyrosyl-DNA-phosphodiesterase I/II
TDP-43	TAR DNA-binding protein 43
TEM	Transmission Electron Microscopy
TMEM230	Transmembrane Protein 230
TOP1	DNA Topoisomerase I
TRIS-HCL	Tris (hydroxymethyl) Aminomethane buffer
UBQLN1/2	Ubiquilin-1/2
UPS	Ubiquitin-Proteasome System

USD	United States Dollars
UTR	Untranslated Region
UV	UltraViolet
VCP	Valosin-Containing Protein

List of Figures

Figure 1: Socio-economic impact of neurodegenerative diseases. 17
Figure 2: Socio-economic reasons behind lack of treatment for NDs. 19
Figure 3: Socio-economic impact of ALS/FTD. 20
Figure 4: Overlapping symptoms and genetics of ALS and FTD. 27
Figure 5: Molecular hallmarks of the most common NDs. 28
Figure 6: Biological consequences of protein aggregation.. 30
Figure 7: DNA damage pathways associated with NDs. 32
Figure 8: Impaired RNA biological roles associated with NDs. 33
Figure 9: Repeat expansion toxicity pathways. 37
Figure 10: TDP-43 main domains structures, relevant ALS/FTD mutations and pathological pathways. 39
Figure 11: Chromosome 9 mutations are responsible for ALS/FTD *via* multiple mechanisms of action. 44
Figure 12: Nucleic acids can arrange into several secondary structures. 47
Figure 13: G-quadruplexes biological relevance. 51
Figure 14: General characteristics of multimolecular G4s versus unimolecular G4s. 53
Figure 15: Biological roles of mG4s. 54
Figure 16: Alternative mechanism of pathological aggregation in ALS/FTD based on mG4 formation.. 58
Figure 17: mG4s vs G4s from a thermodynamic viewpoint. 60
Figure 18: Experimental pipeline to assess the successful denaturing of $(GGGGCC)_{11}$ species. 66
Figure 19: Optimisation of the disassembly step for $(GGGGCC)_n$ prior to annealing. 68
Figure 20: Experimental pipeline to assess the successful mG4-formation protocol of $(GGGGCC)_{11}$ species. 70
Figure 21: mG4 assembly protocol optimisation.. 73
Figure 22: Circular Dichroism melting experiment. 75
Figure 23: $(GGGGCC)_n$ forms higher molecularity species that can be ascribed to mG4s. 78
Figure 24: $(GGGGCC)_n$ can form mG4s at multiple repeat lengths. 81
Figure 25: $(GGGGCC)_n$ can aggregate in absence of proteins.. 84
Figure 26: G4-topology effect on aggregation. 85
Figure 27: Opentrons pipetting robot $(GGGGCC)_n$ mG4-formation protocol optimisation. 86
Figure 28: $(GGGGCC)_n$ aggregation analysed with the Opentrons protocol. 88

Figure 29: FAM-labelled $(GGGGCC)_n$ can aggregate *in vitro*. 90
Figure 30: $(GGGGCC)_6$ aggregates in presence of NMM. 93
Figure 31: $(GGGGCC)_n$ aggregates possess autofluorescence abilities. 95
Figure 32: $(GGGGCC)_{11}$ aggregates melting experiment. 96
Figure 33: $(GGGGCC)_n$ mutation analysis. 106
Figure 34: $(GGGGAA)_6$ and $(GGGGTT)_6$ mG4-formation and aggregation properties. 108
Figure 35: $(GGGGC)_6$, $(GGGGCCC)_6$ and $(CCCCGG)_6$ mG4-formation and aggregation properties. 111
Figure 36: DMSO effect on aggregation of $(GGGGCC)_n$. 114
Figure 37: PDS solubility tests. 116
Figure 38: PDS effect on $(GGGGCC)_{10}$ mG4-formation and aggregation abilities. 118
Figure 39: Antisense treatment effect on $(GGGGCC)_8$ mG4 formation and aggregation abilities. 121
Figure 40: *r*$(GGGGCC)_n$ aggregation properties. 129
Figure 41: *r*$(GGGGCC)_6$ aggregates *via* mG4-formation. 130
Figure 42: TDP-43 properties. 133
Figure 43: TDP-43 kinetics of aggregation.. 136
Figure 44: EC50 of aggregation of TDP-43 in presence/absence of DNA. 139
Figure 45: TDP-43 and $(GGGGCC)_{11}$ co-localisation in the aggregates.. 141
Figure 46: TDP-43 and $(GGGGCC)_{11}$ aggregates quantification *via* NMM-staining. 143
Figure 47: Confocal images quantification analysis.. 144
Figure 48: TDP-43 interaction with denatured $(GGGGCC)_{11}$. 146
Figure 49: TDP-43 and $(GGGGCC)_{11}$ aggregates subjected to NMM-photooxidation. 147
Figure 50: TEM images of TDP-43 containing samples. 149
Figure 51: Live IPSC ALS/FTD neuronal cells with the C9orf72 mutation. 151
Figure 52: Confocal microscopy images of C9N1 and C9N6 fixed cell lines. 153
Figure 53: Confocal microscopy images of fixed IPSC patient derived neuronal cells stained with anti-β-tubulin antibody. 154
Figure 54: FACS data for the NMM staining studies. 156
Figure 55: NMM staining of RNA foci in ALS/FTD cell lines.. 157
Figure 56: The three main findings of the study. 161

List of Tables

Table 1: List of current treatments for NDs. 22
Table 2: List of current treatments for ALS and FTD and their biological target. 23
Table 3: Symptoms and clinical observation of neurodegeneration for the most common ND pathologies. 25
Table 4: List of repeat expansions usually associated with pathological consequences. 35
Table 5: ssDNA sequences used for the assembly of the dsDNA control. 92
Table 6: Samples tested with the Proteostat™ dye over three days. 134

Chapter 1

Introduction

1.1 Neurodegenerative diseases

1.1.1 Socio-economic impact

As of the latest statistics released by the World Health Organisation up to one billion people are affected by neurodegenerative diseases (NDs), with a related annual estimate of 6.8 million deaths[1]. NDs are commonly known as a heterogeneous class of diseases which involve neuron degeneration. As the risk of developing NDs increases with age, the rapidly ageing world population will lead to an inevitable rise in cases. For instance, in Europe, 16% of people above > 65 y.o. is affected by NDs. But this percentage is predicted to increase to 25% by 2030 and to double by 2040[2].

NDs have been medically described since the early 19th century with the depiction of Parkinson's disease (PD) in 1817[3], Amyotrophic Lateral Sclerosis (ALS) in 1869[4], Frontotemporal Dementia (FTD) in 1892[5] and Alzheimer's disease (AD) in 1906[6], but the way they develop remains widely unknown and no cure is currently available.

NDs affect a patient's nervous system, with implications for their communication skills, cognition, memory, behaviour, sensory and motor functions, tragically affecting their quality of life progressively throughout the disease's consecutive stages. As NDs develop slowly, patients require external care and lose a large part of their independence in consequence of the neurodegeneration, constituting an increasingly high cost for society. For instance, the cost for formal care for an AD patient annually is on average 27,672 USD[7], which tends to get higher as the disease progresses over the course of 8-10 years. To put this number into perspective, the UK National Health Service (NHS) spends around 112 billion GBP annually on neurological conditions[8] compared to 5 billion GBP on cancer care[9] and 7.4 billion GBP on cardiovascular diseases[10] (***Figure 1c***). Although cardiovascular diseases are still the leading causes of death worldwide[11], in 2022 the UK's highest number of deaths is attributed to dementia and AD, likely due to the increasing impact of NDs on an ageing population[12] (***Figure 1b***). The UK numbers are staggering and representative of what the larger world picture looks like (***Figure 1d***[13]), which reveals a similar effect in the rest of Europe and the US. Although the life expectancy of people affected by NDs has increased, the number of new deaths grows by millions yearly due to a lack of strategies for therapeutic intervention (***Figure 1e***[14]).

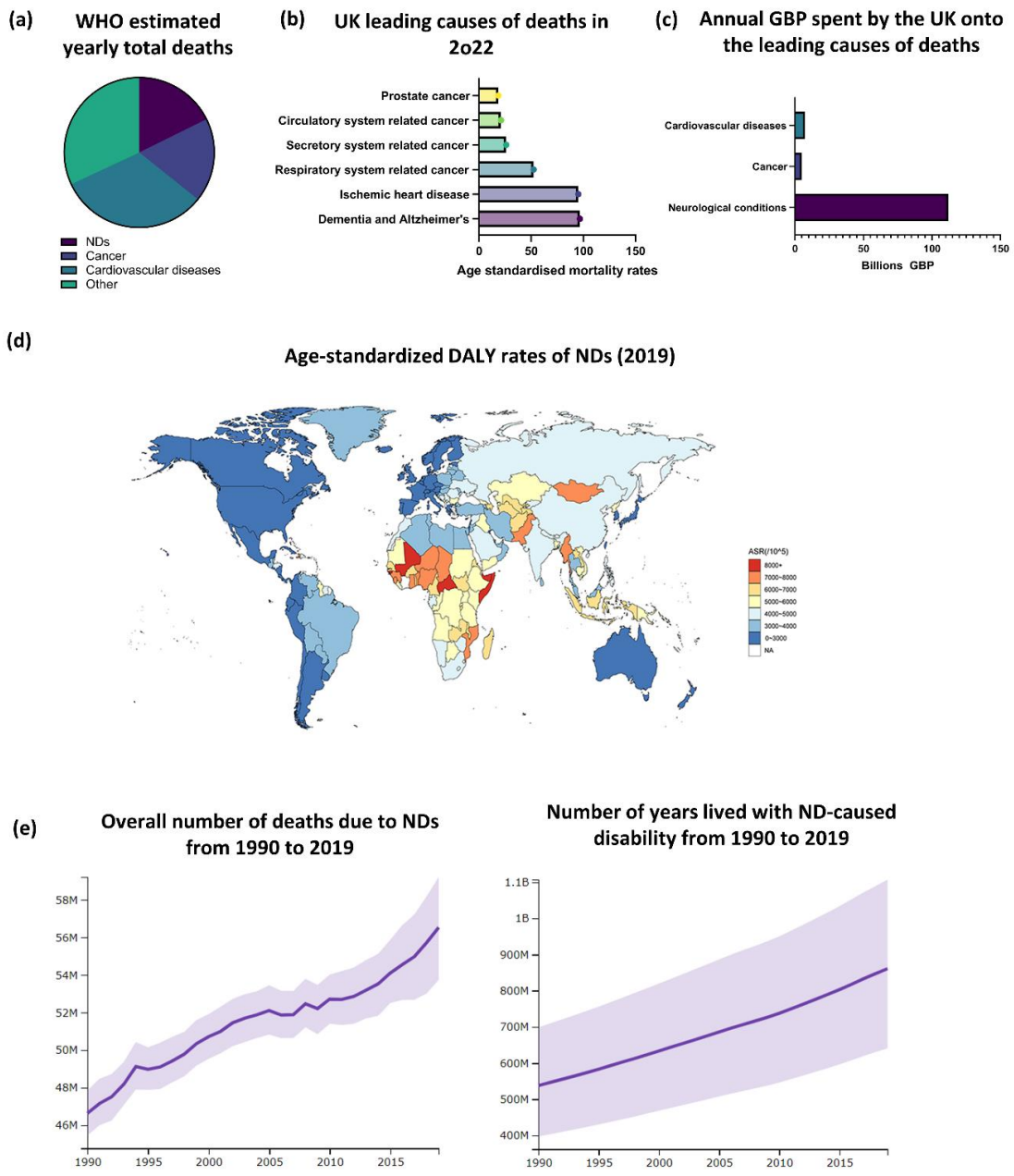


Figure 1: Socio-economic impact of neurodegenerative diseases. (a) Number of deaths by NDs in compared to cancer and cardiovascular diseases, the other two leading causes of death worldwide (55.4 million deaths worldwide) (as estimated by the WHO) **(b)** Leading causes of deaths in the UK in 2022 according to Census data. **(c)** Compared to the money spent by countries for ND vs cancer vs cardiovascular diseases. **(d)** World map cases of NDs. DALY (loss of the equivalent of one year of full health). Figure taken from [15]. **(e)** Overall number of deaths due to NDs from 1990 to 2019 compared to the number of years lived with the ND disease prior to death. Calculated using the GBD online data[14].

These figures show how the search for a treatment for NDs is one of the biggest challenges and priority of the 21st century. But although the government invests so much money into these diseases already, no viable cure has yet been found.

Indeed, although many resources are put at the service of NDs, most of the funds are utilised for long-term care of the patients. It has been estimated that for each 1 USD spent on NDs, only 1% goes towards research[2]. After taking these expenses into account, the US research-only budget for NDs ends up being 1.8 billion USD, less than the resources for cardiovascular diseases (2.2 billion USD) and cancer (6.0 billion USD)[2]. Lack of funds for research implies less workforce, lesser access to clinical data and resources and a lesser understanding of the underlying mechanism for NDs: all necessary for the development of an effective treatment. Furthermore, whilst in other cases – such as the COVID-19 epidemic[16] – pharmaceutical companies could easily approach the market and competitively offer treatments to the wider public, the development of a cure for NDs is too time consuming and uncertain to be a "cost-effective" investment for pharmaceutical companies (although 70% of the total clinical trials for AD are still sponsored or cosponsored by the pharmaceutical industries in the US[2]).

And research costs are not the only hurdle.

"Neurodegenerative disease" is a very general definition for a class of illnesses that include over 600 pathologies[17]. They all share the common neurodegeneration feature, but their causes, symptoms and development are, albeit somewhat interlinked, incredibly varied. Each pathology needs to be addressed singularly and a complete model for the disease is required for an effective treatment development (***Figure 2***).

Consequently, current ND research is incomplete and occasionally contradictory. Indeed, in June 2021, the drug Aducanumab was approved by the FDA in the US *via* the fast-track approval route for AD, despite the lack of strong scientific evidence to support its effectiveness[18]. Despite this, the FDA deemed necessary approving what was then the only treatment for a deadly disease, but their view was not shared by the European Medicines Agency, which denied its commercialisation in the EU continent[19].

In summary, NDs are the leading cause of death in most industrialised countries but remain underfunded and untreatable.

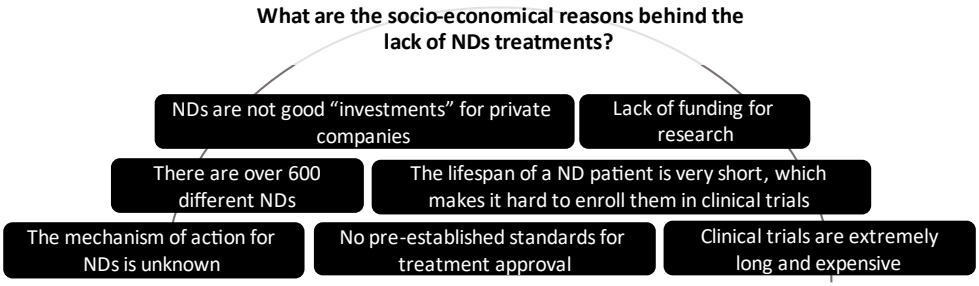


Figure 2: Socio-economic reasons behind lack of treatment for NDs. The underlying reasons behind the lack of treatment for NDs are given by a combination of lack of funding for research and poor understanding of the mechanisms of action of the diseases.

This work will focus specifically on Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD), two fatal neurodegenerative diseases that affect the motor system and cognitive behaviours with consequences for execution of functions, behaviours and language abilities[20,21] . ALS affects the motor neurons in the frontal cortex, brainstem and spinal cord, while FTD is related to motor neuron disorder, the most common type of dementia in people below 65[22]. ALS is found in approximately 2.1 cases per 100,000 people[23], whereas FTD has an incidence of 4 cases per 100,000[24] (***Figure 3a***).

ALS and FTD, much like other NDs, lack characterisation and are an increasingly heavy economic and social burden. The average life expectancy after disease onset is 1-4 years for ALS and 2-20 years for FTD[25,26].

Depending on the region of provenience, the total annual cost for an ALS patient is on average 69,475 USD, which in the USA sums up to be 279-472 million USD yearly[24] (***Figure 3b***). This figure has been estimated to be significantly higher than other NDs due the invasive nature of the motor neurodegeneration and the length of the disease. In fact, much like in the case of other NDs, there is currently no cure for ALS. The more severe the disease stage is, the more care the patient will require and the costlier their care will be. But what puts ALS over the edge in terms of annual costs, is a combination of direct and indirect costs, which are both higher than for other NDs. The direct costs are "expenses requiring payment" such as medications, medical equipment, home and vehicle adaptations, medical devices, mobility aids and services[25]. These resources can be provided either through direct contributions from the patient or through intermediaries like government agencies or health insurance providers. Indirect costs are "collateral" costs linked to the money "lost" by the patient or the patients' immediate relations in consequence of the disease. As the pathology progresses, ALS patients lose independence and eventually mental capabilities, which impairs their

ability to live on their own or (if applicable) to continue their jobs. Their long-term caretakers are often close relations, who also have been shown to lose career progression opportunities, number of hours at work and who require on average more health care in consequence of their constant state of stress. In fact, it is very common for ALS patients' families to employ a second full-time caretaker to release some of the burden, which is an additional cost[25]. But although the indirect cost has perhaps a harsher social impact, it has been calculated that most of the ALS expense relates to medical equipment, especially assisted respiratory equipment, which is necessary in the latest stages of the diseases[25]. ALS is considered a fairly rare disease, not due to its incidence (which is higher than other NDs), but due to the low life-expectancy there are not a lot of people living with ALS on a yearly basis. Despite that, a big step in raising awareness for the disease has been taken in 2014 with the Ice Bucket Challenge, the social media phenomenon in the USA which generated over 100 million USD for the ALS American Association (compared to the usual annual 2.8 million USD)[27].

FTD is the most common type of dementia in people < 60 y.o. As it tends to affect people in a younger category and for a longer period, it has been estimated to constitute a social and economic burden higher than Alzheimer's disease and ALS[26]. In fact, the total cost of an FTD patient is on average 119,653 USD, which varies depending on the age at which the disease is developed[26]. Much like in the case of ALS, the costs of FTD can be divided into direct and indirect costs. And whilst the direct medical expenses for FTD patients tend to be less significant than ALS, the indirect burden is higher. This is a result of the average age of onset of the diseases, a time where patients are usually in the prime years of their careers and have families. Caretakers sometimes need to be employed not just for the patient, but for their children too. It has been estimated that the cost of a young-onset of dementia is almost double what an old-onset of dementia would be[26].

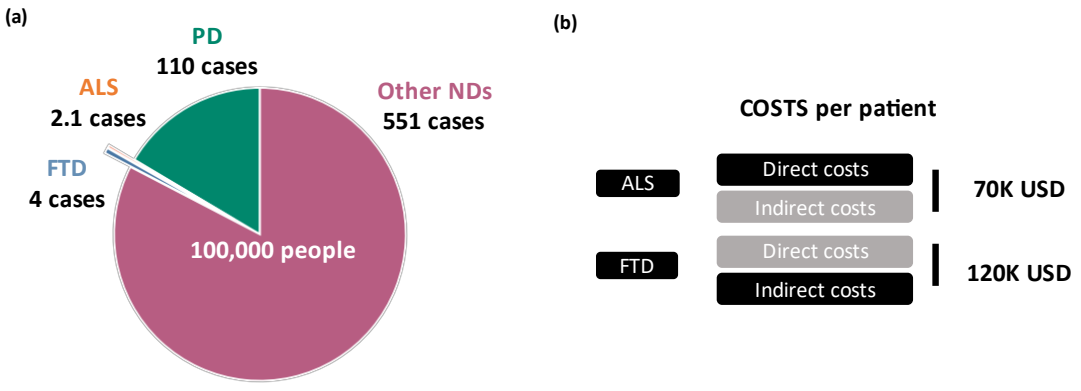


Figure 3: Socio-economic impact of ALS/FTD. (a) Incidence of ALS/FTD per 100,000 people compared to all forms of dementia[28]. **(b)** ALS/FTD costs per patient. Highlighted in dark are the most significant contributors to the final cost depending on the disease estimations.

1.1.2 Current treatments

Although no definitive treatment has been developed for any neurodegenerative disease, the field has shown significant improvements in increasing the lifespan of people affected by the pathologies.

The development of a final treatment for NDs is extremely challenging for a multitude of reasons, ranging from poor understanding of the underlying diseases mechanisms to the expensive long clinical trials required for the drugs acceptance[29,30]. Indeed, it has been shown that drugs that affect the symptoms do not necessarily address the root of the diseases or vice versa.

To aid in the approval of effective drugs, the European Medicine Agency has developed guidelines for drug-development for NDs, defining as a "disease modifying effect" the action of a treatment on a specific ND hallmark biomarker with consequent reduction of symptoms[29,30]. Of course, given that no biomarker has been officially confirmed in humans, these regulatory restrictions are a hard barrier to surpass.

Nevertheless, there are several drugs currently in clinical trials which target one or more of the recognised ND hallmarks with reduction of symptomatic behaviour (***Table 1***). For instance, Dextromethorphan, an anti-tussive agent in cough medicines, has shown reduction in glia-mediated neuroinflammation *via* inhibiting production of inflammatory mediators and is currently in phase II clinical trials[31]. Donepezil, Rivastigmine, and Galantamine are all acetylcholinesterase inhibitors, which prevent cholinesterase enzyme from breaking down neurotransmitter acetylcholine[32] and are used to manage symptoms of AD patients[33]. In all cases, the drugs slow the progression of the disease, but they do not constitute cures. Other drugs have been commercialised for other purposes but have shown slowing of neurodegenerative symptoms such as Dextromethorphan (cough suppressant approved by the FDA in 1958 with an unknown mechanism of action), Palmitoylethanolamide (present in egg yolks, peanuts and soybeans and has shown neuroinflammation protective mechanisms), Citalopram (an antidepressant which works as a serotonin reuptake inhibitor), Resveratrol (an antioxidant present in red-wine shown to have anti neuroinflammatory effects), and others[33].

Possibly the greatest breakthrough in the field in the past twenty years was the FDA approval and commercialisation of the drug Aducanumab for AD in 2021, the first ever treatment which claimed to stop the progression of the disease[34]. Aducanumab is an amyloid-beta directed monoclonal antibody which targets aggregated soluble oligomers and the insoluble fibrils formed by meaning the Aβ

plaques in AD affected neurons. Two clinical trials were performed prior to the fast-track approval by the FDA, but they brought forward conflicting results. In fact, while a reduction of plaques was observed, there was no consensus of the symptomatic effects. Indeed after the controversial approval, three FDA advisory board members resigned[34].

Only a year later, in January 2023, the second ever treatment for AD was approved, Lecanemab – still using the fast-track approval route by the FDA. Lecanemab is also a monoclonal antibody and has shown more robust evidence of slowing dementia progression *via* robust clinical trials for patients in early stage AD[35]. In both cases, it will be a few years before the general population will understand the effectiveness of these drugs as no long-term clinical trial has been performed.

Although there have been several developments and promising innovations for the treatment of NDs, no officially internationally recognised solution has been found.

Drug name	Target	Disease modifying ?	International approval	Disease
Dextromethorphan	Neuroinflammation	TBD	Pending	NDs
Donepezil and similar	Acetylcholinesterase inhibitors	No	Yes	AD
Dextromethorphan	Unknown	No	Yes (for other purposes)	NDs
Palmitoylethanolamide	Neuroinflammation	No	Yes (for other purposes)	NDs
Citalopram	Serotonin reuptake inhibitor	No	Yes (for other purposes)	NDs
Resveratrol	Neuroinflammation	No	Yes (for other purposes)	NDs
Aducanumab	Plaques	TBC	No	AD
Lecanemab	Plaques	TBC	No	AD

Table 1: List of current treatments for NDs. For each disease are highlighted their biological target, whether they have a recognised disease modifying effect, if they are internationally recognises and which disease they have been developed for.

Much like in the case of other NDs, no cure has yet been approved for therapeutic purposes for ALS/FTD patients. However, a series of drugs are available on the market/in clinical trials with the aim to slow down the disease progression and aid with symptoms management (***Table 2***).

Currently, only two drugs have been approved by the FDA for treatment of ALS and FTD, and in both cases they are only successful in slowing down the symptom's progression. Riluzole was approved in

1995 as the first ever ALS-modifying drug and it has been shown to extend life expectancy by 2-3 months per year[36]. Although the exact mechanism of action is not fully understood, Riluzole's main pathway involves blocking the glutamatergic excitotoxicity and interaction with other neurotransmitters[36]. Over two decades later, the American FDA approved a second treatment, Edaravone, which was also shown to slow the disease progression. The mechanism of action involves neuroprotection *via* scavenging of free radicals (antioxidant properties)[36]. Other drugs currently between Phase II and Phase III clinical trials are: Arioclomol (targets TDP-43 aggregation amongst other issues), Ceftriaxone (an antibiotic with induction of glutamate transporters able to cross the blood brain barrier), Masitinib and cannabinoids (anti-inflammatory action), Tirasemtiv (affects calcium homeostasis) and stem-cell based approaches[36].

More recently, several treatments have been developed by using $(GGGGCC)_{n,}$ a common mutation in ALS/FTD, as therapeutic target. In particular, antisense oligonucleotides targeting of the gene was shown to reduce the RNA-nuclear *foci* associated with the ALS/FTD phenotype in patients' cells[37,38] leading to pre-clinical studies in humans[39,40]. One patient presenting $(GGGGCC)_{2400}$ was treated with antisense therapy over a course of several months and showed an 80% decrease in poly(GP) biomarker (showing expression of C9orf72)[37]. Whether this is connected to complete remission from the symptoms is still uncertain, but it shows promise for therapy and treatment.

Drug name	Target	FDA approval
Riluzole	Glutamate toxicity	Yes
Edaravone	Oxygen free radicals	Yes (USA)
Arioclomol	TDP-43	No
Ceftriaxone	Glutamate transporters	No
Masitinib	Anti-inflammatory	No
Tirasemtiv	Calcium homeostasis	No
Antisense therapeutics	C9orf72	No

Table 2: List of current treatments for ALS and FTD and their biological target.

1.1.3 Classification of NDs from clinical observations

Although the mechanistic progression of NDs is not fully understood, clinical observations of symptomatic patients and post-mortem biopsies have allowed doctors to classify NDs based on common features. Of over 600 ND pathologies[17], the most common NDs with recognisable hallmarks are Alzheimer's disease (AD), Parkinson's Disease (PD), Frontotemporal Dementia (FTD) and Amyotrophic Lateral Sclerosis (ALS), Huntington Disease (HD), Spinocerebellar ataxias (SCA), Lewy body dementia (LBD) and Multisystem Atrophy (MSA) (both known as synucleinopathies), prion disease (PrD), Multiple Sclerosis (MS), Chronic Traumatic Encephalopathy (CTE), Traumatic Brain Injury (TBI), stroke and Spinal Cord Injury (SCI)[41].

The NDs' primary features are highlighted in ***Table 3***. CTE, TBI, stoke and SCI are not explored in detail as, although presenting neurodegeneration, the triggering event is external and recognisable. Interestingly, however, the final stages of neurodegeneration-affected biopsies of these conditions are very similar to previously mentioned diseases. For instance, TBI and AD both share the presence of β-amyloid deposition and the cerebrovascular pathology element[42], which led questioning how two completely different causes, genetic mutations and environmental factors, can lead to the same neurodegeneration effect.

ND name	Early symptoms	Clinically observed neurodegeneration in the neurons	Brain regions affected
Alzheimer's disease	- Language function difficulties - Memory impairment	- Senile plaques - Neurofibrillary tangles - Brain atrophy	 Starts in the hippocampus
Parkinson's disease	- Tremors - Change in the ability to walk	- Loss of substantia nigra - Lewy bodies	 Basal ganglia
Huntington disease	- Memory impairment - Changes in behavior - Lack of movement control	- Accumulation of lipofuscin in neurons and glial cells	 Basal ganglia
Spinocerebellar ataxia	- Problems with balance and coordination - Muscle stiffness and cramps - Memory loss	- PolyQ aggregates	 Cerebellum
Multiple System Atrophy	- Dysregulation of body autonomic functions - Lack of motor control	- General degeneration in the striatum	 Cerebellum
Prion Diseases	- Rapidly developing dementia - Hallucinations - Difficulties speaking - Difficulties walking	- Pathological infectious protein aggregates	 Cerebellar cortex
Multiple Sclerosis	- Motor and coordination problems - Mental impairment	- Damage to the myelin sheath surrounding axons	 Periventricular, brainstems, cerebellum
Amyotrophic Lateral Sclerosis	- Muscle twitches - Muscle cramps - Tight and stiff muscles	- Pathological aggregates in the motor neurons - Basal ganglia atrophy	 Primary motor cortex
Frontotemporal Dementia	- Personality changes - Language/speech problems - Memory impairment	- Pathological aggregates in the motor neurons - Small vessel ischemia - Atrophy in the frontal or anterior lobes	 Frontal and temporal lobes

Table 3: Symptoms and clinical observation of neurodegeneration for the most common ND pathologies. The table highlights the early symptoms observed, the first clinically recognized signs of neurodegeneration at the brain level and the initial location of the observed neurodegeneration. Hand-drawn illustration with Procreate.

ALS and FTD present different symptomatic behaviours in the early stages, but both diseases share clinical, neuropathological and genetic features which indicate either a common neurodegenerative pathway or belonging to a same neurodegenerative spectrum[20] (***Figure 4***).

The term Amyotrophic Lateral Sclerosis (ALS) was coined in 1869 by neurologist Jean-Martin Charcot referring to patients with loss of muscle mass (amyotrophic), degeneration on both sides of the spinal cord (lateral) and scar tissue-like consistency in the spinal cord[43,44]. Although ALS has been mainly described as motor neuron disease, it is widely recognised to affect other areas of the brain over time[20]. Most ALS cases are sporadic (non-inherited), about 5% have a family history of ALS (familial ALS)[45]. From a genetic viewpoint, most ALS patients present at least one of the following mutated genes, which are now used by clinicians to help in the diagnosis. These genes include SOD1, C9orf72, FUS and TDP-43 plus another series of less common variants[46]. Importantly, sometimes the downstream effects of ALS are also noticed in absence of the mutations, implying that these might be a consequence in the mechanism and not the causation.

FTD was first observed by the neuropsychiatrist Arnold Pick in 1892, who noticed the presence of spherical argyrophilic inclusions accumulated in the neurons of the frontal lobe of a patient presenting signs of dementia. He crafted the term “Pick Bodies” and called the disease “Pick Diseases”[47]. It was later observed that Pick Bodies were not they only histological hallmark in dementias in the frontotemporal lobe and Pick disease was renamed Frontotemporal Dementia (FTD)[20]. Unlike ALS, about 40% of patients with FTD have a family history of dementia, but familial and sporadic FTD types present the same clinical features[48]. The first symptoms are usually of behavioural nature and involve dysfunction in self-monitoring, evaluation of internal mental states, awareness of emotion, sustaining personality and production of language[49]. From a genetic viewpoint, the majority of genetic FTD is caused by mutations at C9orf72, MAPT or GRN. Occasionally, it is also observed upon mutations at TDP-43, VCP, CHMP2B, SQSTM1, UBQLN1 or TBK1[50]. Much like in the case of ALS, the downstream effects of FTD can occur in absence of these mutations.

The first associations between the two diseases were made in 1981, when it was reported that patients with ALS also showed FTD signs, and in 1987 when FTD patients was reported to share similar morphological aspects with ALS ones[20].

Currently, it has been validated that up to 50% of ALS patients will develop cognitive impairment associated with FTD and up to 30% of FTD will develop motor dysfunction[22]. The link between the two diseases is not only recognised from these clinical observations, but also from overlapping in their genetics. In fact, both diseases present mutations in TANK-Binding Kinase 1 (TBK1)[51], Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 10 (CHCHD10)[52], Valosin Containing Protein (VCP)[53], Fused in Sarcoma (FUS) protein, Sequestosome-1 (SQSTM1)[54], and Chromosome 9 Open Reading Frame 72 (C9Orf72)[55]. In particular, the hexanucleotide expansion $(GGGGCC)_n$ in C9orf72 is the most common

cause of 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)[56] in Europe.

Despite the numerous clinical observations on the two diseases, it is still not clear how these interlink with each other and how they give rise to the dementia symptomatic behaviours.

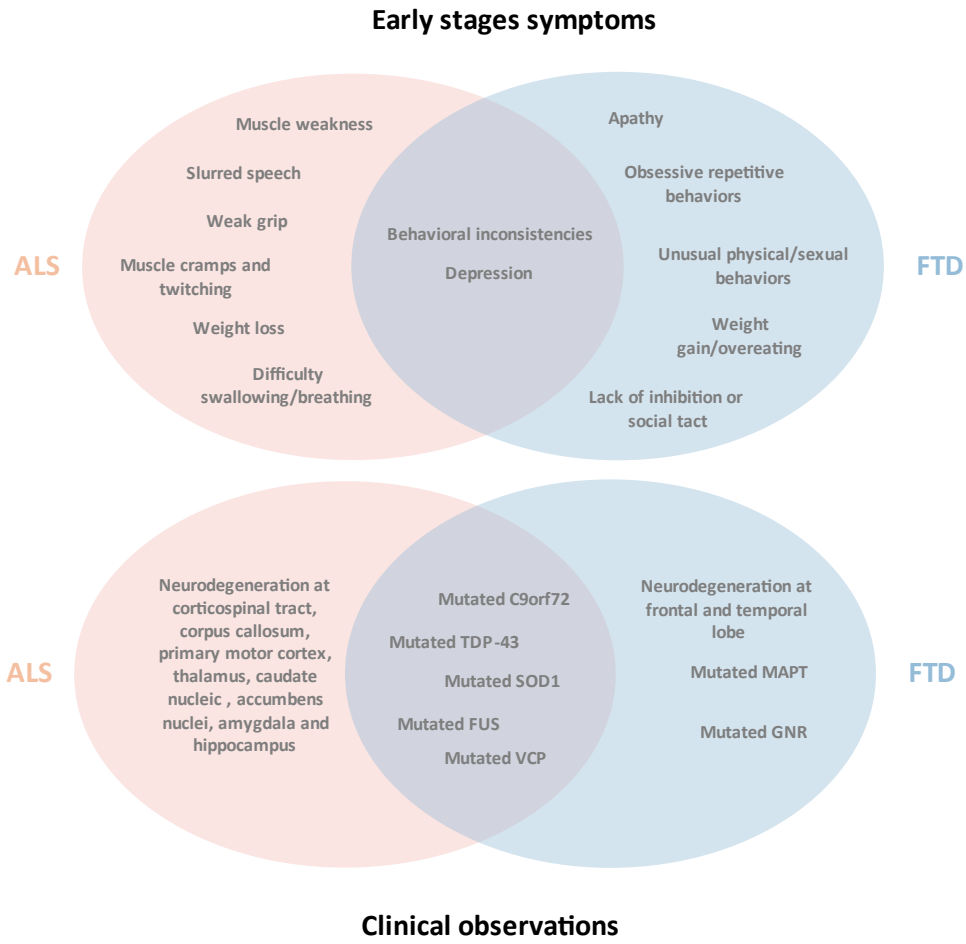


Figure 4: Overlapping symptoms and genetics of ALS and FTD. Although different in their early symptomatic stages, ALS and FTD share symptoms in their later stage. They are considered a neurodegenerative continuum due to the presence of several overlapping genetic features.

1.1.4 Molecular hallmarks of neurodegeneration

Table 3 shows how different regions of the brain which are affected by neurodegenerative pathways for several pathologies. But despite these discrepancies in neurodegeneration locations

from biopsies, it is still possible to find a series of hallmarks on a molecular level that are shared amongst many NDs.

These are: pathological protein aggregation, synaptic and neuronal network dysfunction, aberrant proteostasis, cytoskeletal abnormalities, altered energy metabolism, impaired DNA damage response and impaired RNA biological function, inflammation and neuronal cell death[41] (***Figure 5***). The relevant hallmarks for this work are highlighted in more detail in later paragraphs.

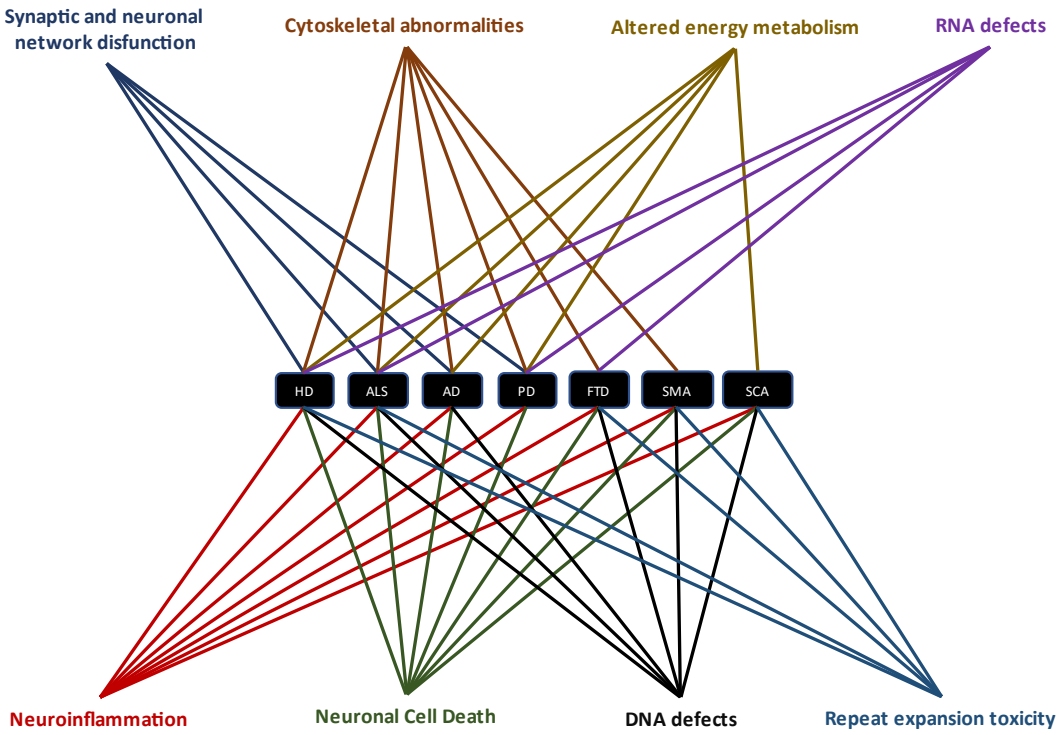


Figure 5: Molecular hallmarks of the most common NDs. As evident from clinical mapping of the hallmarks of most common NDs, these are thoroughly interlinked, making it very complex to understand the roots of the pathology versus the effect.

1.1.4.1 Aberrant proteostasis

Proteins can aggregate under different circumstances and assume several morphologies depending on their role[57]. Aggregation is encouraged when the protein requires to protect its hydrophobic core from exposure to a hydrophilic environment.

The most common protein aggregates morphologies found *in vivo* are: **(1)** amorphous protein aggregates, **(2)** globular protein aggregates and **(3)** well-ordered fibrillar structures (or amyloid fibres)[57]. The formation of amyloid fibres depends on a series of factors, ranging from the amino acid

sequence to environmental effects. For instance, proteins with a β-sheet folding propensity tend to form fibrils with more ease[57]. Instead, proteins with no defined secondary structure which present low complexity domains (LCD), are able to phase-separate due to the presence of unspecific hydrostatic interactions[58]. In fact, it has been hypothesised that most proteins possess the potential to aggregate under the right conditions *via* means of high temperature, high pressure, low pH, alcohols, crowding conditions or centrifugation and samples shaking[58]. *In vivo*, the presence of protein aggregates is often related to pathologies (***Figure 6***). For instance, type II diabetes is correlated with the presence of islet amyloid polypeptide aggregates[59] and amyloid transthyretin cardiopathy presents tetrameric thyroxine transport protein transthyretin amyloid fibrils[60].

In the case of NDs, the presence of pathological aggregates made up of misfolded or accumulating proteins at the neuron level is caused by an alteration of proteostatis. Two of the major regulatory mechanisms of action involved in proteostatis are the ubiquitin-proteasome system (UPS)[61] and the autophagy-lysosome pathway (ALP)[62].

UPS is responsible for ubiquitination of proteins and subsequent removal from the system. If the key elements of the UPS machinery are mutated, accumulation of pathogenic protein aggregates can occur. This is the case for UBQLN2 and VCP proteins for ALS/FTD[63] and Parkin for PD[64].

ALP detects and clears pathogenic protein aggregates and misfunctioning organelles by encapsulating them into the autophagosome. Mutation of autophagy-related genes such as GBA, GBA1, LRRK2, TMEM230, Parkin (PD), GRN (FTD), SNX14 (SCA), APP, PS1(AD), C9orf72, SQSTM1, TBK1 (ALS/FTD), HTT (HD) has been linked to the specified NDs[65].

Notably, that protein aggregation is not necessarily correlated to a pathological phenotype (***Figure 6***). In fact, cells can use aggregation events to sequester damaging proteins blocking their effects, or to provide a scaffold for other biological functions. For instance, functional protein aggregates have been connected to storage of peptide hormones, reproduction and fertilisation, pigmentation, necroptosis, antimicrobial response, adaptation to stress, cellular dormancy, bacterial biofilm formation, regulating fungal-host and fungal-fungal interactions, modulating epigenetic heritable phenotypes in yeast and the persistence of long-term memory in *Drosophila*[66].

How cells trigger protein aggregation and when that leads to pathologies is not fully understood, making it impossible to disentangle the cause/effect of aberrant proteostasis with respect to the pathologies and weather this can be considered a suitable target for treatment of NDs.

Interestingly, the only FDA approved drugs (Aducanumab[34] and Lecanemab[67]) for AD do in fact target protein aggregates (formed by Aβ aggregates at the neuron level), but their efficacy in aiding a patient's recovery is still debated[18].

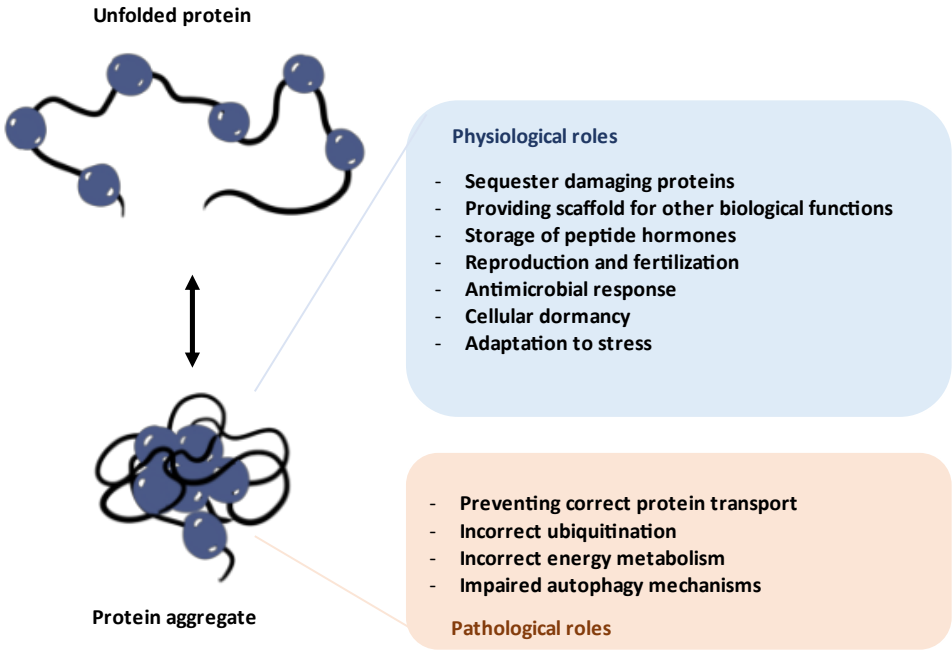


Figure 6: Biological consequences of protein aggregation. On the left – schematic of protein aggregation. **On the right** - highlighted principal pathological and physiological roles of protein aggregates in cell biology. Hand-drawn illustration with Procreate.

1.1.4.2 Impaired DNA damage response

Post-mortem brain biopsies of ND patients show an increased oxidative DNA damage in nuclear and mitochondrial DNA, which led to impaired DNA damage response-mechanisms to be recently added to the list of potential hallmarks for these diseases[68].

When a faulty neuron is damaged, it cannot be replaced. Irreplaceability is why this cell type relies on efficient DNA damage response mechanisms to avoid toxic genomic rearrangements, transcriptional dysregulation and accumulation of unrepaired lesions which could lead to irreversible damage[69]. Five possible damage pathways have been highlighted in literature, showcasing the cause/effect relation of failing DNA damage repairs mechanisms to other NDs hallmarks (***Figure 7***).

Firstly, several NDs patients present mutations in genes that encode for proteins responsible to aid upon DNA-damage, in particular, double stranded (ds) breaks and single stranded (ss) breaks. This occurs for the pathogenic mutation at TDP1, responsible for resolving ss breaks at TOP1 activity sites, which is present in nervous systems deficits such as ataxia, neuropathy and cerebellar atrophy[69]. Similarly, mutations at TDP2, a ds breaks resolver, are found in certain types of SCA[69]. Furthermore, commonly mutated proteins associated with ALS/FTD (TDP-43 and FUS), HD (HTT), SCA (ATXN2) or SMA (SMN1/1) are also involved in the various DNA damage response mechanisms[41].

However, an incorrect DNA damage response in not necessarily introduced by a genetic mutation. In persistent DNA damage (genotoxic stress) a series of signalling cascades causes cell apoptosis to prevent replicating a damaged genome[70].

The third damage pathway involves failure of the transcription machinery, often correlated to NDs *via* the presence of transcription-blocking lesions. In the multi-level DNA damage response mechanism, the RNA polymerases are stalled and transcription is prevented[71]. This mechanism is not limited to NDs and is often observed in types of ataxias and ALS. In particular, in juvenile types of ALS the formation of RNA-DNA hybrid R-loops (or other multimolecular structures) has been correlated to the failure of the transcription machinery[41].

Another pathological pathway involving DNA damage is preceded by mitochondrial disfunction and altered energy homeostasis. In fact, mitochondrial disfunction leads to the formation of reactive oxygen species (ROS) which have the ability to attack DNA to form toxic oxidized bases and abasic sites[69]. The effect of ROS is particularly damaging in neurons as they consume approximately 20% of all the body's available oxygen[72].

Finally, the last pathological pathway relates to the incorrect DNA damage accumulation at the neuron level. In fact, there have been several studies on the positive effect of DNA damage on the plasticity of the neurons, where DNA breaks are suggested to have a functional role in early neuronal activity in mice studies[73,74]. A mis regulation of these effects could lead to toxic loss of function of the neurons and consequently neurodegeneration.

In summary, DNA damage response is an essential element in neurons and lack thereof, either due to genetic mutations or as a downstream effect of other NDs, is linked to neurodegeneration. Targeting DNA damage repair mechanisms has been considered a viable option for new ND therapeutics, especially since similar treatments are now being heavily developed for cancer treatment[75]. However, it is unclear whether erroneous DNA damage response is causing or is a consequence of other NDs hallmarks.

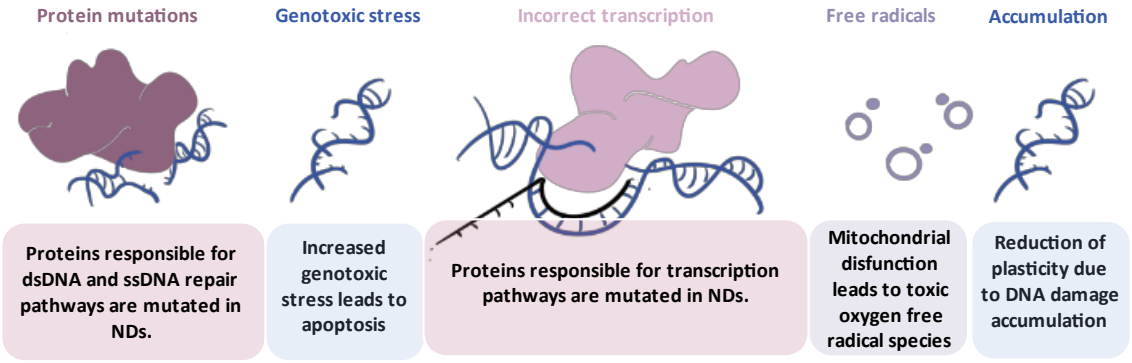


Figure 7: DNA damage pathways associated with NDs. The main five DNA damage pathways by which NDs have been isolated are highlighted in the figure. **1.** Proteins responsible for DNA damage responses are commonly mutated. **2.** ND patients have shown to present an increased genotoxic stress leading to dsDNA breaks. **3.** Proteins responsible in the translation of DNA into RNA are mutated. **4.** Mitochondrial disfunction, often observed in NDs, leads to the formation of toxic oxygen free radicals. **5.** It has been observed that accumulation of toxic dsDNA and ssDNA damaged species leads to a reduction of neuroplasticity and consequently health of the neurons. Hand-drawn illustration with Procreate.

1.1.4.3 Impaired RNA biological functions

Similarly to DNA, impaired RNA biological functions have also been associated with several NDs *via* several pathological pathways. These can be classified in three main classes: **(1)** general RNA toxicity, **(2)** generation of stress granules (SGs)[41] and **(3)** toxicity *via* the formation of repeat sequence networks[76] (***Figure 8***).

General RNA toxicity refers to all the downstream consequences created by the over or under-translation of RNA. For instance, in ALS/FTD, mis-localised TDP-43 leads to altered RNA splicing, stability and transport[77] and mutations of the FUS protein cause mis-localisation into the cytoplasm[78]. In both cases, RNA is mislocalised and its biological function altered, leading to unwanted downstream consequences.

Stress granules can be defined as condensates of ribonucleic proteins and mRNA formed in the cytoplasm from stalled translation initiation in consequence of a stress response[79]. Their functional biological role is to collect the unused mRNA until the translation process can restart outside of the stress response. During neurodegeneration, the neuronal cell is under constant stress due to the other hallmarks, and the stress granules cannot disassemble. It is unclear whether it is a misregulation of the stress granules response to lead to RNA defective pathways or if stress granules are just a consequence of the stress response to other factors. In fact, several RNA-binding proteins often

misregulated in NDs also have defining roles in stress granule generation. This is the case for the previously mentioned TDP-43 and FUS (ALS/FTD), G3DP1 (HD)[80], DJ-1 (PD)[81], and others. As further confirmation of the pathological nature of stress granules in ND patients, these have been examined *via* FRAP studies, presenting a compromised plasticity[82].

The final mechanistic pathway is the most relevant in the context of this work and relates to repeat sequence expansions in neurodegenerative pathologies. In fact, more than 40 neurodegenerative pathologies have been related to the presence of expansion repeats throughout the human genome[83]. Historically, the toxicity of these repeat expansions has been associated to the impaired function of the corresponding translated protein[84]. However, recent evidence has highlighted how the RNA sequences themselves can also have pathogenic implications in forming multimolecular networks.

Although the role of RNA toxicity in the context of other NDs remains unclear, RNA-targeting drugs have been recently put into clinical trials with some success. The aim of these RNA-based therapies is to target toxic RNA-transcripts by binding the Watson-Crick RNA strand and silence their downstream effect[85]. These preliminary results indicate that RNA has a direct effect in neurodegenerative diseases, making it a suitable target for drug discovery in the context of ND.

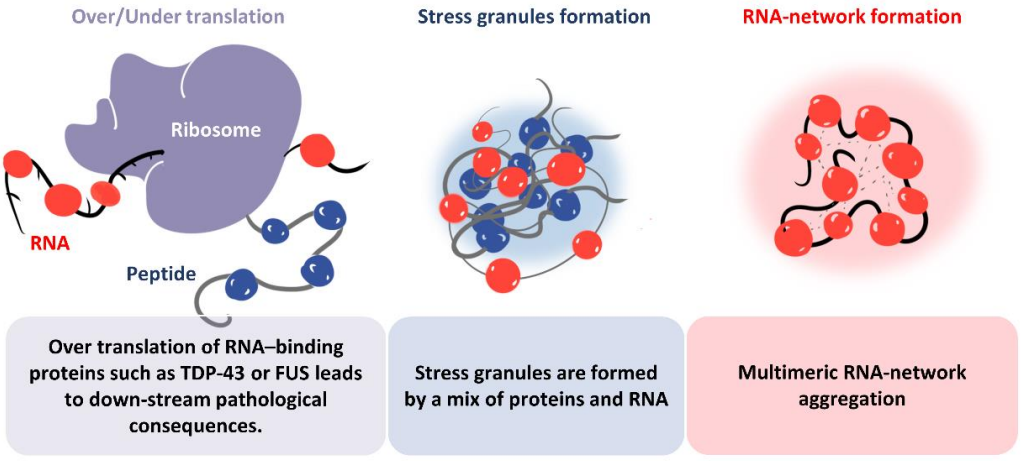


Figure 8: Impaired RNA biological roles associated with NDs. The three main toxicity pathways are illustrated in the figure: **1.** Over/under translation of RNA into proteins can lead to downstream pathological consequences such as mis-localisation of proteins like TDP-43 and FUS in the cytoplasm. **2.** Formation of stress-granules. **3.** Formation of solid multimeric networks due to alternative secondary structure assumed by the RNA sequences. Hand-drawn illustration with Procreate.

1.1.4.4 Repeat Expansion Toxicity

In this work we define as "repeat expansion (RE) toxicity" the toxic gain or loss of function of the RNA transcript due to the presence of mutated repeat expansion sequences in the genome. A RE is characterised by a number of bases B repeated *n* times (*i.e.* $(BBB)_n$). The RE is considered pathological if the number *n* goes above a threshold established by statistical studies.

There are several NDs which are genetically associated to the presence of pathological REs, and, although they differ in composition and size, it is possible to find common features that allow characterisation of this class of mutations as an ND hallmark.

The first REs observed in a pathological context were trinucleotide repeats such as $(CAG)_n$, $(CGG)_n$, $(CTG)_n$, $(GAA)_n$, $(GCC)_n$ and $(GCG)_n$, but it was soon discovered that the number of bases could vary quite significantly, even reaching dodecanucleotide REs[83]. Indeed, these occur in DM1/DM2 (myotonic dystrophy), C9orf72 (ALS/FTD), CAG repeats (HD and other polyQ disorders including ataxia and Fragile X Syndrome) and others[83] (***Table 4***).

But regardless of the number of bases, REs all present the following general characteristics: **(1)** a high degree of polymorphism, **(2)** although generally inheritable they can also be the cause of sporadic cases due to genetic mutation, **(3)** longer repeat lengths (high *n*) tend to correlate to a more severe form of the disease, **(4)** they present a highly variable phenotype due to the large differences in repeat length *n*[83].

It is still debated what are the characteristics that give repeat expansions their toxic behaviour.

Historically, the field has associated the presence of these repeated expansions to their protein translated product and protein toxicity (***Figure 9a***). In fact, $(CAG)_n$, one of the most studied REs, is also responsible for a class of diseases known as polyglutamine (polyQ) diseases[86]. In polyQ diseases, which include SMA, SCA and HD, $(CAG)_n$ has $n > 35$ and leads to the expression of a highly repeating protein glutamine polymer[86]. PolyQ proteins are prone to self-aggregation into amyloid fibrillar structures in neurons, leading to neurodegeneration. Similarly $(GCG)_n$, responsible for muscular dystrophy, is often referred to as a polyA disorder[87].

A second mechanistic hypothesis involves the toxic loss of function of REs translated mRNA (***Figure 9b***). This occurs when RE gene translation is impaired due to the formation of non-canonical structures in the sequences, leading to a reduced yield of mRNA and respective protein encoded. This occurs for

$(CTG)_{n,}$ which is linked to pathological muscular dystrophy when $n > 60$, leading to a decreased expression of the DMPK gene and lowered protein levels, hypothesised to cause its toxic effects[88].

Repeat expansion toxicity can also come from a toxic gain of function of the mRNA product (***Figure 9c***). This occurs when the mRNA is able to stabilise pathological aggregates formed by translocated proteins. As an example, the RE $(GGGGCC)_{n,}$ responsible for ALS/FTD when $n > 24$, has been shown to bind to TDP-43[89], FUS[90] and C9orf72 proteins stabilising pathological protein aggregates formed at the neuronal level.

Mutation	Pathological length n	Diseases
(CAG)n	n > 35	HD, SCAs, atrophies
(CGG)n	n > 62/95	Fragile X
(CTG)n	n > 50	Myotonic dystrophy, Huntington disease-like 2, SCA type 8
(GAA)n	n > 100	Friedreich ataxia
(GCC)n	n > 200	FRAXE mental retardation
(GCG)n	/	Muscular dystrophy
(CCTG)n	n > 75	Myotonic dystrophy
(ATTCT)n	n > 8	SCA type 10
(TGGAA)n	Differs	SCA type 31
(GGCCTG)n	n > 14	SCA type 36
(GGGGCC)n	n > 24	ALS/FTD
(CCCCGCCCCGCG)n	/	EPM1

Table 4: List of repeat expansion usually associated with pathological consequences. [90 91 92]

The final toxicity pathway relies on these sequences' ability to form multimeric networks condensates based on their highly repetitive nature. Indeed, the pathological aggregates found in ND-affected neurons are known to be composed of a mix of proteins and RNA[91]. But while the protein component has been widely investigated as the keystone element of aggregation, RNA has been discarded as a secondary element of stabilisation. This is because while protein can form multimeric networks and phase-separate, ssRNA cannot aggregate in its own right. However, higher molecularity RNA networks have been shown to trigger phase-separation in absence of additional proteins[76] and REs are widely

recognised to be prone to form non-canonical secondary structures – which instead have the potential to form such linkages (***Figure 9d***). For instance, it has been observed that $(GAA)_n$, related to ataxia, is able to form an intramolecular DNA triplex of supercoiled DNA *in vitro*, which could affect efficient DNA repair[92]. $(CAG)_n$ repeats instead, previously mentioned in the context of polyQ diseases, have been shown to arrange into hairpin conformations leading to incorrect gene splicing[93], which could lead or accelerate polyQ aggregation. $(GGGGCC)_n$, has also been associated with the formation of G-quadruplex structures that could also have downstream potential effects in the ALS/FTD[94].

This is not to say that protein aggregates do not contribute to pathologies, there is plenty of clinical evidence to support this, but collateral evidence suggests that RNA aggregates could also be playing an important role in pathological aggregation[76]. The idea that REs possess the ability to form aggregates on their own right could lead the way towards a new class of therapeutics targeting nucleic acids.

Importantly, RE can be detected before the onset of symptoms[83]. Although research in this area has just started, it offers incredible potential in treating diseases before the onset of neurodegeneration (***Figure 9***).

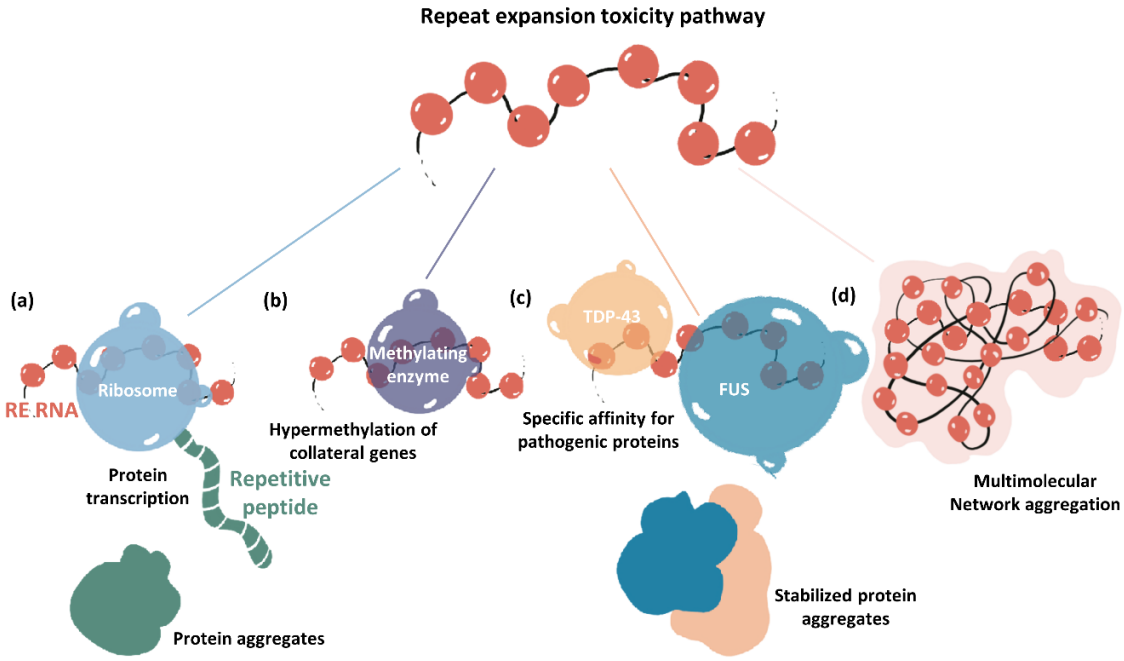


Figure 9: Repeat expansion toxicity pathways. (a) Repeat expansions are translated into proteins with repetitive units prone to aggregation. **(b)** The presence of repeat expansions affects other genes up or downstream. For instance, hypermethylation of collateral genes occurs near pathological repeat expansions. **(c)** Repeat expansions bind specific proteins which have shown the ability to aggregate. **(d)** Repeat expansions can form multimeric networks with the ability to aggregate on their own right. Hand-drawn figure generated with Procreate.

1.1.5 Mechanism behind ALS/FTD onset

ALS and FTD are associated with several interlinking NDs hallmarks, including synaptic dysfunction, altered energy or/and protein homeostasis, neuronal cell death, repeat expansion toxicity, neuroinflammation and cytoskeletal abnormalities.

Many mechanisms have been hypothesised for ALS and FTD, but here I will focus on the mechanistic pathways common for patients with ALS developing FTD in later stages of the diseases or vice versa, which have been found to associated with specific genetic mutations common in both pathologies. These include: MAPT (TAU-protein related)[95], PGRN (Progranulin)[96], CHMP2B (body protein 2B transport protein)[20], SOD1 (antioxidant SOD1 protein)[20], FUS (regulatory protein fuse in sarcoma)[97], TDP-43 and Chromosome 9, which will be explored in more detail in the following paragraphs.

1.1.5.1 TDP-43

TDP-43 is an RNA-binding protein encoded in the chromosome 1p36.22. It is a ubiquitous protein belonging to the hnRNP family, usually present in the nucleus and responsible for RNA regulation (transcriptional regulation, alternative splicing, and mRNA stabilization)[98]. In neurodegenerative pathologies such as ALS, FTD, PD and AD, the protein is accumulated in aggregates in the cytoplasm[98]. Over 38 pathogenic variants have been identified from screening ALS patients[20]. Amongst the most acknowledged, we can find M337V and Q331K, which possibly lead to a toxic gain of function, G290A and G298S, which are associated with both toxic gain and loss of function, G348C and N352S, which are predicted to increase TDP-43 phosphorylation, M337V, N345K and I383V, shown to contribute to neuronal apoptosis, K263E, which leads to neuronal and glial TDP-43 immunoreactive deposits and G294V, which influences TDP-43 exon skipping activity[20].

In terms of its structure, TDP-43 is composed of two RNA recognition domains (N-terminus domain) and a C-terminus low-complexity glycine rich domain which resembles prion-like proteins (***Figure 10***). The RNA binding domains (RBDs) are arranged as two α-helix and six β-sheets[99]. This region of the protein is what is responsible for nucleic acid recognition and consequently pre-mRNA splicing, translational regulation and other downstream effects. In fact, although TDP-43 is widely recognised as an RNA-binding protein, it has also been shown to efficiently bind as a dimer to single stranded and double stranded DNA (ss and ds DNA) and that nucleic acid binding stabilises the protein complex[99]. TDP-43 has also been shown to be able to bind, stabilise and transport non-canonical RNA secondary structures known as G-quadruplexes[101]. Mutations in these regions can therefore have a pathological downstream effect associated with toxic loss of function of the protein, given its inability to efficiently bind nucleic acids.

The C-terminus domain lacks any specific secondary structure due to the presence of a low-complexity sequence enriched in uncharged polar residues. Much like prion-like proteins, the low-complexity domain in TDP-43 is prone to aggregation and affects the solubility and folding ability of the whole protein. Of all the TDP-43 variants found in ALS patients, 85% are at the C-terminus region, indicating that that might be the major region of interest for potential therapeutics development[20]. Mutations at this region tend to alter the protein aggregation homeostasis and facilitate the liquid-solid transition of the protein aggregates, giving rise to pathological consequences and toxic gain of function of the protein. In fact, TDP-43 is the major component of the inclusions in motor neurons in ALS and some FTD cases[97].

Notably, a large number of ALS/FTD cases do not present any pathogenic TDP-43 variant and still develop motor neurodegeneration[102], suggesting that TDP-43 aggregation and mis localisation (regardless of the presence of the mutation) is a consequence of other ND hallmarks. One possible mechanism of action, relies on the presence of this protein within stress granules[98]. In fact, in healthy patients, TDP-43 is known to translocate from the nucleus to the cytoplasm to form stress granules with other proteins and RNA to respond to conditions such as a state of oxidative stress and changes in local biomolecules concentrations. But whilst in healthy patients the stress granules are released upon the end of the stress conditions, in NDs the stress granules become permanent aggregated resident in the cytoplasm. It is possible that even in absence of the TDP-43 mutation, other ND hallmarks cause a stress-response which makes these stress granules unable to disassemble. This evidence proves that although TDP-43 aggregation might play a role in the progression of ALS/FTD it is likely not the primary causation for most of the disease family. Furthermore, TDP-43 aggregation is not only present in ALS/FTD but in other NDs such as AD and PD, where TDP-43 aggregates are present in the cytoplasm[98]. This gives yet another indication that whilst TDP-43 mutations certainly contribute to the ND pathogenesis, they are not the triggering event but the consequence of other underlying issues which are specific to each disease.

