G-Quadruplex Nucleic Acid Structures Influence *DUX4* Expression in Facioscapulohumeral Muscular Dystrophy

By Lukasz Ciszewski

School of Biological Science

Royal Holloway, University of London

A Thesis Presented for the Degree of Doctor of Philosophy

2018

Statement

I hereby declare that this thesis has been composed by myself, that it has not been accepted in any previous application for a degree, that the work of which it is a record has been done by myself, that any personal data have been processed in accordance with the provisions of the Data Protection Act 1998, and that all quotations have been distinguished by quotation marks and the sources of information specifically acknowledged.

Lukasz Ciszewski

July 2018

Abstract

Although Facioscapulohumeral Muscular Dystrophy (FSHD) was first described in 1885, an effective treatment is yet to be developed. FSHD is an autosomal dominant disorder characterized by distinct chromatin changes including DNA hypomethylation of the D4Z4 macrosatellite repeat array on a disease-permissive 4qA allele and aberrant expression of the D4Z4-embedded DUX4 retrogene in skeletal muscle. Compelling evidence indicates that this mutation-induced *DUX4* expression plays a key role in disease pathogenesis.

We initially performed bioinformatic studies which predicted the presence of novel nucleic acid regulatory elements, namely G-quadruplexes (GQs), in the *DUX4* genomic locus and transcript. GQ motifs were identified in transcriptional regulatory elements such as DUX4 myogenic enhancer and promoter regions, as well as near splice sites of *DUX4* transcript. The structural characteristics of these putative GQs were characterised and confirmed using circular dichroism and nuclear magnetic resonance spectroscopy. Using a reporter gene system and cell transfection, mutation of GQ sequences in the *DUX4* promoter led to decreased reporter gene expression indicating a role in transcription. In addition, when expression from the *DUX4* genomic sequence was driven by the CMV promoter (lacking GQs) expression was also downregulated when transfected cells were treated with the GQ specific, small-molecule drug, berberine. The downregulation of DUX4 mRNA expression by berberine treatment was also confirmed in FSHD patient muscle cell cultures. High affinity of berberine binding to the GQ sequences within the *DUX4* enhancer, promoter and transcript, was determined using UV, visible light and fluorescence spectroscopic techniques. Although the specific molecular mechanisms involved remain as yet to be fully unravelled, these data demonstrate for the first time that GQs are present in *DUX4* locus sequences and that

targeting them can reduce *DUX4* expression, and thus that their pharmacological modulation may provide a novel therapeutic strategy for the treatment of FSHD.

Acknowledgement

I would like to express my great appreciation to Prof. George Dickson for providing me with the opportunity to undertake this PhD. To be able to learn analytical thinking and how to approach scientific problems from him was a privilege, and undoubtedly, it had a profound effect on my development as a scientist. I thank Dr Linda Popplewell for providing her constant advice throughout my PhD. Her remarkable ability for creative thinking has greatly improved my work.

My deepest gratitude to everyone in the Dickson and Popplewell lab for creating a fun working environment and assisting me throughout my PhD.

Prof Mark Searle and his lab member from the University of Nottingham were instrumental in helping me with the biophysical aspects described in this thesis. I would like to express my thanks to them for providing me with new invaluable scientific experiences and showing me the best Fish and Chip shop in the UK!

Finally, I would like to thank my family and friends. My mum, Teresa, a 5'1'' lady with the heart of a lion. Your love and care has allowed me to grow as a man and look with courage into the future. You are the greatest mum in the infinite universe.

My sisters, Emilia and Justyna, you were there for me from the beginning. Like three musketeers we hustle through life. Thank you for always sticking up for your little brother.

My father, Tedeusz, as unlikely it might seem, you have shown me which directions to take in life, for which I am eternally grateful.

My true friends: Ryan Han, you were there for me when it mattered the most, I will owe you till the end of time (!); Golnoush Golshirazi, you always catch me when I fall, without your support (and your dog Bella =]) I would not make it through this PhD. There are no words to express my gratitude for you! Friendship above all!

To the vast number of people that are not mentioned here (so I can spare you my sentimental ramblings), but were there for me throughout my PhD journey, I would like to say thank you!

'If I have seen further than others, it is by standing upon the shoulders of giants.'

- **Isaac Newton**

Abbreviations

AAV- adeno-associated virus

AGO2-argonulate-2

AnP- Antarctic phosphatase

Bcl2-B-cell lymphoma 2

CD- circular dichroism

ChIP- chromatin immunoprecipitation

CK- cytosine kinase

CMV- cytomegalovirus

CRISPR/Cas9- Clustered Regularly Interspaced Short Palindromic Repeats/CRISP-associated protein 9

CT-calf thymus

D4P- DUX4 promoter

DMD- Duchenne muscular dystrophy

DMEM- Dulbecco's modified Eagle medium

DNA-deoxyribonucleic acid

DUX4- double homeobox 4

DUX4-fl- DUX4 full length

DUX4-s- DUX4 short

E1- exon1

EDTA- ethylenediaminetetraacetic acid

eGFP- enhanced green fluorescent protein

FANCJ- Fanconi anemia complementation group J

FMRP-Fragile X mental retardation protein

FSHD -Facioscapulohumeral dystrophy

GQ- guanine quadruplex

HIF1α-Hypoxia-inducible factor 1α

ICW- In-Cell Western

iPS- induced pluripotent steam cells

kDa- kilo Dalton

miRNA- microRNA

NHEIII- ease hypersensitive element III

NMD- nonsense mediated decay

NMR- nuclear magnetic resonance

OD- optical density

ORF -open reading frame

PFA- paraformaldehyde

PMO- phosphorodiamidate morpholino oligomer

QGRS-quadruplex-forming G-rich sequences

RD- rhabdomyosarcoma

RNA-ribonucleic acid

SEM- standard error of the mean

shRNA- short hairpin RNA

siRNA- small interfering RNA

SMA- Spinal muscular atrophy

SMCHD1- structural maintenance of chromosomes flexible hinge domain containing 1

SOC- super optimal broth with catabolite

SS1- splice site 1

TEBPβ- telomere ending protein

TF- transcription factor

UTR-untranslated region

UV-Vis- Ultraviolet–visible

VEGEF- Vascular endothelial growth factor

XPB(D)- xeroderma pigmentosum type B(D)

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Facioscapulohumeral muscular dystrophy: clinical description

1.1.1. *Disease discovery*

The first picture published of an affected Facioscapulohumeral (FSHD) patient was reported by Duchenne de Boulogne in his *Album de photographies pathologiques* in 1862 (Duchenne 1862). Later, in 1869 Duchenne released a more detailed description of the disease in *Archives of General Medicine* (Duchenne 1869), which is often regarded as the first reference of FSHD. However, the term Facioscapulohumeral was first devised in 1885, when Landouzy and Dejerine described cases of more advanced stages of the disease (Landouzy and Dejerine 1885). They highlight the progressive muscle atrophy of the face, shoulder girdle and upper arms, followed by weakening of abdominal muscles resulting in lumbar hyperlordosis, and of tibialis anterior muscles with steppage gait. The first modern clinical characterisation of FSHD was performed by Padberg in 1982 that involved 107 individuals from 19 families of which 73 subjects showed clinical hallmarks of FSHD. Padberg's studies give the first evidence for extensive clinical variability of patients with FSHD, even from the same family (Padberg 1982).

1.1.2. *Clinical features*

FSHD is a progressive disease that weakens and destroys skeletal muscles. The typical clinical manifestation at first involve facial weakness that progresses to shoulder girdle, humeral, truncal and anterior leg muscles.

Facial weakness is reported in up to 90% of patients, of which over 50% of cases show to have the facial muscle affected in an asymmetrical fashion- a trait found useful in differential diagnosis of FSHD (Figure 1.1.D) (Padberg 1982). Shoulder girdle weakens is reported as the first symptom by around 80% of patients (Padberg 1982) and it is a result of atrophy of the scapula fixator and pectoralis muscles (Figure 1.1). The scapula weakness is usually asymmetrical and locates to the right side of the body, independently of the handedness (Brouwer et al. 1993). Significant number of patients (up to 30%) do not develop further symptoms past the shoulder weakness. (Padberg 1982). In the next stage of the disease a foot extensor or pelvic girdle weakness is reported in 80% and 20% of cases, respectively (Padberg 1982). Usually, abdominal muscle weakness is reported at this stage that results in lumbar lordosis. Roughly 10% of all the patients and 20% of patients past their 50's will become wheelchair-bound. Extraocular, masticatory and cardiac muscles are not considered to be involved in FSHD and should be considered for other diagnoses (Padberg 1982; Laforet et al. 1998; A. J. van der Kooi et al. 2000). Serum creatine kinase (CK) level is typically less than five times the upper limit. Electromyography (EMG) and muscle histology examination show myopathic changes that are not FSHD specific (Padberg 2004; Dorobek et al. 2013).

Non-muscular disease manifestations, include: scoliosis that is present in around 30% of patients and might lead to respiratory problems; retinal vasculopathy found in approximately 60% of patients that leads to visual loss in less than 1% of cases; high-tone hearing loss which occurs prominently in early-onset patients (Padberg 1982; Padberg et al. 1995; Rogers et al. 2002).

FSHD is an autosomal dominant disease with an age dependent penetrance estimate of >95% at 20 years of age (Lemmers and van der Maarel 1993). The prevalence of the condition was calculated to be around 1 in 8000 persons (Deenen et al. 2014). However, the disease pattern of progression can vary as infantile cases have been reported (Bailey, Marzulo, and Hans 1986). In addition, a gender bias manifested by higher penetrance in males (95%) *vs.* females (65%) that carry the FSHD allele has also been found (Zatz et al. 1998). The underlying reasons why the FSHD does not follow a classical autosomal dominant Mendelian chronology is not clearly understood.

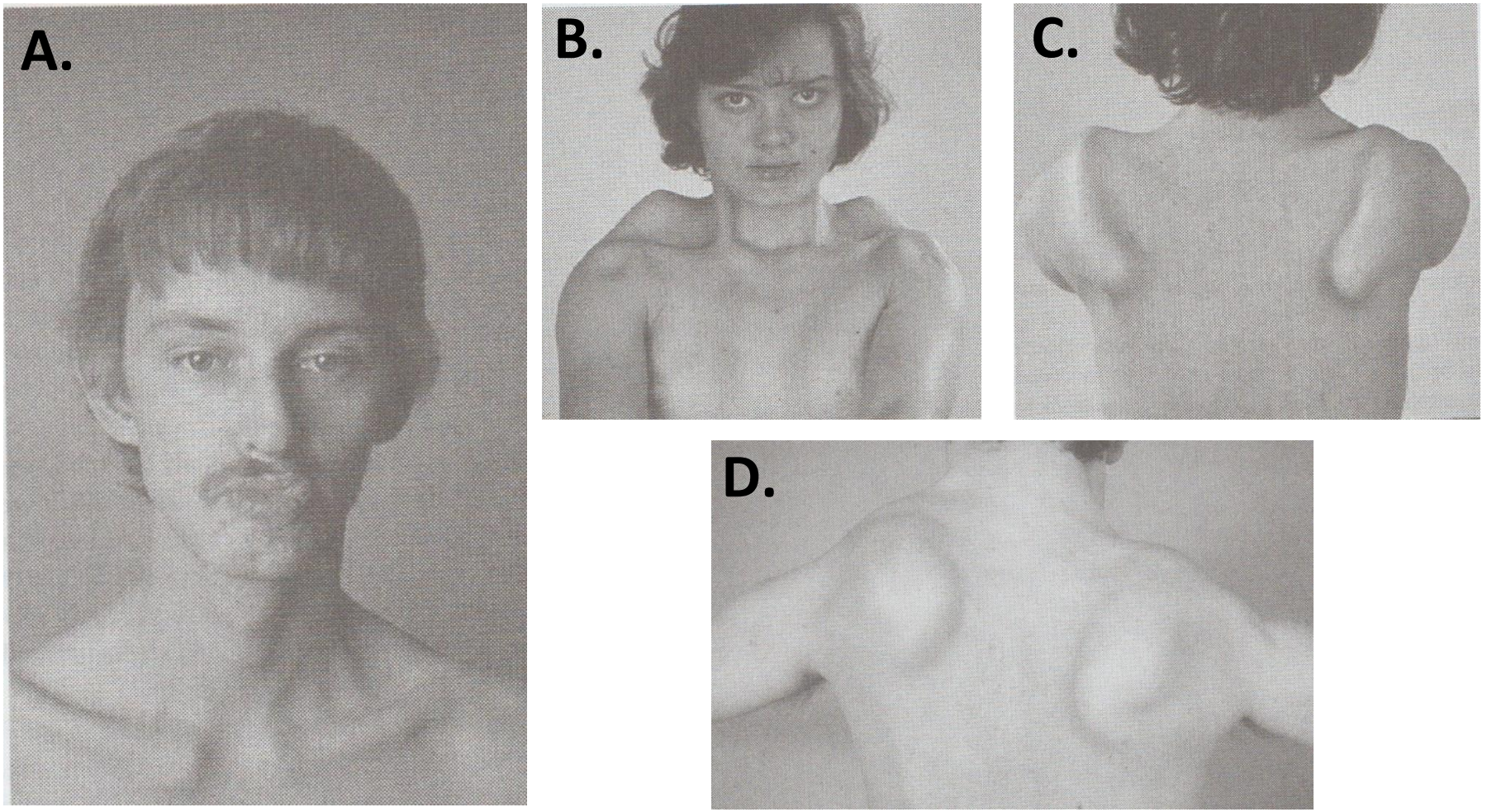


Figure 1.1 Clinical manifestations of FSHD

(**A.**) asymmetrical facial weakness; (**B.**) and (**C.**) elevated scapulae and shoulder girdle weakness; D) asymmetric muscle wasting. Images were acquired from Padberg, 2004

1.2. Genetic basis of FSHD

The genetic mechanisms behind FSHD are as fascinating as they are complex. This autosomal dominant disorder originates from the loss of epigenetic marks within the polymorphic D4Z4 macrosatellite repeat on a FSHD-permissive haplotype located at the telomeric end of chromosome 4q. The most common form of the disease (FSHD1) is caused by a deletion of a large region of the D4Z4 repeat array. In rare cases, mutations in Structural Maintenance of Chromosomes Hinge Domain Containing 1 (*SMCHD1*) gene that methylates the D4Z4 region leads to FSHD2. The mutations in both forms of the disease results in partial relaxation of D4Z4 chromatin that consequently allows expression of a DUX4 transcription factor found to be extremely toxic in skeletal muscle.

1.2.1. *Contraction of D4Z4 repeat array in FSHD1*

The polymorphic D4Z4 macrosatellite repeat array consist of GC-rich 3.3 kb D4Z4 units arranged in head-to-tail fashion in the subtelomeric region of chromosome 4q located around 40kb from the telomere repeat (Figure 1.2) (van Deutekom et al. 1993; Wijmenga et al. 1992; R. J. L. F. Lemmers et al. 2002). In non-affected individuals, the size of the D4Z4 array ranges from 11 to 100 units. A virtually identical repeat array localizes to the subtelomeric region of chromosome 10q. In addition, two major variants of distant chromosome 4q have been identified, namely 4qA and 4qB. However, only contracted 4qA form, with 1-10 repeat units left, can lead to FSHD1. This D4Z4 repeat contraction causes chromatin structure relaxation in somatic cells as demonstrated by the decreased levels of CpG methylation and loss of histone modification markers within

the array (Balog et al. 2012; Zeng et al. 2009). Within each of the D4Z4 unit an open reading frame sequence for a retrogene called DUX4 is present. Interestingly, only the distal unit can produce a stable *DUX4* transcript, since only the last D4Z4 unit is directly proceeded by the pLAM region that provides two facultative introns and a crucial polyadenylation sequence (ATTAAA) that is specific to 4qA variant only. D4Z4 repeat present in other than 4qA chromosome configurations and the homologous sequences located at the chromosome 10q that lack the pLAM sequence, even when contracted, do not express the *DUX4* in somatic skeletal muscle and therefore do not result in FSHD.

1.2.2. ***Role of SMCHD1 in FSHD2***

The FSHD2 term is used to ascribe 5% of FSHD that do not exhibit the D4Z4 repeat contraction in the permissive chromosome 4. Despite normal size of the D4Z4 array, DNA methylation of D4Z4 is strongly reduced in FSHD2, just as it is in FSHD1, indicating a similar epigenetic silencing impairment in both forms of the disease. Whole exome sequencing in selected FSHD2 patients has recently identified mutations in (*SMCHD1*) gene located in the chromosome 18 (R. J. L. F. Lemmers et al. 2012). FSHD2 is a complex digenic disease where the *SMCHD1* mutations are segregated independently from the FSHD-permissive chromosome 4q that encodes stable *DUX4* transcript. Interestingly, *SMUCH1* mutations account for approximately 85% of all FSHD2 patients, indicating that an FSHD3 locus may be yet to be discovered (Winston et al. 2015; Tawil et al. 2014).

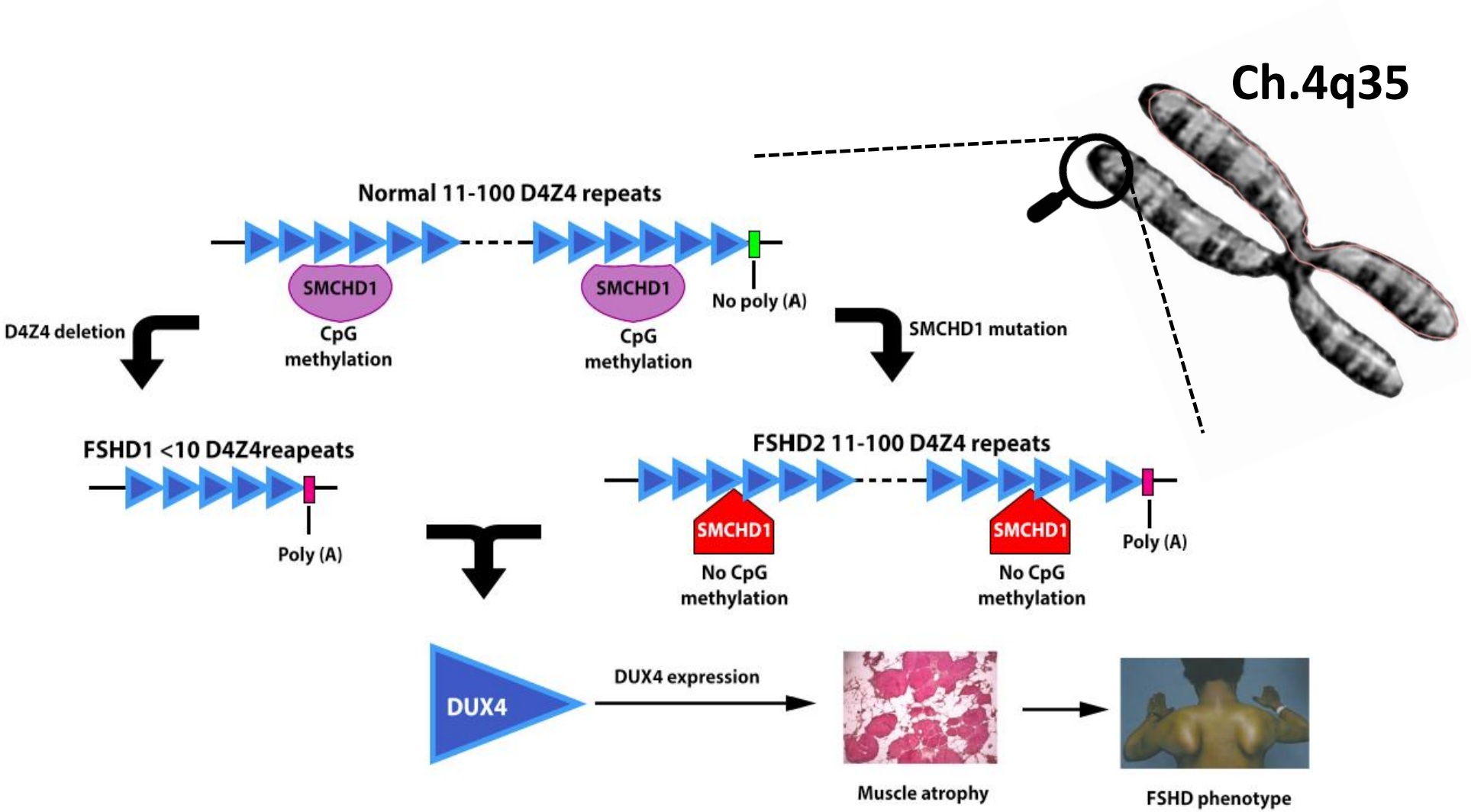
SMCHD1 is a member of a condensin/cohesin protein complex mediating chromatin compaction and has been found to directly bind to D4Z4 repeat array, indicating its important role in epigenetic suppression of the FSHD-permissive chromosome 4q (R. J.

L. F. Lemmers et al. 2012). SMDCH1 shows preferential binding to longer telomeres in human cell lines. A similar mechanism of binding of the protein with affinities dictated by the size of the D4Z4 repeats was proposed (Daxinger et al. 2015). However, the exact molecular mechanism of SMCH1 binding to D4Z4 repeat array is to be established.

Interestingly, recent studies suggest that SMCHD1 can act as a disease severity regulator, especially in FSHD families that carry the upper-limit of 7-10 within their permissive alleles (Sacconi et al. 2013). In these cases, the combined effect of a D4Z4 array contraction with *SMCUH1* mutation results in faster disease progression and earlier onset (Sacconi et al. 2013).

Figure 1.2 The genetic and molecular mechanism behind the FSHD1 and 2

In normal, unaffected cases there are 11-100 tightly condensed D4Z4 microsatellite repeat units (triangles) on the sub-telomeric region of 4q35 chromosome. Large deletion of the D4Z4 repeat units in FSHD1 leaving only 10-1 repeats, leads to chromatin relaxation and allows expression of DUX4 from the last D4Z4 repeat unit. Non-permissive chromosome alleles do not carry a functional poly(A) signal (green bar) that is essential for transcription of a stable *DUX4* mRNA. Only the alleles that carry the functional poly(A) sequence (red box) distal to the repeat array can produce stable *DUX4* transcripts that can be translated into toxic DUX4 transcription factor protein. Methylation of D4Z4 repeat array is regulated by SMCHD1. In FSHD2, the SMCHD1 becomes mutated and unable to methylate the D4Z4 region, resulting in open chromatin conformation and consequently *DUX4* expression. Abbreviations: FSHD, facioscapulohumeral dystrophy; DUX4, double homeobox 4; SMCHD1, structural maintenance of chromosomes flexible hinge domain-containing protein. Images of the patient and muscle histology acquired from the FSH Society website.


Ch.4q35
Normal 11-100 D4Z4 repeats
SMCHD1
CpG methylation
SMCHD1
CpG methylation
No poly (A)
D4Z4 deletion
SMCHD1 mutation
FSHD1 <10 D4Z4reapeats
Poly (A)
FSHD2 11-100 D4Z4 repeats
SMCHD1
No CpG methylation
SMCHD1
No CpG methylation
Poly (A)
DUX4
DUX4 expression
Muscle atrophy
FSHD phenotype

1.3. DUX4 gene, transcript and protein.

FSHD1 and FSHD2 are clinically indistinguishable (de Greef et al. 2010). Despite the differences in the underlying genetic mechanism behind the FSHD1 and FSHD2, they both seem to be a result of an impaired suppression of *DUX4* expression in skeletal muscle (R. J. L. F. Lemmers et al. 2012). However, due to the complex nature of *DUX4* transcription, it took over a decade to determine its expression by RT-PCR after its open reading frame (ORF) was first found within the D4Z4 repeat unit (Hewitt et al. 1994; Dixit et al. 2007a).

1.3.1. *DUX4 gene*

Until relatively recently, repeated DNA elements have been considered as 'junk' DNA. However, due to recent advances in high throughput technologies, a range of RNAs expressed from these repeat elements have been found to play an important role in cell biology (Trofimova and Krasikova 2016). Similarly, initial sequence analysis of D4Z4 region revealed a large ORF of a double homeodomain protein that however lacked a promoter, introns and polyadenylation sequences, indicating that it most likely encoded a non-functional retrotransposed pseudogene (Hewitt et al. 1994; Lyle et al. 1995; Yip and Picketts 2003; Alexiadis et al. 2007). However, subsequent studies have identified a functional promoter within the larger ORF and suggested a smaller double homeobox *DUX4* transcription factor (TF) gene (Gabriels et al. 1999b). Since homeodomain-containing TFs bind DNA and often regulate embryonic development (Gehring et al. 1994), and a large number of genes have been found misregulated in FSHD patient

muscle cells in initial studies, and the *DUX4* gene has become an interesting candidate that could help explain genetic mechanism behind the disease (Tupler et al. 1996).

Detection of a specific *DUX4* mRNA has proven to be extremely difficult due to its high GC content, very low expression in skeletal muscles and high level of sequence alignment with other *DUX* homologous genes distributed across the human genome (Beckers et al. 2001). The first RT-PCR study that could reliably detect *DUX4* expression in FSHD myoblast confirmed that the gene is expressed from the distal D4Z4 unit extending to the pLAM region that contain an intron and a critical polyadenylation sequence (Figure 1.3.1) uu. Further publications have provided a detailed standardised protocol for RT-PCR that has proven useful to others in the field in confirming expression of a stable *DUX4* transcript (Snider et al. 2009a; R. J. L. F. Lemmers et al. 2010). The complexity of *DUX4* detection was also demonstrated by a dilution series experiment showing that a small fraction of 1 in a 1000 FSHD myoblasts express *DUX4 in vitro* (Snider et al. 2010).

Expression of *DUX4* mRNA seems to be induced after initiation of the differentiation process, as all FSHD myotubes and only a fraction of proliferating cells shows to be positive for *DUX4* in proliferating myoblast cultures (Dixit et al. 2007a; Snider et al. 2010a). Interestingly, expression of *DUX4* mRNA could not be found in the patient muscle biopsies (Snider et al. 2010a), most likely confirming a low abundance of the transcript. It was proposed that *DUX4* might be preferentially expressed during regeneration process, in activated satellite cells and their progenitors that constitute a small minority of the total muscle biopsy (Saccone and Puri 2010). Therefore, it would be considerably more challenging to detect *DUX4* in the adult FSHD biopsies, compared

to pathological, highly regenerating or foetal muscles (Richards et al. 2012). Since nucleosomes become disrupted during DNA replication in some of the activated satellite cells and their progenitors, it could contribute to the relaxed chromatin structure that leads to *DUX4* de-repression in FSHD cells (Richards et al. 2012).

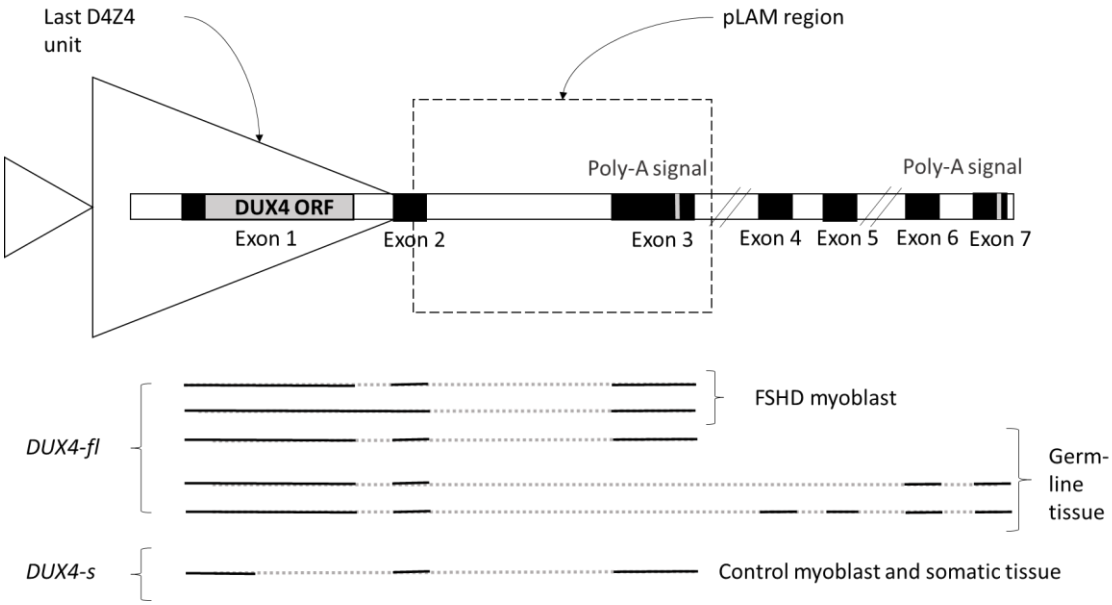


Figure 1.3 DUX4 genomic features and its mRNA variants

(**Top**) Schematic representation of the most distal D4Z4 repeat unit, followed by the adjacent pLAM region and distal exons. The *DUX4* ORF was mapped to the first exon. Two poly-A signals were found to be present in exon 3 and distal exon 7, respectively. The pLAM region provides an intron and a functional poly-A signal on chromosome 4, whereas on chromosome 10 the poly-A sequence is lost. Stop codon maps at the end of the exon 1.

(**Bottom**) Graphic representation of *DUX4* mRNA variants. In all currently investigated FSHD myoblasts, only the full length *DUX4* (*DUX4-fl*) ending in exon 3 could be found. In addition, DUX4-fl was reported in FSHD embryoid bodies, FSHD fibroblast, and in control and FSHD fibroblast-derived IPSC. Germ line tissue also expresses some *DUX4-fl* ending in exon 3 and others ending in exon 7. The shorter *DUX4* (*DUX4-s*) isoform was reported in muscle and other somatic tissues. All of the above *DUX4* mRNA variants derive specifically from the chromosome 4

1.3.2. *DUX4 transcript*

Increasing evidence indicates that *DUX4* transcripts could also play a role in FSHD. It has been found that the D4Z4 repeat array is transcribed in bi-directional fashion that forms sense and antisense *DUX4* mRNA as well as additional small RNAs (Snider et al. 2009a). These types of transcript have been implicated in recruitment of Heterochromatin Protein 1 that could be involved in maintenance of heterochromatic state of the D4Z4 repeat unit (Snider et al. 2009a). Moreover, the bi-directional transcription was also reported in mouse *DUX* paralogue (Clapp et al. 2007).

Location of introns within the 3'UTR of DUX4 could also indicate that levels of the transcript might be regulated by the process of nonsense-mediated decay (NMD), as suggested by the recent experiments showing that NMD endogenously downregulates the *DUX4* mRNA levels in muscle (Feng et al. 2015). An increased level of RNAs with premature stop codon has been found by the RNA-seq analysis in DUX4-induced myoblast cells implying defective NMD. The increased *DUX4* expression has also resulted in downregulation of UPF1 that most likely directly led to the NMD-mediated downregulation (Feng et al. 2015).

Three *DUX4* isoforms have been reported and these include: two full length transcripts (*DUX4-fl*) alternatively spliced in the 3'UTR that encode the complete DUX4 protein; and a short splice variant (*DUX4-s*) that utilises a splice donor site from the exon 1 of the coding sequence (Figure 1.3). The translated *DUX4-s* would produce a protein that contains the homeodomain (i.e., DNA binding domanin), but lacks the transactivation domain of the C-terminal (Figure 1.3)(Dixit et al. 2007; Snider et al. 2009).

The process of transcriptional silencing via siRNA has been well understood over the years (Holoch and Moazed 2015), and has been also found to be relevant in the context of *DUX4* expression (Lim et al. 2015). A number of small RNAs have been mapped upstream of *DUX4* coding sequences and were suggested to mediate silencing of D4Z4 repeat array (Lim et al. 2015; Snider et al. 2009). Targeting of this distal D4Z4 region with exogenous siRNA resulted in augmented H3K9me2 methylation and consequent downregulation of *DUX4* through Argonaut (AGO)-dependent pathway in FSHD cells (Lim et al. 2015). Non-FSHD muscle cells that possess 13 permissive repeat units and were subjected to DICER or AGO2 knockdown become positive for DUX4. However, when the same approach of DICER or AGO2 downregulation was performed in the cells that carried 4 D4Z4 units on a permissive haplotype, there was no apparent induction of *DUX4* expression, further highlighting the importance of D4Z4 endogenous silencing (Lim et al. 2015). This also aligns with the findings showing that SMCHD1 haploinsufficiency results in limited repression of shorter repeat arrays (Lemmers et al. 2012). Since AGO proteins regulate elongation rate of RNA polymerase II that influences alternative splicing (Ameyar-Zazoua et al. 2012), it would be interesting to determine if it also plays a role in splicing regulation of *DUX4-fl* and *-s* isoforms. Indeed, finding these novel silencing processes of D4Z4 provides a platform for development of new siRNA-mediated oligonucleotide therapeutic strategies (Lim et al. 2015).

1.3.3. ***DUX4 protein***

DUX4 protein is around 55kDa in size and was first detected in human testis via immunodetection on a western blot (Snider et al. 2010). After the first specific monoclonal antibody for DUX4 has been developed, studies by Dixit et al., 2007 have

managed to detect the protein in FSHD myoblast but not in control samples. The same antibody was subsequently used to stain nuclear DUX4 using immunofluorescence and the protein has been found to be expressed in approximately 1 in 1000 myoblast, which corresponds to the previous RNA data (Snider et al. 2010). However, since DUX4 protein levels appear to be extremely low, a number of groups have failed to reproduce the above experiments in muscle tissue (personal communication with Dr J. Dumonceaux). Another factor considered to be related to low abundance of DUX4 is its poor stability, as it has been reported that proteasome inhibition greatly facilitated detection of DUX4 (Dixit et al. 2007). Also DUX4-induced cell death seems to be much more pronounced in myoblast rather than in myotubes, leaving virtually no DUX4-positive myoblasts that can be utilised for the protein detection, and few DUX4 expressing nuclei found in differentiated myotubes (Bosnakovski et al. 2008). Finally, there are currently no positive outcomes from the immunohistochemical experiments that have attempted to detect the DUX4 in muscle sections, which makes it difficult to establish whether the expression of *DUX4* is specific to regenerating fibres or satellite cells (Richards et al. 2012).

A lot of scientific effort has recently gone into trying to better understand the biological role of *DUX4* transcript, its encoded protein and how these potentially lead to FSHD pathology. Since DUX4 has been identified as a TF, it was believed that it is most likely involved in deregulation of an array of downstream genes. An example of a gene that becomes upregulated in FSHD patients by binding of DUX4 to its promoter was *PITX1* (paired-like homeodomain transcription factor 1) (Dixit et al. 2007). PITX1 TF has been found to play an important role in limb development and establishment of left/right

symmetry, and is frequently found disturbed in FSHD patients (Figure 1.4 B) (see section 1.1). Furthermore, *DUX4* expression in muscle cell culture leads to increased toxicity thought to be mediated by the caspase-3 upregulation that subsequently leads to cell death (Kowaljow et al. 2007). *MyoD* levels are also downregulated in *DUX4*-transfected C2C12 cells, which results in impaired differentiation of these cells (Bosnakovski et al. 2008). *PAX3* and *PAX7* genes that have been found to regulate muscle regeneration and myogenesis have a highly homologous DNA-binding homeodomain to the DUX4 DNA-binding domain, and it was suggested that these TFs can act as mutual competitors (Figure 1.3.1) (Bosnakovski et al. 2008; Buckingham and Relaix 2007). DUX4 could therefore interfere with *PAX3/7* downstream genes such as *MyoD* and *MYF5* that are responsible for normal myogenic signalling (Bosnakovski et al. 2008). A recent RNA-seq study show that 118 genes are dysregulated in FSHD muscle biopsies, of which 68% are directly regulated by DUX4 expression (Yao et al. 2014), providing a partial explanation how a protein expressed in such an extremely low abundance can have such a dramatic impact on the muscle cell function. Another mechanism explaining DUX4 potency was provided by the study that found the ability of DUX4 TF to diffuse from a single nucleus into a larger portion of the myotube affecting gene expression in neighbouring nuclei (Figure 1.4 A)(Tassin et al. 2013).

Development of FSHD transgenic mouse models that use human equivalent genetic and/or epigenetic mechanism to drive *DUX4* expression has proven to be extremely challenging. However, studies that used vector-induced *DUX4* expression in mice and fish do report myopathic changes in these animals (Snider et al. 2009; Wallace et al. 2011). Interestingly, knock-out mice for *Tp53* shows reduced *DUX4*-related myotoxicity

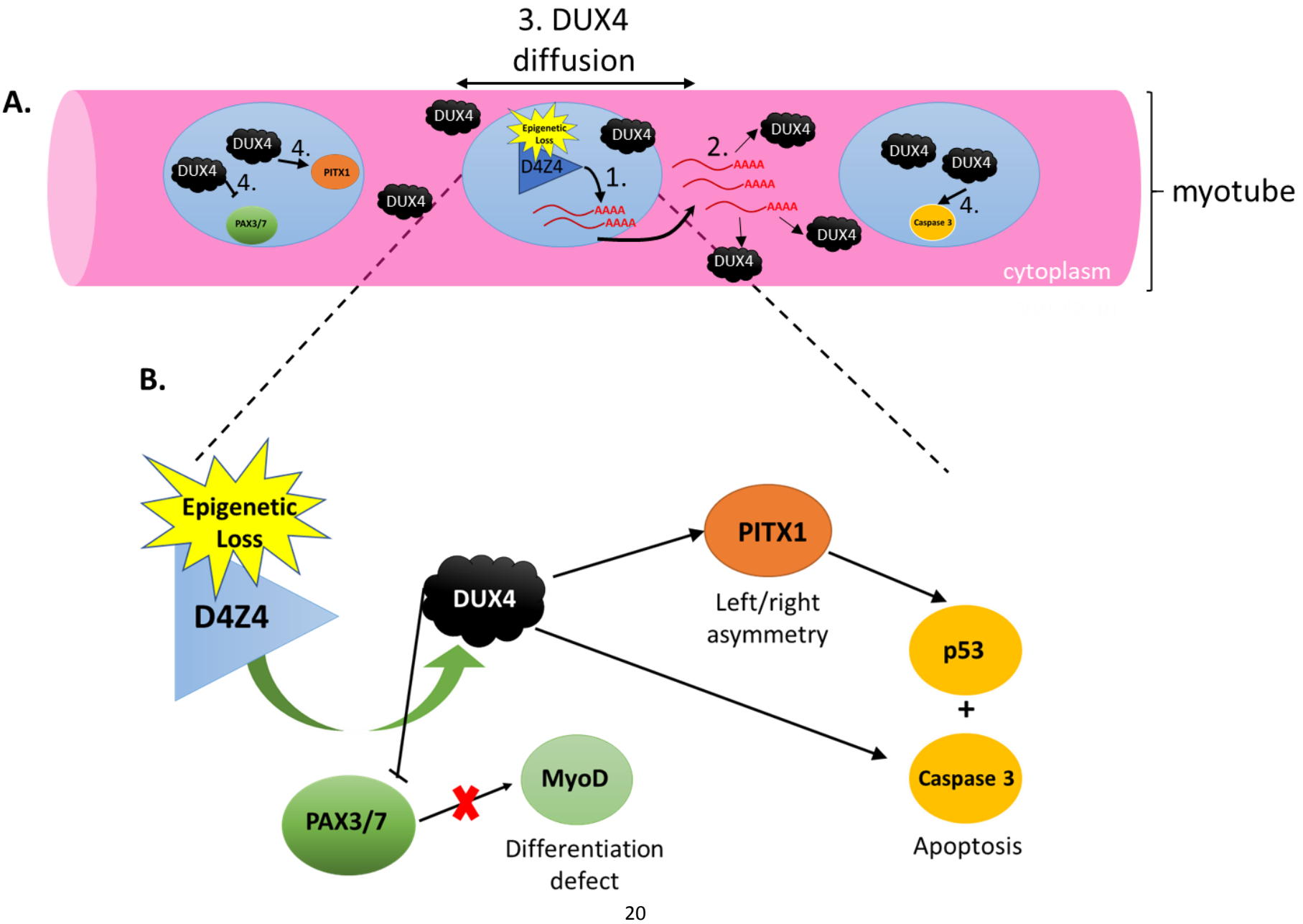
(Wallace et al. 2011). This data aligns with the previous findings showing p53 pathway activation in FSHD muscle tissue (Figure 1.4 B) (Sandri et al. 2001). The p53 as well as being implicated in apoptosis and cell cycle regulation, has also been involved in regulation of muscle atrophy and metabolic homeostasis (Maddocks and Vousden 2011; Schwarzkopf et al. 2006). Recent development of a muscle-specific, doxyciline-induced *DUX4* expression mouse model shows a new promise in better understanding of DUX4 role in postnatal muscle (Bosnakovski et al. 2017). In this model, it is possible to induce a very low-level expression of the toxic gene, thereby recapitulating more the pathophysiological conditions that appear in FSHD patient muscle. Similarly to FSHD patients, these animals shows slow progressive degenerative myopathy and high frequency hearing loss (Bosnakovski et al. 2017).

The C-terminal (transactivating) domain of DUX4 protein is believed to play a major role in mediating cytotoxic effects, as transfected cells with the DUX4c protein that lack the C terminal region do not exhibit myotoxic effects to the same extent as with the full length DUX4 protein (Ansseau et al. 2009). The C-terminal region is also absent in *DUX4-s* isoform, which was previously found in healthy control muscle samples (Snider et al. 2010).

The *DUX4* promoter sequence binding elements such as Sp1, YY1 and E box are characteristic for genes expressed in terminal differentiation (Dixit et al. 2007). However, the fact that half of the Caucasian population lacks the permissive FSHD allele suggests that the DUX4 has a limited or no function in the postnatal stages and might be more relevant during the development (Lemmers et al. 2002).

Figure 1.4 DUX4 transcription factor: process of transcription initiation and propagation cascade

(A.) Myotubes are multinucleated cells sharing common cytoplasm. Each nucleus can express genes independently. In FSHD, DUX4 is expressed in a given nucleus (**1.**) and becomes transcribed from a distal D4Z4 arrays that contains permissive polyA signal which stabilises its mRNA. Multiple copies of DUX4 mRNA are transcribed and transported to the cytoplasm where they become translated into proteins (**2.**). DUX4 is able to diffuse to several adjacent nuclei (**3.**). (**B.**) Each imported DUX4 proteins affects expression of a range of genes. DUX4 can amplify its toxic potency by abnormally upregulating other transcription factors (e.g., PITX1) that can also diffuse into neighbouring nuclei and activate its target genes such as p53. DUX4 expression also leads to downregulation of muscle differentiation master regulator- MyoD transcription fact and upregulate caspase 3 expression -a key regulator of apoptosis. DUX4 cascade in muscle cells leads to atrophy, reduced differentiation potential, and oxidative stress. Figure based on publications by Tassin et al., (2013) and Vanderplank et al., (2011).


3. DUX4
diffusion
A.
DUX4
4.
PITX1
PAX3/7
Epigenetic
Loss
D4Z4
1.
2.
AAAA
Caspase 3
myotube
cytoplasm
B.
Epigenetic
Loss
D4Z4
DUX4
PITX1
Left/right
asymmetry
p53
+
Caspase 3
Apoptosis
PAX3/7
MyoD
Differentiation
defect

1.3.4. *Path to understanding DUX4 biological role*

Evidence that DUX4 might be involved in human development comes from the studies that found *DUX4* expression in human testis (Snider et al. 2010). DUX4 is particularly robustly expressed in germline cells under normal conditions (Figure 1.3)(Snider et al. 2010). However, the transcriptional mechanisms are substantially different to those found in FSHD. Germ line cells express *DUX4* from both 4qA and 4qB arrays as well as the D4Z4 array in the chromosome 10 locus (10q), and instead of the pLAM, the polyadenylation signal located within exon 7 is utilised, which extends the *DUX4* transcript over 4 additional untranslated exons (Figure 1.3). The fact that the *DUX4* transcript was also found to be expressed in induced pluripotent stem cells (iPS) cells derived from control individuals, further indicates that *DUX4* most likely plays a role in human development (Snider et al. 2010). This idea was also supported by the study that found over 74 genes aberrantly expressed in *DUX4*-induced cultures that are directly involved in stem and germ cell function (Geng et al. 2012).

Since the *DUX4* retrogene is conserved exclusively in primates, it has been challenging to study a normal physiological role in development of the gene in standard animal models (Clapp et al. 2007). Current approaches focused on induced expression of *DUX4* globally, which resulted in significant defects in zebrafish or was fatal in mouse (Mitsuhashi et al. 2013; Bosnakovski et al. 2017). Understanding the complex spatio-temporal expression of *DUX4* during development, and recapitulation of these mechanism in an animal model, could lead to better understanding of *DUX4* function in the future.

1.4. Therapeutic strategies to tackle FSHD

The definitive cure for FSHD has not been found yet. Current therapies focus on symptom management and slow down or halting of disease progression. The non-pharmacological interventions include; orthopaedic intervention, surgical scapular fixation; physical therapy; respiratory insufficiency management and aerobic exercise (Tawil 2008). Current pharmacological intervention focus on use of corticosteroids that reduce inflammation apparent in dystrophic patients (Munsat et al. 1972). Furthermore, compounds that could protect muscle mass and function like albuterol and creatine monohydrate and are also frequently recommended to FSHD patients (Benson et al. 1991; Kemp et al. 1993; Tawil 2008). Since impaired methylation of the permissive D4Z4 array repeat and synthesis of the toxic DUX4 TF are the molecular hallmarks of the FSHD, substantial pharmacological research has focused on controlling methylation of the D4Z4 and inhibition of *DUX4* expression.

1.4.1. *Attempts to increase methylation of D4Z4 repeats*

Since it has been well established that the hypomethylation of the D4Z4 repeat array is one of the main features of FSHD1 and FSHD2 needed for the expression of *DUX4* from the permissive 4q chromosome (section 1.2), one rational approach to reverse the FSHD pathogenesis would be to promote methylation of the D4Z4 region. Indeed, a pilot study by Van der Kooi et al., 2006, used folic acid and methionine treatment on FSHD patients to test this hypothesis. Both folic acid and vitamin B12 are required for production of methionine, which is essential for maintenance of DNA methylation (E. L. van der Kooi et al. 2006). Although concentration of serum folate optimal for DNA methylation augmentation was achieved, no such an effect was demonstrated in the FSHD or control

patients following 12-week treatments. It has been proposed that a dose increase of folic acid combined with B12 treatment could be more beneficial (E. L. van der Kooi et al. 2006).

More promising experimental results in epigenetic repression of the D4Z4 region was achieved using CRISPR/Cas9 method (Himeda et al. 2016). Authors have modified the nuclease-deficient component of the genome engineering tool (dCas9) by fusing it to the transcriptional effector domain known as Kruppel associated box (KRAB) that can suppress up to 99% of the target gene in human cells (Gilbert et al. 2013). When the dCas9-KRAB system was recruited to the promoter of *DUX4,* it has led to a significant downregulation of the toxic TF in the FSHD myocytes (Himeda et al. 2016). In addition, the ChIP analysis showed that the KRAB domain has increased recruitment of the HP1α HP1β that had a repressive effect on the chromatin of the D4Z4 region (Himeda et al. 2016).

1.4.2. ***Silencing of DUX4 expression***

There are a number of tools are available that can be utilised to aid gene silencing, including: small hairpin RNA (shRNA), micro RNA (miRNA), small interfering RNA (siRNA) and antisense oligonucleotide chemistries (Wallace et al. 2011; Marsollier et al. 2016).

shRNA, miRNA and siRNA are essentially double stranded, negatively charged RNA that can lead to target gene silencing through independent pathways (Meister and Tuschl 2004). siRNA species was successfully used to downregulated *DUX4* expression by targeting the 3' untranslated region of the gene (Vanderplanck et al. 2011). Since shRNA and miRNA show to be activated and stable in the nucleus, it is believed that these RNA species can have a longer lasting effect and require lower doses to mediate gene

silencing (Bao et al. 2016). In the study by Wallace et al., *DUX4*-targeting miRNA was delivered via adeno-associated viral (AAV) vector into a mouse model expressing AAV-induced *DUX4*, demonstrated a considerable reduction of 90% in DUX4 protein and 64% in *DUX4* transcript levels (Wallace et al. 2011).

1.4.3. ***Muscle stem cell therapy***

Intramuscular transplantation of cultured myoblasts has been considered as a therapeutic strategy for several muscular dystrophies (Tawil 2008). An interesting finding comes from the Vilquin et al. study (2005), where it has been found that myoblasts derived from FSHD biopsies showing healthy histology have unaffected processes of division and differentiation (Vilquin et al. 2005). Therefore, it was proposed that these myoblasts could prove useful for autologous cell therapy in FSHD.

Mesoangioblasts, another type of myogenic mesodermal stem cell, occurring within perivascular tissue of skeletal muscle, have been demonstrated to improve muscle function and morphology as shown in dystrophic animal models injected with these cells (Morosetti et al. 2007). Mesoangioblasts derived from FSHD patient muscle tissue, show to have abnormal morphology and impaired differentiation (Morosetti et al. 2007). Interestingly, it was possible to obtain biopsies from FSHD mosaic patients that produced mesoangioblast populations that were morphologically normal, and functioning virtually as control cells (Morosetti et al. 2007). This raises an opportunity to use mesoangioblasts for autologous cell therapy. An advantage of using mesoangioblasts over myoblasts is the fact that these cells can be delivered systematically, whereas the intramuscular injection of myoblasts have a less practical use clinically (Berry 2015).

1.5. G-quadruplexes: structure, function and roles as therapeutic target.

1.5.1. *G-quadruplex nucleic acid motifs*

All known living organisms contain genetic information in the form of long linear chains of copolymers composed of many nucleotide units (Koonin and Novozhilov, 2009). These molecules, known as DNA and RNA, are typically formed from five different types of nucleotide bases: the pyrimidines - cytosine (C), thymine (T) and uracil (U); and purines - adenine (A) and guanine (G) (Nordhoff et al. 1996). While the T is specific to DNA, the U is characteristic for the RNA polymers. The most commonly occurring complimentary base pairing of nucleotides in the biological systems occurs via hydrogen bonding of A to T or U, and G to C. The A-T (-U) and G-C base pairs are held together by two and three hydrogen bonds, respectively (Nordhoff et al. 1996). In the contexts of DNA, this classical complementary base paring of nucleotides results in the formation of two anti-parallel strands that form the well-known, right handed canonical double helix (B-form) secondary structure, which was first proposed by James Watson and Francis Crick in 1953 (Watson and Crick, 1953). While the ability of the DNA to form secondary structure has been well characterised in the recent past, the discovery that in addition to storing and passing on information, the DNA is also capable of G-quadruplex (GQ) structure formation, has suggested a novel role of DNA in biology (Yang and Okamoto 2010).

Before the B-form DNA structure was discovered, it has been found that at high concentrations, G nucleotides form gels that were not characteristic for any other nucleobases (Bang 1910). Over 50 years later, it appeared that the gelatinous substance was composed of G-tetrad structures (Gellert et al. 1962). In the G-tetrad, four G

residues hydrogen bonded together are held in a square planar array via Hoogsteen pairings as opposed to the classical Watson-Crick hydrogen bonding found in the B-form duplex (Huppert 2010). Further stabilisation of the G-tetrad is mediated by the presence of monovalent cations, usually K^+ and Na^+ (Sen and Gilbert 1990). Since the intracellular concentration of K^+ (140 mM) is considerably higher than of Na^+ (10 mM), it is considered physiologically more relevant in the context of GQ formation and stabilisation (Sen and Gilbert 1990). The centrally positioned monovalent cation within the G-tetrad coordinates to O6 atom of each guanine via strong negative electrostatic interaction (Figure 1.5 A). The linear array assembly running through the centre of the core G-tetrad is promoted by the polarizable aromatic surface of guanine, this creates large π surfaces with a high propensity to stack (Huppert 2008). A core of two or more π-π G-tetrads are considered to be sufficient to form a GQ structure.

1.5.2. *GQ topology*

In both DNA and RNA, the G-tetrads are held together by intervening sequences that are variable in length and nucleotide composition. These stretches form loops and run on the exterior edges of the core. Similarly to the amino acid side-chains in proteins, the loops are the key element that define the structural variability in GQs. The diversity in loop sequences results in highly variable (and sometimes flexible) cavities on the exterior of GQs that can form part of ligand binding sites (Balasubramanian et al.2011).

In addition, the presence of different cations can promote formation of diverse GQ conformations of the same G-rich sequence (Yang and Okamoto 2010). For example, the conserved human telomeric DNA sequence consists of tandem repeats, and thus lacks this sequence variability. This elegantly evolved telomeric sequence is able to form

various GQ structures with low energy differences and therefore convey intrinsic structure polymorphism.

Another factor that influences the GQ formation is the number of strands used to form the structure. In theory, the GQ can be formed from four separate DNA/RNA strands (tetramolecular), two distinct DNA/RNA strands (bimolecular) or a single, continuous strand (unimolecular or intramolecular) (Figure 1.5 B). The unimolecular GQ has been considered to be the most biologically relevant and its structure extensively studied, especially in gene promoters and telomere regions (Balasubramanian et al. 2011). A unimolecular GQ consist of four G-tracts (continuous or discontinuous), three or more loops and two flanking segments. Two types of G-tract can be distinguished based on their direction: parallel and anti-parallel (Figure 1.5 B). The parallel and anti-parallel strand directionality is associated with specific sugar glycosidic conformation. The possible glycosidic conformations that form between the guanine base and its sugar backbone are *syn* and *anti* (Figure 1.5 C). The guanines from the parallel G-tract have the same glycosidic conformation, whereas the antiparallel G-tract will adopt the opposite glycosidic conformation. Loops form three conformations: strand reversal (joining adjacent parallel strands), lateral (connecting adjacent antiparallel strands) and diagonal (coupling together diagonal antiparallel strands on opposite sides of the tetrad) (Figure 1.5 B).

The first biologically relevant GQs were found in eukaryotic chromosomal telomeric DNA (E. Henderson et al. 1987). Human telomeres consist of 5-10 kb d(TTAGGG)¬n tandem sequences with a 50-600 nt single stranded 3’ overhang at chromosome ends that are capable of forming DNA GQs (de Lange 2005). The presence of the DNA GQs in the

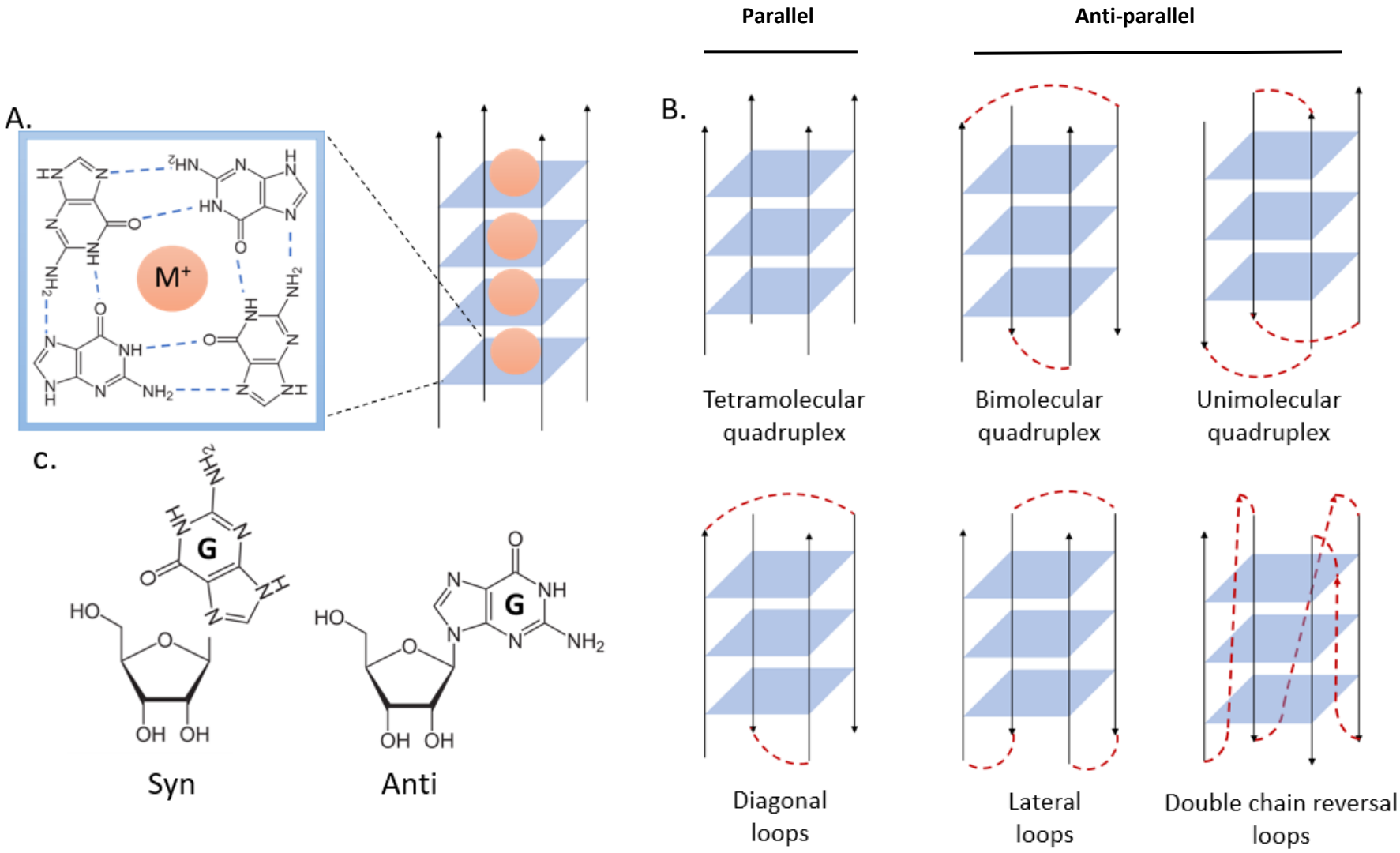
human telomere has been demonstrated to diminish activity of telomerase, which is over activated in 80 to 85% of cancer, making it an attractive drug target (Kim et al. 1994). Structure of the telomeric GQ has been found to be rather complex. It has been shown that in the presence of K^+, it forms two coexisting hybrid structures that possess mixed parallel and anti-parallel G-tracts making the GQ structure (Ambrus et al. 2006a). Since the composition of the G-tracts is the same, it is the position of the strands that allows distinction of the two hybrid structures (Figure 1.5 B). Flanking sequences and loop content is the main contributor for the formation of these hybrid forms. Presence of these two hybrid structures helps explain the multimeric formation at the human telomere 3' overhang. It was also suggested that the structural polymorphism, caused by the asymmetry of the telomeric sequence forming GQs, could provide a platform for protein recognition, and hence control of telomere biology (Dai, Carver, and Yang 2008).

In recent years there was a lot of effort to try and describe structure of GQs, especially those present in a number of oncogene promoters (Balasubramanian et al. 2011). Compared to the invariable human telomeric sequence, the GQs present within the gene promoter regions are more diverse in sequence, possessing different length G-tracts and flanking sequences. Several GQ structures have been described in oncogene promoters. These include genes encoding c-MYC, B cell lymphoma 2 (BCL-2), hypoxia-inducible factor 1 α (HIF1α), vascular endothelial growth factor (VEGF), retinoblastoma protein 1 (RB1), the transcription factor MYB, human telomerase reverse transcriptase and platelet-derived growth factor α polypeptide (PDGFA) (Huppert 2010). Out of all of the above listed genes, the c-MYC was the first and most thoroughly investigated system for the promoter GQ formation (Seenisamy et al. 2004). Within the promoter, the GQ motif maps to the nuclease hypersensitivity element III (NHE III) that consist of 33

nucleotides and contains six G-tracts of unequal lengths. The predominant GQ formed within the c-MYC promoter is a parallel-stranded structure. The G-tetrads are connected by two single nucleotide loops and one two nucleotide loop. Interestingly, these parallel-stranded structures appear to be the most common type of GQ structure in promoter regions (Agrawal et al. 2014).

Figure 1.5 GQ structure topology

(**A.**) Four guanines (G) can self-assemble into stable square planar arrays known as the G-quartet (left panel). Guanines are held together by four hydrogen bonds on each guanine (shown by the dashed red lines) and may be further stabilized by the presence of monovalent cations (orange ball). Stacking of G-quartets generates a GQ (right panel). (**B.**) A basic classification of GQ structures by the number of oligonucleotide strands used to form a quadruplex structure. Tetramolecular structures are generated from four separate strands; bimolecular structures are generated from two separate strands while unimolecular structures are folded structures derived from a single guanine-rich strand. Looping sequences are highlighted in red and strand polarity is indicated by arrow directions. (**C.**) The glycosidic bond angle of G-quartets will change depending of the polarity of strands in a GQ.


A.
M+
B.
Parallel
Anti-parallel
Tetramolecular quadruplex
Bimolecular quadruplex
Unimolecular quadruplex
Diagonal loops
Lateral loops
Double chain reversal loops
C.
G
HO
OH OH
Syn
Anti

1.5.3. ***Detection, visualization and mapping of GQs in cells***

1.5.3.1. Antibody and small molecule based approach to localize GQs in cells

Computational analysis is a useful tool to predict putative GQ motifs across genomes. The potential and application of these algorithms are discussed in more detail in section 3.2.1. However, since the computational methods predict consensus GQ sequences only, potential structures that are formed from non-canonical sequences could be potentially omitted (Schiavone et al. 2014). Furthermore, long repeated DNA sequences are not available in current sequence databases, which could lead to underestimation of potential genomic GQ (Rhodes and Lipps 2015). Importantly, current algorithms predict the GQ motifs to be particularly enriched in non-random, functional locations of the genome such as telomeres, promoters and first introns (Konig et al 2010). Indeed, it is important that the potential role(s) of GQs are studied in a biologically relevant context.

One of the current methods that aids visualisation of DNA structures in cells, utilises structure-specific molecular probes. Antibodies are an extremely powerful tool in detecting molecular structures of a particular protein with an astonishing specificity (Hansel-Hertsch et al. 2017). These can be also synthesised by either immunization or by affinity *in vitro* selection to recognize particular DNA or RNA structures. For example, a GQ-specific single chain variable fragment (scFV) of an antibody, has allowed first visualisation of a biologically relevant GQ formation in the telomeres of the *Stylonychia lemnae* ciliate (Schaffitzel et al. 2001). Two additional GQ-selective antibodies, 1H6 and BG4, have been produced and used to visualize these motifs using immunofluorescence microscopy in human cells (A. Henderson et al. 2014; Biffi et al. 2013). Each antibody

was tested by separate groups in a number of fixed human cell lines, showing cell cycle dynamic GQ formation with the maximum number of the structures being formed in the S-phase (Biffi et al. 2013). Interestingly, the number of GQ nuclear foci increased when the live cells were treated with the structure specific ligands, including TMPyP4, Phen-DC3 and pyridostatin (PDS), suggesting that these compounds have the potential to stabilise the motifs (Henderson et al. 2014; Biffi et al. 2013). Furthermore, depletion of GQ-specific helicase Fanconi anaemia group J protein (FANCJ) resulted in an increase of GQ foci in chicken DT40 cells treated with the stabilising ligand (A. Henderson et al. 2014).

Modified small molecules that can bind GQ also have been used to detect these structures. A fluorophore fused PDS (PDSα) molecule staining has shown a significant accumulation of its signal with the GQ-helicase integration frequency peptide 1 (PIF1) in osteosarcoma cells (U20S), further supporting the role of PIF1 in regulation of GQ structures in human cells (Rodriguez et al. 2012). When other intrinsically fluorescent ligands, DAOTA-M2 and BMVC, were used to visualise GQs, it was found that the motifs are significantly enriched in some cancer cell lines compared to the control cells (Huang et al. 2015; Shivalingam et al. 2015).

Both GQ-specific small molecules and antibodies can potentially influence stability of the motifs through simple binding interactions. Therefore, it is important to supplement the probe-based studies with observation of normal biological processes (e.g., cell cycle) or perform functional analysis by perturbation of key enzymes that regulate GQ stability, to ensure that probe binding would act independent of these processes.

1.5.3.2. Mapping of genomic GQs in cells

One approach to map GQs in a purified single stranded human DNA, used next generation sequencing and GQ-induced DNA polymerase stalling (Chambers et al. 2015a). In this method the isolated DNA is first sequenced under conditions that do not favour folding of GQ, following by sequencing of the same DNA that is sequenced in the presence of ions (e.g., K^+) or ligands (e.g., PDS) that promote quadruplex formation. The GQ dependent DNA polymerase stalling events are compared between the two sequencing procedures. As a result, over 700,000 GQ were identified. This number exceeds that predicted by the bioinformatic estimations, which can be partially explained by false positive results caused by the non-guanine sequences forming bulges or formation of considerably longer loop sequences that are not accounted for in the standard algorithms (Huppert and Balasubramanian 2005). This data suggest that the human genome sequence has the potential to form GQs in numbers vastly exceeding previous estimations (Bedrat et al. 2016). It also confirms that the GQs were particularly enriched within the promoters, 5' untranslated regions and splice sites as well as cancer related genes (Chambers et al. 2015a).

The main pitfall of analysing GQs using bioinformatics and the high throughput sequencing methods is the fact that the DNA is predominately double-stranded (except for telomeric overhangs and transcripts) and condensed in the form of chromatin. In recent reports, detection of GQ utilising GQ-specific antibodies to probe the motif structures using ChIP-seq method, has been explored (Rodriguez et al. 2012; Hansel-Hertsch et al. 2016). They have compared fixed chromatin derived from primary human epidermal keratinocyte (NHEK) cells to the immortalised pre-cancerous keratinocytes (HaCaT) (Hansel-Hertsch et al. 2016). It has been found that only around 10,000 and

1,000 GQ structures could be detected in the chromatin of HaCaT and NHEK cells, respectively. This constitutes around 1% of the GQs found by the sequencing experiments or bioinformatic analysis, indicating that the GQs are supressed in the chromatin state. Proteins might play a vital role in regulating GQ formation, as it has been found that GQs are particularly enriched within regions of chromatin that are deprived from nucleosomes- a highly transcribed sites (Hansel-Hertsch et al. 2016). In addition, the histone deacetylase inhibitor treatment that leads to chromatin condensation also resulted in decreased GQ detection by Chip-seq, consistent with the idea that GQ formation is correlated with transcriptionally active genomic sequences. Interestingly, this study also has showed that the identified sequences significantly overlap with binding sequences of transcription helicases, including xeroderma pigmentosum group B- complementing protein (XPB) and XPD (Gray et al. 2014). In addition, the ChIP-seq-based analysis has demonstrated that various proteins map to the regions that have been predicted to be significantly enriched in GQ forming sequences, including: α-thalasemia/ mental retardation syndrome X-linked human helicases; the Pif1 helicase from yeast and the RAP1-interacting factor 1 telomere protein (Whitehouse and Hughes 2010; Kanoh et al. 2015). Despite the fact that the helicases are able to bind to a range of genomic sequence, increasing evidence suggests that these might also play an important role in GQ regulation.

1.5.4. *Formation and regulation of GQs in cells*

Since GQs formed from short artificial stretches of nucleotides are shown to be extremely stable and with their thermal stability significantly exceeding physiological range, it is important to understand how folding and unfolding kinetics of these motifs

are regulated in real cellular situations. This is especially important, considering that cellular ionic composition, consisting typically of 5-15 nM Na^+, 140 nM K^+ and 0.5-2 mM Mg^{2+} (pH 7.2), is particularly favourable for GQ formation (Davis 2004). The assumption that there are cellular proteins regulating GQ structures becomes increasingly supported by the accumulating number of reports suggesting that chaperone and helicase proteins are responsible for GQ folding and unfolding, respectively (Rhodes and Lipps, 2015).

Predominant evidence showing that chaperones control GQ folding comes from the studies of telomeres. It has been demonstrated two decades ago that the double stranded DNA binding protein, Rap1, induces telomeric GQ formation in *S. cerevisiae* (Rhodes and Giraldo 1995). Yeast telomerase regulatory subunit Est1 has also been shown to bind and stabilise GQ, as well as the human telomeric binding protein TRF2 (Biffi et al. 2012). A convincing *in vivo* experiment has demonstrated that the ciliate telomere ending protein (TEBPβ), regulates GQ folding in a cell cycle dependent fashion (Fang and Cech 1993). In addition, TEBPβ has been shown to increase GQ formation by 10^5 - 10^6 fold *in vitro*. Another example of a human protein that binds to GQs is MutSα, which was previously implicated in recognising DNA mismatches (Larson et al. 2005). Nucleophosmin (NPM1), a frequently mutated gene in acute myeloid leukaemia that normally is involved in ribosome maturation processes, has also been found to bind ribosomal DNA that forms GQs both *in vitro* and *in vivo* (Chiarella et al. 2013). A well-established GQ binding protein, nucleolin, has been recently shown to be sequestered by the presence of an aborted RNA transcript sequence which are carrying expanded hexanucleotide repeat (GGGGCC)n that forms GQs (Dempsey et al. 1999).

