UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Department of Molecular Medicine

Ph.D. COURSE IN: Molecular Medicine

CURRICULUM: Biomedicine

SERIES XXXI

INVESTIGATION OF THE ROLE OF DNA G-QUADRUPLEX STRUCTURES IN HUMAN TRANSCRIPTION, CANCER AND VIRAL INFECTIONS

Thesis written with the financial contribution of DiaSorin S.p.A.

Coordinator: Prof. Stefano Piccolo

Supervisor: Prof. Sara Richter

Ph.D. Student: Sara Lago

"Assemble a hundred thoughts and discard ninety."

Dale Carnegie

"Science only proceeds after a long series of failures [..]."

Matteo Nadai

CONTENTS

CONTENTS 1
ABSTRACT 1
RIASSUNTO 3
1 INTRODUCTION 5
1.1 DNA secondary structures 5
1.2 G-quadruplex biophysical properties 6
1.3 G-quadruplexes biological functions 8
1.3.1 G-quadruplexes in Human 9
1.3.1.1 Transcription 10
1.3.1.2 Involvement in human diseases 14
Cancer and soft tissue sarcoma 15
1.3.2 G-quadruplexes in viruses 17
1.3.2.1 HIV-1 17
1.3.2.2 HSV-1 19
1.4 G-quadruplexes as therapeutic target and tool 21
1.5 Main challenges in the *in vivo* investigation of G-quadruplexes 23
2 AIM 25
3 MATERIALS AND METHODS 26
3.1 Oligonucleotides and compounds 26
3.2 QGRS prediction of putative G4s 26
3.3 Circular Dichroism Spectroscopy (CD) 26
3.4 Dimethylsulphate footprinting (DMS) 27
3.5 Nuclear proteins extraction, pull-down and Mass-Spectroscopy (MS) 28
3.6 Purification of BG4 antibody 29
3.7 G4 Chromatin Immunoprecipitation and library preparation 30
3.8 G-quadruplex ChIP–qPCR 31
3.9 RNA extraction and cDNA library preparation 31
3.10 Data analysis 31
3.11 Cytotoxicity assay 32
3.12 Electrophoretic Mobility Shift Assay 32

3.13 Surface Plasmon Resonance (SPR) 33
3.13.1 SPR for the measurement of AS1411 affinity 34
3.13.2 SPR for the optimization of 1H6 antibody immobilization 34
Covalent immobilization through Standard Amine Coupling of 1H6 antibody 34
Mouse Antibody Capture kit mediated immobilization of 1H6 antibody 35
G-quadruplex complex-mediated immobilization of 1H6 antibody 35
Kinetic analysis and data evaluation 35
3.13.3 SPR for the determination of 1H6 affinity for HSV-1 G4s 37
4 RESULTS AND DISCUSSION 38
DNA G-quadruplexes in the human genome 38
4.1 Discovery and characterization of a G4 structure in a relevant gene for liposarcoma 38
4.1.1 Computational prediction of G4s in the *MDM2* inducible promoter 38
4.1.2 Biophysical characterization of the most probable *MDM2* G4 40
4.1.3 *MDM2* G4 interaction with nuclear proteins 43
4.2 Genome-wide relationship between G4s and transcription in LPS 46
4.2.1 Mapping of DNA G4s in LPS through BG4 ChIP-seq 47
4.2.2 Transcriptional state of G4-containing genes 49
4.2.3 Effect of G4 genomic position on the transcriptional outcome 51
4.2.4 Comparison of the G4 state in different cell lines 53
4.2.5 Altering gene expression through a G4-stabilizing ligand 55
DNA G-quadruplexes in viruses 59
4.3 HIV-1 59
4.3.1 Understanding the molecular features necessary for NCL-G4 interaction 60
4.3.2 The G4 aptamer AS1411 as anti HIV-1 agent 71
4.4 HSV-1 73
4.4.1 Development of a new SPR immobilization strategy for 1H6 anti-G4 antibody 73
4.4.2 Screening of 1H6 affinity for HSV-1 and human G4s 79
5 CONCLUSIONS 82
6 ACKNOLEDGEMENTS 85
7 APPENDIX 88
7.1 List of recurrent abbreviations 88
7.2 List of used oligonucleotides 90

7.3 Index of figures and tables 93

8 BIBLIOGRAPHY 95

ABSTRACT

G-quadruplexes (G4) are among the most studied nucleic acids secondary structures. They are involved in several fundamental biological functions both in human and other organisms. On one hand the G4 deregulation can lead to different pathologies, on the other hand their peculiar conformation provides a specific target for therapeutic drugs.

In the present work we addressed different points of G4 biology both in human and viruses. We explored the formation of G4 structures in a rare type of hardly curable human cancer: the well-differentiated liposarcoma (WD-LPS). Through a computational tool (QGRS) we identified several putative G4 sequences in the inducible promoter of the LPS related oncogene *MDM2*. By circular dichroism (CD) spectroscopy and DMS protection assay we assessed the folding of the key putative G4-forming region. By a pull-down assay we identified several proteins that specifically interact with *MDM2* G4. The vicinity of this structure to transcription factor binding sites suggests an involvement in the modulation of *MDM2* transcription initiation. Moreover, the high number of interacting proteins with helicase function reflects the importance in maintaining the unfolded state of *MDM2* G4 in the LPS cancer environment. This evidence and the unique structure of *MDM2* G4 makes it a valuable target for the hard challenge to cure LPS.

Expanding the analysis to the genome-wide level we mapped about 4800 G4s in the WD-LPS genome by ChIP-seq. We found an overrepresentation of G4s in promoter regions. An active transcriptional state and high expression were found in association with the presence of G4s, especially in gene promoters. The stabilization of G4s upon treatment with the potent G4-ligand H-NDI-NMe_2PhAm caused the downregulation of most G4-containing genes together with a massive alteration of the LPS transcriptome. Moreover, G4 folding appeared to be cell-type dependent, helping in the establishment of cell-specific transcriptome and identity. These results are one of the firsts attempts to characterize position and function of G4s in an *in vivo* system and move a step forward in the understanding of G4 involvement in the regulation of gene transcription in a cancer context.

In addition, our group recently discovered the formation of G4 structures in the genome of two worldwide spread viruses: HIV-1 and HSV-1. We found that the cellular protein NCL interacts with HIV-1 LTR G4s repressing viral transcription. To understand the G4 structural feature necessary for NCL binding we applied a systematic EMSA-based experimental approach. We found that long loops in the G4 structure are the main required feature for high affinity

recognition. This information can be useful for the rational design of new anti-HIV-1 drugs mimicking NCL action and repressing the provirus replication. Here we also demonstrated that NCL is a valuable antiviral target also at the level of viral entry. NCL is indeed a co-receptor for viral attachment to permissive cells and we showed that the G4-forming aptamer AS1411, already tested in phase II clinical trials for anticancer purposes, is able to sequester membrane NCL, thus inhibiting HIV-1 entry. We therefore propose AS1411 as a highly tolerated alternative approach to the current entry inhibitors.

Finally, we developed a new surface plasmon resonance (SPR) immobilization method. This strategy provides several advantages for the oriented immobilization of easy-to-degrade ligands and can be theoretically applied to each molecule that does not successfully immobilize through the standard covalent couplings. Employing this strategy, we measured with high reliability the affinity of the anti-G4 monoclonal antibody 1H6 for different HSV-1 and human G4s. The measured high affinity of the antibody to the most represented G4s in the HSV-1 genome was exploited to visualize the viral G4s in infected cells. We therefore provided one of the first evidence of the biological modulation of HSV-1 G4s in a biologically relevant context.

RIASSUNTO

I G-quadruplex (G4) sono tra le strutture secondarie del DNA attualmente più studiate. Tali strutture sono coinvolte in molte funzioni biologiche sia nell'uomo che in altri organismi. La deregolazione dei G4 può portare all'insorgenza di diverse patologie e la loro conformazione unica ne permette l'utilizzo come bersagli terapeutici specifici.

Questo lavoro si occupa di diversi aspetti della biologia dei G4 sia nell'uomo che nei virus. È stata esplorata la formazione di strutture G4 in un tumore raro e difficilmente curabile: il liposarcoma ben differenziato (WD-LPS). Utilizzando l'algoritmo QGRS, sono state identificate più sequenze che presentano la possibilità di formare un G4 nel promotore inducibile di un gene coinvolto nella patogenesi del liposarcoma: *MDM2*. Attraverso dicroismo circolare (CD) e il saggio di protezione con DMS, è stata verificata la formazione del G4 di una sequenza chiave fra quelle individuate. Sono state inoltre identificate alcune proteine che interagiscono in modo specifico con il G4 di *MDM2*. La vicinanza del G4 in analisi con più siti di legame di fattori trascrizionali suggerisce il coinvolgimento nella regolazione della trascrizione. Inoltre, l'elevato numero di elicasi in grado di legare il G4 di *MDM2* evidenzia l'importanza di mantenere il suo stato dispiegato nel contesto del liposarcoma. Queste evidenze, assieme alla sua conformazione unica, rendono il G4 di *MDM2* un interessante bersaglio per la cura del liposarcoma.

Ampliando l'analisi a livello genomico, è stata impiegata la tecnica di immunoprecipitazione della cromatina seguita da sequenziamento (ChIP-seq). Questa tecnica ha permesso di mappare circa 4800 G4 in cellule di WD-LPS. Analizzando la localizzazione dei G4, è stato osservato un arricchimento nelle regioni promotoriali. La presenza dei G4 in queste regioni ha inoltre dimostrato una correlazione con lo stato attivo della trascrizione e un'elevata espressione genica. Stabilizzando i G4 attraverso il composto H-NDI-NMe2PhAm è stata ottenuta la repressione della maggior parte dei geni contenenti G4 e l'alterazione massiva del trascrittoma cellulare. Una diversa modulazione dello stato dei G4 è stata osservata in linee cellulari diverse con conseguente variazione dell'espressione dei geni corrispondenti. Questi risultati costituiscono una delle prime evidenze della funzione dei G4 *in vivo* e migliorano la comprensione del ruolo dei G4 nella regolazione della trascrizione genica in un contesto tumorale.

Il gruppo di ricerca coordinato dalla professoressa Richter ha inoltre recentemente individuato delle sequenze formanti G4 nel genoma di due virus a diffusione mondiale: HIV-1 e HSV-1. È

stato evidenziato che la proteina cellulare NCL interagisce con i G4 promotoriali di HIV-1, causando la repressione della trascrizione virale. Un approccio sperimentale basato sulla tecnica EMSA è stato applicato con l'obiettivo di individuare le caratteristiche strutturali necessarie per l'interazione NCL-G4. In questo modo è stata dimostrato che la lunghezza dei *loop* del G4 è il fattore determinante per il legame fra le due molecole. Questa informazione può essere sfruttata per il disegno ragionato di nuovi farmaci contro HIV-1 che mimino l'azione di NCL. In questo lavoro è stata anche dimostrata la validità di NCL come bersaglio antivirale all'ingresso di HIV-1 nelle cellule ospiti. NCL svolge infatti anche una funzione di co-recettore per l'ingresso virale. Questo lavoro dimostra che, l'aptamero formante G4 AS1411 è in grado di sequestrare la NCL di membrana, inibendo l'ingresso virale. Tale aptamero rappresenta perciò un approccio alterativo agli attuali farmaci inibitori dell'ingresso virale nella terapia contro l'HIV-1.

Infine, un nuovo metodo di immobilizzazione è stato ideato ed applicato alla tecnica di SPR. Questa tecnica fornisce diversi vantaggi, permettendo l'immobilizzazione orientata di molecole facilmente degradabili e applicabilità a qualsiasi molecola dove le tradizionali tecniche di immobilizzazione covalente non sono efficaci. Sfruttando questo metodo, è stata misurata l'affinità dell'anticorpo 1H6 per diversi G4 cellulari e virali. Avendo dimostrato una maggiore affinità per i G4 di HSV-1, tale anticorpo si è dimostrato ottimale per la visualizzazione dei G4 virali in cellule infettate e ha fornito un'importante evidenza biologica della modulazione dei G4 di HSV-1 in un contesto biologicamente rilevante.

1 INTRODUCTION

1.1 DNA secondary structures

DNA is today known as the prevalent genetic material in living organisms and repository of hereditable information. The genetic information is encoded in DNA into two different forms: as linear nucleotide sequence and as a structural code, regulating its accessibility. Over 60 years ago, Watson and Crick described the predominant and more energetically favoured form in which DNA can be found: the right-handed double helix[1]. However, it became soon clear that other conformations and localized secondary structures can also be adopted by the most prevalent nucleic acid. Two different forms of right-handed double helix are possible through the base pairing modality described by Watson and Crick in the presence of high or low water hydration, respectively called B-DNA and A-DNA[2]. Ten years later a different base pairing modality has been described by Karst Hoogsteen, in which the purine base is rotated and the C1'-C1' distance is smaller with respect to the standard geometry[3] **(Figure 1-1)**. Despite being rarer, Hoogsteen hydrogen bonding is fundamental for the formation of important double helix variations such as triple helices and G-quadruplexes which come on top of other known biologically relevant structures like bubbles, Z-DNA, cruciforms, slipped loops and *i*-motifs[4–7] **(Figure 1-1)**. The next paragraphs will provide a more detailed survey about G-quadruplex (G4) structural properties.

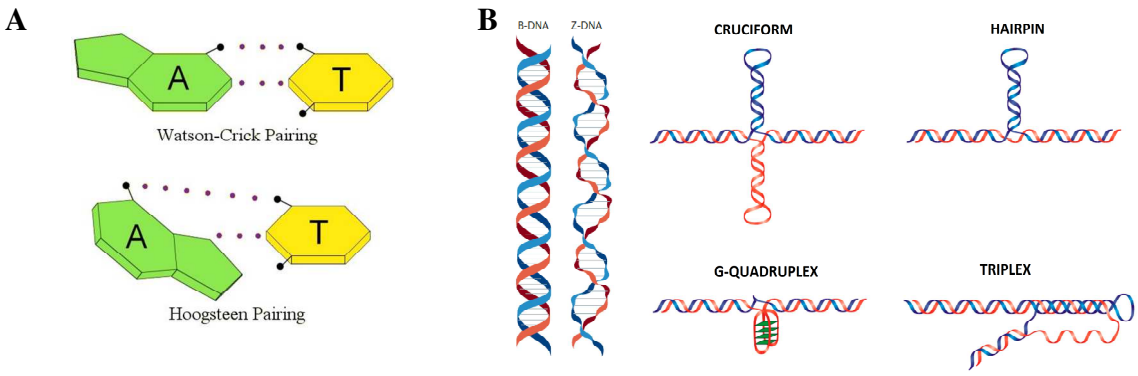

Figure 1-1. ***Alternative forms of DNA.***

A) *Comparison of nucleoside spatial arrangement and bonding in Watson and Crick or Hoogsteen base pairing.* ***B)*** *Examples of non canonical DNA double helix and local secondary structures compared to the prevalent form of B-DNA*[8,9].

1.2 G-quadruplex biophysical properties

Since 1910 it was known that high concentrations of guanine bases (G) can form viscous gels in water. This behaviour found an explanation only 50 years later, when X-ray diffraction data showed that guanosine moieties are able to arrange in a tetrameric organization linked by Hoogsteen hydrogen bonds[10]. Adjacent runs of guanines (G-tracts) in nucleic acid sequences (DNA and RNA) can form planar associations of four Gs (G-quartets), coordinated with a central monovalent cation (generally K^+ or Na^+). A G-quadruplex (G4) results from the stacking, one upon the other, of two or more G-quartets **(Figure 1-2 A)**. Stretches of nucleotides connecting adjacent G-tracts are called loops and, depending on the length and base composition of loops, the way a G4 can fold presents a high degree of polymorphism. Gs in each quartet can adopt *syn* or *anti* glycosidic bond angles, G-tracts can orient in the same or opposite 5'-3' direction one respect to the other, forming parallel, antiparallel or hybrid conformations **(Figure 1-2 B)**. Depending on the conformation, grooves between G-tracts are of variable size. For intramolecular G4s the strand polarity is strictly related to the loop nature. Loops can adopt lateral, diagonal or propeller orientation **(Figure 1-2 B).** G4s can assemble also from G-tracts belonging to two or more separated molecules, forming bimolecular or higher order G4 structures.

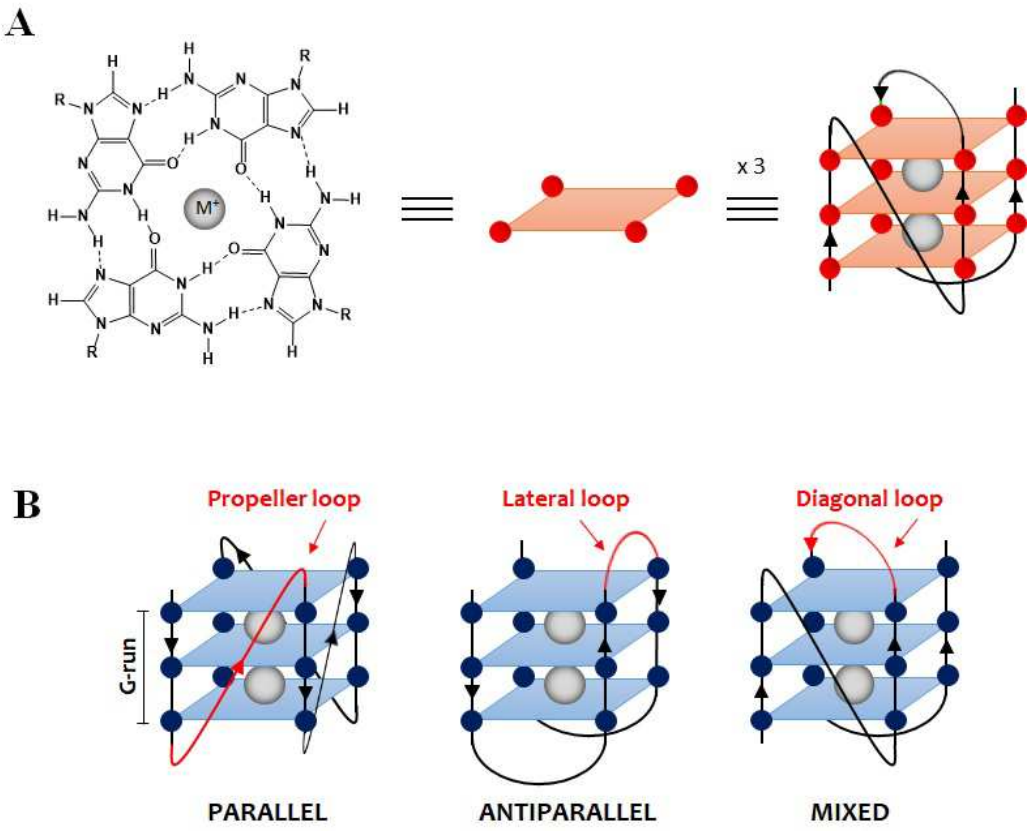

Figure 1-2. ***G-quadruplex polymorphism.***

A) *Guanines arrangement into a G-quadruplex. Hoogsteen bond are represented as dotted lines. A central metal ion favors the self association of the structure into a G-quartet and more stacked G-quartets form the G-quadruplex (adapted from Tassinari et al.* [11]*).* ***B)*** *G4 loop types. Highlighted in red from left to right: propeller, lateral and diagonal loop. Adapted from Ruggiero et al.* [12]*).*

The elevated plasticity of G4s has always exerted great curiosity in dissecting the factors determining their precise conformation, spatial arrangement and stability. Nowadays, a deep knowledge and advanced technologies are available for the biophysical characterization of G4 structures. What pushed these investigations is the opportunity to exploit the high polymorphism of G4s for the selective recognition and interaction with binding partners. A special interest comes from the pharmaceutical field, seeing in G4s a highly specific target for the development of antitumoral and antimicrobial drugs.