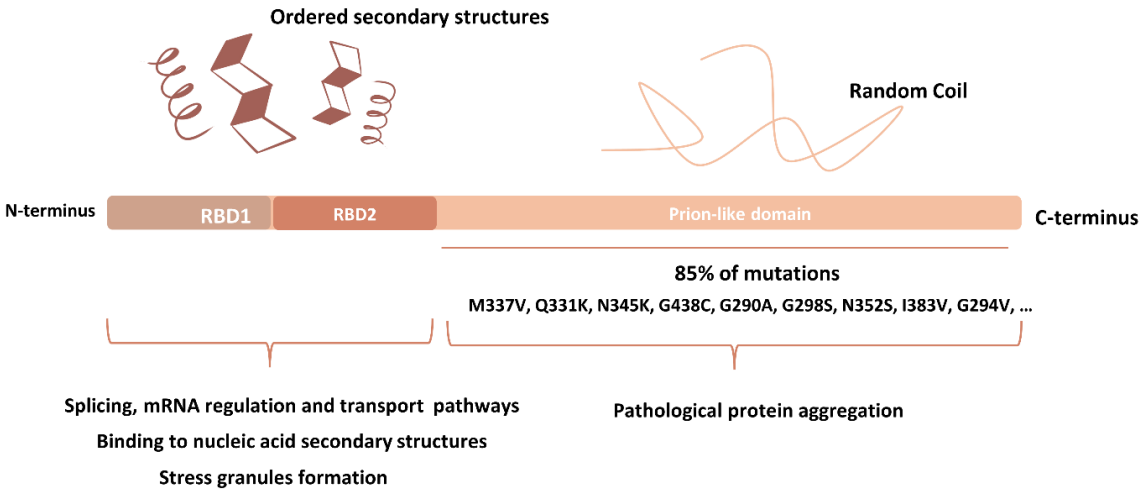


Figure 10: TDP-43 main domain structures, relevant ALS/FTD mutations and pathological pathways. TDP-43 is composed of two ordered RNA binding domains which are responsible for a series of biological functions including splicing and mRNA regulation, RNA transport, binding to nucleic acid secondary structures and stress granules regulation. The C-terminus domain is composed primarily of charged residues which resemble those of prion-proteins. They are unstructured and prone to aggregation. Most of mutations linked to ALS/FTD are found in this region.

1.1.4.2 Chromosome 9

Although ALS and FTD have been observed for over a century, it is only in 2006 that mutations at chromosome 9 started to be associated to familial forms of these diseases[103]. Several mutations at chromosome 9 have been found in linkage studies between familial ALS and FTD, indicating that this has potential for therapeutic targeting for both diseases. A first mutation has been noted at 9p13.3, the gene encoding for valosin-containing protein (VCP)[104]. VCP is part of the AAA-ATPase protein family responsible for several cellular pathways including DNA-damage response (ND hallmark) and ubiquitin-dependent protein degradation (altered protein homeostasis hallmark)[20]. A second mutation has been observed at the intraflagellar transport 74 protein gene (IFT74) at 9p21.2, which is linked to transport at the neuronal axons and dendrites (synaptic dysfunction hallmark)[105]. Furthermore, the most common mutation in familial cases of ALS/FTD is a mutation at the chromosome 9 opening reading frame 72 (C9orf72), which has been linked to the pathologies *via* various mechanisms of action interlinking most ND hallmarks. The fact that this particular mutation is so prevalent in most familial cases in both diseases and in some sporadic ones, makes it a perfect candidate to characterise the mechanism of action.

1.1.5.3 C9orf72

An incredible breakthrough in understanding the pathogenesis of ALS and FTD was made a decade ago, in 2011, with the identification of the hexanucleotide repeat expansion $(GGGGCC)_n$ in the non-coding region of chromosome 9 open reading frame 72 (C9orf72)[103]. In most people, the repeat length *n* is 2-8, but patients suffering from ALS/FTD will present hundred to thousands of repeats. Generally speaking, the pathological threshold is considered at around 24[56], but it is not clear how the length of the repeat relates to clinical observations on the progression of the disease. Interestingly, in REs larger *n* tends to correlate to either an earlier onset of the pathology or a more aggressive form of the disease but this is not necessarily true in ALS/FTD. Studies have shown both presence and absence of correlation between repeat length and degree of aggressiveness of the pathology[103]. Nevertheless, this mutation is considered a key to understanding disease progression due to the fact that it is the most common genetic hallmark of ALS/FTD, accounting for about 40% of familial patients and 7% of sporadic ALS/FTD cases[56].

However, after over ten years of investigation, the exact pathogenic role of mutations at the C9orf72 gene in ALS/FTD is still under discussion but it has been widely recognised that it has to do with a combination a of loss-of-function effect of the C9orf72 protein and a toxic gain-of-function of the RNA repeat. But before going into the details of the mechanisms by which C9orf72 has been implicated to cause neurodegeneration, it is important to describe the typical biological functions of C9orf72 in healthy individuals.

The C9orf72 gene includes two non-coding exons and 10 coding exons which can be translated into three recognised coding variants (V1,V2 and V3)[103]. When these are alternatively spliced, two different isoforms are obtained: one C9-short of 24 kDa and another C9-long of 54 kDa). Since this gene is highly conserved in primates it would point towards the idea that it is necessary for the correct biological functions of the evolved species[103]. The transcripts of C9orf72 have been isolated in most tissues, including brain region and spinal cord, while protein expression has been highly detected in the brain, spinal cord and immune system (primarily in the cytoplasm but with nuclear *punctate*). Specifically, C9-long is found in the cytoplasm as speckles while C9-short is primarily located in the nuclear membrane (at least in post-mortem samples[103]). Amongst the functions that C9-proteins have been related to, the prevalently studied include: interaction with proteins involved in the nucleocytoplasmic import (Ran-GTPase and Importin b1), stress-granule homeostasis, regulation of autophagy (via interaction with p62), axon growth, mitochondrial regulation and synaptic function (via activation of the OXPHOS complexes), membrane trafficking (by activation of RabGTPase)[103]. Furthermore in 2016, several studies started to report the effect of the C9-protein onto immune system regulation (especially in myeloid cells)[106]. But it is not just the C9-protein to have an essential regulatory effect. In fact, C9orf72 mRNA transcript interacts with several aforementioned proteins such as TDP-43 and FUS, aiding in maintaining protein homeostasis and neuronal health in the CNS[107].

Overall, there are three general mechanisms of action which have been hypothesised to link C9orf72 to ALS/FTD onset: **(1)** C9orf72 gain-of-function through the accumulation of toxic dipeptide repeat proteins, **(2)** loss-of-function through haploinsufficiency and **(3)** RNA toxic gain-of-function[108] (***Figure 11***).

$(GGGGCC)_n$ RE can lead to toxic gain of function by translation into dipeptide protein repeats (***Figure 11b***). These are obtained when a highly repetitive RNA sequence is translated into a highly repetitive amino acid sequence with several low-complexity domains prone to aggregation. In fact, these have been shown to form cytoplasmatic inclusions which are negative for TDP-43 but positive for p62 (autophagy marker)[109]. Dipeptide repeats have also been shown to inhibit DNA double strand break repair mechanisms, which is one of the common hallmarks of neurodegeneration[110].

For the first following two mechanisms, the pathogenic consequences of $(GGGGCC)_n$ come from the sequence's inherent polymorphism adopted on the basis of the external environment (***Figure 11b***). In fact, $(GGGGCC)_n$ has been shown to be able to form stable DNA and RNA structures non-canonical secondary structures including G-quadruplexes[111], hairpins[112] and R-loops[113], which all have different downstream consequences.

At the genomic level, these structures can hinder transcription and cause the accumulation of aborted transcripts[113]. These aborted transcripts will have an increased length and local concentration the longer the original expansion was in the gene, and they have been shown to sequester RNA-binding proteins involved in transcription and splicing (such as TDP-43, FUS and nucleolin[113]). These proteins, usually necessary in the nucleus, are translocated into the cytoplasm forming pathological aggregates known in NDs as RNA *foci*, a typical feature in ALS/FTD patients, leading to a combination of toxic loss-of function of the proteins in the nucleus and toxic gain-of-function in the cytoplasm. Although the mechanism of formation of the aggregates is fundamentally unclear, it has been shown that these are formed by a mix of translocated proteins and $(GGGGCC)_n$ transcript (both in the sense and antisense direction) and that they are present in the nuclei of the motor and frontal cortex neurons, hippocampus, cerebellum and spinal cord[114,115]. Furthermore, as $(GGGGCC)_n$ is located in the intron of the C9orf72 gene, the presence of such a large repeat length forming additional structures leads to downregulation of the gene itself. In fact, the expansion repeat has been shown to bind trimethylated histones in ALS/FTD brain autopsies[116]. As C9orf72 is a mitochondrial inner-membrane-associated protein regulating cellular energy homeostasis *via* its critical role in the control of oxidative phosphorylation[117], when its expression is downregulated the energy metabolism of the cell is impacted.

Moreover, $(GGGGCC)_n$, much like other REs, has the potential to condensate in absence of proteins in by forming multimeric networks. In fact, due to its high Guanine content, it is possible it could form G-quadruplexes or even multimeric G-quadruplexes networks which could potentially trigger the pathology.

Importantly, in biological systems the formation and disassembly of DNA and RNA secondary structures is tightly regulated by a series of proteins involved in DNA/RNA damage repair pathways[118] (*e.g.* helicases for unwinding G-quadruplexes). There where these proteins are downregulated or mutated, pathogenic consequences have the potential to arise. Indeed, helicase DDX3X, which is downregulated in patients with ALS/FTD and linked to the ability to resolve G4s[119], has been shown *via* CRISPR-Cas9 screen to repress pathological RAN-translation. In fact, elevating the levels of DDX3X was shown to decrease DPR levels, rescue nucleocytoplasmic transport abnormalities, and improve

survival of patient iPSC-differentiated neurons, further suggesting that the presence of non-canonical secondary structures is strongly linked to the pathology. Although a few other G4-resolving proteins linked to ALS/FTD have been proven to be downregulated in NDs[120], it is worth noticing that alternative nucleic acid secondary structures have only recently become a point of attention in the field. Therefore, further studies are required to fully understand how these structures impact DNA/RNA-damage repair mechanisms.

Importantly, mutations at the C9orf72 gene are the major genetic cause of ALS/FTD, not only present in familial cases, but also in a smaller percentage of sporadic events. $(GGGGCC)_n$ repeat expansion has been directly linked to the presence of pathological protein aggregates (altered protein homeostasis), autophagy (neuronal cell death), altered energy metabolism and to affecting DNA damage repair mechanisms, interlinking most of the widely recognised pathways for neurodegeneration. Therefore, although the exact mechanistic pathway is still under discussion, it is clear that C9orf72 $(GGGGCC)_n$ repeat expansion plays a fundamental role in the progression – and perhaps development - of the ALS/FTD phenotype, making it a very attractive target for therapeutics development.

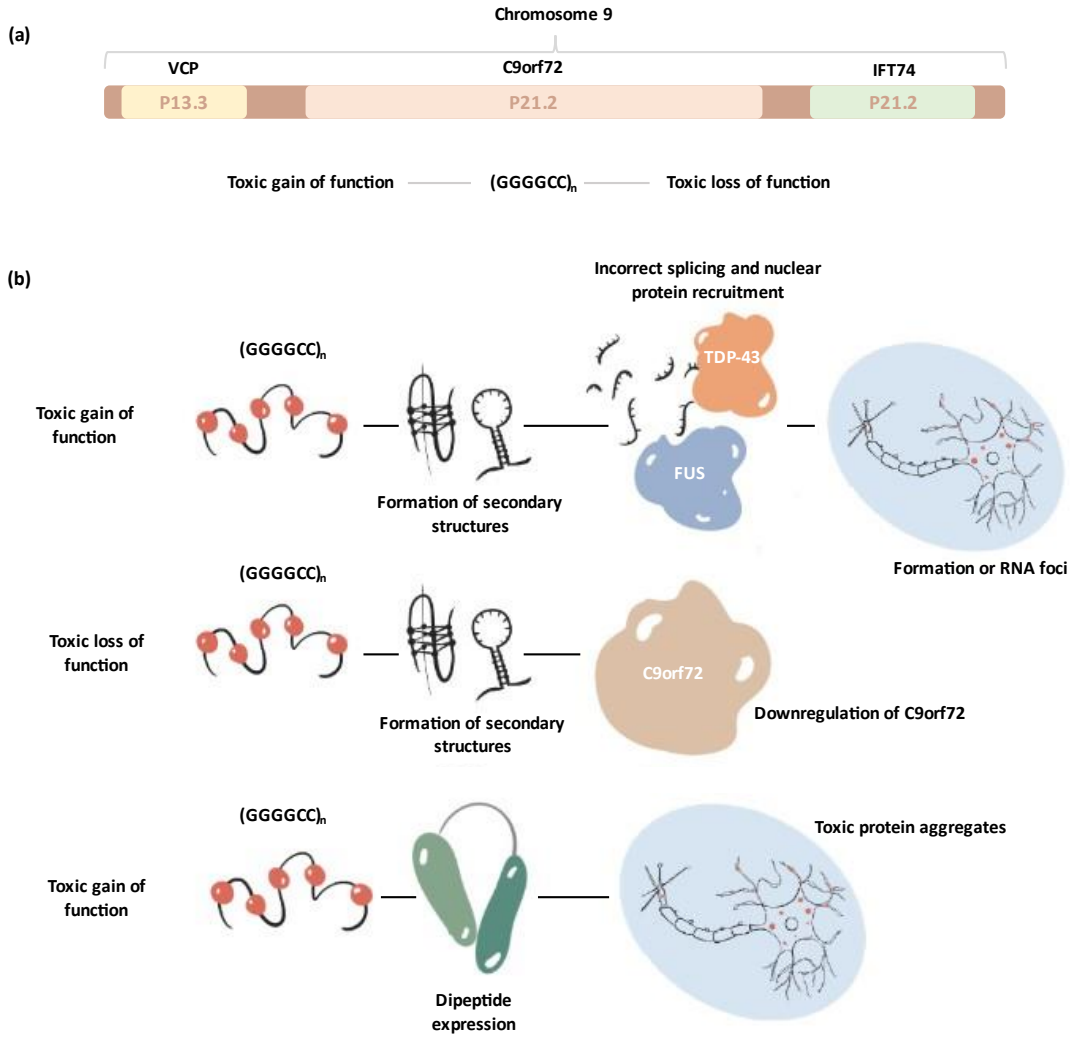


Figure 11: Chromosome 9 mutations are responsible for ALS/FTD *via* multiple mechanisms of action. (a) Chromosome 9 main mutated sites in ALS and FTD. The three mainly recognized mutations of chromosome 9 are at the VCP, C9orf72 and IFT74 sites. Figure is not up to scale with respect to the relative size of the domains. **(b)** Main pathological mechanisms of action caused by C9orf72 repeat expansion. Three main mechanisms have been identified to cause the pathology from the hexanucleotide repeat expansion: **1)** toxic gain of function *via* the formation of alternative RNA secondary structures which alter the splicing mechanism and lead to the formation of RNA foci, **2)** toxic loss of function due to the formation of alternative RNA secondary structures and consequent downregulation of C9orf72 protein and **3)** toxic gain of function *via* the expression of dipeptide proteins prone to aggregation. In the illustration, two peptides (in green) are held together by a peptide bond (gray line). Hand-drawn illustration with Procreate.

1.2 G-quadruplexes

1.2.1 Nucleic Acids: from helices to G4s

A deep dive into the mechanisms of pathogenesis of ALS/FTD has revealed an increasing relevance of nucleic acid alternative secondary structures. In particular, the RE $(GGGGCC)_n$ in the C9orf72 gene has the potential to arrange into self-condensing multimolecular networks *via* the formation of G-quadruplexes structures.

Nucleic acids possess a remarkable level of complexity and are influenced by various environmental factors. The 3D conformation of helical DNA was first discovered in 1953 by James Watson and Francis Crick (together with the often-forgotten Rosalind Franklin), which led a fundamental step forward in our understanding of many biological processes. Structural analysis highlighted how the two strands of DNA were held together by hydrogen bonding (H-bonding) strong forces, later named Watson-Crick H-bonding, between complementary DNA bases (***Figure 12a***). The specificity of this binding was not only fundamental in understanding biological recognition processes, but also opened doors for new science outlooks where DNA could be tuned and used as an artificial building block. In fact, based on the principle that complementary bases are strongly driven to bind, the new-born field of DNA nanotechnology developed tools to design artificial DNA sequences of microscopic scale with artificial functionality[121]. Artificially designed DNA has now been used for a variety of different applications, including designing membranelles compartments in artificial cells[122], artificial signalling processes[123] and others[124].

It soon became evident that the dsDNA helical structure predicted by Watson and Crick was not the only secondary structure that nucleic acids could assume *in vivo*. Already in 1955, the existence of cruciform DNA was hypothesised[125], to then be confirmed *in vitro* in 1981 in palindromic sequences[126]. In cruciform DNA, the canonical Watson-Crick Base pairing is respected but due to the self-complementarity of the sequences, intra-strands interactions occur and give rise to the shape of a cross (***Figure 12b***). Short-after, nucleic acids hairpins were observed *in vitro* and their importance was recognised for both DNA and RNA biological functions[127] (***Figure 12b***). Similarly, as for the cruciform form, nucleic acid hairpins are observed when the sequence is self-complementary *via* Watson and Crick base pairing. But it was already in 1959, only a few years after the discovery of Watson-Crick base pairing, that questions started to arise on whether non-canonical H-bonding interactions were also possible[128]. This was found to be the case by Karst Hoogsteen, under whose name these forces were subsequently named. A first example of non-canonical Hoogsteen base pairing was observed in triplex-DNA, where instead of only combining two strands, the bases utilise additional H-bonding

donors and acceptors sites to add an additional strand[129] (***Figure 12b***). The second pioneering observation was made in 1962, where it was discovered that nucleic acids can also adopt alternative, four-stranded helical structures in sequences that are rich in Guanines[130]. It has now been established that these non-canonical four stranded structures called G-quadruplexes (G4s) can be formed under physiological conditions *in vitro* and in cells[131]. The core scaffold of a G4 is the G-tetrad, which is constituted by four Guanines arranged in a planar tetrad held together by Hoogsteen H-bonding between additional H-bonding donor and acceptor sites in Guanine (N7) (***Figure 12b***). The 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^+)[132] and more tetrads are held together by π-stacking interactions. G4s can be formed within the same DNA sequence strand (unimolecular G4) or between distinct DNA sequences (multimolecular G4)[131]. They are not thermodynamically stable structures, but are often the kinetically favourable conformation in G-rich nucleic acid strands[133]. Their ability to associate and dissociate depends on many environmental factors such as pH changes, length and concentration of the G4-forming sequence and presence of monovalent cations. In cells, such conditions are fluctuating, leading to an expected dynamicity of the structure. G4s were first observed in physiological conditions whilst analysing DNA sequences extracted from the human genome in telomeric DNA, but they have now been widely recognised in several other regions, in particular at promoters, suggesting a variety of biological functions. In recent years, G4s were implicated in a number of diseases including cancer, accelerated ageing disorders such as Cockayne Syndrome, neurodegenerative diseases such as ALS and FTD, and rare-genetic diseases such as Fragile X syndrome[131].

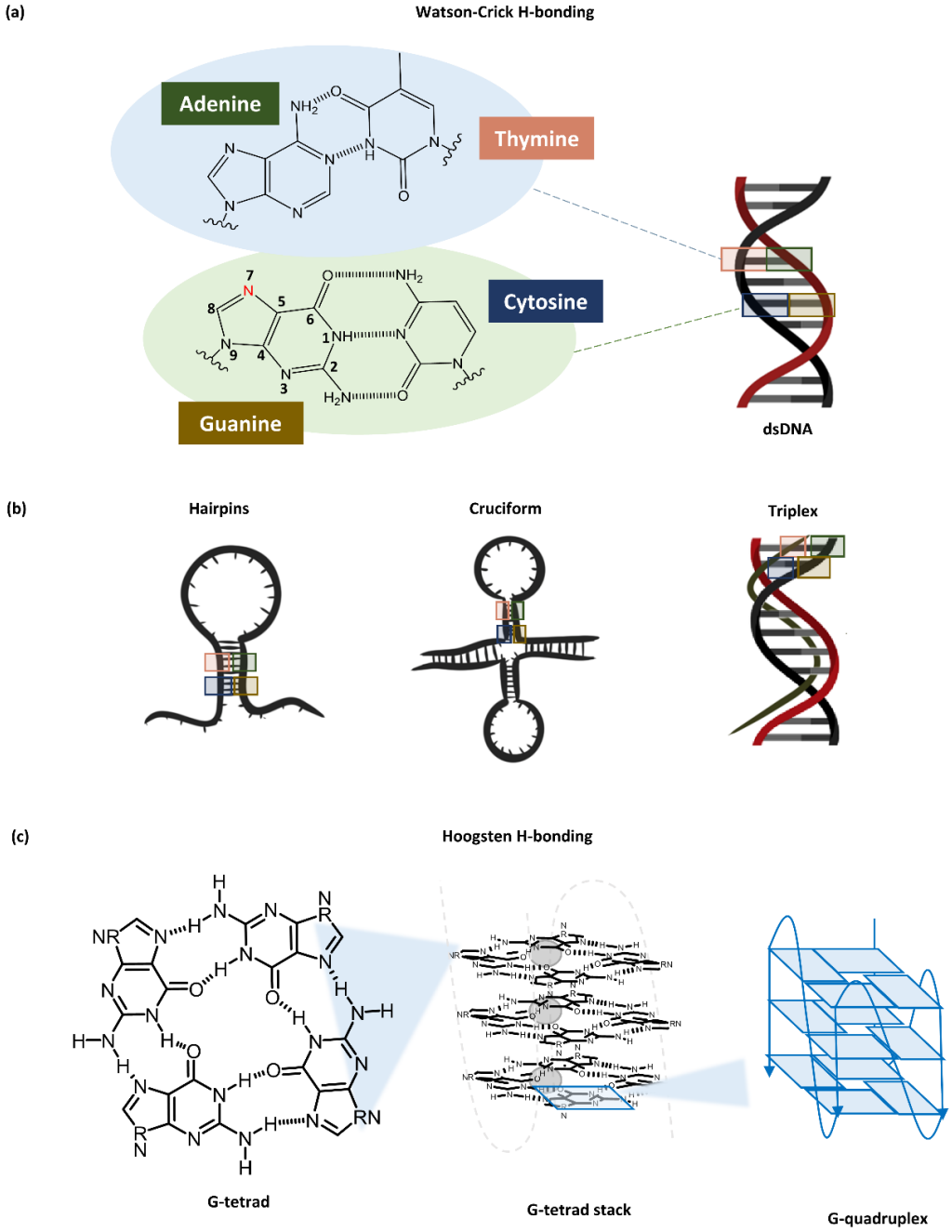


Figure 12: Nucleic acids can arrange into several secondary structures. (a) Canonical Watson and Crick base pairing between bases in DNA with highlighted H-bonding donor and acceptor sites. **(b)** Alternative secondary structures conformations that rely on Watson and Crick base pairing. **(c)** Non-canonical Hoogsten base pairing in Guanine rich sequences and formation of the G-tetrad and G-tetrads stack. Hand-drawn illustration with Procreate.

1.2.2 Biological relevance of G4s

Sections adapted from **Raguseo** et al, 2020, ChemComm

The biological relevance of G4s has been proven by a combination of computational tools and chemical-biology probes showing a direct correlation between *in vitro* and *in vivo* studies (***Figure 13a***). It is important to notice, however, that each method developed so far is subject to limitations and might lead to an over or under-estimation of the biological importance of these structures.

Computational algorithms have been used for several years to predict the formation of G4s based on the guanine content and distribution within a sequence[134]. The initial computational algorithms developed for the purpose only took in account the canonical motif of a G4 ($G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$), without considering G-rich sequences that defy this consensus motif. Although computational tools have now been implemented to account for more complex topologies and use machine learning algorithms to improve accuracy (Quadron)[135], the original consensus motif was useful in proving that the human genome had *at least* the potential to form G4s non-randomly. In fact, it was observed that these were enriched at regions regulating important cellular processes such as translation and transcription.

The first experimental evidence showing that predicted G4s from computational algorithms could in fact form *in vitro* was presented by *Chambers et al*. for DNA[136] and *Kwok et al*. for RNA[137]. They allowed to detect the presence of 716314 G4-structures in the human genome, almost twice as many than what observed from computational predictions. In particular, G4s were identified in important cancer-related genes such as BRCA1, BRCA2 and MAP3K8 and rG4s were proved to form at physiological conditions in mRNA at 5'-UTR and 3'-UTR regions.

But although *in vitro* the ability of certain genomic locations to arrange into G4s has been indisputable for several years, there are still controversial opinions on the G4-presence in cells. That is because, *in vitro* studies are generally performed on single stranded nucleic acid, whilst in cells DNA is wrapped around proteins known as histones. These DNA-protein complexes are known as nucleosomes and are in turn arranged into chromatin. Chromatin architecture is known to regulate gene expression by altering the accessibility of DNA to enzymes, implying that this could alter the ability to form G4s. Several approaches have been taken to test for G4-presence in the chromatin context.

The most well-known approach involves an antibody-based G4 chromatin immunoprecipitation followed by high-throughput sequencing (BG4-ChIP)[138]. BG4-chip showed that only 1% of the total number of G4s detected *in vitro* by G4-Seq on single-stranded DNA are actually formed and detected

in a chromatin context. This suggests that chromatin structure is key to expose the formation of certain G4s over all the available structures present in the genome. But whilst there are still questions on the biological relevance of DNA G4s, rG4s might even be more disputed due to the lack of specific tools for their targeting.

A second approach, relies on using G4 small molecules ligands in cells and observe the downstream effects of other biological markers such as for DNA damage control (non- fluorescent ligands) or localisation of fluorescent G4-*foci* (fluorescent ligands). A big advantage of using G4-binding ligands is the ability to image cells in absence of fixing, leading to a true representation of the cellular dynamic system.

Indeed, new selective small-molecule ligands are continuously being developed for G4s (G4 Ligand Database 2.1[139]). 3000 different G4-targeted ligands have been identified, although different in selectivity, affinity, binding sites, and cellular permeability and they all share two typical structural features: **(1)** an aromatic core (which allows for π-π stacking interactions with the G-tetrad) and **(2)** one or more positive moieties (for electrostatic interactions with the negative backbone of RNA/DNA)[140]. Usually, a small molecule probe is defined as a G4-ligand if its dissociation constant is lower than 10^{-6} M and it presents selectivity over ssDNA/RNA and dsDNA[140]. Some of the most common G4-ligands are illustrated in ***Figure 13b***. G4-ligands can be differentiated into fluorescent and non-fluorescent probes, depending on their ability to emit fluorescence upon binding to nucleic acids.

One of the most commonly known non-fluorescent ligands is Pyridostatin (PDS (4-(2-aminoethoxy)-*N*^{2},*N*^{6}-bis(4-(2-aminoethoxy)quinolin-2-yl)pyridine-2,6-dicarboxamide)] and its derivatives [*e.g.*, PyPDS (4-(2-aminoethoxy)-*N*^{2},*N*^{6}-bis(4-(2-(pyrrolidin-1-yl)ethoxy)quinolin-2-yl)pyridine-2,6-dicarboxamide)] which has shown outstanding G4-binding affinity and specificity abilities[141]. These molecules not only bind G4s selectively, but have also been shown to stabilise the structures *via* melting assays[142]. Both PDS and PyPDS have been employed with different technologies to investigate the G4-interactome in cells and provide further G4-genomic accessibility. Another common non-fluorescent G4-ligand is PhenDC3 (3,3'-[1,10-Phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methylquinolinium] 1,1,1-trifluoromethanesulfonate (1:2)) and its derivatives[143]. Much like PDS, PhenDC3 presents a planar aromatic central scaffold with collateral positive moieties for efficient G4-binding, but it is characterised by a U-shape which allows the scaffold to have a better spacial overlap with the G-tetrad. For this reason, it binds very efficiently to G4s and has even been shown to induce refolding of certain human telomeric sequences into G4s *in vitro*[143]. Other commonly known non-fluorescent G4-ligands are TMPyP4 and others[140], which have been proved useful for specific purposes but show a lower binding affinity with respect to PDS and PhenDC3. It is important to notice, that

although PDS and derivatives are not fluorescent on their own, they have been shown to successfully conjugate to other fluorescent probes for detection of G4s in cells. In particular, SiR-PyPDS was developed to visualise G4s in cells on a single-molecule level[144].

Although very selective for G4s over dsDNA, SiR-PyPDS still presents some inevitable off-site binding which could lead to false positives in the detection of G4s in cells. One way to overcome this limitation, is to have a probe that is selectively detected (via fluorescence) only in presence of the G4 scaffold.

This is the case for NMM (N-methyl mesoporphyrin IX) a porphyrin-based probe that is fluorescent upon binding to G4s[145]. Usually, porphyrin and its derivatives are naturally fluorescent molecules[146], but in the case of NMM, this fluorescence is self-quenched in bulk. This is caused by the NMM planar cores stacking on top of each other, leading to self-quenching effects. In presence of G4s, NMM binds to the planar tetrad, allowing fluorescence to return. The one limitation of NMM is that – to limited extents – it still is able to intercalate with ds and ssDNA, leading to potential off-site binding.

One solution to this selectivity issue with fluorescent probes was proposed by *Shivalingam et al*, who observed that fluorescent probes have a different lifetime when bound to G4s versus other biomolecules, even ss and dsDNA[147]. This is the case for DAOTA-M2, which has shown signature fluorescence lifetime spectra for DNA and RNA G4s[148].

Importantly ligand probes have a stabilising effect on G4s themselves, which led to questioning whether the structure is effectively biologically relevant in absence of the probe. Addressing this limitation represents the main challenge for the design of the next-generation probes for G4-detection in living cells.

The combination of *in vitro*, *in vivo* and live cell imaging probes developed so far point towards the fact that these structures are directly involved in fundamental biological processes and pathogenesis and have the potential to constitute future drug development targets, but further studies need to be performed to optimise the tools specificity (***Figure 13***).

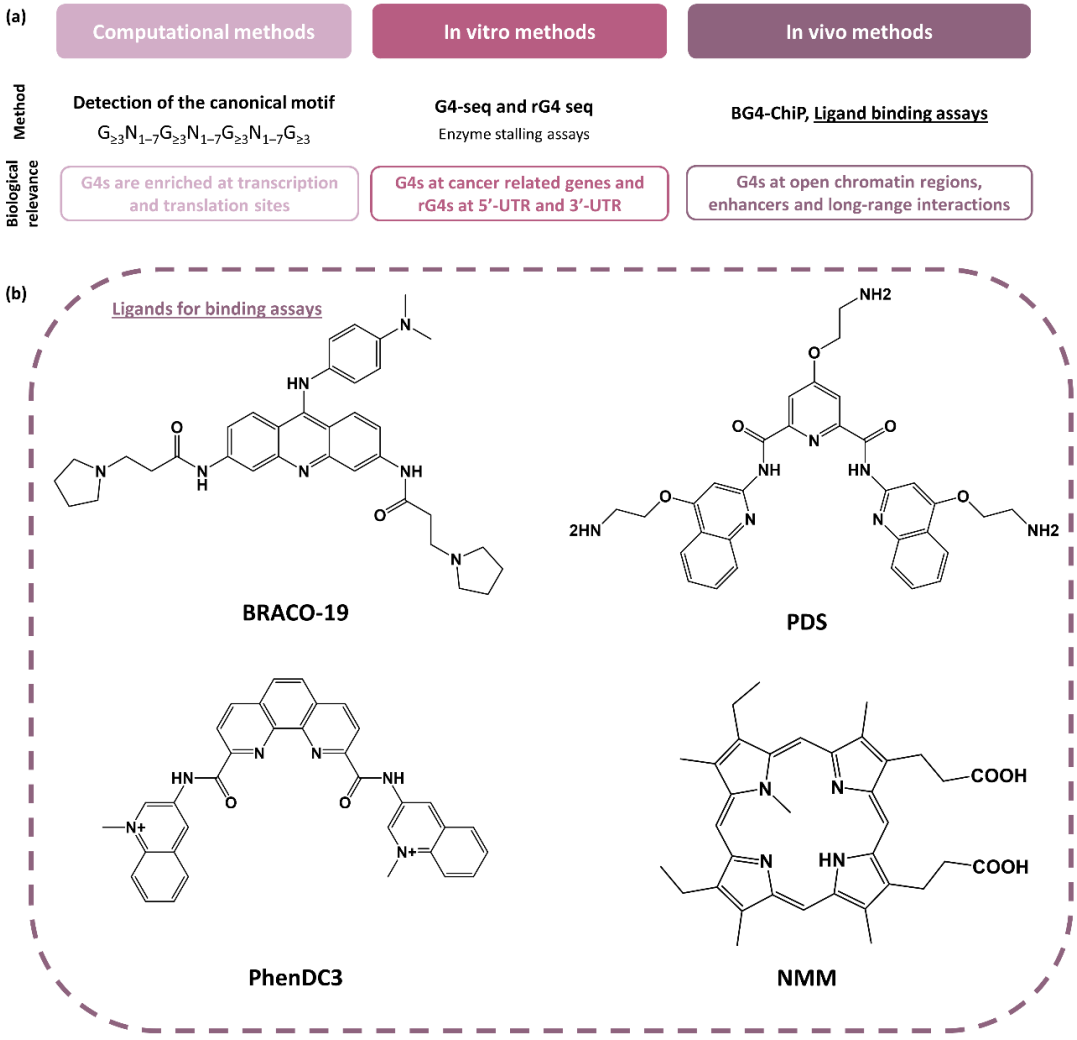


Figure 13: G-quadruplexes biological relevance. (a) Summary of the methods employed to prove the biological relevance of G4s in humans and their main respective findings. These include computational tools that only account for canonical motifs; G4-seq and rG4-seq, limited to in vitro screening; BG4-ChIP, which relies on fixated cells and unpredictable BG4 binding; small fluorescent light up probes for live cell imaging**. (b)** Chemical structures of the main G4-ligand probes mentioned in the text. Obtained with ChemDraw.

1.2.3 Potential biological relevance of mG4s

Adapted from Liano, Monti, Chowdhury, **Raguseo** et al, 2022, Chem Comm

Thanks to numerous chemical biology tools developed to detect G4s, these structures are now increasingly being recognised as potential targets for a series of transcription/translation mediated mechanisms and overall chromatin architecture. But whilst the biological relevance of unimolecular

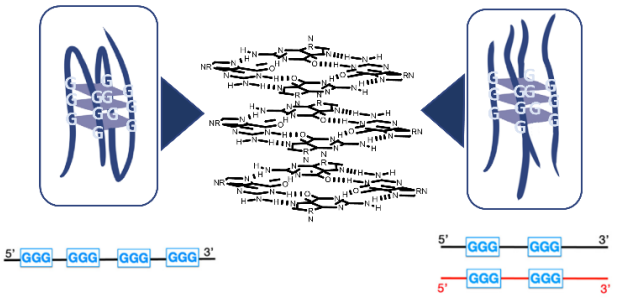
G4s has been explored for decades, it is only in recent years that multimolecular G4s (mG4s) have started to gain more and more interest[149].

Multimolecular G4s are G-quadruplexes containing more than one strand. Their specific terminology depends on the number of strands involved: bimolecular or dimeric (two), trimolecular or trimeric (three) and tetramolecular or tetrameric (four)[150] (***Figure 14***). They can form primarily *via* two mechanisms: one where different strands come together linked by guanines Hoogsten H-bonding and another where multiple unimolecular G4s are interconnected *via* π-stacking interactions.

Although unimolecular species form more readily and tend to be the kinetic product, mG4s are usually more thermodynamically stable in cell crowding conditions[151]. Multiple unimolecular species coming together leads to a reduction in entropy of the system, even if the final product is more thermodynamically stable than the unimolecular species. To overcome the energy barrier caused by the reduction in entropy, one can either **(1)** promote crowding conditions or **(2)** allow the system to equilibrate for longer. Cell-crowding promotes multimolecular structures by bringing the strand closer together, while allowing a longer equilibration time simply permits the system to shift towards the thermodynamic product. Although this makes mG4s more challenging to be assembled *in vitro*, it should also highlight how the crowded live-cell environment could promote their formation.

The initial hypothesis of mG4s in a biological context emerged within telomeric DNA, right there were G4s were first observed. *Moye et al* identified a human telomerase able to recognise and extend telomeric DNA when folded into an mG4s[152]. A few years later, *Liano et al* showed a similar selective mG4 helicase activity (CSB – relevant for Cockayne Syndrome cases) for bimolecular G4s, implying a biologically relevant effect of these structures[153].

Unfortunately, aside from a few studies, there is very little evidence of the fact that mG4s are relevant in a biological context. The scarcity of evidence does not arise from the absence of associations between these structures and biological phenomena, but rather from the lack of specific probes capable of distinguishing mG4s from G4s in detection methods. In fact, 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[154], PhenDC3[155], TMPyP4[156], NMM[157]) have not been thoroughly studied in the context of mG4s, making differentiation in live-cell experiment impossible. In fact, even the newest computational tools are not equipped to deal with complex mG4s[158].



Unimolecular G4s	Properties	Multimolecular G4s
Anti-parallel, parallel and hybrid	Topologies	Anti-parallel, parallel and hybrid
> 4 G-runs and at least 2 Gs	Sequence length	1 G-run (tetramolecular) 2 G-runs (bimolecular)
1-7	Ideal loop length	1-7
Faster formation and unfolding	Kinetics	Slower formation and unfolding

Figure 14: General characteristics of multimolecular G4s versus unimolecular G4s. While unimolecular G4s are the kinetic product and form more readily given a canonical G4 sequence, multimeric G4s are the thermodynamic product. Although they take a longer time to form, they have an additional final stability. The table illustrates the main differences and properties of both conformations.

1.2.4 mG4s role in phase-separation events

Adapted from sections of Robinson, **Raguseo** et al, 2021, Nucleic Acid Research

Aside from the more historically recognised biological roles of G4s in the genome that see these structures as acting as transcriptional "on-off" blocks, there has been increasing evidence of the fact that G4s could also have an organisational role within the cell by utilising multimeric interactions, specifically in **(1)** allowing long-range interactions at enhancers regions and **(2)** in liquid-liquid phase separation events[159].

Indeed, it has been shown that G4s are enriched at super enhancers regions[160], which is thought to promote long-range DNA interactions. Transcription of a gene is regulated not only by its adjacent promoter, but by the use collateral docking sites to increase/decrease the process based on epigenetic factors. These additional genomic sites are known as enhancers and enhancers-mediated control has been shown to increase transcription as high as 100-fold. Interestingly, this control of transcription can occur over hundreds of thousands of base pairs, indicating that distal interactions in DNA may be

facilitated by the folding of DNA into secondary structures such as loops or mG4s to allow normally separated sequences to meet[159].

Another interesting biological process involving mG4s relates to liquid-liquid phase separation or aggregation events. In recent years it has been discovered that the expression of genes can be mediated by phase separation events that allow for the concentration of transcriptional machinery within membrane-less organelles inside the nucleus[159]. Although many of the studies considering phase-separation have historically focused on the protein component of the aggregate, recent work has revealed that certain nucleic acid structures can also act as nucleation sites for phase-separation, even in the absence of proteins[159]. In particular G4s, much like proteins, are biopolymers, are able to form multimolecular, non-covalent, allosteric interaction between strands which could lead to aggregation. This was exemplified in work directly associating G4s with phase-separation in live cells within short root RNA[161]. In these studies, it was shown that G4-triggered phase separation is dependent on exposure of the system to G4-favouring conditions, such as increasing potassium concentration, local concentration of nucleic acids and relative guanine content (***Figure 15***).

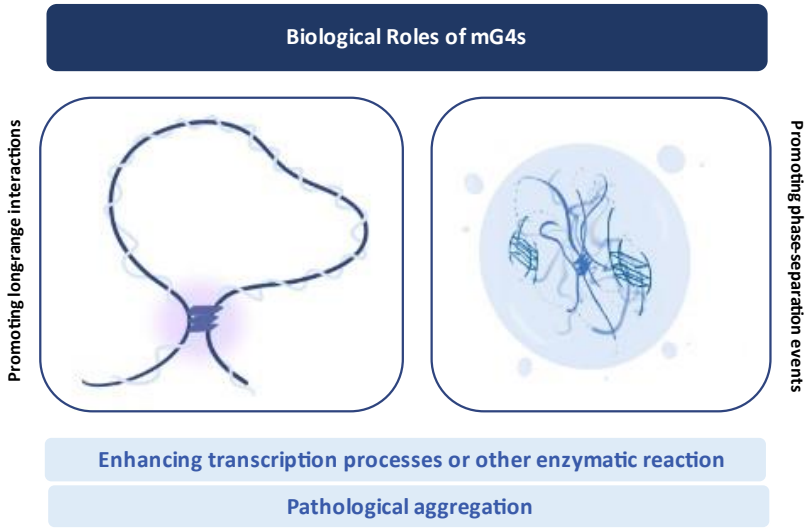


Figure 15: Biological roles of mG4s. The two main mechanisms of action by which mG4s have been thought to gain biological relevance is by promoting **(1)** long range interactions at super-enhancers regions of the genome and **(2)** phase separation events, as in the case of neurodegenerative diseases. Hand-drawn illustration with Procreate.

1.2.5 G4s as therapeutic targets

Ever since G4s enrichment has been observed at telomeric regions in cancer cells and linked to other diseases, the development of stabilising and destabilising drugs for these structures has been taken in consideration for future therapeutic pathways. However, two main issues are related to the development of G4-based therapies: **(1)** it is unclear whether G4s have purely a "good" or "bad" physiological effect. In fact, given the evidence that G4s are so intertwined in a varied of regulation mechanisms, it is possible that their enrichment is present as a defence mechanism against a pre-existing pathology and trying to disassemble them could lead to ineffective treatments or even disastrous consequences; **(2)** it is not clear whether a selective probe for a specific G4 can be developed. The absence of specific probes hinders the ability to selectively target and modulate individual G4s, thereby increasing the risk of unintended and harmful effects on normal biological processes. Furthermore, to this day there is a limited number of chemical probes able to destabilise G4s and none that can do so selectively, which would be a major step in resolving pathologies linked to these structures[149]. One interesting novel approach to address these two concerns was proposed by *Chowdhury et al* who showed how LNA probes could be utilised to disrupt G4s[162]. In the study, it was demonstrated that short LNA-modified probes can disrupt G4s responsible for stalling of DNA polymerases, suggesting that these probes have the potential to rescue the progression of the stalled polymerases.

G4-probes have been explored for years as anti-cancer treatments but, whilst there has been some success in cells, these probes have never shown effectiveness in cancer patients on their own[163]. Aside from anti-cancer agents, G4-probes have been hypothesized to ameliorate pathological signatures of ALS/FTD models *via* targeting the G4s formed by the mRNA $(GGGGCC)_n$ toxic repeat[164]. According to the mechanistic hypothesis, G4s have the capability to stabilize toxic protein aggregates. However, when a G4-ligand is present, it is believed to displace the protein, leading to the destabilization of the pathological aggregate. Nevertheless, the precise mechanism behind this process remains a subject of debate and is still not fully understood.

In conclusion, G4s are desirable therapeutic targets for a multitude of diseases. It is important to notice that although no treatment targeting G4s has been FDA approved, these structures are a relatively novel target, and only recently it has been understood the variety of mechanisms of actions that they can be involved in.

Chapter 2

Project Aims

Many mechanistic hypotheses for the formation of pathological aggregates in ALS/FTD revolve around the presence of pathological protein/RNA aggregates at the neuronal level. Historically, proteins have been looked at as the potential trigger for aggregation. But despite numerous attempts of developing probes designed to target and mitigate the accumulation of pathological protein aggregates, this has yielded limited results.

A ground-breaking discovery was made only a little over ten years ago, when a mutation at the gene C9orf72 was associated with 40% of familial and 7% of sporadic ALS/FTD cases[20]. C9orf72 is characterised by the presence of the hexanucleotide repeat expansion $(GGGGCC)_n$, which could be affect neuronal health *via* a variety of pathways ranging from the translation in highly repetitive proteins, stabilisation of pre-existing protein-based aggregates, mis regulation of collateral genes and the formation of RNA multimolecular networks[103]. The latter, in particular, would challenge the notion that aggregation-prone proteins are the biomolecules responsible for pathological aggregation in neurodegenerative diseases, a notion that sees RNA structures only involved in secondary stabilisation events. Indeed, the ability of the RNA $(GGGGCC)_n$ repeat to form multimolecular networks gives it the potential to aggregate in a similar fashion to proteins[111,112,113]. In particular, such networks could be formed by higher complexity structures such as hairpins, triplex DNA, cruciform DNA and – in Guanine-rich sequences – G-quadruplexes.

$(GGGGCC)_n$ has repeatedly been proven able to arrange into G4 structures *in vitro*[94,165] and in live cells[164] and it has been shown that treating ALS/FTD affected cells with G4-binding ligands can ameliorate pathological aggregates[164]. Furthermore, some of the proteins associated with pathological aggregates (such as FUS and TDP-43) have both been reported to be able to binding G4s efficiently[101,166], confirming that these structures are associated with pathological aggregation.

However, to date it is unclear as to how exactly the G4s could contribute to aggregates pathogenesis. The leading hypothesis is that they could contribute to stabilising pre-existing protein aggregates in their RNA form, which is supported by the fact that pathological aggregates have been shown to colocalise with RNA-staining dyes at RNA *foci* locations and separate studies showing that those same proteins are able to bind G4s[167].

Although feasible, this mechanism presents several inconsistences.

Firstly, although protein aggregation is considered one of the hallmarks of NDs, it has been shown that targeting the protein aggregates in many cases does not lead to a reduction of symptoms – suggesting that the cause of neurodegeneration might be another one[35,168]. Secondly, this leading hypothesis does not take into account the fact that RNA, just like the proteins, is present at high concentrations in the neurons and has the potential to form multimolecular networks *via* assembling into G-quadruplexes.

My hypothesis is that $(GGGGCC)_n$ can aggregate on its own right in absence of proteins by generating a multimolecular G4 network (***Figure 16***).

Although no studies have been reported on the topic, this mechanistic pathway is supported by some indirect evidence. Firstly, *Fay et al.* proved that there is a correlation between formation of rG4 secondary structures and phase-separation in a different system[169]. They did not go into the specifics of what type of multimolecular network was giving rise to the phase-separation, but their study offered a starting point to think of G4s as trigger for aggregation in cells. Secondly, *Disney et al.* have shown how altering the secondary structure of $(GGGGCC)_n$ *via* the use of hairpin-stabilising probes can modify the disease progression[112], once again proving that preventing $(GGGGCC)_n$ to form multimolecular G4s has a direct effect on aggregation. Lastly, the fact that targeting the protein component does seem to affect symptoms progression[35] would suggest that there must be something else triggering the aggregation in the first place. According to this hypothesis, the protein part could only be a secondary stabilisation element in the RNA G4-driven aggregation.

The objective of this work is to demonstrate formation of mG4s within the C9Orf72 expansion repeat, to correlate this to the ability of the RNA molecule to aggregate in the absence of proteins and to understand how these structures could interact with other biomolecules in the context of ALS/FTD neurons.

In order to address these points, I have set a number of key aims which I will discuss in these respective chapters:

- In Chapter 1 I aim to investigate if multimolecular G4s can trigger aggregation of $(GGGGCC)_n$ in absence of proteins. Although it has been shown that $(GGGGCC)_n$ can arrange into G-quadruplexes, the molecularity and exact topologies of the sequence remain unknown. By subjecting the sequences at different repeat lengths to an optimised mG4-formation protocol I aim to probe aggregation *in vitro* in absence of proteins.

- In Chapter 2 I will be studying the molecularity of the $(GGGGCC)_n$ system and its relevance to aggregation. I aim to discuss the effect of mutating the sequence, adding G4-specific probes and antisense strands on the aggregation abilities of the $(GGGGCC)_n$ repeat.
- Finally, in Chapter 3 I strive to investigate the biological relevance of $(GGGGCC)_n$ aggregation; the biological relevance of the system was assessed by analysing the RNA version of $(GGGGCC)_n$, in the presence of proteins G4-probes. I will also look for evidence of G4-based aggregation in ALS/FTD patient derived neuronal cells.

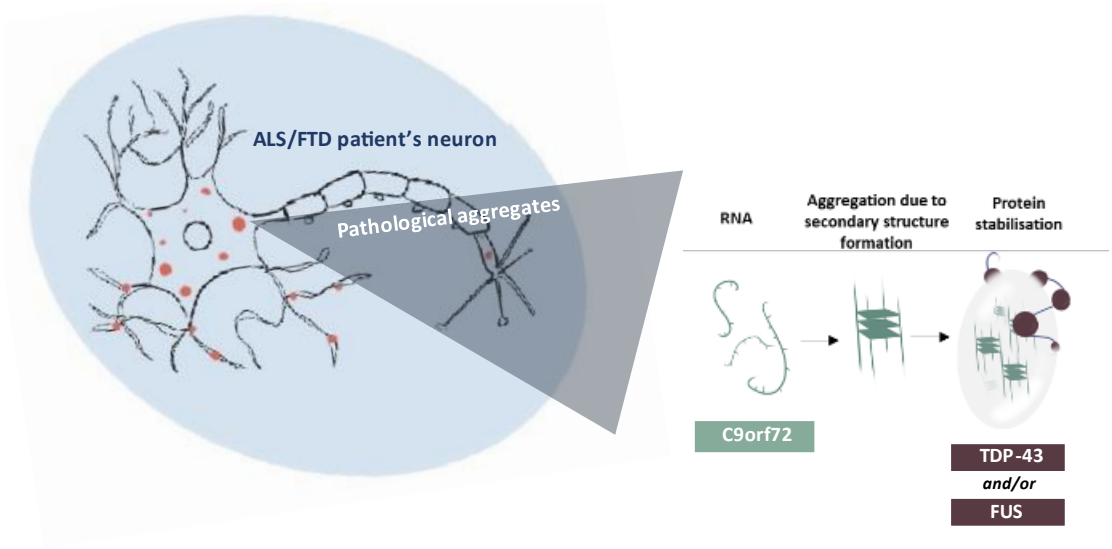


Figure 16: Alternative mechanism of pathological aggregation in ALS/FTD based on mG4 formation. Aggregation is triggered by the formation of mG4s thanks to the G-rich $(GGGGCC)_n$ sequence in the C9orf72 gene. ALS/FTD commonly mutated proteins are then recruited and aided in aggregating by the mG4s. Hand-drawn illustration with Procreate.

Chapter 3

Multimolecular G4s can trigger aggregation of $(GGGGCC)_n$ in absence of proteins

3.1 Background and Aims

Over 40% of familial cases of ALS/FTD are defined by the hexanucleotide repeat sequence $(GGGGCC)_n$ ($n > 24$)[20], which has previously shown the ability to form G4s both in its RNA and DNA form at various repeat lengths[170,171]. But whilst there is a general consensus on the ability of C9orf72 to form G4s, the topology and molecularity of the species is hard to predict *a priori*. Indeed, G-molecularity (where "multimolecular" refers to multiple unimolecular G4s interlinked together) is strongly dependent on sample preparation which tends to be different in each study. Such conditions include the presence of G4-stabilising ions (K^+, Li^+ and Na^+) and/or the rate and the oligonucleotide concentration during annealing. For instance, while $r(GGGGCC)_4$ was shown to adopt prevalently an intramolecular parallel G4 conformation by annealing the sample in a 30 mM K^+ buffer overnight at low μM concentrations[115], the very same sequence arranged into higher molecularity species in presence of KCl if annealed in 100 mM K^+ and instant cooling to 0 °C. Furthermore, it has been specifically shown that the G4-topology of $(GGGGCC)_n$ is affected by annealing conditions[111]. As there is no consensus on the G4-formation conditions, discrepancies on the topologies and especially molecularities are to be expected and further studies and careful consideration of annealing conditions are necessary to achieve a comprehensive understanding of the G4 properties in relation to $(GGGGCC)_n$ and their potential involvement in ALS and FTD pathologies.

Multimolecular G4s (mG4s) present similar general characteristics to their unimolecular (G4s) counterparts (***Figure 17a***). These include increased stability in presence of K^+ ions due to stabilisation of the G-tetrad and a higher probability of formation in higher G-content sequences. However, while unimolecular species are often the kinetic product and form more readily, mG4s are formed more slowly and are generally more thermodynamically stable, especially under cell crowding conditions (***Figure 17b***). This has been confirmed by several biophysical characterisation studies comparing melting temperatures of mG4s *vs* G4s[172] .

Thus, to favour the formation of $(GGGGCC)_n$ mG4s *in vitro* it is possible to either **i)** use crowding conditions or **ii)** allow the system to equilibrate for longer. Cell-crowding promotes multimolecular structures by bringing the strand closer together, while allowing a longer equilibration time simply permits the system to shift towards the thermodynamic product. Although this makes mG4s more

challenging to be assembled *in vitro,* it also highlights how the crowded live-cell environment might promote their formation.

In this chapter, I questioned if $(GGGGCC)_n$ could arrange into multimolecular G4s, and whether the formation of these structures could contribute trigger aggregation of the nucleic acids in the absence of proteins.

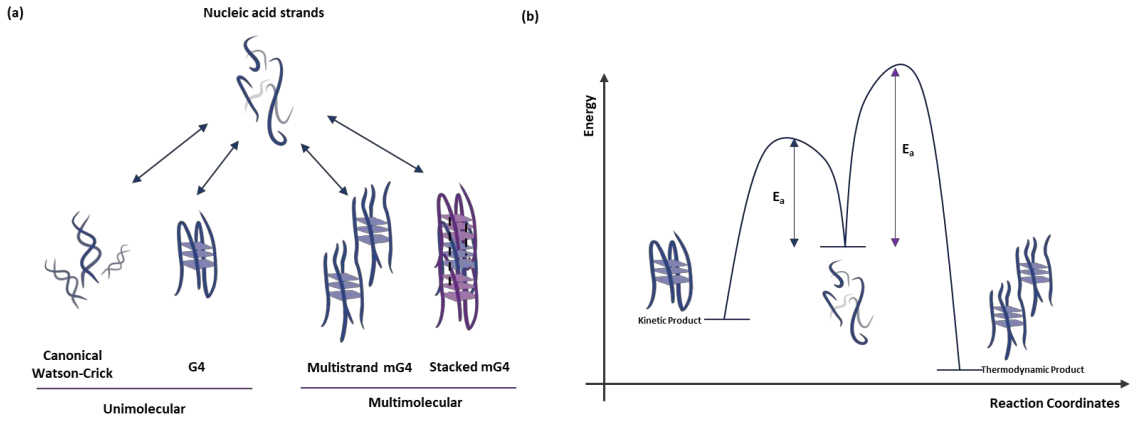


Figure 17: mG4s vs G4s from a thermodynamic viewpoint. The equilibrium between unimolecular species and multimolecular species is dictated by a balance between the loss of entropy of the system (loss of degrees of freedom due to multiple entities coming together) and a gain in enthalpy (stacking interactions). **(a)** Some of the possible conformations adopted by guanine rich nucleic acid strands, divided by molecularity. **(b)** Energy diagram for guanine rich nucleic acid strands able to arrange into G4s and mG4s. Whilst G4s are typically the kinetic product and therefore have a lower activation energy barrier, mG4s tend to be more thermodynamically stable upon formation, making them the thermodynamic product.

Note on thermodynamic versus kinetics stability

The thermodynamic properties of a system are defined by state functions which link to its overall properties and composition of a system in its equilibrium state. Indeed, state functions are not defined by the path taken by a system component to reach its final state, only by that final state itself. For instance, enthalpy, Gibbs free energy, entropy and others are all considered state functions. The thermodynamic product of a reaction process is the structure where the overall energy is minimised to assume its most stable conformation.

The kinetics property of a system instead describes the rate of a reaction. The kinetic product of a reaction is the structure that is reached faster after crossing the reaction energy barrier, regardless of whether that its absolute most stable conformation.

In the context of this work, mG4s are considered the thermodynamic product as, although they do not assemble as fast as G4s, they show a higher degree of thermodynamic stability. For instance, they are more resistant to melting with respect to their equivalent unimolecular G4s, indicating that stronger forces are holding the structure together.

3.2 Experimental Techniques

3.2.1 Thermal Annealing

Thermal annealing is a technique by which nucleic acids are thermally denatured by heating the samples to the annealing temperature (temperature at which all pre-existing H-bonding interactions can be denatured) and slowly cooled down to allow the system to equilibrate in its most stable conformation[122]. In this text, I set "annealing temperature" at 95 ℃, which is the maximum temperature which can be used to heat biological samples in an aqueous buffer prior to evaporation. Indeed, whilst the annealing temperature of a sample containing canonical dsDNA can be approximated, H-bonding holding together G4s are harder to estimate. To ensure that all non-canonical conformations are disassembled, in a typical G4 annealing experiment, samples are heated to 95 ℃ and cooled down to room temperature to allow for the formation of the G4s. Slower cooling rates and cell crowding conditions of the samples allow for a better rearrangement of the system into mG4s.

3.2.2 Microscopy

In this work "aggregates", are defined as a species observable in the microns scale which can be qualitatively observed with the use of a microscope. Given the microscopy set-up used for these experiments (Leica SP8 Inverted Confocal microscope – details in the Methods section), the resolution R value should correspond to about 0.7 µm (calculated as in **Equation 1**), meaning that any aggregate smaller than that cannot be visualised. Furthermore, other factors, such as the quality of the optics and the signal-to-noise ratio, can also affect the effective resolution of the microscope. Therefore, in order to ensure that the aggregates were accurately characterised, a DNA sample was considered containing aggregation only if containing elements of visible size > 1 µm which would not be confused with noise or other impurities in the solutions.

$$\mathbf{R(xy) = \lambda/2NA}$$

***Equation 1*:** Confocal microscope resolution in the xy plane. NA is the numerical amplitude and λ is the wavelength of excitation (588 nm) and NA is 0.4. Therefore, the resolution of the microscope used for these experiments is around 735 nm.

3.2.2.1 Confocal Microscopy

Confocal microscopy provides a high spatial resolution and optical sectioning, allowing for the visualization of three-dimensional structures within a specimen[173]. In these experiments confocal microscopy is used to detect $(GGGGCC)_n$ fluorescent aggregates and to perform G4-specific photo-oxidation assays.

3.2.2.2 Fluorescence Recovery After Photobleaching (FRAP)

FRAP allows the study of the dynamics of molecules within living cells or aggregating species. It involves selectively bleaching a small region of a fluorescently labelled sample, which in these experiment case was done with a 488 nm laser, and then monitoring the recovery of fluorescence intensity in the bleached area over time[174]. This recovery reflects the mobility and turnover of the labelled molecules. Indeed, whilst liquids are expected to have very fast recovery due to the fast diffusion coefficient, the molecules in solid particles have very limited movement and will not present any recovery in the timeframe of a few hours.

3.2.3 Agarose Gel Electrophoresis

Gel electrophoresis is used to separate biomolecules of different sizes based on their relative charge[175]. This is done by applying a voltage to an agarose gel matrix and allowing the biomolecules to pass through upon subjection to the electric current. In the case of nucleic acids, the negatively charged phosphate backbone will allow the DNA to travel down the agarose gel matrix. The larger the fragment, the further down the gel it will be able to travel. DNA on agarose gels can be visualized by staining with fluorescent DNA-binders such as Gel Red and SYBR Safe, which become fluorescent on the gel matrix upon stacking to the DNA bases[176]. The relative sizes of DNA fragments can usually be calculated by comparison with standard DNA ladders. This is not the case with non-canonical DNA conformations such as G-quadruplexes as they are more compact and travel faster down the matrix than their ssDNA equivalents[177].

3.2.3 Circular Dichroism

Circular dichroism (CD) is a spectroscopic technique used to study the differential absorption of left- and right-circularly polarized light by optically active molecules[178]. It provides information about the structural and electronic properties of chiral molecules, such as proteins, nucleic acids, and small organic molecules. In CD spectroscopy, circularly polarized light is passed through a sample, and the difference in absorbance between left- and right-circularly polarized light is measured. This difference in absorbance is typically expressed as the ellipticity (Θ) in degrees or molar ellipticity (Θ/M)[179]. The magnitude and sign of the CD signal provide insights into the secondary and tertiary structures, conformational changes, and interactions of biomolecules. In nucleic acids, the CD spectrum in the ultraviolet (UV) region is primarily linked to the base-pair stacking interactions in the helical structure[180]. Changes in the nucleic acid CD spectrum can indicate conformational transitions, such as the melting of double-stranded DNA into single-stranded DNA or to non-canonical secondary structures. G4s, in particular, present a very specific CD trace which varies depending on the topology of the quadruplex. Indeed, parallel G4s will tend to show a prominent positive peak at 263 nm and a negative one at 240 nm, while antiparallel G4s will show a positive peak at 295 nm and a negative one

at 260 nm[171]. CD melting is used to study the thermal stability and conformational changes of biomolecules such as proteins or nucleic acids. It involves monitoring changes in the CD signal of a protein as a function of temperature. A characteristic peak is taken in examination and the melting temperature (T_m) is extrapolated by calculating the temperature at which 50% of the sample is unfolded or denatured[179]. The higher the melting temperature (T_m), the more stable the structure was to begin with. In the case of G4s, upon heating, the H-bonding holding the structure together are weakened and the secondary structure disassembles.

3.3 Results and Discussion

3.3.1 The C9orf72 repeat expansion $(GGGGCC)_n$ can arrange into mG4s

To promote mG4-formation for $(GGGGCC)_n$ *in vitro*, where conditions like cell-crowding and other biomolecules interactions are absent, the nucleic acid sequences were subjected to a thermal annealing process. In thermal annealing, pre-existing interactions are broken *via* heating the sample to a certain temperature (usually 95 ℃ for G4s). However, it soon became evident that some pre-existing aggregates in the $(GGGGCC)_n$ stocks were resistant to a typical annealing step. Indeed, solid aggregates were visible under the microscope after annealing in different buffer systems, indicating lack of solvation of the oligomers into solution (***Appendix 1***).

Therefore, a heating denaturing step was deemed necessary prior to the addition of salts or other G4-forming conditions to prevent having pre-formed structures in solution (***Figure 18***).

3.3.1.1 Optimisation of the denaturing step

Several heating denaturing conditions were tested by diluting 100 µM $(GGGGCC)_{11}$ in a 10 mM TRIS-HCl (pH 7.4) buffer system (see Methods and ***Figure 18*** for details). Each sample was tested in presence and in absence of KCl salt. The presence of salts in the buffer was added as a positive control to prevent the disassembly of the preformed aggregates as it had previously been reported as a tool to stabilise DNA-based condensates[181].

Denaturing conditions	Buffer
4 h r.T.	TRIS-HCl
4 h r.T.	TRIS-HCl + 500 mM KCl
3 h 95℃ + 1 h cooling	TRIS-HCl
3 h 95℃ + 1 h cooling	TRIS-HCl + 500 mM KCl
1 h 95℃ + (-0.25 ℃/min) cooling	TRIS-HCl
1 h 95℃ + (-0.25 ℃/min) cooling	TRIS-HCl + 500 mM KCl

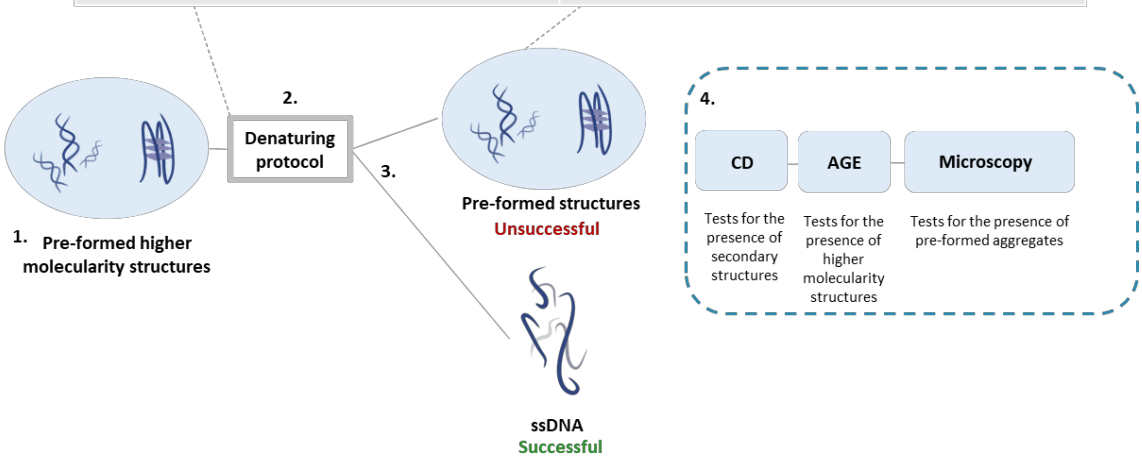


Figure 18: Experimental pipeline to assess the successful denaturing of $(GGGGCC)_{11}$ species. Firstly, the sample is diluted in TRIS-HCl in absence of KCl (or in presence of KCl in the case of the controls). It is then subjected to one of the annealing protocols and the results analysed *via* CD, AGE and microscopy. If higher molecularity secondary structures are still present in the samples, the denaturing process is considered unsuccessful. In the Table are highlighted the different heating denaturing conditions used for disassembling $(GGGGCC)_{11}$ pre-formed structures.

The success of the denaturing steps was judged *via* three orthogonal techniques: optical microscopy, agarose gel electrophoresis and circular dichroism (***Figure 18***).

Firstly, optical microscopy was used to assess the presence of fluorescent macroscopic aggregates for denatured $(GGGGCC)_{11}$ FAM-labelled (Carboxyfluorescein tag) sequence. Upon successful denaturing, $(GGGGCC)_{11}$ was expected to be fully dissolved in solution in its ssDNA form, not presenting any solid aggregates visible under microscopy. Given the added presence of the fluorescent tag, the micrograph should appear as a homogeneous fluorescent mix in the case of successful denaturing.

When incubated for 4 h at room temperature, $(GGGGCC)_{11}$ aggregates showed only a partial disassembly behaviour, with fewer and smaller aggregates with respect to the salt-containing control (***Figure 19a***). This likely occurred due to lack of salt stabilisation of the pre-formed DNA structures, which was enough to partially disassemble the stock solution. Samples heated at 95 ℃ for 3 h showed a significant amount of evaporation in the well, with the accumulation of crystallised DNA fragments

present at the bottom (***Figure 19a***). This implied that the conditions used to disassemble the sample were not compatible with the buffer system and DNA had crystallised upon evaporation. When $(GGGGCC)_{11}$ was heated at 95℃ for 1 h and then slowly cooled down to room temperature (-0.25 ℃/min), no aggregates were visible in the final sample and no significant evaporation was observed, indicating a successful denaturation. For comparison, the salt-containing control sample still contained fluorescent DNA aggregates (***Figure 19a***).

Next, agarose gel electrophoresis (AGE) was employed to detect sub-microscopic multimolecular structures. Upon successful denaturation, the DNA in solution was expected to be in its lower molecularity ssDNA form, manifesting as a distinct lower band at around 66 bases for $(GGGGCC)_{11}$ (***Figure 19b***).