As mentioned above, helicases are very likely to play a major role in GQ unwinding. Helicase enzymes are part of a larger ATP-dependent family of proteins that are involved in unwinding of nuclei acids and are also implicated in genome stability regulation (Singleton et al. 2007). An important insight into the role of helicases in GQ structure regulation and genome stability came from studies of genetic diseases that are caused by mutations of these enzymes (Rhodes and Lipps 2015). The conserved domain of RQC present in human WRN and BLM helicases show high affinity binding to GQs, and when mutated it results in Werner and Blooms syndrome, respectively (Lipps and Rhodes 2009). Interestingly, the DOG-1 helicase found in *C.elegans* seems to be essential to maintain the genomic G-rich sequences, especially the ones that have been predicted to form GQs (Cheung et al. 2002). The human orthologue of DOG-1, FANCJ helicase, when mutated has been found to result in the heritable cancer susceptibility disorder, Fanconi anaemia (London et al. 2008). The patient cell lines have large stretches of DNA deleted that are mapped to GQ forming regions. The *in vitro* experiments showing preferential unwinding of GQs over the double stranded DNA by the FANCJ, suggest that just like the DOG-1, these helicases are involved in resolving of transcriptional barriers formed by the DNA GQs (London et al. 2008). Also, RETL1, a DNA helicase regulating telomeres length, when mutated, it increases susceptibility to certain cancers and was also implicated in GQ unwinding, given its high homology to FANCJ and DOG-1 (Vannier et al. 2012). However, due to lack of the biochemical evidence of the RETL1 on GQ unwinding, more work is required to support this idea. PIF1 DNA helicase is another protein that is highly conserved and shows GQ unwinding ability (Paeschke et al. 2013). As previously mentioned, Pif1 of *S.cerevisiae* not only binds GQ sequences as shown by ChIP-seq analysis, but also when deprived from cells, it induces DNA double strand

breaks (Ribeyre et al. 2009; Lopes et al. 2011; Paeschke et al. 2013). It appears that when a GQ unwinding helicase is absent or dysfunctional, a number of nucleases can become involved in deletion of G-tracts that could be potentially involved in the quadruplex formation. Example of such nucleases, include: the yeast Kem1 and human DNA2, FEN1 and EXO1 (Vallur and Maizels 2008; Liu and Gilbert 1994). All of these have been found to cleave GQs *in vitro*. Normally, FEN1 and EXO1 are involved DNA replication and telomere maintenance. In addition, when these nucleases are absent, the telomeres are not maintained correctly and become dysfunctional (Saharia et al. 2008). Another protein that has an important role in telomere maintenance, RPA, has been demonstrated to drive the kinetic equilibrium from the folded to unfolded state of quadruplex structures *in vitro* (Safa et al. 2014).

On the level of translation, an DEAH-box human RNA helicase RHAU has been identified to bind and resolve RNA GQs into linear forms (Lattmann et al. 2011). Over 100 RNAs have been identified to be bound by the helicase at sites that potentially form GQs *in vivo*. Furthermore, it has been found *in vitro* and *in* vivo that binding of RHAU to one of its targets which is the human telomerase RNA TER, depends on the presence of a stable GQ structure within the 5'-region of the TER's transcript. Additionally, disruption of the RHAU helicase lead to incorrect telomerase assembly and telomer extension (Booy et al. 2012). Together, these findings strongly indicate the there is a vast array of proteins mediating GQ unfolding that are involved in preventing DNA breaks, and interruption of replication and translational processes.

1.5.5. *Role of GQs in transcription and translation.*

Considering that over 50% of human gene promoters were found to potentially form GQs within their sequence, it is tempting to speculate that these motifs might play an important role in gene expression regulation (Huppert and Balasubramanian 2005). Interestingly, it has been predicted that the promoters of oncogenes and regulatory genes, for instance transcription factors, are particularly likely to contain GQs compared to other gene types, such as promoter tumour suppressor and housekeeping genes (Huppert and Balasubramanian 2005; Nakken et al. 2009). This pattern of GQs presence seems to be conserved in other organisms, for example bacteria, yeast and plants (Capra et al. 2010; Hershman et al. 2008; Todd et al. 2005). Furthermore, the GQs in humans are predominately found in the non-template strand and also to congregate at 5' of the 5'UTR (Huppert 2008). In bacteria transformed with GQ encoding plasmid on the non-template strand, a loop structure formation was detected on the complementary strand, indicating formation of the GQs in cells during transcription (Duquette et al. 2004). GQ formation in this case could release the template strand from the double stranded conformation in order to allow uninterrupted, high level transcription (Duquette et al. 2004). Therefore, since it is generally believed that the GQ formation on the non-template strand has an enhancing effect on transcription, it has been also suggested that the GQ present on the template strand could inhibit transcription by potentially blocking the transcription machinery (Qin and Hurley 2008) (Figure 1.6).

One of the first and best studied examples of GQ role in transcription regulation comes from the study of the *c-MYC* oncogene promoter (Huppert 2010). MYC is a cell proliferation-associated transcription factor that has been found upregulated in 80% of the human cancer cells (Pelengaris et al. 2002). The *c-MYC* promoter region contains GQ

forming sequence that has been shown to control over 80% of the gene's expression (Simonsson et al. 1998). The reporter studies comparing gene expression driven by the wild-type *c-MYC* promoter to the one carrying sequence mutation disrupting quadruplex formation, have demonstrated that presence of the GQ within the promoter sequence supresses gene expression (Siddiqui-Jain et al. 2002). In addition, treatment with a GQ stabilising ligand, TMPyP4, has also shown decrease in *c-MYC* transcription in lymphoma cell lines as well as antitumor effects in mice (Grand et al. 2002). However, since the TMPyP4 is not GQ specific and it shows binding affinity also to double stranded DNA in multiple regions of the genome, more work needs to be done in order to establish the exact mode of action of this chemistry. Interestingly, a highly selective compound for the *c*-MYC promoter-related GQ known as GQC-05 has been found. Its treatment of Burkit's lymphoma cell line has led to significant reduction of *c-MYC* mRNA expression, further highlighting the potential of rational drug design for the development of highly specific GQ chemistries (Brown et al. 2011).

Previously mentioned GQ-binding protein, nucleolin, is one of the most abundant cellular proteins and has been found to bind to the promoter sequence of *c-MYC*. What is more, over-expression of nucleolin leads to dose-dependent reduction of *MYC* transcription (González and Hurley 2010). It has been suggested that the nucleolin-mediated oncogene suppression is mediated by binding and stabilising of the *c-MYC*-related GQ, which prevents binding of the transcription activating factors (e.g., SP1 and CNBP) (González et al. 2009).

GQs have been found vastly overrepresented within 5'UTRs of a large number of genes, suggesting their important role in translation regulation (Bugaut and Balasubramanian

2008). It was initially speculated that presence of the extremely thermodynamically stable GQs within the 5'UTR would result in an inhibition of translation, and indeed this has been found to be the case for number of genes, including: *Bcl-2*, *ESR1, FMR1*, *NRas*, *TRF2*, *Zic-1* and *Mt3-MMP* (Bugaut and Balasubramanian, 2008). However, cases where translation becomes enhanced by the presence of the 5' UTR GQs have also been demonstrated for fibroblast growth factor 2 and vascular endothelial growth factor transcripts (Bonnal et al. 2003; Morris et al. 2010). Another example demonstrating that mRNA GQs can regulate translation, comes from the studies of fragile X mental retardation, where the transcript expansion of the (CGG)n repeat causes the mental retardation protein (FMRP) to bind to its own transcript. It has been demonstrated that the expanded transcript variant forms GQ stable structures that promote binding to the FMRP (Brown et al. 2011). This FMRP binding to the GQ-forming transcript lead to abnormal mRNA processing and consequently result in the disease pathology (Darnell et al. 2001). Furthermore, GQs have also been found present within the 3'UTR of transcripts where they mediate alternative polyadenylation and transcript shortening (Beaudoin et al. 2014).

Identification of RNA helicases such as RHAU that bind GQs with high affinity, also indicates that GQs could be instrumental in regulation of translation (Lattmann et al. 2011).

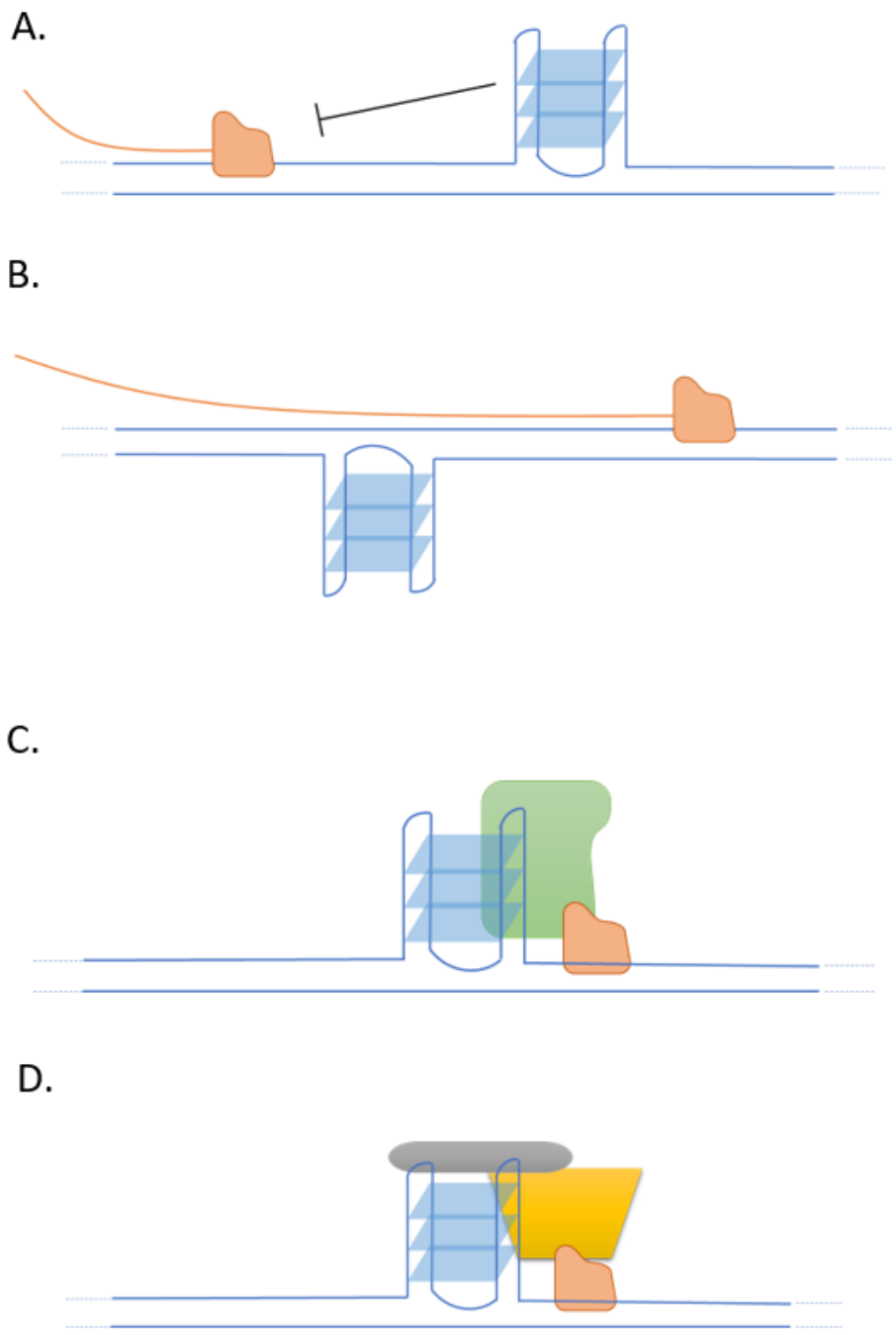


Figure 1.6 Potential roles of GQs in regulating transcription

(**A.**) GQs were suggested to act as physical barrier on the template strand that stalls or inhibits polymerase (orange) that consequently prevents transcription. (**B.**) Formation of the GQ on the coding strand may facilitate transcription by keeping the template strand in the open conformation. (**C.**) GQs could recruit transcriptional activator proteins (green) that promote transcription. (**D.**) Alternatively, GQs can recruit transcriptional suppressor proteins (grey, yellow) that inhibit transcription.

1.5.6. *DNA and RNA GQs as a target for small-molecules compounds.*

Design and development of effective GQ-binding small molecules still presents many challenges. These limitations particularly concern specificity of binding to GQs over duplex DNA structures, selectivity of various GQ types distributed across the genomic loci (e.g., telomeric *vs.* promoter DNA), and binding selectivity to different nucleic acid chemistries, e.g., DNA *vs.* RNA. Although GQs show a vast diversity in their structures, there are many common characteristics in their conformation that can be utilised for development of small molecule chemistries. These strategies were particularly explored in the targeting of telomere, promoter and 5'UTR related GQs.

1.5.6.1. GQ targeting: telomeric DNA

The first investigation looking at the effects of small-molecule binding to GQs was performed in the context of telomerase activity, which can become inhibited upon treatment that stabilises telomeric GQ structures (Sun et al. 1997). Consequently, a number of novel compounds have been developed that show binding to the telomeric DNA GQs. However, despite development of a large library of these compounds (listed in online 'G-quadruplex Ligands Database'), only a handful of these molecules have been tested in cell-based systems, of which very few have reached clinical trials (De Cian et al. 2007). When designing small molecules to bind GQs with high affinity and specificity, three main chemical properties typically are maintained:

a) A large aromatic core that ensure maximal π stacking interaction of the G-tetrad.

b) Presence of a positive charge that uses the negative charges of the negatively charged phosphate backbone to enhance interactions.
c) Introduction of sidechain groups that increase specific interactions with specific loops/grooves or individual bases of a particular GQ structure.

Existing methods that help accelerated rational design of small molecules with optimal binding properties to a given GQ structures combine high-resolution NMR and X-ray crystallography with computational molecular modelling methods (Li et al. 2013). The candidate molecules are subsequently assessed for quadruplex-ligand binding using a range of biophysical techniques, including circular dichroism-UV, fluorescent resonance energy transfer or electrospray ionisation mass spectrometry. Many small molecule chemistries have been developed using the above approach with many showing preferential binding to GQ over duplex DNA structures, and specificity for telomeric quadruplexes over other types of the motif structure found for example in the promoter sequences (De Cian et al. 2007).

When screening test compounds, it is important to choose appropriate cell lines in order to assess their toxicity and ensure it is relatively low. Recently, the most comprehensive testing of telomeric DNA quadruplex binding in cell-based and *in vivo* systems has been performed for BRACO19, RHSP4 and telomestatin (Burger et al. 2005; Phatak et al. 2007; Tauchi et al. 2006). Despite a range of the compounds being synthesised and tested for their efficacy and safety, only one candidate has reached clinical trials, namely Quarfloxin. The ineffectiveness of these compounds is often attributed to their ability to bind off-target protein, RNA and DNA species. It has to be noted that the d(TTAGGG) quadruplex-forming telomeric sequence, apart from the telomeric ends, can also be

found in a number of regions within the genome (Todd and Neidle 2011). Therefore, any future quadruplex-drug design has to consider any potential off-target sites that the potential chemistry might bind inside the cells.

1.5.6.2. GQ targeting: promoter DNA

During the transcriptional process, the double-stranded DNA structure becomes transiently opened, which allows for the formation of a quadruplex within the G-rich stand and an i-motif on the complimentary C-rich strand (Brooks et al. 2010). This event presents an opportunity to develop small molecules that target and stabilise either of these structures to regulate transcription. The main limitation behind the quadruplex/i-motif formation within the promoter sequences, is the stability of the double stranded DNA. However, presence of the hypersensitive elements (e.g., found in *c-MYC* promoter) and susceptibility of promoter sequences to transcription-induced supercoiling, can dramatically facilitate formation of these secondary structures over the duplex form (Sun and Hurley 2009).

When performing bioinformatic identification of promoter GQs, it has to be noted that that the general algorithms that identify typical GQs of 3-5 G-runs linked by 1-7 nucleotides, might omit promoter quadruplexes that can contain linking sequences of up to 26 nucleotides and extended G-runs as reported in the PDGF-A promoter (Todd et al. 2005). Therefore, selection of a correct sequence that will result in biologically relevant quadruplex forming structure is critical for rational drug design and development. The first example of gene expression at the transcription comes from a study of TMPyP4 ligand on the expression of *c-MYC* oncogene (Siddiqui-Jain et al. 2002). The TMPyP4-mediated effects leads to gene expression inhibition, and combined with

mutagenesis studies of the quadruplex sequence, it has been confirmed that the presence the stable motif within the promoter is responsible for inhibition of the *c-MYC* expression (Siddiqui-Jain et al. 2002). Additional examples of promoter forming GQs were found within promoters of BCL2, c-KIT, VEGF and HIF1-α, where their presence has been found to be generally inhibitory, except for the VEGF promoter in which the presence of the quadruplex has been shown to have an enhancing effect on the rate of transcription (Figure 1.7 B)(Dash et al. 2008; De Armond et al. 2005; Nambiar et al. 2011; Sun et al. 2008). As for the telomeric quadruplexes, the same limitations exist when it comes to designing small molecules targeting these structures. Despite the promising specificity for quadruplex over duplex that can be achieved with current chemistries, targeting of quadruplex structures for specific promoters still remains a major challenge.

1.5.6.3. GQ targeting: mRNA 5'UTR

As previously mentioned, bioinformatic analysis has found a significant enrichment of GQs present within 5'UTR of mRNAs, indicating that these motifs might play a role in gene expression regulation (Huppert and Balasubramanian 2005). The first demonstration of function of 5'UTR GQs comes from the study of NRAS proto-oncogene, where it has been found that the formation of the stable motif structure within the 5'UTR leads to the gene downregulation (Figure 1.7 C) (Kumari et al. 2007). It has been suggested that the gene expression downregulation is caused by the quadruplex-induced steric blocking of the ribosome complex and inhibition of translation initiation (Kumari et al. 2007). One advantage of targeting 5'UTR quadruplexes is the fact that the mRNA is single-stranded, and there is no complimentary strand that competes with the motif for the duplex formation.

Another example of RNA GQ structure involvement in disease pathogenesis comes from the study of Fragile X mental retardation syndrome, where the disease-causing proteins, FMRP1 and FMRP2, have been demonstrated to bind and stabilise GQs of 5'UTR of a variety of gene transcripts (e.g., PP2 and MAPB1)(Castets et al. 2005; Darnell et al. 2001). In addition, the FMRP1 has been found to also bind directly to quadruplexes within its own coding sequence and within the mRNA of AAP gene (Westmark and Malter 2007). The FMRP1 quadruplex forming sequence within the coding region of the gene has been suggested to be a potential exon splice enhancer, since the FMRP1 and 2 have been demonstrated to bind to these motifs and affect splicing patters of the affected mRNA (Didiot et al. 2008). Interestingly, the FMRP1 has been also shown to bind RNA GQ targets within 3'UTR of a number of genes, however the potential role of these motifs is yet to be determined (Darnell et al. 2001).

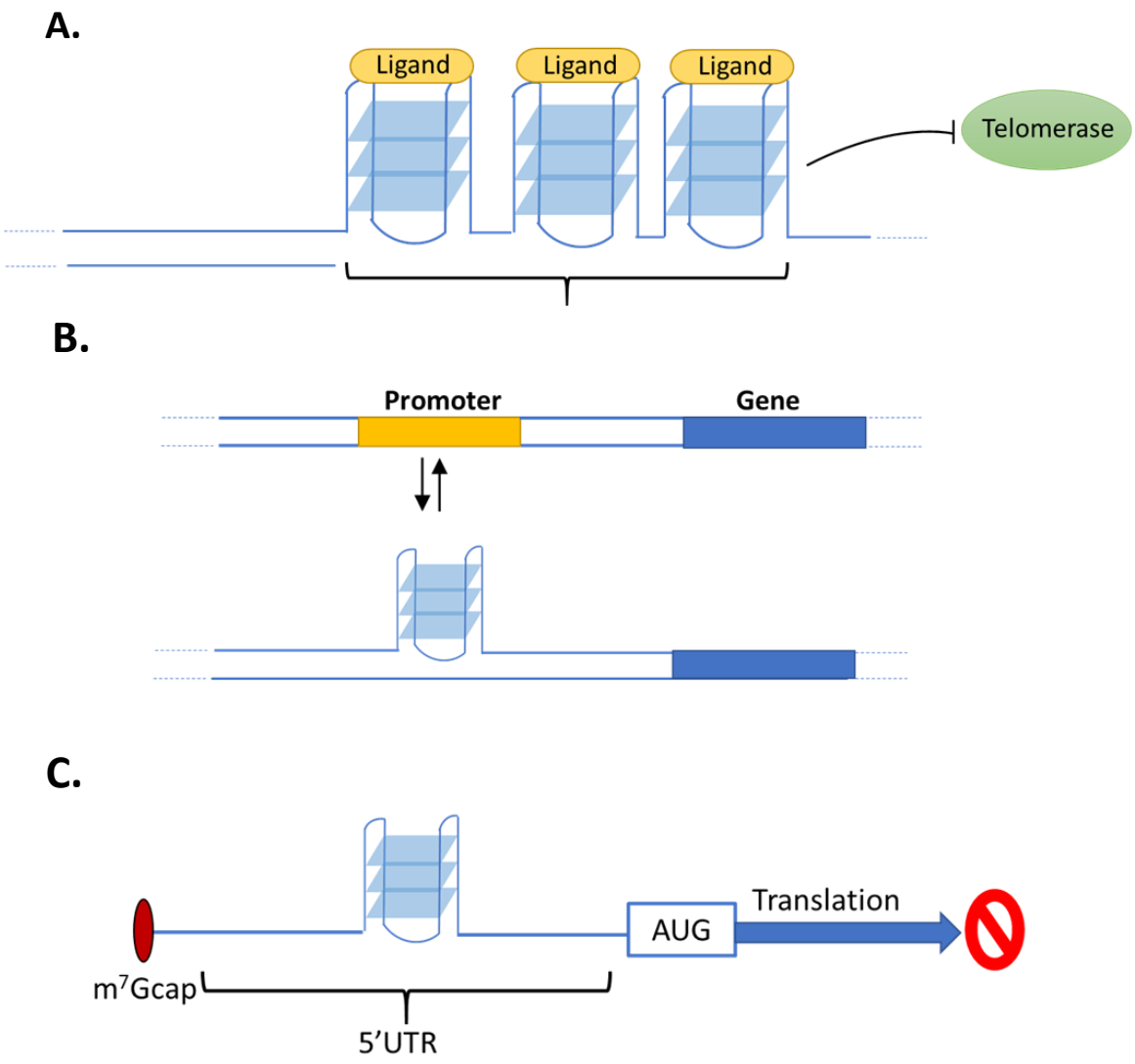


Figure 1.7 Potential roles of GQs in gene expression regulation

Genome wide bioinformatic analysis have demonstrated the presence of non-randomly distributed GQs in functional genomic loci such as telomeres, promoters and 5'UTRs of mRNAs. These combined with protein- and ligand-GQ interaction studies provide an insight into possible roles of these motifs in gene expression. (**A.**) GQs forming at the telomeric 3' overhangs, especially when stabilised with ligands, can prevent telomerase from regulating length of telomeres. (**B.**) Before the transcription initiation process, the double stranded helix becomes unwound and allows stable GQ formation. This masks transcription factors from binding sites, leading to transcription downregulation. (**C.**) Stabilisation of GQ at the 5'UTR of pre-mRNA prevents progress of the ribosome complex, resulting in downregulation of translation.

1.5.7. ***Berberine***

Traditional medicines, often derived from plants, have been used as remedies to effectively treat physical and mental disorders long before modern medicine has been established (Guamán Ortiz and Scovassi 2013). As modern science starts to understand the biological and chemical properties of these traditional treatments, they have become increasingly recognised and recommended by the scientific communities (including World Health Organization) as an alternative treatment for a number of different diseases (Kohler and Baghdadi-Sabeti 2011).

Secondary metabolites from plants, such as flavonoids, tannins saponins, steroids and alkaloids have been shown to regulate many biological effects, including: anti-inflammatory, anti-bacterial, immunomodulating, antioxidant and even anticancer (Teiten et al. 2013; Wang and Chen 2013). One example of a plant alkaloid widely studied for its range of pharmacological properties is berberine.

Berberine belongs to a group of protoberberine compounds known as benzylisoquinoline alkaloids ($C_{20}H_{18}NO_4^+$) and is found in a number of plants, including: *Coptis chinensis*- Chinese goldthread, *Mahonia aquifolium* – Oregon grape, *Berbis aristata*- tree turmeric, *Hydrastis canadensis*- goldenseal, *Berbis vulgaris*-barberry, *Xanthorhiza simplicissima*-yellowroot, *Argemone Mexicana*-prickly poppy and *Phellodendron amurense*- Amur cork tree (Tillhon et al. 2012). The use of berberine containing plants for medical use dates to 200 A.D where in the work of "*The Herbal Classic of the Divine Plowman"* by Shen Nong Ben Cao Jing, it was used to treat gastrointestinal infections (Lian 1986). Later, in 500 A.D. it was first realised that the berberine containing plant, *Rhizoma Coptidis,* shows anti-diabetic properties as

reported by Hongjing Tao in his *"Note of Elite Physicians"*(S. Li 1996). Other medical applications of berberine include treatment of coronary artery disease, hyperlipidemia, hypertension, obesity, polycystic ovary disease, diabetes and Alzheimer's (Heidarian et al. 2014; Tang et al. 2013; Zhang et al. 2013). Interestingly, berberine has been recently demonstrated to inhibit proliferation and migration as well as promote apoptosis in a number of cancer cell lines, suggesting its potential as a novel anti-cancer treatment (Li et al. 2013; Liu et al. 2013; Zhang et al. 2010). Since berberine shows therapeutic benefits in a range of distinct diseases, it is clear that the compound has a complex mechanism of action that involves targeting of multiple pathways. Some of the berberine protein targets discovered to date, include p53, NF-kB, MMP, telomerase and estrogenic receptors (Li et al. 2013; Liu et al. 2013; Tillhon et al. 2012). Importantly, berberine has been shown to strongly bind DNA and RNA (Bhadra and Kumar 2011b), resulting in double-strand breaks and DNA topoisomerase binding inhibition (Li et al. 2013; Qin et al. 2007). Berberine was also found to prevent transcription by binding TATA box sequence in gene promoters, which leads to displacement of enhancer proteins (Jiao Wang and Jiang 2012).

Berberine's particularly strong binding to secondary DNA structure such as GQ over duplex DNA also has been previously demonstrated (Bhadra and Kumar 2011b). Human telomeric sequences are bound by berberine with high affinity, and was suggested to block the abnormal telomerase-mediated chromosome elongation that leads to cellular tumorigenesis and immortalization. (Bazzicalupi et al. 2013; Bernardes de Jesus and Blasco 2013). This GQ binding ligand has also been shown to have antifibrotic properties in cardiac fibroblast and rats with induced myocardial fibrosis (Gu et al. 2012). Berberine upregulates relaxin-1, a protein that can prevent and reverse cardiac fibrosis, in a dose

depend manner *in vivo* and *in vitro.* The antifibrotic mode of action of berberine was suggested to be a result of the ligand binding to GQ forming structures within the promoter of relaxin-1 (Gu et al. 2012).

Despite the issue of berberine having a range of non-specific binding to various protein and nucleic acid targets, its structural properties, in particular the aminoalkyl side chains, allows for chemical modifications that can address the issue of poor specificity. For example, a 9-0-substituted berberine with extended side chains and terminal amino groups shows a significant increase in specificity for GQs over any other DNA structures (Ma et al. 2008; Zhang et al. 2007). In addition, linking of two berberine moieties with a polyether linker has demonstrated a 76-508 fold increase of binding to GQ structures compared to native berberine structure (Z.-Q. Li et al. 2017).

Proven safety of berberine and its therapeutic potential in a range of diseases models as well as its capacity for chemical modification to increase its GQ specificity, makes it an attractive molecular candidate for initial proof-of-concept testing to: (i) determine any positive improvement of molecular disease signature, and (ii) determination of potential target pathways of berberine action that mediate these changes at either protein or nucleic acid level.

1.6. GQs and DUX4 genomic locus.

The D4Z4 repeat array contains a high CG content (73%) and in unaffected individuals it provides a large CpG island platform for the formation of methylated, DUX4 repressive closed chromatin state (Tsumagari et al. 2008). A lot of effort has gone into understanding how the impairment of the methylation process leads to the FSHD

pathology (Section 1.5.1). However, virtually no attention has been paid into the potential presence of GQ-mediated *DUX4* expression regulation. The single mention of potential GQ-forming sequences in D4Z4 locus has been published by Tsumagari et al., 2008. These have been mapped to the region of putative *DUX4* promoter sequences and proximal site of the D4Z4 repeat (Tsumagari et al. 2008). In this work the GQ-predicting algorithm that was used discriminates any potential motif-forming sequences with loops of >7 nucleotides. The idea that the GQs do not form stable genomic structures has been challenged with recent findings of *hTERT* and *BCL-2* gene promoter GQs that contain 26-nuleotide and 13-nucleotides loops, respectively (Agrawal et al. 2014; Palumbo et al 2009). In addition, at the time of the study the DUX4 myogenic enhancer 1 and 2 had not been established (Himeda et al. 2014). Therefore, these elements as well the DUX4 transcript, should be analysed using bioinformatic tools that allow for extensive GQ loops and allow prediction of RNA GQ sequences.

1.7. Aims

We aimed to advance our understanding of molecular mechanisms controlling expression of *DUX4* by identifying, and assessing the role of, novel GQ regulatory motifs within enhancer, promoter and transcript elements of the gene. This was achieved through the following objectives:

I. Prediction of GQ formation within the analysed sequences with bioinformatic tools, and assessment of the candidate sequences for the formation of GQ structures using biophysical techniques.

II. Testing of the biophysical interaction between berberine and the selected sequences in solution, and *in vitro* effects of the ligand on the *DUX4* expression to evaluate the GQ motifs as potential therapeutic targets.

2. Materials & Methods

2.1. General laboratory reagents

All general chemical reagents were purchased from Sigma, Invitrogen, BDH or VWR (unless stated otherwise) with standard chemical purity graded as AnalaR (Analytical Reagents for analysis applications). More commonly used chemicals are listed below, whereas the regents required for only more specialised assays are described in the relevant materials section. All of the reagents were dissolved in double de-ionized H_20 (ddH20), unless stated otherwise. The majority of the solutions required for tissue culture use were autoclaved at 121°C for 15 min, except for the solutions containing protein, detergents or glucose which required passing through a 0.22 µm filter (Falcon). All solutions were stored at room temperature unless indicated otherwise.

Table 2.1 List of common reagents

Reagents	Manufacturer:
Agar	Sigma
Agarose	Invitrogen
Berberine chloride	Sigma
Calcium chloride ($CaCl_2$)	Sigma
Dimethyl sulphoxide (DMSO)	Sigma
EDTA	Sigma
Ethanol (EtOH)	Sigma
Glacial acetic acid	WVR
Glucose	Sigma
Goat serum	Sigma
Hydrochloric acid (HCl)	Sigma
Lysogeny broth (LB)	BDH
Magnesium chloride ($MgCl_2$)	Sigma
NaOH	BDH

Na_2HPO_4	Sigma
Paraformaldehyde (PFA)	Sigma
Potassium chloride (KCl)	Sigma
Phosphate buffered saline (PBS) pH 7.3	Sigma
Sodium chloride (NaCl)	Oxoid Ltd
Triton X-100	Sigma
Tween-20	Sigma
Trizma Base	Sigma
Trizma hydrochloride (HCl)	Sigma

2.2. Bacterial culture and storage

2.2.1. *Materials*

- Ampicillin (Sigma): prepared as 1000x stock in ddH2O at 20 mg/ml concentration, filtered through 0.22 µm filter (Falcon) and stored at -20°C.
- Kanamycin (Sigma): prepared as 200x stock in ddH2O at 10 mg/ml concentration, filtered through 0.22 µm filter (Falcon) and stored at -20°C.
- Agar (Sigma)
- Glycerol (Sigma)
- Lysogeny Broth (LB) (Sigma).
- LB (1.5%) agar: 1.5% w/v agar was added to LB solution and autoclaved. The LB-agar solution was cooled to 45°C before addition of desired antibiotic. ~25ml of LB-agar was added into 100 mm petri dish in proximity to the Bunsen burner. Once solidified the plates were sealed in cling film and stored at 4°C
- SOC medium: 100 µl of 1 M $MgSO_4$ and 20 µl of 1 M glucose were added to 10 ml LB medium.

- TOP 10 (*E. coli*) competent cells (Invitrogen)

- $CaCl_2$ (Sigma): prepared as 100 mM stock by adding 5.5 g of $CaCl_2$ into 500 µl of ddH_2O. The solution was filtered through 0.22 µm filter (Falcon) and stored at -20°C.

- 100 mM $CaCl_2$ in (15%) glycerol: 5.6 g CaCl2 and 75 ml glycerol were made up to 500 ml with ddH_2O, filtered through 0.22 µm filter (Falcon) and stored at -20°C.

2.2.2. *Preparation of TOP10 bacteria*

The bacterial cells (TOP 10, Invitrogen) were streaked on LB-agar plate without antibiotic and incubated overnight at 37°C. The following day, a single colony was selected and added to 5 ml of LB media (no antibiotic). The LB media containing the colony was then incubated at 37°C overnight with shaking (200 rpm). 5 ml of the overnight culture was added to 500 ml of LB media (no antibiotic) and incubated at 37°C with shaking (200 rpm) until the OD260/280 was between 0.2 and 0.5. The cells were pelleted at 3000 rpm for 5 min at 4°C. The supernatant was removed and the pellet was resuspended in 50 ml of 100 mM $CaCl_2$. The resuspended cells were left on ice for 20 min. Cells were pelleted again at 4000 rpm for 5 min at 4°C. The supernatant was removed and cells resuspended in 5 ml (1/10 original vol) ice-cold 100 mM $CaCl_2$/15% glycerol. Resulting cells suspension was divided into 50 or 250 ul aliquots on dry ice submerged in methanol and stored at -80°C.

2.2.3. *Bacterial culture and storage*

A single bacterial clone was selected from the LB selective plate and incubated overnight at 37°C with shaking (200 rpm) in 5 ml of antibiotic containing LB medium. 700 µl of the overnight culture was resuspended in 300 µl of 80% glycerol and stored at -80°C.

2.3. Expansion and purification of DNA plasmid

2.3.1. *Materials*

- QIAprep spin Miniprep kit (Qiagen).
- SOC medium: 100 µl of 1 M $MgSO_4$ and 20 µl 1 M glucose were added to 10 ml LB medium.
- TE: 10 mM Tris pH 8.0 and 1 mM EDTA pH 8.0
- Ampicillin: prepared as 1000x stock in ddH_2O at 20 mg/ml concentration, filtered through 0.22 µm filter (Falcon) and stored at -20°C.
- DNA plasmids (for cloning procedures see section 2.4): stored in TE or ddH_20 at -20°C.

2.3.2. *DNA plasmid transformation of bacteria*

10 ng of plasmid DNA in a volume of 1 µl of ddH_2O was added to 25 µl of thawed chemically competent cells. The cells were heat-shocked at 45°C for 30 seconds and 250 µl of SOC media was added. The culture was incubated for at least 1 hour at 37°C with shaking (200 rpm) and plated on agar plates containing appropriate antibiotic. Next day at least one single colony was screened for the presence of the plasmid DNA by placing it in 5 ml LB media (with appropriate antibiotic) and incubating overnight with shaking (200 rpm). The following day, miniprep DNA extraction and purification was performed (Section 2.3.3).

2.3.3. *Qiagen plasmid extraction and purification*

For DNA plasmid extraction using QIAprep spin Miniprep kit, the 5 ml starter culture (Section 2.3.2) was centrifuged at 6000g for 3 minutes at room temperature. The supernatant was aspirated and the pellet was resuspended in 250 µl of P1 buffer. The resuspended cells were transferred into a microcentrifuge tube and 250 µl of P2 buffer was added. The solutions were mixed by inverting the tube several times. 350 µl of N3

buffer was added and mixed by inverting the tubes several times. The tubes were centrifuged for 10 minutes at 1300 rpm. The resulting supernatant was transferred (up to 800 µl) into a QIAprep spin column that was placed in a 2 ml collection tube. The column was centrifuged for 30 seconds at 13000 rpm. The flow through was discarded and 500 µl of buffer PB was added to the spin column. The column was centrifuged for 30 seconds at 1300 rpm and the flow through discarded. 750 µl of PE buffer was added and the column was centrifuged for 30 seconds at 13000 rpm. The flow through was discarded and the column centrifuged for 1 minute at 1300 to remove any residual buffer. The column was placed in a new DNase-free microcentrifuge tube and 30 µl of DNase-free water was added to the centre of the column. After 1 min incubation time the DNA was eluted from the column by 1 minute centrifugation at 13000 rpm. The purity and concentration of the extracted DNA was assessed using NanoDrop spectrophotometry (Thermo Fisher).

2.3.4. *Preparing plasmids for sequencing*

Components required for sequencing, included: 600 ng of plasmid DNA and 4 µM of the M13 forward and reverse primers (Invitrogen). The total volume of the plasmid and primer solution was adjusted with water to a total of 15 µl.

2.4. Molecular cloning

2.4.1. *Materials*

2.4.1.1. Reagents

- 50x TAE: 242 g Tris Base, 57.1 ml glacial acetic acid, 100 ml of 0.5 M EDTA pH 8.0 were brought to a total volume of 1 L with ddH_2O.
- 5x Loading buffer (Bioline).
- DNA molecular weight markers (Bioline): Hyperladder I.
- Restriction enzymes and buffers (New England Biolabs).
- Antarctic Phosphatase (New England Biolabs).
- T4 DNA ligase (New England Biolabs).
- Qiaquick gel extraction kit (Qiagen).
- 1000x SYBR Safe DNA gel stain (Invitrogen).
- UltraPure Agarose (Invitrogen)
- QIAquick Gel Extraction Kit (Qiagen)

2.4.1.2. Plasmids

pUC57.Promoter_DUX4

The pUC57 plasmid is a commercial vector (GenScript) that carried the synthesised full promoter sequence of the DUX4 promoter and 5'UTR sequence. The synthesised promoter sequence included flanking restriction enzyme sites for directional cloning of

the sequence into donor plasmids, i.e.: pC1.Promoter_DUX4.GFP.3'UTR_DUX4 and pAAV.Promoter_DUX4.DUX4 plasmid.

pC1.CMV.GFP.

The pC1.CMV.GFP.3'UTR_DUX4 plasmid was kindly donated by Dr Julie Dumonceaux, UCL, U.K.

pC1.Promoter_DUX4.GFP.

The pC1.Promoter_DUX4.GFP. was created by replacing the CMV promoter sequence from the pC1.CMV.GFP. vector backbone with the DUX4 promoter sequence derived from the pUC57.Promoter_DUX4.

pAAV.CMV.DUX4.

The pAAV.CMV.DUX4 was a kind donation from Dr Scott Harper, The Ohio State University College of Medicine, U.S.

pAAV.Promoter_DUX4.DUX4.

The pAAV.Promoter_DUX4.DUX4 was created by replacing the CMV promoter sequence from the pC1.CMV.GFP. vector backbone with the DUX4 promoter sequence derived from the pUC57.Promoter_DUX4.

2.4.2. *Restriction enzyme digest and gel electrophoresis*

For the diagnostic digest 1 µg of DNA product was subjected to a restriction enzyme reaction under conditions recommended by the manufacturer. 5x loading buffer was used to load the digested products onto an agarose gel. The 0.5-2% agarose gel was made with 1x TAE and 1x SYBR Safe to allow visualisation of the DNA bands with UV

light. The appropriate maker was chosen depending on the expected bands size. The loaded gel was electrophoresed at 60-100 V submerged in 1x TAE buffer.

2.4.3. *Antarctic phosphatase dephsphorylation*

Antarctic phosphatase (AnP) was used to catalyse the removal of phosphates from DNA 5';3' and blunt-ends extensions. Dephosphorylation was performed directly after the restriction digest reaction was completed by adding 5 U of AnP per 1 µg of the digested DNA. The reaction was incubated for 15 minutes at 37°C for blunt-ends and 5' strand extensions. Inactivation was performed for 5 minutes at 70°C. The resulting product was verified on 1x SYBR Safe-stained, 1% agarose gel and run at 60-100 V in 1x TAE buffer. The expected band was excised from the gel and purified Qiaquick gel extraction kit according to manufacturer's instructions.

2.4.4. *DNA ligation*

To catalyse the formation of a phosphodiester bond between juxtaposed 5' phosphate and 3' hydroxyl in duplex DNA, the T4 DNA ligase was used. Typically, 1:3 or 1:5 vector to insert molar ration was used for the ligation reaction. The reaction was incubated at 16°C overnight, following heat inactivation at 65°C for 10 minutes, and stored at 4°C ready for transfection.

2.4.5. *Cloning of GFP gene into the plasmid backbone (i.e., pC1.Promoter_DUX4)*

The DUX4 promoter sequence was excised form the donor pUC57 vector with the PciI and NheI restriction enzymes under reaction conditions suggested by the enzyme manufacturer (NEB). To create compatible restriction enzyme sites for directional cloning between the DUX4 promoter insert and the acceptor plasmid, the pC1.CMV.GFP

plasmid was subjected to the same digestion reaction as the pUC57 vector, so the CMV promoter sequence could be removed from the pC1.CMV.GFP vector backbone (Figure 2.1). Both reaction products were run on 1% agarose gel to separate the plasmid backbones from their corresponding promoter constructs. The bands corresponding to pC1.GFP. vector and DUX4 promoter sequences were excised from the gel and purified using gel purification kit (Qiagen).

The purity and concentration of the excised DNA fragments were assessed using nanodrop spectrophotometer (Thermo Fisher). The vector and insert were ligated using the T4 DNA ligase (NEB) using 1:3 and 1:5 vector to insert ratio. The ligated product was transformed into chemically competent TOP10 *E.coli* cells. The cells were streaked onto the ampicillin-containing LB agar selection plates. Following day, 4 positive colonies were selected and grown overnight in 1X ampicillin LB media shaking at 200 rpm. The plasmid DNA was purified from the suspended cell culture using miniprep kit according to the manufacturer's instructions (Qiagen). The quality and concentration of the purified plasmid was assessed with the nanodrop (Thermo Fisher). Diagnostic restriction enzyme digest was performed to ensure presence of the correct backbone and insert. The presence of the promoter insert was verified by sequencing (MWG) using primers listed in **table 2.2**. The sequencing reaction was prepared in barcoded tubes provided

by MWG. The reagent submitted for sequencing, included: 15 µl of 10 ng/ µl DNA plasmid and 2 µl primer of 10 µM concentration.

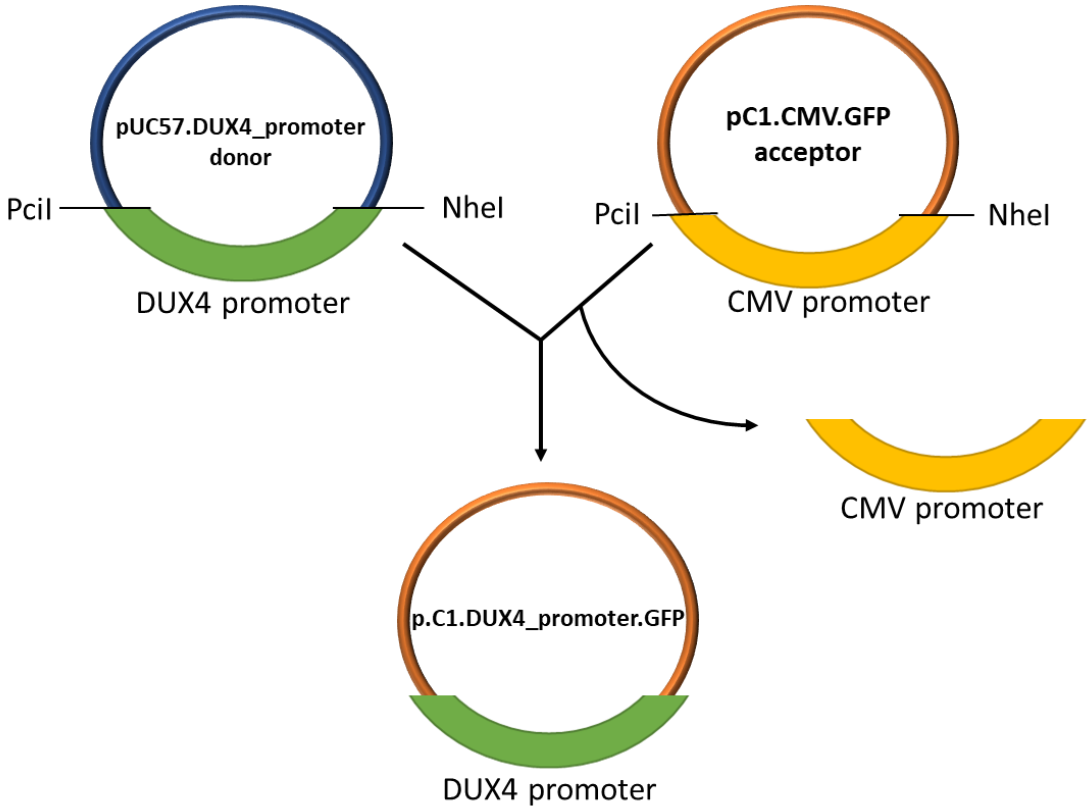


Figure 2.1 Simplified schematic showing cloning of the DUX4 promoter construct into the pC1.CMV.GFP. reporter plasmid

The DUX4 promoter sequence was excised from the donor pUC57 vector using PciI and Nhe-I restriction enzymes cleaving at the 5'- and 3'-end of the promoter construct. The CMV promoter sequence was removed from the acceptor pC1.GFP. plasmid using the same PciI and NheI- restriction enzymes that also cleaved the CMV sequence at the 5'- and 3'-end, respectively. This produced compatible sticky ends between the DUX4 promoter insert and pC1.GFP., which were utilised for directional cloning process.

2.4.6. *Cloning of the DUX4 promoter sequence into the plasmid backbone (i.e., pAAV.Promoter_DUX4.DUX4)*

The restriction enzymes used to digest the donor and acceptor plasmids were SpeI and Nhe-I (Figure 2.2). The cloning process for the pAAV.Promoter_DUX4.DUX4 plasmid was the same as for the pC1.Promoter_DUX4.GFP. (Section 2.4.5).

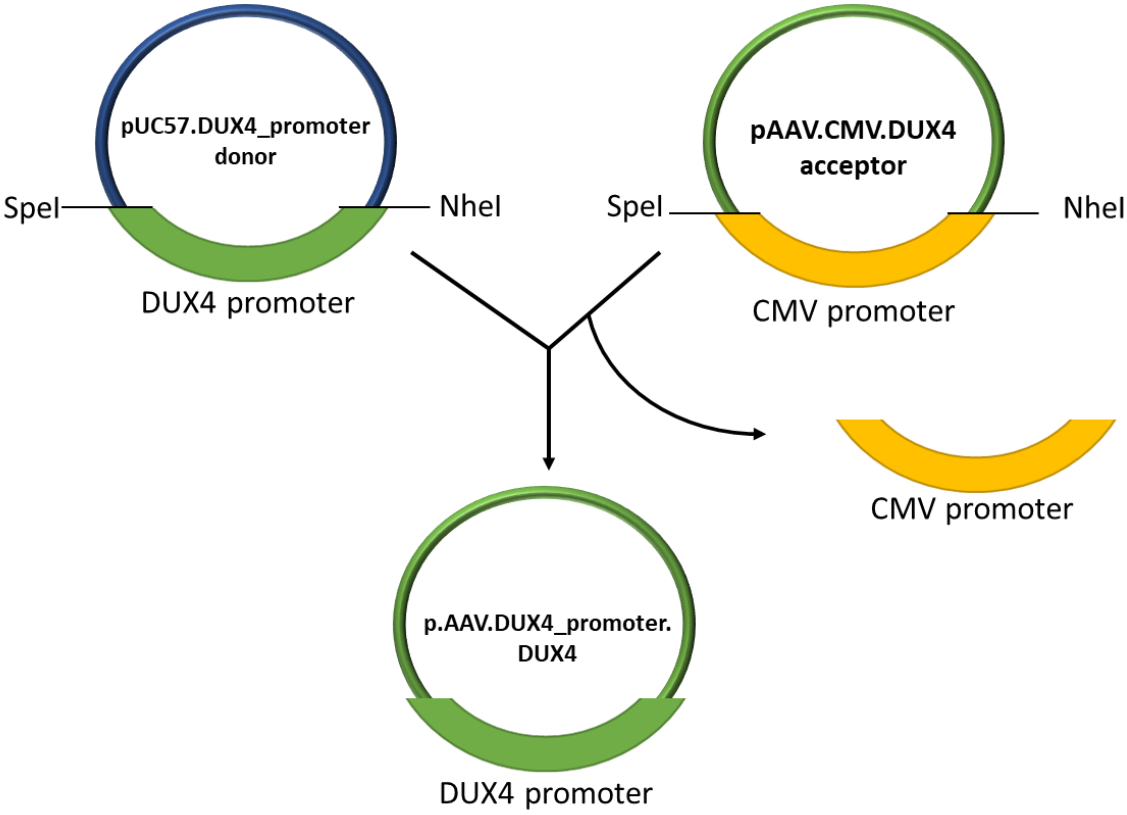


Figure 2.2 Simplified schematic showing cloning of the DUX4 promoter construct into the DUX4-expressing pAAV.CMV.DUX4 vector

The DUX4 promoter sequence was excised from the donor pUC57 vector using SpeI and Nhe-I restriction enzymes cleaving at the 5'- and 3'-end of the promoter construct. The CMV promoter sequence was removed from the acceptor pAAV.DUX4. plasmid using the same SpeI and NheI restriction enzymes that also cleaved the CMV sequence at the 5'- and 3'-end, respectively. This produced compatible sticky ends between the DUX4 promoter insert and pAAV.DUX4. plasmid, which were utilised for directional cloning process.

.

Table 2.2 Primer used in sequencing DUX4 promoter sequence from pC1 and pAAV plasmid backbones

Target sequence	Primer sequence	Tm	GC%
DUX4 promoter in pC1 plasmid	5'-GCCTATGGAAAAACGCCAGC-3'	60	55
DUX4 promoter in pAAV plasmid	5'-GTGAATTCCCTTCCGGGGTG-3'	61	60

2.5. Plasmid DNA transfection

2.5.1. *Materials*

- Lipofectamine 3000 (Invitrogen).

2.5.2. *Protocol*

$2.5x10^5$ Rhabdomyosarcoma cells (RD) (ATCC® CCL-136™) were seeded in 2 mL of RD DMEM growth media per well on 6-well plates. The cells were at least 80% confluent before transfection. An hour prior to transfection the overnight growth media was replaced with fresh RD DMEM growth media. Optimal DNA to Lipofectamine transfection ratio was 1 µg DNA to 3.75 µl Lipofectamine per well in total volume of 250 µl of serum and antibiotic free DMEM media. The DNA/Lipofectamine mixture was added dropwise to each well to make up the final volume of 2000 µl of RD DMEM growth media per well. 24 hours post-transfections, the cells were analysed under fluorescent microscope (Zeiss) and harvested for RNA extraction or flow cytometry (FACScan II) analysis.

2.6. Tissue culture techniques

2.6.1. *Materials*

- Dulbecco's Modified Eagle's Medium (DMEM): high glucose (4.5 g/l), GlutaMAXTM-I (862 mg/l), pyruvate (110 mg/l) and phenol red (15 mg/l) (ThermoFisher). Stored at 4°C.

- Medium 199. Stored at 4°C.

- Foetal bovine serum Serum (FBS), certified heat inactivated (Invitrogen). Stored at -20°C in 50 ml aliquots.

- Gentamicin (50 mg/ml) (Thermo Fisher).

- FSHD skeletal muscle growth medium: 4 vols of DMEM, 1 vol 199 medium, 20% FBS, 50 µg/ml gentamicin, 5 µg/ml insulin, 0.2 µg/ml dexamethasone, β-FGF , 5 ng/ml hEGF, 25 µg/ml fetuine. Stored at 4°C.

- FSHD skeletal muscle differentiation medium (PromoCell GmbH): 10 µg/ml insulin. Stored at 4°C.

- RD growth medium: DMEM supplemented with 10% FBS. Stored at 4°C.

- Methylthiazolyldiphenyl-tetrazolium bromide (MTT, 5mg/ml): 250 mg of MTT was dissolved in 50 ml of filtered using 0.22 µl filter (Falcon), stored at -20°C in the dark.

- Trypsin-EDTA (10x)- 0.5% trypsin, 0.2% EDTA in PBS. Stored at -20°C

- Zeiss Vert.A1 microscope with Axiocam 503 monocamera (Zeiss)

- Basic cell culturing equipment: 37°C, 5% CO_2 incubator; Class 2 microbiology safety cabinet; 37°C water bath, low speed centrifuge and haemocytometer.

2.6.2. *Cell lines*

- **Rhabdomyosarcom**a (RD) cells (ATCC® CCL-136™): Stored in vapour phase of liquid nitrogen at 500,000 cells/vial in growth media plus 10% DMSO.

- **DUX4 expressing FSHD patient immortalised myoblasts** (kindly made available by Dr Vincent Mouly, Institute of Myology, Paris). Stored in vapour phase or liquid nitrogen at 500,000 cells/vial in growth media plus 10% DMSO.

2.6.3. *Culture of human RD CCL-136 and FSHD immortalized myoblasts*

All work was performed using sterile plastic inside grade 2 laminar flow hood. Cells were incubated at 37°C with 5% CO_2.

$1x10^6$ RD and $0.5x10^6$ FSHD immortalised myoblast cells were seeded in 25 ml of their corresponding growth media per T175 flask. When 80% confluent, the media was aspirated and the cells were washed with sterile 5 ml of 1x PBS. PBS was replaced with 5 ml of 1x Trypsin/EDTA solution and incubated at 37°C for 2-3 minutes to detach the cells from the flask. 10 ml of growth media was added to neutralise the trypsin and the cell suspension was transferred into a 50 ml Falcon tube (Corning) and spun at 3000 rpm for 5 minutes. The resulting supernatant was aspirated and the pelleted cells were resuspended in their growth media to reach concentration of $1x10^6$ cells/ml. The resuspended cells were placed in a fresh T175 flask containing 25 ml of their corresponding media and incubated for a 3-4 days until 80% confluency was reached again.

2.6.4. *Freezing and thawing process of cells.*

To freeze the cells, $0.5x10^6$ cells suspended in their corresponding growth media were transferred into a 1 ml sterile cryovial. DMSO was added to a final concentration of 10% to make up freezing media. The cell-containing cryovials were placed into cryo-freezing container and placed in the -80°C freezer overnight. The cryovials were transferred into the vapour phase of a liquid nitrogen storage facility.

The cells were swiftly defrosted by placing cell-containing cryovial into 37°C water bath. As soon as they became defrosted, the cells were suspended in 9 ml of growth media and spun at 3000 rpm for 5 minutes. The supernatant was aspirated and the resulting pellet was resuspended in 1 ml of growth media. The cells were seeded in T75 flask containing 15 ml of growth media. When 80% confluency was reached, the cells were transferred into T175 flasks containing 25 ml of growth media.

2.6.5. *Differentiation of FSHD immortalized myoblast cells.*

$1x10^5$ of FSHD immortalized myoblast cells were seeded in 2 ml of their corresponding growth media per well on 6 well plates. When >95% confluency was reached, the growth media was replaced with the differentiation media. The cells were fully differentiated 4 days after the media was changed.

2.6.6. *Cell viability assessment*

For assessment of the cell viability MTT assay was used. To each well containing cells MTT was added evenly in 1:10 ratio (MTT:media). The treated cells were incubated for 4 hours at 37°C with 5% CO_2. Media was gently aspirated and plates were air dried at room temperature. 1 ml of DMSO was added per well and the plates were placed on a

shaker (200 rpm) to dissolve the crystals. 75 µl from each well was transferred into 96 well plate. The OD at 570 nm was recorded with the GloMax-Microplate Multidote Reader (Promega). The absorbance was normalised to blank containing DMSO (no MTT) and compared to untreated cells.

2.6.7. *Berberine treatment in cell culture*

Berberine chloride was dissolved in ddH_2O to produce stock solution of 1 mM. To increase solubility of berberine chloride, the solution was placed in the 37°C water bath and vortexed several times until fully dissolved.

The seeding process of the RD cells on 6-well plates was described in section 2.5. RD cells were treated with berberine chloride at the time of transfection (Section 2.5). The berberine concentration dose range used was between 0 and 100 µM in the final volume of 2000 µl of the fresh growth medium per well. The treated and untreated controls were harvested and analysed 24 hours after the exposure to the compound.

FSHD immortalizes myoblast cells were treated with berberine chloride on the second day of differentiation (Section 2.6.5). The concentration rage of berberine chloride used to tread the cells was 0-25 µM in the total volume of 2000 µl of the fresh differentiation medium per well. The cells were harvested for analysis 48 hours after treatment.

2.7. RNA extraction

2.7.1. *Materials*

- QIAshredder kit (Qiagen Ltd.): Stored at room temperature.
- RNeasy Mini Kit (Qiagen Ltd.): Stored at room temperature.

- RNase-free DNase-I (Qiagen Ltd.): Stored at 4°C. Once reconstituted, stored in 50 µl aliquots at -20°C.

- ND-1000 NanoDrop spectrophotometer (Thermo Fisher).

2.7.2. *RNA extraction protocol*

The RNA extraction was performed from cells seeded on 6 well plates using RNeasy mini kit and QiaShredder columns. The cell culture medium was carefully aspirated. 350 µl of RLT lysis buffer was added into each well. The buffer was carefully pipetted several times across the well to ensure complete lysis. The cell lysate was transferred into Qiashredder column and placed into a 2 ml collection tube. The lysate containing column was centrifuged for 2 minutes at full speed. The flow through could be stored at -80°C at this point,and defrosted at 37°C for 10 minutes when needed . 350 µl (1 volume) of 70% ethanol was added to the flow through and pipetted up and down several times to mix. 700 µl of the mixture was transferred to an RNeasy mini column placed in 2 ml collection tube. Closed tube was centrifuged for 30 seconds at ≥8,000xg. The flow through was discarded and 350 µl of buffer RW1 was added to the RNeasy column. The column was centrifuged for 30 seconds at ≥8,000xg and the flow through was discarded. 10 µl of reconstituted RNase-free DNase and 70 µl of RPE buffer were mixed together and carefully pipette the 80 µl on to the Rease column. The column was incubated at room temperature for 15 minutes. After incubation was completed, 350 µl of RW1 buffer was added to the column and centrifuged for 30 seconds at ≥8,000xg. The flow through was discarded and 500 µl of RLP buffer (with ethanol added as indicated on the label of the bottle) the column. The column was centrifuged for 30 seconds at ≥8,000xg. The flow though was discarded and another 500 µl of RLP buffer was added to the column. The

column was spun for 2 minutes at ≥8,000xg. The column was then placed into an RNase-free; DNase-free 1.5 ml Eppendorf tube and 30 µl of RNase-free; DNase-free water was added directly into the membrane of the column. The column was then spun for 1 minute at ≥8,000xg. Quality and concentration of the extracted RNA was assessed using ND-1000 NanoDrop spectrometer. Samples were stored in a -80°C freezer.

2.8. cDNA synthesis

2.8.1. *Materials*

- GoScript reverse transcription system kit (Promega): Stored at -20°C.
- Random primers (Invitrogen)
- Oligo(dT) primers (Promega)
- Thermal cycler (VWR)

2.8.2. *Protocol*

Fort the first strand cDNA synthesis, the GoScript kit (Promega) was used. Before setting up the reaction, all the components were thawed on ice except for the reverse transcriptase that was kept in the Labtop cooler to maintain temperature close to -20°C. 600 ng of RNA and 500 ng of both random primers and oligo(dT) were added to a RNase-free PCR tube. Total volume of RNA/primer mixture was adjusted to 10 µl with RNase-free water. To ensure resolution of any potential secondary structures formed by the RNA and effective primer annealing to the RNA, the reaction was placed in a PCR machine and a single cycle was run for 5 minutes at 70°C followed by at least 5 minutes

at 4°C. The resulting samples were stored on ice until the reverse transcription master mix was ready.

Reverse transcription reaction master mix was prepared on ice in the order listed in **table 2.3**

Table 2.3 Composition of the reverse transcription reaction master mix

Components	**Volume**
GoScript 5x Reaction Buffer	4 µl
$MgCl_2$ (25 mM)	2 µl
dNTP (5 mM)	2 µl
GoScript Reverrse Transcriptase	1 µl
Nuclease-Free water	6 µl
Final volume	**15 µl**

The 10 µl RNA/primer mix was combined with 15 µl reverse transcription mix and placed in the PCR heat block. The PCR program, included: annealing for 5 minutes at 25°C, extension at for 1 hour at 42°C and heat inactivation for 15 minutes at 70°C. Samples were kept at -20°C for long term storage.

2.9. Reverse transcriptase polymerase chain reaction (RT-PCR)

2.9.1. *Materials*

- cDNA generated using Method 3.13.
- Platinum Green Hot Start PCR Master Mix (Invitrogen). Stored at -20°C.

- Gene-specific primers and housekeeping primers designed using PrimerQuest Tool at https://eu.idtdna.com/PrimerQuest/Home/Index (IDT).
- Thermal cycler (WVR)

2.9.2. *RT-PCR amplification protocol*

To amplify GC-rich sequences such as *DUX4* mRNA, Platinium Green Hot Start PCR Master Mix (2x) kit was used. Primers used are listed in **table 2.4.**

Table 2.4 Primers used in RT-PCR

Target gene	Primer	Accession no.	Sequence (5'-3')	Location	Product size
DUX4	Forward:	HQ266760	AGGCGCAACCTCTCCTAGAAAC	Exon1	368/504 bp
	Reverse	HQ266761	TCCAGGAGATGTAACTCTAATCCA	Exon3	
B2M	Forward:	NM_004048.2	CTCTCTTTCTGGC-CTGGAGG	Exon1	67 bp
	Reverse	NM_004048.2	TGCTGGATGACGTGAGTAAACC	Exon2	

All the components of the PCR mix and primers were thawed and then combined on ice in order listed in **table 2.5**

Table 2.5 RT-PCR reaction set up

Component	Volume
Platinum Green Hot Start PCR 2x Master Mix	12.5 µl
Forward and reverse *DUX4-all* primers (10 µM each)	0.5 µl
Forward and reverse *B2M* primers (5 µM each)	0.5 µl
Platinum GC Enhancer (optional)	5 µl
Nuclease-Free water	2.5 µl
Final volume	**21 µl**

21 µl of the master mix reaction was added to each PCR tube. 4 µl of cDNA template was added to the reaction mix. At least one negative control was set up that lacked the cDNA, which was replaced with nuclease-free water. The PCR tubes were sealed, placed in the PCR machine and programmed to run the single denaturation step for 2 minutes at 92°C followed by 30 cycles of PCR amplification (i.e., denaturation for 1 minute at 94°C, annealing for 1 minute at 55°C and extension for 45 seconds at 72°C), and one final cycle of extension for 5 minutes at 72°C. The PCR program was ended by putting the reaction on hold at 4°C until ready for analysis on agarose gel electrophoresis. Otherwise, samples were stored at -20°C.

2.10. Semi-quantitative densitometric analysis of gene expression

2.10.1. *Materials*

- Agarose. Stored at room temperature.
- 10x TAE buffer: diluted to a 1X solution. containing 40 mM Tris, 40 mM acetate, and 1 mM EDTA, pH ~8.3. Store at room temperature.
- SYBR Safe (x10,000) DNA gel stain (Invitrogen). Store at room temperature.
- DNA Hyperladder V (Bioline UK Ltd.). Stored at 4°C.
- Horizontal electrophoresis system (Bio-Rad).
- E-BOX VX2 gel documentation system (PeqLab).
- Access to GeneTools software (Syngene) for densitometric analysis.

2.10.2. *Protocol*

To examine expression of target and housekeeping genes, the specific PCR products were separated by agarose gel electrophoresis. For the detection of the DUX4 mRNA detection, the PCR products were run on a horizontal 3% agarose. The gel was prepared by dissolving 3 g of agarose in 100 ml of 1x TAE buffer and heating until transparent, crystal-free solution was formed. When the gel was cooled to 60°C, 15 µl of 1000x SYBR safe was added and mixed into the gel solution. The gel was poured into a casting tray and an appropriate comb was inserted. The gel was allowed to set for 20 minutes at room temperature. The comb was removed and the gel in the tray was placed in the gel tank. The gel was submerged with 1x TAE to around 1 cm in depth. Since the PCR reaction mix contains the loading dye, each sample was directly loaded into each well at consistent volume. 7 µl of Hyperladder V was used as a verification marker of PCR product size and loaded at each side of the gel. The loaded gel was run at 90 volts for 45 minutes and visualised under UV light using Ebox VX2 imaging system. The semi-quantitative analysis of the DUX4 mRNA expression levels was done by normalising its expression to the levels of the house keeping gene (i.e., B2M) expression. The analysis was performed using GeneTools software. The formula below was used to calculate expression of DUX4 relative to B2M:

$$DUX4 \text{ expression } = \frac{(DUX4 - \text{background})}{(B2M - \text{background})} \quad \text{(Eq.1)}$$

2.11. Quantitative RT-PCR (RT-qPCR)

2.11.1. *Materials*

- cDNA generated using Method 3.13.

- Housekeeping realtime primers and gene-specific primers (Primetime qPCR primers from IDT) or designed using PrimerQuest software (IDT).
- 3LightCycler 488 SYBR Green I Master (Roche). Stored at -20°C.
- LightCycler 480 (Roche) for quantitative PCR

2.11.2. ***Protocol***

Very low levels in human muscle and extremely GC-rich sequence of DUX4 (Snider et al., 2010), made the direct detection of the DUX4 with qPCR extremely challenging. C2C12 mouse myoblast cells transfected with DUX4 show to deregulate a range of the downstream genes (Krom et al. 2013). Expression profile of the genes downstream of DUX4 provided a bench standard allowing for indirect measurement of the DUX4 expression.

The described protocol has been optimised for use with the ROCHE SyBr Green master-mix and the ROCHE LightCycler480 machine using a 386 well plate. Sequences of primers used for RT-qPCR are listed in **table 2.6**. All the procedures were carried out in a UV-treated, nucleic acid-free safety cabinet.

Table 2.6 Primers used in RT-qPCR

Target gene	Accession no	Primer	Sequence (5'-3')	Location	Product size
B2M	NM_004048.2	Forward Reverse	CTCTCTTTCTGGCCTGGAGG TGCTGGATGACGTGAGTAAACC	Exon1 Exon2	67 bp
MBD3L2	NM_144614.3	Forward Reverse	CGTTCACCTCTTTTCCAAGC AGTCTCATGGGGAGAGCAGA	Exon1 Exon2	142 bp
TRIM43	NM_138800.1	Forward Reverse	ACCCATCACTGGACTGGTGT CACATCCTCAAAGAGCCTGA	Exon6 Exon7	100 bp
ZSCAN4	NM_152677.2	Forward Reverse	CTGGAGCAGTTTATGATTGG AGCTTCCTGTCCCTGCATGT	Exon3 Exon4	162 bp

The synthesised cDNA (Section 2.8) was diluted in nuclease free water. The cDNA from untreated samples were used to set up standard serial dilutions (i.e., 1:50, 1:500, 1:5,000 and 1:50,000) for calculation of primer binding efficiency. The final volume of each dilution was at least 50 µl. A qPCR plate set up by including the four standard serial dilution standards and one negative control that had nuclease-free water instead of cDNA. In each qPCR run at least three biological replicates of treated and untreated samples were included. Each of the biological replicates had 3 technical replicates to ensure consistent reading.

The master mix for each of the DUX4 downstream genes and corresponding housekeeping gene was set up as outlined in **table 2.7**, excluding the cDNA. 6 µl of the master mix was added to its corresponding wells, followed by addition of 4 µl of cDNA.

Table 2.7 Composition of RT-qPCR reaction

Components	Initial concentration	Volume	Final cincetration
SYBR Green	2x	5 µl	1x
Primer concentration	10 µM	0.5 µl	0.5 µM/primer
cDNA	n.a.	4 µl	n.a.
Nuclease-free water	n.a.	0.5 µl	n.a.
Fianl volume/well	n.a.	10 µl	n.a.

The plate was sealed using the provided cover and spun briefly to ensure all the reagent have reached the bottom. The plate was placed into the machine and the programme outlined in **table 2.8** was initiated. The melting curve programme was enabled to asses any off target primer binding (Figure 2.3). When there was no off target primer binding or contamination present in the sample, a single peak was present for each primer set (Figure 2.3).

Table 2.8 RT-qPCR cycling parameters

PCR programme	No. of cycles	Temp. (°C)	Time
Preincubation	1	94	5 minutes
Amplification	45	94	10 seconds
		60	10 seconds
		72	10 seconds

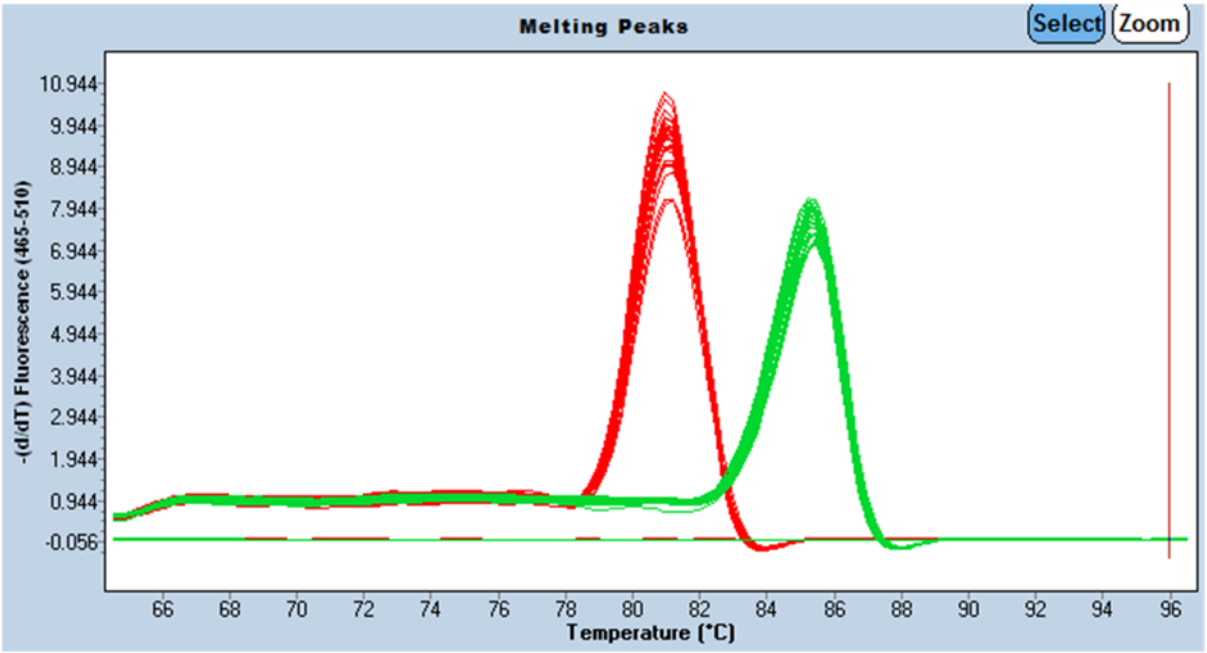


Figure 2.3 Example of melting peaks acquired from the Roche LightCycler480

Two distinct peaks specific for ZSCAN4 (green) and B2M (red) housekeeping gene indicate that the primers are amplifying a single product and no off-target amplification or sample contamination is present.