Several studies suggest that the loop sequence and length strongly influence G4s conformation, folding dynamic and binding ability to proteins and small molecules[13,14]. Through single molecule FRET and Circular Dichroism (CD) experiments, Tippana et al. showed that the

presence of single thymine loops is sufficient for inducing a parallel G4 conformation, even in presence of very long neighbour loops (6-9 bases)[13]. For non-thymine single nucleotide loops (A or C), the same behaviour is only confirmed for G4s with remaining loops not longer than 3 bases[15]. Steric constraint can be the explanation for these observations. The presence of all loops longer than two bases, instead, gives rise to G4s with mixed conformations and alters the whole folding dynamic: while single base loops fold very fast into a parallel topology, longer loops firstly acquire an antiparallel conformation and can then rearrange in a less defined and slower path.

Interestingly there is a direct correlation between G4 stability and the number of single nucleotide loops. Through UV melting experiments Bugaunt et al. demonstrated that G4s with longer loops have a lower thermodynamic stability. Moreover, single nucleotide substitutions in the loops appear to strongly affect G4s stability. In particular pyrimidines (T, C) are preferred over adenine, both in single base and longer loops[16]. This observation is an important proof of the fact that the loop sequence, not only influence G4 topology, but also stability.

This information is the starting point of one of the projects that we addressed in the present work: i.e. is the characterization of G4 structural determinants necessary for the recognition and binding of the cellular protein nucleolin (NCL). Since previous researches reported that the length and sequence of G4 loops were critical factors influencing G4 biophysical properties and thus G4 binding to other molecules, these aspects should be considered when studying both proteins and small molecules ligands interacting with G4s.

1.3 G-quadruplexes biological functions

G4s are characterized by a scaffold pattern consisting of at least four stretches of two or more guanines, separated by a variable number of other bases. This conserved motif can be reduced to a consensus sequence: $G_{\geq 2}\ X_{1\text{-}13}\ G_{\geq 2}\ X_{1\text{-}13}\ G_{\geq 2}\ X_{1\text{-}13}$ where G are guanines and X are tracts of any residue. Several computational algorithms are available[17], that allowed the initial identification of putative G4 structures in the genomes of several organisms, like yeast[18] and other eukaryotes[19,20] along with plants[21], prokaryotes[22,23] and viruses[24]. Biophysical studies first and functional biology assay then, revealed the implication of G4s in many fundamental functions for cell maintenance[25].

The present work treats different aspects of G4s in humans and viruses: i) the regulation of human gene expression and the role in a specific human cancer (liposarcoma); ii) the modulation and visualization of G4s in selected viruses (HIV-1 and HSV-1).

1.3.1 G-quadruplexes in Human

G4s are dynamic structures which primarily form in single stranded nucleic acid regions. This feature is found in mRNA and telomeric ends, as well as transient single stranded intermediate occurring during DNA replication, transcription and repair. The regulatory role of G4s in these fundamental processes is described below in more details, with a particular focus on transcription, which is object of investigation in the present work **(Figure 1-3)**.

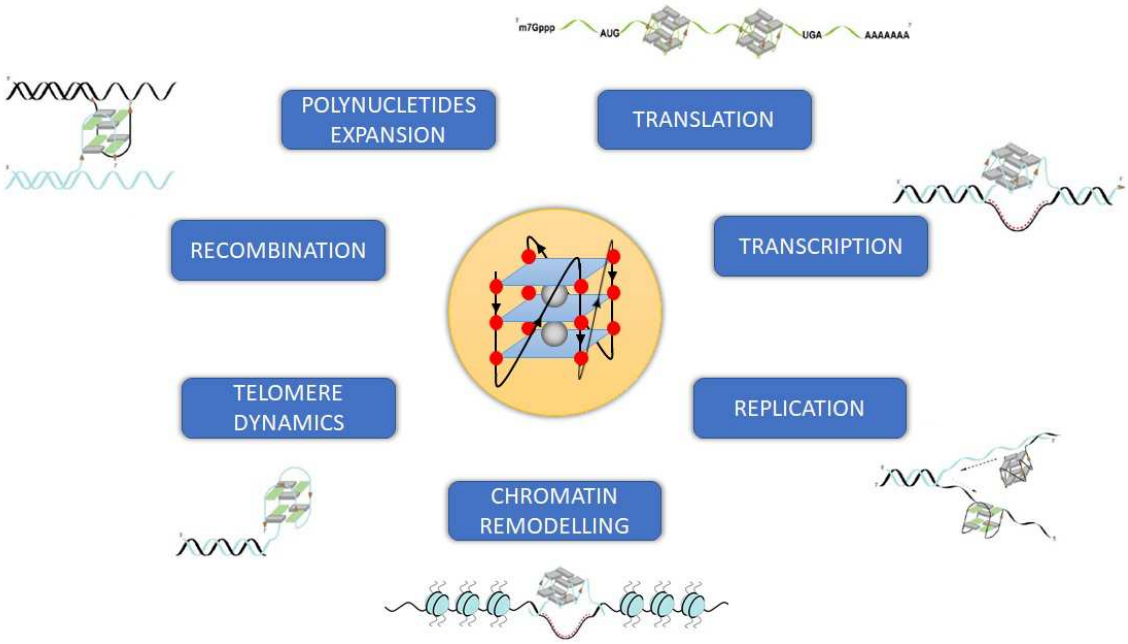

Figure 1-3. ***Role of G4 structures in the genome.***

Formation of a G4 at telomeric end. Chromosome ends alignment mediated by an intermolecular G4. During transcription G4s can form in the non template strand or on the template, producing DNA/RNA duplex (R-loop). Intermolecular G4 are putative sites for DNA recombination and crossing. Formation of G4s during replication con lead to replication fork stalling. Formation of G4s within mRNAs block translation[40].

A portion of the human genome which can easily fold into G4 is the telomeric 3' overhang characterized by the tandem repeat of the GGGTTA sequence[26–28]. The proposed function of telomeric G4s is to provide an alternative form of telomere capping, protecting chromosome

ends from being recognized as unwanted DNA breaks and favouring telomeric heterochromatinization through TERRAs[29,30].

In addition to telomers, stably single stranded nucleic acids are also found at the mRNA level. RNA is relatively unconstrained and thus prone to adopt secondary and tertiary structures. Several examples of G4s folding at the 5'UTR level have been reported to play an important role in the post-transcriptional regulation of gene expression, being the target of translation regulatory proteins or favouring IRES (internal ribosome entry sites) activation[31–34]. Moreover, transient single stranded DNA portions form as intermediates during the processing of the double helix: principally during transcription, replication and recombination. In particular there is evidence of G4 involvement in the selection of firing replicative origins[35–37] and obstruction of the mismatch repair proteins, leading to genomic recombination and repeat expansion[38–40]. Finally, G4s are involved in transcription regulation with both enhancing and repressing effects. Being this aspect of G4 biology one of the subjects of the present work, we dedicated a more detailed survey on G4s implication in gene transcription in the next paragraph.

1.3.1.1 Transcription

The high frequency of putative G4s at gene promoters immediately suggested a role of such structures in the regulation of gene expression. The first confirmation of this hypothesis came from studies on the nuclease hypersensitive element (NHE) III_1 of the oncogene *C-MYC*, followed by several other studies supporting the implication of G4s in the fine tuning of gene transcription[41].

The *C-MYC* NHE-III_1 controls 80 to 90% of the gene *C-MYC* transcription and its ability to fold into a G4 structure has been deeply characterized *in vitro*[42]. G4s alone are not however sufficient to decide for the fate of a gene in terms of expression. Indeed, it is through the interplay with specific proteins that G4s affect transcription. In one of the proposed models, upon helicase mediated unwinding of the double-stranded DNA helix, the temporarily single-stranded G-rich filament of *C-MYC* is free to assume a G4 conformation. The complementary C-rich strand is consequently exposed and recognized by transcription factors (e.g. hnRNPK) which in turn recruit the RNA polymerase machinery to start transcription[41]. On the contrary, under nuclear stress conditions, the phosphoprotein NCL binds and stabilizes the *C-MYC* G4 causing a repressive effect on transcription[43,44] **(Figure 1-4)**. Adding a further level of complexity, NCL-G4 interaction does not always lead to the same effect on transcription. For

instance, when bound to *VEGF* and *NPGPx* promoter-G4s it causes transcription enhancement by keeping the double helix open[45,46].

The apparently controversial role of G4s in the modulation of transcription suggests an important contribution of the cellular environment in defining the conditions for a gene positive or negative regulation.

It must be considered that under physiological conditions G4s are not always formed in the double helix, since sufficient kinetic energy is necessary to promote their folding. A source of this energy is for example the negative superhelicity accumulated behind the RNA polymerase machinery during its progression. Consequently, G4s can act as transcriptional modulators or recruit helicases and topoisomerases to be resolved and reconstitute a stable DNA double helix[47–49].

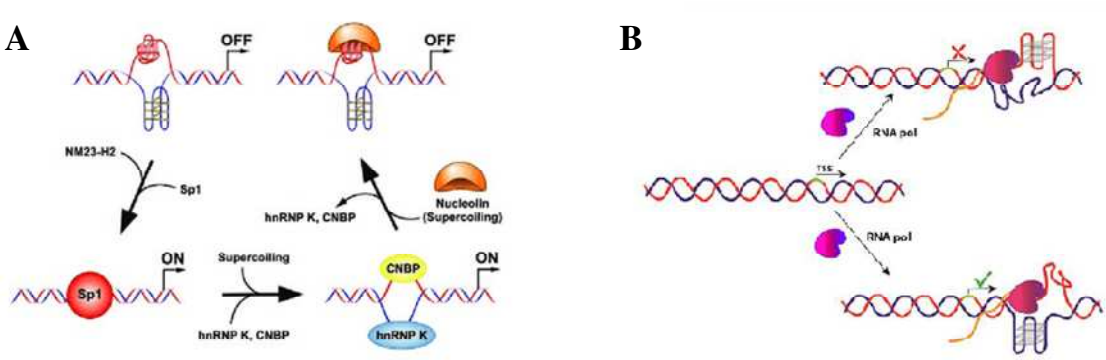

Figure 1-4. ***G4 mediated regulation of transcription.***

A) *A model for c-Myc transcription regulation: transcription is inhibited or activated depending on the proteins that are recruited on the promoter, which in turn depends on the folded or unfolded state of the NHE-III$_1$ G4*[50]. ***B)*** *The effect of G4 folding in promoter region can be opposite depending on which is the G-rich DNA strand*[41].

Another issue that influences the outcome of G4 formation is the position of the G-rich sequence within a gene (intron, exon or outside) and the strand (sense or antisense) the G-rich sequence is in **(Figure 1-4)**. A genome-wide analysis revealed that putative G4 forming sequences are predominant on the coding strand and downstream of the TSS, and that they generally inhibit transcription. However, an opposite effect has been observed when G4s are formed in the template strand, as in the case of the Bloom and Werner Syndromes, in which the

loss of function of specific helicases leads to an up-regulated transcription of G4-associated genes, in contrast to the expected down-regulation[51–53].

An additional consideration in understanding the complex mechanism of G4-mediated regulation of transcription is the high polymorphism that characterizes the G4 structure, as discussed in the previous section. Despite the common scaffold, G4s can indeed adopt a variety of spatial conformations, with different stabilities and kinetics. This affects not only their recognition by specific proteins, but also the consequent outcome of this binding. Moreover, G4s are often found in regions with multiple G-stretches allowing the formation of several distinct G4s, which can be either exclusive or contemporarily folded. The heterogeneity of G4 clusters not only contributes to an intricate regulation of gene transcription but is also hard to be correctly interpreted. Since *in vitro* studies not always reproduce the exact physiological conditions and G4s are highly sensitive to the environment, a G4 studied *in vitro* can be different from the one that actually forms *in vivo* and the obtained results can be misleading.

In this concern, one aspect that should not be neglected for a comprehensive study of G4 involvement in transcription is epigenetics. Gene transcription and expression is a complicated mechanism which must be strictly regulated in a cell-type specific manner and with the correct timing during development and the different phases of cells life. Chromatin compaction, DNA methylation and histone modifications are the main heritable features contributing to gene expression modulation and establishment of cellular identity during development but also in adult organisms. G4 involvement in epigenetic modifications has been reported in several studies[54]. Specifically, the correlation between DNA methylation and nucleosome occupancy has been investigated.

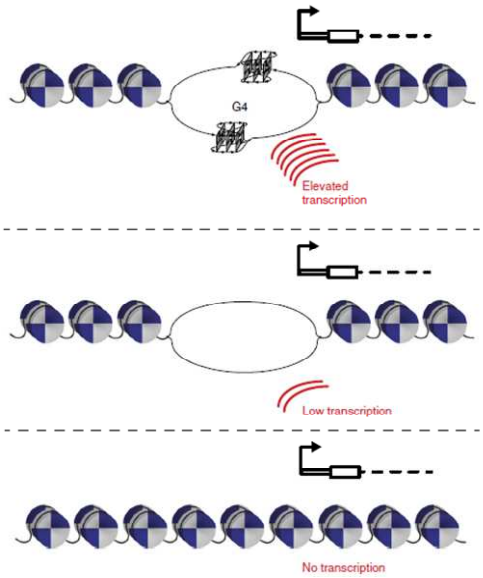

Figure 1-5. ***G-quadruplex structures role in regulatory chromatin.***

DNA G4s tends to form in openchromatin regions and are associated to a high level transcription [55].

Cytosine methylation at CpG islands is one of the major epigenetic mechanisms for chromatin modelling. This modification is enriched at promoters and non-coding DNA regions, having as an effect the silencing of transcription[56]. The formation of non-canonical DNA secondary structures is known to affect the positioning of epigenetic marks and there are specific DNA methyltransferases (DNMTs) that preferentially methylate unusual DNA forms. Interestingly, a comprehensive study on 12 different tissues from the Human Epigenome Project revealed a conserved negative correlation between CpGs methylation and G4 motifs: CpGs with low methylation are enriched in G4 motifs, while high methylated CpGs are depleted. G4 formation can therefore be exploited by the cell to regulate the non-random distribution and inheritance of cytosine methylation, having important implications on the transcriptome[56].

The occurrence of G4 conserved sequences near promoter transcription start sites (TSS) also correlates with nucleosome depletion, apparently giving selective regulatory benefits, even in comparison with open chromatin TSS lacking G4s[57] **(Figure 1-5)**. This evidence suggests that G4 structures work as epigenetic marks for highly transcriptionally active genes.

The role of G4s in epigenetic programming has been recently consolidated demonstrating that small-molecules mediated stabilization of a specific G4 can induce heritable changes in histone

marks and cytosine methylation, providing a tool for locus-specific epigenetic reprogramming[58,59].

Thus, despite accumulating evidence of a linkage between G4s and gene transcription, a clear rationalization of the mechanism is still missing. In addition to the intrinsic complexity of G4 effects, a limiting factor in the advancement of these studies is the difficulty in analysing G4 in specific cellular contexts and to separate the direct transcriptional output from the consequent cascade of events, involving all the phases of gene expression and genome maintenance. Recent progresses have been made in this direction, thanks to the development of G4 specific antibodies and the easier and easier accessibility to high throughput and genome-wide technologies. In the present work, the relationship between G4s and transcription will be analysed in the context of the human liposarcoma (LPS) cancer, with an attempt to dissect both the general mechanism of G4-mediated regulation and the involvement of the LPS diagnostic oncogene *MDM2*.

1.3.1.2 Involvement in human diseases

The involvement of G4s as regulatory elements in the described crucial cellular processes requires a very precise control of their formation in terms of time and position. Consequently, improper folding, stabilization or unwinding of G4s can lead to severe disfunction of fundamental molecular pathways, with pathological outcomes.

The connection between G4s and human diseases was initially prompted by the following two issues: i) The presence of helicase activity at G4 enriched regions. DNA helicases are part of a family of important proteins responsible for the unwinding of nucleic acids structures. Several helicases implicated in the unfolding of G4 structures and restoring of the DNA double helix have been characterized[60]. The disfunction of these proteins is linked to many human diseases characterized by: accelerated senescence (Werner syndrome, Hoyeraal-Hreidarsson syndrome and Seckel syndrome)[61–63], immunodeficiency (Bloom's syndrome)[64], anemia[65,66] and neurodegenerative alterations (Fragile X Syndrome or Fragile X-associated tremor ataxia syndrome, Amyotrophic Lateral Sclerosis and Frontotemporal Dementia[60,67,68])

ii) The overrepresentation of G4s in proto-oncogene promoters. Being cancer one of the firsts causes of death in the developed world, the possible implication of G4s in its pathogenesis has always attracted great interest and is also part of the present research.

Cancer and soft tissue sarcoma

The understanding and treatment of cancer is a difficult goal, due to the heterogeneity and complexity of this abnormality. At the beginning of the century Hanahan and Weinberg defined six hallmarks that describe all different types of oncogenic malignancies[69]. A linkage between G4s and cancer appeared with the observation that each of the hallmark categories was represented by at least one G4-containing gene: self-sufficiency for growth signal (*C-MYC*, *C-KIT*, *KRAS*), insensitivity to anti-growth signal (*pRB*), evasion of apoptosis (*BCL-2*), sustained angiogenesis (*VEGF-A*), limitless replicative potential (*hTERT*), and tissue invasion and metastasis (*PDGF-A*)[70] **(Figure 1-6 A)**.

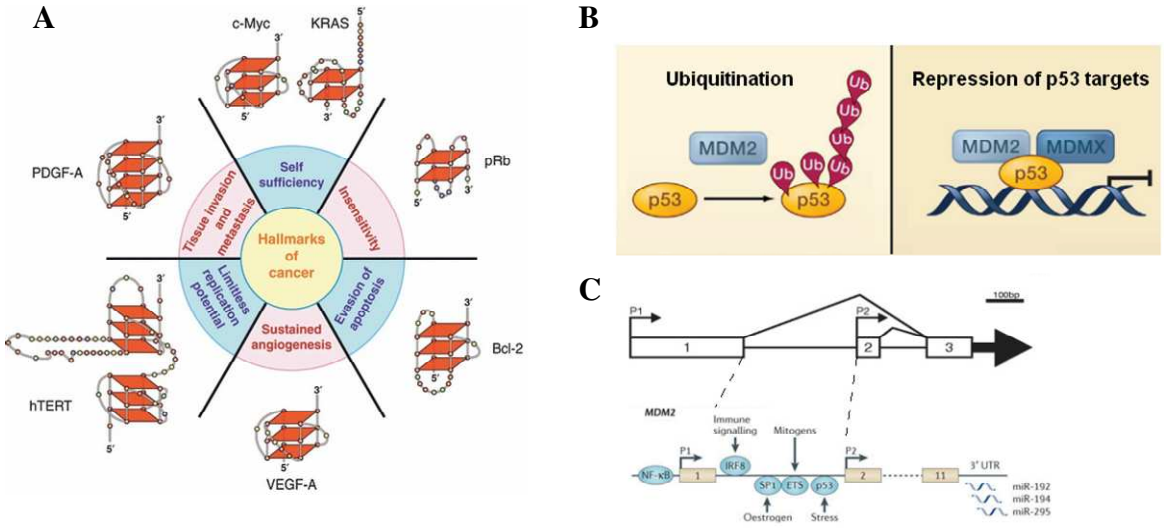

Figure 1-6. ***G-quadruplexes in the hallmarks of cancer and LPS.***

A)*The six hallmarks of cancer in relation with G4 containing genes*[71]*.* ***B)*** *MDM2 main ways for inhibiting p53. MDM2 mediates the export of p53 in the cytoplasm where it is ubiquitinated and degraded by the proteasome (on the left). MDM2 binding to p53 impedes p53 transcriptional activity (on the right)*[72]*.* ***C)*** *MDM2 promoter organization. P1 and P2 transcriptional products include different 5'-UTRs. Transcription factors regulating the two promoters are indicated*[73,74]*.*

The function of G4s in the mentioned genes has been extensively studied and promisingly appeared that their stabilization with G4 ligands has as a common effect the downregulation of transcription of the corresponding genes, with potential anti-tumoral effects. In this light, the

study of G4s in relation with cancer becomes even more attractive and motivated our interest in investigating their involvement in soft tissue sarcomas, for which there is a limited biological knowledge and pressing necessity for new specific treatments.