Both samples (KCl and non-KCl containing) incubated for 4 h at room temperature presented two bands, one at approximately 60 base pairs (bp) and another below the 50 bp mark, likely representing $(GGGGCC)_{11}$ in its unimolecular and bimolecular state (***Figure 19c***) and indicating an unsuccessful denaturation. When subjected to heating at 95 ℃, both in the case of the 3 h and 1 h heating times, the higher molecularity band disappeared, signifying the successful disassembly of higher molecularity species. The samples heated for 3 h, however, presented very faint bands and a significant amount of smearing in the wells, most likely due to the excessive evaporation in the sample. Notably, the lower band, which should have been located at approximately 66 bp for the unimolecular species, was quite broad and migrated to lower molecular weights than predicted (around 50 bp). This was probably due to the presence of a mix of ssDNA and unimolecular G4s in solution which caused the band to be broader and more shifted than expected. Quantification of the respective bands in the AGE could not be performed as unequal amounts of DNA were loaded onto the wells post-annealing. These differences were due to the partial aggregation of the samples, which led to different effective concentrations of DNA in solution.

The presence of multiple species was further confirmed by the weak and uncharacteristic signal in the circular dichroism spectra of $(GGGGCC)_{11}$ in absence of KCl (**Figure 21*d***)[182].

The combination of optical microscopy, AGE and CD confirmed that the final denaturing protocol was successfully disassembling any preformed $(GGGGCC)_{11}$ structures. This was of fundamental importance for the following steps, as a successful denaturing step allowed us to ensure the controlled formation of multimolecular species *in vitro*. It is unclear whether previous studies performed on the C9orf72 repeat sequence have gone through the same denaturing process as it has never been explicitly mentioned in the relative protocols.

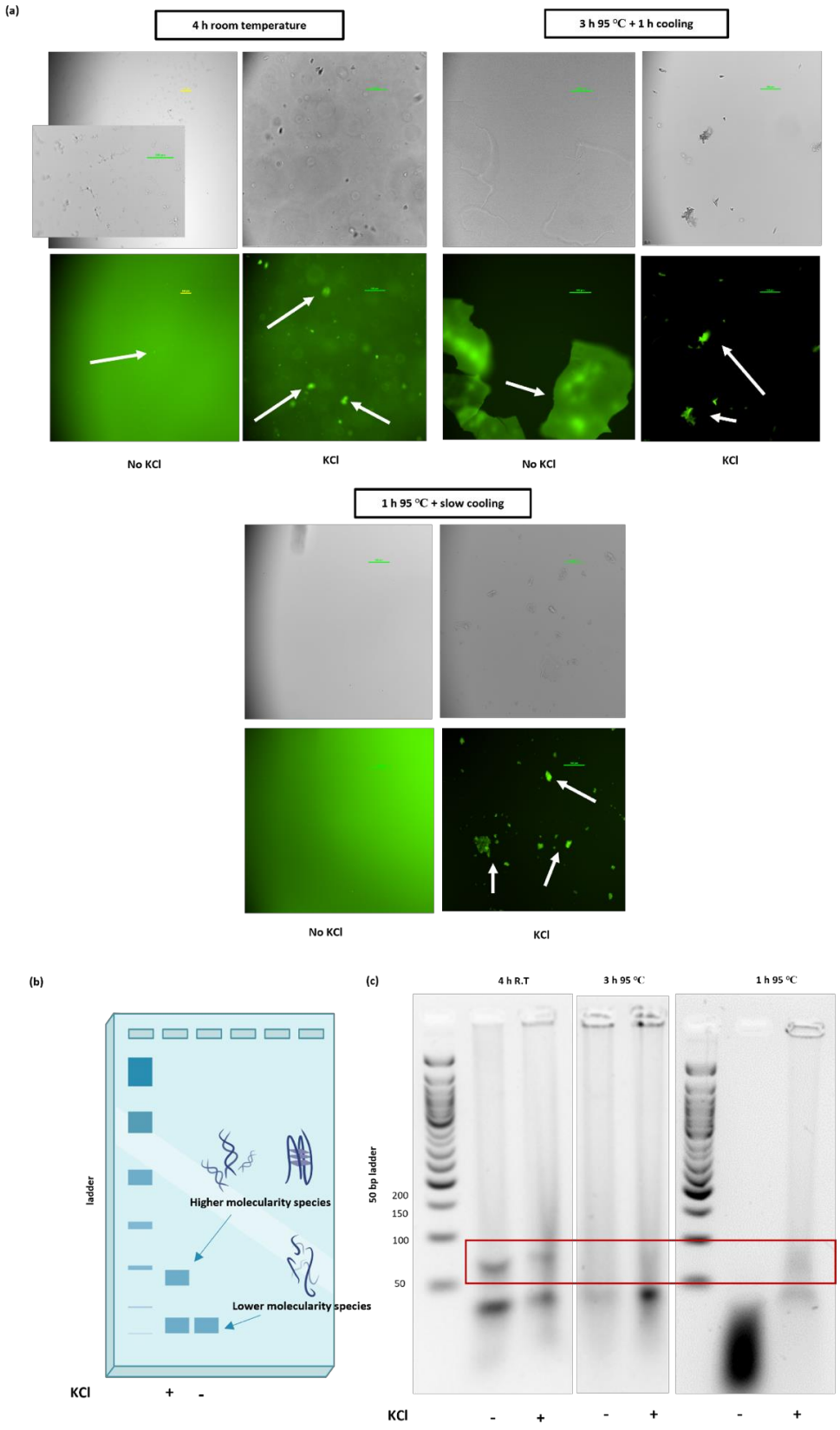


Figure 19: Optimisation of the disassembly step for $(GGGGCC)_n$ prior to annealing. (a) Optical and fluorescence microscopy of $(GGGGCC)_n$ upon different denaturing conditions. **T**he samples are unable to disassemble if KCl is present in any of the

conditions tested. In absence of salt, incubation at room temperature presents partial disassembly behaviour. Heating at 95 °C for 3 h leads to significant evaporation of the sample and dried out DNA is shown clearly in the fluorescent channel. A more moderate 1 h heating temperature followed by a slow colling successfully disassembled all the pre-formed aggregates. The optimised denaturing conditions were tested for $(GGGGCC)_{11}$ in biological triplicates (**Appendix 2**). **(b)** Expected agarose gel electrophoresis experiment outcome for the denaturing of $(GGGGCC)_n$. Samples containing KCl are expected to not disassemble and to maintain the pre-formed structures of mixed molecularity from the stock. Samples successfully denatured should only present a lower molecularity band. **(c)** Agarose Gel electrophoresis of $(GGGGCC)_{11}$ subjected to different denaturing conditions. As evident from the gel, the structures are pre-arranged into a mixed molecularity state with a higher band in between 50 and 100 bp and another below 50. Upon heating at 95°C for 1 h, the higher molecularity species disassemble completely, and only lower mixed molecularity states are left. Notably, samples are denatured at 100 µM and diluted to 10 µM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE, on ice at 50 V for 2.30 h. The samples contain a FAM label and are stained with SYBR Safe. Optimised denaturing conditions were tested for several repeat lengths in biological triplicates (**Appendix 2**).

3.3.2.2 Optimisation of mG4 forming protocol

After melting $(GGGGCC)_{11}$ into its ssDNA form, re-assembly into G-quadruplexes structures was triggered using 5 different protocols to test, which had the potential to promote multimolecular species over the unimolecular ones (for details see Methods and ***Figure 20***).

	mG4 forming conditions	Buffer
1.	Isothermal assembly	TRIS-HCl
	Isothermal assembly	TRIS-HCl + 500 mM KCl
2.	1 h 95°C (-0.25°C/min) cooling	TRIS-HCl
	1 h 95°C (-0.25°C/min) cooling	TRIS-HCl + 500 mM KCl
3.	1 h 95°C + (-0.25 °C/min) cooling + PEG	TRIS-HCl
	1 h 95°C + (-0.25 °C/min) cooling + PEG	TRIS-HCl + 500 mM KCl
4.	1 h 95°C + (-0.01 °C/min) cooling + PEG	TRIS-HCl
	1 h 95°C + (-0.01 °C/min) cooling + PEG	TRIS-HCl + 500 mM KCl
5.	1 h 95°C + (-0.02 °C/min) cooling + PEG	TRIS-HCl
	1 h 95°C + (-0.02 °C/min) cooling + PEG	TRIS-HCl + 500 mM KCl

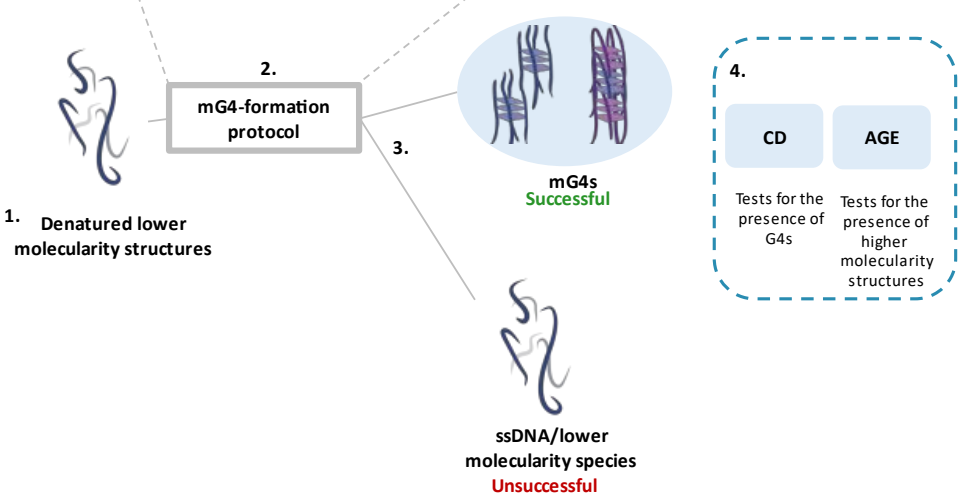


Figure 20: Experimental pipeline to assess the successful mG4-formation protocol of $(GGGGCC)_{11}$ species. Firstly, the sample is diluted in TRIS-HCl in presence of KCl (or in absence of KCl in the case of the controls). It is then subjected to one of the annealing protocols highlighted in the Table and the results analysed *via* CD and AGE. If higher molecularity secondary structures with a G4-trace are present in the samples, the mG4-formation process is considered successful. In the Table are highlighted the different mG4-forming protocols used in the text by number.

For all protocols, the presence of multimolecular species and their arrangement into G4s was assessed respectively *via* AGE and circular dichroism and each condition was tested in presence and absence of KCl. The samples without KCl lacked the additional G4-stabilising conditions that should promote mG4 formation, working as a negative control for the experiment.

MG4s should appear as higher molecularity species in the gel with respect to the denatured ssDNA product (***Figure 21a***). In protocol **(1)**, KCl was added to the ssDNA denatured product to probe mG4 formation and, consequently, a shift in the molecularity. This transition was monitored over 6 days by AGE observing the location of the higher molecularity band with respect to its negative control.

Indeed, in ***Figure 21b***, the ssDNA-corresponding band slowly shifted to higher molecular weights over time, as the system reassembled into bimolecular G4s (approximately between the 100 and 200 base pairs mark). Notably, the bands were broad and did not correspond to exact molecular weights because the G4s formed travelled through the agarose gel matrix at different rates than the corresponding DNA ladder used as reference. Isothermal incubation of the ssDNA denatured produced in KCl containing buffer was then performed at higher temperatures (36, 45 and 60 ℃), providing the system with additional thermal energy to probe the transition to the higher molecularity species. As in the previous case, the progress was monitored for 6 days *via* AGE. In ***Figure 21c***, it is possible to distinguish the difference between samples incubating the G4-favouring KCl salt (higher molecularity bands in the gel matrix) and the control denatured samples (lower molecularity bands). Although a higher thermal energy provided to the system was expected aid mG4-formation and give a more pronounced shift in molecularity of the system, this was not observed *via* electrophoresis, where all samples ran similarly.

Samples were then subjected to circular dichroism to validate formation of a G4-structure. As shown in ***Figure 21d***, all samples from protocol 1 contained a weak G4-trace of mixed topology (positive peaks at 295 nm and 263 nm and negative peak at 240 nm).

In protocol 2 (1 h 95 ℃ and cooling with a temperature ramp of – 0.25 ℃ min^{-1}), the system was heated to higher temperatures and cooled with slow rates to favour multimolecular species. The kinetic product (unimolecular G4s) should still be able to form, but the high temperatures and slower cooling time conditions should allow the system to rearrange into the multimolecular product. AGE (***Figure 21e***) showed an expected shift in the population to higher molecular species in presence of KCl. The CD spectrum presented evidence of parallel G4 species (positive peak at 263 nm and negative peak at 240 nm) in KCl containing samples (***Figure 21d***).

In protocol 3 (1 h 95 ℃ and cooling with a temperature ramp of – 0.25 ℃ min^{-1} in presence of PEG) the system was additionally provided 30% PEG (w/v) to simulate cell crowding conditions. Multimolecular species are favoured by local high concentrations and cell crowding promotes a good environment for the event to occur. AGE (***Figure 21e***) presented a shift towards higher molecular species in presence of KCl, while circular dichroism analysis proved a prevalence for the parallel topology (***Figure 21d***).

Collectively, AGE and CD showed that slower cooling steps and the presence of a crowding agent favoured the formation of multimeric structures, likely by allowing time and energy conditions for the system to equilibrate at higher molecularity. Therefore, we questioned whether further decreasing the cooling rate could further promote the process.

In Protocols 4 and 5 the system is allowed to cool down with a slow cooling rate in presence of a crowding agent (respectively – 0.01 ℃ min^{-1} and – 0.02 ℃ min^{-1} cooling rate). As expected, relative band intensities in the AGE in ***Figure 21e*** suggested that the yield of higher molecularity species was better for the slower cooling step (– 0.01 ℃min^{-1}) as the system was given more time to equilibrate. Although the band intensity was a good indication of relative concentrations, it is important to notice that this is not a quantifiable effect as G4 species and canonical DNA conformations cannot be stained with DNA-dyes in similar manners. The samples in protocol 4 were analysed *via* CD and confirmed the presence of the typical parallel topology G4-trace (***Figure 21d***).

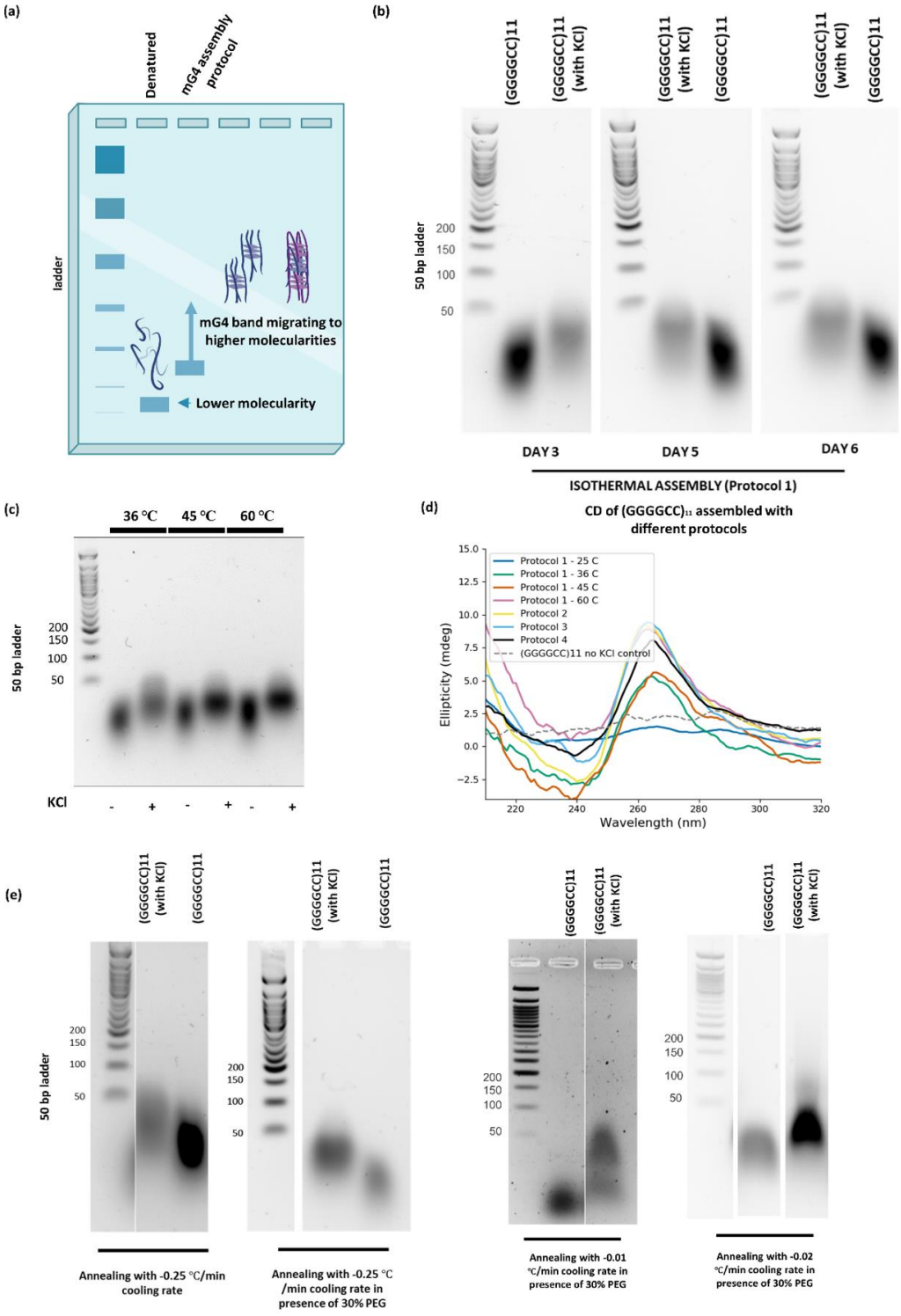


Figure 21: mG4 assembly protocol optimisation. (a) Expected agarose gel electrophoresis features of a successful mG4 assembly protocol. In presence of mG4 favouring conditions, the lower molecularity species rearrange into higher molecularity structures causing a shift into higher bands. **(b)** Agarose gel electrophoresis of $(GGGGCC)_{11}$ samples under protocol 1. In absence of KCl in the buffer, the sample does not assemble into higher molecular species. However, after 3 days, the population starts to shift towards higher molecular species and on day 6 a faint band at 66 base pairs (bimolecular

species) can be identified. **(c)** Agarose gel electrophoresis of $(GGGGCC)_{11}$ samples subjected to protocol 1 for 6 days. In absence of KCl in the buffer, the sample does not assemble into higher molecular species. However, after 6 days, the population starts to shift towards higher molecular species. There not seems to be any difference between isothermal assembly at different temperatures. **(d)** Circular Dichroism spectra taken at 25°C for 10 μM $(GGGGCC)_{11}$ samples in TRIS HCL buffer (10 mM pH 7.5 + 500 mM KCl). The typical G4 trace can be observed for samples except samples annealed *via* Protocol 1 (25, 36, 45°C) and the control non containing KCl. The spectra have been subjected to a baseline correction when necessary. CD spectra were acquired as technical triplicates by selecting accumulation = 3 on the spectrometer. Biological repeats were taken for the protocols that showed the highest yield of multimolecular species (protocols 2,3,4) and for the negative no-KCl controls **(Appendix 3). (e)** Agarose gel electrophoresis of $(GGGGCC)_{11}$ samples under protocols respectively 2,3, 4 and 5 (from right to left). In absence of KCl in the buffer, the sample does not assemble into higher molecular species, regardless of the cooling rate or other annealing or other annealing conditions. In presence of KCl salt, the population shifts to higher molecular species. Notably, for all gels reported the samples are denatured at 100 μM and diluted to 10 μM for AGE. The gels running conditions are the following: 3% agarose in 1X TBE, on ice at 50 V for 2.30 h. The samples are stained with SYBR Safe. The optimised mG4 forming conditions were tested in biological duplicates to ensure repeatability (**Appendix 4**).

3.3.2.3 Abundance of stable $(GGGGCC)_n$ mG4 structures dependence on annealing protocol

After testing $(GGGGCC)_n$ ability to form multimeric species under different annealing protocols, I questioned whether not only the yield, but also the stability of the resulting mG4s was affected by the different formation steps. CD melting was performed on $(GGGGCC)_{11}$ annealed samples to assess which annealing protocol populated the most stable structure. The resulting curves showed that the G4s obtained *via* protocol 4 (heating at 95 °C for 1 h and cooling with – 0.01 °C min^{-1} temperature ramp) had the highest stability resulting into a higher melting temperature (T_m = 87 °C). In order of increasing melting temperature, the other annealing protocols followed as: protocol 2 (heating at 95 °C for 1 h and cooling with – 0.25 °C min^{-1} temperature ramp)($\boldsymbol{T_m}$ = 82 °C) < protocol 3 (heating at 95 °C for 1 h and cooling with – 0.25 °C min^{-1} with PEG) ($\boldsymbol{T_m}$ = 86 °C) < protocol 4 (heating at 95 °C for 1 h and cooling with – 0.01 °C min^{-1} temperature ramp) ($\boldsymbol{T_m}$ = 87 °C) (***Figure 22***).

The combination of AGE and CD melting curves results showed that when the $(GGGGCC)_{11}$ system was subjected to cell crowding conditions and/or slower annealing cooling rates it yielded more stable G4s and higher molecularity species. As the G4s formed *via* protocol 4 (-0.01℃/min cooling rate in presence of PEG) were the most stable, I decided to use this methodology for the following experiments. However, to avoid long experimental procedures, I chose a faster cooling rate. Indeed, protocol 5 (-0.02 ℃/min cooling rate in presence of PEG), although showed a less prominent formation of higher molecularity *via* AGE (***Figure 21e***), offered a much faster experimental procedure allowing multiple repeats and a higher throughput approach.

It is important to note the melting values are obtained *via* an estimation of the melting data as the G4 structure was not completely disassembled even at 95 °C, the bottom plateau of the sigmoidal curve is not reached, and fitting of the raw data is not accurate (***Equation 2***). This leads to a higher variance in the extrapolation of the T_m values, indicating that the difference of a few degrees between melting curves may simply be due to experimental errors. Given the results from this first set of experiments and literature knowledge that mG4s are known to have melting temperatures of > 95 °C in KCl-containing buffers[195], CD melting was subsequently considered an inaccurate method to establish the relative stability of the structures formed by the annealing protocol and was not repeated.

$$\boldsymbol{\theta(\mathrm{mdeg}) = \ \theta_{min} + \frac{(\theta_{max} - \theta_{min})}{1 + \exp\left(\frac{T_m - T}{\theta'}\right)}}$$

Equation 2: Boltzmann sigmoid for fitting melting curve where Θ is the normalised ellipticity [156].

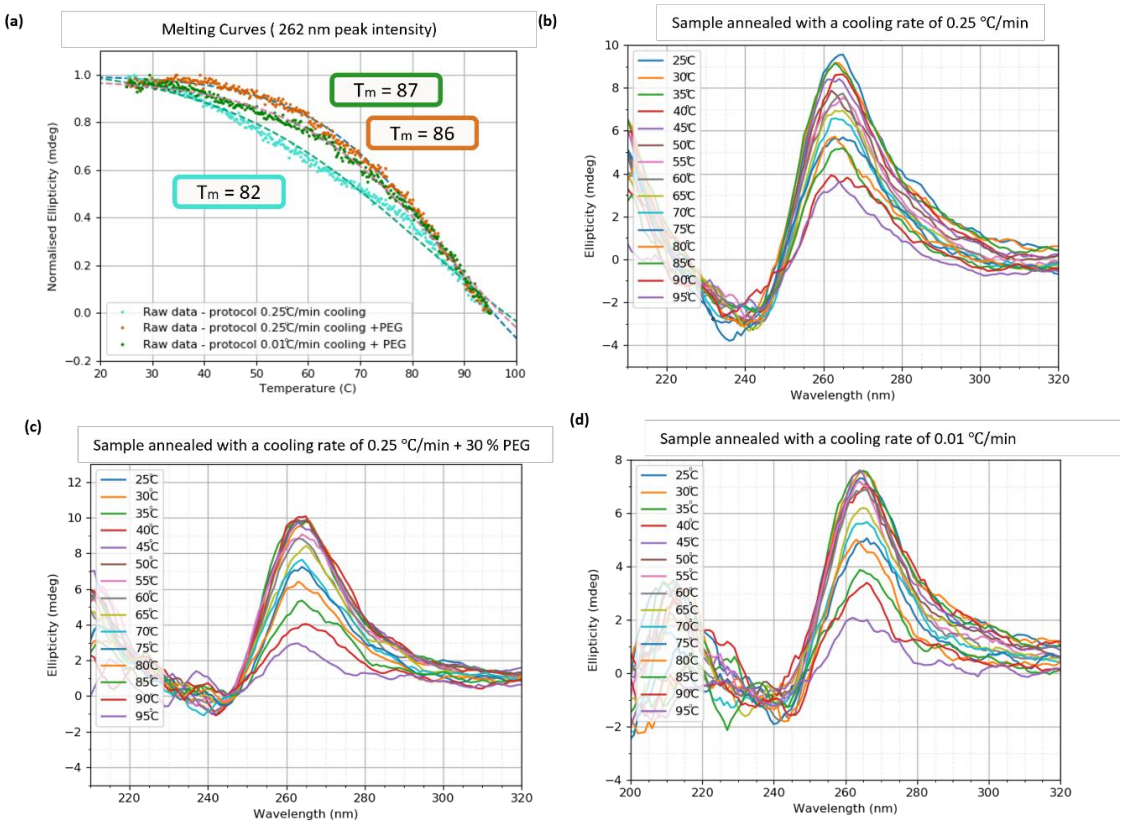


Figure 22: Circular Dichroism melting experiment. (a) Melting curve for G4s obtained in protocols 2, 3, 4. The temperature ramp used for the experiment is of 1 ℃/min with data points collected every 0.2 ℃. The individual spectra at each temperature are shown in **(b)** for protocol 2 **(c)** for protocol 3 and **(d)** for protocol 4. In all cases samples were annealed at 100 µM in 10 mM TRIS-HCl and diluted to 10 µM for spectra acquisition.

3.3.2. $(GGGGCC)_n$ mG4s characterisation

Although previous CD experiments showed that the bulk solution was prevalently composed of stable G4s upon annealing and AGE confirmed the presence of a multimolecular stoichiometry, further experiments were needed to confirm that the higher molecularity species observed were in fact mG4s and not a consequence of other nucleic acids interactions. Given the high G/C content of the sequence, other higher molecularity structures could originate from canonical Watson-crick H-bonding interactions. Indeed, $(GGGGCC)_n$ has previously been observed to arrange into hairpin conformations which, much like G4s, also can form multimeric species[112].

To confirm that the multimolecular species observed in the agarose gel could be ascribed to mG4s, $(GGGGCC)_6$ (50 µM) was subjected to annealing conditions in presence of different KCl concentrations (0-500 mM). As G4-formation is highly dependent on the concentration of K^+, we expected to see a change only in the species containing the non-canonical secondary structure (***Figure 23a***). Indeed, AGE experiments revealed more intense slowly migrating bands at higher K^+ concentrations, with no high-molecularity species being observed at sufficiently low K^+, confirming the role of G4s in mediating the formation of multimeric species (***Figure 23a***). The lower bands were likely to be a mixture of unimolecular/lower molecularity G4s and unstructured ssDNA sequences. In fact, although to a lesser extent with respect to the top band, the lower band was also migrating upwards with increasing concentration of potassium, indicating that the lower population was too affected by the presence of stabilising K^+ ions.

Further confirmation of the presence of G4s into the higher molecularity band was given by staining the gel with NMM[183]. NMM is small molecule, nucleic acid-specific fluorescent dye with a high selectivity for G4 structures. NMM fluorescence is enhanced in presence of G4s due to its π-stacking abilities on top of the G4-tetrad preventing NMM-NMM self-quenching. A staining protocol was developed to utilise the NMM-specific fluorescent ability to stain AGE gels and observe the presence of mG4s. In the protocol, the $(GGGGCC)_6$ sample is annealed and run in the exact same conditions as for previously analysed AGE. However, instead of inserting the SYBR Safe dye into the gel, this is subjected to 1 h incubation in an NMM and 1X TBE solution after running. The gel is imaged on the EtBr channel (526 nm excitation and 610 nm emission), as commercial gel imagers are not able to selectively detect the NMM emission range (600-650 nm) and/or excite at the appropriate wavelength (399 nm)[183]. If the higher molecularity bands did indeed present a high G4 content as the other experiments had suggested, this band should efficiently be stained by the NMM dye and visible on the

fluorescent channel. Similarly, the lower band, should contain G4s of a lower molecularity (albeit to a lesser extent) and be susceptible to NMM-staining. Even though the poor AGE quality in ***Figure 23b*** does not allow for a clear distinction of the multimolecular bands, the gel shows the expected outcome. The low quality of the stained gel comes from the specific fluorescent properties of NMM. In fact, whilst nucleic acid-staining dyes such as SYBR Safe, EtBr and Gel Red have been optimised for agarose gel use[184], NMM is usually utilised for other purposes. In fact, not only commercial gel imagers are not able to excite and observe NMM in its specific channels, but the dye itself has quite a low quantum yield (0.64)[185] compared to other DNA-stains such as SYBR Green (0.8)[186]. Furthermore, as the gel imager used for acquisition was designed for EtBr (exc. 301 nm and emi. 603 nm)[187], NMM (exc. 399 nm emi. 610 nm) detection was limited due to the different excitation and emission curves overlap. However, it is possible to observe that the dye selectively stains the $(GGGGCC)_6$ top band more efficiently than the lower one and does not fluoresce upon contact with the non G4-control or the DNA ladder. As a control, the same samples were stained with SYBR Safe (***Figure 23b***), showing clearly where the control band and the ladder were expected to be seen.

The combination of KCl titration experiments and NMM-staining confirmed that the higher molecularity species that were previously observed upon subjecting $(GGGGCC)_6$ to mG4 annealing conditions were in fact mG4s, confirming that the optimised protocol was successful.

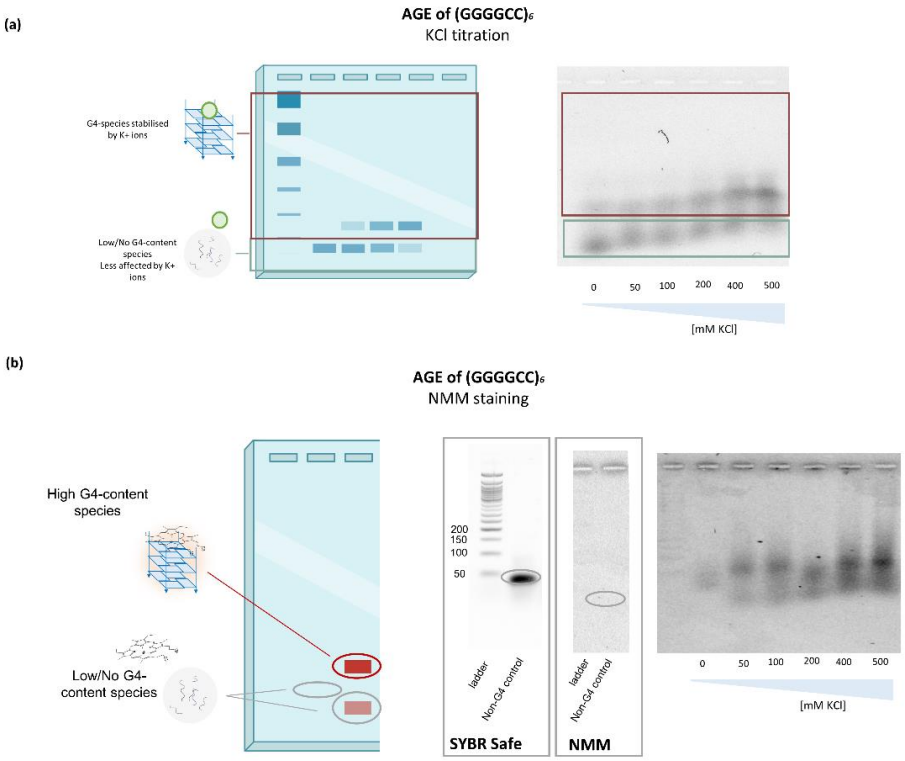


Figure 23: $(GGGGCC)_n$ forms higher molecularity species that can be ascribed to mG4s. (a) Agarose gel electrophoresis of $(GGGGCC)_6$ in different KCl conditions. **On the left.** Expected gel outcome at different KCl concentrations. If the top band is due to higher molecularity species dependant on G4s, these should only appear at higher KCl concentration and exist in a multiple population state in between. **On the right**. Experimental AGE of $(GGGGCC)_6$ annealed at different KCl concentrations. The top band, characteristic of higher molecularity species, is only present at higher concentrations of KCl. The samples are annealed at 50 µM and diluted to 10 µM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE, on ice at 65 V for 80 mins. The samples are labelled with FAM. The KCL titration experiment was performed in biological duplicates (**Appendix 5**). **(b)** Agarose gel electrophoresis of $(GGGGCC)_6$ stained with G4-specific dye NMM. **On the left.** Expected gel outcome upon staining with NMM dye. The non-G4 control sequence is expected to not stain due to lack of G4s, while the bands containing $(GGGGCC)_6$ are expected to present fluorescence. The experiment is confirming the hypothesis that the higher band contains a superior G4 amount if the dye preferentially stains the top band. **On the right.** Experimental AGE for the control sequences and KCl titration stained with NMM. As expected, both higher and lower bands in $(GGGGCC)_6$ are stained by the dye, indicating the presence of a mix of higher molecularity and lower molecularity G4s. The samples are annealed at 50 µM and diluted to 10 µM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE, on ice at 65 V for 80 mins. The experiment was performed as a single biological replicate but was repeated later in the text for different repeat lengths.

3.3.2.5 $(GGGGCC)_n$ (n = 1-12) can arrange into mG4s

Once established reliable conditions to generate mG4 species for $(GGGGCC)_{11/6,}$ the same protocol was used for a range of other repeat lengths. Notably, $(GGGGCC)_n$ is correlated to ALS/FTD pathogenic aggregates in a length dependent fashion (e.g. higher repeat lengths lead to the pathogenesis)[169], suggesting the importance of demonstrating that different repeat lengths presented mG4-forming abilities.

$(GGGGCC)_n$ (n = 1-12) sequences were subjected to the optimised mG4 annealing protocol (annealing concentration of 250 µM) in a TRIS-HCl buffer containing 500 mM KCl. CD spectroscopy was performed on $(GGGGCC)_n$ (n = 4,6,9,12) to assess bulk formation of the G4 species. As shown in ***Figure 24a*** the repeat lengths tested presented the typical trace of a parallel G4. The G4-trace proved that the mG4-forming protocol led to bulk formation of G4-species. Notably, the relative intensities of the G4-trace for different samples could not be quantified due to differences in concentrations in the samples due to aggregation.

To test for the presence of multimeric DNA species, $(GGGGCC)_n$ (n = 1-12) post-annealing were analysed *via* AGE. As a control, an additional annealing condition was tested in parallel, where 500 mM LiCl salt was added instead of KCl. The LiCl conditions were used to test that the higher molecularity species observed where not caused by unspecific interaction pushed by the high salt concentration, but by the formation of mG4s. In fact, whilst LiCl and KCl solutions have the same ionic strength, K^+ ions are able to stabilise G4s to a greater extent[188]. Therefore, the expected outcome of the control was to have a lesser presence of multimeric species compared to their relative same-n sample (***Figure 24b***). A first gel was run for all the repeat sequences annealed either in KCl or LiCl containing buffers (***Figure 24c***) at 250 µM. Each $(GGGGCC)_n$ sample had a higher molecularity band which was more intense in KCl conditions and a lower molecularity band which was stronger in the LiCl buffer. Interestingly, $(GGGGCC)_1$, $(GGGGCC)_2$ and $(GGGGCC)_3$ showed very little differences between KCl and LiCl conditions. Indeed, these sequences were too short to form consensus unimolecular G4 structures and could solely arrange into multimolecular species. The AGE results provided further confirmation that KCl was able to promote the formation of multimeric over monomeric species. LiCl, although still offering some degree of G4-stabilisation to the system, only allowed the sequences to arrange into their lower molecularity product. When the sequences were too short to form anything but the multimeric stoichiometry, then the LiCl stabilisation was enough to push the system towards higher molecularities species.

To confirm that the multimeric species observed contained G4s, the $(GGGGCC)_n$ (n = 1-11) (250 µM) samples were once again annealed in mG4-forming conditions in a KCl-containing buffer and NMM-stained AGE was performed on the resulting sequences. In ***Figure 24d***, the equivalent SYBR Safe and

NMM stained gels are analysed side by side for comparison. Each one of $(GGGGCC)_n$ samples presented a higher and lower molecularity band which could be stained with SYBR Safe – used to stain all DNA-containing species. Both bands, however, could also be stained with the G4-specific NMM, additionally indicating the presence of G4s (***Figure 24d***).

In conclusion, the mG4-forming protocol successfully led to the formation of higher molecularity species, ascribed to mG4s, for multiple repeat lengths. These results confirmed that the ALS/FTD responsible hexanucleotide repeat sequence $(GGGGCC)_n$ has the ability arrange into mG4s *in vitro* in absence of proteins.

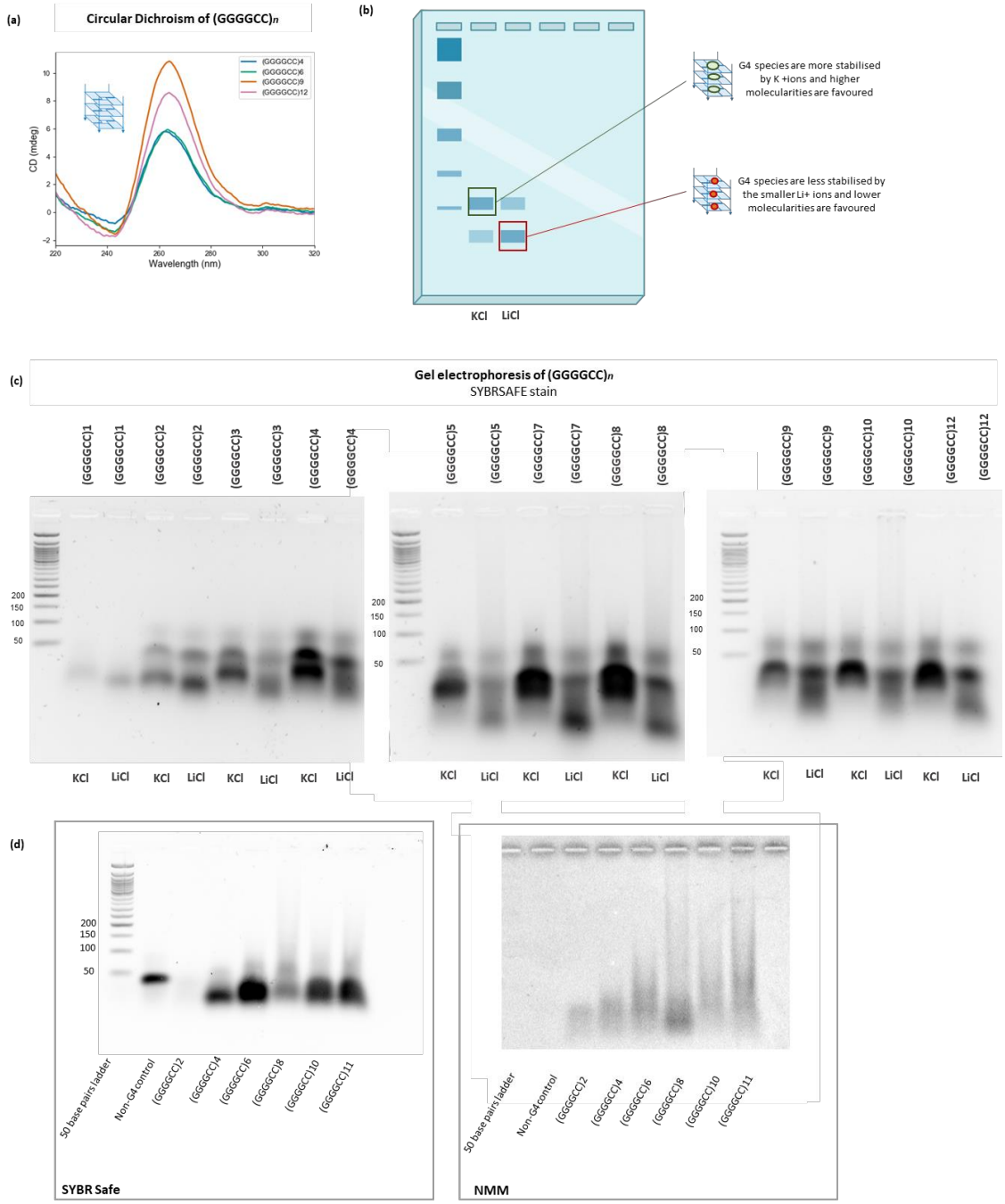


Figure 24: (GGGGCC)$_n$ can form mG4s at multiple repeat lengths. (a) (GGGGCC)$_n$ CD spectra - (GGGGCC)$_n$ (n = 4, 6, 9, 12) were annealed at 500 µM, 250 µM, 100 µM and 50 µM under mG4-forming conditions and further diluted to 5 µM for CD. All the samples present a positive peak at about 263 nm and a negative peak at 240 nm, both associated to a parallel G4 conformation. The spectra are acquired by the spectrometer as technical triplicates (accumulation = 3). The experiment was repeated as biological replicates for selected repeat lengths (**Appendix 6**). **(b)** Theoretical AGE highlighting the differences between LiCl control and KCl buffer systems for annealing. LiCl is known to stabilised G4s to a lesser extent, and it is therefore expected to promote higher molecularity species to a lesser extent. **(c)** Experimental AGE gel of (GGGGCC)$_n$ for n = 1-12. KCl containing samples are more prone to form higher molecularity species, as evident by the stronger higher bands. The samples are annealed at 250 µM and diluted to 10 µM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE,

on ice at 65 V for 80 mins. The samples are stained with SYBR Safe. Experiment was performed in biological duplicates (**Appendix 7**). **(d)** $(GGGGCC)_n$ agarose gel electrophoresis – SYBRS Safe (left) and NMM (right) stains. 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. The samples are annealed at 250 μM and diluted to 10 μM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE, on ice at 65 V for 80 mins. Experiment was performed in biological duplicates (**Appendix 8**).

3.3.2 $(GGGGCC)_n$ can aggregate in vitro in absence of proteins

3.3.2.1 $(GGGGCC)_n$ phase diagram

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, I employed brightfield microscopy to screen for the emergence of aggregates.

Initially, $(GGGGCC)_n$ (n = 1-12) sequences were subjected to mG4-forming conditions upon annealing at 350 μM, a concentration in the same order of magnitude to that used to acquire the AGE in ***Figure 24c***. On the hypothesis that mG4s were a sufficient characteristic of the system to probe aggregation, a selected number of repeat lengths were observed under brightfield microscopy.

Imaging showed that $(GGGGCC)_n$ (n = 3, 4, 5, 7, 8, 11) sequences had ability to aggregate in a length dependant fashion (***Figure 25a***). The length dependence trend in the aggregation was not surprising as sequences with a higher n are more prone to arrange into mG4s as they have **(1)** a higher guanine content and therefore a better ability to form G4s and **(2)** a higher number of conformations that the mG4s can assume. In fact, even though the samples were annealed at the same concentrations, aggregates were visualised on brightfield only at repeat length $n > 5$, suggesting a direct proportionality trend between aggregation extent and the sequence ability to fold into mG4s. Since aggregation seemed to be more favourable in conditions that promoted mG4s, these results constituted the first piece of indirect evidence that these multimolecular structures were the prevalent species driving the aggregation. Importantly, this aggregation was visualised *in vitro* in absence of additional aggregation-prone proteins, suggesting that $(GGGGCC)_n$ might not need other biomolecules to trigger pathological aggregation in ALS/FTD.

To further expand these studies, $(GGGGCC)_n$ (n = 1-12) sequences were subjected to mG4-forming conditions at different annealing concentrations (0-500 μM). In the experiment, aggregation dependence was assessed in function of repeat length and annealing concentration with the scope to

link aggregation to mG4-formation. In fact, as mG4-formation is driven by high annealing concentration, the aggregation extent was expected to increase upon higher DNA presence.

The completed phase diagram depicting $(GGGGCC)_n$ aggregation at different repeat lengths and annealing concentrations is shown in ***Figure 25b*** and **Appendix *9***. As predicted, aggregation was more significant for higher *n* and/or higher annealing concentrations, indicating a correlation between aggregation extent and the sequence mG4-forming ability. Another phase diagram was formulated by expressing the DNA concentration in µg/µL rather that molar concentrations. The µg/µL conversion was performed to confirm that aggregation was not obtained when a certain number of guanines where present in solution, but it was observed with a sequence specific length dependence. The findings, depicted in ***Figure 25c***, confirmed that there was no fixed threshold of µg of DNA concentration which led to aggregation. In fact, below 1000 µg/µL only higher repeat lengths showed aggregation abilities. Notably, the aggregation extent could only be qualitatively assessed from the micrographs. In order to perform a more in-depth analysis to quantify the aggregation extent, fluorescent staining of the aggregate and several technical repeats would be required to minimise the error coming from image analysis software such as ImageJ. Such large-scale experiment requires the use of high-throughput plating, annealing and imaging facilities which were not available at the time of the experiment.

The microscopy results collectively showed that once subjected to a mG4 annealing protocol, $(GGGGCC)_n$ can aggregate *in vitro* in absence of protein with a length and annealing concentration dependence, suggesting a correlation with the presence of the secondary structures themselves.

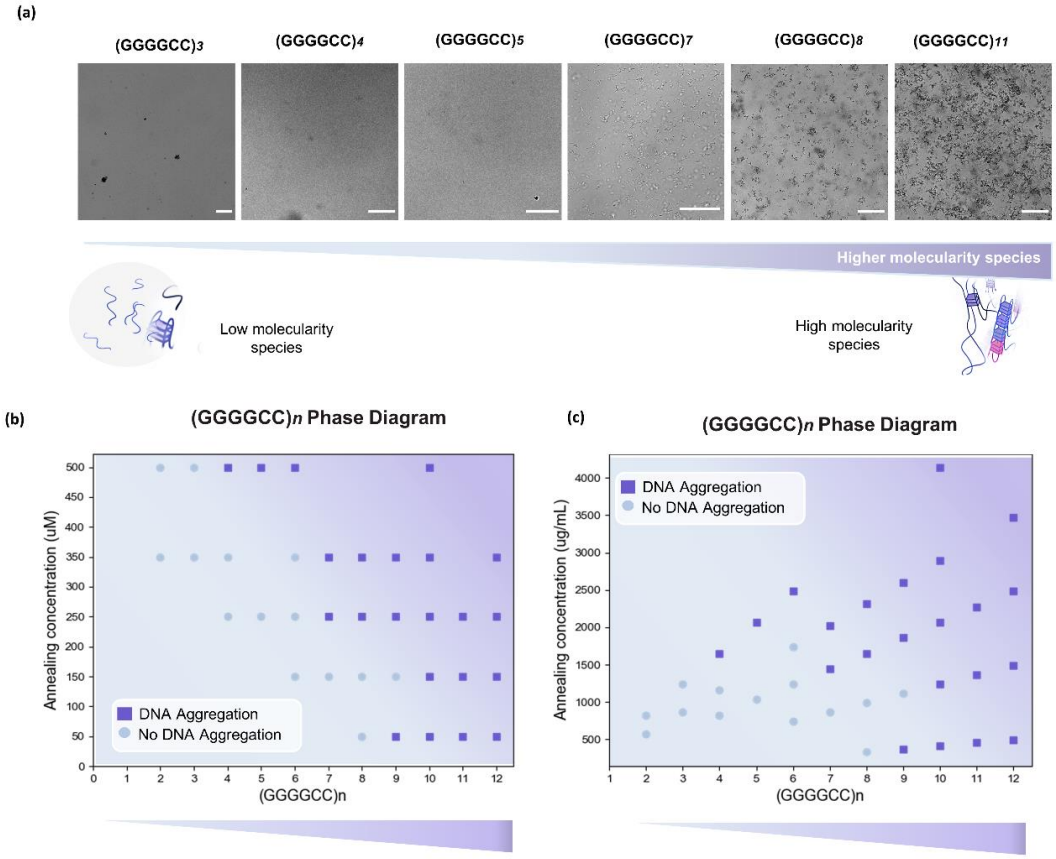


Figure 25: (GGGGCC)$_n$ can aggregate in absence of proteins. (a) Brightfield images of (GGGGCC)$_n$ subjected to mG4-forming conditions. Samples were annealed at 350 µM and directly observed under the microscope. Aggregation is present for higher repeat lengths. Scale bar is 100 µm. **(b)** Phase diagram of (GGGGCC)$_n$ subjected to mG4-forming conditions. Each sample was defined as "aggregating" or "non-aggregating" based on the visualisation of aggregates > 1 µm which could be easily distinguished from the background noise. **(c)** Phase diagram of (GGGGCC)$_n$ subjected to mG4-forming conditions where the DNA concentration is expressed in mass molarity. The phase diagram was obtained by imaging two biological duplicates at selected concentrations (**Appendix 10**).

3.3.2.2 (GGGGCC)$_n$ G4 topologies influence on aggregation

I then questioned whether microscopic aggregation was a consequence of a different G4 topology forming in the bulk solution. Therefore, several datapoints were selected around the threshold of aggregation in the phase diagram as shown in ***Figure 26a***, and CD was performed on each sample. The resulting spectra in ***Figure 26b*** show a prevalently parallel conformation for all samples, regardless of their aggregation extent. Therefore, it is unlikely that the G4-topology was playing a significant role in aggregation.

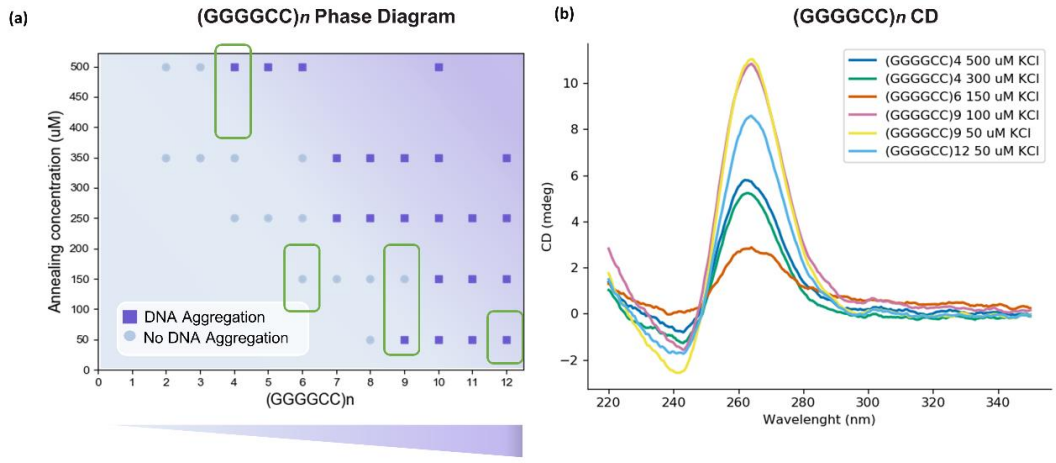


Figure 26: G4-topology effect on aggregation. (a) $(GGGGCC)_n$ phase diagram with highlighted concentrations taken for the CD experiment. The data points were selected below and above the aggregation threshold for several annealing concentrations to test for different topologies based on aggregation. **(b)** Circular Dichroism spectra of $(GGGGCC)_n$ (n = 4, 6, 9, 12). The DNA samples were annealed under mG4-annealing conditions at the reported concentrations (50 – 500 µM) and diluted to 5 µM for CD. All samples show prevalently a parallel conformation in the bulk, showing no significant difference between aggregating and non-aggregating samples. The CD spectra are acquired by the spectrometer as three technical replicates (accumulation = 3). No biological repeats were acquired as the spectra already constitutes a biological repeat of previously reported data.

3.3.2.3 $(GGGGCC)_n$ phase diagram high throughput protocol optimisation

One of the biggest limitations of this study involved repeatability of the experiments. In fact, small experimental errors such as slightly erroneous concentrations, an excessive degree of evaporation during annealing or altered salt amounts brought conflicting results. Even when evaporation was not significant, exact salt DNA sample concentrations remained constant, sample-to-sample variability remained significant. To address this concern, multiple repeats for each sample at each concentrations had to be tested. As each replicate of the experiment was time consuming and expensive, the mG4-formation protocol was automatized using the Opentrons Pipetting robot (OT2). Furthermore, aside from reducing waste of materials and time, using an automated protocol would additionally reduce the risk of contamination. Importantly, aggregation can be triggered by external

dust particles or other impurities that contaminate the samples prior to annealing, impacting the reliability of the results. Automation would minimize the chance of contamination by utilizing sterile disposable tips and incorporating measures to prevent sample carryover or aerosol formation.

Aware of the several advantages offered by an automatized annealing step, the protocol was coded to reproduce the phase diagram and test its repeatability. $(GGGGCC)_n$ (n = 1-12) sequences were diluted at selected annealing concentrations for each repeat length (0-500 µM) based on the previously obtained results (***Figure 25***). As in previous experiments, crowding agent PEG and KCl salt were added to the system to probe mG4s during annealing. To avoid excessive evaporation to affect the actual concentration in the system, the wells at the edge of the plate were left empty as shown in ***Figure 27***, and the plate was covered with tin foil and a heated lid during the annealing step.

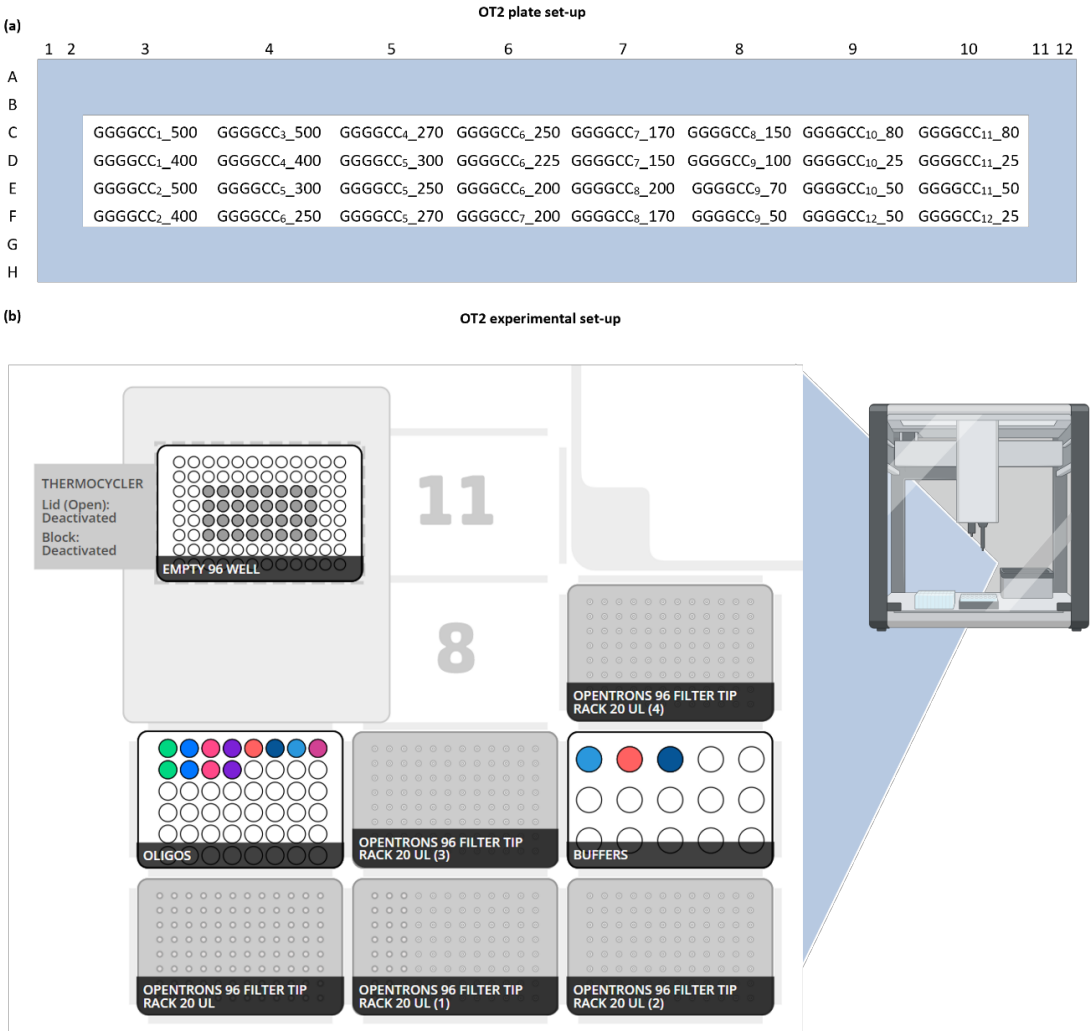


Figure 27: Opentrons pipetting robot $(GGGGCC)_n$ mG4-formation protocol optimisation. (a) Plate set-up. In blue are highlighted the wells left empty due to potential evaporation. **(b)** Experimental set up to automatize the development of a phase-diagram for the $(GGGGCC)_n$ system. The areas in blue on the plate set-up were left empty to avoid excessive evaporations of the wells. Each well in the plate contained 60 µL of solution.

Upon completed annealing, each sample was analysed *via* AGE to test for the presence of multimeric species at the concentrations tested. As shown in **Figure *28a***, most of the sequences behaved as expected, with a lower and a higher molecularity band appearing at higher concentrations. Some of the species, however, did not present any signal, indicating that no DNA was present in the wells. This likely occurred due to an error of the robot when transferring the oligomer solution out of the tube at the dilution stage. Possibly, the high viscosity of the stock solutions led to unprecise measurements during automatic pipetting. Although this issue could be addressed by manually cutting the pipette tips ends before loading them into the robot, this would nullify their sterile condition and introduce other potential errors into the experiment.

Optical microscopy images were then taken directly from the glass-bottomed plate to avoid transferring of the samples. The results are reported in the OT2 phase diagram in **Figure *28b*** and **Appendix *11***. Although a very similar trend was observed for the samples annealed *via* the OT2 protocol, there was a discrepancy between some of the data points (highlighted in red). This discrepancy likely came from inaccurate volumes of oligomer picked up by the robot due to the viscosity of the solutions. Furthermore, although the plate was sealed with tin foil, a significant amount of evaporation was observed for many of the samples (some of them even had up to 50% less of their initial volume) making measurements inaccurate.

In conclusion, although using the OT2 robot allowed for a faster production of replicates for the phase diagram, it was not accurate enough to be considered a reproducible method. For this reason, the following experiments were performed with the original manual protocol as, even though slower, it offered more reliable results.

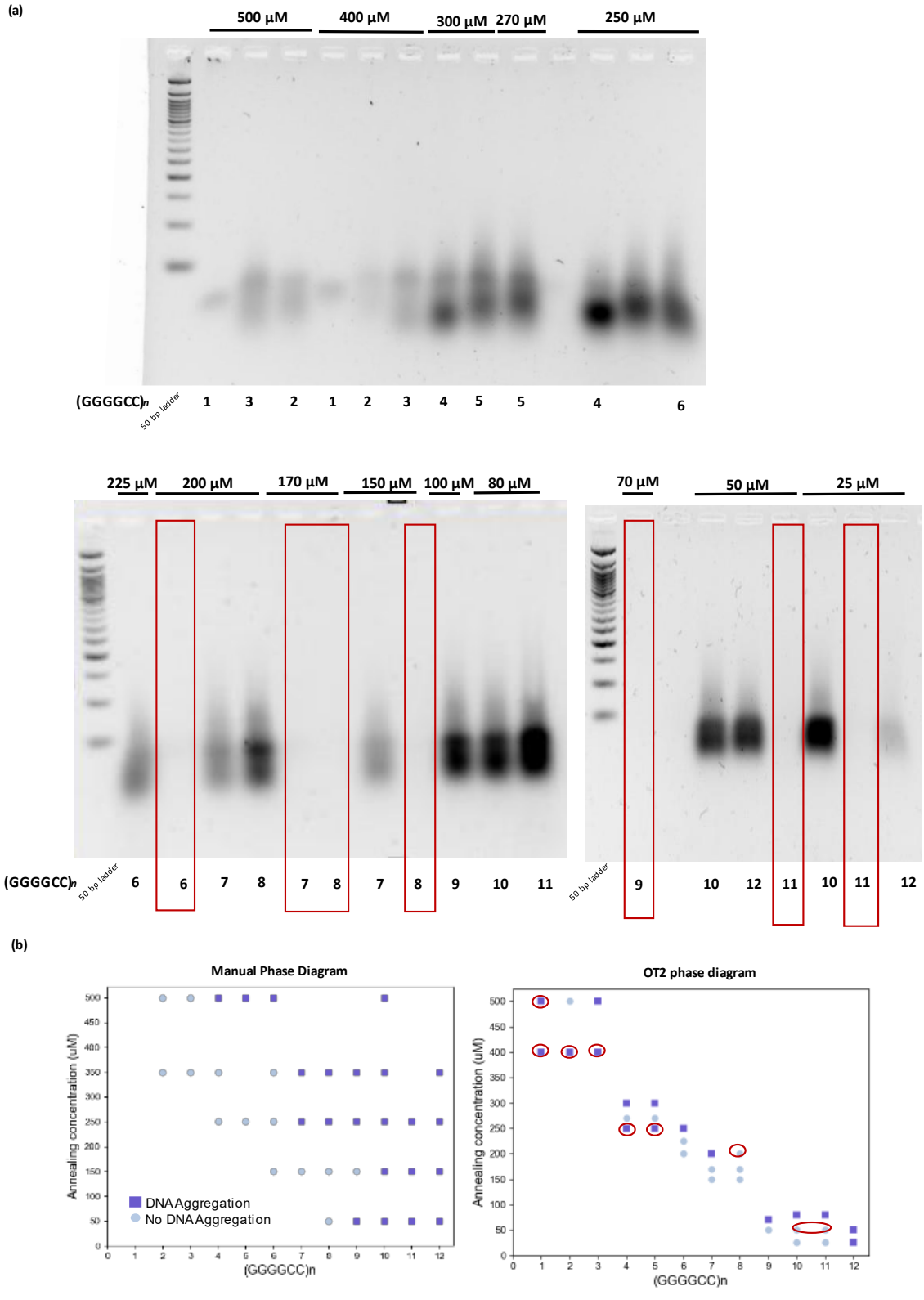


Figure 28: $(GGGGCC)_n$ aggregation analysed with the Opentrons protocol. (a) AGE of $(GGGGCC)_n$ annealed at different concentrations in mG4-forming conditions *via* the Opentrons high-throughput protocol. Highlighted in red are the samples that showed no DNA in the wells, most likely due to a mistake in extracting the liquid from the tubes by the pipetting robot.

The samples are diluted to 10 µM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE, on ice at 65 V for 80 mins and stained with SYBR Safe for imaging. **(b)** Comparison between the manually obtained phase diagram for $(GGGGCC)_n$ annealed under mG4-forming conditions (**on the left**) and the OT2 phase diagram (**on the right**). In red are highlighted the data points that vary from expectation. Microscopy images are in **Appendix 11**. Notably, the experiment was repeated once more and significant evaporation was observed in all the wells. The results are not reported as the Opentron protocol was subsequently not utilized.

3.3.3 $(GGGGCC)_n$ aggregates are formed via mG4 self-assembly

3.3.3.1 $(GGGGCC)_n$ aggregates characterisation

Once demonstrated that $(GGGGCC)_n$ DNA sequences could aggregate in a length and concentration dependent fashion, the aggregation process was further characterised to assess which structural species were present in the aggregates. In fact, although CD spectroscopy and AGE had demonstrated mG4s presence in the bulk solution upon annealing $(GGGGCC)_n$ under mG4-forming conditions, direct evidence of the presence of the G4s within the aggregates had not yet been provided.

Firstly, a fluorescent labelling experiment was set up to assess that the $(GGGGCC)_n$ strands were mixing homogeneously in the aggregates. This would **(1)** confirm that the denaturing step had allowed all strands to be in the ssDNA form prior to the formation of multimeric structures and **(2)** that the DNA strands were homogeneously distributed in the multimeric network which formed the aggregates.

$(GGGGCC)_{10}$ was annealed at 250 µM in mG4-forming conditions and subsequently diluted to 10 µM for easier imaging. 10% of the DNA strands were fluorescently labelled with Carboxyfluorescein (FAM), while the rest was non-tagged. As only 10% of the DNA strands were fluorescently labelled, homogeneous mixing of the components should lead to a homogeneously distributed fluorescent signal. As evident from the confocal image in ***Figure 29a***, fluorescent and non-fluorescent strands were homogeneously distributed, indicating complete mixing of the denatured strands.

Further characterisation of the amorphous aggregates involved understanding their physical state. To test this, fluorescent recovery after photobleaching (FRAP) was employed on the $(GGGGCC)_{10}$ (10% FAM labelled) annealed samples. In the FRAP experiment, a portion of the $(GGGGCC)_{10}$ aggregate was irradiated with a high power 488 laser until all the FAM fluorophore was bleached and no fluorescence emission was detected. As evident from ***Figure 29c***, although the fluorescence channel appeared dark in the region of interest (ROI), the brightfield remained unaffected. This implied that although the fluorescent tag had been successfully bleached, the integrity of the aggregate remained unaffected.

After 30 minutes, no recovery was observed, suggesting that the molecules within the aggregate was in a solid state. Interestingly, protein-RNA aggregates in neurodegenerative diseases (*e.g.* TDP-43 inclusions) have shown a prevalently "non-recovering" solid nature[189].