2.12. Immunohistochemistry

2.12.1. *Materials*

- Extracellular matrix (ECM) gel from Engelbreth-Holm-Swarm murine sarcoma (SigmaAldrich). Diluted with DMED to 1 mg/ml and stored in aliquots at -80°C.
- Triton X (SigmaAldrich). Stored at room temperature.

- Goat serum (SigmaAldrich). Stored in aliquots at -20°C.
- 4% paraformaldehyde (SigmaAldrich) in 1xPBS pH 7.0. Stored in aliquots at-20°C.
- Tween-20 (SigmaAldrich). Stored at -20°C.
- DAPI (SigmaAldrich) in ddH_2O at 20 mg/ml. Stored at -20°C.
- Odyssey CLx Imaging System (Licor)
- Primary antibodies:
    - Mouse anti MF20 (DSHB, MF20). Stored in aliquots at -20°C
    - Rabbit anti-DUX4 (Abcam, E5.5). Stored in aliquots at -20°C
- Secondary antibodies:
    - Goat anti-mouse (Licor, IRDye 680RD). Stored in aliquots at -20°C
    - Goat anti-rabbit (Licor, IRDye 680RD). Stored in aliquots at -20°C
    - Goat anti-rabbit (ThermoFisher, AlexaFluor488). Stored in aliquots at -20°C

2.12.2. *In-cell Western of FSHD cell culture*

6-well plates were coated with 500 µl of ECM (1 mg/ml) to ensure cell adhesion. The seeding of the FSHD patient myoblast, differentiation and berberine treatment were described in sections 2.6.5 and 2.6.7, respectively. At the end of incubation period, the culture medium was removed and cells rinsed with 1xPBS, followed by fixing in ice cold 4%PFA for 15 minutes at room temperature. Cells were washed three times with 1xPBS for 5 minutes. Permeabilization of cell was performed in 0.5% Triton X-100 in 1xPBS for 10 minutes at room temperature. 10% goat serum in 1xPBS was used to perform 1 hour blocking. The primary antibodies were diluted 1:100 in 5% goat serum, 1xPBS and added

to each well for overnight incubation at 4°C. Each well was washed three times for 5 minutes with 1xPBST (0.1% Tween-20). Secondary antibody was diluted 1:800 in 5% goat serum, 1xPBS, Cells were incubated for 1 hour with the diluted secondary antibodies. 1xPBST washes were preformed three times for 5 minutes. After the final wash, any remaining solution was removed from the wells completely. The plates were used for the analysis using Odyssey CLx Imaging System. Plates can be stored if 1xPBS was added to each well and kept at 4°C.

2.12.3. *Fusion index of FSHD patient cells*

6-well plates were coated with 500 µl of ECM (1 mg/ml) to ensure cell adhesion. The seeding of the FSHD patient myoblast, differentiation and berberine treatment were described in sections 2.6.5 and 2.6.7, respectively. The cell culture medium was removed and cells were washed with 1xPBS once. The cells were fixed with ice cold 4% PFA for 10 min at room temperature. The fixative was washed away by rising once in 1xPBS. The cells were permeabilised in 0.5% Triton X-100, 1xPBS for 10 minutes at room temperature. Blocking was performed with 10% goat serum in 1xPBS for 1 hour at room temperature. The mouse anti-MF30 primary antibody was diluted 1:1000 in 5% goat serum, 1xPBS. The primary antibody incubation was performed overnight at 4°C. Next day cells were washed in 1xPBS three times for 5 minutes. The goat anti-mouse , AlexaFluor488 secondary antibody was diluted 1:200 in 5% goat serum, 1xPBS. The secondary antibody incubation was performed for 1 hour at room temperature. The cells were washed three times for 5 minutes with 1xPBS. DAPI stock was diluted 1:100 in 1xPBS, added to each well and incubated at room temperature for 10 minutes. The cells were rinsed twice in 1xPBS and kept in 1xPBS at 4°C protected from

light. Fluorescent microscope at 10x magnification was used to acquire images for analysis. The fusion index was calculated by counting the number if nuclei in MF20 - positive myotubes containing 3 or more nuclei, and expressed as a percentage of total number of nuclei present in the captured field.

2.13. Flow cytometry

2.13.1. *Materials*

- Paraformaldehyde (PFA). 4% PFA in 1x PBS. Stored at -20°C.
- 5 ml snap cap round bottom tubes (Falcon)
- FACSCantoII flow cytometer (BD Bioscience)
- FACSDiva Software (BD Bioscience)

2.13.2. *Analysis of GFP expression in transfected RD CCL-136 cells by flow cytometry*

RD cells were trypsinised 24 hours post-transfection/berberine treatment (Section 2.5 describing transfection; Section 2.6.7 describing the berberine treatment), transferred into 5 ml snap cap round bottom tubes and pelleted by centrifugation at 3000 rpm for 3 minutes. The supernatant was aspirated and the cells were resuspended in 300 µl of 4% PFA. A mock transduced (i.e., no plasmid) control population was also included in every experiment analysing GFP expression.

To discriminate debris from the viable cell population FSC-A/SSC plot was selected and the P1 gate was drawn over the cells (Figure 2.4.A). Single viable cell population were selected from doublets by selecting them using polygonal on the FSC-H/FSH-A plot (Figure 2.4.B). The GPF positive cell population was displayed in a histogram using FITC channel. The mock population was analysed first to set the gate of false GFP positive to <1%. At least 10 000 cells were analysed for each condition.

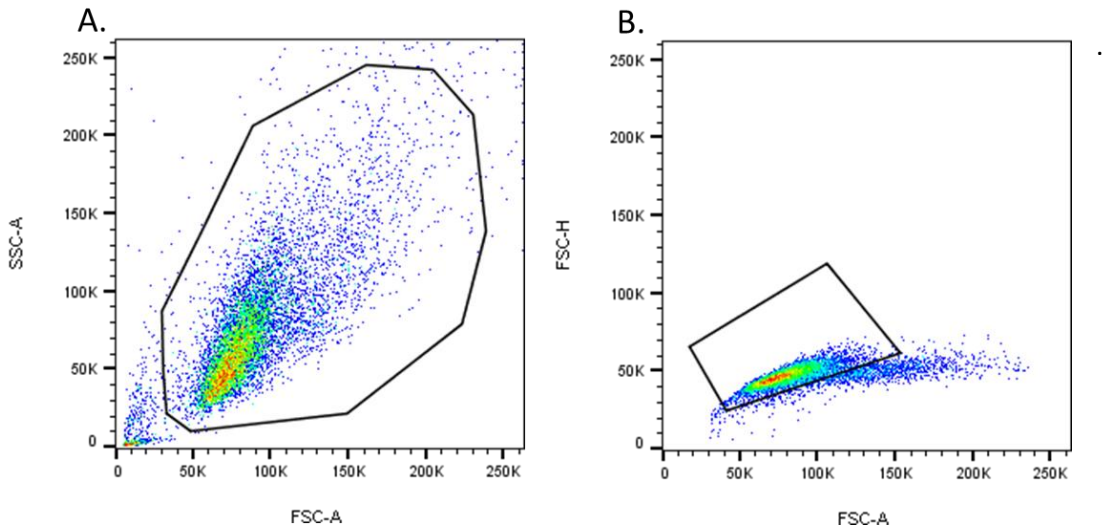


Figure 2.4 Example of gating to select viable (A.) and single (B.) in flow cytometric analysis (**A.**) Forward scatter-area vs. side scatter-area plots allowed for selection of viable RD cells as shown by the polygonal gate. Cells outside the gate were consider debris. (**B.**) Forward scatter- area vs. forwards scatter-height was subsequently selected in order to select single viable cells for the analysis. Cells outside the polygonal gate are considered to be doublets.

2.14. Spectroscopic analysis of nucleic acid secondary structures and ligand binding

2.14.1. *Materials*

- KP buffer: 100 mM KCl, 10 mM K_2HPO_4/KH_2PO_4, pH 7.0, ddH_2O, filtered using 0.22 µm filter (Flacon).

- Tris buffer: 10 mM Tris-HCl, 100 mM KCl, pH 7.0, ddH_20, filtered using 0.22 µm filter (Flacon).
- Chirascan qCD, circular dichroism spectrometer (Applied Photophysics Ltd).
- Nuclear magnetic resonance spectrometer (800 MHz; Bruker Avance III).
- Fluorescent spectrometer (Perkin Elmer).
- UV/Vis spectrophotometer (Jenway 7305).
- 1 mM Berberine chloride stock: berberine was dissolved in either KP or Tris buffer and heated at 37°C until fully dissolved. Stored at room temperature.
- DNA and RNA oligonucleotides (IDT) (Table 2.9).

Table 2.9 List of oligonucleotides used in spectroscopic analysis

Oligonucleotides synthesised at 0.2 µm and HPLC purified by IDT. Underlined guanine residues are predicted to form putative GQs. Adenosine residues in bold indicate guanine substitution to disrupt GQ and hairpin structures in M1-3M2DGCb_GQ, oligonucleotides. Abbreviations: DMEM_GQ, DUX4 myogenic enhancer 1 G-quadruplex; DGCb_GQ, DUX4 GC box; M1DGCb_GQ, mutant 1 DUX4 GC box; CSS, cryptic splice site; E1, enhancer 1; SS1, splice site 1; H.telomere, Human telomere.

Construct name:	Construct sequence:	Amount (µmol)	Purification method
DME1_GQ	CAG**GGG**ATGGT**GGG**GCTG**GGG**TTGAGTGAT**GGG**C	0.2	**HPLC**
D4P_ GQ	CGGGGTGGGGCGGGCTGTCCCAGGGGGGCT	0.2	**HPLC**
M1DGCb_GQ (NO GQ)	CGA**AAA**TGGGGCGGGCTGTCCC**AAAAAA**CT	0.2	**HPLC**
M1DGCb_GQ (NO GQ)	CGA**AAA**TGGGGCGGGCTGTCCC**GGGGGG**CT		
M3DGCbQ (no Hairpin)	CGGGGTGGGGCGGGCTGT**AAA**AGGGGGGCT	0.2	**HPLC**
CSS_GQ	A**GGG**CCAGGCACCC**GGG**ACA**GGG**UGGCA**GGG**C	0.2	**HPLC**
E1_GQ	AGGGGAGUCCGUGGUGGGGCUGGGGCCGGGGU	0.2	**HPLC**
SS1_GQ	C**GGG**GUU**GGG**ACG**GGG**UC**GGG**U	0.2	**HPLC**
H. telomere_GQ	A**GGG**TTA**GGG**TTA**GGG**TTA**GGG**	0.2	**HPLC**

2.14.2. *Circular dichroism and nuclear magnetic resonance spectroscopy*

DNA and RNA oligonucleotides for circular dichroism (CD) and nuclear magnetic resonance (NMR) were prepared from the HPLC purified samples by resuspending them in KCl buffer to a final concentration of 2–4 μM.

CD spectra were acquired using a Chirascan qCD spectrophotometer (Applied Photophysics Ltd), equipped with a LTD6G circulating water bath (Grant Instruments, UK) and thermoelectric temperature controller (Melcor, USA). Samples were heated in the cell to 95^{o}C for a total period of 15 minutes, samples were then annealed by allowing to cool to room temperature for a minimum period of 4 hours. CD spectra were recorded over a wavelength range of 215– 340 nm using a 1 cm path length strain-free quartz cuvette and at the temperatures indicated. Data points recorded at 1 nm intervals. A bandwidth of 3 nm was used and 5000 counts acquired at each point with adaptive sampling enabled. Each trace is shown as the mean of three scans (±SD). CD temperature ramps were acquired at 265 nm corresponding to the band maxima of the folded quadruplex species. Ranges between 5 and 99^{o}C were used, with points acquired at 0.5^{o}C intervals with a 120–180 s timestep between 0.5^{o}C increments. Points were acquired with 10 000 counts and adaptive sampling enabled.

CD was recorded in units of absorbance known as ellipticity [θ]. Data was normalised to molar concentration of the repeating unit, where the repeating unit represents number of bases in each of the DNA/RNA oligonucleotide. The mean residue weight (MRW) was calculated from the molecular weight (MW) of each oligonucleotide divided by the number of bases (N) minus 1 (i.e., [MRW=MW/N-1]). To calculate the molar residual ellipticity (M.R.E) the following equation was applied:

$$M.R.E = \frac{\theta * MRW}{10 * d * c} \ [deg.cm^2.dmol^{-1}] \qquad \text{(Eq. 2)}$$

Where:

θ= observed molar ellipticity
d= path length in cm
c= concentration in g/L

The molar residue ellipticity equation (Eq. 2) was used to normalise the molar ellipticity spectrum in order for the measured value to be independent of the nucleic acid polymer length (Ishtikhar et al. 2014).

NMR spectra (^{1}H) were collected at 800 MHz using a Bruker Avance III spectrometer with a triple resonance cryoprobe. Standard Bruker acquisition parameters were used. Data were collected using Topspin (v. 3.0) and processed in CCPN Analysis (v. 2.1).

2.14.3. *UV/Vis and fluorescence spectroscopy binding and analytic techniques*

DNA and RNA oligonucleotides for UV/Vis spectrophotometric and fluorescent spectrometric analysis stock solutions were prepared from the HPLC purified samples by resuspending them in Tris buffer to a final concentration of 500 µM. To anneal the oligonucleotides, the samples were placed on a heat block at 950 for 10 minutes and annealed by allowing to cool overnight to room temperature.

Annealed oligonucleotides were titrated in 0-10 µM range in both UV/Vis and florescent spectrometry readouts. The berberine concertation was kept constant at 10 and 5 µM in UV/Vis and fluorescent spectroscopic analysis, respectively. Tris buffer was used as a blank. The UV/Vis spectrophotometer was set to record spectra from 300-550nm. Both types of spectra were recorded at room temperature. Fluorescent spectra were measured at $\lambda_{ex}/\lambda_{em}$=355/530 nm. To measure the binding constant (Ka) data was

plotted into hyperbolic function using KaleidaGraph software following the equation below:

$$\Delta F = \left(\frac{\Delta Fmax}{2[L]_0}\right)\{\left([L]_0 + [Q] + \frac{1}{K_a}\right) - \sqrt{\left([L]_0 + [\mathrm{Q}] + \frac{1}{K_a}\right)^2 - 4[L]_0[Q]}\} \quad \text{(Eq. 3)}$$

Where:

ΔF=F-F_0 and ΔFmax= Fmax-F_0.

F_0 and F- initial and subsequent fluorescent intensities

$[L]_0$-berberine concertation

[Q]- oligonucleotide concertation.

Ka-binding constant

The Eq. 3 is a quadratic velocity equation for tight-biding substrates to determine the affinity binding between berberine and quadruplex. The Eq. 3 was previously used to determine the binding affinity of berberine to the human telomeric DNA quadruplex structure by utilising the fluorescent spectroscopic analysis (Arora et al. 2008).

2.15. Bioinformatics

2.15.1. *Online resources*

- NCBI GenBank (www.ncbi.nlm.nih.gov/genbank/)
    - *Genomic sequences*
- Ensembl (www.ensemble.org)
    - *Transcript sequences*
- *DUX4* secondary structure was predicted with mFold software:
    - *Duplex DNA and RNA formation prediction*
- QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/index.php)

- *DNA and RNA GQ formation prediction*

2.15.1. ***DNA and RNA sequences***

The DUX4 sequences enhance and promoter sequences were acquired from the GenBank database (gene ID: AF117653). The DUX4 transcript isoform sequence were taken from the Ensembl database with their ID outlined in **table 2.10**. Sequences were exported in FASTA format form both databases and used to be directly inputted into a web server of software of interest.

Table 2.10 Details of GenBank and Ensemble entries used

DUX4 gene feature	Database	ID
Enhancer and promoter	GenBank	AF117653
Transcript 1	Ensembl	ENST00000565211.1
Transcript 2	Ensembl	ENST00000569241.5
Transcript 3	Ensembl	ENST00000616166.1

2.15.1. ***MFold***

The RNA folding form was set to default conditions. Folding temperature was at 37°C, in 1 M NaCl. Maximum loop sizes were set to 30 bases and there was no limit on the distance between base pairs. The inputted sequences were no longer that 40 bases. The output structures were saved as pdf for later analysis.

2.15.2. ***QGRS Mapper***

The data could be submitted in the input box as raw FASTA nucleotide sequence, by providing the NCBI gene ID or Ensemble transcript ID. The GQ max length was set to 30

base and loop size to 36 bases. The output view was provided as a FASTA sequence view with highlighted potential GQ-forming tetrads. The JavaScript enable internet browser (e.g., Internet Explorer) allowed for graphical view of putative GQs formation. GQ predicted sequences with a G-score of above 30 were considered for further analysis.

2.16. Statistical analysis

All statistical analysis was performed using GraphPad Prism 6 software.

The data from the RT-PCR, RT-qPCR flow cytometry and spectrophotometry was analysed using one-way Anova unless stated otherwise.

3. Analysis and determination of GQ formation within DUX4 enhancer, promoter and transcript sequences

3.1. Introduction

3.1.1. *Computational prediction of GQs*

Computational analysis investigating GQs permits a thorough genome wide analysis of mammalian genes. These methods already allowed detailed surveillance of the GQ structures in the human genome (Huppert and Balasubramanian, 2006). However, a user friendly software that is available to a wider public has not existed until the release of QGRS mapper (Kikin et al. 2006). Particular focus of this software was to analyse alternatively processed mammalian pre-mRNA sequence of human and mouse genes (Kostadinov et al. 2006). As a result it has been found that GQ motifs are particularly enriched at or near alternative splice sites and poly(A) regions (Kostadinov et al. 2006). Furthermore, the QGRS Mapper provides comprehensive information on the configuration and distribution of potential GQ forming sequences that can be either manually submitted or retrieved from the NCBI databases through options provided in the software (Kikin et al. 2006). One of the core strengths of the software is the fact that it is able to analyse mammalian pre-mRNA sequences that have been alternatively processed i.e., have different splice isoforms or have been alternatively polyadenylated. The software provides flexibility in allowing the user to define the parameters of a number of tetrads, length of the quadruplex structure and content as well as size of the loops (Kikin et al. 2006). Selected genomic sequences such as enhancers, promoters,

telomeric regions as well as RNA sequences can be submitted in FASTA format for mapping unimolecular GQ motifs. In addition, QGRS Mapper is a very useful tool in predicting oligonucleotide structures. To retrieve desired genomic or transcript structures for the analysis, the software provides direct access to the NCBI Gene Entrez, GenBank and RefSeq databases. The output is represented in the form of an interactive graphic that can be also viewed in the form of a raw sequences and visualised distribution of the quadruplex motifs across all alternative RNA transcripts of a gene. Therefore, the QGRS Mapper allows functional analysis of the GQ motifs that might be involved in the alternative processing of a gene. The basic motif that QGRS uses to predict putative GQs is the following:

$$\mathbf{G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x}$$

Where **x**= number of guanine tetrads in the GQ and $\mathbf{y_1}$, $\mathbf{y_2}$, $\mathbf{y_3}$,= length of the loops connecting the guanine tetrads (Kikin et al. 2006).

3.1.2. *Circular dichroism and nuclear magnetic resonance as experimental methods to study GQ structures and ligands*

Phenomenon of circular dichroism (CD) arises when chiral molecules interact with circularly polarized electromagnetic rays (Woody, 1995). In biological research, the CD has proven to be particularly useful in determining secondary structures of DNA and/or RNA due to their asymmetric sugar backbone and the helical structure resulting from a specific arrangement of its constituent nucleobases. Typical CD analysis uses UV light within 200-320 nm range of the spectrum that detects electronic transition of DNA (or RNA) bases (Gray et al. 1995). Although the resulting CD spectra of the nucleic acid chemistries can be calculated using complex quantum mechanical methods, in practice

the spectra patterns are interpreted empirically (Vorlickova et al. 2012). CD spectroscopy is a convenient method to analyse conformational changes in DNA structure that might be affected by environmental cues, such as pH, temperature, counter ions concentration or presence of crowding agents.

Since the discovery of GQ motifs in the human genome and the fact that these structures might play an important biological role in processes related to aging and cancer development, extensive effort has gone into understanding of the biological role of these guanine-rich nucleic acids and potential molecules that regulate their stability. One of the most widely used techniques to interrogate the conformational changes and ligand interactions within GQs is CD. Three major DNA GQ conformations have been described using CD, such as parallel, antiparallel or mixed (Gray et al. 2008). Since RNA molecules lack the anti and possess only the syn glycosidic bond geometry, they form only parallel quadruplexes. Different types of quadruplex folding can be distinguished by characteristic pattern of CD spectra. The parallel GQs can be distinguished by the pronounced positive 260 nm band, while a positive band at 295 nm and a negative band at 260 confirm presence of an antiparallel quadruplex topology (Karsisiotis et al. 2011). Currently CD spectral patterns cannot be used to unambiguously determine the exact topology of the GQ structure. Since the CD analysis limits the number of possible structures that can be distinguished, defining exact GQ topology should also be supported by nuclear magnetic resonance (NMR) or X-ray crystallography data and techniques.

NMR spectroscopy is a powerful tool that allows atomic resolution of GQ structures (Campbell and Parkinson, 2007). In addition, NMR is the only technique providing such

a high resolution that also allows structural analysis and ligand interaction in solution. Presence of the GQ structure can be determined by detection of the hydrogen bonds formed between guanines (Hoogsteen base pairing) within the G-tetrad. These characteristic guanine imino protons (^{1}H) show a specific chemical shift in the region of 10-12 nm of the spectrum, whereas the typical Watson-Crick base pairing can be detected within the range of 13-14 nm (Webba da Silva 2007). NMR has proven to be extremely useful in determining kinetic and dynamics of the interaction of GQ structures with their specific ligands, especially in the context of telomeric and oncogene promoter GQs. Furthermore, NMR can be also used to characterise 3D structures of an individual GQ as well as GQ-ligand complexes (Haider et al. 2002). The fact that the structure is investigated in solution serves as an advantage over the static crystallography approach.

CD and NMR are currently considered be state-of-the-art scientific techniques that tremendously increased our understanding of GQ structure topologies as well as allowing us to test an extensive array of GQ interacting ligands that might serve as a potential treatment for many devastating diseases, including cancer (Balasubramanian et al. 2011).

3.1.3. *Berberine binding affinity for RNA and DNA secondary structures measured with UV-Vis and fluorescent spectroscopy*

Photochemical properties of berberine were widely utilised to study its binding affinity to different types of nucleic acid secondary structures, including double stranded DNA, tRNA, GQs and triple helical nucleic acids (Basu et al. 2013; Nandi et al. 1990; Park et al. 2004; Zhang et al. 2007). Berberine is a planar molecule with a large aromatic surface

that contains an extensive π-delocalised system and partial positive charge on N7 (Figure 3.1) (Grycová et al. 2007).

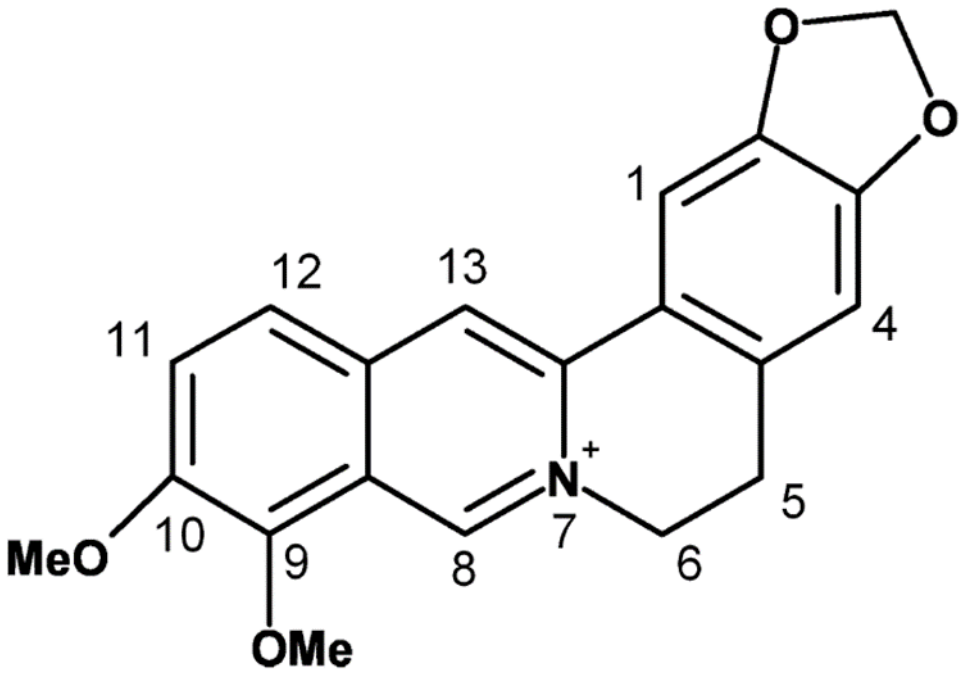


Figure 3.1 Chemical structure of berberine

These chemical characteristic properties of berberine result in a specific absorbance pattern of the molecule. The absorption maximum of berberine has two characteristic peaks at 340 and 420 nm that change when the compound is bound to nucleic acid sequences (Arora et al. 2008; Li et al. 2017). Generally, the red shift in absorption peak indicates end stacking for GQs or intercalative properties for duplex DNA or RNA. The hypochromic shifts indicate reduction of the berberine absorption due to its binding to DNA/RNA. Furthermore, when the hypochromicity occurs without any red shift, it indicates groove binding rather the above-mentioned end stacking or intercalating interactions of berberine to the nuclei acid.

In addition, berberine shows significant fluorescent emission enhancement at 525 nm when bound to DNA/RNA and excited at 355 nm wavelength, while being essentially non-fluorescent while unbound. By utilising the UV-Vis and fluorescence spectroscopy techniques berberine binding affinity and recognition of nucleic acid secondary structures can be studied.

3.1.4. *Objectives of the chapter*

The study of *DUX4*-related GQs was initiated by *in silico* analysis of the DUX4 myogenic enhancer 1 and 2, promoter and transcript elements to determine presence of the putative motifs within the analysed sequences. The candidate quadruplex-forming sequences were then subjected to the CD and NMR spectroscopy to determine their secondary structure topology in solution. Finally, the binding properties of berberine to the selected DNA and RNA oligonucleotide was assessed by the UV-Vis and fluorescence spectroscopy methods.

3.2. Results

3.2.1. *Bioinformatic analysis of GQ formation within the enhancer, promoter and transcript of DUX4*

The full promoter and enhancer sequences were manually selected and submitted into the QGRS Mapper software in FASTA format (Figure 3.2 and 3.3). The output result for the DUX4 myogenic enhancer 1 and 2 (DME1 and 2) has predicted 11 and 1 GQ motifs, respectively. Out of those predicted motifs only 1 has scored the value of ≥30 G-Score in the DME1 sequence, which is considered to likely form GQs *in vitro* (Prof M. Searle personal communication). The QGRS Mapper analysis of the DUX4 promoter (D4P) sequence has found 4 putative GQ motifs, from which only 1 has a G-Score of ≥30.

The whole DUX4 transcript sequence (ENST00000569241.5) contains 30 putative GQ motifs according to the GQRS Mapper analysis (Figure 3.2.3). However, only 4 sequences have been predicted to have a G-Score of ≥30.

All of the predicted sequences with the score of ≥30 were mapped to their corresponding positions within enhancer, promoter or transcript sequence regions as shown in Figure 3.4. From the schematic, it can be deduced that GQs with the G-Score ≥30 cluster particularly with the coding sequence (CDS) and exon/intron boundaries of the transcript, indicating their possible biologically relevant role. The overlap of the GQ at the cryptic splice site (CSS), suggest a role in alternative splicing. There were no GQ forming sequences predicted within the exon 3 where the polyadenylation signal resides.

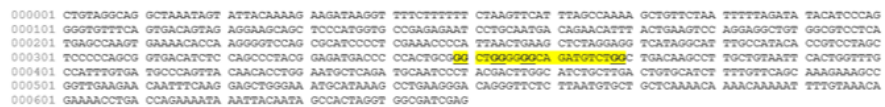
A particularly strongly scoring GQ (G-score: 36) motif has been also identified in the 5' end of the DUX4 promoter (Figure 3.4). The motif spans through the conserved GGGGCG sequence known as GC box that has been previously found to be an important *cis*-element recognised by common transcription factors such as Sp1 (Rettion et al. 2009).

A.

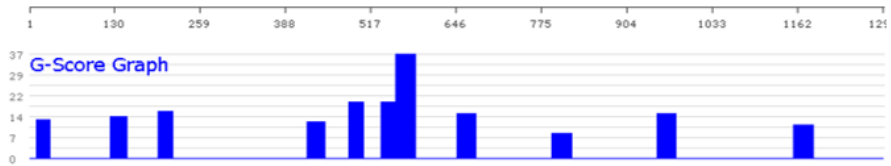
QGRS Mapper summary	
Sequence name:	DUX4 myogenic enhancer 1 (**DME1**)
Number of predicted GQ motifs:	11
No. of motifs with G-score ≥ 30	1

QGRS Mapper summary	
Sequence name:	DUX4 myogenic enhancer 2 (**DME2**)
Number of predicted GQ motifs:	1
No. of motifs with G-score ≥ 30	0

B.



C.



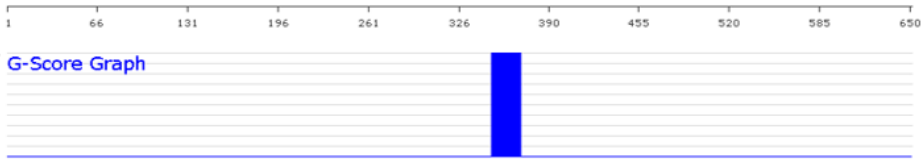


Figure 3.2 QGRS Mapper-based prediction of G-quadruplexes within DUX4 enhancer sequences

Top and bottom graphic show predictions of GQs within the sequence of DUX4 myogenic enhancer 1 and 2 (DM E1 and 2), respectively. (**A**.) The summary table showing the total number of GQs predicted and a number of motifs with a G-score ≥ 30 (i.e., highly likely to form GQs in vitro) present within the analysed sequence. (**B.**) Full FASTA sequence for each analysed genomic element with potential GQ sequences highlighted in yellow. (**C.**) A graphical output form QGRS Mapper showing all the potential GQ motifs forming within the analysed sequences and their relative position.

A.

QGRS Mapper summary	
Sequence name:	DUX4 Promoter
Number of predicted GQ motifs:	4
No. of motifs with G-score ≥ 30	1

B.

```
000001 CCTTCCGGGG TGGGGCGGGC TGTCCCAGGG GGGCTCACCG CCATTCATGA AGGGGTGGAG CCTGCCTGCC TGTGGGCCTT TACAAGGGCG GCTGGCTGGC
000101 TGGCTGGCTG GCTGTCCGGG CAGGCCTCCT GGCTGCAC
```

C.

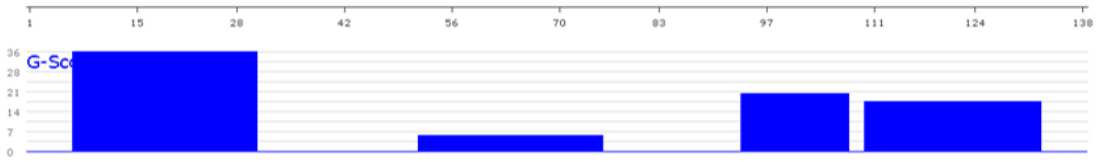


Figure 3.3 QGRS Mapper-based prediction of GQs within DUX4 promoter sequence

(**A.**) The summary table showing the total number of GQ predicted and a number of motifs with the G-score ≥ 30 (i.e., highly likely to form GQs *in vitro*) present within the analysed sequence. (**B.**) Full FASTA sequence for each analysed genomic element with potential GQ sequences highlighted in yellow. (**C.**) A graphical output from QGRS Mapper showing all the potential GQ motifs forming within the analysed sequences and their relative position. The sequences were acquired from the study by Himeda et al., (2014) that first identified DUX4 myogenic enhancers. Promoter sequence was acquired from GenBank (Accession number: AF117653).

A.

QGRS Mapper sumamry	
Sequence name:	DUX4 transcript
Number of predicted GQ motifs:	30
No. of motifs with G-score ≥ 30	4

B.

```
000001 CTGCCGCAGT GCACAGTCCG GCTGAGGTGC ACGGGAGCCC GCCGGCCTCT CTCTGCCCGC GTCCGTCCGT GAAATTCCGG CCGGGGCTCA CCGCGCCATG
000101 GCCCTCCCGA CACCCTCGGA CAGCACCCTC CCCGCGGAAG CCCGGGACG AGGACGGCGA CGGAGACTCG TTTGGACCCC GAGCCAAAGC GAGGCCCTGC
000201 GAGCCTGCTT TGAGCGGAAC CCGTACCCGG GCATCGCCAC CAGAGAACGG CTGGCCCAGG CCATCGGCAT TCCGGAGCCC AGGGTCCAGA TTTGGTTTCA
000301 GAATGAGAGG TCACGCCAGC TGAGGCAGCA CCGGCGGGAA TCTCGGCCCT GGCCCGGGAG ACGCGGCCCG CCAGAAGGCC GGCGAAAGCG GACCGCCGTC
000401 ACCGGATCCC AGACCGCCCT GCTCCTCCGA GCCTTTGAGA AGGATCGCTT TCCAGGCATC GCCGCCCGGG AGGAGCTGGC CAGAGAGACG GGCCTCCCGG
000501 AGTCCAGGAT TCAGATCTGG TTTCAGAATC GAAGGGCCAG GCACCCGGGA CAGGGTGGCA GGGCGCCCGC GCAGGCAGGC GGCCTGTGCA GCGCGGCCCC
000601 CGGCGGGGGT CACCCTGCTC CCTCGTGGGT CGCCTTCGCC CACACCGGCG CGTGGGGAAC GGGGCTTCCC GCACCCCACG TGCCCTGCGC GCCTGGGGCT
000701 CTCCCACAGG GGGCTTTCGT GAGCCAGGCA GCGAGGGCCG CCCCCGCGCT GCAGCCCAGC CAGGCCGCGC CGGCAGAGGG GATCTCCCAA CCTGCCCCGG
000801 CGCGCGGGGA TTTCGCCTAC GCCGCCCCGG CTCCTCCGGA CGGGGCGCTC TCCCACCCTC AGGCTCCTCG GTGGCCTCCG CACCCGGGCA AAAGCCGGGA
000901 GGACCGGGAC CCGCAGCGCG ACGGCCTGCC GGGCCCCTGC GCGGTGGCAC AGCCTGGGCC CGCTCAAGCG GGGCCGCAGG GCCAAGGGGT GCTTGCGCCA
001001 CCCACGTCCC AGGGGAGTCC GTGGTGGGGC TGGGGCCGGG GTCCCCAGGT CGCCGGGGCG GCGTGGGAAC CCCAAGCCGG GGCAGCTCCA CCTCCCCAGC
001101 CCGCGCCCCC GGACGCCTCC GCCTCCGCGC GGCAGGGGCA GATGCAAGGC ATCCCGGCGC CCTCCCAGGC GCTCCAGGAG CCGGCGCCCT GGTCTGCACT
001201 CCCCTGCGGC CTGCTGCTGG ATGAGCTCCT GGCGAGCCCG GAGTTTCTGC AGCAGGCGCA ACCTCTCCTA GAAACGGAGG CCCCGGGGGA GCTGGAGGCC
001301 TCGGAAGAGG CCGCCTCGCT GGAAGCACCC CTCAGCGAGG AAGAATACCG GGCTCTGCTG GAGGAGCTTT AGGACGCGGG GTTGGGACGG GGTCGGGTGG
001401 TTCGGGGCAG GGCGGTGGCC TCTCTTTCGC GGGGAACACC TGGCTGGCTA CGGAGGGCGG TGTCTCCGCC CCGCCCCCTC CACCGGGCTG ACCGGCCTGG
001501 GATTCCTGCC TTCTAGGTCT AGGCCCGGTG AGAGACTCCA CTCCGCGGAG AACTGCCTTT CTTTCCTGGG CATCCCGGGG ATCCCAGAGC CGGCCCAGGT
001601 ACCAGCAGGT GGGCCGCCTA CTGCGCACGC GCGGGTTTGC GGGCAGCCGC CTGGGCTGTG GGAGCAGCCC GGGCAGAGCT CTCCTGCCTC TCCACCAGCC
001701 CACCCCGCCG CCTGACCGCC CCCTCCCCAC CCCCACCCCC CACCCCCGGA AAACGCGTCG TCCCCTGGGC TGGGTGGAGA CCCCCGTCCC GCGAAACACC
001801 GGGCCCCGCG CAGCGTCCGG GCCTGACACC GCTCCGGCGG CTCGCCTCCT CTGCGCCCCC GCGCCACCGT CGCCCGCCCG CCCGGGCCCC TGCAGCCTCC
001901 CAGCTGCCAG CACGGAGCGC CTGGCGGTCA AAAGCATACC TCTGTCTGTC TTTGCCCGCT TCCTGGCTAG ACCTGCGCGC AGTGCGCACC CCGGCTGACG
002001 TGCAAGGGAG CTCGCTGGCC TCTCTGTGCC CTTGTTCTTC CGTGAAATTC TGGCTGAATG TCTCCCCCCA CCTTCCGACG CTGTCTAGGC AAACCTGGAT
002101 TAGAGTTACA TCTCCTGGAT GATTAGTTCA GAGATATATT AAAATGCCCC CTCCCTGTGG ATCCTATAG
```

C.

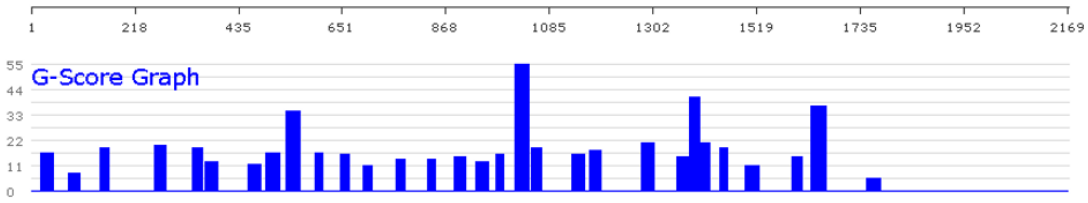
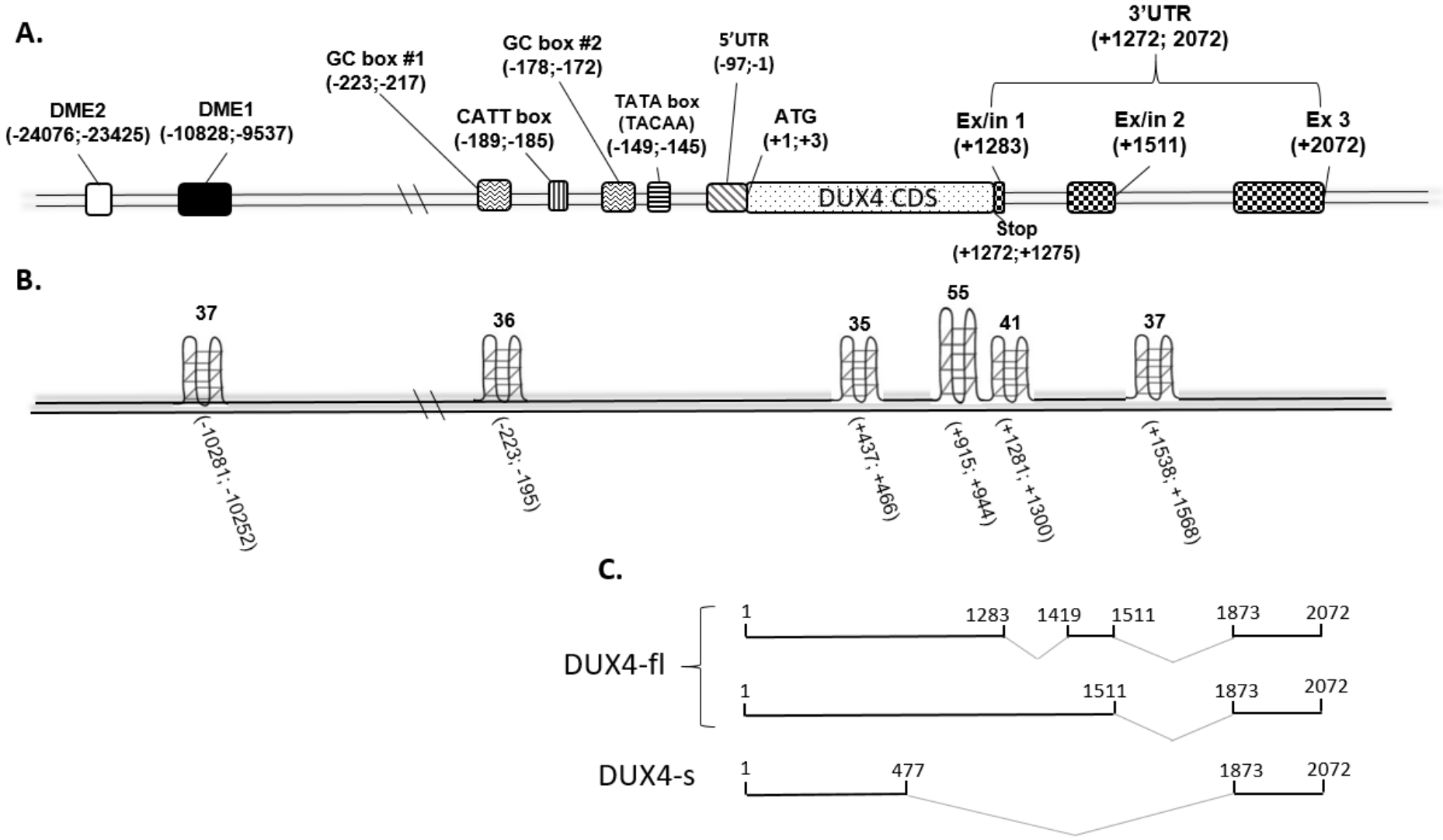


Figure 3.4 QGRS Mapper-based prediction of GQs within DUX4 transcript sequence

(**A.**) The summary table showing the total number of GQ predicted and a number of motifs with the G-score ≥ 30 (i.e., highly likely to form GQs *in vitro*) present within the analysed sequence. (**B.**) Full FASTA sequence for each analysed genomic element with potential GQ sequences highlighted in yellow. (**C.**) A graphical output from QGRS Mapper showing all the potential GQ motifs forming within the analysed sequences and their relative position. Coding and 3'UTR sequence was acquired from Ensembl (ENSG00000260596), whereas the 5'UTR was derived from GenBank (Accession number: AF117653).

Figure 3.5 Schematics of GQ motifs prediction scores (B.) aligned with the biologically relevant genetic signal elements of DUX4 genomic locus (A) and transcript (C)

Score prediction of the putative GQ structures was preformed using QGRS Mapper software. Location of the genomic elements relative to the start codon (ATG) was mapped based on the data acquired form GenBank (Accession number: AF117653) and publications by Himeda et al., (2014) and Dixit et al., (2007). The transcript sequences were derived from Ensembl (ENSG00000260596). Abbreviations: DME1 and 2, DUX4 myogenic enhancer 1 and 2; DUX4-fl, DUX4 full length; DUX4-s, DUX4 short; Ex, exon; In, intron; CDS, coding DNA sequence; UTR, untranslated region; GQ, G-quadruplex.


A.
DME2
(-24076;-23425)
DME1
(-10828;-9537)
GC box #1
(-223;-217)
CATT box
(-189;-185)
GC box #2
(-178;-172)
TATA box
(TACAA)
(-149;-145)
5'UTR
(-97;-1)
ATG
(+1;+3)
DUX4 CDS
Stop
(+1272;+1275)
Ex/in 1
(+1283)
Ex/in 2
(+1511)
Ex 3
(+2072)
3'UTR
(+1272; 2072)
B.
37
36
35
55
41
37
(-10281; -10252)
(-223; -195)
(+437; +466)
(+915; +944)
(+1281; +1300)
(+1538; +1568)
C.
DUX4-fl
DUX4-s
1
1283
1419
1511
1873
2072
1
1511
1873
2072
1
477
1873
2072

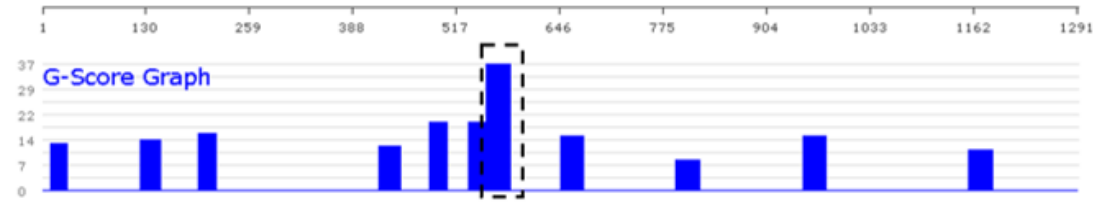
Since the GQ forming sequences often contain cytosine residues, it is important to determine whether these will contribute to potential hairpin formations that are more thermodynamically stable than the corresponding GQ (Fay et al. 2017). To assess potential for formation of other secondary structures by the selected GQ sequences from enhancer, promoter and transcript sequences, the Mfold software was used. The isolated oligonucleotide GQ forming sequences (as predicted by the QGRS Mapper), that have ΔG <-10 kcal/mol were considered to have a higher driving force for hairpin formation structures over the GQ motifs. The ΔG prediction for DME1 G-quadruplex (GQ) and D4P GQ sequence was -3.31 and 0.33 kcal/mol, indicating a very low potential for these oligonucleotide sequences to fold into hairpin structures (Figure 3.6). The sequences with highest G-Score present within exon 1 (E1 GQ), has a ΔG= -7.30 kcal/mol that is relatively unfavourable for the hairpin formation (Figure 3.7). The relatively high ΔG=-0.66 kcal/mol predicted for the GQ predicted sequence motif located in the proximity of the splice site 1 (SS1_GQ), indicated low probability for this sequence to form hairpin structure (Figure 3.7). The ΔG for the remaining GQ forming sequences within CSS and splice site 2 (SS2 GQ) scored -14.70 and -10.00 kcal/mol, respectively, indicating that these sequences have a significantly higher potential to form hairpin structures rather than the GQ motifs (Figure 3.7).

Figure 3.6 Mfold analysis of GQ forming sequence derived from the DUX4 myogenic enhancer 1 (DME1) and promoter (D4P) regions

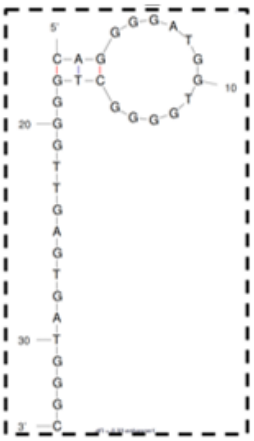
(A.) QGRS mapper graphic score highlighting sequence that is highly likely to form GQ *in vitro* (dashed box; G-score ≥ 30). (B.) Mfold analysis of the GQ forming sequence. The free energy (ΔG) is indicated.

DME1 sequence analysis

A.



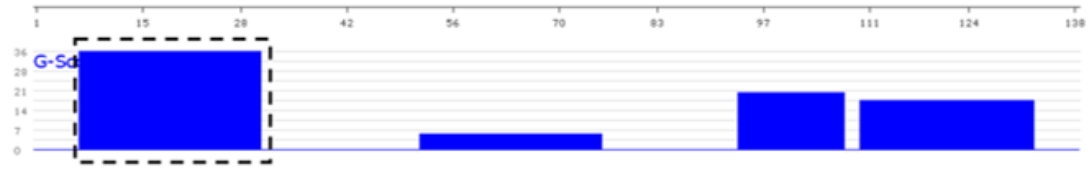
B.



ΔG= 0.33

D4P sequence analysis

A.



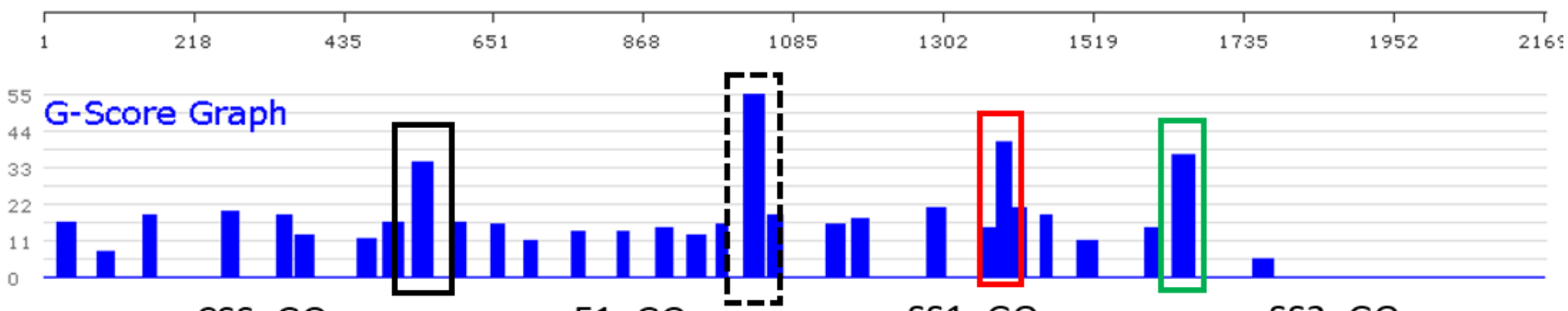
B.

ΔG= -3.32

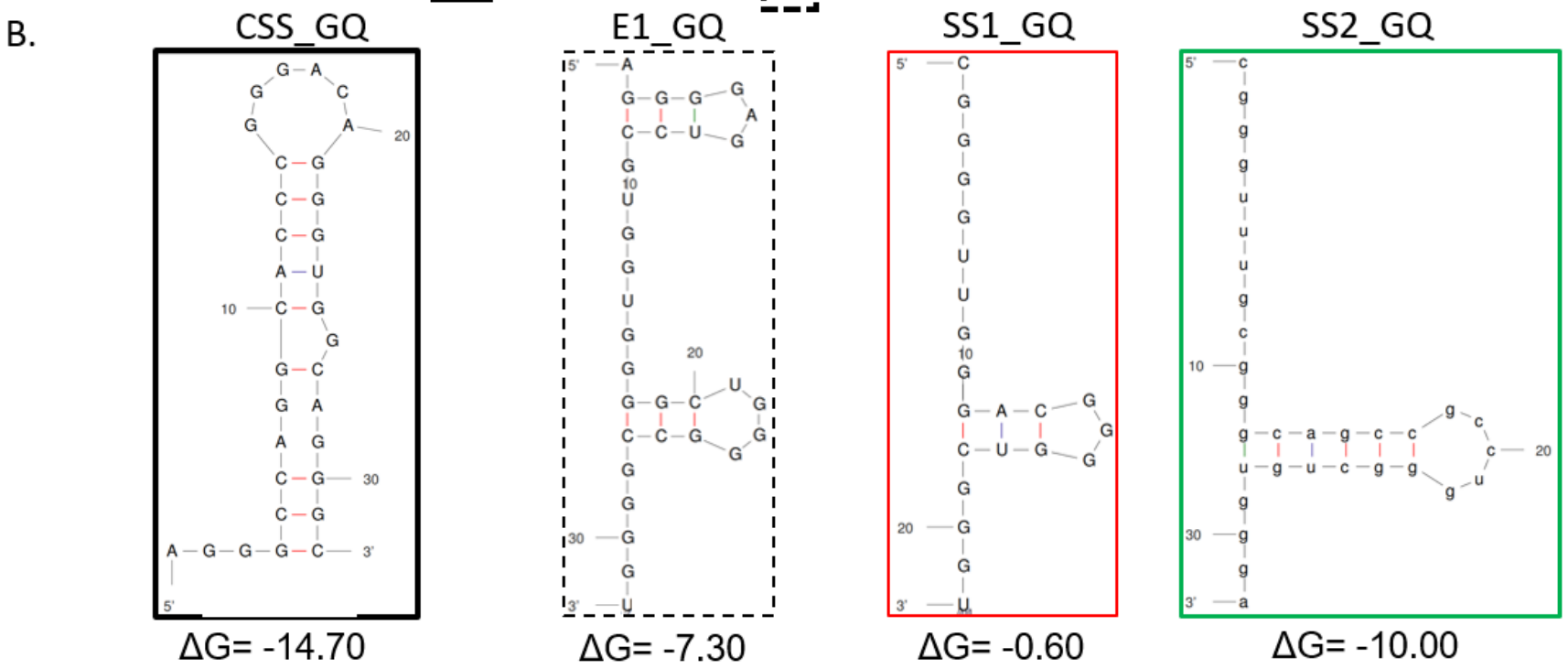
Figure 3.7 Mfold analysis of GQ forming sequence derived from the DUX4 transcript region

(**A.**) QGRS mapper graphic score highlighting the sequences that are highly likely to form GQ *in vitro* with the G-score ≥ 30. The putative GQ sequences that lay within the regions of cryptic splice site (CSS; black solid box), exon 1 (E1; black dashed box), splice site 1 (SS1; red box) and splice site 2 (SS2; green box) of the DUX4 transcript. (**B.**) Mfold analysis of the detailed GQ forming sequences. The free energy (ΔG) is indicated.

A.


DUX4 transcript analysis
1
218
435
651
868
1085
1302
1519
1735
1952
G-Score Graph
55
44
33
22
11
0

B.


CSS_GQ
E1_GQ
SS1_GQ
SS2_GQ
ΔG= -14.70
ΔG= -7.30
ΔG= -0.60
ΔG= -10.00

3.2.2. *CD and NMR structural analysis to determine GQ formation of selected DNA and RNA oligonucleotides*

To confirm whether the *in silico* predicted GQ-forming sequences (Figure 3.5) derived from the enhancer, promoter and transcript elements of *DUX4*, form in solution, CD and NMR analyses were performed on corresponding synthetic oligonucleotides. In addition, to determination of GQ presence, the CD and NMR were used to establish secondary topology of the selected DNA and RNA oligonucleotides. To generate a stable GQ structures the oligonucleotides were annealed in KP buffer with added 100 mM KCl that promotes formation of the motif. Exact sequence of each oligo is outlined in (Table 2.9). All of the selected oligonucleotides carry 4 runs of three Gs that have potential to form 3 tetrads within the GQ structure. The runs of Gs are interspaced by varying number of nucleotides that will form loops in the potential GQ structure. Each oligonucleotide sequence was selected so the nucleotide at the end of each sequence is different than G to prevent stacking of the GQ molecules.

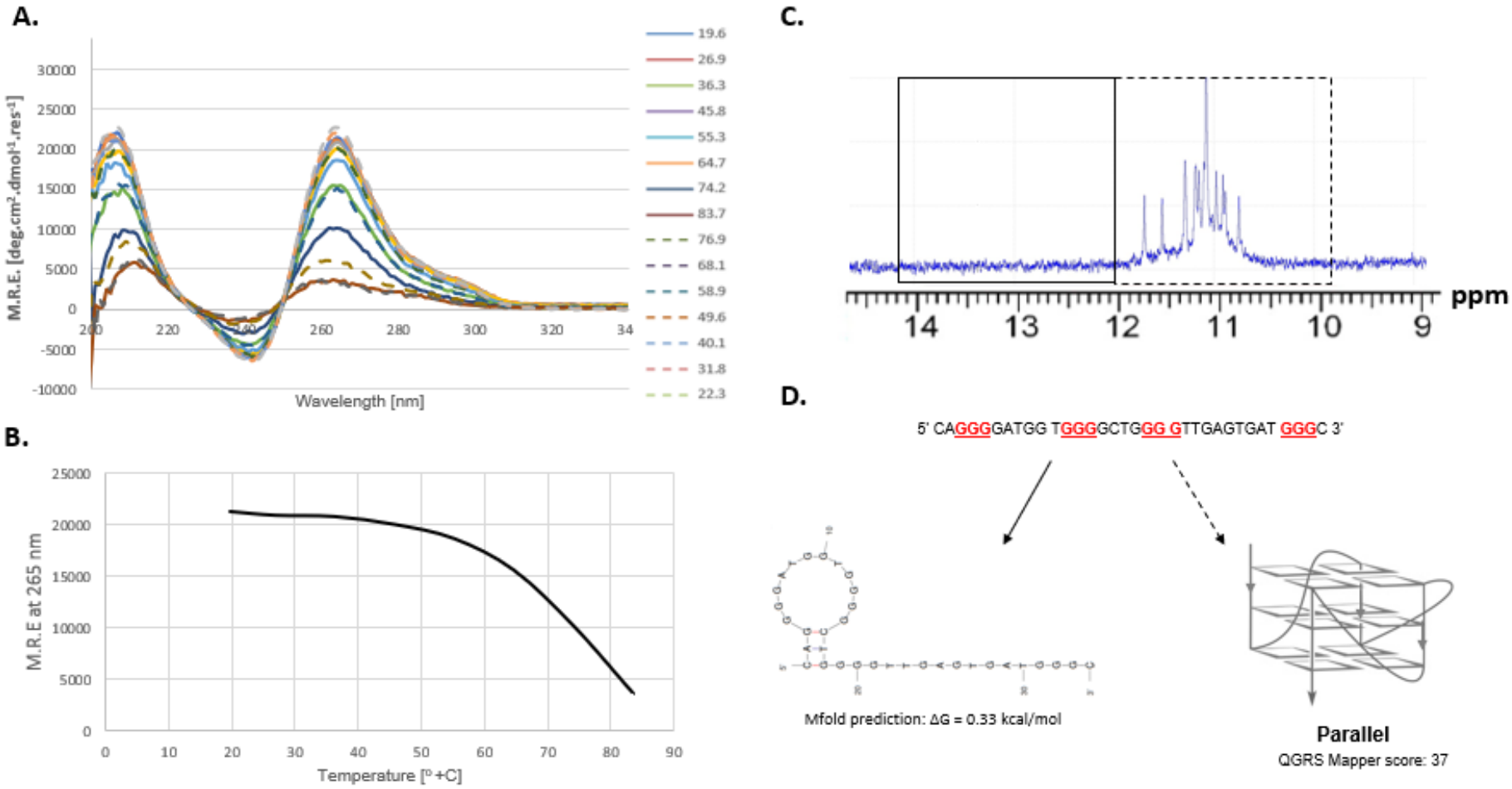
3.2.2.1. CD and NMR structural analysis of GQ formation by the DNA oligonucleotide sequences derived from DME1 and D4P of DU4

The DME1 DNA GQ oligo (5' CA**GGG**GATGG T**GGG**GCTG**GG G**TTGAGTGAT **GGG**C 3') shows a characteristic CD spectrum pattern of an ellipticity maximum around 260 nm and negative ellipticity minimum at 240 nm at room temperature indicative of a parallel GQ structure (Figure 3.8 A). The NMR spectrum confirmed the presence of GQ formation by the oligo as detected by the strong signal at 10 and 12 ppm, which is specific to the Hoogsteen hydrogen boding. The NMR readout also shows no signal detection at the >12 ppm range, indicating no other secondary structure formation by the oligonucleotide (Figure 3.8 C). CD spectra shows weak absorbance at the 290 nm

ellipticity, suggesting presence of external loops formed by bimolecular parallel GQ, rather than the tetramolecular (four stranded) structure that has a single positive peak at 260 nm ellipticity (Paramasivan, Rujan, and Bolton 2007)(Figure 3.8 A). The melting transition appears smooth, further indicating that there is one predominant structure of the oligonucleotide formed (Figure 3.8 B). The relatively low melting temperature of ~70°C is most likely caused by the presence of potential long loop-forming sequences within the motif. Upon thermal renaturation of the DNA oligo, the absorbance pattern returns to the original position suggesting no transition to other stable topologies or secondary structures than the parallel GQ (Figure 3.8 A).

The CD and NMR spectral data, supplemented by the bioinformatic analysis, showing a low probability of hairpin formation (ΔG=0.33 kcal/mol) and relatively high score for quadruplex formation (≥30) strongly suggest that the DME1 DNA oligonucleotide forms a single species GQ of bimolecular parallel structure in solution (Figure 3.8 D).

Figure 3.8 Biophysical characterisation of DNA secondary structure of the predicted GQs in the DUX4 myogenic enhancer 1 (DME1) sequence (**A.**) Far-UV circular dichroism spectrum at 25°C reveals ellipticity maximum at 265 nm indicative of a parallel GQ structure formation by the oligonucleotide. The temperature at which each absorption spectra was obtained is expressed in degrees Celsius. (**B.**) Sigmoidal CD melting curves for the oligonucleotide shows transitional mid-point at 70°C indicating modest stability of the quadruplex structure. (**C.**) ^{1}H NMR spectrum of the oligonucleotide recorded at 800 MHz and 298K showing groups of resonances characteristic for Watson-Crick (solid line box highlighting non-quadruplex structure signal) and Hoogsten H-bonded bases (dash line box highlighting quadruplex structure). As highlighted, the DME1 sequence forms a single species of quadruplex structure. (**D.**) Bioinformatic prediction of possible secondary structures formed by the sequence using mFold (solid arrow) or QGRS Mapper (dashed arrow). Abbreviations: M.R.E - molar residue ellipticity

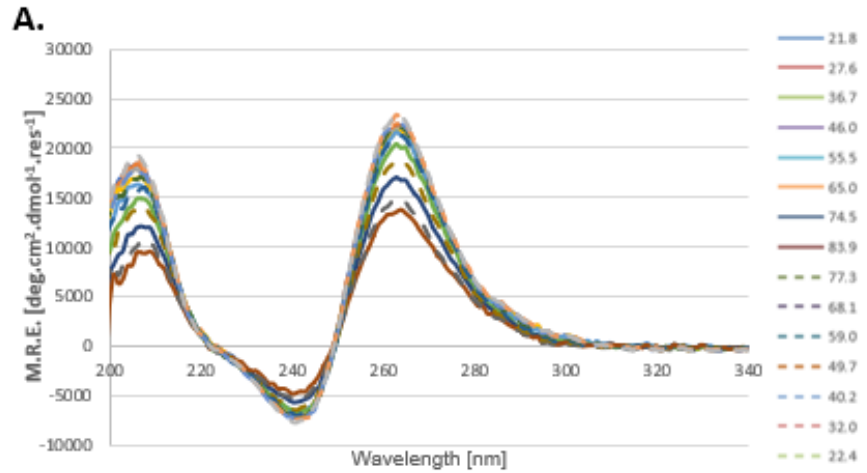

A.
M.R.E. [deg.cm².dmol⁻¹.res⁻¹]
Wavelength [nm]
19.6
26.9
36.3
45.8
55.3
64.7
74.2
83.7
76.9
68.1
58.9
49.6
40.1
31.8
22.3
B.
M.R.E at 265 nm
Temperature [°+C]
C.
ppm
D.
5' CAGGGGATGG TGGGGCTGGG GTTGAGTGAT GGGC 3'
Mfold prediction: ΔG = 0.33 kcal/mol
Parallel
QGRS Mapper score: 37

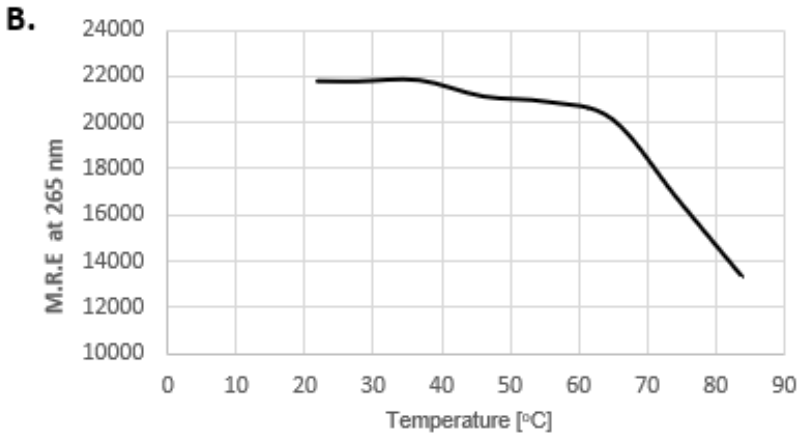
The D4P DNA GQ oligonucleotide *(*5’ CG**GGG**T**GGGGCGGG**CT**GTCCC**A**GGG**GGGCT 3’) has a characteristic pattern of absorbance at the ellipticity maximum at 260 nm and ellipticity minimum at 240nm for a parallel GQ as shown by the CD (Figure 3.9 A). Lack of apparent absorbance at 290 nm indicates that the structural topology is tetramolecular lacking external loops (Figure 3.9 A). The NMR spectrum displays distinctive signals characteristic of both GQ as well as double-stranded hairpin structure (Figure 3.9 C). The double transition of the melting curve indicates presence of two structures with quite different stabilities (i.e., hairpin and quadruplex). There is no significant transition in the CD spectra upon thermal renaturation of the oligonucleotide, since both hairpin and parallel GQs have a strongly overlapping ellipticity maxima at 260nm (Kypr et al. 2009). The negative ΔG of -3.32 kcal/mol for the oligonucleotide indicates that there is a possibility of a hairpin and GQ folding might exist in thermodynamic equilibrium, which is supported by the CD and NMR readouts (Figure 3.9 D).

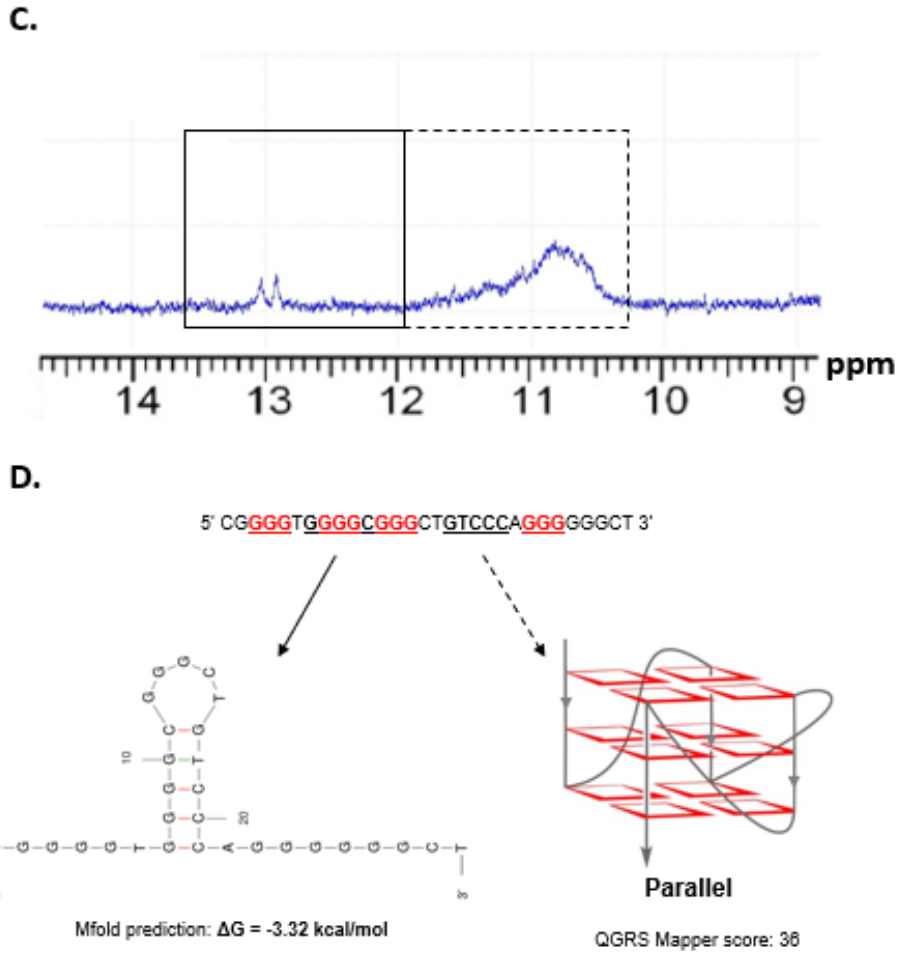
.

