The University Of Sheffield.

The role of G-Quadruplexes in regulating gene expression and *Arabidopsis thaliana* development

Colette Lisa Baxter

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy The University of Sheffield Faculty of Science Department of Molecular Biology and Biotechnology

March 2019

Abstract

G-Quadruplexes (G4s) are RNA and DNA secondary structures that form in guanine-rich sequences. Increasing evidence identify that G4s play important roles within fundamental biological processes, however there is little known about the role of G4s in plants.

Here we demonstrate that treatment with G4 stabilising ligands significantly affects overall plant growth with particularly strong effects on the primary root and root hairs. We propose that these phenotypes may be caused by G4 ligands causing a significant downregulation of expression in the *EXTENSIN (EXT)* gene family. The proline rich motifs within *EXTs* results in two tetrad G4s on the template strand of the gene body. We suggest that these G4s within *EXTs* may impact on RNA Polymerase II processing during transcription and splicing. The reduction in EXT expression may result in the plants compromised cell walls and therefore cause the root hair morphology and growth defects observed

This research indicates novel roles of G4s in fundamental biological processes, which may also occur in many other organisms, including mammals. Furthermore, we identify potential G4 dependent growth mechanisms in plants, the knowledge of which could facilitate the production of plants with greater yields, use fewer resources, or utilised for herbicide production.

Acknowledgements

Firstly, I would like to thank my supervisor, Karim Sorefan, for giving me this fantastic opportunity and for his constant support and guidance throughout this whole experience. I would like to thank Stuart Casson for effectively being my second supervisor throughout my PhD and for all the help and advice he has given me. Thank you to Julie Gray and Rebecca Barnes for their guidance as my academic advisors.

Thank you to the Department of Molecular Biology and Biotechnology, The University of Sheffield for funding my PhD. Professor Christoph Ringli for the EXT mutant lines and Dr Liam Dolan for the *rsl4* mutant. Thank you to the University of Sheffield, light microscopy facility for allowing me to use their microscopes.

I would like to thank the members of the Casson and Sorefan labs who made the lab a wonderful place to be. Mathew Parker for being my G-sQuad partner, doing the great bioinformatics that continually pushed the project forward. Manoj Kumar Vallaru and Giulia Arsuffi for kicking off this research with their microarray. Peter Venn for his scientific advice and ideas. James Thackery for being the best lab neighbour I could have wished for. Tinashe Mawodza for somehow being cheery every day. Hannah Sewell for encouraging me to apply for things that I would have wimped out of. Jim Rowe for his musical education and cheering me up after rounds of failing PCR. Rachel Denley Bowers and Jordan Brown for the laughter, ranting and film nights. Also, Nicholas Zoulias who went above and beyond for me, as a mentor and friend, no matter the circumstance (and for doing the CD Spec).

I would like to thank my family and friends. Jess keen, Hannah Jones, Nicole Ashworth and Liz Ruane, for enabling me to escape science and have fun occasionally. Chris Baxter, Emma Baxter, Georgia Lloyd Baxter, Oliver Lloyd Baxter, Jane Hope and Matt Hope for their constant support and encouragement. A huge thank you to Joshua Hope, my amazing true friend and partner, for being there for me since the beginning, through highs and lows, making sacrifices while still making me smile, laugh and beyond happy. Finally, a special thank you to Julie Baxter, Gill Baxter and Sam Baxter who I could never thank enough for everything they have done for me; for your love, support and belief in me, I am eternally grateful.

Contents

List of Figures

.

Figure of Tables

Acronyms and Abbreviations

Acronym	Definition
ALS	Amyotrophic Lateral Sclerosis
AS1	Asymmetric Leaves 1
AS2	Asymmetric Leaves 2
ATR	*Ataxia Telangiectasia-Mutated and Rad-3 Related Gene*
ATRX	ATP-*D*ependent Helicase
ARF7	Auxin Responsive Factor 7
ARF19	Auxin Responsive Factor 19
BCL	B Cell Lymphoma
BLM	Bloom
BP	Brevipedicellus
C9ORF72	Chromosome 9 Open Reading Frame 72
CD	Circular Dichroism
CHX	Chyclohexamide
DHP	3,4-Dehydro-L-Proline
DMS	Dimethylsulphate
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
EIF4A	Eukaryotic Initiation Factor 4A
EIL1	EIN3 Like – 1
EIN3	Ethylene Insensitive 3
EMSA	Electrophoretic mobility shift assay
EXT	Extensin
FADS2	Fatty Acid Desaturase 2
FANCJ	Fanconi Anaemia Complementation Group J
FGF-2	Fibroblast Growth Factor 2
FRET	Forster Resonance Energy Transfer
FTD	Frontotemporal Dementia
G4	G-Quadruplex
GTL-1	GT-2 Like 1
GUS	Beta-Glucuronidase
HDG11	Homeodomain GLABRA 2-like protein 11
HRGP	Hydroxyproline Rich Glycoprotein
HYP	Hydroxyproline
KAN4	Kanadi 4
KNOX	KNOTTED – Like Homeobox
LRR	Leucine Rich Receptor

LRX	Leucine-rich Repeat Extensins
LTR	Long Terminal Repeat
MED25	Mediator 25
MYC	Myelocytomatosis Viral Oncogene
MS	Murashige & Skoog basal salt mixture
NHIII	Nuclease Hypersensitive Element III
NMM	N-Methyl-Mesoporphyrin
NMR	Nuclear Magnetic Resonance Spectrometry
NRAS	Neuroblastoma RAS Viral (V-Ras) Oncogene
PCR	Polymerase Chain Reaction
PERK	Proline-rich Extensin-like Receptor Kinase
PIC	Preinitiation Complex
PFT1	Phytochrome and Flowering Time 1
PIF1	Petite Integration Frequency 1
PG4	Putative G - Quadruplex
POLII	RNA Polymerase II
RHD6	Root Hair Defective 6
RNA	Ribonucleic Acid
RSL1	RHD6 Like - 1
RSL4	RHD6 Like - 4
RT	Reverse Transcription
SEQ	Sequencing
SMXL4/5	Suppressor of Max2 1-Like4/5
SP	Serine - Hydroxyproline
TBS	Tris-Buffered Saline
TFIIH	Transcription Initiation Factor Complex
UTR	Untranslated Region
VEGF	Vascular Endothelial Growth Factor
WRN	Werner Syndrome Protein
XPB	Xeroderma Pigmentosum Type B
XPD	Xeroderma Pigmentosum Type D

1. Introduction

1.1 G-quadruplex Structure

In 1910, almost 50 years before Watson, Crick and Franklins's discovery of the double helix, Ivar Bang discovered that guanine solutions could associate to form gels (Bang, 1910). However, the molecular basis of this interaction was not determined until 1962 when x-ray crystallography was used to show that guanosine monophosphate could form what is now known as a G-quadruplex (G4)(Gellert et al, 1962). G4s are secondary structures that form in nucleic acid sequences rich in guanine residues. The guanine nucleotides interact through Hoogsteen base pairing rather than the classical Watson-Crick base pairing, in which the N1-H and N2-H of one guanine molecule act as hydrogen donors to the O6 and N7 atoms of the adjacent guanine (Fig 1.1). Through Hoogsteen bonding, four guanines associate to form a planar structure called a G-quartet (Gellert et al., 1962; Sen and Gilbert, 1988; Sundquist and Klug, 1989a). The G-quartet can then stack through hydrophobic interaction to form a G4 (Fig 1.1, B).

Within the G4, the oxygen molecules of the guanines face inwards towards each other, which could result in electrostatic repulsion. To prevent this repulsion, mono or divalent cations sit in the central channel of the G4 and form dipole interactions with the oxygen atoms to neutralise the repulsion. Calculations and structural studies have shown that cation binding is essential for G4 formation. *In vitro* and in the presence of K^+, the human telomeric sequence preferentially forms a G4 over Watson-Crick base

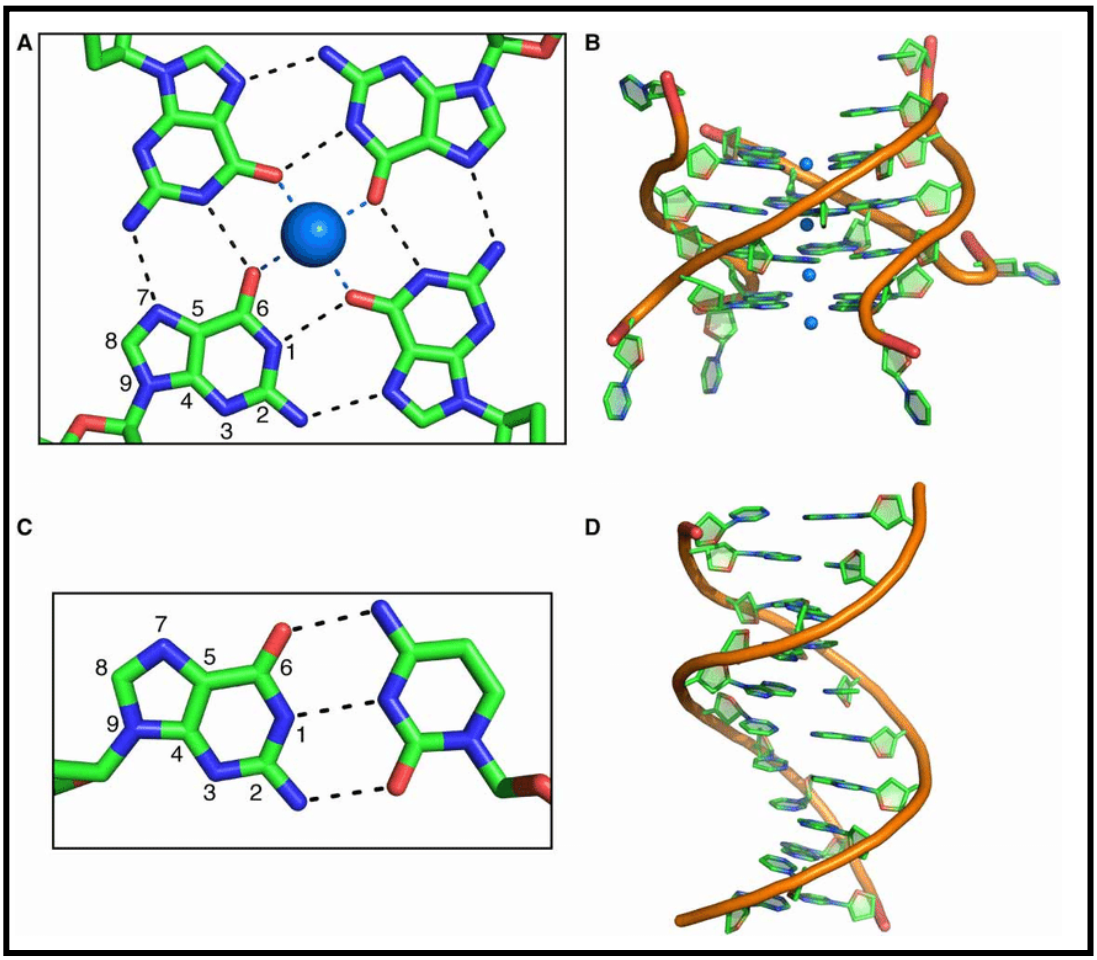

Figure 1.1 Hoogstein vs Watson - Crick base pairing. (A) Hoogstein base pairing of 4 guanines forming a tetrad, stabilised by a central cation. (B) Structure of a parallel G4 (PDB: 244D). (C) Watson–Crick base pair of guanine and cytosine. (D) Structure of a DNA double-helix (PDB: 1BNA). This figure is taken from (Weldon et al., 2016)

pairing up to temperatures of 30^{o}C (Clay and Gould, 2005; Dingley et al., 2005; Hardin et al., 1991; van Mourik and Dingley, 2005; Schultze et al., 1999; Spačková et al., 1999). K^{+} and Na^{+} are abundant within cells, therefore their role stabilising G4s has been extensively studied. However, many other cations have been found to stabilise G4, generally in the order of $Sr^{2+} > K^{+} > Ca^{2+} > Na^{+} > Mg^{2+} > Li^{+}$, with Sr^{2+} resulting in the

strongest stabilisation of G4s and Li+ the least in most cases (Hardin et al., 1992; Kankia and Marky, 2001; Worlinsky and Basu, 2009). The cations can form dipole interactions with up to 8 oxygen molecules from two stacked G-quartets to stabilise the G4 and hence there is commonly one less cation molecule than the number of G-quartets in a G4 (Deng et al., 2001; Rovnyak et al., 2000; Wu et al., 2003).

G4s form in both DNA and RNA sequences, in a wide range of conformations. Studies have shown that G4s can form with the DNA/RNA strands in the same direction (parallel), opposite directions (anti-parallel) or a combination (hybrid) (Ambrus et al., 2006; Parkinson et al., 2002; Phan and Patel, 2003). Additionally, G4s occur within a single molecule (intramolecular) or from up to four molecules (intermolecular) (Fig 1.2) (Burge et al., 2006; Hazel et al., 2006; Marušič and Plavec, 2015; Merkina and Fox, 2005; Víglaský et al., 2010). However, intermolecular G4s are thought to rarely occur *in vivo* due to the low DNA stand concentration meaning that it would be unusual for the required sequences to be in close enough proximity to form the structure (Huppert and Balasubramanian, 2005).

The conformation of an intramolecular G4 can be partially be determined by its sequence. Intramolecular G4 are generally thought to have a canonical sequence of $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ where G represents the four contiguous runs of 3 or more guanines within a tract that form the pillars of the G4 and N represents the loop regions which can consist of up to seven nucleotides of any base that connect the quartets (Huppert and Balasubramanian, 2005). The loops can connect the guanine pillars

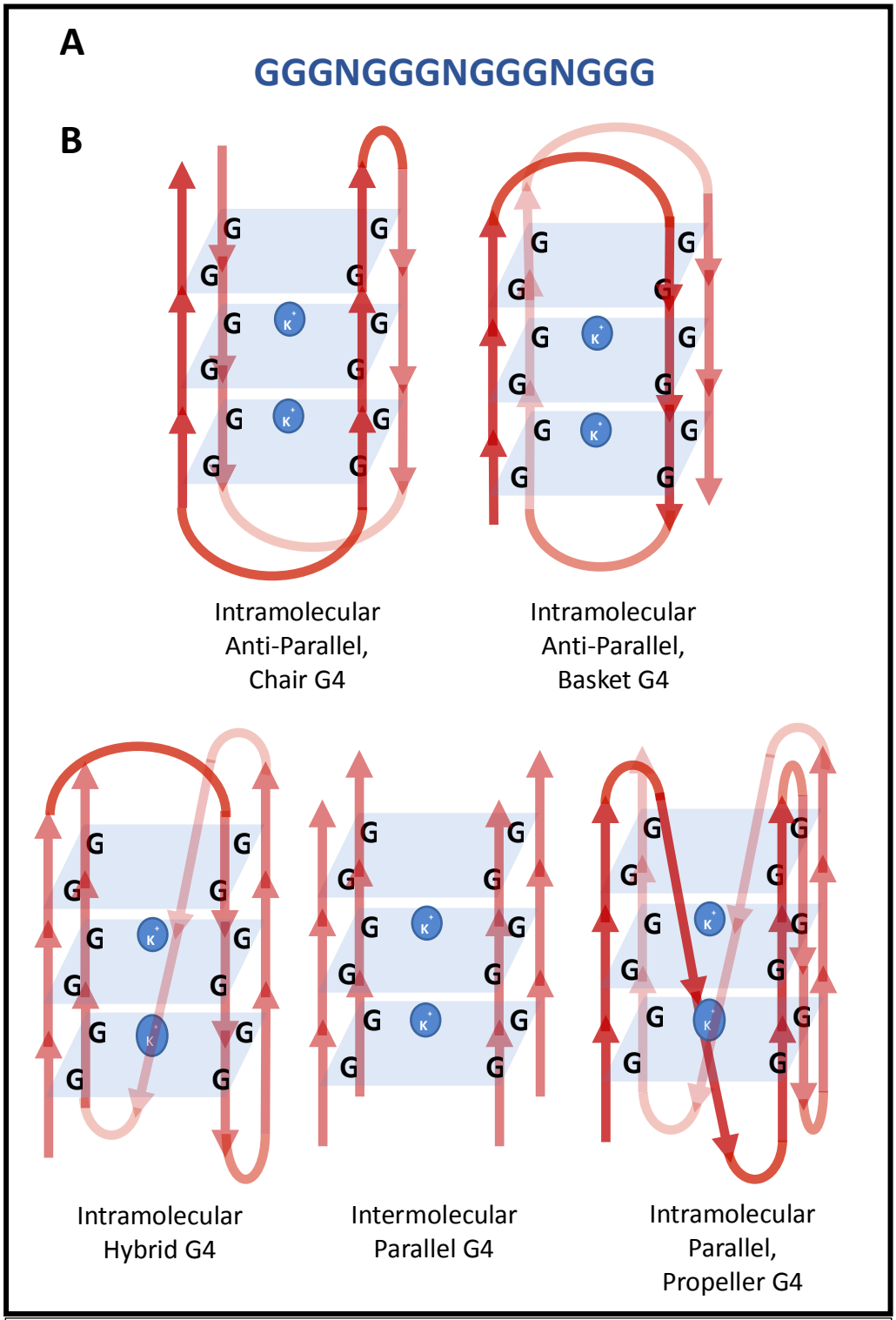

Figure 1.2 G-Quadruplex conformations. **(A)** A potential G4 forming sequence with four consecutive runs of three guanines separated by, N, loops and **(B)** several of the possible conformations that the sequence could form.

laterally or diagonally across the tetrad, with G4s only containing lateral loops referred to as “chair” like and G4s containing diagonal loops referred to as “basket” like. Alternatively, the loops can form a “propeller” like fold and result in a parallel G4 (Fig 1.2). However, G4 conformation is not solely dependant on sequence, it has been shown that the cations present, flanking sequences and G4 ligands can also alter the conformation formed within a sequence. Alterations in G4 conformation within a set sequence has been seen in the human telomeric sequence, which forms an antiparallel basket G4 in the presence of $Na^{+,}$ whereas in the presence of K^{+} the G4 formed is a mixture of the antiparallel basket conformation and a parallel propeller conformation (Ambrus et al., 2006; Gaynutdinov et al., 2008; Parkinson et al., 2002; Phan and Patel, 2003). Additionally, the number of contiguous guanines within the sequence accounts for how many G-quartets can form and stack up to assemble an intermolecular G4. Furthermore, the number of quartets also alters the stability of the G4. The more G-quartets stacked to form a G4, the more thermostable the G4 is (Petraccone et al., 2005)

1.2 G-Quadruplex Discovery Methods

1.2.1 G-Quadruplex prediction

In 2005 Huppert and Balasubramanian designed the “Quadparser” canonical sequence, $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+,}$ to aid in the identification of G4s based on their sequence requirements (Huppert and Balasubramanian, 2005). This canonical sequence has successfully been used to predict the prevalence of G4 in many species and significantly advanced G4 research. However, in designing the canonical sequence several assumptions were made.

The first was that *in vivo* most G4s are likely to be intramolecular. Although it has been shown that G4s can form between multiple DNA/RNA molecules (Merkina and Fox, 2005; Víglaský et al., 2010), the concentration of DNA/RNA within cells is much lower than concentrations used i*n vitro* and therefore, makes intermolecular G4s unfavourable in a cellular environment.

The second assumption made is concerning the number of quartets within a G4. The canonical sequence contains four contiguous runs of 3 or more guanines, as it's these runs that form the pillars of the G4 and determines the number of quartets. This therefore assumes a G4 has a minimum number of three quartets. This assumption is based on the lower stability of two quartet G4s although there is evidence that two quartet G4 can form under biological conditions (Bock et al., 1992; Fry and Loeb, 1994; Petraccone et al., 2005; Smirnov and Shafer, 2000). Single quartet G4s were also identified in the initial *in vitro* experiments by Gellert et al in which the quartets were first discovered; however, this was at mM concentrations of guanine, much higher than the concentration within the cell (Gellert et al., 1962).

A third assumption was that the G4 requires four contiguous guanine runs although structures have been identified where G-quartets can be formed containing other bases (Caceres et al., 2004; Escaja et al., 2003; Matsugami et al., 2001; Patel and Hosur, 1999; Silva, 2003), 'slipped' structures where quartets of two G4 slip to form a larger dimerized G4 (Krishnan-Ghosh et al., 2004) or sequences with bases within the contiguous runs of guanines resulting in short bulges within the G4 (Mukundan and Phan, 2013).

The final assumption made was the length of the loops within the G4. The Quadparser canonical sequence suggests loop lengths of 1-7 bases as it has been shown that G4 stability usually decreases with increasing loop length (Hazel et al., 2004). However, G4s with much longer loop sequences of up to 30 nucleotides have been identified to be stable above physiological temperatures (Guédin et al., 2010).

These assumptions cause a significant amount of false negative results with Chambers et al finding that approximately 70% of G4s in the human genome (observed using the G4-seq technique, discussed later) do not conform to the canonical sequence (Chambers et al., 2015). Therefore, several adjustments have been made to the canonical sequence to allow for 2 quartet G4s, longer loop lengths and short sequences within the contiguous guanine runs. However, these predictive tools can still produce false positive and false negative results. Therefore, while prediction tools are extremely useful in the first instance of identifying whether a G4 has the potential to form, further structural studies are required to determine whether a G4 forms for certain.

1.2.2 G-Quadruplex structural analysis

Studies of G4 structures have identified several biophysical characteristics that can be utilised to allow identification of G4s with a high degree of certainty. However, how accurate these techniques are for predicting G4s that form under biological conditions is questionable. All of these techniques use oligos under experimental conditions and therefore do not take in to account the impact of molecular crowding, the competition of other secondary structures, supercoiling and in the case of the DNA, what is happening on the other strand. Therefore, additional studies are required under physiological conditions or *in vivo* to confirm biological relevance. Despite their

limitations, these techniques have provided a breadth of knowledge on the sequence specificity and required conditions for G4 formation.

These methods include electrophoretic mobility shift assay (EMSA), dimethylsulphate (DMS) foot printing and Förster resonance energy transfer (FRET). Although, UV or Circular Dichroism (CD) melting curves and nuclear magnetic resonance spectrometry (NMR) tend to be the gold standard for determining whether sequence can form a G4 structure. CD looks at how molecules absorb circularly polarised light, of which G4s have a distinct absorption pattern. Parallel G4s produce a spectrum with a peak at ~264nm and a trough at ~245nm, whereas anti-parallel G4s have a peak at ~295nm max and a trough at ~260nm min and finally the hybrid G4 has a peak at ~295nm max and ~260nm max and a trough at ~245 nm min (Chen, 1992; del Villar-Guerra et al., 2018). On the other hand, NMR detects proton exchange within the structure and utilises the protected guanines of the G4 structures to identify them (Wang and Patel, 1992; Webba da Silva, 2007).

1.2.3 G-Quadruplex ligands

Several G4 ligands have been identified, or designed, for use in research and as potential drugs. As G4s have many possible conformations, there are also many potential ligands with different binding mechanisms and different affinities to particulars G4s. The G4 ligands usually bind by either: external stacking (binding to the hydrophobic surface of the outer tetrad), intercalating (binding in the groove between the tetrads) or external groove binding (binding in the groove of the G4 or interacting with the loop regions) (Monchaud and Teulade-Fichou, 2008). There are now more than 800 known small

molecules that bind G4s, which can be found in an online database (Li et al., 2013), some of which I have detailed below (Table 1). There are complications with using ligands to study G4s as it has been shown that the ligands can induce as well as stabilise the G4; therefore, the binding of a ligand does not mean that a G4 would form in its absence. However, in combination with the several other G4 identification techniques we are now able to identify G4 structures, enabling the study of their functions and further facilitating the design of drugs to target them.

Table 1.1. G4 binding ligands

LIGAND	CLASS	BINDING TYPE	REFERENCES
TMPYP4	Porphyrin	External stacking, intercalating	(Arora et al., 2008; Haq et al., 1999; Nagesh et al., 2010; Phan et al., 2005)
N-METHYL-MESOPORPHYRIN (NMM)	Porphyrin	External stacking	(Nicoludis et al., 2012)
BERBERINE	Ellipticine	External stacking	(Bazzicalupi et al., 2013; Franceschin et al., 2006; Li et al., 2017; Wu et al., 1999)
GQC-05	Ellipticine	External stacking	(Brown et al., 2011; Ghosh et al., 2013)
PYRIDOSTATIN	Quinolinyl pyridine dicarboxamide	External stacking	(Müller et al., 2010; Rodriguez et al., 2012a)
QUARFLOXIN	Fluoroquinolones	External stacking	(Brooks and Hurley, 2010; Drygin et al., 2009; Kim et al., 2003)
BRACO-19	Acridine	external groove binding	(Moore et al., 2005; Schultes et al., 2004; Yi-Meen Chou et al., 1999)

1.3 G-Quadruplex Biological functions

The canonical sequence has been used to predict G4s in the genomes of many organisms and has suggested the presence of over 700,000 potential G4s in the human genome (Chambers et al., 2015) and 40,000 in Arabidopsis (Mullen et al., 2010). The predicted G4s within these genomes were universally enriched in known functional areas of the genome such as promoters, origins of replication and most significantly telomeres (Capra et al., 2010; Garg et al., 2016a; Hershman et al., 2008; Huppert, 2010; Nakken et al., 2009). This conservation of G4s and their enrichment within the genome strongly suggested G4s may have biological functions and there are now many studies highlighting these functions.

1.3.1 G-quadruplexes and telomere maintenance

Telomeres are repetitive DNA sequences with single-stranded 3' overhangs, found at the end of chromosomes that are vital in protecting the chromosomes from degradation, end to end fusions and inappropriate activation of a double-strand break repair (Makarov et al., 1997; Moyzis et al., 1988; Zakian, 2012). The telomere 3' overhang is G-rich and was one of the first biological sequences shown to form a G4 within several organisms using non- denaturing PAGE gels and G4 specific antibodies (Henderson et al., 1987; Schaffitzel et al., 2001; Sen and Gilbert, 1988; Sundquist and Klug, 1989b).

There are also several proteins found associated with telomeres, several of which have been found to bind G4s. These include the Telomere End-Binding Proteins (TEBPs) which protect the telomeres from being recognised by the DNA damage response

pathways. TEBPα and TEBPβ in *S. lemnae* and TEBP and Rap1 in *S. cerevisiae,* have been shown by structural studies and immunofluorescence to facilitate the formation of G4s (Giraldo and Rhodes, 1994; Juranek and Paeschke, 2012; Moye et al., 2015; Paeschke et al., 2005, 2008; Postberg et al., 2012). On the other hand, TELOMERASE, RECQ helicases and PROTECTION OF TELOMERIC PROTEIN 1 (POT1) are all involved in elongating telomeres, and in humans, these proteins have been found to resolve G4s to allow telomere extension (Moye et al., 2015; Zaug et al., 2005). These combined findings suggest that G4s may be involved in regulating the elongation and protection of telomeres. Understanding this mechanism is of particular interest as telomeres are crucial for genome stability as well as the cell life cycle. The effect of G4s on telomerase has been of particular interest as telomerase is an enzyme that replicates and elongates telomeres and therefore can increase the cell lifespan and is found upregulated in 85% of cancers (Shay and Wright, 2011). Studies have shown that G4 have the ability to allow or block telomerase activity depending on their conformation (Oganesian et al., 2007; Zahler et al., 1991), meaning telomeric G4s could be a novel target in treating cancer.

1.3.2 G-quadruplexes in DNA replication

During DNA replication, the DNA double helix is unwound by helicases to form single-stranded DNA to enable replication. It is hypothesised that the presence of single-stranded DNA during this process provides increased potential for G-rich regions to form G4s. Immunostaining using the BG4 antibody which preferentially binds to intramolecular G4s has shown the amount of G4s increases during the S-phase of the cell cycle in which the DNA is replicated (Biffi et al., 2013). This finding suggests that more G4s may form while the DNA is single-stranded during replication. However, it has

been well established that G4s are capable of stalling polymerases *in vitro* (Han et al., 1999a; Howell et al., 1996; Sun and Hurley, 2010; Takahashi et al., 2017; Woodford et al., 1994) and there is increasing evidence that, in the absence of helicases to resolve them, G4s also stall polymerases *in vivo* (Johnson et al., 2010a; Kruisselbrink et al., 2008; London et al., 2008; Rodriguez et al., 2012a; Van Wietmarschen and Lansdorp, 2016). Therefore, for replication to occur the G4s would need to be unwound. As previously mentioned, there have been suggestions of helicases that can unwind telomeric G4s and further studies have shown that the helicases: PETITE INTEGRATION FREQUENCY 1 (PIF1), Fanconi anaemia complementation group J (FANCJ), WERNER SYNDROME PROTEIN (WRN) and BLOOM (BLM) resolve G4 structures (Kruisselbrink et al., 2008; Singleton et al., 2007; Sun et al., 1998a; Wu and Spies, 2016). Loss of these helicases causes increased DNA breaks, deletions, and slowed DNA replication potentially due to DNA polymerase stalling and replication fork collapse (Davies et al., 2007; Kruisselbrink et al., 2008; Paeschke et al., 2011; Piazza et al., 2015; Ribeyre et al., 2009; Rodriguez et al., 2012a; Van Wietmarschen and Lansdorp, 2016; WU, 2007). Genomic deletions have also been seen in human patients with mutations in these helicases (German, 1993; London et al., 2008).

Conversely, G4s may also have a role in promoting DNA replication. It has been identified that the origins of replication in humans are mainly GC rich sequences 67% of which conformed to a consensus G4 sequence and 91% conformed to a G4 sequence with an extended loop length of up to 15bp (Besnard et al., 2012). It was also identified that the more G4 motifs within a region the more commonly that region is used as an origin of replication (Besnard et al., 2012). Additionally, it was found that point

mutations in two G4s, found at two origins of replication, completely abolished replication origin formation and hence inhibited DNA replication further suggesting a role for G4s in the promotion of DNA replication (Valton et al., 2014).