Sarcomas are a rare and heterogeneous type of tumours, predominantly arising from mesenchymal cells. Among the most common forms there are bone sarcomas and soft tissue sarcomas, with about 28.000 new cases per year in Europe[75]. Treatment for sarcomas is based on surgical resection and radiotherapy, however conventional chemotherapy has a controversial efficacy; therefore, new therapeutic approaches are required for a most efficient treatment[76].

A common feature between osteosarcomas and liposarcomas (one of the most frequent type of soft tissue sarcoma) is the amplification of the chromosome 12 region comprising the proto-oncogenes Cycline Dependent Kinase 4 (*CDK4*) and Mouse double minute 2 (*MDM2*).

In addition to this somatic alteration, *MDM2* presents two germline single-nucleotide polymorphisms (SNPs), SNP309 and SNP285, which are related to cancer predisposition and accelerated tumorigenesis[77]. From a molecular point of view, MDM2 is the principal regulator of the tumor suppressor gene *p53*. In particular, MDM2 is involved in a negative feedback loop in which it inactivates p53 by translocating it from the cytoplasm to the nucleus, ubiquitinating and targeting it for proteasomal-degradation, hindering p53 ability to interact with the transcriptional machinery and with target promoters. On the other side p53 binds to two adjacent p53 recognition elements in the first intron of *MDM2*, thus promoting its transcription **(Figure 1-6 B)**[72]. *MDM2* contains two promoters: the first (P1) is located upstream to the gene and is constitutively active even in the absence of p53, the second (P2) resides within the first intron and is induced by p53 binding **(Figure 1-6 C)**. Transcripts resulting from the P2 promoter lack the first exon of the normal mRNA and present an enhanced ability to be translated in to full-length MDM2 protein due to the presence of an alternative open reading frame[73,78].

It is to note that p53 is not the only transcription factor able to induce the P2 promoter activation: a number of other responsive elements have been identified in this region: a thyroid hormone response element, ETS and combined AP1-ETS binding site recognized by growth factors of the ras-raf-MEK-MAPK-signaling pathway and ERα[79].

Since *MDM2* gene amplification is a key event in about one third of sarcoma patient's pathogenesis, it has been exploited for the rational design of small molecules acting as *MDM2* inhibitors. To date the most successful results have been obtained with RG7388 (Idasanutlin), a derivative of Nutlin3a, which has entered a phase III clinical trial[80]. The Nutlin cis-imidazolines

class of small molecules target a hydrophobic cleft in MDM2 N-terminal domain which is the precise binding region of p53, having as effect the increase in apoptosis and reduction of proliferative capacity[77].

Interestingly, we noticed that the P2 promoter of *MDM2* gene is particularly rich in guanines, and that it contains a peculiar $(nnGGGGC)_5$ reapeat sequence. The integrity of the latter appeared to be necessary for P2 activity in $ER\alpha^{-ve}$ breast cancer cells[79].

In the present project we checked for the ability of the *MDM2*-P2 promoter, and more specifically of the $(nnGGGGC)_5$ repeat sequence, to form G4 structures and we investigated their role in the modulation of P2 transcriptional activity.

1.3.2 G-quadruplexes in viruses

While G4s have been extensively studied in the human genome, viral genomes have started to be investigated for the presence of such secondary structures only recently. Interestingly, G4 harbouring sequences have been found in important regulatory regions for the life cycle of several viruses and G4-ligands showed potential therapeutic activity, raising the interest of a deeper investigation in this field.

The presence of G4s has been experimentally reported in many *Herpesviridae*: the Herpes Simplex Virus 1 (HSV-1), Epstein-Barr Virus (EBV), Kaposi's Sarcoma Herpesvirus (KSHV) and Human Herpesvirus 6 (HHV6)[81–84]. Preliminary studies on the formation and role of G4s have been conducted on various other DNA viruses: The Human Papilloma Virus (HPV), Hepatitis B Virus (HBV), Adeno-associated Viruses (AAV); and RNA viruses: the Human Immunodeficiency Virus 1 (HIV-1), Severe Acute respiratory Syndrome Virus (SARS-CoV), Hepatitis C Virus (HCV), Zika Virus (ZIKV), Ebola Virus (EBOV) and Marburg Virus (MARV)[24].

In the present work, the attention is mainly focused on HIV-1 and HSV-1.

1.3.2.1 HIV-1

The Human Immunodeficiency Virus HIV-1 was identified for the first time in 1983 as the etiologic agent of the Acquired Immune Deficiency Syndrome (AIDS)[85]. It is a Lentivirus of the

Retroviridae family that possesses two identical copies of a single stranded RNA genome. HIV-1 replicative cycle starts with the recognition of $CD4^+$ T-helper cells by the envelope protein gp120[86]. The viral attachment to the cell is mediated by two main co-receptors, CCR5 or CXCR4 respectively for M-tropic and T-tropic lymphocytes. After fusion of the viral pericapside with the cellular membrane, the viral genome is released into the cytoplasm, where the viral reverse transcriptase promotes the synthesis of a complementary cDNA containing two identical Long Terminal Repeats (LTRs). The LTRs are necessary for viral integration in the host genome and function as promoter for viral transcription. Alternatively, the LTR promoter in the provirus can become silent and thus a latent infection ensues with the virus, that evades immune system detection[87] **(Figure 1-7 A)**.

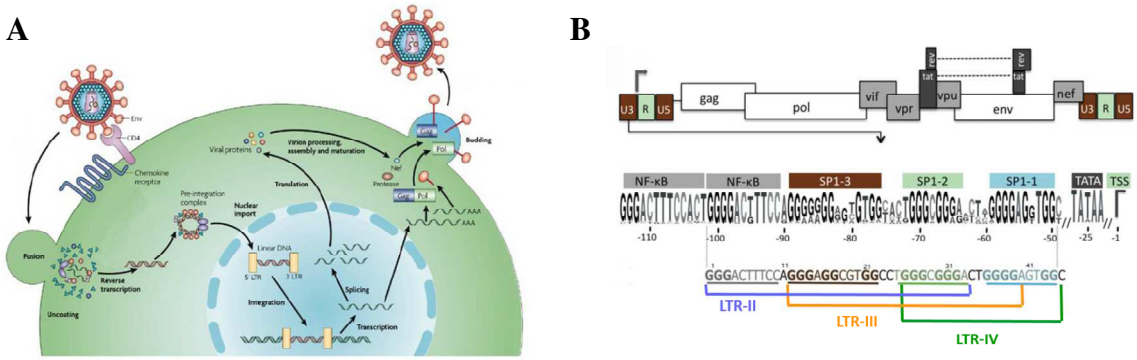

Figure 1-7. ***HIV-1 life cycle, genome organization and G4s.***

A) *Schematic representation of the different phases of the viral life cycle*[88]*.* ***B)*** *Organization of HIV-1 reverse transcribed genome and focus on the G4 forming sequence, subdivided into three main G4s: LTR-II, LTR-III and LTR-IV (adapted from:*[40]*).*

Despite the highly active anti-retroviral therapy (HAART) allows to keep the disease under control, a cure is still missing, and patients still die from HIV-related causes. Currently, more than 37 million people worldwide are infected by HIV and the pandemic continues to spread[89].

G4 structures have been detected at the HIV-1 RNA and DNA level and appeared to be involved in the regulation of many phases of the viral life cycle. A first event in which G4s have been found to play a critical role is viral RNA dimerization and recombination. During reverse transcription, the two copies of the viral RNA interact by a dimer initiation site (DIS). G-rich sequences, able to form bi-molecular G4s, are present near the DIS (in the gag region)

and are thought to be responsible for the stabilization of the dimer, allowing recombination between the two copies[90]. It has been proposed that the retroviral nucleocapsid protein (Ncp) is involved in this recombination mechanism enhancing bi-molecular RNA G4 formation[91]. Other G4 forming regions have been found in the Nef reading frame that overlaps with the 3'-LTR. Nef is an accessory protein implicated in immune evasion through the down-regulation of $CD4^+$ and MHC cells. The stabilization of Nef G4s leads to repression of Nef expression and decreases viral replication[92]. Recently, our and another group demonstrated that the U3 region in the LTR unique promoter of HIV-1 contains a 45 nucleotides long G-rich sequence, upstream the transcription start site (TSS), that is able to fold into a dynamic cluster of G4s[93,94] **(Figure 1-7 B)**. Interestingly, some of the different G4s in this region (i.e. LTR-II, LTR-III and LTR-IV) are in a mutually exclusive equilibrium, suggesting a role in the regulation of the promoter activity. In this light, LTR-III and LTR-IV have been deeply characterized, since the first is the prevalent species under physiological conditions and the second is only formed upon induction with G4-ligands[95]. The two structures were characterized by nuclear magnetic resonance (NMR) revealing very peculiar conformations, which exert opposite effects on LTR promoter activity. LTR-III silences, while LTR-IV enhances viral transcription[96]. This effect is modulated by G4 interaction with cellular proteins: NCL, which stabilizes the LTR G4 with a preference for LTR-III, thus suppressing the promoter activity; and the human ribonucleoproteins hnRNP A2/B1, which unwinds the LTR-G4s promoting viral transcription[97,98]. Since NCL was already reported to interact with some, but not all G4s structures, we here aimed at understanding the G4 structural features necessary for NCL binding. In addition to improve the understanding of HIV-1 employment of cellular resources, this point opens the possibility to exploit NCL as alternative antiviral target.

1.3.2.2 HSV-1

Herpes Simplex Virus 1 (HSV-1) is one of the nine types of *Herpesviridae* which can infect humans. HSV-1 infections manifest as orolabial vesicular lesions of the mucosa with high recurrence. The current anti-HSV-1 therapies are highly effective, but do not save the patients from the reactivation of the virus. Moreover, a consistent percentage of HIV-1 infected patients and patients who underwent prolonged administration of anti-herpetic drugs, develop resistances to the treatment. The search for new anti-HSV-1 drugs is therefore still a priority[99].

HSV-1 entry into permissive cells can occur through fusion of the viral envelope with the plasma membrane or through endocytosis. The viral particle is then transported along cellular microtubules and translocated inside of the nucleus, where transcription, replication and

assembly of the new virions take place. HSV-1 has a long linear double stranded DNA genome, consisting of two units, long (U_L) and short (U_S), separated by a set of inverted repeats, TR_L–IR_L and IR_S–TR_S. Many enzymes involved in the nucleotide metabolism are encoded in the viral genome and are activated in a cascade, according to which they are classified into: immediate early, early and late. In the absence of protein biosynthesis, the viral genome is maintained in a latent circular form inside the nucleus of neurons and ganglia innervating infection susceptible epithelium, from which it can be reactivated through still unclear mechanisms[99] (**Figure 1-8 A**).

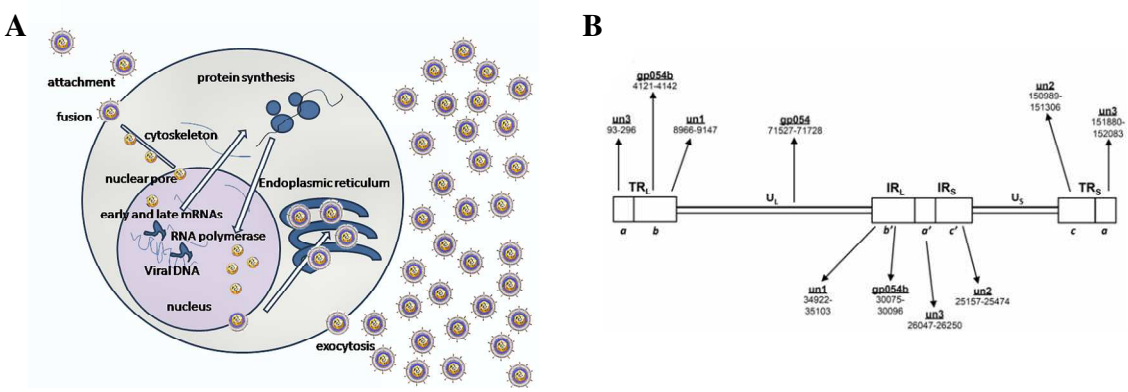

Figure 1-8. ***HSV-1 life cycle, genome organization and G4s.***

A) *Schematic representation of the different phases of the viral life cycle. Credit: GrahamColm.* ***B)*** *Organization of HSV-1 genome,position of the identified G4s is indicated*[100]

The HSV-1 genome has an incredibly high GC-content (about 68%) especially clustered in the repeat regions and thus has the potential to form numerous G4 structures. Biophysical approaches were used to demonstrate the formation of six G4s in the *gp054* gene (named *gp054a-f*), encoding for the most important viral tegument protein, and four more extremely stable G4s clustering at the terminal and internal repeats (named *unb, un1-3*)[101] (**Figure 1-8 B**). Due to the abundance of these repeated G4s in the HSV-1 genome, we aimed at visualizing them in infected cells through a G4-specific antibody (1H6) at different times post-infections. In the present work we describe the development of a tailored SPR strategy for the assessment of 1H6 antibody specificity and affinity for a series of viral and cellular G4s. This made it possible to confidently follow the HSV-1 G4s during the different phases of infection. In particular, viral G4s were shown to form in a virus cycle-dependent fashion, with a peak during viral replication and different localization in cellular compartments according to the viral genome movements[102].

1.4 G-quadruplexes as therapeutic target and tool

The involvement of G4s in key biological functions and their elevated polymorphism has motivated the interest in such structures as specific therapeutic targets. One of the most intriguing fields is anti-cancer therapy. Computational and experimental evidence supports the enrichment of G4 structures in proto-oncogene promoters and telomers, both involved in hallmarks of many cancers. In this context, the most successful target is *C-MYC* G4: the fluoroquinolone-derivative Quarfloxin **(Figure 1-9)** has reached phase II clinical trial[103]. Its mechanism is based on targeting *C-MYC* rDNA G4s, inhibiting RNA polymerase I transcription and NCL interaction, which is translocated in the nucleoplasm triggering a stress response that results in induction of apoptosis and cancer cell death[50,103]. Downregulation of *C-MYC* expression has also been observed with the ellipticine derivative GQC-05:NSC338258 and a 11-piperazynquindiline[104,105]. However, while in the first case *C-MYC* inhibition was a direct effect of G4 stabilization, in the second, G4 targeting was not directly involved in the observed effects. Therefore, assuming a direct cause and effect mechanism between *in vitro* and cellular observations can sometimes be dangerous.

Two other well investigated G4 targets are *BCL-2* and *C-KIT*[106,107].

BCL-2 expression was shown to be downregulated at mRNA and protein level by a library of mono-substituted quindoline derivatives and the tetra-substituted naphthalene diimide MM41 **(Figure 1-9).** Dose dependent reduction of C-KIT expression was observed in a human gastric cancer cell line treated with a mono-substituted benzo[a]phenoxazine **(Figure 1-9)**

Besides oncogene promoters, also telomeric G4s represent an interesting target for anti-cancer therapy. Several G4 stabilizing small-molecules have been reported to bind and affect telomere function. The G4 ligand BRACO-19 **(Figure 1-9)** has anti-tumoral activity ascribable to DNA damage induction at telomers. It has been hypothesized that G4 stabilization leads to displacement of telomere associated proteins and the exposed DNA extremities activate a cascade of DNA damage response. Deficiencies in DNA repair processes are common in many cancers, therefore their stimulation can lead to cell death[110].

Several others G4 stabilizing compounds with anti-cancer activity have been reported and targeting a single quadruplex target among the many that can fold in the human genome is object of continuous research. However, absolute specificity for a unique G4 is a hard

challenge, due to the common core that all G4s share. The occurrence of G4s in cancer related genes offers anyway the opportunity of a multi-target strategy, which can be even an advantage, considering the heterogeneity and complexity of most solid tumors[111].

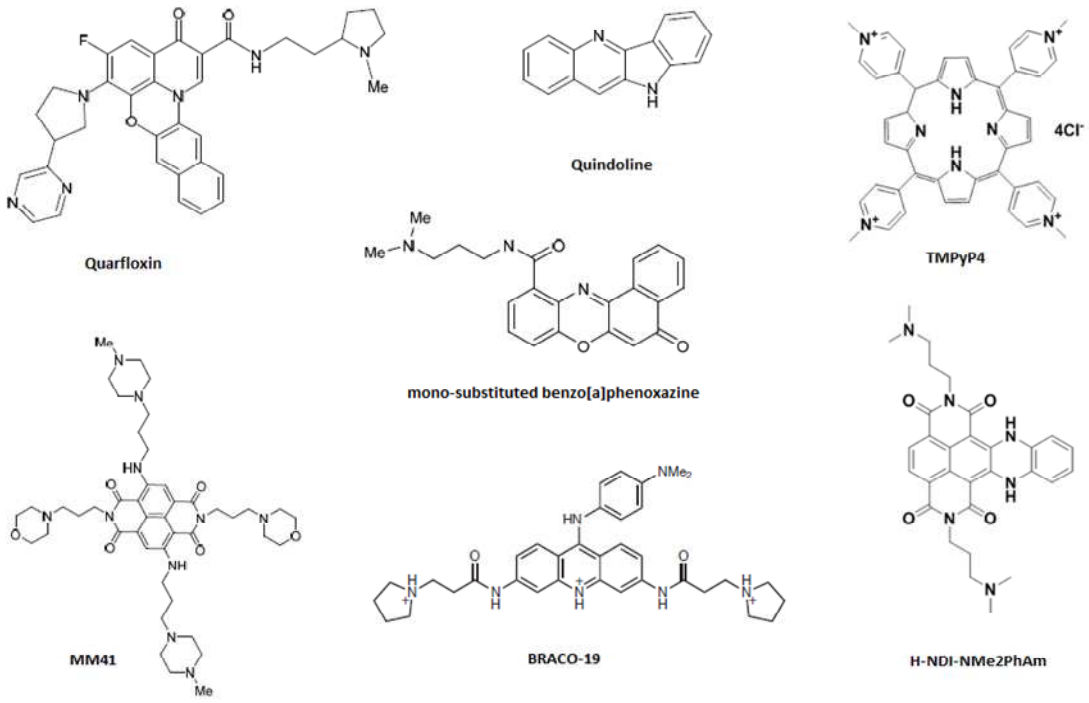

Figure 1-9. ***Chemical structures of some G4 ligands.***

Quarfloxin was designed as c-Myc stabilizer and is the first G4 ligand entering clinical trials. MM41 and Quindline derivatives is a tetra-substituted naphthalene diimide used to suppress Bcl-2 transcription. Mono-substituted benzo[a]phenoxazine was shown to have anti-cancer activity upon stabilization of c-Kit G4. BRACO-19 has both anti-tumoral and antiviral activity, has been tested as telomeric G4 stabilizer in huma[108,109]*n, and as antiviral in HIV-1 and HSV-1. TMPyP4 has been tested both as antitumoral and antiviral compound. H-NDI-NMe2Pham was found through the screening of a small library of c-exNDIs, it preferentially binds to viral G4s (HIV-1 and HSV-1) with respect to human telomere.*

The recent exploration of viral genomes for G4 forming sequences has provided the chance to develop new antiviral drugs with an innovative mechanism of action and great potential, given the high conservation of G4 genomic regions. For what concerns HIV-1, our group showed that targeting the HIV-1 LTR and Nef G4 with known G4 binders, such as BRACO-19 and TMPyP4 inhibited virus replication (**Figure 1-9**)[92,93]. Therefore, we are trying to achieve higher specificity for HIV.1 LTR G4s by screening and designing new compounds. Good results have been obtained with the core-extended Naphtalene Diimides (c-exNDIs) H-NDI-NMe_2PhAm

(**Figure 1-9**), which discriminate the HIV-1 promoter G4s from human G4s, inhibiting viral replication[111].