Figure 3.9 Biophysical characterisation of DNA secondary structure of the predicted GQs in the DUX4 promoter (DP) sequence

(**A.**) Far-UV circular dichroism spectrum at 25°C reveals ellipticity maximum at 265 nm indicative of a parallel GQ structure formation by the oligonucleotide. The temperature at which each absorption spectra was obtained is expressed in degrees Celsius. (**B.**) However, the melting curve shows a double transition, indicating that two different structures i.e., hairpin and quadruplex exist with different stabilities. Overall melting temperature was recorded at 78°C (**C.**) ^{1}H NMR spectrum of the oligonucleotide recorded at 800 MHz and 298 K showing groups of resonances characteristic for Watson-Crick (solid line box highlighting non-quadruplex structure signal) and Hoogsten H-bonded bases (dash line box highlighting quadruplex structure). Recorded NMR signal confirms presence of hairpin (solid line box) and quadruplex structure (dashed line box). (**D.**) Bioinformatic prediction of possible secondary structures formed by the sequence using Mfold (solid arrow) or QGRS Mapper (dashed arrow). Abbreviations: M.R.E - molar residue ellipticity


A.
M.R.E. [deg.cm².dmol⁻¹.res⁻¹]
Wavelength [nm]
21.8
27.6
36.7
46.0
55.5
65.0
74.5
83.9
77.3
68.1
59.0
49.7
40.2
32.0
22.4


B.
M.R.E at 265 nm
Temperature [°C]

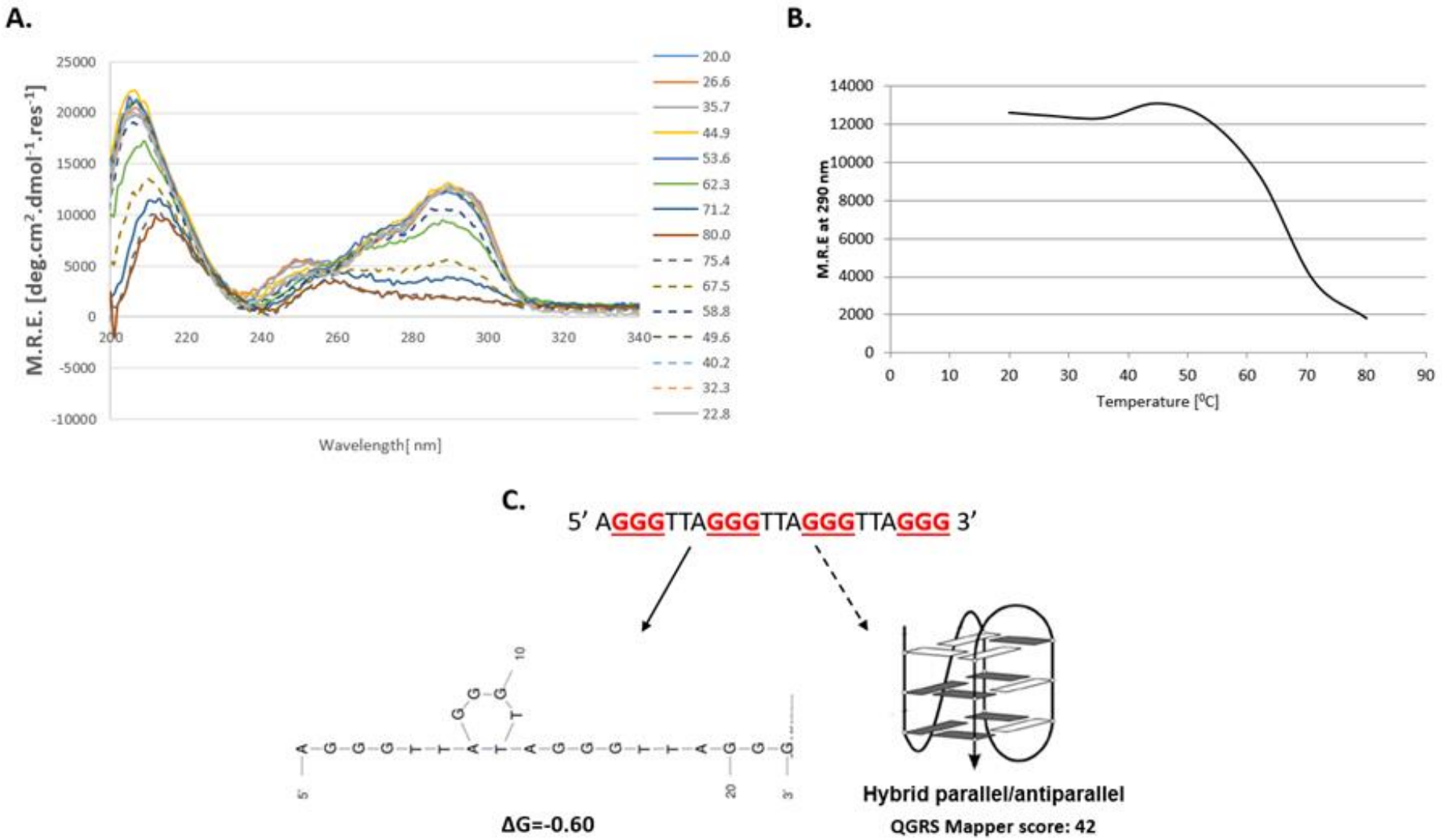

C.
ppm
D.
5' CGGGGTGGGGCGGGCTGTCCCAGGGGGGCT 3'
Mfold prediction: ΔG = -3.32 kcal/mol
Parallel
QGRS Mapper score: 36

.

The DNA oligonucleotide sequence derived from the human telomere (H. Telomer GQ) (5'A**GGG**TTA**GGG**TTA**GGG**TTA**GGG** 3') was used as a positive control. The classic CD absorption spectra for an antiparallel GQ has the ellipticity maxima at 290 nm and ellipticity minima 260 nm. Although the measured H. Telomere GQ oligonucleotide shows maximum absorbance at 290 nm, there is a shoulder peak at around 250 nm and ellipticity minimum at 238 nm (Figure 3.10 A). This absorption patterns is indicative of a hybrid type unimolecular parallel/antiparallel GQ structure as it has been previously reported for this sequences (Ambrus et al. 2006b). The melting temperature of the GQ structure was recorded at around 60°C using the CD analysis (Figure 3.10 B). The CD data confirms the bioinformatic data predicting no hairpin structure and high G-Score (i.e., 42) for quadruplex formation as shown by the Mfold and GQRS Mapper, respectively (Figure 3.10 C).

Figure 3.10 Biophysical characterisation of DNA secondary structure of the human telomeric sequence

(**A.**) Far-UV circular dichroism spectrum at 25°C reveals ellipticity maximum at 290 nm, shoulder peak at 250 and ellipticity minimum at 238 nm indicative of a hybrid parallel/antiparallel GQ structure formation by the oligonucleotide. The temperature at which each absorption spectra was obtained is expressed in degrees Celsius. (**B.**) Overall melting temperature was recorded at 60°C (**C.**) Bioinformatic prediction of possible secondary structures formed by the sequence using mFold (solid arrow) or QGRS Mapper (dashed arrow). Abbreviations: M.R.E - molar residue ellipticity


A.
M.R.E. [deg.cm².dmol⁻¹.res⁻¹]
Wavelength[nm]
20.0
26.6
35.7
44.9
53.6
62.3
71.2
80.0
75.4
67.5
58.8
49.6
40.2
32.3
22.8
B.
M.R.E at 290 nm
Temperature [⁰C]
C.
5’ AGGGTTAGGGTTAGGGTTAGGG 3’
ΔG=-0.60
Hybrid parallel/antiparallel
QGRS Mapper score: 42

3.2.2.2. GQ formation by RNA oligonucleotides derived from DUX4 transcript analysed by CD and NMR

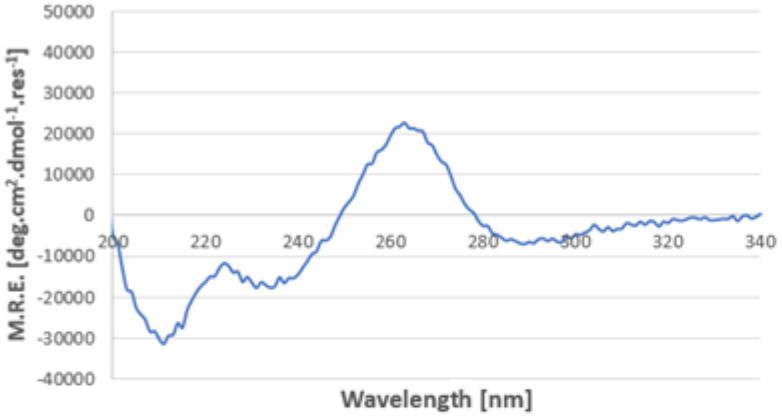
The first RNA oligonucleotide of interest that was tested by the CD and NMR for the formation of GQ was the sequence derived from the region overlapping the cryptic splice site (CSS), since it could be involved in alternative splicing of the *DUX4* transcript and yielded non-toxic DUX4 (short) isoform (Snider et al. 2010). As the RNA GQcan only form parallel structures due to the lack of the *anti*-glycosidic bond geometry and the fact the hairpin structures show overlapping ellipticity maximum absorbance at 260 nm with parallel quadruplexes, it is important to supplement the CD data with the NMR and bioinformatic analysis to discriminate between those two secondary structures.

As predicted, the CSS GQ oligonucleotide (5' C**GGG**GUU**GGG**ACG**GGG**UC**GGG**U 3') measured by the CD shows ellipticity maxima at 260 nm and ellipticity minimum at around 240 nm that would normally indicate that it is a parallel GQ motif (Figure 3.11 A). However, presence of a second ellipticity minima at 210 nm suggest that the A-form (i.e., hairpin) structure is formed by this sequence (Figure 3.11 A). The clear signal at >12ppm from the NMR also shows presence for hairpin specific hydrogen bonds, whereas no GQ specific signal was recorded (Figure 3.11 C). The smooth melting curve indicates a single hairpin structure in the solution (Figure 3.11 B). The mFold predicts a very thermodynamically stable hairpin formation due to presence of multiple C residues within the analysed sequences (Figure 3.11 D). Despite the high G-Score of 41, the NMR and CD analysis does not confirm quadruplex formation by the analysed sequences, but confirms the hairpin formation as predicted by the Mfold analysis.

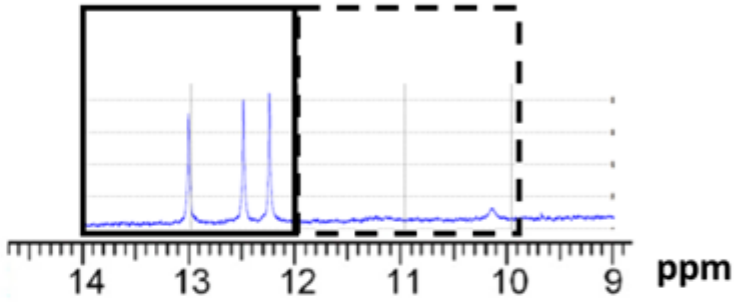
Figure 3.11 Biophysical characterisation of RNA secondary structure of the predicted GQs in the cryptic splice site (CSS) sequence

(**A.**) Far-UV circular dichroism spectrum at 25°C reveals the ellipticity maximum at 265 nm and the ellipticity minimum at 240 and 210 nm characteristic for a hairpin structure. (**B.**) ^{1}H NMR spectrum of the oligonucleotide recorded at 800 MHz and 298 K showing groups of resonances characteristic for Watson-Crick (solid line box highlighting non-quadruplex structure signal) and Hoogsten H-bonded bases (dash line box highlighting quadruplex structure). As highlighted, the CSS sequence forms a single species of hairpin structure. (**C.**) Bioinformatic prediction of possible secondary structures formed by the sequence using mFold (solid arrow) or QGRS Mapper (dashed arrow). Abbreviations: M.R.E - molar residue ellipticity

A.

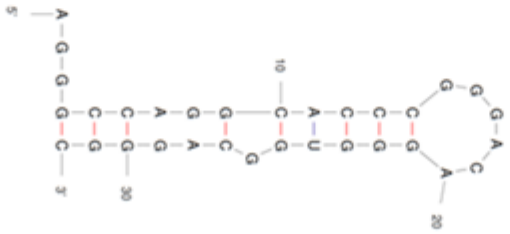


B.



C.

5′ C<u>GGG</u>GUU<u>GGG</u>ACG<u>GGG</u>UC<u>GGG</u>U 3′



ΔG=-14.70

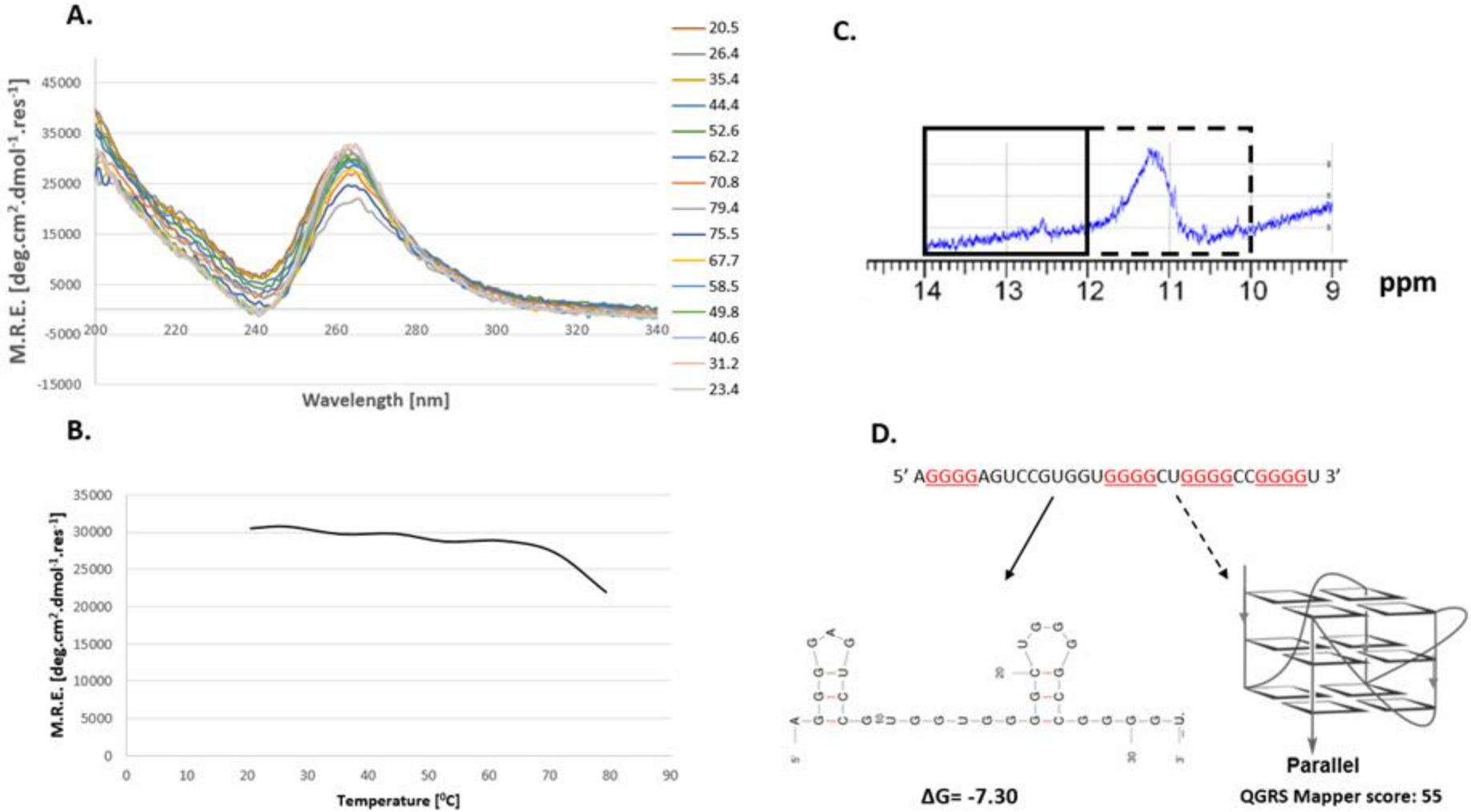
QGRS Mapper score: 41

The highest scoring sequence for the formation of GQ out of all of the analysed sequences was present downstream of the CSS in the exon 1. The exon 1 GQ (E1 GQ) had a G-Score of 55 as predicted by the GQRS Mapper, which is higher than the well-studied human telomeric quadruplex sequence that scored 44 when analysed with the same software (Figure 3.7; Figure 3.10 C).

The RNA E1 GQ sequences (5' AGGGGAGUCCGUGGUGGGGCUGGGGCCGGGGU 3') analysed by the CD shows the ellipticity maximum at 260 nm and ellipticity minimum at 240 nm, which is a characteristic pattern of absorption for a parallel GQ (Figure 3.12 A). The oligonucleotide shows high thermal stability as the full melting profile could not be achieved at the near 80°C (Figure 3.12 B). Despite the relatively large intervening loops between the G-tetrads, the analysed sequence shows to be particularly thermodynamically stable, which aligns with the QGRS Mapper output predicting the high G-Score of 55 for the RNA oligo. The NMR shows a strong, clear signal at the 10-12 ppm range indicating presence of quadruplex-specific Hoogsteen hydrogen bonding (Figure 3.12 C). No Watson-Crick hydrogen bonding could be detected by the NMR, confirming that no hairpin structures are formed by this sequence in the solution (Figure 3.12 C). The ΔG of <10 kcal/mol from the Mfold output indicates small probability of a stable hairpin formation by the E1 GQ sequences (Figure 3.12 D).

Figure 3.12 Biophysical characterisation of RNA secondary structure of the predicted GQ in the exon 1 (E1) sequence

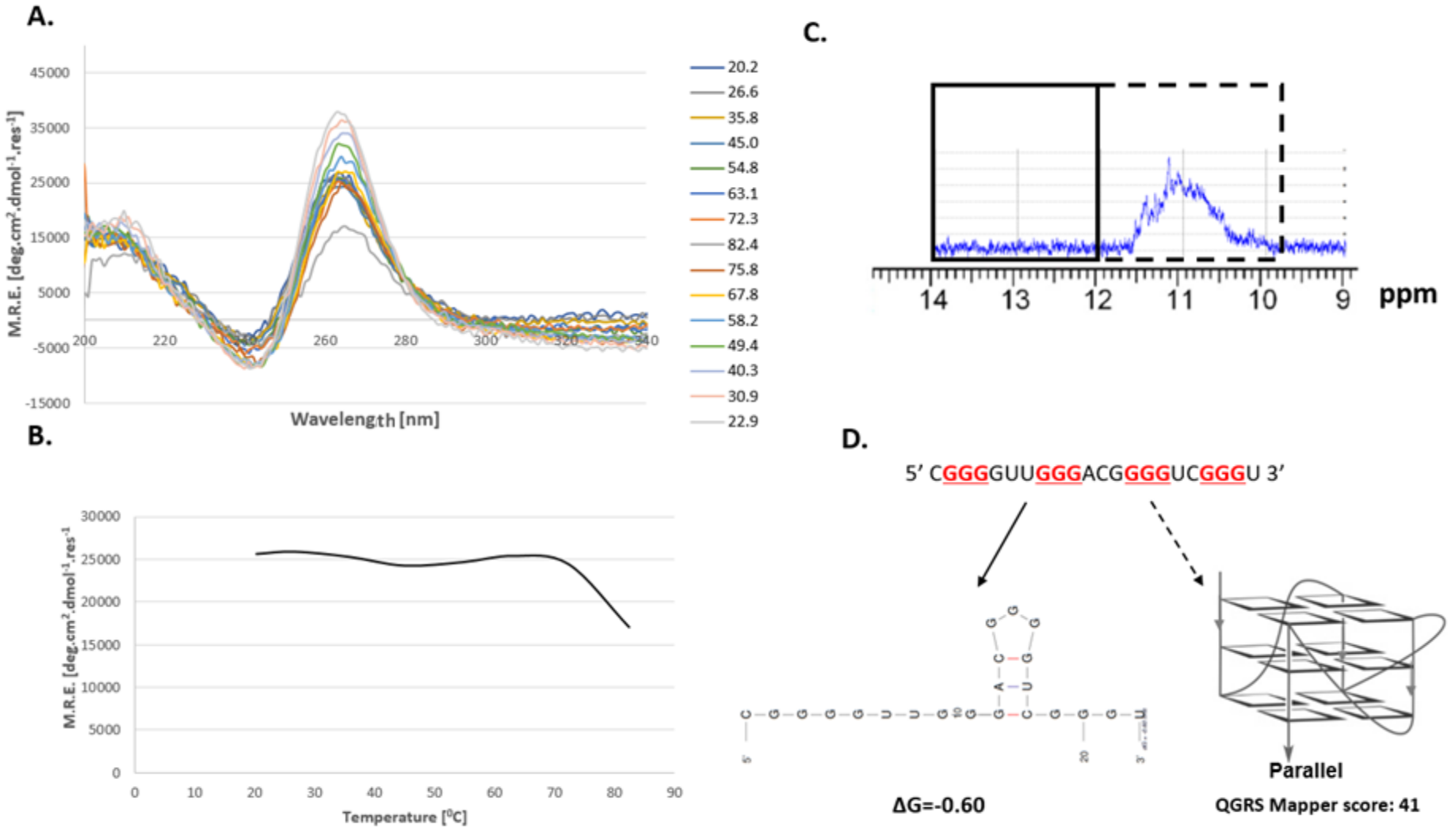
(**A.**) Far-UV circular dichroism spectrum at 25°C reveals the ellipticity maximum at 265 nm and the ellipticity minimum at 240 typical for parallel GQ structure. The temperature at which each absorption spectra was obtained is expressed in degrees Celsius. (**B.**) Sigmoidal CD melting curve was not fully formed in the assessed temperature range; therefore, no Tm was assessed. (**C.**) ^{1}H NMR spectrum of the oligonucleotide recorded at 800 MHz and 298 K showing groups of resonances characteristic for Watson-Crick (solid line box highlighting non-quadruplex structure signal) and Hoogsten H-bonded bases (dash line box highlighting quadruplex structure). As highlighted, the E1 GQ sequence forms a single species of quadruplex structure. (**D.**) Bioinformatic prediction of possible secondary structures formed by the sequence using mFold (solid arrow) or QGRS Mapper (dashed arrow). Abbreviations: M.R.E - molar residue ellipticity


A.
M.R.E. [deg.cm^2.dmol^{-1}.res^{-1}]
45000
35000
25000
15000
5000
-5000
-15000
200
220
240
260
280
300
320
340
Wavelength [nm]
20.5
26.4
35.4
44.4
52.6
62.2
70.8
79.4
75.5
67.7
58.5
49.8
40.6
31.2
23.4
B.
M.R.E. [deg.cm^2.dmol^{-1}.res^{-1}]
35000
30000
25000
20000
15000
10000
5000
0
0
10
20
30
40
50
60
70
80
90
Temperature [^{0}C]
C.
14
13
12
11
10
9
ppm
D.
5' AGGGGAGUCCGUGGUGGGGCUGGGGCCGGGGU 3'
Parallel
ΔG= -7.30
QGRS Mapper score: 55

The final RNA oligonucleotide that had the G-score ≥30 lies at the exon1/intron1 boundary, in the region of the first splice site (SS1). The RNA oligonucleotide SS1 GQ sequences (5' C**GGG**GUU**GGG**ACG**GGG**UC**GGG**U 3'), when subjected to the CD analysis, shows ellipticity maximum at 265 nm and ellipticity minimum at 240 nm, which is characteristic of a parallel GQ (Figure 3.13 A). The complete thermal denaturation of the structure could not be achieved within the available temperature limits, showing high stability of the analysed sequences (Figure 3.13 A). As a result, full melting curve profile could not be achieved for this sequence and therefore no Tm could be derived (Figure 3.13 B). Interestingly, the peak at 265 nm becomes more pronounced after the renaturation process, suggesting that the process of denaturation promoted formation of additional GQ motifs that failed to form during the initial annealing process (Figure 3.13 A). In the NMR analysis, the oligonucleotide sequences have been found to produce a signal within the 10-12 ppm rage which is GQ specific (Figure 3.13 A). No signal for Watson-Crick, hairpin specific signal was recorded in the NMR readout (Figure 3.13 C). In support of the CD and NMR analysis, the Mfold predicted a potential hairpin formation with a modest stability as suggested by the low ΔG of -0.60 kcal/mol (Figure 3.13 D).

Figure 3.13 Biophysical characterisation of RNA secondary structure of the predicted GQs in the splice site 1 (SS1) sequence

(**A.**) Far-UV circular dichroism spectrum at 25°C reveals the ellipticity maximum at 265 nm and the ellipticity minimum at 240 nm typical for parallel GQ structure. The temperature at which each absorption spectra was obtained is expressed in degrees Celsius. (**B.**) Sigmoidal CD melting curve was not fully formed in the assessed temperature range, therefore no Tm was assessed. (**C.**) ^{1}H NMR spectrum of the oligonucleotide recorded at 800 MHz and 298 K showing groups of resonances characteristic for Watson-Crick (solid line box highlighting non-quadruplex structure signal) and Hoogsten H-bonded bases (dash line box highlighting quadruplex structure). As highlighted, the SS1 GQ sequence forms a single species of quadruplex structure. (**D.**) Bioinformatic prediction of possible secondary structures formed by the sequence using mFold (solid arrow) or QGRS Mapper (dashed arrow). Abbreviations: M.R.E - molar residue ellipticity


A.
45000
35000
25000
15000
5000
-5000
-15000
200
220
260
280
300
320
340
M.R.E. [deg.cm².dmol⁻¹.res⁻¹]
Wavelength [nm]
20.2
26.6
35.8
45.0
54.8
63.1
72.3
82.4
75.8
67.8
58.2
49.4
40.3
30.9
22.9
C.
14
13
12
11
10
9
ppm
B.
30000
25000
20000
15000
10000
5000
0
0
10
20
30
40
50
60
70
80
90
M.R.E. [deg.cm².dmol⁻¹.res⁻¹
Temperature [⁰C]
D.
5’ CGGGGUUGGGACGGGGUCGGGU 3’
ΔG=-0.60
Parallel
QGRS Mapper score: 41

3.2.3. *Effect of berberine on secondary structure of selected DNA and RNA oligonucleotides*

Having established that all of the selected oligonucleotide sequences from the enhancer, promoter and transcript elements (except for the hairpin-forming CSS sequences) form GQs in solution, the next step was to investigate the effect of berberine on conformation of these secondary structures. The CD spectra was used to assess structural changes of the DME1 GQ, D4P GQ, CSS GQ and SS1 GQ oligonucleotides (5 μM) in the presence of increasing concentrations of berberine (0-50 μM). Each of the sequences was also subjected to thermal denaturation process in the presence of the ligand at the maximum concentration (50 μM) in order to study berberine's binding properties.

First, the DME1 GQ DNA oligonucleotide was investigated (Figure 3.14). Upon increasing concentration of the compound, no apparent change of the ellipticity maximum at 265 nm, ellipticity minimum at 240 nm and the shoulder peak at 290 nm indicative of bimolecular parallel GQ structure has occurred (Figure 3.14 A). This indicated that no further induction of the GQ structure has been promoted, and that there is no structural shift towards other types of GQ topologies (e.g., antiparallel) induced by the presence of berberine. The melting curve analysis reveals melting transition increase from ~44 to 55°C for the oligonucleotide with the berberine added compared with the oligonucleotide solution that did not contain the compound, indicating potential binding of the compound to the GQ. The double transition in the melting curve for the oligonucleotide solution containing berberine, might indicate the temperature-induced dissociation of the compound from the GQ prior to the denaturation of the secondary structure (Figure 3.14 B). Furthermore, the double transition might be also indicative of

a weak binding of berberine, leading to equilibrium between bound and unbound state. However, due to the limitations of the apparatus, it was not possible to reach temperatures above ~80°C that would allow formation of the full melting curve profile for both of the analysed oligonucleotides (Figure 3.14 B). Therefore, an accurate Tm could not be calculated that in turn would allow for assessment of berberine's binding affinity to the DME1 GQ DNA oligonucleotide.

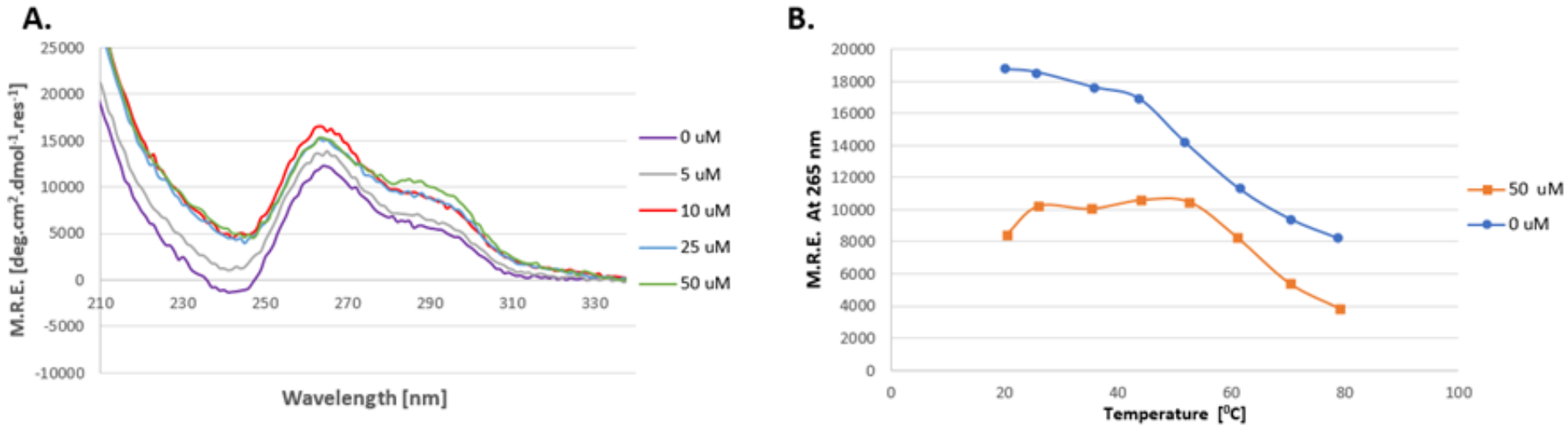


Figure 3.14 Effect of increasing berberine concentrations on secondary structure formation of DME1 GQ sequence

(**A.**) Far-UV spectra were recorded using CD at each concentration of berberine (0-50µM) added to the fixed concentration of the oligonucleotide (5 µM). Each measurement was performed at room temperature. Increasing concentration of berberine has not changed the intensity or transition the 265 or 290nm peaks noticeably, indicating no change in the parallel GQ structure of the oligonucleotide. (**B.**) CD spectra recorded at 265 nm absorbance of the DME1 GQ oligo without and with 50µM berberine added. The absorbance was expressed as a function of temperature ranging from ~20 to 80°C (±2°C). Melting of the oligonucleotide structure is increased in the presence of berberine from ~45 to ~55°C. Abbreviations: M.R.E (molar residue ellipticity); DME1 GQ (DUX4 myogenic enhancer 1 G-quadruplex).

When the D4P GQ DNA oligonucleotide was analysed in the presence of berberine, a very significant change in ellipticity was recorded (Figure 3.15 A). The ellipticity maximum at 265 nm has decreased as the concentration of the ligand was increased. In addition, the absorbance peak at 290 nm has increased as the berberine was added to the D4P GQ oligonucleotide solution. At the highest concentration of the ligand, the 290 nm absorbance has reached ellipticity maximum and has become more pronounced than the 265 nm absorbance. This 265 to 290 nm shift of absorbance is mediated by end stacking of berberine to the GQ structures that leads to transition from the parallel to the antiparallel topology (Figure 3.15 A). The berberine-induced change in structural plasticity of the D4P GQ is also apparent from the melting profile (Figure 3.15 B). The early increase in temperature (~20-50°C) initially displaced the ligand from the GQ and lead to the initial return from the antiparallel to the original (i.e., ligand free) parallel structure conformation with the ellipticity peak shifting from 290 to 265 nm. As the temperature was increased to ~60°C, the secondary structure of the D4P GQ started to denature (Figure 3.15 B). The full melting profile of the measured GQ structure could not be achieved due to the limitations of the machine. Therefore, an accurate Tm value for binding affinity calculations could not be derived.

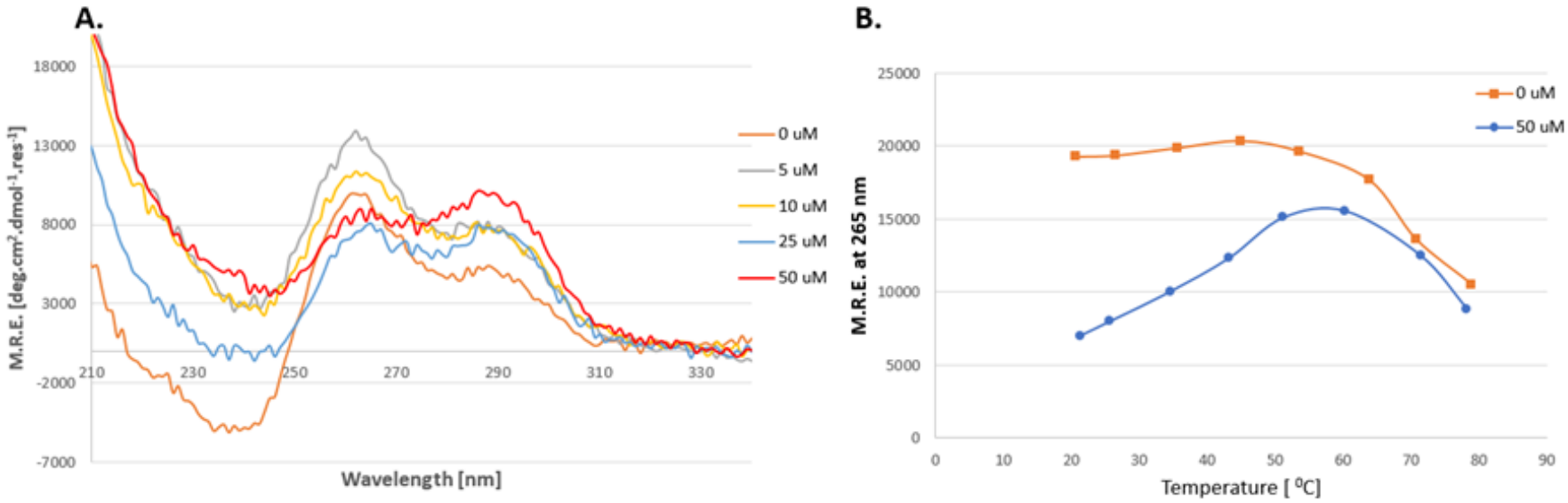


Figure 3.15 Effect of increasing berberine concentrations on secondary structure formation of D4P GQ sequence

(**A.**) Far-UV spectra was recorded using circular dichroism at each concentration of berberine (0-50 µM) added to the fixed concentration of the oligo (5 µM). Each measurement was performed at room temperature. Increasing concentration of berberine induces a transition of 265 nm to 290 nm peak, indicating shift of the GQ structure from parallel to anti-parallel form. (**B.**) CD spectra recorded at 265 nm absorbance of the D4P GQ oligo without and with 50 µM berberine added. The absorbance was expressed as a function of temperature ranging from ~20 to 80°C ($\pm 2^0$). In the presence of berberine the 265 nm peak increases as the temperature is raised. Melting of the secondary structure of the oligonucleotide initiates between 50 and 60°C. Abbreviations: M.R.E., molar residue ellipticity; D4P GQ, DUX4 promoter G-quadruplex.

Introduction of berberine to the RNA CSS GQ solution has not induced any major changes in the ellipticity of the oligonucleotide's secondary structure (Figure 3.16). This suggests that this particular hairpin conformation is a weak binder of berberine.

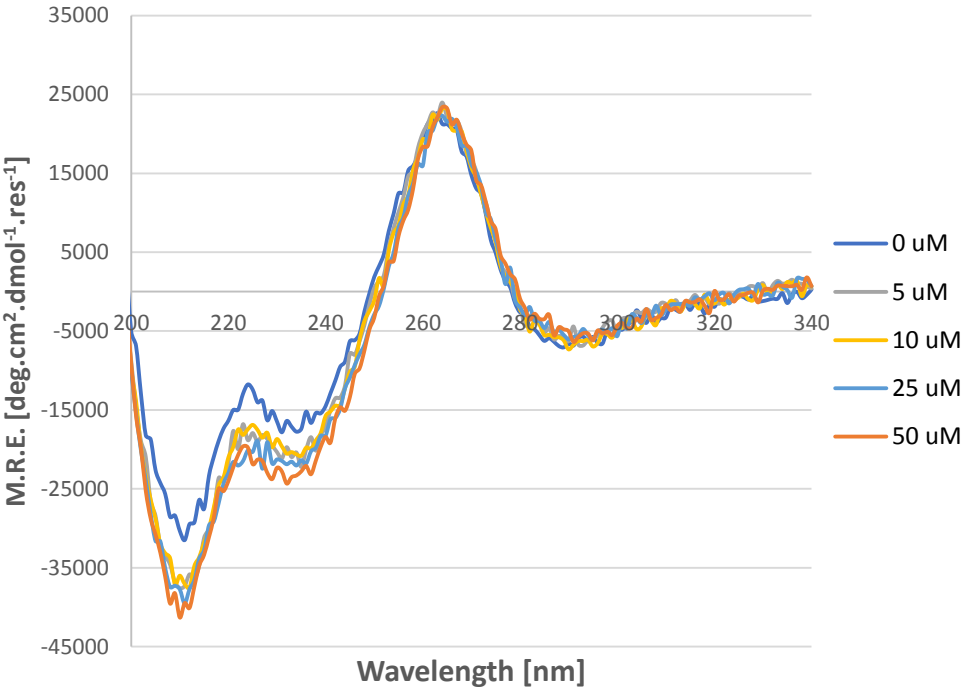


Figure 3.16 Effect of increasing berberine concentrations on secondary structure formation of CSS GQ sequence

Far-UV spectra was recorded using circular dichroism at each concentration of berberine (0-50 µM) added to the fixed concentration of the oligonucleotide (5 µM). Each measurement was performed at room temperature.

.

No significant changes in ellipticity was reordered for the RNA E1 GQ oligonucleotide upon increasing concentrations of berberine in solution (Figure 3.17 A). As a result, no apparent effects of berberine on E1 GQ structure was reported. The melting profile for E1 GQ in the presence and absence of the ligand shows to be similar and has estimated melting transition point of ~65°C and ~70°C, respectively (Figure 3.17 B). The estimated melting point indicates a relatively high stability of the GQ structure. At the maximum temperature of ~80°C, that could be achieved by the CD spectroscopy, the secondary structure was not fully dissociated as the full meting curve was not formed (Figure 3.17 B). Therefore, it was not possible to establish accurate Tm values.

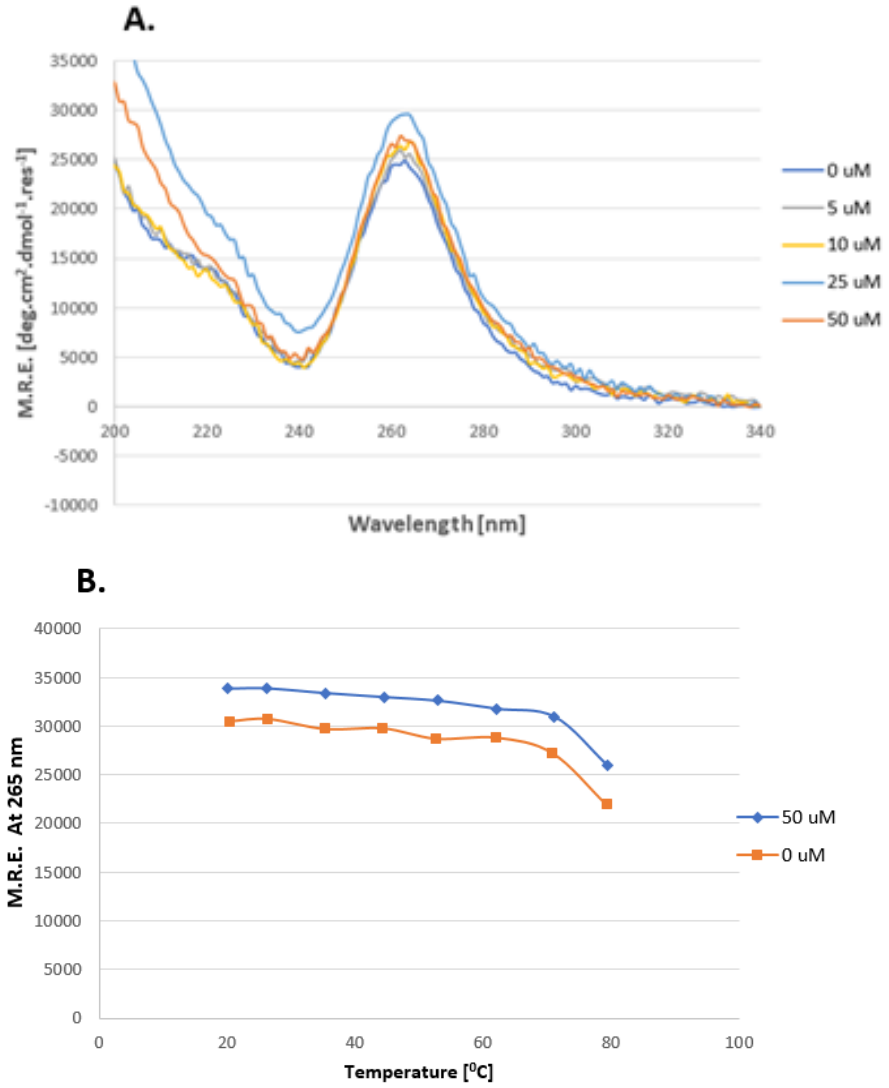
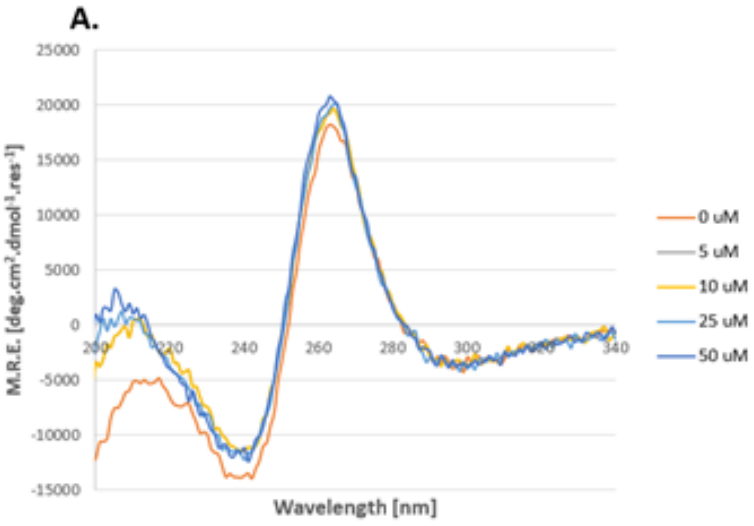


Figure 3.17 Effect of increasing berberine concentrations on secondary structure formation of the E1 GQ sequence

(**A.**) Far-UV spectra was recorded using circular dichroism at each concentration of berberine (0-50 μM) added to the fixed concentration of the oligonucleotide (5 μM). Each measurement was performed at room temperature. Increasing concentration of berberine has not changed the intensity or transition the ellipticity maximum at 265 nm and ellipticity minimum at 210 nm, indicating no change in parallel GQ structure of the oligonucleotide. (**B.**) CD spectra recorded at 265 nm absorbance of the E1 GQ oligonucleotide without and with 50 μM berberine added. The absorbance was expressed as a function of temperature ranging from ~20 to 80°C ($\pm 2^0$). Melting curve for the oligonucleotide in presence and absence of the ligand is incomplete within the limits of temperature range used. Abbreviations: M.R.E molar residue ellipticity; E1 GQ, exon 1 G-quadruplex

.

Finally, the RNA SS1 GQ oligonucleotide shows no change in the CD spectra in the presence of increasing concentrations of berberine (Figure 3.18 A). This suggests that the ligand does not have a major effect on the topology of the SS1 quadruplex. Overlapping melting profiles of the SS1 GQ in presence and absence of berberine, indicates weak binding of the compound to the structure (Figure 3.18 B). The GQ structure shows to have a modest thermodynamic stability as its melting transition begins at ~65°C. The full melting profile could not be achieved due to the device limitations (Figure 3.18 B). Therefore, the Tm for the oligonucleotide could not be measured.



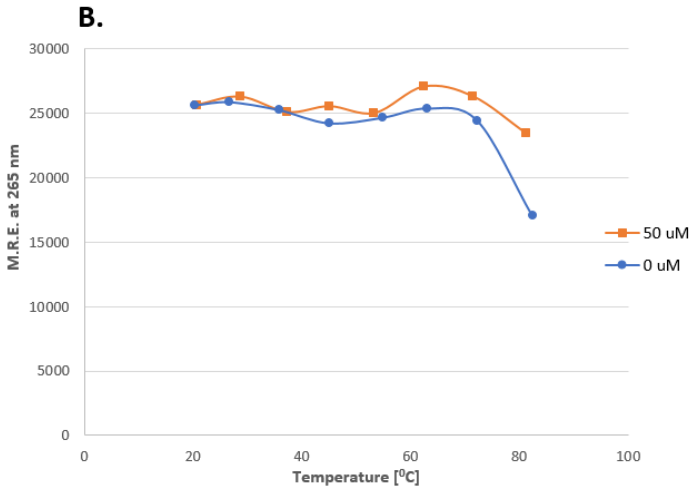


Figure 3.18 Effect of increasing berberine concentrations on secondary structure formation of the SS1 GQ sequence

(A.) Far-UV spectra was recorded using circular dichroism at each concentration of berberine (0-50 μM) added to the fixed concentration of the oligonucleotide (5 μM). Each measurement was performed at room temperature. Increasing concentration of berberine has not changed the intensity or transition the ellipticity maximum at 265 nm and ellipticity minimum at 240 nm, indicating no change in the parallel GQ structure of the oligonucleotide. (**B.**) CD spectra recorded at 265 nm absorbance of the SS1 GQ oligonucleotide without and with 50 μM berberine added. The absorbance was expressed as a function of temperature ranging from ~20 to 80°C (±2°C). Melting curve for the oligonucleotide in presence and absence of the ligand is incomplete within the limits of temperature range used. Abbreviations: M.R.E' molar residue ellipticity; DME1 GQ, DUX4 myogenic enhancer GQ. Abbreviations: M.R.E., molar residue ellipticity; SS1 GQ, splice site 1 G-quadruplex.

.

3.2.4. *Berberine binding affinity to selected DNA and RNA oligonucleotide sequences measured using UV-Vis and fluorescent spectroscopy*

3.2.4.1. UV-Vis spectroscopic analysis of berberine binding to DNA and RNA oligonucleotides

Addition of DME1 GQ and D4P GQ oligonucleotides into berberine solutions resulted in considerable λ_{max} hypochromicity of ~26 and 24% at the highest concentration of the oligonucleotides, respectively (Figure 3.19). A red shift of ~6 nm at 320 nm and ~8 nm at 420 nm for the DME1 GQ oligo was recorded, whereas the D4P GQ oligonucleotide induced a shift of ~4 nm at 320 nm and ~8 nm at 420 nm (Figure 3.19). This absorbance pattern change is indicative of end-stacking of the ligand (i.e., berberine) to the G-quadruplex structures, which further supports the data obtained from the CD and NMR analysis of these sequences.

The DNA control sequences included calf thymus (CT) DNA, long sequence oligonucleotide (LS) (26-mer) and short sequence oligonucleotide (SS) (12-mer) were annealed to form a double stranded DNA conformation. For the CT, LS and SS there was a noticeable λ_{max} hypochromicity recorded of 14.4; 13.7; and 7.0%, respectively. No red shift was recorded for any of the negative DNA controls, indicating that the berberine is not a good intercalator but rather a duplex groove binder (Figure 3.19).

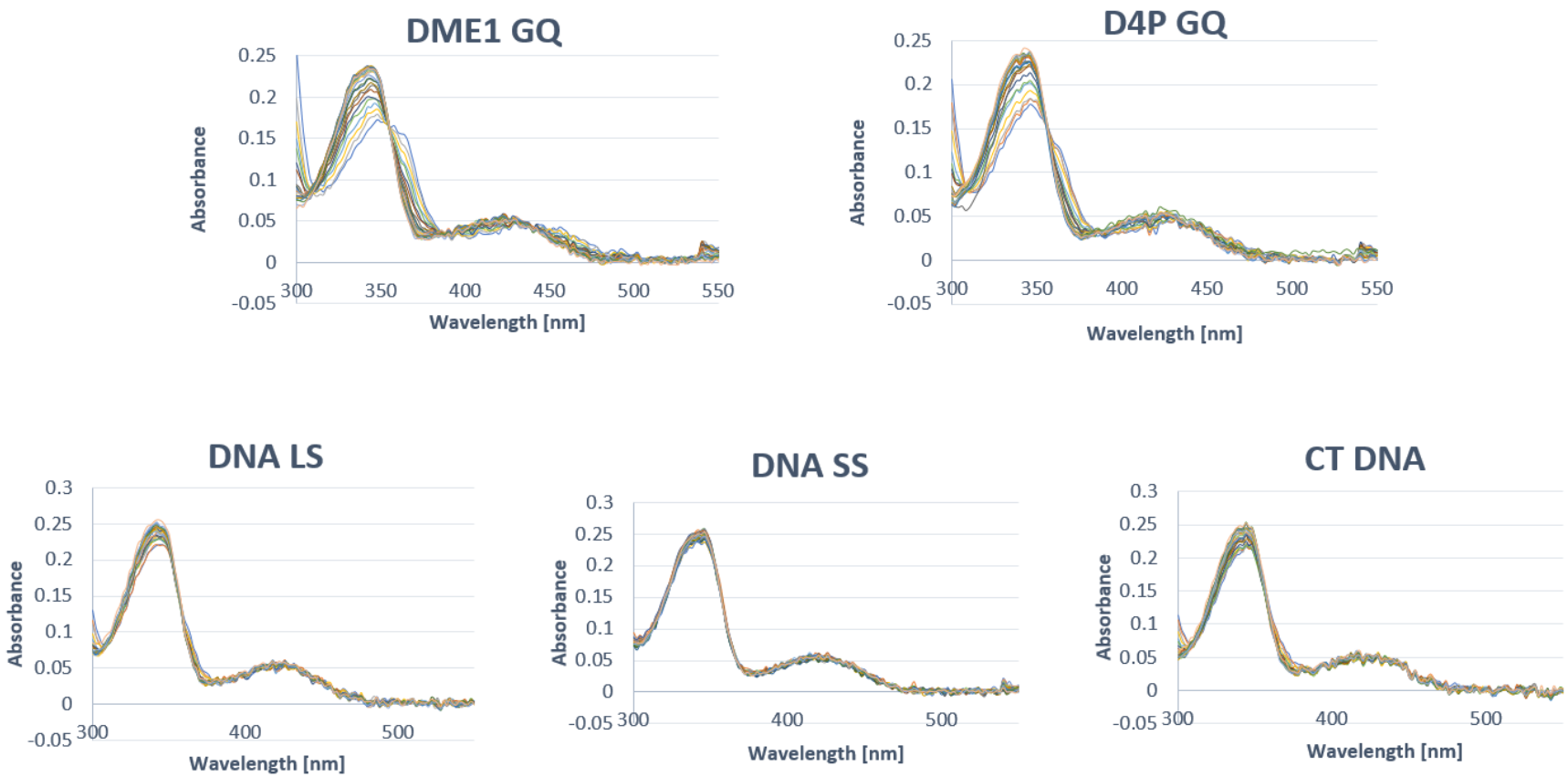


Figure 3.19 UV-Vis absorbance spectra of putative G-quadruplex forming oligonucleotides

UV-Vis absorbance spectra of 10 µM berberine in presence of DNA oligonucleotides at a range of concentrations (0-10 µM) in Tris-HCl (10 mM), KCl (100 mM) pH 7.0, measured at room temperature. The GQ forming DNA oligonucleotide sequences, include: DME1 GQ and D4P GQ. The Negative controls used are CT DNA, DNA LS and DNA SS. Abbreviations: CT, calf thymus; LS, long sequence; SS, short sequence; DME1, DUX4 myogenic enhancer 1; D4P, DUX4 promoter; GQ, G-quadruplex.

The binding mode of berberine towards CSS, E1 and SS1 GQ RNA oligonucleotides was also investigated. At the highest oligonucleotide concentration, the λ_{max} hypochromicity for the CSS, E1 and SS1 was up to 20.4; 28.0 and 30.0%, respectively (Figure 3.20). The red shift at the highest concentration of each oligonucleotide was ~6 nm at 340 nm and ~12 nm at 420 nm for the CSS GQ; ~6 at 340 nm and 0 nm at 420 for the E1 GQ; ~4 nm at 340 nm and ~12 nm at 420 nm for the SS1 GQ (Figure 3.20). The recorded absorbance signal at 420 nm measured in the presence of the E1 GQ was particularly noisy that most likely resulted in an error producing no red shift for berberine at this particular wavelength.

The RNA oligonucleotides that were used as a negative control were the RNA LS (26-mer) and SS (12-mer) that were annealed to form a duplex structure. The RNA LS and SS show a negligible λ_{max} hypochromicity of 7.8 and 8.9%, respectively (Figure 3.20). The red shift of berberine absorbance was not induced by presence of neither of the negative control oligonucleotide. This suggests that berberine is a weak groove binder of the annealed negative control RNA sequences.

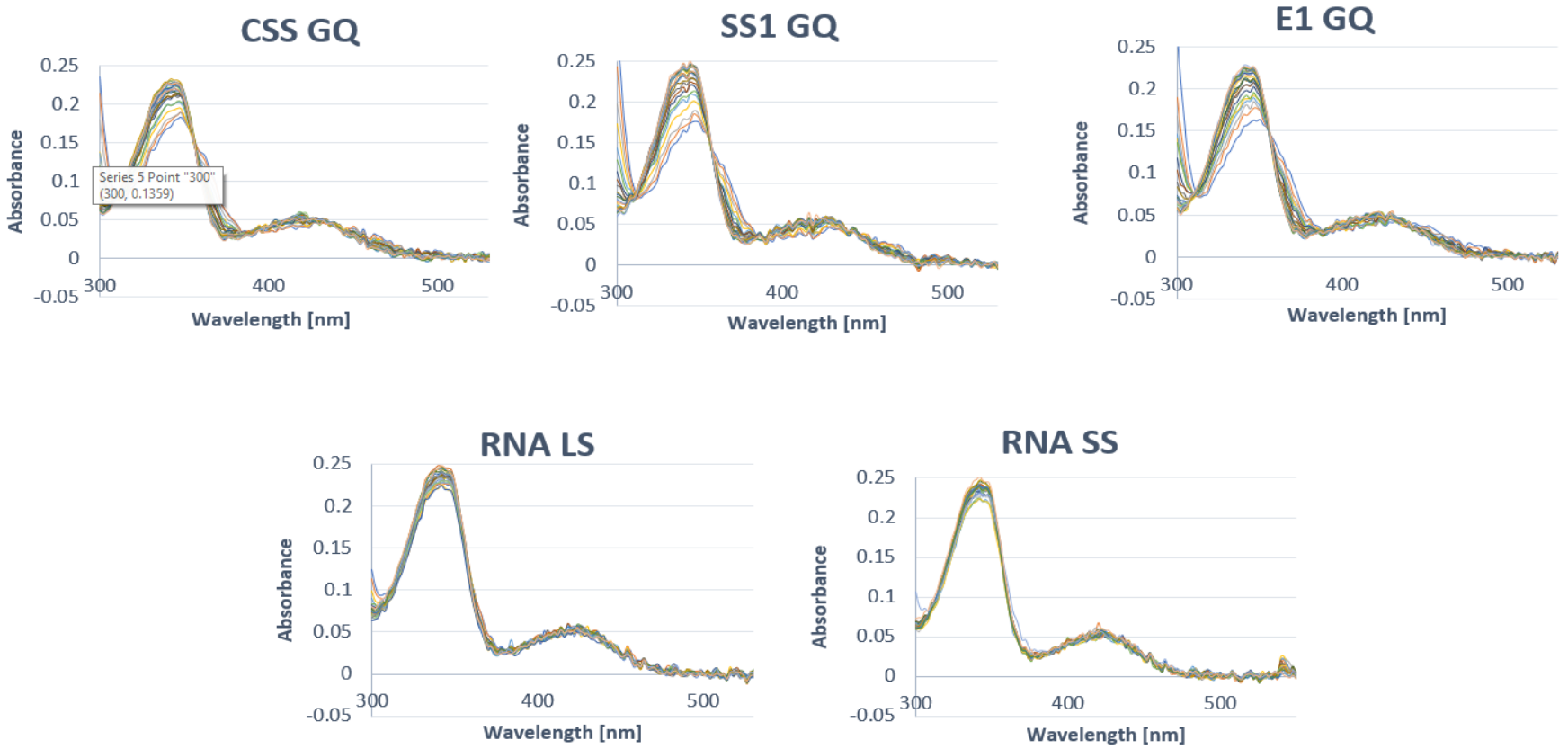


Figure 3.20 UV-Vis absorbance spectra of RNA oligonucleotides

UV-Vis of 10 µ M berberine in presence of RNA oligonucleotides at a range of concentrations (0-10 µM) in Tris-HCl (10 mM), KCl (100 mM) pH 7.0, measured at room temperature. The GQ forming DNA oligonucleotide sequences, include: CSS GQ, E1 GQ and SS1 GQ. The Negative controls used are RNA LS and RNA SS. Abbreviations: LS, long sequence; SS, short sequence; CSS, cryptic splice site; SS1, splice site 1; E1, exon 1; GQ, G-quadruplex.

3.2.4.2. Fluorescence spectroscopic titration to assess binding kinetics of berberine to secondary structures of selected DNA and RNA

The berberine alone in Tris-HCL buffer was essentially non-fluorescent. The change in fluorescent emission intensity at 525 nm was recorded as each of the oligonucleotides was titrated gradually at increasing concentrations. The gradual addition of the DME1 and D4P GQ DNA oligonucleotides resulted in up to ~7 and ~2.5-fold fluorescent increase of berberine fluorescent emission, respectively (Figure 3.21). The negative control DNA duplexes such as CT, LS and SS DNAs show minimal fluorescent emission of 0.5; 0.4 and 0.2, respectively (Figure 3.21). The results collected for the DME1 and D4P GQ sequences were plotted as a hyperbolic function using the equation listed in section 2.14.3, to calculate the binding constant (Ka). When plotted, the data points from the DME1 and D4P GQ titration forms a clear hyperbolic curves indicative of binding of berberine to these conformations (Figure 3.22 A). The calculated Ka for DME1 and D4P GQ are 1.9 ± 0.2 x 10^6 M^{-1} and 4.0 ± 0.3 x 10^5 M^{-1}, respectively (Figure 3.22 B). All of the titrated duplex DNA negative control tested (i.e., CT, LS and SS DNA) form a straight-line function indicative of low or no binding activity of berberine to these conformations (Figure 3.22 A). The calculated Ka values of ~1 ± 0.2 x 10^5 M^{-1} for the CT and LS DNA oligonucleotides indicate that these duplexes are weak berberine binders (Figure 3.22 B). The DNA SS has considerably lower binding affinity with a large error, suggesting that it does not bind berberine effectively (Figure 3.22 A).

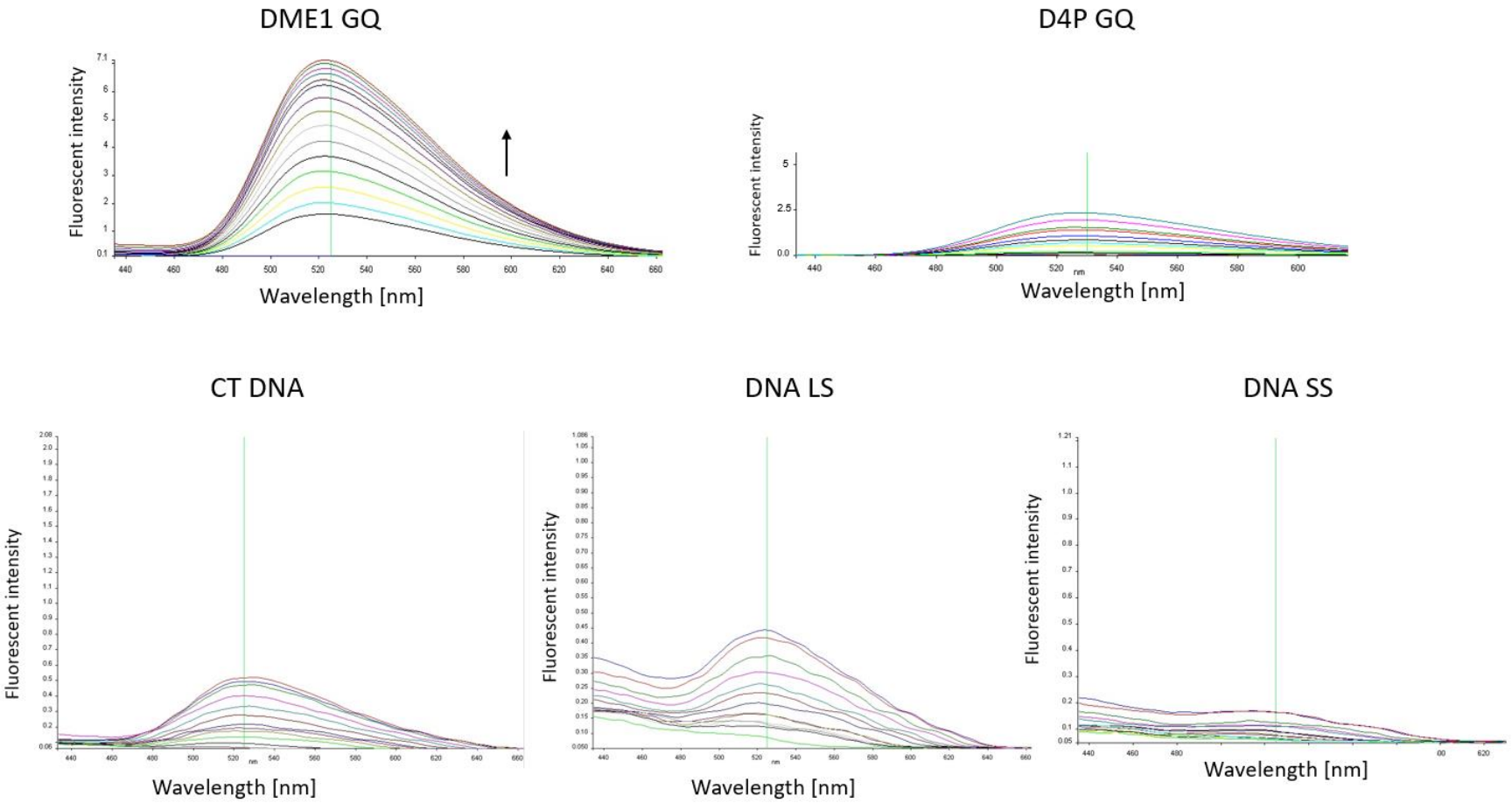
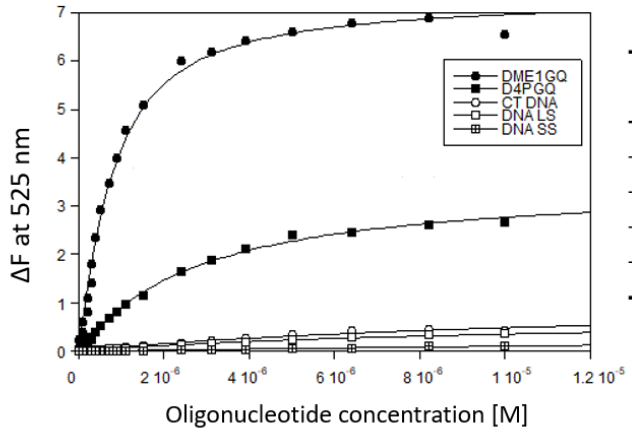


Figure 3.21 Fluorescent spectra of DNA oligonucleotides

Fluorescent emission spectra of (0.5 µM) berberine in the presence of increasing concentrations (0-10 µM) of DME1 GQ, D4P GQ, CT DNA, DNA LS and DNA SS oligos in Tris-HCl (10 mM), KCl (100 mM), measured at pH 7.0 and room temperature. Arrow indicates increase of fluorescent emission at 525 nm as the increasing amount of the oligo was titrated. Abbreviations: CT, calf thymus; LS, long sequence; SS, short sequence; DME1, DUX4 myogenic enhancer 1; D4P, DUX4 promoter; GQ, G-quadruplex

A.



B.

Oligo name:	Binding constant (Ka)
DME1 GQ	1.9 ± 0.2 x 10^6 M^{-1}
D4P GQ	4.0 ± 0.3 x 10^5 M^{-1}
CT DNA	0.9 ± 0.2 x 10^5 M^{-1}
DNA LS	9.8 ± 1.3 x 10^4 M^{-1}
DNA SS	3.0 ± 16 x 10^3 M^{-1}

Figure 3.22 Binding affinity of berberine to DNA oligonucleotides

(**A.**) Plot of ΔF emission at 522 nm of (0.5 µM) berberine versus DNA oligo concentrations (0-10 µM). The DNA oligonucleotide measured include DME1 GQ, D4P GQ, CT DNA, DNA LS and DNA SS sequence. (**B.**) To calculate the binding affinity for the plotted data, the equation from section 2.3.13 was used. Abbreviations: CT, calf thymus; LS, long sequence; SS, short sequence; DME1, DUX4 myogenic enhancer 1; D4P, DUX4 promoter; GQ, G-quadruplex.

The CSS, E1 and SS1 GQ RNA oligonucleotides were gradually titrated at increasing concentrations, resulting in an increase in fluorescent emission of 2, 3 and 4-fold, respectively (Figure 3.23). The fluorescent emission data was plotted as a hyperbolic function using the same method as the above-mentioned DNA counterparts using the equation described in section 2.143. These RNA sequences show to have a comparable Ka of 4.3-3.3 x 10^5 M^{-1} and a relatively low error of ±0.2-0.3 x 10^5 M^{-1} (Figure 3.24 B). Furthermore, the RNA secondary structure motifs show to have a comparable berberine binding affinity to the DNA D4P GQ conformation (Figure 3.22 B and 3.24 B). The highest berberine binding affinity was calculated for the DME1 DNA GQ, which was ~5-fold higher than for second highest GQ-forming sequence (i.e., D4P GQ). The relatively low Ka value calculated for the CSS GQ, could be dictated by the fact that the sequence forms a hairpin rather than the GQ structure.

The RNA SS negative control was plotted as a flat straight line and its Ka was calculated to be particularly low and with high error, indicating that this RNA duplex has a very low binding potential to berberine (Figure 3.24 B).

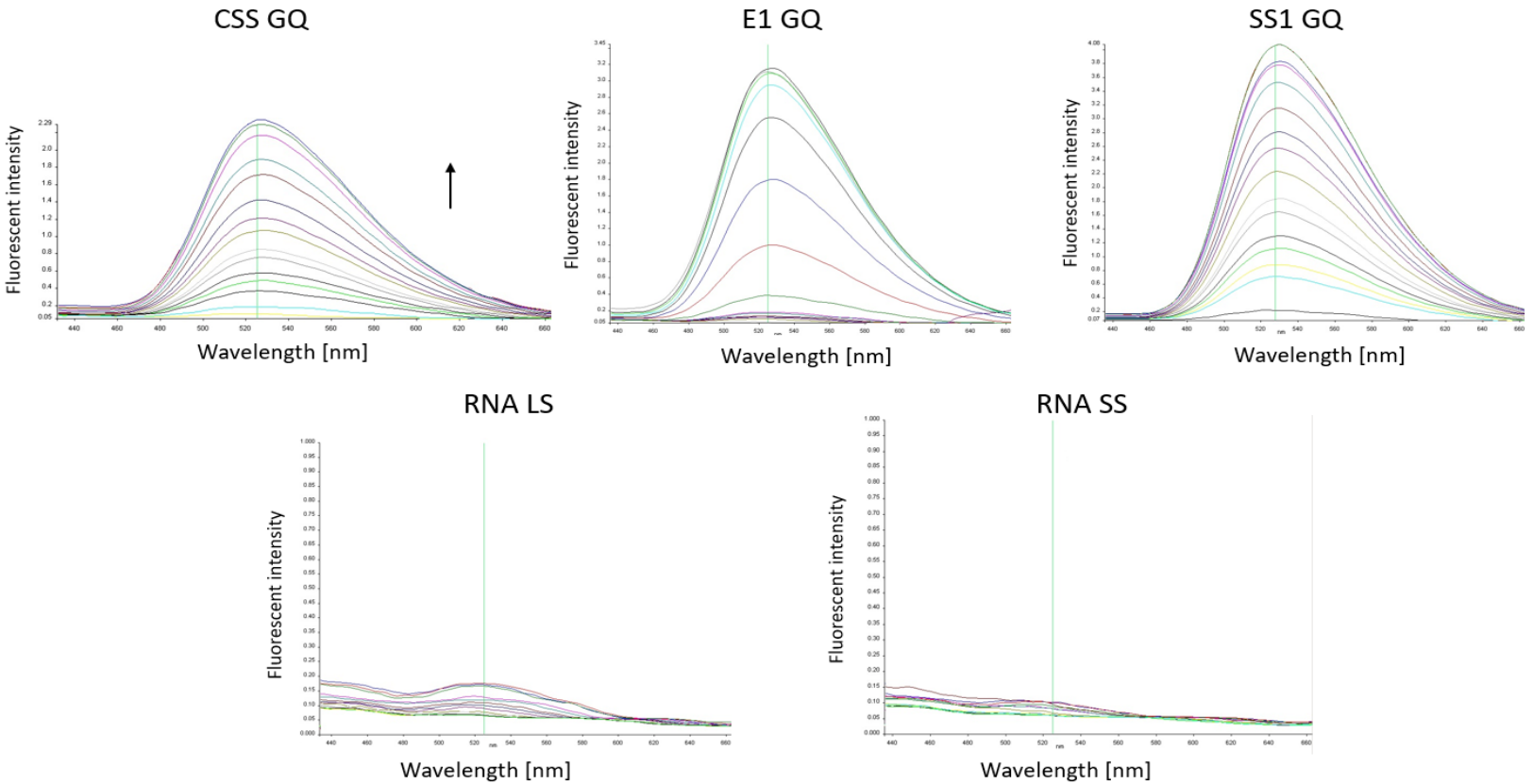
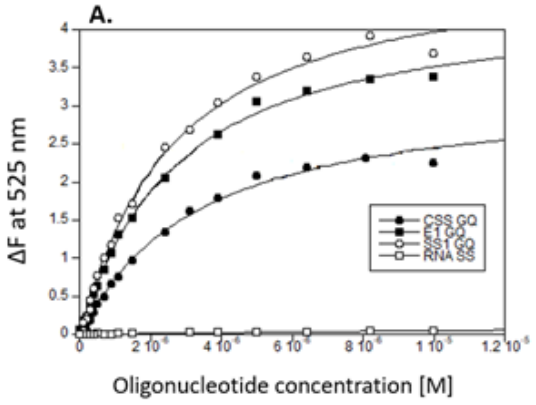


Figure 3.23 Fluorescent spectra of RNA oligonucleotides

Fluorescent emission spectra of (0.5 µM) berberine in the presence of increasing concentrations of (0-10 µM) CSS GQ, E1 GQ, SS1 GQ, RNA LS and RNA SS oligonucleotides in Tris-HCl (10 mM), KCl (100 mM) pH 7.0, measured at room temperature. Arrow indicates increase of fluorescent emission at 525 nm as the increasing amount of the oligo was titrated. Abbreviations: LS, long sequence; SS, short sequence; CSS, cryptic splice site; SS1, splice site 1; E1, exon 1; GQ, G-quadruplex.