1.3.3 G-quadruplexes in transcription

It has been hypothesised that G4s can potentially affect transcription in several ways depending on the location of the G4. The expression of genes with G4s in the promoter have been found to be significantly regulated when treated with G4 ligands (Cogoi and Xodo, 2006; Cogoi et al., 2014; David et al., 2016; Siddiqui-Jain et al., 2002; Yu et al., 2012). Hence, it has been hypothesised that G4s in promoters may allow the binding of transcription factors, repressors or mediate the transition of DNA from double-stranded to single-stranded to regulate transcription (Lam et al., 2013). Bioinformatics analysis suggests that 40% of human promoters contain predicted G4s with a prominent enrichment in oncogenes and depletion in tumour suppressor genes (Eddy and Maizels, 2006; Huppert and Balasubramanian, 2007). Additionally, it has been shown that G4s are associated with areas of open chromatin and the promoters of actively transcribed genes (Hänsel-Hertsch et al., 2016; Huppert and Balasubramanian, 2007). This finding may suggest an important role of G4s in the correct transcription of genes. Genome-wide expression studies in yeast and humans have shown that G4 stabilising ligands, and G4 antibodies can cause global changes in gene expression (Fernando et al., 2009; Hershman et al., 2008; Verma et al., 2008). One of the best examples of G4s altering gene expression is in the oncogene *c-MYC* (Siddiqui-Jain et al., 2002). The nuclease hypersensitive element III (NHE III) within the c-MYC promoter

controls 80% of c-MYC transcription and was shown to form G4s in vitro (Ambrus et al., 2005; Seenisamy et al., 2004; Siddiqui-Jain et al., 2002; Simonsson et al., 1998). It has been demonstrated that G4 ligands cause a decrease in *c-MYC* mRNA levels; while mutating out the G4 within the NHE III element causes an increase in expression (Brown et al., 2011; Siddiqui-Jain et al., 2002). Two NHE III binding factors, Nucleolin and NM23-H2, have since been shown to bind to the G4 thought to regulate *c-MYC* expression (Kopylov et al., 2015) (González et al., 2009). Additionally, overexpression of NUCLEOLIN has been shown to downregulate c-MYC expression by promoting G4 formation (González and Hurley, 2010; González et al., 2009).

While G4s may provide binding sites for specific promoter proteins, there is also potentially a more global role that G4s play in the initiation of transcription. The transcriptional initiation factor complex (TFIIH) is a vital protein complex for both transcription and nucleotide excision repair (Compe and Egly, 2016). XPD and XPB are two components of TFIIH that have been shown to bind G4s with a higher affinity than double-stranded DNA and XPD can also unwind G4s (Gray et al., 2014a). Additionally, XPD and XPB have been shown, by ChIP-seq and G4 antibody ChIP-seq, to be enriched at transcription start sites that contain G4s (Gray et al., 2014a; Hänsel-Hertsch et al., 2016). Suggesting G4s may enable recruitment of the TFIIH and facilitate transcription initiation.

G4s can also form within the coding sequence of a gene. Genes that have G4 on the coding strand tend to have higher expression than those without (Du et al., 2008), it has therefore been proposed that G4s on the coding strand will increase the stability of

the single-stranded DNA and facilitate the binding of PolII to increase transcription (Rhodes and Lipps, 2015). Alternatively, if the G4 forms on the template strand then this is proposed to decrease PolII processivity by blocking the passage of PolII down the coding sequence and blocking the gene from being transcribed (Bochman et al., 2012). The ability of G4 to stall polymerases has been well established (Han et al., 1999a; Howell et al., 1996; Sun and Hurley, 2010; Takahashi et al., 2017; Woodford et al., 1994) and this coincides with a finding that genes with a higher GC content have slower PolII elongation rate (Veloso et al., 2014) and the result of transcription-dependent DNA damage caused by G4 ligands (Rodriguez et al., 2012a).

1.3.4 G-quadruplexes in splicing

Pre-mRNA splicing can be a co-transcriptional process that can be effected by PolII processing speeds (Jonkers and Lis, 2015; Mata et al., 2010). Therefore, G4 are likely to affect alternative splicing through G4 dependent stalling PolII during transcription. However, G4s may play a more direct role in regulating splicing. Studies have shown that 50% of splicing boundaries in humans contain a PG4s within the first 100bp of intronic sequence (Eddy and Maizels, 2006) and several of the hnRNP protein family were shown to be enriched at these G-rich sequences, regulate splicing and bind to G4s (Conlon et al., 2016; Dominguez et al., 2010; Liu et al., 2017; Samatanga et al., 2013; Xiao et al., 2009). This is well studied within the apoptosis-regulating gene *BCL-X*. *BCL-X* has two isoforms, *BCL-XL*, which is anti-apoptotic and the shorter form, *BCL-XS*, which is pro-apoptotic (Boise et al., 1993). Alternative splicing of *BCL-X* was previously identified as being regulated by hnRNP F/H binding to two G-rich regions, one upstream

of the *BCL-XL* donor site, and another downstream of the *BCL-XS* donor site (Garneau et al., 2005). Both of the G-rich regions in *BCL-X* were shown to form G4s *in vitro* and it was found that G4 formation at the *BCL-XS* donor site removed a long stem-loop structure that inhibited donor site usage, while the G4 formation at the *BCL-XL* donor site actually overlapped with the donor site and therefore blocked its usage (Weldon et al., 2018a, 2018b). Thus, G4 formation in *BCL-X* gene may act as a complex molecular switch to control splicing.

1.3.5 G-quadruplexes in translation

It has been suggested that G4s may have more significant roles within RNA than DNA as they have been shown to be more stable due to the additional hydroxyl groups of the ribose (Collie et al., 2010). G4s are particularly enriched within the 5' untranslated region (UTR) of genes. G4s have been identified in the 5' UTR of many genes including the Arabidopsis *ATAXIA TELANGIECTASIA-MUTATED AND RAD-3 RELATED GENE (ATR)*, the human *ZIc-1* gene and several human oncogenes (*NRAS*, *BCL-2*, *FGF-2*, and *VEGF*). In most cases, the presence of the 5'UTR G4 has been found to inhibit translation when folded (Arora et al., 2008; Kumari et al., 2007; Kwok et al., 2015; Morris et al., 2010; Shahid et al., 2010). It is thought that the reduced translation caused by G4s in the 5' UTR may be due to the G4 preventing recognition of the translation start site by the preinitiation complex (PIC), an important factor in the initiation of translation (Bugaut and Balasubramanian, 2012). Also, it has been found that the RNA helicase Eukaryotic Initiation Factor 4a (EIF4A) is necessary for the translation of mRNA transcripts containing G4 forming motifs within their 5' UTR (Wolfe et al., 2014), suggesting that the cell may have a mechanism to overcome G4 dependent downregulation of translation.

There is also evidence of G4s forming within RNA to promote translation. It has been demonstrated that the G4 within the 5' UTR of *VEGF* is required for initiation of translation (Morris et al., 2010). Additionally, G4s are also found enriched in the 3' UTRs of mRNA. One study of a G4 in the 3' UTR of *FADS2* mRNA found that G4 formation prevented microRNA binding resulting in an upregulation of *FADS2* (Rouleau et al., 2017).

1.3.6 G-quadruplexes and epigenetics

Epigenetics is thought to be important in regulating the formation of G4s. G4s are shown to be over-represented at sites of low histone occupancy and relaxation of chromatin by histone deacetylase inhibitors caused an increase in G4 formation (Hänsel-Hertsch et al., 2016; Hegyi, 2015; Hershman et al., 2008). Additionally, ATRX is a protein that has both helicase and chromatin remodelling capabilities. ATRX stabilises chromatin by depositing H3.3 histones, which is thought to maintain the DNA double helix and decrease the presence of alternative DNA secondary structures (Clynes et al., 2013). It has also been identified that ATRX binds specifically to G-rich tandem repeats and can bind to G4s and resolve them (Law et al., 2010). Mutating ATRX caused the silencing of the α-globulin gene cluster (Law et al., 2010). The silencing of α-globulin may be due to the formation of G4s within the coding sequencing and blocking RNA polymerase II (PolII) from transcribing the gene. Supporting this theory is the demonstration that ATRX deficiency is linked to reduced H3.3 and PolII stalling (Levy et al., 2015). Therefore, it is thought that ATRX is essential in maintaining genome stability by allowing guanine-rich DNA to be replicated and transcribed through the incorporation of H3.3 to prevent the formation of secondary structures.

Conversely, G4s have also been shown to have an impact on the maintenance of epigenetic marks. Histones are recycled during DNA replication to maintain epigenetic marks (De Koning et al., 2007). However, as previously discussed G4s can cause stalling of DNA replication and it has been found that this stalling also leads to the uncoupling of histone recycling, causing new unmethylated histone to be incorporated into the chromatin. The uncoupling of histone recycling also resulted in gene expression changes (Eddy et al., 2014; Guilbaud et al., 2017; Sarkies et al., 2010, 2012, Schiavone et al., 2014, 2016). Furthermore, there is evidence that double strand breaks, caused by G4 formation, induce phosphorylation of histone H2AX (Ayrapetov et al., 2014; Rodriguez et al., 2012a; Shanbhag et al., 2010).

1.4 G-Quadruplexes in Plants

As discussed, there is now a large body of evidence as to the potential biological roles of G4s, but these studies were primarily in mammals or yeast. In contrast, there is little known about the role of G4s within plants however there is a growing interest in this field.

The density of putative G4s (PG4s) in plants is generally lower than in mammals, for example, *Arabidopsis thaliana* has a density of 9.3 PG4/Mb compared to 115 PG4/Mb in *Homo sapiens* (for three quartet PG4s), which may make Arabidopsis a good model to use to study G4 function (Mullen et al., 2010). Monocots have been found to have more predicted PG4s than dicots (Garg et al., 2016b), which is likely linked with the finding that monocots have a higher GC content within their DNA (Šmarda et al., 2014). Additionally, Mullen et al noted that Arabidopsis had 43000 two quartet PG4s, which

they postulate may be stable at the environmental temperature ranges that Arabidopsis is usually found (Mullen et al., 2010, 2012). Furthermore, G4s containing two tetrads were conserved in all dicot species in 1331 genes (Garg et al., 2016b). Enrichment of PG4s was found at telomeres, transposable elements, the transcription start site of genes (particularly on the template strand), intergenic regions and 5' untranslated regions (UTR) and in several species of plants (Andorf et al., 2014; Garg et al., 2016b; Lexa et al., 2014; Mullen et al., 2010; Takahashi et al., 2012).

Similar to the vertebrates studied, Arabidopsis telomeres are G rich consisting of the repetitive sequence TTTAGGG (Richards and Ausubel, 1988). The telomeric sequence of both Arabidopsis and *O. sativa* binds the telomerase inhibitor telomestatin, known to bind to telomeric G4s, resulting in telomere shortening (Zhang et al., 2006). These results suggest that plant telomeres are also likely to form G4s that are important in telomere protection.

71% of non-telomeric G4s in *Zea mays* were found in repetitive sequences (Andorf et al., 2014) and PG4s are particularly enriched in the highly repetitive long terminal repeat (LTR) retrotransposons (Lexa et al., 2014). Younger LTRs commonly contain longer guanine runs, which could suggest a role of G4s in the lifecycle of LTRs (Tokan et al., 2018). Additionally, LTR sequences are commonly incorporated into different regions of the plant genome through recombination; this has been suggested as a mechanism of introducing novel G4s into the genome which may evolve functions (Kejnovsky and Lexa, 2014).

One proposed G4 function of particular interest is in plant stress response. Mullen et al identified that two quartet G4s are particularly enriched within the genes body and they suggest that these two quartet G4s are only stable at high K^+ concentration (Mullen et al., 2010, 2012). This gives rise to the hypothesis that G4 could act as sensors of high salt concentrations which occurs when the plant is experiencing drought stress. 16% of the Arabidopsis genome is responsive to drought, and of these genes, 45% were found to contain PG4s in the annotated cDNA sequence which is significantly more than the 31% occurrence in all genes (Garg et al., 2016a). Also, in *Z. mays* PG4s were enriched in hypoxia and nutrient deprivation response gene ontology classes.

Some of the PG4s identified in 5' UTRs have been shown to form a G4 in RNA. A G4 in the 5' UTR of the *ATR* gene in Arabidopsis forms within the RNA and was shown to inhibit translation within a transient reporter gene assay (Kwok et al., 2015). Additionally, a G4 in the 5' UTR of *SMXL4/5* (a gene involved in phloem differentiation) was found to be conserved within all vascular plants analysed. The G4 within the 5' UTR of *SMXL4/5* was found to bind a protein called JULGI which was found to inhibit translation and phloem development in a G4 dependent manner (Cho et al., 2018).

G4s have been further linked to plant development in a study by Nakagawa et al. Arabidopsis seedling treated with the G4 stabilisers NMM and Berberine, displayed leaf polarity defects (Nakagawa et al., 2012). Additionally, Arabidopsis with a double mutation in *ASYMMETRIC LEAVES 1* and *2* displayed more severe leaf phenotypes when

treated with NMM and Berberine (Nakagawa et al., 2012). These findings suggest a novel function of G4s in plant development.

1.5 Extensins and the cell wall

The suggested role of G4s in plant growth and development and unpublished data by Parker. M from our lab has shown that cell wall genes, in particular the *EXTENSIN* (EXT) gene family, are enriched for two tetrad G4s and are strongly downregulated by G4 stabilisation by NMM and Berberine (Parker, 2018).

1.5.1 The cell wall in growth and development

Plants have two primary mechanisms of growth and development; cell division and cell expansion. Plant cell division is initiated from two regions of stem cells, one of which is at the tip of the shoot and hence called the shoot apical meristem and the other is at the primary root tip and called the root apical meristem. Within these meristematic zones, the stem cells divide, both maintaining the stem cell population while also generating daughter cells that differentiate and this is controlled by the cell cycle and plant hormones. After cell division and differentiation further growth occurs through cell elongation. The cell wall structure majorly controls cell elongation. The plant cell wall is a rigid outer barrier to the cell that allows them to withstand the turgor pressure created by the vacuole within the cell. Hence, the cell wall limits cell elongation and thus to allow cell elongation the cell wall structure must be altered.

1.5.2 Hydroxyproline-Rich glycoprotein superfamily

EXTs form one of the major groups of the hydroxyproline-rich glycoprotein (HRGP) superfamily along with arabinogalactans, solanaceous lectins and proline-rich proteins (Allen and Neuberger, 1973; Averyhart-Fullard et al., 1988; Chan Hong et al., 1987; Chen

and Varner, 1985; Fincher et al., 1974; Lamport et al., 2011). In Arabidopsis, bioinformatic analysis identified 59 EXT related proteins, 18 Proline-rich proteins and four arabinogalactan – extensin hybrids (Johnson et al., 2017; Showalter et al., 2016). HRGPs were identified as a critical component in the primary cell wall in 1960 (LAMPORT and NORTHCOTE, 1960). The enrichment of hydroxyproline (HYP) suggested a potential role for HRGPs as structural elements of the cell wall (Shoulders and Raines, 2009). HYP is important in structural proteins as it restricts side chain rotation and the hydroxyl group of HYP can be involved in carbohydrate attachment. Lamport proposed that the HYP may control protein crosslinking within the cell wall and the ability of the cell to extend during cell growth, hence EXTs were named initially due to their proposed role in cell extension (Lamport, 1963).

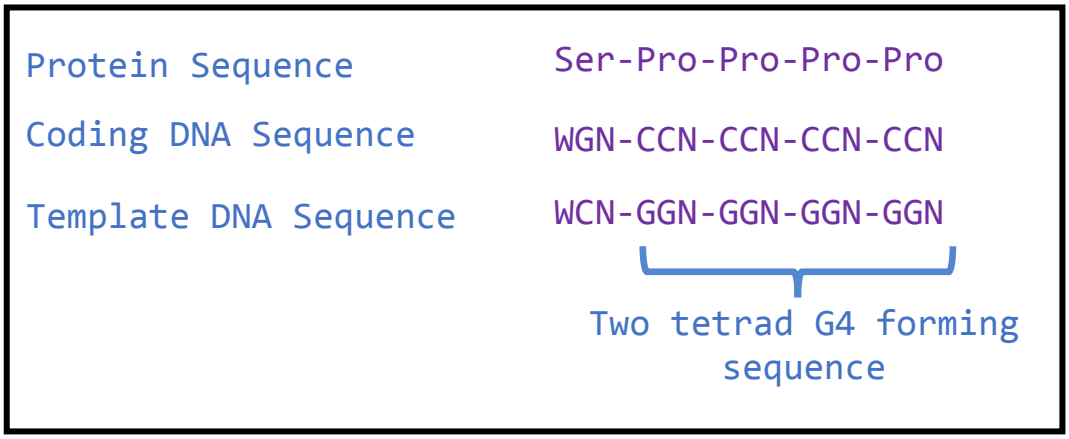

Figure 1.3. The Extensin SP4 motif. The protein sequence, coding sequence and template sequence of the EXT SP4 motif demonstrating the hardcoded two tetrad G4 forming sequence.

1.5.3 Extensins and cell wall integrity

There are several subclasses of EXTs, but the majority can typically be classed as classical EXTs, Leucine-rich repeat extensins (LRXs) and proline-rich extensin-like

receptor kinase (PERKs). The classical EXTs are typically composed almost entirely of the X-HYP_n EXT motif (Fig 1.3), where X is most commonly serine and n is typically three or 4 but can be up to 6 (Campargue et al., 1998; Qi et al., 1995). Whereas the LRXs have three distinct domains with an N-terminal leucine-rich receptor (LRR) domain, followed by a typical EXT HYP-rich domain and then a C-terminal domain. The role of EXTs is visible in root hairs with mutations in *LRX1* resulting in several abnormal root hair morphologies (Baumberger et al., 2001, 2003; Ringli, 2010) and mutants in *EXT6*, *EXT7*, *EXT10*, *EXT11* and *EXT12* display short root hairs (Velasquez et al., 2011a, 2011b). However, it is postulated that EXTs may be genetically redundant as only mutants defective in *EXT3* displays a near lethal phenotype (Cannon et al., 2008).

1.5.4 Control of extensin expression

Several transcription factors have been linked to controlling *EXT* expression (Fig 1.4). One of the most studied is the basic helix-loop-helix transcription factor ROOT HAIR DEFECTIVE 6 LIKE-4 (RSL4). Several EXTs have been seen to be co-expressed with *RSL4*; *EXT12*, *EXT14*, *EXT18* are seen to be downregulated in the *rsl4-1* loss of function mutant and upregulated in *CaMV35S: RSL4* overexpression lines (Yi et al., 2010). Additionally, EXTs have root hair-specific cis elements that act as RSL4 binding domains and binding of RSL4 has been confirmed for the *LRX1* and *PRP3* promoters and their expression upregulated in response (Hwang et al., 2017). *RSL4* is negatively regulated by tri-helix transcription factor GT-2 LIKE 1 (GTL1) and its homolog DF1 and therefore has been seen to downregulate *EXT* expression and represses root hair growth (Shibata et al., 2018). In addition, ETHYLENE INSENSITIVE 3 (EIN3) and ROOT HAIR DEFECTIVE 6 (RHD6) have also been shown to interact and regulate *RSL4* expression, although, RHD6 and EIN3 are also

thought to be able to regulate some EXT genes independently of *RSL4* through interactions with (RSL1) and EIN3 respectively (Feng et al., 2017).

Additionally, the HD-ZIP transcription factor HDG11, and related proteins MEDIATOR 25 (MED25) and PHYTOCHROME AND FLOWERING TIME 1 (PFT1) regulate the expression of several EXTs in root hairs. EXTs have been found to have an HD-binding cis element which has been shown to bind HDG11 and an *HDG11* gain of function mutant was found to upregulate *EXTENSINS* and have increased primary root length, which Xu et al suggests may be due to increased loosening of the cell wall (Xu et al., 2014). Also, *med25/pft1* mutants display shorter root hairs and downregulation in EXT expression (Sundaravelpandian et al., 2013).

Furthermore, there is evidence of *EXT* gene expression being regulated by hormones. It is known that *RSL4* expression is mediated by auxin and ethylene to result in differential root hair growth (Feng et al., 2017; Yi et al., 2010). It has been shown that ethylene root hair growth is dependent on the EIN3/EIL1 and RHD6/RSL1 complex (Feng et al., 2017). Also high auxin concentration has been found to mediate the interaction between MED25/PFT1 and ARF7/ARF19 to cause downstream signalling that promotes transcription of lateral root genes (Ito et al., 2016; Raya-González et al., 2014), it has therefore been hypothesised due to the role of MED25/PFT1 in EXT expression that a similar mechanism could exist in root hairs to regulate EXT expression (Marzol et al., 2018).

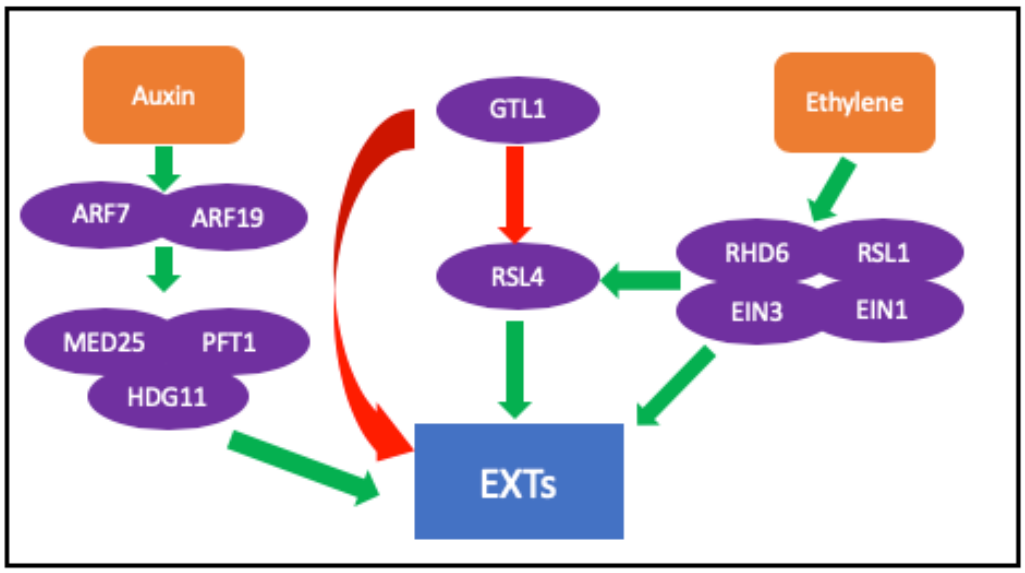

Figure 1.4. Transcriptional regulation of extensins. The proteins involved in regulating EXT transcription with red arrows indicating a negative regulation and green arrows indicating a positive regulation.

1.6 Summary

It was once thought that G4s were simply *in vitro* artefacts; however, there is now a breadth of evidence that G4s are involved in many fundamental biological processes and have implications in several diseases in humans. On the other hand, relatively little is known about the function of G4s within plants. The decreased number of pG4s within Arabidopsis makes it an excellent organism to further study the role of G4s within fundamental biological mechanisms. Also, the indication of G4s affecting plant growth and development provides interesting new potential targets to increase crop yields or identify novel herbicides. Within this thesis, we explore the role of G4s in the growth and development of the model plant species Arabidopsis and how these effects arise from G4s effecting fundamental biological processes.

2. Materials and methods

2.1 Plant Strains

All *A. thaliana* lines used were of the Columbia (Col-0) ecotype. The *lrx1* loss of function*, lrx2* loss of function*, lrx1lrx2 double mutant*, *proLRX1::GUS cosisting of the LRX1* promoter fused to the gus reporter gene, *mycLRX1ΔE153, mycLRX1ΔE90* and *mycLRX1ΔE14* were kindly donated by Prof Ringli (Baumberger et al., 2001, 2003; Ringli, 2010). The *rsl4* line was generously donated by Prof Dolan (Datta et al., 2015; Yi et al., 2010). The helicase mutant plant lines used were obtained from the Nottingham Arabidopsis Stock Centre; *recq4a* (N653278), *recqsim* (N661902), *uvh6* (N6375).

2.2 Plant growth conditions

For all experiments, seeds were surface sterilised by immersing seed in 70% ethanol for 1min, 10% bleach with 0.1% Tween for 5mins, and wash three times in autoclaved water. The seed, immersed in autoclaved water, were stratified for 2-3 days at 4°C and then sown on sterile vertical plates containing Murashige & Skoog (MS) media consisting of; 1x Murashige and Skoog salts plus vitamins (MS; Duchefa Biochemie, M0222), 1% sucrose (Sigma, S0389), 0.05% 4-Morpholineethanesulfonic acid (Sigma, M3671), 0.8% agar and adjusted to pH5.7 using NaOH. The plates were sealed using micropore tape to allow gas exchange, while preventing contamination, then were placed vertically in a Versatile Environmental Test Chamber MLR 350-HT (Sanyo, Japan) with 16 hours photoperiod and an irradiance of 120µmol m^{-2} s^{-1} and a constant temperature of 23°C (experimental growth time refers to the time in growth cabinet after stratification).

Seedlings used to analyse root phenotypes were grown for 4 days on MS agar and then transferred to soft MS agar (with agar content reduced to 0.3%) containing the desired treatment and grown for a further 3 days before analysis. Seedlings used for expression analysis by qRT-PCR were grown on MS plates for 7 days, treated for 6 hours (with the exception of the time course experiment) by flooding the plate with MS liquid media (no agar) containing the treatment. Following treatment, roots and shoots are harvested separately by slicing with a razor blade and instantly freezing in liquid nitrogen. Rosette area and shoot phenotype experiments were grown on flat plates containing MS media for 14 days.

Chemical treatments used were n-methyl mesoporphyrin (NMM) (Frontier Scientific, NMM580), Berberine (Sigma, B32517), ML216 (Sigma, SML0661), 3,4-dehydro-L-proline (DHP) (Sigma, D4893), cycloheximide (CHX) (Sigma, 01810) and the naturally occurring auxin, Indole-3-acetic acid (IAA) (Sigma, I3750). All chemicals were dissolved in either ddH_2O (Fisher, W/0100/21), DMSO (Fisher, BP231), or EtOH (Fisher, E/0650DF/17) and hence control samples were treated with the corresponding solution without the chemical dissolved in it and grown at the same time and under the same conditions as the test samples.

2.3 Root hair phenotype and length observations

Plants used for root hair measurements and morphology experiments were either imaged by light or confocal microscopy. For light microscopy the roots were imaged directly on the plates they were grown on using the STEMI DV4 ZEISS 47 50 22 stereomicroscope with the iOptron iE3000 Electron Eyepiece (#8430) camera or the SP300F Brunel microscope, and processed using Future WinJoe software (Future Optics)

Motic Images Advanced 3.2 Software (Motic) and ImageJ (NIH). For confocal experiments, the roots were removed from vertical plates and stained with 10µg/µL propidium iodide for 1min, rinsed and mounted in water on a microscope slide and imaged using the FV1000 Confocal microscope (Olympus) at the University of Sheffield Light Microscopy Facility. Root, root hair and rosette measurements were then taken using ImageJ software. The propidium iodide stain was visualised using a 514 nm argon laser beam causing the stain to emit between 580-610 nm. Images were analysed using FV10-ASW viewer (Olympus) or Image J.

2.4 RNA extraction

Plant tissue was collected into sterile 2mL tubes containing a 5mm steel ball bearing. Seedlings with metal balls were snap-frozen in liquid nitrogen and then placed in a TissueLyser (Qiagen, 85300) to grind seedlings into a fine powder. Total nucleic acid isolation was then carried out by phenol-chloroform extraction using a protocol adapted from White and Kaper (1989). Approximately 500mg frozen ground seedlings were transferred to ice, 600µL of extraction buffer was added (100mM Glycine, 10mM EDTA, 100mM NaCl, 2% SDS, pH9.5) and vortexed. 600µL of Tris-buffered phenol was then added to the samples, vortexed and then centrifuged 10 minutes at 16,000 × g at 4°C. The upper phase was transferred to a sterile tube containing 600µL of 25:24:1 phenol:chloroform:isoamyl alcohol, vortexed and then centrifuged for 10 minutes at 16,000×g at 4°C. The upper phase was transferred to a sterile tube containing 600µL of 24:1 chloroform:isoamyl, vortexed and centrifuged for 5 min 16,000 × g at 4°C. The upper phase containing the TNA was then transferred to a sterile tube and then precipitated by adding 40µL of 4M sodium acetate pH5.2 and 800µL absolute ethanol,

incubating for 15 mins on ice and centrifuging for 15 minutes at 16,000 × g at 4°C. The ethanol was removed, and the pellet then was washed with 70% ethanol and centrifuged again for 5mins at 16,000 × g at 4°C. The wash ethanol was removed, and the pellet was then allowed to air dry. The resulting pellets were resuspended in sterile water. The RNA concentration and quality were checked using the NanoDrop 1000 Spectrophotometer (ThermoFisher). DNA contamination was then removed using the DNase I (Sigma, AMPD1) by adding 1 unit DNase 1 and DNase buffer to 2µg of TNA, incubating for 15mins at room temperature, DNase inactivation agent was then added and incubated or 10mins at 70°C. Purified RNA was stored at -80°C.

2.5 cDNA synthesis

Table 2.1 Reverse Transcription Reaction Mix

Component	Volume
10x RT Buffer	2µL
25x dNTP Mix (100mM)	0.8µL
10x RT Random Primers	2.0µL
Multiscribe Reverse Transcriptase	1.0µL
Nuclease-Free H2O	4.2µL
RNA	10µL

Approximately 2µg of DNase treated RNA was reverse transcribed to produce cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied biosystems, 4368814). Reactions were set up as shown in table 2.1 and placed in a thermal cycler to be incubated at 25°C for 10mins, 37°C for 2hours and 80°C for 5mins. For all

experiments, a control sample with no reverse transcriptase added was used to check for contaminating DNA.