Remarkable antiviral activity has also been observed by using G4 ligands against HSV-1 infection. Both BRACO-19 and the c-exNDI H-NDI-NMe2-PhAm have been tested by our group[101,111].

The massive accumulation of viral G4s in the host nucleus is probably the reason of the significant activity of these compounds, leading to reduced viral DNA synthesis.

In addition to targets, G4s have also been validated as therapeutic means. For instance, two G4 oligonucleotide aptamers that entered phase II clinical trials are the thrombin-binding aptamer (TBA) and AS1411. The latter represents a potential anti-cancer drug with anti-proliferative activity and works by targeting NCL[112]. Several others G4 forming aptamers have also been studied as anti-HIV agents. The main mechanism of action of antiretroviral G4 aptamers is the inhibition of HIV-1 reverse transcriptase, virus spread among cells mediated by V3 loop binding (ISIS 5320), replication and integrase (T30177) and viral entry (T30923)[113–116]. We recently tested the AS1411 aptamer as antiviral agent, obtaining inhibition of viral entry. Here we will show evidence that AS1411 acts as antiviral agent by the subtraction of NCL availability as HIV-1 coreceptor[117].

In general, G4-mediated antiviral strategies represent an innovative and valuable alternative to current treatments, allowing to overcome the problem of resistances and giving hope for the development of a cure for latent infections.

1.5 Main challenges in the *in vivo* investigation of G-quadruplexes

Much of the earlier work on G4s was restricted to biophysical, biochemical and functional studies on synthetic oligonucleotides and *in vitro* environment. *In vitro* results can sometimes be misleading, not representing the real physiological conditions. Recently, significant advances have been reached in the detection of G4s in cells and in the whole genome context, supporting the importance of these non-canonical structures for many fundamental biological functions and therapeutic application[118]. Further experiments are however necessary for a more robust and complete understanding of G4 biology.

A first challenge that should be addressed is the investigation of the G4 state and regulation in different cell types and context. Recently, independent laboratories produced two anti-G4 antibodies: BG4 and 1H6[119,120]. Both antibodies and synthetic G4 ligands (PDSα, BMVC and DAOTA-M2) have been used to visualize G4s in cells by immunofluorescence[121–123].

The BG4 antibody has also been used to map G4s through ChIP-sequencing in an immortalized cell line of keratinocytes and the obtained data were integrated with quantitative transcriptome analysis[55]. More data are however necessary for a comprehensive understanding of how G4s are distributed and modulated in different cell lines, and during different phases of the cell cycle and development.

A second important challenge in investigating G4 function *in vivo* is the intrinsic difficulty to extrapolate the direct effect of G4 formation from the consequent cascade of molecular pathways. Manipulation of G4s can be achieved by insertion of mutations disrupting the structure or by stabilizing ligands. Both strategies provide a potent mean, however, being G4s involved in many of the vital cellular processes, the consequences of their alteration can lead to a number of direct and indirect effects modifying the whole cellular metabolism. In such conditions, researchers must be very cautious in drawing conclusions.

Despite the key role of G4s in the maintenance of vital cell functions, the variety of their implication and consequences of their alteration make the G4s field in constant evolution. A clear understanding of G4 mechanisms of actions is however a valuable opportunity for a wider knowledge of cellular mechanisms regulation and therapeutic targeting to treat lethal diseases as cancer and viral infections.

With the present work we tried to contribute to the increasing understanding of G4s implication in the regulation of gene expression. At the same time, we explored the biology of LPS, a highly lethal human cancer, for which G4s represent a promising new therapeutic target, possibly overcoming the problem of LPS resistance to the classical chemotherapies. We also focused on G4s in two worldwide spread viruses: HIV-1 and HSV-1. We explored the requirements for G4s interaction with the cellular protein NCL, known to downregulate HIV-1 G4-mediated transcription and establish the basic assumptions necessary for the monitoring of HSV-1 G4s in infected cells.

2 AIM

The present work investigates G4s in two very different contexts: humans and viruses. In the first part of this production we aim at exploring the formation of DNA G4 structures in a human cell line of WD-LPS (93T449). Very little is known about the biology of LPS and the mechanisms leading to its strong resistance to classical chemotherapies. Employing 93T449 cell line as a model, we not only aim at identifying previously unreported G4s in LPS relevant genes, which can be exploited as alternative therapeutic targets in the challenging treatment of LPS, but also expand the analysis to the genome-wide level, in the wider attempt to clarify the complex involvement of G4s in the regulation of the gene transcription process.

In the second part, we studied G4s in two worldwide-spread viruses, for which the search for new treatments is still a priority: HIV-1 and HSV-1. Our group previously discovered that the high affinity interaction between the cellular protein NCL and the HIV-1 LTR G4s leads to suppression of viral transcription. Given the reported ability of NCL to interact also with other but not with all G4s, we here aim at defining the G4 structural requirements for this interaction. Understanding how NCL interacts with G4s will provide useful information for the future design of therapeutic compounds, having this complex as target. Indeed, here we also tested NCL as antiviral target employing the phase II clinical trial G4 forming aptamer AS1411.

HSV-1 was instead chosen as subject of investigation for the huge number of G4s formed *in vitro* in its genome. This feature is highly suitable for visualizing the physiological modulation of viral G4s in infected cells by 1H6 antibody mediated IF. We here addressed the prerequisite to reliably perform the latter experiment: i.e. analysis of 1H6 specificity and affinity for the most abundant HSV-1 G4s with respect to the cellular ones.

In general, in this work we studied G4s in two different organisms (human and viral). In this attempt we employed two different anti-G4 antibodies (one used in ChIP-seq and the other in IF, for which we here prove the prerequisite) to break down the bigger obstacle in the understanding of G4 biology: the monitoring of their formation in *in vivo* systems. Moreover, we aimed at establishing G4s as new targets and tools for antitumoral and antiviral therapy.

3 MATERIALS AND METHODS

3.1 Oligonucleotides and compounds

Desalted oligonucleotides were purchased from Sigma-Aldrich (Milan, Italy) and qPCR primers from Eurofins (Munich, Germany). A detailed list of oligonucleotides name and sequence is available in the **Appendix 7.2**. BRACO-19 was obtained from ENDOTHERM GmbH (Saarbruecken, Germany), H-NDI-NMe_2PhAm and PyPhen$_2$PropNMe_2 were synthesized and provided by Freccero et al. (University of Pavia, Italy).

3.2 QGRS prediction of putative G4s

The first intron of human *MDM2* gene (GRCh38.p2, NC_000012.12) was analyzed by QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/index.php) for the prediction of G4 forming sequences[124]. The putative G4s were identified by the motif $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$, where x is the number of guanine (G) tetrads and N_y the length of loops connecting the G tetrads. The following restrictions have been applied: i) the number of tetrads had to be ≥ 2; ii) maximum length of QGRS was set to 40 bases; iii) at most one of the loops was allowed to be of zero length; iv) loop size from 0 to 15. The found QGRS were ranked based on the G-score, which is the likelihood to form a stable G4, according to the following principles: a) shorter loops are more common than longer loops; b) G4s tend to have loops roughly equal in size; c) the greater the number of G tetrads, the more stable the G4.

3.3 Circular Dichroism Spectroscopy (CD)

For assessing the folding of tested DNA oligonucleotides into G4 structures and determine their conformation and melting temperature, CD spectroscopy was performed. Oligonucleotides were diluted to a final concentration of 4 µM in lithium cacodylate buffer (10 mM, pH 7.4) with or without KCl or NaCl as indicated in text. 20% of PEG200 was added where indicated to test G4 conformational changes in molecular crowding conditions. After heat denaturation at 95 °C for 5 minutes, samples were folded at room temperature overnight. For experiment in the presence of BRACO-19, the compound was added after 4h from oligonucleotide heat

denaturation in saturating concentration of 16 μM. CD spectra were recorded on a Chirascan-Plus (applied Photophysocs, Leatherhead, UK) equipped with a Peltier temperature controller using a quartz cell of 5 mm optical path length. The temperature was set at 20 °C and the spectra were recorded over a wavelength range of 230-320 nm. For Tm determination, spectra were recorded over a temperature range of 20-90 °C, with temperature increase of 5 °C/min. Tm were calculated according to the van't Hoff equation, applied for a two-state transition from a folded to unfolded state. Acquired spectra were baseline corrected for signal contribution due to the buffer and the observed ellipticities converted to mean residue ellipticity (θ) = deg x cm2 x dmol-1 (molar ellipticity).

3.4 Dimethylsulphate footprinting (DMS)

The DNA substrate of interest was purified before use by running on a 20 % denaturing polyacrylamide gel until separation of the synthesis sub-products is obtained. Bands corresponding to the DNA product of interest were excised and eluted in water solution overnight at RT shaking. Eluates were recovered and concentrated by using Amicon Ultra-3k Centrifugal Filter Unit (Millipore). The purified oligonucleotides were then 5'-end-labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase for 30 min at 37 °C, purified by using MicroSpin G-25 columns (GE Healthcare Europe, Milan, Italy), resuspended in lithium cacodylate buffer 10 mM pH 7.4 with or without KCl 100 mM, heat-denatured for 5 min at 95 °C and folded at RT overnight. Sample solutions were then treated with dimethylsulfate (DMS, 0.5% in ethanol) for 4 min and stopped by addition of 10% glycerol and β-mercaptoethanol. Samples were loaded onto a 15% native polyacrylamide gels (acrylamide/bis solution 19:1) and run until the desired resolution was obtained. DNA bands were localized via autoradiography, excised and eluted overnight. The supernatants were recovered, ethanol-precipitated and treated with piperidine 1 M for 30 min at 90 °C. Samples were dried in a speed-vac, washed with water, dried again, and resuspended in formamide gel loading buffer. Reaction products were analyzed on 20% denaturing polyacrylamide gels (acrylamide/bis solution 19:1) containing 8 M urea, visualized by phosphorimaging analysis on a Typhoon FLA 9000 (GE Healthcare Europe,Milan, Italy), and quantified by ImageQuant TL software (GE Healthcare Europe, Milan, Italy).

3.5 Nuclear proteins extraction, pull-down and Mass-Spectroscopy (MS)

The 93T449 (ATCC® CRL-3043™) retroperitoneal WD-LPS cell line was used to detect proteins interaction with *MDM2* G4. 93T449 cells were grown in RPMI 1640 (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% heat-inactivated FBS. Nuclear extracts were obtained by using NXTRACT kit (Sigma-Aldrich, Milan, Italy) and quantified using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific, Monza, Italy).

Biotinylated oligonucleotides (Sigma-Aldrich, Milan, Italy) were diluted to 3 µM in phosphate buffer pH 7.4 20 mM supplemented with KCl 100 µM (PB-KCl buffer), heat denatured for 5 min at 95 °C and slowly cooled to RT to allow secondary structures folding. 50 µl of streptavidin coated magnetic beads (Dynabeads® M-280 Streptavidin, Thermo Fisher Scientific, Monza, Italy) were washed with PB-KCl buffer and incubated at RT for 30 min in the presence of 150 pmols of folded oligonucleotides *(Mdm2-biot*, *G-ss-biot* as a single stranded G-rich control and *C-ss-biot* as C-rich single stranded control) or PB-KCl buffer only as negative control. The excess oligonucleotide was washed with PB-KCl buffer and PBS-BSA 0.01% to reduce non-specific protein binding. After that, 25 µg nuclear extracts were added and incubated at 4 °C for 2-3h. Samples were washed three times with PB-KCL buffer and protein were eluted first in 20 µl NaCl 2 M and second in 2X Laemmli buffer (4% SDS, 80% glycerol, 120 mM Tris-HCl pH 6.8, 200 mM DTT, 0.02% bromophenol blue). Eluted proteins were separated on a 8% SDS-PAGE and stained overnight with colloidal Coomassie staining (0.02% w/v CBB G-250, 5% w/v aluminum sulfate-(14-18)-hydrate, 10% v/v ethanol, 2% g/v orthophosphoric acid). After destaining procedure (10% ethanol and 2% orthophosphoric acid) bands corresponding to different ranges of molecular weights were cut and in-gel digested. Briefly, they were first washed with 50% CH_3OH and 2.5% acetic acid, dehydrated with CH_3CN, and then reduced with 30 µL of DTT (10 mM in 100 mM NH_4HCO_3) for 30 min at room temperature. The excess of DTT was eliminated before treating the bands with 30 µL of iodoacetamide (50 mM in 100 mM NH_4HCO_3) for 30 min at room temperature in order to alkylate cysteine residues. Bands were washed with 100 mM NH_4HCO_3, dehydrated with CH_3CN twice, and then digested. Bands were washed with 100 mM NH_4HCO_3, dehydrated with CH_3CN twice, and then digested with 1 µg of MS-grade trypsin (ThermoFisher Scientific, Waltham, MA, USA) in 50 µL of 50 mM NH_4HCO_3 for 30 min on ice. The excess of trypsin was eliminated and substituted with 20 µL of 50 mM NH_4HCO_3 and the sample were incubated overnight at 37°C. Peptides were extracted twice with 5% formic acid and two more times with

50% CH_3CN, 5% formic acid, the peptide mixture was further concentrated in SpeedVac (Hetovac VR-1, Heto Lab Equipment, Denmark) to 1 µL and resuspended in 30 µL of 1% formic acid. The peptide mixture was spun down 10 min at 10,000 x g in order to eliminate any solid particles and analyzed by LC-MS. The injection was automatically performed by a low pressure Acquity H-class bioQuaternary Solvent Manager UPLC (Waters, Manchester, UK) system and a Jupiter proteo® RP12 (1.0 × 150 mm, 4 µm, 90Å) (Phenomenex, Torrance, CA, US) chromatographic column. The detection was performed at the beginning of the gradient ramp for 30 minutes with a mass analyzer Xevo G2-XS QTof mass spectrometer (Waters, Manchester, UK). Instrument control, data acquisition and data processing were performed with MassLynx 4.1 software (Waters Corp.). Parent ions having the charge state 4^+, 3^+ and 2^+ and signals more intense than 200 counts and the related MS^e fragments ion signals more intense than 50 counts were employed to perform a Mascot Database Search to identify their parent protein. The matched proteins were deemed as being positively identified when two or more peptides provided a mascot score greater than 20 and 2-fold higher than the controls.

3.6 Purification of BG4 antibody

BG4-encoding plasmid (kindly provided by Professor Shankar Balasubramanian, University of Cambridge, UK) was transformed into BL21(DE3) competent cells (Stratagene) which were cultured in TY medium (1.6% tryptone peptone, 1% yeast extract and 0.5% NaCl) and 50 µg/ml kanamycin. Transformed cells were grown at 37°C 160 rpm to an OD_{600} of 0.7-0.8._BG4 antibody expression was induced with 0.85 mM isopropyl β-D-1-thiogalactopyranoside overnight at RT. The cells were pelleted for 25 min at 25,000g at 4 °C, resuspended in lysis buffer (20 mM Tris-Cl pH 8.0, 50 mM NaCl, 5% Glycerol, 1% Triton and 100 µM Phenylmethanesulfonylfluoride solution) and lysed through 5 cycles of freezing and thawing. After centrifugation at 10,000g at 4 °C for 20 min, the supernatant was filtered (0.45 µm) and purified on a Protino Ni-NTA-Agarose Affinity column (Machery-Nagel, Germany) according to the manufacturer instructions. The column was washed in 20 mM imidazole in 20 mM Tris HCl pH 8.0 and 300 mM NaCl, and BG4 antibody 1.5 ml fractions eluted in 250 mM imidazole in 20 mM Tris HCl pH 8.0 and 300 mM NaCl. BG4 antibody containing fractions were checked on a Coomassie-stained SDS-PAGE and concentrated in Amicon Ultra-3k Centrifugal Filter Unit (Millipore). The concentration of BG4 was determined using Thermo Scientific Pierce BCA Protein Assay kit, and the antibody was stored at −20 °C.

3.7 G4 Chromatin Immunoprecipitation and library preparation

The 93T449 (ATCC® CRL-3043™) cells were grown to 80 % confluence in RPMI 1640 (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% heat-inactivated FBS. After trypsinization, 2 million cells were fixed in RPMI containing 1 % (v/v) formaldehyde and 10 % (v/v) FBS for 10 min at RT. After 5 min quenching with 125 mM glycine, cells were pelleted and washed twice with PBS containing 10 % FBS. The flash frozen pellets were lysed for 5 min on ice in 100 µl of 50 mM tris-HCl pH 8.0, 10 mM EDTA, 0.5 % SDS and protease inhibitor cocktail. Samples were sonicated using the Covaris E220 to shear chromatin to an average size of 100-500 bp (2 % duty cycle, 105 W peak incident power, 200 cycles per burst, 25 min). Sheared chromatin was diluted 1:5 in IP-buffer (10 mM Tris-HCl pH 7.5, 1mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1 % SDS, 0.1 % Na-deoxycholate and 140 mM NaCl) supplemented with protease inhibitor cocktail. After centrifuging 10 min at 13,000g at 4 °C, the supernatant containing soluble chromatin fraction was recovered and incubated with 0.7 mg/ml RNase A (ThermoFisher) for 30 min at 37 °C. For chromatin immunoprecipitation 10 µl protein-G magnetic beads (Pierce™ ThermoFisher Scientific) were washed in IP-buffer and incubated with 1 µg Anti-FLAG Ab (Sigma-Aldrich) for 1 h at 4 °C on a rotating wheel. 50 µl of RNA digested chromatin were incubated with 250 ng BG4 Ab (or without for the mock negative control) for 1h at 16 °C. The anti-FLAG coated beads were washed with IP-buffer and incubated with chromatin-BG4 complex for 3 h at 4 °C on a rotating wheel. Beads were washed 4 times with IP-buffer and once in wash buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA). Elution of immonoprecipitates and chromatin crosslink reversal were performed incubating beads with 70 µl elution buffer (10 mM Tris-HCl pH 8.0, 5 mM EDTA, 300 mM NaCl and 0.5 % SDS) containing 0.3 mg/ml RNase A (Thermo Fisher) for 30 min at 37 °C followed by the addition of 0.5 mg/ml proteinase K for 1h at 55 °C and 8 h at 65 °C shaking. Supernatant was then recovered and incubated for 1 additional h at 55 °C in the presence of 0.25 mg/ml proteinase K (Therno Fisher). The eluate was finally purified with SPRI AMPure XP beads (Thermo Fisher). For each technical replicate, eluted DNA from two ChIP reactions were combined and the pool subjected to Nextera library preparation as described by the manufacturer (NEBNext Ultra II DNA library Prep Kit for Illumina, NEB). The quality and size of libraries and chromatin shearing fragments were checked by Agilent Bioanalyzer using Agilent DNA High Sensitivity Chips (Agilent Technologies). Samples were sequenced on an Illumina HiSeq 1500 platform in single-end using 50-bp reads. The experiment was repeated twice.