B.

Oligo name:	Binding constant (Ka)
CSS GQ	$3.3 \pm 0.3 \times 10^5\ M^{-1}$
E1 GQ	$4.2 \pm 0.2 \times 10^5\ M^{-1}$
SS1 GQ	$4.3 \pm 0.3 \times 10^5\ M^{-1}$
RNA SS	$2.2 \pm 9.3 \times 10^4\ M^{-1}$

Figure 3.24 Binding affinity of berberine to DNA oligonucleotides

(**A.**) Plot of ΔF emission at 522 nm of (0.5 µM) berberine versus RNA oligonucleotide concentrations (0-10 µM). The RNA oligonucleotide measured include CSS GQ, E1 GQ, SS1 GQ and RNA SS sequence. (**B.**) To calculate the binding affinity for the plotted data, the equation from section 2.3.13 was used. Abbreviations: SS, short sequence; CSS, cryptic splice site; SS1, splice site 1; E1, exon 1; GQ, G-quadruplex.

.

3.3. Discussion

QGRS Mapper proved to be a very intuitive, easy to use software that allowed prediction with high confidence of six novel GQ: two of these were found in DME1 and D4P, respectively and 4 in the transcript coding sequence of DUX4. Interestingly, the D4P GQ and three of the transcript predicted GQ motifs (i.e., the CSS the E1 and SS2 GQ) contain a loop that is >7 nucleotide long.

The biophysical analyses confirm GQ formation in solution within the selected putative GQ sequences except for the CSS sequence that was found to form a hairpin. Therefore, the original GQ predicting algorithm assuming that any potential GQ forming sequences with loops longer than 7 are redundant should be re-evaluated, especially since there is a body of accumulating experimental evidence showing that extensive loops sizes do not prevent formation of stable GQs. For example, a nine-nucleotide propeller loop was found within the stable GQ structure of the human CEB25 mini-satellite locus as determined by NMR studies (Amrane et al. 2012). Furthermore, a very long 26-nucleotide loop, stabilised with a GC-based hairpin has been found to form within a promoter sequence of *hTERT* gene (Palumbo et al. 2009). Finally, the human *BCL-2* gene contains GQ-forming sequences containing a 13-nuclotide central loop (Agrawal et al. 2014). Since these restrictive algorithm were originally used to predict total number of 350000 putative GQ within the human genome, it highly likely that this number could be much greater and potentially needs to be revised (Huppert and Balasubramanian 2005). The development of a whole genome GQ screening assay that utilises combination of next-generation sequencing and the polymerase stop assay, identifies

~720000 potential GQ motifs, which further demonstrates the redundancy issue of GQ prediction by the early bioinformatic software (Chambers et al. 2015b). Indeed, the recent updates of the QGRS mapper software allows for more extensive loop analysis of up to 36 nucleotides. This could also explain why the previous study could not predict presence of the putative GQ motifs that we present in this work (Tsumagari et al. 2008).

Another limitation of the GQRS mapper is the fact that it does not account for the formation of competing secondary structures other than the GQ motifs. For instance, we have found a run of CCC within a central loop of the putative CSS GQ oligonucleotide. Although there are reports demonstrating that the secondary structures forming within the loops of GQs present within *hTERt* and *c-KIT* promoter sequences, the cytosine present in the CSS GQ sequences have been found to thermodynamically favour hairpin formation rather than a GQ (Palumbo et al. 2009; Phan et al. 2007). The addition of berberine to the CSS GQ solution did not influence the topology in favour of GQ formation, suggesting that the sequence in fact forms a highly thermodynamically stable hairpin structure. This study also demonstrates the importance of the secondary structures of loops and flanking sequences that could potential affect formation of the GQ structures. The study by Beauodin et al., has demonstrated that GQ structures are strongly inhibited by the presence of flanking or loop-related runs of CCC, particularly when present within RNA sequences. As a result, the group derived a new scoring system taking into account the neighbouring sequences containing C-tracks that can form hairpin structures and inhibit GQ formation (Beaudoin et al. 2014).

Particularly interesting was the fact that both the *DUX4* DME1 enhancer and promoter show presence of GQs within their sequence. It has been previously suggested that

hybrid GQs can be actually formed between enhancer and promoter sequences, where each of the *cis*-elements contributes half of the GQ structure bringing these two regulatory elements together (Hegyi, 2015). Since the DUX4 enhancer elements are separated a considerable distance of up to 20 kb from the target promoter sequences, it can be hypothesised that both of these elements come into proximity by the looping out of the intervening DNA sequences between them (Dean 2011). While the looping out theory has previously been supported experimentally, it is not clear if the looping out process happens first and is followed by binding of transcription factors or *vice versa*. It has been demonstrated that in the immunoglobulin heavy-chain locus, the looping-out and nuclear migration to the site of transcription occurs first and is then followed by transcription (Guo et al. 2011). As it is not clear what are the molecular mechanisms that regulate enhancer-promoter interactions to drive DUX4 expression, it would be interesting to investigate a potential role GQs in mediating this process.

Stabilisation and folding pathways for the DNA and RNA sequences into GQ structures depend upon among other things, length and composition of the loops as well as concentration of stabilising ions present (eg., K^+ and/or Na^+), and number of G-tetrads (Mullen et al. 2012; Pandey et al. 2013). Although we could not achieve a full melting curve profiles for the majority of the analysed sequences using CD, the difference in the melting transition points was apparent. Therefore, the hypothetical speculation on the thermal stability of the analysed secondary structures was made.

Initial reports investigating the GQ structure stability in the motifs containing loops ranging from 1-4 nucleotide in size, have found that the longer the loop size the less stable is the GQ structure (Guédin et al. 2010). However, in our analysed DNA GQ-

forming sequences the shorter loops did not correlate with a higher thermal stability of the analysed secondary structures. For example, the human telomeric quadruplex-forming sequences containing of 3-nucleotide loops shows the main melting transition point at ~50°C, whereas the D4P GQ sequences that contains 8-nucleotide loops has a melting transition point of ~65°C. In addition, the DME1 GQ with the 9-nucleotide loops appears to start melting at ~60°C, indicating higher thermal stability compared to the human telomer GQ structure. Indeed, the recent findings looking at the loop length stability could not determine an upper limit of GQ loop sizes (Guédin et al. 2010). In their study, the analysed GQ structures containing loops of up to 30-nucleotides still show to be thermodynamically stable and the size of the loops and Tm trends were found to be an independent variable (Guédin et al. 2010). In addition, the C-track present in the D4P GQ nucleotide could potential form a C:G stabilised secondary structure, which would contribute to the overall stability of the GQ. The hairpin forming, C:G stabilised loops have been previously reported in the context of GQ structure found in the promoter of the *hTERT* (Palumbo et al. 2009).

When comparing the thermal stability of the DNA and RNA GQs, it was found that the RNA GQ structures appear to be more stable in general. The melting transition for all of the analysed RNA GQs appears to begin at ~70°C, which is a considerably higher temperature compared to the DNA counterparts which show melting transition at ~50-65°C. In addition, the melting profile is more complete at the maximum melting temperature (~82°C) for the DNA GQs, when compared to the melting curves of the RNA GQs. These findings confirm the previous studies showing that RNA GQs are typically more stable than analogous DNA sequences (Zhang et al. 2010). For instance, the ΔG and Tm for the RNA GQs composed of 3 G-tetrads and loops of a single nucleotide, in 15

mM of cationic solution are 9 kcal/mol and 86°C, respectively, compared to 5 kcal/mol and 77°C for the DNA analogue (Lane 2012). It was suggested that the difference in the stability seen between the DNA and RNA quadruplexes is dictated by the differences in the chemistry of the two nucleic acids, rather than their folding topology (Joachimi, et al. 2009; Zhang et al. 2010). The presence of 2'-OH-group in the RNA sugar plays crucial roles in water molecules binding to the groves of the quadruplex structure, formation of H-bonds, and conformation of the sugar residue (Zhang et al. 2010). In addition, molecular modelling studies demonstrate that the 2'OH-groups provide additional intramolecular contacts with different hydrogen bond acceptors, for example, within the phosphate and backbone oxygens, 04' atoms of the ribose as well as the polar exocyclic groups of the bases (i.e, the NH_2-group), which leads to the increased stability of the parallel RNA GQ structures (Collie et al. 2010).

Although the loop size can play a role in stability of RNA and DNA GQ structures, we suggest that the composition of intervening sequences is also an important factor and that runs of cytosine residues within the RNA GQ sequence especially influence stability of the motif. The size of the largest loop within both DME1 and SS1 GQ oligonucleotides consists of 10-nucleotides. However, the DME1 GQ forms a very thermodynamically stable GQ structure, whereas the CSS1 GQ appears to form a hairpin conformation as shown by the CD and NMR analyses. It seems that the presence of the C-track within the loop sequence is a major factor contributing to the inhibition of GQ formation and that the relatively long loop did not interfere with the DME1 GQ formation significantly. C-tracks within the RNA sequences have indeed been shown previously to strongly inhibit GQ formation (Beaudoin et al 2014). Interestingly, In the D4P GQ oligonucleotide a run of cytosine residues has not interfered with the formation of the GQ structure, indicating

that perhaps the RNA-forming GQ sequences are more prone to C-track mediated inhibition of quadruplex formation.

In the absence of the GQ-binding ligand, all of the tested oligonucleotides showed to form parallel GQ structures as demonstrated by the CD and NMR analyses, expect for the CSS GQ and h. telomere GQ that formed a hairpin and an anti-parallel GQ motif, respectively. As expected, both RNA GQ forming sequences (i.e., E1 and SS1 GQ) form a parallel GQ topology due to presence of the restrictive 2'OH-group that allows for only the anti- conformation in guanosines. The addition of berberine did not influence changes in the topology and thermal stability of the RNA GQ and the analysed hairpin motif (i.e., CSS GQ). Previous attempts also failed to modulate topology of the parallel RNA GQ structures by introducing: small-molecule ligands; changes in loop configuration and composition of monovalent ions (K^+ and Na^+), further confirm that the highly stable parallel topology of the RNA GQs is mainly due to the chemical properties of the ribonucleotide residues (Collie et al. 2011). On the other hand, the DNA GQ structures show much greater flexibility in their topology due to their ability to adapt syn- and anti-glycosidic bond conformation, which allows for formation of both parallel and antiparallel GQ motifs.

The flexibility of the DNA GQ motifs was particularly apparent in the context of the D4P GQ structure when the berberine was introduced. Using the CD analysis, we have demonstrated that berberine can recognise and stabilise the antiparallel D4P GQ conformation over the parallel topology. Since the loop sequences are the key elements in dictating the preferred GQ topology, it can be suggested that connecting loops of the GQ structure play an important role in small-molecule binding. Indeed, it has been

previously demonstrated that the RHPS4 ligand has affinity for lateral and diagonal loops, which promotes formation of an antiparallel conformation of the human telomer GQ structure (Garner et al. 2009). Although it should be noted that the GQ groove-specific recognition also have been exploited as a target for drug design, the loop-mediated targeting of the GQ structures serves as an attractive alternative strategy for drug discovery (White et al. 2007).

In order to study the ligand-GQ binding interaction in terms of affinity and sequence specific recognition, it is important to support the high-resolution CD and NMR studies with thermodynamic analysis. Especially, since the melting curves from the CD data for each of the analysed oligonucleotides in the presence or absence of berberine were unable to form and therefore their thermal stability and interaction could not be thoroughly analysed. By utilising UV-Vis and fluorescent spectroscopy techniques, we could assess the thermodynamic binding profiles of berberine to the secondary structures of the selected sequences.

The UV-Vis absorption titration has demonstrated a clear red shift for all sequences that have tested positively for GQ formation (i.e., DME1, D4P, E1 and SS1 GQs) in CD and NMR analysis. Previous studies indicate that this characteristic change in absorbance is specific for GQ-ligand end stacking interaction (Arora et al. 2008; Li et al. 2017). In addition, the red absorbance shift is also associated with intercalating properties, explaining why the titration of the hairpin-forming CSS GQ also shows similar pattern of absorbance to the GQ forming sequences. The λ_{max} hypochromicity for these samples was in the 20-30% range, which is comparable to that of berberine binding to the human telomeric GQ, reported to be ~35% (Arora et al. 2008). The Arora et al., 2008 study has

estimated 1:1 berberine to GQ motif binding ratio through the end stacking interactions, suggesting that the partial binding mode and stoichiometry of berberine to the GQ-forming sequences analysed by us is also similar. The 2:1 stoichiometry of ligand (i.e., porphyrin) binding to parallel GQ structure had been previously demonstrated to have λ_{max} hypochromicity of 56-60% (Wei et al. 2006). Therefore, it can be speculated that the hypochromicity of berberine (roughly half that seen for the porphyrin-GQ binding) would be indicative of 1:1 binding ratio through end stacking interaction. The significant λ_{max} hypochormicity (~14%) combined with the lack of any red shift as recorded for the CT DNA titration, indicates groove binding interaction as it has been previously suggested by the work of Li et al., 2017.

However, the accuracy of the UV-Vis read out could be questioned by the fact the absorption curve is not perfectly smooth, indicating potential noise in the recorded signal. This could perhaps explain why the E1 GQ, despite showing a clear red shift at 340 nm, no red shift at 420 nm was produced. One explanation for the poor resolution readout could be the fact that the bandwidth of the UV-Vis apparatus (Jenway 7305) used in our experiments was set and limited to 5 nm. Other UV-Vis spectrometry devices (e.g., UV-2450), where the bandwidth can be reduced to 0.1, have up to 10 times higher resolution (Soares and Costa 1999). Therefore, a UV-Vis spectrophotometer with a higher accuracy bandwidth capability, would perhaps allow formation a smoother more accurate absorbance readout at lower berberine concentrations giving more reliable indication of berberine binding to the titrated GQ structures.

The gradual increase in the fluorescence emission of berberine upon addition of increasing concentrations of GQ, suggests transfer of the ligand from the aqueous to the

hydrophobic environment. Therefore, the end stacking on the end terminals of GQs is more likely, since the outside berberine stacking would allow continuing quenching of the chromophore fluorescence by the solvent. (Bhadra and Kumar 2011a). Indeed, molecular modelling studies show that berberine stacks on the G-tetrad terminal end human telomeric quadruplex structure (Moraca et al. 2017).

The relatively high binding affinity of berberine to DME1 GQ ($1.9 \pm 0.2 \times 10^6$ M^{-1}) was found to be comparable to the one established for the human telomeric GQ ($1.2 \pm 0.1 \times 10^6$ M^{-1}), which has also been determined using the fluorescence spectroscopy (Arora et al. 2008). However, the berberine binding affinity to the DNA GQ (i., e D4P GQ) and RNA GQ (i.e., E1 and SS1 GQ) was ~5 times lower than to the DME1 GQ, suggesting a moderate binding but no preference between DNA and RNA GQ binding. Chemical modification such as polyether-tethered berberine dimers shows a remarkably high binding to GQ structures with Ka values $>10^8$ M^{-1} (Z.-Q. Li et al. 2017). In other studies, 9-O-N-aryl/arylalkyl amino carbonyl methyl substituted berberine analogues shows 10 times higher binding affinity towards tRNA sequences that berberine. RNA GQ-specific small-molecule high affinity binders such as tetrandrine, fangchinoline and cepharanthine have also been recently established (Cui et al. 2012). This highlights that the literature on natural and synthetic small-molecule and their thermodynamic interaction with GQ sequences is expanding. Combined with the accumulating high-resolution GQ-ligand structural data, the prospect of future development of potentially highly specific therapeutics to treat a range of diseases (including FSHD) remains optimistic.

4. Chapter 4: Role of DUX4 promoter GQ (D4P GQ)

4.1. Introduction

4.1.1. *Evidence suggesting presence of GQs within the DUX4 promoter*

The identification of the *DUX4* promoter was discovered by chance in the study by Ding et al., (1998), where the binding sites of helicase-like transcription factor (HLTF) was investigated. Surprisingly, one of the HLTF binding sites had 87% homology with a region derived from the D4Z4 repeat region (Ding et al. 1998). Subsequently this led to identification of the putative promoter mapped directly upstream of the ORF of *DUX4* gene (Gabriels et al. 1999). Functionality of the putative DUX4 promoter was initially doubted since its TATAA box carries sequence mutation (i.e., TACAA). Furthermore, the lack of the introns in the gene's coding sequence and the presence of a canonical polyA signal indicated that the *DUX4* might be a pseudogene. To assess the activity of the promoter, its 191 bp fragment was cloned upstream of the luciferase reporter gene (Gabriels et al. 1999). The study showed that the DUX4 promoter is up to 30 times more active in human rhabdomyosarcoma TE671 cells compared to non-muscle HeLa cells (Gabriels et al. 1999). Further studies on the functionality of the promoter has revealed that a 2 bp mutation in the GC box sequence can significantly reduce its activity (Gabriels et al 1999). It has been also discovered that the *cis*-elements of the promoter is recognised and bound to by a multiprotein complex composed of YY1, HMGB2 and nucleolin, which have been suggested to act as a transcriptional suppressor element (Gabellini et al. 2002). YY1 is a trans-factor known to activate or repress transcription

depending on the cell type and promoter context (Bauknecht et al. 1996). In undifferentiated myoblasts YY1 represses expression of a number of muscle specific genes, whereas in differentiated myotubes the transcription factor is subjected to proteolytic degradation allowing upregulation of these genes (Walowitz et al. 1998). For example, the *DMD* gene promoter has the YY1 binding site which drives low expression of a reporter gene in C2C12 cells, and is increased five times once the cells become differentiated (Galvagni et al. 1998). HMGB2 is one of the three members of high mobility group (HMG) proteins. These proteins show affinity to DNA as well as other proteins and are thought to facilitate multiprotein complexes assembly on DNA (Bianchi and Beltrame, 1998). Interestingly, HMGB1 (the most studied protein member) interacts with human telomeric GQ DNA (Pagano et al. 2015). Nucleolin is an abundant nuclear protein and has been associated with chromatin structure regulation, rRNA maturation, rRNA transcription, ribosome assembly and cytoplasmic transport (Ginisty et al. 1999). Most importantly, nucleolin was identified by affinity chromatography to bind to the *c-MYC* promoter GQ (Gonzalez et al. 2009). Furthermore, nucleolin stabilises the promoter quadruplex and subsequently leads to repression of *c-MYC* expression *in vitro* (Gonzalez et al. 2009). The nucleolin mediated transcription repression is mediated through prevention of Sp1 activator protein from binding to its recognition sequence in the c-MYC promoter (Gonzalez et al. 2009). Since the above mentioned *DUX4* promoter associated proteins have also been found to have affinity for the GQ structures, it suggests that interaction of these *in-trans* elements could be regulated by the formation of the potential GQ motifs within the promoter sequence.

The GQ-forming sequences have been predicted, using bioinformatics, to form upstream of the functional promoter region of *DUX4* on the antisense (non-coding)

strand (Tsumagari et al. 2008). However, at the time of the analysis the outdated GQ predicting algorithm omitted any potential GQ-forming sequences with loops longer than 7-nucleotide in length (Tsumagari et al. 2008)(Discussed in Section 3.3). Furthermore, the role of the identified GQ motifs has not been studied.

4.1.2. *Study of function and targeting of promoter GQs*

The first evidence of GQ formation comes from the study of *c-MYC* oncogene expression (detailed in Section 1.5.1). The study of *c-MYC* promoter GQ served as a paradigm for subsequent research that identified a number of other GQ motifs located within the promoters of human oncogenes and the genes expressed in tumour cells , including: *BCL-2*, *h-RAS*, *b-RAF*, *HIF*, *c-KIT*, *VEGF*, *HSP90*, *RET* (Agrawal et al. 2014; Hsu et al. 2009; Kuryavyi, Phan, and Patel 2010; Ohnmacht et al. 2012; Sun et al. n 2011; Sun and Hurley 2009; Tong et al. 2011; Zidanloo et al. 2016). Examples of non-cancer related genes that also had GQ motifs recognised within their promoters, include: *HIV-1* genes and *tyrosine hydroxylase* gene that is linked with number of neurological disorders such as schizophrenia and Parkinson's (Amrane et al. 2014; Farhath et al. 2015).

The canonical approach to study promoter GQs typically involves: a) identification of the motifs using a bioinformatics approach; b) CD and NMR analysis of secondary structures of oligonucleotide sequences derived from promoter regions containing four consecutive G-tracts; c) plasmid reporter construct development that contain G-rich promoter regions or their fragments to give a better indication of the motif's role in gene expression *in vitro* (Balasubramanian et al. 2011; Małgowska et al. 2014; Podbevšek and Plavec 2016). Mutagenesis of the GQ-forming sequences was also frequently performed to disrupt formation of the motif (Dolinnaya et al. 2016) . Finally, the early proof-of-

concept experiments demonstrating *c-MYC* expression downregulation by small-molecule (e.g, porphyrin) treatment has served as an example of a new potential quadruplex-targeted therapeutic strategy (Siddiqui-Jain et al. 2002). Although the ligand-mediated, promoter GQ targeting predominately has been shown to lead to target gene downregulation, several cases exist where upregulation of gene expression was reported. Most notably, the GQ stabilisation in the promoter of *relaxin* (encoding an antifibrotic protein in rat cardiac fibroblasts) with the berberine treatment has led to upregulation of the gene (Gu et al. 2012).

Since DUX4 protein is expressed at extremely low levels and it has a high sequence homology with number of other double homeobox transcription factors (eg., DUX4c and DUX1) expressed in muscle tissue, targeting it at the protein levels might be extremely challenging (Ansseau et al. 2016; Snider et al. 2010). Therefore, targeting the DUX4 at the gene level, in principle, serves as an alternative strategy to regulate expression of the toxic transcription factor.

In this work we have already demonstrated presence and formation of novel GQ-forming sequences within the proximal region of the DUX4 promoter using CD and NMR analysis on the sense (coding) strand (Section 3.2.2). In order to better understand the potential role of the motif formation *in vitro*, we have introduced the isolated *DUX4* promoter sequence containing the GQ motif and inserted it upstream of the eGFP reporter gene. In addition, the GQ formation was either interrupted by mutagenesis of the GQ-forming sequence or stabilised by addition of a small-molecule ligand- (i.e., berberine). The effect of the mutagenesis of the GQ-forming oligonucleotide sequences was also assessed using CD analysis.

4.2. Results

4.2.1. *Design and bioinformatic analysis of the D4P GQ sequence variants*

In order to begin evaluation of the biological role of the D4P GQ, we assessed how the substitution of the oligonucleotide G-tetrad sequences would affect the formation of the motifs using the bioinformatic and biophysical analysis.

First, G-A base pair substitutions at 5’ and 3’- ends (M1D4P) or 3’-end only (M2D4P) of the GQ forming sequence was introduced (Figure 4.1 A). The QGRS mapper analysis of the D4PGQ disrupted sequences shows that the G-A substitutions introduced at each end of the oligonucleotide eliminate formation of the putative GQ motif completely as the G-score for the M1D4P oligonucleotide was zero. Interestingly, the G-A substitution of the G-tract at the 5’ of the M2D4P oligonucleotide did not result in a significant disruption of the putative GQ formation as the predicted G-score was 34 (Figure 4.1 A). The 3’-end of the M2D4PGQ oligo consists of 6 G run that can serve as two separate G-tracts providing a platform to rescue formation of the GQ structure. The C-A of the hairpin forming C-tract did not show any potential effect on the GQ formation within the D4PGQ sequence (Figure 4.1 A). The introduced G-A substitutions introduced in the M1-(ΔG=-3.91 kcal/mol) and M2D4P (ΔG=-3.80 kcal/mol) oligonucleotides was shown not to have a major influence on the stability of the hairpin formation compared to the wild-type sequences (ΔG=-3.32 kcal/mol) as shown by the mFold analysis (Figure 4.1 C).

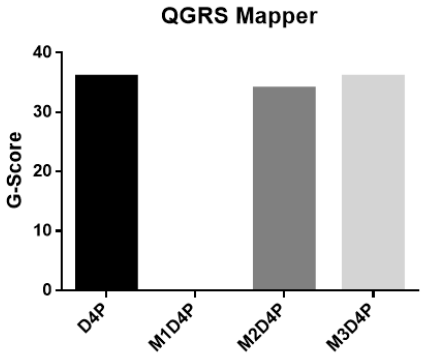
In addition, a mutation variant was created where the C-tract was also disrupted to eliminate the potential hairpin formation (M3D4P) (Figure 4.1 A). The C-tract sequence disruption by the C-A substitution showed predicted elimination of a strong hairpin

formation by the M3D4PGQ sequence (Figure 4.1 C). However, the 2 Cs present at the 3' end of the M3D4P oligonucleotide interspaced by the 13 nucleotides could still serve as a platform for a weak hairpin formation with ΔG=0.16 kcal/mol (Figure 4.1 C).

A.

Oligo name	Sequences (5'-3')
D4PGQ	CG**GGG**TG**GGG**C**GGG**CTGTCCCA**GGG**GGGCT
M1D4P	CAAAATG**GGG**C**GGG**CTGTCCCAAAAAAACT
M2D4P	CAAAATG**GGG**C**GGG**CTGTCCCA**GGG**GGGCT
M3D4P	CG**GGG**TG**GGG**C**GGG**CTGTAAAA**GGG**GGGCT

B.



C.

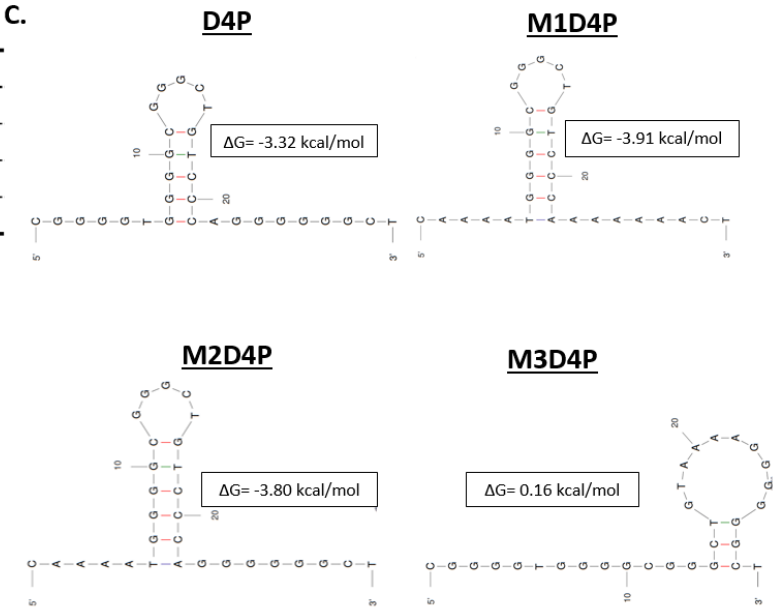


Figure 4.1 Mutagenesis of DP4 GQ oligonucleotide sequences required for secondary structure formation

(**A.**) Table listing the analysed sequences. Underlined sequences indicate substituted nucleotides from the native D4P sequence. Three altered variants were created: M1D4P, carries G to A substitution of the tetrad forming sequences at the 5' and 3' of the sequence; M2D4P, has a G to A substitution of the tetrad forming sequence at the 3' end of the oligo; M3D4P, C to G substitutions was introduced to disrupt hairpin formation. G-tetrad forming residues highlighted in bold. (**B.**) Bar chart summarising G-Scores of each of the oligonucleotide listed in table A). Oligonucleotides were analysed using QGRS Mapper. G-score is a GQ scoring system of the QGRS Mapper. Sequences with higher G-Score (≥30) make better GQ candidates. (**C.**) Mfold analysis of each sequence listed in table A). Free energy (ΔG) is indicated for each oligonucleotide. Sequences with lower ΔG have higher chance of forming hairpin structures.

4.2.2. *CD analysis of D4PGQ and its sequence variants*

The structural analysis of the D4P GQ sequences using CD revealed that it forms a stable parallel GQ structure in KP buffer solution containing of 100 mM KCl (Section 3.2.2). Each spectrum was recorded at ~20-85°C temperature range to achieve unfolding for each oligo. Refolding from ~85-20°C was also recorded for each oligo to assess their potential to form alternative secondary structures. This would show how the introduced sequence alternation of the oligonucleotide might interfere with the formation of the GQ motif.

In the M1D4P oligo where the G-tracts were substitute with runs of As, a complete absence of the parallel GQ absorbance pattern was observed; instead an ellipticity maximum at 260 nm and ellipticity minimum at 210 nm was recorded, which is indicative of a hairpin conformation (Figure 4.2). As previously shown by the mFold analysis, the run of Cs in the middle of the M1D4P oligonucleotide sequence could bond to the proximal G residues and promoter hairpin structure formation (Figure 4.1). Therefore, the QGRS Mapper and mFold bioinformatic analysis of the M1D4P, demonstrating lack of GQ structure and strong hairpin-forming potential, respectively, are strongly supported experimentally by the CD data.

The 5’ end A to G sequence substitution in the M2D4PGQ oligo did not result in a major disruption of the native GQ structure as the ellipticity maximum and minimum recorded by the CD was at 265 and 240 nm, respectively (Figure 4.2). This indicated that the parallel GQ fold could form due to the presence of the run of 6 G residues in the 3’ end of the oligonucleotide. Thus, the CD data aligns with the QGRS Mapper analysis of the M2D4PGQ predicting formation of the GQ by the oligo sequence. What is more, the

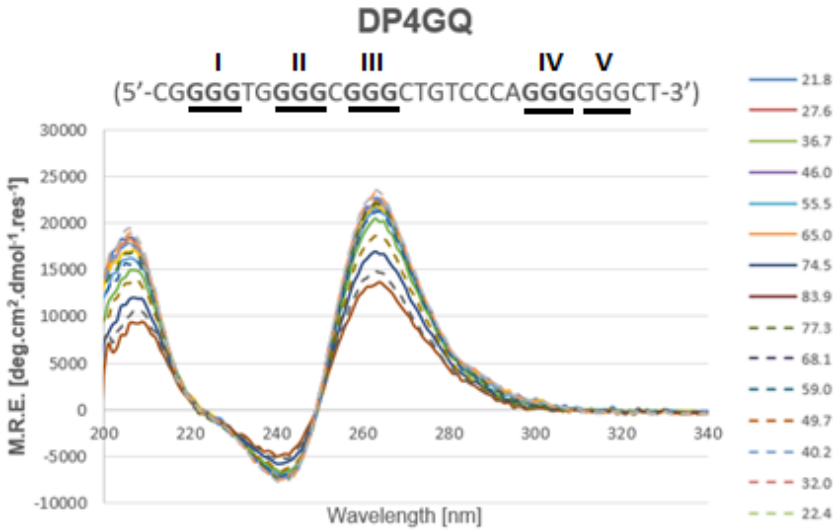
absorbance maxima of the M2D4PGQ was marginally reduced, compared with the absorption spectrum of the wild-type D4P GQ sequence (Figure 4.2). Interestingly, the G-score of QGRS Mapper for the M2D4P sequence was also only marginally lower (i.e., by 2 points) compared to the D4P GQ sequence (Figure 4.1 B), suggesting that the software can serve as a powerful tool for prediction of GQ formation *in vitro*.

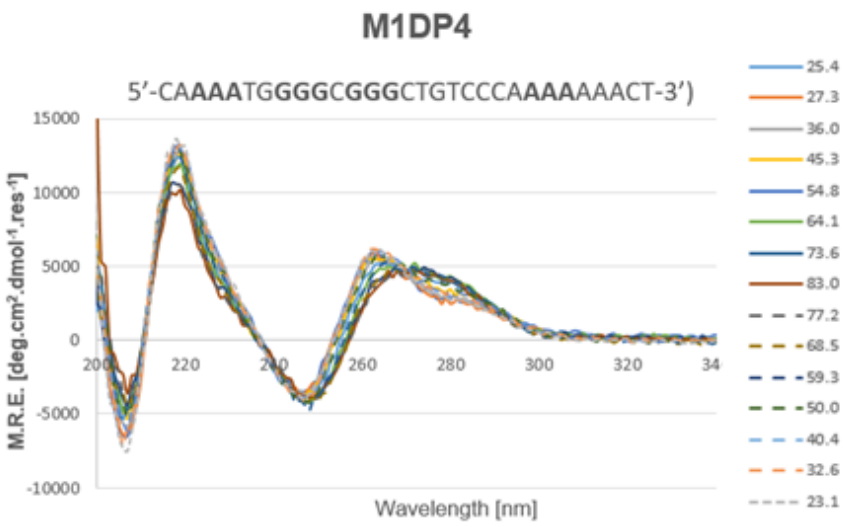
During the process of M2D4P denaturation, an additional absorbance minimum started to form at 210 nm, suggesting that potential structure transition from quadruplex to hairpin was occurring. It is possible that the introduced G to A substitutions reduced the thermodynamic stability of the GQ-motif and indirectly promoted formation for the hairpin formation upon increasing temperatures.

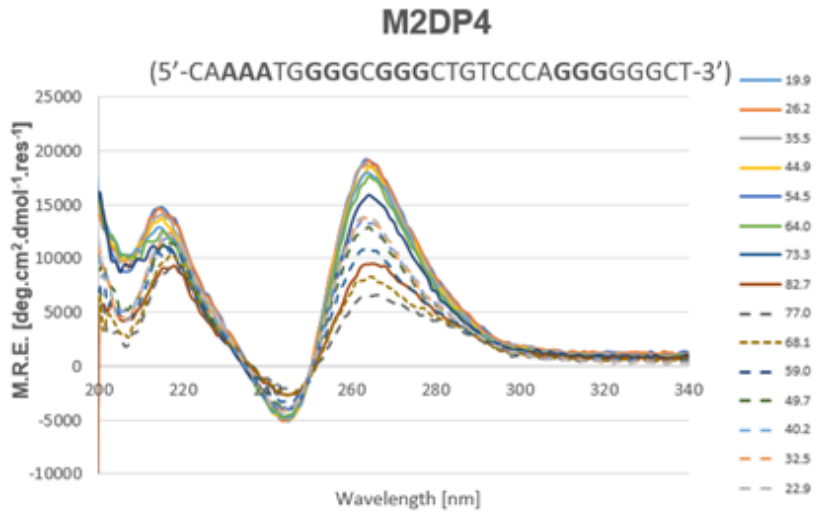
The M3D4P oligonucleotide with the intact GQ-forming sequence and C-A substitution which disrupts the hairpin structure, had a strong ellipticity maximum and minimum at 265 nm and 240 nm, indicating a parallel quadruplex formation by the oligonucleotide. The absorbance pattern of the M3D4PGQ strongly overlapped with the absorbance of the D4PGQ sequence at all measured temperatures, which could suggest that the thermal stability of the measured GQ motif was not affected by the introduced sequence alterations (Figure 4.2). Lack of negative ellipticity absorbance recorded at 210 nm further reinforces the idea that the M3D4P sequence did not form a hairpin structure

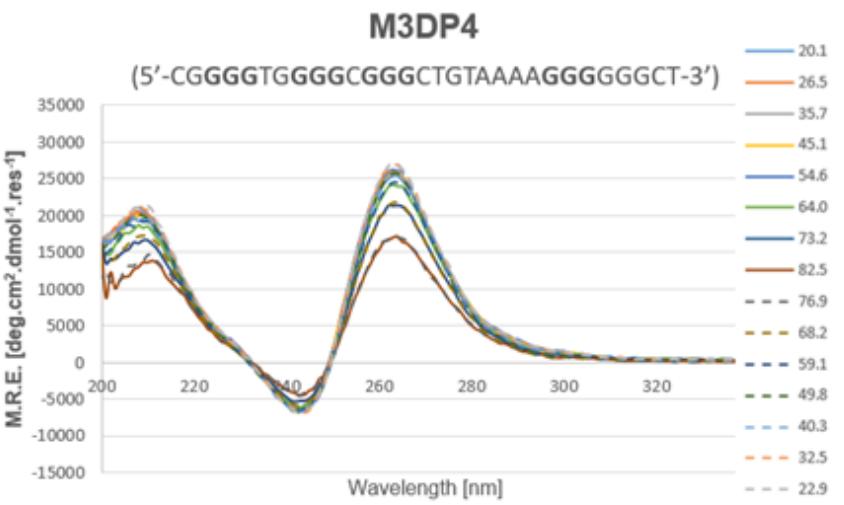
Figure 4.2 Assessment of secondary structures of the D4P and its variant sequences using far-UV CD spectra

20 µM of each DNA oligonucleotide dissolved in KP buffer with 100 mM KCl, pH 7.0 was analyzed using CD at ~20-85°C temperature range (solid lines). The refolding was assessed by reverse temperature ramp at ~85-20°C (dashed lines). The ellipticity maximum at 265 nm indicative of a parallel GQ structure formation was recorded for the D4P GQ oligonucleotide. The sequence substitution of G-tetrad forming sequences in the M1D4P variant resulted in formation of the 210 ellipticity minimum indicative of a hairpin structure, whereas the M2D4P sequence substitution preserved but thermodynamically weakened (dashed lines) the native GQ parallel structure. Disruption of the hairpin forming sequence (i.e., C-tract) strengthen the thermodynamic stability of the native motif (solid and dashed lines). The number of G-tract in the native were underlined and numbered I-V. Abbreviations: M.R.E - molar residue ellipticity


DP4GQ
I II III IV V
(5′-CGGGGTGGGGCGGGCTGTCCCAGGGGGGCT-3′)
M.R.E. [deg.cm².dmol⁻¹.res⁻¹]
Wavelength [nm]
21.8
27.6
36.7
46.0
55.5
65.0
74.5
83.9
77.3
68.1
59.0
49.7
40.2
32.0
22.4


M1DP4
5′-CAAAATGGGGCGGGCTGTCCCAAAAAACT-3′)
M.R.E. [deg.cm².dmol⁻¹.res⁻¹]
Wavelength [nm]
25.4
27.3
36.0
45.3
54.8
64.1
73.6
83.0
77.2
68.5
59.3
50.0
40.4
32.6
23.1


M2DP4
(5′-CAAAATGGGGCGGGCTGTCCCAGGGGGGCT-3′)
M.R.E. [deg.cm².dmol⁻¹.res⁻¹]
Wavelength [nm]
19.9
26.2
35.5
44.9
54.5
64.0
73.3
82.7
77.0
68.1
59.0
49.7
40.2
32.5
22.9


M3DP4
(5′-CGGGGTGGGGCGGGCTGTAAAAGGGGGGCT-3′)
M.R.E. [deg.cm².dmol⁻¹.res⁻¹]
Wavelength [nm]
20.1
26.5
35.7
45.1
54.6
64.0
73.2
82.5
76.9
68.2
59.1
49.8
40.3
32.5
22.9

The recorded CD absorbance at 265 nm (major absorbance point) for the D4P GQ sequence and its variants was expressed as a function of temperature ranging from ~20 to 85°C (Figure 4.3), in order to assess the effects of the sequence substitutions on the stability of the M1-M3D4P oligonucleotides. The full melting curve profiles could not be produced due to the limited maximum heating capacity (~85°C) of the equipment. Therefore, accurate Tm values indicative of thermal stability could not be produced. However, the considerable differences in melting transitions recorded between the oligonucleotides, suggest the presence of distinct thermodynamic profiles characteristic for each of the analysed sequence.

The native D4P GQ sequence displayed a main melting transition temperature at ~65°C and a double transition point indicating heterogeneous population of secondary structures (Figure 4.3). Supported by the bioinformatic analysis and the CD absorbance spectra, the proposed secondary structures responsible for the initial melting transition could be caused by the existence of hairpin-forming oligonucleotide species present in the solution alongside the quadruplex-forming sequences.

The melting profile of the M1D4P sequence expressed as a straight line, indicates that the oligonucleotide forms a hairpin structure derived from a single strand of nucleic acid with significantly weaker stability compared to the other analysed GQ-forming sequences (Figure 4.3).

The A-G substitution at the 3'-end of the M2D4P did not prevent the quadruplex motif formation as the double melting transition was recorded, indicative of an additional hairpin-forming sequences being present. Although the comparable melting profiles were produced for D4PGQ and M2D4PGQ sequences, the oligonucleotide carrying the

GQ disrupting sequence substitutions has a steeper and earlier descending melting curve slope, suggesting a potentially lower Tm value and therefore decreased stability compared to the native D4P GQ sequence.

A to C substitution of the hairpin-forming C-tract sequence in the M3D4P oligonucleotide, clearly eliminated a double melting transition pattern that has been recorded for the D4P GQ and M2D4P oligos (Figure 4.3). This suggests that only a homogenous population of quadruplex structure is present in the analysed solution. Furthermore, the more gradual descending melting slope of the M3D4PGQ, indicates possibly higher Tm value indicative of higher stability of the secondary structure formed by the oligo compared to the D4P GQ and M2D4P sequences.

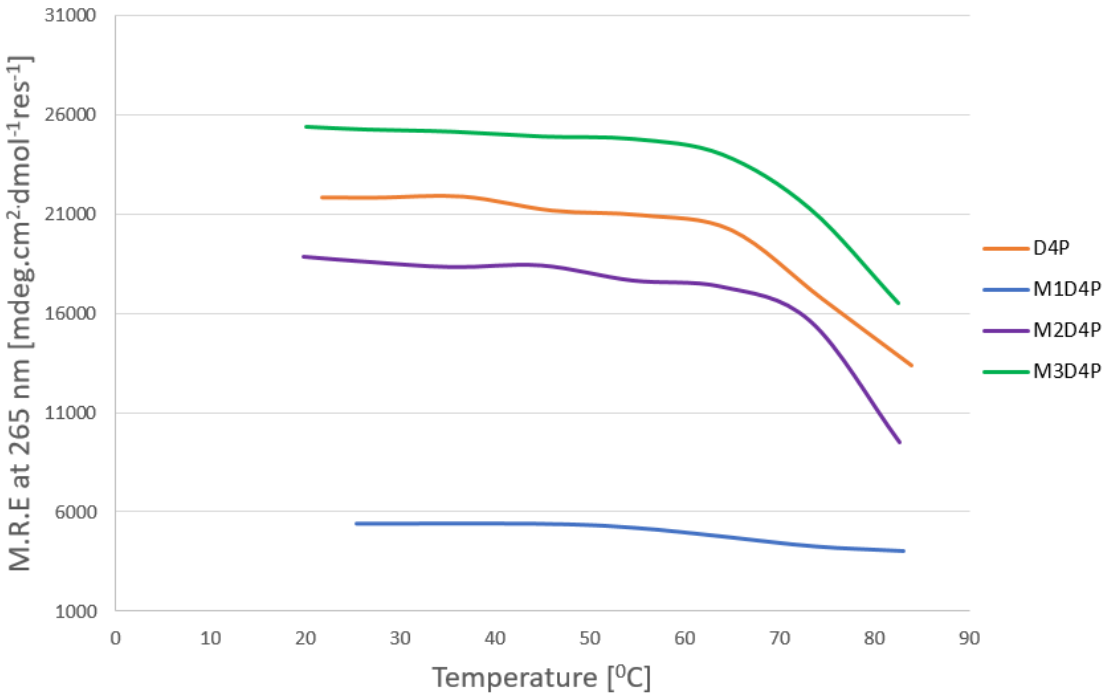


Figure 4.3 Effects of D4P sequence substitution on the melting profile

The far-UV spectra from circular dichroism at 265 nm absorbance was expressed as a function of temperature ranging from 20 to 85°C. Each DNA oligonucleotide was measured at 20 µM in KP buffer, 100 KCl, pH 7.0. Arrows indicate melting transition points. Abbreviations: M.R.E - molar residue ellipticity.

4.2.3. *NMR analysis of the D4PGQ oligo and its variants*

The CD spectra for the parallel quadruplexes and hairpin structures have a very similar pattern of absorption (Kypr et al. 2009). Since quadruplexes have a characteristic NMR spectrum that is separate from the resonances formed by other forms of nucleic acids, such as single strands, duplexes (e.g., hairpins) and other secondary structures, it can serve as a useful tool to assess the secondary structure differences induced by the base substitutions introduced to the D4P GQ sequence (Webba da Silva 2007).

The NMR analysis of the D4P GQ oligo showed a strong resonance signal detection in the 10-12 ppm range, indicating formation of the quadruplex structure by the sequence (Figure 4.4). However, a weak resonance signal was also detected between the 12-14 ppm, suggesting presence of a double strand nucleic acid structure existing within the solution. Combined with the mFold analysis, this alternative secondary structure formed by the D4P GQ sequence was most likely a hairpin motif (Figure 4.1).

The 5’ and 3’ end G to A substitutions introduced to the M1D4P demonstrated a significant reduction in the GQ-specific resonance signal between 10-12 ppm (Figure 4.4). Interestingly, a weak quadruplex specific resonance remained present, suggesting that the oligonucleotide could still perhaps form an intramolecular GQ structure composed of two oligonucleotide residues. The hairpin-specific imino proton signal appeared to be more pronounced in the M1D4P compared to the native D4P GQ sequence (Figure 4.4). Therefore, the dominant secondary structure formed by the M1D4P in solution was highly likely to be the hairpin motif.

The NMR readout for the M2D4P sequence shows only the quadruplex specific resonance signal at 10-12 ppm range (Figure 4.4). Lack of the hairpin-specific NMR

signal, suggest that the parallel GQ structures were likely to form a heterogenous population of inter- and intramolecular quadruplex motifs, rather than a mixture of hairpin and quadruplex structures as proposed by the CD analysis (Figure 4.2).

C-tract substitution with A residues in the M3D4P oligonucleotide eliminated formation of the hairpin structure as the NMR readout demonstrated presence of the imino pronton signal resonance characteristic for quadruplex secondary structure only (Figure 4.4). The NMR findings combined with the CD and bioinformatic analysis strongly indicate that the M3D4P sequence was highly likely to form a stable parallel GQ structure in solution.

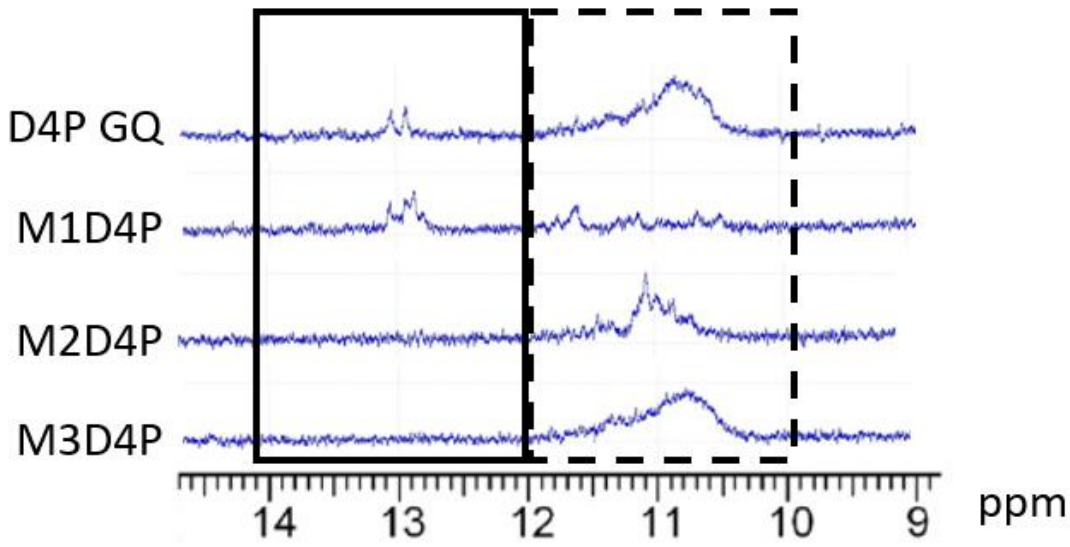


Figure 4.4 NMR analysis of the effects of D4P GQ sequence substitution on the secondary structure of the oligo

The 1D ^{1}H NMR spectra was acquired at the 9-15 ppm range to detect imino proton signal relating to hydrogen bonded bases. The Hoogsten hydrogen bonding corresponding to bonded Gs within the G-tetrad are recorded at 10-12 ppm (dashed box), whereas the Watson-Crick hydrogen boding within hairpin structure are present at >12 ppm (solid box). The DNA oligonucleotide samples were measured in 20 µM in KP buffer, 100 mM, pH 7.0 at 298 K

4.2.4. *In vitro function assessment of the DUX4 promoter fragment containing the D4PGQ sequence in the reporter plasmid system*

4.2.4.1. Sequence analysis of the DUX4 promoter fragment and 5'UTR

Plasmid constructs with inserted G-rich promoter regions or their fragments upstream of a reporter gene, have been previously found to be a useful tool in assessing the *in vivo* role of the motifs on gene expression (Sun and Hurley 2009). The DNA GQ structure can be formed by either the sense or antisense strand of the promoter sequence, while the 5'UTR can only be formed in the mRNA coded by the positive strand (Agarwal et al. 2014). Previously, the GQ-forming sequence was predicted to form ~400-500 bp upstream of DUX4 start codon on the antisense strand (Hewitt et al. 1994; Tsumagari et al. 2008). However, since the previously used algorithms to predict GQ-sequences within the *DUX4* gene would discriminate any potential candidates with loops larger than 7 nucleotides (Tsumagari et al. 2008), we asked the question whether the updated software (i.e., QGRS Mapper) would predict any GQs forming closer to the transcriptional start site (TSS) or within the 5'UTR of the *DUX4*. Therefore, we have analysed the 235 bp promoter region (including the 97 bp *DUX4* 5'UTR region) upstream of the *DUX4* start codon (Figure 4.5). The closest GQ forming sequence predicted by the QGRS Mapper was located on the sense strand, upstream of the previously determined functional *DUX4* TACAA (equivalent to TATTA box) and GC box as well as the 5'UTR sequence (Figure 4.5) (Gabriels et al. 1999; Dixit et al. 2007). No GQ motifs were predicted on the antisense strand of the *DUX4* promoter fragment or the 5'UTR (Figure 4.5).

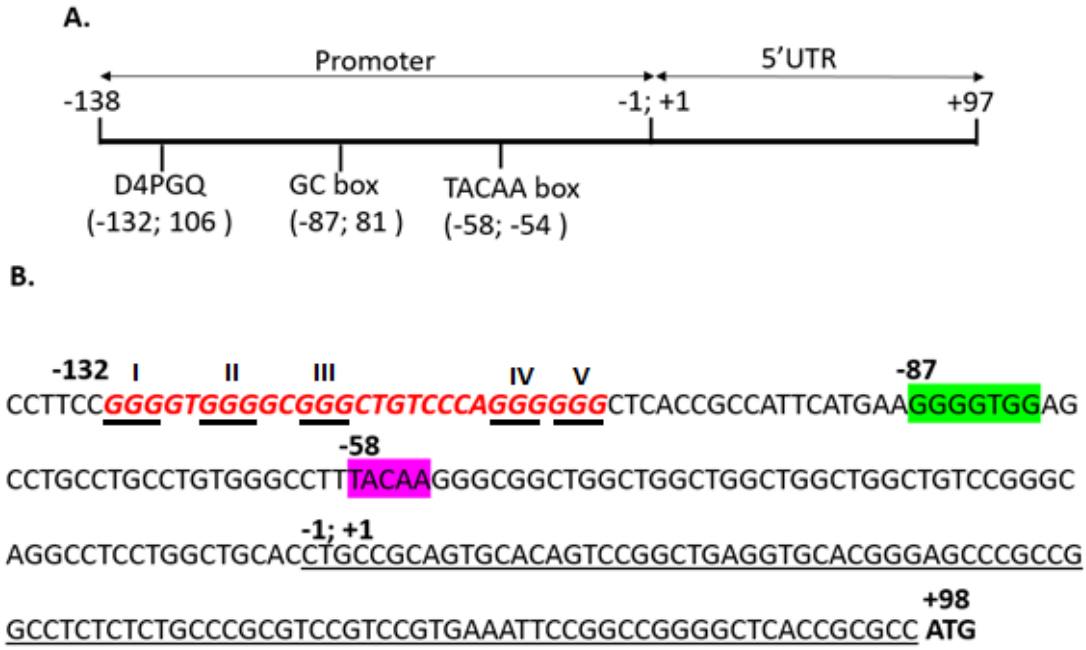


Figure 4.5 Schematic primary structure and sequence of *DUX4* promoter and 5'UTR

(**A.**) Schematic of the *DUX4* promoter and position of its key elements. 5'UTR and the transcriptional start site (TSS; -1;+1 position) are indicated. Nucleotide positions of each element are numbered relative to the TSS. (**B.**) The DNA primary sequence of *DUX4* promoter and 5'UTR. The most proximal GQ-forming sequence (D4P GQ; red font) to the TSS on the sense strand is shown. No quadruplex-forming sequences on the antisense strand of the analyzed sequence could be predicted. The key promoter elements such as GC and TACAA box are highlighted in green and purple, respectively. The 5'UTR sequence is underlined. Putative G-tracts of the D4P GQ sequence are numbered I-V. ATG denotes start codon. Nucleotide positions of each element are numbered relative to the TSS (-1;+1).

4.2.4.2. Development of expression plasmid constructs containing the 5'UTR and the DUX4 promoter fragment containing D4PGQ sequence

We have observed the presence of the GQ forming sequence in *DUX4* promoter fragment but not in the 5'UTR region (Figure 4.5). In addition, the promoter GQ-forming oligonucleotide sequence (D4P GQ) was confirmed to form the GQ motif in solution as shown by the CD and NMR analysis (section 4.2.2 and 4.2.3). The substitutions in the D4P GQ sequence were purposely introduced to prevent or weaken the GQ formation (M1-M2D4P) (section 4.2.2 and 4.2.3). In another D4P GQ oligonucleotide variant (i.e., M3D4P), the hairpin-forming sequence substitution was preformed to assess the influence of other secondary structure formation within the GQ forming region.

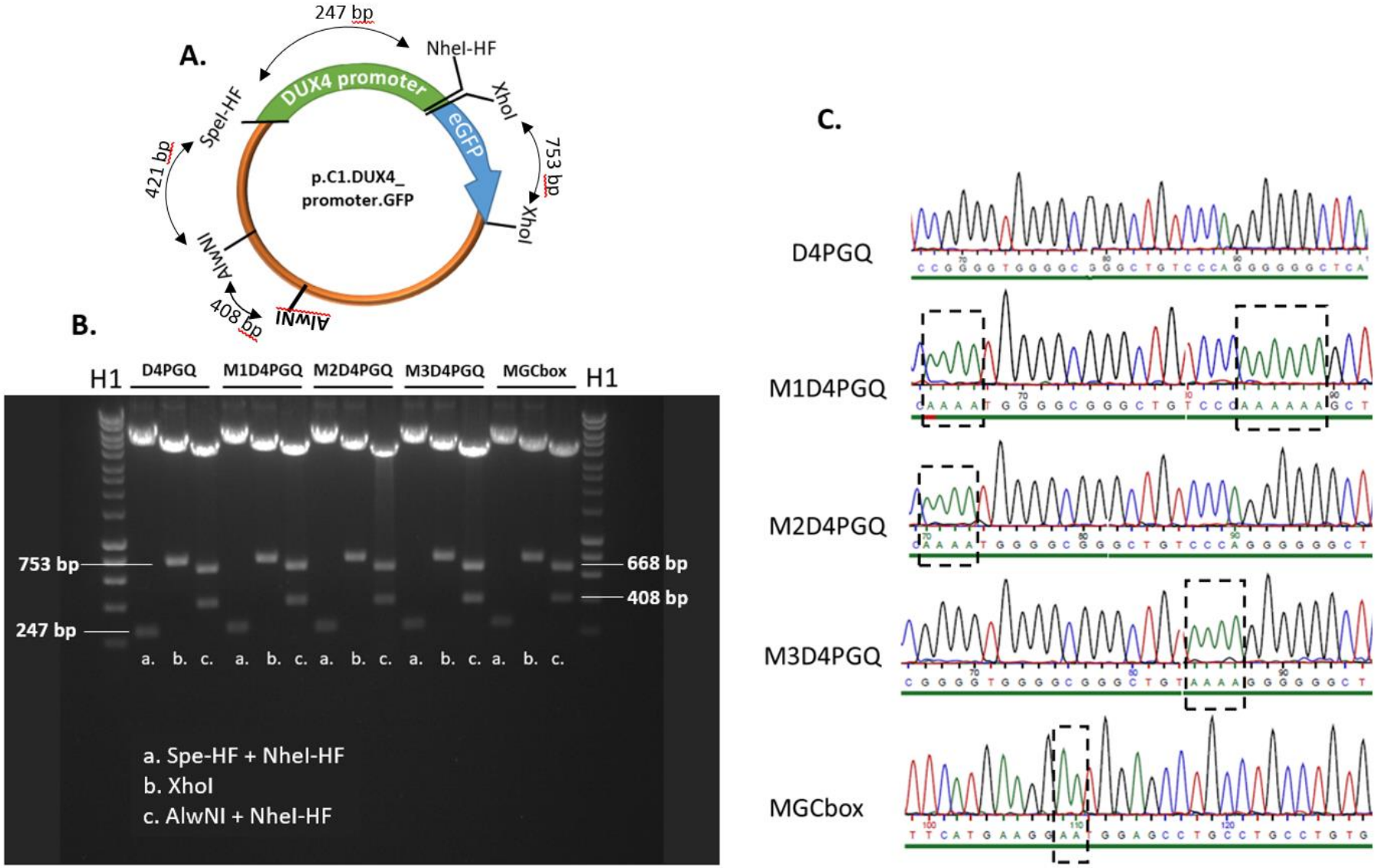
Once we have observed and confirmed the presence of GQ structure formation in the D4P GQ sequence and its variants (M1-M3D4PGQ), we then asked whether these structures or their lack may affect the expression of *DUX4*. To address this question, the wild-type, D4PGQ-containing 235 bp promoter fragment that also includes the 5'UTR of *DUX4* were designed and then synthesised by GenScript (Figure 4.5). The promoter sequence variants (M1-3D4PGQ) containing mutations to the promoter GQ forming sequence (as outlined in figure 4.1 A) were also synthesised. An additional mutation construct (MGCbox) of the *DUX4* promoter sequence was designed where the GC box sequence mutation was introduced (GGGGTGG to GGAATGG), since it has been previously shown to strongly affect the *DUX4* promoter activity (Gabriels et al. 1999).

The *DUX4* promoter fragment constructs were directionally cloned from the commercial plasmids (pUC57, GenScript) directly upstream of the eGFP-containing plasmid (Figure 4.6 A). The resulting plasmid constructs were subjected to restriction enzyme diagnostic

digest reaction to determine: (i) DUX4 promoter sequence presence using the SpeI-HF and NheI-HF restriction enzymes; (ii) correct orientation of the cloned fragment using AlwNI and NheI-HF enzymes; and (iii) presence of the eGFP reporter sequence by performing the XhoI digest (Figure 4.6 B). All of the plasmid constructs tested positive for presence and correct orientation of the inserted DUX4 promoter plasmid construct and eGFP in the restriction enzyme diagnostic digest analysis (Figure 4.6 B). Subsequently, each of the cloned DUX4 promoter fragments were sequenced to confirm whether the cloned inserts are present at the correct orientation within the plasmid and to ensure that the mutagenesis of the D4PGQ and GC box sequences has been carried out correctly. The sequencing results confirm the results of the diagnostic restriction digest reaction and show that the specific mutation of the D4PGQ and GC box sequences are present within their corresponding plasmid backbones (Figure 4.6 A)

Figure 4.6 Mutagenesis of G-quadruplex sequence within *DUX4* promoter driving eGFP expression construct

(**A.**) Schematic plasmid map demonstrating restrictive enzyme sites used for cloning and diagnostic digest of the *DUX4* promoter sequence upstream of the eGFP. Sizes of the digested products are indicated in base pairs (bp) (**B.**) Restriction enzyme digest of the p.C1.DUX4_promoter.eGFP plasmid. All samples were run on 1% agarose gel (w/v) in TAE, stained in with 10,000x SYBRSafe (Invitrogen). (**C.**) Partial DNA sequencing graph showing successful mutations of the GQ sequence within the *DUX4* promoter (M1-M3D4PGQ). MGCbox represents mutation within the GC box sequence of *DUX4* promoter. Native DUX4 promoter sequence corresponds to D4PGQ. Dashed box highlights the G to A substitutions. Abbreviations: H1, Hyperladder I (Bioline); D4PGQ, DUX4 promoter G-quadruplex; M1-3D4PGQ, mutation1-3 of D4PGQ; MGCbox, mutation of GC box.


A.
247 bp
NheI-HF
DUX4 promoter
eGFP
XhoI
753 bp
XhoI
AlwNI
408 bp
AlwNI
421 bp
SpeI-HF
p.C1.DUX4_
promoter.GFP
B.
H1
D4PGQ
M1D4PGQ
M2D4PGQ
M3D4PGQ
MGCbox
H1
753 bp
247 bp
668 bp
408 bp
a. b. c. a. b. c. a. b. c. a. b. c. a. b. c.
a. Spe-HF + NheI-HF
b. XhoI
c. AlwNI + NheI-HF
C.
D4PGQ
M1D4PGQ
M2D4PGQ
M3D4PGQ
MGCbox

4.2.4.3. Optimisation of DNA amount for optimal transfection for the DUX4 promoter fragment-containing reporter plasmid in RD CCL-136 cells

After developing reporter plasmids containing *DUX4* promoter fragment and its variants, it was important to derive optimal DNA plasmid amounts for transient transfection conditions using Lipofectamine 3000 (Invitrogen) reagent to assess the activity of the promoter *in vitro.* The human rhabdomyosarcoma (RD) CCL-136 cell have been used for the transection experiments as it has been demonstrated that the *DUX4* promoter is activated at significantly higher rates in these cells (i.e., 10-30 fold increase) compared to non-muscle tissue such as HeLa cells (Gabriels et al. 1999). In the transient transfection reaction, the plasmid containing wild-type *DUX4* promoter was used over a range of concentrations from 0.5 to 2.5 µg (Figure 4.7). 24 hours after transfection, cells were harvested for flow cytometry analysis. The eGFP protein expression detected by the flow cytometry has shown to be present at the highest levels in cells transfected with the 1 µg of the plasmid (Figure 4.7). Since the maximum transfection efficiency at 1 µg of the transfected plasmid could only achieve 37% (± 1%) of cells positive for the GFP protein expression, it can be suggested that the *DUX4* promoter activity in muscle tissue is relatively weak under the tested conditions (Figure 4.7). Other amounts of plasmid than the 1 µg resulted in significantly lower transfection efficiency (Figure 4.7). Therefore, the transfection conditions utilising 1 µg of plasmid were used in all of the subsequent transient transfection experiments involving *DUX4* promoter plasmid and its variants. Since the concentration of the Lipofectamine regent was kept constant during the experiment, the decreasing transfection efficiency could be attributed to the increasing plasmid concentration. Apart from the GFP protein, the transfected plasmid also encodes ampicillin resistance gene driven by a strong CMV promoter. Therefore,

other than the GFP protein, the cells are very likely to express large amounts of other proteins. This alone could be inducing a protein unfolded response leading to endoplasmic reticulum stress-induced cell death (Egger et al. 2007) and/or depletion of the histidine pool (essential amino acid that is limiting in tissue culture)(Salazar et al. 2016).

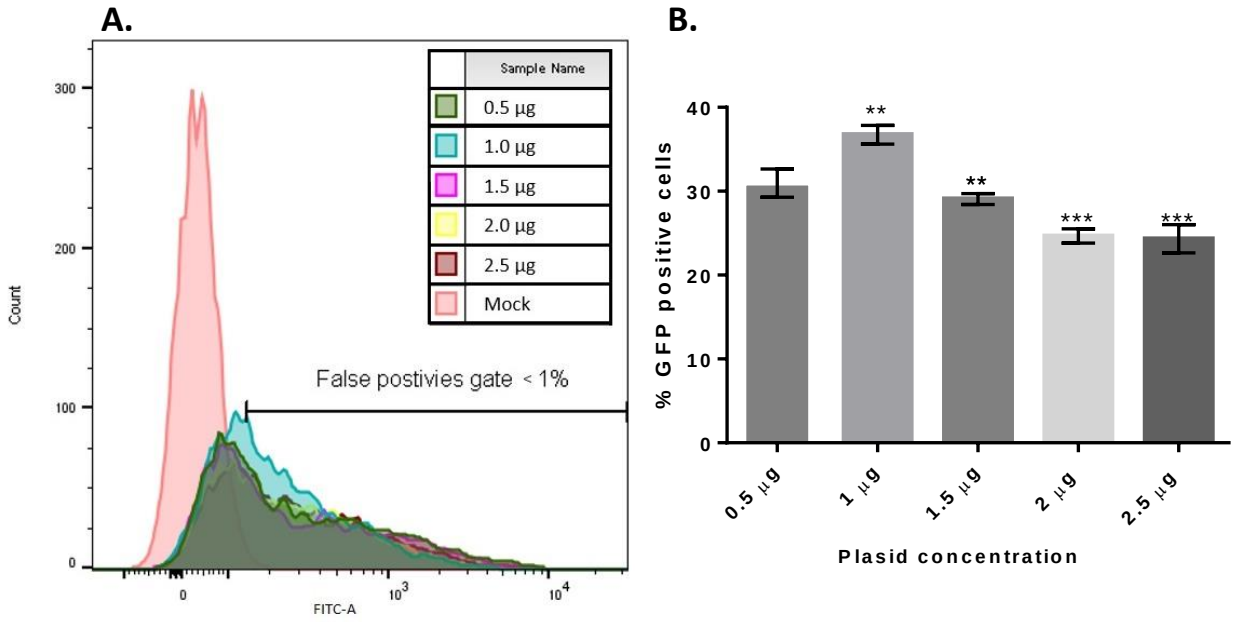


Figure 4.7 Transcriptional activity of the *DUX4* promoter

Increasing amounts of GFP expressing plasmid, driven by *DUX4* promoter was transfected into RD CCL-136 cells using Lipofectamine 3000 (Invitrogen). (**A.**) Flow cytometry plots showing count of GFP-expressing RD CCL-136 cells. Background fluorescence was set using mock samples and was below 1%. The transient GFP activity was measured in harvested cells 24 hours after transfection with flow cytometry using FITC-A channel. (**B.**) Statistical analysis of the flow cytometry data comparing effects of varying plasmid concentrations on GFP expression in RD CCL-136 cells. A one-way Anova test was performed (*p<0.05; **p<0.01 ***p<0.001); N=3

4.2.4.4. D4PGQ motif present in DUX4 promoter weakly modulates the expression of a reporter gene

Next, the presence of the sequences responsible for the formation of secondary structures (e.g., GQ or hairpin) within the *DUX4* promoter was assessed for their role in the regulation of the reporter gene expression. Five of the generated reporter plasmids, including: the wild-type promoter fragments sequence containing the GQ motif (i.e ,D4PGQ); mutated variants of the GQ-forming sequence (M1-M2D4PGQ) and hairpin-forming sequence (M3D4PGQ); as well as the GC box mutated sequence of the DUX4 promoter, were transiently transfected into the RD CCL-136 cells

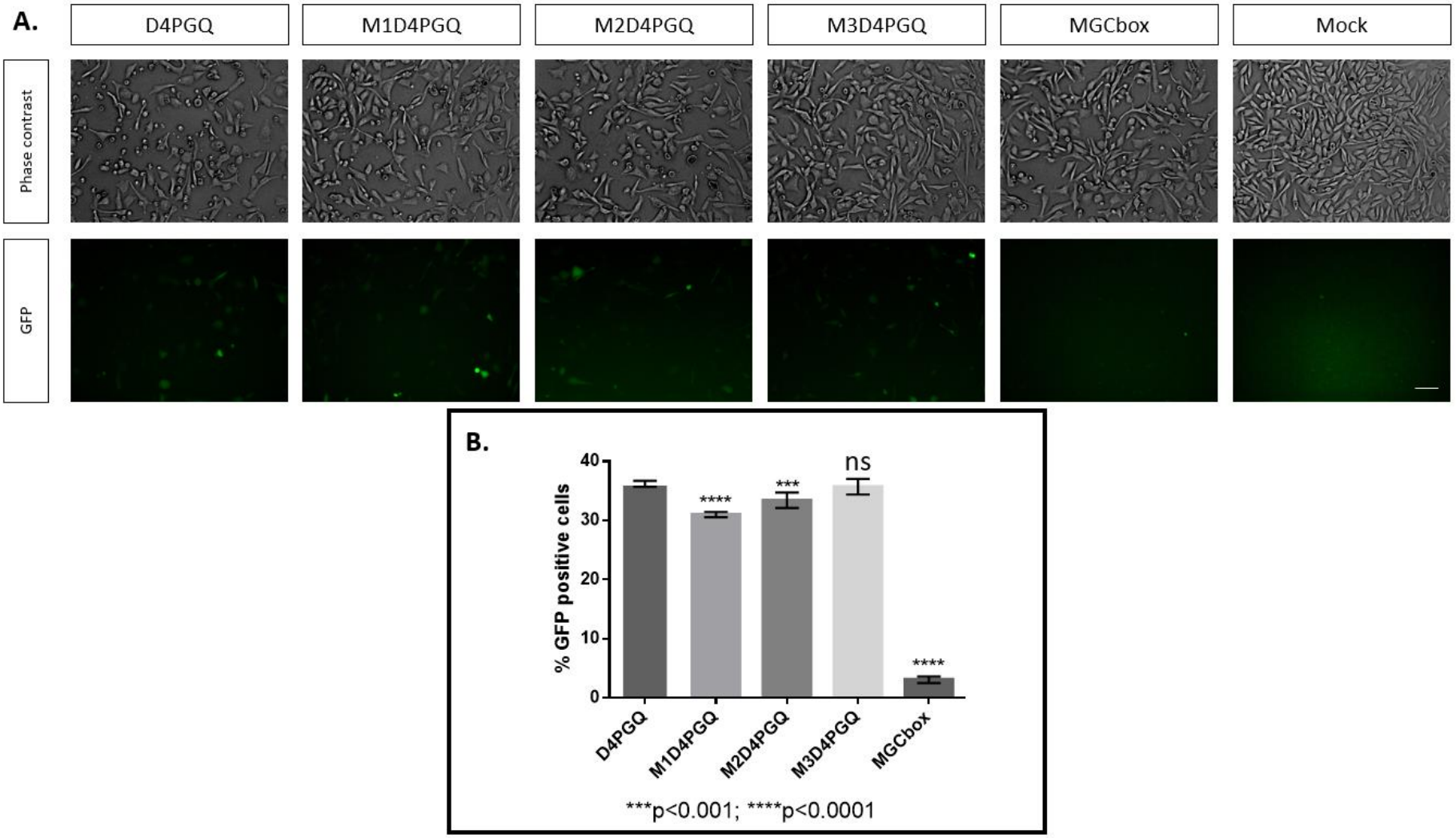
The reporter gene expression driven by the wild-type DUX4 promoter, transfected into the RD CCL 136 cells using the optimised transfection conditions, resulted in 36% (± 2%) of the cells positively expressing GFP protein as recorded by the flow cytometry (Figure 4.8 B). The plasmid construct containing the M1D4PGQ sequences mutation that significantly disrupts formation of the GQ motif as shown by the CD and NMR analysis, demonstrated a statistically significant decrease ($p<0.0001$; 31% (±2%) GFP positive cells) in the reporter gene expression compared to the wild-type construct (Figure 4.8 B). Although the M2D4PGQ plasmid carries the sequence mutations that perhaps weaken thermodynamic stability but not completely abolishes the GQ formation within the DUX4 promoter, there is a still statistically significant reduction ($p<0.001$) in the number of cells positively expressing the GFP (i.e. 33±2%), compared to the D4P GQ-containing promoter construct. Interestingly, the statistical significance of the GFP expression was found to be lower for the M1D4PGQ construct, compared to the M2D4PGQ plasmid variant, suggesting that the severity of the GQ disruption potentially

correlates with the activity of the promoter (Figure 4.8 B). Mutation of the hairpin forming sequence in M3D4PGQ construct variant did not result in any significant change of the *DUX4* promoter activity compared to the wild-type plasmid and indicates that it does not play a major role in regulating expression of the reporter gene (Figure 4.8 B). The plasmid construct carrying the mutation of the GC box sequences, downstream of the GQ forming region within the *DUX4* promoter fragment, led to a dramatic reduction in GFP expression of 30-fold compared to the wild-type plasmid construct (Figure 4.8 B).