2.6 Semi-quantitative real-time polymerase chain reaction (qRT-PCR)

The qPCR protocol is adapted from the SYBR® Green JumpStart™ Taq ReadyMix™ (Sigma-Aldrich Cat. No. S4438) using the reaction mix described in Table 2.2. The reactions were placed in an optical 96 well plate (Star Labs) which was covered with an optical adhesive film (Bio-Rad, #MSB-1001) and centrifuged to remove air bubbles. The reactions were run using the Mx3005p qPCR System (Agilent Technologies) with MxPro software (Agilent Technologies). Thermal cycling conditions consisted of a denaturation step at 94°C for 2 minutes, 45 cycles of 15-second denaturation at 94°C and 1-minute extension at 60°C, then a final step to obtain a dissociation curve step of 2 minutes at 94°C, 1 minute at 60 °C and then a gradual increase in temperature for 5mins.

Table 2.2. qPCR Reaction Mix

Component	Volume
2X SYBR® GREEN JUMPSTART™ TAQ READYMIX™	7.5µL
Forward Primer (10mM)	1.0µL
Reverse Primer (10mM)	1.0µL
Nuclease-Free H2O	4.5µL
10x Diluted cDNA	2µL

The data was analysed using the MxPro software which automatically determines the cycle threshold, and then the ΔCT method was used to compare expression using the following equations:

ΔCT = (CT of the normalisation control - CT of the gene of interest)

ΔΔCT = - [(ΔCT) sample 1 - (ΔCT) sample 2)]

UBC10 is constitutively expressed and was used as a normalisation control (Xia et al., 2014). Three biological and three technical repeats were performed for all the experiments

2.7 Western blot

Seedlings were placed in a 2ml tube with a 5mm ball bearing, frozen in liquid nitrogen and ground using the Tissuelyser. Cells were lysed by adding 6x laemmli buffer (1 µL/µg of ground seedling) to the ground tissue which consists of 375 mM Tris-HCl, 9% SDS, 50% glycerol, 9% beta-mercatoethanol and 0.03% bromophenol blue. The tissue was then mixed, boiled at 95°C for 5mins and then centrifuged at 14000rpm for 5mins. The supernatant was then run on a 10% SDS-PAGE gel (Table 2.3 and 2.4) alongside a colour prestained protein standard ladder (NEB, P7712S) in running buffer (25 mM Tris, 192 mM Glycine, 0.1% SDS, pH 8.3) at 150V until the bromophenol blue front had just passed out of the gel. The samples were then transferred from the gel to a nitrocellulouse membrane (ThermoFisher, 88018) via the wet transfer method in transfer buffer (48 mM Tris, 39 mM glycine (Fisher), 10% w/v SDS, 10% (v/v) methanol) at 300mA for 1 hour. After transfer, the membrane was stained with Ponceau S stain solution (Sigma, P7170) to use as a loading control and washed in 1X Tris-Buffered Saline (TBS), which consists of 50 mM Tris, ph7.5 and 150mM NaCl in H2O. The membrane was

then blocked in 5% milk (w/v) in TBS overnight, incubated in anti-myc antibody (Cell Signalling, DB4C12) at 1:500 v/v in 5% milk for two hours, rinsed in TBS, incubated in anti-rabbit antibody (abcam, ab6721) at 1:1000 v/v in 5% milk for 1 hour. After rinsing in TBS, the membrane was imaged using Peirce ECL western blotting substrate (ThermoFisher, 32106) and Chemidoc XRS+ imager (Biorad).

Table 2.3 Western Blot gel recipe

Separating Gel

COMPONENT	AMOUNT
30% Acrylamide	2.67mL
1.5m Tris pH8.8	2mL
10% SDS	80µL
10% APS	80µL
TEMED	8µL
H2O	3.2mL

Stacking Gel

COMPONENT	AMOUNT
30% Acrylamide	1mL
0.5m Tris pH6.8	1.25mL
10% SDS	50µL
10% APS	50µL
TEMED	5µL
H2O	2.6mL

2.8 Electrophoretic mobility shift assay

Nuclear Proteins were extracted from 2 grams of 14-day old seedlings. The seedlings were frozen in liquid nitrogen and ground in a pestle and mortar while keeping frozen until a fine pale green powder. The nuclear protein was then extracted using the crude extraction protocol from the CellLytic PN isolation/extraction kit (Sigma, CELLYTPN1) which consists of adding the ground seedling in cell lysis buffer to lyse the cells, then the nuclei are extracted through centrifugation in nuclei isolation buffer. Finally, nuclear proteins are extracted using extraction buffer, dithiothreitol and protease inhibitor cocktail. The protein concentration was quantified using Nanodrop. The Lightshift Chemiluminescent EMSA kit (ThermoFisher, 20148) was used for the EMSA and carried out as per manufacturer's instructions. 5µg of extracted protein was

then incubated with 1µM biotin labelled oligos (found in table 2.4),1X binding buffer (10 mM Tris 50mM KCL, 1mM DTT, pH7.5) and 1 µL of 1µg/µL Poly (dI:dC) at 4°C for 30mins. The reactions were then run on a 5% non-denaturing polyacrylamide gel, in 0.5x TBE (450mM Tris, 450mM boric acid, 10mM EDTA, pH 8.3) at 100V until the bromophenol front is around ¾ down the gel, transferred to a nylon membrane using the wet transfer method in 0.5X TBE at 100V for 30mins, UV crosslinked to the membrane and then imaged using Peirce ECL western blotting substrate and Chemidoc XRS+ imager.

2.9 Polymerase Chain Reaction

20 µL reactions were used containing 2µL of cDNA that has been diluted 10X, 0.5mM of forward and reverse primers and 1x REDTaq® ReadyMix™ PCR Reaction Mix (Sigma-Aldrich, P0982). Thermocycling conditions varied depending on product size and primer temperature but was usually an initial denaturation at 95^{O}C for 2mins, then 35 cycles of; 95^{O}C for 30s, 55 ^{O}C for 1min, 72^{O}C for 1min per Kb of product, and a final extension at 72^{O}C for 5mins.

2.10 Gel electrophoresis

As REDTaq® ReadyMix contains loading dye, PCR products were loaded directly into 1% or 2% agarose gels prepared using 1xTAE (40mM Tris/ 20mM Acetate/ 1mM EDTA) buffer containing ethidium bromide (3µL of 10mg/mL per 100mL of gel). 5µL of 100 bp or 1 kb DNA ladder (NEB) was loaded alongside samples. Gels were run in 1xTAE Buffer at 100V for the desired length of time and imaged using a gel imaging system (UVP gel doc).

2.11 Gel extraction

The DNA band of interest was excised from the agarose gel using a clean scalpel blade over a UVP gel doc UV-transilluminator. The DNA was then extracted from the agarose gel slice using the QIAquick Gel Extraction Kit (Qiagen, #28704).

2.12 Cloning and sequencing

Gel extracted DNA fragments were ligated into the CloneJet vector using the PCR CloneJET kit (ThermoFisher, K1232) according to the manufacturer's instructions and then 5µL of the ligation reaction was added to 50µL of DH5α chemically competent *E. coli* cells for transformation. The transformation was carried out by heat shocking in a 42°C water bath for 20secs. The *E.coli* were then added to 1mL of liquid LB (ThermoFisher, 12780029) for 1hour before spreading on a LB agar (ThermoFisher, 22700041) plate containing 100µg/mL ampicillin and allowing colonies to grow overnight. Colonies were then selected by colony PCR using the primers used in the PCR of the gene fragment (Table 2.3), the colonies containing the desired products were then bulked up overnight in LB broth. The plasmids from the cultures were then extracted using a Plasmid Mini Prep kit (Qiagen, 27104). The extracted plasmid containing the EXT transcripts were then sent to the Core Genomic Facility (The University of Sheffield) for Sanger sequencing using both the forward and reverse CloneJet sequencing primers (Table 2.4).

2.13 Primers

Genomic DNA or mRNA sequences were obtained from TAIR (https://www.arabidopsis.org/). Primers were designed using primer-BLAST from The National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers and oligos used can be seen in Table 2.4.

Table 2.4 Primers and Oligos

PRIMER	SEQUENCE
CLONEJET F	CGACTCACTATAGGGAGAGCGGC
CLONEJET R	AAGAACATCGATTTTCCATGGCAG
LRX1 5' F qPCR	CCTCTTGCTCTCCTCCGTTT
LRX1 5' R qPCR	GAGCAAACATCTGAACCATTCCA
LRX1 3' F qPCR	GGCCATCCACCACAAGCAACA
LRX1 3' R qPCR	CGCTTGGCATCGGTGGAAAGT
EXT13 5' F qPCR	GGAGCGAGTGCATTGCATGGT
EXT13 5' R qPCR	TGGGATGAGTAAGGGTATGCTGTCA
EXT13 3' F qPCR	ATGTCTACAGCTCCCCACCACCTC
EXT13 3' R qPCR	GGGTGTCTTTGAGACATAAGGTGCT
GUS F qPCR	GGCATCCGGTCAGTGGCAGT
GUS R qPCR	CCTTTGCCACGCAAGTCCGC
RSL4 F qPCR	AGGCAAAACTAGAGCCACCA
RSL4 R qPCR	ATCGACTTTTGTCCCGTTTG
GTL1 F qPCR	ATGGAATTGTTTGAAGGTTTGG
GTL1 R qPCR	GACATGACCTCGTGTTCTCG
RHD6 F qPCR	TCACGAGAGCTTTCCTCCTC
RHD6 R qPCR	TGAAGCCGTAGCTCATGTTG
LRX1 EMSA Oligo	AGAGGTGGTGGTGGTATG

mtLRX1 EMSA Oligo	AGAGGTGATGGTGGTATG
LRX1 5' F RT-PCR 1	CCTCTTGCTCTCCTCCGTTT
LRX1 5' R RT-PCR 1	CGCATATAGGAGGACGACCG
LRX1 3' F RT-PCR 1	AAGAACATCGATTTTCCATGGCAG
LRX1 3' R RT-PCR 1	TGCTAATGAGCAGCTCTGTGTTA
LRX1 3' F RT-PCR 2	CTTGAAGTCCGGACACTTCC
LRX1 3' R RT-PCR 2	CGCTTGGCATCGGTGGAAAGT
EXT9 RT-PCR F	TGCCCATCTCGTCTACGCTA
EXT9 RT-PCR R	CGAAGGTGAATGTGAAGCTTAGT

2.14 Data Analysis

All graphs were produced in Graphpad Prism7 and statistical test were also carried out on Graphpad Prism7 using either a t-test or ANOVA with a post analysis TUKEY tests to compare individual means. Variance was considered statistically significant when p = ≤0.05, asterisks indicate significance; * = p ≤0.05, ** = p ≤0.005, *** = p ≤0.0005), **** = p ≤0.0001.

3. The role of G4s on plant growth and development

3.1 Abstract

G4s have a potential role in plant development as a study found that Arabidopsis grown on media containing G4 stabilising ligands developed shoot defects (Nakagawa et al., 2012). Here, we have further investigated the effect of G4 ligands on Arabidopsis development and also taken a transcriptomic approach to investigate the impact of these G4 ligands on global gene expression. While we did not identify shoot phenotypes in plants grown on the G4 ligands; NMM and Berberine, we did see a significant decrease in the overall growth of the plant and a bulbous root hair phenotype. We identify that plants treated with G4 ligands have major changes in the expression of members of the EXTENSIN (EXT) gene family, which are structural components of the cell wall (Johnson et al., 2017; Ringli, 2010; Showalter et al., 2016; Xu et al., 2014). We propose that the hydroxyproline-rich EXTs are enriched for PG4s on the template strand, which may explain the expression loss caused by G4 ligands. This loss of EXT gene expression may lead to a defective cell wall and cause growth defects and bulbous root hairs.

3.2 Introduction

While several biological functions of G4s have been suggested in mammals and yeast, very little is known of their functional significance in plants. In a screen of natural plant products that impact shoot development, it was found that Berberine treatment resulted in plants with pin-shaped leaves and in more severe cases, an absence of true

leaves (Nakagawa et al. 2012). As Berberine affected shoot morphology, the expression of key shoot development genes was investigated. It was determined that KANADIA 4 (KAN4) and BREVIPEDICELLUS (BP), which have roles in leaf polarity, were significantly affected. However, Berberine did not affect ASYMMETRIC LEAVES1 (AS1) and ASYMMETRIC LEAVES 2 (AS2) which are downstream of KAN4 and BP in the known polarity pathway. This finding led to the hypothesis that G4s may have a role in a novel shoot development pathway.

Berberine is an isoquinoline alkaloid that has been used in Chinese medicine for centuries as an antimicrobial and is sold as a homoeopathic medicine for diabetes, cancer, high blood pressure and high cholesterol (for which there is increasing understanding of the mechanisms of action) (Affuso et al., 2010; Caliceti et al., 2016; He et al., 2016; Kim et al., 2009; Liu et al., 2013; Peng et al., 2015; Tian et al., 2018; Wei et al., 2016; Wu et al., 1999; Yin et al., 2008). However, Berberine is known to bind DNA secondary structures including G4s (Bazzicalupi et al., 2013; Franceschin et al., 2006; Li et al., 2017; Ren and Chaires, 1999). Therefore, to determine whether the phenotypic effects were due to its G4 binding function or an alternative mechanism, a second G4 ligand NMM was also tested. The phenotypes seen in NMM treated plants were similar to that seen with Berberine treatment, and it was therefore proposed that G4s may regulate shoot development through a novel mechanism (Nakagawa et al., 2012).

These limited studies implicate G4s as having a role in shoot development. To investigate this further, this chapter explores the developmental effect of G4 binding ligands on the model plant species Arabidopsis thaliana and investigates whether changes in gene expression cause the observed phenotypes.

3.2.1 Aim

In this chapter, the aim is to analyse the role, if any, of G4 on plant development

3.2.2 Objectives

1) To perform phenotypic analysis of plants grown on G4 stabilising drugs.

2) To analyse the impact of these G4 stabilising drugs on gene expression and correlate this with the presence of G4s in target genes.

3) To perform a phenotypic analysis of of plants with mutations in G4-rich genes.

3.3 Results

3.3.1 Treatment with G4 stabilising drugs causes reduced plant growth

To begin to understand the effect of G4 stabilisation on plant growth and development we analysed how G4 ligands affected the overall growth of Arabidopsis. To investigate this, plants were germinated on MS agar plates containing increasing concentrations of the G4 ligands, NMM and Berberine. NMM is a porphyrin which have implications in effecting iron metabolism and photosynthesis (Lecerof et al., 2000; Mauzerall, 1977). However, NMM and Berberine are structurally different and have alternative mechanisms of G4 binding, therefore, both ligands were used to give some certainty that the effects seen on the plants were due to G4 stabilisation and not off-target effects of the ligands. As Berberine was diluted in DMSO, for this experiment NMM was also diluted in DMSO and the equivalent volume of DMSO (0.1% v/v) was added to the control plates. It was determined that plants grown on NMM and Berberine show a visible decrease in overall growth (Fig. 3.1, A and B). To quantify the impact on seedling growth, measurements of the rosette area were taken and concentrations as low as

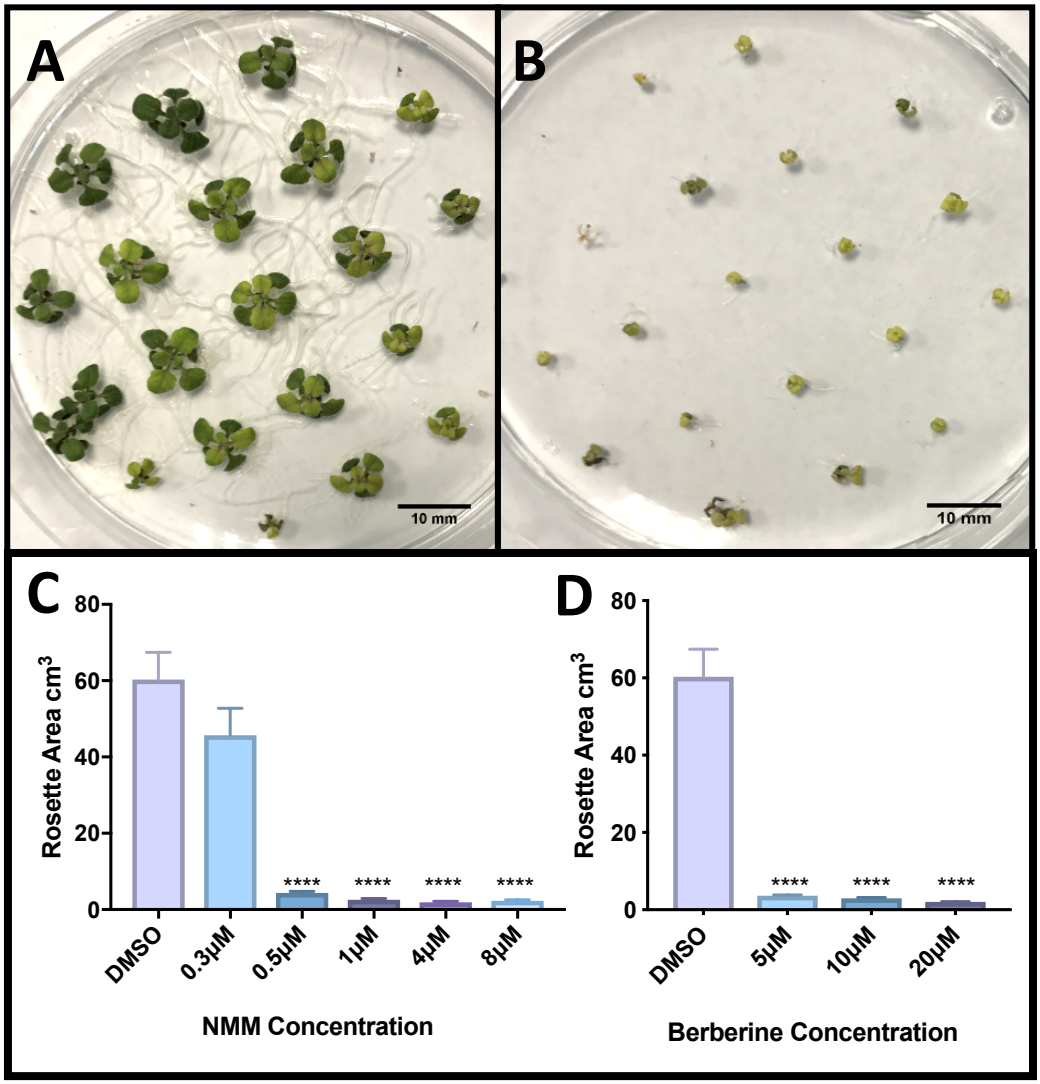

Figure 3.1. G4 NMM and Berberine affect overall plant growth. Arabidopsis seedlings were germinated on MS media containing **(A)** 0.1% v/v DMSO and **(B)** 0.5µM NMM and grown for 14 days post stratification. The graphs show rosette area measurements of 14 day old seedlings grown on varying concentrations of **(C)** NMM and **(D)** Berberine compared to the control containing 0.1% v/v DMS0. **** indicates a p-value of > 0.0001 compared to DMSO control, from analysis by one-way ANOVA with multiple comparisons, n > 30.

0.5µM NMM caused significant decreases in the rosette area compared to the DMSO control (Fig.3.1, C). Similar results were also observed with Berberine (Fig. 3.1, D). During this analysis pin shaped leaves or loss of true leaves was not observed, however, the overall decrease in growth seen in plants treated with these G4 ligands may suggest a more global G4 dependant growth mechanism in plants than that suggested by Nakagawa et al. (2012) in shoot development.

3.3.2 G4 Stabilisation effects organ growth and development

We next investigated the role of G4s in shoot development given that a previous study suggested that growing plants on the G4 ligands, NMM and Berberine, can disrupt shoot development in Arabidopsis (Nakagawa et al. 2012). We began by repeating the previous experiments; growing seedlings on 2µM and 3µM of NMM (diluted in water) or untreated agar for 14 days. We found that the leaves were smaller in the NMM treatment, which agrees with the previous rosette measurements. However, after analysing 100 seedlings grown on each of the conditions, we did not see the leaf developmental defects that had previously been reported (Nakagawa et al. 2012). We identified one seedling with a leaf patterning phenotype (Fig 3.2, C) and one which had delayed growth of the true leaves at 2µM NMM, which is a much lower incidence of shoot phenotypes than reported, where 40% of seedling grown on 2µM NMM had no true leaves (Nakagawa et al. 2012).

While investigating the role of G4 stabilisation on shoot development it was observed that the ligands had a strong effect on the primary root of seedlings and was therefore further investigated. Seeds were germinated on vertical plates containing MS

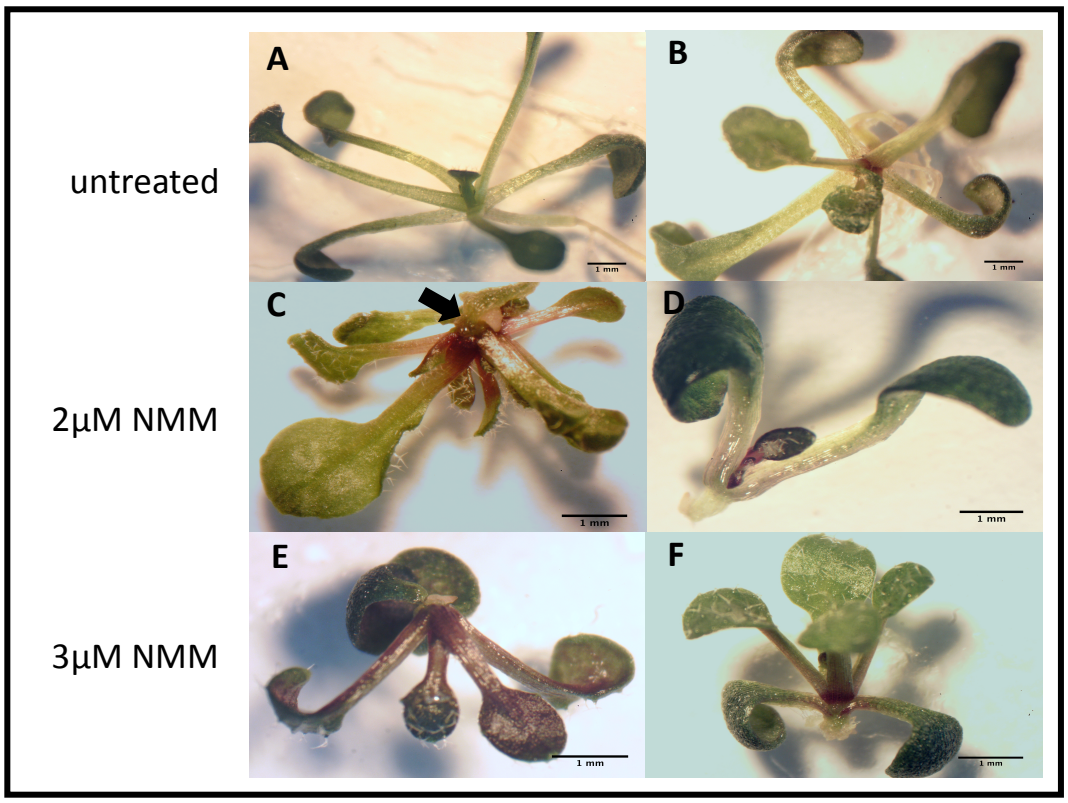

Figure 3.2. Effect of NMM on shoot development. Representative 14 day old Arabidopsis seedlings germinated on **(A,B)** untreated, **(C,D)** 2µM NMM and **(E,F)** 3µM NMM media and analysed for shoot phenotypes. Arrow indicates leaves developing on the underside if the seedling. n>100

agar with increasing concentrations of NMM. NMM and Berberine caused a severe and significant reduction in primary root growth (Fig 3.3, A,B) and also a bulbous root hair phenotype (Figure 3.3, C,D). However, the root hair phenotype was variable depending on whether the root hairs were directly in contact with the agar containing the drug. To investigate whether direct contact with the drug was required for it to affect root hair; seedlings were grown on softer agar (0.3% rather than 0.8%) and plates were angled at approximately 75°, which resulted in the roots growing into the media. Additionally, it was found that the root hairs of control plants at the root/hypocotyl junction frequently had unusual phenotypes (bulbous, branching or longer length) and therefore when the

primary root was drastically stunted by NMM it was plausible that the only root hairs seen were from this root junction region. Also, there was a possibility that the decrease in root growth that was previously measured may have been due to an effect on the germination of the plant rather than root growth mechanisms. To address these possibilities, the experimental approach was optimised to ensure that any observed phenotypes could be attributed to direct impacts on root growth and morphology. Seedlings were germinated and grown on untreated 0.8% agar for 4 days post stratification so that the primary root had initiated, began normal growth and normal root hairs were present. The seedlings were then transferred to agar containing the treatment.

The optimised assay showed that treatment with either NMM or Berberine resulted in significant reductions in primary root growth (Fig. 3.3, A and B). The effect of G4 ligands on root length was concentration dependent, though concentrations of 1µM, 2µM and 4µM of NMM were also tested but displayed almost complete primary growth inhibition. In addition, NMM treatment resulted in a striking bulbous root hair phenotype at 0.3µM NMM, while concentrations higher than 0.5µM commonly resulted in no root hairs being present (Fig. 3.3, C and D). Similarly, the cells at the primary root tip also appeared to have distorted bulbous cells following NMM treatment (Fig. 3.3, E and F), although this was not seen with Berberine. This suggests that G4s may play an important role in both primary root and root hair development.

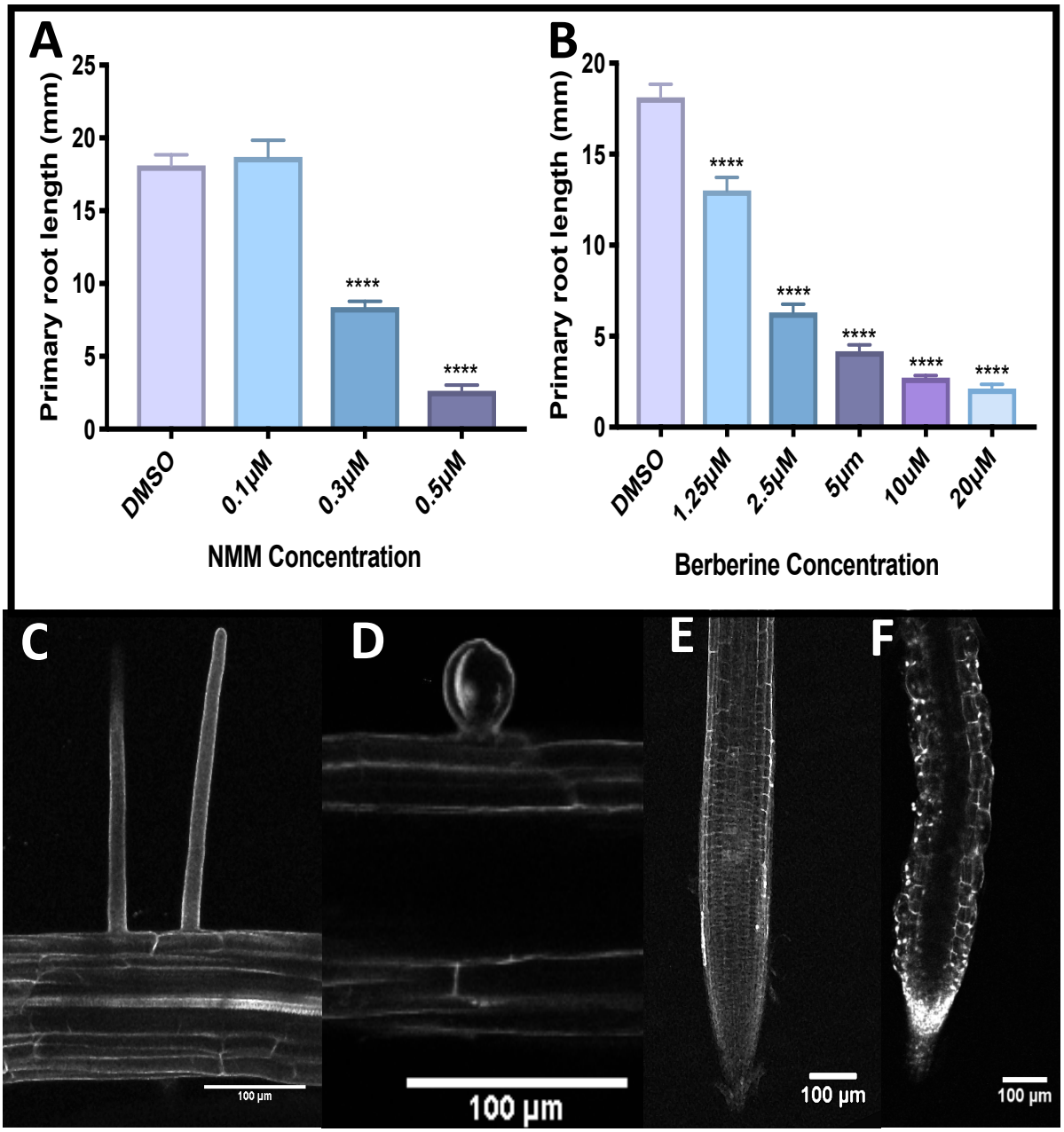

Figure 3.3. NMM treatment decrease primary root length and root hair morphology. Primary root length of 7 day old Arabidopsis seedlings with increasing concentration of **(A)** NMM and **(B)** Berberine. Seedlings were transferred to treatment plates at 4 days post stratification. Confocal images show root hairs of 7 day old seedlings grown on **(C)** untreated and **(D)** 0.3µM NMM media, n>30. Also, the primary root tip of 7 day old seedlings grown on **(E)** untreated and **(F)** NMM treated seedlings, n>30. Statistics done using multiple comparison ANOVA compared to DMSO treated, **** indicates a p-value <0.0001.