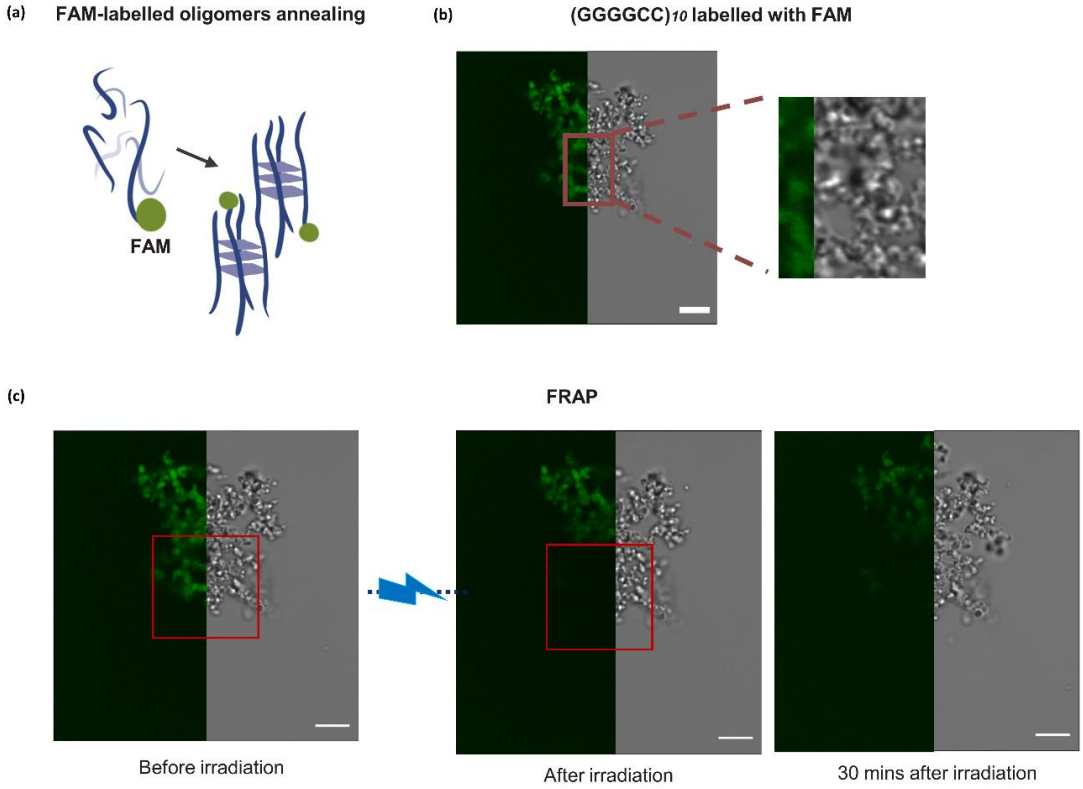


Figure 29: FAM-labelled $(GGGGCC)_n$ can aggregate in vitro. **(a)** Schematics of mG4s formation when a percentage of the oligonucleotides is fluorescently labelled. The green ball represents the Carboxyfluorescein label (FAM). **(b)** Confocal image of $(GGGGCC)_{10}$ (10% FAM labelled), annealed at 250 µM in mG4-forming conditions and diluted to 10 µM for imaging. The confocal image shows on the left the fluorescence channel and on the right the brightfield image. **(c)** Confocal images of $(GGGGCC)_{10}$, annealed at 250 µM (10% FAM labelled) in mG4-forming conditions and diluted to 10 µM. Upon laser excitation FAM labelled strands photobleached, leading to a local reduction of fluorescence intensity. The photobleached strands did not diffuse within the reset of the aggregate after 30 minutes, indicating a solid-like state. 10 µm scalebar. Both Carboxyfluorescein labelling and FRAP were performed in biological duplicates for a different repeat length (**Appendix 12**).

3.3.3.2 $(GGGGCC)_n$ aggregates are fluorescent upon staining with NMM

I then questioned if the $(GGGGCC)_n$ solid aggregates contained mG4s and whether G4-based interactions were key in holding the structure together. Indeed, whilst in the previous paragraphs it

was illustrated how the bulk solution was able to arrange into mG4s (CD and AGE results) and $(GGGGCC)_n$ was shown to aggregate *in vitro* (Phase Diagram), there was no direct correlation between the two sets of experiments was present. As the $(GGGGCC)_n$ sequence could also interact *via* G-C H-bonding, other non-G4-mediated interactions could also be the cause of multimeric networks and aggregation. Indeed, $(GGGGCC)_n$ has previously been reported to arrange into hairpin conformations[112].

The first element of direct proof of G4-involvement in aggregation came from NMM-staining experiments. As mentioned in earlier paragraphs, NMM is a G4-specific fluorescent dye. In solution, NMM self-stacks due to its high planarity leading to a fluorescence self-quenching effect. In presence of G4s, however, the NMM preferentially binds to the G4-tetrad and fluorescence increases. Earlier, I used NMM to stain AGE gels, proving that the multimeric species observed were indeed G4-containing. The same principle was applied to the aggregates *via* confocal microscopy to prove the presence of G4s within the microscopic structures.

$(GGGGCC)_6$ was annealed under mG4s forming conditions and observed under brightfield microscopy to assess the presence of the aggregates. Once ensured that the aggregates were present before addition of the NMM, the samples were diluted to 10 µM and NMM was added in a 1:1 ratio and incubated for 30 minutes. As shown in ***Figure 30a*****,** the aggregates resulted fluorescent on the NMM channel, indicating the presence of G4s. To ensure that NMM could not bind to other multimolecular structures given by canonical Watson-Crick base pairing, its selectivity was assessed by staining dsDNA condensates. The dsDNA condensates were achieved by annealing a combination of ssDNA strands (**Table *1***) in glass capillaries to probe the formation of artificial microscopic dsDNA constructs. As evidence from ***Figure 30b***, the condensates were not able to bind to NMM in the same order of magnitude as the $(GGGGCC)_6$ aggregates, proving selectivity of the dye. Importantly, NMM staining does not *per se* imply a structural role of G4s, but simply demonstrates their presence within the aggregates, for which further experiments need to be conducted.

Name	Sequence (5'-3')
N4_ 1	**GATCGC**CGCCGCAATCACGCGCGTGCTCGGCGCCAGCAGTCCTGGCG
N4_ 2	**GATCGC**CGCCAGGACTGCTGGCGCCGTCGCTTCTCTTCATAACAACG
N4_3	**GATCGC**CGTTGTTATGAAGAGAAGCGTCGCTCTGGCACAGGTGTACG
N4_4	**GATCGC**CGTACACCTGTGCCAGAGCGTGCACGCGCGTGATTGCGGCG
L_1	**GCGATC**CGCAAACCAGCAAGCTCACG
L_2	**GCGATC**CGTGAGCTTGCTGGTTTGCG

Table 5: ssDNA sequences used for the assembly of the dsDNA control. Further information on the annealing in the Methods section.

3.3.3.3 $(GGGGCC)_n$ aggregates disassemble upon NMM-photooxidation

Aside from selective fluorescent staining of G4s, NMM can also be used as a selective G4-disassembly probe in specific light conditions. Indeed, 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[190]. This phenomenon has been previously observed to not occur for dsDNA[190], making it a highly G4-specific disassembly route (***Figure 30c***). In fact, this property was previously leveraged to achieve light-controlled disassembly of mG4s in engineered DNA nanostructures[191].

To assess whether G4s played a structural role within the $(GGGGCC)_6$ aggregates, NMM photo-oxidation was performed. $(GGGGCC)_6$ (250 µM) were subjected to mG4-annealing conditions and stained with NMM as previously described. Subsequently, a portion the aggregates was irradiated to trigger photo-oxidation and disassembly of the aggregates was immediately observed (***Figure 30c***). 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 could be observed. The selectivity of the assay was tested on NMM-stained dsDNA condensates, used as a negative control, which did not show any significant change upon irradiation (***Figure 30c***).

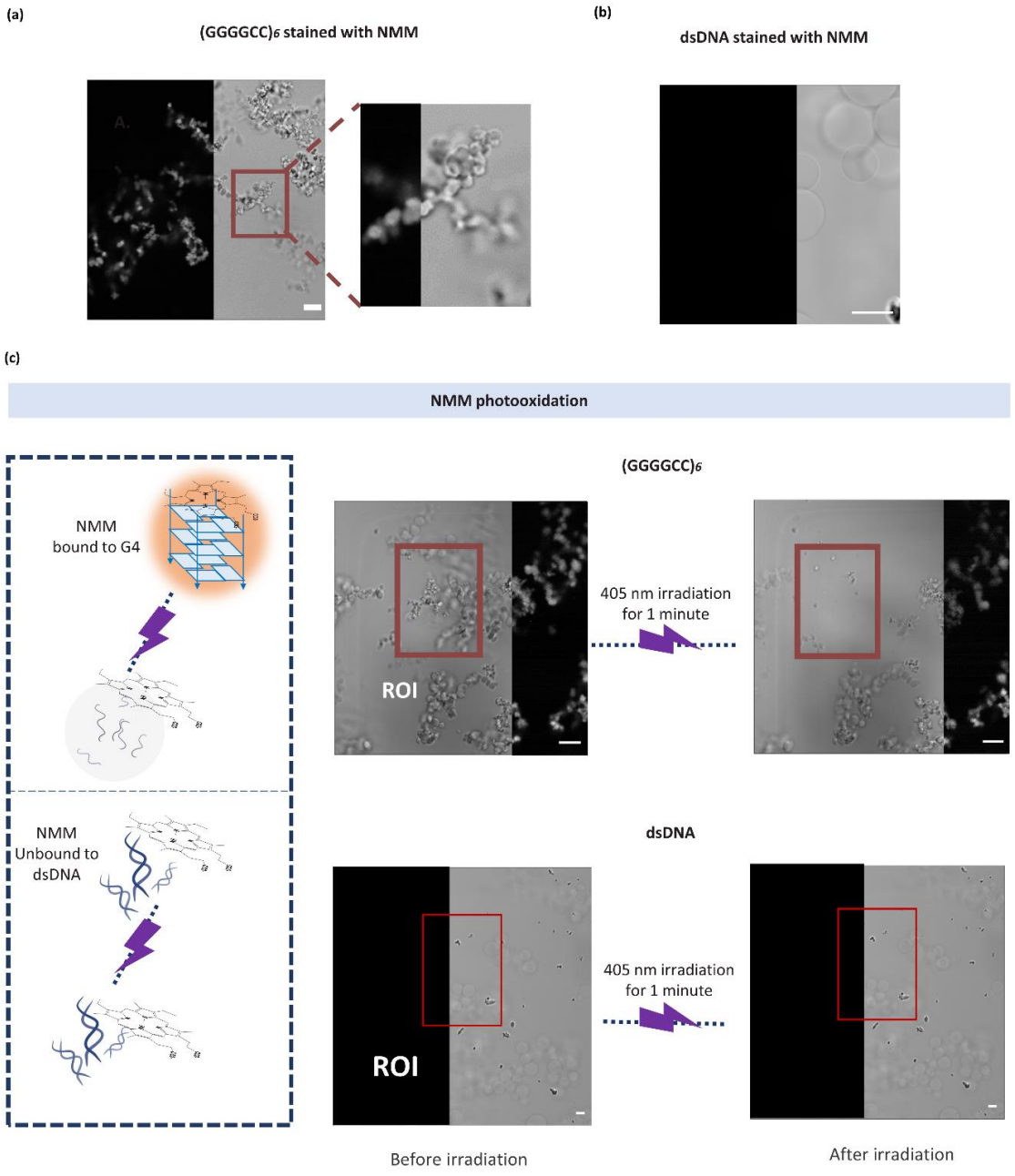


Figure 30: $(GGGGCC)_6$ aggregates in presence of NMM. (a) $(GGGGCC)_6$ (250 µM) was annealed under mG4s forming conditions and stained with NMM. The image is split to show the brightfield channel (on the right) and the NMM fluorescence channel (on the left). **(b)** Confocal microscopy imaging of dsDNA condensates in the brightfield and fluorescent channel. No NMM fluorescence is observed in presence of dsDNA. **(c)** NMM photooxidation experiment for mG4-containing aggregates and dsDNA condensates. Whilst the mG4-based aggregates (top panel) disassemble upon irradiation, the dsDNA condensates are unaffected. NMM-photooxidation of $(GGGGCC)_6$ was performed in biological triplicates (**Appendix 13**).

3.3.3.4 $(GGGGCC)_n$ aggregates present G4-like autofluorescence properties

Although NMM-photooxidation demonstrated the presence of G4s in the aggregates, I sought confirmation of the phenomenon with a different experimental set up which did not require the presence of external fluorescent probes: testing the aggregates for G4-autofluorescent abilities.

Autofluorescence refers to the intrinsic fluorescence emitted by endogenous molecules in biological samples without the addition of exogenous fluorophores[192]. It is a phenomenon observed for several biomolecules and biological tissues, in particular planar protein residues in proteins such as tryptophan[193]. As this property is related to the protein residues π-stacking interactions, G4s, being composed of planar structures stacked at the top of each other's, have also been investigated as auto fluorescent biomolecules. Indeed, it has been reported that parallel G4s have a detectable 3.5 fold higher fluorescence than ssDNA and dsDNA[194]. Although the G4-autofluorescence study was performed *via* plate reader on monomeric G4s, I hypothesised that mG4s would present a similar behaviour.

To test this hypothesis, $(GGGGCC)_{11}$ was annealed at 350 µM in mG4-forming conditions and diluted to 10 µM for confocal imaging. In the previously mentioned study, autofluorescence of nucleic acid species was observed between 300-500 nm emission by exciting the sample at 260 nm and 300 nm[194]. Due to the limitations of confocal microscopy, the $(GGGGCC)_n$ test for autofluorescence in the aggregates was set up with slight modifications. Excitation was performed at 405 nm and emission observed was between 430 and 700 nm. As estimated from ***Figure 31a***, these wavelengths should have still been enough to observe a significant difference with the other DNA species. Indeed, as shown in ***Figure 31b***, whilst $(GGGGCC)_{11}$ aggregates presented significant autofluorescence behaviours, the dsDNA control did not have any signal on the same channel.

This experiment not only offered a further confirmation that G4 species were present into $(GGGGCC)_n$ aggregates, but also suggested the possibility of detecting G4s in absence of external probes. Indeed, whilst there are several small molecule probes developed to detect G4s, it is not clear whether they were stabilising the structures and introducing bias into the results[131]. This method proved that G4s were present into the aggregates without the risk of triggering their formation.

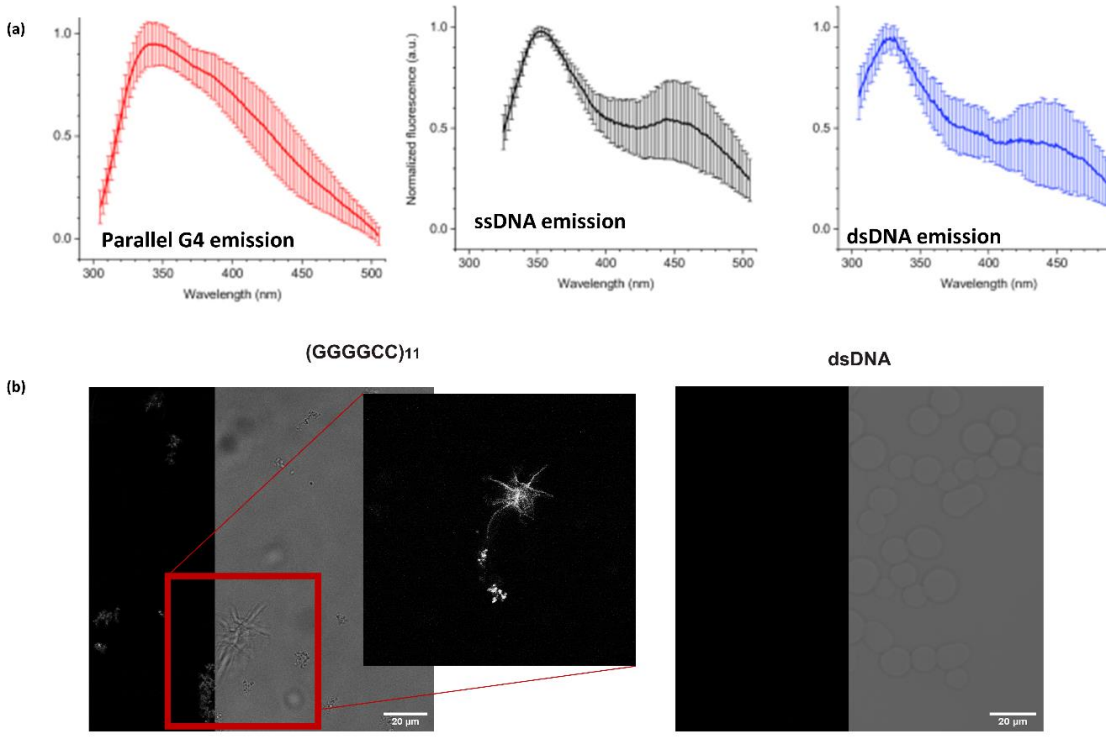


Figure 31: $(GGGGCC)_n$ aggregates possess autofluorescence abilities. (a) G4, ssDNA and dsDNA autofluorescence emission spectra adapted from ref[194]. **(b)** $(GGGGCC)_{11}$ autofluorescence behaviour observed by irradiating the sample at 405 nm and observing the emission from 430 to 700 nm. The same settings used on dsDNA constructs yield no fluorescence. Experiment was performed as three biological replicates (**Appendix 14**).

3.4.3.5 $(GGGGCC)_n$ aggregates present mG4-like melting behaviours upon heating

As mentioned earlier, mG4s are the thermodynamic product of the annealing process of $(GGGGCC)_n$ sequences. The extreme stability of the thermodynamic product has been observed in previous literature[195], as well as *via* my CD melting curves experiments in ***Figure 22***. In both instances, mG4s failed to fully melt even at the highest annealing temperature of 95℃, above which the aqueous buffer would have evaporated.

On the assumption that the $(GGGGCC)_n$ aggregates are held together by mG4s, they should present a high thermal stability. Therefore, $(GGGGCC)_{11}$ was annealed at 80 µM in mG4 forming conditions in glass capillaries and observed under a heating stage on an optical microscope. Glass capillaries were chosen for the experimental set-up as they could be easily sealed with glue and would prevent any significant evaporation upon heating. As shown in **Figure *32***, the aggregates were heated from 25 ℃ to 95 ℃ with a temperature ramp of 10 ℃ min^{-1}. The sample was kept at the highest temperature for

3 hours, but significant evaporation was observed after 20 minutes. At the 15 minutes mark – prior to evaporation - no melting of the aggregates was observed. The melting experiment results confirmed that mG4s were the structure holding the aggregates together as **(1)** no other DNA multimeric secondary structure has been observed to have such a high stability (hairpins melt between 30-60 °C[196], triplex DNA melts between 10-90 °C[197]) and **(2)** they provided a correlation between the biophysical observations of the $(GGGGCC)_{11}$ sequences inability to melt completely under CD and the nature of the aggregates.

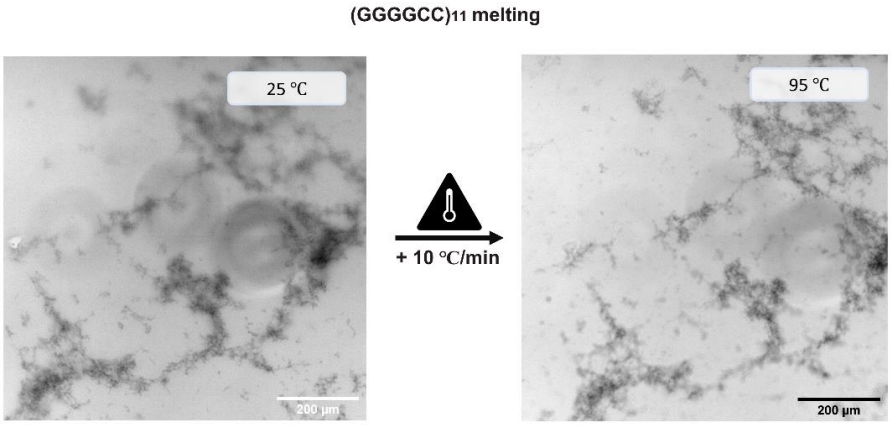


Figure 32: $(GGGGCC)_{11}$ aggregates melting experiment. Optical microscopy of $(GGGGCC)_{11}$ annealed at 80 µM in mG4-forming condition in glass capillaries. A first picture was obtained at 25°C showing the sample aggregating as previously described. By then subjecting the sample to a temperature ramp of 10 °C/min up to 95°C (second image), no significant melting was observed. Experiment was performed in one biological replicate.

3.4 Conclusions

The biological relevance of multimolecular G4s is still a widely unexplored topic. In this chapter, I have highlighted how $(GGGGCC)_n$ is not only able to arrange into mG4s conformations *in vitro*, but that the formation of these structures is able to trigger aggregation in absence of additional proteins. This is not only highlighting the potential for biological relevance of mG4s, but also puts into question the recurrent narrative that puts protein aggregation as the main pathological hallmark of neurodegeneration in ALS/FTD.

These statements are supported by substantial experimental evidence, proving that mG4s are the leading structure generating the aggregates. Firstly, the optimal *in vitro* system aimed at simulating cell crowding conditions was optimised for $(GGGGCC)_n$. This optimisation involved ensuring that any pre-formed G4-structures in the stock solutions were fully disassembled before proceeding in defining an mG4-specific annealing protocol. Interestingly, no literature has ever reported the need for a denaturing step before working with $(GGGGCC)_n$, despite using buffer systems which would have stabilised pre-formed G4s, which might affect the validity of some observations published in this space. Upon ensuring complete denaturation of the sequences, several mG4s-forming protocols were tested with the aim to find the most efficient and effective one in forming mG4s. Not surprisingly, the resulting protocol involved using a crowding agent and a slow cooling step, both methods which are employed to aid the system in crossing the activation energy barrier towards the thermodynamic product. The successful formation of the mG4s was assessed in the bulk *via* a series of biophysical assays. CD spectroscopy showed the typical G4-parallel trace, AGE stained with SYBR Safe allowed for the visualisation of higher molecularity bands and NMM-staining of those same gels ensured that those higher molecularity species were due to the presence of G4s. As collateral evidence, a KCl titration was also used to confirm the higher molecularity band's dependence on G4-favouring conditions. Once assessed that mG4s were forming in solutions, their ability to aggregate was confirmed *via* microscopy. $(GGGGCC)_n$ was able to aggregate in absence of protein with a length dependence, interestingly presenting a similar trend to the ALS/FTD pathology itself. The aggregates were amorphous and solid in nature, as reported from FRAP experiments, once again drawing a parallel with other biomolecular condensates found in cells. The presence of G4s in the aggregates was assessed *via* NMM staining, which succeeded in showing fluorescence upon binding of the aggregates to an external probe, and *via* G4-autofluorescence testing. Furthermore, I aimed to confirm that G4s were the prevalent structure holding the aggregates together, for which I performed NMM-photooxidation and melting experiments, which were both successful in asserting the presence of G4-structures.

In conclusion, these results bring forward a new potential mechanism of action by which $(GGGGCC)_n$ could lead to pathological aggregation in ALS/FTD patients, one that does not rely on protein aggregation.

This study, however, still leaves some unaddressed questions. Firstly, all the biophysical studies were performed on DNA, whilst RNA tends to be the more biologically relevant species. This was done on the assumption that aggregation was a consequence of a structural effect, and the nature of the nucleic acid should not impact this, which is something that needed to be validated. Secondly, our *in vitro* system, although useful for understanding the basic mechanisms, may not fully represent the physiological status and mechanism occurring in living cells. Our DNA annealing concentrations are in the micromolar region, and our buffer conditions have very high KCl concentrations. Such extreme conditions had to be used to trigger aggregation in shorter repeat length (n = 1-12) as longer repeat lengths – which are more biologically relevant – could not be obtained. In fact, the pathological aggregates are only observed in patients with $n > 24$. It is likely that longer repeat length would require lower concentrations (which is a conclusion which could also be extrapolated from the phase diagram) and lower KCl amounts. Furthermore, once proven that the system is in fact able to aggregate, the question arises on how to prevent this phenomenon for future therapeutics applications.

These questions will be addressed in the following chapters of this thesis.

3.5 Materials and Methods

All DNA and RNA sequences were purchased via the external provider Integrated DNA Technologies (IDT). IDT provides oligonucleotide quality control via free electrospray ionization (ESI) mass spectrometry analysis that is provided as standard with every oligo. HPLC (ensured >85% purity) and standard desalting purifications were both initially tested to assess the effect of purity on the mG4-forming experiments. As both showed equal behaviours at high repeat lengths, standard desalting was chosen for purification in all experiments as cheaper and led to higher yields. Although it is possible that smaller truncated products were present in the samples at higher repeat lengths, all oligos were tested via agarose gel prior to aggregation studies to ensure their behaviour was repeatable.

3.5.1 Optimisation of the annealing conditions

3.5.1.1 Optimisation of different denaturing steps

$(GGGGCC)_{11}$ DNA sequence was purchased from Integrated DNA Technologies (IDT) in lyophilised form and dissolved in MilliQ water (final concentration of 1 mM). Samples were then diluted to the annealing concentration of 0.1 mM in 10 mM TRIS-HCl (pH 7.4) and heated according to the protocols illustrated in **Figure *18***: **(1)** Incubation for 4 hours at room temperature; **(2)** Incubation for 3 h at 95℃ followed by cooling at room temperature (around 25 ℃) for 1 h and **(3)** Incubation for 1 h at 95℃ and a slow cooling down process with a temperature ramp of (-0.25 °C min^{-1}) (all conditions are summarised in). 8% of the samples presented a Carboxyfluorescein (FAM) tag at the 5' end. The controls samples were heated in presence of 500 mM KCl, with the aim of preventing the denaturing. Heating was performed in the Thermal cycler C1000™ with heated lid to prevent evaporation.

3.5.1.2 Optimised denaturing step

$(GGGGCC)_n$ (n=1-12) DNA sequences were purchased from Integrated DNA Technologies (IDT) in lyophilised form and dissolved in MilliQ water (final concentration of 1 mM). Samples were then diluted to the annealing concentration of 0.1 mM in 10 mM TRIS-HCl (pH 7.4), heated to 95℃ for 1 h and slowly cooled to room temperature with a cooling rate of -0.25 ℃/min. Heating was performed in the Thermal cycler C1000™ with heated lid to prevent evaporation.

3.5.1.3 Optimisation of mG4s formation protocol

$(GGGGCC)_{11}$ DNA sequence was purchased from Integrated DNA Technologies (IDT) in lyophilised form and dissolved in MilliQ water (final concentration of 1 mM). Samples were then diluted to the annealing concentration of 0.1 mM in 10 mM TRIS-HCl (pH 7.4) and denatured as of the optimised denaturing step in duplicates. 500 mM KCl was then added to half of the samples whilst buffer was added to the rest as a control. The samples were then subjected to different formation protocols as

listed in **Figure *20*** or as follows: **(1)** isothermal assembly (25°C, 36°C, 45°C, 60°C) upon addition of 500 mM KCl; **(2)** annealing process with cooling step of 0.25 °C min^{-1}; **(3)** annealing process with cooling step of 0.25 °C min^{-1} in presence of crowding agent PEG; **(4)** annealing process with cooling step of 0.01 °C min^{-1} in presence of crowding agent PEG; **(5)** annealing process with cooling step of 0.02 °C min^{-1} in presence of crowding agent PEG. Heating was performed in the Thermal cycler C1000™ with heated lid to prevent evaporation.

3.5.1.4 Optimised mG4s formation protocol

$(GGGGCC)_{11}$ DNA sequence was purchased from Integrated DNA Technologies (IDT) in lyophilised form and dissolved in MilliQ water (final concentration of 1 mM). Samples were then diluted to the annealing concentration of 0.1 mM in 10 mM TRIS-HCl (pH 7.4) and 30% PEG (BioUltra 200) and denatured as in the optimised denaturing step. 500 mM KCl was then added the samples, which were subsequently heated to 95℃ for 2h and cooled down to room temperature with a cooling rate of –0.02 ℃/min. Heating was performed in the Thermal cycler C1000™ with heated lid to prevent evaporation.

3.5.2 Optical Microscopy

Pre-annealed samples were diluted down to 10 µM in the corresponding buffer and transferred into PDMS wells for microscopy. Micrographs were acquired by using Nikon Ti2 microscope with a 20X0.5 Nikon Plan Fluor dry objective and 40X0.75 Nikon Plan Fluor dry objectives. The images were analysed in Nikon Element Viewer. Specific details of the experiments are provided in the subsequent paragraphs.

3.5.2.1 FAM-labelled samples

DNA samples with a FAM label were annealed as per their specific applications and visualised by utilising the GFP channel pre-sets.

3.5.2.1 $(GGGGCC)_{11}$ *melting experiment*

The annealing sample solution was prepared according to standard mG4-optimised protocol. A volume of 70 µL of the sample 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 Hellmanex 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 their ends sealed with epoxy glue (Araldite) on a coverslip. Once annealed as according

to protocol, the sample was transferred onto a heating stage on the Nikon Optical microscope. There, the aggregates were heated from 25 to 95℃ with a temperature ramp of 10 ℃/min.

3.5.3 Agarose gel electrophoresis (AGE)

3.5.3.1 SYBR Safe staining

The agarose gel matrix was prepared by mixing 1.5 g of agarose in 50 µL 1X TBE buffer and 3 µL of SYBR Safe stain, heated by microwaving and left to polymerise for 20 minutes prior to running. Samples were diluted to 10 µM and loaded in the gel matrix with 2 µL 50% glycerol. The gel was run at 50 V for 2.30 h at room temperature or at 65 V for 85 mins on ice (depending on application) in 1X TBE and imaged on the SynGene Gel Reader. Image analysis was performed on ImageJ (Fiji).

3.5.3.2 NMM staining

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. Samples were diluted to 10 µM and loaded in the gel matrix with 2 µL 50% glycerol. The gel was run at 50 V for 2.30h at room temperature or at 65 V for 85 mins on ice (depending on application) in 1X TBE. The gel was then incubated in a solution of 50 µL 2mM NMM and 50 mL 1X TBE for 1h and 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. Image analysis was performed on ImageJ (Fiji). Notably, the following non-G4 sequence was used as control: 5' TTTCACCGACGACACCAACCACGGCACCTATTATTTATTTATAATTAT 3'.

3.5.4 Circular Dichroism of DNA samples

3.5.4.1 Wavelength scan at room temperature

Each DNA sample was diluted to a concentration of 5-10 µM in the buffer of choice for a final volume of 150 µL. The sample was then transferred into a sealed 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). The data was collected as txt files and analysed in Jupiter Notebook (Python 3) with the matplotlib library.

3.5.4.2 CD melting curves

Pre-annealed DNA samples were diluted to the desired concentration (5-20 µM) in TRIS HCl buffer (10 mM pH 7.5 500 mM KCl). Control samples did not contain KCl. Circular Dichroism was performed at 25°C recorded on a JASCO J-810 circular dichroism spectrophotometer using a 1 mm path length

quartz cuvette. Melting curves were obtained by adding a 1°C min^{-1} temperature ramp to the sample with measurements every 0.2°C at 262 nm and a full spectrum every 5°C. The data was collected as txt files and analysed in Jupiter Notebook (Python 3) with the matplotlib library (curve fit).

3.5.5 Confocal microscopy

3.5.5.1 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.

3.5.5.2 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 a 510-550 nm emission filter band.

3.5.5.3 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. (10 µM) NMM was then added to the sample and left to incubate for 10 minutes in a dark environment. The sample was then imaged with a 600-560 nm emission filter band.

3.5.5.4 Fluorescence Recovery After Photobleaching (FRAP)

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 a 600-560 nm emission filter band. 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.

3.5.5.5 NMM photo-oxidation

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 with a 600-560 nm emission filter band.

3.5.5.6 G4 auto-fluorescence

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. Emission was observed in the 430-700 nm emission channel.

3.5.6 dsDNA condensates controls

Condensates made from dsDNA strands were prepared by mixing four 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 a final core concentration of 1 µM and linkers concentration of 2 µM. A volume of 70 µL 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 ***Table 5***.

Chapter 4

The molecularity of the $(GGGGCC)_n$ system affects its aggregation abilities

4.1 Background and Aims

In the previous chapter I showed how arranging $(GGGGCC)_n$ into multimolecular G4 (mG4s) allowed the sequence to aggregate in absence of proteins. Several G4 specific assays were performed to assess that mG4s were the prevalent species generating the aggregates, without excluding that other intermolecular forces (such as canonical Watson-Crick base pairing) could also be playing a role in the aggregation.

In this chapter, I aim to disentangle the different intermolecular forces which can give rise to the presence of the aggregates in the $(GGGGCC)_n$ system. This has two main aims: **(1)** further characterising the aggregation mechanism which might have a biological relevance in ALS/FTD and **(2)** understanding how the changing the molecularity of the system could be used as a tool to prevent or disrupt the aggregates themselves.

4.2 Experimental Techniques

4.2.1 Mutation Analysis

Mutation analysis is a technique used to study and analyse genetic variations, or mutations, in DNA sequences. It involves the examination of changes in the nucleotide sequence of a gene or a specific region of the genome, which can provide insights into the functional and structural implications of these alterations[198].

In the $(GGGGCC)_n$ system, mutation analysis was used to investigate the G-C interaction contribution (typical of dsDNA) to mG4-driven aggregation. In fact, whilst G4s rely on G-G interactions, in dsDNA G interacts selectively with C according to canonical Watson-Crick base pairing (**Figure *33a***)[131]. The effects of mutating the wild type sequence were analysed by CD spectroscopy, AGE and optical microscopy.

4.2.2 PDS treatment

Several G4 ligands have been developed through the past fifty years, all sharing the same main characteristics, namely a planar backbone which can interact with the G-tetrad and some positively charged residues which form electrostatic interactions with the negatively charged phosphate backbone[199]. As previous studies had been performed on the effect of G4-ligands on the ALS/FTD phenotype[200], I selected the well-known G4-ligand Pyridostatin (PDS) (**Figure *36a***)[199], which has been used in *in vivo* assays and present a high selectivity over dsDNA, to test its effects on $(GGGGCC)_n$ aggregation. The effects of PDS on the molecularity of the $(GGGGCC)_n$ system was assessed *via* CD spectroscopy, AGE and optical microscopy.

4.2.3 Antisense treatment

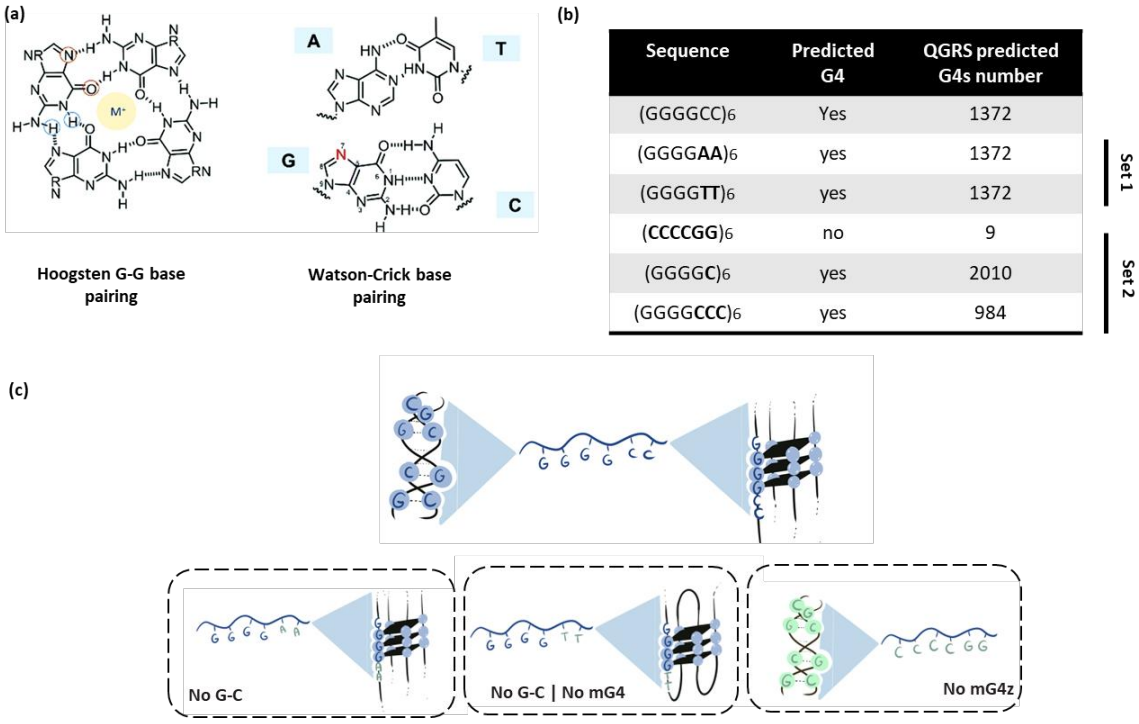
Antisense oligonucleotides are synthetic strands of nucleic acids that are designed to specifically target and bind to complementary RNA/DNA sequences, leading to a range of effects such as gene silencing, splicing modulation, inhibition of RNA translation or – in the case of addition to the mG4 $(GGGGCC)_n$ system – can be used to achieve competitive formation of dsDNA over other non-canonical structures[201]. The effects of the antisense strand on the molecularity of the $(GGGGCC)_n$ system was tested *via* CD spectroscopy, AGE and optical microscopy.

4.3 Results and Discussion

4.3.1 $(GGGGCC)_n$ mutation analysis

Previous experiments demonstrated that the primary forces holding together the structures were mG4s but did not exclude a G-C contribution. Therefore, two sets of mutant sequences (n = 6) were developed, one where the C in $(GGGGCC)_n$ were substituted with different bases (Adenine and Thymine) and one where the G/C ratio and loop-length in $(GGGGCC)_n$ was modified (**Figure *33b***).

In order to understand which other intermolecular forces played a role in the aggregation of the $(GGGGCC)_n$ system, I designed a series of mutated hexanucleotide repeat expansions with the aim to disentangle the individual bases contributions (***Figure 33b***).



Sequence	Predicted G4	QGRS predicted G4s number	
$(GGGGCC)_6$	Yes	1372	
$(GGGG\mathbf{AA})_6$	yes	1372	Set 1
$(GGGG\mathbf{TT})_6$	yes	1372	
$(\mathbf{CCCCGG})_6$	no	9	
$(GGGG\mathbf{C})_6$	yes	2010	Set 2
$(GGGG\mathbf{CCC})_6$	yes	984	

Figure 33: $(GGGGCC)_n$ mutation analysis. (a) H-bonding interactions between bases in nucleic acids. On the left, the Hoogsten H-bonding in G4s between guanines with highlighted interacting H-bond donors and acceptors. On the left, canonical Watson-Crick interactions between A-T and G-C. Adapted from [131]. **(b)** Table of sequences used for the mutation studies. **(c)** Potential conformations assumed by the sequences in set 1. The A-mutant can only assume G4 conformations, the T-mutant is disfavoured from forming higher molecularity species and the complementary sequence cannot form G4s. Hand-drawn illustration with Procreate.

4.3.1.1 $(GGGGCC)_n$ mutation analysis – substituting C with A or T

With the hypothesis that G-C interactions did contribute to the aggregation of the system, two mutant sequences $(GGGGAA)_6$ and $(GGGGTT)_6$ were rationally designed. Both sequences were analysed by QGRS (a G4-predictor) and showed the same number of putative G4s as the wild type $(GGGGCC)_6$, indicating that their ability to form G4s (at least unimolecular G4s) should not be impacted by the C-substitution[202]. However, given the lack of C, no canonical base pairing interaction can take place in either instance. Interestingly, whilst $(GGGGAA)_6$ is expected to form multimeric species to the same extent as the original $(GGGGCC)_6$, the presence of thymine in the G4-loops has previously been shown to favour mono-molecular G4s over multimolecular ones[203]. Therefore, T-substitution was designed to test whether unimolecular G4s could also lead to aggregation.

Therefore, $(GGGGCC)_6$, $(GGGGAA)_6$ and $(GGGGTT)_6$ were all annealed at 250 µM under mG4-forming conditions to observe differences in their aggregation behaviours. Following annealing, the samples were diluted to 5 µM and analysed by CD spectroscopy to assess the formation of G4-structures. As shown in **Figure *34a***, the CD spectra from all expansions presented peaks that are typical of G4s, including a positive peak at 263 nm and 295 nm and a negative peak at 240 nm. Importantly, the species observed differed in the exact topology ratios (*e.g.,* antiparallel vs parallel), indicating that the lack of C could modify the overall topology of the system. The mutant's ability to form multimeric structures was assessed by AGE. The pre-annealed samples were diluted to the running concentration of 10 µM and run in two sets of conditions. A first gel was stained using SYBR Safe, a general DNA stain, whilst the second gel was incubated with NMM, a G4-specific binding dye. As shown in **Figure *34b***, all the tested sequences were able to form higher molecularity species, which are represented by the top band in the SYBR Safe stained gel. This band could be ascribed to be G4-containing given the fluorescence emission observed upon incubation with NMM. As the mutant sequences $(GGGGAA)_6$ and $(GGGGTT)_6$ could not form any canonical base pairing interactions, this experiment confirmed that the higher molecularity species were primarily caused by Hoogsten's interactions between the guanines in the sequences. Interestingly, the intensity of the top band in **Figure *34b*** was not the same for all the mutant sequences. Indeed, $(GGGGAA)_6$ seemed to be significantly more prone to form multimeric species, followed by the wild type $(GGGGCC)_6$ and lastly $(GGGGTT)_6$. This result agreed with the nature of the bases interactions. In fact, $(GGGGTT)_6$ presents thymine loops which promote unimolecular G4s formation. The wild type sequence $(GGGGCC)_6$ is prone to Watson-Crick G-C interactions which might be acting in competition with the formation of multimeric species, leading to a less intense higher molecularity band. Finally, $(GGGGAA)_6$ can only form G4s, and it is not disfavoured in forming higher molecularity species, leading to a more favourable formation of mG4s.

Once the sequences' ability to form mG4s had been confirmed, microscopy was utilised to assess their aggregation ability. If the aggregates in $(GGGGCC)_n$ were predominantly due to the presence of mG4s, the expectation for the mutants was to see aggregation only for those sequences which had shown a high mG4-formation ability *via* CD and AGE. Upon annealing the sequences at 350 µM no significant aggregation was detected for $(GGGGTT)_6$, which predominantly forms unimolecular G4s, while $(GGGGAA)_6$ and $(GGGGCC)_6$, which have shown to be able to arrange into mG4s – albeit to different extents - revealed significant aggregation (**Figure *34c***). Interestingly, this result confirmed that it was not the presence of G4s *per se* that led to aggregation, but it was specifically multimeric G4s which allowed the phenomenon to occur.

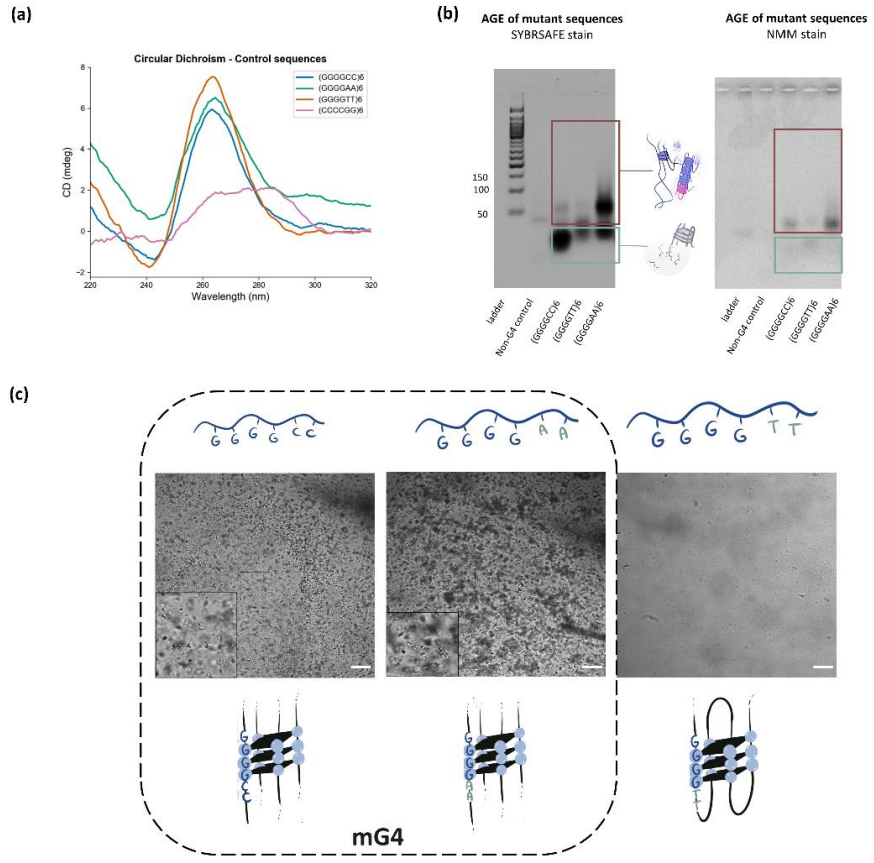


Figure 34: $(GGGGAA)_6$ and $(GGGGTT)_6$ mG4-formation and aggregation properties. (a) CD spectra of $(GGGGXX)_6$ (AA or TT) rationally designed mutants. All sequences were annealed at 250 µM and diluted to 5 µM to run the spectra at room temperature. All structures show the typical peaks of G4-species. CD spectra were acquired as three technical replicates by selecting accumulation = 3 on the spectrometer. The experiment was repeated at different concentrations and repeat length to ensure repeatability of the parallel conformation of $(GGGGAA)_n$ and $(GGGGTT)_n$ (**Appendix 15**). **(b)** AGE of the mutant sequences. The top band is more prominent in $(GGGGAA)_6$, indicating its tendency to arrange into a multimolecular species, but it is also present for the wild type $(GGGGCC)_6$ and $(GGGGTT)_6$. By staining the AGE with NMM, it is possible to notice how the top band has a high G4 composition. The samples are annealed at 250 µM and diluted to 10 µM for AGE. The gel running

conditions are the following: 3% agarose in 1X TBE, on ice at 65 V for 80 mins and stained with SYBR Safe or NMM for imaging. SYBR Safe-stained AGE gels were repeated in biological duplicates for different repeat lengths and concentrations for both mutants. (**Appendix 16**). **(c)** Brightfield imaging of the mutant sequences 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. Aggregation experiment was performed in biological duplicates (**Appendix 17**).

4.3.1.2 $(GGGGCC)_n$ mutation analysis – changing G/C ratio and loop length

The results discussed in the previous paragraph highlighted how mG4s are the necessary species for aggregation to occur and proved that G-C interaction in the wild type $(GGGGCC)_6$ reduced its ability to form higher molecularity species compared to the mutant $(GGGGAA)_6$. This led to design of the following set of mutants, where the G/C ratio was changed to reduce and/or increase the sequence's ability to form canonical base pairing.

The sequences designed are shown in the set 2 of ***Figure 33*** and were subjected to a similar analysis as to what performed for the other mutants. Firstly, the sequence's ability to form G4s was assessed by using the QGRS mapper[202]. The new sequences were unsurprisingly predicted to be able to arrange into G4s. However, the number of putative G4s was dependant on the G/C ratio: the higher the number of cytosines in the sequences, the less G4s were predicted to form. This is possibly due to a combination of factors, including the lower number of G-G conformations that the sequences can assume and the competing G-C canonical G-C interactions taking place.

The sequences were then annealed at 250 µM under mG4-forming conditions to observe differences in their aggregation behaviours. CD was performed to test for their ability to form G4s *in vitro* under our experimental conditions, confirming the QGRS predictions. In fact, as shown in ***Figure 35a***, both wild type and $(GGGGC)_6$ showed the typical peaks of G4s, while the other sequences, which had a lower likelihood of forming G4s due to the high C-content, contained a mixture of species. To further characterise the species obtained, AGE was performed on the annealing products. As shown in ***Figure 35b***, while the wild type $(GGGGCC)_6$ sequence behaved as predicted showing a top and lower molecularity band (respectively mG4s and lower molecularity species), the rest of the sequences primarily remained into a lower molecularity population. The only exception was $(GGGGC)_6$, which showed some smearing going towards higher molecularity species. The gel was then stained with NMM to assess the presence of G4s, and fluorescence was observed with $(GGGGCC)_6$ and $(GGGGC)_6$ – the sequences with the higher number of predicted G4s *via* QGRS and CD.

These results showed that by either changing the G/C ratio and loop length, the system was shifted towards the formation of lower molecularity species. In the case of the sequences with a higher C-content (GGGGCCC and CCCCGG) lower molecularity species are formed due to their lesser ability to form G4s in consequence of G-C competition. Instead, the formation of lower molecularity species for GGGGC can be justified with two main reasons, despite the higher G-content: **(1)** by removing one C from the original sequence, not only the G/C content was modified, but also the length of the G4-loop which the structure was reliant upon to form. This could affect the structures' ability to form mG4s, even if the G-C binding influence was decreased[204]; **(2)** the shorter loop might favour unimolecular species. Given that the CD showed a typical G4-trace and that the corresponding band on the AGE gel presented fluorescence upon NMM incubation, the answer probably resides in the formation of unimolecular species.

The mutant sequences were then analysed *via* brightfield microscopy to assess their ability to aggregate (***Figure 35c***). As expected, $(GGGGCC)_6$, the only mG4-forming sequence, showed ability to aggregate. With regards to the other sequences – unable to form stable G4s - they were also unable to aggregate. In particular $(GGGGC)_6$, similarly to $(GGGGTT)_6$, showed a preferential unimolecular arrangement, and did not present aggregation. The other two sequences did not form G4s and could only interact *via* canonical base pairing and did not present any aggregation. These results once again showed how mG4s-interactions are necessary for the aggregation. Importantly, $(CCCCGG)_6$ can form G-C canonical interactions to the same extent as the wild type $(GGGGCC)_6$, but is unable to form stable G4s, indicating that aggregation is not driven by Watson-Crick interactions.

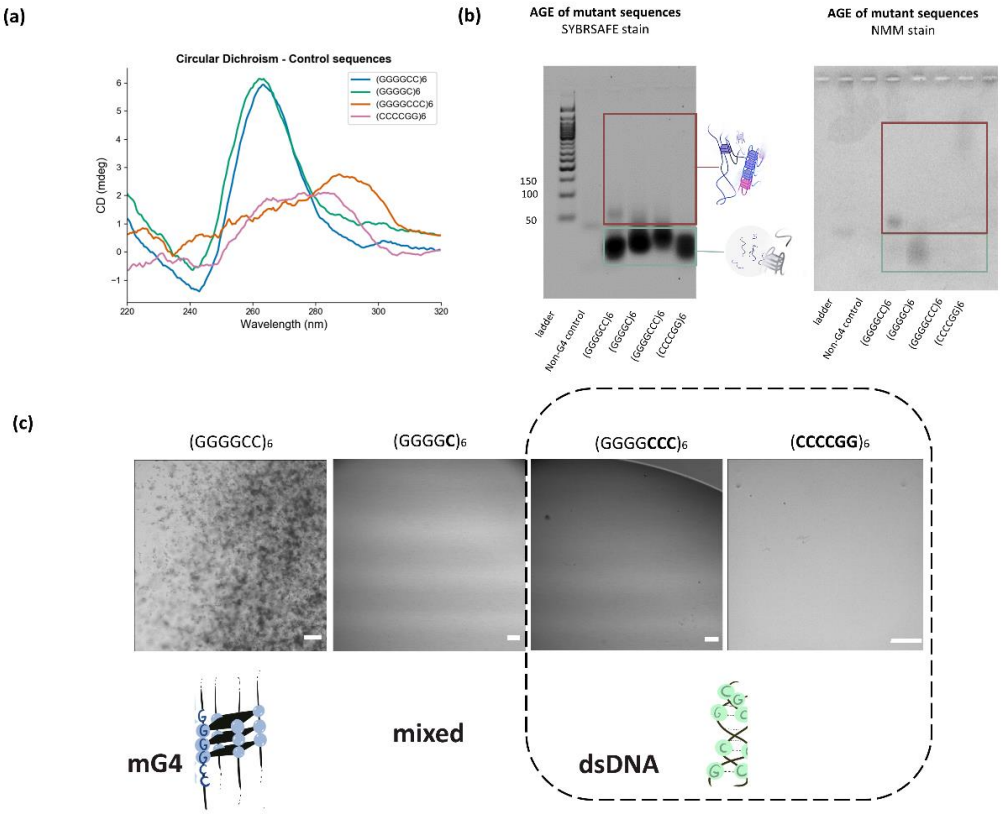


Figure 35: $(GGGGC)_6$, $(GGGGCCC)_6$ and $(CCCCGG)_6$ mG4-formation and aggregation properties. (a) CD spectra of $(G_nC_n)_6$ rationally designed mutants. Samples were all annealed at 250 µM and diluted to 5 µM to run the spectra at room temperature. CD spectra were acquired as three technical replicates by selecting accumulation = 3 on the spectrometer. The experiment was repeated at different concentrations and repeat length to ensure repeatability of the parallel conformation of $(GGGGC)_n$ and $(GGGGCCC)_n$ (Appendix 15). **(b)** AGE of the mutant sequences. The top band is only significantly prominent for the original sequences, with some smearing coming from the $(GGGGC)_6$ mutant. By staining the AGE with NMM, it is possible to notice how $(GGGGC)_6$ and $(GGGGCC)_6$ are the only G4-forming sequences. The samples are annealed at 250 µM and diluted to 10 µM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE, on ice at 65 V for 80 mins and stained with SYBR Safe or NMM for imaging. SYBR Safe-stained AGE gels were repeated in biological duplicates for different concentrations and repeat lengths of the mutants (**Appendix 16**). **(c)** Mutant sequences were annealed under mG4-forming conditions (250 µM for $(GGGCC)_6$ and $(CCCCGG)_6$ and 300 µM for $(GGGGC)_6$ and $(GGGGCCC)_6$. Aggregates were only clearly distinguishable for $(GGGGCC)_6$ and $(GGGGC)_6$ – the only sequences G4s. The different morphology of the latter, however, indicates that the species might be a mixture of other interactions. 100 µm scalebar. Hand-drawn illustration with Procreate. Aggregation experiment was performed as biological duplicates (**Appendix 17**).

The rational design of specific mutants of the $(GGGGCC)_n$ system allowed us to disentangle the different contributions of intermolecular forces contributing to aggregation. In the first set of mutants, the cytosines in $(GGGGCC)_n$ are substituted for adenine or thymine, ridding the system to any G/C contribution. Whilst aggregation was still observed for the A-mutant, this was not the case for the T-

mutant, which had been shown to primarily arrange into unimolecular G4s. Therefore, not only this proved that the G/C contribution was not necessary to aggregation, but neither were unimolecular G4s. The second set of mutants looked at varying the G/C ratio rather than substituting the bases altogether. Interestingly, by increasing the C-content of the sequences, these were not able anymore to arrange into G4s and no aggregation was observed. Instead, by reducing the C-content, the unimolecular G4-species was favoured, and aggregation could not be observed.

These results revealed two interesting points:

(1) Variation of G/C content and loop length from the wild type hexanucleotide repeat expansion $(GGGGCC)_n$ leads to loss of is aggregation abilities by forming non-mG4 structures.

(2) Affecting the molecularity of the system affects the sequences' ability to form aggregates, even when most of the sequence is left unchanged.

These observations led to question whether it was possible to prevent/change aggregation extent by using molecular probes to prevent/disturb mG4-formation of the $(GGGGCC)_n$ system. The question is addressed in later paragraphs.

4.3.2 $(GGGGCC)_n$ aggregation is prevented by G4-ligand addition

G4-targeting ligands have previously been employed to target ALS/FTD pathological aggregates in live-cells, showing an amelioration of the aggregation phenotype over time[200]. In past studies, three G4-interacting small molecules were identified and tested for their ability to bind to $(GGGGCC)_n$ RNA and DNA *in vitro* and then used to treat C9orf72 patient iPSC-derived motor and cortical neurons. A significant reduction of the RNA *foci* was observed, indicating that the ligands interaction with the G4s in cells was preventing aggregation. The authors' hypothesis was that by targeting G4 they could **(1)** Displace G4-protein interactions leading to aggregation and **(2)** Prevent the translation of toxic dipeptide protein species. In both cases, it was concluded that G4s had a direct effect on the aggregation phenomena in neuronal ALS/FTD cells.

Therefore, I decided to investigate the direct effect that the G4-ligand Pyridostatin (PDS) (***Figure 36a***)[199] had on $(GGGGCC)_n$, aiming to understand whether it would affect mG4-formation and thus preventing or delaying aggregation. As the compound was not compatible with an aqueous buffer system, it required a minimum percentage of DMSO to be solubilised. Given that DMSO had been observed to stabilise G4s in its own right[205], a first round of optimisation was performed to test for the

maximum concentration of DMSO that could be used without affecting the aggregation properties of the system.

4.3.2.1 DMSO controls optimisation

To understand how the addition of a percentage of DMSO could influence the extent of aggregation in the $(GGGGCC)_n$ system, the sequences were annealed at different concentrations (25-500 μM) under mG4-forming conditions in presence of 0.75% DMSO. For each repeat length, a series of annealing concentrations were chosen around the phase boundary in the $(GGGGCC)_n$ phase diagram (***Figure 36b***). The selection was carried out to avoid wasting excess material whilst still being able to confidently highlight whether a shift in the phase boundary could be observed. Each sample was monitored *via* optical microscopy for aggregation. Samples containing 0.75% DMSO presented a higher aggregation extent and a significant shift in the phase boundary. For example, whilst $(GGGGCC)_5$ required around 500 μM as annealing concentration in absence of DMSO, in presence of 0.75% DMSO aggregation was observed even at 25 μM. A similar effect was seen for other repeat lengths (***Figure 36b***). As DMSO is a G4-stabiliser[205], the increased aggregation extent of $(GGGGCC)_n$ in its presence offers further confirmation of the G4s involvement in the phenomenon. The same experiment was repeated by using a lower percentage of DMSO (0.125%), with the aim of leaving the system unaffected whilst still being able to solubilise G4-ligands for further applications. The samples post-annealing were analysed *via* brightfield microscopy and the results are reported in the phase diagram in ***Figure 36b***. The phase boundary remained unaffected with the exception of two data points circled in green. Given that the rest of the diagram did not present any difference and that significant evaporation had been observed in the well, those points were discarded as outliers and 0.125% DMSO was considered as an acceptable solvent system to use for G4-ligands experiments.

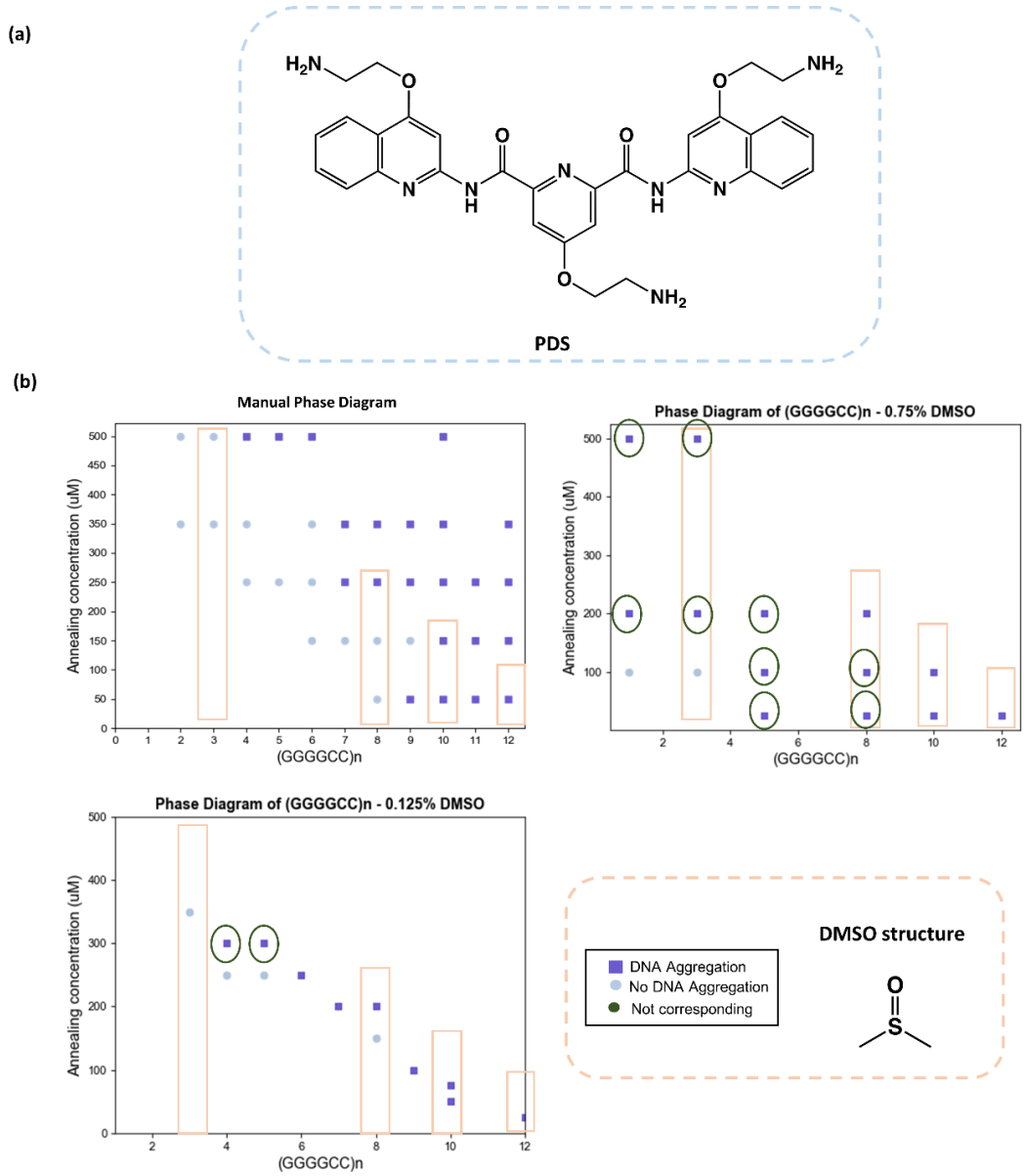


Figure 36: DMSO effect on aggregation of $(GGGGCC)_n$. (a) Chemical structure of the G4-ligand PDS (Pyridostatin). **(b)** Comparison of phase diagrams obtained by adding respectively 0.75% (**Appendix 18**) and 0.125% (**Appendix 19**) of DMSO. $(GGGGCC)_n$ sequences were annealed under mG4-forming conditions at different annealing concentrations and their aggregation extent was analysed *via* confocal microscopy. Adding 0.75% DMSO significantly shifted the phase boundary whilst 0.125% seemed to have minimal effect. To assess repeatability of the experiment and effective shift in the phase boundary, several concentrations were tested for several repeat lengths.

4.3.2.2 PDS solubility in mG4-buffer system conditions

Prior to adding the G4-ligand PDS to the mG4-annealing protocol, its solubility properties were tested in the TRIS-HCL buffer system. Indeed, PDS has a high solubility in DMSO (> 20.85 mg/mL) and in water upon gentle heating (> 9.66 mg/mL)[206], but had never been tested for 10 mM TRIS-HCl pH 7.4. Solubility issues could prevent the PDS to interact correctly with the G4s in the $(GGGGCC)_n$ system and consequently provide incorrect or an incomplete analysis. In order to assess PDS solubility, 10 µM PDS was diluted in different buffer conditions (TRIS-HCl buffer + 0.1% DMSO or TRIS-HCL + 0.1% water) in triplicates and its absorbance at 595 nm was recorded by plate reader (***Figure 37a***). Absorbance at 595 nm is commonly used to test for bacteria growth (OD600)[207], but can be employed to detect aggregation by measuring light scattering similarly to DLS (dynamic light scattering). It provides faster but less accurate information. In the assay, the plate is irradiated with a laser at 595 nm and the absorbance recorded. Since PDS does not absorb in the 595 nm region, the changes observed in the absorbance values are representative of the amount of light scattered. The higher the value of the absorbance, the higher the scattering observed in the samples and the more microscopic precipitation is expected. Upon complete dissolution of PDS in the buffer system, no difference of scattering should be observed between the PDS-containing or non-containing samples. As shown in ***Figure 37b***, in absence of PDS (negative controls), the absorbance value recorded was always below 0.04, whilst in presence of the small molecule this increased to above 0.05. The absorbance values were slightly lower in TRIS-HCl + DMSO compared to TRIS-HCl +water, reflective of the higher solubility of PDS in DMSO. Although there was an increase in the scattering with respect to the negative control, this was considered negligible compared to other previous solubility studies performed within the group on other small molecule ligands (studies by Thomas Maher). Therefore, PDS was considered soluble at 10 µM in the TRIS-HCl + 0.1% DMSO buffer system and could used for $(GGGGCC)_n$ aggregation assays.

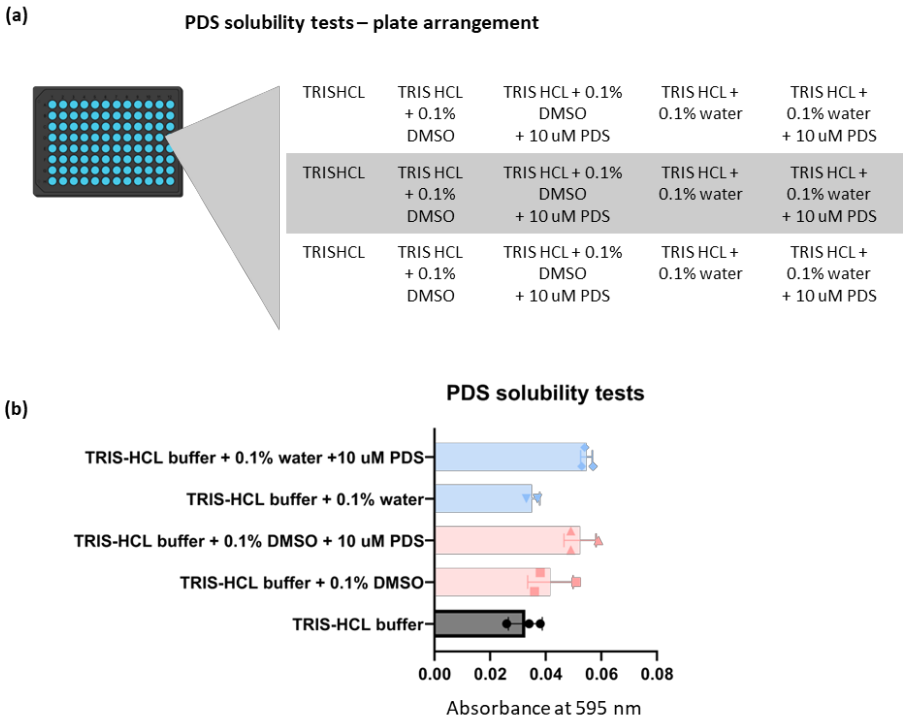


Figure 37: PDS solubility tests. (a) Experimental set up for PDS solubility tests. Each buffer condition was tested in triplicates in a 96-well plate with and without PDS to assess its solubility. **(b)** Results from the plate reader absorbance at 595 nm. In absence of PDS, there is a background scattering of < 0.04 which is present for all buffer conditions samples. In presence of PDS, this value rises above 0.05. The experiment was performed in technical triplicates.