The microscopic analysis of the GFP expression of each plasmid construct also confirms the weak activity of the wild-type expression construct demonstrated by the dim fluorescence of the protein in the transfected cells (Figure 4.8 A). The qualitative difference of the reporter gene expression between the D4P GQ and the M1-M3D4PGQ are virtually indistinguishable (Figure 4.8 A). Furthermore, the activity of the *DUX4* promoter shows to be almost completely deactivated by the mutation introduced to its GC box sequence as no GFP expression could be detected microscopically in the cells transfected with the MGCbox construct (Figure 4.8 A).

Figure 4.8 Effect of sequence mutagenesis of DUX4 promoter *in-cis* elements of gene expression

RD CCL 136 cells were transfected with the *eGPF*-containing plasmids driven by the native *DUX4* promoter sequence and its variants containing GQ sequence mutations (M1-M3D4PGQ). The construct containing mutation within the GC box of the DUX4 prompter was also analyzed (MGCbox). Lipofectamine 3000 (Invitrogen) was used to transfect 1 μg of each plasmid. Mock sample contains non-transfected cells. (**A.**) Expression of GFP was photographed using FITC channel. Scale bar: 100 μm. (**B.**) Flow cytometry measuring number of GFP positive cells expressed as a percentage of the total cell count. The FITC-A channel was used for the GFP positive cell detection. Background fluorescence was set using mock samples and was below 1%. A one-way Anova test was performed (*p<0.05; **p<0.01; ***p<0.001; ****p<0.0001;); N=6.


A.
D4PGQ
M1D4PGQ
M2D4PGQ
M3D4PGQ
MGCbox
Mock
Phase contrast
GFP
B.
% GFP positive cells
40
30
20
10
0
****
***
ns
****
D4PGQ
M1D4PGQ
M2D4PGQ
M3D4PGQ
MGCbox
***p<0.001; ****p<0.0001

4.2.4.5. Effects of berberine on DUX4 promoter activity

We have previously observed a strong affinity binding and stabilising effect of berberine on the GQ-forming D4P GQ oligonucleotide sequence in solution (section 3.2.4). Assuming that berberine also binds to the D4P GQ sequence in the *DUX4* promoter fragment *in vitro*, we asked the question how the small-ligand binding to the *DUX4* promoter GQ structure would affect gene expression. When the D4P GQ expression construct was transiently transfected and the cells were treated with berberine, there was a dramatic dose-dependent increase in the green fluorescence recorded by flow cytometry (Figure 4.9). However, when the ligand was added to the cells in the absence of the *eGFP* expressing plasmid, there was a significant fluorescent emission detected by the FITC-A channel of the flow cytometry (Figure 4.9). At the highest ligand concentration, the recorded emission was comparable to the emission from the group cells that were treated with the same ligand-concentration in the presence of the reporter plasmid (Figure 4.9). This indicates that the GFP and berberine have a very similar excitation and emission spectra. Therefore, the signal produced by the ligand interferes with the GFP readout produced by the reporter system. As a result, it is very difficult to draw any conclusions regarding the effects of berberine treatment on the activity of the *DUX4* promoter fragment using this reporter system.

The microscopy images confirm the flow cytometry data as there is a clear green fluorescence increase, especially at higher berberine concentrations (e.g., 50 and 100 µM), in groups of cells untransfected with the reporter plasmid (Figure 4.9). Although the cells transfected with the reporter plasmid do show a clear increase in the green fluorescence emission between the 5 and 20 µM berberine concentration range

compared to the cells treated with the same concentrations of the ligand in the absence of the plasmid, it is difficult to conclude whether the increase in the recorded emission is a result of the ligand-mediated upregulation of the *DUX4* promoter activity or the combined emission of berberine and GFP has caused the qualitative increase in the recorded fluorescent intensity (Figure 4.10).

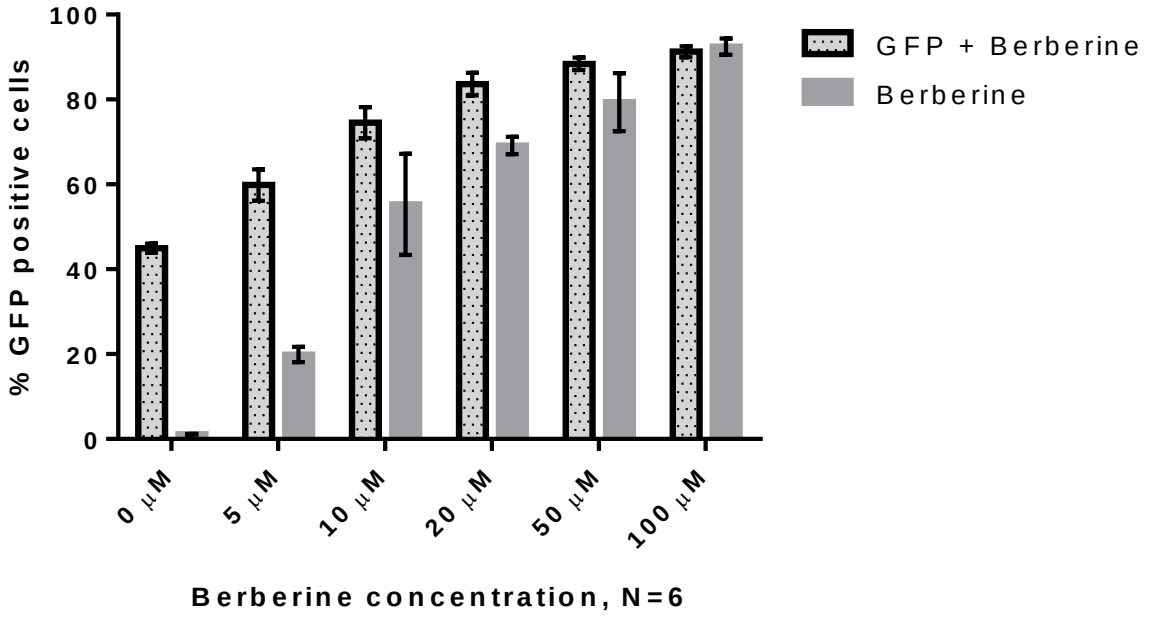
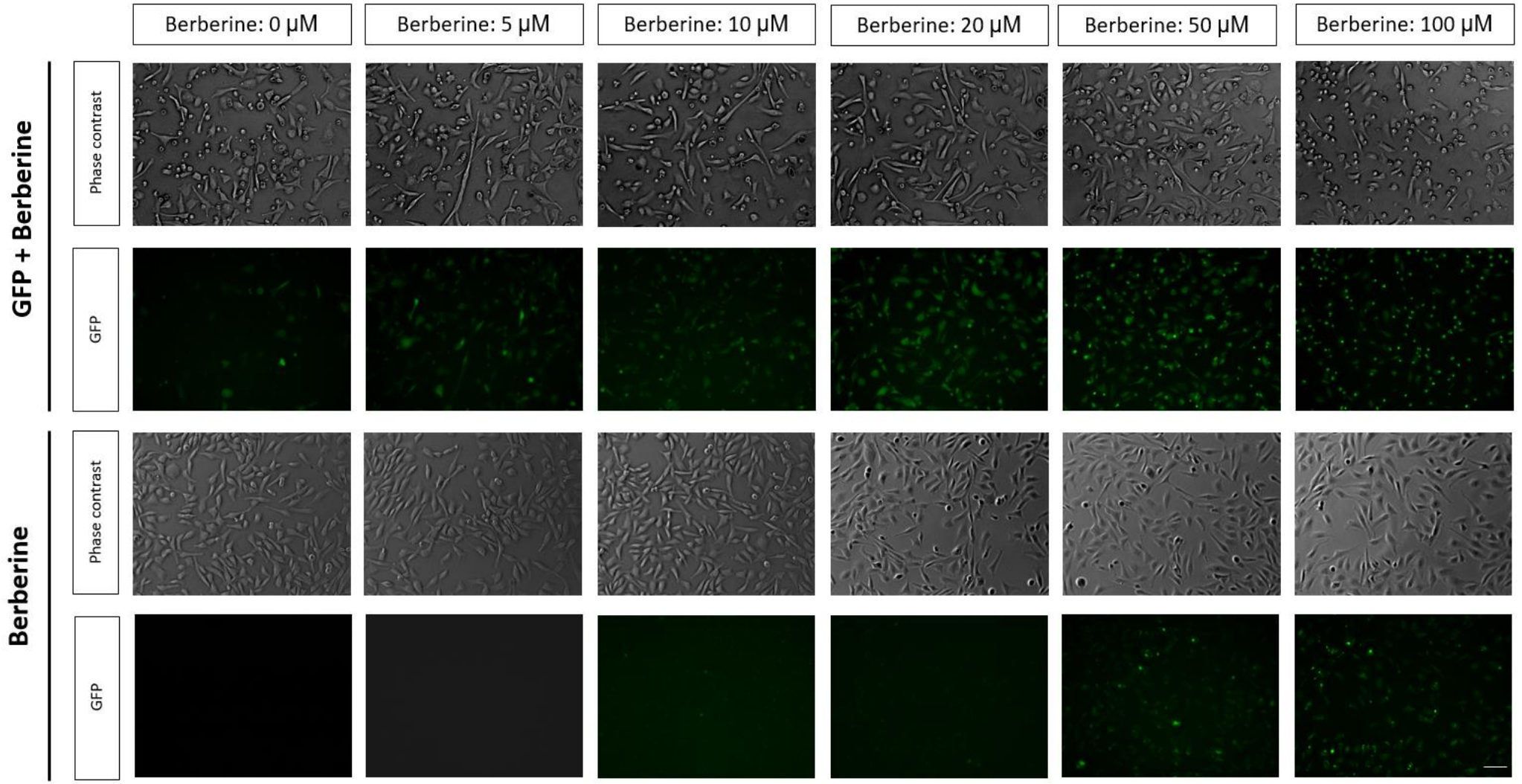


Figure 4.9 Analysis of *DUX4* promoter driven eGFP expression in presence of berberine

RD CCL 136 cells were transfected with the eGPF plasmids driven by the native *DUX4* promoter sequence (D4PGQ) in presence of increasing concentrations of berberine. Lipofectamine 3000 (Invitrogen) was used to transfect 1 μg of the plasmid. Mock sample contains non-transfected cells. Flow cytometry measuring number of eGFP positive cells expressed as a percentage of the total cell count. The FITC-A channel was used for the eGFP positive cell detection. Cells treated with berberine only were used to detect background noise. Background fluorescence produced by the cells was set using mock samples and was below 1%. A one-way Anova test was performed (*p<0.05; **p<0.01 ****p<0.0001;); N=3.

Figure 4.10 Analysis of eGFP expression driven by the DUX4 promoter sequence in presence of berberine

RD CCL 136 cells were transfected with the eGPF-containing plasmids driven by the native *DUX4* promoter sequence (D4PGQ) in presence of increasing concentrations (0-100 µM) of berberine. Lipofectamine 3000 (Invitrogen) was used to transfect 1 µg of the plasmid. Non-transfected cells treated with berberine were used to established background fluorescent. Expression of eGFP was photographed using FITC channel. Scale bar: 100 µm.


Berberine: 0 μM
Berberine: 5 μM
Berberine: 10 μM
Berberine: 20 μM
Berberine: 50 μM
Berberine: 100 μM
GFP + Berberine
Phase contrast
GFP
Berberine
Phase contrast
GFP

4.3. Discussion

In this chapter, the biophysical (i.e., CD and NMR) and bioinformatic analysis in combination with *in vitro* transient transcription assay were used to explore the potential role of GQ structure in the *DUX4* promoter. The novel GQ-forming sequence was predicted to form closely upstream of the TSS on the sense strand of the promoter sequence. Different variants of GQ-related sequence substitutions were preformed, and each was assessed for its ability to form the motif before it was cloned into the reporter plasmid. The CD and NMR of the analysed GQ-forming sequence showed to have more than four G-tracts (i.e., 

) contributing to the formation of more than one discrete quadruplexes depending on what combination of G-tracts was used. The analysed D4P GQ oligonucleotide was revealed to have potentially five G-tracts that could contribute to the GQ formation, since when the 5' end was G-A substituted, the motif structure was still able to form as shown by the bioinformatic and the biophysical analysis. Therefore, various GQ conformations are formed by the D4P GQ sequence that could potentially exist in dynamic equilibrium with each other.

This idea was previously explored in the context of the P1 promoter region of the *Bcl*-2 gene, where a 39 nucleotide G-rich sequence (i.e., Pu39), containing six G-tracts of 3-5 guanosine residues was analysed for the GQ formation as well as the effect of loops and flanking regions on the thermodynamic stability of the motif folding (Sun et al. 2014). A range of oligonucleotides with overlapping sequences that contain four G-tracts were selected for the analysis (Sun et al. 2014). Interestingly, it has been shown that the

individual short GQ-forming sequences derived from the Pu39 sequence form heterogeneous GQ structures. In addition, the GQ topology of the investigated sequences would have a mixed structure, rather than a parallel conformation , commonly found in promoter sequences (Dolinnaya et al. 2016). However, when the full 39 nucleotide Pu39 sequence was analysed, the dominant structure found *in vitro* in the folded sequence was a parallel quadruplex with two single and one 13 nucleotide loops. Interestingly, the long nucleotide sequence was the most thermodynamically stable compared to its oligonucleotide variants (Sun et al. 2014).

In addition, the ability to form a variety of GQ topologies by the same sequence regions could be instrumental in regulating gene expression by fine-tuning binding affinity to the different quadruplex structures by *in-trans*acting factors. The *Bcl-2* promoter region, contains another 28- nucleotide GQ-forming region, that can form two parallel GQ motifs under physiological conditions and exist in an equilibrium state with each other. Both motifs form protruding loops that assume hairpin conformation and can be selectively recognised by small molecules (Onel et al. 2016). Similarly, our findings suggest the D4P GQ structure that has the potential to form both a parallel GQ and hairpin structure, providing a unique conformation that distinguishes it from other promoter-based parallel GQ structures and could potentially provide a platform for recognition by specific proteins or small molecules. These unique structural characteristics of the D4P GQ sequence combined with the findings that its topology can be changed by small-molecule binding interaction (i.e., berberine), suggests an unexplored platform for development of novel, highly specific compounds that can effectively target and regulate nucleic acid secondary structure formation.

In addition to the idea that D4P GQ sequence can form single GQ confirmation and/or multiple GQ structure formed by overlapping motifs, it can be speculated that other models of *DUX4* promoter GQ structures can exist. For example, it has been previously found that upstream of the *DUX4* promoter region (i.e., in hhspm3 region) there is a GQ forming sequence (Tsumagari et al. 2008), suggesting that multiple (or at least two) GQ structures can be formed within the promoter region separated by a long nucleotide stretch. There is a well-documented example of two GQ structures forming within the promoter of the *c-KIT* gene separated by over 30-nucleotide long sequence (Hsu et al. 2009). Although the structure topology for both of these structure have been well characterised and have been found to be an overall parallel structure, the exact combined role of these two motifs is not well understood (Hsu et al. 2009). It is therefore important to analyse a larger upstream portion of the *DUX4* gene in the D4Z4 satellite region, to assess the presence of other potential GQ sequences that could play a role in the *DUX4* promoter regulation.

Although the understanding of *DUX4* promoter interaction with regulatory proteins is not yet clear, there is an accumulating number of studies shedding light on the role of *in-trans* acting elements in regulating the *DUX4* expression (Dixit et al. 2007; Himeda et al. 2014; Sharma et al. 2016). A number of well know transcription factors that regulate gene expression have been found to specifically bind to quadruplex structures (Brázda et al. 2014). One prominent example of promoter GQ binding protein is poly [ADP-ribose] polymerase 1 (PARP1). PARP1 is an zinc-finger protein, abundantly found in nucleosomes and has been implicated in DNA damage repair, chromatin remodelling and gene expression (Benjamin and Gill 1980; Soldatenkov et al. 2002). Interestingly, it has been found to bind *DUX4* promoter fragment and strongly activate its activity

leading to increased *DUX4* expression in FSHD myoblasts (Sharma et al. 2016). In addition the same report have found that the PARP1 regulates the *DUX4* promoter function in conjunction with a catalysing enzyme, the DNA methyltransferase 1 (DNMT1)(Sharma et al. 2016). Both PARP1 and DNMT1 are known for their affinity for GQ binding (Cogoi et al. 2010; Cree et al. 2016). For example, the PARP1 has also been found to be an activator of the KRAS promoter that recognises parallel quadruplex DNA (Cogoi et al. 2010). Furthermore, the PARP-1 becomes catalytically activated upon binding to GQ derived from the *c-KIT* promoter and binds the motif with high affinity. Interestingly, in a separate report, the DNMT1 has also been found to play a vital role in genome methylation and co-localised at the sites of DNA damage (Cree Simone L. et al. 2016). The DNMT1 also shows strong binding affinity to GQ structures in gene promoters *in vitro*, but becomes catalytically inhibited by the RNA GQs that specifically targets the enzyme's active site (Cree et al. 2016; Zhang et al. 2015).

Nucleolin is another nuclear protein found to play multiple cellular roles, including: ribosome biogenesis, chromatin remodelling and apoptosis (Angelov et al. 2006; Ginisty et al. 1999; He et al. 1998). Overexpression of nucleolin was also correlated with *c-MYC* promoter inhibition as measured by the luciferase expression construct assays in MCF10A cells (Cogoi et al. 2010). It has been demonstrated that the nucleolin also binds to the *c-MYC* GQ motif with high affinity and its binding to the quadruplex structure was also confirmed i*n vivo* (Cogoi et al. 2010). In addition, the nucleolin protein has been found to bind the D4Z4 repeat unit as a multiprotein complex and mediate repression of the genes expresses proximal to the D4Z4 (Gabellini et al. 2002). However, the role of nucleolin on *DUX4* expression is not clear as the mechanisms describing binding to the promoter of the toxic gene have not been investigated yet.

MyoD is a master regulator transcription factor that plays a key role in differentiation process of myoblast cells (Ishibashi et al. 2005). MyoD transcriptional activity is promoted by the formation of a heterodimer with E-proteins that recruits the complex to E-box motifs present in promoter sequences that regulate muscle gene expression (Wendt, Thomas, and Ellenberger 1998). Interestingly, the MyoD shows a high affinity binding to promoter GQ structures, which leads to significant gene expression upregulation compared to the gene expression driven by the promoter that lacks GQ-forming sequences (Shklover et al. 2010). It has been suggested by the author that the GQ helps to recruit the MyoD transcription factor to the site of the promoter (near the E-box sequence) without activating transcription (Figure 4.11). Subsequent heterodimer formation between the MyoD and E-proteins, weakens the MyoD association with the quadruplex structure which then able to bind to the neighbouring E-boxes to activate the gene expression (Figure 4.11) (Shklover et al. 2010). Interestingly, the *DUX4* promoter fragment does contain the E-box motif and MyoD has been found to bind the promoter sequence (Dixit et al. 2007). Since the D4P GQ sequence is present closely upstream of the E-box of the *DUX4* promoter, it could be speculated that MyoD may potentially also be recruited to the promoter sequence aided by the presence of the GQ structure. This is supported by the fact that the mutation of the GQ-forming sequence downregulates expression of the reporter gene, but does not completely eliminates it, indicating that the MyoD can still perform its role once it forms a heterodimer with the E-proteins and is recruited to the E-box. Conversely, the presence of the GQ structure in the *DUX4* promoter sequence increases gene expression levels as MyoD is more effectively recruited to the promoter target region. Whether MyoD has a strong binding affinity to *DUX4* promoter GQ structures is still to be established.

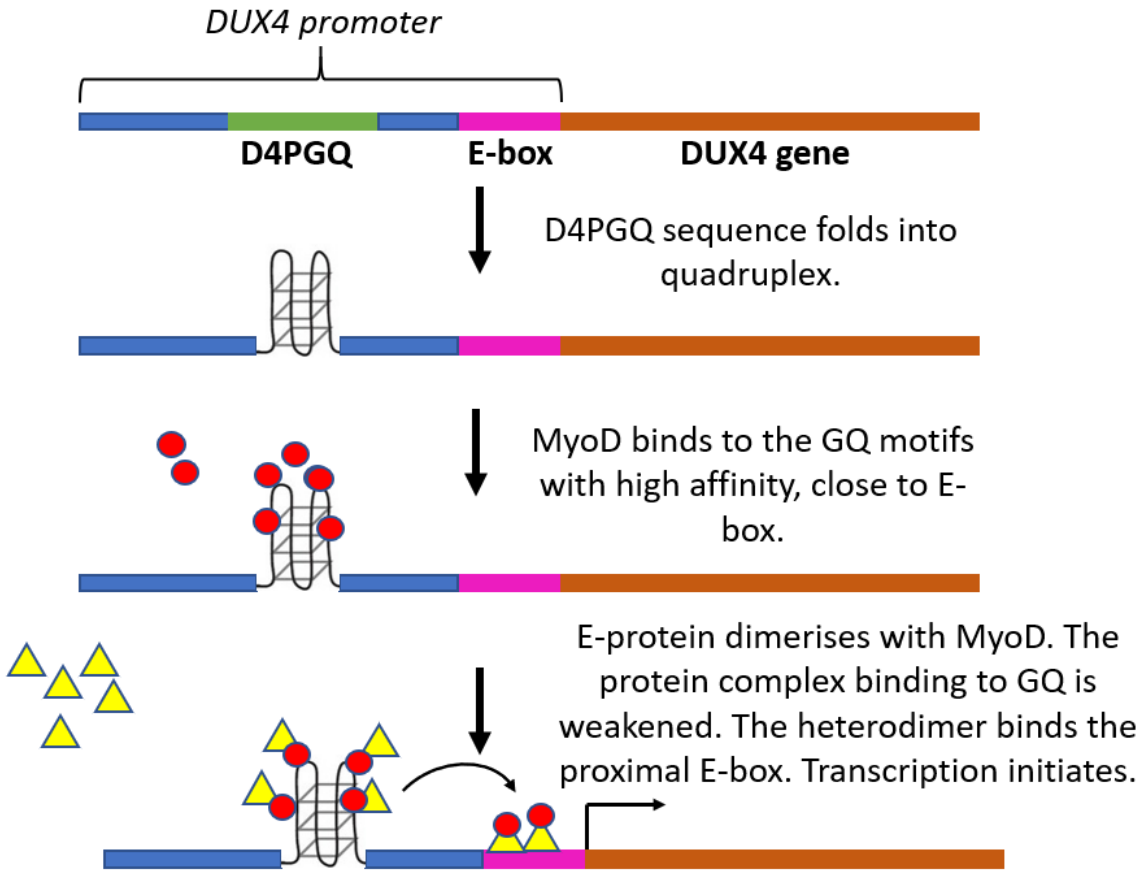


Figure 4.11 Hypothetical model of transcriptional enhancement by promoter quadruplex structure

MyoD myogenic transcription factor show high binding affinity towards promoter GQs. The MyoD GQ-mediated binding near E-box sequence promotes dimerization between MyoD and E-proteins that together activate transcription upon binding to the promoter sequence. Since MyoD and E-box binding sites have been previously mapped in *DUX4* promoter sequences (Dixit et al. 2007), the presence of the GQ within the promoter of the toxic transcription factor could have an enhancing effect on gene expression.
Red circles represent MyoD, yellow triangles represent E-protein.

Mutation of the non-canonical GC box sequence (GGGGTG -> GGAATG) in the promoter of *DUX4* has resulted in a potent inhibition of the reporter gene expression, confirming the importance of this *in-cis* element on gene expression (Gabriels et al. 1999). The

consensus GC box motif (GGGGCG) typically serves as double-stranded binding site for the SP1 transcription factor. Since the *DUX4* promoter lacks the minimal canonical GC box sequence, it was not clear how the SP1 can effectively bind to the *in-cis* acting element. Interestingly, the evidence from the genome-wide SP1 ChIP analysis showed that around 36% of the SP1 binding sites lack the consensus GC box sequences (Raiber et al. 2012). Furthermore, the majority of the SP1 binding fragments were predicted to form putative GQ structures. The SP1 strong binding affinity to GQ motifs was confirmed in the promoter sequence of the *c-KIT* gene (Raiber et al. 2012). Collectively these findings show that the SP1 transcription factor can interact with promoter sequences through interaction with the double-stranded GC box consensus sequence and/or through interaction with the GQ-forming sequences within the target promoter. However, in the context of the *DUX4* promoter, it is not clear how the potential GQ could form the non-canonical GC box sequence, since the G-tract of the *in-cis* element would be separated from the nearest GQ-forming sequences by a relatively long loop of 13 nucleotides. Nevertheless, stable GQ-structure with loops longer than 13 nucleotide have been reported, for example in the *Bcl-2* gene (Sun et al. 2014). Therefore, it would be an interesting prospect to determine whether the *DUX4* promoter GC box sequence could contribute to GQ motif formation and contribute to regulation of the promoter's activity by recruitment of SP1.

Structural motifs can form in either the sense (i.e., coding) or antisense (i.e., template) strands of the double stranded DNA. A striking asymmetric difference of putative GQ structure distribution between sense/antisense DNA strands have been reported across the human genome in the proximity (± 500b) of the TSS (Du et al. 2008). Additionally, GQ motifs have been more often found in the sense strand than the antisense strand

(Du et al. 2008). Indeed, in the *DUX4* promoter fragment only one GQ motif was predicted closely upstream of the TSS in the sense strand. Disruption of the GQ motif by the sequence mutagenesis resulted in modest downregulation of gene, suggest that the presence of the GQ in the sense strand of the *DUX4* promoter may act as an activator in gene expression. This model was supported by the fact that RNA polymerase II (RNAP II) occupancy (an important factor directly associated with gene expression) highly correlates with the putative GQ-forming regions proximal to the TSS found in the sense strand (Du et al. 2008). In the proposed model, the presence of GQ in the sense strands promotes opening of double-stranded DNA structure, making RNAP II binding more accessible and subsequently resulting in higher rates of transcription (Figure 4.12). Furthermore, lack of the GQ motifs in the antisense strands eliminates a barrier that would potentially have an inhibitory effect on RNAP II progression during mRNA synthesis (Figure 4.12)(Du et al. 2008). Indeed, the inhibitory effect of GQ formation in the antisense strand, particularly downstream of TSS, has been demonstrated experimentally (Agarwal et al. 2014). This collective data further strengthens the idea that the presence of GQ structures in promoter regions could play an important role in fine tuning gene regulation.

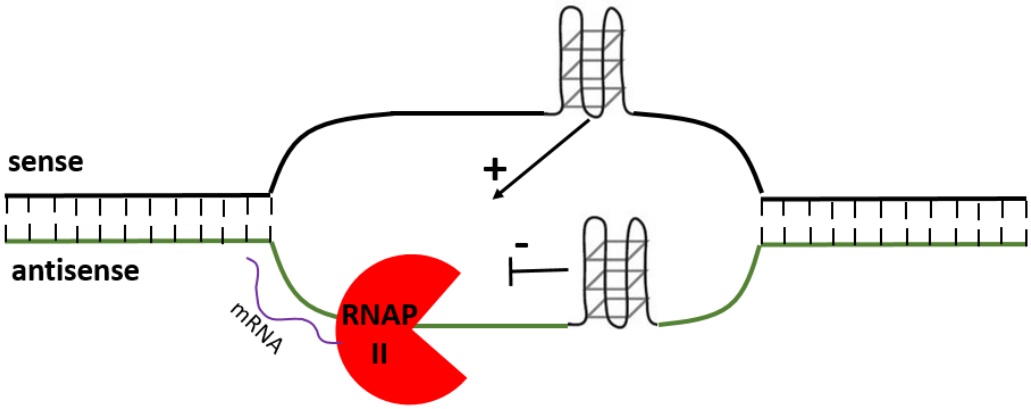


Figure 4.12 Role of G-quadruplex strand asymmetry on transcription

Quadruplex structure formed on the sense strand promotes transcription by holding the DNA helix in an open conformation, making RNAP II more accessible. GQ forming on the antisense strand creates a physical barrier for RNAP II progression leading to transcription inhibition.

It is also important to note that while the GQ motif is formed on one strand, on the complementary C-rich strand, an i-motif may form (Kuryavyi et al. 2010). Although there is a strong evidence that C-rich sequences can form i-motifs, their *in vivo* role is still a matter of an ongoing scientific debate (Zeraati et al. 2018). The fact that folding of i-motifs requires acidic conditions, resulted in major doubts of the *in vivo* role of the secondary motif (Jin et al. 2009). However, a recent discovery of antibody fragments that bind the i-motifs with high affinity and selectivity enabled visualisation of the motifs formation in nuclei of human cells (Zeraati et al. 2018). Furthermore, the small-molecule specific targeting of the i-motif in the upstream region of the *Bcl-2* promoter, has resulted in significant gene expression upregulation, supporting the argument that these secondary structures could play a major role in the regulation of expression(Cui et al. 2014). The *Bcl-2* example also supports our model where the mutagenesis of the GQ sequence in the *DUX4* promoter would also eliminate the C-rich sequence that could

potentially form an i-motif, subsequently leading to downregulation of the reporter gene expression. Although the idea of a potential i-motif regulating *DUX4* expression is highly hypothetical at this stage, it is an undoubtedly and interesting prospect for the future studies.

Although mutagenesis of GQ-forming promoter sequences was commonly applied as a tool to disrupt the motif's secondary structure formation in order to study its function, its obvious limitation is that the primary sequence of the promoter becomes changed. Therefore, it is not always clear if the changes of the reporter gene expression are predominately related to changes to the secondary structure of the primary sequence of the promoter. For this reason the studies of promoter GQ are often complimented by the use of small-molecules that can target and stabilise GQ structures with relatively high specificity and affinity. Unfortunately, the small-molecule berberine used in this study has demonstrated to have an overlapping excitation and emission spectra with the GFP readout. This introduced a significant background signal noise that made it virtually impossible to measure the GQ-stabilising effects of berberine on the reporter gene expression. Since we are interested in investigating the role of berberine effects on *DUX4* expression, changing the reporter system to luciferasefor example, would eliminate the issues of background signal noise interference with the protein readout. Alternatively, the *DUX4* expression could be also assessed on the mRNA level using RT-qPCR. Despite the above limitations, the finding of the novel GQ structure within the *DUX4* promoter provides a new platform for a scientific debate about its potential role in gene expression regulation that contributes to the field of FSHD and GQ biology.

5. Effects of berberine on *DUX4* expression

5.1. Introduction

5.1.1. *Emerging roles of RNA GQs in health and disease*

In addition to the finding that the putative DNA GQ structures are highly prevalent in the human genome regions such as telomeres, gene promoter recombination hotspots and ribosomal DNA, the bioinformatic analysis also have found GQ enrichment in 5'end of the 1st introns as well as the 5' and 3' UTRs of pre-mRNAs, suggesting an important role of these motifs in mRNA synthesis, processing and function (Huppert and Balasubramanian 2005). This bioinformatic data is further supported by the *in vitro* experiments showing that the RNA GQs could play an important role in cellular functions (Agarwala et al. 2015; Simone et al. 2015), including: gene expression regulation and telomere homeostasis. The mRNA-associated GQs are now becoming widely recognised as a crucial in *cis*-acting elements regulating processing of pre-mRNA (e.g., polyadenylation and splicing), mRNA targeting and turnover as well as translation. The genome wide sequencing provided further evidence of GQ overrepresentation in 5'UTR and introns of mRNA sequences, implying important regulatory functions (Chambers et al. 2015b). GQs also have been linked to pre-microRNA and long non-coding RNA sequences, suggesting a novel role of GQ in miRNA biogenesis and *in-trans* post-transcription gene expression regulation, respectively (Jayaraj et al. 2012; Arachchilage

et al. 2015). Interestingly, RNA GQs have been implied in immunoglobulin class switch regulation, indicating that the motif plays a role in the humeral immune system pathway (Zheng et al. 2015).

One of the major mechanism through which the RNA GQs mediate their function in cells, involves binding of protein factors such as RNA-binding proteins (RBP) that regulate topology of the secondary structure and/or recruit other protein regulators (review Brázda et al. 2014). It has been demonstrated *in vivo* that the predominant role of the RBPs is unfolding of the eukaryotic RNA GQs. Furthermore, the RNA GQ associated RBPs were also implicated in the DNA-related processes such as recombination (e.g., immunoglobulin class switch) or elongation of telomeres (Guo and Bartel 2016; Takahama et al. 2013; Zheng et al. 2015). RNA and DNA complementary sequences can also interact with each other to form hybrid GQ sequences that have been suggested to regulate transcription termination (Skourti-Stathaki et al. 2011). RBPs such as the Senataxin helicase and Xrn2 endonuclease can interact with and unwind the hybrid RNA/DNA GQs, subsequently leading to 3'end RNA product cleavage and release of RNA polymerase II from DNA (Skourti-Stathaki et al. 2011).

An increasing body of evidence suggests that impaired biological function of RNA GQ structures could lead to diseases pathogenesis as demonstrated in cancer and several neurological disorders (review: Maizels, 2015). Reports demonstrating that RNA GQs act as *cis*-regulatory elements involved in mRNA expression regulation in several cancer-related genes, include: the angiogenic factor, *VEGEF*; the tumour suppressor *TP53* and the oncogene *NRAS* (Cammas et al. 2015; Kumari et al. 2007; Marcel et al. 2011). In the context of neurological disorders, a non-coding expansion of the GQ-forming

hexanucleotide repeat (GGGGCC) in the first intron of *C9orf72* gene leads to frontotemporal dementia and/or amyotrophic lateral sclerosis (Reddy et al. 2013; Su et al. 2014). In addition, RNA GQs have been found to play a role in controlling microbial pathogenesis (Métifiot et al. 2014). For example, Epstein-Barr virus (EBV) encodes the RNA GQ binding protein, EBNA1. It has been suggested that EBNA1 is involved in an autoregulatory feedback loop that finely tunes its expression during the viral cell cycle process.

The above outlined examples along with others suggest that the RNA GQs play an important role in regulating cell biology. Advancing our understanding of biological processes regulating RNA GQ function could help to uncover mechanisms behind the pathogenesis of many diseases and contribute to the development of potential treatments.

5.1.2. *DUX4 transcript GQs and study objectives*

In this chapter we have explored the role of *DUX4* transcript GQs on gene expression. In theory, formation of the RNA GQ *in vivo* is a more likely possibility compared to the DNA GQ motifs due to their increased thermal stability in the folded state in comparison to the DNA counterparts (Dolinnaya et al. 2016). In addition, the fact that the RNA is single-stranded, means it does not have to compete with the hybridising forces of the complementary strand. Using bioinformatic and biophysical analysis, two stable RNA GQ motifs have been identified within the exon 1 and intron 1 of the *DUX4* transcript previously (section 3.2.1). Furthermore, both of the analysed sequences show a relatively strong binding affinity (Ka) towards a GQ-stabilising small-molecule i.e.,

berberine. Therefore, to investigate the effect of the *DUX4* transcript GQs on gene expression, berberine was tested in the FSHD immortalised patient cells as well as cells transiently transfected with DNA plasmids containing the *DUX4* transcript specific constructs sequence. Using RT- and RT-qPCR we demonstrate for the first time that the stabilisation of the novel *DUX4* transcript GQ motifs could have a potentially downregulatory effect on *DUX4* expression, providing a new platform for therapeutic strategy to treat the FSHD.

5.2. Results

5.2.1. *Berberine downregulates expression of DUX4 mRNA in FSHD cells*

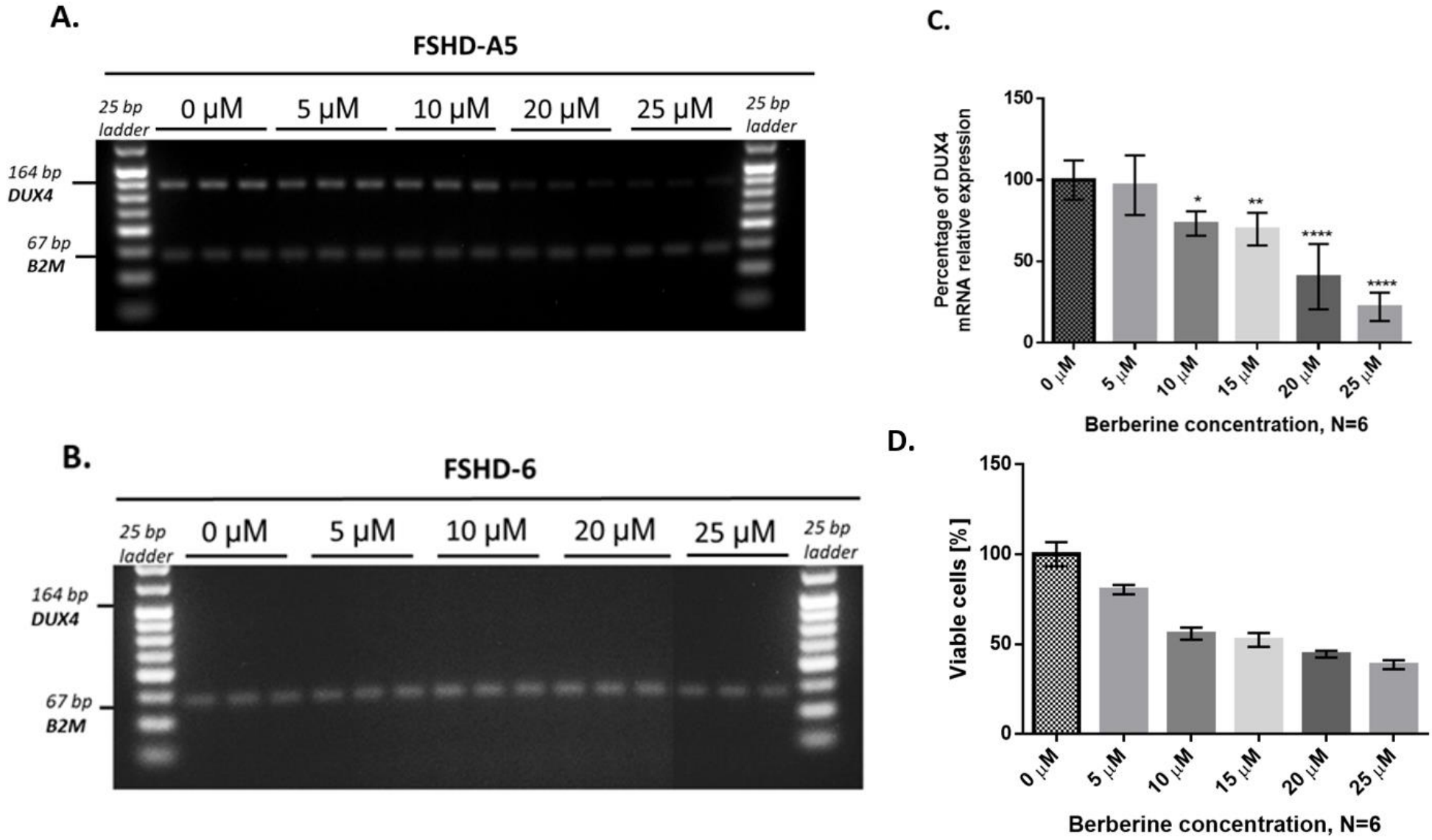
The semi-quantitative analysis of the *DUX4* expression in the FSHD-A5 cell line in the presence of berberine, demonstrated a dose dependent downregulation of the transcription factor (Figure 5.1 A.). At the lowest tested dose (5 μM), the recorded *DUX4* expression was reduced by 4% and the difference was not statistically significant (Figure 5.1 C). However, at the highest ligand concentration (25 μM), the reduction of *DUX4*-specific band was clearly apparent ($p<0.0001$) and it was quantified to be decreased by 78% compared to the untreated samples (Figure 5.1 C). No *DUX4* expression could be detected in the control cell line, FSHD-6 (Figure 5.1 B).

To assess the effects of berberine on viability of FSHD immortalised patient cells the MTT assay was performed. The results indicated that berberine significantly reduced cell viability of FSHD-A5 myotubes in a dose-dependent manner (Figure 5.1 D). However, a large proportion of the cells did not demonstrate any cytotoxic effects leaving 40% (±

2.6%) of cells viable at the highest ligand concentration (Figure 5.1 D). Therefore, normalising DUX4 expression to a well-established housekeeping gene of FSHD immortalised patient cells such a B2M was necessary to ensure that the berberine-mediated effect on DUX4 expression was not due to the cellular growth inhibition. Despite the berberine treatment and increased cytotoxicity, the treated cell populations still appear to express high levels of the house keeping gene, whereas the *DUX4* expression becomes significantly downregulated (Figure 5.1 A, B). Therefore, this strongly suggests that the effect of berberine treatment is predominately *DUX4* specific and not due to the cytotoxic effect of the ligand.

Figure 5.1 DUX4 expression in FSHD-A5 and -6 cells, and cytotoxic effects of berberine treatment

Immortalized FSHD-A5 (**A.**) and -6 (**B.**) myotubes were treated with berberine in a dose-dependent manner on the 2nd day of differentiation and incubated for an additional 48 hours before harvesting the total RNA. (**A.**) Representative RT-PCR analysis using primers detecting all DUX4 full-length isoforms in the presence of berberine. B2M was used as the housekeeping gene. The DUX4 mRNA expression was represented as a percentage of band intensities normalised to their corresponding housekeeping bands (**C.**). The RT-PCR gel image (A.) was semi-quantitatively analysed using GeneTools showing a significant decrease of DUX4 mRNA level in a dose dependent manner (Syngene). (**D.**) Immortalized FSHD myotubes (which clone) were treated with berberine at the day 2nd of differentiation. The MTT cell viability assay was performed 48 hours after treatment. Treated cells are compared to control cell with no berberine added. Independent treatment at each concentration was performed, N=6 (*p<0.05; **p<0.01; ***p<0.001, ****p<0.0001 one -way ANOVA). Error bars represent standard error of the mean (SEM).


A.
FSHD-A5
25 bp ladder
0 µM
5 µM
10 µM
20 µM
25 µM
25 bp ladder
164 bp DUX4
67 bp B2M
B.
FSHD-6
25 bp ladder
0 µM
5 µM
10 µM
20 µM
25 µM
25 bp ladder
164 bp DUX4
67 bp B2M
C.
Percentage of DUX4 mRNA relative expression
150
100
50
0
*
**
****
****
0 µM
5 µM
10 µM
15 µM
20 µM
25 µM
Berberine concentration, N=6
D.
Viable cells [%]
150
100
50
0
0 µM
5 µM
10 µM
15 µM
20 µM
25 µM
Berberine concentration, N=6

5.2.2. *Berberine treatment leads to downregulation of DUX4 downstream genes*

DUX4 is a transcription factor that affects expression of several downstream genes. Since DUX4 expression levels in muscle tissue are very low and hard to detect, the signature expression of these downstream genes can serve as a useful DUX4 biomarker (Marsollier et al. 2016). The signature downstream genes that become upregulated by DUX4 have been previously determined and include: *ZSCAN4*, *TRIM43* and *MBD3L2* (Ferreboeuf et al. 2014; Geng et al. 2012). These three genes have been selected for the RT-qPCR analysis to determine the effects of berberine on their expression. In the FSHD positive cell line all of the analysed downstream genes showed a significant downregulation at the lowest (5 μM) berberine dose, which was not observable when the *DUX4* had been detected directly by the RT-PCR (Figure 5.2). The *ZSCAN4* was significantly downregulated in a dose-dependent manner. In addition, all of the tested genes showed a highly statistically significant ($p < 0.0001$) expression downregulation at 20 and 25 μM berberine dose. Interestingly, although the expression levels of *TRIM43* were significantly reduced at 10 μM compared to the control, when compared to 5 μM, there was an increase in expression of the downstream gene (Figure 5.2). Similarly, the *MBD3L2* showed no statistically significant change at 10 μM, whereas it appears to be downregulated at a lower dose of 5 μM (Figure 5.2). No detectable levels of expression in any of the analysed downstream genes could be detected in the FSHD-6 negative control cell line (data not shown).

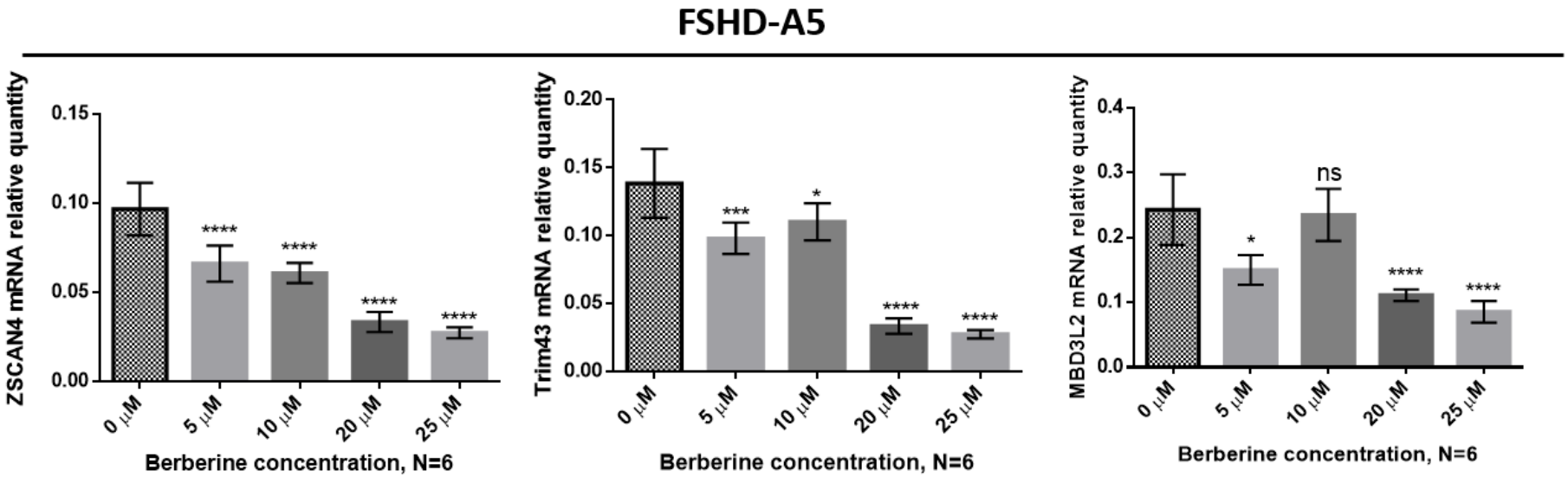


Figure 5.2 Expression of the genes downstream of DUX4 in FSHD-A5 cells in the presence of berberine

Expression levels of (**A.**) *ZSCAN4*, (**B.**) *TRIM43* and (**C.**) *MBD3L2* were measured by RT-qPCR in FSHD-A5 and -6 myotubes treated with the GQ ligand on the 2nd day of and harvested on the 4th day of differentiation. B2M was used as a housekeeping gene. Independent treatment at each concentration was performed, N=6 (*p<0.05; **p<0.01; ***p<0.001; ****p<0.0001, one -way ANOVA). Error bars represent SEM.

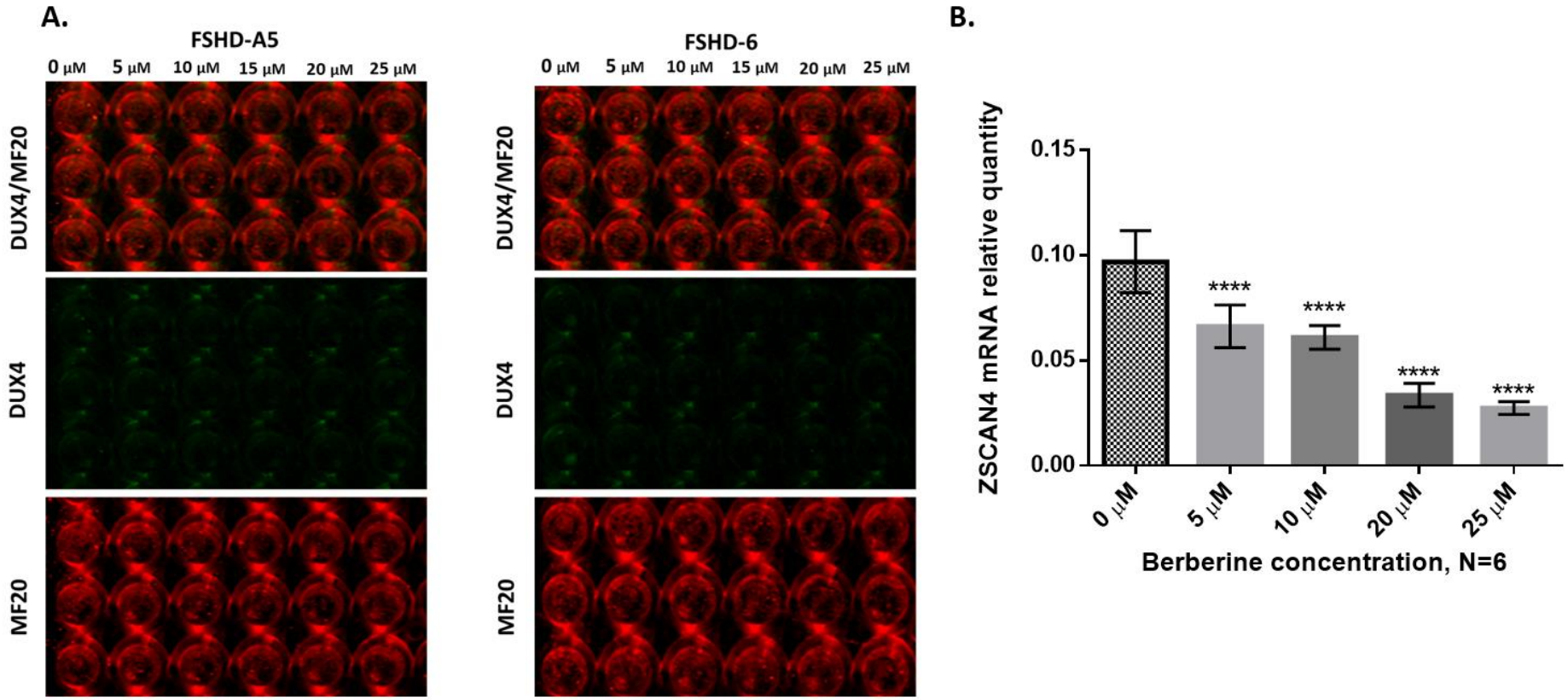
5.2.3. *In-cell western assay to measure DUX4 protein expression*

Although the DUX4 protein is expressed at extremely low levels, it was estimated to be present in 1/1000 or 1/200 in primary patents myoblast and myotubes, respectively (Snider et al. 2010; Tassin et al. 2013). The first developed 9A12 mouse monoclonal antibody against DUX4 has shown to also strongly react with another highly homologues protein such as DUX4c (Dixit et al. 2007). More recently, a monoclonal DUX4 rabbit antibody (E5.5) targeting the DUX4-specific C-terminus have been developed and tested showing positive results in the C2C12 DUX4 transfected cells and cre-inducible DUX4 transgenic mouse model (Geng et al. 2011; Jones and Jones 2018). Therefore, we asked a question whether the E5.5 rabbit monoclonal antibody can also successfully detect the DUX4 expression in the FSHD immortalised patient myotubes by using the in-cell western assay (ICW) as a readout. ICW is a quantitative immunostaining technique that allows to study levels of protein of interest (i.e., DUX4 using the E5.5 mouse monoclonal antibody in this case) in cells (Egorina et al. 2006). The normalising signal of the corresponding cells was produced by antibody staining of the myosin heavy chain (i.e., MF20). Unfortunately, the DUX4 expression could be detected in the FSHD positive (A5) myotubes as shown by the Odyssey Imaging system readout (Figure 5.3 A). As expected, there was no apparent expression of DUX4 present in the FSHD-6 myotubes either (FSHD 5.3 A). Although quantification of the DUX4 expression could be detected as a weak signal in the FSHD-A5 cells, it was not statistically significantly different from the intensities of the DUX4 expression present in the negative control cell line, indicating that the measured emission is mostly a background signal noise (Figure 5.3 B).

In order to acquire a higher image resolution, to check for the potential DUX4 expression at the nuclear level within the FSHD immortalized myotubes, fluorescent microscopy analysis was performed. Despite using a relative high antibody concentration (1:100 dilution), no positive DUX4 nuclei could be found in the patient myotubes using the anti-DUX4 rabbit monoclonal antibody (E5.5) (Figure 5.4 A).

Figure 5.3 In-cell western measuring DUX4 protein levels in FSHD patient cells treated with GQ ligand

Immortalized FSHD-A5 (DUX4 positive) and FSHD-6 (DUX4 negative) myotubes were treated with berberine in a dose dependent manner on the 2nd day of differentiation. Treated cells were fixed and permeabilized on the 4th day of differentiation. (**A.**) Representative in-cell western image where DUX4 was measured using rabbit anti-DUX4 (E5.5) primary antibody followed by the detection with Goat anti-Rabbit secondary antibody. Normalization to number of differentiated cells was performed by staining myosin heavy chain with the mouse IRDye 800CW anti-MF20 primary antibody. The plate was scanned on Oddysey Classic Infrared Imaging System (Resolution: 100 µM; Quality: medium; Focus offset: 4.0 mm; Intensity: 7 and 8 for 700 and 800 Channel, respectively) (**B.**) Values of DUX4 intensity were normalised to the MF20, demonstrating no statistically significant change in DUX4 expression upon berberine treatment in each cell group. Independent treatment at each concentration was performed, N=8 (one -way ANOVA). Error bars represent standard error of the mean (SEM).


A.
FSHD-A5
FSHD-6
0 μM
5 μM
10 μM
15 μM
20 μM
25 μM
DUX4/MF20
DUX4
MF20
B.
ZSCAN4 mRNA relative quantity
0.15
0.10
0.05
0.00
****
Berberine concentration, N=6

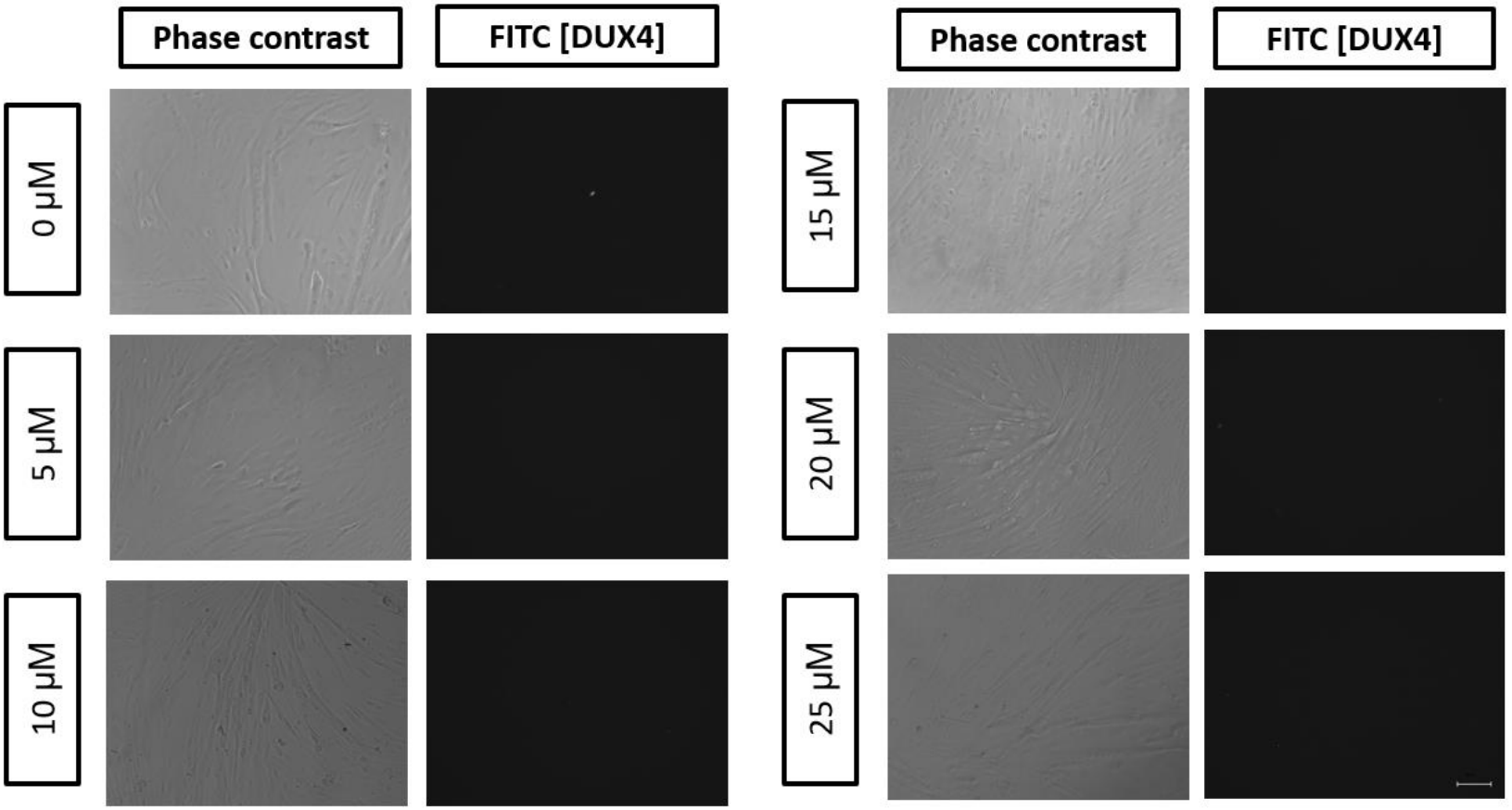


Figure 5.4 Immunocytochemistry to detect DUX4 protein expression in FSHD-A5 cells treated with berberine

Immortalized FSHD-A5 (DUX4 positive) myotubes were treated with berberine in a dose dependent manner on the 2nd day of differentiation. Treated cells were fixed and permeabilized on the 4th day of differentiation. DUX4 stained using rabbit anti-DUX4 (E5.5) primary antibody followed by detection with AlexaFluor488 Goat anti-Rabbit secondary antibody. FITC channel was used to detect the DUX4. No DUX4 was detected. Scale bar: 100 µm.

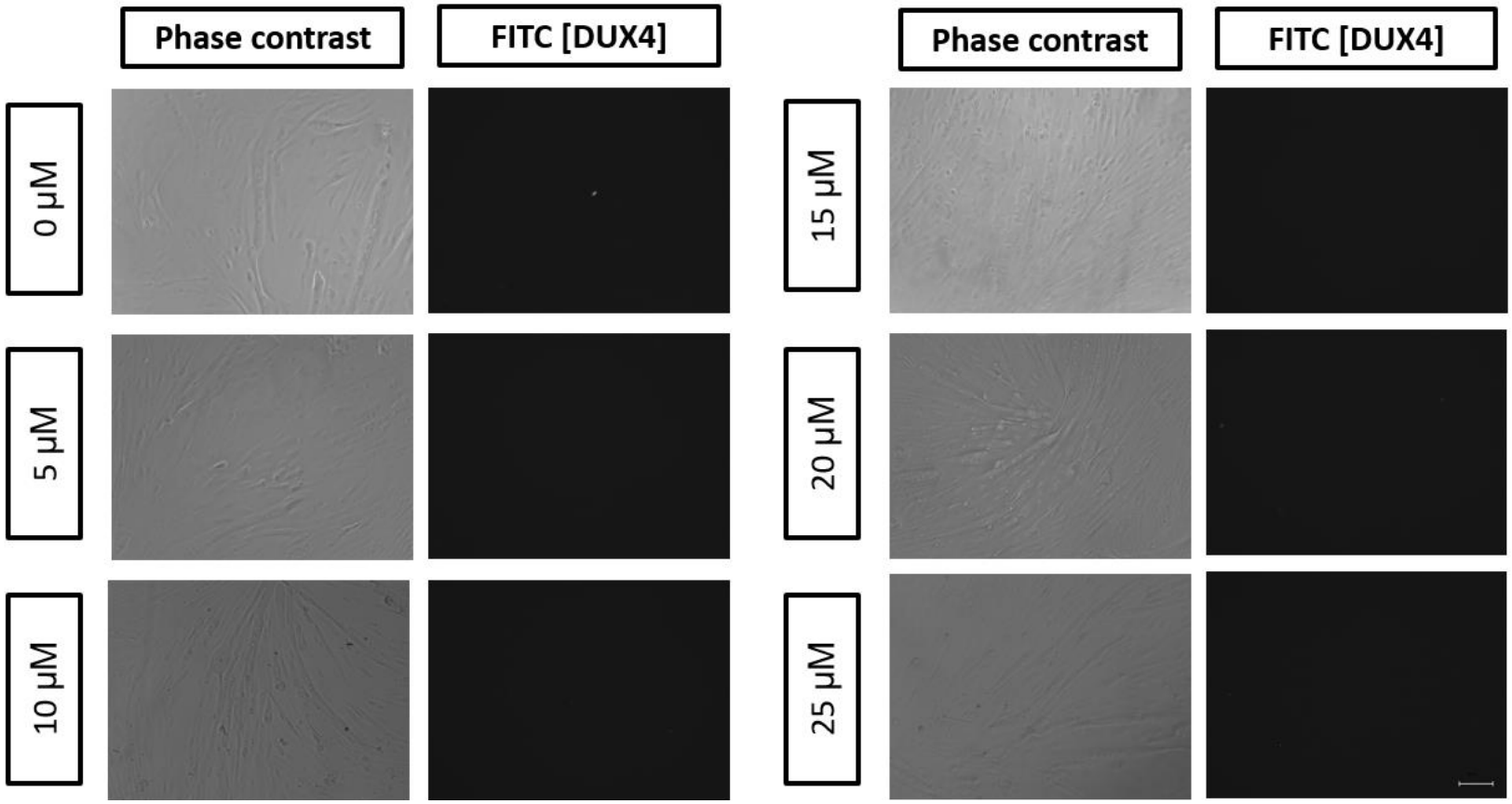


Figure 5.5 Immunocytochemistry to detect DUX4 protein expression in FSHD-6 cells treated with berberine

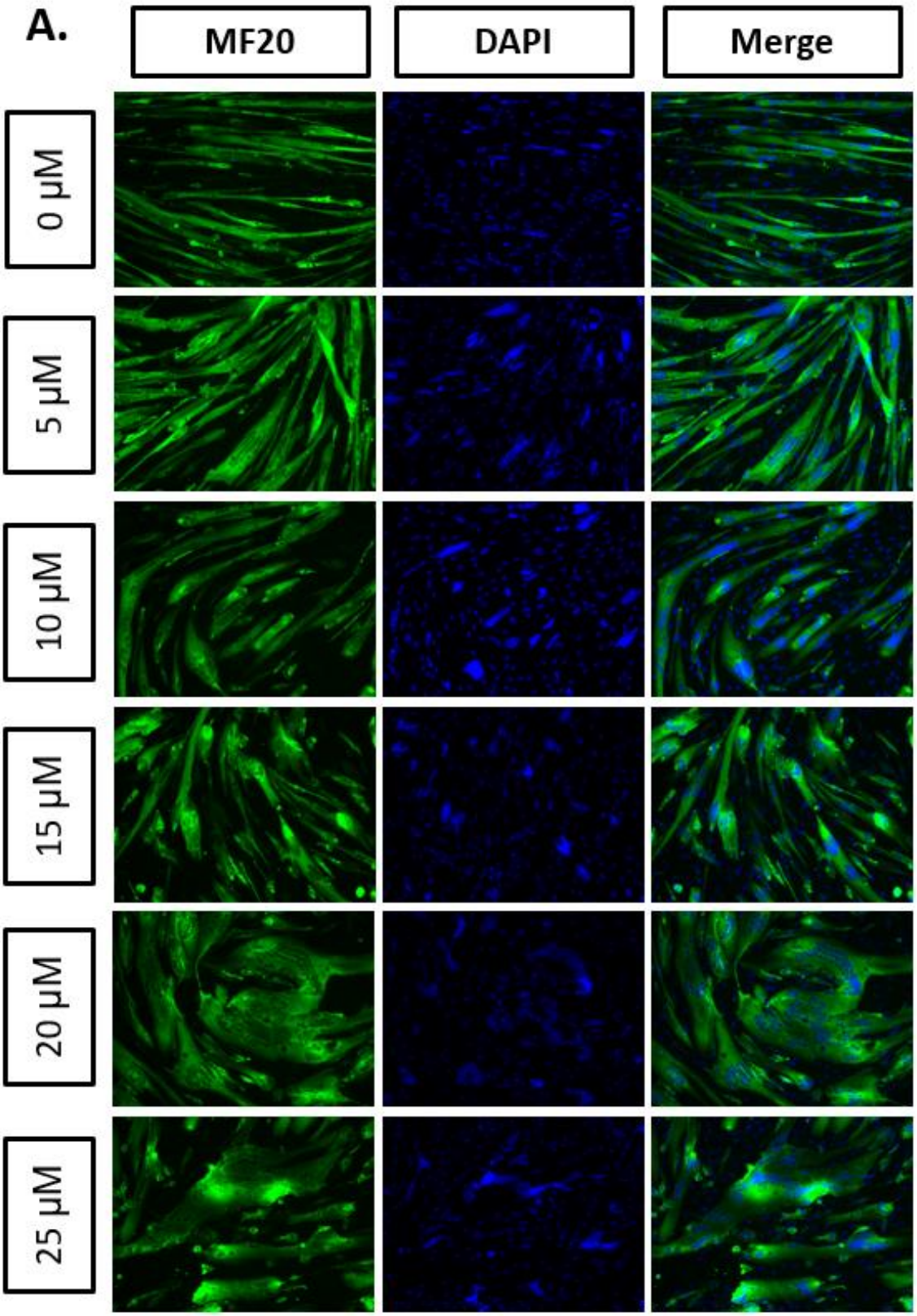
Immortalized FSHD-6 (DUX4 negative) myotubes were treated with berberine in a dose dependent manner on the 2nd day of differentiation. Treated cells were fixed and permeabilized on the 4th day of differentiation. DUX4 stained using rabbit anti-DUX4 (E5.5) primary antibody followed by detection with AlexaFluor488 Goat anti-Rabbit secondary antibody. FITC channel was used to detect the DUX4. No DUX4 was detected. Scale bar: 100 µm.

5.2.4. *Berberine increases fusion index of FSHD cells and induces a phenotypic switch from atrophic to disorganised myotube morphology*

One of the mechanism by which a rare protein such as DUX4 can cause myopathy is its ability to diffuse across the cytoplasm of myotubes to the neighbouring nuclei and lead to aberrant activation of many downstream genes (Ferreboeuf et al. 2014).Therefore, in order to ensure that the berberine downregulation of the measured DUX4 downstream genes (i.e., ZSCAN4, TRIM43 and MBD3L2) was not an indirect effect of the myotube formation defect caused by the presence of the ligand, the fusion index of the FSHD myotubes (clones –A5 and -6) was assessed. The berberine treated cells were immunostained with the antibody specific for the myosin heavy chain (MF20), whereas the nuclei were visualised using DAPI (Figure 5.6 A). The nuclei present in the MF20 positive myotubes were counted and divided by the total number of nuclei present in all cells of each filed. The fusion index was expressed as the percentage ratio of myotube -related nuclei vs. total number of nuclei (Figure 5.6 B). A statistically significant increase in the index fusion was observed across all tested concentrations of berberine, suggesting that the downregulation of the DUX4 downstream genes was not caused by the impairment of the myotube formation, but it was rather *DUX4* specific. This idea was further supported by the fact that the negative control cells did not show any statistically significant change in the calculated fusion index (Figure 5.7 B).