Treatment with G4 ligands causes growth and developmental phenotypes in Arabidopsis. Therefore, it was hypothesised that these phenotypes may be due to treatment with G4 ligands causing differential gene expression in genes required for correct growth and development. To investigate this hypothesis Dr Manoj Valluru and Giulia Arsuffi from the lab performed a whole transcriptome microarray analysis of seedlings treated with NMM for 6 hours, and Mathew Parker carried out an initial analysis of the microarray data. Their analysis found 858 and 1098 genes were differentially upregulated and downregulated respectively using a 2-fold change threshold (Parker, 2018). Additionally, genes differentially expressed in the NMM microarray correlated with those found to be differentially expressed in response to Berberine treatment in a previous study (Pearsons R: 0.43, Spearman's: 0.44) (Nakagawa et al., 2012). Considering the different conditions in which these experiments were performed and the use of a different G4 ligand, this strongly suggests that these gene expression changes are a result of G4 binding. This supports the hypothesis that stabilisation of G4s using ligands can significantly alter gene expression within Arabidopsis.

A major question arising from these analyses was whether any of these differentially regulated genes could explain the growth and developmental changes we see in NMM treated plants. The Berberine study found more than a 2-fold increase in expression of the class 1 KNOX genes, *BP* and *KAN4*, and hence proposed that the shoot developmental defects that they see may be caused by the changes in expression of these genes (Nakagawa et al., 2012). However, in the NMM microarray KAN and BP1 expression was not significantly changed. Further analysis of the genes differential

expressed in the NMM microarray with a enrichment PG4 enrichment identified a strong correlation with genes associated with the plasma membrane and cell wall. One gene ontology group of notable interest was GO:0005199, which contained 992 PG4s in only 32 genes and 50% of the expressed genes in GO:0005199 were downregulated by NMM. The gene ontology GO:0005199 consist primarily of genes from the *EXT* family (29 out of the 32 genes within the family) (Parker, 2018).

EXTs are thought to have a role in the structural integrity of the cell wall and have highly repetitive HYP regions. As the codon for proline is CCN, this gives the sequence GGN on the template strand of the gene, causing a highly G-rich sequence that could potentially form a G4 (Fig. 3.4 A). Therefore, these repetitive HYPs within *EXTs* gene bodies result in many PG4 in the template strand (Fig. 3.4, B).

It was also identified (work by Mathew Parker) that the *EXTs* differentially regulated by NMM were mainly of the SP4 or SP5 class of EXTENSIN, meaning they were highly enriched for motifs consisting of Serine followed by 4 or 5 HYP, rather than the SP3 class (containing three HYP) (Parker, 2018). This may be due to the 4 or 5 HYP of the SP4/SP5 motif providing at least 4 consecutive runs of guanines in the template strand which would be predicted to form a G4, whereas the SP3 EXTs only have 3 consecutive runs of guanines and so would require further guanines within the gene to form a G4.

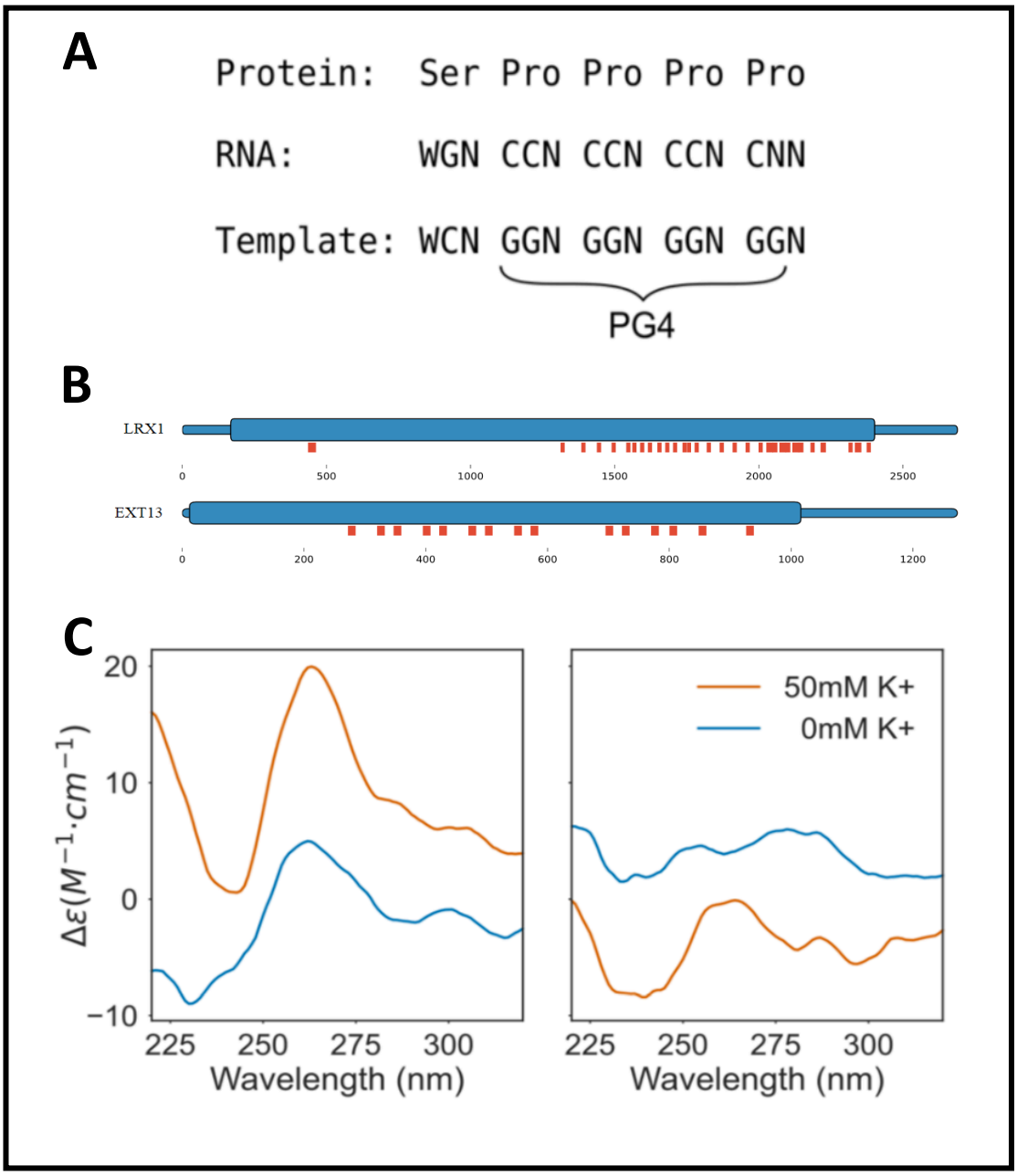

Figure 3.4. The EXT SP4 motif forms a G-Quadruplex *in vitro* (A) Schematic of how the EXT SP4 protein motif results in a two tetrad PG4 into the template strand of the coding sequence. **(B)** annotation of the extensin genes *LRX1* and *EXT13* in blue with the PG4 indicted below in red. **(C)** CD spectroscopy of an EXT SP4 sequence (left) and a mutated sequence not predicted to form a G4 (right). Figure adapted from (Parker, 2018)

To identify whether the EXT SP4 repeat can form a G4, Dr Nicholas Zoulias performed CD spectroscopy using an oligo resembling the EXT SP4 motif (AGAGGTGGTGGTGGTATG) sequence and a mutated sequence where one of the guanines is changed to an adenine which is predicted to remove the formation of the G4 (mutated sequence: AGAGGTG**A**TGGTGGTATG). The CD spectroscopy of the SP4 oligo demonstrated an absorption spectrum with a peak at 260nm and trough at 240nm (Fig. 3.4, C) which is typical of a parallel G4 structure (Chen, 1992; del Villar-Guerra et al., 2018). Furthermore, when the absorbance was measured in the absence of K+ (which should cause the G4s to no longer be stabilised) the G4 spectrum was reduced (Fig. 3.4, C) and in the mutated sequence the spectrum is absent (Fig 3.4, C). This is consistant with the EXT SP4 motif being capable of forming a G4 *in vitro*. Hence the PG4s formed by the SP4 motif may be the target of the G4 ligands and the cause of the strong downregulation in gene expression caused by G4 stabilisation with these ligands. qRT-PCR was used to confirm that several members of the *EXT* family are downregulated by NMM treatment (Fig, 3.5, C). *EXTs* are more severely downregulated with increasing NMM concentration and increasing time in the presence of the NMM as can be seen in the expression of *EXT13* in Fig. 3.5, A and B. Additionally, downregulation of *EXT* gene expression in roots appears to be stronger or quicker compared to shoot tissue (Fig. 3.5 A), which may explain the more severe root phenotypes observed.

Next, it was investigated whether the decrease in *EXT* expression caused a decrease in EXT protein levels. Western blots were performed using anti-EXT antibodies (PlantProbes), however, perhaps due to the post translation modifications of the EXTs and the selectivity of the antibody to all EXTs, we struggled to obtain an interpretable

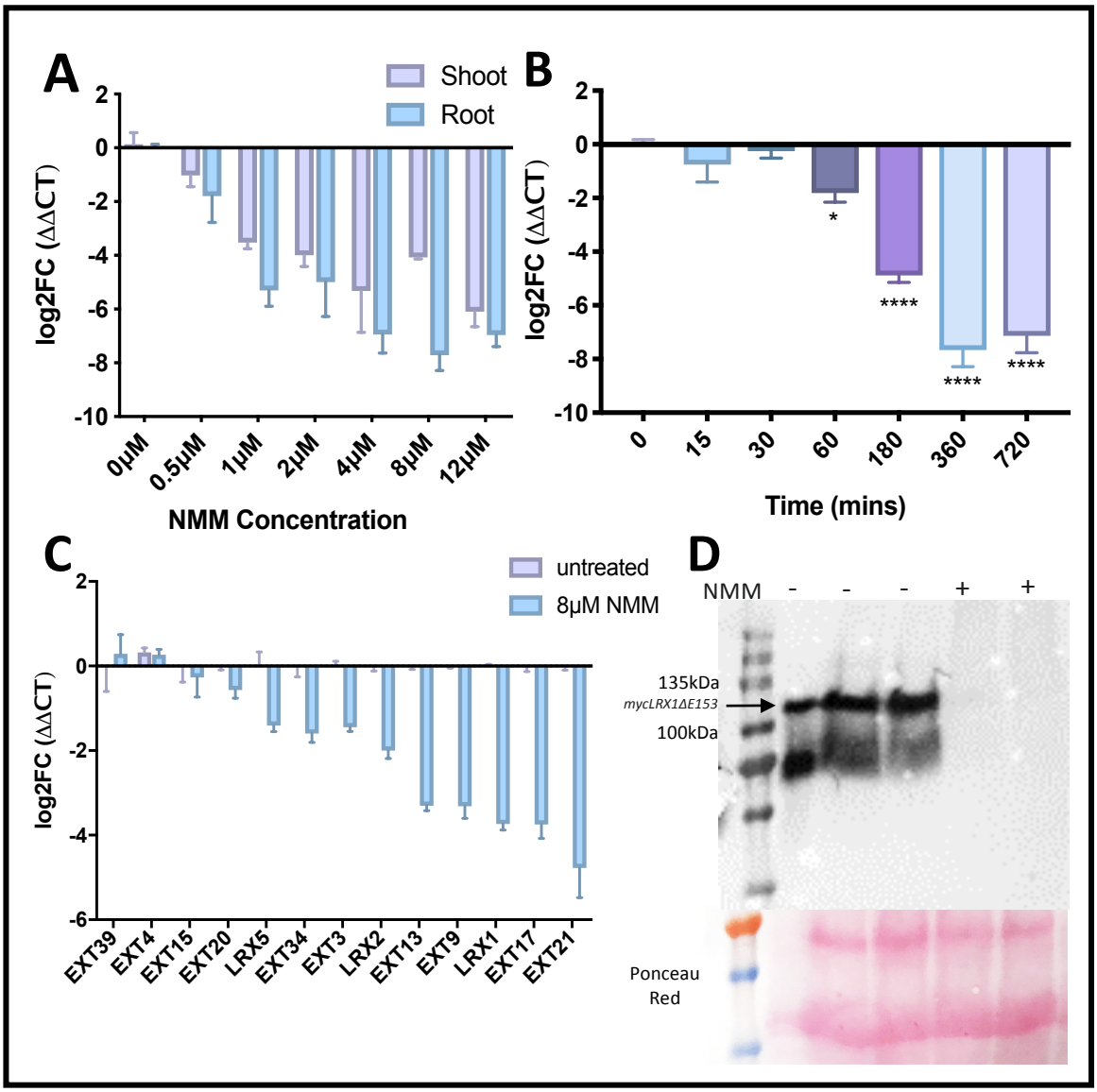

Figure 3.5. G4 stabilisation causes decreased expression of EXTs. (A) Change in expression of the *EXT13* gene when treated with different concentration of NMM for 6 hours relative to untreated, in root and shoot tissue**. (B)** Change in *EXT13* expression when treated with 8μM NMM for different periods of time relative to untreated, in root tissue. **(C)** The change in expression of several members of the EXT gene family upon treatment with 8μM NMM for 6 hours, relative to untreated. **(D)** Western blot showing myc-tagged LRX1 protein expression with and without NMM and ponceau red staining to show equal protein loading. Statistics done using students T-test, **** indicates a p-value <0.0001, ***<0.0005, **<0.005, *<0.05. n=3

blot. Given the *LRX1* expression is significantly downregulated, the protein levels of LRX1 was investigated using a plant line, kindly donated to us by Christoph Ringli, *mycLRX1ΔE153* (Ringli, 2010), which contains a MYC tagged LRX1 construct with a small truncation. It was therefore possible to use a myc antibody to selectively detect this protein. We found that NMM caused a decrease in the protein levels of this transgene (Fig. 3.5 D), which suggests that the down regulation of EXT expression caused by NMM is also likely to effect protein levels.

3.3.4 Downregulation of EXTs causes root hair phenotypes

The *EXT* family are a group of structural cell wall genes, rich in hydroxyproline residues, that are thought to be involved in controlling the rigidity of the cell wall. The identification of the *EXT* family being both enriched for PG4s and strongly down-regulated by NMM was of particular interest as mutations of genes from this family have previously been seen to have root hair phenotypes similar to those observed when we treat with NMM (Baumberger et al., 2001, 2003, Velasquez et al., 2011a, 2011b).

To investigate this further, a phenotypic analysis was performed of root hairs of the mutant *lrx1*, the double mutant *lrx1lrx2* and plants treated with DHP, which inhibits the synthesis of hydroxyproline and therefore the synthesis of EXT proteins (Xu et al, 2011). Root hairs within the *lrx1* mutant appeared to have some bulging morphology, primarily at the base where the root hairs meet the root (Fig. 3.6, D). The *lrx1lrx2* mutant demonstrated a more severe root hair phenotype with some root hairs appearing globular in shape but some just displaying a broad base (Fig. 3.6, E). However, DHP treated plants had a phenotype most similar to that of NMM with the root hair

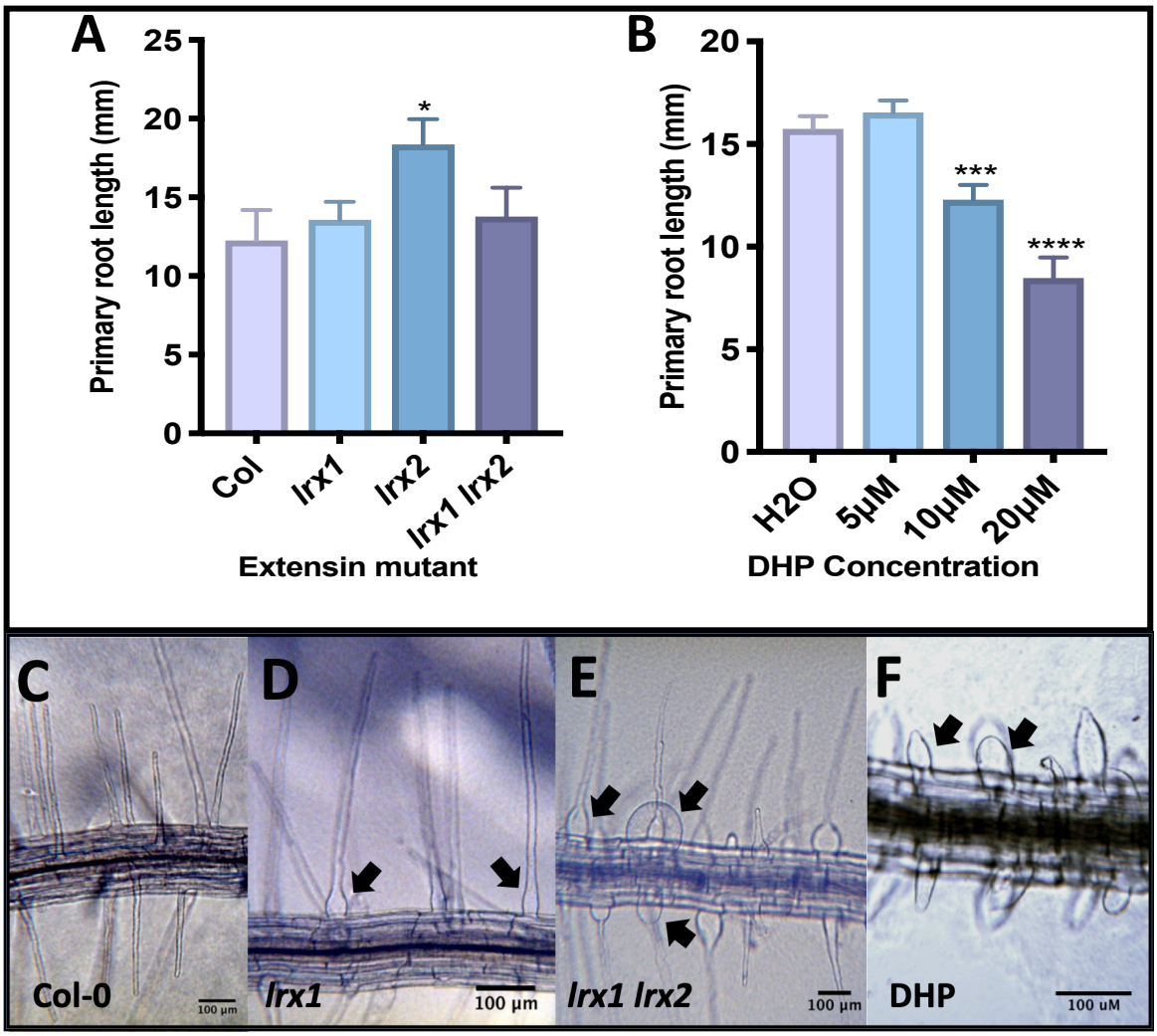

Figure 3.6. Loss of EXTs causes root hair morphology. (A) Primary root length of 7 day old EXT mutant lines and **(B**) Col-0 seedlings treated with the EXT inhibitor DHP. Microscopy images show root hairs of **(C)** Col-0 seedlings, **(D)** *lrx1*, **(E)** *lrx1 lrx2* and **(F)** 10uM DHP treated seedlings. Statistics done using students T-test, **** indicates a p-value <0.0001, ***<0.0005, **<0.005, *<0.05. Arrows indicate bulbous morphology.

appearing to be unable to extend and having a short bulbous phenotype (Fig. 3.6, F). This supports the hypothesis that a decrease in EXT expression, through mutation or DHP treatment, causes changes in root hair morphology and hence the downregulation of EXTs caused by NMM is also likely to be the cause of the root hair phenotype we see.

It is also possible that the EXTs may play a role beyond that of root hair development and changes in *EXT* gene expression may also be the cause of the decrease in primary root growth that was observed upon treatment with NMM. Mutation in *lrx1* or *lrx2* did not appear to stunt primary root growth and in fact, *lrx2* mutants displayed an increase in primary root length (Fig. 3.6 A), though the lack of correlation with NMM inhibition of root growth may be due to redundancy of some *EXT* genes (Baumberger et al., 2003; Borassi et al., 2015). However, a significant decrease was observed in primary root length when plants were treated with DHP, which would be expected to have a global effect on all EXT synthesis (Fig. 3.6 B). Although, the effect of DHP on primary root length was not as strong as that seen with NMM suggesting there may be other factors involved; nonetheless, this suggests EXTs or related hydroxyproline-rich proteins may play a role in primary root growth and G4 stabilisation could therefore regulate root growth via these genes.

3.4 Discussion

Here, it has been determined that treating plants with G4 stabilising ligands causes a decrease in overall plant growth (Fig.3.1) suggesting a major role of G4s in regulating plant growth and development beyond the shoot development phenotypes previously described (Nakagawa et al. 2012).

In the previous study, it was found that 40% and 60% of seedlings treated 2µM and 3µM NMM respectively had no true leaves (Nakagawa et al. 2012). However, in this study, these phenotypes were not replicated (Fig 3.2), and there was no significant change in *BP* or *KAN4* expression. It is possible that photoperiod differences in the growth conditions could account for the differences between the two studies with this

study growing plants in a 24h photoperiod as opposed to the 16h photoperiod used previously (Nakagawa et al. 2012). As light plays an essential role in shoot development and the growth rate of plants, it is possible that the differences in photoperiod or impacts on the circadian clock could explain the lack of shoot defects in this study. Therefore, further studies in to this may reveal another role of G4s in light signalling or polarity. On the other hand, despite not seeing the phenotypes observed in the previous study, there was a strong correlation in gene expression changes between our NMM microarray and the Berberine microarray from the shoot development study (Pearsons R: 0.43, Spearman ’s: 0.44), suggesting that the gene expression changes observed are G4 dependent and not off target effects (Nakagawa et al., 2012).

Additionally, this study suggests that G4 stabilisation using ligands resulted in changes in root growth and development. NMM and Berberine treated plants had significantly stunted primary root growth, bulbous root hairs and bulbous root tip cells (Fig 3.3). Additionally, we identified an increase in lateral and adventitious roots which could be further studied to understand whether these phenotypes were a result of the plant trying to increase its root network as the primary root was being stunted by the G4 ligands, or alternatively whether G4s also play a role in lateral and adventitious root initiation and development. As previously mentioned, NMM is a porphyrin, which has been known to effect photosynthesis and iron metabolism, however neither of these pathways has previously been linked to the phenotypes observed. Additionally, Berberine is not a porphyrin and hence we would not predict it to effect photosynthesis and iron metabolism, suggesting the phenotypes and gene expression changes observed are likely to be due to a G4 effect.

We propose that the bulbous root hair phenotype is caused by the G4 ligands downregulating the expression of EXT family genes. EXTs are significantly downregulated by NMM in the microarray and this has been confirmed by qRT-PCR (Fig 3.5, A-C). In addition, the down-regulation in gene expression is also translated into a downregulation in protein levels of EXT as shown by western blot (Fig 3.5, D). However this western blot was carried out using the myc-tagged EXT, *mycLRX1ΔE153,* which is truncated to consist of only half of the EXT domain of LRX1 (Ringli, 2010). While utilising the *mycLRX1ΔE153* gives a good indication that EXT protein levels are reduced by NMM, more evidence would be required for certainty. For example, repeating the experiment with a full length myc-tagged EXT (preferably several genes including LRX's and classical EXTS) or using an antibody that targets the EXT domain, such as Jim20 (Xu et al., 2011; Zhang et al., 2014), however this antibody can be difficult to use and interpret due to the glycosylation of the EXT causing them to be very large.

The EXT gene family is strongly enriched for PG4s, which are hardcoded in EXT genes due to HYP enrichment, as you cannot remove the G4 coding sequence without changing the proline codon. The SP4 motif is capable of forming a G4 *in vitro* and therefore this may explain why the EXTs are so strongly downregulated by the G4 ligands (Fig 3.4). Furthermore, the SP4 motif was identified to form a parallel G4 *in vitro,* which is NMMs preferred substrate and may explain why NMM treatment results in a stronger downregulation of EXT expression and more severe phenotypes than Berberine. Therefore, it would be interesting to investigate the effect of other G4 ligands, such as CuII-4'-(p-tolyl)-2,2':6',2'' (Marchand et al., 2016), which preferentially binds anti-parallel G4s to see how this effects the phenotypes observed and if there is a

different gene expression profile. Additionally, to prove the expression changes caused by G4 ligands are G4 dependent, most G4 studies mutate one or more of the guanines within the G4-forming sequence to prevent formation. However, in the EXTs if we mutate the guanines then it will change the codon and therefore the amino acid sequence of the protein, which could cause other effects within the plant. However, it has been shown that the bases in the loop sequences can affect the stability of the G4s (Guédin et al., 2008), therefore we could potentially mutate the wobble base of the codon, CCN, within the EXTs to make it unfavourable to form G4s which would help to prove that the effects seen when using G4 ligands is G4 dependent.

Finally, we confirm that EXT downregulation can result in the phenotypes that we see in NMM and Berberine treated plants by analysing plants with mutations in EXT genes or by treatment with the hydroxyproline inhibitor DHP. As previously shown, mutations in a single EXT gene, *lrx1,* causes some bulging root hair morphology (Baumberger et al., 2001, 2003). The loss of a single *EXT* is unlikely to cause a severe phenotype as it has been previously shown that there is some redundancy within the *EXT* family (Baumberger et al., 2003; Borassi et al., 2015). Consequently, when several EXTs are targeted by treatment with DHP, severe bulbous phenotypes and a reduction in primary root length is observed (Fig 3.6). This supports our hypothesis that downregulation of EXTs can cause similar root phenotypes to those we observe in plants treated with G4 ligands.

We propose that the G4 ligands, NMM and Berberine bind the PG4s formed within the SP4/SP5 motifs of EXTs and through a yet to be determined mechanism, cause a down regulation in expression of *EXTs*. As EXTs are suggested to play a role in

cell wall structure, downregulation may cause the cell wall of root hairs to be unable to withstand the turgor pressure of the cell to elongate and hence the cell collapses to form the characteristic bulbous shape. Moreover, as the structural integrity of the cell wall is important for cell elongation and division, the two main mechanisms of plant growth, hence the loss of EXTs expression may also be responsible for the overall decrease in growth seen by G4 stabilising drugs. In conclusion, these results suggest a novel role of G4s in plant growth mechanisms, potentially through the modulation of *EXT* expression.

4. How G-Quadruplexes effect the transcription of EXTENSINs

4.1 Abstract

G4s on the template strand of the coding sequence have been shown to inhibit transcription by stalling RNA Polymerase II (PolII). Therefore, we propose that EXTENSIN (EXT) G4s may block PolII during transcription. Meta-analysis of PolII occupancy within EXTs demonstrates decreased PolII occupancy over the pG4 dense region of EXTs and EXTs appear to have aberrant transcripts indicative of PolII stalling. There may be mechanisms to regulate G4 formation or allow EXT expression. We find that an oligo, with the sequence based on the most frequent PG4 in the EXTs, may be capable of binding nuclear proteins and that the helicase RecQsim may be a potential candidate for regulating G4 formation.

4.2 Introduction

G4s are predicted to form in *EXTs* within the gene body, on the template strand. A proposed consequence of G4s on the template strand of the coding sequence is the inhibition of transcription by the stalling of RNA polymerase II (PolII). Several studies of G4s in target genes show stalling of both RNA and DNA polymerases using *in vitro* polymerase stop assays in which ligand and cation-dependent G4 formation inhibit *in vitro* transcription or single cycle PCR (Caliceti et al., 2016; Chambers et al., 2015; Cogoi and Xodo, 2006; Dexheimer et al., 2006; Han et al., 1999b; Kwok et al., 2016; Siddiqui-Jain et al., 2002). It has been demonstrated that there are several helicases capable of binding and unwinding G4s, some of which have been shown to facilitate transcription

elongation (Gray et al., 2014b; Mohaghegh et al., 2001; Sun et al., 1998b). This supports the idea that G4s can form when the DNA is single-stranded during transcription, and there may be cellular mechanisms to overcome G4 formation.

This chapter explores whether the hypothesis that NMM treatment of plants causes decreased EXT expression by stabilising G4s and inhibiting transcription through the stalling of PolII, is explored. Additionally, it is investigated whether there are mechanisms within plants to regulate G4 formation such as helicases or hormones.

4.2.1 Aim

In this chapter, the aim is to determine whether G4s affect the transcription of *EXTs*.

4.2.2 Objectives

1) To analyse the impact of G4s on PolII processivity of *EXTs*

2) To establish whether G4 ligands affect *EXT* expression directly

3) To identify whether the *EXT* G4s is capable of binding nuclear proteins

4) To investigate the potential of helicases to resolve G4 structures

4.3 Results

4.3.1 G4s effect on PolII processivity

The literature suggests that G4s within the template strand of the gene body can result in stalling of PolII. G4s are predicted to form within this region in the *EXTs* and hence it was hypothesised that transcriptional downregulation of the *EXTs* by NMM may be due to PolII stalling. To investigate PolII stalling within the *EXT* genes under physiological conditions, Mathew Parker used publicly available PolII chip-chip data (Chodavarapu et al., 2010) to analyse the occupancy of PolII over *EXT* genes compared

to the average of all genes from *A. thaliana*. In this analysis (Fig 4.1, A), on average for all genes, PolII occupancy gradually increases towards the 3' end of the gene; however, in *EXTs* there is a large peak of PolII at the 5' end of the gene followed by a severe drop off in occupancy (Parker, 2018). Additionally, the mean PG4 distribution predicted using the Quadparser prediction method (but limited to two tetrad G4s) inversely correlates with the PolII occupancy within EXTs, suggesting that G4s formation could be associated with a mechanism that results in PolII stalling in EXTs. To further investigate whether PolII inhibition could result in *EXT* gene downregulation and similar root hair phenotypes to that seen with NMM, the PolII inhibitor α-amanitin was used. Plants were grown on media containing 0.1μM, 0.5μM, 1μM and a maximum of 5μM of α-amanitin as 5μM has previously been shown to inhibit PolII in *Arabidopsis* (Shin and Chekanova, 2014). No defects in root hair morphology were observed in α-amanitin treated plants. At the highest concentration of α-amanitin used (5μM), an increase in LRX1 expression was found, although this was not significant compared to controls (Fig 4.1, B). These findings may suggest that PolII inhibition does not cause the decrease in EXT expression, although further investigations should be carried out to verify the effectiveness of amanitin treatment including analysis of other *EXT*s and further concentrations of α-amanitin.