4.3.2.3 PDS prevents aggregation in $(GGGGCC)_n$

Once assessed that PDS could be solubilised into the TRIS-HCl buffer system in presence of a small percentage of DMSO which did not affect aggregation on its own, its effect on the aggregation ability of $(GGGGCC)_{10}$ was assessed.

150 µM $(GGGGCC)_{10}$ was annealed under mG4-forming conditions in presence of 10 µM PDS and 0.1% DMSO to solubilise the ligand. Two negative controls were added to the experiments, one without the ligand and another without the ligand and DMSO. The samples were first tested for the presence of the G4-trace *via* CD spectroscopy, as reported in ***Figure 38a***. All samples showed the characteristic peaks of a parallel G4-trace, with peaks at 263 and 240 nm. Interestingly, the intensity of the samples non-containing PDS was significantly lower. Likely, the effect was due to the extensive aggregation of the samples (which will be observed later) causing inaccurate concentrations in the CD cuvette.

The formation of multimeric species was monitored by AGE (***Figure 38b***). In absence of PDS, $(GGGGCC)_{10}$ behaved as previously observed, reporting two bands (one corresponding to the mG4s and one for the lower molecularities species) together with a significant presence of smearing in the well due to the presence of higher molecularity species disassembling onto the gel matrix. Interestingly, in presence of PDS, the smearing disappeared, indicating that lower molecularity populations were present in the sample after annealing.

By looking at recent literature, I investigated whether PDS was known to preferentially bind to lower molecularity species and could therefore lead to a shift in the population in the $(GGGGCC)_n$ system. But although extensively characterised in the context of unimolecular G4s, PDS interaction with mG4s has not been thoroughly explored. In one study, however, it was shown that PDS bound more efficiently to dimeric G4s with respect to higher molecularity species by sitting in between stacked tetrads[199]. This would explain why adding PDS to the $(GGGGCC)_n$ higher molecularity system was able to shift the equilibrium towards lower molecularity species: PDS would bind to unimolecular or lower molecularity species first, preventing the system to cross the activation energy barrier to reach the multimeric conformation (***Figure 38c***). Although the mechanism of action of PDS interacting with G4s in general is still source of speculation, the experimental results from AGE and CD led to the conclusion that PDS favoured lower molecularity species in the $(GGGGCC)_n$ system.

Given the results of the mutation analysis, a higher aggregation extent was expected only in presence of a multimolecular species. Indeed, as shown in the brightfield images in ***Figure 38d***, the $(GGGGCC)_{10}$ sample containing 10 µM PDS presented a lower extent of aggregation with respect to the negative controls. This indicated that by shifting the molecularity of the system to lower molecularity species, $(GGGGCC)_n$ mG4-driven aggregation could be prevented. Interestingly, what was initially a counter-intuitive argument brough forward by *Simone* et al, proved to be in agreement with our mechanistic explanation of pathological aggregation in ALS/FTD. Indeed, just like they observed amelioration of pathological aggregates in neuronal cells by the addition of a G4-ligand, so did we observe reduction of mG4-driven aggregation *in vitro*, drawing a clear mechanistic parallel.

Of course, since *Simon et al* use a different G4-ligand, it is still possible that different mechanisms are taking place altogether.

Therefore, I questioned whether the PDS behaviour in presence of $(GGGGCC)_n$ could be generalised to other G4-ligands. I looked at PhenDC3[199] and NMM[190], two other well-known G4-stabilisers, but whilst PDS could be made compatible with the buffer system, both PhenDC3 and NMM were not successfully solubilised in presence of 0.1% DMSO in TRIS-HCl and could not be examined for these studies.

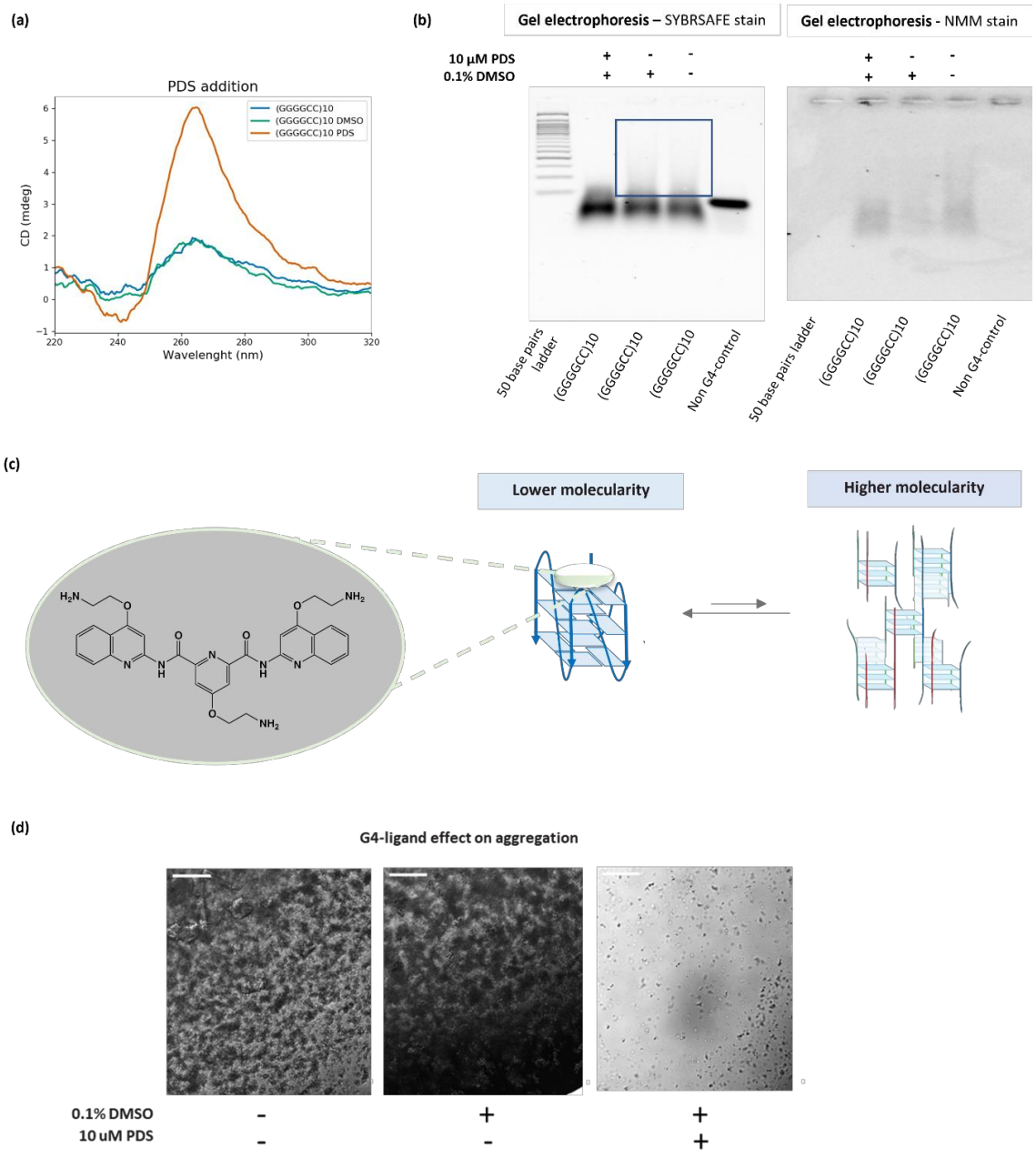


Figure 38: PDS effect on $(GGGGCC)_{10}$ mG4-formation and aggregation abilities. (a) CD of $(GGGGCC)_{10}$ annealed at 150 µM in mG4-forming conditions (in presence or absence of PDS) and were then diluted to 5 µM for CD experiments. The spectra show the typical G4-trace for all samples. CD spectra were acquired as three technical replicates (accumulation = 3). **(b)** AGE gel of $(GGGGCC)_{10}$ annealed in presence or absence of PDS. On the left, the gel is stained with SYBR Safe, whilst on the right with the G4-specific NMM. In absence of PDS, lower molecularity species are favoured as highlighted by the absence of smearing in the well. The samples are annealed at 150 µM and diluted to 10 µM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE, on ice at 65 V for 80 mins. The samples are stained with SYBR Safe or NMM. The experiment

was run in biological duplicates (**Appendix 20**). **(c)** Mechanistic hypothesis of PDS interacting with the mG4-system. By binding to the lower molecularity species (kinetic product) the equilibrium is shifted. **(d)** Brightfield images of $(GGGGCC)_{10}$ annealed in presence or absence of PDS. In presence of the ligand, aggregation is unfavoured due to the lack of mG4-species. 100 µm scalebar. The experiment was performed in technical duplicates (**Appendix 20**).

4.3.3 $(GGGGCC)_n$ antisense treatment

The results reported in the previous paragraphs provided preliminary evidence that shifting the molecularity of the system from mG4s to G4s had the ability to prevent aggregation with potential therapeutic applications. Therefore, I questioned whether it was possible to obtain a similar effect by shifting the population from mG4s to dsDNA species *via* the addition of an antisense oligonucleotide.

Ever since the C9orf72 gene was recognised as one of the main genetic causes of ALS/FTD several research outlooks focused on using antisense oligonucleotides to target the hexanucleotide repeat expansion. Outstandingly, *Tran* et al showed in Nature Medicine in 2021 that antisense therapy aimed at $(GGGGCC)_n$ administered in a patient with ALS/FTD over a course of several days could reduce pathological protein aggregation[208]. I pondered whether aggregation could be prevented not just by silencing the gene, but also by directly interacting with the species involved in the pathological aggregate formation (*e.g.,* mG4s).

Therefore, an experiment was designed to treat the mG4s aggregates with an antisense strand in the same timescales to see if the strand could affect the molecularity of the system in a similar way to G4-ligands (***Figure 39a***).

Two samples of $(GGGGCC)_8$ were annealed at 350 µM (above the aggregation threshold for this repeat) under mG4-forming conditions and observed *via* brightfield microscopy to ensure the presence of aggregates (***Figure 39e***). 330 µM of antisense strand $(GGCCCC)_8$ were then added to one of the samples, whilst the equivalent volume of buffer was added to the other as a negative control. The samples were left to incubate at room temperature for 16 days and observed after a 10 days' time interval. The long incubation time was chosen to emulate the studies in mice performed by *Tran* et al[208]. As shown in ***Figure 39b***, already from visual inspection it was observed that the sample containing the antisense strand was clearer with respect to the background, indicating a lesser aggregation extent. The samples were then analysed by CD spectroscopy to assess whether the system has shifted to its dsDNA conformation in presence of the antisense. As shown in ***Figure 39c***, the signal for the $(GGGGCC)_8$ sequence was very noisy and almost absent, whilst that of the sample incubated with the antisense showed traces of a mixed population. Indeed, unlike previously analysed CD spectra

for the same sample showing a typical parallel G4-trace with a positive peak at 263 nm and a negative one at 240 nm, here the sample presented a broad peak between 260 and 300 nm. AGE was then performed to assess the molecularity of the species. When stained with SYBR Safe, $(GGGGCC)_8$ in presence of the antisense showed a distinct band at approximately twice the molecular weight of $(GGGGCC)_8$ on its own, indicating the formation of duplex DNA. The dsDNA band did not stain with NMM as it did not contain G4s. Furthermore, the higher molecularity species present in the $(GGGGCC)_8$ negative control sample which could be visualised in the NMM-stained gel, did not appear in presence of the antisense strand, indicating a shift to lower molecularities. The combination of CD spectroscopy and AGE demonstrated that $(GGGGCC)_8$ in presence of its antisense strand would arrange into dsDNA and was unable to form mG4s. When the aggregation ability of the sequences was analysed, the samples incubated with the antisense strand presented a significant reduction in aggregation after 10 days (***Figure 39e***), confirming the role of mG4s over dsDNA in the aggregation phenomenon.

These results collectively show that using an antisense approach will successfully disassemble the mG4-based aggregates formed by $(GGGGCC)_n$ ALS/FTD relevant system *in vitro* by probing the formation of dsDNA species.

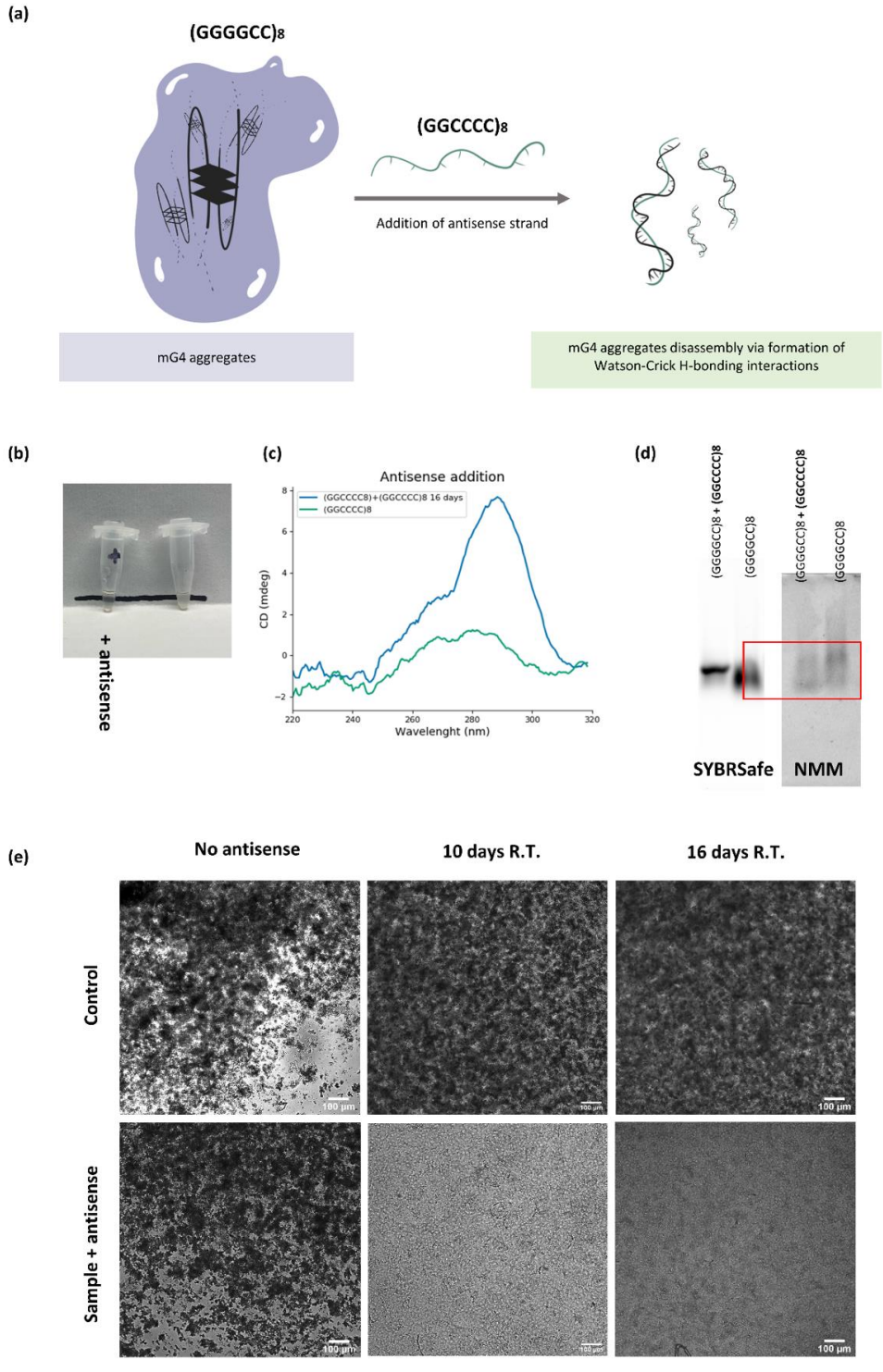


Figure 39: **Antisense treatment effect on (GGGGCC)$_8$ mG4 formation and aggregation abilities.** **(a)** Schematic of how the antisense addition would disassemble the mG4 aggregates by driving the system towards the dsDNA arrangement. **(b)** Photography of the samples before and after incubation with the antisense. The sample containing the antisense strand is visibly less opaque by comparison with the background. **(c)** CD spectra of the samples before and after incubation with the

antisense strand. The sample incubated with the antisense strand does not present a typical G4-trace. The CD spectra were acquired as technical triplicates (accumulation = 3). **(d)** AGE of $(GGGGCC)_8$ before and after incubation with the antisense strand stained with SYBR Safe (on the left) and NMM (on the right). Upon incubation with the antisense, the molecularity is shifted to a lower band. The samples are annealed at 350 µM and diluted to 10 µM for AGE. The gel running conditions are the following: 3% agarose in 1X TBE, on ice at 65 V for 80 mins. **(e)** Brightfield microscopy of $(GGGGCC)_8$ before and after incubation with the antisense strand. Both AGE and microscopy present an internal control to monitor the addition of the antisense strand. Due to the time-consuming nature of the experiment and the presence of the internal control to ensure repeatability, the experiment was performed in a single biological replicate.

4.4 Conclusions

In this chapter, I have confirmed the ability of $(GGGGCC)_n$, the most common genetic mutation in ALS/FTD patients, to aggregate *in vitro via* a mG4-based mechanism. In particular, I showed, through a detailed mutation analysis, that G/C base pairing does not play a strong role in the aggregation phenomena and that unimolecular G4s are not sufficient to lead to aggregation. It is indeed multimeric species that cause the phenomenon. Whilst G4s have been previously implicated in contributing to protein aggregation in neurodegenerative diseases, this is the first study to suggest the potential implications of mG4s in ALS/FTD.

Upon understanding that the molecularity of the system could affect its aggregation extent, two different methods were used to perturb aggregation.

The first method involved using a well-known G4-ligand, PDS. PDS had previously been shown to favour binding to lower molecularity species over higher molecularity ones. This was confirmed experimentally in the system $(GGGGCC)_n$, where upon addition of the small molecule during the annealing process, the molecularity of the system was shifted towards lighter species and the sequences' aggregation ability was reduced. Importantly, this permitted to draw a parallel with an ALS/FTD live cell study performed by *Simone et al* where the RNA *foci* in the pathological aggregates were ameliorated upon binding to G4-ligands[209].

The second method involved the use of the antisense strand $(GGCCCC)_n$. In this case, the hypothesis was to shift the molecularity by driving the formation of the equivalent dsDNA over G4s. Successfully, aggregates disassembled upon disruption of the mG4s species.

Interestingly, antisense therapy has already been partially successfully used in ALS/FTD patients to silence the C9orf72 gene[210]. Possibly, the antisense strand could also interact with mG4s species at the RNA foci, directly disassembling the aggregates over time.

Our results allowed us to draw more and more parallels with pathological aggregation in ALS/FTD, but still needed further validation to prove that mG4-driven aggregation could be biologically relevant, which will be explored in more details in the following chapters.

4.5 Materials and methods

4.5.1 Annealing of $(GGGGCC)_n$ and its mutant sequences

4.5.1.1 mG4-forming protocol

$(GGGGCC)_6$, $(GGGGAA)_6$, $(GGGGTT)_6$, $(GGGGC)_6$, $(GGGGCCC)_6$ and $(CCCCGG)_6$ DNA sequences were purchased from Integrated DNA Technologies (IDT) in lyophilised form and dissolved in MilliQ water (final concentration of 1 mM). Samples were then diluted to the annealing concentration of 0.1 mM in 10 mM TRIS-HCl (pH 7.4), heated to 95℃ for 1 h and slowly cooled to room temperature with a cooling rate of -0.25 ℃/min. Heating was performed in the Thermal cycler C1000™ with heated lid to prevent evaporation. 500 mM KCl was then added the samples, which were subsequently heated to 95℃ for 2h and cooled down to room temperature with a cooling rate of – 0.02 ℃/min. Heating was performed in the Thermal cycler C1000™ with heated lid to prevent evaporation.

4.5.1.2 mG4-forming protocol in presence of PDS

$(GGGGCC)_n$ DNA sequences were purchased from Integrated DNA Technologies (IDT) in lyophilised form and dissolved in MilliQ water (final concentration of 1 mM). Samples were then diluted to the annealing concentration of 0.1 mM in 10 mM TRIS-HCl (pH 7.4), heated to 95℃ for 1 h and slowly cooled to room temperature with a cooling rate of -0.25 ℃/min. Heating was performed in the Thermal cycler C1000™ with heated lid to prevent evaporation. 500 mM KCl was then added the samples, together with 10 µM PDS, which were subsequently heated to 95℃ for 2h and cooled down to room temperature with a cooling rate of – 0.02 ℃/min. To note is that PDS is solubilised by having a final concentration of 0.1% DMSO in the samples. Heating was performed in the Thermal cycler C1000™ with heated lid to prevent evaporation.

4.5.2 PDS solubility assay

10 µM PDS was added to 3 different buffer solutions to assess its solubility (1. 10 mM TRIS-HCL pH 7.4, 2. 10 mM TRIS-HCL pH 7.4 + 0.1% DMSO, 3. 10 mM TRIS-HCL pH 7.4 + 0.1% water). 100 µL of each sample were distributed in a 96-well plate. Absorbance at 595 nm was recorded in triplicates with the EnVision micro plate reader (Perkin Elmer). The data was collected and analysed in Prism GraphPad.

4.5.3 Antisense treatment assay

Two solutions of $(GGGGCC)_8$ were annealed at 350 µM under mG4-forming conditions. 330 µM of $(GGCCCC)_8$ antisense was then added to one of the solutions and the equivalent volume in buffer was added to the control and left to incubate for 16 days.

Chapter 5

Biological relevance of $(GGGGCC)_n$ aggregation

5.1 Background and Aims

In the previous chapters I illustrated how DNA $(GGGGCC)_n$ at different repeat length *n* was able to aggregate *in vitro* in absence of additional proteins *via* an mG4-driven mechanism. However, to prove that mG4-driven aggregation played a role in ALS/FTD cells, more evidence was required to underpin the biological relevance of our findings.

In this following chapter, my aim is three-fold: **(1)** understanding whether the biologically relevant RNA equivalent of $(GGGGCC)_n$ *r*$(GGGGCC)_n$ was also able to arrange into mG4s and aggregate; **(2)** disentangling the effect of the ASL/FTD relevant protein TDP-43 on the $(GGGGCC)_n$ aggregation mechanism **(3)** investigate the presence of G4-aggregatesin ALS/FTD patient derived neuronal cells.

5.2 Experimental Techniques and reagents

5.2.1 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gel electrophoresis (PAGE) is a widely used technique in molecular biology and biochemistry for separating proteins or nucleic acids based on their size and charge. Similarly to AGE, this method involves subjecting the sample to an electric field, causing the molecules to migrate through a polyacrylamide gel matrix. Smaller molecules move faster and travel further, resulting in distinct bands or "lanes" that can be visualized using dyes or stains[211]. Unlike AGE, however, PAGE allows to perform the analysis under denaturing conditions. This implies that all non-covalent interactions in the species, including H-bonding, can be broken.

Proteins are usually analysed *via* SDS-PAGE (Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis)[212], which can separate proteins based on their size. In SDS-PAGE, proteins are denatured and coated with SDS to give them a uniform negative charge which allows them to run through a porous polyacrylamide gel matrix. As a result, proteins migrate through the gel at different rates according to their molecular weight, allowing for protein separation and subsequent visualization. In the following experiments, the purity of TDP-43 was assessed *via* SDS-PAGE gel.

5.2.2 Electrospray ionization mass spectrometry (ESI-MS)

Electrospray ionization mass spectrometry (ESI-MS) is an analytical technique used to determine the molecular composition and mass of ions in a biological sample[213]. It works by generating a spray of charged droplets from a liquid sample, typically in a solution, which is then introduced into the mass spectrometer. In the process, the ions are de-solvated and directed into the mass analyser, where their mass-to-charge ratios are measured with high precision, providing valuable information about the composition and structure of molecules in the sample. This is often utilised to test for protein identity after protein expression, as in the case of these experiment for TDP-43.

5.2.3 Proteostat™ molecular rotor

Proteostat™ is a molecular rotor dye which is self-quenching in absence of protein-binding due to its ability to rotate on its own axis[214]. When the dye is bound to a protein aggregate, its rotation is prevented causing the dye to fluoresce. Whilst other commercial molecular rotors are usually sensitive to the environment, especially the degree of exposure of hydrophobic residues, this is not the case for Proteostat™, making it very sensitive to detect protein aggregation. Although the exact structure of the molecule is not available due IP, ***Figure 43a*** shows a schematic of a prototype dye readily available from the Proteostat™ provider[215]. In these experiments, the molecular rotor was utilized to monitor the kinetics of TDP-43 protein aggregation with/without additional DNA strands.

5.2.4 Transmission Electron Microscopy (TEM)

TEM is a powerful imaging technique used to examine the ultrastructure of materials at the nanoscale[216]. By passing an electron beam through a thin specimen, TEM reveals the details of internal structures, such as crystal lattices, nanoparticles, and biological specimens. Its high resolution, typically down to sub-nanometre levels, is employed in various fields, including nanotechnology, materials science, and biology[216]. In these experiments, TEM is employed to observed sub-microscopic structures formed by TDP-43.

5.2.5 IPSC-derived neuronal cells (iPSCs)

IPSC-derived neuronal cells (iPSCs) refer to neurons that are generated from induced pluripotent stem cells, a new type of pluripotent cells which are obtained by reprogramming differentiated cells into various cell types. The process of generating iPSC-derived neurons involves inducing the iPSCs to differentiate into neural precursor cells, followed by further maturation and specialization to obtain functional and mature neurons. These iPSC-derived neurons closely resemble native human neurons, making them an ideal platform for disease modelling and drug testing, as they allow researchers to investigate disease mechanisms and evaluate the effects of potential therapies in a human-specific context[217]. The iPSC-cell line generated were the C9N6 and C9N1 ALS/FTD cell lines with the C9-mutation object of these studies.

5.2.6 Fluorescence-activated cell sorting (FACS)

Fluorescence-activated cell sorting (FACS) is used in cell biology and immunology to sort and analyse individual cells based on their fluorescence properties[218]. It involves tagging cells with fluorescent markers (which in the case of our experiment will be an G4-specific probe, NMM) and passing them through a flow cytometer, where a laser excites the fluorophores, allowing for the identification, separation and counting of specific cell populations.

5.2.7 Anti β-tubulin cells staining

Anti β-tubulin is a fluorescent antibody which specifically binds to β-tubulin, one of the heterodimers of the protein tubulin, in turn a sub-unit of the protein of microtubules[219]. This antibody has been shown specifically to bind in neurons, proving their successful differentiation[220]. In these experiments, the antibody will be used to ensure correct differentiation of the C9-cell lines.

5.3 Results and discussion

5.3.1 The biologically relevant r(GGGGCC)$_n$ behaves similarly to (GGGGCC)$_n$

The studies performed on the ALS/FTD system presented so far have focused on the DNA sequence (GGGGCC)$_n$. However, it is the RNA equivalent which is considered pathologically relevant in the context of the ALS/FTD diseases. Indeed, *r*(GGGGCC)$_n$ has been found to be present into the pathological aggregates in ALS/FTD affected cells[114]. DNA has been used in our studies for two main reasons: **(1)** ssDNA and RNA are both nucleic acids biomolecules with similar properties and are able to arrange into G-quadruplexes with equivalent behaviours[221]; **(2)** large quantities of the sequences were required for a thorough optimisation of the system and its aggregation properties. DNA is easier to handle, less environment-sensitive and cheaper than RNA, and it was therefore established that it could be used as a substitute. However, RNA-G4s (*r*G4s) and DNA-G4s (G4s) do present some differences, including rG4s preferential arrangement into parallel G4 conformation and their higher stability with respect to the DNA equivalent[221]. Importantly these differences are expected to have an effect the formation of mG4-aggregates.

5.3.1.1 *r*(GGGGCC)$_n$ aggregates *in vitro*

mG4-forming conditions were tested on *r*(GGGGCC)$_n$ to examine if its biophysical properties differed from (GGGGCC)$_n$. *r*(GGGGCC)$_n$ (*n* = 6,10) sequences were purchased from IDT in their lyophilised form and redissolved in MilliQ water stocks. *r*(GGGGCC)$_{6\text{-}10}$ were annealed at different concentrations (0.5-250 µM) under the same mG4-forming conditions previously utilised for the DNA equivalents to study their aggregation behaviour. Importantly, the mG4-formation protocol was preserved but the RNAase free version of all components and sterile conditions were used to prevent RNA degradation. After annealing, the *r*(GGGGCC)$_n$ samples were observed with brightfield imaging to assess the presence of aggregates. Gratifyingly, aggregation was observed in a similar fashion to the corresponding DNA samples (***Figure 40***). In particular, *r*(GGGGCC)$_n$ aggregated with a length and annealing concentration dependence, presenting an amorphous shape similar in nature to that of DNA aggregates (***Figure 40a***). Interestingly, however, the annealing concentrations required to see aggregation for a same repeat length was significantly lower for the RNA equivalents. For instance, while DNA (GGGGCC)$_6$ did not present aggregates unless annealed at 250 µM and above, *r*(GGGGCC)$_6$ aggregated at 100 µM. Given that *r*G4s are known to be more stable than G4s, the greater aggregation ability could be a further indication that the phenomenon is G4-driven.

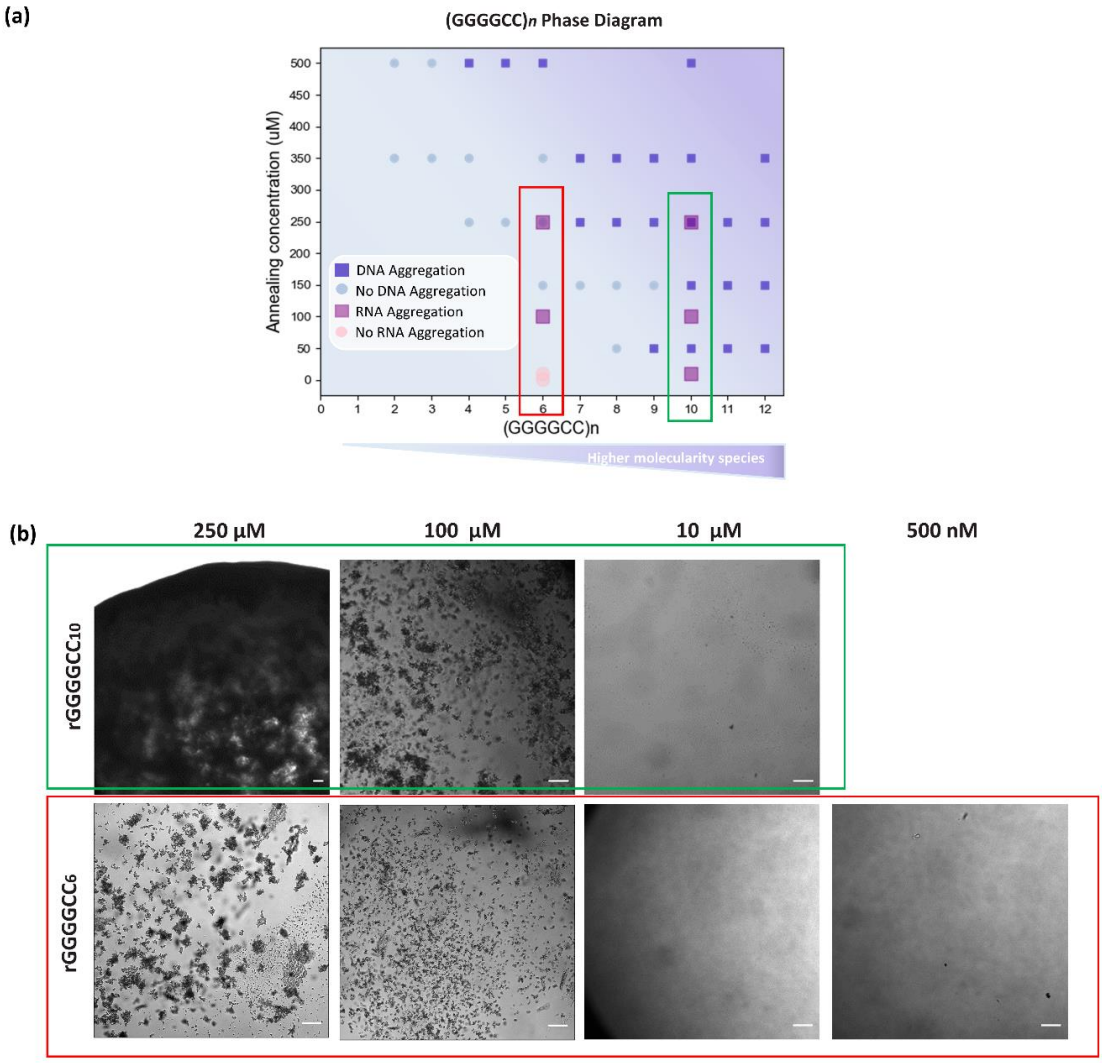


Figure 40: *r*(GGGGCC)$_n$ aggregation properties. (a) (GGGGCC)$_n$ phase diagram including the RNA equivalent data-points. like for DNA, RNA aggregates with a length and an annealing concentration dependence. Importantly, aggregation occurs at lower annealing concentrations with respect to DNA. **(b)** Brightfield images of r(GGGGCC)$_n$ sequences upon mG4-annealing as reported in the phase diagram. 100 µm scalebar. The experiment was performed in biological duplicates for *r*(GGGGCC)$_6$ (**Appendix 21**) but not *r*(GGGGCC)$_{10}$ (one biological replicate) due to unavailability of materials.

5.3.1.2 *r*(GGGGCC)$_n$ aggregates formation is G4-driven

Once confirmed that *r*(GGGGCC)$_n$ could aggregate in a similar trend to its DNA equivalents, I pondered whether aggregation was similarly mG4-driven. Firstly, *r*(GGGGCC)$_6$ (100 µM) was annealed under mG4-forming conditions and its autofluorescence properties were evaluated. If the aggregates contained G4s, these should fluoresce upon irradiation at 405 nm, much like the DNA (GGGGCC)$_n$

equivalents had been observed to do before. As shown in **Figure *41a***, fluorescence was observed in the emission range previously reported for G4-autofluorescence (450 – 700 nm), indicating that G4s were present in the aggregates.

To further confirm the presence of G4s, the aggregates were the stained with NMM, a G4-specific probe. As shown in **Figure *41b,*** the aggregates were fluorescent on the NMM-emission channel, indicating NMM fluorescence upon binding to the G4s in the structures.

Furthermore, *r*(GGGGCC)$_6$ aggregates were exposed to NMM photooxidation to ensure that mG4s were the predominant species forming the aggregates. A portion of the annealed *r*(GGGGCC)$_6$ aggregates (ROI) was subjected to irradiation at 405 nm, leading to the aggregate complete disassembly (**Figure *41c***) – thus confirming that the forces holding the structure together were G4-based.

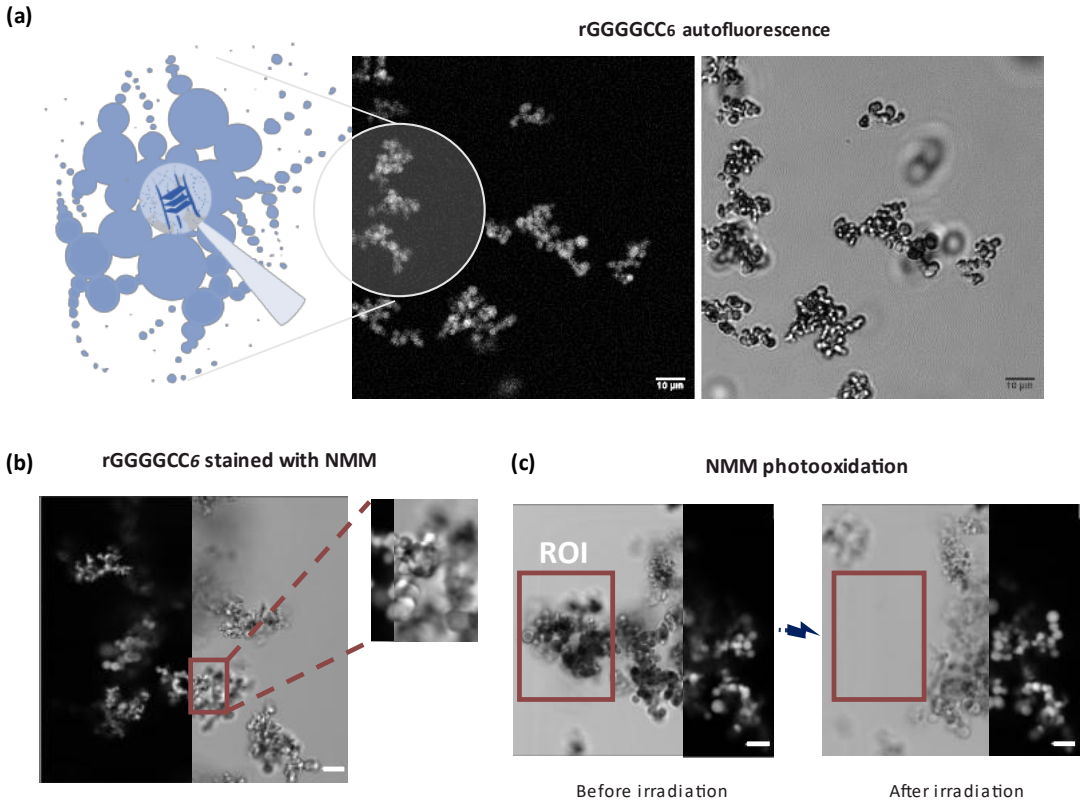


Figure 41: *r*(GGGGCC)$_6$ aggregates *via* mG4-formation. (a) *r*(GGGGCC)$_6$ was annealed at 100 µM under mG4-forming conditions is observed under confocal microscopy. The sample was irradiated at 405 nm and autofluorescence was observed on the fluorescence emission channel (450 – 700 nm). The experiment was performed as biological duplicates (**Appendix 21**). **(b)** *r*(GGGGCC)$_6$ NMM staining experiment - *r*(GGGGCC)$_6$ aggregates are 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). **(c)** *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. The experiment was performed as biological duplicates (**Appendix 21**).

5.3.2 TDP-43 enhances mG4-driven aggregation of $(GGGGCC)_n$

In the previous section, we rendered our *in vitro* system more biologically relevant by utilising RNA sequences, which are the biomolecules present in the disease aggregates[114]. But whilst this system was good as a proof of concept, it was still far from allowing in-depth characterisation of the mechanistic behaviours of the pathological aggregates in live cells. Indeed, these aggregates are known to be made up not only of RNA, but of proteins as well[97,222], leading to question which was the role of the latter into the aggregation process.

Previous literature on NDs mechanisms of formation of pathological aggregates described the protein component as the "keystone" of the aggregation phenomenon[223], but in this work I propose a different mechanism of action, one by which the nucleic acids are able to arrange on their own into aggregates. Importantly, this idea does not deny the added role of the protein into the phenomenon, but it simply shifts the focus on what could be an aggregation trigger.

In order to review the role of proteins in ALS/FTD pathological aggregation, the protein TDP-43 (TAR DNA-binding protein 43) was added to the mG4 $(GGGGCC)_n$ system. TDP-43 was chosen as **(1)** it has been shown to be present in pathological aggregates in ALS/FTD and other NDs[222] and **(2)** it has been described as a G-quadruplexes-binder and transporter in cells[101], giving it the potential to interact with the main mG4 system.

5.3.2.1 TDP-43 RNA binding domains characterisation

TDP-43 is a 414 amino acid protein sub-divided into three domains: two RNA-binding domains and one Glycine rich, prion-like domain at the C-terminus[222] (***Figure 42***). In the literature, the C-terminus is known to be very prone to aggregation due to its highly repetitive sequence of charged residues[224], making the expression of full-length TDP-43 difficult. As I was primarily interested in the interaction of the protein with nucleic acid, my studies were conducted by expressing solely the 2 RNA-binding domains (RRM1-2), which retained the protein nucleic acid binding abilities with a more accessible expression protocol.

Expression and purification of TDP-43 RRM1-2 was carried out as previously described[225] by collaborator Anouk Huyghebaert from the Di Aprile group at Imperial College London and protein identity and purity assessed *via* polyacrylamide gel electrophoresis (PAGE) and mass spectrometry (**Appendix *22***).

To ensure that the shorter version of the protein retained similar aggregation properties to its full-length version and could be used as a reliable model for interactions with the $(GGGGCC)_n$ mG4 system, TDP-43 RRM1-2' kinetics of aggregation was assessed.

While protein aggregation can be monitored with several biophysical assays, ranging from size-exclusion chromatography, gel electrophoresis, dynamic light scattering, or turbidity measurements, these do not give insights on the nature or kinetics of the aggregation, which is especially of interest for neurodegenerative diseases. Indeed, it is not just the presence of the protein aggregates which has been linked to NDs, but the nature of the aggregates themselves, whether in their fibrillar or monomeric state, which also has an impact on the disease progression[226].

In these studies, I employed a molecular rotor, Proteostat™, to quantitatively analyse the aggregation extent and kinetics of TDP-43 RRM1-2. In the assay, Proteostat™ becomes fluorescent when its freedom of rotation is restricted (*e.g.,* in presence of protein aggregates). By tracking the fluorescence trend, one can analyse the kinetics of aggregation of different proteins. Indeed, most amyloid proteins, often found in NDs, will aggregate following a nucleation-dependent fibril formation process[227], following the trend highlighted in **Figure *43b***. The protein will first misfold, leading to the formation of oligomers (Lag Phase). There will then be an intermediate stage when the protein arranges into protofibrils (Growth Phase) to then reach the full fibril arrangement (Saturation Phase).

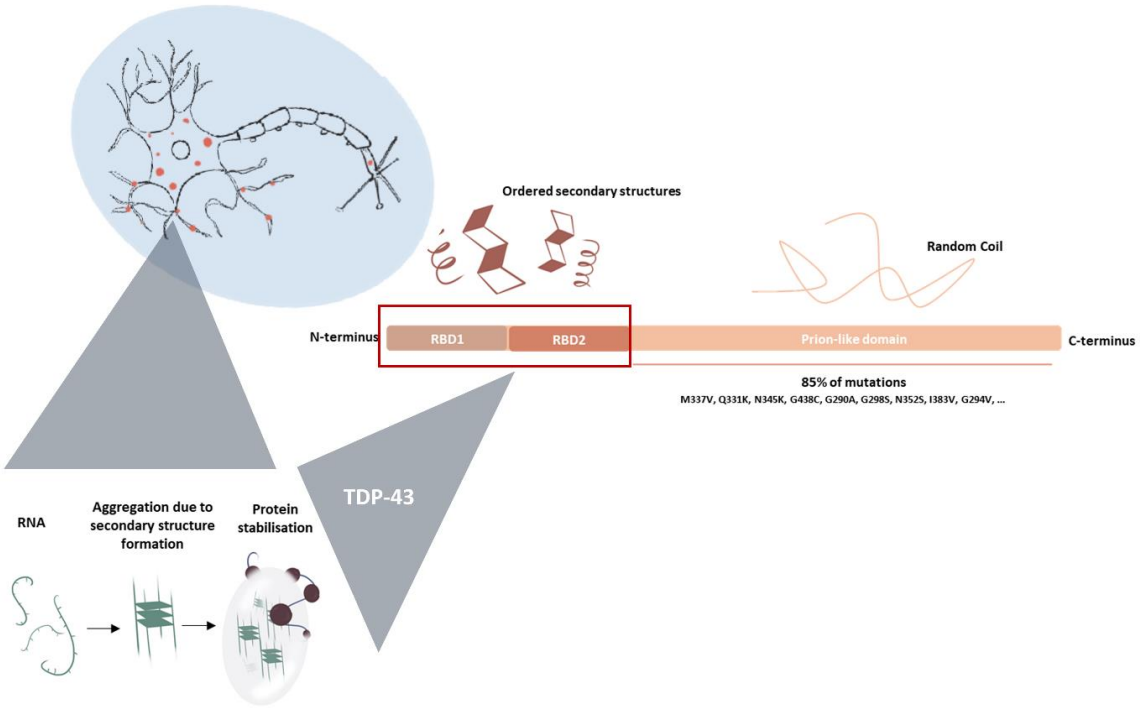


Figure 42: TDP-43 properties. Pathological aggregates in ALS/FTD neurons are composed of a nucleic acid and a protein component. One of the commonly found proteins is TDP-43. TDP-43 is made up primarily of 3 domains, two RNA-binding domains and a C-terminus responsible for prion-like aggregation. We are interested in how the RNA binding domains interact with mG4s.

In the case of TDP-43, it's been previously reported in the Aprile Group (Imperial College London) that the full-length protein does not present a Lag Phase and it arranges directly into fibrils. Therefore, when testing for the kinetics of aggregation of the TDP-43 RNA binding domains, I anticipated to solely observe a Growth and Saturation phase if the shorter version of the protein behaved similarly to full length TDP-43.

The experiment was designed so that fluorescence intensity of Proteostat™ could be detected with a ClarioStar plate reader over 3 days. Notably, all Proteostat™ experiments were designed and performed in collaboration with Anouk Huyghebaert from the Di Aprile group at Imperial College London.

100 µL of solution was prepared for each well containing 2 µL of 1M Proteostat™ dye and 10 µM TDP-43 fragments, in duplicates. Formation of amyloid-like structures was followed at 37 °C by recording the fluorescence intensity at 550 and 600 nm every 4-8 min. The resulting data is reported in ***Figure 43c*** showing as expected only a Saturation and Growth phase for TDP-43. The EC50 was extrapolated from this graph to estimate when 50% of the protein was in its aggregated state (EC50 27.05 h (26.61

to 27.47 h) 95% CI). As these results agreed with what previously reported for TDP-43 full-length, TDP-43 RRM1-2 (from now on referred to as TDP-43) was considered a good approximation for the following studies.

Therefore, I proceeded to test how the aggregation kinetics of the TDP-43 RRM1-2 was affected by its interaction with the $(GGGGCC)_n$ mG4 system.

$(GGGGCC)_{11}$ was pre-annealed under mG4-forming conditions (250 µM) and diluted to 100 µM in a 100 µL solution of 2 µL Proteostat™ and 10 µM TDP-43 (1:10 protein:DNA). The solution was transferred in duplicates in a 96-well plate and imaged on the Proteostat™ settings in the ClarioStar plate reader. Several controls were added to the experiment to ensure an accurate reading of the results (***Table 6***). Indeed, control samples were prepared by:

(1) incubating TDP-43 with denatured $(GGGGCC)_{11}$ (in its linear/lower molecularity state); in this control, the $(GGGGCC)_{11}$ sequence was subjected to optimised denaturing conditions to ensure the absence of multimeric species (as reported in ***Figure 19***). It was employed to test for TDP-43 aggregation in relation to $(GGGGCC)_{11}$ sequence rather than the mG4 secondary structure.

(2) Incubating TDP-43 with a DNA sequence known to not form G4s; this control was employed to ensure that TDP-43 interaction with $(GGGGCC)_{11}$ was not a consequence of a general DNA-related effect, but specific for this sequence.

(3) Incubating the DNA samples on their own; this control was added to test for unspecific effects of mG4 aggregates/solvated DNA on the molecular rotor.

Sample	DNA:TDP-43
TDP-43	0:1
$(GGGGCC)_{11}$ mG4	10:0
$(GGGGCC)_{11}$ denatured	10:0
Non G4-sequence	10:0
TDP-43:$(GGGGCC)_{11}$ mG4	10:1
TDP-43:$(GGGGCC)_{11}$ denatured	10:1
TDP-43: non-G4	10:1

Table 6: Samples tested with the Proteostat™ dye over three days.

The results of the kinetics experiments performed on the DNA-only control revealed that Proteostat™ was not effective in the presence of mG4-aggregates. Indeed, in the control sample containing $(GGGGCC)_{11}$ mG4s (***Figure 43d***), an almost constant high fluorescence was recorded in absence of the protein, indicating that the molecular rotor was interacting with the DNA-aggregates. This was confirmed by the fluorescence trend observed in the sample containing TDP-43 and mG4-aggregates (***Figure 43d***), where, instead of showing an increasing fluorescence due to increased aggregation, the opposite effect was observed. This decrease in fluorescence could imply a decrease in protein aggregation levels however, given the significantly higher initial fluorescence value of X compared to when TDP-43 was on its own, indicates that the molecular rotor was prevented to bind to TDP-43 due to off-site binding to the mG4s.

DNA-aggregation interference with the Proteostat™ molecular rotor was further confirmed by the fact that reasonable trends were obtained in those samples where DNA was not predicted to aggregate (denatured $(GGGGCC)_{11}$ control and the non-G4 sequence). Indeed, although some Proteostat™/DNA binding was still present in the DNA-only samples, upon addition of the protein, TDP-43 was shown to aggregate with a typical Growth and a Saturation Phases with either denatured $(GGGGCC)_{11}$ or the non-G4 sequence (***Figure 43d***).

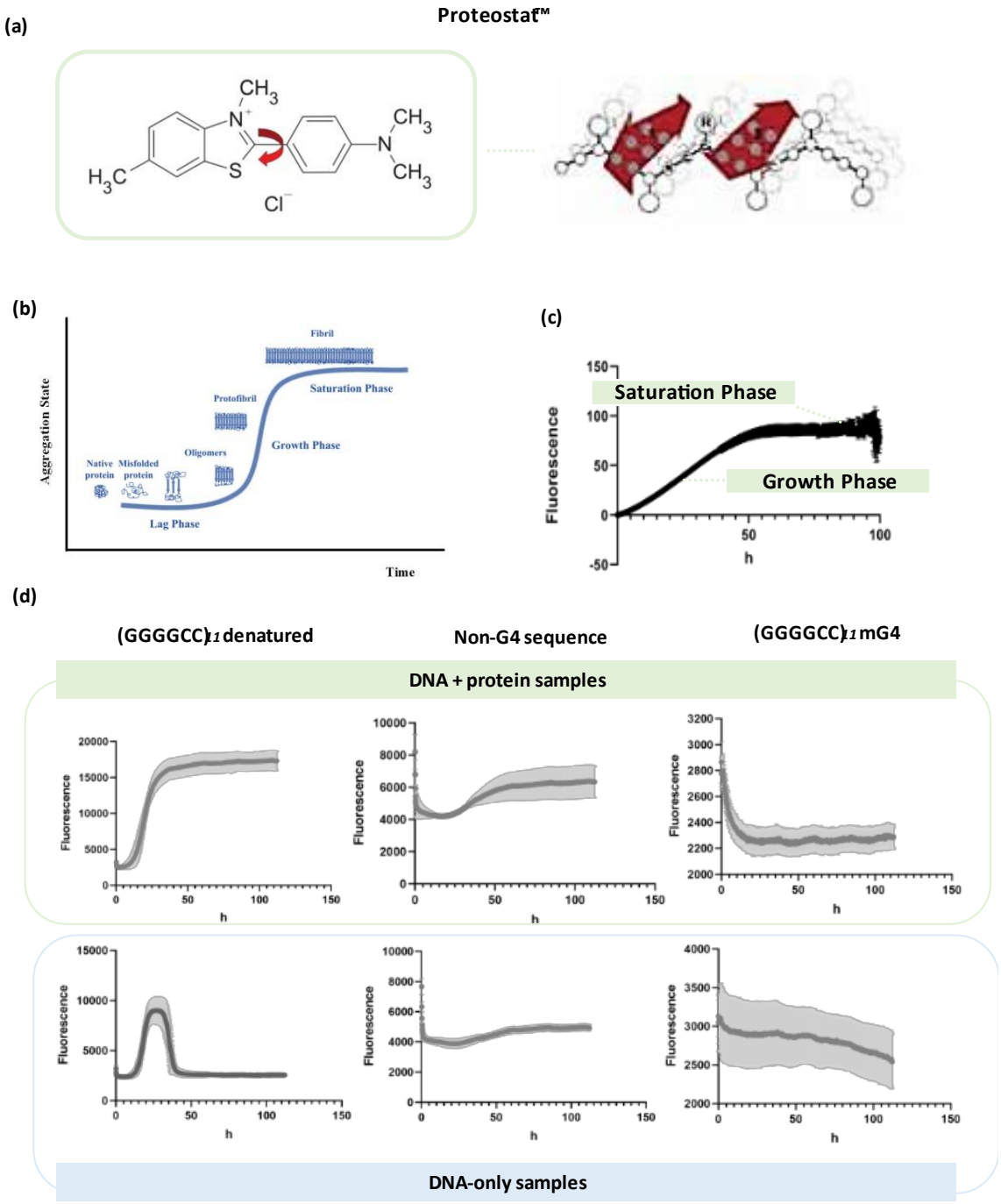


Figure 43: TDP-43 kinetics of aggregation. (a) Structure of a derivative of the molecular rotor Proteostat™ and its mechanism of rotation around its rotational axis. **(b)** Typical aggregation state graph over time for fibrillar proteins like TDP-43 (adapted from [227]). **(c)** Proteostat™ fluorescence data for TDP-43. As evident from the graph, only the growth phase and the saturation phase are present for this protein. **(d)** Proteostat™ results raw data. On the top are reported the samples containing both DNA and TDP-43, whilst on the bottom there are the same samples but in absence of the protein. All Proteostat™

experiments were designed and performed in collaboration with the Di Aprile group at Imperial College London. The experiment was performed in technical triplicates.

Although TDP-43 aggregation curve in presence of DNA had similar characteristics (*e.g.*, Saturation and Growth Phase) to the curve in absence of the nucleic acid species, I pondered whether the respective kinetics of aggregation were quantifiably different. Kinetics of proteins aggregating with a sinusoidal trend, is quantified by fitting the normalised Proteostat™ data with an asymmetric sigmoidal fitting to calculate the EC50 (*e.g.,* concentration of protein at which half of the system is in its aggregated state) (***Figure 44***).

TDP-43 in presence of denatured $(GGGGCC)_{11}$ (ssDNA or lower molecularity species) showed an EC50 of 20.20 h (20.03-20.38 h CI 95%), which interestingly was lower than the EC50 of the protein on its own (27.05 h). The lower EC50 suggested that denatured $(GGGGCC)_{11}$ enhanced protein aggregation by several hours.

I then questioned whether this effect could be generalised to all DNA sequences by testing an additional non-G4 forming sequence. In this case, TDP-43 aggregation increased to 40.74 h, more than double the time it would have taken for the protein to aggregate in absence of the nucleic acid. The higher EC50 implied that the negatively charged DNA sequences stabilised the oligomeric form of the protein, preventing the initiation of the growth phase. Indeed, whilst TDP-43 on its own has no Lag Phase, an initial plateau can be observed in presence of the non G4-DNA sequence (***Figure 44***).

I then interrogated how the mG4 arrangement of $(GGGGCC)_{11}$ impacted the kinetics of aggregation of the protein, but this could not be investigated *via* the Proteostat™ assay due to the lack of characteristic sigmoidal curve and – consequently – impossibility of calculating an EC50 for the system.

These initial results provided a more in-depth understanding of the kinetics of aggregation of TDP-43 in presence of nucleic acids. Importantly, I demonstrated that the aggregation kinetics of TDP-43 in presence of denatured $(GGGGCC)_{11}$ lowered the protein's aggregation EC50 with respect to the protein in absence of nucleic acids. Given that both $(GGGGCC)_n$ and TDP-43 have been observed in pathological aggregates at the ALS/FTD neuronal cells, it is reasonable to assume that this synergistic effect to promote aggregation could be important for pathological aggregation. Additionally, I tested TDP-43 kinetics in presence of a DNA sequence which could not assemble into G4s to assess whether the enhanced aggregation had been caused by a generic nucleic acid stabilisation. In this case, the EC50 of TDP-43 aggregation was much higher, suggesting that the nucleic acid was preventing aggregation by keeping the protein solvated. Although TDP-43 selectivity for mG4s could not be

tested, these kinetics experiments highlighted how the protein interacted with denatured $(GGGGCC)_n$ to increase its aggregation abilities, which did not occur for another DNA sequence.

Using the Proteostat™ molecular rotor to detect protein aggregation in a DNA/protein mixed system revealed several limitations of the assay. Indeed, while in protein-only samples the m Proteostat™ rotation is only affected by the increasing aggregation extent of the protein, in presence of DNA there is a series of competing events which can take place. Firstly, Proteostat™ might bind to DNA species *via* π-stacking interactions. Specifically, Proteostat™'s parent structure is very structurally similar to ThT, a well-known DNA (and specifically G4s)-binder[228]. This was indeed confirmed by analysing DNA-only samples in presence of the probe as shown in ***Figure 43d***, where an almost constant fluorescent intensity is recorded for a non-G4 forming DNA sequence and an mG4-assembled sequence. Interestingly, in the case of denatured $(GGGGCC)_{11}$, an initial increase in fluorescence followed by a sharp drop is observed, potentially linked to the ability of this sequence to form unimolecular G4s and mG4s over time. However, in absence of orthogonal assays it is impossible to predict the dynamics of Proteostat™'s specific interaction with the DNA species. Due to the unpredictability of the probe binding to other species in solution, the resulting trends and EC50 calculated might not be reflective of the effect of DNA on protein aggregation.

An alternative assay that could be used to investigate the extent of aggregation by distinguishing the effects of DNA and protein component would be fluorescence polarisation (FP) (which has previously been optimised for G4-systems)[229]. In FP, the biomolecule of interest is labelled with a fluorophore. In absence of additional binding or aggregation, the biomolecule is free to tumble in solution. When the fluorophore is excited by a light source, fluorescence will be emitted in all directions and the detector will be able to pick up a fraction of that light which is plane polarised in its direction. Upon binding (or in presence of aggregation when the complex is larger) the biomolecule's rotation in solution will be limited and the plane polarised light picked up by the detector will be higher (**Figure 44c**).

By applying FP to the TDP-43/mG4 aggregating system, several events could be distinguished. By fluorescently labelling the DNA component, the extent of DNA aggregation could be monitored over time and in a range of concentration upon addition of the protein. In parallel, the same experiment can be performed by fluorescently tagging the protein and observe protein aggregation. By combining the results of both experiments, the kinetics of aggregation of either biomolecules can be compared.

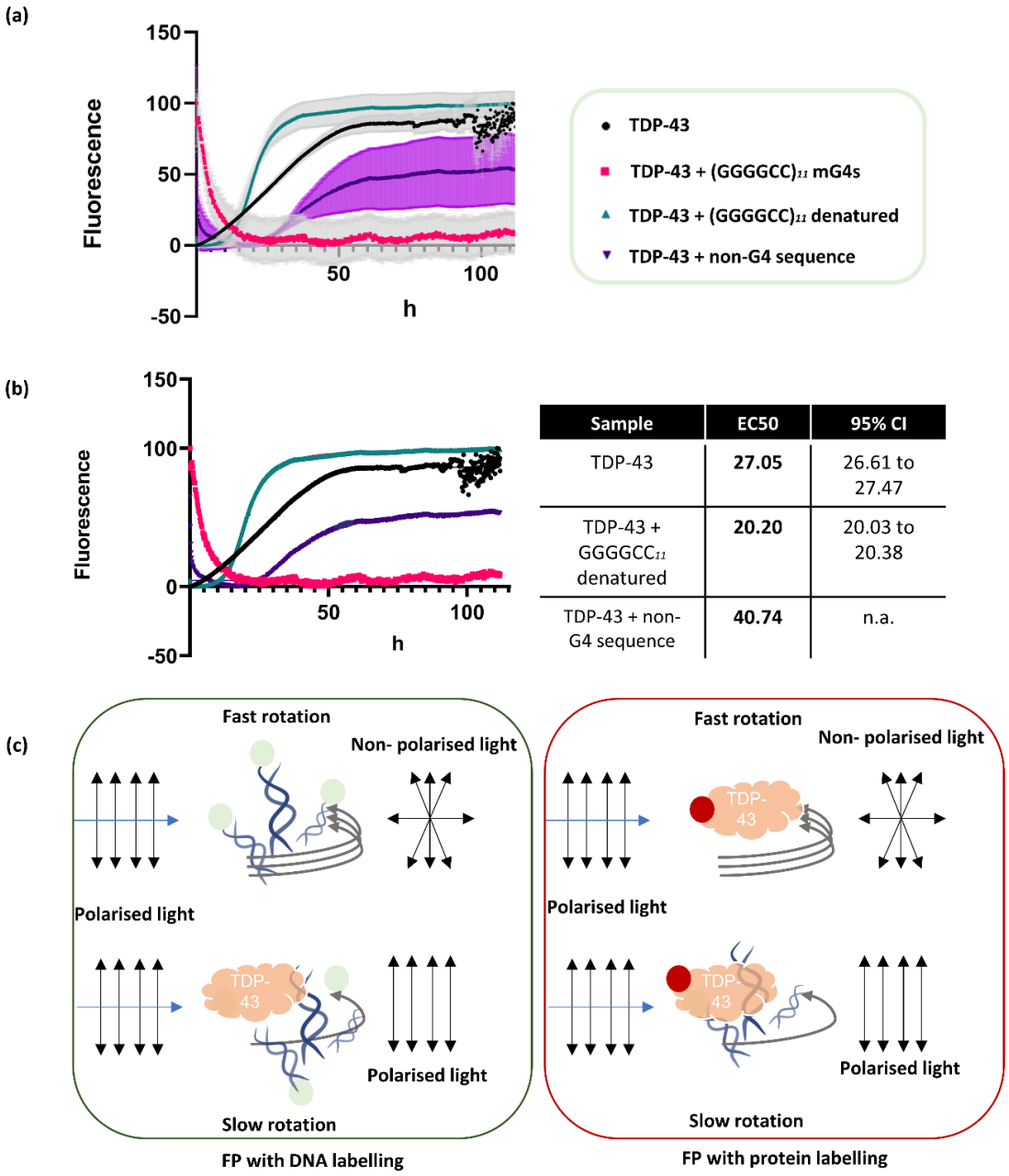


Figure 44: EC50 of aggregation of TDP-43 in presence/absence of DNA. (a) Normalised fluorescence increase curves for samples: TDP-43 only, TDP-43 + $(GGGGCC)_{11}$ mG4s, TDP-43 + denatured $(GGGGCC)_{11}$ and TDP-43 + non-G4 DNA. **(b)** Same samples are reported in absence of the error bars for clarity by highlighting the EC50 cut-off. On the right, a table summarises the EC50 data as calculated by fitting the curves with an Asymmetric Sigmoidal Fitting on GraphPad. The experiment was performed as technical triplicates (error bars in panel (a)). The CI for the TDP-43 + non-denatured G4 sequence could not be calculated due to the poor curve fitting with the Asymmetric Sigmoidal Fitting parameters. **(c)** Schematics of a potential substitute experiment to the molecular rotor to understand the kinetics of aggregation of the TDP-43 and mG4 system. In a fluorescence polarisation experiment either the DNA (first frame) or the protein (second frame) are labelled with a fluorophore. Depending on the presence of binding/aggregation, the rotation speed of the biomolecule complex will be different, and consequently so will the readout of the polarised light.

5.3.2.2 $(GGGGCC)_{11}$ mG4-driven aggregation is affected by TDP-43

The Proteostat™ results indicated that TDP-43 aggregation was affected by the presence of $(GGGGCC)_{11,}$ but could not **1)** provide information on the nature of the aggregates and **2)** report the effect of the mG4s into the system.

5.3.2.2.1 $(GGGGCC)_{11}$ mG4s co-localise with TDP-43 during aggregation

To address the first point on the nature of the aggregates, a new incubation protocol was optimised to allow observation of aggregates formed by DNA and/or protein *via* confocal microscopy (***Figure 45a***). The $(GGGGCC)_{11}$ sequence (10% FAM labelled) was annealed under mG4-forming conditions at 50 µM (a concentration above the aggregation threshold according to the phase diagram in ***Figure 26***). The sequence was then incubated with 10 µM TDP-43 (fluorescently labelled with ALEXA-633) for 3 days (***Figure 45b***). Notably, by adding the volume of protein solution to the sample, the final concentration of DNA was decreased to 31.25 µM (still above the aggregation threshold). The sample was then analysed *via* confocal microscopy over multiple channels to visualise the biomolecules co-localization in the complex. As shown in ***Figure 45c*** aggregates were observed to fluoresce both in the FAM (DNA) and the ALEXA-633 channel (TDP-43), indicating co-localization of both biomolecules. The FAM channel showed a significant amount of background fluorescence, indicating that much of the DNA was in solution. Interestingly, the aggregate showed some dark patches within the fluorescent structures, which I ascribed to either **(1)** Carboxyfluorescein self-quenching at high concentrations or **(2)** photobleaching of the dye over time.

To discern between the two, Fluorescein (the equivalent fluorophore of FAM non-covalently attached to the DNA) was tested for both behaviours. Fluorescein was incubated at different concentration in an aqueous buffer system and its fluorescence was recorded over time (**Appendix *24***). While the fluorescence intensity was not affected by photobleaching over time, it significantly reduced at high concentrations, indicating self-quenching. Notably, self-quenching was only observed at high fluorescein concentrations (> 10 mg/mL) and not under the conditions used for the experiments (µM range). Likely, this was a consequence of local concentrations in the aggregate themselves, much higher than the estimated DNA concentration in solution. This hypothesis was not only based on the

results from the fluorescein self-quenching assay, but also on previous literature reporting this behaviour[230].

The aggregates in ***Figure 45c,*** which presented fluorescence on both FAM and ALEXA-633 channels, were ascribed to a biomolecular aggregate made up of co-localised nucleic acid $(GGGGCC)_{11}$ and protein TDP-43.