Two phenotypes of FSHD primary myotubes were previously reported: the atrophic and disorganised (Barro et al. 2010; Tassin et al. 2012). The atrophic phenotype was characterised by a narrow elongated myotubes with neatly aligned nuclei, whereas the disorganised myotubes were described as giant structures containing large clusters of

nuclei. Furthermore, an antisense-mediated downregulation of DUX4 have been demonstrated to lead to the prevention of the atrophic, but not disorganised myotube phenotype formation in FSHD primary cell culture (Ansseau et al. 2017). Interestingly, berberine treatment led to phenotypic switch from atrophic to disorganised state in FSHD immortalised myotubes (Figure 5.2.6 A). The clustering of nuclei in the berberine treated FSHD negative control cells was not as apparent, suggesting that the phenotypic switch could be DUX4 related (Figure 5.2 7 A).


A.
MF20
DAPI
Merge
0 µM
5 µM
10 µM
15 µM
20 µM
25 µM

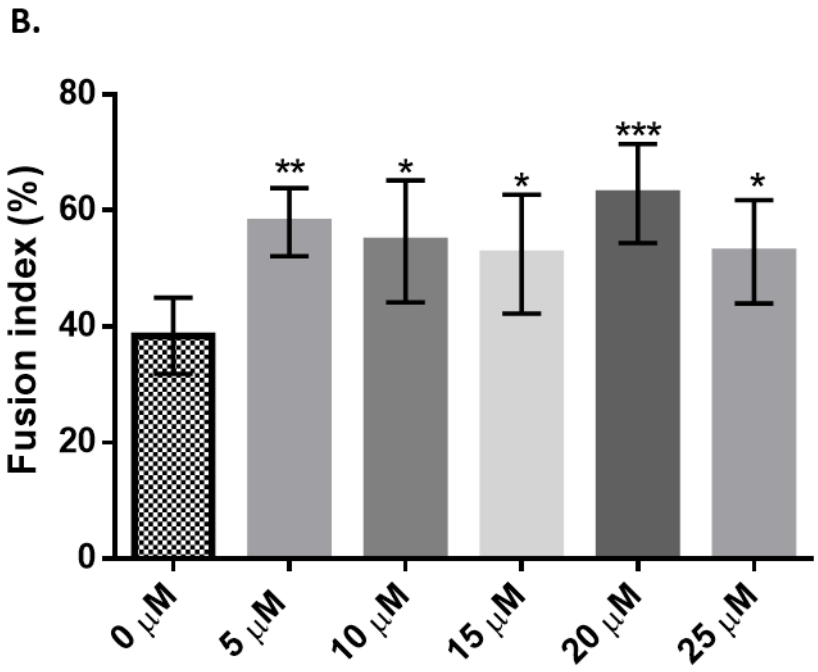
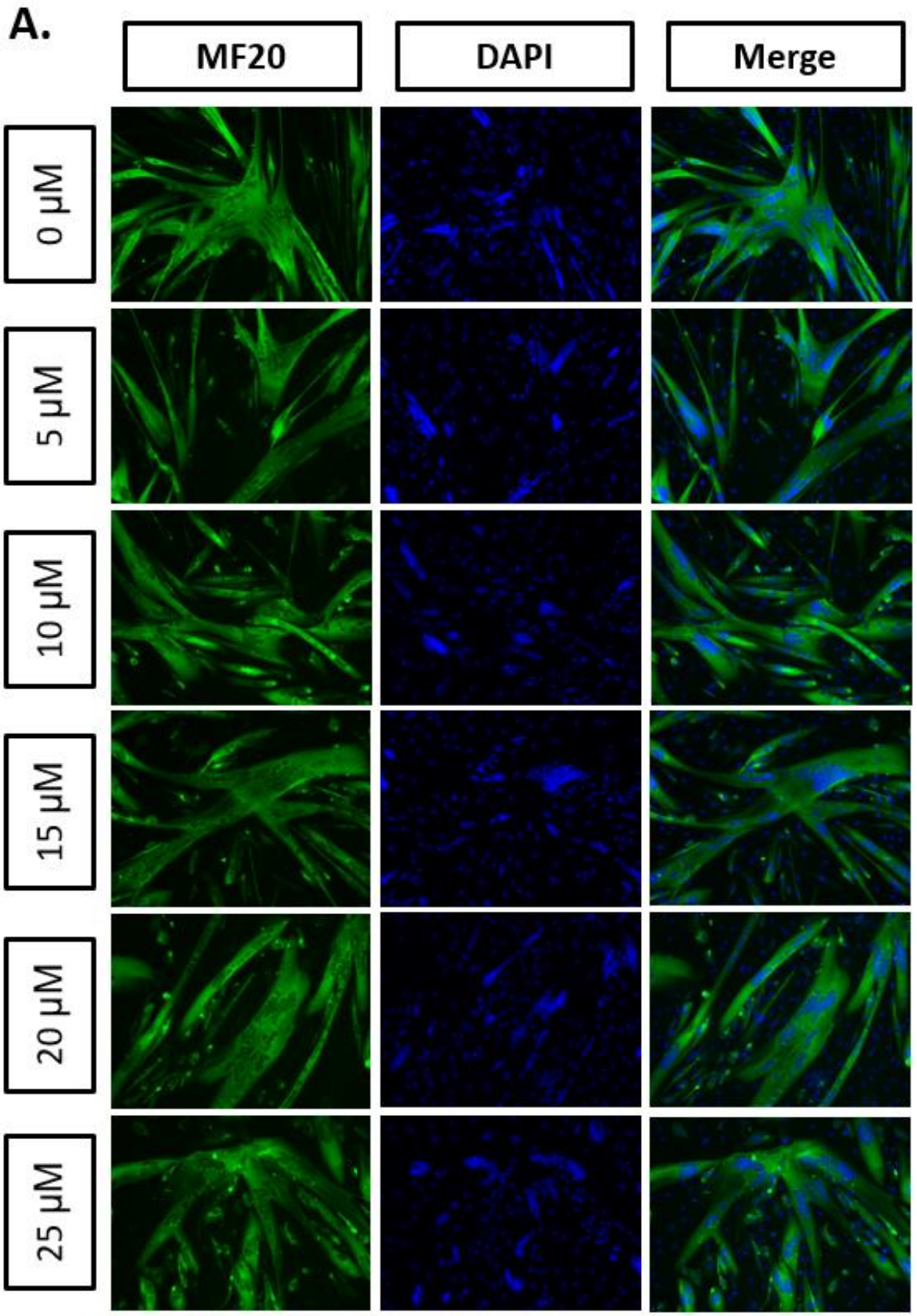


Figure 5.6 Fusion index of immortalized FSHD-A5 myotubes treated with berberine

(**A.**) Immortalized FSHD myotubes were treated with berberine at the day 2nd of differentiation. Myosin heavy chain and nuclei were stained at the 4th day of differentiation with the MF20 antibody and DAPI, respectively. (**B.**) For fusion index calculation, nuclei were counted in myotubes containing two or more nuclei and were expressed as a percentage of total nuclei present. Independent treatment at each concentration was performed, N=8 (*p<0.05; **p<0.01; ***p<0.001, one -way ANOVA). Error bars represent standard error of the mean (SEM).


A.
MF20
DAPI
Merge
0 µM
5 µM
10 µM
15 µM
20 µM
25 µM

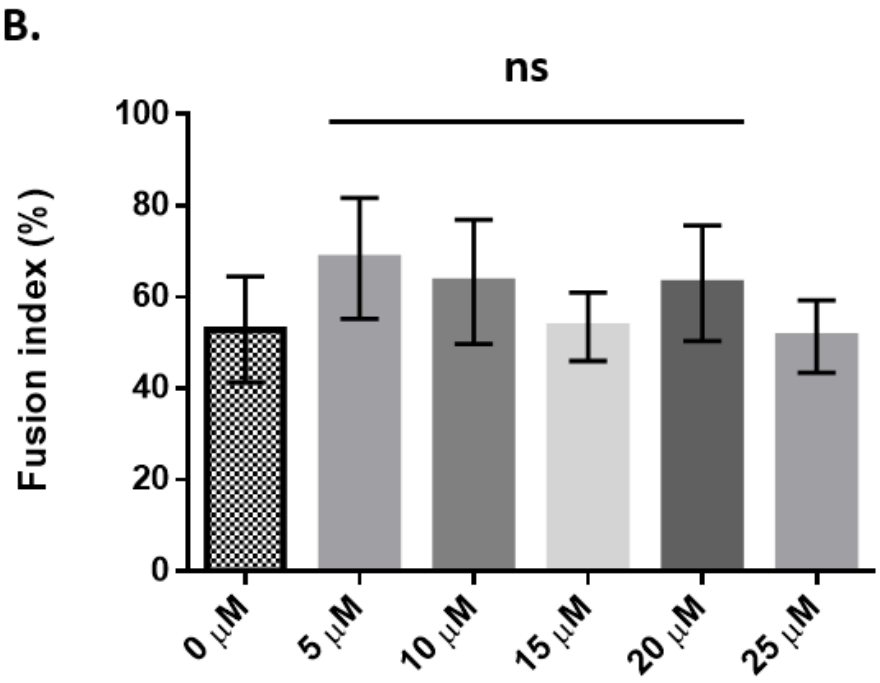


Figure 5.7 Fusion index of immortalized FSHD-6 myotubes treated with berberine

(**A.**) Immortalized FSHD myotubes were treated with berberine at the day 2nd of differentiation. Myosin heavy chain and nuclei were stained at the 4th day of differentiation with the MF20 antibody and DAPI, respectively. (**B.**) For fusion index calculation, nuclei were counted in myotubes containing two or more nuclei and were expressed as a percentage of total nuclei present.

Independent treatment at each concentration was performed, N=8 (*p<0.05; **p<0.01; ***p<0.001, one -way ANOVA). Error bars represent standard error of the mean (SEM).

.

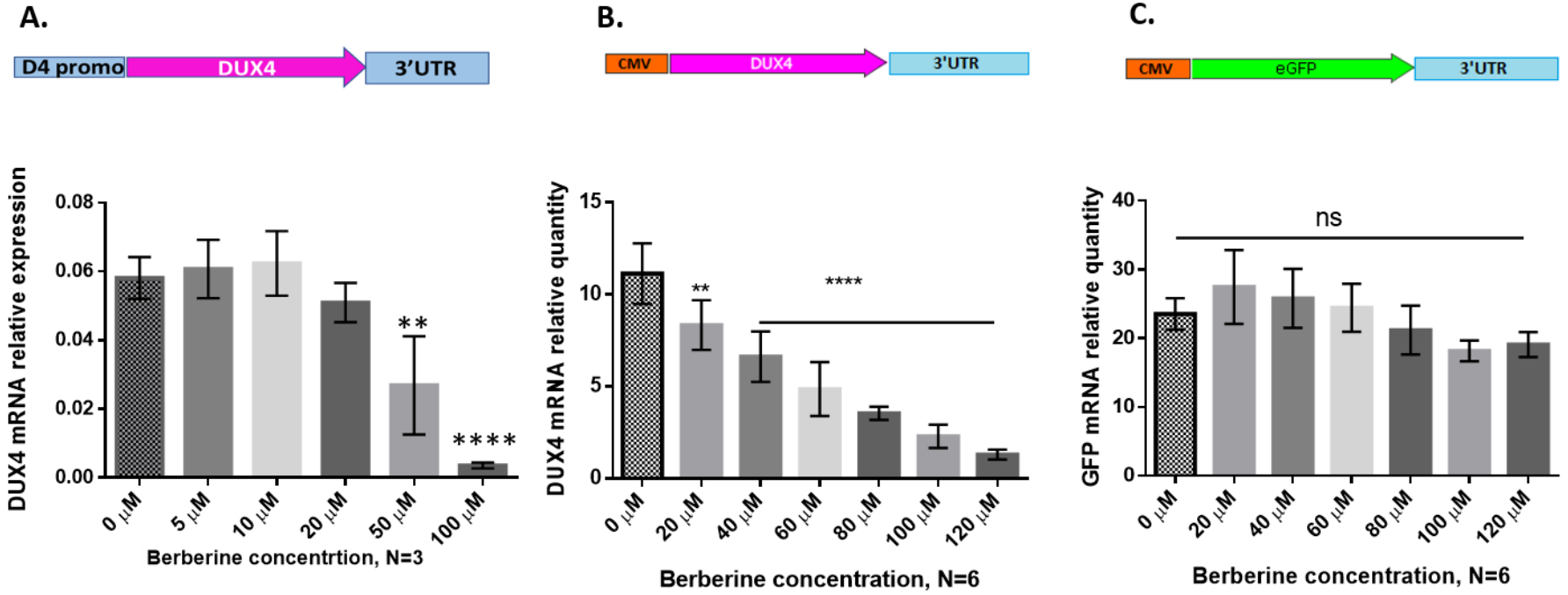
5.2.5. *Berberine-mediated downregulation is DUX4 transcript specific*

Since GQ structures have been predicted to prevalently form across the human genome, one must bear in mind that berberine could have potential DUX4-unspecific, off target effects. Indeed, this work has demonstrated that stable GQ motifs form within multiple genomic loci and transcript of *DUX4* sequence that could serve as potential targets for the ligand (Section 3.2). Furthermore, we have demonstrated that the predicted GQ structure within the D4P sequence has a relatively weak effect on regulating the gene expression in a transient expression reporter system (Section 4.2.4). Therefore, the next question to address was whether berberine-mediated strong inhibition of *DUX4* expression in FSHD patient cells (Figure 5.1 A) was the result of transcript-specific targeting by the ligand. To answer this issue, two plasmid constructs were developed, where the *DUX4* transcript sequence expression was driven by the endogenous D4P fragment or the non-native CMV promoter (Figure 5.8 A and B). In addition, a CMV driven, eGFP expressing plasmid was used as a control to determine if the berberine treatment could affect transfection reaction or influence expression from a plasmid deprived of GQ sequences (Figure 5.8 C). The berberine treatment was applied at the day of transfection to RD CCL 136 cells for 24 hours before the total RNA was harvested for the RT-qPCR analysis. The plasmid DUX4 expression construct driven by the D4P begins to show a statistically significant downregulation of the gene expression at 50 μM of the ligand concentration (Figure 5.8 A). The CMV-driven construct shows a significant downregulation of DUX4 expression at lower drug concentration of 20 μM (Figure 5.8 B). Furthermore, the *DUX4*-downregulation from the CMV-containing plasmid continued to decrease in dose dependent manner (Figure 5.8 B). Since the CMV

promoter does not contain any GQ structures, the effect of berberine could be related to binding of secondary structures present within the mRNA and/or the coding sequence of the DUX4. This idea is supported by the fact that berberine shows strong binding affinity towards the *DUX4* RNA GQs (e.g., E1 and SS1 RNA GQs) derived from the transcript sequence of the gene (section 3.2.4). The potential mechanisms how berberine could downregulate the DUX4 expression on the level of the mRNA are discussed in detail in section 5.3. The expression of the negative control plasmid shows no change in CMV-driven eGFP at any of the tested concentrations of the ligand, further strengthening the idea that the berberine-mediated downregulation of *DUX4* is specifically related to the gene's transcript and/or gene coding sequence (Figure 5.8 C).

Figure 5.8 Levels of transiently expressed *DUX4* transcript at increasing berberine concentrations in RD CCL 136 cells

Expression levels of *DUX4* mRNA was measured using RT-qPCR in rhabdomyosarcoma (RD) CCL 136 cells transfected with 1µg of the pAAV.DUX4 using Lipofectamine^{TM} 3000 (Invitrogen) and simultaneously treated with berberine. Total RNA was harvested 24 hours after transfection and treatment. *GAPDH* was used as a housekeeping gene. (**A.**) Top: schematic presentation of the transfected plasmid containing *DUX4* driven by its native promoter sequence (D4P). Bottom: Downregulatory effect of the berberine on DUX4 expression from the transfected plasmid (**B.**) Top: Schematic presentation of the transfected plasmid containing *DUX4* driven by a CMV promoter. Bottom: Downregulatory effect berberine on DUX4 expression from the transfected plasmid (**C.**) Top: schematic of the negative control plasmid expression *eGFP* driven by the CMV promoter. Berberine treatment in (**A.**), (**B.**) and (**C.**) was performed independently with N=3; N=6 and N=6, respectively (**p<0.01; ****p<0.0001, one - way ANOVA). Error bars represent SEM.


A.
D4 promo
DUX4
3'UTR
DUX4 mRNA relative expression
0.08
0.06
0.04
0.02
0.00
**
****
0 μM
5 μM
10 μM
20 μM
50 μM
100 μM
Berberine concentrtion, N=3
B.
CMV
DUX4
3'UTR
DUX4 mRNA relative quantity
15
10
5
0
**
****
0 μM
20 μM
40 μM
60 μM
80 μM
100 μM
120 μM
Berberine concentration, N=6
C.
CMV
eGFP
3'UTR
GFP mRNA relative quantity
40
30
20
10
0
ns
0 μM
20 μM
40 μM
60 μM
80 μM
100 μM
120 μM
Berberine concentration, N=6

5.3. Discussion

Our understanding of the FSHD pathophysiology has advanced significantly in recent years (review: (Tawil et al. 2014). The current consensus recognises that the postnatal expression of the *DUX4* is a key factor causing the disease. The novel finding showing that the GQ motifs present within the coding sequence of the gene, could potentially lead to further understanding the molecular mechanism that govern expression of the DUX4. In this study, we have demonstrated that berberine leads to downregulation of the *DUX4* mRNA levels in FSHD immortalised patient cells. In addition, the ability of berberine to downregulate *DUX4* expression could be specifically related to its coding/transcript sequence as demonstrated by the reporter studies.

Since RNA GQs are implicated in mRNA transcription and processing, it is possible that the berberine treatment mediates downregulation of *DUX4* mRNA by interference with these processes. For example, during transcription, the newly produced pre-mRNA sequence strand can fold into an intermolecular GQ structure with the coding (sense) DNA strand (Zhang et al. 2014). The RNA transcript sequence and the coding DNA strand require as little as two tandem G-tracks to form a stable GQ structure (Duquette et al. 2004). Formation of such a hybrid RNA/DNA GQs has been predicted and confirmed *in vitro* using T7 RNA polymerase transcription, site specific mutagenesis and reporter-based transient transcription assays (Wanrooij et al. 2012; Zheng et al. 2013). These studies conclude that formation of RNA/DNA hybrid GQ structures act as potential *in-cis* elements mediating transcription inhibition. The hybrid GQ structures are thought to be also involved in transcription termination as they have been observed to pause RNA

polymerase II transcription (Gromak et al. 2006; Skourti-Stathaki et al. 2011). The RNA/DNA GQ structures have been predicted to form downstream of poly(A) signals and since these motifs are very stable, a helicase Senataxin (SETX) is required to resolve these. Binding of SETX, promotes of endonucleolytic cleavage at the poly(A) and finally release of the newly synthesised pre-mRNA from RNA polymerase II (Gromak et al. 2006; Skourti et al. 2011). Therefore, berberine binding to RNA/DNA hybrid motifs formed from the *DUX4* transcript and DNA coding strands could potentially prevent functions of the specific GQ helicases that ensure the resolution of these structures and consequently lead to premature transcription termination and *DUX4* downregulation.

In addition, the transcription termination is tightly linked to the polyadenylation process at the 3'end (Mandel, Bai, and Tong 2008). Briefly, the mammalian canonical polyadenylation involves pre-mRNA 3'UTR signals recognition (e.g., AAUAAA) by a multisubunit protein complex (CFI, CFII, CPSF and CstF). In the case of cellular stress and DNA damage, the CstF factor becomes sequestered and the polyadylation regulatory complex becomes inactive (Shi and Manley 2015). In the pre-mRNA of *TP53*, a GQ structure proximal to the polyadenylation sequence binds a hnRNP H/F slicing factor that in turn recruits and protects CstF from sequestration and allows efficient polyadenylation to occur (Decorsière et al. 2011). Perhaps *DUX4* 3'UTR sequences could also aid the process of polyadenylation of the *DUX4* transcript, even under cellular stress conditions induced by the toxic transcription factor itself. Moreover, the berberine binding to the *DUX4* RNA GQs could potential prevent formation of the polyadenylation complex by interfering with binding of its components (e.g., nhRNP H/F).

Three *DUX4* transcript isoforms have been detected in primary FSHD myoblasts and myotubes (Dixit et al. 2007; Snider et al. 2009). In this study, only the two full-length *DUX4* transcripts PCR that are thought to be responsible for the formation of the functional toxic DUX4 protein have been detected using RT-PCR-. However, there are a growing number of reports indicating that the *DUX4* can also be alternatively spliced into a short isoform (*DUX4-s*) that would form a truncated protein with a functional DNA binding domain but lacking the effector domain (Jones and Jones 2018; Snider et al. 2009). Therefore, the *DUX4-s* could potentially have an inhibitory effect on the function of the full length toxic protein variant by competing for the DNA target binding site. Interestingly, it has been demonstrated that the intronic and exonic RNA quadruplexes can be involved in the regulation of alternative splicing (review: Fay et al. 2017). A non-exhaustive list of physiologically relevant genes that utilise quadruplex motifs to regulate splicing of their transcripts, include: B-tropomyosin, hTERT, PAX9, p53, BACE-1 and FMR1 (Didiot et al. 2008; Fisette et al. 2012; Marcel et al. 2011; Ribeiro et al. 2015). Generally, the GQ formation regulates the splicing process by either masking binding sites of the target proteins or recruiting GQ-specific biding proteins (e.g., hnRNP protein family) (Dominguez et al. 2010). For example, formation of different Fragile X mental retardation protein isoforms is regulated by to the exonic GQ structures that provide a binding site for a splicing enhancer (Bensaid et al. 2009). By analogy, the GQ structures present within the exon 1 of the *DUX4* transcript could potentially promote the expression of the *DUX4-s* isoform, which would not be detected by the primers used to detect the full-length variant here. Although it is not clear how berberine targeting GQ or how these motifs could regulate splicing by of the *DUX4* transcript, these are interesting questions that could be addressed in the future.

The main hallmark of the FSHD molecular mechanism is loss of the epigenetic marks within the D4Z4 region leading to chromatin de-repression and *DUX4* expression. An important epigenetic modificator, EZH2, has been previously found to be enriched within the D4Z4 repeat array in healthy primary myocytes, but not in the cells derived from the FSHD patients (Cabianca et al. 2012). EZH2 is one of the components of the PRC2 complex that acts as an epigenetic transcriptional repressor shown to regulate developmental processes and cancer (Sellers and Loda 2002). Interestingly, the EZH2 has been found to bind to RNA GQs with significantly higher affinity compared to its DNA counterparts (X. Wang et al. 2017). It has been now well demonstrated that RNA GQs can play an important function in disease pathogenesis by sequestering important RNA-binding proteins (Cammas et al. 2015). For example, a large hexanucleotide repeat expansion within intron 1 that leads to formation of stable RNA GQ structures within the pre-mRNA of *C9orf72* gene (Reddy et al. 2013). The *C9orf72* is a key factor behind the pathogenesis of ALS and its transcript's RNA GQ has been demonstrated to bind and sequester important splicing and polyadenylation regulator (e.g., nhRNP H/F), leading to a global deregulation of RNA processing evens (Wang et al. 2015). Therefore, an interesting hypothesis arises suggesting that the *DUX4* transcript could contribute to the disease pathogenesis by interfering with the epigenetic pathway regulating D4Z4 array repressive chromatin state. Targeting of the *DUX4* RNA GQ with small molecules could potentially prevent sequestration of important epigenetic modificators (e.g., EZH2) and consequently result in *DUX4* expression suppression.

RNA GQ motifs found in open reading frames have also been found to act as 'roadblocks', inhibiting progression of ribosomes along the mRNA, leading to significant decrease in protein synthesis efficiency (Endoh and Sugimoto 2016). The

RNA GQ downstream of start codon (downstream of 5'UTR) have been found to be particularly effective in blocking translation, as demonstrated in the context of *KMT2A* and *MLL* protooncogenes, where the *in line* probing combined with the G-A mutagenesis and the luciferase assay showed over 75% reduction in protein synthesis to be RNA GQ structure related (Thandapani et al. 2015). Presence of stable RNA GQ structures within the coding sequence of *DUX4* could perhaps also negatively affect its translation, providing a potential explanation for the low levels of DUX4 protein present in the patient cells.

Berberine treatment of the FSHD immortalised patient cells has led to significant decrease in cell viability. Since the immortalised clones were produced by retrotransduction of the *hTERT* gene, it has been anticipated that berberine could have an anti-proliferative effect on these cells (Krom et al. 2012). HTERT is an important enzyme involved in lengthening of the telomeric chromosome ends and maintaining chromosomal stability (Thorley et al. 2016). However, when overexpressed, it leads to cell immortalization and eventually results in wide range of cancers (Cong, Wright, and Shay 2002). It has been demonstrated that berberine strongly downregulates expression of *hTERT* and binds to GQ-forming telomeric ends sequences, resulting in growth inhibition of cancer cell lines (Fu et al. 2013). Therefore, the primary cause of berberine-mediated decrease of viability of the patient cells could be a result of interference of the ligand with the immortalization pathways. It would be interesting to perform the same berberine treatment on primary FSHD patient cell lines to establish whether the cytotoxic effects of berberine are specifically attributed to the targeting of the *hTERT*-related pathway present in the FSHD immortalised patient cells.

Berberine treatment has been demonstrated to lead to *DUX4* mRNA downregulation in FSHD immortalised patient cells and in RD cells transiently expressing *DUX4*. Furthermore, the shift of the atrophic to disorganised myotube phenotype, could further suggest that the treatment is *DUX4*-specific as it has been previously implied (Ansseau et al. 2017). However, since berberine has been found to have a range of pharmacological activities, including anti-inflammatory, antimicrobial and anti-tumour effects, it is inevitable that berberine has *DUX*-nonspecific off-target effects (Ganesan and Xu 2017). This is also further supported by the fact that berberine, apart from the *DUX4* transcript, it has a high affinity binding towards the GQ-forming sequences within its own enhancer and promoter regions (section 3.2). There is currently no feasible method to target specific GQs within the genome, having said that, new small molecule chemistries have begun to emerge that have high binding selectivity to GQ over duplex DNA or in some cases can even discriminate between RNA and DN GQ structures. Two small-molecule ligands have been recently developed able to target RNA GQs, namely carboxy pyridostatin and RGB1 (Di Antonio et al. 2012; Katsuda et al. 2016). Since the *DUX4* transcript's RNA GQ are demonstrated be a likely target of berberine that leads to its mRNA downregulation, it would be interesting to evaluate these RNA GQ specific-ligands and their ability to suppress expression of the toxic transcription factor. The RBG1 is especially interesting candidate molecule, due to its ability to specifically bind RNA GQ not only over the DNA counterparts, but also other RNA structures.

One promising strategy that could potentially address the specificity issue of small-molecule RNA GQ targeting, is the use of antisense oligonucleotide (AO) chemistries. AOs can be designed to target GQ structures based on their sequence, which is the level of specificity currently unattainable by any other types of chemical compounds. Indeed,

disruption of RNA GQ folding of the *H2AFY*'s gene mRNA in human cell line has shown to interfere with its translational processes regulating the gene's expression (Rouleau et al. 2015). A number of strategies, utilising various AO chemistries have been deployed over the years to regulate mRNA processing, including: interference with 5'-cap formation; splicing regulation, degradation of the transcript by the RNase H1-induced cleavage or interference with 3'UTR elements; and translation inhibition by interference with ribosome access (DeVos and Miller 2013; Marsollier et al. 2016). However, the strategy of modulating RNA GQ structures of *DUX4* transcript using an AO approach to regulate its expression has not been studied. Since the antisense technology shows a very promising clinical potential as a treatment for muscular dystrophies, it provides an interesting avenue of research for development of novel therapeutic strategy to target the *DUX4* expression.

6. Final discussion, evaluation and future directions

6.1. Summary of results

The paramount objective of this thesis was to advance our understanding of the molecular mechanisms that lead to FSHD pathophysiology. More specifically, it aimed to investigate the strength and possible roles of GQ motifs forming within the genomic loci (e.g, enhancer and promoter) and transcript of *DUX4*. These genetic regions of *DUX4* were subjected to bioinformatic analyses to assess the potential formation of GQ structures. Novel putative GQ-forming sequences were predicted within the enhancer and promoter regions. In addition, the *DUX4* transcript has been found to be particularly enriched in putative GQ motifs. Aided by these bioinformatic results, candidate oligonucleotide sequences were selected and analysed for the formation of the GQ structures in solution using CD and NMR. The biophysical analysis confirmed that the enhancer-, promoter-, and transcript-related putative GQ-forming oligonucleotides had indeed formed GQs is solution. Next, a binding interaction of berberine, a small-molecule known to interact with GQs, with the GQ-forming candidate sequences was assessed. Berberine was shown to have a particular high binding affinity towards the *DUX4* locus enhancer GQ-related sequence. The RNA GQ sequences appear to be weaker binders of berberine compared to the DNA counterparts.

Due to an extensive focus of the scientific literature on promoter-related GQs (Balasubramanian et al. 2011), particularly in the context of oncogene expression, the

next task undertaken in this thesis was to investigate the role of the GQ-forming sequence within the promoter fragment of *DUX4*. First, substitution of the motif's sequences was performed to disrupt the secondary structure and confirm GQ presence. Since the GQ- forming sequences were also predicted to have the potential to form alternative non-GQ hairpin structures, both of these putative motif types were disrupted by the sequence substitution as shown by the CD and NMR analysis. In addition, the mutagenised *DUX4* promoter variants with the disrupted secondary structures were cloned upstream of an eGFP reporter cassette, to assess their activity to drive gene expression. The results demonstrated a statistically significant downregulation of gene expression driven by the promoter, with the deprived GQ-forming sequence, whereas the presence of hairpin-forming sequence has demonstrated to play no significant role. The experimental attempt to demonstrate that stabilisation of GQ-structures of the promoter using berberine resulted in enhanced expression was thwarted by the fact that berberine and GFP have overlapping emission/excitation wavelengths, which produced a significant signal noise in the flow cytometry readout. However, the mutagenesis studies, combined with previous literature (Armas et al. 2017) strongly suggest that due to the position of the GQ within the *DUX4* promoter fragment (i.e., coding/positive strand and upstream of the TATA box), this structure could have a weak positive effect on transcription by contributing to double helix unwinding, which allows easier access of the transcriptional machinery binding. Although we concluded that the investigated promoter GQ sequence is a most likely weak, finely tuning *in-cis* element regulating DUX4 expression, its potential to form an intramolecular GQ structure with its corresponding enhancer sequence to control the interaction between these two crucial transcriptional elements is a very interesting but unexplored possibility.

Finally, in chapter 5 the effects of berberine on *DUX4* mRNA expression were investigated in FSHD immortalised patient cells. The small molecule was demonstrated to have a strong downregulatory effect on expression of the toxic transcription factor as well as its downstream target genes, without a negative effect on the fusion index processes of the treated cells. Furthermore, it has been demonstrated that the potent downregulation of the DUX4 expression is likely to be related to berberine-mediated interferences of mechanisms mediating transcription and/or transcript processing as demonstrated in the reporter-based assay. Considering that the DUX4 transcript is enriched in GQ-forming sequences, bound with high affinity by berberine, strongly suggest the small-molecule mediated downregulation of DUX4 is a result of its binding to the *DUX4*-specific GQ motifs. Although the molecular mechanisms of berberine-mediated DUX4 downregulation are not clear at this stage, the potential GQ-related pathways of the compound have been discussed in section 5.3.

6.2. Evaluation and future directions

The field of GQ epigenetics of FSHD has been previously virtually unexplored. This work provides the first comprehensive set of evidence indicating the presence of GQs in DUX4 genomic loci, and transcript, that could potentially play an important role in the regulation of gene expression. The understanding of GQ biology is ever-expanding and the novel techniques to evaluate their role are constantly being developed. Furthermore, cellular and animal models that help recapitulate FSHD are also emerging. Taking advantage of the recent advancements in these scientific fields could significantly advance understanding of the FSHD-related GQ function.

Since it has been found that the DUX4 DME1 enhancer sequence forms a very stable GQ structure that binds berberine with high affinity, it would be interesting to investigate the role of the motif in a similar way to that used to examine the promoter elements. For example, the enhancer elements could be included in a reporter gene system and its intramolecular GQ-forming sequences disturbed through mutagenesis, and/or stabilised by small-molecule (e.g. berberine) binding, to determine the potential role of the motif on the gene expression regulation. Clearly, if berberine was to be used to assess the DME1 enhancer GQ function, an alternative cassette expressing a protein would be required with different excitation/emission to spectra to berberine to avoid the interference of the signal readout from the flow cytometry seen with eGFP. Alternatively, the luciferase reporter system could be used, where the signal is a result of an enzymatic reaction, and the reporter product serves as a substrate and therefore would not result in formation of a fluorescent activity by berberine.

Although protein binding to the *DUX4* promoter fragment has been previously assessed (Sharma et al. 2016), studies describing *DUX4* enhancer-associated in-trans factors are not currently available. It would be therefore interesting to investigate DUX4 promoter/enhancer-associated proteins and determine, whether the GQ-forming motifs influence these interactions by for example performing ChIP analysis. RNA-binding proteins associated with the *DUX4* transcript could also be investigated and evaluated for their potential to interact with the RNA GQ structures.

The biophysical tools used in this thesis show the formation of GQ structures in isolated short sections of a given sequence. However, the bioinformatic analysis has demonstrated that the transcript sequence in particular is enriched in multiple

(potentially overlapping) RNA GQ sequences. Recently, a method has been developed where long stretches of RNA sequences can be footprinted for the presence of GQ by performing the 7-deazaguanine substitution (Weldon et al. 2017). The 7-deazaguanine-substituted RNA is not able to form GQs, which allows differential identification of the structures in the control RNA sequences. Therefore, it would be interesting to confirm the *DUX4* transcript GQ structures identified by bioinformatic and biophysical analyses in our studies using this novel foot-printing method.

Accurate mapping of the *DUX4* RNA GQ could also aid rational design of specific AOs that could target and potentially disrupt the structures by competitive hybridisation and interference with structural components (e.g., G-tetrad sequences) that are essential for the formation of the motif. The AO-mediated targeting of the *DUX4* RNA GQs could help understand their potential role in regulating the transcript processing such as splicing and/or polyadenylation.

Development of animal models for the FSHD has proven difficult over the years. The main hurdle behind modelling of FSHD is the fact that the D4Z4 microsatellite repeat region containing the *DUX4* gene is conserved only to Old World primates, which has led to the absence of naturally-arising models of the disease in common animal species used in laboratories (Leidenroth and Hewitt, 2010). Attempts to create a mouse animal model that carries the contracted FSHD1 or full wild-type D4Z4 allele was performed (Krom et al. 2012). However, despite achieving a genetic signature equivalent to the one seen in the FSHD patient (i.e. detectable levels of DUX4 and dysregulation of its downstream genes) no disease-specific phenotype could be recapitulated in these animals (Krom et al. 2012). Recently, a cre-inducible DUX4 transgenic mouse has been created that

manifests the FSHD -related myopathy (Jones and Jones, 2018). However, the promoter and enhancer elements that have been introduced to drive DUX4 expression in this model, were not endogenous to the transcription factor, which eliminates the possibility of studying the GQ structures related to these transcriptional elements.

Another approach to model the FSHD pathogenesis has utilised adeno-associated virus mediated delivery of myopathic *DUX4* gene constructs in vivo (Wallace et al. 2012). At the time of writing of this thesis, it has been decided to deliver the AAV vector expressing the DUX4 (kindly provided by Dr Harper, The Ohio State University) into mouse (wild-type C57BL/6) muscle tissue and investigate if berberine treatment could potentially prevent the disease phenotype. Interestingly, the initial results indicate that berberine treatment reverses the DUX4- induced histopathological changes of the muscle tissue by significantly reducing the central nucleation and fibrotic markers (Dr. Lu-Nguyen, Personal communications). In addition, the berberine treatment also reduces the DUX4 protein levels as well as significantly improving muscle function as demonstrated by the improved muscle specific force of the berberine-treated tissue (Dr. Lu-Nguyen, Personal communications). It is important to note the DUX4 expression from the AAV vector was driven by the non-endogenous CMV promoter that lacks functional GQ motifs (Huang et al. 2012; Salvati et al. 2014). Therefore, it can be speculated that the berberine-mediated DUX4 downregulation is specific to stabilisation of the GQs forming in the transcript rather that to the promoter/enhancer-related GQs. This was supported by the finding that berberine shows strong binding affinity towards the DUX4 RNA GQs (e.g., E1 and SS1 RNA GQs) (section 3.2.4). Even though berberine does not induce structural changes to these RNA GQs, as reported in the case of the DUX4 promoter-related GQ, strong binding of berberine to the *DUX4* RNA GQ structures alone could lead to

downregulation of gene expression. The suggested mechanisms explaining DUX4 downregulation mediated by the berberine biding to the gene's transcript were outlined in section 5.3.

Considering the vast growth in the scientific advancements of both FSHD and GQ fields, multiple exciting opportunities arise for the future research that can combine and further explore the two domains of study.

6.3. Concluding remarks

The molecular mechanisms that underlie FSHD pathogenesis have been investigated for over two decades. Here we make a scientific contribution to the field by showing that the GQ structures are novel epigenetic elements involved in regulating *DUX4* expression. Future work on GQ function in FSHD may facilitate the identification of new drug targets and the development of new treatments for the disease as well as fully confirm authenticity of the observed GQ structures as important components in the molecular pathology of FSHD.

7. References

Agarwal, Tani, Saumya Roy, Santosh Kumar, Tushar Kanti Chakraborty, and Souvik Maiti. 2014. "In the Sense of Transcription Regulation by G-Quadruplexes: Asymmetric Effects in Sense and Antisense Strands." *Biochemistry* 53 (23): 3711–18. https://doi.org/10.1021/bi401451q.

Agarwala, Prachi, Satyaprakash Pandey, and Souvik Maiti. 2015. "The Tale of RNA G-Quadruplex." *Organic & Biomolecular Chemistry* 13 (20): 5570–85. https://doi.org/10.1039/C4OB02681K.

Agrawal, Prashansa, Clement Lin, Raveendra I. Mathad, Megan Carver, and Danzhou Yang. 2014. "The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-Nt Loop in K+ Solution." *Journal of the American Chemical Society* 136 (5): 1750–53. https://doi.org/10.1021/ja4118945.

Alexiadis, Vassilios, Mary E Ballestas, Cecilia Sanchez, Sara Winokur, Vettaikorumakankav Vedanarayanan, Mary Warren, and Melanie Ehrlich. 2007. "RNAPol-ChIP Analysis of Transcription from FSHD-Linked Tandem Repeats and Satellite DNA." *Biochimica et Biophysica Acta* 1769 (1): 29–40. https://doi.org/10.1016/j.bbaexp.2006.11.006.

Ambrus, Attila, Ding Chen, Jixun Dai, Tiffanie Bialis, Roger A Jones, and Danzhou Yang. 2006a. "Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution." *Nucleic Acids Research* 34 (9): 2723–35. https://doi.org/10.1093/nar/gkl348.

Ameyar-Zazoua, Maya, Christophe Rachez, Mouloud Souidi, Philippe Robin, Lauriane Fritsch, Robert Young, Nadya Morozova, et al. 2012. "Argonaute Proteins Couple Chromatin Silencing to Alternative Splicing." *Nature Structural & Molecular Biology* 19 (10): 998–1004. https://doi.org/10.1038/nsmb.2373.

Amrane, Samir, Michael Adrian, Brahim Heddi, Alexandre Serero, Alain Nicolas, Jean-Louis Mergny, and Anh Tuân Phan. 2012. "Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite." *Journal of the American Chemical Society* 134 (13): 5807–16. https://doi.org/10.1021/ja208993r.

Amrane, Samir, Abdelaziz Kerkour, Amina Bedrat, Brune Vialet, Marie-Line Andreola, and Jean-Louis Mergny. 2014. "Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development." *Journal of the American Chemical Society* 136 (14): 5249–52. https://doi.org/10.1021/ja501500c.

Angelov, Dimitar, Vladimir A. Bondarenko, Sébastien Almagro, Hervé Menoni, Fabien Mongélard, Fabienne Hans, Flore Mietton, et al. 2006. "Nucleolin Is a Histone Chaperone with FACT-like Activity and Assists Remodeling of Nucleosomes." *The EMBO Journal* 25 (8): 1669–79. https://doi.org/10.1038/sj.emboj.7601046.

Ansseau, Eugénie, Jocelyn O. Eidahl, Céline Lancelot, Alexandra Tassin, Christel Matteotti, Cassandre Yip, Jian Liu, et al. 2016. "Homologous Transcription Factors DUX4 and DUX4c Associate with Cytoplasmic Proteins during Muscle Differentiation." *PLOS ONE* 11 (1): e0146893. https://doi.org/10.1371/journal.pone.0146893.

Ansseau, Eugenie, Dalila Laoudj-Chenivesse, Aline Marcowycz, Alexandra Tassin, Celine Vanderplanck, Sebastien Sauvage, Marietta Barro, et al. 2009. "DUX4c Is Up-Regulated in FSHD. It Induces the MYF5 Protein and Human Myoblast Proliferation." *PloS One* 4 (10): e7482. https://doi.org/10.1371/journal.pone.0007482.

Ansseau, Eugénie, Céline Vanderplanck, Armelle Wauters, Scott Q. Harper, Frédérique Coppée, and Alexandra Belayew. 2017. “Antisense Oligonucleotides Used to Target the DUX4 MRNA as Therapeutic Approaches in FaciosScapuloHumeral Muscular Dystrophy (FSHD).” *Genes* 8 (3). https://doi.org/10.3390/genes8030093.

Arora, Amit, Chandramouli Balasubramanian, Niti Kumar, Saurabh Agrawal, Rajendra P. Ojha, and Souvik Maiti. 2008. “Binding of Berberine to Human Telomeric Quadruplex - Spectroscopic, Calorimetric and Molecular Modeling Studies.” *The FEBS Journal* 275 (15): 3971–83. https://doi.org/10.1111/j.1742-4658.2008.06541.x.

Bailey, R. O., D. C. Marzulo, and M. B. Hans. 1986. “Infantile Facioscapulohumeral Muscular Dystrophy: New Observations.” *Acta Neurologica Scandinavica* 74 (1): 51–58.

Balasubramanian, Shankar, Laurence H Hurley, and Stephen Neidle. 2011a. “Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy?” *Nature Reviews. Drug Discovery* 10 (4): 261–75. https://doi.org/10.1038/nrd3428.

Balog, Judit, Peter E Thijssen, Jessica C de Greef, Bharati Shah, Baziel G M van Engelen, Kyoko Yokomori, Stephen J Tapscott, Rabi Tawil, and Silvere M van der Maarel. 2012. “Correlation Analysis of Clinical Parameters with Epigenetic Modifications in the DUX4 Promoter in FSHD.” *Epigenetics* 7 (6): 579–84. https://doi.org/10.4161/epi.20001.

Bang, I. 1910. “Untersuchungen Über Die Guanylsäure.” *Biochemische Zeitschrift*, no. 26: 293–311.

Bao, Bo, Rika Maruyama, and Toshifumi Yokota. 2016. “Targeting MRNA for the Treatment of Facioscapulohumeral Muscular Dystrophy.” *Intractable & Rare Diseases Research*. https://doi.org/10.5582/irdr.2016.01056.

Barro, Marietta, Gilles Carnac, Sébastien Flavier, Jacques Mercier, Yegor Vassetzky, and Dalila Laoudj-Chenivesse. 2010. “Myoblasts from Affected and Non-Affected FSHD Muscles Exhibit Morphological Differentiation Defects.” *Journal of Cellular and Molecular Medicine* 14 (1–2): 275–89. https://doi.org/10.1111/j.1582-4934.2008.00368.x.

Bazzicalupi, Carla, Marta Ferraroni, Anna Rita Bilia, Francesca Scheggi, and Paola Gratteri. 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 Research* 41 (1): 632–38. https://doi.org/10.1093/nar/gks1001.

Beaudoin, Jean-Denis, Rachel Jodoin, and Jean-Pierre Perreault. 2014. “New Scoring System to Identify RNA G-Quadruplex Folding.” *Nucleic Acids Research* 42 (2): 1209–23. https://doi.org/10.1093/nar/gkt904.

Beckers, M, J Gabriels, S van der Maarel, A De Vriese, R R Frants, D Collen, and A Belayew. 2001. “Active Genes in Junk DNA? Characterization of DUX Genes Embedded within 3.3 Kb Repeated Elements.” *Gene* 264 (1): 51–57.

Bedrat, Amina, Laurent Lacroix, and Jean-Louis Mergny. 2016. “Re-Evaluation of G-Quadruplex Propensity with G4Hunter.” *Nucleic Acids Research* 44 (4): 1746–59. https://doi.org/10.1093/nar/gkw006.

Benjamin, R. C., and D. M. Gill. 1980. “Poly(ADP-Ribose) Synthesis in Vitro Programmed by Damaged DNA. A Comparison of DNA Molecules Containing Different Types of Strand Breaks.” *The Journal of Biological Chemistry* 255 (21): 10502–8.

Bensaid, Mounia, Mireille Melko, Elias G. Bechara, Laetitia Davidovic, Antonio Berretta, Maria Vincenza Catania, Jozef Gecz, Enzo Lalli, and Barbara Bardoni. 2009. “FRAXE-Associated Mental Retardation Protein (FMR2) Is an RNA-Binding Protein with High Affinity for G-Quartet RNA Forming Structure.” *Nucleic Acids Research* 37 (4): 1269–79. https://doi.org/10.1093/nar/gkn1058.

Benson, D W, T Foley-Nelson, W T Chance, F S Zhang, J H James, and J E Fischer. 1991. “Decreased Myofibrillar Protein Breakdown Following Treatment with Clenbuterol.” *The Journal of Surgical Research* 50 (1): 1–5.

Bernardes de Jesus, Bruno, and Maria A. Blasco. 2013. “Telomerase at the Intersection of Cancer and Aging.” *Trends in Genetics: TIG* 29 (9): 513–20. https://doi.org/10.1016/j.tig.2013.06.007.

Berry, Suzanne E. 2015. “Concise Review: Mesoangioblast and Mesenchymal Stem Cell Therapy for Muscular Dystrophy: Progress, Challenges, and Future Directions.” *Stem Cells Translational Medicine* 4 (1): 91–98. https://doi.org/10.5966/sctm.2014-0060.

Bhadra, Kakali, and Gopinatha Suresh Kumar. 2011a. “Interaction of Berberine, Palmatine, Coralyne, and Sanguinarine to Quadruplex DNA: A Comparative Spectroscopic and Calorimetric Study.” *Biochimica et Biophysica Acta (BBA) - General Subjects* 1810 (4): 485–96. https://doi.org/10.1016/j.bbagen.2011.01.011.

Biffi, Giulia, David Tannahill, and Shankar Balasubramanian. 2012. “An Intramolecular G-Quadruplex Structure Is Required for Binding of Telomeric Repeat-Containing RNA to the Telomeric Protein TRF2.” *Journal of the American Chemical Society* 134 (29): 11974–76. https://doi.org/10.1021/ja305734x.

Biffi, Giulia, David Tannahill, John McCafferty, and Shankar Balasubramanian. 2013. “Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells.” *Nature Chemistry* 5 (3): 182–86. https://doi.org/10.1038/nchem.1548.

Bonnal, Sophie, Céline Schaeffer, Laurent Créancier, Simone Clamens, Hervé Moine, Anne-Catherine Prats, and Stéphan Vagner. 2003. “A Single Internal Ribosome Entry Site Containing a G Quartet RNA Structure Drives Fibroblast Growth Factor 2 Gene Expression at Four Alternative Translation Initiation Codons.” *The Journal of Biological Chemistry* 278 (41): 39330–36. https://doi.org/10.1074/jbc.M305580200.

Booy, E. P., M. Meier, N. Okun, S. K. Novakowski, S. Xiong, J. Stetefeld, and S. A. McKenna. 2012. “The RNA Helicase RHAU (DHX36) Unwinds a G4-Quadruplex in Human Telomerase RNA and Promotes the Formation of the P1 Helix Template Boundary.” *Nucleic Acids Research* 40 (9): 4110–24. https://doi.org/10.1093/nar/gkr1306.

Bosnakovski, Darko, Sunny S K Chan, Olivia O Recht, Lynn M Hartweck, Collin J Gustafson, Laura L Athman, Dawn A Lowe, and Michael Kyba. 2017. “Muscle Pathology from Stochastic Low Level DUX4 Expression in an FSHD Mouse Model.” *Nature Communications* 8 (1): 550. https://doi.org/10.1038/s41467-017-00730-1.

Bosnakovski, Darko, Zhaohui Xu, Eun Ji Gang, Cristi L Galindo, Mingju Liu, Tugba Simsek, Harold R Garner, et al. 2008. “An Isogenetic Myoblast Expression Screen Identifies DUX4-Mediated FSHD-Associated Molecular Pathologies.” *The EMBO Journal* 27 (20): 2766–79. https://doi.org/10.1038/emboj.2008.201.

Brázda, Václav, Lucia Hároníková, Jack C. C. Liao, and Miroslav Fojta. 2014. “DNA and RNA Quadruplex-Binding Proteins.” *International Journal of Molecular Sciences* 15 (10): 17493–517. https://doi.org/10.3390/ijms151017493.

Brooks, Tracy A., Samantha Kendrick, and Laurence Hurley. 2010. “Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters.” *The FEBS Journal* 277 (17): 3459–69. https://doi.org/10.1111/j.1742-4658.2010.07759.x.

Brouwer, O F, C Wijmenga, R R Frants, and G W Padberg. 1993. “Facioscapulohumeral Muscular Dystrophy: The Impact of Genetic Research.” *Clinical Neurology and Neurosurgery* 95 (1): 9–21.

Brown, Robert V., Forest L. Danford, Vijay Gokhale, Laurence H. Hurley, and Tracy A. Brooks. 2011. “Demonstration That Drug-Targeted down-Regulation of MYC in Non-Hodgkins Lymphoma Is Directly Mediated through the Promoter G-Quadruplex.” *The Journal of Biological Chemistry* 286 (47): 41018–27. https://doi.org/10.1074/jbc.M111.274720.

Buckingham, Margaret, and Frederic Relaix. 2007. “The Role of Pax Genes in the Development of Tissues and Organs: Pax3 and Pax7 Regulate Muscle Progenitor Cell Functions.” *Annual Review of Cell and Developmental Biology* 23: 645–73. https://doi.org/10.1146/annurev.cellbio.23.090506.123438.

Bugaut, Anthony, and Shankar Balasubramanian. 2008. “A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes.” *Biochemistry* 47 (2): 689–97. https://doi.org/10.1021/bi701873c.

Burger, Angelika M., Fangping Dai, Christoph M. Schultes, Anthony P. Reszka, Michael J. Moore, John A. Double, and Stephen Neidle. 2005. “The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function.” *Cancer Research* 65 (4): 1489–96. https://doi.org/10.1158/0008-5472.CAN-04-2910.

Cabianca, Daphne S., Valentina Casa, Beatrice Bodega, Alexandros Xynos, Enrico Ginelli, Yujiro Tanaka, and Davide Gabellini. 2012. “A Long NcRNA Links Copy Number Variation to a Polycomb/Trithorax Epigenetic Switch in FSHD Muscular Dystrophy.” *Cell* 149 (4): 819–31. https://doi.org/10.1016/j.cell.2012.03.035.

Cammas, Anne, Alexandre Dubrac, Benjamin Morel, Assala Lamaa, Christian Touriol, Marie-Paule Teulade-Fichou, Hervé Prats, and Stefania Millevoi. 2015. “Stabilization of the G-Quadruplex at the VEGF IRES Represses Cap-Independent Translation.” *RNA Biology* 12 (3): 320–29. https://doi.org/10.1080/15476286.2015.1017236.

Capra, John A., Katrin Paeschke, Mona Singh, and Virginia A. Zakian. 2010. “G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces Cerevisiae.” *PLoS Computational Biology* 6 (7): e1000861. https://doi.org/10.1371/journal.pcbi.1000861.

Castets, Marie, Céline Schaeffer, Elias Bechara, Annette Schenck, Edward W. Khandjian, Sylvie Luche, Hervé Moine, Thierry Rabilloud, Jean-Louis Mandel, and Barbara Bardoni. 2005. “FMRP Interferes with the Rac1 Pathway and Controls Actin Cytoskeleton Dynamics in Murine Fibroblasts.” *Human Molecular Genetics* 14 (6): 835–44. https://doi.org/10.1093/hmg/ddi077.

Chambers, Vicki S, Giovanni Marsico, Jonathan M Boutell, Marco Di Antonio, Geoffrey P Smith, and Shankar Balasubramanian. 2015a. “High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome.” *Nature Biotechnology* 33 (8): 877–81. https://doi.org/10.1038/nbt.3295.

Chambers, Vicki S, Giovanni Marsico, Jonathan M Boutell, Marco Di Antonio, Geoffrey P Smith, and Shankar Balasubramanian. 2015b. “High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome.” *Nature Biotechnology* 33 (8): 877–81. https://doi.org/10.1038/nbt.3295.

Cheung, Iris, Michael Schertzer, Ann Rose, and Peter M Lansdorp. 2002. “Disruption of Dog-1 in Caenorhabditis Elegans Triggers Deletions Upstream of Guanine-Rich DNA.” *Nature Genetics* 31 (4): 405–9. https://doi.org/10.1038/ng928.

Chiarella, Sara, Antonella De Cola, Giovanni Luca Scaglione, Erminia Carletti, Vincenzo Graziano, Daniela Barcaroli, Carlo Lo Sterzo, et al. 2013. “Nucleophosmin Mutations Alter Its Nucleolar Localization by Impairing G-Quadruplex Binding at Ribosomal DNA.” *Nucleic Acids Research* 41 (5): 3228–39. https://doi.org/10.1093/nar/gkt001.

Clapp, Jannine, Laura M Mitchell, Daniel J Bolland, Judy Fantes, Anne E Corcoran, Paul J Scotting, John A L Armour, and Jane E Hewitt. 2007. “Evolutionary Conservation of a Coding Function for D4Z4, the Tandem DNA Repeat Mutated in Facioscapulohumeral Muscular Dystrophy.” *American Journal of Human Genetics* 81 (2): 264–79. https://doi.org/10.1086/519311.

Cogoi, Susanna, Manikandan Paramasivam, Alexandro Membrino, Kazunari K. Yokoyama, and Luigi E. Xodo. 2010. “The KRAS Promoter Responds to Myc-Associated Zinc Finger and Poly(ADP-Ribose) Polymerase 1 Proteins, Which Recognize a Critical Quadruplex-Forming GA-Element.” *The Journal of Biological Chemistry* 285 (29): 22003–16. https://doi.org/10.1074/jbc.M110.101923.

Collie, Gavin W., Shozeb M. Haider, Stephen Neidle, and Gary N. Parkinson. 2010. “A Crystallographic and Modelling Study of a Human Telomeric RNA (TERRA) Quadruplex.” *Nucleic Acids Research* 38 (16): 5569–80. https://doi.org/10.1093/nar/gkq259.

Collie, Gavin W., Silvia Sparapani, Gary N. Parkinson, and Stephen Neidle. 2011. “Structural Basis of Telomeric RNA Quadruplex−Acridine Ligand Recognition.” *Journal of the American Chemical Society* 133 (8): 2721–28. https://doi.org/10.1021/ja109767y.

Cong, Yu-Sheng, Woodring E. Wright, and Jerry W. Shay. 2002. “Human Telomerase and Its Regulation.” *Microbiology and Molecular Biology Reviews: MMBR* 66 (3): 407–25, table of contents.

Cree Simone L., Fredericks Rayleen, Miller Allison, Pearce F. Grant, Filichev Vyacheslav, Fee Conan, and Kennedy Martin A. 2016. “DNA G-quadruplexes Show Strong Interaction with DNA Methyltransferases in Vitro.” *FEBS Letters* 590 (17): 2870–83. https://doi.org/10.1002/1873-3468.12331.

Cui, Xiaojie, Sen Lin, Jiang Zhou, and Gu Yuan. 2012. “Investigation of Non-Covalent Interaction of Natural Flexible Cyclic Molecules with Telomeric RNA G-Quadruplexes by Electrospray Ionization Mass Spectrometry.” *Rapid Communications in Mass Spectrometry: RCM* 26 (16): 1803–9. https://doi.org/10.1002/rcm.6295.

Cui, Yunxi, Deepak Koirala, HyunJin Kang, Soma Dhakal, Philip Yangyuoru, Laurence H. Hurley, and Hanbin Mao. 2014. “Molecular Population Dynamics of DNA Structures in a Bcl-2 Promoter Sequence Is Regulated by Small Molecules and the Transcription Factor HnRNP LL.” *Nucleic Acids Research* 42 (9): 5755–64. https://doi.org/10.1093/nar/gku185.

Dai, Jixun, Megan Carver, and Danzhou Yang. 2008. “Polymorphism of Human Telomeric Quadruplex Structures.” *Biochimie* 90 (8): 1172–83. https://doi.org/10.1016/j.biochi.2008.02.026.

Darnell, J. C., K. B. Jensen, P. Jin, V. Brown, S. T. Warren, and R. B. Darnell. 2001. “Fragile X Mental Retardation Protein Targets G Quartet MRNAs Important for Neuronal Function.” *Cell* 107 (4): 489–99.

Dash, Jyotirmayee, Pravin S. Shirude, and Shankar Balasubramanian. 2008. “G-Quadruplex Recognition by Bis-Indole Carboxamides.” *Chemical Communications (Cambridge, England)*, no. 26 (July): 3055–57. https://doi.org/10.1039/b806042h.

Davis, Jeffery T. 2004. “G-Quartets 40 Years Later: From 5’-GMP to Molecular Biology and Supramolecular Chemistry.” *Angewandte Chemie (International Ed. in English)* 43 (6): 668–98. https://doi.org/10.1002/anie.200300589.

Daxinger, Lucia, Stephen J Tapscott, and Silvere M van der Maarel. 2015. “Genetic and Epigenetic Contributors to FSHD.” *Current Opinion in Genetics & Development* 33 (August): 56–61. https://doi.org/10.1016/j.gde.2015.08.007.

De Armond, Richard, Stacey Wood, Daekyu Sun, Laurence H. Hurley, and Scot W. Ebbinghaus. 2005. “Evidence for the Presence of a Guanine Quadruplex Forming Region within a Polypurine Tract of the Hypoxia Inducible Factor 1alpha Promoter.” *Biochemistry* 44 (49): 16341–50. https://doi.org/10.1021/bi051618u.

De Cian, Anne, Gael Cristofari, Patrick Reichenbach, Elsa De Lemos, David Monchaud, Marie-Paule Teulade-Fichou, Kazuo Shin-Ya, Laurent Lacroix, Joachim Lingner, and Jean-Louis Mergny. 2007. “Reevaluation of Telomerase Inhibition by Quadruplex Ligands and Their Mechanisms of Action.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (44): 17347–52. https://doi.org/10.1073/pnas.0707365104.

Dean, Ann. 2011. “In the Loop: Long Range Chromatin Interactions and Gene Regulation.” *Briefings in Functional Genomics* 10 (1): 3–10. https://doi.org/10.1093/bfgp/elq033.

Decorsière, Adrien, Anne Cayrel, Stéphan Vagner, and Stefania Millevoi. 2011. “Essential Role for the Interaction between HnRNP H/F and a G Quadruplex in Maintaining P53 Pre-MRNA 3’-End Processing and Function during DNA Damage.” *Genes & Development* 25 (3): 220–25. https://doi.org/10.1101/gad.607011.

Deenen, Johanna C.W., Hisse Arnts, Silvère M. van der Maarel, George W. Padberg, Jan J.G.M. Verschuuren, Egbert Bakker, Stephanie S. Weinreich, André L.M. Verbeek, and Baziel G.M. van Engelen. 2014. “Population-Based Incidence and Prevalence of Facioscapulohumeral Dystrophy.” *Neurology* 83 (12): 1056–59. https://doi.org/10.1212/WNL.0000000000000797.

Dempsey, L A, H Sun, L A Hanakahi, and N Maizels. 1999. “G4 DNA Binding by LR1 and Its Subunits, Nucleolin and HnRNP D, A Role for G-G Pairing in Immunoglobulin Switch Recombination.” *The Journal of Biological Chemistry* 274 (2): 1066–71.

Deutekom, J C van, C Wijmenga, E A van Tienhoven, A M Gruter, J E Hewitt, G W Padberg, G J van Ommen, M H Hofker, and R R Frants. 1993. “FSHD Associated DNA Rearrangements Are Due to Deletions of Integral Copies of a 3.2 Kb Tandemly Repeated Unit.” *Human Molecular Genetics* 2 (12): 2037–42.

DeVos, Sarah L., and Timothy M. Miller. 2013. “Antisense Oligonucleotides: Treating Neurodegeneration at the Level of RNA.” *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics* 10 (3): 486–97. https://doi.org/10.1007/s13311-013-0194-5.

Di Antonio, Marco, Giulia Biffi, Angelica Mariani, Eun-Ang Raiber, Raphaël Rodriguez, and Shankar Balasubramanian. 2012. “Selective RNA versus DNA G-Quadruplex Targeting by in Situ Click Chemistry.” *Angewandte Chemie (International Ed. in English)* 51 (44): 11073–78. https://doi.org/10.1002/anie.201206281.

Didiot, Marie-Cécile, Zhaoxia Tian, Céline Schaeffer, Murugan Subramanian, Jean-Louis Mandel, and Hervé Moine. 2008. “The G-Quartet Containing FMRP Binding Site in FMR1 MRNA Is a Potent Exonic Splicing Enhancer.” *Nucleic Acids Research* 36 (15): 4902–12. https://doi.org/10.1093/nar/gkn472.

Ding, H., M. C. Beckers, S. Plaisance, P. Marynen, D. Collen, and A. Belayew. 1998. “Characterization of a Double Homeodomain Protein (DUX1) Encoded by a CDNA Homologous to 3.3 Kb Dispersed Repeated Elements.” *Human Molecular Genetics* 7 (11): 1681–94.

Dixit, Manjusha, Eugenie Ansseau, Alexandra Tassin, Sara Winokur, Rongye Shi, Hong Qian, Sebastien Sauvage, et al. 2007a. “DUX4, a Candidate Gene of Facioscapulohumeral Muscular Dystrophy, Encodes a Transcriptional Activator of PITX1.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (46): 18157–62. https://doi.org/10.1073/pnas.0708659104.

Dolinnaya, N. G., A. M. Ogloblina, and M. G. Yakubovskaya. 2016. “Structure, Properties, and Biological Relevance of the DNA and RNA G-Quadruplexes: Overview 50 Years after Their Discovery.” *Biochemistry. Biokhimiia* 81 (13): 1602–49. https://doi.org/10.1134/S0006297916130034.

Dominguez, Cyril, Jean-François Fisette, Benoit Chabot, and Frédéric H.-T. Allain. 2010. “Structural Basis of G-Tract Recognition and Encaging by HnRNP F Quasi-RRMs.” *Nature Structural & Molecular Biology* 17 (7): 853–61. https://doi.org/10.1038/nsmb.1814.

Dorobek, Malgorzata, Elzbieta Szmidt-Salkowska, Katarzyna Rowinska-Marcinska, Malgorzata Gawel, and Irena Hausmanowa-Petrusewicz. 2013. “Relationships between Clinical Data and Quantitative EMG Findings in Facioscapulohumeral Muscular Dystrophy.” *Neurologia i Neurochirurgia Polska* 47 (1): 8–17.

Du, Zhuo, Yiqiang Zhao, and Ning Li. 2008. “Genome-Wide Analysis Reveals Regulatory Role of G4 DNA in Gene Transcription.” *Genome Research* 18 (2): 233–41. https://doi.org/10.1101/gr.6905408.

Duchenne, GB. 1862. *Album de Photographies Pathologiques*. Paris J. B. Bailli re et fils. 1869. *Album de Photographies Pathologiques.* Paris Adrien Delahaye.

Duquette, Michelle L., Priya Handa, Jack A. Vincent, Andrew F. Taylor, and Nancy Maizels. 2004. “Intracellular Transcription of G-Rich DNAs Induces Formation of G-Loops, Novel Structures Containing G4 DNA.” *Genes & Development* 18 (13): 1618–29. https://doi.org/10.1101/gad.1200804.

Egger, L., D. T. Madden, C. Rhême, R. V. Rao, and D. E. Bredesen. 2007. “Endoplasmic Reticulum Stress-Induced Cell Death Mediated by the Proteasome.” *Cell Death and Differentiation* 14 (6): 1172–80. https://doi.org/10.1038/sj.cdd.4402125.

Egorina, E. M., M. A. Sovershaev, and B. Østerud. 2006. “In-Cell Western Assay: A New Approach to Visualize Tissue Factor in Human Monocytes.” *Journal of Thrombosis and Haemostasis: JTH* 4 (3): 614–20. https://doi.org/10.1111/j.1538-7836.2005.01781.x.

Endoh, Tamaki, and Naoki Sugimoto. 2016. “Mechanical Insights into Ribosomal Progression Overcoming RNA G-Quadruplex from Periodical Translation Suppression in Cells.” *Scientific Reports* 6 (March): 22719. https://doi.org/10.1038/srep22719.

Fang, G, and T R Cech. 1993. “The Beta Subunit of Oxytricha Telomere-Binding Protein Promotes G-Quartet Formation by Telomeric DNA.” *Cell* 74 (5): 875–85.

Farhath, Mohamed M., Matthew Thompson, Sujay Ray, Abby Sewell, Hamza Balci, and Soumitra Basu. 2015. “G-Quadruplex-Enabling Sequence within the Human Tyrosine Hydroxylase Promoter Differentially Regulates Transcription.” *Biochemistry* 54 (36): 5533–45. https://doi.org/10.1021/acs.biochem.5b00209.

Fay, Marta M., Shawn M. Lyons, and Pavel Ivanov. 2017. “RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms.” *Journal of Molecular Biology* 429 (14): 2127–47. https://doi.org/10.1016/j.jmb.2017.05.017.

Feng, Qing, Lauren Snider, Sujatha Jagannathan, Rabi Tawil, Silvere M van der Maarel, Stephen J Tapscott, and Robert K Bradley. 2015. “A Feedback Loop between Nonsense-Mediated Decay and the Retrogene DUX4 in Facioscapulohumeral Muscular Dystrophy.” *ELife* 4 (January). https://doi.org/10.7554/eLife.04996.

Ferreboeuf, Maxime, Virginie Mariot, Bettina Bessières, Alexandre Vasiljevic, Tania Attié-Bitach, Sophie Collardeau, Julia Morere, et al. 2014. “DUX4 and DUX4 Downstream Target Genes Are Expressed in Fetal FSHD Muscles.” *Human Molecular Genetics* 23 (1): 171–81. https://doi.org/10.1093/hmg/ddt409.

Ferreboeuf, Maxime, Virginie Mariot, Denis Furling, Gillian Butler-Browne, Vincent Mouly, and Julie Dumonceaux. 2014. “Nuclear Protein Spreading: Implication for Pathophysiology of Neuromuscular Diseases.” *Human Molecular Genetics* 23 (15): 4125–33. https://doi.org/10.1093/hmg/ddu129.

Fisette, Jean-François, Daniel R. Montagna, Mihaela-Rita Mihailescu, and Michael S. Wolfe. 2012. “A G-Rich Element Forms a G-Quadruplex and Regulates BACE1 MRNA Alternative Splicing.” *Journal of Neurochemistry* 121 (5): 763–73. https://doi.org/10.1111/j.1471-4159.2012.07680.x.

Fu, Lingyi, Wangbing Chen, Wei Guo, Jingshu Wang, Yun Tian, Dingbo Shi, Xiaohong Zhang, et al. 2013. “Berberine Targets AP-2/HTERT, NF-KB/COX-2, HIF-1α/VEGF and Cytochrome-c/Caspase Signaling to Suppress Human Cancer Cell Growth.” *PloS One* 8 (7): e69240. https://doi.org/10.1371/journal.pone.0069240.

Gabriels, J, M C Beckers, H Ding, A De Vriese, S Plaisance, S M van der Maarel, G W Padberg, et al. 1999a. “Nucleotide Sequence of the Partially Deleted D4Z4 Locus in a Patient with FSHD Identifies a Putative Gene within Each 3.3 Kb Element.” *Gene* 236 (1): 25–32.

Gabriels, J, M C Beckers, H Ding, A De Vriese, S Plaisance, S M van der Maarel, G W Padberg, et al. 1999b. “Nucleotide Sequence of the Partially Deleted D4Z4 Locus in a Patient with FSHD Identifies a Putative Gene within Each 3.3 Kb Element.” *Gene* 236 (1): 25–32.

Ganesan, Kumar, and Baojun Xu. 2017. “Telomerase Inhibitors from Natural Products and Their Anticancer Potential.” *International Journal of Molecular Sciences* 19 (1). https://doi.org/10.3390/ijms19010013.

Garner, Thomas P., Huw E. L. Williams, Katarzyna I. Gluszyk, Stephen Roe, Neil J. Oldham, Malcolm F. G. Stevens, John E. Moses, and Mark S. Searle. 2009. “Selectivity of Small Molecule Ligands for Parallel and Anti-Parallel DNA G-Quadruplex Structures.” *Organic & Biomolecular Chemistry* 7 (20): 4194–4200. https://doi.org/10.1039/B910505K.

Gehring, W J, Y Q Qian, M Billeter, K Furukubo-Tokunaga, A F Schier, D Resendez-Perez, M Affolter, G Otting, and K Wuthrich. 1994. “Homeodomain-DNA Recognition.” *Cell* 78 (2): 211–23.

GELLERT, M, M N LIPSETT, and D R DAVIES. 1962. “Helix Formation by Guanylic Acid.” *Proceedings of the National Academy of Sciences of the United States of America* 48 (December): 2013–18.

Geng, Linda N., Ashlee E. Tyler, and Stephen J. Tapscott. 2011. “Immunodetection of Human Double Homeobox 4.” *Hybridoma* 30 (2): 125–30. https://doi.org/10.1089/hyb.2010.0094.