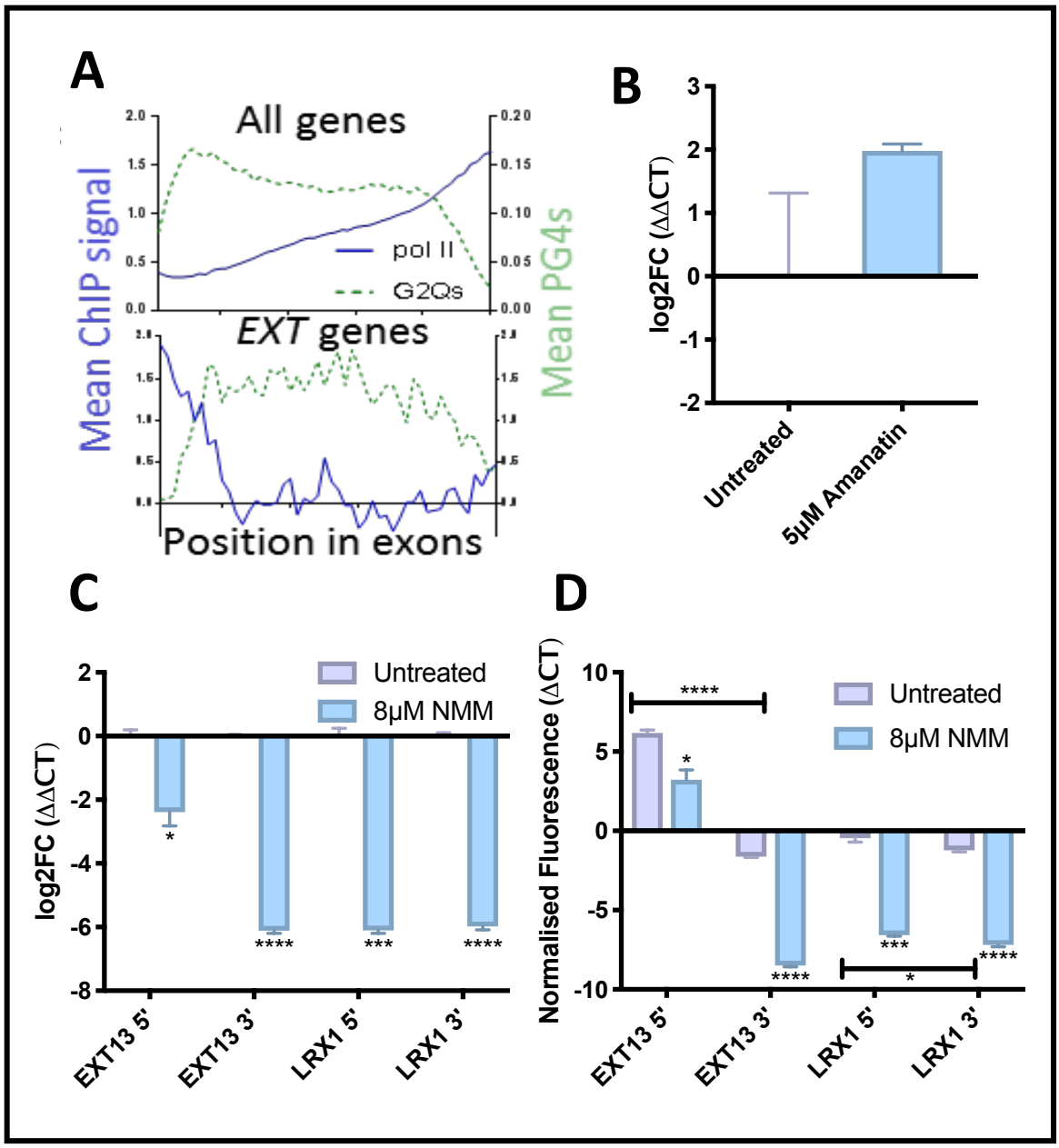

Figure 4.1. G4 pausing of RNA Polymerase II. (A) Analysis of PolII ChIP:ChIP data showing the mean PolII occupancy (blue) and PG4s (green) within the coding sequence of all *Arabidopsis* genes and *EXT* genes. **(B)** Log2 fold change in *LRX1* expression in roots, 7 days post stratification, treated with amanitin. The Log2 fold change **(C)** and normalised fluorescence **(D)** of the 5' and 3' region of *LRX1* and *EXT13* in NMM treated roots. Statistics carried out using a Student's T-test, * indicates a p-value of > 0.05, *** > 0.0005, **** > 0.0001, n > 3.

Stalling of PolII could increase the production of nascent or truncated mRNA, and, therefore, it would be predicted that treatment with NMM could result in a stronger downregulation at the 3' end of *EXT* genes, than the 5'. The effect of NMM on the relative expression of the 5' and 3' regions of *EXT* genes was therefore investigated by RT-qPCR. Some *EXTs* genes, such as *EXT13*, showed a stronger decrease in expression of the 3' region than the 5', supporting the hypothesis of PolII stalling and abortive transcripts (Fig 4.1, C). However, the 5' region of *EXT13* is down-regulated by NMM, and some *EXTs* such as *LRX1* have a similar downregulation of both 5' and 3' regions of the gene when treated with NMM (Fig 4.1, C). This effect is not explained by PolII stalling, which may suggest NMM is having multiple effects on EXTs or acting through a more complex mechanism. Though, analysis of the raw fluorescence values of the RT-qPCR, normalised using UBC10 as a housekeeping gene and the primer efficiency (the normalised fluorescence), consistently identify less 3' transcripts than 5' in both treated and untreated samples in all tested genes (Fig 4.1, D). This difference in 5' and 3' expression could be due to the G4s forming at physiological conditions to cause abortive transcription. Therefore, PolII stalling as a result of G4 stabilisation may cause the production of truncated or nascent transcripts in EXTs, in some cases, but does not explain the full effect of NMM treatment.

4.3.2 NMM downregulates regulators of EXT expression

The 5' region of *EXT* genes is downregulated by NMM and as the 5' region of *LRX1* contains only a single PG4, it is unlikely that the significant downregulation observed in the 5' region is caused by PolII stalling. Therefore, we investigated whether

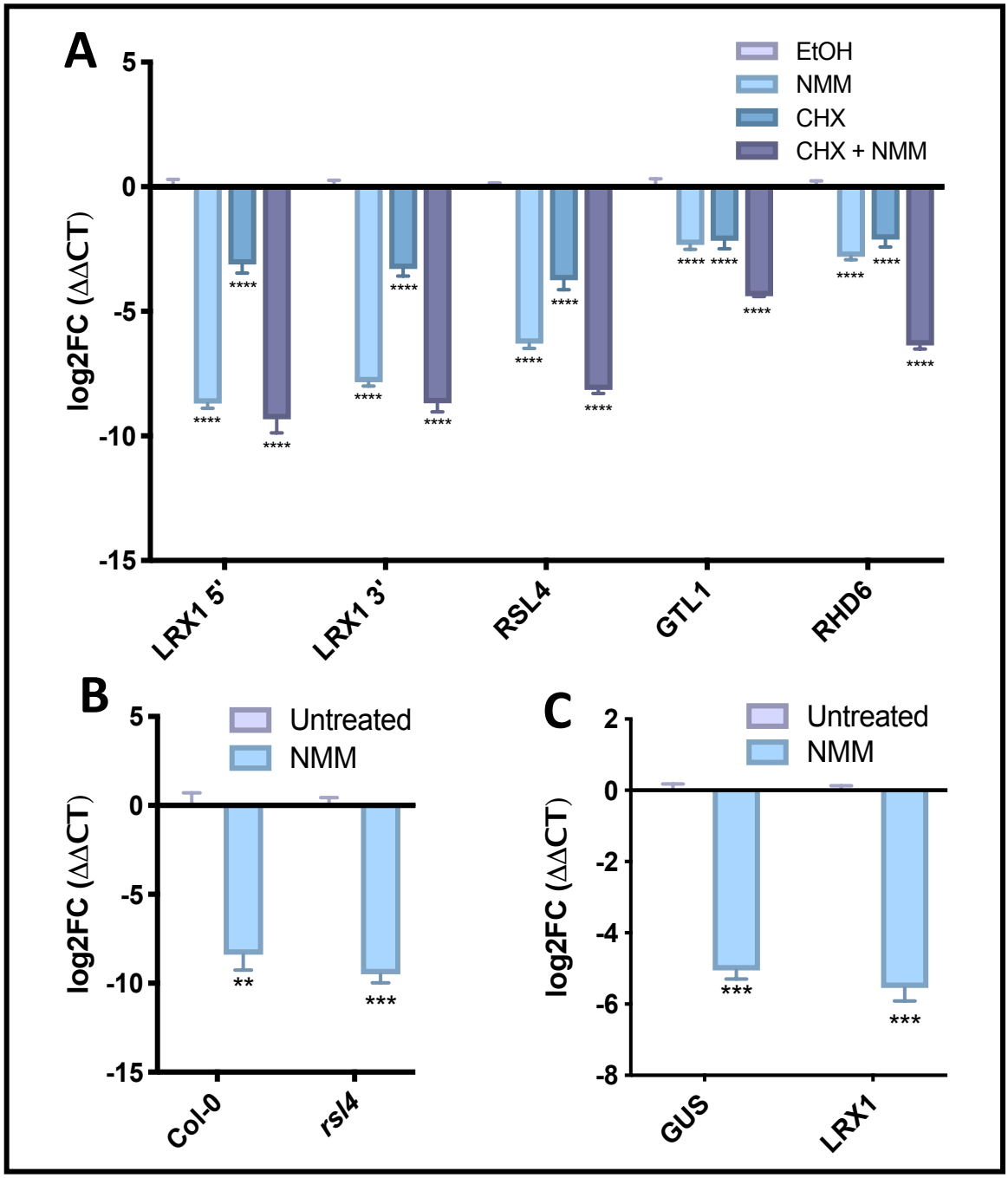

Figure 4.2. The effect of NMM on EXT promoter and transcription factors. (A) Log2 fold change in expression of *LRX1* and known EXT transcription factors in roots treated with NMM and CHX. **(B)** Log 2 fold change of *LRX1* expression in roots of NMM treated Col-0 and *rsl4* plant lines. (C) Log2 fold change of GUS and LRX1 expression in pLRX1:GUS plant line. Statistics carried out using a student's T-test, ** indicates a p-value of > 0.005, *** > 0.0005, **** > 0.0001, n > 3.

NMM could act through the promoter or an upstream regulator of *EXT* genes. As an indication of whether NMM was acting directly on EXTs, plants were treated with the translation inhibitor, cycloheximide (CHX), to determine whether active translation was required for NMM downregulation of *EXTs*. The plants treated with CHX were treated for 6 hours in media containing 10μM CHX and then transferred to media containing CHX and NMM or just further CHX treatment for another 6 hours. These treatment times where chosen as it has been shown that this is enough time for CHX to inhibit *de novo* protein synthesis and for one of the key EXT regulators RSL4 (Kim et al., 2013) to be almost entirely degraded before treating with NMM (Datta et al., 2015). We found that CHX significantly downregulated *LRX1* expression and that NMM further reduced expression in the presence of CHX (Fig 4.2, A). This finding implies that NMM is likely to have a direct effect on *LRX1* and does not require translation of a short-lived intermediary factor. However, it was also found that NMM reduces the expression of several positive and negative regulators of *EXT* genes including the key regulator, *RSL4*. The downregulation of *EXT* regulators caused by NMM occurs in combined treatment with CHX (Fig 4.2, A) implying a direct effect. We identified that within the coding region of the gene *RSL4* there was one PG4 on the template strand and four on the coding strand, and, *GTL1* contains seven PG4s on both the template and coding stand. Therefore, the G4s found within these EXT regulator genes may be the cause of the direct down-regulation by NMM. However due to the complications of a dual treatment experiment; such as not knowing how the two treatments may interact, how long each treatment takes to affect the cell and how long the effects last when combined, means further evidence is required to support this finding.

RSL4 is known to be a key regulator in EXT expression (Feng et al., 2017; Yi et al., 2010). Therefore, if NMM causes downregulation of *EXTs* by affecting an upstream regulator, then RSL4 is a potential target. To explore this hypothesis, *EXT* expression was analysed within the *rsl4* mutant following NMM treatment (Datta et al., 2015). NMM downregulates LRX1 expression in the absence of *RSL4* and this downregulation is comparable to the response in the WT (Fig 4.2, B). While further studies are required to exclude other regulators, this data supports a mechanism where NMM is having a direct effect on *EXT* and does not act through RSL4 alone.

To further investigate whether NMM can act through EXT promoters, the *pLRX1::GUS* plant line was used (Baumberger et al., 2001),which contains a transcriptional fusion of the *LRX1* promoter to the GUS (*uidA*) reporter gene. NMM significantly downregulated *GUS* expression in the *pLRX1::GUS* line (Fig 4.2, C), suggesting that NMM may have a significant effect on the *LRX1* promoter or an upstream regulator of LRX1, going against the findings from the CHX and *rsl4* experiments.

4.3.3 EXT G4 binds nuclear proteins

The PG4s in the coding region of the EXTs should be able to form under physiological conditions without a stabilising drug such as NMM. However, if NMM stabilisation of G4s causes downregulation of EXTs then this indicates that there may be cellular mechanisms to overcome G4 formation and allow regular expression of *EXTs*. Therefore, there is most likely a nuclear protein that binds and regulates G4 formation.

We investigated whether the most frequent PG4 sequence in *EXTs* (GGTGGTGGTGG) could bind nuclear proteins, which could regulate G4 formation or

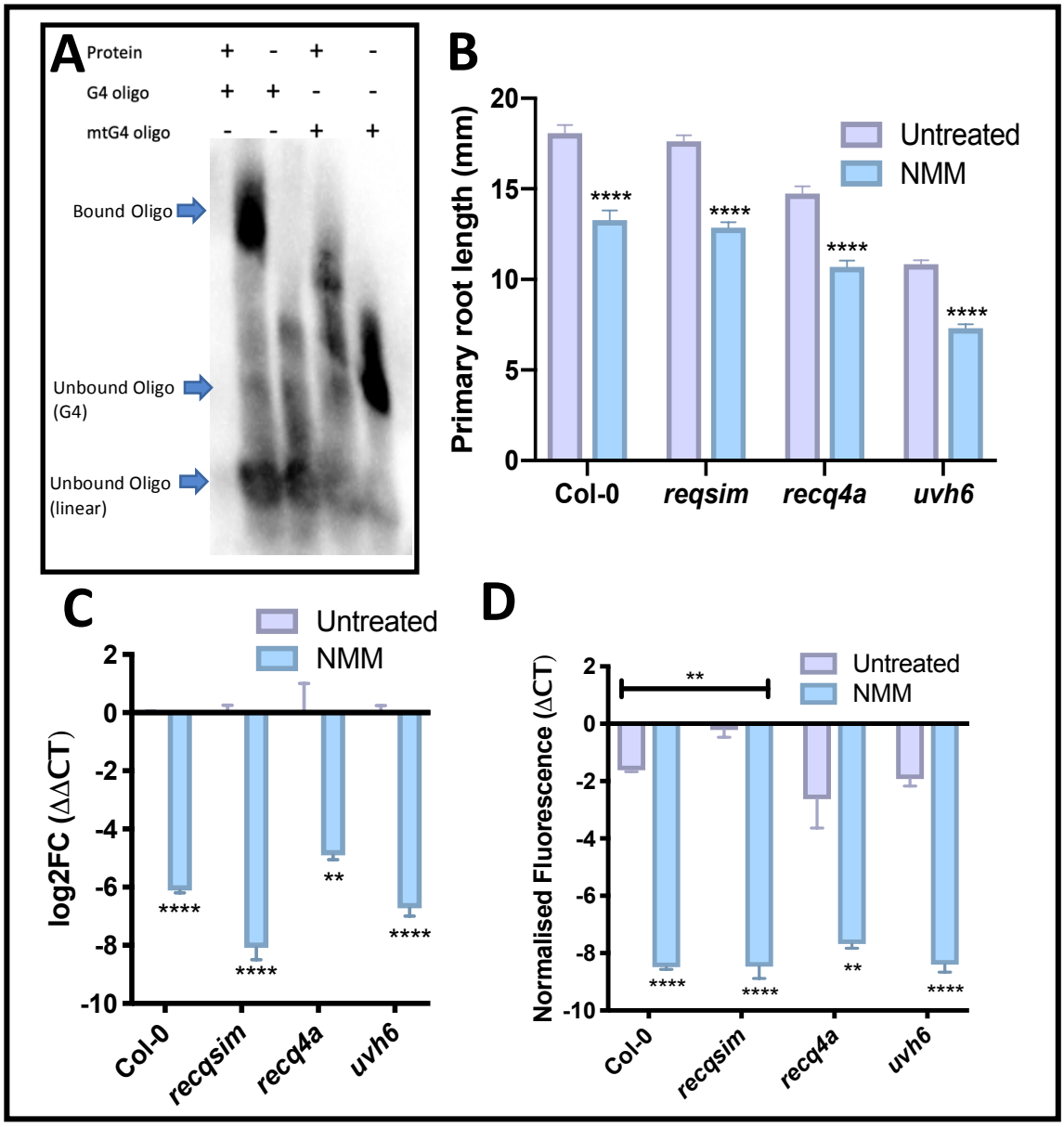

Figure 4.3. G4 protein binding and regulation by helicases. (A) Nuclear protein binding EMSA using biotinylated EXT G4 and mutated G4 oligo. **(B)** The effect of NMM on primary root length in helicase mutants. The log2 fold change **(C)** and normalised fluorescence **(D)** of EXT13 expression in Col-0 and helicase mutant lines. Statistics carried out using a student's T-test, ** indicates a p-value of > 0.005, **** > 0.0001, n > 3 for qPCR and EMSA, n>100 for root measurements.

that use G4s as binding sites. This was investigated using an electrophoretic mobility shift assay, in which biotin labelled oligos of the *EXT* repeat and a mutated *EXT* repeat were incubated with or without *Arabidopsis* nuclear protein extract. The oligos were then run on a gel, and any protein binding should cause a shift due to reduce mobility of the oligo-protein complex. The *EXT* G4 oligo shifted when incubated with protein suggesting protein binding (Fig 4.3, A). A shift was also seen in the mutated oligo, although not the same shift as with the EXT G4, therefore, this may be caused by the oligo forming an intermolecular G4 or alternative structure or binding a different protein. The binding of nuclear proteins to the EXT G4 forming sequence, which may suggest a potential mechanism to overcome the G4 or a functional role within the cell. Further analysis of the putative binding factors is required before further conclusions can be made.

4.3.4 The role of helicases as G4 regulators

This mechanism to overcome G4s may be through helicases which have been previously shown to bind and unwind G4s to facilitate transcription in mammals (Davies et al., 2007; Kruisselbrink et al., 2008; Paeschke et al., 2011; Piazza et al., 2015; Ribeyre et al., 2009; Rodriguez et al., 2012b; Singleton et al., 2007; Sun et al., 1998a; Van Wietmarschen and Lansdorp, 2016; Wu and Spies, 2016; WU, 2007). Therefore, plant lines with mutations in; *recqsim* (a WRN helicase homologue), *recq4a* (a BLM helicase homologue) and *uvh6* (a XPD homologue) were analysed. These are all homologues of helicases previously shown to act upon G4s in other systems (Davies et al., 2007; Gray et al., 2014a; Johnson et al., 2010b; Liu, 2003; Mohaghegh et al., 2001; Popuri et al., 2008; Sun et al., 1998a). The helicase mutant lines all had a similar change in primary root

length upon NMM treatment as Col-0 however *recq4a*, and *uvh6* have significantly shorter roots in untreated plants than Col-0 which may suggest these helicases are involved in root growth. Analysis of *EXT13* expression was used to investigate the plant's sensitivity to NMM and demonstrated that *recqsim* mutant is more sensitive to NMM than Col-0 (Fig 4.3, C). Therefore, this helicase may have a role in negating the NMM response, potentially through unwinding EXT G4s. However, the normalised fluorescence analysis of *recqsim* indicates an increased transcript level *EXT13* in untreated plants compared to Col-0 (Fig 4.3, D) suggesting that RecQsim is not required for *EXT13* expression. These results may suggest a more complex or non-direct mechanism through which RecQsim effects EXT expression. Alternatively, recq4a is less sensitive to NMM compared to Col-0 (Fig 4.3, C), potentially due to the G4 already being stable in these plants.

4.4 Discussion

We propose that G4s may form naturally under physiological conditions, without the presence of a G4 ligand and cause PolII stalling during the transcription of *EXT* genes. PolII ChIP-ChIP data shows altered PolII occupancy that correlates with the presence of PG4s (Fig 4.1, A) suggesting the G4s within EXTs may cause PolII stalling as has been previously shown (Chambers et al., 2015; Cogoi and Xodo, 2006; Dexheimer et al., 2006; Han et al., 1999b; Kwok et al., 2016; Siddiqui-Jain et al., 2002). However, analysis of the PolII inhibitor, amanitin, caused a slight increase in *EXT* expression, although not significant (Fig 4.1, B). Unfortunately, we were unable to test a large concentration gradient of amanitin and while 5µM treatment has previously shown changes in gene expression within *Arabidopsis* (Shin and Chekanova, 2014) other studies have indicated

that a ten-fold higher concentration may be required (de Jimenez et al., 1984). Therefore, amanitin may cause root hair defects and lead to lower *EXT* expression if a wider concentration range was tested and more *EXT*s analysed.

PolII stalling by G4s has previously been suggested to lead to truncated or aberrant transcripts. Therefore, the 5', and 3' levels of *EXT*s was analysed using RT-qPCR (reverse transcription was done using random primers to avoid 3' bias) as the presence of aberrant transcripts would cause decreased 3' levels. NMM caused a stronger downregulation of 3' transcripts than 5' transcripts in some *EXTs* such as *EXT13*, supporting the hypothesis of the production of aberrant transcripts. Although, in some *EXT* genes, such as LRX1, NMM causes a significant downregulation of both the 5' and 3' regions, suggesting that PolII stalling cannot entirely explain the *EXT* downregulation caused by NMM. The raw RT-qPCR fluorescence values were analysed by normalising to UBC10 (housekeeping gene) and balancing for any differences in primer efficiencies between the 5' and 3' primers (normalised fluorescence). Interestingly, the normalised fluorescence, of all EXTs analysed, have a decreased expression of the 3' region in untreated samples which could be caused by G4s forming under physiological conditions, without the presence of a ligand, and resulting in truncated transcripts.

NMM reduced expression of both the 5' and 3' region of *LRX1* similarly which would not be predicted to be caused by PolII stalling. This finding led to the hypothesis that there may be a more complicated mechanism involved in *EXT* downregulation by NMM, in which NMM causes a downregulation by an upstream regulator of *EXTs* as well as PolII stalling. Therefore, it was investigated whether active translation was required for NMM downregulation of *LRX1*, suggesting a non-direct effect. The samples were pre-

treated with the translation inhibitor, CHX, for 6 hours which has been shown to be long enough to stop most de novo protein synthesis (Kim et al., 2013) and for any existing RSL4 protein to have been degraded (Datta et al., 2015). NMM downregulated both the 5' and 3' region of *LRX1* in the presence of CHX (Fig 4.2, A), implying that NMM has a direct effect on the 5' and 3' region of *LRX1* and, therefore, may not be caused by a feedback mechanism. However, NMM directly downregulates several transcription factors known to regulate *EXT* expression, including RSL4 in the presence of CHX (Fig 4.2, A). As RSL4 is a key transcriptional regulator of *EXT* genes, and is significantly downregulated by NMM, the requirement of RSL4 for NMM downregulation of *LRX1* was investigated by analysis of an *rsl4* mutant. NMM significantly downregulates *LRX1* in the absence of RSL4 (Fig 4.2, B) suggesting that NMM likely does not act on or upstream of RSL4 in *EXT* transcriptional regulation. However, *pLRX1::GUS* plants treated with NMM were analysed to further determine whether the *LRX1* promoter is involved in NMM downregulation and GUS was significantly downregulated by NMM (Fig 4.2, C). In contrast with the CHX and *rsl4* data, this implies that NMM affects the *LRX1* promoter or upstream regulators. The *LRX1* promoter contains two PG4 on the coding and template strand which overlap with the predicted binding sites of several transcription factors including MYC3 which has been shown to affect root growth through jasmonic acid signalling (Schweizer et al., 2013) and LBD16 which affects lateral root development through auxin signalling (Goh et al., 2012). However, interpretation of the data is complicated by the fact that the GUS gene contains seven PG4s on the coding strand and 3 PG4s on the template stand which could also cause downregulation upon NMM

treatment of the *pLRX1::GUS* line. Therefore, further studies are required to clarify the role of the *LRX1* promoter in EXT regulation.

If NMM causes EXT downregulation through G4 stabilisation and the G4s within EXTs form under physiological conditions, then there may be a cellular mechanism to overcome G4s. A potential regulatory mechanism of G4s would most likely involve binding of a nuclear protein. Therefore, we used an EMSA to explore the protein binding capacity of the EXT G4 sequence, which showed a shift in oligo, indicative of protein binding when the EXT G4 oligo was incubated with *Arabidopsis* nuclear protein extract. This result suggests that the EXT G4 may bind a nuclear protein that could regulate its formation, although the shift seen in the mutated G4 oligo may suggest this binding is sequence specific binding rather than structural, meaning that further research is required to confirm G4 dependent binding and identify the specific proteins binding to the EXT G4.

Several studies demonstrate the role of helicases in binding and unwind G4s, they are potential candidates for regulating EXT G4 formation. Investigation of candidate helicases identified that, *recqsim,* the WRN homologue, when mutated has increased sensitivity to NMM (Fig 4.3, A) which may suggest the requirement of *RecQsim* in regulating G4s and EXT expression, but the *recqsim* plants have higher basal *EXT13* expression (Fig 4.3, B) which cannot be explained by loss of G4 unwinding and may suggest a more complex action of RecQsim such as unwinding the DNA double helix which would facilitate G4 formation. However, *recq4a* has lower *EXT* expression in untreated plants and is less affected by NMM, which may suggest that G4s are stable in untreated *recq4a* plants, making *recq4a* a potential regulator of G4s in *Arabidopsis*.

In conclusion, our data indicate a link between G4s, PolII stalling and production of aberrant transcripts in some EXTs. NMM may have a direct effect on *EXT* genes despite regulating several key transcription factors of *EXTs*. Plus, we identify the capability of the EXT G4s to bind nuclear proteins. A candidate *EXT* G4 binding protein is RecQ4a for further study into its role in regulating EXT expression.

5. The Effect of G4 on Alternative Splicing

5.1 Abstract

Previous RNA-Seq data suggests that several *EXT* genes are alternatively spliced despite being annotated as single exon genes. As splicing can be a co-transcriptional process, this led to the hypothesis that G4 dependent stalling of PolII may cause alternative splicing of *EXTs.* In this chapter the presence of alternatively spliced *EXT* isoforms is confirmed by Sanger sequencing of RT-PCR products. NMM appears to decrease the number of splice variants in *Arabidopsis*, suggesting that G4s may play a role in alternative splicing. Also, it is identified that altering the number of EXT domains within *LRX1,* as would be found in the splice variants, affects root hair development.

5.2 Introduction

Several studies have demonstrated that splicing is a co-transcriptional process and it has been shown that the speed of PolII during transcription can affect whether a gene is alternatively spliced (Ameur et al., 2011; Girard et al., 2012; Khodor et al., 2011; Oesterreich et al., 2010; Tilgner et al., 2012; Windhager et al., 2012). Therefore, it has been suggested that the slowing or stalling of PolII allows the assembly of the spliceosome at weak splice sites, whereas fast PolII processing can result in splice junctions not being detected (Jonkers and Lis, 2015; Mata et al., 2010).

A previous study that carried out RNA-Seq of Arabidopsis roots identified that *EXT9* has the most splice isoforms in the *Arabidopsis* transcriptome and a large number of the genes within the *EXT* family have several splice isoforms (Li et al., 2016a). This finding was particularly interesting as the majority of *EXTs* are annotated as consisting of

a single exon, suggesting that the *EXTs* undergo splicing. Therefore, this chapter investigates the hypothesis that slowing of PolII, as a result of G4 formation, may result in alternative splicing of *EXTs*. In this chapter, it is investigated whether *EXTs* are alternatively spliced, whether splicing of *EXTs* could affect root hair development and if G4 stabilisation by NMM affects *EXT* splicing.

5.2.1 Aim

In this chapter, the aim was to identify the role of G4s in the alternative splicing of *EXTs*

5.2.2 Objectives

1) To validate whether *EXTs* are alternatively spliced.

2) To identify whether alternatively spliced *EXTs* could have a functional role in root hair development.

3) To establish whether *EXTs* are differentially spliced when plants are treated with NMM.