3.8 G-quadruplex ChIP–qPCR

Purified and sonicated DNA (as described above) was used to quantify G4 enrichment via qPCR, using Fast SYBR PCR mix (Applied Biosystems), with a LightCycler 480 (Roche) quantitative PCR machine. Cycling conditions were 95 °C for 20 s followed by 50 cycles of 3 s at 95 °C and 30 s at 60 °C. We employed primer pairs that target G4 ChIP positive and negative regions **(Appendix 7.2)**. Relative enrichments were derived with respect to their inputs.

3.9 RNA extraction and cDNA library preparation

Total RNA for RNA–seq experiments was extracted from 80 % confluent cells. For the H-NDI-NMe_2-Pham treated cells 1 million cells were seeded in a T75 flask, grown for 24 h and treated with 1 µM compound for additional 24 h. Then, 1 million cells were trypsinized, pelleted and resuspended in 1 ml TRIreagent (Sigma Aldrich). Total RNA was extracted by phenol-chloroform and purified through the RNA Clean and Concentrator-25 kit (Zymo Research), following the manufacturer's instructions. Ribosomal RNA was depleted and purified using Ribo-Zero rRNARemoval Kit (Illumina) and RNA Clean and Concentrator-5 (Zymo Research). The quality of extracted RNA was checked by Agilent Bioanalyzer on Agilent RNA 6000 Pico Chips (Agilent technologies) both before and after rRNA depletion. RNA-seq libraries were generated using the NEBNext Ultra Directional RNA Library Prep kit for Illumina (NEB). Agilent DNA High Sensitivity Chips (Agilent Technologies) were used to check library size and quality. Samples were sequenced on an Illumina HiSeq 1500 platform in single-end using 50-bp reads. The experiment was done in three independent biological replicates.

3.10 Data analysis

Raw FASTQ reads were trimmed to remove adaptor contamination and aligned to the human reference genome version hg38 using Bowtie1 (http://bowtie-bio.sourceforge.net/index.shtml). Reads with more than 2 mismatches and multimapped reads were excluded from further analysis. G4–ChIP peaks were mapped using HOMER http://homer.ucsd.edu/homer/index.html) considering only peaks with at least 2 fold more normalized tags count in the target experiment with respect to the input (used as control), disabled local tag count and poisson p-value threshold of 0.0001. HOMER was also used to

associate peaks with the nearby gene, determine the genomic annotation of the region occupied by the peak and merge peaks from replicates.

RNA–seq reads were aligned to the human reference genome with TopHat and filtered by using samtools[125] to remove alignments with quality lower than 20, not primary alignments and PCR duplicates.

Gene expression levels were quantified as transcripts per million (TPM). Genes differentially expressed between: i) 93T449 cells and HaCaT cells and (s-value < 0.1; fold change > 1.0); ii) 93T449 cells treated with H-NDI-NMe_2-Pham and untreated 93T449 cells (s-value < 0.25; fold change > 0.5); were identified using the Bioconductor package DESeq2 and 'apeglm' for LFC shrinkage[126].

3.11 Cytotoxicity assay

Cytotoxicity of H-NDI-NMe_2PhAm, BRACO-19 and $PyPhen_2PropNMe_2$ in 93T449 cells was determined by MTT assay. 7500 cells/well were plated in a 96-wells plate and grown for 24 h. Serial dilutions of c-exNDI (0.7–12.8 μM) were dispensed to the cells in triplicate. After 24 h from treatment, cells were supplemented with freshly diluted 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (Sigma-Aldrich, Milan, Itlay) solution (5 mg/mL) and incubated O/N at 37 °C. MTT crystals were solubilized with the addition of 100 μl solubilization solution (10 % SDS and 0.01 M HCl) for 8 h at 37 °C and absorbance was measured by Sunrise Tecan plate reader (Mannendorf, Switzerland) at 620 nm. The percentage of cell survival was calculated as follows: cell survival = (Awell − Ablank)/(Acontrol − Ablank) × 100, where blank denotes the medium without cells and control the untreated cells. The 50% cytotoxic concentration (CC50) was defined as the concentration of compound required to reduce cell grown by 50

3.12 Electrophoretic Mobility Shift Assay

EMSA was used to detect G4-NCL interaction and the stoichiometry of G4 folding. The human recomabinant NCL was purchased from Origene Technologies, Rockville, USA. Oligonucleotides were 5′-end labeled with [γ-32P]ATP by T4 polynucleotide kinase for 30 min at 37 °C. After ethanol precipitation, DNA labeled species were resuspended in lithium cacodylate buffer (10 mM, pH 7.4, KCl 100 mM or NaCl 100 mM and 20% (w/v) PEG200 where indicated), heat denatured for 5 min at 95 °C and gradually cooled at room temperature

to achieve proper G4 folding. For the determination of G4 stoichiometry samples were run in 20% non-denaturing polyacrylamide gels (acrylamide/bis solution 29:1) for ~24 h at 60 V. Gels were fresh impressed and visualized by phosphorimaging (Typhoon FLA 9000, GE Healthcare Europe, Milan, Italy). For NCL binding to DNA, reactionswere performed in binding buffer (Tris–HCl 20 mM pH 8, KCl 30 mM, MgCl2 1.5 nM, DTT 1 mM, glycerol 8%, protease inhibitor cocktail – Sigma Aldrich, Milan, Italy – 1%, NaF 5mM,Na3VO4 1mM), labeled oligonucleotides (40 nM), human recombinant NCL (300 ng) and incubated for 1 h at 37 °C. Binding reactions were supplemented with 20% (w/v) PEG200 where indicated. To load the same amount of DNA in each reaction, counts per second emitted by the radiolabeled oligonucleotides were measured by a Geiger counter (Mini 900 Scintillation Monitor, Thermo Scientific™, Monza, Italy). Samples were loaded on 8% non-denaturing polyacrylamide gels for ~15 h at 40 V. Gels were dried using a gel dryer (Bio Rad Laboratpries, Milan, Italy) and visualized by phosphorimaging (Typhoon FLA 9000, GE Healthcare Europe,Milan, Italy). Quantification of bound DNA molecules was performed by ImageQuant TL Software (GE Healthcare Europe, Milan, Italy). Each gel was performed twice to confirm quantification results; only results consistent throughout the repeated gels were retained and reported.

3.13 Surface Plasmon Resonance (SPR)

SPR biosensor analysis was conducted on a Biacore T100 platform with dextran-functionalized CM5 Series S sensor chip (GE Healthcare, Life Science, Milan, Italy). Amine Coupling kit and Mouse Antibody Capture kit (GE Healthcare, Life Science, Uppsala, Sweden) were respectively exploited for standard amine coupling and the two capturing mediated immobilization strategies.

SPR technology was used to address different objectives in the present work. Given the diversity of the used conditions and for clarity reasons we described below the different used applications separately.

3.13.1 SPR for the measurement of AS1411 affinity

The human recombinant protein NCL (OriGene Technologies Inc., Rockville, MD) and the HIV-1_{IIIB} recombinant gp120 (ImmunoDiagnostics Inc., Woburn, MA) were immobilised by amine coupling in HEPES-NaCl running buffer (10 mM HEPES pH 7.4,150 mM NaCl, 3 mM EDTA). NCL and gp120 were diluted in sodium acetate buffer (GE Healthcare, Milan, Italy) at different pHs (4.0, 4.5, 5.0, 5.5) and a pH-scouting procedure was performed as described by the manufacturer to asses the better pH for immobilization to the dextran chip-surface. Based on the preconcentration curves NCL and gp120 were then diluted in sodium acetate buffer pH 4.0 at 15 ng/µL and 20 ng/µL, respectively, and were injected to reach 1500 resonance units (RU). The flow cell 1 was left free for reference subtraction. Binding analysis was performed at 20 µL/min with contact and dissociation times of 120 s in HEPES-KCl buffer (10 mM HEPES pH 7.4, 200 mM KCl, 3 mM EDTA).Test oligonucleotides **(Appendix 7.2)** were diluted to 1 µM in HEPES-KCl buffer, denatured at 95 °C for 5 min and gradually cooled at room temperature to allow secondary structures folding. Sensorgrams were obtained in the concentration range of 15.6–1000 nM. The chip surface was regenerated with 1 M KCl solution. Kinetic analysis was performed through the BIAevaluation software (GE Healthcare). Double correction for blank (running buffer injection) and reference flow cell was performed and the corrected sensorgrams fitted through a 1:1 binding model.

3.13.2 SPR for the optimization of 1H6 antibody immobilization

Covalent immobilization through Standard Amine Coupling of 1H6 antibody

Covalent amine coupling immobilization of the mouse monoclonal 1H6 antibody (kindly provided by P. M. Lansdorp)[119] was performed according to the manufacturer use. The flow cell was activated by injection of 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and N-hydroxysuccinimide (NHS) mixture. After pH-scouting at different pHs (4.0, 4.5, 5.0, 5.5), 1H6 (MW ~155 kDa) was diluted to 15 ng/µl in sodium acetate buffer pH 5.0 and coupled to the active flow cell surface via exposed primary amines to reach an immobilization level ranging between 1000-1200 RU. Finally, ethanolamine was injected to block the unreacted ester groups. A control flow cell was activated and blocked without ligand injection to allow reference subtraction. HEPES-NaCl (HEPES 10 mM pH 7.4, NaCl 150 mM, EDTA 3 mM) was used as immobilization buffer. To allow the correct folding of secondary structures, each

oligonucleotide was diluted to 20 μM in HEPES-KCl buffer (HEPES 10 mM pH 7.4, KCl 200 mM, EDTA 3 mM), denatured fro 5 minutes at 95 °C and then slowly cooled to RT.

Mouse Antibody Capture kit mediated immobilization of 1H6 antibody

Apolyclonal anti-mouse IgG was used to mediate 1H6 coupling the the sensor chip surface. Immobilization of the anti-mouse IgG (GE Healthcare, Life Science, Uppsala, Sweden) was performed by amine coupling in HEPES-NaCl buffer at a flow rate of 5 μl/min both on the reference and active flow cell. The EDC/NHS mixture was injected for 420 sec to activate the chip surface. The anti-mouse IgG diluted to 40 ng/μl in HEPES-NaCl buffer was injected for 600 sec and the unreacted ester groups were then blocked by 420 sec injection of ethanolamine. Finally, the unbound ligand was washed away through three consecutive injections of 10 mM glycine, pH 1.7 (60 sec, 50 μl/min). Immobilization levels between 3000 and 6000 RU were typically obtained and then primed with HEPES-KCl buffer. 1H6 was diluted to 15 ng/μl in the same buffer and captured by 600 sec injection at a flow rate of 5 μl/min on the active anti-mouse IgG immobilized flow cell. Finally, 1 M KCl was injected (20 sec, 30 μl/min flow rate) to dissociate the non-specifically bound ligand. Typical capturing levels ranged between 800-1200 RU.

G-quadruplex complex-mediated immobilization of 1H6 antibody

The anti-mouse IgG was immobilized on reference and active flow cells as described in the section above. After 1 h incubation at RT in the presence of 10 folds of G4 folded oligonucleotide (*Oxy2*), 1H6 was captured on the active flow cell. The final injection of 1 M KCl (20 sec, 30 μl/min flow rate) was used to dissociate *Oxy2* from the captured 1H6. Typical capturing levels ranged between 800-1200 RU.

Kinetic analysis and data evaluation

Each analyte was tested in the concentration range 0.250 - 8 μM. For *Oxy2*, we tested two additional low concentrations (0.125 and 0.062 μM) to cover the whole kinetic points. Samples were injected for 340 sec at a flow rate of 20 μl/min and dissociated for 340 sec. HEPES-KCl buffer was used as running buffer and for samples dilution. Surface regeneration at the end of each analyte concentration was performed by 20 sec injection of 1 M KCl at a flow rate of 30 μl/min. Complete regeneration of 1H6 for successive immobilizations was instead performed with 180 sec injection of 10 mM glycine pH 1.7 (GE Healthcare, Uppsala, Sweden). BiaEvaluation Software was used for data analysis. The likelihood of fittings was assessed

through the statistical parameters of Chi^2 and U-value[127,128]. The percentage of active surface was measured as a proportion of the maximum observed response ($R_{max}OBS$) to the theoretically expected maximum response ($R_{max}EXP$), where:

$$R_{max}EXP = (MW_A / MW_L) \; x \; (\delta n/\delta C)_A / (\delta n/\delta C)_L \; x \; RU_I \; x \; S \qquad (1)$$

$R_{max}EXP$ is the expected maximal response of the interacting molecules expressed in RU; MW_A and MW_L are the molecular weights of the analyte and ligand, respectively; RU_I is the amount of immobilized ligand expressed in RU; S is the stoichiometry of the interaction; $(\delta n/\delta C)_A$ and $(\delta n/\delta C)_L$ are the refractive index increments (RII) for the analyte and ligand, respectively. The RII value is necessary to correct the discrepancies in the expected and observed R_{max}. RII depends on the refractive index at the surface (n) and the concentration (C) of the binding partners. For protein-nucleic acids interactions, RII can be approximated to 1 and discrepancies from the expected R_{max} can be ascribed to the physico-chemical properties of the immobilized SPR chip surface[129]. Often, differences between the composition of the analyte stock buffer and the SPR running buffer used for diluting the injected analytes generate the so called bulk effect [37, 39], which appears as a sharp jump of the RU response at the very beginning and after the end of the injection cycle. To obtain an appropriate fitting of experimental curves, a standard value of 1/5 of the experimental responses is normally used to correct experimental data where a bulk effect is detected[130]. To obviate this problem, $R_{max}OBS$ has been measured 7 sec after the injection stop, approximating the analyte binding level at a kinetic point not affected by bulk shifts[131,132].

Kinetic constants were obtained from 1:1 Langmuir fitting (*un3* and *bcl*-2) or heterogeneous analyte fitting (Oxy2) of sensorgrams, applying default bulk correction, after double subtraction of buffer and reference flow cell responses. Each immobilization strategy and kinetic analysis was repeated at least three times. Baseline stability and influences of small baseline deviations on the goodness of fitting have been measured as:

$$R\varepsilon i / R_{BL} \; x \; 100 \quad \text{where} \quad R\varepsilon i = | Ri - R_{BL} | \qquad (2\text{-}3)$$

R_{BL} is the measured baseline before injection of the analyte and Ri is the baseline measurement after each regeneration.

3.13.3SPR for the determination of 1H6 affinity for HSV-1 G4s

To screen 1H6 binding and affinity for a series of viral and cellular G4s with respect to a range of non-G4 controls **(Appendix 7.2)** the G4-complex mediated immobilization procedure described above was employed. The polyclonal anti-mouse IgG (GE Healthcare, Life Science, Uppsala, Sweden) and the viral G4 oligonucleotide *un3* were used to allow the correct orientation of 1H6 binding site on the chip. Flow cell 1 was blank immobilized with anti-mouse IgG to permit reference subtraction. Sensorgrams were obtained in the concentration range of 0.250-8 μM. After each oligonucleotide injection, the chip surface was regenerated with KCl 1M solution. All sensorgrams were corrected by reference subtraction of blank flow cell response and buffer injection response. Data were fitted to a global 1:1 binding model using BIAevaluation software (GE Healthcare). The stability of 1H6 binding was evaluated by measuring RU response 20 sec before end of the dissociation phase.

4 RESULTS AND DISCUSSION

DNA G-quadruplexes in the human genome

4.1 Discovery and characterization of a G4 structure in a relevant gene for liposarcoma

The vast majority of G4s that were found in the human genome are within promoters of cancer related genes. Among the better characterized ones we can mention: *C-MYC*[43], *C-KIT*[133] and *BCL-2*[134]. The understanding of G4 functions in such genes can be of extreme utility, not only for gaining a deeper knowledge of cancer biology, but also as alternative therapeutic targets.

The present project is focused on the exploration of G4s in liposarcoma (LPS), a class of soft tissue sarcomas which is particularly resistant to the classical chemotherapies and is characterized by a high mortality rate, especially when occurring in the retroperitoneal cavity[76]. The primary and most frequent form into which LPS can manifest is the well-differentiated liposarcoma (WD-LPS), characterized by the amplification of chromosome 12q13-15 portion, that comprises the proto-oncogenes *MDM2*, *CDK4* and *HMGA2*[135,136]. *MDM2* is a key gene in the regulation of apoptosis, being responsible for the inactivation of p53[72]. There are two different promoters regulating *MDM2* transcription: the first (P1) is constitutive and responsible for the basal expression of the gene, while the second (P2) is inducible, allows a faster processing of transcripts and is the main promoter responsible for *MDM2* overexpression in different cancers[73]. For this reason, we decided to investigate for the presence of putative G4s in the first intron of *MDM2*, containing the P2 promoter.

4.1.1 Computational prediction of G4s in the *MDM2* inducible promoter

Putative G4 forming regions were analyzed in the G-rich P2 promoter-containing the intron of *MDM2* (GRCh38.p2, NC_000012.12) by using QGRS. QGRS is an online computational tool which predicts and maps putative G4 sequences. A score (G-score) is associated to each identified sequence reflecting the likelihood to form a stable G4. We performed G4 prediction

on both forward and reverse strand of the selected genomic region finding a total of 11 putative G4 forming sequences of which 10 on the forward strand and 1 on the revers strand (**Table 4-1**).

Interestingly, the putative G4 with the highest score (Mdm119) is located immediately adjacent to an E-box DNA motif (CACGTG). This type of element is known to work as enhancer for the recruitment of transcription factors and thus initiate gene transcription[137]. Moreover, Mdm119 is contained in the so called $(nnGGGGC)_5$ repeated sequence, whose integrity has been reported to be necessary for p53-independent activation of the P2 promoter[79].

Table 4-1. ***Putative G4 sequences identified by QGRS-Mapper.***

Position	Length	Putative G4 sequence	G-Score	Strand
21	28	**GGTCACTTTTGGGTCTGGGCTCTGACGG**	28	F
100	16	**GGTTCGTGGCTGGGGG**	27	F
119	25	**GGGGCGCGGGGCGCGGGGCATGGGG**	93	F
160	23	**GGTTTTGTTGGACTGGGGCTAGG**	28	F
195	14	**GGGAGGAGGGCGGG**	31	F
214	21	**GGACGGCTCTCGCGGCGGTGG**	26	F
236	11	**GGTGGGGGTGG**	31	F
262	34	**GGGAGTTCAGGGTAAAGGTCACGGGGGCCGGGGG**	58	F
311	16	**GGCGCGGGAGGTCCGG**	30	F
377	26	**GGGCGGGATTGGGCCGGTTCAGTGGG**	53	F
434	39	**GGCTGCGAACGGGCAGAGGCTGGGAACCAGCGATAGAGG**	27	R

The indicated position corresponds to distance from the first base of MDM2 intron 1. Gs predicted to be involved in G4 formation are underlined.