Geng, Linda N, Zizhen Yao, Lauren Snider, Abraham P Fong, Jennifer N Cech, Janet M Young, Silvere M van der Maarel, et al. 2012a. “DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy.” *Developmental Cell* 22 (1): 38–51. https://doi.org/10.1016/j.devcel.2011.11.013.

Gilbert, Luke A, Matthew H Larson, Leonardo Morsut, Zairan Liu, Gloria A Brar, Sandra E Torres, Noam Stern-Ginossar, et al. 2013. “CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes.” *Cell* 154 (2): 442–51. https://doi.org/10.1016/j.cell.2013.06.044.

Ginisty, H., H. Sicard, B. Roger, and P. Bouvet. 1999. “Structure and Functions of Nucleolin.” *Journal of Cell Science* 112 (Pt 6) (March): 761–72.

González, Verónica, Kexiao Guo, Laurence Hurley, and Daekyu Sun. 2009. “Identification and Characterization of Nucleolin as a C-Myc G-Quadruplex-Binding Protein.” *The Journal of Biological Chemistry* 284 (35): 23622–35. https://doi.org/10.1074/jbc.M109.018028.

González, Verónica, and Laurence H. Hurley. 2010. “The C-MYC NHE III(1): Function and Regulation.” *Annual Review of Pharmacology and Toxicology* 50: 111–29. https://doi.org/10.1146/annurev.pharmtox.48.113006.094649.

Grand, Cory L., Haiyong Han, Rubén M. Muñoz, Steve Weitman, Daniel D. Von Hoff, Laurence H. Hurley, and David J. Bearss. 2002. “The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo.” *Molecular Cancer Therapeutics* 1 (8): 565–73.

Gray, Lucas T, Aarthy C Vallur, Johanna Eddy, and Nancy Maizels. 2014. “G Quadruplexes Are Genomewide Targets of Transcriptional Helicases XPB and XPD.” *Nature Chemical Biology* 10 (4): 313–18. https://doi.org/10.1038/nchembio.1475.

Greef, J C de, R J L F Lemmers, P Camano, J W Day, S Sacconi, M Dunand, B G M van Engelen, et al. 2010. “Clinical Features of Facioscapulohumeral Muscular Dystrophy 2.” *Neurology* 75 (17): 1548–54. https://doi.org/10.1212/WNL.0b013e3181f96175.

Gromak, Natalia, Steven West, and Nick J. Proudfoot. 2006. “Pause Sites Promote Transcriptional Termination of Mammalian RNA Polymerase II.” *Molecular and Cellular Biology* 26 (10): 3986–96. https://doi.org/10.1128/MCB.26.10.3986-3996.2006.

Gu, Hui-Ping, Sen Lin, Ming Xu, Hai-Yi Yu, Xiao-Jun Du, You-Yi Zhang, Gu Yuan, and Wei Gao. 2012. “Up-Regulating Relaxin Expression by G-Quadruplex Interactive Ligand to

Achieve Antifibrotic Action." *Endocrinology* 153 (8): 3692–3700. https://doi.org/10.1210/en.2012-1114.
Guamán Ortiz, L.M., and A.I. Scovassi. 2013. "Traditional Medicine: An Ancient Remedy Rediscovered." 2 (1).
Guédin, Aurore, Julien Gros, Patrizia Alberti, and Jean-Louis Mergny. 2010. "How Long Is Too Long? Effects of Loop Size on G-Quadruplex Stability." *Nucleic Acids Research* 38 (21): 7858–68. https://doi.org/10.1093/nar/gkq639.
Guo, Changying, Tatiana Gerasimova, Haiping Hao, Irina Ivanova, Tirtha Chakraborty, Roza Selimyan, Eugene M. Oltz, and Ranjan Sen. 2011. "Two Forms of Loops Generate the Chromatin Conformation of the Immunoglobulin Heavy-Chain Gene Locus." *Cell* 147 (2): 332–43. https://doi.org/10.1016/j.cell.2011.08.049.
Guo, Junjie U., and David P. Bartel. 2016. "RNA G-Quadruplexes Are Globally Unfolded in Eukaryotic Cells and Depleted in Bacteria." *Science (New York, N.Y.)* 353 (6306). https://doi.org/10.1126/science.aaf5371.
Hansel-Hertsch, Robert, Marco Di Antonio, and Shankar Balasubramanian. 2017. "DNA G-Quadruplexes in the Human Genome: Detection, Functions and Therapeutic Potential." *Nature Reviews. Molecular Cell Biology* 18 (5): 279–84. https://doi.org/10.1038/nrm.2017.3.
Hansel-Hertsch, Robert, Dario Beraldi, Stefanie V Lensing, Giovanni Marsico, Katherine Zyner, Aled Parry, Marco Di Antonio, et al. 2016. "G-Quadruplex Structures Mark Human Regulatory Chromatin." *Nature Genetics* 48 (10): 1267–72. https://doi.org/10.1038/ng.3662.
He, T. C., A. B. Sparks, C. Rago, H. Hermeking, L. Zawel, L. T. da Costa, P. J. Morin, B. Vogelstein, and K. W. Kinzler. 1998. "Identification of C-MYC as a Target of the APC Pathway." *Science (New York, N.Y.)* 281 (5382): 1509–12.
Heidarian, Esfandiar, Mahmoud Rafieian-Kopaei, Abolfazle Khoshdel, and Morteza Bakhshesh. 2014. "Metabolic Effects of Berberine on Liver Phosphatidate Phosphohydrolase in Rats Fed on High Lipogenic Diet: An Additional Mechanism for the Hypolipidemic Effects of Berberine." *Asian Pacific Journal of Tropical Biomedicine* 4 (Suppl 1): S429-435. https://doi.org/10.12980/APJTB.4.2014C474.
Henderson, Alexander, Yuliang Wu, Yu Chuan Huang, Elizabeth A Chavez, Jesse Platt, F Brad Johnson, Robert M Jr Brosh, Dipankar Sen, and Peter M Lansdorp. 2014. "Detection of G-Quadruplex DNA in Mammalian Cells." *Nucleic Acids Research* 42 (2): 860–69. https://doi.org/10.1093/nar/gkt957.
Henderson, E, C C Hardin, S K Walk, I Jr Tinoco, and E H Blackburn. 1987. "Telomeric DNA Oligonucleotides Form Novel Intramolecular Structures Containing Guanine-Guanine Base Pairs." *Cell* 51 (6): 899–908.
Hershman, Steve G., Qijun Chen, Julia Y. Lee, Marina L. Kozak, Peng Yue, Li-San Wang, and F. Brad Johnson. 2008. "Genomic Distribution and Functional Analyses of Potential G-Quadruplex-Forming Sequences in Saccharomyces Cerevisiae." *Nucleic Acids Research* 36 (1): 144–56. https://doi.org/10.1093/nar/gkm986.
Hewitt, J E, R Lyle, L N Clark, E M Valleley, T J Wright, C Wijmenga, J C van Deutekom, F Francis, P T Sharpe, and M Hofker. 1994. "Analysis of the Tandem Repeat Locus D4Z4 Associated with Facioscapulohumeral Muscular Dystrophy." *Human Molecular Genetics* 3 (8): 1287–95.
Himeda, Charis L., Céline Debarnot, Sachiko Homma, Mary Lou Beermann, Jeffrey B. Miller, Peter L. Jones, and Takako I. Jones. 2014. "Myogenic Enhancers Regulate Expression of the Facioscapulohumeral Muscular Dystrophy-Associated DUX4 Gene." *Molecular and Cellular Biology* 34 (11): 1942–55. https://doi.org/10.1128/MCB.00149-14.
Himeda, Charis L, Takako I Jones, and Peter L Jones. 2016a. "CRISPR/DCas9-Mediated Transcriptional Inhibition Ameliorates the Epigenetic Dysregulation at D4Z4 and

Represses DUX4-Fl in FSH Muscular Dystrophy." *Molecular Therapy : The Journal of the American Society of Gene Therapy* 24 (3): 527–35. https://doi.org/10.1038/mt.2015.200.

Holoch, Daniel, and Danesh Moazed. 2015. "RNA-Mediated Epigenetic Regulation of Gene Expression." *Nature Reviews. Genetics* 16 (2): 71–84. https://doi.org/10.1038/nrg3863.

Hsu, Shang-Te Danny, Peter Varnai, Anthony Bugaut, Anthony P. Reszka, Stephen Neidle, and Shankar Balasubramanian. 2009. "A G-Rich Sequence within the c-Kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics." *Journal of the American Chemical Society* 131 (37): 13399–409. https://doi.org/10.1021/ja904007p.

Huang, Wei-Chun, Ting-Yuan Tseng, Ying-Ting Chen, Cheng-Chung Chang, Zi-Fu Wang, Chiung-Lin Wang, Tsu-Ning Hsu, et al. 2015. "Direct Evidence of Mitochondrial G-Quadruplex DNA by Using Fluorescent Anti-Cancer Agents." *Nucleic Acids Research* 43 (21): 10102–13. https://doi.org/10.1093/nar/gkv1061.

Huang, Weiwei, Philip J. Smaldino, Qiang Zhang, Lance D. Miller, Paul Cao, Kristin Stadelman, Meimei Wan, et al. 2012. "Yin Yang 1 Contains G-Quadruplex Structures in Its Promoter and 5'-UTR and Its Expression Is Modulated by G4 Resolvase 1." *Nucleic Acids Research* 40 (3): 1033–49. https://doi.org/10.1093/nar/gkr849.

Huppert, Julian L. 2010. "Structure, Location and Interactions of G-Quadruplexes." *The FEBS Journal* 277 (17): 3452–58. https://doi.org/10.1111/j.1742-4658.2010.07758.x.

Huppert, Julian L, and Shankar Balasubramanian. 2005a. "Prevalence of Quadruplexes in the Human Genome." *Nucleic Acids Research* 33 (9): 2908–16. https://doi.org/10.1093/nar/gki609.

Huppert, Julian Leon. 2008. "Four-Stranded Nucleic Acids: Structure, Function and Targeting of G-Quadruplexes." *Chemical Society Reviews* 37 (7): 1375–84. https://doi.org/10.1039/b702491f.

Ishibashi, Jeff, Robert L. Perry, Atsushi Asakura, and Michael A. Rudnicki. 2005. "MyoD Induces Myogenic Differentiation through Cooperation of Its NH2- and COOH-Terminal Regions." *The Journal of Cell Biology* 171 (3): 471–82. https://doi.org/10.1083/jcb.200502101.

Ishtikhar, Mohd, Shawez Khan, Gamal Badr, Amany Osama Mohamed, and Rizwan Hasan Khan. 2014. "Interaction of the 5-Fluorouracil Analog 5-Fluoro-2'-Deoxyuridine with 'N' and 'B' Isoforms of Human Serum Albumin: A Spectroscopic and Calorimetric Study." *Molecular BioSystems* 10 (11): 2954–64. https://doi.org/10.1039/c4mb00306c.

Jayaraj, Gopal Gunanathan, Satyaprakash Pandey, Vinod Scaria, and Souvik Maiti. 2012. "Potential G-Quadruplexes in the Human Long Non-Coding Transcriptome." *RNA Biology* 9 (1): 81–86. https://doi.org/10.4161/rna.9.1.18047.

Jin, Kyeong Sik, Su Ryon Shin, Byungcheol Ahn, Yecheol Rho, Seon Jeong Kim, and Moonhor Ree. 2009. "PH-Dependent Structures of an i-Motif DNA in Solution." *The Journal of Physical Chemistry B* 113 (7): 1852–56. https://doi.org/10.1021/jp808186z.

Joachimi, Astrid, Armin Benz, and Jörg S. Hartig. 2009. "A Comparison of DNA and RNA Quadruplex Structures and Stabilities." *Bioorganic & Medicinal Chemistry* 17 (19): 6811–15. https://doi.org/10.1016/j.bmc.2009.08.043.

Jones, Takako, and Peter L. Jones. 2018. "A Cre-Inducible DUX4 Transgenic Mouse Model for Investigating Facioscapulohumeral Muscular Dystrophy." *PLoS ONE* 13 (2). https://doi.org/10.1371/journal.pone.0192657.

Kanoh, Yutaka, Seiji Matsumoto, Rino Fukatsu, Naoko Kakusho, Nobuaki Kono, Claire Renard-Guillet, Koji Masuda, et al. 2015. "Rif1 Binds to G Quadruplexes and Suppresses Replication over Long Distances." *Nature Structural & Molecular Biology* 22 (11): 889–97. https://doi.org/10.1038/nsmb.3102.

Katsuda, Yousuke, Shin-Ichi Sato, Lisa Asano, Yoshitaka Morimura, Tomoyuki Furuta, Hiroshi Sugiyama, Masaki Hagihara, and Motonari Uesugi. 2016. “A Small Molecule That Represses Translation of G-Quadruplex-Containing MRNA.” *Journal of the American Chemical Society* 138 (29): 9037–40. https://doi.org/10.1021/jacs.6b04506.

Kemp, G J, D J Taylor, J F Dunn, S P Frostick, and G K Radda. 1993. “Cellular Energetics of Dystrophic Muscle.” *Journal of the Neurological Sciences* 116 (2): 201–6.

Kim, N W, M A Piatyszek, K R Prowse, C B Harley, M D West, P L Ho, G M Coviello, W E Wright, S L Weinrich, and J W Shay. 1994. “Specific Association of Human Telomerase Activity with Immortal Cells and Cancer.” *Science (New York, N.Y.)* 266 (5193): 2011–15.

Kohler, J.C., and G. Baghdadi-Sabeti. 2011. *Traditional Medicines: Global Situation, Issues and Challenges. The World Medicines Situation 2011*. 3rd ed. WHO: Geneva, Switzerland.

Konig, Sebastian L B, Amanda C Evans, and Julian L Huppert. 2010. “Seven Essential Questions on G-Quadruplexes.” *Biomolecular Concepts* 1 (2): 197–213. https://doi.org/10.1515/bmc.2010.011.

Kooi, A J van der, M C Visser, N Rosenberg, R van den Berg-Vos, J H Wokke, E Bakker, and M de Visser. 2000. “Extension of the Clinical Range of Facioscapulohumeral Dystrophy: Report of Six Cases.” *Journal of Neurology, Neurosurgery, and Psychiatry* 69 (1): 114–16.

Kooi, E L van der, J C de Greef, M Wohlgemuth, R R Frants, R J G P van Asseldonk, H J Blom, B G M van Engelen, S M van der Maarel, and G W Padberg. 2006. “No Effect of Folic Acid and Methionine Supplementation on D4Z4 Methylation in Patients with Facioscapulohumeral Muscular Dystrophy.” *Neuromuscular Disorders : NMD* 16 (11): 766–69. https://doi.org/10.1016/j.nmd.2006.08.005.

Koonin, Eugene V., and Artem S. Novozhilov. 2009. “Origin and Evolution of the Genetic Code: The Universal Enigma.” *Iubmb Life* 61 (2): 99–111. https://doi.org/10.1002/iub.146.

Kowaljow, Valeria, Aline Marcowycz, Eugenie Ansseau, Cecilia B Conde, Sebastien Sauvage, Christel Matteotti, Cristina Arias, et al. 2007. “The DUX4 Gene at the FSHD1A Locus Encodes a Pro-Apoptotic Protein.” *Neuromuscular Disorders : NMD* 17 (8): 611–23. https://doi.org/10.1016/j.nmd.2007.04.002.

Krom, Yvonne D., Julie Dumonceaux, Kamel Mamchaoui, Bianca den Hamer, Virginie Mariot, Elisa Negroni, Linda N. Geng, et al. 2012. “Generation of Isogenic D4Z4 Contracted and Noncontracted Immortal Muscle Cell Clones from a Mosaic Patient: A Cellular Model for FSHD.” *The American Journal of Pathology* 181 (4): 1387–1401. https://doi.org/10.1016/j.ajpath.2012.07.007.

Kumari, Sunita, Anthony Bugaut, Julian L. Huppert, and Shankar Balasubramanian. 2007. “An RNA G-Quadruplex in the 5’ UTR of the NRAS Proto-Oncogene Modulates Translation.” *Nature Chemical Biology* 3 (4): 218–21. https://doi.org/10.1038/nchembio864.

Kuryavyi, Vitaly, Anh Tuân Phan, and Dinshaw J. Patel. 2010. “Solution Structures of All Parallel-Stranded Monomeric and Dimeric G-Quadruplex Scaffolds of the Human c-Kit2 Promoter.” *Nucleic Acids Research* 38 (19): 6757–73. https://doi.org/10.1093/nar/gkq558.

Kypr, Jaroslav, Iva Kejnovská, Daniel Renčiuk, and Michaela Vorlíčková. 2009. “Circular Dichroism and Conformational Polymorphism of DNA.” *Nucleic Acids Research* 37 (6): 1713–25. https://doi.org/10.1093/nar/gkp026.

Laforet, P, C de Toma, B Eymard, H M Becane, M Jeanpierre, M Fardeau, and D Duboc. 1998. “Cardiac Involvement in Genetically Confirmed Facioscapulohumeral Muscular Dystrophy.” *Neurology* 51 (5): 1454–56.

Landouzy and Dejerine, J. 1885. “De La Myopathie Atrophique Progressive.” *Rev Med Francaise* 5 (81).

Lane, Andrew N. 2012. “The Stability of Intramolecular DNA G-Quadruplexes Compared with Other Macromolecules.” *Biochimie* 94 (2): 277–86. https://doi.org/10.1016/j.biochi.2011.08.004.

Lange, Titia de. 2005. “Shelterin: The Protein Complex That Shapes and Safeguards Human Telomeres.” *Genes & Development* 19 (18): 2100–2110. https://doi.org/10.1101/gad.1346005.

Larson, Erik D, Michelle L Duquette, W Jason Cummings, Raphael J Streiff, and Nancy Maizels. 2005. “MutSalpha Binds to and Promotes Synapsis of Transcriptionally Activated Immunoglobulin Switch Regions.” *Current Biology : CB* 15 (5): 470–74. https://doi.org/10.1016/j.cub.2004.12.077.

Lattmann, Simon, Michael B Stadler, James P Vaughn, Steven A Akman, and Yoshikuni Nagamine. 2011. “The DEAH-Box RNA Helicase RHAU Binds an Intramolecular RNA G-Quadruplex in TERC and Associates with Telomerase Holoenzyme.” *Nucleic Acids Research* 39 (21): 9390–9404. https://doi.org/10.1093/nar/gkr630.

Lemmers, Richard J L F, Peggy de Kievit, Lodewijk Sandkuijl, George W Padberg, Gert-Jan B van Ommen, Rune R Frants, and Silvere M van der Maarel. 2002. “Facioscapulohumeral Muscular Dystrophy Is Uniquely Associated with One of the Two Variants of the 4q Subtelomere.” *Nature Genetics* 32 (2): 235–36. https://doi.org/10.1038/ng999.

Lemmers, Richard J L F, Rabi Tawil, Lisa M Petek, Judit Balog, Gregory J Block, Gijs W E Santen, Amanda M Amell, et al. 2012. “Digenic Inheritance of an SMCHD1 Mutation and an FSHD-Permissive D4Z4 Allele Causes Facioscapulohumeral Muscular Dystrophy Type 2.” *Nature Genetics* 44 (12): 1370–74. https://doi.org/10.1038/ng.2454.

Lemmers, Richard J L F, Patrick J van der Vliet, Rinse Klooster, Sabrina Sacconi, Pilar Camano, Johannes G Dauwerse, Lauren Snider, et al. 2010. “A Unifying Genetic Model for Facioscapulohumeral Muscular Dystrophy.” *Science (New York, N.Y.)* 329 (5999): 1650–53. https://doi.org/10.1126/science.1189044.

Lemmers, Richard JLF, Daniel G. Miller, and Silvere M. van der Maarel. 1993. “Facioscapulohumeral Muscular Dystrophy.” In *GeneReviews®*, edited by Margaret P. Adam, Holly H. Ardinger, Roberta A. Pagon, Stephanie E. Wallace, Lora JH Bean, Karen Stephens, and Anne Amemiya. Seattle (WA): University of Washington, Seattle. http://www.ncbi.nlm.nih.gov/books/NBK1443/.

Li, Jiansha, Lubing Gu, Hailong Zhang, Tao Liu, Dan Tian, Muxiang Zhou, and Sheng Zhou. 2013. “Berberine Represses DAXX Gene Transcription and Induces Cancer Cell Apoptosis.” *Laboratory Investigation; a Journal of Technical Methods and Pathology* 93 (3): 354–64. https://doi.org/10.1038/labinvest.2012.172.

Li, Qian-Jin, Xia-Jing Tong, Yi-Min Duan, and Jin-Qiu Zhou. 2013. “Characterization of the Intramolecular G-Quadruplex Promoting Activity of Est1.” *FEBS Letters* 587 (6): 659–65. https://doi.org/10.1016/j.febslet.2013.01.024.

Li, S. 1996. *The History of Traditional Chinese Medicine (in Chinese)*. Science Press, Beijing.

Li, Zi-Qi, Ting-Cong Liao, Cheng Dong, Jian-Wei Yang, Xiao-Jie Chen, Lihong Liu, Yuan Luo, Yuan-Yuan Liang, Wen-Hua Chen, and Chun-Qiong Zhou. 2017. “Specifically Targeting Mixed-Type Dimeric G-Quadruplexes Using Berberine Dimers.” *Organic & Biomolecular Chemistry* 15 (48): 10221–29. https://doi.org/10.1039/c7ob02326j.

Lian, Huang. 1986. *Chinese Materia Medica Dictionary.* Shanghai: Shanghai Scientific & Technical Publishers.

Lim, Jong-Won, Lauren Snider, Zizhen Yao, Rabi Tawil, Silvere M Van Der Maarel, Frank Rigo, C Frank Bennett, Galina N Filippova, and Stephen J Tapscott. 2015. “DICER/AGO-Dependent Epigenetic Silencing of D4Z4 Repeats Enhanced by Exogenous SiRNA Suggests Mechanisms and Therapies for FSHD.” *Human Molecular Genetics* 24 (17): 4817–28. https://doi.org/10.1093/hmg/ddv206.

Lipps, Hans J, and Daniela Rhodes. 2009. “G-Quadruplex Structures: In Vivo Evidence and Function.” *Trends in Cell Biology* 19 (8): 414–22. https://doi.org/10.1016/j.tcb.2009.05.002.

Liu, Shiguo, Yue Fang, Huiling Shen, Wenlin Xu, and Hao Li. 2013. “Berberine Sensitizes Ovarian Cancer Cells to Cisplatin through MiR-21/PDCD4 Axis.” *Acta Biochimica Et Biophysica Sinica* 45 (9): 756–62. https://doi.org/10.1093/abbs/gmt075.

Liu, Z, and W Gilbert. 1994. “The Yeast KEM1 Gene Encodes a Nuclease Specific for G4 Tetraplex DNA: Implication of in Vivo Functions for This Novel DNA Structure.” *Cell* 77 (7): 1083–92.

London, Timothy B C, Louise J Barber, Georgina Mosedale, Gavin P Kelly, Shankar Balasubramanian, Ian D Hickson, Simon J Boulton, and Kevin Hiom. 2008. “FANCJ Is a Structure-Specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts.” *The Journal of Biological Chemistry* 283 (52): 36132–39. https://doi.org/10.1074/jbc.M808152200.

Lopes, Judith, Aurele Piazza, Rodrigo Bermejo, Barry Kriegsman, Arianna Colosio, Marie-Paule Teulade-Fichou, Marco Foiani, and Alain Nicolas. 2011. “G-Quadruplex-Induced Instability during Leading-Strand Replication.” *The EMBO Journal* 30 (19): 4033–46. https://doi.org/10.1038/emboj.2011.316.

Lyle, R, T J Wright, L N Clark, and J E Hewitt. 1995. “The FSHD-Associated Repeat, D4Z4, Is a Member of a Dispersed Family of Homeobox-Containing Repeats, Subsets of Which Are Clustered on the Short Arms of the Acrocentric Chromosomes.” *Genomics* 28 (3): 389–97. https://doi.org/10.1006/geno.1995.1166.

Ma, Yan, Tian-Miao Ou, Jin-Qiang Hou, Yu-Jing Lu, Jia-Heng Tan, Lian-Quan Gu, and Zhi-Shu Huang. 2008. “9-N-Substituted Berberine Derivatives: Stabilization of G-Quadruplex DNA and down-Regulation of Oncogene c-Myc.” *Bioorganic & Medicinal Chemistry* 16 (16): 7582–91. https://doi.org/10.1016/j.bmc.2008.07.029.

Maddocks, Oliver D K, and Karen H Vousden. 2011. “Metabolic Regulation by P53.” *Journal of Molecular Medicine (Berlin, Germany)* 89 (3): 237–45. https://doi.org/10.1007/s00109-011-0735-5.

Maizels, Nancy. 2015. “G4-Associated Human Diseases.” *EMBO Reports* 16 (8): 910–22. https://doi.org/10.15252/embr.201540607.

Małgowska, Magdalena, Dorota Gudanis, Anna Teubert, Grażyna Dominiak, and Zofia Gdaniec. 2014. “REVIEW PAPER
How to Study G-Quadruplex Structures.” *BioTechnologia* 93 (4): 381–90. https://doi.org/10.5114/bta.2012.46592.

Mandel, C. R., Y. Bai, and L. Tong. 2008. “Protein Factors in Pre-MRNA 3’-End Processing.” *Cellular and Molecular Life Sciences: CMLS* 65 (7–8): 1099–1122. https://doi.org/10.1007/s00018-007-7474-3.

Marcel, Virginie, Phong L. T. Tran, Charlotte Sagne, Ghyslaine Martel-Planche, Laurence Vaslin, Marie-Paule Teulade-Fichou, Janet Hall, Jean-Louis Mergny, Pierre Hainaut, and Eric Van Dyck. 2011. “G-Quadruplex Structures in TP53 Intron 3: Role in Alternative Splicing and in Production of P53 MRNA Isoforms.” *Carcinogenesis* 32 (3): 271–78. https://doi.org/10.1093/carcin/bgq253.

Marsollier, Anne-Charlotte, Lukasz Ciszewski, Virginie Mariot, Linda Popplewell, Thomas Voit, George Dickson, and Julie Dumonceaux. 2016a. “Antisense Targeting of 3’ End Elements Involved in DUX4 MRNA Processing Is an Efficient Therapeutic Strategy for Facioscapulohumeral Dystrophy: A New Gene-Silencing Approach.” *Human Molecular Genetics* 25 (8): 1468–78. https://doi.org/10.1093/hmg/ddw015.

Meister, Gunter, and Thomas Tuschl. 2004. “Mechanisms of Gene Silencing by Double-Stranded RNA.” *Nature* 431 (7006): 343–49. https://doi.org/10.1038/nature02873.

Métifiot, Mathieu, Samir Amrane, Simon Litvak, and Marie-Line Andreola. 2014. “G-Quadruplexes in Viruses: Function and Potential Therapeutic Applications.” *Nucleic Acids Research* 42 (20): 12352–66. https://doi.org/10.1093/nar/gku999.

Mirihana Arachchilage, Gayan, Arosha C. Dassanayake, and Soumitra Basu. 2015. “A Potassium Ion-Dependent RNA Structural Switch Regulates Human Pre-MiRNA 92b Maturation.” *Chemistry & Biology* 22 (2): 262–72. https://doi.org/10.1016/j.chembiol.2014.12.013.

Mitsuhashi, Hiroaki, Satomi Mitsuhashi, Taylor Lynn-Jones, Genri Kawahara, and Louis M Kunkel. 2013. “Expression of DUX4 in Zebrafish Development Recapitulates Facioscapulohumeral Muscular Dystrophy.” *Human Molecular Genetics* 22 (3): 568–77. https://doi.org/10.1093/hmg/dds467.

Moraca, Federica, Jussara Amato, Francesco Ortuso, Anna Artese, Bruno Pagano, Ettore Novellino, Stefano Alcaro, Michele Parrinello, and Vittorio Limongelli. 2017. “Ligand Binding to Telomeric G-Quadruplex DNA Investigated by Funnel-Metadynamics Simulations.” *Proceedings of the National Academy of Sciences* 114 (11): E2136–45. https://doi.org/10.1073/pnas.1612627114.

Morosetti, Roberta, Massimiliano Mirabella, Carla Gliubizzi, Aldobrando Broccolini, Cristina Sancricca, Mario Pescatori, Teresa Gidaro, et al. 2007. “Isolation and Characterization of Mesoangioblasts from Facioscapulohumeral Muscular Dystrophy Muscle Biopsies.” *Stem Cells (Dayton, Ohio)* 25 (12): 3173–82. https://doi.org/10.1634/stemcells.2007-0465.

Morris, Mark J., Yoichi Negishi, Cathy Pazsint, Joseph D. Schonhoft, and Soumitra Basu. 2010. “An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES.” *Journal of the American Chemical Society* 132 (50): 17831–39. https://doi.org/10.1021/ja106287x.

Mullen, Melissa A., Sarah M. Assmann, and Philip C. Bevilacqua. 2012. “Toward a Digital Gene Response: RNA G-Quadruplexes with Fewer Quartets Fold with Higher Cooperativity.” *Journal of the American Chemical Society* 134 (2): 812–15. https://doi.org/10.1021/ja2096255.

Munsat, T L, D Piper, P Cancilla, and J Mednick. 1972. “Inflammatory Myopathy with Facioscapulohumeral Distribution.” *Neurology* 22 (4): 335–47.

Nakken, Sigve, Torbjørn Rognes, and Eivind Hovig. 2009. “The Disruptive Positions in Human G-Quadruplex Motifs Are Less Polymorphic and More Conserved than Their Neutral Counterparts.” *Nucleic Acids Research* 37 (17): 5749–56. https://doi.org/10.1093/nar/gkp590.

Nambiar, Mridula, G. Goldsmith, Balaji T. Moorthy, Michael R. Lieber, Mamata V. Joshi, Bibha Choudhary, Ramakrishna V. Hosur, and Sathees C. Raghavan. 2011. “Formation of a G-Quadruplex at the BCL2 Major Breakpoint Region of the t(14;18) Translocation in Follicular Lymphoma.” *Nucleic Acids Research* 39 (3): 936–48. https://doi.org/10.1093/nar/gkq824.

Nordhoff, E, F Kirpekar, and P Roepstorff. 1996. “Mass Spectrometry of Nucleic Acids.” *Mass Spectrometry Reviews* 15 (2): 67–138. https://doi.org/10.1002/(SICI)1098-2787(1996)15:2<67::AID-MAS1>3.0.CO;2-8.

Ohnmacht, Stephan A., Marialuisa Micco, Vanessa Petrucci, Alan K. Todd, Anthony P. Reszka, Mekala Gunaratnam, Marta A. Carvalho, Mire Zloh, and Stephen Neidle. 2012. “Sequences in the HSP90 Promoter Form G-Quadruplex Structures with Selectivity for Disubstituted Phenyl Bis-Oxazole Derivatives.” *Bioorganic & Medicinal Chemistry Letters* 22 (18): 5930–35. https://doi.org/10.1016/j.bmcl.2012.07.065.

Onel, Buket, Megan Carver, Guanhui Wu, Daria Timonina, Salil Kalarn, Marti Larriva, and Danzhou Yang. 2016. “A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription.” *Journal of the American Chemical Society* 138 (8): 2563–70. https://doi.org/10.1021/jacs.5b08596.

Padberg, G W. 1982. “Facioscapulohumeral Disease.” University of Leiden.
2004. “Facioscapulohumeral Muscular Dystrophy: A Clinician’s Experience.” In *Facioscapulohumeral Muscular Dystrophy: Clinical Medicine and Molecular Cell Biology.* Garland Science/BIOS Scientific Publishers, Oxon, United Kingdom.
Padberg, G W, O F Brouwer, R J de Keizer, G Dijkman, C Wijmenga, J J Grote, and R R Frants. 1995. “On the Significance of Retinal Vascular Disease and Hearing Loss in Facioscapulohumeral Muscular Dystrophy.” *Muscle & Nerve. Supplement* 2: S73-80.
Paeschke, Katrin, Matthew L Bochman, P Daniela Garcia, Petr Cejka, Katherine L Friedman, Stephen C Kowalczykowski, and Virginia A Zakian. 2013. “Pif1 Family Helicases Suppress Genome Instability at G-Quadruplex Motifs.” *Nature* 497 (7450): 458–62. https://doi.org/10.1038/nature12149.
Palumbo, SunMi L., Scot W. Ebbinghaus, and Laurence H. Hurley. 2009. “Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the HTERT Core Promoter with Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands.” *Journal of the American Chemical Society* 131 (31): 10878–91. https://doi.org/10.1021/ja902281d.
Pandey, Satyaprakash, Prachi Agarwala, and Souvik Maiti. 2013. “Effect of Loops and G-Quartets on the Stability of RNA G-Quadruplexes.” *The Journal of Physical Chemistry B* 117 (23): 6896–6905. https://doi.org/10.1021/jp401739m.
Paramasivan, Sattanathan, Iulian Rujan, and Philip H. Bolton. 2007. “Circular Dichroism of Quadruplex DNAs: Applications to Structure, Cation Effects and Ligand Binding.” *Methods*, Quadruplex DNA, 43 (4): 324–31. https://doi.org/10.1016/j.ymeth.2007.02.009.
Pelengaris, Stella, Mike Khan, and Gerard Evan. 2002. “C-MYC: More than Just a Matter of Life and Death.” *Nature Reviews. Cancer* 2 (10): 764–76. https://doi.org/10.1038/nrc904.
Phan, Anh Tuân, Vitaly Kuryavyi, Sarah Burge, Stephen Neidle, and Dinshaw J. Patel. 2007. “Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-Kit Promoter.” *Journal of the American Chemical Society* 129 (14): 4386–92. https://doi.org/10.1021/ja068739h.
Phatak, P, J C Cookson, F Dai, V Smith, R B Gartenhaus, M F G Stevens, and A M Burger. 2007. “Telomere Uncapping by the G-Quadruplex Ligand RHPS4 Inhibits Clonogenic Tumour Cell Growth in Vitro and in Vivo Consistent with a Cancer Stem Cell Targeting Mechanism.” *British Journal of Cancer* 96 (8): 1223–33. https://doi.org/10.1038/sj.bjc.6603691.
Podbevšek, Peter, and Janez Plavec. 2016. “KRAS Promoter Oligonucleotide with Decoy Activity Dimerizes into a Unique Topology Consisting of Two G-Quadruplex Units.” *Nucleic Acids Research* 44 (2): 917–25. https://doi.org/10.1093/nar/gkv1359.
Qin, Yong, and Laurence H. Hurley. 2008. “Structures, Folding Patterns, and Functions of Intramolecular DNA G-Quadruplexes Found in Eukaryotic Promoter Regions.” *Biochimie* 90 (8): 1149–71. https://doi.org/10.1016/j.biochi.2008.02.020.
Qin, Yong, Ji-Yan Pang, Wen-Hua Chen, Zhong-Zhen Zhao, Liang Liu, and Zhi-Hong Jiang. 2007. “Inhibition of DNA Topoisomerase I by Natural and Synthetic Mono- and Dimeric Protoberberine Alkaloids.” *Chemistry & Biodiversity* 4 (3): 481–87. https://doi.org/10.1002/cbdv.200790040.
Raiber, Eun-Ang, Ramon Kranaster, Enid Lam, Mehran Nikan, and Shankar Balasubramanian. 2012. “A Non-Canonical DNA Structure Is a Binding Motif for the Transcription Factor SP1 in Vitro.” *Nucleic Acids Research* 40 (4): 1499–1508. https://doi.org/10.1093/nar/gkr882.
Reddy, Kaalak, Bita Zamiri, Sabrina Y. R. Stanley, Robert B. Macgregor, and Christopher E. Pearson. 2013. “The Disease-Associated r(GGGGCC)n Repeat from the C9orf72 Gene Forms Tract Length-Dependent Uni- and Multimolecular RNA G-Quadruplex

Structures." *The Journal of Biological Chemistry* 288 (14): 9860–66. https://doi.org/10.1074/jbc.C113.452532.
Rhodes, D, and R Giraldo. 1995. "Telomere Structure and Function." *Current Opinion in Structural Biology* 5 (3): 311–22.
Rhodes, Daniela, and Hans J Lipps. 2015. "G-Quadruplexes and Their Regulatory Roles in Biology." *Nucleic Acids Research* 43 (18): 8627–37. https://doi.org/10.1093/nar/gkv862.
Ribeiro, Mariana Martins, Gleidson Silva Teixeira, Luciane Martins, Marcelo Rocha Marques, Ana Paula de Souza, and Sergio Roberto Peres Line. 2015. "G-Quadruplex Formation Enhances Splicing Efficiency of PAX9 Intron 1." *Human Genetics* 134 (1): 37–44. https://doi.org/10.1007/s00439-014-1485-6.
Ribeyre, Cyril, Judith Lopes, Jean-Baptiste Boule, Aurele Piazza, Aurore Guedin, Virginia A Zakian, Jean-Louis Mergny, and Alain Nicolas. 2009. "The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences in Vivo." *PLoS Genetics* 5 (5): e1000475. https://doi.org/10.1371/journal.pgen.1000475.
Richards, Mark, Frederique Coppee, Nick Thomas, Alexandra Belayew, and Meena Upadhyaya. 2012. "Facioscapulohumeral Muscular Dystrophy (FSHD): An Enigma Unravelled?" *Human Genetics* 131 (3): 325–40. https://doi.org/10.1007/s00439-011-1100-z.
Rodriguez, Raphael, Kyle M Miller, Josep V Forment, Charles R Bradshaw, Mehran Nikan, Sebastien Britton, Tobias Oelschlaegel, Blerta Xhemalce, Shankar Balasubramanian, and Stephen P Jackson. 2012. "Small-Molecule-Induced DNA Damage Identifies Alternative DNA Structures in Human Genes." *Nature Chemical Biology* 8 (3): 301–10. https://doi.org/10.1038/nchembio.780.
Rogers, Mark T, Fei Zhao, Peter S Harper, and Dafydd Stephens. 2002. "Absence of Hearing Impairment in Adult Onset Facioscapulohumeral Muscular Dystrophy." *Neuromuscular Disorders : NMD* 12 (4): 358–65.
Rouleau, Samuel G., Jean-Denis Beaudoin, Martin Bisaillon, and Jean-Pierre Perreault. 2015. "Small Antisense Oligonucleotides against G-Quadruplexes: Specific MRNA Translational Switches." *Nucleic Acids Research* 43 (1): 595–606. https://doi.org/10.1093/nar/gku1311.
Saccone, Valentina, and Pier Lorenzo Puri. 2010. "Epigenetic Regulation of Skeletal Myogenesis." *Organogenesis* 6 (1): 48–53.
Sacconi, Sabrina, Richard J L F Lemmers, Judit Balog, Patrick J van der Vliet, Pauline Lahaut, Merlijn P van Nieuwenhuizen, Kirsten R Straasheijm, et al. 2013. "The FSHD2 Gene SMCHD1 Is a Modifier of Disease Severity in Families Affected by FSHD1." *American Journal of Human Genetics* 93 (4): 744–51. https://doi.org/10.1016/j.ajhg.2013.08.004.
Safa, Layal, Emmanuelle Delagoutte, Irina Petruseva, Patrizia Alberti, Olga Lavrik, Jean-Francois Riou, and Carole Saintome. 2014. "Binding Polarity of RPA to Telomeric Sequences and Influence of G-Quadruplex Stability." *Biochimie* 103 (August): 80–88. https://doi.org/10.1016/j.biochi.2014.04.006.
Saharia, Abhishek, Lionel Guittat, Sandra Crocker, Adeline Lim, Martin Steffen, Shashikant Kulkarni, and Sheila A Stewart. 2008. "Flap Endonuclease 1 Contributes to Telomere Stability." *Current Biology : CB* 18 (7): 496–500. https://doi.org/10.1016/j.cub.2008.02.071.
Salazar, Andrew, Michael Keusgen, and Jörg von Hagen. 2016. "Amino Acids in the Cultivation of Mammalian Cells." *Amino Acids* 48: 1161–71. https://doi.org/10.1007/s00726-016-2181-8.
Salvati, Erica, Pasquale Zizza, Angela Rizzo, Sara Iachettini, Chiara Cingolani, Carmen D'Angelo, Manuela Porru, et al. 2014. "Evidence for G-Quadruplex in the Promoter of Vegfr-2 and Its Targeting to Inhibit Tumor Angiogenesis." *Nucleic Acids Research* 42 (5): 2945–57. https://doi.org/10.1093/nar/gkt1289.

Sandri, M, A H El Meslemani, C Sandri, P Schjerling, K Vissing, J L Andersen, K Rossini, U Carraro, and C Angelini. 2001. “Caspase 3 Expression Correlates with Skeletal Muscle Apoptosis in Duchenne and Facioscapulo Human Muscular Dystrophy. A Potential Target for Pharmacological Treatment?” *Journal of Neuropathology and Experimental Neurology* 60 (3): 302–12.

Schaffitzel, C, I Berger, J Postberg, J Hanes, H J Lipps, and A Pluckthun. 2001. “In Vitro Generated Antibodies Specific for Telomeric Guanine-Quadruplex DNA React with Stylonychia Lemnae Macronuclei.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (15): 8572–77. https://doi.org/10.1073/pnas.141229498.

Schiavone, Davide, Guillaume Guilbaud, Pierre Murat, Charikleia Papadopoulou, Peter Sarkies, Marie-Noelle Prioleau, Shankar Balasubramanian, and Julian E Sale. 2014. “Determinants of G Quadruplex-Induced Epigenetic Instability in REV1-Deficient Cells.” *The EMBO Journal* 33 (21): 2507–20. https://doi.org/10.15252/embj.201488398.

Schwarzkopf, Martina, Dario Coletti, David Sassoon, and Giovanna Marazzi. 2006. “Muscle Cachexia Is Regulated by a P53-PW1/Peg3-Dependent Pathway.” *Genes & Development* 20 (24): 3440–52. https://doi.org/10.1101/gad.412606.

Seenisamy, Jeyaprakashnarayanan, Evonne M Rezler, Tiffanie J Powell, Denise Tye, Vijay Gokhale, Chandana Sharma Joshi, Adam Siddiqui-Jain, and Laurence H Hurley. 2004. “The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4.” *Journal of the American Chemical Society* 126 (28): 8702–9. https://doi.org/10.1021/ja040022b.

Sellers, William R., and Massimo Loda. 2002. “The EZH2 Polycomb Transcriptional Repressor--a Marker or Mover of Metastatic Prostate Cancer?” *Cancer Cell* 2 (5): 349–50.

Sen, D, and W Gilbert. 1990. “A Sodium-Potassium Switch in the Formation of Four-Stranded G4-DNA.” *Nature* 344 (6265): 410–14. https://doi.org/10.1038/344410a0.

Sharma, Vishakha, Sachchida Nand Pandey, Hunain Khawaja, Kristy J Brown, Yetrib Hathout, and Yi-Wen Chen. 2016. “PARP1 Differentially Interacts with Promoter Region of DUX4 Gene in FSHD Myoblasts.” *Journal of Genetic Syndromes & Gene Therapy* 7 (4). https://doi.org/10.4172/2157-7412.1000303.

Shi, Yongsheng, and James L. Manley. 2015. “The End of the Message: Multiple Protein–RNA Interactions Define the MRNA Polyadenylation Site.” *Genes & Development* 29 (9): 889–97. https://doi.org/10.1101/gad.261974.115.

Shivalingam, Arun, M Angeles Izquierdo, Alix Le Marois, Aurimas Vysniauskas, Klaus Suhling, Marina K Kuimova, and Ramon Vilar. 2015. “The Interactions between a Small Molecule and G-Quadruplexes Are Visualized by Fluorescence Lifetime Imaging Microscopy.” *Nature Communications* 6 (September): 8178. https://doi.org/10.1038/ncomms9178.

Shklover, Jeny, Pnina Weisman-Shomer, Anat Yafe, and Michael Fry. 2010. “Quadruplex Structures of Muscle Gene Promoter Sequences Enhance in Vivo MyoD-Dependent Gene Expression.” *Nucleic Acids Research* 38 (7): 2369–77. https://doi.org/10.1093/nar/gkp1208.

Siddiqui-Jain, Adam, Cory L. Grand, David J. Bearss, and Laurence H. Hurley. 2002. “Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (18): 11593–98. https://doi.org/10.1073/pnas.182256799.

Simone, Roberto, Pietro Fratta, Stephen Neidle, Gary N. Parkinson, and Adrian M. Isaacs. 2015. “G-Quadruplexes: Emerging Roles in Neurodegenerative Diseases and the Non-Coding Transcriptome.” *FEBS Letters* 589 (14): 1653–68. https://doi.org/10.1016/j.febslet.2015.05.003.

Simonsson, T., P. Pecinka, and M. Kubista. 1998. “DNA Tetraplex Formation in the Control Region of C-Myc.” *Nucleic Acids Research* 26 (5): 1167–72.

Singleton, Martin R., Mark S. Dillingham, and Dale B. Wigley. 2007. “Structure and Mechanism of Helicases and Nucleic Acid Translocases.” *Annual Review of Biochemistry* 76: 23–50. https://doi.org/10.1146/annurev.biochem.76.052305.115300.

Skourti-Stathaki, Konstantina, Nicholas J. Proudfoot, and Natalia Gromak. 2011. “Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination.” *Molecular Cell* 42 (6): 794–805. https://doi.org/10.1016/j.molcel.2011.04.026.

Snider, Lauren, Amy Asawachaicharn, Ashlee E Tyler, Linda N Geng, Lisa M Petek, Lisa Maves, Daniel G Miller, et al. 2009a. “RNA Transcripts, MiRNA-Sized Fragments and Proteins Produced from D4Z4 Units: New Candidates for the Pathophysiology of Facioscapulohumeral Dystrophy.” *Human Molecular Genetics* 18 (13): 2414–30. https://doi.org/10.1093/hmg/ddp180.

Nelson, Rabi Tawil, et al. 2010a. “Facioscapulohumeral Dystrophy: Incomplete Suppression of a Retrotransposed Gene.” *PLoS Genetics* 6 (10): e1001181. https://doi.org/10.1371/journal.pgen.1001181.

Soares, O. D., and J. L. Costa. 1999. “Spectrophotometer Spectral Bandwidth Calibration with Absorption Bands Crystal Standard.” *Applied Optics* 38 (10): 2007–13.

Soldatenkov, Viatcheslav A., Sergey Chasovskikh, Vladimir N. Potaman, Irina Trofimova, Mark E. Smulson, and Anatoly Dritschilo. 2002. “Transcriptional Repression by Binding of Poly(ADP-Ribose) Polymerase to Promoter Sequences.” *Journal of Biological Chemistry* 277 (1): 665–70. https://doi.org/10.1074/jbc.M108551200.

Su, Zhaoming, Yongjie Zhang, Tania F. Gendron, Peter O. Bauer, Jeannie Chew, Wang-Yong Yang, Erik Fostvedt, et al. 2014. “Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in C9FTD/ALS.” *Neuron* 83 (5): 1043–50. https://doi.org/10.1016/j.neuron.2014.07.041.

Sun, D., B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S. Neidle, and L. H. Hurley. 1997. “Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound.” *Journal of Medicinal Chemistry* 40 (14): 2113–16. https://doi.org/10.1021/jm970199z.

Sun, Daekyu, Kexiao Guo, and Yoon-Joo Shin. 2011. “Evidence of the Formation of G-Quadruplex Structures in the Promoter Region of the Human Vascular Endothelial Growth Factor Gene.” *Nucleic Acids Research* 39 (4): 1256–65. https://doi.org/10.1093/nar/gkq926.

Sun, Daekyu, and Laurence H. Hurley. 2009. “The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression.” *Journal of Medicinal Chemistry* 52 (9): 2863–74. https://doi.org/10.1021/jm900055s.

Sun, Daekyu, Wei-Jun Liu, Kexiao Guo, Jadrian J. Rusche, Scot Ebbinghaus, Vijay Gokhale, and Laurence H. Hurley. 2008. “The Proximal Promoter Region of the Human Vascular Endothelial Growth Factor Gene Has a G-Quadruplex Structure That Can Be Targeted by G-Quadruplex-Interactive Agents.” *Molecular Cancer Therapeutics* 7 (4): 880–89. https://doi.org/10.1158/1535-7163.MCT-07-2119.

Sun, Hongxia, Junfeng Xiang, Yunhua Shi, Qianfan Yang, Aijiao Guan, Qian Li, Lijia Yu, et al. 2014. “A Newly Identified G-Quadruplex as a Potential Target Regulating Bcl-2 Expression.” *Biochimica et Biophysica Acta (BBA) - General Subjects* 1840 (10): 3052–57. https://doi.org/10.1016/j.bbagen.2014.07.014.

Takahama, Kentaro, Asami Takada, Shota Tada, Mai Shimizu, Kazutoshi Sayama, Riki Kurokawa, and Takanori Oyoshi. 2013. “Regulation of Telomere Length by G-Quadruplex

Telomere DNA- and TERRA-Binding Protein TLS/FUS." *Chemistry & Biology* 20 (3): 341–50. https://doi.org/10.1016/j.chembiol.2013.02.013.

Tang, Qing-Lin, Min-Ling Lai, Yuan-Fu Zhong, Ai-Min Wang, Jun-Kai Su, and Ming-Qing Zhang. 2013. "Antinociceptive Effect of Berberine on Visceral Hypersensitivity in Rats." *World Journal of Gastroenterology* 19 (28): 4582–89. https://doi.org/10.3748/wjg.v19.i28.4582.

Tassin, Alexandra, Dalila Laoudj-Chenivesse, Celine Vanderplanck, Marietta Barro, Sebastien Charron, Eugenie Ansseau, Yi-Wen Chen, Jacques Mercier, Frederique Coppee, and Alexandra Belayew. 2013a. "DUX4 Expression in FSHD Muscle Cells: How Could Such a Rare Protein Cause a Myopathy?" *Journal of Cellular and Molecular Medicine* 17 (1): 76–89. https://doi.org/10.1111/j.1582-4934.2012.01647.x.

Tassin, Alexandra, Baptiste Leroy, Dalila Laoudj-Chenivesse, Armelle Wauters, Céline Vanderplanck, Marie-Catherine Le Bihan, Frédérique Coppée, Ruddy Wattiez, and Alexandra Belayew. 2012. "FSHD Myotubes with Different Phenotypes Exhibit Distinct Proteomes." *PloS One* 7 (12): e51865. https://doi.org/10.1371/journal.pone.0051865.

Tauchi, T., K. Shin-ya, G. Sashida, M. Sumi, S. Okabe, J. H. Ohyashiki, and K. Ohyashiki. 2006. "Telomerase Inhibition with a Novel G-Quadruplex-Interactive Agent, Telomestatin: In Vitro and in Vivo Studies in Acute Leukemia." *Oncogene* 25 (42): 5719–25. https://doi.org/10.1038/sj.onc.1209577.

Tawil, Rabi. 2008. "Facioscapulohumeral Muscular Dystrophy." *Neurotherapeutics : The Journal of the American Society for Experimental NeuroTherapeutics* 5 (4): 601–6. https://doi.org/10.1016/j.nurt.2008.07.005.

Tawil, Rabi, Silvere M van der Maarel, and Stephen J Tapscott. 2014a. "Facioscapulohumeral Dystrophy: The Path to Consensus on Pathophysiology." *Skeletal Muscle* 4: 12. https://doi.org/10.1186/2044-5040-4-12.

Teiten, Marie-Hélène, François Gaascht, Mario Dicato, and Marc Diederich. 2013. "Anticancer Bioactivity of Compounds from Medicinal Plants Used in European Medieval Traditions." *Biochemical Pharmacology* 86 (9): 1239–47. https://doi.org/10.1016/j.bcp.2013.08.007.

Thandapani, Palaniraja, Jingwen Song, Valentina Gandin, Yutian Cai, Samuel G. Rouleau, Jean-Michel Garant, Francois-Michel Boisvert, et al. 2015. "Aven Recognition of RNA G-Quadruplexes Regulates Translation of the Mixed Lineage Leukemia Protooncogenes." *ELife* 4 (August). https://doi.org/10.7554/eLife.06234.

Thorley, Matthew, Stéphanie Duguez, Emilia Maria Cristina Mazza, Sara Valsoni, Anne Bigot, Kamel Mamchaoui, Brennan Harmon, Thomas Voit, Vincent Mouly, and William Duddy. 2016. "Skeletal Muscle Characteristics Are Preserved in HTERT/Cdk4 Human Myogenic Cell Lines." *Skeletal Muscle* 6 (December). https://doi.org/10.1186/s13395-016-0115-5.

Tillhon, Micol, Luis M. Guamán Ortiz, Paolo Lombardi, and A. Ivana Scovassi. 2012. "Berberine: New Perspectives for Old Remedies." *Biochemical Pharmacology* 84 (10): 1260–67. https://doi.org/10.1016/j.bcp.2012.07.018.

Todd, Alan K., Matthew Johnston, and Stephen Neidle. 2005. "Highly Prevalent Putative Quadruplex Sequence Motifs in Human DNA." *Nucleic Acids Research* 33 (9): 2901–7. https://doi.org/10.1093/nar/gki553.

Todd, Alan K., and Stephen Neidle. 2011. "Mapping the Sequences of Potential Guanine Quadruplex Motifs." *Nucleic Acids Research* 39 (12): 4917–27. https://doi.org/10.1093/nar/gkr104.

Tong, Xiaotian, Wenxian Lan, Xu Zhang, Houming Wu, Maili Liu, and Chunyang Cao. 2011. "Solution Structure of All Parallel G-Quadruplex Formed by the Oncogene RET Promoter Sequence." *Nucleic Acids Research* 39 (15): 6753–63. https://doi.org/10.1093/nar/gkr233.

Trofimova, Irina, and Alla Krasikova. 2016. “Transcription of Highly Repetitive Tandemly Organized DNA in Amphibians and Birds: A Historical Overview and Modern Concepts.” *RNA Biology* 13 (12): 1246–57. https://doi.org/10.1080/15476286.2016.1240142.

Tsumagari, Koji, Lixin Qi, Kesmic Jackson, Chunbo Shao, Michelle Lacey, Janet Sowden, Rabi Tawil, Vettaikorumakankav Vedanarayanan, and Melanie Ehrlich. 2008. “Epigenetics of a Tandem DNA Repeat: Chromatin DNaseI Sensitivity and Opposite Methylation Changes in Cancers.” *Nucleic Acids Research* 36 (7): 2196–2207. https://doi.org/10.1093/nar/gkn055.

Tupler, R, A Berardinelli, L Barbierato, R Frants, J E Hewitt, G Lanzi, P Maraschio, and L Tiepolo. 1996. “Monosomy of Distal 4q Does Not Cause Facioscapulohumeral Muscular Dystrophy.” *Journal of Medical Genetics* 33 (5): 366–70.

Vallur, Aarthy C, and Nancy Maizels. 2008. “Activities of Human Exonuclease 1 That Promote Cleavage of Transcribed Immunoglobulin Switch Regions.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (43): 16508–12. https://doi.org/10.1073/pnas.0805327105.

Vanderplanck, Celine, Eugenie Ansseau, Sebastien Charron, Nadia Stricwant, Alexandra Tassin, Dalila Laoudj-Chenivesse, Steve D Wilton, Frederique Coppee, and Alexandra Belayew. 2011. “The FSHD Atrophic Myotube Phenotype Is Caused by DUX4 Expression.” *PloS One* 6 (10): e26820. https://doi.org/10.1371/journal.pone.0026820.

Vannier, Jean-Baptiste, Visnja Pavicic-Kaltenbrunner, Mark I R Petalcorin, Hao Ding, and Simon J Boulton. 2012. “RTEL1 Dismantles T Loops and Counteracts Telomeric G4-DNA to Maintain Telomere Integrity.” *Cell* 149 (4): 795–806. https://doi.org/10.1016/j.cell.2012.03.030.

Vilquin, J-T, J-P Marolleau, S Sacconi, I Garcin, M-N Lacassagne, I Robert, B Ternaux, B Bouazza, J Larghero, and C Desnuelle. 2005. “Normal Growth and Regenerating Ability of Myoblasts from Unaffected Muscles of Facioscapulohumeral Muscular Dystrophy Patients.” *Gene Therapy* 12 (22): 1651–62. https://doi.org/10.1038/sj.gt.3302565.

Wallace, Lindsay M, Sara E Garwick, Wenyan Mei, Alexandra Belayew, Frederique Coppee, Katherine J Ladner, Denis Guttridge, Jing Yang, and Scott Q Harper. 2011. “DUX4, a Candidate Gene for Facioscapulohumeral Muscular Dystrophy, Causes P53-Dependent Myopathy in Vivo.” *Annals of Neurology* 69 (3): 540–52. https://doi.org/10.1002/ana.22275.

Wang, Jiao, and Yang-Fu Jiang. 2012. “Natural Compounds as Anticancer Agents: Experimental Evidence.” *World Journal of Experimental Medicine* 2 (3): 45–57. https://doi.org/10.5493/wjem.v2.i3.45.

Wang, Jiou, Aaron R. Haeusler, and Eric A. J. Simko. 2015. “Emerging Role of RNA•DNA Hybrids in C9orf72-Linked Neurodegeneration.” *Cell Cycle (Georgetown, Tex.)* 14 (4): 526–32. https://doi.org/10.1080/15384101.2014.995490.

Wang, Peng, and Zhen Chen. 2013. “Traditional Chinese Medicine ZHENG and Omics Convergence: A Systems Approach to Post-Genomics Medicine in a Global World.” *Omics: A Journal of Integrative Biology* 17 (9): 451–59. https://doi.org/10.1089/omi.2012.0057.

Wang, Xueyin, Karen J. Goodrich, Anne R. Gooding, Haroon Naeem, Stuart Archer, Richard D. Paucek, Daniel T. Youmans, Thomas R. Cech, and Chen Davidovich. 2017. “Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines.” *Molecular Cell* 65 (6): 1056-1067.e5. https://doi.org/10.1016/j.molcel.2017.02.003.

Wanrooij, Paulina H., Jay P. Uhler, Yonghong Shi, Fredrik Westerlund, Maria Falkenberg, and Claes M. Gustafsson. 2012. “A Hybrid G-Quadruplex Structure Formed between RNA and DNA Explains the Extraordinary Stability of the Mitochondrial R-Loop.” *Nucleic Acids Research* 40 (20): 10334–44. https://doi.org/10.1093/nar/gks802.

WATSON, J D, and F H CRICK. 1953. “Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid.” *Nature* 171 (4356): 737–38.

Webba da Silva, Mateus. 2007. “NMR Methods for Studying Quadruplex Nucleic Acids.” *Methods (San Diego, Calif.)* 43 (4): 264–77. https://doi.org/10.1016/j.ymeth.2007.05.007.

Wei, Chunying, Guoqing Jia, Jingli Yuan, Zhaochi Feng, and Can Li. 2006. “A Spectroscopic Study on the Interactions of Porphyrin with G-Quadruplex DNAs.” *Biochemistry* 45 (21): 6681–91. https://doi.org/10.1021/bi052356z.

Wendt, Hans, Richard M. Thomas, and Tom Ellenberger. 1998. “DNA-Mediated Folding and Assembly of MyoD-E47 Heterodimers.” *Journal of Biological Chemistry* 273 (10): 5735–43. https://doi.org/10.1074/jbc.273.10.5735.

Westmark, Cara J., and James S. Malter. 2007. “FMRP Mediates MGluR5-Dependent Translation of Amyloid Precursor Protein.” *PLoS Biology* 5 (3): e52. https://doi.org/10.1371/journal.pbio.0050052.

White, Elizabeth W., Farial Tanious, Mohamed A. Ismail, Anthony P. Reszka, Stephen Neidle, David W. Boykin, and W. David Wilson. 2007. “Structure-Specific Recognition of Quadruplex DNA by Organic Cations: Influence of Shape, Substituents and Charge.” *Biophysical Chemistry* 126 (1–3): 140–53. https://doi.org/10.1016/j.bpc.2006.06.006.

Whitehouse, Iestyn, and Tom Owen-Hughes. 2010. “ATRX: Put Me on Repeat.” *Cell* 143 (3): 335–36. https://doi.org/10.1016/j.cell.2010.10.021.

Wijmenga, C, J E Hewitt, L A Sandkuijl, L N Clark, T J Wright, H G Dauwerse, A M Gruter, M H Hofker, P Moerer, and R Williamson. 1992. “Chromosome 4q DNA Rearrangements Associated with Facioscapulohumeral Muscular Dystrophy.” *Nature Genetics* 2 (1): 26–30. https://doi.org/10.1038/ng0992-26.

Winston, Jincy, Laura Duerden, Matthew Mort, Ian M Frayling, Mark T Rogers, and Meena Upadhyaya. 2015. “Identification of Two Novel SMCHD1 Sequence Variants in Families with FSHD-like Muscular Dystrophy.” *European Journal of Human Genetics : EJHG* 23 (1): 67–71. https://doi.org/10.1038/ejhg.2014.58.

Yang, Danzhou, and Keika Okamoto. 2010. “Structural Insights into G-Quadruplexes: Towards New Anticancer Drugs.” *Future Medicinal Chemistry* 2 (4): 619–46.

Yao, Zizhen, Lauren Snider, Judit Balog, Richard J L F Lemmers, Silvere M Van Der Maarel, Rabi Tawil, and Stephen J Tapscott. 2014. “DUX4-Induced Gene Expression Is the Major Molecular Signature in FSHD Skeletal Muscle.” *Human Molecular Genetics* 23 (20): 5342–52. https://doi.org/10.1093/hmg/ddu251.

Yip, Darren J, and David J Picketts. 2003. “Increasing D4Z4 Repeat Copy Number Compromises C2C12 Myoblast Differentiation.” *FEBS Letters* 537 (1–3): 133–38.

Zatz, M., S. K. Marie, A. Cerqueira, M. Vainzof, R. C. Pavanello, and M. R. Passos-Bueno. 1998. “The Facioscapulohumeral Muscular Dystrophy (FSHD1) Gene Affects Males More Severely and More Frequently than Females.” *American Journal of Medical Genetics* 77 (2): 155–61.

Zeng, Weihua, Jessica C de Greef, Yen-Yun Chen, Richard Chien, Xiangduo Kong, Heather C Gregson, Sara T Winokur, et al. 2009. “Specific Loss of Histone H3 Lysine 9 Trimethylation and HP1gamma/Cohesin Binding at D4Z4 Repeats Is Associated with Facioscapulohumeral Dystrophy (FSHD).” *PLoS Genetics* 5 (7): e1000559. https://doi.org/10.1371/journal.pgen.1000559.

Zeraati, Mahdi, David B. Langley, Peter Schofield, Aaron L. Moye, Romain Rouet, William E. Hughes, Tracy M. Bryan, Marcel E. Dinger, and Daniel Christ. 2018. “I-Motif DNA Structures Are Formed in the Nuclei of Human Cells.” *Nature Chemistry*, April, 1. https://doi.org/10.1038/s41557-018-0046-3.

Zhang, Dong-Hao, Takeshi Fujimoto, Sarika Saxena, Hai-Qing Yu, Daisuke Miyoshi, and Naoki Sugimoto. 2010. “Monomorphic RNA G-Quadruplex and Polymorphic DNA G-

Quadruplex Structures Responding to Cellular Environmental Factors." *Biochemistry* 49 (21): 4554–63. https://doi.org/10.1021/bi1002822.

Zhang, Guoqiang, Pierre-Olivier Estève, Hang Gyeong Chin, Jolyon Terragni, Nan Dai, Ivan R. Corrêa, and Sriharsa Pradhan. 2015. "Small RNA-Mediated DNA (Cytosine-5) Methyltransferase 1 Inhibition Leads to Aberrant DNA Methylation." *Nucleic Acids Research* 43 (12): 6112–24. https://doi.org/10.1093/nar/gkv518.

Zhang, Jia-yu, Ke-wei Zheng, Shan Xiao, Yu-hua Hao, and Zheng Tan. 2014. "Mechanism and Manipulation of DNA:RNA Hybrid G-Quadruplex Formation in Transcription of G-Rich DNA." *Journal of the American Chemical Society* 136 (4): 1381–90. https://doi.org/10.1021/ja4085572.

Zhang, Ming, Chun-Mei Wang, Jing Li, Zhao-Jie Meng, Sheng-Nan Wei, Ji Li, Richard Bucala, Yu-Lin Li, and Li Chen. 2013. "Berberine Protects against Palmitate-Induced Endothelial Dysfunction: Involvements of Upregulation of AMPK and ENOS and Downregulation of NOX4." *Mediators of Inflammation* 2013: 260464. https://doi.org/10.1155/2013/260464.

Zhang, Wan-Jin, Tian-Miao Ou, Yu-Jing Lu, Ying-Yu Huang, Wei-Bin Wu, Zhi-Shu Huang, Jin-Lin Zhou, Kwok-Yin Wong, and Lian-Quan Gu. 2007. "9-Substituted Berberine Derivatives as G-Quadruplex Stabilizing Ligands in Telomeric DNA." *Bioorganic & Medicinal Chemistry* 15 (16): 5493–5501. https://doi.org/10.1016/j.bmc.2007.05.050.

Zhang, Xiaoling, Lubing Gu, Jiansha Li, Noopur Shah, Jing He, Lin Yang, Qun Hu, and Muxiang Zhou. 2010. "Degradation of MDM2 by the Interaction between Berberine and DAXX Leads to Potent Apoptosis in MDM2-Overexpressing Cancer Cells." *Cancer Research* 70 (23): 9895–9904. https://doi.org/10.1158/0008-5472.CAN-10-1546.

Zheng, Ke-wei, Zhao Chen, Yu-hua Hao, and Zheng Tan. 2010. "Molecular Crowding Creates an Essential Environment for the Formation of Stable G-Quadruplexes in Long Double-Stranded DNA." *Nucleic Acids Research* 38 (1): 327–38. https://doi.org/10.1093/nar/gkp898.

Zheng, Ke-wei, Shan Xiao, Jia-quan Liu, Jia-yu Zhang, Yu-hua Hao, and Zheng Tan. 2013. "Co-Transcriptional Formation of DNA:RNA Hybrid G-Quadruplex and Potential Function as Constitutional Cis Element for Transcription Control." *Nucleic Acids Research* 41 (10): 5533–41. https://doi.org/10.1093/nar/gkt264.

Zheng, Simin, Bao Q. Vuong, Bharat Vaidyanathan, Jia-Yu Lin, Feng-Ting Huang, and Jayanta Chaudhuri. 2015. "Non-Coding RNA Generated Following Lariat Debranching Mediates Targeting of AID to DNA." *Cell* 161 (4): 762–73. https://doi.org/10.1016/j.cell.2015.03.020.

Zidanloo, Saeedeh Ghazaey, Abasalt Hosseinzadeh Colagar, Hossein Ayatollahi, and Jahan-Bakhsh Raoof. 2016. "Downregulation of the WT1 Gene Expression via TMPyP4 Stabilization of Promoter G-Quadruplexes in Leukemia Cells." *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine* 37 (7): 9967–77. https://doi.org/10.1007/s13277-016-4881-9.

APPENDIX: Scientific publications

Research Article: Antisense targeting of 3'end elements involved in DUX4 mRNA processing is an efficient therapeutic strategy for Facioscapulohumeral Dystrophy: a new gene silencing approach

*Anne-Charlotte Marsollier, **Lukasz Ciszewski**, Virginie Mariot, Linda Popplewell, Thomas Voit, George Dickson, Julie Dumonceaux.*

Hum Mol Genet. 2016 Apr 15;25(8):1468-78

ABSTRACT:

Defects in mRNA 3' end formation have been described to alter transcription termination, transport of the mRNA from the nucleus to the cytoplasm, stability of the mRNA and translation efficiency. Therefore, inhibition of polyadenylation may lead to gene silencing. Here, we choose Facioscapulohumeral Dystrophy (FSHD) as a model to determine whether or not targeting key 3'end elements involved in mRNA processing using antisense oligonucleotide drugs can be used as a strategy for gene silencing within a potentially therapeutic context. FSHD is a gain-of-function disease characterized by the aberrant expression of the DUX4 transcription factor leading to altered pathogenic deregulation of multiple genes in muscles. Here we demonstrate that targeting either the mRNA polyadenylation signal and/ or cleavage site is an efficient strategy to downregulate DUX4 expression and to decrease the abnormally high pathological expression of genes downstream of DUX4. We conclude that targeting key functional 3'end elements involved in pre-mRNA to mRNA maturation with antisense drugs can lead to efficient gene silencing and is thus a potentially effective therapeutic strategy for at least FSHD. More over polyadenylation is a crucial step in the maturation of almost all eukaryotic mRNAs, and thus all mRNAs are virtually eligible for this antisense-mediated knockdown strategy.

Book Chapter: Antisense Oligonucleotide Targeting of 3'-UTR of mRNA for Expression Knockdown

Golnoush Golshirazi, Lukasz Ciszewski, Ngoc Lu-Nguyen, Linda Popplewell

'Exon skipping and Inclusion Therapies' Methods Molecular Biology, Vol 1828, ISBN:978-1-4939-8650-7

Abstract:

With the recent conditional approval of an antisense oligonucleotide (AON) that restores the reading frame of DMD transcript in a subset of Duchenne muscular dystrophy patients, it has been established that AONs sharing similar chemistry have clear clinical potential. Genetic diseases, such as facioscapulohumeral dystrophy (FSHD), can be the result of gain-of-function mutations. Since mRNA processing in terms of termination of transcription, its transport from the nucleus to the cytoplasm, its stability and translation efficiency are dependent on key 3'UTR elements, it follows that targeting these elements with AONs have the potential to induce gene silencing. Aberrant expression of the Double homeobox 4 (DUX4) transcription factor and the downstream consequences of such expression is the hall-mark of FSHD. Here we describe the bioinformatic strategies behind the design of AONs targeting polyadenylation signals and the methodologies relevant to their *in vitro* screening for efficacy and safety, including analysis of expression at the transcript and protein level of the specific target and downstream genes, and measurement of the effect on the fusion index of myotubes. The targeting of permissive DUX4 and MSTN are used as examples. MSTN encodes for myostatin, a negative regulator of myogenesis; the downregulation of MSTN expression has the potential to address the muscular atrophy associated with muscular dystrophies, sarcopenia, cancer and acquired immunodeficiency syndrome.