5.3 Results

5.3.1 EXT genes are alternatively spliced

EXT genes appear to be highly spliced, with EXT9 being the most highly spliced gene in Arabidopsis (Li et al., 2016a). Interestingly, the majority of *EXT*s are annotated as consisting of a single exon, with no predicted splicing. As the *EXTs* contain the repetitive SP4 motif, it is possible that mapping errors of the RNA-Seq data could account for incorrectly annotating these splice variants. To test this, RNA was extracted from roots and reverse transcribed using a Poly-T primer (to reverse transcribe

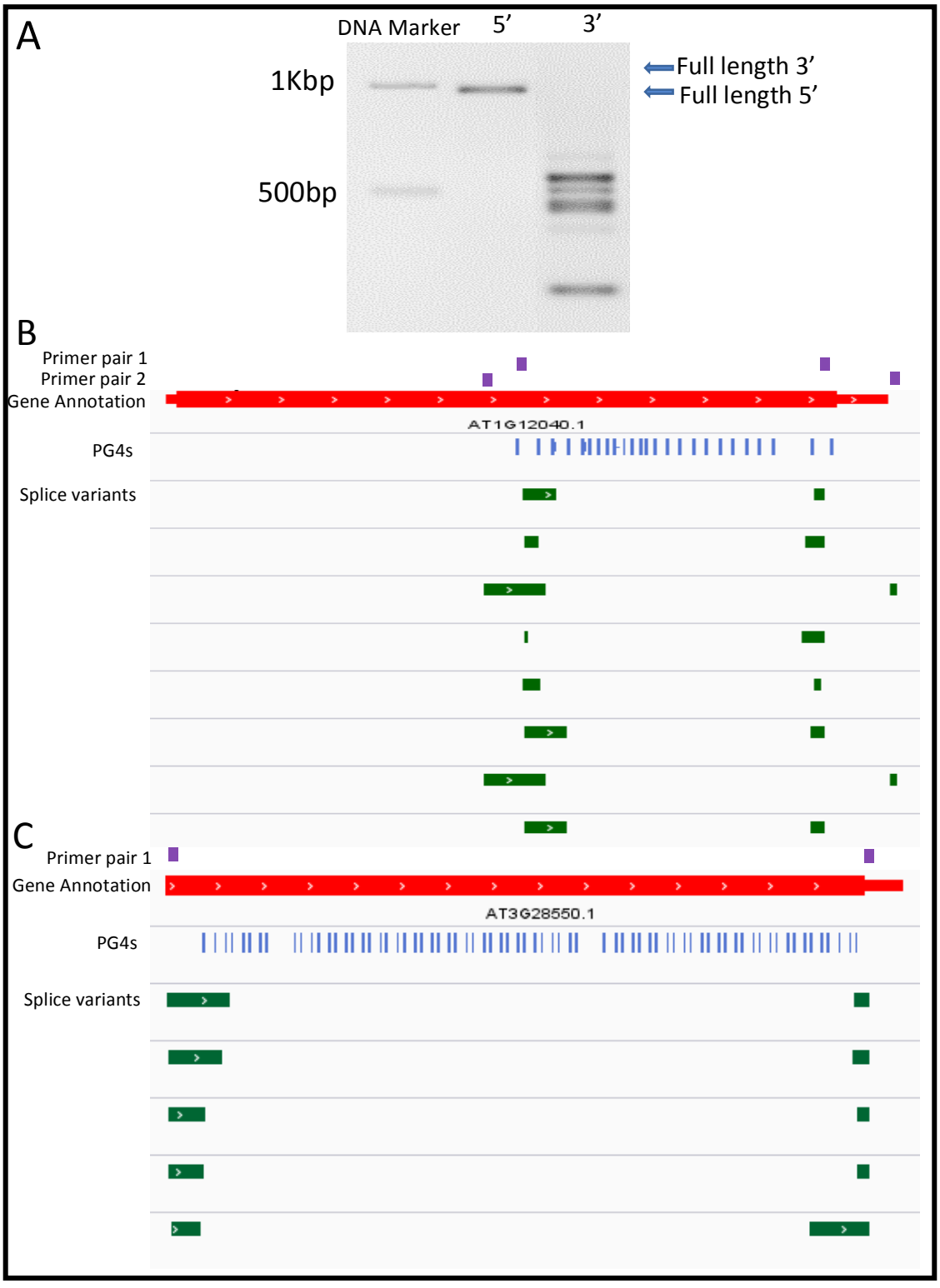

Figure 5.1 LRX1 is spliced over the G4 rich region. (A) RT-PCR of the *LRX1* 5' leucine rich region and 3' *EXTENSIN* G4 rich region**.** Sequenced splice variants of **(B)** LRX1 and **(C)** EXT9**,** sequencing primers shown in purple, gene annotation in red, predicted G4s in blue and sequenced splice variants in green.

only poly-A processed transcripts), and PCR amplified the specific *EXT* genes. The resulting PCR products were then gel purified, cloned and sequenced.

Primers were designed for LRX1, spanning the 5' LRR domain (containing a single PG4) and the 3' EXT domain (containing 39 PG4s), to investigate the potential role of G4s in splicing. In the absence of splicing, it would be anticipated that each primer pair would generate a single product following RT-PCR, assuming that there were no off-target amplification events. The results indicate that the 5' region commonly had a single product corresponding to the full-length PCR product (Fig 5.1, A). On the other hand, the 3' PG4 rich region of LRX1 had several major products, none of which corresponded to the expected size of the full-length 3' PCR product (Fig 5.1, A).

To ensure these products where not due to nonspecific target amplification, individual products were cloned and Sanger sequenced and it was found that the products corresponded to several different LRX1 splice forms (Fig 5.1, B). When the experiment was repeated with another set of primers, several additional alternative splice variants were identified suggesting many independent isoforms exist (Fig 5.1, B). The high abundance of splice variants only in the 3' region of LRX1, where the PG4s are abundant, may support the hypothesis that G4 dependent stalling of PolII causes the alternative splicing, or suggests that the SP4 EXT domain encodes a splice site. The experiment was repeated with EXT9 to investigate the occurrence of alternative splicing in a classical extensin, and confirm the splicing seen within the previous RNA-Seq data set. The primers were designed towards the whole gene, as EXT9 has pG4s throughout the gene body. We determined that EXT9 is also alternatively spliced, confirming the RNA-Seq data (Fig 5.1, C). Due to the repetitive nature of the EXTs the exact splice site of

the sequenced variants tended to be ambiguous and the true splice site could not be mapped, however the variants where the junctions could be mapped did not align to the canonical GT-AG splicing sequence (Figure 5.2). Furthermore, for both genes only short splice variants were sequenced, potentially due to cloning bias of the shorter transcripts; however, this confirms the presence of alternative splice variants within EXTs.

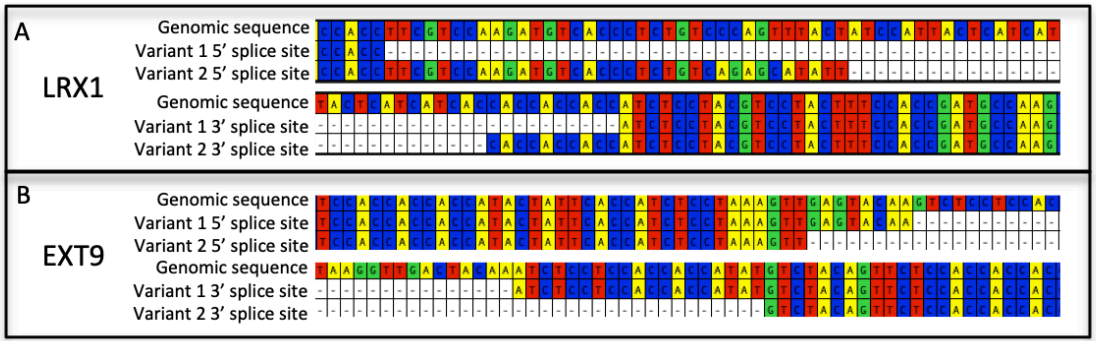

Figure 5.2 LRX1 and EXT9 splice junctions. Aligned splice variant junctions for (A) LRX1 and (B) EXT9 with the top row displaying the genomic DNA sequence and the 5' or 3' splice site for two variants shown below. Sequences manually aligned using MacVector software.

5.3.2 The number of extensin repeats affects root hair development

It has previously been shown that when *LRX1* is truncated to contain a decreased number of extensin repeats, there is a significant effect on root hair morphology. As the various truncated *LRX1* transgenic lines may be similar to the transcripts produced by alternative splicing of *LRX1,* analysis of these lines could give an insight into how splicing isoforms may affect root hair development. Therefore, the truncated LRX1 transgenic lines were analysed for root hair phenotypes (Ringli, 2010). The LRX1 truncation lines have various size truncations of LRX1 with an N-terminal myc tag, expressed in the *lrx1* mutant line (Baumberger et al., 2001). Unfortunately, the full-length *mycLRX1* and *mycLRX1ΔE296* were not available*;* therefore, the lines were compared to the

background ecotype of wildtype Col-0. Analysis of the truncated lines demonstrates that seedlings expressing the longest *LRX1* variant, *mycLRX1ΔE153,* have no clear root hair phenotype (Fig 5.2, B). The shortest variant, *mycLRX1ΔE14,* displayed a severe bulbous root hair phenotype with the root hairs having a bulbous base or appearing globe-shaped (Fig 5.2, D). A comparison of the *mycLRX1ΔE153* and *mycLRX1ΔE90* lines suggests that for correct cell wall development and root hair morphology, EXTs require a minimum number of EXT domains (Fig 5.2, C). However, we predict that the truncated lines would also undergo alternative splicing, meaning we cannot be certain of the

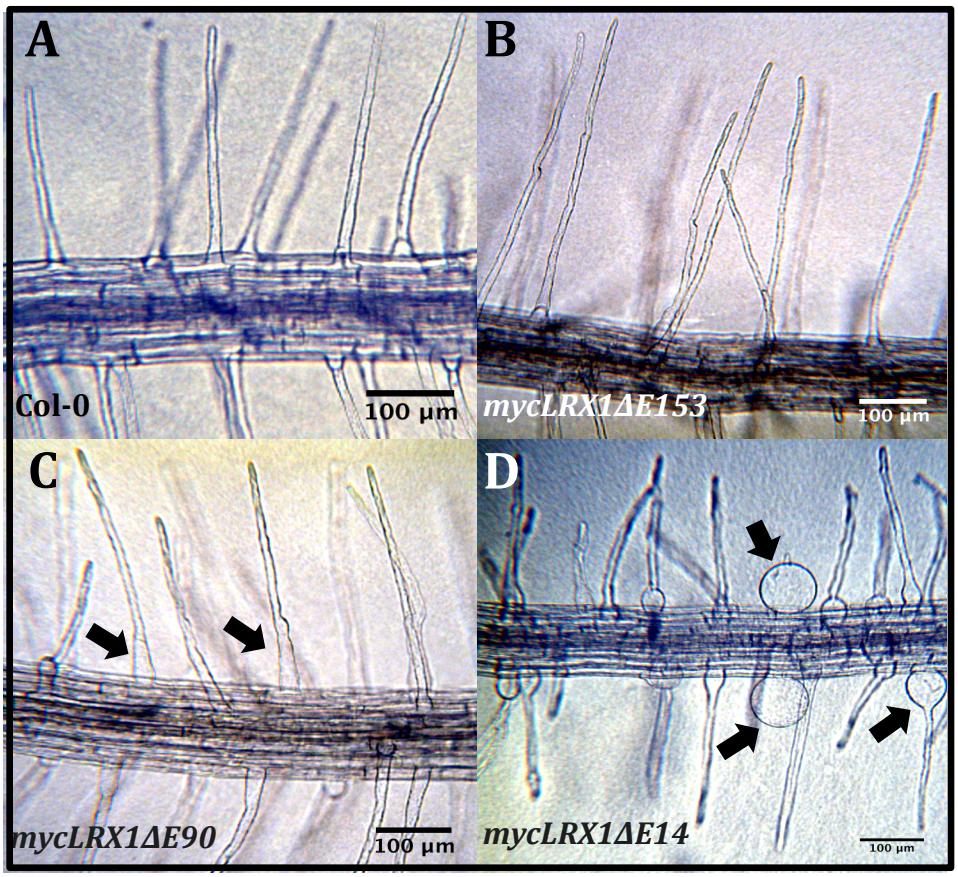

Figure 5.3 Decreased EXT repeats in LRX1 cause bulbous root hair. Representative root hair phenotypes of (A) Col-0, (B) *mycLRX1ΔE153*, (C) *mycLRX1ΔE90*, (D) *mycLRX1ΔE14*. Arrows indicate bulging morphology, n => 30.

number of EXT domains within the transcripts produced from the truncated lines; therefore, the minimum amount of EXT repeats required may be fewer than predicted. However, these results indicate that less than half of the EXT domains encoded for by the full-length LRX1 gene is required for normal root hair morphology and development. Therefore, plants may utilise the alternative splicing mechanism to create greater molecular diversity of EXTs for cell wall construction.

The response of the LRX1 truncation lines to NMM was also investigated. It was predicted that the shortest transcript with the fewest EXT repeats, *mycLRX1ΔE14,* would be less responsive to NMM as it has fewer G4s; therefore, less inhibition of PolII. Analysis of *LRX1* expression by qPCR found no differential response to NMM between

Col-0 and the truncated lines (Fig5.3, A). This result could suggest that the two PG4s in *mycLRX1ΔE14* may be sufficient for NMM to affect transcription. However, this may also be explained by the position of the primers used for the qPCR. As the 3' regions of the truncated lines were all different, the qRT-PCR primers were located within the LRR domain, which is before the majority of the G4s. Therefore, further investigation is required to analyse whether NMM differentially regulates the truncated lines in the 3' region. Additionally, the primary root length in the untreated plant lines does not significantly change; however, the primary root of *mycLRX1ΔE90* and *mycLRX1ΔE14* is less responsive to 0.5µM NMM than Col-0 and *mycLRX1ΔE153* (Fig 5.3, B) indicating that the number of EXT repeats may affect the sensitivity to NMM.

5.3.3 NMM does not affect the alternative splicing of extensins

If the splicing in *EXT*s is caused by G4 formation, this suggests that G4 stabilisation by NMM would affect the splicing of *EXT*s. Therefore, the RT- PCR splicing

analysis was repeated with roots treated with or without NMM for 6 hours. The NMM sample appeared to generate less of the smaller splice products (Fig. 5.4). While it was predicted that NMM would cause more splicing due to increased PolII stalling, NMM may cause all of the G4s within the gene to form stably, whereas in untreated plants different G4s may form at different times resulting in a greater variety of splice variants.

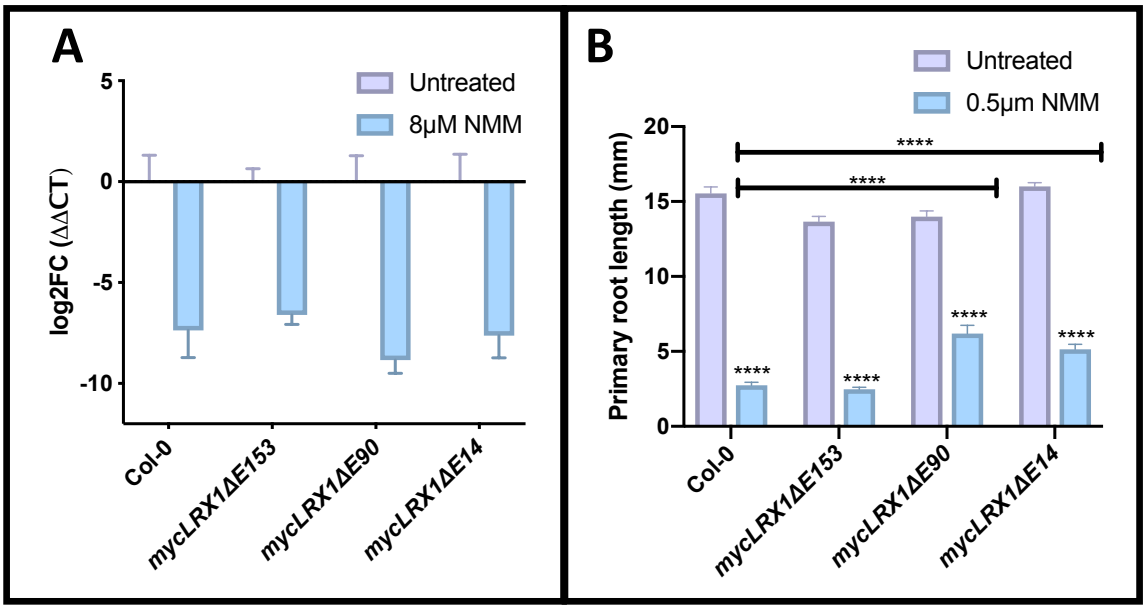

Figure 5.4 NMM sensitivity of LRX1 truncated plant lines. (A) qRT-PCR analysis of *LRX1* expression in untreated and NMM treated, Col-0 and *mycLRX1* truncated lines and (B) primary root lengths of Col-0 and the truncated LRX1 lines. RT-qPCR n = 3, root length n > 30. Error bars show standard error of mean. Statistical test carried out using multiple comparison ANOVA, **** = p-value > 0..0001

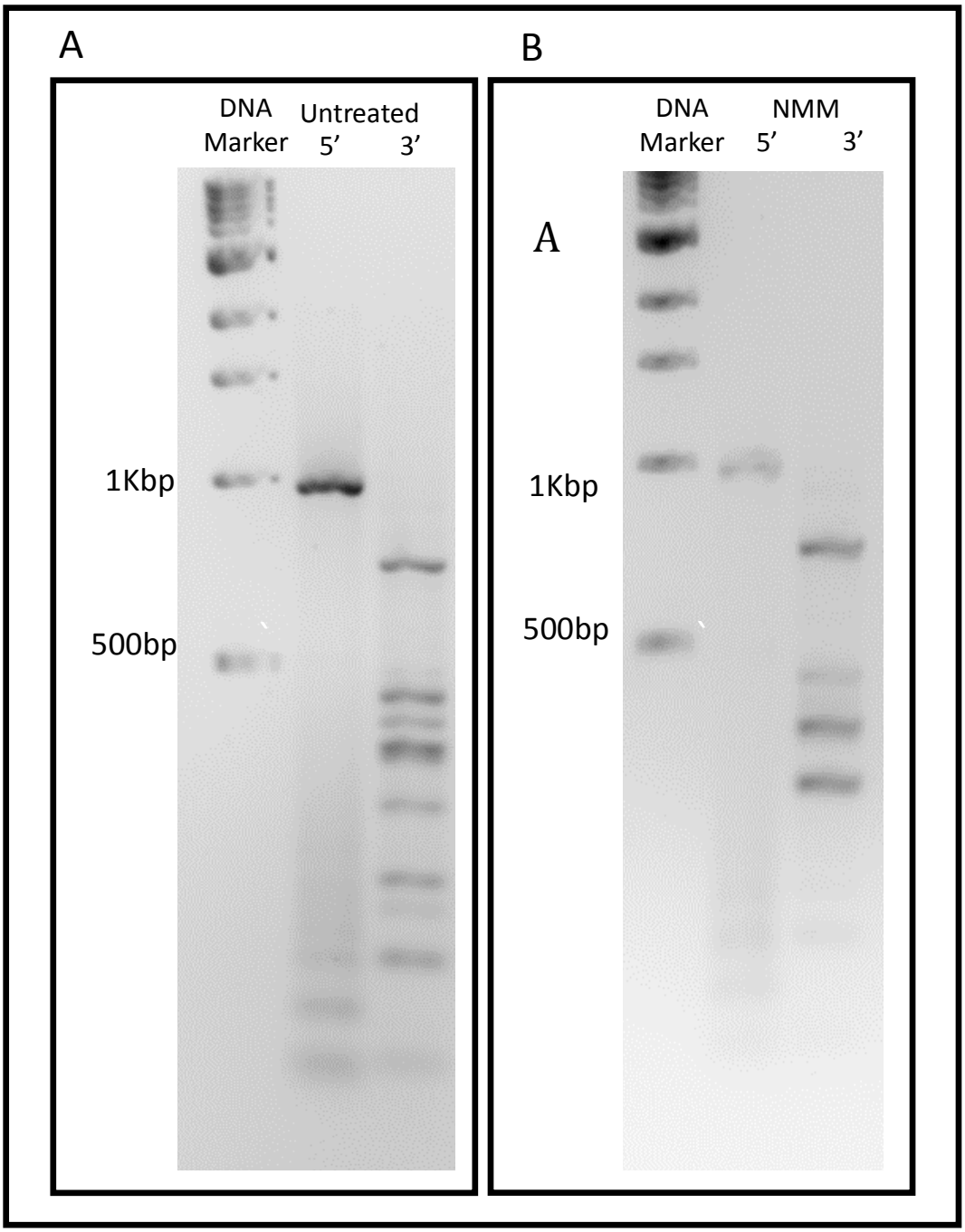

Figure 5.5 The effect of NMM on the splicing of *LRX1*. RT-PCR of the *LRX1* 5' leucine rich region and 3' *EXTENSIN* G4 rich region in **(A)** untreated and **(B)** NMM treated plants. Images where taken under the same exposure.

5.4 Discussion

Previous RNA-Seq data suggested that *EXTs* have many splice isoforms (Li et al., 2016a). However, as *EXTs* are highly repetitive, and the RNA-Seq experiment used paired-end reads of a maximum length of 125bp, there was a possibility that the splice isoforms identified could have been due to mapping errors. Using RT-PCR and Sanger sequencing, the presence of numerous polyadenylated splice isoforms of *EXT9* and *LRX1* was confirmed (Fig.5.1, A-C). Only the sequence of the shortest variants was obtained, potentially due to a bias towards small fragments in the cloning, used to isolate single isoforms. We were unable to determine the exact splice site of the EXTs splice variants

sequencing, the presence of numerous polyadenylated splice isoforms of *EXT9* and *LRX1* was confirmed (Fig.5.1, A-C). Only the sequence of the shortest variants was obtained, potentially due to a bias towards small fragments in the cloning, used to isolate single through Sanger sequencing as their repetitive nature made the exact junction difficult to align. The splice sites that were able to be determined did not align with canonical splice sites (Fig 5.2) and therefore we cannot rule out that the splice variants may be a result of reverse transcriptase skipping the structured regions of the EXTs during cDNA synthesis, emulating splicing which has been previously reported (Cocquet et al., 2006; Houseley and Tollervey, 2010; Zhang et al., 2001). Therefore further analysis by northern blots could be used for several EXT genes to thoroughly analyse the extent of splicing within the EXTs and the full effect of NMM.

The sequenced splice variants were reverse transcribed using a PolyT primer, meaning that all the splice variants are poly-adenylated and could be functional. A previous study found that plants require at least 5 EXT repeats in *LRX1* to form wildtype root hairs (Ringli, 2010). The LRX1 truncated transgenes would be similar to the spliced

isoforms; therefore, it was investigated how these truncated transcripts affect root hair morphology. The most truncated line, *mycLRX1ΔE14,* had a bulbous root hair phenotype (Fig.5.3, D). However, the lines with smaller truncations were able to form mostly wild-type root hairs in appearance (Fig.5.3, A-C). This finding indicates that the plant can still utilise these truncated transcripts (similar to the splice isoforms) although it would suggest that there is a minimum number of EXT repeats that are required for correct growth and development. The requirement of a particular number of EXT may be due to the crosslinking function of the HYPs within the *EXT* domain, up to certain lengths the proteins can still effectively crosslink to form a stable cell wall; however, too few EXT motifs cause inadequate crosslinking and an impaired cell wall.

As the truncated *LRX1* lines also contain less pG4s, it was proposed that they would be less affected by NMM. However, there was no significant difference in expression of *LRX1* before or after NMM treatment (Fig.5.4, A), between the truncated lines and Col-0. The expression change may be due to *mycLRX1ΔE14* still containing enough G4s for NMM to take effect despite being severly truncated, or due to the primers used to measure the expression, as they are in the 5' region of the gene, upstream of the majority of the pG4s. We also analysed the decrease in primary root length caused by NMM in the truncated lines, which while there was no significant difference between untreated plant lines, the lines with the fewest EXT repeats, *mycLRX1ΔE90 and mycLRX1ΔE14,* were significantly less affected by NMM (Fig.5.4, B), further supporting the idea that the effect seen on the primary root is G4 dependant.

Finally, it was investigated whether NMM could affect the splicing of *EXTs*. We proposed that G4 stabilisation by NMM would cause further decreased PolII processing

of the gene and potentially result in increased alternative splicing. However, splicing (in particular smaller splice variants) was decreased in NMM treated plants (Fig.5.5), which may be caused by NMM treatment causing all of the G4s within the gene to form, whereas in untreated tissue there is a variability in which G4s form, which would result in a greater variety of splice variants formed in comparison to NMM treated plants. Although the bands missing in the NMM treated samples were faint in the Col-0; therefore, the loss of bands could potentially be due to an overall decrease in expression rather than loss of splicing. Northern blots could be used to analyse this and potentially use a concentration gradient of NMM treatments to investigate whether the NMM concentration alters the variability of splicing.

In conclusion, we have suggested that *EXT9, EXT6* and *LRX1* are extensively alternatively spliced despite commonly being annotated as single exon genes. Additionally, the splice transcripts may be functional as they are poly-adenylated and truncations in LRX1 can still produce wildtype root hairs. Therefore, the alternative splicing of EXTs could be a mechanism to increase molecular diversity to allow for cell wall assembly. The splicing of EXTs appears to occur in the region of the PG4s and may be altered by NMM treatment suggesting a novel G4 dependent splicing mechanism, although, this should be confirmed using northern blots and using lines that can no longer form G4s to investigate whether splicing still occurs.

6. Discussion

6.1 Study Overview

G4s research is quickly advancing, with G4s having rolesin several fundamental biological processes and multiple diseases. Here we present evidence for a novel G4-dependent growth mechanism within Arabidopsis and provide insight into the transcriptional roles of gene body G4s.

In this study, it is shown that *Arabidopsis* grown on the G4 ligands, NMM and Berberine, causes an overall decrease in growth (Fig 3.1, C,D). The negative impact of these G4 stabilising ligands on overall growth, suggests that G4s may have regulatory roles in plant growth pathways. This finding means that G4 ligands have the potential to be developed as herbicides. Interestingly, Berberine has previously been investigated as a herbicide, with the suggestion that it causes cytotoxicity through cell wall, mitochondria, nuclei and chromatin damage (Wu et al. 2017; X.-H. Zhang et al. 2018). Additionally, by further understanding the plant growth mechanisms targeted by these G4 ligands, there is the potential to manipulate them to improve yields in crops.

Although we see overall growth defects, the previously observed shoot phenotypes (Nakagawa et al. 2012) were not identified. These phenotypes may not have been seen due to differences in photoperiod used, which would require future study. However, under the growth conditions utilised in this work, G4 ligands had an inhibitory effect on root length and also affected root hair morphology (Fig 3.3,A-F). It was identified that the root hairs needed to be in direct contact with the ligand-containing media in order for the phenotype to be present, which may suggest either uptake or

transport issues for the G4 ligands or that the bulbous root hair morphology requires a combination of both a physical obstruction (the media) and the presence of the ligands. Monitoring the uptake and transport of these ligands at the cellular level in plant organs and tissues would be technically challenging. For example, the addition of a fluorescent dye to enable tracking could also have an impact on uptake/transport or on the interaction with nucleic acids, while isotope labelling might not give significant tissue level resolution. It would also be interesting to utilise G4 ligands with preferences for alternative G4 configurations given that NMM has a preference to bind parallel G4s; this may allow mapping of the prevalence of different configurations as well as whether they have different biological roles.

To understand how G4 ligands affect plant growth we analysed how NMM affects gene expression. Analysis of the gene expression in NMM treated plants by microarray revealed significant changes in gene expression which correlated with other transcriptome studies in which Berberine was used (Nakagawa et al., 2012), suggesting that the majority of the changes are due to G4 binding and not the other effects of NMM. We identified from the microarray and confirmed by qPCR that the *EXT* gene family are significantly downregulated by NMM (Fig 3.5, A-C). We also utilised the mycLRX1Δ153 line to show that NMM also decreases protein levels of the EXTs (Fig 36 D). As this is a truncated form of LRX1, it would be good to confirm the downregulation of EXT proteins by repeating using a full-length myc-tagged EXT and preferably other EXT proteins to be certain it is not a single gene effect. Previous studies have mutated the guanines within the sequence to stop G4 formation, to ensure the effect seen on expression is G4 dependent. However, this cannot be done in the EXTs as if the guanines

are altered then this alters the proline codon, causing a change in protein sequence. However, there is evidence suggesting that altering the bases in the loop sequences of a G4 can make it unfavourable to form (Guédin et al., 2008). Therefore, it is feasible that we could alter the loop sequence of the EXT G4, which is the wobble base of the proline codon, making the G4 unfavourable without altering protein sequence. There is the potential that this could affect codon usage and translation, however the creation of these EXT mutant lines would be beneficial in determining that the effects reported in this study are dependent on G4 formation.

The EXTs are highly enriched for PG4s on the template strand due to the HYP repeats and therefore guanine-rich (Fig 3.4, A-B). It has been shown that Arabidopsis has a significant enrichment for two-quartet PGs within the template strand (Mullen et al., 2010, 2012). This has not been highlighted in humans, though this could be due to lack of investigation as many believe two-quartet G4s are not stable at body temperature, despite studies suggesting they remain stable at much higher temperatures (Bock et al., 1992; Fry and Loeb, 1994; Petraccone et al., 2005; Smirnov and Shafer, 2000). However, understanding of how template strand G4s affect the EXTs could provide insight into their roles in the many other genes, within Arabidopsis and in other organisms.