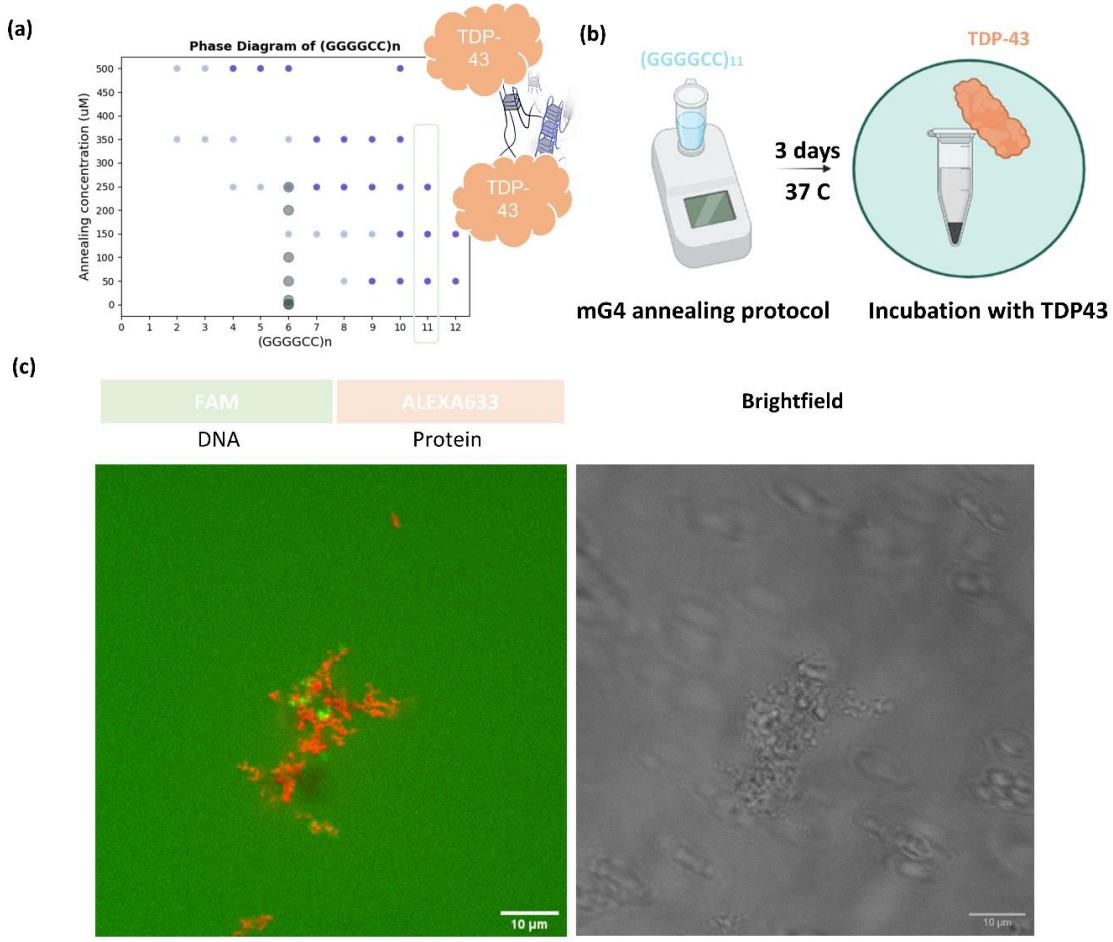


Figure 45: TDP-43 and $(GGGGCC)_{11}$ co-localisation in the aggregates. (a) $(GGGGCC)_n$ phase diagram with highlighted the selected concentrations for incubation with TDP-43. **(b)** Schematic of the protocol utilised for the experiment. $(GGGGCC)_{11}$ is subjected to an mG4 annealing protocol at different concentrations and is then incubated with a fixed concentration of protein for 3 days. **(c)** Confocal microscopy images of an aggregate formed by $(GGGGCC)_{11}$ mG4s (FAM labelled) and TDP-43 (ALEXA633 labelled). As evident from the fluorescence in both channels, both biomolecules are co-localising. The colocalization experiment was performed in biological duplicates (**Appendix 23**).

5.3.2.2.2 $(GGGGCC)_{11}$ mG4s aggregation extent is increased by the presence of TDP-43

Knowing that TDP-43 and $(GGGGCC)_{11}$ could form co-localised aggregates, I then questioned whether **(1)** the newly formed aggregates contained mG4s or ssDNA and **(2)** how the presence of the protein was affecting the extent of aggregation.

To address these questions, 10 µM TDP-43 was incubated with different concentrations of $(GGGGCC)_{11}$ after mG4-annealing (300, 100 and 1 µM – 187.5, 62.5 and 0.62 µM respectively after dilution with the protein sample). As control samples, the following conditions were included: **(1)** different concentrations of $(GGGGCC)_{11}$ incubated in absence of the protein (with the equivalent volume of protein buffer) **(2)** a TDP-43 only-sample in absence of the oligomers (incubated with the equivalent volume of oligo buffer) and **(3)** a denatured $(GGGGCC)_{11}$ control ($(GGGGCC)_{11}$ in its ssDNA or lower molecularity state) to be incubated with the protein. The latter, served to assess TDP-43 aggregating abilities in presence of mG4s in comparison to lower molecularity species given the same sequence. The samples were prepared as biological duplicates and left to incubate for 3 days at 37 ℃.

The incubated samples were imaged *via* confocal microscopy to assess the presence of the aggregates (***Figure 46a***). Qualitatively, it was evident that the samples containing both biomolecules presented a much higher extent of aggregation with respect to the individual components on their own (***Figure 46a***). To quantify this difference, each sample was stained with the G4-specific fluorescent dye NMM in a 1:1 NMM:DNA ratio. After 10 minutes of incubation, the aggregates showed a significant fluorescence on the NMM emission channel, indicating the presence of G4s. Quantification of the images was performed by acquiring three separate frames per biological sample, each maintaining the same gain and magnification settings. The images were then translated into graphics (masks) with ImageJ and the area of individual aggregates was recorded. To note is that aggregates smaller than 1.292 μm^2 were not accounted for as those could be confused with artifacts given by noise (***Figure 46b***).

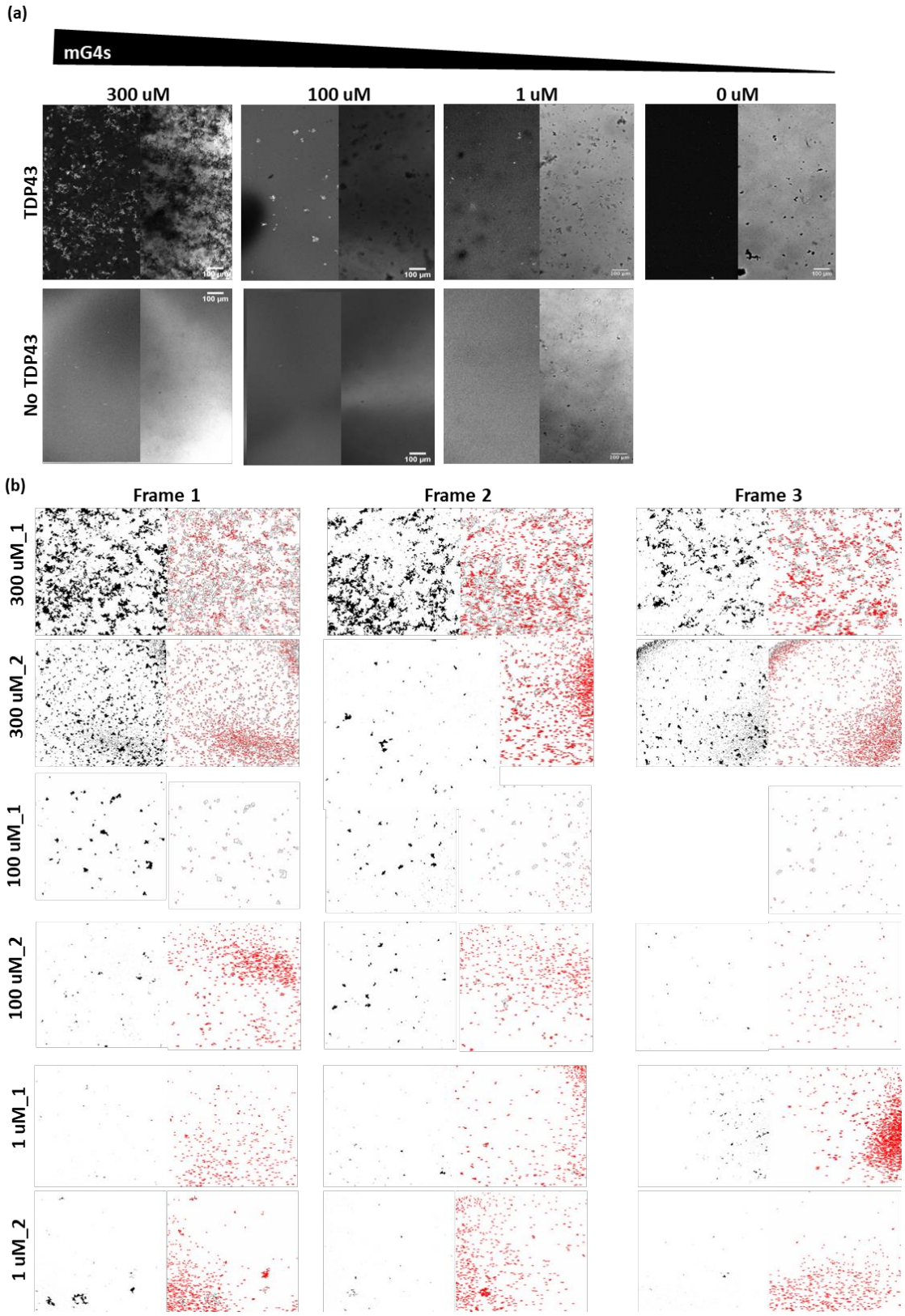


Figure 46: TDP-43 and $(GGGGCC)_{11}$ aggregates quantification *via* NMM-staining. (a) Confocal microscopy images of samples containing varying concentrations of $(GGGGCC)_{11}$ mG4s and 10 µM TDP-43, which have been stained with a 1:1 NMM:DNA ratio. Each picture is split into the brightfield channel (on the right) and the NMM-fluorescent channel (on the left). **(b)** For

each sample, three separate frames were acquired *via* confocal microscopy and quantified with image J. In the figure, each DNA concentration sample was analysed separately from its biological duplicate. Per frame, on the left we can observe the binary mask obtained from the fluorescent channel of each picture and on the right the outlines traced by Fiji based on the binary image. The experiment was performed in biological duplicates and technical triplicates (each frame).

Initially, the average size of an individual aggregate per sample was examined to understand if a higher presence of DNA/protein led to bigger aggregates. As shown in ***Figure 47***, the general trend suggested that a higher concentration of $(GGGGCC)_{11}$ mG4s did in fact lead to larger aggregates when averaging multiple biological duplicates. However, the spread of the measurements was quite broad and the error associate with the median calculation high. This was likely due to the amorphous shape of the aggregates and their spreading across different confocal planes at time of acquisition, leading to unprecise outlining and following calculations. Therefore, the same data was analysed in function of total aggregate area, to measure the extent of the aggregation regardless of average size. The trend in ***Figure 47*** showed clearly that higher concentrations of DNA led to a higher aggregation extent.

The imaging quantification results demonstrated that TDP-43 and $(GGGGCC)_{11}$ were interacting to form aggregates and that the presence of either component significantly increased the extent of aggregation. Indeed, in absence of either the protein or the DNA, the aggregates were either much smaller or undetectable *via* confocal microscopy. Furthermore, the aggregates showed to contain both the DNA and protein component (*via* co-localisation of fluorescently tagged components), but also G4s as they were effectively stained with NMM.

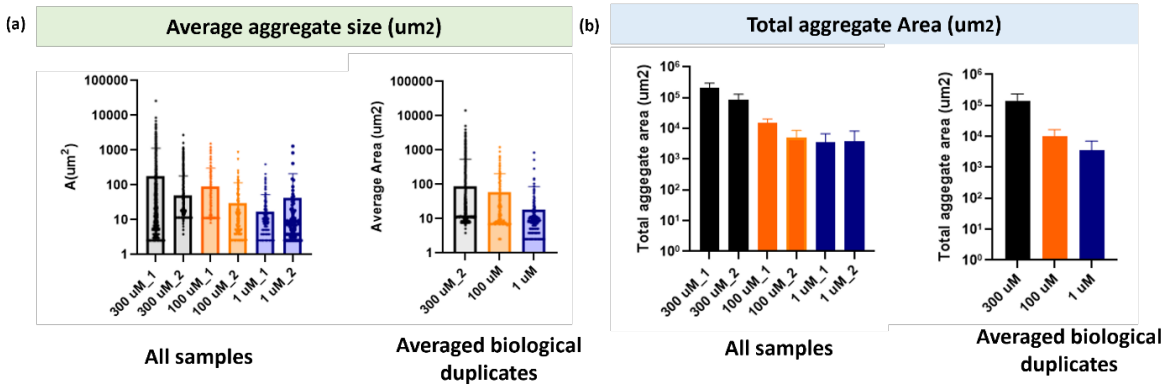


Figure 47: Confocal images quantification analysis. (a) The total average aggregate size was observed either per sample or as an average of biological samples. **(b)** The total aggregate area was observed either per sample or as an average of biological samples.

I then interrogated whether mG4s secondary structures were necessary to see an increase in aggregation or TDP-43 could equally interact with denatured $(GGGGCC)_{11}$ (previously reported to present only lower molecularity species - **Figure *19***).

Denatured $(GGGGCC)_{11}$ was incubated with 10 μM TDP-43 and the resulting aggregation product was analysed *via* confocal microscopy. As no significant aggregation was observed (***Figure 48***), the microscopy results indicated that the presence of the $(GGGGCC)_{11}$ sequence alone with the protein did not impact aggregation, and the mG4 secondary structure was required to see larger aggregates.

To test for the presence of G4-species in solution, samples were subsequently stained with NMM (1:1 NMM to DNA ratio). Staining with NMM showed a weak and diffused fluorescent signal, indicating that a few fluorescent species were dissolved in solution. As NMM fluorescence only derives from binding to G4-species, the weak signal in the sample was likely obtained from unimolecular G4s formed by denatured $(GGGGCC)_{11}$ over time. Indeed, denatured $(GGGGCC)_{11}$ has the potential to arrange into unimolecular G4s even in absence of G4-specific annealing, thanks to the added molecular crowding introduced by TDP-43. However, since unimolecular G4s do not lead to aggregation, the NMM-fluorescence signal was diffused in solution and not localised in solid aggregates. Although a few exceptions were present, these aggregates were too small to be quantified and were considered negligible.

The microscopy results of the denatured $(GGGGCC)_{11}$ control confirmed that the aggregation of biomolecular assemblies constituted by TDP-43 and $(GGGGCC)_{11}$ increased in size and concentration only if $(GGGGCC)_{11}$ is arranged into mG4s. Indeed, when the denatured nucleic acid sequence, which previous studies had reported being arranged into ssDNA/lower molecularity species (**Figure *19***), was subjected to incubation with the protein, no significant aggregation was observed.

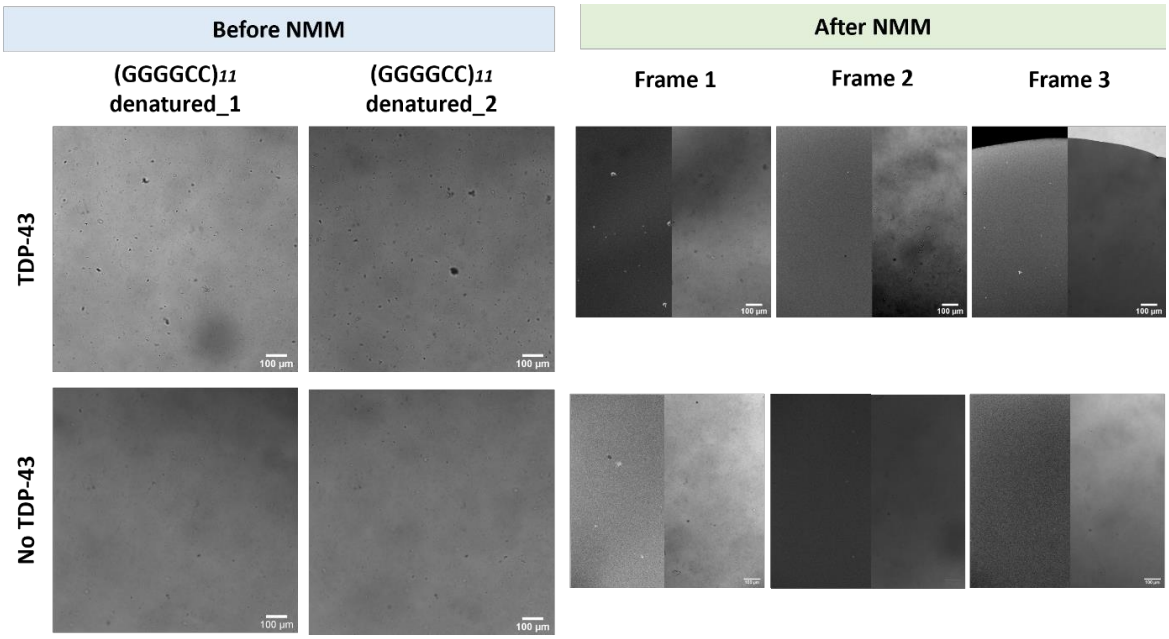


Figure 48: TDP-43 interaction with denatured (GGGGCC)$_{11}$. (GGGGCC)$_{11}$ was denatured at 100 µM and incubated with or without 10 µM TDP-43 over 3 days at 37℃ in two biological samples. After incubation, no significant aggregation was observed. The samples were then stained with 1:1 NMM:DNA concentration. The fluorescent channel is reported on the left in the panel. Very little fluorescence is detected, indicating no significant presence of G4-based aggregates. The experiment was performed in biological duplicates and technical triplicates (each frame).

5.3.2.2.3 (GGGGCC)$_{11}$ mG4s aggregates are stabilised by the presence of TDP-43

I then assessed the effect of TDP-43 on the stability of the DNA/protein aggregates by NMM-photooxidation (***Figure 49a***). In previously reported experiments (**Figure *30*** and **Figure *41***), NMM-photooxidation was used to prove that mG4s were the prevalent species holding the DNA or RNA aggregates together. In RNA/DNA only aggregates, disassembly was observed immediately upon irradiation at 405 nm, the wavelength at which NMM oxidises the Guanines in the G-tetrad leading to the G4-disruption. If the biomolecular aggregates obtained by co-localisation of TDP-43 and (GGGGCC)$_{11}$ were held together by mG4s interacting with the protein, NMM-photooxidation could present two possible outcomes: **(1)** disassembly might require longer time or not occur at all if the protein was offering further stabilisation to the aggregate or **(2)** the DNA/protein aggregates could present the same disassembly behaviours as the nucleic-acid only aggregates if the protein was not influencing the aggregate structural integrity.

(GGGGCC)$_{11}$ was annealed in mG4-forming conditions (100 µM) and incubated for 3 days at 37 ℃ with 10 µM TDP-43. The sample was then diluted to 10 µM for imaging and stained with NMM in a 1:1

NMM:DNA ratio. As shown from the fluorescent channel in the PRE-oxidation frame of ***Figure 49b***, the aggregates resulted fluorescent due to the presence of mG4s. A region of interest (ROI) was then selectively irradiated with a high-power laser at 405 nm for over two minutes to observe disassembly, but the integrity of the aggregate was not affected (***Figure 49b***). Indeed, although the NMM dye photobleached (as evident from lack of fluorescence in the NMM-channel), the brightfield showed that the aggregated was still intact. As no disassembly was observed, the experiment suggested that TDP-43 was offering further stabilisation to the aggregates. The increased aggregate stability could be due to a series of factors which involve the presence of TDP-43 in solution: **(1)** TDP-43 adds a degree of crowding which aids the stability of mG4s, making them more resistant to photooxidation; **(2)** TDP-43 binds to the mG4s, increasing their stability or **(3)** TDP-43 could be partially displacing NMM in binding mG4s, preventing photooxidation.

In all cases, the hypotheses rely on a direct interaction between the TDP-43 and mG4s, further confirming the influence of both biomolecules in the formation of the biomolecular aggregate.

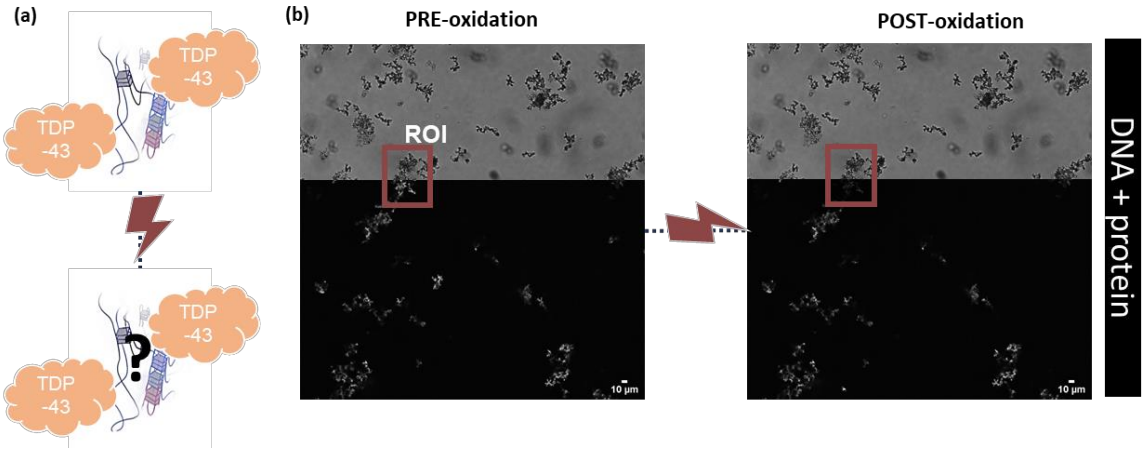


Figure 49: TDP-43 and $(GGGGCC)_{11}$ aggregates subjected to NMM-photooxidation. (a) Schematic of NMM-photooxidation of mG4s in presence of TDP-43. **(b)** $(GGGGCC)_{11}$ was annealed in mG4-forming conditions at 100 µM and incubated with 10 µM of TDP-43 as described in the text. The sample was then diluted to 10 µM for imaging and incubated with 1:1 NMM:DNA. As shown in the PRE-oxidation frame, the samples are fluorescent indicating the presence of G4s. A Region of Interest (ROI) is the selected and irradiated to observe disassembly. But whilst the NMM did photo bleach in the timeframe analysed, he aggregate remains intact as shown in the brightfield. The experiment was performed in biological duplicates (**Appendix 23**).

5.3.2.2.4 The sub-microscopic TDP-43 aggregates are stabilised by the presence of mG4s

Up to this moment the formation of aggregates was assessed by considering as aggregation threshold the sensitivity limit of the confocal microscope. TDP-43 aggregates, however, can be characterised by means of different techniques given their smaller size. Indeed, although the Proteostat™ assay confirmed that TDP-43 was assembling into aggregates, little to no aggregation could be detected *via* confocal microscopy in the protein-only sample in ***Figure 46a***. For the purpose of this study, the microscope limited sensitivity did not constitute an issue as what I was interested in was the ability of TDP-43 to interact with mG4s to create larger and more stable aggregates. However, the sub-microscopic structures in the TDP-43 samples were examined *via* TEM to **(1)** ensure that the protein was behaving as previously reported when on its own and **(2)** to observe $(GGGGCC)_n$/TDP-43 aggregates interaction with a higher resolution.

The TEM experiments were performed in collaboration with Anouk Huyghebaert from the Di Aprile group at Imperial College London. 10 µM TDP-43 was incubated for 3 days at 37 ℃ with/without 100 µL $(GGGGCC)_{11}$ previously annealed under mG4 annealing conditions. As a control, TDP-43 was also incubated with denatured $(GGGGCC)_{11}$ (ssDNA/lower molecularity species). 10 µL of each sample were then transferred onto Fromvar/Carbon coated 300 mesh copper grids and stained with uranyl acetate (2% w/v). These were analysed *via* an electron transmission microscope, and the resulting micrograph are reported in ***Figure 50***.

As previously reported in literature[231], the sample containing only TDP-43 presented aggregates visible in the nano-scale (below 500 nm) even in absence of additional species (***Figure 50***). The size of the aggregates increased significantly when in presence of mG4 $(GGGGCC)_{11}$, with aggregates a few microns large (***Figure 50***), in agreement with the previous confocal microscopy results. Indeed, since the resolution of the confocal microscope was above 1.2 µm, TDP-43-only aggregates could not be observed, while the larger biomolecular aggregates made up by both biomolecules were clearly distinguishable.

Interestingly, aggregates of around 500 nm (in between TDP-43 only aggregates and aggregates formed by TDP-43 and mG4s) were observed for the sample incubated with denatured $(GGGGCC)_{11}$ (ssDNA/lower molecularity species). In previous studies, denatured $(GGGGCC)_{11}$ had been shown to increase TDP-43 aggregation *via* kinetics studies (**Figure *44***), but no distinguishable aggregates had been observed *via* confocal microscopy (**Figure *48***). TEM studies clarified that whilst TDP-43 was still interacting with denatured $(GGGGCC)_{11}$, the aggregates were smaller than the ones observed for mG4 species, below the resolution of a confocal microscope.

The combination of the kinetics studies, quantitative confocal microscopy and transmission electron microscopy allowed me to fully evaluate the interaction of TDP-43 RNA Binding domains with different

nucleic acid species. In particular, I showed how TDP-43 and $(GGGGCC)_{11}$ mG4s can co-localise in biomolecular aggregates which become larger in presence of higher concentrations of DNA and more stable in presence of the protein.

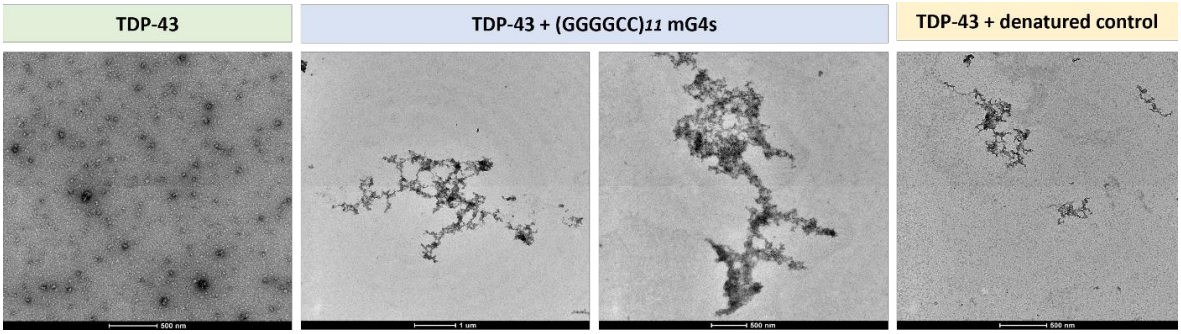


Figure 50: **TEM images of TDP-43 containing samples.** TEM images of 10 µM TDP-43, 10 µM TDP-43 incubated with 100 µM $(GGGGCC)_{11}$ annealed to form mG4s, 10 µM TDP-43 incubated with 100 µM $(GGGGCC)_{11}$ in its denatured form and $(GGGGCC)_{11}$ at 100 µM annealed to form mG4s. Experiment performed in collaboration with Anouk Huyghebaert from the Di Aprile group at Imperial College London. The experiment was performed in three technical replicates (**Appendix 25**).

5.3.3 Evidence of G4-driven aggregation in iPSC derived neuronal cells

While the addition of TDP-43 added depth to our understanding of the ALS/FTD system, the system was still limited to *in vitro* studies and further proof was required to establish that mG4-structures were playing a role in pathological aggregation at the cytoplasmatic inclusions in ALS/FTD.

Therefore, I decided to detect G4s directly in induced pluripotent stem (iPSC) ALS/FTD neuronal cells. iPSC cell lines were selected for G4 visualisation staining for two reasons: **(1)** they are representative of the ALS/FTD pathology but are more readily accessible than brain tissue derived directly from patients[217]. Indeed, the approval for use of post-mortem brain tissue for clinical purposes can only be obtained upon given consent from the late patients, but it is a lengthy bureaucratical process established by the Health Research Authority to prevent waste of the high-demand in brain tissue[232,233]; **(2)** motor neurons – often damaged in several inherited or acquired NDs like ALS/FTD - have been successfully produced from human iPSCs by directed differentiation *in vitro*[234,235], and a protocol employing iPSC cell lines has previously been developed to reveal the presence of nuclear RNA *foci* susceptible to the addition of a G4-ligand. As the goal of these experiments was to stain the RNA *foci* with a G4-stain, it was reasonable to start from well-established protocol developed for a similar purpose[236,237,238].

In my studies, I aimed to use the same approach to test the RNA *foci* for the direct presence of G4s by adding NMM, the very same fluorescent tool exploited for my *in vitro* studies to confirm that these structures have a pathological relevance. Although NMM had never been used for G4-visualisation in cells[239], it had been employed for other purposes, which suggested limited cytotoxicity. Nevertheless, a series of optimisation studies were performed to select the most appropriate conditions to visualise NMM in ALS/FTD affected motor neuronal cells.

5.3.3.1 NMM toxicity and permeability in neuronal cells

iPSC cell lines were produced and maintained by collaborator Jessica Li from the Patani Lab at the Francis Crick Institute. To assess the toxicity of NMM (dissolved in 100% DMSO) into the iPSC neuronal cells, different concentrations of the dye (2-10 μM) were incubated overnight with the live cells (C9N6 cell line). Each condition was plated in triplicates for biological relevance of the results and a red nuclear stain (DRAQ5 Nucleic acid stain) which showed no overlap with NMM was added to ensure correct visualisation of the localisation of the species. The plate was then analysed *via* high through-put confocal microscopy under the different conditions. As shown in ***Figure 51***, in absence of NMM, both the C9orf72 cell-line and its isogenic control only presented nuclear staining. DMSO controls at different percentages were added to ensure viability of the cells. In presence of increasing concentrations of NMM, the cytoplasm of the neurons fluoresced, without showing significant damage to the cells. This confirmed that NMM was non-cytotoxic. The highest concentration of NMM (10 μM) was used for the following studies as it offered the best signal-to-noise ratio while not being cytotoxic to the cells. However, the experiment presented two main limitations: **(1)** NMM did not permeate into the neuron's nuclei, as evident from a qualitative comparison between the localisation of the DRAQ5 and NMM. This was either due to low permeability of the nuclear membrane or a preference of NMM to bind in the cytoplasm, which required further exploration; **(2)** the isogenic control cell-line presented unhealthy cells which could not be used for reliable comparison. Indeed, the neurons showed a strong fluorescent signal into the nuclei, indicating that their nuclear membranes had become strongly permeable – a sign of cellular death (***Figure 51***).

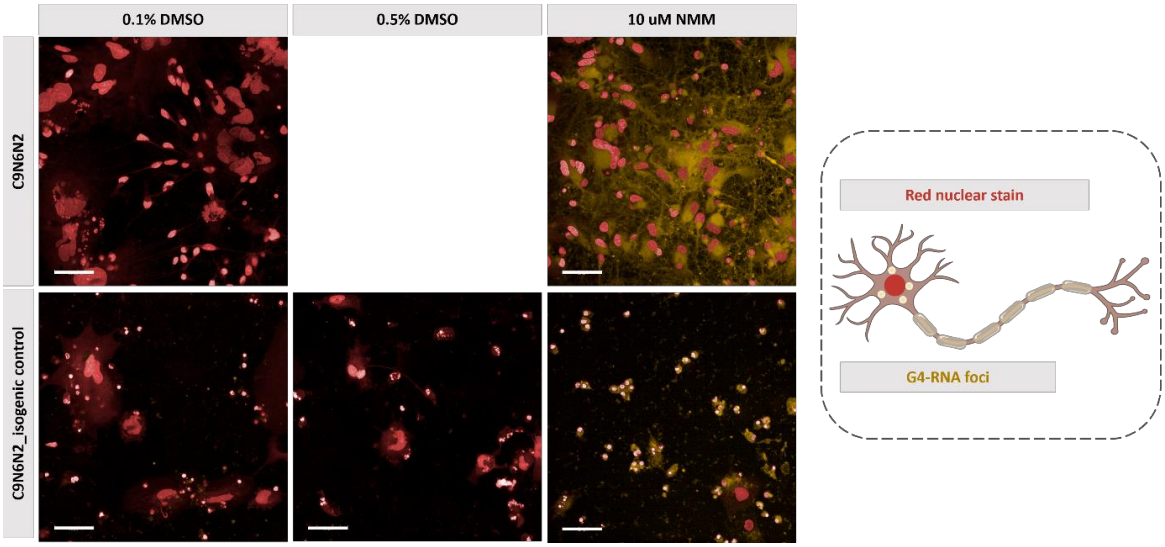


Figure 51: Live IPSC patient derived ALS/FTD neuronal cells with the C9orf72 mutation. In red, DRAQ5 Nucleic acid stain is used to visualise the different cellular entities. In yellow, the NMM channel is selectively excited. Despite the increasing concentration of NMM, the cells remained alive in the C9orf72 cell-line, indicating no/low cytotoxicity. Imaging in live cells indicated NMM accumulation in the cytoplasm. Experiment performed in collaboration with Jessica Li. Scale bar 50 µm. As a preliminary experiment for future screen, the experiment was performed as a single biological replicate.

To increase the permeability of NMM into the nucleus, the cells were first fixed with 4% paraformaldehyde and then permeabilised by sing 0.3% Triton-X. Although the nucleus did in fact start to fluoresce in the NMM-emission channel (***Figure 52***), the signal was still quite weak compared to the cytoplasmatic stain. As most pathological DNA-RNA inclusions have been observed in the cytoplasm of ALS/FTD neurons, the NMM ability to permeate the nuclear membrane was not necessary and therefore not further optimised. The issue of the poorly differentiated isogenic control cell line C9N6_i was tackled by adding an additional cell line control (C9N1). But while C9N6 and C9N1 were both efficiently differentiated in neurons, their respective isogenic controls (C9N6_i and C9N1_i) were not (***Figure 52***).

Notably, isogenic controls are cell-lines differentiated in a nearly identical manner to the C9-cell lines but in absence of the C9-mutations. This is a common experimental practise in genetics as it is useful to understand the true impact of these genetic alterations by having a control group that is as similar as possible in all aspects except for the genetic modification of interest[240]. However, changing the genomics of a cells could lead to downstream unpredicted effects unrelated to the disease, which makes the generation of isogenic cell-lines quite complex. The fact that neither C9N6_i and C9N1_i were not differentiating currently indicated the presence of unpredictable downstream effects of the

C9-mutation. As an alternative control, I decided to employ differentiated healthy motor neurons not presenting the C9-mutation.

In order to assess that the neurons (both the C9-cell line and the healthy control) were differentiating correctly, the cells were stained with anti β-tubulin. The results are reported in the confocal micrographs in ***Figure 53***, where all samples were efficiently stained with the antibody indicating a correct differentiation into neurons.

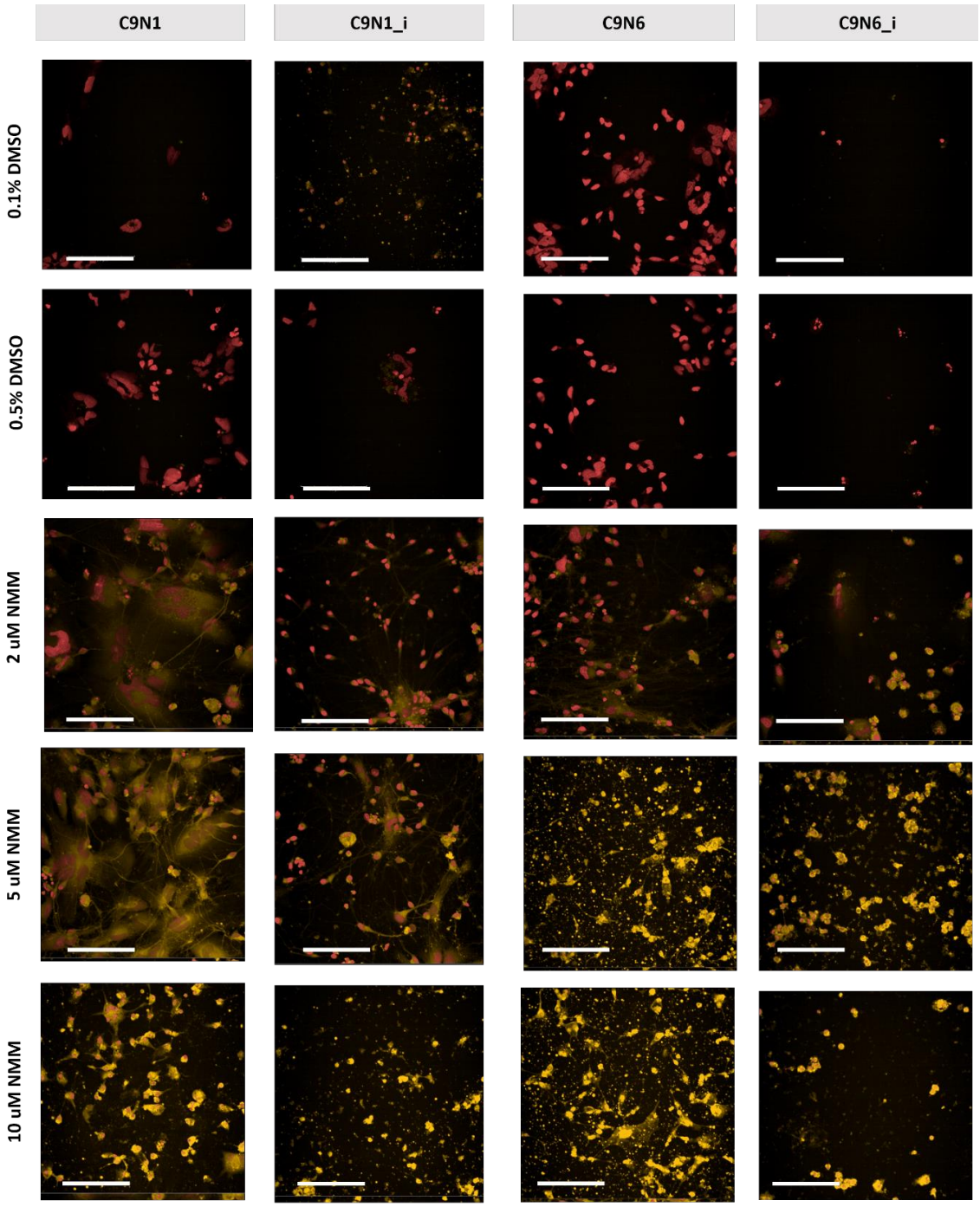


Figure 52: Confocal microscopy images of C9N1 and C9N6 fixed cell lines. Fixed IPSC patient derived ALS/FTD neuronal cells with the C9orf72 mutation. In red, a nuclear stain is used to visualise the different cellular entities. In yellow, the NMM channel is selectively excited. Despite the increasing concentration of NMM, the cells remained alive in the C9orf72 cell-line, indicating no/low cytotoxicity. Imaging in live cells indicated NMM accumulation in the cytoplasm and – to a lesser extent – the nucleus. Data obtained in collaboration with Jessical Li from the Patani group at the Francis Crick Institute. 100 µm scalebar. Triplicates images are reported in **Appendix 26** and **Appendix 27**.

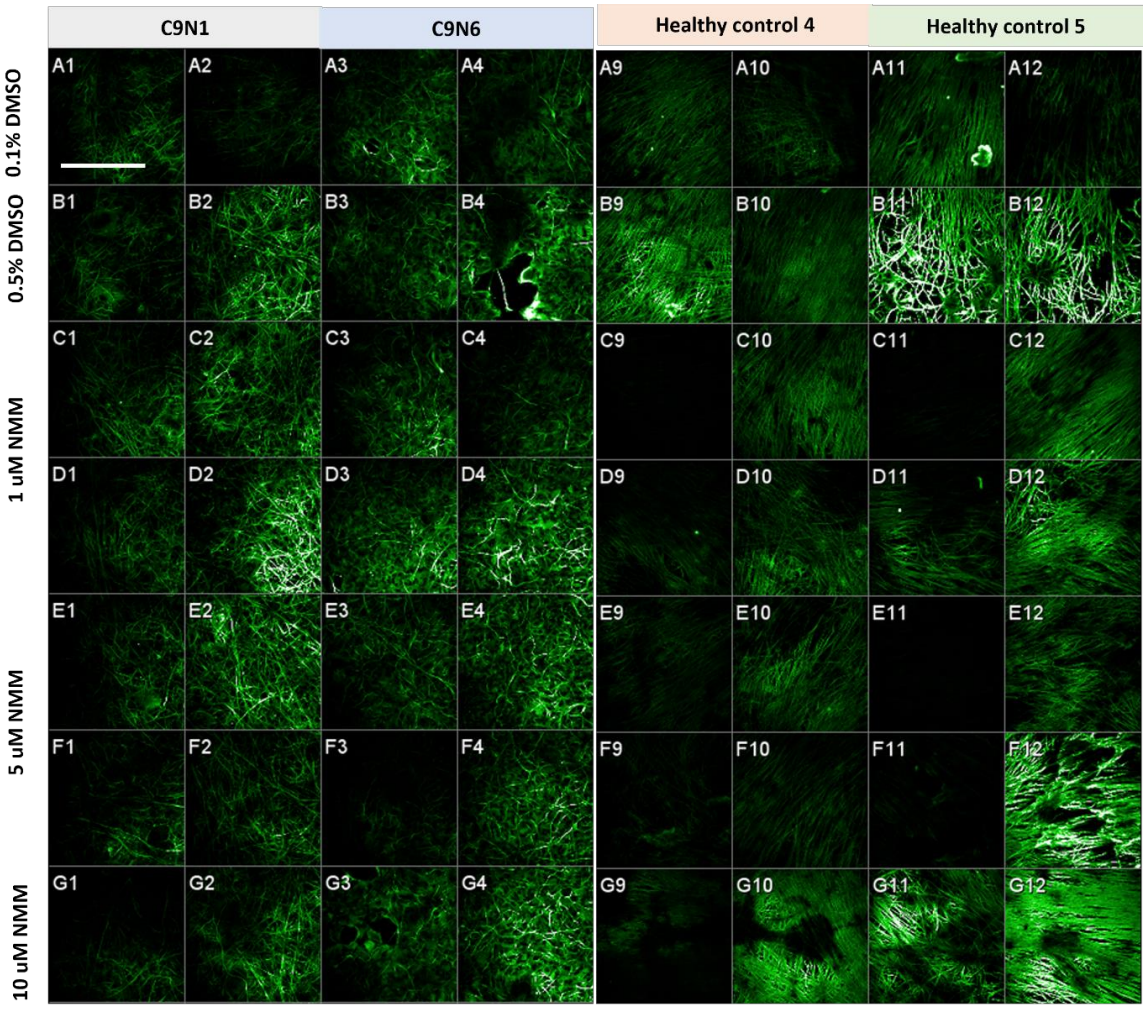


Figure 53: Confocal microscopy images of fixed IPSC patient derived neuronal cells stained with anti β-tubulin antibody. As evident from the micrographs, for the C9N1 and C9N6 pathogenic cell lines and the healthy neuron controls, all samples can be efficiently stained with the antibody indicating a correct differentiation. Data obtained in collaboration with Jessica Li from the Patani Lab at the Francis Crick Institute. 200 µm scalebar.

5.3.3.2 Quantification of RNA *foci via* staining with NMM

After having optimised the ALS/FTD cell system, I sought to quantify the NMM-staining effect on the different cell lines *via* Fluorescence-activated cell sorting (FACS). The aim was two-fold: **(1)** I wanted to understand if there was any significant autofluorescence at the NMM emission range in the DMSO control that could affect the fluorescence of the cells compared to those stained with dye and **(2)** if there was any difference between the C9-cell lines and the healthy neurons controls. FACS was performed in collaboration with Dr Yiran Wang from the Patani Lab at the Francis Crick Institute. The data, presented in ***Figure 54***, showed a significant difference in the number of cells emitting in the

NMM-emission range between the DMSO control and the NMM-stained samples. This implicated that no cell autofluorescence behaviour was occurring in the range which could affect our results.

Both cell lines (C9-mutations and healthy neurons) presented NMM fluorescent upon staining compared to the DMSO control. This was expected as, much like all cells, neurons present G4s for other biological functions unrelated to ALS/FTD pathological aggregates which are expected to fluoresce upon binding to G4s. However, the number of NMM *foci* (overall and within the nucleus) was significantly different between cell lines. Indeed, ***Figure 55b*** shows that the number of *foci* is significantly different between cell lines (data analysis performed by Dr Yiran Wang from the Francis Crick Institute).

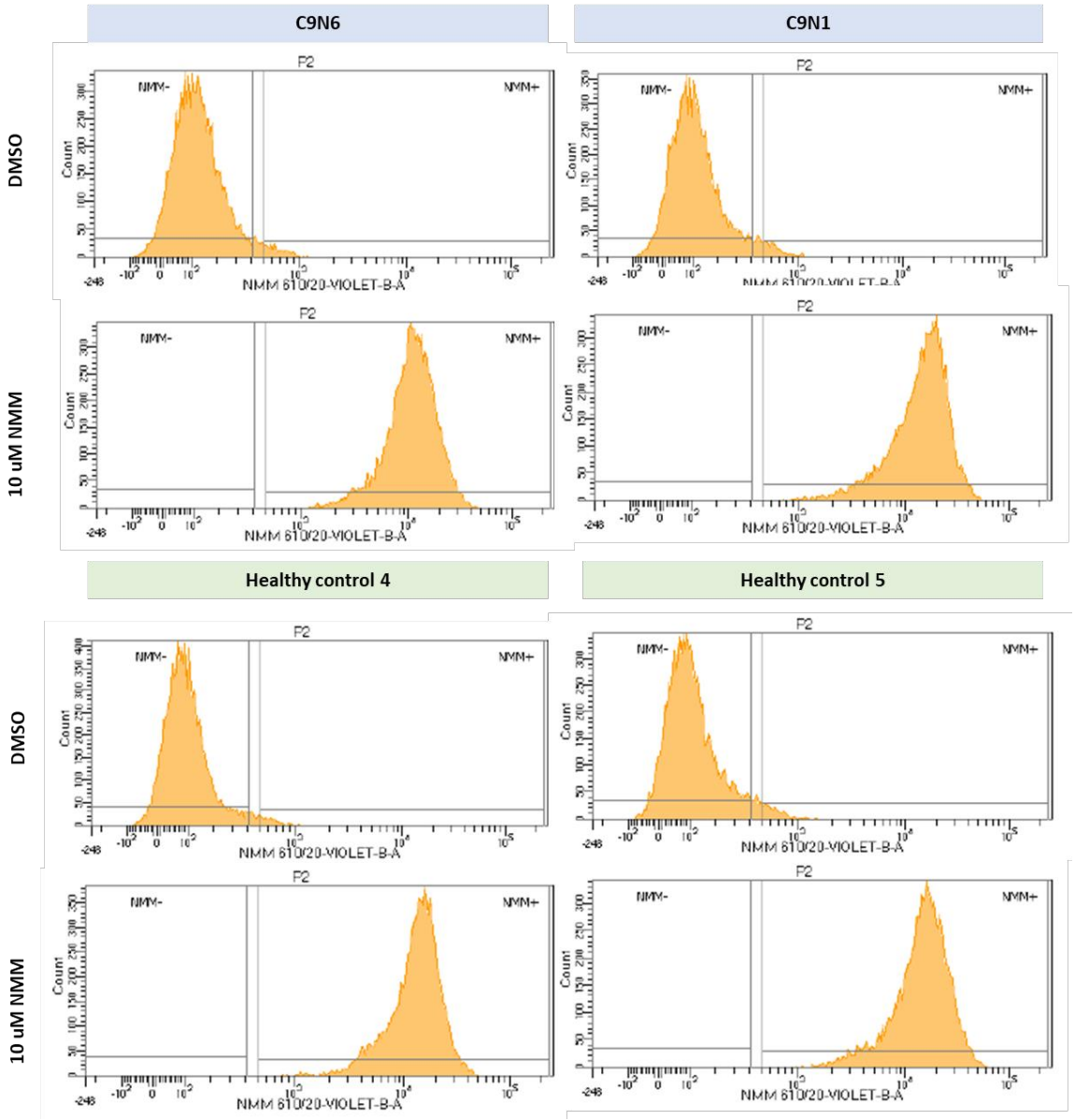


Figure 54: FACS data for the NMM staining studies. The flow cytometry studies clearly show a difference in the cell population emitting in the NMM-range between the DMSO and NMM-stained samples, indicated by the significant shift in the histogram. There is no significant difference between the healthy neurons and the C9-mutated cell-lines. Raw data obtained by Yiran Yang from the Patani Lab at the Francis Crick Institute.

Such difference can also be visually noticed by analysing the individual micrographs to distinguish the presence of NMM-stained *foci* in the C9-cell lines *versus* the healthy neurons. Indeed, whilst NMM-stained RNA *foci* were qualitatively observed in the cytoplasm of all C9-cell lines and could not be found in the healthy neurons' controls (***Figure 55*a**).

This was the first piece of direct evidence that ALS/FTD pathological aggregates contain G4s, suggesting that these structures could be of therapeutical and physiological relevance. Although more validation of these observations should be performed, this data represents indirect validation of a mG4-based aggregation mechanism in ALS/FTD relevant cell lines.

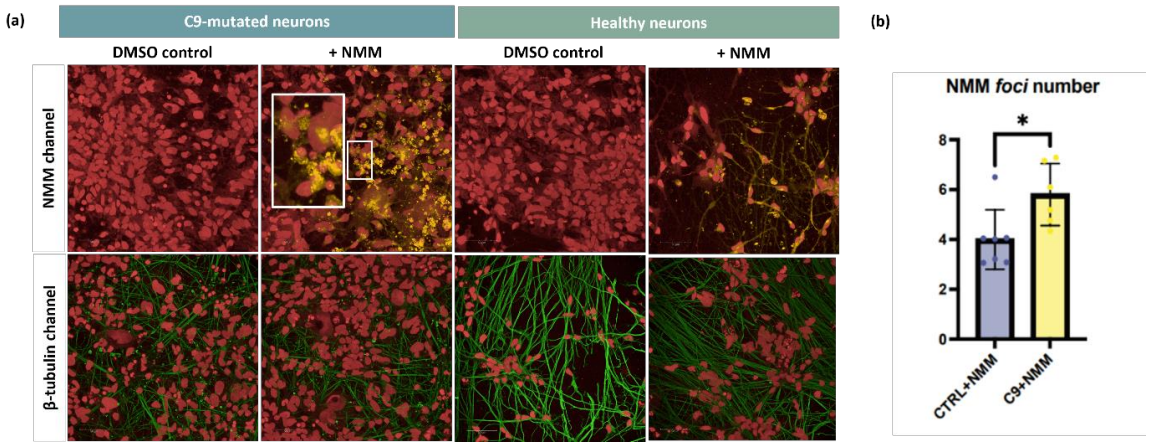


Figure 55: NMM staining of RNA *foci* in ALS/FTD cell lines. (a) Confocal images. Top row. Confocal images of IPSC patient derived neuronal cells compared to healthy neuronal cells stained with NMM. Cells were fixed *via* 4 % formaldehyde, stained with 10 µM NMM and DRAQ5 Nucleic acid stain and left to incubate for 2 days before acquisition. The micrographs show the images taken by overlapping the Nuclear Stain and the NMM channel. As clear from the DMSO control, no autofluorescence in cells can be observed in the NMM emission channel. Upon staining with the dye, the neurons containing the C9-mutation present fluorescent nuclear *foci* in the cytoplasm. In absence of the C9-mutation, this effect is significantly reduced. Notably, staining with NMM can still be observed as G4s will be present in heathy cells as well. **Bottom row.** Confocal images of IPSC patient derived neuronal cells compared to healthy neuronal cells stained with NMM. Cells were fixed *via* 4 % formaldehyde, stained with 10 µM NMM, anti β-tubulin III and DRAQ5 nucleic acid stain and left to incubate for 18 hours before acquisition. The micrographs show the images taken by overlapping the Nuclear Stain and the anti β-tubulin channel. In all cases, the β -tubulin dye efficiently stained the cells, indicating their neuronal nature. Data obtained in collaboration with Jessica Li from the Patani Lab at the Francis Crick Institute. **(b) FACS quantification of the number of nuclei.** Quantification performed by Dr Yiran Wang at the Francis Crick Institute.

5.4 Conclusions

Up to this chapter, the biophysical characterisation of the DNA $(GGGGCC)_n$ repeat sequence had proven that the sequence could arrange into multimolecular G4s and that these structures led to aggregation *in vitro*. Given that the sequence is the most common mutation in ALS/FTD and that the diseases are characterised by pathological aggregates, I hypothesised that these structures could be a contributor in their formation.

In the first section, I proved that the biologically relevant $r(GGGGCC)_n$ sequence could aggregate *in vitro* in absence of proteins in a length and concentration dependant fashion and, much like its DNA equivalents, it did so by arranging into mG4s structures, making all the observations we reported for the DNA repeats relevant for the biologically active RNA species. Interestingly, the concentration of RNA required to see aggregation for repeat $n = 6$ was much lower than for the corresponding DNA, giving an indication that this might be relevant in a cellular context at sub-micromolar concentrations for much higher repeat lengths. Indeed, it is important to remember that the $(GGGGCC)_n$ repeat expansion is only considered pathological at $n > 24$, but the RNA sequences analysed in this work only reach up to $n = 10$ due to traditional synthesis limitations. Given the length/annealing concentration dependence of the sequences, it is reasonable to assume that a higher repeat number would only require low nanomolar concentrations to obtain aggregates – which becomes an even more accessible effect with a lower aggregation threshold for RNA with respect to DNA.

I then aimed to increase the complexity of the nucleic acid-only $(GGGGCC)_n$ system to increase the biological relevance of these studies. For the purpose, I selected the RNA-binding domains of protein TDP-43, a protein commonly mutated in ALS/FTD phenotypes which has been shown prone to form pathological aggregates at the neuron level and is a known G4-binder. The hypothesis was that TDP-43 could bind and stabilise pre-formed mG4-based aggregates. I first analysed how the kinetics of TDP-43 aggregation was affected by the presence of $(GGGGCC)_{11}$. In presence of denatured $(GGGGCC)_{11}$, an increase in TDP-43 aggregation was observed, but the assay did not provide sufficient data for $(GGGGCC)_{11}$ mG4s due to binding of the molecular rotor to the aggregates themselves. Therefore, the mG4/TDP-43 interaction was monitored by observing co-localisation of both biomolecules in the aggregates *via* confocal microscopy. To confirm the synergistic effect of the biomolecules during aggregation, the same experiment was repeated by using different concentrations of both protein and DNA. It was observed that significant aggregation was only observed in presence of both entities, and the higher the concentration of mG4 DNA used, the higher the extent of the aggregation observed.

The aggregates were also stained with NMM, successfully proving the mG4-composition in the species. Interestingly, no significant aggregation was observed in presence of the denatured $(GGGGCC)_{11}$ control (ssDNA/lower molecularity species), confirming that the increase in aggregation was due to the presence of the $(GGGGCC)_{11}$ sequence itself, but it was driven by its arrangement into mG4s. The stability of the newly formed aggregates was assessed *via* NMM-photooxidation. In the experiment, the aggregates were stained with NMM and irradiated in a specific region of interest. Contrarily to what was previously observed for RNA and DNA-only aggregates, these did not disassemble upon irradiation, suggesting that the protein adds stability to the whole complex. Interestingly this indicates that NMM-photooxidation might not be used in a cellular context due to the presence of additional proteins. Finally, the aggregates were also observed *via* TEM to confirm the presence of sub-micromolar structures. Indeed, TDP-43-only samples were able to aggregate, but these species, being in the nanoscale, could not be visualised with confocal microscopy but are picked by more sensitive techniques such as TEM. These aggregates were much larger in the presence of DNA, confirming our previously obtained results.

In last part of this chapter, I further tested my hypothesis of the $(GGGGCC)_n$ system to look at the effect of the mutation directly in IPSC patient derived neuronal cells. To confirm the role of mG4s in development of pathological aggregate at the neuron level, I stained the ALS/FTD cells with NMM – the very same G4 detecting dye used so far in the *in vitro* studies. Firstly, I optimised NMM concentration and other control stains to currently visualise the location of the RNA *foci* and neuron health. The final NMM concentration chosen was 10 μM, which gave a good signal-to-noise ratio in the confocal micrographs and was not cytotoxic in the timescales experimented. Healthy neurons were used as control cell-lines instead of the isogenic control due to the poor differentiation abilities of the latter. To ensure correct differentiation, the neurons were stained with anti β-tubulin – which proved their neuronal nature – and with DRAQ5 nuclear stain – which allowed us to distinguish the nuclei. The number of cells per culture was counted *via* FACS by assessing fluorescence at the NMM emission channel. These results confirmed that **(1)** No autofluorescence was present in absence of the dye and **(2)** That there was no significant difference in the amount of staining between the C9-mutated cells and the healthy controls. This was particularly important as it was showing that, although the healthy neurons still possess G4s for their normal biological functions, they lack presence of RNA *foci* indicative of the pathology. Indeed, individual micrographs were then analysed showing a distinct fluorescent signal at the RNA *foci* in the pathological cells – which was not present in the controls. Importantly, this study was limited to 1 cell-line repeat for the healthy neurons control and would require a larger sample size and a higher microscope definition to quantify the number and size of

RNA *foci*. Altogether, my observations strongly support a model by which RNA aggregation by means of mG4-formation could be of relevance to the physiopathology of ALS/FTD.

5.5 Methods

5.5.1 RRM1–2 expression and purification

Purification of the TDP-43 RRMs fragment was carried out as previously described[241]. Briefly, the protein construct was expressed in a pET-SUMO expression vector containing a kanamycin resistance gene. The plasmid was expressed by heat-shock transformation in BL21 *Escherichia coli* (*E. Coli*) as a protein fused with a SUMO solubilization tag and a 6×His tag. Cells were grown in Luria-Bertani (LB) with 50 µg/mL kanamycin at 37 °C with 200 rpm shaking until they reached an OD (600 nm) of 0.7. Protein expression was induced with 0.5 mM IPTG, and cells were grown overnight at 18 °C with 200 rpm shaking. Cells were then collected by centrifugation at 4000 rcf for 20 min at 4°C and resuspended in a lysis buffer (10 mM potassium phosphate buffer pH 7.2, 150 mM KCl, 5mM imidazole, 5% v/v glycerol, 1mg/ml lysozyme, complete^{TM} EDTA-free Protease Inhibitor tablet by Roche, 1 µg/ml DNase I and 1 µg/ml RNaseA). The resuspended cells were sonicated on ice for 20 minutes (15s on and 45s off pulses, 20% amplitude) and soluble proteins were recovered by centrifugation at 18,000 rcf for 45 min at 4°C. The supernatant was filtered using a 0.22 µm filter and loaded onto a HisTrap HP 5 ml column (Cytiva) which was pre-equilibrated with 10 mM potassium phosphate buffer (pH 7.2) supplemented with 15 mM imidazole (binding buffer). The 6xHis-SUMO-construct was eluted with phosphate buffer supplemented with 300 mM imidazole (elution buffer). After measuring the concentration of RRMs on a Nanodrop (Thermo Fisher Scientific), the eluate was dialysed overnight at 4 °C against phosphate buffer in the presence of Tobacco Etch Virus (TEV) protease (1:10 TEV:protein construct molar ratio) to remove the 6xHis-SUMO tag. A second nickel-affinity chromatography was performed, and the flowthrough was loaded onto a HiTrap Heparin column (Cytvia) which was pre-equilibrated with 10 mM phosphate buffer. The protein was eluted with high-salt phosphate buffer (10 mM potassium phosphate buffer pH 7.2, 1.5 M KCl) and applied on a HiLoad 16/60 Superdex 75 prep grade column (Cytvia) pre-equilibrated with 10 mM phosphate buffer for size-exclusion chromatography. Protein concentration was measured on a Nanodrop (Thermo Fisher Scientific), and protein identity and purity were checked by PAGE and mass spectrometry. All protein purification steps were carried out using an AKTA Pure (Cytvia).

5.5.2 SDS-PAGE

Samples were prepared in 4X LDS sample buffer and 10X reducing agent and boiled at 95 °C for 5 min. Samples were run on 4-12% Bis-Tris NuPAGE gels (Thermo Fisher Scientific) at 180 V for 35 min. Gels were stained with InstantBlue Coomassie Protein Stain (Abcam) for 15 mins, washed in water for 30 mins and imaged on an ImageQuant LAS 4000.

5.5.3 Electrospray ionization mass spectrometry (ESI-MS)

ESI-MS was performed on purified protein samples (50 µM), buffer-exchanged into HPLC grade water to confirm molecular weight and sample purity. ESI-MS was performed by Malgorzata Puchnarewicz using the Chemistry Mass Spectrometry facilities available at the Molecular Sciences Research Hub, Department of Chemistry, Imperial College London.

5.5.4 Proteostat Binding Assays

Aggregation kinetics were followed using a BMG CLARIOstar Plus microplate reader in Greiner Bio-One CELLSTAR non-binding black 96-well half-area plates. Proteostat Solution was prepared as per manufacturer instructions (Enzo Life Sciences). To each well, 2 µL of Proteostat Solution was added in the centre followed by 98 µL of sample. The excitation and emission wavelengths were set at 550 nm and 600 nm respectively. Formation of amyloid-like structures was followed at 37 °C by recording the fluorescence intensity of Proteostat every 4-8 min using spiral averaging (3 mm diameter). Data was derived from three technical repeats, buffer-subtracted and plotted using GraphPad Prism version 8.0.1. Notably, the sequence for the non-G4 DNA control was the following: 5' TTTCACCGACGACACCAACCACGGCACCTATTATTTATTTATAATTAT 3'.

5.5.5 RRM1–2 labelling with ALEXA 633

Purified RRMs were fluorescently tagged using Alexa Fluor™ 633 C_5 Maleimide (Thermo Fisher Scientific). The dye was added to a sample of purified protein (1:10 protein:dye molar ratio) in 10 mM potassium phosphate buffer (pH 7.2) and incubated at 25 °C for 4 hours on a roller mixer with slow rotation. The solution was then dialysed against 10 mM potassium phosphate buffer (pH 7.2) overnight to remove unbound dye.

5.5.6 DNA/protein aggregates formation

$(GGGGCC)_{11}$ was annealed at different concentrations (300, 100 and 1 µM) as per mG4-annealing protocol. The samples were then incubated for 3 days at 37℃ with 10 µM of TDP-43 RBD.

5.5.7 Fluorescein self-quenching tests

Fluorescein powder (SIGMA) was diluted to 150 µL in the buffer system of choice at different concentrations (100 mg/mL – 1 µg/mL) and transferred in triplicates in a 96 black-bottomed well plate. Fluorescence readings were performed by irradiating at 475 – 490 nm and recording emission at 520 – 520 nm. Emission was recorded over 1.5 minutes. Data was analysed on Prism GraphPad.

5.5.8 Aggregates quantification analysis from microscopy

Each picture was rasterised as a binary map and analysed with the quantification tool on Fiji ImageJ. The chosen aggregates threshold size was 1.291 um^2.

5.5.9 Transmission electron microscopy (TEM)

To prepare TEM grids, 10 µL of sample was pipetted on Formvar/Carbon coated 300 mesh copper grids (Agar Scientific) and left to incubate for 2 mins. Excess sample was removed with Whatman filter paper. The grids were then washed with dH_2O and stained with uranyl acetate (2% w/v) for 2 mins. Excess uranyl acetate was removed, and the grid was left to dry for at least 5 mins. For each sample, images were taken in at least 5 locations on the grid to ensure consistent structures were present across the grid. Grids were imaged on a T12 Spirit electron microscope.

5.5.10 IPSC patient derived neuronal cell culture

Experimental protocols were all carried out according to approved regulations and guidelines by UCL Hospitals National Hospital for Neurology and Neurosurgery and UCL Institute of Neurology joint research ethics committee (09/0272). iPSCs were maintained with Essential 8 Medium media (Life Technologies) on Geltrex (Life Technologies) at 37°C and 5% carbon dioxide. iPSCs were passaged when reaching 70% confluency using EDTA (Life Technologies, 0.5 mM). iPSC cultures underwent differentiation into spinal cord motor neurons as previously described[236]. Briefly, iPSCs were plated to 100% confluency and then differentiated to a spinal neural precursor fate by sequential treatment with small molecules, day 0–7: 1 µM Dorsomorphin (Tocris Bioscience), 2 µM SB431542 (Tocris Bioscience), and 3.3 µM CHIR99021 (Tocris Bioscience), day 7–14: 0.5 µM retinoic acid (Sigma Aldrich) and 1 µM Purmorphamine (Sigma Aldrich), day 14–18: 0.1 µM Purmorphamine. After neural conversion and patterning, 0.1 µM Compound E (Bio-techne) was added for terminal differentiation into spinal cord motor neurons.

5.5.11 Immunocytochemistry (ICC)

Cells were seeded on clear bottom 96 well plates (Falcon) at a density of 30,000 cells per well. After 24 hours, Compound E (Bio-techne) was added for terminal differentiation into spinal cord motor neurons. At day 6 of terminal differentiation, DRAQ5 (Merck) was applied to the cells diluted at 1:1000 in media and incubated for 10 minutes to stain nuclei. The cells were then fixed in 4% paraformaldehyde for 15-20 minutes at room temperature. The cells were then permeabilized using 0.3% Triton-X containing PBS (NMM) or blocked with 5% BSA in PBST (antibodies) for 60 minutes. Either 10 µM NMM (Cambridge Bioscience) and DMSO diluted in 0.15% Triton-X containing PBS or a primary antibody was then applied. The NMM/DMSO was incubated for 40 hours at 4°C or the primary antibody anti-tubulin beta-III (#802001, BioLegend), diluted at 1:2000, was left overnight at 4°C. Plates treated with NMM/DMSO were imaged directly. Secondary antibody incubation was then performed

using Alexa Fluor-conjugated secondary antibody 568 nm (anti-rabbit) at 1:1000 dilution in PBS at room temperature. Cells were imaged with Opera Phenix (PerkinElmer) at 40X, a Z-stack of images were used. Standard imaging and acquisition settings were applied.

5.5.12 Fluorescence-activated cell sorting (FACS)

Motor neurons were dissociated into single cells using Accutase (Life Technologies). The cell suspensions were fixed using 4% paraformaldehyde for 15-20 minutes and then permeabilized by incubation with 0.3% Triton-X in PBS for 1 hour at room temperature. Subsequently, they were treated with either 10 µM NMM (Cambridge Bioscience) or 0.5% DMSO diluted in 0.15% Triton-X overnight at 4°C. The samples (10,000 cells/sample) were examined using a BD LSRFortessa analyser operated by the FACSDiva software (BD), and the results were analysed using the FlowJo software.

6. General Conclusions and Future Outlooks

Historically, neurodegenerative diseases have been linked to the presence of pathological aggregates at the neuron's levels. But whilst it is well known that the aggregates are constituted by proteins and nucleic acids, it is still unclear what their mechanism of formation entails or whether the aggregation trigger could be used as a potential therapeutic target[223]. As of 2020, the year I originally started my work, only a couple of papers looked at the role of nucleic acids in the aggregation process – perceiving it as a secondary stabilisation effect to that of proteins.