The putative G4 Mdm262 comprises a Sp1 binding site and SNP285G>C locus (rs117039649), which forms a particular haplotype together with SNP309T>G (rs2279744). The C variant of SNP285C has been shown to reduce the strength of Sp1 binding and thus counteract the effect of SNP309G, the latter instead extended a second Sp1 binding site leading to increased transcription of *MDM2*[138,139]. A high-score putative G4 structure (Mdm311) has been predicted to overlap with this second Sp1 binding site and an ETS response element, suggesting a potential G4-mediated regulatory mechanism. The only predicted G4 on the reverse strand

(Mdm434) is also located in a key region, containing one of the two p53 response elements of *MDM2* P2 promoter[78]. In summary, 11 putative G4 forming sequences have been identified in the inducible promoter of *MDM2* through a computational algorithm. Of these sequences, at least 4 (Mdm119, Mdm262, Mdm311 and Mdm434) are located in or immediately adjacent to important transcription factor binding sites (**Figure 4-1**). This observation suggests a potential role of the predicted G4s in the regulation of *MDM2* P2 promoter activation, but further investigation is necessary to assess this statement: first, the actual ability of the putative G4s has to be confirmed.

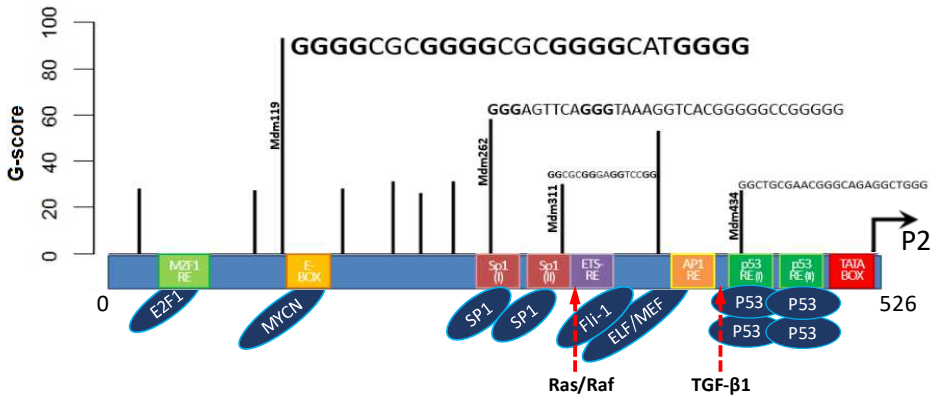

Figure 4-1. ***Position and score of the putative G4s predicted by QGRS-Mapper.***

The principal transcription factor biding sites and regulatory regions in the P2 promoter are reported, as well as four putative G4 sequences located in key regions for the transcriptional modulation of the promoter. Numbers at left and right of the promoter scheme are the bp count of P2 promoter sequence.

4.1.2 Biophysical characterization of the most probable *MDM2* G4

We choose to initially focus on the predicted putative G4 with the highest G-score (Mdm119) for assessing its actual folding into a G4 conformation. Mdm119 is embedded in the $(nnGGGGC)_5$ repeated sequence that can theoretically fold into a dynamic G4. Therefore, we choose to extend the QGRS predicted sequence to comprise the entire five tandem G-tracts, together with few surrounding bases, in order to represent a more physiological situation. The exact nucleotidic sequence of the selected region is reported in **Appendix 7.2** and from now on is called *Mdm2-G4* for the sake of simplicity.

CD spectroscopy was applied to analyze the selected sequence in the presence of 0 mM, 20 mM and 100 mM KCl, since monovalent cations and in particular K^+ are known to stabilize G4s[140]. Each of the recorded spectra corresponded to a typical G4 topology, with increasing stability in the presence of higher K^+ concentration **(Table 4-2 and Figure 4-2)**. Notably, in the absence of K^+ ions, *Mdm2-G4* adopted a hybrid parallel/antiparallel conformation with two positive peaks around 260 nm and 290 nm and a negative peak at 240 nm. By adding K^+ the antiparallel conformation prevails, showing two positive peaks at 290 nm and 240 nm and a negative peak at 260 nm. The antiparallel conformation is the rarest folding assumed by known G4s, making *Mdm2-G4* a good candidate for selective drug targeting.

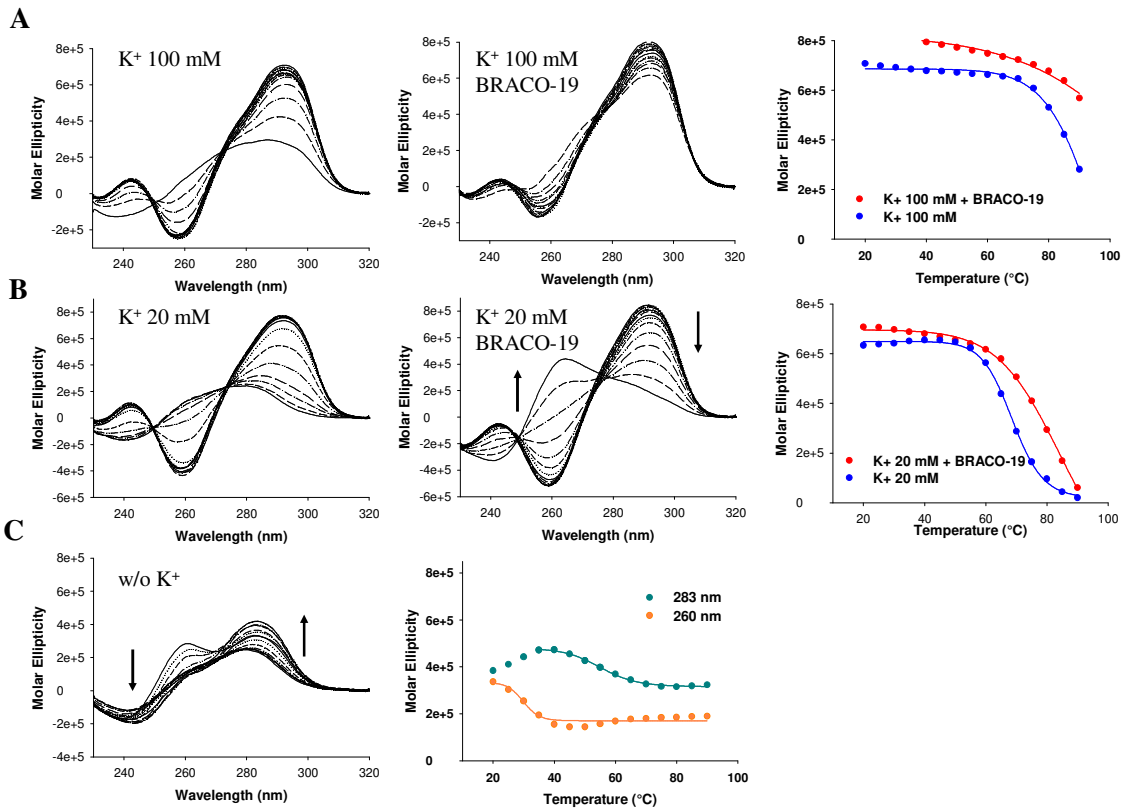

Figure 4-2. ***CD characterization of Mdm2-G4 folding and thermal stability.***

A) *In the presence of K^+ 100 mM and BRACO-19.* ***B)*** *In the presence of K^+ 20 mM and BRACO-19.* ***C)*** *in the absence of K^+.*

A further proof of the ability of the *Mdm2-G4* sequence to fold into a G4 is given by the G4-ligand dependent stabilization of the latter. Several compounds have been demonstrated to stabilize or induce G4 folding. Here we selected the trisubstituted acridine BRACO-19[141] **(Figure 1-9)** to test its effect on *Mdm2-G4* in the presence of 20 mM and 100 mM K^+. In both

conditions a stabilization of the antiparallel G4 conformation was observed (Figure 4-2) but only for 20 mM K^+ the ΔTm could be calculated, since K^+ 100 mM alone was sufficient to induce a G4 stability higher than 90 °C, which is the upper limit of the instrument sensitivity. In the presence of 20 mM K^+, instead, BRACO-19 generated a stabilization of 17.0 ± 2.7 °C (**Table 4-2**).

Table 4-2. ***CD topological and thermal analysis of Mdm2-G4 oligonucleotide.***

K^+ concentration	Drug added	Tm (°C)	ΔTm (°C)	Topology
0 mM	/	29.8 ± 1.3 / 55.1 ± 0.6	/	Hybrid
20 mM	/	68.7± 0.3	/	Antiparallel
20 mM	BRACO-19	85.7± 2.4	17.0 ± 2.7	Antiparallel
100 mM	/	> 90	/	Antiparallel
100 mM	BRACO-19	> 90	/	Antiparallel

Mdm2-G4 melting temperatures in the presence or absence of the G4-ligand BRACO-19 are reported at different K^+ concentrations. When possible the stabilizing effect of the compound is indicated.

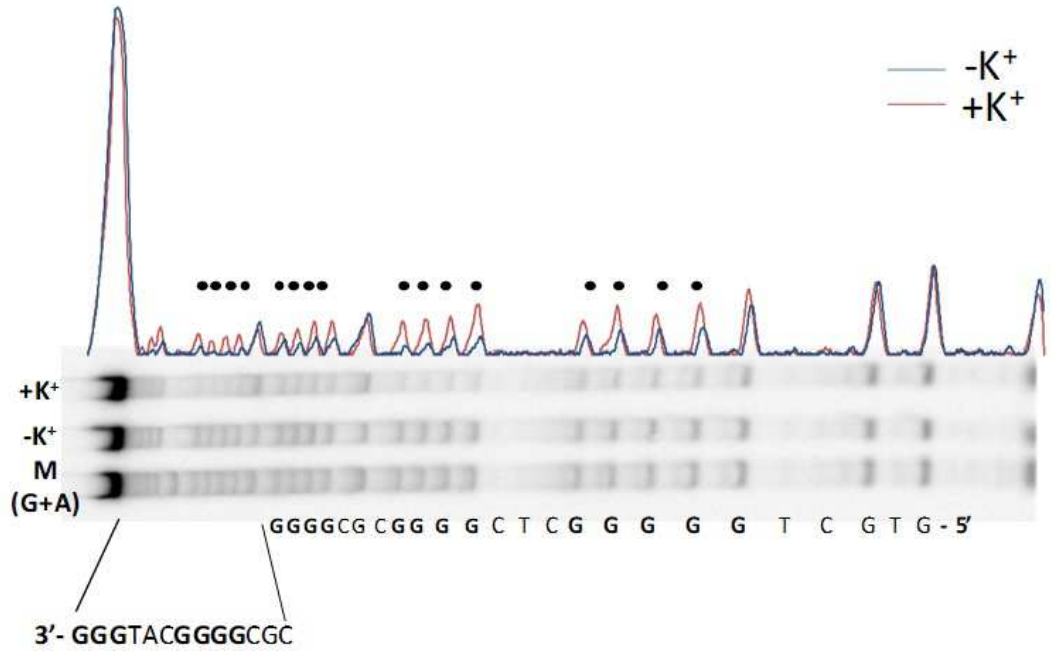

Figure 4-3. ***DMS protection assay.***

The M (G+A) lane corresponds to the Mdm2-G4 sequence cleaved in correspondence of G and A after treatment with formic acid and hot piperidine. DMS and formic acid has been instead used to treat Mdm2-G4 and identified protected guanines in the presence or absence of K^+ 100

mM. Protected guanines are indicated with black dots. The visible oligonucleotide sequence is reported at the bottom of the figure.

A protection assay was next performed to identify the guanines involved in the G4 structure. *Mdm2-G4* exceeds indeed the minimum number of guanines required for the formation of an intramolecular G4, therefore more equilibria with a predominant species are expected. DMS has been exploited to this aim, as alkylator of nitrogen in position 7 of guanines. *Mdm2-G4* folded in the presence or absence of K^+ 100 mM were compared. While both conditions allowed formation of a G4 topology, the absence of K^+ ions led to a different and significantly less stable conformation, therefore a stronger G protection was expected in the presence of K^+ 100 mM. Indeed, each guanine in the three internal G-tracts and four guanines in the first G-tract (5'-3' direction) in the *Mdm2-G4* sequence appeared protected in DMS assay (**Figure 4-3**). Less easy is the interpretation of protection of the fifth G-tract, for the evaluation of which an NMR investigation would be required. Nevertheless, this result indicates that the studied G4 forms four stacked quartets, excluding the first guanine in the first G-tract, thus favoring a shorter loop.

Together these results confirm the ability of *Mdm2-G4*, derived from the computationally predicted Mdm119 putative G4, to fold into an antiparallel topology with four stacked quartets and to be greatly stabilized by the interaction with a G4-ligand. The high stability (> 90°C) of *Mdm2-G4* structure at physiological K^+ concentration is an indication of its potential ability to fold also in a cellular environment and to play a biological role in the regulation of *MDM2* gene.

4.1.3 *MDM2* G4 interaction with nuclear proteins

The most common way through which G4s work as gene regulators is by the interaction with specific proteins. Several classes of proteins are reported to bind to DNA and RNA G4s with opposing effects. On one hand they can promote the formation or stabilize tetraplex structures (e.g. NCL, reported to bind to c-Myc promoter and HIV-1 LTR[44,93] and HOP1, involved in chromosome recombination[142]). On the other hand, they can destabilize and catalyze G4 unwinding (e.g. the protein families of helicases and hnRNPs[143–145]).

MDM2 protein is known to be overexpressed in human LPS. Published results demonstrate that *MDM2* transcripts derived from the P2 promoter activation have a higher translational potential

with respect to transcript derived from the constitutive promoter P1[73]. It would be therefore of interest to investigate if *Mdm2-G4* is a regulator of P2 dependent transcription and which are the proteins that interact with it.

To this aim we performed a preliminary analysis searching for proteins that recognized the *Mdm2-G4* structure in a selective way. In particular, we used nuclear protein extracts from the WD-LPS cell line 93T449 to perform a magnetic beads pull-down assay followed by MS identification. To focus on proteins which specifically binds to *MDM2* G4 folding, we used several controls: a single stranded non-G4 G-rich sequence (*G-ss*), a single stranded C-rich (*C-ss*) and beads alone. Among the pulled-down proteins displaying specificity for *Mdm2-G4*, we detected an enrichment in helicase and unwinding proteins (**Table 4-3**). The protein with the highest score was hnRNP A3: this protein was already reported to recognize single stranded telomeric DNA, acting as negative regulator of telomere length maintenance and suggesting a G4 mediated function[146]. Interestingly, hnRNP A3 was found to be sequestered in intranuclear inclusion by accumulated C9orf72 RNA containing the G4-forming hexanucleotide GGGGCC in ALS and FTD patients[147]. The studied *Mdm2-G4* is actually constituted by a very similar sequence $(GGGGC)_5$, supporting the specificity of this protein for G4s with this peculiar GC-composition.

More in general, hnRNPs are a family of ribonucleoproteins, many of which have been reported in association with G4s, with a destabilizing activity[148,149]. Beside hnRNP A3, we found three other proteins belonging to this family and complexed with *Mdm2-G4:* hnRNP F, hnRNP Q and hnRNP H3. While several studies investigated the function of hnRNP F-G4 complex[150–152], less is known about hnRNP Q and H3.

Besides from hnRNPs, other relevant proteins were pulled-down by *Mdm2-G4*. Ku70 and Ku80 proteins, respectively encoded by *XRCC6* and *XRCC5* genes, form a dimer and together are part of the DNA helicase II complex, involved in double-strand breaks repair non-homologous end joining (NHEJ) and V(D)J recombination[153–156]. Ku70 was already reported to interact with the G4s of *ERBB2* proto-oncogene[157], *KRAS*[158] and telomeric G4[159] in different types of human cancers. Its partner Ku80 was not always detected in association with the G4, suggesting that Ku70 is the factor that mediates the binding to DNA. This hypothesis is also supported by the higher score and sequence coverage of Ku70 (XRCC6) in the reported experiment (**Table 4-3**).

Another protein with helicase function that has been found in association with *Mdm2-G4* is DDX3X. It is an RNA helicase that can also unwind partially single stranded DNA portions and is involved in different steps of gene expression, among which transcription, with either

enhancing or silencing effect, depending on the target. When involved in translation, DDX3X was found to specifically recognize complex secondary structures that require helicase activity to be processed (https://www.genecards.org/cgi-bin/carddisp.pl?gene=DDX3X). Two DDX3X closely related helicases (DDX5 and DDX17) were reported to interact with GC-rich sequences with the ability to fold into G4s in cooperation with hnRNP H/F and work as slicing regulators[160]. Interestingly, hnRNP F was also detected in our experiment.

While the proteins described until now prevalently share a G4-destabilizing activity, we also identified a protein with G4 stabilizing effect: EWS is involved in various cellular processes among which transcriptional activation. Interestingly, EWS shares with several hnRNPs the RGG DNA-binding domain, which was demonstrated to recognize G4 structures with strong specificity[161,162].

Lastly, MYH9, involved in maintenance of platelet structure and cytolytic granule exocytosis, was previously reported to recognize the *C-MYC* promoter G4, but the specific role of this complex was not further investigated[44,163].

Altogether, the high number of G4-helicases and destabilizing proteins that we found in association with *Mdm2-G4* suggests the importance for the cell to strictly regulate the folding of the *MDM2* G4 structure and its possible involvement in gene expression. In particular, the Ku70-Ku80 complex can be activated after DNA damage response, which is the same stress stimulus that induces *MDM2*-P2 transcription. This observation suggests a possible role of Ku70-Ku80 helicase in the activation of P2 mediated *MDM2* transcription[164]. On the other hand, EWS protein can counteract the activity of helicases and hnRNPs, stabilizing the G4. Further experiments are however necessary for consolidating these observations.

Table 4-3. ***MS identification of specific Mdm2-G4 binding proteins.***

Name	Function	Score	Seq Coverage	Also found on	Reported on G4s
hnRNP A3	Trafficking of RNA	125	6%	*C-ss* with similar score	[165–167]
ACBL	involved in various types of cell motility	123	10%	/	
XRCC6	ss DNA-dependent ATP-dependent helicase	89	12%	*G-ss*	[157–159]
XRCC5	ss DNA-dependent ATP-dependent helicase	57	4%	*G-ss*	[164]

hnRNP F	Binds G-rich sequences in pre-mRNAs keeping it in an unfolded state	57	6%	/	150,151,168,169
DDX3X	Multifunctional ATP-dependent RNA helicase	29	2%	/	160
MYH9	cytokinesis, cell shape, secretion and capping	29	0%	/	44
EWS	transcriptional repressor	28	2%	/	26,161,162
CALX	interacts with newly synthesized glycoproteins in the ER	28	1%	/	
hnRNP Q	implicated in mRNA processing mechanisms	24	3%	*G-ss* / *C-ss*	170
hnRNP H3	early heat shock-induced splicing arrest	22	4%	/	165,171

The reported proteins are those identified with sufficient confidence based on the indicated score and which appear to bind Mdm2-G4 in a specific way. When the proteins were also found in the control oligonucleotides, the score difference was at least double, allowing to presume higher preference for the target structure. The reported proteins are ordered based on their score.

The results obtained so far allowed the discovery and characterization of a previously unreported G4 in the inducible promoter of the human proto-oncogene *MDM2*. At physiological K^+ salts concentration it folds into a highly stable antiparallel structure and dilutions of salts allow to measure a strong stabilization of the structure after incubation with a known G4 ligand. The observed structural features of *Mdm2-G4*, i.e. the uncommon antiparallel conformation and the presence of four stacked quartets, makes *MDM2* G4 an attractive therapeutic target. This observation, together with the identification of specific binding proteins with opposing effects, supports the biological implication of the studied *MDM2* G4 in the regulation of various gene processes and the possibility to therapeutically act on them with potential anticancer activity.