We demonstrate using CD spectroscopy that the EXT SP4 motif forms a parallel G4 *in vitro* (Fig 3.4, D), which may explain the significant effect of NMM on their expression. Additionally, plants with mutations in the *EXT* genes *LRX1* and *LRX2* have bulbous root hair phenotypes similar to NMM treated plants although less severe (Baumberger et al. 2003)(Fig 3.6, C+D). The EXT mutants do not have significant changes in root length, which along with the less severe root hair phenotype, may potentially be

due to the redundancy of the EXTs. *EXT* genes are suggested to be highly redundant as despite the Arabidopsis genome encoding 59 *EXTs*, only a mutation in EXT3 causes a near-lethal phenotype, and it has been shown the double or triple mutations in EXTs are usually required for a phenotypic effect (Sede et al. 2018; Zhao et al. 2018; Cannon et al. 2008). However, plants with the HYP synthesis inhibitor, DHP, which has been shown to reduce global HYP-rich glycoproteins (Xu et al. 2011; X. Zhang et al. 2014) had severe bulbous root hair morphology and decreased primary root length (Fig 3.6, B+F). Therefore, we propose that NMM binds to the G4s formed in the SP4 motifs to cause a downregulation in *EXT* expression. Loss of *EXT* expression causes cell wall defects causing the cell to be unable to extend and withstand turgor pressure, leading to bulbous root hair and decreased growth. The severe root phenotypes seen within NMM treated plants could cause a defective root network meaning that the plant cannot efficiently take up nutrients and water, which could cause the overall decrease in growth, seen in NMM treated plants. Further studies could be carried out to understand whether the exact cause of the overall reduction in growth is due to decreased nutrient uptake through the reduced root system or due to cell wall defects by analysing the cell wall composition in roots and shoots.

In the literature, the proposed function of G4s within the coding region of a gene, on the template strand is to affect transcription through inhibiting PolII processing of the gene. G4s have been shown to stall both RNA and DNA polymerases and can inhibit transcription *in vitro* (Han, Hurley and Salazar, 1999; Siddiqui-Jain *et al.*, 2002; Cogoi and Xodo, 2006; Dexheimer, Sun and Hurley, 2006; Chambers *et al.*, 2015; Caliceti *et al.*, 2016; Kwok *et al.*, 2016) and interestingly our analysis of PolII ChiP-Chip data

(Chodavarapu et al., 2010) suggests *EXTs* have decreased PolII occupancy within the gene body which also correlates with the positioning of PG4s (Fig 4.1, A). This is also supported by our qPCR data that shows that the 3' region of EXT genes has lower expression in untreated plants than the 5' (Fig 4.1, D). We propose that further stalling of PolII may be responsible for the decrease in *EXT* expression found in NMM treated plants, which is supported by the 3' region of some EXTs being more affected by NMM than the 5' region. However, this was not the case within all EXTs (Fig 4.1, C).

PolII stalling cannot fully explain the downregulation of EXTs by NMM treatment as our results show that NMM downregulates the 5' region of EXTs. This finding led to the hypothesis that NMM was affecting a transcription factor or the promoter of EXTs in addition to stalling PolII. Therefore, we investigated whether NMM had a direct effect on EXTs using a translation inhibitor, CHX, to prevent *de novo* protein synthesis. NMM was found to downregulate EXTs in the presence of CHX, indicating a direct interaction (Fig 4.2, A). However, dual treatment experiments can have complications as it is uncertain how the treatments may interact with each other. Therefore, we also investigated the effect of NMM in a plant line with a mutation in the key EXT regulator RSL4 (Datta et al., 2015; Feng et al., 2017; Hwang et al., 2017; Shibata et al., 2018; Yi et al., 2010). We found that NMM can affect EXT expression without RSL4 (Fig 4.2, B) suggesting NMM does not downregulate EXTs specifically through RSL4, though we cannot discount a role for other upstream regulators.

To test whether NMM regulates EXTs via promoter or gene body effects, we also used the pLRX1::GUS line (Baumberger et al., 2001). This reporter lacks any EXT coding sequence, and the hypothesis was that any impact on GUS expression would, therefore,

be more likely to be promoter driven. Indeed, we found that NMM did significantly downregulate GUS expression suggesting that NMM can cause downregulation through the EXT promotor. However, the validity of this conclusion was complicated by the fact that on later analysis, we did identify that the GUS gene contains PG4s within the coding and template sequence, which could also explain the downregulation by NMM. Further evidence is required to understand the role of the promoter within NMM downregulation, which could be done using an alternative reporter gene to GUS that does not contain any PG4s, or analysis of expression using an alternative promoter that does not contain G4s to drive EXT expression. Understanding how exactly G4s affects EXT gene expression could allow us to manipulate this pathway not only to enable better plant growth, but this mechanism could also be used in human cells to regulate gene expression, it has just not been identified yet as two-quartet G4s are often disregarded.

If G4s form *in vivo* without the presence of G4 ligands, then the cell is likely to have a mechanism to prevent G4 formation when they are not beneficial. The literature suggests that the primary candidates for G4 regulation are helicases, of which several have been shown to bind and unwind G4s, and affect transcription through this action (Davies et al., 2007; Kruisselbrink et al., 2008; Paeschke et al., 2011; Piazza et al., 2015; Ribeyre et al., 2009; Rodriguez et al., 2012b; Singleton et al., 2007; Sun et al., 1998a; Van Wietmarschen and Lansdorp, 2016; Wu and Spies, 2016; WU, 2007). We identify using an EMSA that the EXT G4 can bind nuclear proteins (Fig 4.3, A), though this needs further investigation to demonstrate that this is G4 dependent binding. However, this method in combination with mass-spectrometry could be used to identify G4 binding

factors, which could lead on to determining whether such factors regulate G4 formation or facilitate the role of the G4 in a biological process. Additionally, we investigated some candidate helicases which are homologues of helicases previously shown to act upon G4s in humans (Davies et al., 2007; Gray et al., 2014a; Johnson et al., 2010b; Liu, 2003; Mohaghegh et al., 2001; Popuri et al., 2008; Sun et al., 1998a). This analysis revealed a potential candidate helicase in RecQ4a which was found to have shorter roots, lower expression of *LRX1* in untreated mutant plants and be less affected by NMM (Fig 4.3, B-D), suggesting that this mutant may be defective in resolving G4s forming under physiological conditions. However, further binding studies and mutant analysis is required to confirm this finding. Also, the decrease in expression in untreated *recq4a* mutants was not as severe as NMM treated plants, which may suggest the involvement of other helicases or proteins in regulating the EXTs G4 *in vivo*.

Finally, EXTs have been identified as being highly spliced, with *EXT9* having the most splice isoforms of any *Arabidopsis* gene (Li et al., 2016b). It has been shown that changes in PolII speed can affect the detection and choice of splice sites within a gene (Ameur et al., 2011; Girard et al., 2012; Khodor et al., 2011; Oesterreich et al., 2010; Tilgner et al., 2012; Windhager et al., 2012). As we propose that G4s cause stalling of PolII this led to the hypothesis that the stalling of PolII may also be causative of the excessive splicing seen within EXTs. RT-PCR and Sanger sequencing results suggest that LRX1 and EXT9 are alternatively spliced (Fig 5.1, A-C). Primers spanning the 5' LRR domain and the 3' EXT domain of LRX1, reveal that the majority of splicing occurs in the G4 dense EXT domain, supporting our hypothesis that G4s promote splicing, potentially through PolII splicing. However, there is the potential that the EXT domain encodes a

splice site, although no canonical splice site was identified. Additionally, NMM treatment of plants appears to alter the splicing of LRX1, decreasing the number of splice isoforms (Fig 5.2). The effect of NMM on splicing may potentially be due to all of the G4s being stabilised in NMM treated plants, whereas in untreated plants the G4s forming continually change resulting in a greater variety of splice variants. There is also the question of what the biological relevance of different EXT splice forms might be. The EXTs are thought to crosslink within the cell wall to strengthen the structure. Therefore, the alternative splicing of EXTs, giving rise to proteins of varying length, may be used within plants to allow for great molecular diversity, to alter the properties of the cell wall. Analysis of transgenic plants expressing truncated LRX1 lines, which would be similar to some of the LRX1 splice variants, supports their functionality as some of the truncated forms (*mycLRX1ΔE14)* can complement the *lrx1* mutant to allow normal root hair development (Fig 5.2 A-D) (Baumberger et al., 2001). In addition, G4s have only previously been implicated in splicing through their formation in RNA, hence this research suggests a novel function of G4s in the template strand of DNA. However, the splice site sequence of the splice variants could not be confirmed as a canonical splice site and there for it cannot be ruled out that these variant may be caused by an artefact or reverse transcription of the mRNA.

6.2 Future work

6.2.1 Short term questions

Are the proposed mechanisms of EXT downregulation G4 dependant?

We demonstrate that both NMM and Berberine cause EXT downregulation, reduced growth and root hair morphology. Changes in PolII occupancy and splicing was

observed in untreated plants, which we hypothesise is due to G4s forming physiologically in the absence of ligands. This could be confirmed by designing several oligos of the most common EXT motif with mutated loop sequences and analyse them by CD spec to try and identify loop sequences that prevent G4 formation. Transgenic plant lines could then be made using EXT mutant lines and inserting the mutated EXT sequence that no longer forms G4s. These lines could then be investigated by PolII-ChIP qPCR and RT-PCR to investigate whether PolII is stalled and whether splicing still occurs. Additionally, PolII ChIP-Seq could be carried out on G4 ligand treated plants to investigate whether G4 stabilisation alters PolII occupancy and further understand the EXT downregulation caused by G4 ligands. Finally, northern blots could be used to obtain a better picture of how much splicing occurs within the EXTs and how G4 ligands affect this.

How G4 ligands affect EXT promoters and transcription factors?

The *plrx1::GUS* experiment may suggests that G4 ligands may downregulate EXTs by acting through the promoter. We also find that several known EXT transcription factors are downregulated by NMM. Therefore, the effect of G4 ligands on EXTs promoters and transcription factors should be further studied. This could be done using a different EXT promotor construct, attached to a gene without any PG4s. Also, EXT constructs could be made that are transcribed under an alternative promoter such as the CamV 35S promoter to investigate whether NMM can still downregulate the gene in the absence of the promoter. Additionally, the effect of G4 ligands on the EXT transcription factors should be further studied using other mutant lines to investigate whether they are required for NMM downregulation of EXTs. Plus, NMM

downregulation of other transcription factors should also be tested within these mutant lines and bioinformatic analysis could also be used to find the enrichment and location of PG4s to build a hypothesis as to how NMM causes a downregulation in all the EXT transcription factor tested.

What mechanisms do Arabidopsis have to overcome G4s?

Extensive studies have identified that most organisms have proteins that regulate G4 formation, whether to allow the G4 to perform a particular function or alternatively to prevent them from being problematic. The most common G4 binding proteins identified are helicases. We propose RECQ4A as a potential regulator of EXT G4s based on our mutant line analysis; however, the evidence is quite weak and further analysis of RECQ4a could be carried out by using NMR or X-ray crystallography to find whether RECQ4a binds to the EXT motif, additionally RECQ4a ChIP-qPCR could be used to identify whether there is an enrichment of RECQ4a and EXT genes. We identified that the EXT motif binds nuclear proteins suggesting that Arabidopsis does have G4 binding proteins. Mass spec could be used to aid the identification of the nuclear proteins found by the EMSA the allowing binding assays and ChIP with the identified protein.

6.2.2 Long term questions

Can G4s be targeted to produce crops with better yields?

G4 stabilising ligands cause a decrease in the overall growth of Arabidopsis, which we hypothesise is due to EXT downregulation. However, if the limiting factor of growth, the G4s, was removed from the EXT genes then this could potentially increase plant growth. Analysis of the conservation of EXTs, other HRGPs and PG4s within crop species alongside looking at how G4 ligands affect their growth would provide an understanding of how well the proposed mechanisms of G4-dependant EXT expression

would translate into crop species. Then utilising loop sequences that prevent G4 formation or potentially using overexpression of helicases that have been identified to unwind the EXT G4s, the effect of growth could be investigated in the absence of G4 formation, within Arabidopsis and crop species.

Are the EXT G4s effected by drought and can this be utilised to make drought-resistant crops?

It has been identified that there is an enrichment of G4s within drought-responsive genes within Arabidopsis (Mullen et al., 2010, 2012). The effect of drought on G4s is particularly interesting as in drought conditions the cation concentration within cells increases, and as cations stabilise G4s, you would predict there to be more forming under these conditions. Therefore, you may predict that if increased cation concentrations in drought cause G4s to become stabilised within the EXTs, then we would see EXT downregulation, causing decreased growth. This could be investigated by comparing EXT expression in drought conditions to non-droughted plants and look for growth difference and root hair morphology. Then also compare a plant with EXTs that can no longer form G4s (altered loop sequences) to see whether the response seen in drought conditions is affected. This could also be investigated in crop species to potentially produce crops with higher drought resistance.

6.3 Closing Remarks

Overall our research indicates a functional role of two tetrad G4s within the template strand of gene bodies and a novel G4-dependent growth mechanism in *Arabidopsis.* The exact mechanism may be complex and requires further study, although evidence suggests that G4s may cause stalling of PolII to result in decreased gene

expression and alternative splicing. Also, G4s may regulate *EXT* gene expression through a mechanism involving the promoters of *EXTs*, which is likely to be through a separate action to the PolII stalling. The stabilisation of G4s using ligands causes EXT downregulation through these two mechanisms, leading to decreased cell wall rigidity, causing reduced growth and bulbous root hairs.

Further research into this area has direct implications in creating pesticides, manipulating these mechanisms to create crops with greater growth rates, or improving root systems to allow more efficient growth as well as increasing understanding of G4s within transcriptional process within all organisms.

7. Bibliography

Affuso, F., Mercurio, V., Fazio, V., and Fazio, S. (2010). Cardiovascular and metabolic effects of Berberine. World J. Cardiol. *2*, 71–77.

Allen, A.K., and Neuberger, A. (1973). The Purification and Properties of the Lectin from Potato Tubers, a Hydroxyproline-Containing Glycoprotein.

Ambrus, A., Chen, D., Dai, J., Jones, R.A., and Yang, D. (2005). Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization. Biochemistry *44*, 2048–2058.

Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A., and Yang, D. (2006). Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic Acids Res. *34*, 2723–2735.

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., and Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat. Struct. Mol. Biol. *18*, 1435–1440.

Andorf, C.M., Kopylov, M., Dobbs, D., Koch, K.E., Stroupe, M.E., Lawrence, C.J., and Bass, H.W. (2014). G-Quadruplex (G4) Motifs in the Maize (Zea mays L.) Genome Are Enriched at Specific Locations in Thousands of Genes Coupled to Energy Status, Hypoxia, Low Sugar, and Nutrient Deprivation. J. Genet. Genomics *41*, 627–647.

Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., and Kurreck, J. (2008). Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. RNA *14*, 1290–1296.

Averyhart-Fullard, V., Datta, K., and Marcus, A. (1988). A hydroxyproline-rich protein in the soybean cell wall. Proc. Natl. Acad. Sci. U. S. A. *85*, 1082–1085.

Ayrapetov, M.K., Gursoy-Yuzugullu, O., Xu, C., Xu, Y., and Price, B.D. (2014). DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. Proc. Natl. Acad. Sci. *111*, 9169–9174.

Bang, I. (1910). Untersuchungen über die Guanylsäure. Biochem. Z. *26*, 293–231.

Baumberger, N., Ringli, C., and Keller, B. (2001). The chimeric leucine-rich repeat/extensin cell wall protein LRX1 is required for root hair morphogenesis in Arabidopsis thaliana. Genes Dev. *15*, 1128–1139.

Baumberger, N., Steiner, M., Ryser, U., Keller, B., and Ringli, C. (2003). Synergistic interaction of the two paralogous Arabidopsis genes LRX1 and LRX2 in cell wall formation during root hair development. Plant J. *35*, 71–81.

Bazzicalupi, C., Ferraroni, M., Bilia, A.R., Scheggi, F., and Gratteri, P. (2013). The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. Nucleic Acids Res. *41*, 632–638.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., and Lemaitre, J.-

M. (2012). Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat. Struct. Mol. Biol. *19*, 837–844.

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

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet. *13*, 770–780.

Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., and Toole, J.J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. Nature *355*, 564–566.

Boise, L.H., González-García, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nuñez, G., and Thompson, C.B. (1993). bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell *74*, 597–608.

Borassi, C., Sede, A.R., Mecchia, M.A., Salgado Salter, J.D., Marzol, E., Muschietti, J.P., and Estevez, J.M. (2015). An update on cell surface proteins containing extensin-motifs.

Brooks, T.A., and Hurley, L.H. (2010). Targeting MYC Expression through G-Quadruplexes. Genes Cancer *1*, 641–649.

Brown, R. V, Danford, F.L., Gokhale, V., Hurley, L.H., and Brooks, T.A. (2011). Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. J. Biol. Chem. *286*, 41018–41027.

Bugaut, A., and Balasubramanian, S. (2012). 5'-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. *40*, 4727–4741.

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. *34*, 5402–5415.

Caceres, C., Wright, G., Gouyette, C., Parkinson, G., and Subirana, J.A. (2004). A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl+/Na+ ions. Nucleic Acids Res. *32*, 1097–1102.

Caliceti, C., Franco, P., Spinozzi, S., Roda, A., and Cicero, A.F.G. (2016). Berberine: New insights from pharmacological aspects to clinical evidences in the management of metabolic disorders. Curr. Med. Chem.

Campargue, C., Lafitte, C., Esquerré-Tugayé, M.-T., and Mazau, D. (1998). Analysis of Hydroxyproline and Hydroxyproline-Arabinosides of Plant Origin by High-Performance Anion-Exchange Chromatography–Pulsed Amperometric Detection. Anal. Biochem. *257*, 20–25.

Cannon, M.C., Terneus, K., Hall, Q., Tan, L., Wang, Y., Wegenhart, B.L., Chen, L., Lamport, D.T.A., Chen, Y., and Kieliszewski, M.J. (2008). Self-assembly of the plant cell wall requires an extensin scaffold. Proc. Natl. Acad. Sci. U. S. A. *105*, 2226–2231.

Capra, J.A., Paeschke, K., Singh, M., and Zakian, V.A. (2010). G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. PLoS Comput. Biol. *6*, e1000861.

Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P., and Balasubramanian, S. (2015).

High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. *33*, 877–881.

Chan Hong, J., Nagao, R.T., and Key, J.L. (1987). Characterization and Sequence Analysis of a Developmentally Regulated Putative Cell Wall Protein Gene Isolated from Soybean*.

Chen, F.-M. (1992). Sr2+ Facilitates Intermolecular G-Quadruplex Formation of Telomeric Sequences.

Chen, J., and Varner, J.E. (1985). Isolation and characterization of cDNA clones for carrot extensin and a proline-rich 33-kDa protein. Proc. Natl. Acad. Sci. U. S. A. *82*, 4399–4403.

Cho, H., Cho, H.S., Nam, H., Jo, H., Yoon, J., Park, C., Dang, T.V.T., Kim, E., Jeong, J., Park, S., et al. (2018). Translational control of phloem development by RNA G-quadruplex–JULGI determines plant sink strength. Nat. Plants *4*, 376–390.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y. V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., et al. (2010). Relationship between nucleosome positioning and DNA methylation. Nature *466*, 388–392.

Clay, E.H.., and Gould, I.R. (2005). A combined QM and MM investigation into guanine quadruplexes. J. Mol. Graph. Model. *24*, 138–146.

Clynes, D., Higgs, D.R., and Gibbons, R.J. (2013). The chromatin remodeller ATRX: a repeat offender in human disease. Trends Biochem. Sci. *38*, 461–466.

Cocquet, J., Chong, A., Zhang, G., and Veitia, R.A. (2006). Reverse transcriptase template switching and false alternative transcripts. Genomics *88*, 127–131.

Cogoi, S., and Xodo, L.E. (2006). G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. *34*, 2536–2549.

Cogoi, S., Shchekotikhin, A.E., and Xodo, L.E. (2014). *HRAS* is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. Nucleic Acids Res. *42*, 8379–8388.

Collie, G.W., Haider, S.M., Neidle, S., and Parkinson, G.N. (2010). A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. Nucleic Acids Res. *38*, 5569–5580.

Compe, E., and Egly, J.-M. (2016). Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. Annu. Rev. Biochem. *85*, 265–290.

Conlon, E.G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N.A. 8, and Manley, J.L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions 4 and sequesters hnRNP H to disrupt splicing in ALS brains.

Datta, S., Prescott, H., and Dolan, L. (2015). Intensity of a pulse of RSL4 transcription factor synthesis determines Arabidopsis root hair cell size. Nat. Plants *1*, 15138.

David, A.P., Margarit, E., Domizi, P., Banchio, C., Armas, P., and Calcaterra, N.B. (2016). G-quadruplexes as novel cis-elements controlling transcription during embryonic development. Nucleic Acids Res. *44*, 4163–4173.

Davies, S.L., North, P.S., and Hickson, I.D. (2007). Role for BLM in replication-fork restart and suppression

of origin firing after replicative stress. Nat. Struct. Mol. Biol. *14*, 677–679.
del Villar-Guerra, R., Trent, J.O., and Chaires, J.B. (2018). G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. Angew. Chemie Int. Ed. *57*, 7171–7175.
Deng, J., Xiong, Y., and Sundaralingam, M. (2001). X-ray analysis of an RNA tetraplex (UGGGGU)(4) with divalent Sr(2+) ions at subatomic resolution (0.61 A). Proc. Natl. Acad. Sci. U. S. A. *98*, 13665–13670.
Dexheimer, T.S., Sun, D., and Hurley, L.H. (2006). Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the *bcl-2* P1 Promoter. J. Am. Chem. Soc. *128*, 5404–5415.
Dingley, A.J., Peterson, R.D., Grzesiek, S., and Feigon, J. (2005). Characterization of the Cation and Temperature Dependence of DNA Quadruplex Hydrogen Bond Properties Using High-Resolution NMR. J. Am. Chem. Soc. *127*, 14466–14472.
Dominguez, C., Fisette, J.-F., Chabot, B., Frédéric, &, and Allain, H.-T. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi RRMs.
Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C.B., Proffitt, C., Trent, K., Whitten, J.P., et al. (2009). Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. Cancer Res. *69*, 7653–7661.
Du, Z., Zhao, Y., and Li, N. (2008). Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. Genome Res. *18*, 233–241.
Eddy, J., and Maizels, N. (2006). Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res. *34*, 3887–3896.
Eddy, S., Ketkar, A., Zafar, M.K., Maddukuri, L., Choi, J.-Y., and Eoff, R.L. (2014). Human Rev1 polymerase disrupts G-quadruplex DNA. Nucleic Acids Res. *42*, 3272–3285.
Escaja, N., Gelpí, J.L., Orozco, M., Rico, M., Pedroso, E., and González, C. (2003). Four-Stranded DNA Structure Stabilized by a Novel G:C:A:T Tetrad.
Feng, Y., Xu, P., Li, B., Li, P., Wen, X., An, F., Gong, Y., Xin, Y., Zhu, Z., Wang, Y., et al. (2017). Ethylene promotes root hair growth through coordinated EIN3/EIL1 and RHD6/RSL1 activity in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. *114*, 13834–13839.
Fernando, H., Sewitz, S., Darot, J., Tavaré, S., Huppert, J.L., and Balasubramanian, S. (2009). Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. Nucleic Acids Res. *37*, 6716–6722.
Fincher, G.B., Sawyer, W.H., and STONE, B.A. (1974). Chemical and Physical Properties of an Arabinogalactan-Peptide from Wheat Endosperm.
Franceschin, M., Rossetti, L., D'Ambrosio, A., Schirripa, S., Bianco, A., Ortaggi, G., Savino, M., Schultes, C., and Neidle, S. (2006). Natural and synthetic G-quadruplex interactive berberine derivatives. Bioorg. Med. Chem. Lett. *16*, 1707–1711.
Fry, M., and Loeb, L.A. (1994). The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. Proc. Natl. Acad. Sci. U. S. A. *91*, 4950–4954.

Garg, R., Aggarwal, J., Thakkar, B., Phan, A.T., Kuryavyi, V., Patel, D.J., Bochman, M.L., Paeschke, K., Zakian, V.A., Todd, A.K., et al. (2016a). Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. Sci. Rep. *6*, 28211.

Garg, R., Aggarwal, J., and Thakkar, B. (2016b). Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. Sci. Rep. *6*, 28211.

Gaynutdinov, T.I., Neumann, R.D., and Panyutin, I.G. (2008). Structural polymorphism of intramolecular quadruplex of human telomeric DNA: effect of cations, quadruplex-binding drugs and flanking sequences. Nucleic Acids Res. *36*, 4079–4087.

Gellert, M., Lipsett, M.N., and Davies, D.R. (1962). Helix formation by guanylic acid. Proc. Natl. Acad. Sci. U. S. A. *48*, 2013–2018.

German, J. (1993). Bloom syndrome: a mendelian prototype of somatic mutational disease. Medicine (Baltimore). *72*, 393–406.

Ghosh, S., Kar, A., Chowdhury, S., and Dasgupta, D. (2013). Ellipticine Binds to a Human Telomere Sequence: An Additional Mode of Action as a Putative Anticancer Agent? Biochemistry *52*, 4127–4137.

Giraldo, R., and Rhodes, D. (1994). The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. EMBO J. *13*, 2411–2420.

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nat. Commun. *3*, 994.

Goh, T., Joi, S., Mimura, T., and Fukaki, H. (2012). The establishment of asymmetry in Arabidopsis lateral root founder cells is regulated by LBD16/ASL18 and related LBD/ASL proteins. Development *139*, 883–893.

González, V., and Hurley, L.H. (2010). The C-Terminus of Nucleolin Promotes the Formation of the c- *MYC* G-Quadruplex and Inhibits c- *MYC* Promoter Activity. Biochemistry *49*, 9706–9714.

González, V., Guo, K., Hurley, L., and Sun, D. (2009). Identification and Characterization of Nucleolin as a c- *myc* G-quadruplex-binding Protein. J. Biol. Chem. *284*, 23622–23635.

Gray, L.T., Vallur, A.C., Eddy, J., and Maizels, N. (2014a). G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. Nat. Chem. Biol. *10*, 313–318.

Gray, L.T., Vallur, A.C., Eddy, J., and Maizels, N. (2014b). G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. Nat. Chem. Biol. *10*, 313–318.

Guédin, A., De Cian, A., Gros, J., Lacroix, L., and Mergny, J.-L. (2008). Sequence effects in single-base loops for quadruplexes. Biochimie *90*, 686–696.

Guédin, A., Gros, J., Alberti, P., and Mergny, J.-L. (2010). How long is too long? Effects of loop size on G-quadruplex stability. Nucleic Acids Res. *38*, 7858–7868.

Guilbaud, G., Murat, P., Recolin, B., Campbell, B.C., Maiter, A., Sale, J.E., and Balasubramanian, S. (2017). Local epigenetic reprogramming induced by G-quadruplex ligands. Nat. Chem. *9*, 1110–1117.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, P.G., Kim, M.-S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades

of disease. Nature *507*, 195–200.

Han, H., Hurley, L.H., and Salazar, M. (1999a). A DNA polymerase stop assay for G-quadruplex-interactive compounds. Nucleic Acids Res. *27*, 537–542.

Han, H., Hurley, L.H., and Salazar, M. (1999b). A DNA polymerase stop assay for G-quadruplex-interactive compounds. Nucleic Acids Res. *27*, 537–542.

Hänsel-Hertsch, R., Beraldi, D., Lensing, S. V, Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., et al. (2016). G-quadruplex structures mark human regulatory chromatin. Nat. Genet. *48*, 1267–1272.

Haq, I., Trent, J.O., Chowdhry, B.Z., and Terence C. Jenkins (1999). Intercalative G-Tetraplex Stabilization of Telomeric DNA by a Cationic Porphyrin1.

Hardin, C.C., Henderson, E., Watson, T., and Prosser11, J.K. (1991). Monovalent Cation Induced Structural Transitions in Telomeric DNAs: G-DNA Folding Intermediates.

Hardin, C.C., Watson, T., Corregan, M., and Bailey, C. (1992). Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG)t.

Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. J. Am. Chem. Soc. *126*, 16405–16415.

Hazel, P., Parkinson, G.N., and Neidle, S. (2006). Predictive modelling of topology and loop variations in dimeric DNA quadruplex structures. Nucleic Acids Res. *34*, 2117–2127.

He, Q., Mei, D., Sha, S., Fan, S., Wang, L., and Dong, M. (2016). ERK-dependent mTOR pathway is involved in berberine-induced autophagy in hepatic steatosis. J. Mol. Endocrinol.

Hegyi, H. (2015). Enhancer-promoter interaction facilitated by transiently forming G-quadruplexes. Sci. Rep. *5*, 9165.

Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I., and Blackburn, E.H. (1987). Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. Cell *51*, 899–908.

Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.-S., and Johnson, F.B. (2008). Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. Nucleic Acids Res. *36*, 144–156.

Houseley, J., and Tollervey, D. (2010). Apparent Non-Canonical Trans-Splicing Is Generated by Reverse Transcriptase In Vitro. PLoS One *5*, e12271.

Howell, R.M., Woodford, K.J., Weitzmann, M.N., and Usdin, K. (1996). The chicken beta-globin gene promoter forms a novel tetrahelical structure. J. Biol. Chem. *271*, 5208–5214.

Huppert, J.L. (2010). Structure, location and interactions of G-quadruplexes. FEBS J. *277*, 3452–3458.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res. *33*, 2908–2916.

Huppert, J.L., and Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. *35*, 406–413.

Hwang, Y., Choi, H.-S., Cho, H.-M., and Cho, H.-T. (2017). Tracheophytes Contain Conserved Orthologs of a Basic Helix-Loop-Helix Transcription Factor That Modulate ROOT HAIR SPECIFIC Genes. Plant Cell *29*, 39–53.

Ito, J., Fukaki, H., Onoda, M., Li, L., Li, C., Tasaka, M., and Furutani, M. (2016). Auxin-dependent compositional change in Mediator in ARF7- and ARF19-mediated transcription. Proc. Natl. Acad. Sci. U. S. A. *113*, 6562–6567.

de Jimenez, E.S., Aguilar, R., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (1984). Protein Synthesis Patterns : Relevance of Old and New Messenger RNA in Germinating Maize Embryos. PLANT Physiol. *75*, 231–234.

Johnson, J.E., Cao, K., Ryvkin, P., Wang, L.-S., and Johnson, F.B. (2010a). Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. Nucleic Acids Res. *38*, 1114–1122.