My PhD work aimed to explore the mechanism of pathological aggregation in ALS/FTD, currently considered deadly diseases, with particular focus on the role of RNA aggregation. In particular, my objective was to highlight how the repeat expansion sequence $(GGGGCC)_n$, which is the most common mutation in ALS/FTD[48], could aggregate *via* formation of multimeric nucleic acid-only networks by arranging into multimeric G-quadruplexes structures, offering an alternative to the narrative that sees proteins as aggregation triggers. To demonstrate this, I strived to achieve three main goals: **(1)** proving that $(GGGGCC)_n$ can aggregate via formation of mG4s in absence of additional proteins, **(2)** showing that these aggregates can be disassembled by tuning the molecularity of the system and **(3)** confirming that $(GGGGCC)_n$ mG4s can interact with ALS/FTD relevant proteins and are present in neurons (***Figure 56***).

Throughout my studies, I have successfully optimised a multimolecular G-quadruplexes-annealing protocol to probe the formation of these secondary structures *in vitro* for the $(GGGGCC)_n$ sequence. I then subjected the hexanucleotide repeat expansion to an extensive biophysical characterisation to confirm its potential to arrange into multimolecular G-quadruplexes and to test the sequence's aggregation abilities. Indeed, $(GGGGCC)_n$ (both in DNA and RNA form) was shown to aggregate *in vitro* in absence of proteins with a length dependence reminiscent of the pathology itself. I then successfully related the aggregation event to the formation of multimeric G-quadruplexes *via* a series of indirect (*e.g.,* melting experiments, mutation analysis) and direct (NMM-staining, G4-autofluorescence) experiments.

The extensive biophysical characterisation of the $(GGGGCC)_n$ system allowed me to understand that disassembly of the aggregates could be observed by tuning the molecularity of the species in solution. Consequently, $(GGGGCC)_n$ aggregates were disassembled first by Pyridostatin addition, a G4-ligand which shifted the system to a preferential unimolecular G4 arrangement, and then *via* antisense treatment, which allowed the mG4s to reassemble into double stranded DNA. Importantly, both G4-

ligand and antisense treatments have been shown to ameliorate ALS/FTD pathological inclusions, drawing a clear parallel with their ability to disassemble mG4-based aggregates.

In the last part of my studies, I explored the biological relevance of my system by introducing the protein level of complexity and by assessing the presence of G4s in iPSC ALS/FTD neurons. Firstly, TDP-43, an ALS/FTD relevant protein, was shown to interact with $(GGGGCC)_{11}$ *via* kinetics experiments, demonstrating that the presence of both biomolecules increased aggregation. I then illustrated how incubating TDP-43 in presence of mG4-$(GGGGCC)_{11}$ led to an increase in aggregation extent (quantitative confocal microscopy) and stability (NMM-photooxidation) with respect to the DNA/RNA only aggregates, suggesting that both biomolecules were co-operating in the forming the aggregates. As both TDP-43 and $(GGGGCC)_n$ have been previously observed in pathological aggregates in ALS/FTD patients' neurons, this synergistic effect is likely to be relevant in the disease context.

Indeed, in order to confirm mG4s relevance in the ALS/FTD system, I visualised them in iPSC patient derived neuronal cells *via* high throughput microscopy by staining with a G4-specific probe (NMM), demonstrating the structures biological relevance in the pathology.

Importantly, these studies – albeit promising in showing how mG4s are related to ALS/FTD, do present some limitations.

Firstly, my biophysical characterisation of the $(GGGGCC)_n$ sequence is limited to (n = 12), which is below the pathological threshold recognised for ALS/FTD development (n > 24). This comes from limitations of solid-state synthesis for the production of repetitive, G-rich oligonucleotides from commercial providers. This limitation was addressed by employing high annealing concentrations and high concentrations of salts for shorter repeat lengths n < 13 to trigger aggregation. Indeed, as aggregation occurred with a direct proportionality between repeat length n and annealing concentration, even though n > 24 could not be examined, the $(GGGGCC)_n$ aggregation behaviour could be approximated by using high annealing concentrations for shorter repeat lengths. Further studies should be performed to confirm $(GGGGCC)_{n>24}$ ability to arrange into mG4s under conditions which are more physiologically relevant (lower DNA and salt concentrations). As solid-state synthesis is not an option, alternative approaches could be employed for the purpose, perhaps utilising an enzymatic approach to link individual shorter strands together or by purifying the repeat directly from affected neurons.

Furthermore, while from my results there are no doubts that *in vitro* the $(GGGGCC)_n$ can arrange into mG4s, the experiments in live cells only offer a proof-of-concept that G4s are present in the cytoplasmatic inclusions. Whether these can be disassembled *via* G4-specific tools such as the ones

employed for the *in vitro* studies or if said tools are able to cross the blood-brain-barrier, is yet to be explored. An expansion of this study would offer a deeper understanding as to whether mG4s can be used as therapeutic targets for ALS/FTD.

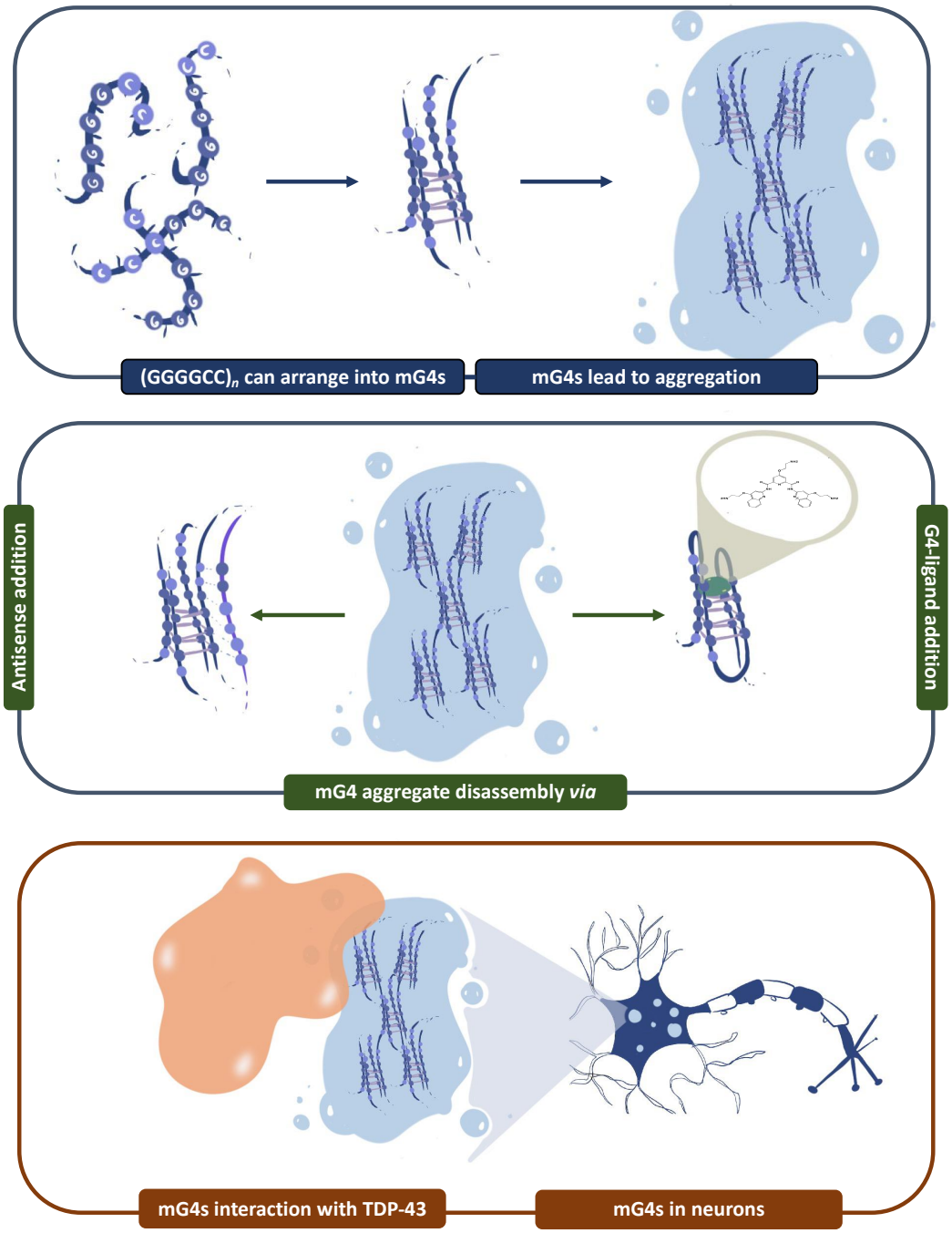


Figure 56: The three main findings of the study.

Overall, my studies demonstrate mG4 biological relevance in the context of neurodegenerative diseases and, in general, highlight how these structures – often overlooked due to their long formation processes – might be relevant in live cells for other biological processes. Importantly, this brings forward the question of whether nucleic acid-driven mechanism of pathological aggregation *via* the formation of multimeric structures could be at the root for other neurodegenerative pathologies or if it is specific to ALS/FTD. Specifically, other NDs which are associated with repeat expansions such as ataxia (related to $(GAA)_n$ which is able to form intramolecular DNA triplex of supercoiled DNA *in vitro*)[92] or polyQ diseases (related to $(CAG)_n$ which can form hairpins)[93], should be tested to confirm the hypothesis.

Nevertheless, my PhD work constitutes the first step in understanding mG4s as a new target for ALS/FTD, offering new outlooks for what is now a deadly disease.

References

1. Neurological disorders affect millions globally: Who report [Internet]. *World Health Organization.* Available from: https://www.who.int/news/item/27-02-2007-neurological-disorders-affect-millions-globally-who-report
2. *Bioeconomy for Sustainable Development*. (Springer, 2020). doi:10.1007/978-981-13-9431-7.
3. Goetz, C. G. The History of Parkinson's Disease: Early Clinical Descriptions and Neurological Therapies. *Cold Spring Harb Perspect Med* **1**, a008862 (2011).
4. Helping You Understand ALS. [Internet]. *The ALS association.* Available from: https://www.als.org/understanding-als. [cited 2023 Dec 13]
5. Alzheimer's Disease and Dementia. [Internet]. Available from: https://alz.org/alzheimers-dementia/what-is-dementia/types-of-dementia/frontotemporal-dementia. [cited 2023 Dec 13]
6. Milestones. [Internet]. *Alzheimer's Disease and Dementia* https://alz.org/alzheimers-dementia/research_progress/milestones. [cited 2023 Dec 13]
7. Rice, D. P. *et al.* Prevalence, costs, and treatment of Alzheimer's disease and related dementia: a managed care perspective. *Am J Manag Care* **7**, 809–818 (2001).
8. Cost of brain disorders to the UK estimated at £112 billion a year | Imperial News | Imperial College London. [Internet]. *Imperial News.* Available from: https://www.imperial.ac.uk/news/126466/cost-brain-disorders-uk-estimated-112/ (2013). [cited 2023 Dec 13]
9. 2010 to 2015 government policy: cancer research and treatment. [Internet]. *GOV.UK. Available from:* https://www.gov.uk/government/publications/2010-to-2015-government-policy-cancer-research-and-treatment/2010-to-2015-government-policy-cancer-research-and-treatment. [cited 2023 Dec 13]
10. Health matters: preventing cardiovascular disease. [Internet]. *GOV.UK.* Available from: https://www.gov.uk/government/publications/health-matters-preventing-cardiovascular-disease/health-matters-preventing-cardiovascular-disease. [cited 2023 Dec 13]
11. The top 10 causes of death. [Internet]. Available from: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. [cited 2023 Dec 13]
12. Monthly mortality analysis, England and Wales. [Internet]. *Office for National Statistic*s. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/monthlymortalityanalysisenglandandwales/june2022. [cited 2023 Dec 13]

13. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* **18**, 459–480 (2019).
14. GBD Results. [Internet]. *Institute for Health Metrics and Evaluation.* Available from: https://www.healthdata.org/data-visualization/gbd-results (2022). [cited 2023 Dec 13]
15. Ding, C. *et al.* Global, regional, and national burden and attributable risk factors of neurological disorders: The Global Burden of Disease study 1990–2019. *Frontiers in Public Health* **10**, (2022).
16. Heled, Y., Rutschman, A. S. & Vertinsky, L. The problem with relying on profit-driven models to produce pandemic drugs. *J Law Biosci* **7**, lsaa060 (2020).
17. Brown, R. C., Lockwood, A. H. & Sonawane, B. R. Neurodegenerative Diseases: An Overview of Environmental Risk Factors. *Environ Health Perspect* **113**, 1250–1256 (2005).
18. Mullard, A. Controversial Alzheimer's drug approval could affect other diseases. *Nature* **595**, 162–163 (2021).
19. Aduhelm: Withdrawn application. European Medicines Agency. [Internet]. *EMA*. Available from: https://www.ema.europa.eu/en/medicines/human/withdrawn-applications/aduhelm (2022). [cited 2023 Dec 13]
20. Ferrari, R., Kapogiannis, D., Huey, E. D. & Momeni, P. FTD and ALS: a tale of two diseases. *Curr Alzheimer Res* **8**, 273–294 (2011).
21. Achi, E. Y. & Rudnicki, S. A. ALS and Frontotemporal Dysfunction: A Review. *Neurology Research International* vol. 2012 e806306 https://www.hindawi.com/journals/nri/2012/806306/ (2012).
22. Abramzon, Y. A., Fratta, P., Traynor, B. J. & Chia, R. The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Front. Neurosci.* **14**, (2020).
23. Chiò, A. *et al.* Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. *Neuroepidemiology* **41**, 118–130 (2013).
24. Ratnavalli, E., Brayne, C., Dawson, K. & Hodges, J. R. The prevalence of frontotemporal dementia. *Neurology* **58**, 1615–1621 (2002).
25. Gladman, M. & Zinman, L. The economic impact of amyotrophic lateral sclerosis: a systematic review. *Expert Rev Pharmacoecon Outcomes Res* **15**, 439–450 (2015).
26. Galvin, J. E., Howard, D. H., Denny, S. S., Dickinson, S. & Tatton, N. The social and economic burden of frontotemporal degeneration. *Neurology* **89**, 2049–2056 (2017).
27. Ice Bucket Challenge dramatically accelerated the fight against ALS. [Internet]. *The ALS Association.* Available from: https://www.als.org/stories-news/ice-bucket-challenge-dramatically-accelerated-fight-against-als. [cited 2023 Dec 13]

28. Javaid, S. F., Giebel, C., Khan, M. A. & Hashim, M. J. Epidemiology of Alzheimer's disease and other dementias: rising global burden and forecasted trends. Preprint at https://doi.org/10.12688/f1000research.50786.1 (2021).
29. Clinical investigation of medicines for the treatment Alzheimer's disease - Scientific guideline. [Internet]. *EMA*. Available from: *European Medicines Agency* https://www.ema.europa.eu/en/clinical-investigation-medicines-treatment-alzheimers-disease-scientific-guideline (2018). [cited 2023 Dec 13]
31. Madeira, J. M., Schindler, S. M. & Klegeris, A. A new look at auranofin, dextromethorphan and rosiglitazone for reduction of glia-mediated inflammation in neurodegenerative diseases. *Neural Regen Res* **10**, 391–393 (2015).
32. Čolović, M. B., Krstić, D. Z., Lazarević-Pašti, T. D., Bondžić, A. M. & Vasić, V. M. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Curr Neuropharmacol* **11**, 315–335 (2013).
33. Ajenikoko, M. K., Ajagbe, A. O., Onigbinde, O. A., Okesina, A. A. & Tijani, A. A. Review of Alzheimer's disease drugs and their relationship with neuron-glia interaction. *IBRO Neurosci Rep* **14**, 64–76 (2022).
34. Padda, I. S. & Parmar, M. Aducanumab. in *StatPearls* (StatPearls Publishing, 2023).
35. Reardon, S. FDA approves Alzheimer's drug lecanemab amid safety concerns. *Nature* **613**, 227–228 (2023).
36. Liscic, R. M., Alberici, A., Cairns, N. J., Romano, M. & Buratti, E. From basic research to the clinic: innovative therapies for ALS and FTD in the pipeline. *Molecular Neurodegeneration* **15**, 31 (2020).
37. Tran, H. *et al.* SUPPRESSION OF MUTANT C9ORF72 EXPRESSION BY A POTENT MIXED BACKBONE ANTISENSE OLIGONUCLEOTIDE. *Nat Med* **28**, 117–124 (2022).
38. Lagier-Tourenne, C. *et al.* Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc Natl Acad Sci U S A* **110**, E4530–E4539 (2013).
39. Liu, Y. *et al.* Variant-selective stereopure oligonucleotides protect against pathologies associated with C9orf72-repeat expansion in preclinical models. *Nat Commun* **12**, 847 (2021).
40. Liu, Y. *et al.* Preclinical evaluation of WVE-004, an investigational stereopure oligonucleotide for the treatment of C9orf72-associated ALS or FTD. *Molecular Therapy - Nucleic Acids* **28**, 558–570 (2022).
41. Wilson, D. M. *et al.* Hallmarks of neurodegenerative diseases. *Cell* **186**, 693–714 (2023).

42. Ramos-Cejudo, J. *et al.* Traumatic Brain Injury and Alzheimer's Disease: The Cerebrovascular Link. *EBioMedicine* **28**, 21–30 (2018).
43. Kumar, D. R., Aslinia, F., Yale, S. H. & Mazza, J. J. Jean-Martin Charcot: The Father of Neurology. *Clin Med Res* **9**, 46–49 (2011).
44. Bede, P., Iyer, P. M., Finegan, E., Omer, T. & Hardiman, O. Virtual brain biopsies in amyotrophic lateral sclerosis: Diagnostic classification based on in vivo pathological patterns. *Neuroimage Clin* **15**, 653–658 (2017).
45. Wijesekera, L. C. & Nigel Leigh, P. Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases* **4**, 3 (2009).
46. Mejzini, R. *et al.* ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? *Frontiers in Neuroscience* **13**, (2019).
47. Kertesz, A. & Kalvach, P. Arnold Pick and German Neuropsychiatry in Prague. *Archives of Neurology* **53**, 935–938 (1996).
48. Ljubenkov, P. A. & Miller, B. L. A Clinical Guide to Frontotemporal Dementias. *Focus (Am Psychiatr Publ)* **14**, 448–464 (2016).
49. Schroeter, M. L., Raczka, K., Neumann, J. & von Cramon, D. Y. Neural networks in frontotemporal dementia--a meta-analysis. *Neurobiol Aging* **29**, 418–426 (2008).
50. Familial FTD. *Memory and Aging Center* https://memory.ucsf.edu/genetics/familial-ftd.
51. Cirulli, E. T. *et al.* Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436–1441 (2015).
52. Bannwarth, S. *et al.* A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain* **137**, 2329–2345 (2014).
53. Johnson, J. O. *et al.* Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68**, 857–864 (2010).
54. Fecto, F. *et al.* SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol* **68**, 1440–1446 (2011).
55. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
56. Iacoangeli, A. *et al.* C9orf72 intermediate expansions of 24–30 repeats are associated with ALS. *Acta Neuropathologica Communications* **7**, 115 (2019).
57. Pastore, A. & Temussi, P. Protein aggregation and misfolding: good or evil? *J. Phys.: Condens. Matter* **24**, 244101 (2012).
58. Lee, J., Cho, H. & Kwon, I. Phase separation of low-complexity domains in cellular function and disease. *Exp Mol Med* **54**, 1412–1422 (2022).

59. Mukherjee, A., Morales-Scheihing, D., Butler, P. C. & Soto, C. Type 2 Diabetes as a Protein Misfolding Disease. *Trends Mol Med* **21**, 439–449 (2015).
60. Saelices, L. *et al.* Uncovering the Mechanism of Aggregation of Human Transthyretin. *J Biol Chem* **290**, 28932–28943 (2015).
61. Schmidt, M. F., Gan, Z. Y., Komander, D. & Dewson, G. Ubiquitin signalling in neurodegeneration: mechanisms and therapeutic opportunities. *Cell Death Differ* **28**, 570–590 (2021).
62. Finkbeiner, S. The Autophagy Lysosomal Pathway and Neurodegeneration. *Cold Spring Harb Perspect Biol* **12**, a033993 (2020).
63. Bajc Česnik, A., Motaln, H. & Rogelj, B. The Impact of ALS-Associated Genes hnRNPA1, MATR3, VCP and UBQLN2 on the Severity of TDP-43 Aggregation. *Cells* **9**, 1791 (2020).
64. Quinn, P. M. J., Moreira, P. I., Ambrósio, A. F. & Alves, C. H. PINK1/PARKIN signalling in neurodegeneration and neuroinflammation. *Acta Neuropathologica Communications* **8**, 189 (2020).
65. Levine, B. & Kroemer, G. Biological Functions of Autophagy Genes: A Disease Perspective. *Cell* **176**, 11–42 (2019).
66. Sheng, J., Olrichs, N. K., Gadella, B. M., Kaloyanova, D. V. & Helms, J. B. Regulation of Functional Protein Aggregation by Multiple Factors: Implications for the Amyloidogenic Behavior of the CAP Superfamily Proteins. *Int J Mol Sci* **21**, 6530 (2020).
67. McDade, E. *et al.* Lecanemab in patients with early Alzheimer's disease: detailed results on biomarker, cognitive, and clinical effects from the randomized and open-label extension of the phase 2 proof-of-concept study. *Alzheimers Res Ther* **14**, 191 (2022).
68. Coppedè, F. & Migliore, L. DNA damage in neurodegenerative diseases. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **776**, 84–97 (2015).
69. Welch, G. & Tsai, L. Mechanisms of DNA damage-mediated neurotoxicity in neurodegenerative disease. *EMBO Rep* **23**, e54217 (2022).
70. Bourseguin, J. *et al.* Persistent DNA damage associated with ATM kinase deficiency promotes microglial dysfunction. *Nucleic Acids Research* **50**, 2700–2718 (2022).
71. Gregersen, L. H. & Svejstrup, J. Q. The Cellular Response to Transcription-Blocking DNA Damage. *Trends Biochem Sci* **43**, 327–341 (2018).
72. Leandro, G. S., Sykora, P. & Bohr, V. A. The Impact of Base Excision DNA Repair in Age-Related Neurodegenerative Diseases. *Mutat Res* **776**, 31–39 (2015).
73. Stott, R. T., Kritsky, O. & Tsai, L.-H. Profiling DNA break sites and transcriptional changes in response to contextual fear learning. *PLoS One* **16**, e0249691 (2021).

74. Madabhushi, R. *et al.* Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **161**, 1592–1605 (2015).
75. Institute, U. D. U. D. R. Targeting DNA repair. [Internet]. *UK DRI: UK Dementia Research Institute.* Available from: https://ukdri.ac.uk/news-and-events/targeting-dna-repair-for-dementia-treatments (2023). [cited 2023 Dec 13]
76. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
77. Arnold, E. S. *et al.* ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. *Proceedings of the National Academy of Sciences* **110**, E736–E745 (2013).
78. Zhou, Y., Liu, S., Öztürk, A. & Hicks, G. G. FUS-regulated RNA metabolism and DNA damage repair. *Rare Dis* **2**, e29515 (2014).
79. Protter, D. S. W. & Parker, R. Principles and Properties of Stress granules. *Trends Cell Biol* **26**, 668–679 (2016).
80. Gutiérrez-Garcia, R. *et al.* G3BP1-dependent mechanism suppressing protein aggregation in Huntington's models and its demise upon stress granule assembly. *Hum Mol Genet* **32**, 1607–1621 (2023).
81. Repici, M. *et al.* The Parkinson's Disease-Linked Protein DJ-1 Associates with Cytoplasmic mRNP Granules During Stress and Neurodegeneration. *Mol Neurobiol* **56**, 61–77 (2019).
82. Wolozin, B. & Ivanov, P. Stress granules and neurodegeneration. *Nature Reviews Neuroscience* **20**, 649–666 (2019).
83. Paulson, H. Repeat expansion diseases. *Handb Clin Neurol* **147**, 105–123 (2018).
84. Lieberman, A. P., Shakkottai, V. G. & Albin, R. L. Polyglutamine Repeats in Neurodegenerative Diseases. *Annu Rev Pathol* **14**, 1–27 (2019).
85. Ayala, Y. M. & Nguyen, A. D. RNA-Based Therapies for Neurodegenerative Diseases. *Mo Med* **118**, 340–345 (2021).
86. Minakawa, E. N. & Nagai, Y. Protein Aggregation Inhibitors as Disease-Modifying Therapies for Polyglutamine Diseases. *Frontiers in Neuroscience* **15**, (2021).
87. Shoubridge, C. & Gecz, J. POLYALANINE TRACT DISORDERS AND NEUROCOGNITIVE PHENOTYPES. in *Madame Curie Bioscience Database [Internet]* (Landes Bioscience, 2013). [cited 2023 Dec 13]
88. Belzil, V. V., Gendron, T. F. & Petrucelli, L. RNA-mediated toxicity in neurodegenerative disease. *Mol Cell Neurosci* **0**, 406–419 (2013).

89. Warraich, S. T., Yang, S., Nicholson, G. A. & Blair, I. P. TDP-43: a DNA and RNA binding protein with roles in neurodegenerative diseases. *Int J Biochem Cell Biol* **42**, 1606–1609 (2010).
90. Nakaya, T., Alexiou, P., Maragkakis, M., Chang, A. & Mourelatos, Z. FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA* **19**, 498–509 (2013).
91. Shmookler Reis, R. J. *et al.* "Protein aggregates" contain RNA and DNA, entrapped by misfolded proteins but largely rescued by slowing translational elongation. *Aging Cell* **20**, e13326 (2021).
92. Khristich, A. N., Armenia, J. F., Matera, R. M., Kolchinski, A. A. & Mirkin, S. M. Large-scale contractions of Friedreich's ataxia GAA repeats in yeast occur during DNA replication due to their triplex-forming ability. *Proceedings of the National Academy of Sciences* **117**, 1628–1637 (2020).
93. Xu, P., Pan, F., Roland, C., Sagui, C. & Weninger, K. Dynamics of strand slippage in DNA hairpins formed by CAG repeats: roles of sequence parity and trinucleotide interrupts. *Nucleic Acids Res* **48**, 2232–2245 (2020).
94. Geng, Y. *et al.* Crystal structure of parallel G-quadruplex formed by the two-repeat ALS- and FTD-related GGGGCC sequence. *Nucleic Acids Res* **49**, 5881–5890 (2021).
95. Strong, M. J., Donison, N. S. & Volkening, K. Alterations in Tau Metabolism in ALS and ALS-FTSD. *Front Neurol* **11**, 598907 (2020).
96. Rhinn, H., Tatton, N., McCaughey, S., Kurnellas, M. & Rosenthal, A. Progranulin as a therapeutic target in neurodegenerative diseases. *Trends in Pharmacological Sciences* **43**, 641–652 (2022).
97. Nolan, M., Talbot, K. & Ansorge, O. Pathogenesis of FUS-associated ALS and FTD: insights from rodent models. *Acta Neuropathologica Communications* **4**, 99 (2016).
98. 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).
99. Koehler, L. C. *et al.* TDP-43 Oligomerization and Phase Separation Properties Are Necessary for Autoregulation. *Frontiers in Neuroscience* **16**, (2022).
100. Kuo, P.-H., Doudeva, L. G., Wang, Y.-T., Shen, C.-K. J. & Yuan, H. S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res* **37**, 1799–1808 (2009).
101. 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).
102. Gijselinck, I. *et al.* Neuronal inclusion protein TDP-43 has no primary genetic role in FTD and ALS. *NEUROBIOLOGY OF AGING* **30**, 1329–1331 (2009).
103. Smeyers, J., Banchi, E.-G. & Latouche, M. C9ORF72: What It Is, What It Does, and Why It Matters. *Frontiers in Cellular Neuroscience* **15**, (2021).

104. Koppers, M. *et al.* VCP mutations in familial and sporadic amyotrophic lateral sclerosis. *Neurobiol Aging* **33**, 837.e7–13 (2012).
105. Bhogaraju, S. *et al.* Molecular Basis of Tubulin Transport Within the Cilium by IFT74 and IFT81. *Science* **341**, 1009–1012 (2013).
106. Rizzu, P. *et al.* C9orf72 is differentially expressed in the central nervous system and myeloid cells and consistently reduced in C9orf72, MAPT and GRN mutation carriers. *Acta Neuropathologica Communications* **4**, 37 (2016).
107. Ferreira, P. A. The coming-of-age of nucleocytoplasmic transport in motor neuron disease and neurodegeneration. *Cell Mol Life Sci* **76**, 2247–2273 (2019).
108. Babić Leko, M. *et al.* Molecular Mechanisms of Neurodegeneration Related to C9orf72 Hexanucleotide Repeat Expansion. *Behav Neurol* **2019**, 2909168 (2019).
109. Freibaum, B. D. & Taylor, J. P. The Role of Dipeptide Repeats in C9ORF72-Related ALS-FTD. *Front Mol Neurosci* **10**, 35 (2017).
110. Andrade, N. S. *et al.* Dipeptide repeat proteins inhibit homology-directed DNA double strand break repair in C9ORF72 ALS/FTD. *Molecular Neurodegeneration* **15**, 13 (2020).
111. Brčić, J. & Plavec, J. ALS and FTD linked GGGGCC-repeat containing DNA oligonucleotide folds into two distinct G-quadruplexes. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1861**, 1237–1245 (2017).
112. 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).
113. Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200 (2014).
114. Mizielinska, S. *et al.* C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol* **126**, 845–857 (2013).
115. Fratta, P. *et al.* C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* **2**, 1016 (2012).
116. Belzil, V. V. *et al.* Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathol* **126**, 895–905 (2013).
117. Wang, T. *et al.* C9orf72 regulates energy homeostasis by stabilizing mitochondrial complex I assembly. *Cell Metab* **33**, 531-546.e9 (2021).
118. Wang, G. & Vasquez, K. M. Impact of Alternative DNA Structures on DNA Damage, DNA Repair, and Genetic Instability. *DNA Repair (Amst)* **19**, 143–151 (2014).

119. Dai, Y., Teng, X., Zhang, Q., Hou, H. & Li, J. Advances and challenges in identifying and characterizing G-quadruplex–protein interactions. *Trends in Biochemical Sciences* **48**, 894–909 (2023).
120. Kim, N. G4-interacting proteins endangering genomic stability at G4 DNA-forming sites. *Biochem Soc Trans* **51**, 403–413 (2023).
121. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
122. Malouf, L. *et al.* Sculpting DNA-based synthetic cells through phase separation and phase-targeted activity. 2023.03.17.533162 Preprint at https://doi.org/10.1101/2023.03.17.533162 (2023).
123. Walczak, M. *et al.* A synthetic signalling network imitating the action of immune cells in response to bacterial metabolism. 2023.02.02.526524 Preprint at https://doi.org/10.1101/2023.02.02.526524 (2023).
124. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nat Rev Mater* **3**, 1–23 (2017).
125. Platt, J. R. POSSIBLE SEPARATION OF INTERTWINED NUCLEIC ACID CHAINS BY TRANSFER-TWIST. *Proc Natl Acad Sci U S A* **41**, 181–183 (1955).
126. Panayotatos, N. & Wells, R. D. Cruciform structures in supercoiled DNA. *Nature* **289**, 466–470 (1981).
127. Svoboda, P. & Cara, A. Di. Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.* **63**, 901–908 (2006).
128. Hoogsteen, K. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Cryst* **12**, 822–823 (1959).
129. Bansal, A., Kaushik, S. & Kukreti, S. Non-canonical DNA structures: Diversity and disease association. *Frontiers in Genetics* **13**, (2022).
130. Gellert, M., Lipsett, M. N. & Davies, D. R. HELIX FORMATION BY GUANYLIC ACID. *Proc Natl Acad Sci U S A* **48**, 2013–2018 (1962).
131. 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).
132. 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).
133. Hardin, C. C., Perry, A. G. & White, K. Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* **56**, 147–194 (2000).
134. Puig Lombardi, E. & Londoño-Vallejo, A. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res* **48**, 1–15 (2020).

135. Sahakyan, A. B. *et al.* Machine learning model for sequence-driven DNA G-quadruplex formation. *Scientific Reports* **7**, 14535 (2017).
136. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* **33**, 877–881 (2015).
137. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods* **13**, 841–844 (2016).
138. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nature Protocols* **13**, 551–564 (2018).
139. Li, Q. *et al.* G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Research* **41**, D1115–D1123 (2013).
140. 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).
141. Rodriguez, R. *et al.* A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *J. Am. Chem. Soc.* **130**, 15758–15759 (2008).
142. Liu, L.-Y., Ma, T.-Z., Zeng, Y.-L., Liu, W. & Mao, Z.-W. Structural Basis of Pyridostatin and Its Derivatives Specifically Binding to G-Quadruplexes. *J. Am. Chem. Soc.* **144**, 11878–11887 (2022).
143. Ghosh, A., Trajkovski, M., Teulade-Fichou, M.-P., Gabelica, V. & Plavec, J. Phen-DC3 Induces Refolding of Human Telomeric DNA into a Chair-Type Antiparallel G-Quadruplex through Ligand Intercalation. *Angewandte Chemie International Edition* **61**, e202207384 (2022).
144. Di Antonio, M. *et al.* Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* **12**, 832–837 (2020).
145. Varizhuk, A. *et al.* The expanding repertoire of G4 DNA structures. *Biochimie* **135**, 54–62 (2017).
146. Teixeira, R. *et al.* Fluorescence Spectroscopy of Porphyrins and Phthalocyanines: Some Insights into Supramolecular Self-Assembly, Microencapsulation, and Imaging Microscopy. *Molecules* **26**, 4264 (2021).
147. Shivalingam, A. *et al.* The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat Commun* **6**, 8178 (2015).
148. Summers, P. A. *et al.* Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat Commun* **12**, 162 (2021).
149. Liano, D., Monti, L., Chowdhury, S., Raguseo, F. & Antonio, M. D. Long-range DNA interactions: inter-molecular G-quadruplexes and their potential biological relevance. *Chem. Commun.* (2022) doi:10.1039/D2CC04872H.

150. Kolesnikova, S. & Curtis, E. A. Structure and Function of Multimeric G-Quadruplexes. *Molecules* **24**, 3074 (2019).
151. Merkina, E. E. & Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophysical Journal* **89**, 365–373 (2005).
152. Moye, A. L. *et al.* Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat Commun* **6**, 7643 (2015).
153. Liano, D. & Antonio, M. D. Cockayne Syndrome B protein selectively interacts and resolves intermolecular DNA G-quadruplex structures. *bioRxiv* 2021.03.25.436565 (2021) doi:10.1101/2021.03.25.436565.
154. 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).
155. 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).
156. 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).
157. Sabharwal, N. C. *et al.* N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J* **281**, 1726–1737 (2014).
158. 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).
159. 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).
160. Hou, Y. *et al.* Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics* **14**, 894–911 (2019).
161. Zhang, Y. *et al.* G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Research* **47**, 11746–11754 (2019).
162. Chowdhury, S., Wang, J., Nuccio, S. P., Mao, H. & Di Antonio, M. Short LNA-modified oligonucleotide probes as efficient disruptors of DNA G-quadruplexes. *Nucleic Acids Research* **50**, 7247–7259 (2022).
163. Miglietta, G., Marinello, J., Russo, M. & Capranico, G. Ligands stimulating antitumour immunity as the next G-quadruplex challenge. *Molecular Cancer* **21**, 180 (2022).
164. 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).

165. Conlon, E. G. *et al.* The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *eLife* **5**, e17820 (2016).
166. Ishiguro, A., Lu, J., Ozawa, D., Nagai, Y. & Ishihama, A. ALS-linked FUS mutations dysregulate G-quadruplex-dependent liquid–liquid phase separation and liquid-to-solid transition. *Journal of Biological Chemistry* **297**, (2021).
167. McEachin, Z. T., Parameswaren, J., Raj, N., Bassell, G. J. & Jiang, J. RNA-mediated toxicity in C9orf72 ALS and FTD. *Neurobiol Dis* **145**, 105055 (2020).
168. Lashuel, H. A. Rethinking protein aggregation and drug discovery in neurodegenerative diseases: Why we need to embrace complexity? *Current Opinion in Chemical Biology* **64**, 67–75 (2021).
169. 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).
170. 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).
171. 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).
172. Merkina, E. E. & Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophysical Journal* **89**, 365–373 (2005).
173. *Handbook Of Biological Confocal Microscopy*. (Springer US, 2006). doi:10.1007/978-0-387-45524-2.
174. Sprague, B. L. & McNally, J. G. FRAP analysis of binding: proper and fitting. *Trends Cell Biol* **15**, 84–91 (2005).
175. Tantray, J. A., Mansoor, S., Wani, R. F. C. & Nissa, N. U. Chapter 24 - Agarose gel electrophoresis. in *Basic Life Science Methods* (eds. Tantray, J. A., Mansoor, S., Wani, R. F. C. & Nissa, N. U.) 103–106 (Academic Press, 2023). doi:10.1016/B978-0-443-19174-9.00024-6.
176. Hall, A. C. A comparison of DNA stains and staining methods for Agarose Gel Electrophoresis. 568253 Preprint at https://doi.org/10.1101/568253 (2020).
177. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res* **36**, 5482–5515 (2008).
178. Circular Dichroism: Principles and Applications, 2nd Edition | Wiley.
179. Woody, R. W. [4] Circular dichroism. in *Methods in Enzymology* vol. 246 34–71 (Academic Press, 1995).

180. Bishop, G. R. & Chaires, J. B. Characterization of DNA structures by circular dichroism. *Curr Protoc Nucleic Acid Chem* **Chapter 7**, 7.11.1-7.11.8 (2003).
181. Do, S., Lee, C., Lee, T., Kim, D.-N. & Shin, Y. Engineering DNA-based synthetic condensates with programmable material properties, compositions, and functionalities. *Science Advances* **8**, eabj1771 (2022).
182. Kypr, J., Kejnovská, I., Renčiuk, D. & Vorlíčková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res* **37**, 1713–1725 (2009).
183. 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).
184. DNA Stains. [Internet]. *ThermoFisher.* Available from: https://www.thermofisher.com/uk/en/home/life-science/dna-rna-purification-analysis/nucleic-acid-gel-electrophoresis/dna-stains.html. [cited 2023 Dec 13]
185. Zheng, W., Li, Y., Zhao, L., Li, C. & Wang, L. Label-free fluorescence aptasensor for chloramphenicol based on hybridization chain reaction amplification and G-quadruplex/N-methyl mesoporphyrin IX complexation.
187. Spectrum [Ethidium Bromide]. [Internet]. *AAT Bioquest.* Available from: https://www.aatbio.com/fluorescence-excitation-emission-spectrum-graph-viewer/ethidium_bromide. [cited 2023 Dec 13]
188. You, J. *et al.* Effects of monovalent cations on folding kinetics of G-quadruplexes. *Bioscience Reports* **37**, BSR20170771 (2017).
189. Farrawell, N. E. *et al.* Distinct partitioning of ALS associated TDP-43, FUS and SOD1 mutants into cellular inclusions. *Sci Rep* **5**, 13416 (2015).
190. 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).
191. 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).
192. Kolenc, O. I. & Quinn, K. P. Evaluating Cell Metabolism Through Autofluorescence Imaging of NAD(P)H and FAD. *Antioxid Redox Signal* **30**, 875–889 (2019).
193. Vivian, J. T. & Callis, P. R. Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophysical Journal* **80**, 2093–2109 (2001).
194. Zuffo, M., Gandolfini, A., Heddi, B. & Granzhan, A. Harnessing intrinsic fluorescence for typing of secondary structures of DNA. *Nucleic Acids Res* **48**, e61 (2020).

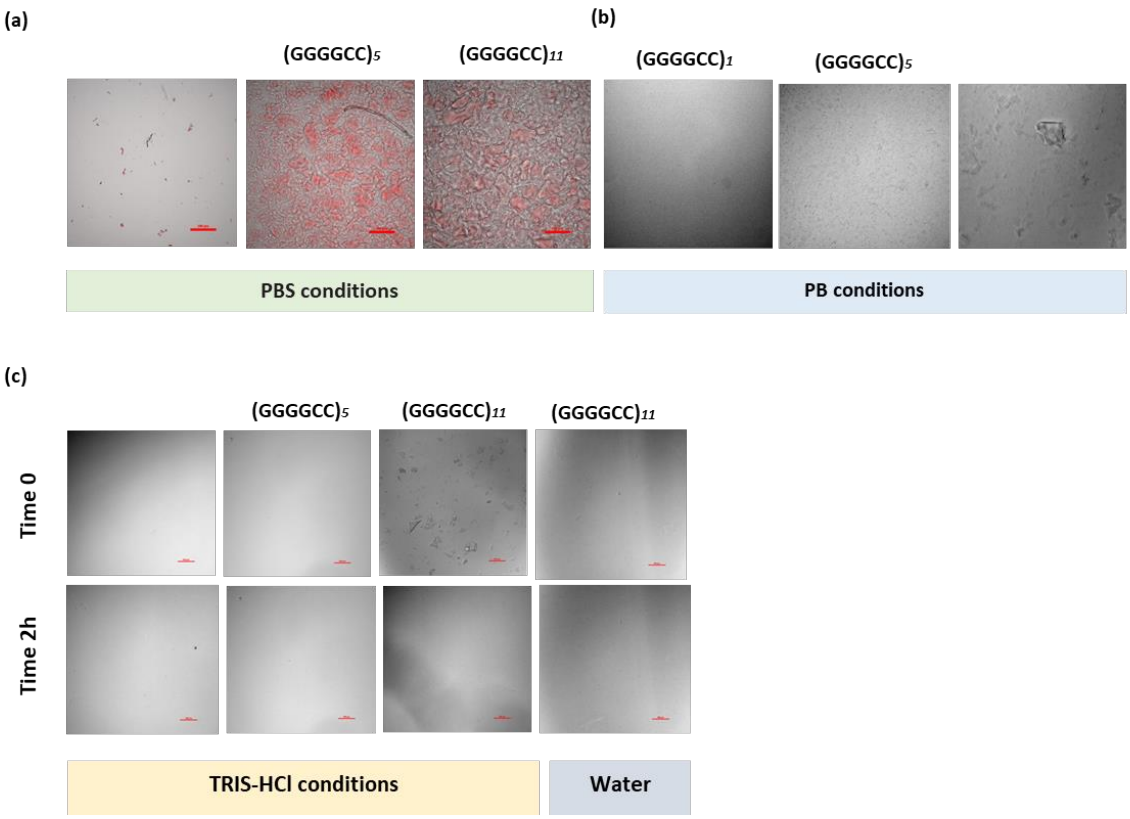
195. Merkina, E. E. & Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophysical Journal* **89**, 365–373 (2005).
196. Vallone, P. M. *et al.* Melting studies of short DNA hairpins: influence of loop sequence and adjoining base pair identity on hairpin thermodynamic stability. *Biopolymers* **50**, 425–442 (1999).
197. Barone, F., Chirico, G., Matzeu, M., Mazzei, F. & Pedone, F. Triple helix DNA oligomer melting measured by fluorescence polarization anisotropy. *Eur Biophys J* **27**, 137–146 (1998).
198. Definition of mutation analysis - NCI Dictionary of Genetics Terms. [Internet]. *NCI*. Available from: https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/mutation-analysis (2012). [cited 2023 Dec 13]
199. 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).
200. 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).
201. Crooke, S. T. Molecular Mechanisms of Antisense Oligonucleotides. *Nucleic Acid Ther* **27**, 70–77 (2017).
202. Kikin, O., D'Antonio, L. & Bagga, P. S. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research* **34**, W676–W682 (2006).
203. 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).
204. Guédin, A., Gros, J., Alberti, P. & Mergny, J.-L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res* **38**, 7858–7868 (2010).
205. Tariq, N. *et al.* Dimethyl sulfoxide (DMSO) is a stabilizing co-solvent for G-quadruplex DNA. *Biophysical Chemistry* **282**, 106741 (2022).
207. Biochemistry of Bacterial Growth, 3rd Edition: Edited by J Mandelstam, K McQuillen and I Dawes. pp 449. Blackwell Scientific Publications.
208. Tran, H. *et al.* Suppression of mutant C9orf72 expression by a potent mixed backbone antisense oligonucleotide. *Nat Med* **28**, 117–124 (2022).
209. 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).
210. Fernandes, S. A., Douglas, A. G. L., Varela, M. A., Wood, M. J. A. & Aoki, Y. Oligonucleotide-Based Therapy for FTD/ALS Caused by the C9orf72 Repeat Expansion: A Perspective. *J Nucleic Acids* **2013**, 208245 (2013).

211. Funaki, M. & Janmey, P. A. Chapter 23 - Technologies to Engineer Cell Substrate Mechanics in Hydrogels. in *Biology and Engineering of Stem Cell Niches* (eds. Vishwakarma, A. & Karp, J. M.) 363–373 (Academic Press, 2017). doi:10.1016/B978-0-12-802734-9.00023-8.
212. Nowakowski, A. B., Wobig, W. J. & Petering, D. H. Native SDS-PAGE: High Resolution Electrophoretic Separation of Proteins With Retention of Native Properties Including Bound Metal Ions. *Metallomics* **6**, 1068–1078 (2014).
213. Ho, C. *et al.* Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin Biochem Rev* **24**, 3–12 (2003).
214. Oshinbolu, S. *et al.* Evaluation of fluorescent dyes to measure protein aggregation within mammalian cell culture supernatants. *J Chem Technol Biotechnol* **93**, 909–917 (2018).
215. ENZO Life Sciences. [Internet]. Proteostat Assay. Available from: https://www.enzolifesciences.com/ENZ-51023/proteostat-protein-aggregation-assay/. [cited 2023 Dec 13]
216. Reimer, L. & Kohl, H. *Transmission Electron Microscopy: Physics of Image Formation*. (Springer Science & Business Media, 2008).
217. Medvedev, S. P., Shevchenko, A. I. & Zakian, S. M. Induced Pluripotent Stem Cells: Problems and Advantages when Applying them in Regenerative Medicine. *Acta Naturae* **2**, 18–28 (2010).
218. Herzenberg, L. A. *et al.* The History and Future of the Fluorescence Activated Cell Sorter and Flow Cytometry: A View from Stanford. *Clinical Chemistry* **48**, 1819–1827 (2002).
219. Guo, J., Walss-Bass, C. & Ludueña, R. F. The β isotypes of tubulin in neuronal differentiation. *Cytoskeleton (Hoboken)* **67**, 431–441 (2010).
220. Siddiqui, S., Aamodt, E., Rastinejad, F. & Culotti, J. Anti-tubulin monoclonal antibodies that bind to specific neurons in Caenorhabditis elegans. *J Neurosci* **9**, 2963–2972 (1989).
221. Yang, D. G-Quadruplex DNA and RNA. *G-Quadruplex Nucleic Acids* **2035**, 1–24 (2019).
222. 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).
223. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat Med* **10 Suppl**, S10-17 (2004).
224. Berning, B. A. & Walker, A. K. The Pathobiology of TDP-43 C-Terminal Fragments in ALS and FTLD. *Frontiers in Neuroscience* **13**, (2019).
225. Zacco, E., Martin, S. R., Thorogate, R. & Pastore, A. The RNA-Recognition Motifs of TAR DNA-Binding Protein 43 May Play a Role in the Aberrant Self-Assembly of the Protein. *Frontiers in Molecular Neuroscience* **11**, (2018).

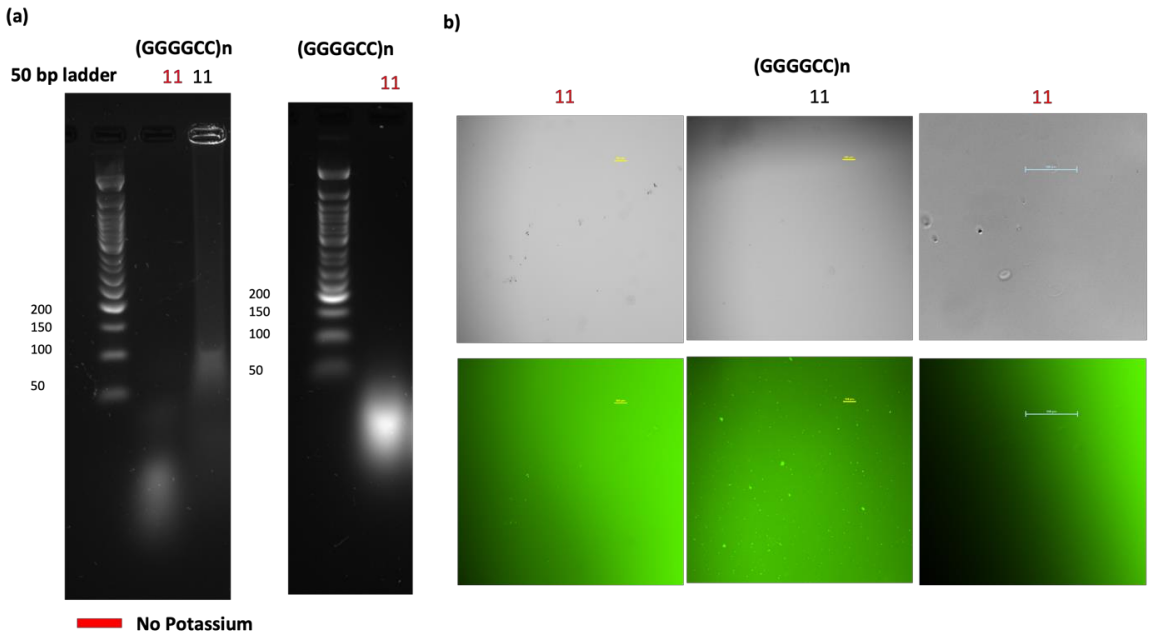
226. Furukawa, Y. & Nukina, N. Functional diversity of protein fibrillar aggregates from physiology to RNA granules to neurodegenerative diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1832**, 1271–1278 (2013).
227. Iannuzzi, C., Maritato, R., Irace, G. & Sirangelo, I. Misfolding and Amyloid Aggregation of Apomyoglobin. *International Journal of Molecular Sciences* **14**, 14287–14300 (2013).
228. Renaud de la Faverie, A., Guédin, A., Bedrat, A., Yatsunyk, L. A. & Mergny, J.-L. Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res* **42**, e65 (2014).
229. Minard, A. *et al.* A short peptide that preferentially binds c-MYC G-quadruplex DNA. *Chem. Commun.* **56**, 8940–8943 (2020).
230. Chen, R. F. & Knutson, J. R. Mechanism of fluorescence concentration quenching of carboxyfluorescein in liposomes: energy transfer to nonfluorescent dimers. *Anal Biochem* **172**, 61–77 (1988).
231. Arseni, D. *et al.* Structure of pathological TDP-43 filaments from ALS with FTLD. *Nature* **601**, 139–143 (2022).
232. Waldvogel, H. J. *et al.* The collection and processing of human brain tissue for research. *Cell Tissue Bank* **9**, 169–179 (2008).
233. Use of human tissue in research. [Internet]. *Health Research Authority.* Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/use-tissue-research/. [cited 2023 Dec 13]
234. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol* **27**, 275–280 (2009).
235. Karumbayaram, S. *et al.* Directed differentiation of human-induced pluripotent stem cells generates active motor neurons. *Stem Cells* **27**, 806–811 (2009).
236. Hall, C. E. *et al.* Progressive Motor Neuron Pathology and the Role of Astrocytes in a Human Stem Cell Model of VCP-Related ALS. *Cell Rep* **19**, 1739–1749 (2017).
237. 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).
238. Liu, J. *et al.* c9orf72 Disease-Related Foci are Each Composed of One Mutant Expanded Repeat RNA. *Cell Chem Biol* **24**, 141–148 (2017).
239. Zyner, K. G. *et al.* G-quadruplex DNA structures in human stem cells and differentiation. *Nat Commun* **13**, 142 (2022).
240. Soldner, F. *et al.* Generation of Isogenic Pluripotent Stem Cells Differing Exclusively at Two Early Onset Parkinson Point Mutations. *Cell* **146**, 318–331 (2011).

241. Zacco, E., Martin, S. R., Thorogate, R. & Pastore, A. The RNA-Recognition Motifs of TAR DNA-Binding Protein 43 May Play a Role in the Aberrant Self-Assembly of the Protein. *Front Mol Neurosci* **11**, 372 (2018).

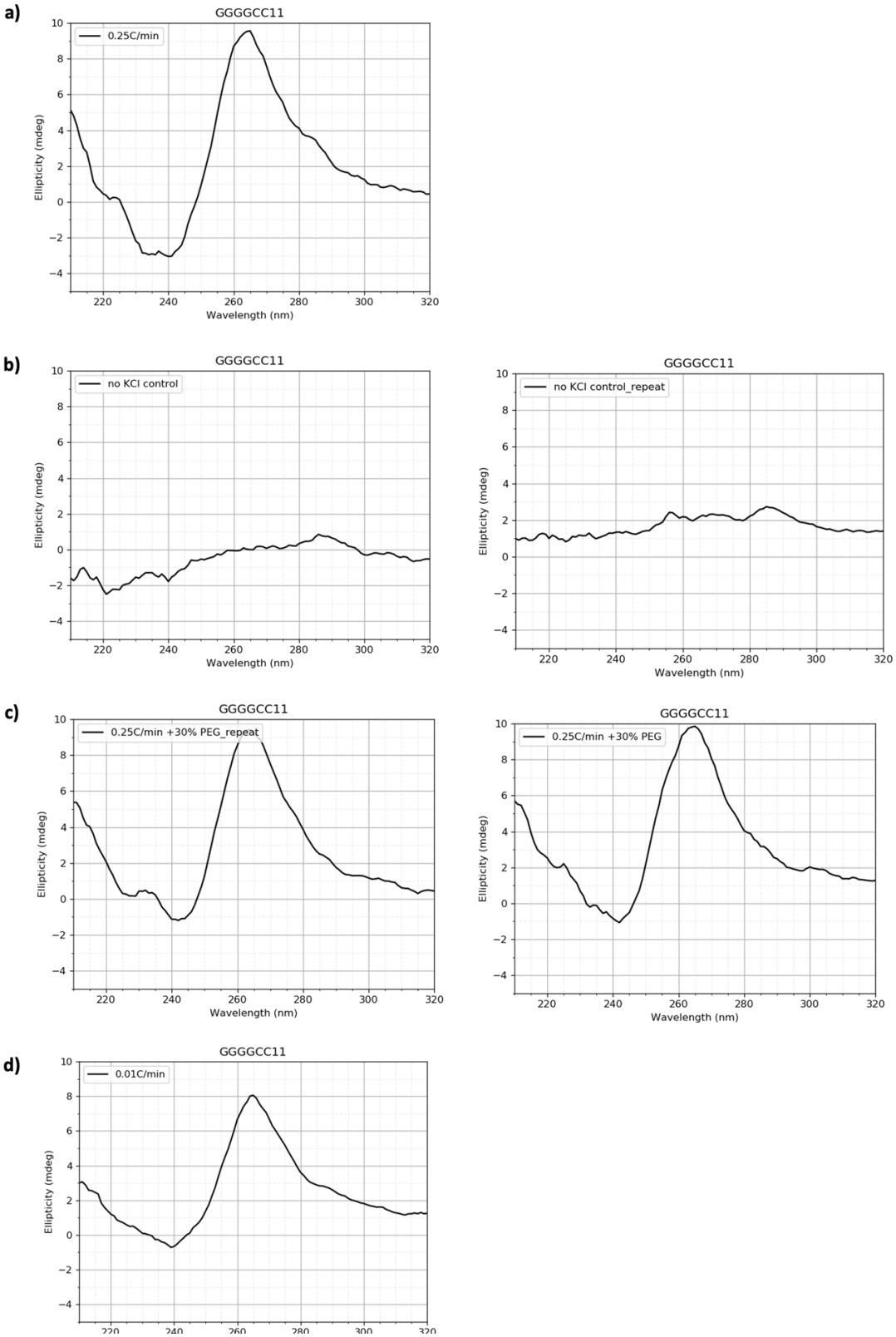
Appendix



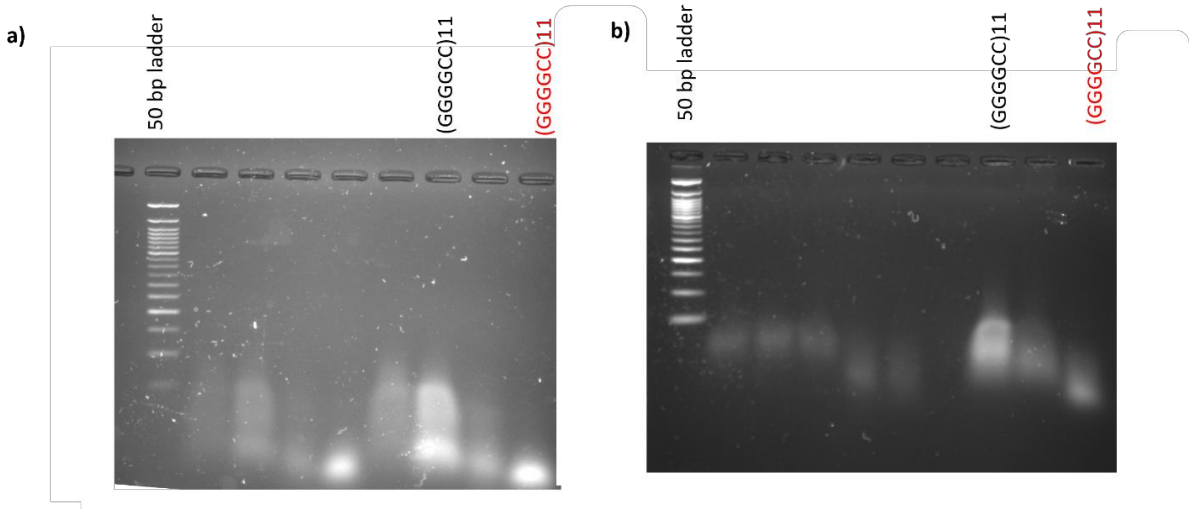
Appendix 1: Optical microscopy images of $(GGGGCC)_n$ (n = 1,5,11) pre-annealed stock in **(a)** PBS, (b) PB and **(c)** TRIS-HCl conditions. In this panel, samples are reported before and after incubation for 2h. Notably, samples images in panel **(a)** were additionally stained with 1:1 NMM: DNA and imaged on the fluorescent channel (shown in overlay with respect to the brightfield).



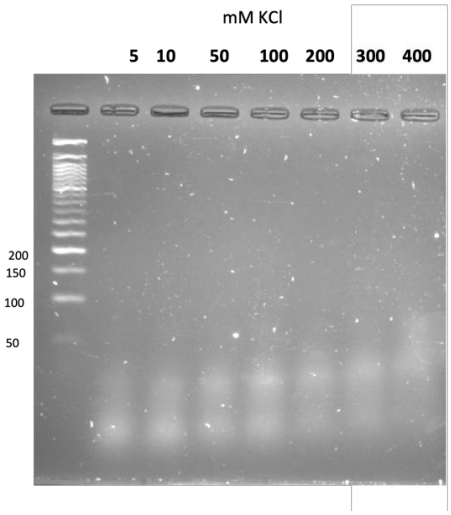
Appendix 2: Biological replicate for optimisation of denaturing conditions **Figure 19**. **(a)** Agarose Gel electrophoresis of $(GGGGCC)_{11}$ subjected to the optimised denaturing conditions. The samples contain a FAM label and are stained with SYBR Safe. **(b)** Optical and fluorescence microscopy of $(GGGGCC)_{11}$ after the optimised denaturing conditions. The DNA are FAM-labelled.



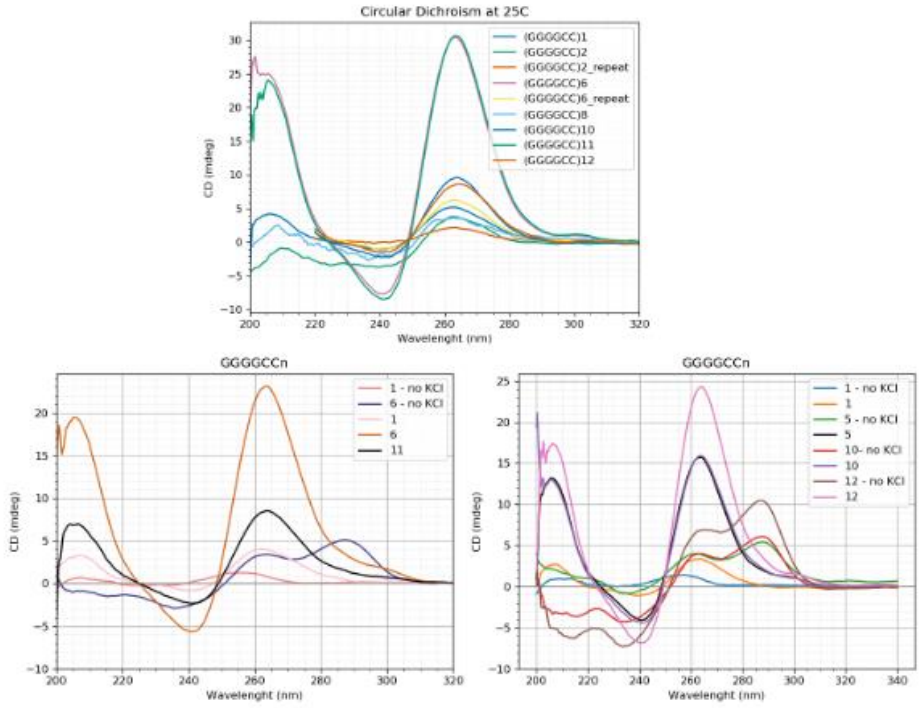
Appendix 3: Biological replicates for optimisation of denaturing conditions and negative control (no KCl) as in **Figure 21**. CD spectra of $(GGGGCC)_{11}$ samples subjected to different annealing conditions taken at 25°C for 10 µM $(GGGGCC)_{11}$ samples in TRIS HCL buffer (10 mM pH 7.5 + 500 mM KCl). Conditions tested: **(a)** heating to 95°C and slow cooling to -0.25°C /min; **(b)** negative control (no KCl) annealed with a -0.01°C /min cooling rate; **(c)** heating to 95C and slow cooling to -0.25°C /min in presence of 30% PEG and **(d)** heating to 95°C and slow cooling to -0.01 °C /min.



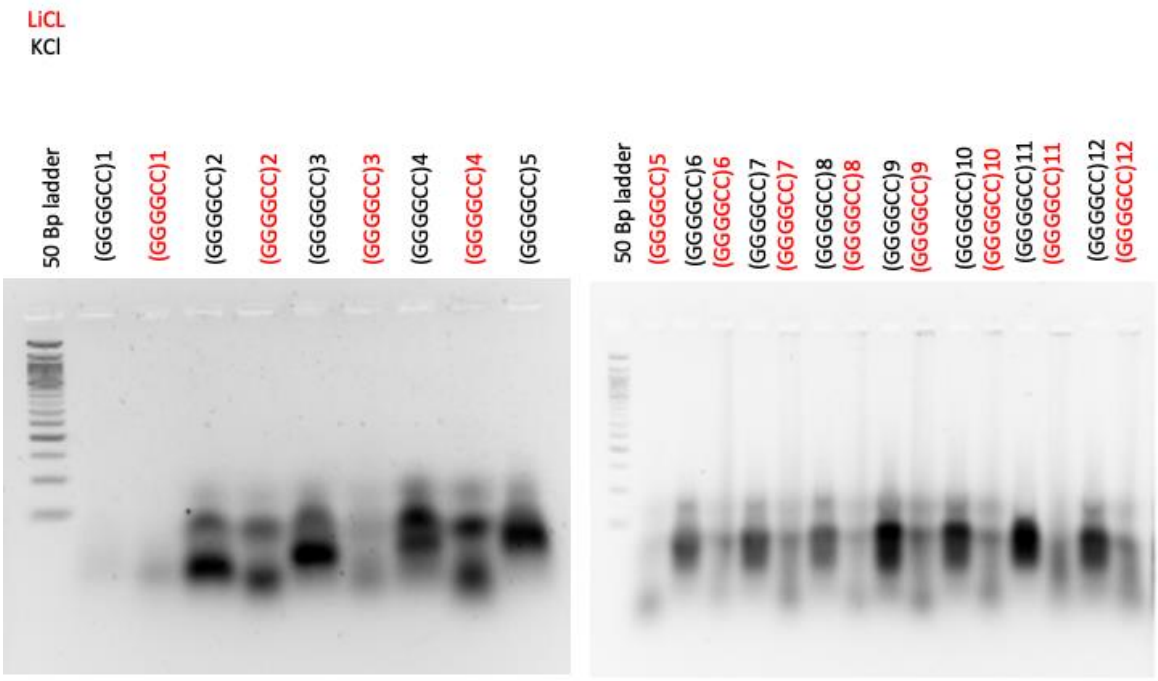
Appendix 4: Biological replicates of the AGE for optimised mG4-forming conditions of $(GGGGCC)_{11}$ as of **Figure 21**.



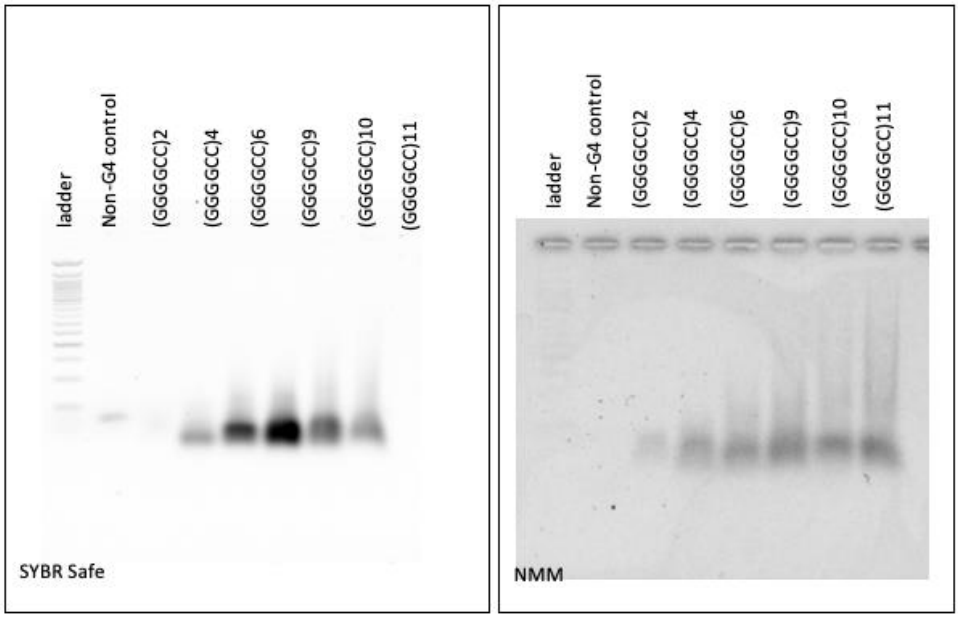
Appendix 5: Biological replicate of KCl titration experiment for sample $(GGGGCC)_6$ annealed under mG4 forming conditions (as of **Figure 23a**).