4.2 Genome-wide relationship between G4s and transcription in LPS

While accumulating biophysical characterization and *in vitro* studies on G4s are accumulating, more challenging is the understanding of their folding and function in a genome-wide cellular context. Computational algorithms and a G4-seq DNA polymerase stalling experiment gave some preliminary results about the genome-wide distribution of over 700,000 G4 forming

sequences[172,173]. However, to add biological significance to these results, there is the pressing need to generate explicit *in vivo* data. In the last years, evidence of *in vivo* G4s started to be consolidated. The main used approaches are fluorescent microscopy G4 visualization, mediated by G4-specific probes or antibodies[119,121,174–176]. Chemical probes were reported to have the ability to stabilize or induce the folding of G4s manipulating their natural state, whereas antibody-based approaches have the advantage to recognize only physiologically folded G4s. The anti-G4 scFv antibody BG4 was recently used to map G4 structures in the fixed chromatin of normal (NHEK) and immortalized human epidermal keratinocytes (HaCaT)[55]. Surprisingly, a striking difference in the number of detected G4s was observed, supporting a cell specific modulation of G4 folding.

In the present study we applied BG4 ChIP-seq to map G4 regions in the LPS cell line 93T449. We then combined these results with RNA-seq data to investigate the G4 role in the cell-specific regulation of gene transcription.

4.2.1 Mapping of DNA G4s in LPS through BG4 ChIP-seq

The FLAG-tagged single chain variable fragment (scFv) antibody BG4 was previously employed to map G4 structures genome-wide in the chromatin of the pre-oncogenenic immortalized keratinocytes cell line HaCaT and its normal counterpart (NHEK)[55,177]. We adapted the published ChIP-seq protocol to the LPS cell line 93T449 in order to identify the cell-type specific G4s. Prior to the sequencing, ChIP-qPCR was used to assess the enrichment of G4 positive regions with respect to the total sheared chromatin (input) **(Figure 4-4 A)**. The chosen G4 positive control regions were conserved genes previously reported to be recognized by the BG4 antibody. *MDM2* and *CDK4* are LPS-related genes, the first of which folds *in vitro* into a G4 structure, as demonstrated in the present study. Negative controls were regions never reported to adopt a G4 structure and were not detected by BG4-ChIP or in the published cell lines.

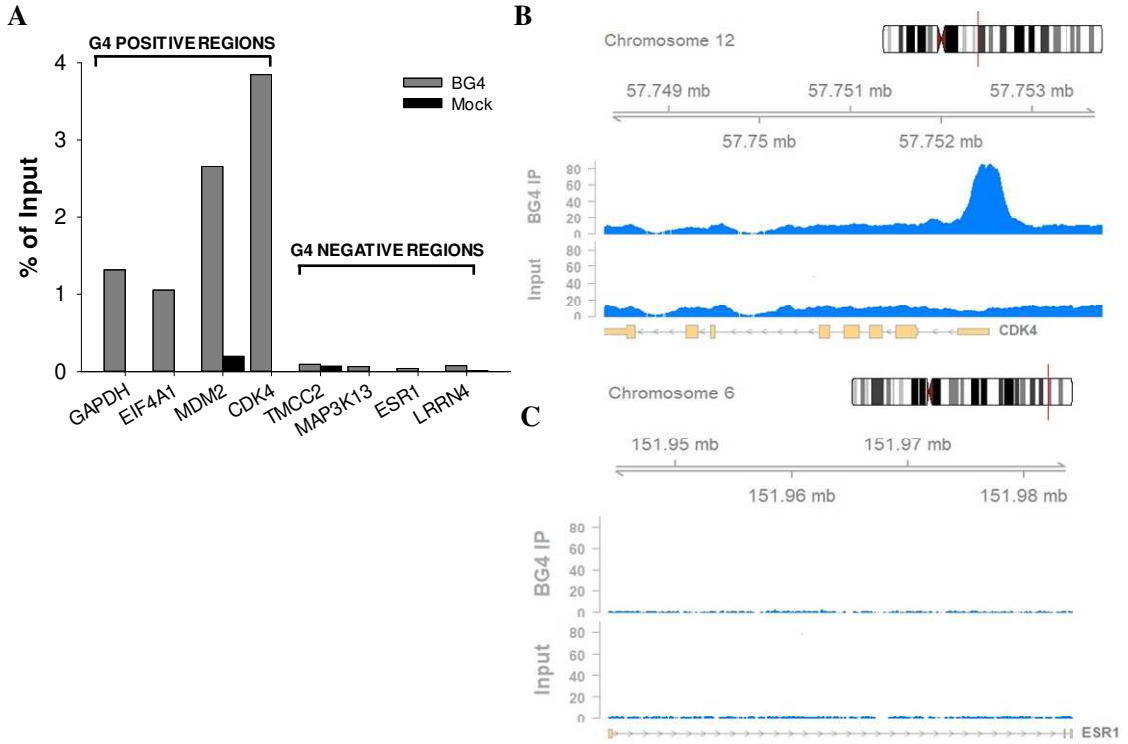

Figure 4-4. ***BG4 chromatin immunoprecipitation of G4 structures.***

A) *ChIP-qPCR of four G4 positive and four G4 negative regions. Displayed values are calculated from the Cp of the IP over Input, normalized for the Input dilution factor and reported as percentages of enrichment.* ***B)*** *ChIP-seq alignments of a G4 positive example region (CDK4 gene). BG4-IP and Input control are shown. C) ChIP-seq alignments of a G4 negative example region (ESR1 gene). BG4-IP and Input control are shown.*

ChIP-qPCR results **(Figure 4-4 A)** showed a clear enrichment in G4 positive regions with respect to the negative controls. Sequencing of the immunoprecipitated chromatin revealed about 4800 G4 peaks. Example regions are reported **Figure 4-4 B and C**.The nonrandom distribution all over the genome of the observed peaks is a first indication of a biological control of G4 formation **(Figure 4-5 A)**. To deeper explore the genomic position of G4s we applied HOMER tool (http://homer.ucsd.edu/homer/ngs/annotation.html) to annotate the peaks underneath's region. Interestingly, we found a striking enrichment of G4s in promoter TSS regions (79 %) with respect to the others functional genomic entities: transcription termination site (TTS), 5' UTR, 3' UTR, introns, exons, intergenic and non-coding portions. This result confirms what previously observed through computational predictions of G4s in the human genome[178,179] and supports the involvement of G4 structures in the regulation of gene transcription at the genome-wide level **(Figure 4-5 B)**.

The relation between G4s and transcription has been already explored in *in vitro* systems of isolated genes and mainly through gene reporter assays. In most cases ligand-mediated stabilization of the G4 leads to the repression of the transcriptional output, while less coherent are the effects of the G4 structure disruption by point mutations: sometimes improving, sometimes repressing transcription[180–184]. A clear picture of how G4s participate in transcription regulation is therefore still missing.

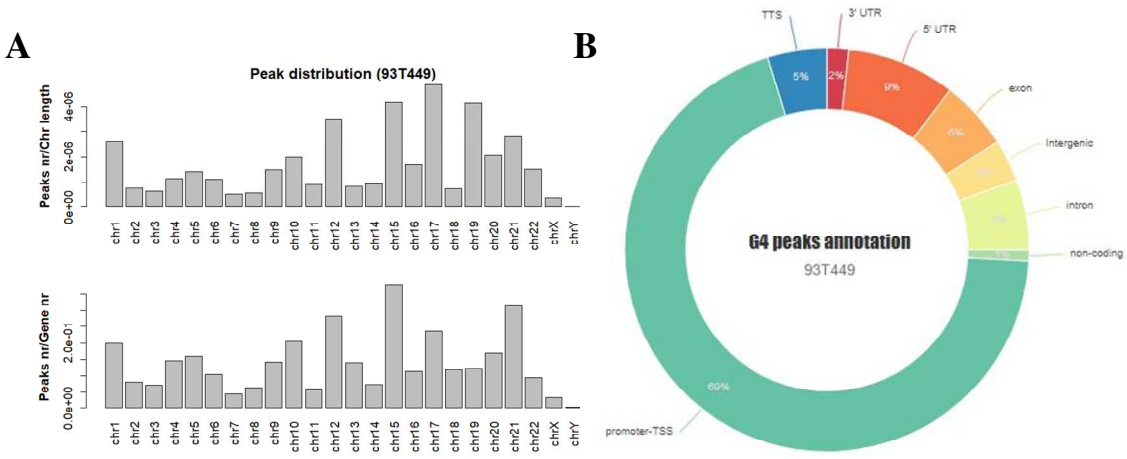

Figure 4-5. ***LPS G4s position and annotation.***

A) *G4 abundance in human chromosomes normalized for chromosome length (upper) and gene number (lower).* ***B)*** *G4 peaks annotation. The relative abundance of G4s in each category is normalized for the genomic percentage of the corresponding functional region.*

4.2.2 Transcriptional state of G4-containing genes

In the attempt to better define G4 structure involvement in transcription regulation, we combined the G4-ChIP-seq data with gene expression data measured by RNAseq. The G4-containig genes were selected from the ChIP-seq results and their transcriptional state evaluated from two points of view: first, if the gene is transcriptionally active or silent, and second, to which level the gene is expressed.

To address the first point, we established a gene expression threshold to 1 transcript per million (TPM) and calculated the percentage of expressed genes for each category, corresponding to the genomic position of the G4 inside of the gene (promoter-TSS, TTS, 5' UTR, 3' UTR,

introns, exons, intergenic and non-coding) **(Figure 4-6 A)**. Interestingly, independently from the G4 position, the probability to be expressed was always higher for G4-containing genes compared to G4-depleted genes (no G4). In particular, genes with G4s in promoters and in 5'-UTRs (regions that are often overlapping in the pre-transcribed genome) displayed a strikingly higher percentage of actively expressed genes (> 90%), corresponding to the almost totality of the genes in those categories. On the contrary, less than 50 % of the G4-depleted genes were expressed. Recently, G4s have been proposed as proper epigenetic marks influencing chromatin packaging and DNA epigenetic modification[57]. In the present work, we reported for the first time G4 structures as positive DNA marks for active transcription.

In addition to the on/off transcriptional state of G4-containing genes, we were also interested in exploring how the G4 state was related to the level of gene expression. To address this second point, we divided the expressed genes (TPM > 1) into three categories (low, medium and highly expressed genes) on the basis of the quantile expression distribution of all genes in the LPS cell line. In particular, the lower quantile was associated to low expression, the mid quantile to medium expression and the upper one to high expression **(Figure 4-6 B)**. The percentage of G4-containing genes in each category was evaluated and further compared to the percentage of G4s in non-expressed genes. An additional parameter that we introduced in this analysis was the effect of the G4 position: we firstly considered only genes with G4s within 1 kbp from the TSS, so mainly inside of the promoter region; then G4s comprised between 1-15 kbp from the TSS, representing the mean length of a human gene; finally, G4s farther than 15 kbp, so mainly in intergenic regions **(Figure 4-6 B)**. As shown in Figure 4-6 C, we could unambiguously observe a positive correlation between the presence of G4s in gene promoters and high transcriptional level. The observed trend was however no more visible when moving farther from the promoter region. This interesting result supports the involvement of G4s in the enhancement of transcription of the corresponding gene only when they are formed in the promoter. G4s in different genomic position are instead probably involved in different regulatory functions, such as recombination or splicing[185,186] but do not significantly affect transcription.

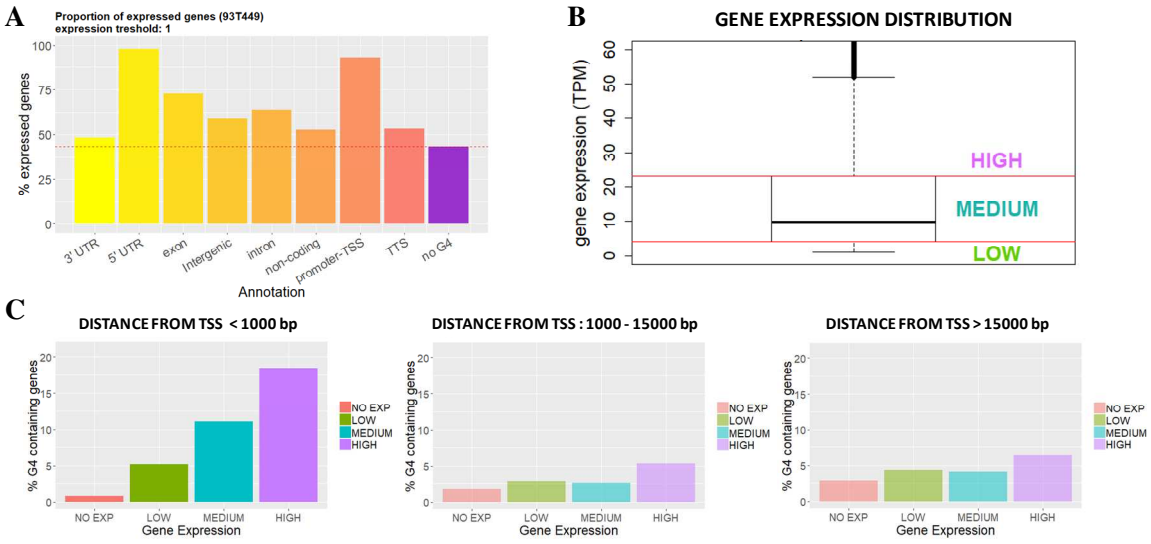

Figure 4-6. ***Transcriptional state of G4-containing genes:***

A) *Proportion of expressed genes divided by the genomic position of G4s and compared to G4 depleted genes.* ***B)*** *Subdivision of expressed genes in the LPS cell line 93T449 into low, medium and highly expressed on the basis of the quantile gene expression distribution.* ***C)*** *Percentage of G4-containing genes in relation with the expression level of the corresponding gene. Difference sitances from the TSS of the nearest gene were evaluated as indicated in the three panels.*

4.2.3 Effect of G4 genomic position on the transcriptional outcome

In the previous paragraph we have shown that G4s promote active transcription of the corresponding gene and that high transcriptional levels are associated to the presence of G4 in the promoter. To confirm the dependency of the transcriptional output from the position of the G4 in the different genomic functional entities, we looked in more details at the transcriptional intensity of LPS genes subdivided for the G4 location in the corresponding gene **(Figure 4-7 A)**. In line with what observed before, a significantly higher expression level was visible for genes with G4s in their promoter or 5' UTR, while less defined was the effect of G4s in other functional regions. 5' UTR annotated regions correspond to those genomic portions that are immediately upstream from the initiation codon and once transcribed into mRNA regulate the initiation of translation[187]. They normally span between the TSS and the AUG initiation codon, therefore overlapping with what we consider the promoter region. G4s in this position are with high probability involved in modulation of the translational potential of mRNAs. A deeper

investigation of this aspect is surely worthy of interest but goes beyond the principal aim of this work.

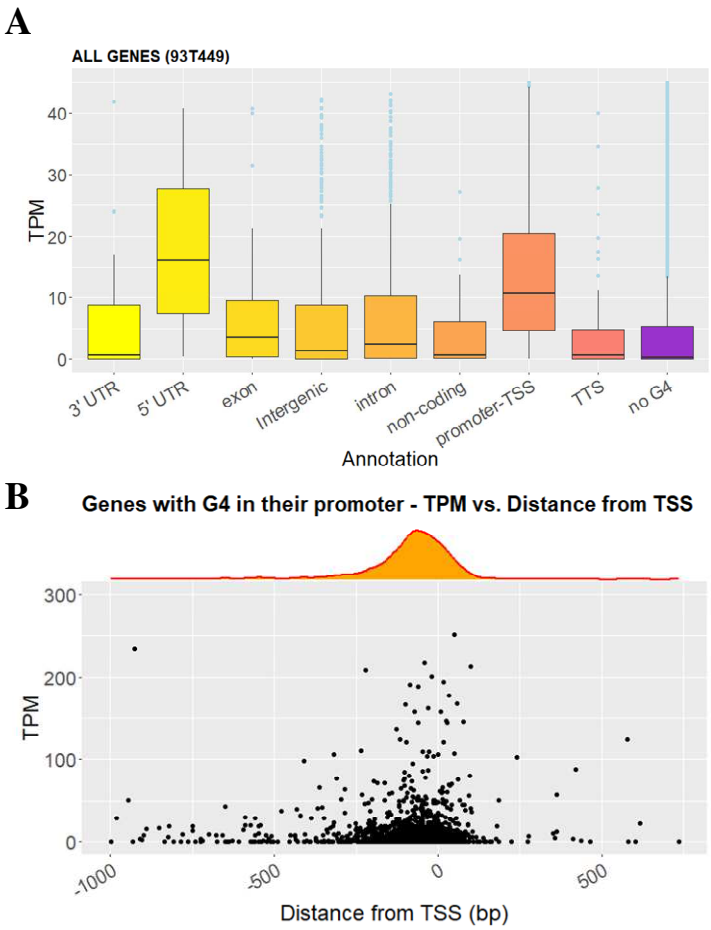

Figure 4-7. ***Transcriptional effect of G4 position.***

A) *G4-containing genes expression distribution in function of the genomic position.* ***B)*** *Gene expression in function of G4 distance from the TSS. The upper part of the graph is the density plot of the x axis.*

Being interested in characterizing the involvement of DNA G4s in the modulation of gene transcription, we focused on the promoter regions and explored them in more details. By plotting the gene expression level (expressed in TPM) in function of the distance of the corresponding TSS-proximal G4, we observed an accumulation of G4s within 250 bp from the TSS. Moreover, we observed that the closer the G4 to the TSS, the higher the transcriptional output of the corresponding gene **(Figure 4-7 B)**. Despite transcription enhancement is not the function most commonly attributed to G4 formation based on *in vitro* experiments, there is

already some experimental evidence that supports this claim[41,45,55]. One hypothesized mechanism through which G4s can facilitate transcription is by preventing the DNA double helix from reassembling after the opening of the single-stranded transcription bubble. In this light and based on our observations, the closer the G4 to the TSS, the easier would be for the RNA polymerase to reach the TSS and synthetize the nascent RNA. On the other hand, G4 formation on the template strand and downstream of the TSS can help in the promoter escape process, allowing the release of the early transcription complex from the promoter region[188]. Therefore, despite G4s are commonly considered structural obstacles to the RNA polymerase complex processivity, there are different steps of transcription in which they could facilitate the process, explaining the high transcriptional potential of G4-conatining genes revealed by our results.

4.2.4 Comparison of the G4 state in different cell lines

To consolidate the evidence that G4 structures are involved in the modulation of gene expression and that are specifically linked to high expression levels, we compared the G4 state in two different cell lines. Specifically, we compared the LPS cell line 93T449 with the immortalized pre-oncogenic keratinocytes HaCaT[55]. Comparison of the RNAseq data of the two cell lines yielded a huge number of differentially expressed genes (5895 genes with higher expression in the liposarcoma cells and 3642 with higher expression in keratinocytes).

We reasoned that if G4s are used as a mean for transcription regulation, their folding state should be differently modulated in genes with the same sequence and different expression level. Therefore, we divided all the differentially expressed genes into groups depending on their fold change (FC) expression with respect to the reference cell line. Considering the 93T449 cell line first, we plotted for each FC group the percentage of genes that contain at least one G4 uniquely in the promoter of that cell line **(Figure 4-8 B)**. Interestingly we observed enrichment in the presence of G4s in genes with higher expression in 93T449 cells than in HaCaT cells. Moreover, there was a positive trend in G4 occurrence with the increasing of the expression FC. On the contrary, genes that were less expressed in 93T449 with respect to HaCaT, also displayed low frequency of G4s, being generally depleted for the presence of these structures. As further proof of the robustness of this observation, when looking at the same genes in HaCaT cells, the opposite trend was observed: G4s were enriched in genes that were less expressed in 93T449 cells, while they were less abundant in genes with higher expression in

93T449. To summarize these data demonstrate that when the G4 is folded, the corresponding gene is susceptible to a high expression level, when it is unfolded, also the gene is repressed.