Johnson, J.E., Cao, K., Ryvkin, P., Wang, L.-S., and Johnson, F.B. (2010b). Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. Nucleic Acids Res. *38*, 1114–1122.

Johnson, K.L., Cassin, A.M., Lonsdale, A., Bacic, A., Doblin, M.S., and Schultz, C.J. (2017). Pipeline to Identify Hydroxyproline-Rich Glycoproteins. Plant Physiol. *174*, 886–903.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 167–177.

Juranek, S.A., and Paeschke, K. (2012). Cell cycle regulation of G-quadruplex DNA structures at telomeres. Curr. Pharm. Des. *18*, 1867–1872.

Kankia, B.I., and Marky, L.A. (2001). Folding of the Thrombin Aptamer into a G-Quadruplex with Sr 2+ : Stability, Heat, and Hydration.

Kejnovsky, E., and Lexa, M. (2014). Quadruplex-forming DNA sequences spread by retrotransposons may serve as genome regulators. Mob. Genet. Elements *4*, e28084.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.-H.A., Marr, M.T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. *25*, 2502–2512.

Kim, M.-Y., Duan, W., Gleason-Guzman, M., and Laurence H. Hurley (2003). Design, Synthesis, and Biological Evaluation of a Series of Fluoroquinoanthroxazines with Contrasting Dual Mechanisms of Action against Topoisomerase II and G-Quadruplexes.

Kim, W.-C., Ko, J.-H., Kim, J.-Y., Kim, J., Bae, H.-J., and Han, K.-H. (2013). MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in Arabidopsis. Plant J. *73*, 26–36.

Kim, W.S., Lee, Y.S., Cha, S.H., Jeong, H.W., Choe, S.S., Lee, M.-R., Oh, G.T., Park, H.-S., Lee, K.-U., Lane, M.D., et al. (2009). Berberine improves lipid dysregulation in obesity by controlling central and peripheral AMPK activity. Am. J. Physiol. - Endocrinol. Metab.

De Koning, L., Corpet, A., Haber, J.E., and Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. Nat. Struct. Mol. Biol. *14*, 997–1007.

Kopylov, M., Bass, H.W., and Stroupe, M.E. (2015). The Maize (Zea mays L.) Nucleoside Diphosphate Kinase1 (ZmNDPK1) Gene Encodes a Human NM23-H2 Homologue That Binds and Stabilizes G-Quadruplex DNA.

Krishnan-Ghosh, Y., Liu, D., and Balasubramanian, S. (2004). Formation of an Interlocked Quadruplex Dimer by d(GGGT).

Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D.B., Cuppen, E., and Tijsterman, M. (2008). Mutagenic Capacity of Endogenous G4 DNA Underlies Genome Instability in FANCJ-Defective C. elegans. Curr. Biol. *18*, 900–905.

Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. Nat. Chem. Biol. *3*, 218–221.

Kwok, C.K., Ding, Y., Shahid, S., Assmann, S.M., and Bevilacqua, P.C. (2015). A stable RNA G-quadruplex within the 5′-UTR of Arabidopsis thaliana ATR mRNA inhibits translation. Biochem. J. *467*, 91–102.

Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., and Balasubramanian, S. (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat. Methods *13*, 841–844.

Lam, E.Y.N., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013). G-quadruplex structures are stable and detectable in human genomic DNA. Nat. Commun. *4*, 1796.

Lamport, D.T.A. (1963). Oxygen Fixation into Hydroxyproline of Plant Cell Wall Protein.

Lamport, D.T.A., Kieliszewski, M.J., Chen, Y., and Cannon, M.C. (2011). Update on the Extensin Superfamily Role of the Extensin Superfamily in Primary Cell Wall Architecture 1. PLANT Physiol. *156*, 11–19.

LAMPORT, D.T.A., and NORTHCOTE, D.H. (1960). Hydroxyproline in Primary Cell Walls of Higher Plants. Nature *188*, 665–666.

Law, M.J., Lower, K.M., Voon, H.P.J., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A., et al. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. Cell *143*, 367–378.

Lecerof, D., Fodje, M., Hansson, A., Hansson, M., and Al-Karadaghi, S. (2000). Structural and mechanistic basis of porphyrin metallation by ferrochelatase. J. Mol. Biol. *297*, 221–232.

Levy, M.A., Kernohan, K.D., Jiang, Y., and Bérubé, N.G. (2015). ATRX promotes gene expression by facilitating transcriptional elongation through guanine-rich coding regions. Hum. Mol. Genet. *24*, 1824–1835.

Lexa, M., Kejnovsky, E., Steflova, P., Konvalinova, H., Vorlickova, M., and Vyskot, B. (2014). Quadruplex-forming sequences occupy discrete regions inside plant LTR retrotransposons. Nucleic Acids Res. *42*, 968–978.

Li, Q., Xiang, J.-F., Yang, Q.-F., Sun, H.-X., Guan, A.-J., and Tang, Y.-L. (2013). G4LDB: a database for discovering and studying G-quadruplex ligands. Nucleic Acids Res. *41*, D1115–D1123.

Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N. (2016a). High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Dev. Cell *39*, 508–522.

Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N. (2016b). High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Dev. Cell *39*, 508–522.

Li, Z.-Q., Liao, T.-C., Dong, C., Yang, J.-W., Chen, X.-J., Liu, L., Luo, Y., Liang, Y.-Y., Chen, W.-H., and Zhou, C.-Q. (2017). Specifically targeting mixed-type dimeric G-quadruplexes using berberine dimers. Org. Biomol. Chem. *15*, 10221–10229.

Liu, Z. (2003). Arabidopsis UVH6, a Homolog of Human XPD and Yeast RAD3 DNA Repair Genes, Functions in DNA Repair and Is Essential for Plant Growth. PLANT Physiol. *132*, 1405–1414.

Liu, X., Ishizuka, T., Bao, H.-L., Wada, K., Takeda, Y., Iida, K., Nagasawa, K., Yang, D., and Xu, Y. (2017). Structure-Dependent Binding of hnRNPA1 to Telomere RNA. J. Am. Chem. Soc. *139*, 7533–7539.

Liu, Y., Zhang, L., Song, H., and Ji, G. (2013). Update on berberine in nonalcoholic fatty liver disease. Evidence-Based Complement. Altern. Med.

London, T.B.C., Barber, L.J., Mosedale, G., Kelly, G.P., Balasubramanian, S., Hickson, I.D., Boulton, S.J., and Hiom, K. (2008). FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. J. Biol. Chem. *283*, 36132–36139.

Makarov, V.L., Hirose, Y., and Langmore, J.P. (1997). Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. Cell *88*, 657–666.

Marchand, A., Strzelecka, D., and Gabelica, V. (2016). Selective and Cooperative Ligand Binding to Antiparallel Human Telomeric DNA G-Quadruplexes. Chem. - A Eur. J. *22*.

Marušič, M., and Plavec, J. (2015). The Effect of DNA Sequence Directionality on G-Quadruplex Folding. Angew. Chem. Int. Ed. Engl. *54*, 11716–11719.

Marzol, E., Borassi, C., Bringas, M., Sede, A., Rodríguez Garcia, D.R., Capece, L., and Estevez, J.M. (2018). Filling the Gaps to Solve the Extensin Puzzle. Mol. Plant *11*, 645–658.

Mata, M. de la, Lafaille, C., and Kornblihtt, A.R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. RNA *16*, 904–912.

Matsugami, A., Ouhashi, K., Kanagawa, M., Liu, H., Kanagawa, S., Uesugi, S., and Katahira, M. (2001). An intramolecular quadruplex of (GGA)4 triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. J. Mol. Biol. *313*, 255–269.

Mauzerall, D. (1977). Porphyrins, Chlorophyll, and Photosynthesis. In Photosynthesis I, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 117–124.

Merkina, E.E., and Fox, K.R. (2005). Kinetic Stability of Intermolecular DNA Quadruplexes. Biophys. J. *89*, 365–373.

Mohaghegh, P., Karow, J.K., Brosh, R.M., Bohr, V.A., and Hickson, I.D. (2001). The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. Nucleic Acids Res. *29*, 2843–2849.

Monchaud, D., and Teulade-Fichou, M.-P. (2008). A hitchhiker's guide to G-quadruplex ligands. Org. Biomol. Chem. *6*, 627–636.

Moore, M.J.B., Schultes, C.M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanious, F.A., Wilson, W.D., and

Stephen Neidle (2005). Trisubstituted Acridines as G-quadruplex Telomere Targeting Agents. Effects of Extensions of the 3,6- and 9-Side Chains on Quadruplex Binding, Telomerase Activity, and Cell Proliferation.

Morris, M.J., Negishi, Y., Pazsint, C., Schonhoft, J.D., and Basu, S. (2010). An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. J. Am. Chem. Soc. *132*, 17831–17839.

van Mourik, T., and Dingley, A.J. (2005). Characterization of the Monovalent Ion Position and Hydrogen-Bond Network in Guanine Quartets by DFT Calculations of NMR Parameters. Chem. - A Eur. J. *11*, 6064–6079.

Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L., and Bryan, T.M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. Nat. Commun. *6*, 7643.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., and Wu, J.R. (1988). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc. Natl. Acad. Sci. U. S. A. *85*, 6622–6626.

Mukundan, V.T., and Phan, A.T. (2013). Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. J. Am. Chem. Soc. *135*, 5017–5028.

Mullen, M.A., Olson, K.J., Dallaire, P., Major, F., Assmann, S.M., and Bevilacqua, P.C. (2010). RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles. Nucleic Acids Res. *38*, 8149–8163.

Mullen, M.A., Assmann, S.M., and Bevilacqua, P.C. (2012). Toward a Digital Gene Response: RNA G-Quadruplexes with Fewer Quartets Fold with Higher Cooperativity. J. Am. Chem. Soc. *134*, 812–815.

Müller, S., Kumari, S., Rodriguez, R., and Balasubramanian, S. (2010). Small-molecule-mediated G-quadruplex isolation from human cells. Nat. Chem. *2*, 1095–1098.

Nagesh, N., Buscaglia, R., Dettler, J.M., and Lewis, E.A. (2010). Studies on the site and mode of TMPyP4 interactions with Bcl-2 promoter sequence G-Quadruplexes. Biophys. J. *98*, 2628–2633.

Nakagawa, A., Takahashi, H., Kojima, S., Sato, N., Ohga, K., Cha, B.Y., Woo, J.-T., Nagai, K., Horiguchi, G., Tsukaya, H., et al. (2012). Berberine enhances defects in the establishment of leaf polarity in asymmetric leaves1 and asymmetric leaves2 of Arabidopsis thaliana. Plant Mol. Biol. *79*, 569–581.

Nakken, S., Rognes, T., and Hovig, E. (2009). The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. Nucleic Acids Res. *37*, 5749–5756.

Nicoludis, J.., Miller, S.., Jeffrey, P.., Barrett, S.., Rablen, P.., Lawton, T.., and Yatsunyk, L.A. (2012). Optimized End-Stacking Provides Specificity of N-Methyl Mesoporphyrin IX for Human Telomeric G-Quadruplex DNA. J.Am.Chem.Soc. *134*, 20446–20456.

Oesterreich, F.C., Preibisch, S., and Neugebauer, K.M. (2010). Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. Mol. Cell *40*, 571–581.

Oganesian, L., Graham, M.E., Robinson, P.J., and Bryan, T.M. (2007). Telomerase recognizes G-quadruplex and linear DNA as distinct substrates. Biochemistry *46*, 11279–11290.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H.J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. Nat. Struct. Mol. Biol. *12*, 847–854.
Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D., and Lipps, H.J. (2008). Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. Nat. Struct. Mol. Biol. *15*, 598–604.
Paeschke, K., Capra, J.A., and Zakian, V.A. (2011). DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. Cell *145*, 678–691.
Parker, M. (2018). G-Quadruplexes and Gene Expression in Arabidopsis thaliana.
Parkinson, G.N., Lee, M.P.H., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. Nature *417*, 876–880.
Patel, P.K., and Hosur, R. V (1999). NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by Saccharomyces cerevisiae telomere repeats. Nucleic Acids Res. *27*, 2457–2464.
Peng, L., Kang, S., Yin, Z., Jia, R., Song, X., Li, L., Li, Z., Zou, Y., Liang, X., Li, L., et al. (2015). Antibacterial activity and mechanism of berberine against Streptococcus agalactiae. Int. J. Clin. Exp. Pathol. *8*, 5217–5223.
Petraccone, L., Erra, E., Duro, I., Esposito, V., Randazzo, A., Mayol, L., Mattia, C.A., Barone, G., and Giancola, C. (2005). Relative stability of quadruplexes containing different number of G-tetrads. Nucleosides. Nucleotides Nucleic Acids *24*, 757–760.
Phan, A.T., and Patel, D.J. (2003). Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics. J. Am. Chem. Soc. *125*, 15021–15027.
Phan, A.T., Kuryavyi, V., Gaw, H.Y., and Patel, D.J. (2005). Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. Nat. Chem. Biol. *1*, 167–173.
Piazza, A., Adrian, M., Samazan, F., Heddi, B., Hamon, F., Serero, A., Lopes, J., Teulade-Fichou, M.-P., Phan, A.T., and Nicolas, A. (2015). Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. EMBO J. *34*, 1718–1734.
Popuri, V., Bachrati, C.Z., Muzzolini, L., Mosedale, G., Costantini, S., Giacomini, E., Hickson, I.D., and Vindigni, A. (2008). The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. J. Biol. Chem. *283*, 17766–17776.
Postberg, J., Tsytlonok, M., Sparvoli, D., Rhodes, D., and Lipps, H.J. (2012). A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication. Gene *497*, 147–154.
Qi, X., Behrens, B.X., West, P.R., and Mort, A.J. (1995). Solubilization and partial characterization of extensin fragments from cell walls of cotton suspension cultures. Evidence for a covalent cross-link between extensin and pectin. Plant Physiol. *108*, 1691–1701.
Raya-González, J., Ortiz-Castro, R., Ruíz-Herrera, L.F., Kazan, K., and López-Bucio, J. (2014). PHYTOCHROME AND FLOWERING TIME1/MEDIATOR25 Regulates Lateral Root Formation via Auxin Signaling in Arabidopsis. Plant Physiol. *165*, 880–894.

Ren, J., and Chaires, J.B. (1999). Sequence and Structural Selectivity of Nucleic Acid Binding Ligands†.

Rhodes, D., and Lipps, H.J. (2015). G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. *43*, 8627–8637.

Ribeyre, C., Lopes, J., Boulé, J.-B., Piazza, A., Guédin, A., Zakian, V.A., Mergny, J.-L., and Nicolas, A. (2009). The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. PLoS Genet. *5*, e1000475.

Richards, E.J., and Ausubel, F.M. (1988). Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. Cell *53*, 127–136.

Ringli, C. (2010). The hydroxyproline-rich glycoprotein domain of the Arabidopsis LRX1 requires Tyr for function but not for insolubilization in the cell wall. Plant J. *63*, 662–669.

Rodriguez, R., Miller, K.M., Forment, J. V, Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012a). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. *8*, 301–310.

Rodriguez, R., Miller, K.M., Forment, J. V, Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012b). Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. *8*, 301–310.

Rouleau, S., Glouzon, J.-P.S., Brumwell, A., Bisaillon, M., and Perreault, J.-P. (2017). 3′ UTR G-quadruplexes regulate miRNA binding. RNA *23*, 1172–1179.

Rovnyak, D., Baldus, M., Wu, G., Hud, N. V., Juli Feigon, and Robert G. Griffin (2000). Localization of 23Na+ in a DNA Quadruplex by High-Field Solid-State NMR.

Samatanga, B., Dominguez, C., Jelesarov, I., and Allain, F.H.-T. (2013). The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. Nucleic Acids Res. *41*, 2505–2516.

Sarkies, P., Reams, C., Simpson, L.J., and Sale, J.E. (2010). Epigenetic Instability due to Defective Replication of Structured DNA. Mol. Cell *40*, 703–713.

Sarkies, P., Murat, P., Phillips, L.G., Patel, K.J., Balasubramanian, S., and Sale, J.E. (2012). FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. Nucleic Acids Res. *40*, 1485–1498.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J., and Plückthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proc. Natl. Acad. Sci. U. S. A. *98*, 8572–8577.

Schiavone, D., Guilbaud, G., Murat, P., Papadopoulou, C., Sarkies, P., Prioleau, M.-N., Balasubramanian, S., and Sale, J.E. (2014). Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. EMBO J. *33*, 2507–2520.

Schiavone, D., Jozwiakowski, S.K., Bailey, L.J., Sale, J.E., Doherty, A.J., Romanello, M., Guilbaud, G., and Guilliam, T.A. (2016). PrimPol Is Required for Replicative Tolerance of G Quadruplexes in Vertebrate Cells Molecular Cell PrimPol Is Required for Replicative Tolerance of G Quadruplexes in Vertebrate Cells. Mol. Cell *61*, 161–169.

Schultes, C.M., Guyen, B., Cuesta, J., and Neidle, S. (2004). Synthesis, biophysical and biological evaluation of 3,6-bis-amidoacridines with extended 9-anilino substituents as potent G-quadruplex-binding telomerase inhibitors. Bioorg. Med. Chem. Lett. *14*, 4347–4351.

Schultze, P., Hud, N. V, Smith, F.W., and Feigon, J. (1999). The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G(4)T(4)G(4)). Nucleic Acids Res. *27*, 3018–3028.

Schweizer, F., Fernández-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G., Lewsey, M.G., Ecker, J.R., Solano, R., and Reymond, P. (2013). Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. Plant Cell *25*, 3117–3132.

Seenisamy, J., Rezler, E.M., Powell, T.J., Tye, D., Gokhale, V., Joshi, C.S., Siddiqui-Jain, A., and Hurley, L.H. (2004). The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. J. Am. Chem. Soc. *126*, 8702–8709.

Sen, D., and Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature *334*, 364–366.

Shahid, R., Bugaut, A., and Balasubramanian, S. (2010). The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. Biochemistry *49*, 8300–8306.

Shanbhag, N.M., Rafalska-Metcalf, I.U., Balane-Bolivar, C., Janicki, S.M., and Greenberg, R.A. (2010). ATM-Dependent Chromatin Changes Silence Transcription In cis to DNA Double-Strand Breaks. Cell *141*, 970–981.

Shay, J.W., and Wright, W.E. (2011). Role of telomeres and telomerase in cancer. Semin. Cancer Biol. *21*, 349–353.

Shibata, M., Breuer, C., Kawamura, A., Clark, N.M., Rymen, B., Braidwood, L., Morohashi, K., Busch, W., Benfey, P.N., Sozzani, R., et al. (2018). GTL1 and DF1 regulate root hair growth through transcriptional repression of ROOT HAIR DEFECTIVE 6-LIKE 4 in Arabidopsis. Development *145*, dev159707.

Shin, J.-H., and Chekanova, J.A. (2014). Arabidopsis RRP6L1 and RRP6L2 Function in FLOWERING LOCUS C Silencing via Regulation of Antisense RNA Synthesis. PLoS Genet. *10*, e1004612.

Shoulders, M.D., and Raines, R.T. (2009). Collagen structure and stability. Annu. Rev. Biochem. *78*, 929–958.

Showalter, A.M., Keppler, B.D., Liu, X., Lichtenberg, J., and Welch, L.R. (2016). Bioinformatic Identification and Analysis of Hydroxyproline-Rich Glycoproteins in Populus trichocarpa. BMC Plant Biol. *16*, 229.

Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., and Hurley, L.H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc. Natl. Acad. Sci. U. S. A. *99*, 11593–11598.

Silva, M.W. da (2003). Association of DNA Quadruplexes through G:C:G:C Tetrads. Solution Structure of d(GCGGTGGAT).

Simonsson, T., Pecinka, P., and Kubista, M. (1998). DNA tetraplex formation in the control region of c-myc.

Nucleic Acids Res. *26*, 1167–1172.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu. Rev. Biochem. *76*, 23–50.

Šmarda, P., Bureš, P., Horová, L., Leitch, I.J., Mucina, L., Pacini, E., Tichý, L., Grulich, V., and Rotreklová, O. (2014). Ecological and evolutionary significance of genomic GC content diversity in monocots. Proc. Natl. Acad. Sci. U. S. A. *111*, E4096-102.

Smirnov, I., and Shafer, R.H. (2000). Effect of Loop Sequence and Size on DNA Aptamer Stability.

Spačková, N., Berger, I., and Sponer, J. (1999). Nanosecond Molecular Dynamics Simulations of Parallel and Antiparallel Guanine Quadruplex DNA Molecules.

Sun, D., and Hurley, L.H. (2010). Biochemical Techniques for the Characterization of G-Quadruplex Structures: EMSA, DMS Footprinting, and DNA Polymerase Stop Assay. In Methods in Molecular Biology (Clifton, N.J.), pp. 65–79.

Sun, H., Karow, J.K., Hickson, I.D., and Maizels, N. (1998a). The Bloom's syndrome helicase unwinds G4 DNA. J. Biol. Chem. *273*, 27587–27592.

Sun, H., Karow, J.K., Hickson, I.D., and Maizels, N. (1998b). The Bloom's syndrome helicase unwinds G4 DNA. J. Biol. Chem. *273*, 27587–27592.

Sundaravelpandian, K., Chandrika, N.N.P., and Schmidt, W. (2013). PFT1, a transcriptional Mediator complex subunit, controls root hair differentiation through reactive oxygen species (ROS) distribution in Arabidopsis. New Phytol. *197*, 151–161.

Sundquist, W.I., and Klug, A. (1989a). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature *342*, 825–829.

Sundquist, W.I., and Klug, A. (1989b). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature *342*, 825–829.

Takahashi, H., Nakagawa, A., Kojima, S., Takahashi, A., Cha, B.-Y., Woo, J.-T., Nagai, K., Machida, Y., and Machida, C. (2012). Discovery of novel rules for G-quadruplex-forming sequences in plants by using bioinformatics methods. J. Biosci. Bioeng. *114*, 570–575.

Takahashi, S., Brazier, J.A., and Sugimoto, N. (2017). Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase. Proc. Natl. Acad. Sci. U. S. A. *114*, 9605–9610.

Tian, Y., Cai, J., Gui, W., Nichols, R.G., Koo, I., Zhang, J., Anitha, M., and Patterson, A.D. (2018). Berberine Directly Impacts the Gut Microbiota to Promote Intestinal Farnesoid X Receptor Activation. Drug Metab. Dispos.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res. *22*, 1616–1625.

Tokan, V., Puterova, J., Lexa, M., and Kejnovsky, E. (2018). Quadruplex DNA in long terminal repeats in maize LTR retrotransposons inhibits the expression of a reporter gene in yeast. BMC Genomics *19*, 184.

Valton, A.-L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintome, C., Riou, J.-F., and Prioleau, M.-N. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J. *33*, 732–746.

Velasquez, S.M., Ricardi, M.M., Dorosz, J.G., Fernandez, P. V, Nadra, A.D., Pol-Fachin, L., Egelund, J., Gille, S., Harholt, J., Ciancia, M., et al. (2011a). O-Glycosylated Cell Wall Proteins Are Essential in Root Hair Growth. Science (80-.). *332*, 1401–1403.

Velasquez, S.M., Iusem, N.D., and Estevez, J.M. (2011b). Root hair sweet growth. Plant Signal. Behav. *6*, 1600–1602.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. *24*, 896–905.

Verma, A., Halder, K., Halder, R., Yadav, V.K., Rawal, P., Thakur, R.K., Mohd, F., Sharma, A., and Chowdhury, S. (2008). Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis -regulatory elements in human and related species. J. Med. Chem. *51*, 5641–5649.

Víglaský, V., Bauer, L., and Tlučková, K. (2010). Structural Features of Intra- and Intermolecular G-Quadruplexes Derived from Telomeric Repeats. Biochemistry *49*, 2110–2120.

Wang, Y., and Patel, D.J. (1992). Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. Biochemistry *31*, 8112–8119.

Webba da Silva, M. (2007). NMR methods for studying quadruplex nucleic acids. Methods *43*, 264–277.

Wei, S., Zhang, M., Yu, Y., Lan, X., Yao, F., Yan, X., Chen, L., and Hatch, G.M. (2016). Berberine Attenuates Development of the Hepatic Gluconeogenesis and Lipid Metabolism Disorder in Type 2 Diabetic Mice and in Palmitate-Incubated HepG2 Cells through Suppression of the HNF-4α miR122 Pathway. PLoS One.

Weldon, C., Eperon, I.C., and Dominguez, C. (2016). Do we know whether potential G-quadruplexes actually form in long functional RNA molecules? Biochem. Soc. Trans. *44*, 1761–1768.

Weldon, C., Dacanay, J.G., Gokhale, V., Boddupally, P.V.L., Behm-Ansmant, I., Burley, G.A., Branlant, C., Hurley, L.H., Dominguez, C., and Eperon, I.C. (2018a). Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. Nucleic Acids Res. *46*, 886–896.

Weldon, C., Dacanay, J.G., Gokhale, V., Boddupally, P.V.L., Behm-Ansmant, I., Burley, G.A., Branlant, C., Hurley, L.H., Dominguez, C., and Eperon, I.C. (2018b). Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. Nucleic Acids Res. *46*, 886–896.

Van Wietmarschen, N., and Lansdorp, P.M. (2016). Bromodeoxyuridine does not contribute to sister chromatid exchange events in normal or Bloom syndrome cells. Nucleic Acids Res. *44*, 6787–6793.

Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., et al. (2012). Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome Res. *22*, 2031–2042.

Wolfe, A.L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V.K., Sanghvi, V.R., Mavrakis, K.J., Jiang, M.,

Roderick, J.E., Van der Meulen, J., et al. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. Nature *513*, 65–70.

Woodford, K.J., Howell, R.M., and Usdin, K. (1994). A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. J. Biol. Chem. *269*, 27029–27035.

Worlinsky, J.L., and Basu, S. (2009). Detection of Quadruplex DNA by Luminescence Enhancement of Lanthanide Ions and Energy Transfer from Lanthanide Chelates. J. Phys. Chem. B *113*, 865–868.

Wu, C.G., and Spies, M. (2016). G-quadruplex recognition and remodeling by the FANCJ helicase. Nucleic Acids Res. *44*, 8742–8753.

Wu, G., Wong, A., Gan, Z., and Davis, J.T. (2003). Direct detection of potassium cations bound to G-quadruplex structures by solid-state 39K NMR at 19.6 T. J. Am. Chem. Soc. *125*, 7182–7183.

Wu, H.L., Hsu, C.Y., Liu, W.H., and Yung, B.Y.M. (1999). Berberine-induced apoptosis of human leukemia HL-60 cells is associated with down-regulation of nucleophosmin/B23 and telomerase activity. Int. J. Cancer *81*, 923–929.

WU, L. (2007). Role of the BLM helicase in replication fork management. DNA Repair (Amst). *6*, 936–944.

Xia, W., Liu, Z., Yang, Y., Xiao, Y., Mason, A.S., Zhao, S., and Ma, Z. (2014). Selection of reference genes for quantitative real-time PCR in *Cocos nucifera* during abiotic stress. Botany *92*, 179–186.

Xiao, X., Wang, Z., Jang, M., Nutiu, R., Wang, E.T., and Burge, C.B. (2009). Splice site strength–dependent activity and genetic buffering by poly-G runs. Nat. Struct. Mol. Biol. *16*, 1094–1100.

Xu, C., Takáč, T., Burbach, C., Menzel, D., and Šamaj, J. (2011). Developmental localization and the role of hydroxyproline rich glycoproteins during somatic embryogenesis of banana (Musa spp. AAA). BMC Plant Biol. *11*, 38.

Xu, P., Cai, X.-T., Wang, Y., Xing, L., Chen, Q., and Xiang, C.-B. (2014). HDG11 upregulates cell-wall-loosening protein genes to promote root elongation in Arabidopsis. J. Exp. Bot. *65*, 4285–4295.

Yi-Meen Chou, Ming Chih Lai, Tsai-Mian Hwang, and Chi Wi Ong (1999). Synthesis and antitumor activity of bithienyl-pyrimidine derivatives with electrostatic binding side chains. Bioorg. Med. Chem. Lett. *9*, 2643–2646.

Yi, K., Menand, B., Bell, E., and Dolan, L. (2010). A basic helix-loop-helix transcription factor controls cell growth and size in root hairs. Nat. Genet. *42*, 264–267.

Yin, J., Xing, H., and Ye, J. (2008). Efficacy of berberine in patients with type 2 diabetes mellitus. Metabolism. *57*, 712–717.

Yu, Z., Gaerig, V., Cui, Y., Kang, H., Gokhale, V., Zhao, Y., Hurley, L.H., and Mao, H. (2012). Tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. J. Am. Chem. Soc. *134*, 5157–5164.

Zahler, A.M., Williamson, J.R., Cech, T.R., and Prescott, D.M. (1991). Inhibition of telomerase by G-quartet DNA structures. Nature *350*, 718–720.

Zakian, V.A. (2012). Telomeres: the beginnings and ends of eukaryotic chromosomes. Exp. Cell Res. *318*, 1456–1460.

Zaug, A.J., Podell, E.R., and Cech, T.R. (2005). Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. Proc. Natl. Acad. Sci. U. S. A. *102*, 10864–10869.

Zhang, L., Tamura, K., Shin-ya, K., and Takahashi, H. (2006). The telomerase inhibitor telomestatin induces telomere shortening and cell death in Arabidopsis. Biochim. Biophys. Acta *1763*, 39–44.

Zhang, X., Ma, H., Qi, H., and Zhao, J. (2014). Roles of hydroxyproline-rich glycoproteins in the pollen tube and style cell growth of tobacco (Nicotiana tabacum L.). J. Plant Physiol. *171*, 1036–1045.

Zhang, Y.-J., Pan, H.-Y., and Gao, S.-J. (2001). Reverse Transcription Slippage over the mRNA Secondary Structure of the LIP1 Gene. Biotechniques *31*, 1286–1294.