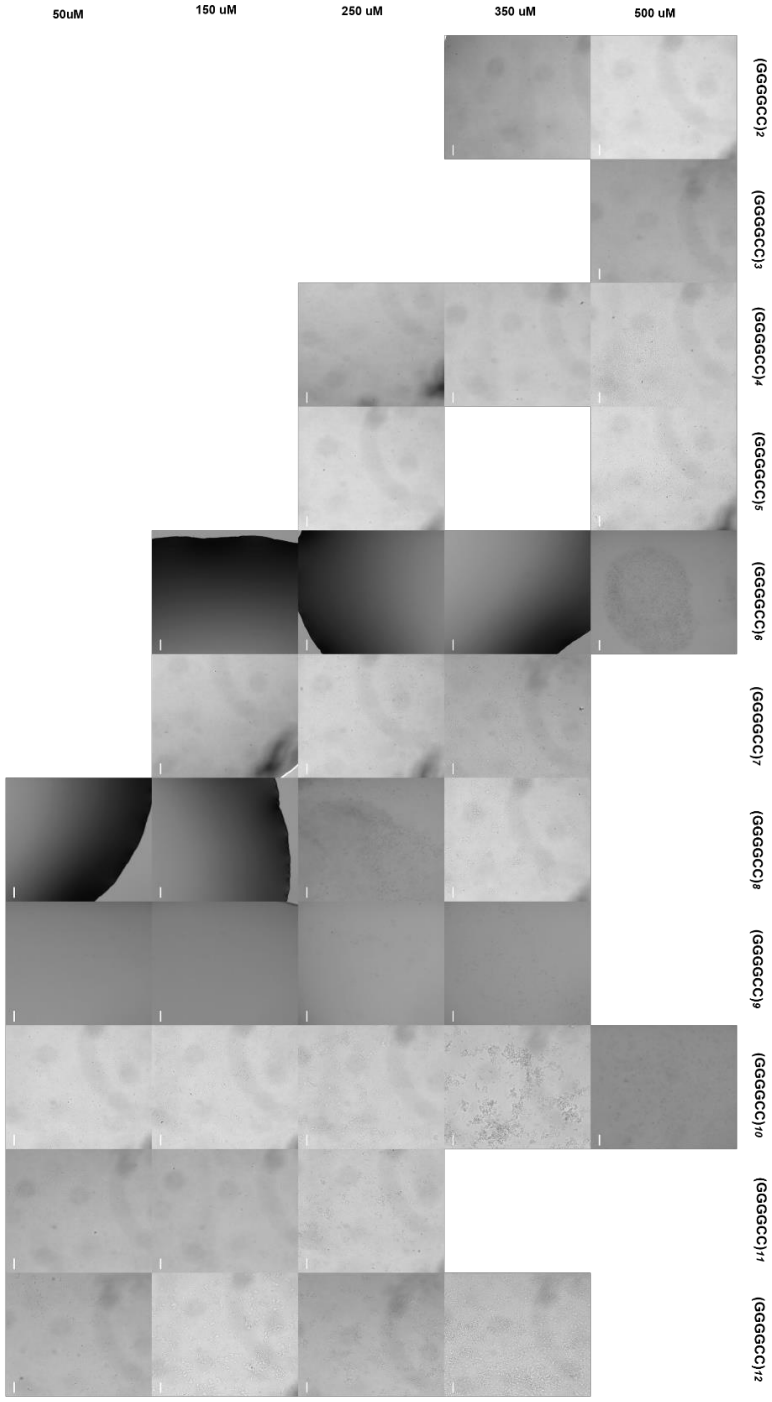
Appendix 6: Biological replicates of $(GGGGCC)_n$ CD spectra - $(GGGGCC)_n$ (n = 1, 2, 5, 6, 7, 8, 10, 11, 12). All samples were annealed under mG4-forming conditions and further diluted to 5 µM for CD (**Figure 24**).



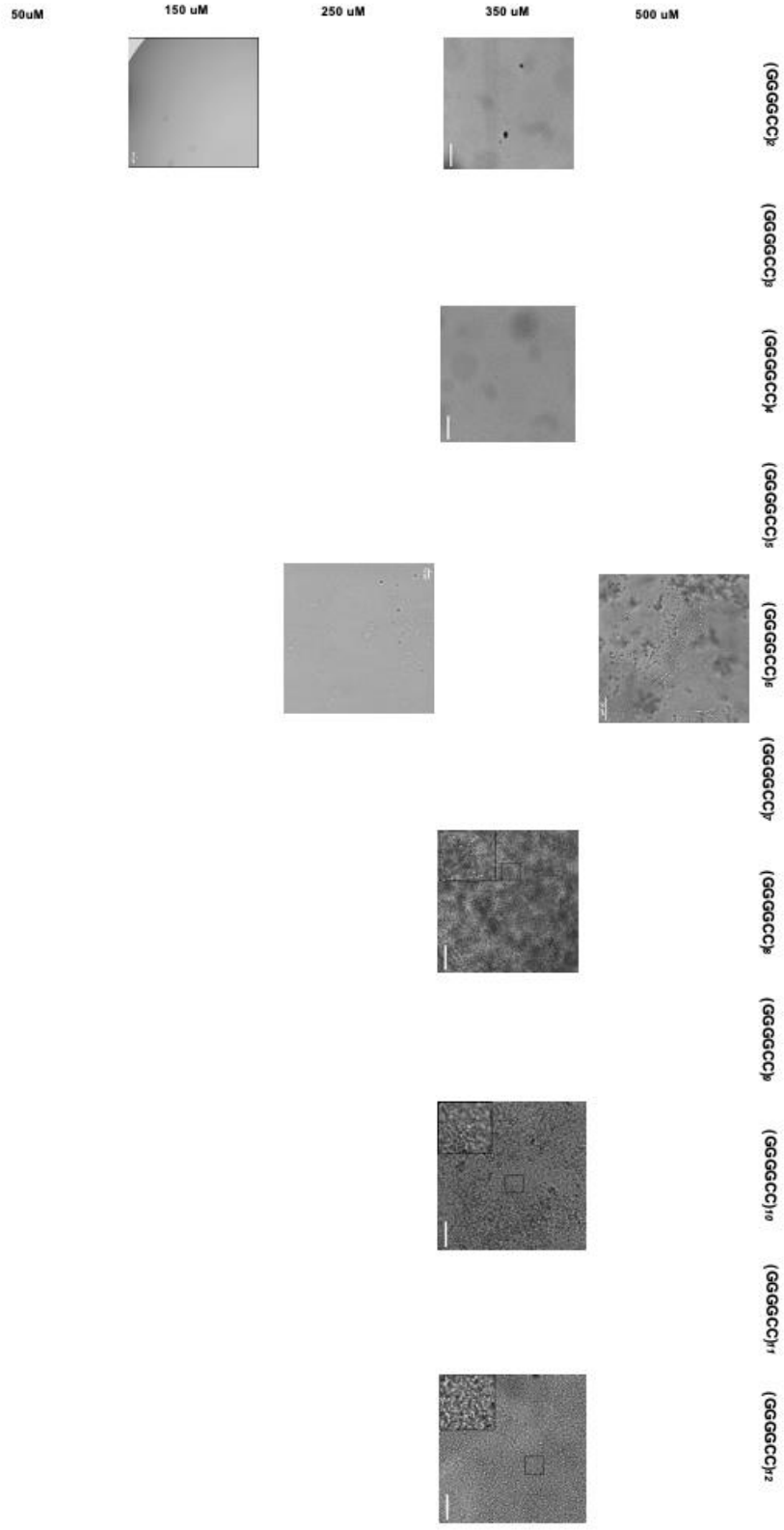
Appendix 7: Biological replicate of AGE representing $(GGGGCC)_n$ for n = 1-12 annealed in different buffer conditions as of **Figure 24**. KCl-containing samples are more prone to form higher molecularity species, as evident by the stronger higher bands (in black) with respect to the LiCl-containing sample (in red). The samples are annealed at 350 µM (n = 1-6) and 150 µM (n = 1-12) and diluted to 10 µM for AGE.



Appendix 8: Biological replicate of AGE representing $(GGGGCC)_n$ stained with G4-specific dye NMM compared to SYBR Safe (as of **Figure 24**).

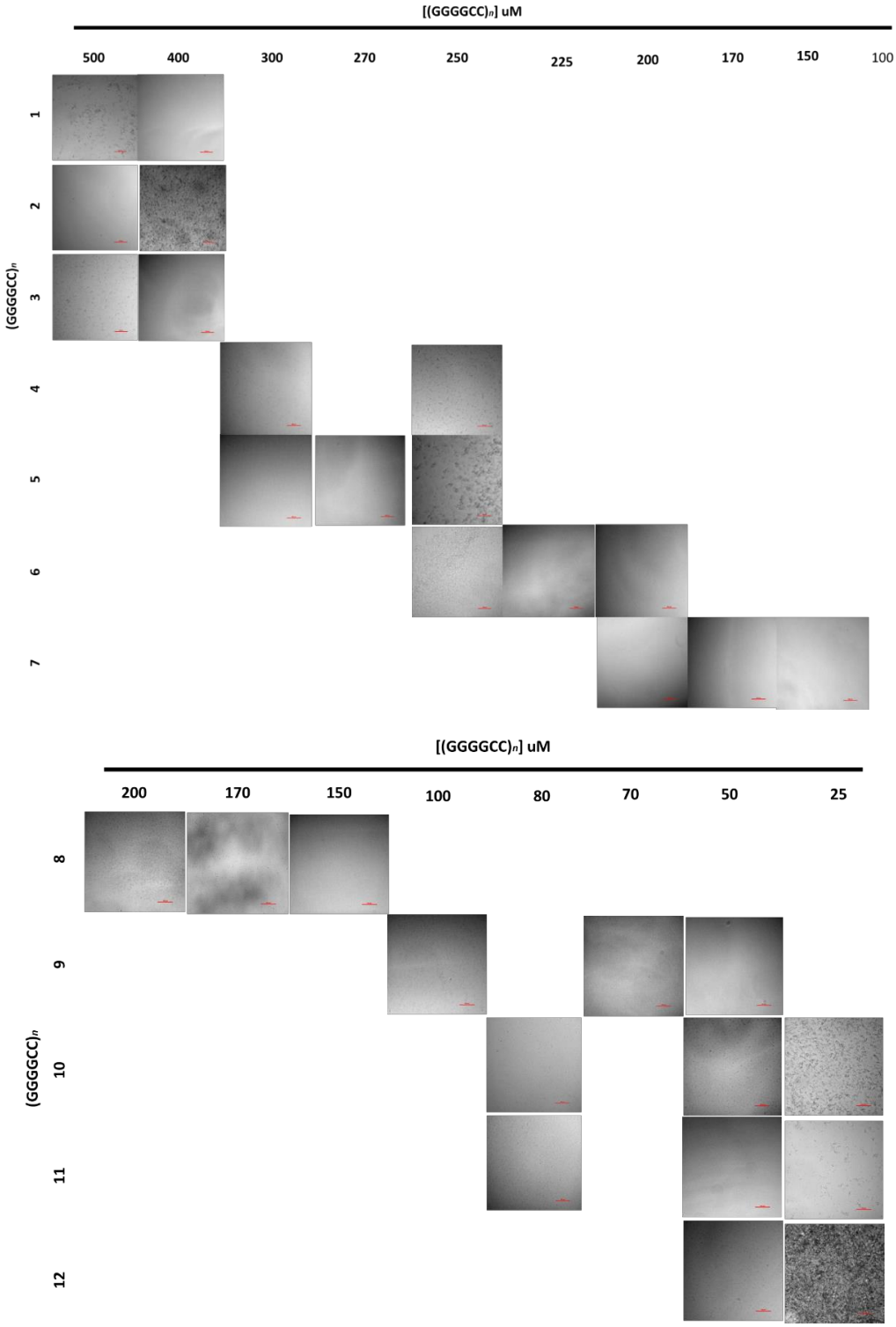


Appendix 9: Confocal Images for the Phase Diagram represented in **Figure 25**.

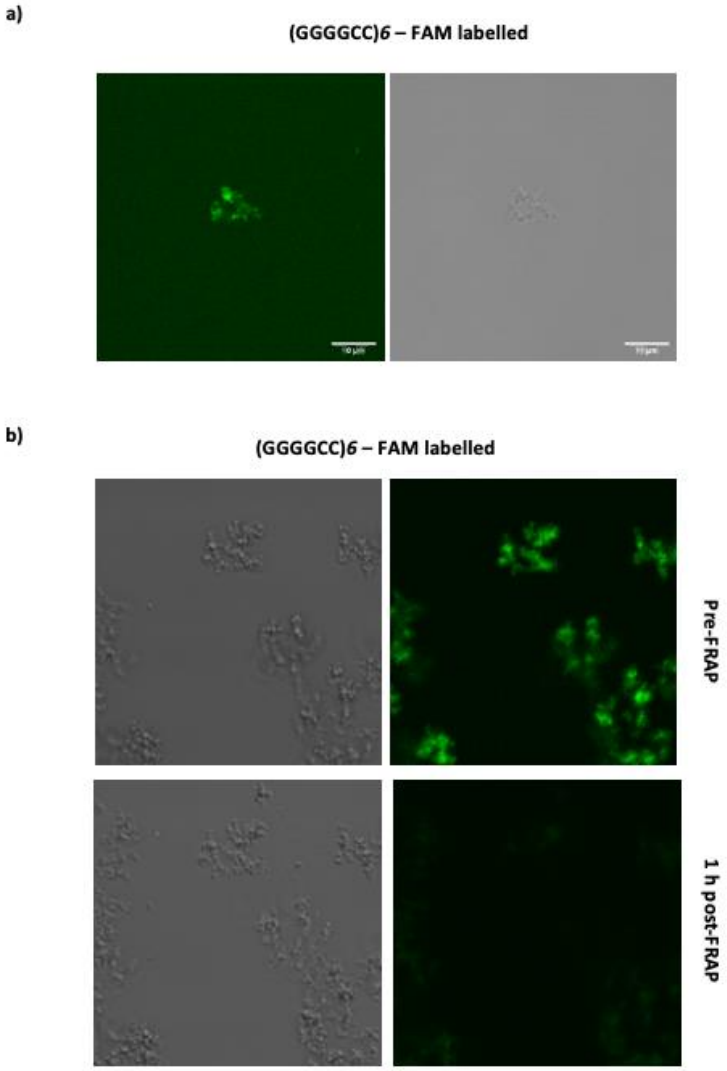


Appendix 10: Confocal Images for the Phase Diagram represented in **Figure 25**, second biological replicate for selected repeat lengths. Not all concentrations were tested due to the time-consuming and expensive nature of the experiment. However,

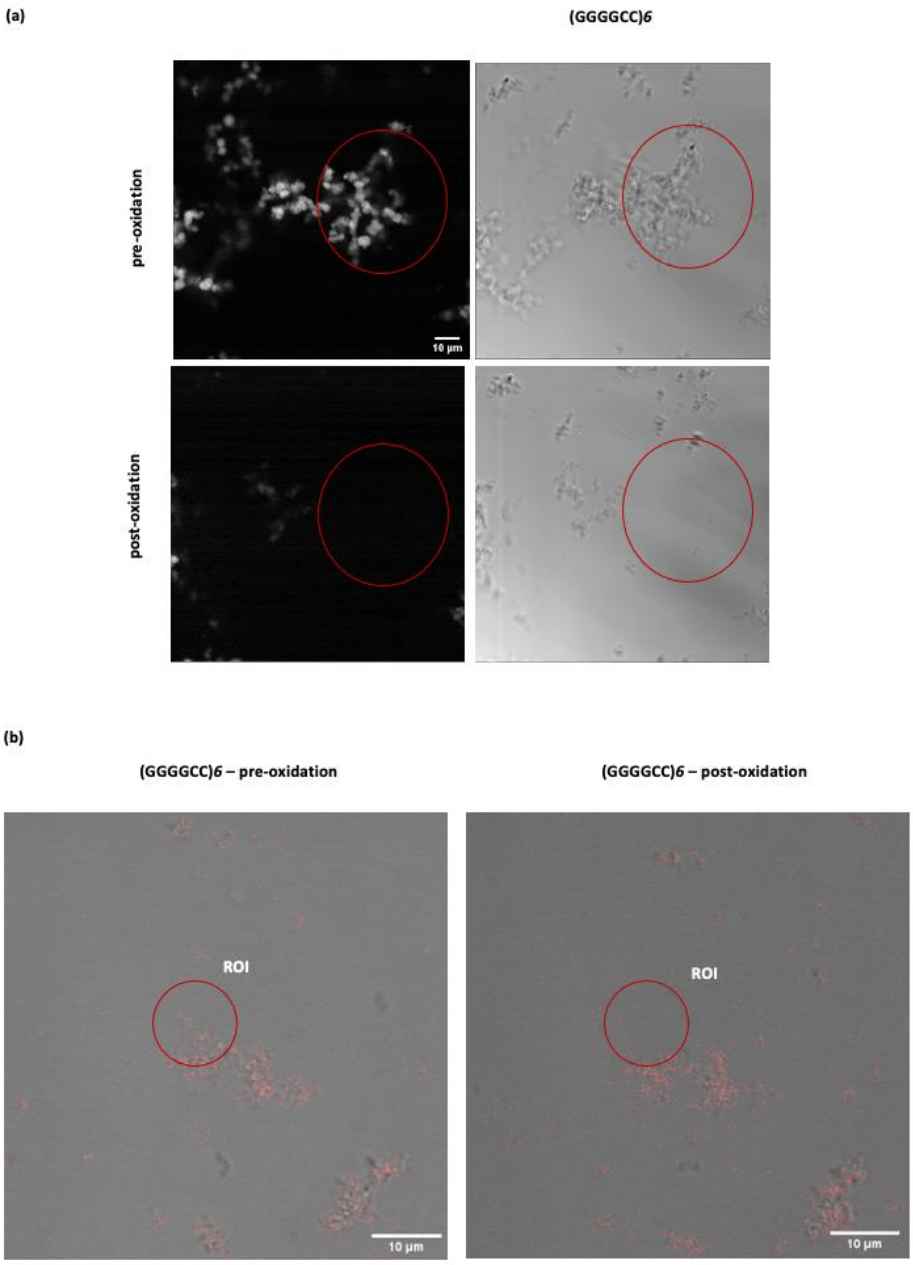
the protocol was successfully repeated independently by other lab-members and with different DNA stocks, indicating the repeatability of the optimised mG4 annealing protocol.



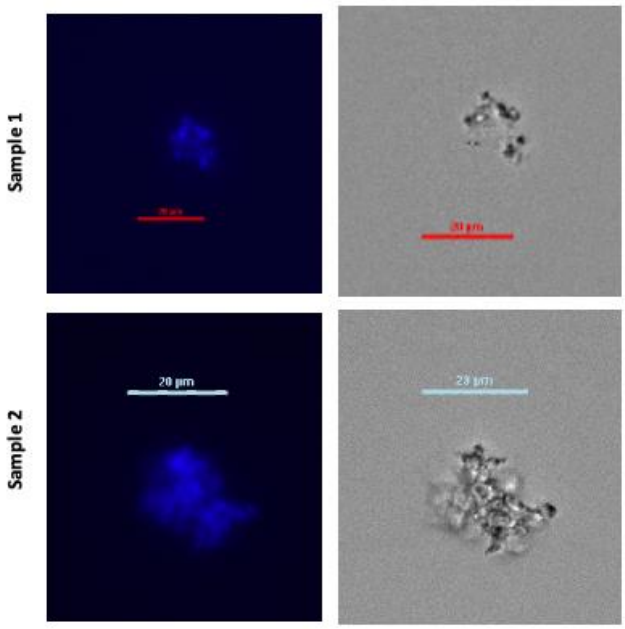
Appendix 11: Optical Microscopy Images for Phase Diagram in **Figure 28b**.



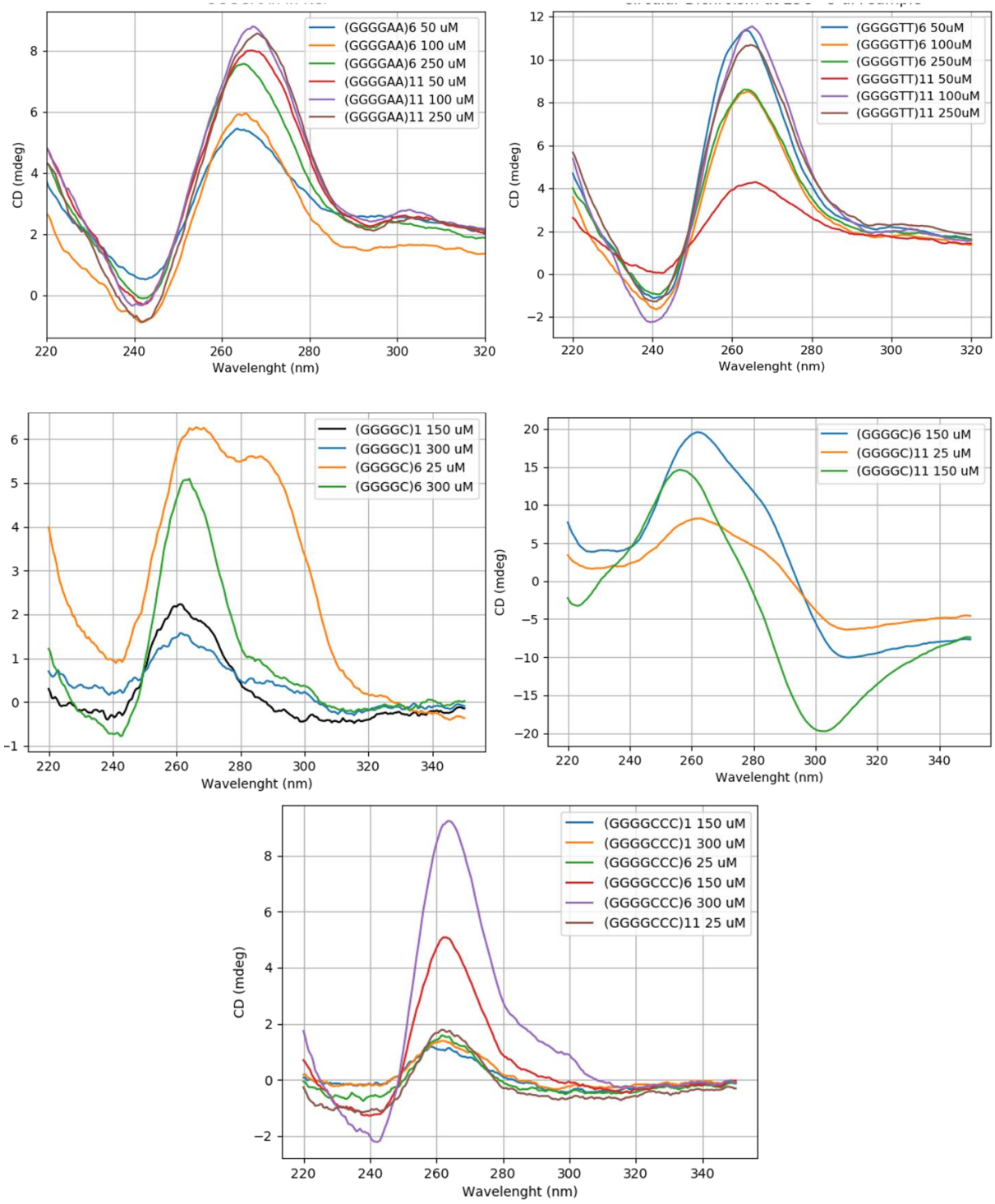
Appendix 12: Biological replicate of micrographs of FAM-labelled $(GGGGCC)_6$ as of **Figure 29. (a)** Confocal image of $(GGGGCC)_6$ (10% FAM labelled), annealed in mG4-forming conditions. **(b)** Upon laser excitation FAM labelled strands photobleached, leading to a local reduction of fluorescence intensity. The photobleached strands did not diffuse within the reset of the aggregate after 1h, indicating a solid-like state.



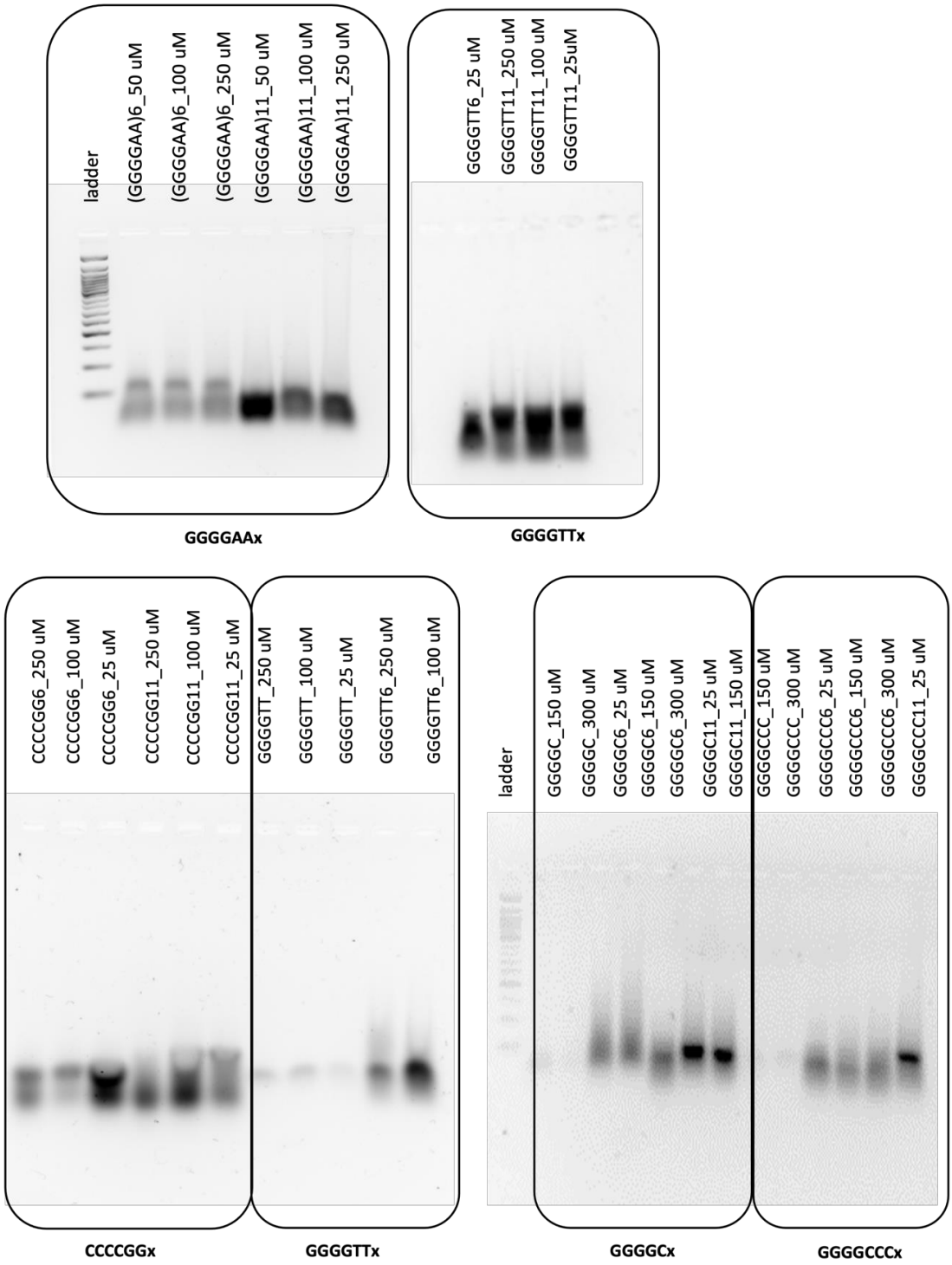
Appendix 13: Biological replicates of NMM-photo oxidation experiment for mG4-containing aggregates as of **Figure 30.** The mG4-based aggregates disassemble in the region of interest (ROI) which has been subjected to irradiation. **(a)** was performed with the same parameters as the experiment presented in the main text whilst **(b)** is reported as an overlay of the NMM and brightfields channel in order to highlight the spacial organisation of all components before and after irradiation.



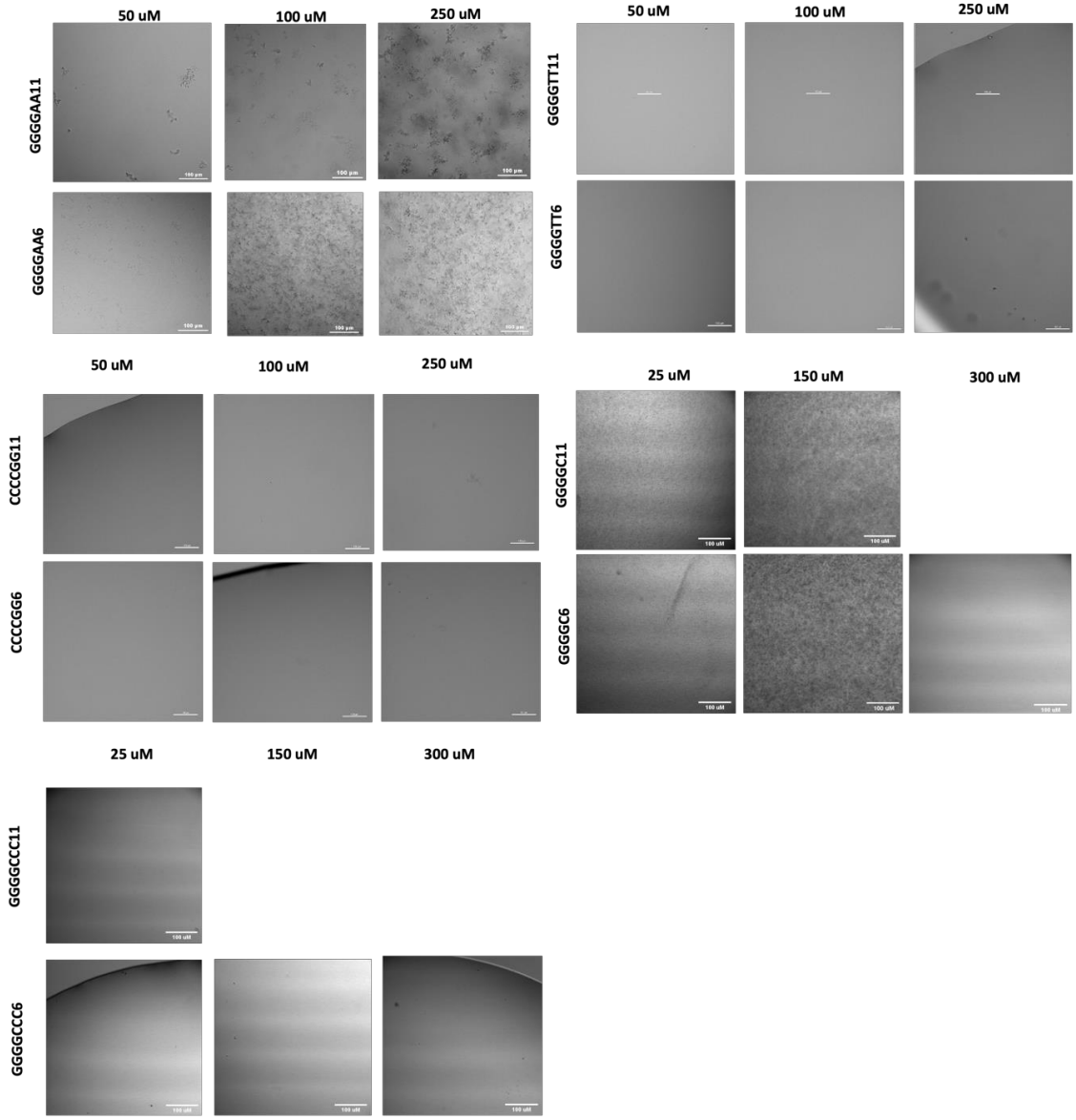
Appendix 14: Biological replicate for mG4-autofluprescence experiment as of **Figure 31**. Optical imaging of mG4-based aggregates formed by $(GGGGCC)_{11}$ observed by irradiating the sample at 405 nm. On the right, brightfield images are shown for two different samples. On the left, autofluorescence is observed on the DAPI channel.



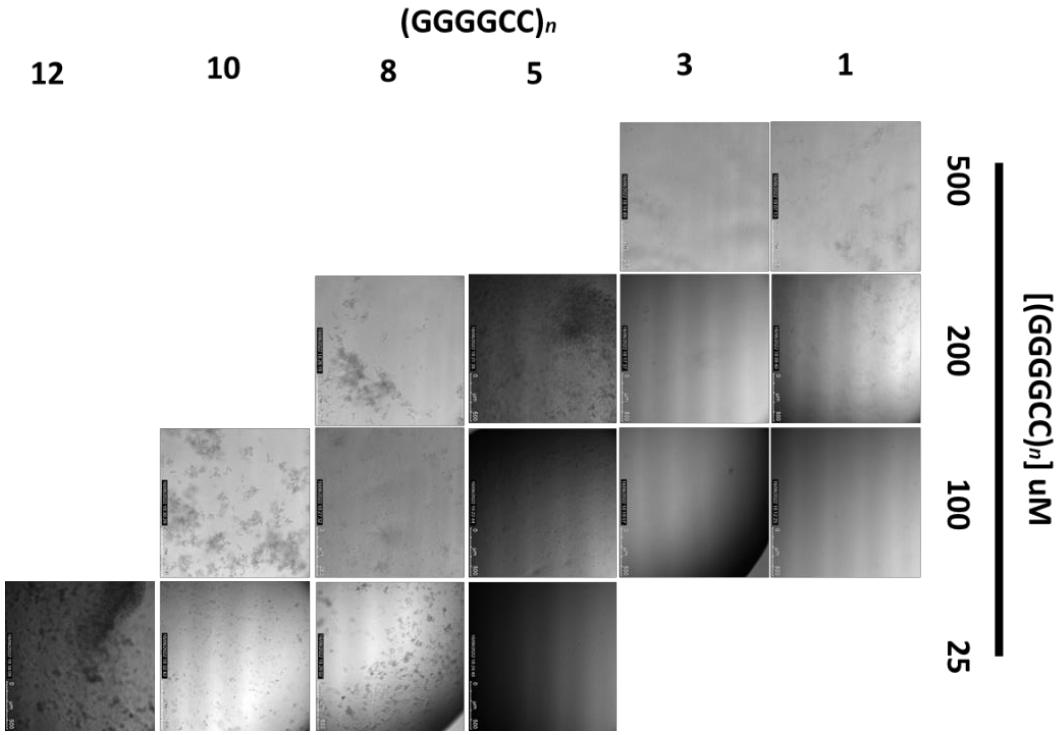
Appendix 15: Biological replicates of CD spectra of rationally designed mutants as of **Figure 34, 35.** All sequences were annealed under mG4 forming conditions and diluted to 5 µM to run the spectra at room temperature.



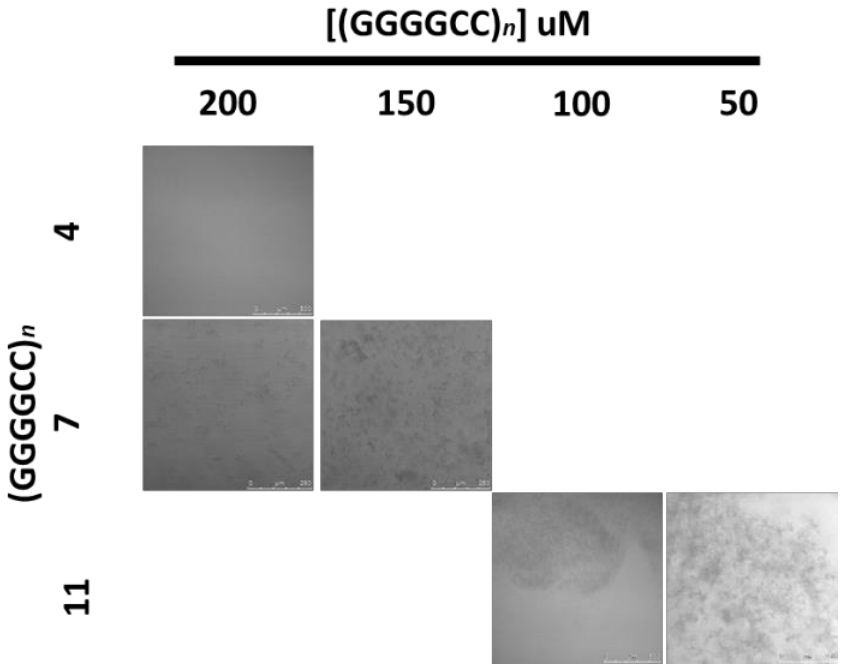
Appendix 16: Biological replicate of AGE of rationally designed mutants as of **Figure 34, 35**. All sequences were annealed under mG4 forming conditions.



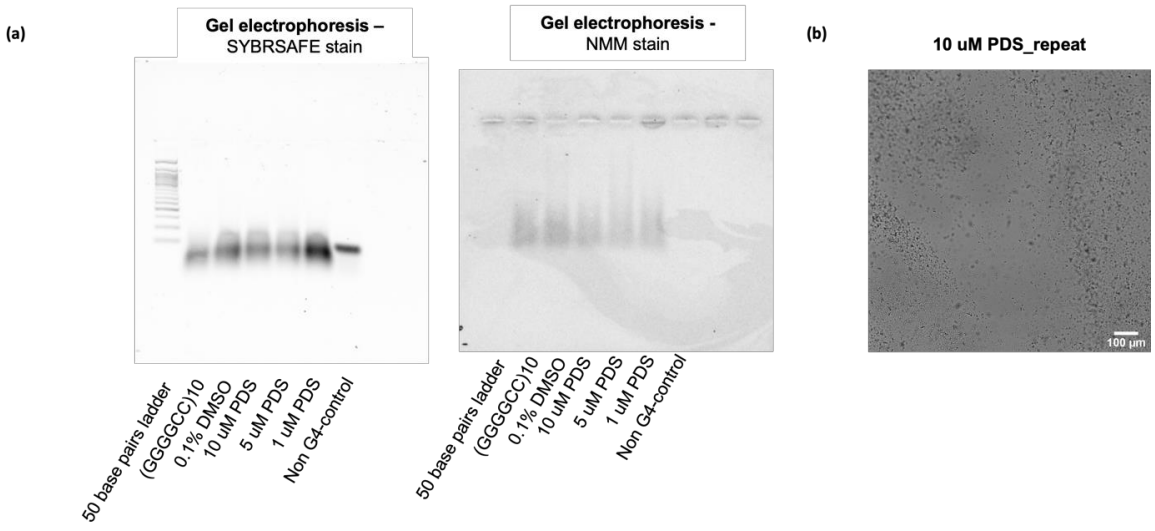
Appendix 17: Biological replicates of micrographs of rationally designed mutants as of **Figure 34,35.** All sequences were annealed under mG4 forming conditions.



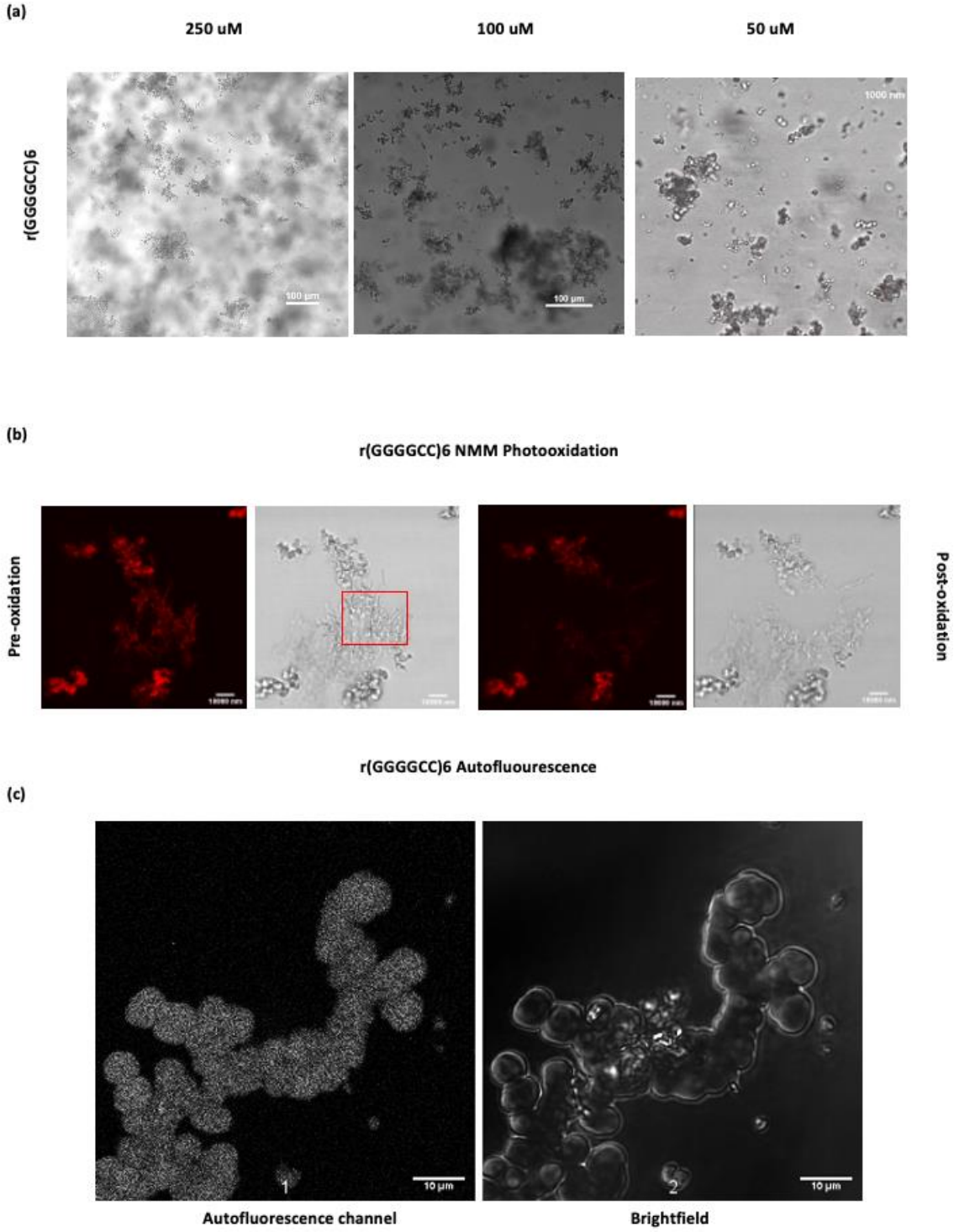
Appendix 18: Brightfield images for the phase diagram in **Figure 36b** 0.75% DMSO.



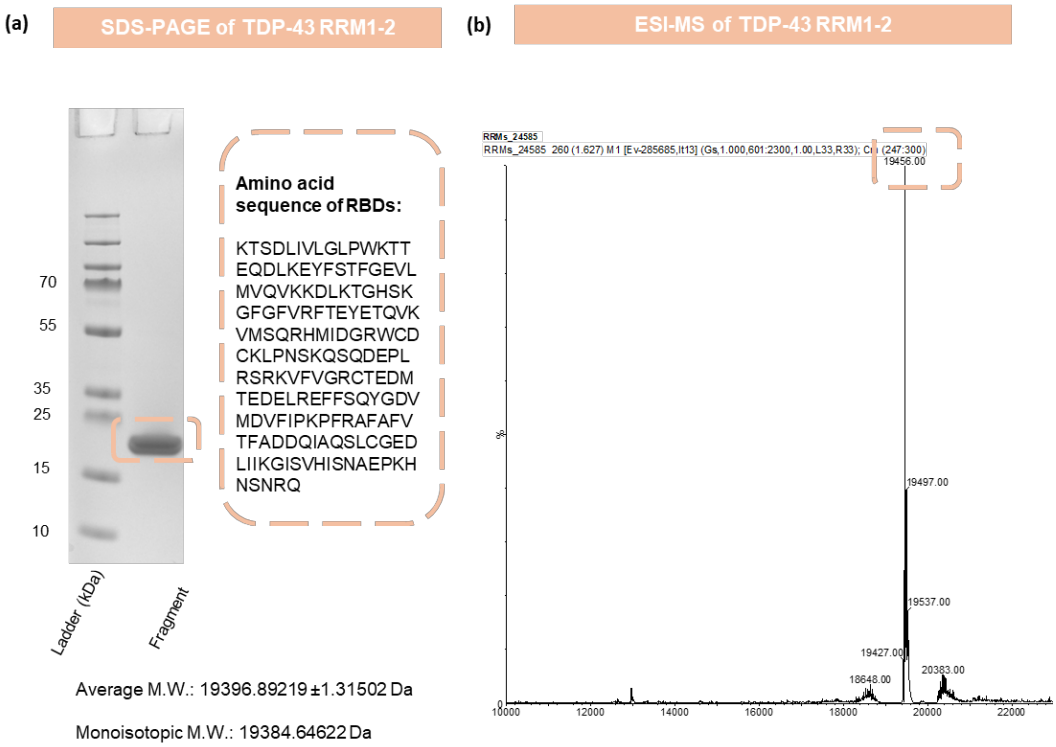
Appendix 19: Brightfield images for samples in the phase diagram in **Figure 36b** 0.125% DMSO.



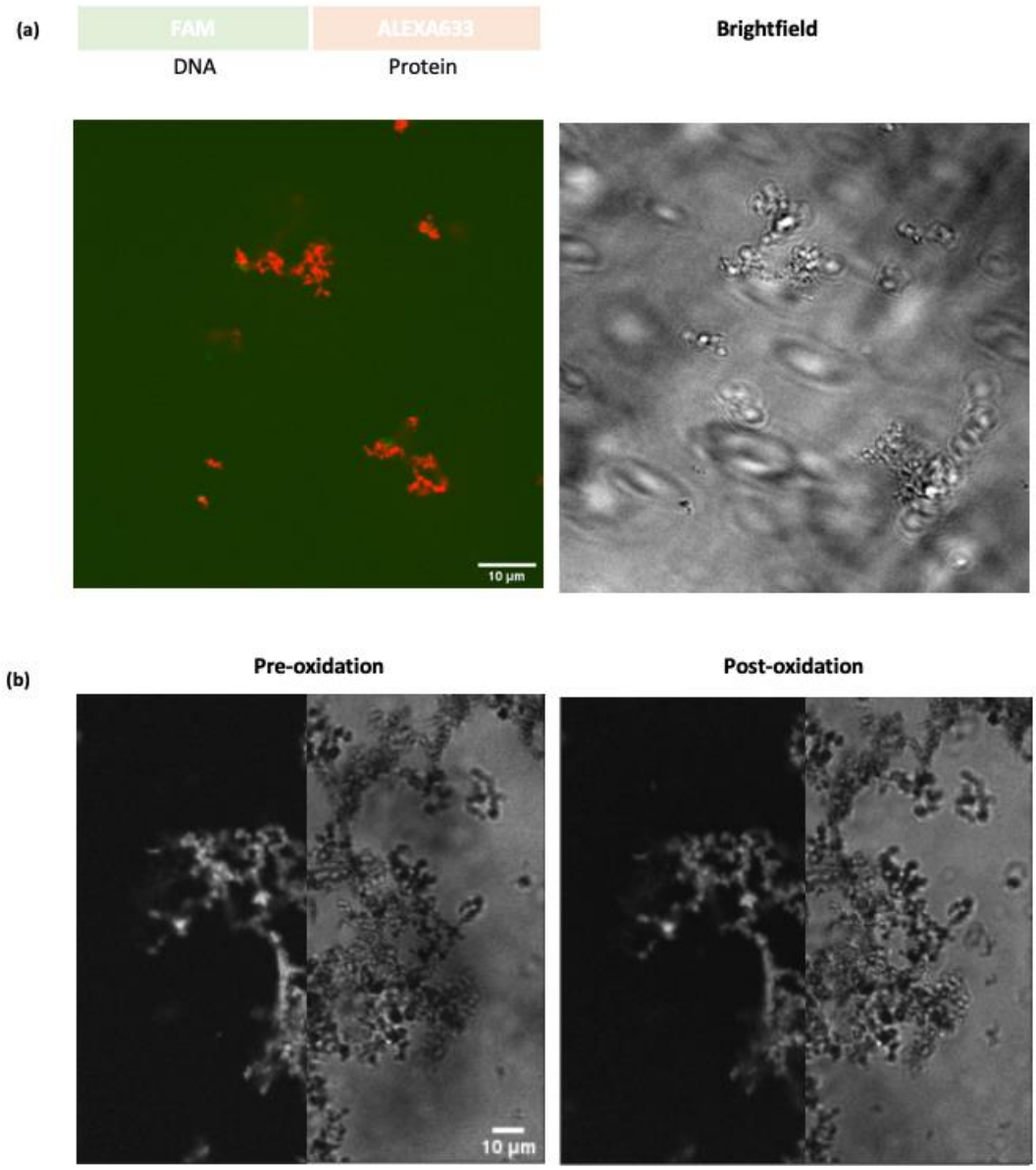
Appendix 20: Biological duplicate of PDS addition experiment in **Figure 38**. **(a)** AGE gels stained with SYBSSAFE and NMM. **(b)** Brightfield microscopy.



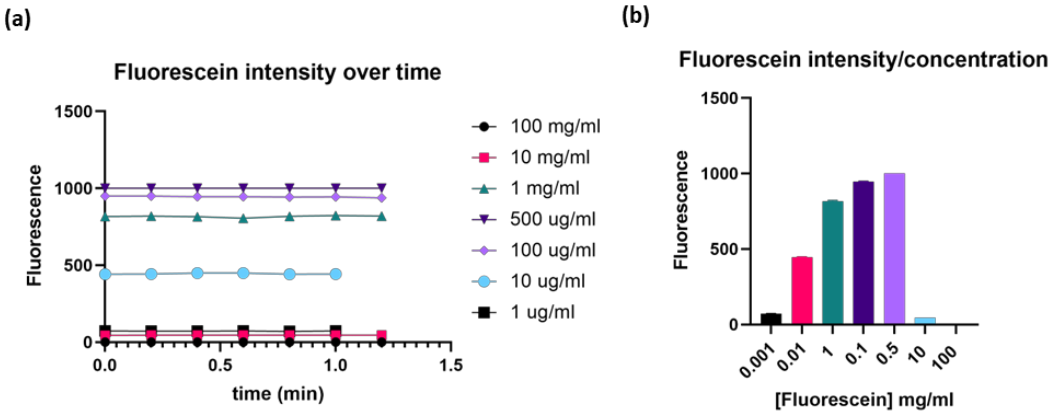
Appendix 21: Biological Replicates of RNA experiments as of **Figure 41.** **(a)** Brightfield microscopy of r(GGGGCC)$_6$ aggregates at different concentrations. Biological replicates of Figure 40. **(b)** NMM-photooxidation of r(GGGGCC)$_6$ aggregates. **(c)** Autofluorescence of r(GGGGCC)$_6$ aggregates.



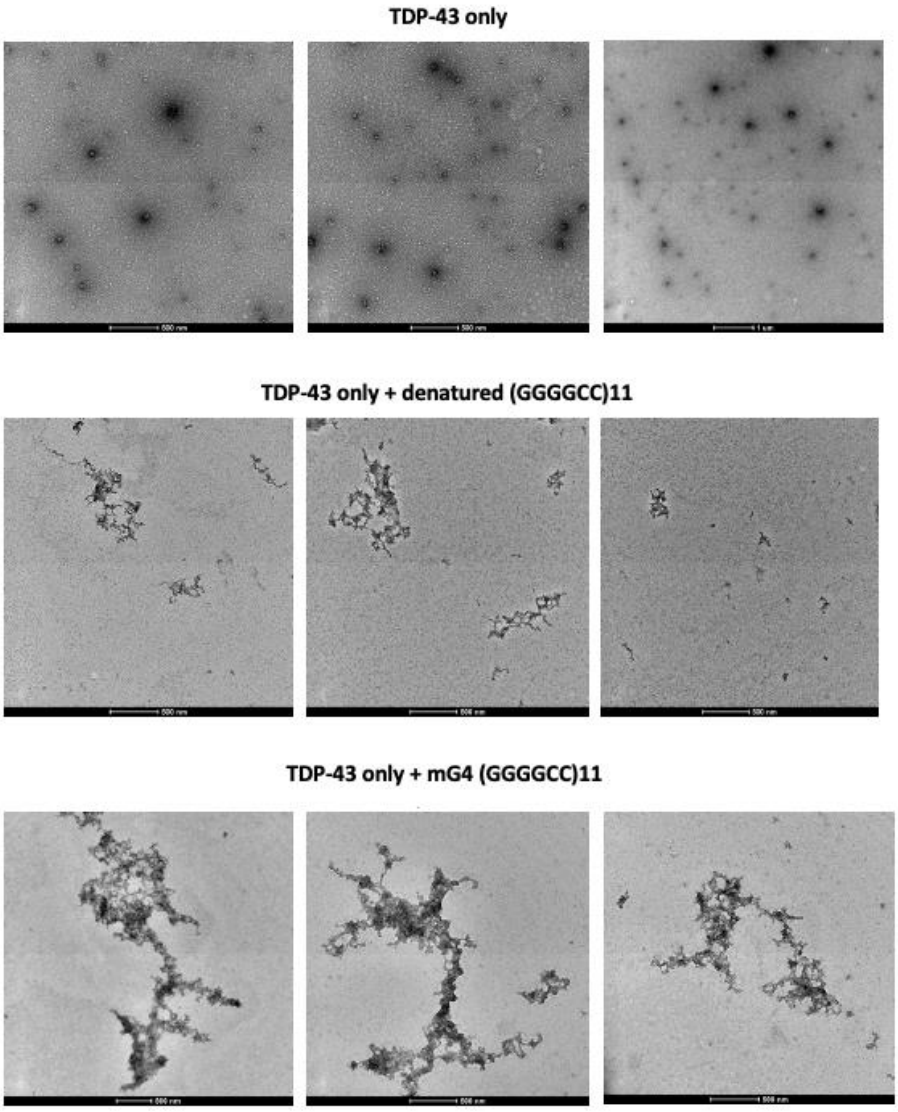
Appendix 22: TDP-43 RRM1-2 protein purification assessment. (a) SDS-PAGE gel to assess purity of the expressed fragments, provided by collaborator Anouk Huyghebaert. **(b)** Mass spectrometry to assess the purity of the express fragments, provided by collaborator Anouk Huyghebaert.



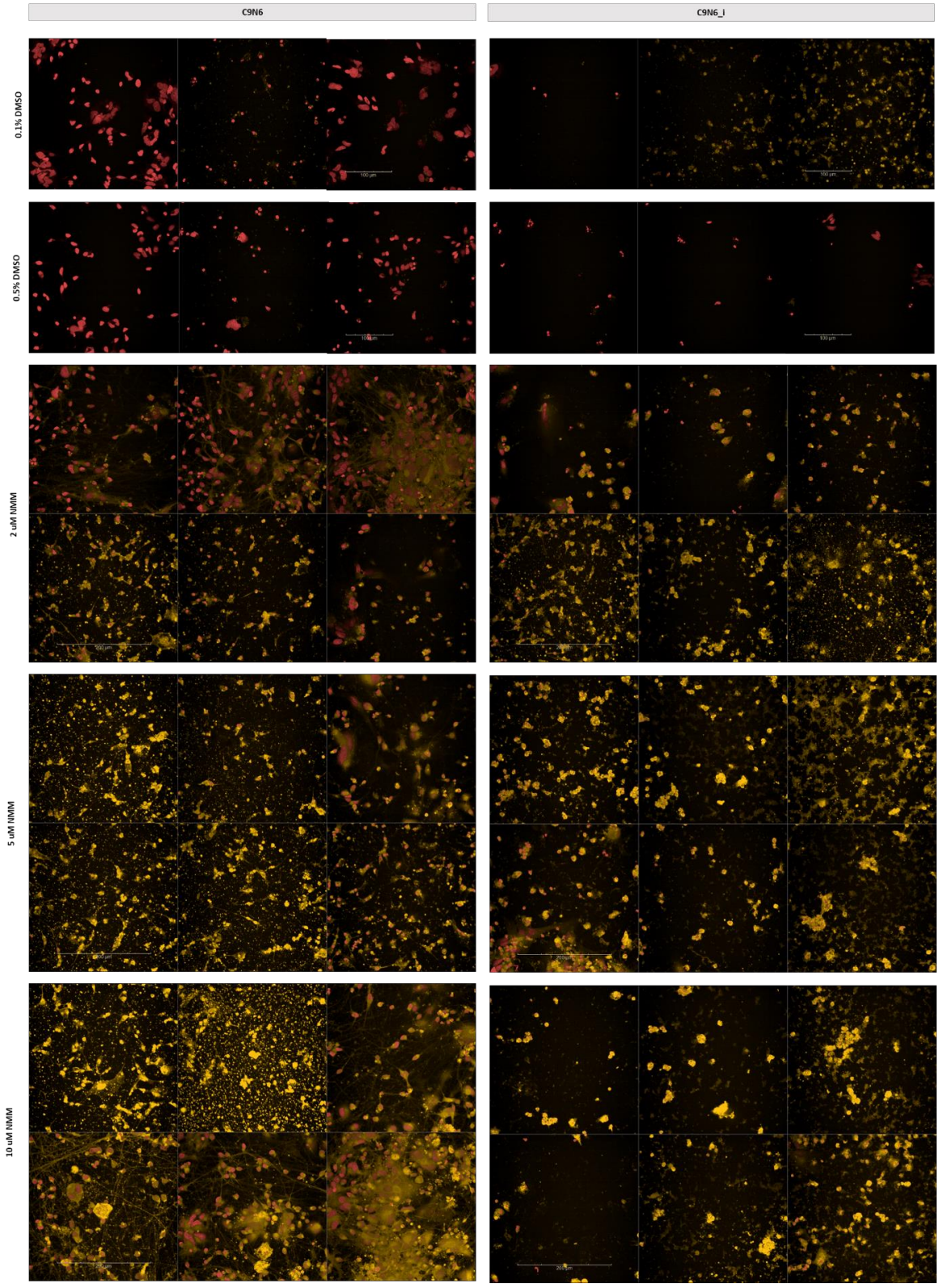
Appendix 23: Biological replicate of TDP-43 experiments as of **Figure 49.** **(a)** co-localization of TDP-43 and mG4s in the aggregates. Biological replicate of Figure 45. **(b)** NMM-photooxidation of TDP-43 and mG4s aggregates.



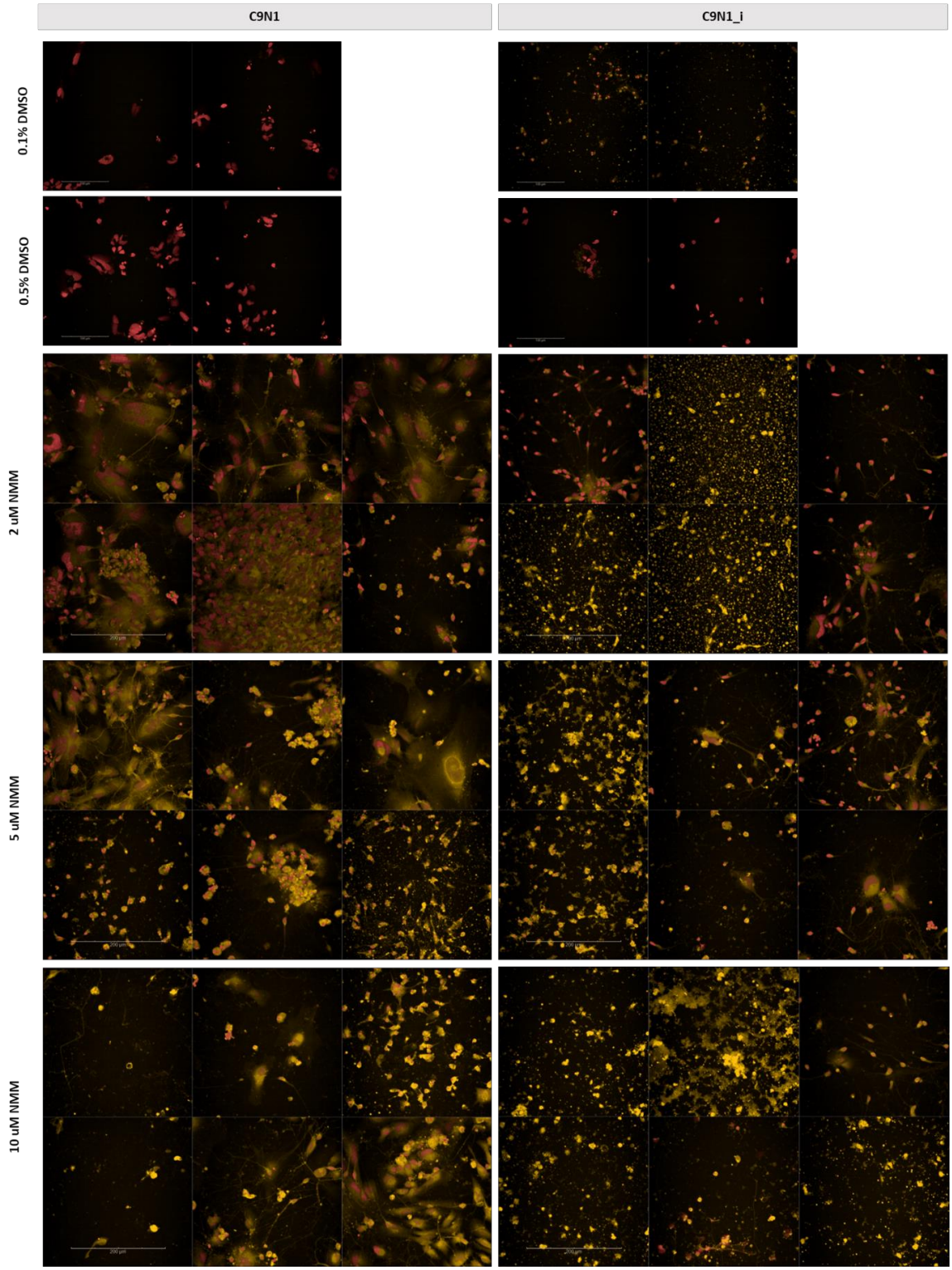
Appendix 24: (a) Fluorescein intensity over time was tested to assess its resistance to photobleaching. Raw data generated by Juan Ivars Minana. The samples did not present any photobleaching behaviours. **(b)** Fluorescein intensity was assessed at different concentrations. At high concentrations, fluorescein self-quenches. Raw data generated by Juan Ivars Minana from the Elani Group at Imperial College London.



Appendix 25: TEM imaging of TDP-43, TDP-43 in presence of denatured $(GGGGCC)_{11}$ and TDP-43 in presence of mG4s. Data obtained by Anouk Huyghebaert from the Di Aprile group at Imperial College London. Technical triplicates for **Figure 50**.



Appendix 26: Triplicates images for C9N6 cell line. Fixed IPSC patient derived ALS/FTD neuronal cells with the C9orf72 mutation. In red, a nuclear stain is used to visualise the different cellular entities. In yellow, the NMM channel is selectively excited. Despite the increasing concentration of NMM, the cells remained alive in the C9orf72 cell-line, indicating no/low cytotoxicity. Imaging in live cells indicated NMM accumulation in the cytoplasm and – to a lesser extent – the nucleus.



Appendix 27: Triplicates images for C9N1 cell line. Fixed IPSC patient derived ALS/FTD neuronal cells with the C9orf72 mutation. In red, a nuclear stain is used to visualise the different cellular entities. In yellow, the NMM channel is selectively excited. Despite the increasing concentration of NMM, the cells remained alive in the C9orf72 cell-line, indicating no/low cytotoxicity. Imaging in live cells indicated NMM accumulation in the cytoplasm and – to a lesser extent – the nucleus.

Figures copyright

The copyright of the figures generated in thesis rests with the author.

Figures generated with BioRender are subject to BioRender's Academic License Terms and Industry License Terms permits users to sublicense completed graphics containing BioRender content to publishers for use in various publications.

Figure 1d – is freely available for non-commercial use according to the Creative Commons CC-BY-NC.

Figure 31a – is freely available for non-commercial use according to the Creative Commons CC-BY-NC.

Acknowledgements

My PhD would have not had such a successful outcome without the contributions of my colleagues, fellow scientists from other institutions and my own support group of family and friends.

Firstly, I would like to thank my supervisors, Marco Di Antonio and Lorenzo Di Michele for their unwavering support throughout my studies. Without them to motivate me when experiment failed, giving me tips and suggestions on how to improve and providing me with a network on which to rely on, I would have not been able to complete this work.

Secondly, I would like to thank my collaborators from the Patani group at the Francis Crick and the Aprile group at Imperial, who provided me with excellent training in fields I was not familiar and aided me in completing the biological relevance studies for my thesis and my first author paper.

I would also like to acknowledge the members of the Di Antonio, the Di Michele and the Elani groups for always been there for me whenever I needed help within or outside the lab. I would like to especially mention Ludovica Monti, Silvia Galli, Jenna Robinson, Sabrina Nuccio, Gem Flint and Thomas Maher, who always managed to make my lab days brighter.

Finally I would like to thank all my family and friends who supported me in the past three years.

It did not matter the time of the day at which I called, my mum and dad and my little brother Francesco never failed to pick up once. Even a country away they were able to provide me with unconditional love and support during a stressful period of my life.

Matt, who was my constant emotional support and source of laughter after every long lab day, every bad meeting and interview.

My flatmate Edoardo, who has been dealing with me complaining about lab-life for over 7 years, and never once made my fears/sorrows feel inappropriate.

My friend Marco, who never failed to make me laugh in the midst of a stressful week by showing me his own misfortunes.

My childhood friends, Maria ed Enrico, who were always there for me whenever I went back home.

My gym buddies, Aglaia and Camilla, who would meet me at the gym after a long day in the lab with relatable work stories and made me laugh at failed experiments.

To all of you I owe a massive thank you, as this work would have not been possible in your absence.