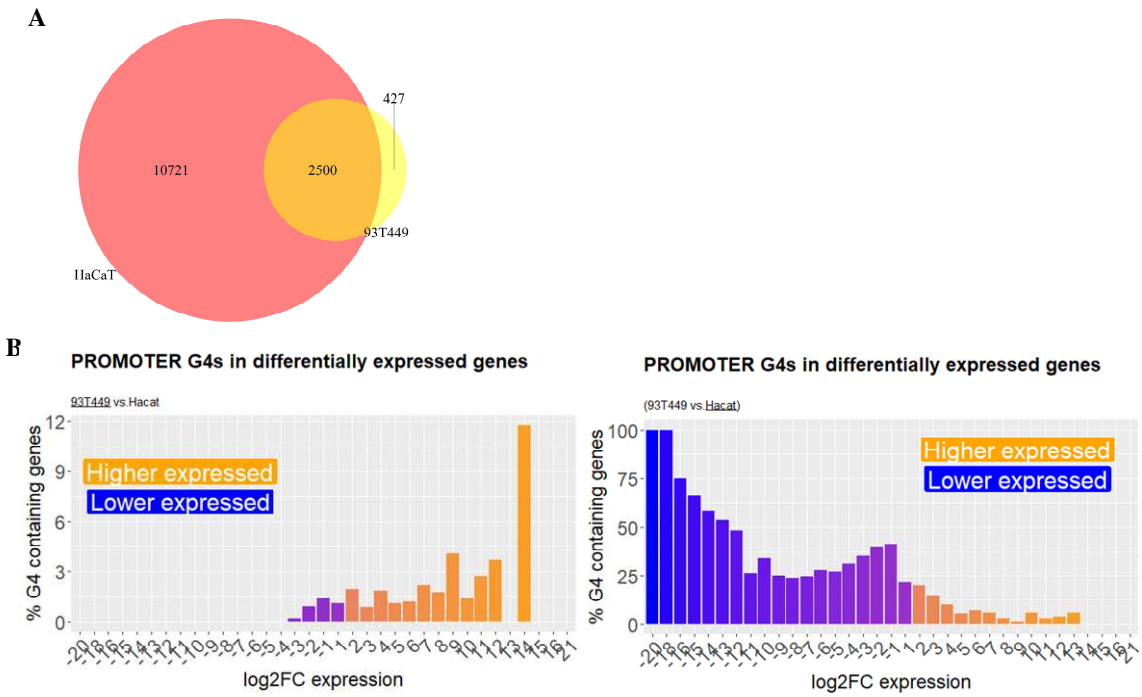

Figure 4-8. ***G4 comparison in 93T449 vs. HaCaT cell lines.***

A) *Venn diagram of G4-conteining genes in 93T449 and HaCaT cell lines.* ***B)*** *Percentage G4 occurrence in promoters of genes with the indicated fold change expression with respect to the opposite cell line. In the two panels the same genes are displayed and the percentage of G4 in 93T449 (right) or HaCaT (left) is calculated. In orange are indicated the bars corresponding to genes with higher expression in 93T449, and in blue genes with higher expression in HaCaT.*

These data support our previous evidence that G4s are associated to a high expression level and suggest that their folding can be modulated by cell-specific factors to regulate transcription. G4 formation can therefore be employed by different cells as a mean to establish their transcriptome and consequently their identity. However, it is not clear yet how the G4 folding is influenced by the chromatin state and the presence of epigenetic modifications. G4s have been indeed reported to correlate with open chromatin regions[55]. Open chromatin is a typical condition of transcriptionally active regions. Thus, we still need to establish if G4s are formed as a consequence of open chromatin state and therefore found in association with high transcription, or if the induction of a G4 determine a modification of the chromatin compaction.

4.2.5 Altering gene expression through a G4-stabilizing ligand

The combination of ChIP-seq and RNAseq data under physiological conditions revealed a strong and consistent correlation between formation of DNA G4s in gene promoters and active and high transcription of the corresponding gene. It is still unclear, however, how the G4 structure intervenes in the transcriptional process to enhance DNA to mRNA conversion. To add robustness and in the attempt to clarify the latter issue we perturbed the physiological G4 dynamics by stabilizing them through a potent G4-ligand (H-NDI-NMe_2PhAm)[189] **(Figure 1-9)**. At the same time, we tested G4s stabilization as anticancer strategy against LPS. The compound H-NDI-NMe2PhAm was chosen among two other compounds with a different chemical core: BRACO-19 which is a well-known G4 stabilizing agent[190] and $PyPhen_2PropNMe_2$, a recently synthesized compound that we tested as G4 binder **(Figure 1-9)**. The cytotoxic effect of the three compounds was tested on 93T449 cells by MTT assay **(Figure 4-9 A)**. After 24 h treatment only the H-NDI-NMe2PhAm displayed strong cytotoxicity at the tested concentration range, with CC_{50} of 1 µM. BRACO-19 and $PyPhen_2PropNMe_2$ compounds were instead discharged for further experiments: the first did not display enough cytotoxicity to be employed as antitumoral agent, the second precipitated at concentrations between 10-50 µM and was not cytotoxic at lower concentrations.

RNAseq of 93T449 cell treated for 24 h in the presence of 1 µM H-NDI-NMe_2PhAm was performed and gene expression levels compared to the control in untreated conditions **(Figure 4-9 B)**. About 800 genes with altered gene expression were identified suggesting a massive alteration of the transcriptional pattern as a consequence of G4s stabilization.

This result is not surprising given the high abundance and diversity of G4s in the genome of the studied cells. The perturbation of the natural plasticity of G4s rationally leads to a cascade of events altering several cellular pathways and probably leading to cells death. For the same reason it is also not surprising that a high number of genes that do not contain G4s in the control condition were altered upon treatment, probably due to indirect effects. Among the G4 containing genes, the main effect was downregulation of gene expression, consistently observed among replicates **(Figure 4-9 C)**. The specificity of the compound in cells was also demonstrated by the significant overrepresentation of G4-containing genes among the deregulated ones. In particular, 21 % of the downregulated genes contained at least one G4. This percentage is highly significant when considering that in the genome of 93T449 cells G4s were found in only 11 % of all genes **(Figure 4-9 D)**.

Repression of transcription after G4 stabilization is the most commonly observed effect in *in vitro* assays. Different explanations of this effect have been formulated: one is that the presence of a G4 structure in the double helix represents a spatial obstacle for the processivity of the RNA polymerase complex[53,143]; a second explanation is that G4s are formed as a compensation mechanism after the accumulation of negative superhelicity behind the transcriptional machinery: blocking G4s into their conformation leads to a reduction in the DNA intrinsic flexibility and consequently impairs RNA polymerase processivity[47,48].

Despite downregulation of transcription is the main observed consequence of G4-ligand mediated stabilization of the G4 structures, some of the G4-containing genes were also upregulated or left unaltered. This observation suggests a more complex mechanism through which G4s are implicated in the regulation of gene transcription, for which further targeted experiments and analysis are necessary.

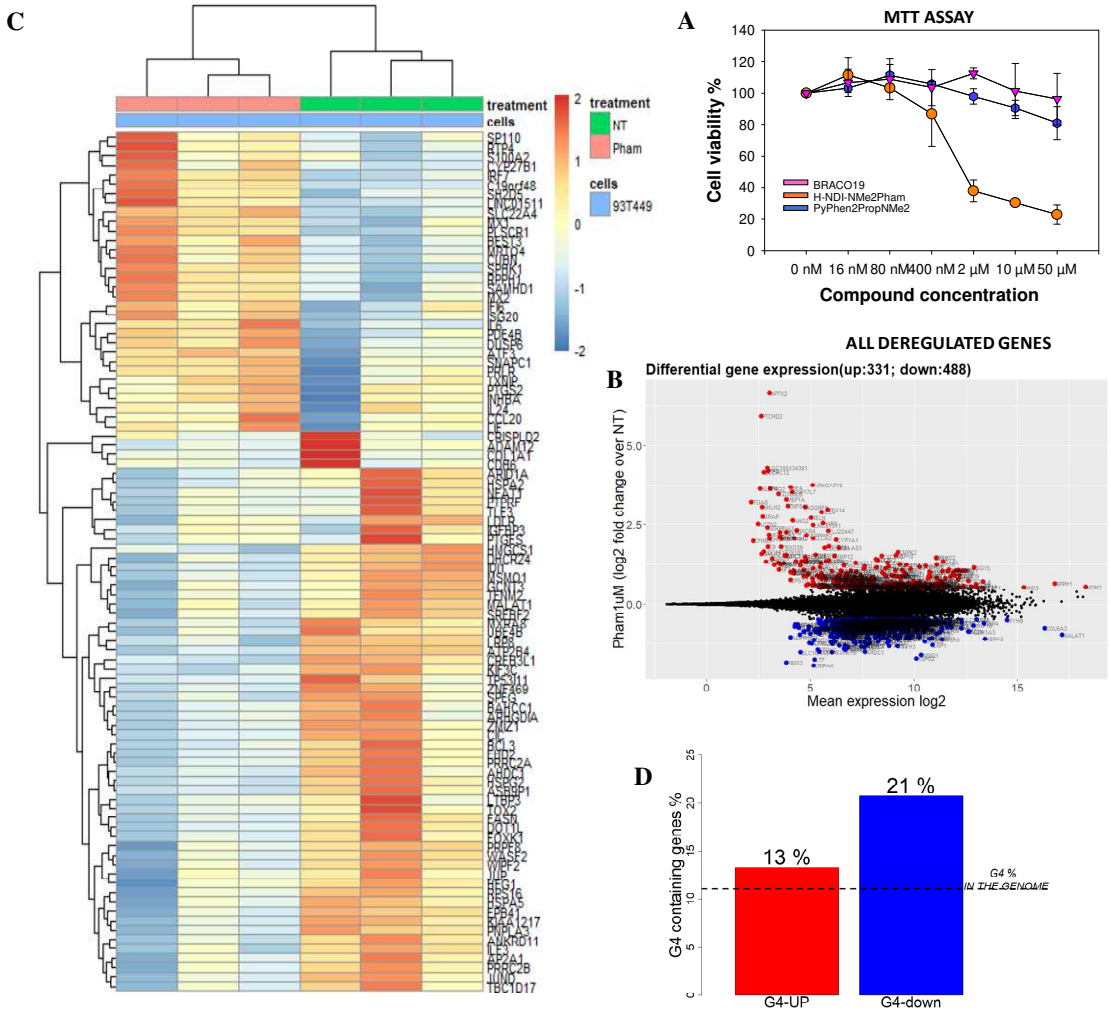

Figure 4-9. ***Figure. Analysis of gene expression after treatment with H-NDI-NMe$_2$Pham G4 ligand.***

A) *MTT assay for 24 h treatment with the G4 ligands: H-NDI-NMe$_2$Pham, BRACO-19 and PyPhen2PropNMe2* ***B)*** *Transcriptionally deregulated genes after 24 h treatment in the presence of the compound H-NDI-NMe$_2$Pham. Upregulated genes are shown in red and downregulated genes in blue.* ***C)*** *Row normalized heatmap Representing the expression level of G4-containing deregulated genes.* ***D)*** *Percentage of up/downregulated G4-contaiing genes with respect to all the identified deregulated genes. The dotted line represents the genomic abundance of G4s in the LPS cell line genome. Values above the line are non random overrepresentations.*

In summary, we were able to map G4s genome-wide in a WD-LPS cell line revealing a striking enrichment of these non-canonical DNA structures in TSS-proximal regions. The presence of G4s also appeared to be associated to active transcriptional genes and to high transcriptional level. G4 stabilization through a potent G4-ligand led to massive alteration of the

transcriptome, mainly downregulating the expression of G4-containing genes and leading to cancer cells death. These data represent one of the first attempts to characterize the position of G4s and their role in the modulation of gene transcription on genome-wide scale in an *in vivo* biologically relevant system.

DNA G-quadruplexes in viruses

4.3 HIV-1

With about 37 million people worldwide chronically living with AIDS, the search for a cure against HIV-1 is still a pressing issue. Our group has previously shown that the unique promoter of HIV-1 dynamically folds into a cluster of G4s and that G4-ligands are able to suppress viral transcription towards the stabilization of this structured region, while point mutations disrupting G4 are associated with increased transcriptional activity[93]. Physiologically, G4 folding and unfolding equilibria are regulated by the interaction with specific proteins and protein complexes. These premises motivated the search for host cellular proteins that could be involved in the G4-mediated modulation of the integrated provirus. Pull-down/MS combined experiments allowed us to identify at least two cellular proteins (NCL and hnRNP A2/B1) that interact with the LTR G4s regulating its folding and thus affecting viral transcription extent[97,98]. NCL, in particular, is the protein most often reported for its biological functions upon G4 recognition. It is widely believed that NCL plays a chaperone role by helping the correct folding of complex nucleic acids structures. Indeed, NCL has been shown to display a marked preference for both endogenous and exogenous G-rich sequences that can fold into G4[112]. In the case of HIV-1, we observed NCL-dependent inhibition of transcription through the stabilization of some of the G4s formed in the LTR promoter. Augmentation of the G4 basal effects was already observed as a consequence of NCL stabilization of some cellular promoter G4s, like *C-MYC* and *VEGF*[44,45]. Investigating the specific G4 features that makes NCL binding selective for some of these structures will allow a deeper understanding of NCL-G4 mediated mechanism of regulation of biological processes and the development of novel therapeutic strategies.

Here therefore we tried to unravel the structural features of G4s that are required to determine NCL recognition of some G4s over others. Moreover, we focused on new therapeutic strategies. We studied the ability of a phase II clinical trial G4 forming aptamer AS1411 to block HIV-1 entry in permissive cells, by sequestering the cell surface NCL.

4.3.1 Understanding the molecular features necessary for NCL-G4 interaction

NCL binding to G4s in relation to G-tracts number. To investigate the determinants required for efficient NCL binding with G4s, we tested through EMSA experiments NCL/oligonucleotide binding with ^{32}P-labelled nucleic acids **(Appendix 7.2)**. The ability of each oligonucleotide to fold into a G4 structure was assessed by CD spectroscopy. Moreover, the used oligonucleotides were chosen with the aim to test the contribution of oligonucleotide length, G4 loop, conformation and sequence to NCL binding.

To this aim, we initially tested two sets of telomeric sequences: found in humans (GGGTTA repeats), and in some ciliates (GGGTTT repeats)[191]. The telomeric sequences were chosen since the RGG domain, also possessed by NCL, was already reported to interact with them and because, being tandem repeats, they are quite simpler compared to the G4 sequences in oncogenes and HIV-1 promoters[162].

The first feature that we investigated was sequence length. We tested telomeric sequences with a different number of G tracts ($(GGGTTA)_n$ and $(GGGTTT)_n$) i.e. n = 4, 6 and 8. NCL binding to the G4 oligonucleotides was visualized as two bands running slower than the free unbound DNA. Thus the two bands were ascribed to the binding of the full length (upper) and the N-terminus-minus (lower) NCL protein to the G4s[97]. In these conditions, NCL binding was directly proportional to the length of the telomeric sequences and the difference in loop sequence between human and ciliate G4s did not modify this trend **(Figure 4-10)**.

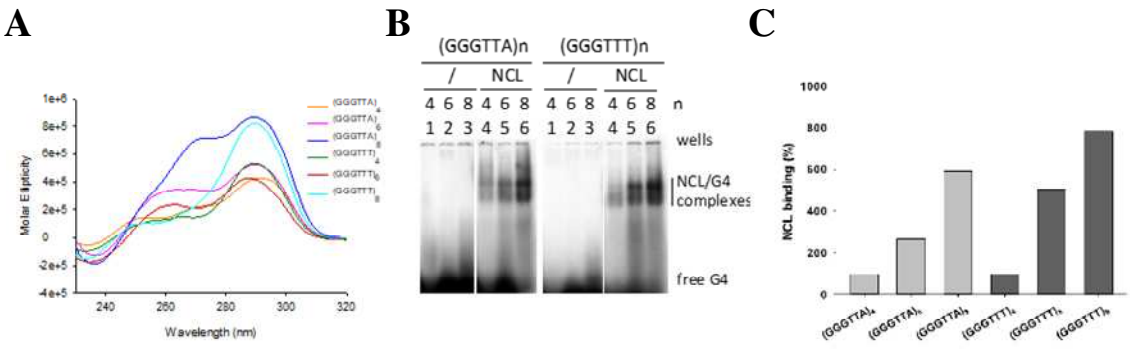

Figure 4-10. ***NCL binding to G4s with increasing number of G-tracts.***

A) *CD spectra of telomeric sequences $(GGGTTA)_n$ and $(GGGTTT)_n$, where n = 4, 6 and 8.* ***B)*** *EMSA analysis of NCL binding to telomeric sequences $(GGGTTA)_n$ - lane 4, 5, 6 left - and $(GGGTTT)_n$ - lane 4, 5, 6 right - . Lanes 1-4 are control reactions withouth NCL.* ***C)*** *Quantification of NCL/G4 complex bands obtained in the EMSA shown in panel B). Binding of*

lanes 4-6 left and 4-6 right was normalized for $(GGGTTA)_4$ and $(GGGTTT)_4$ NCL/G4 band intensity, respectively.

Long loops favor NCL binding to G4s. While the sequences with 4 and 8 G-tracts can fold into 1 and 2 stacked G4s, respectively, sequences with 5 and 6 G-tracts can fold into one G4 structure leaving one or two spare G-tracts. To test if the oligonucleotides length or the different structural conformation improved NCL binding, we first tested whether there was a preferred involvement of G-tracts in these conformationally dynamic sequences. DMS analysis of the $(GGGTTT)_5$ oligonucleotide showed that all Gs were equally protected during folding (Fig. 2A) indicating the equilibrium of multiple G4 structures that involved alternative G-tracts. Similar results have been previously reported[134]. In particular, in the case of the $(GGGTTT)_5$ oligonucleotide, 5 alternative G4 structures, may be envisaged.

Next, mutant oligonucleotides where each G-tract was replaced by Ts were employed **(Appendix 7.2)**. The $(GGGTTT)_4$ oligonucleotide was used as a control since it displays the same sequence and number of G-tracts as the mutant oligonucleotides, but lacks extra T bases in the loops or at the 5′- or 3′-end. All mutant sequences were shown to maintain the G4 folding by CD analysis, mainly displaying a hybrid/mixed type topology as the wild type sequence **(Figure 4-11)**. A slightly lower thermal stability was instead observed by CD thermal unfolding (**Table 4-4**). EMSA analysis of NCL binding to the mutant sequences showed mild preferential binding towards oligonucleotides where internal G-tracts were mutated **(Figure 4-11)**. Binding to sequences with external Ts was consistently lower, but higher than that to the $(GGGTTT)_4$ control. The $(GGGTTT)_5$ oligonucleotide displayed an average binding to all other mutant sequences. These data indicate that NCL preferentially binds G4s with at least one loop longer than 3 nts (e.g. 9 nts, compare n5M2, n5M3 and n5M4 binding, lanes 4–6 NCL, to n5M1 and n5M5, lanes 3 and 7 NCL **(Figure 4-11)** and with free nts at the 5′ or 3′ termini (compare n5M1, lane 3 NCL, and n5M5, lane 7 NCL, to (GGGTTT)4, lane1 NCL,**Figure 4-11**).

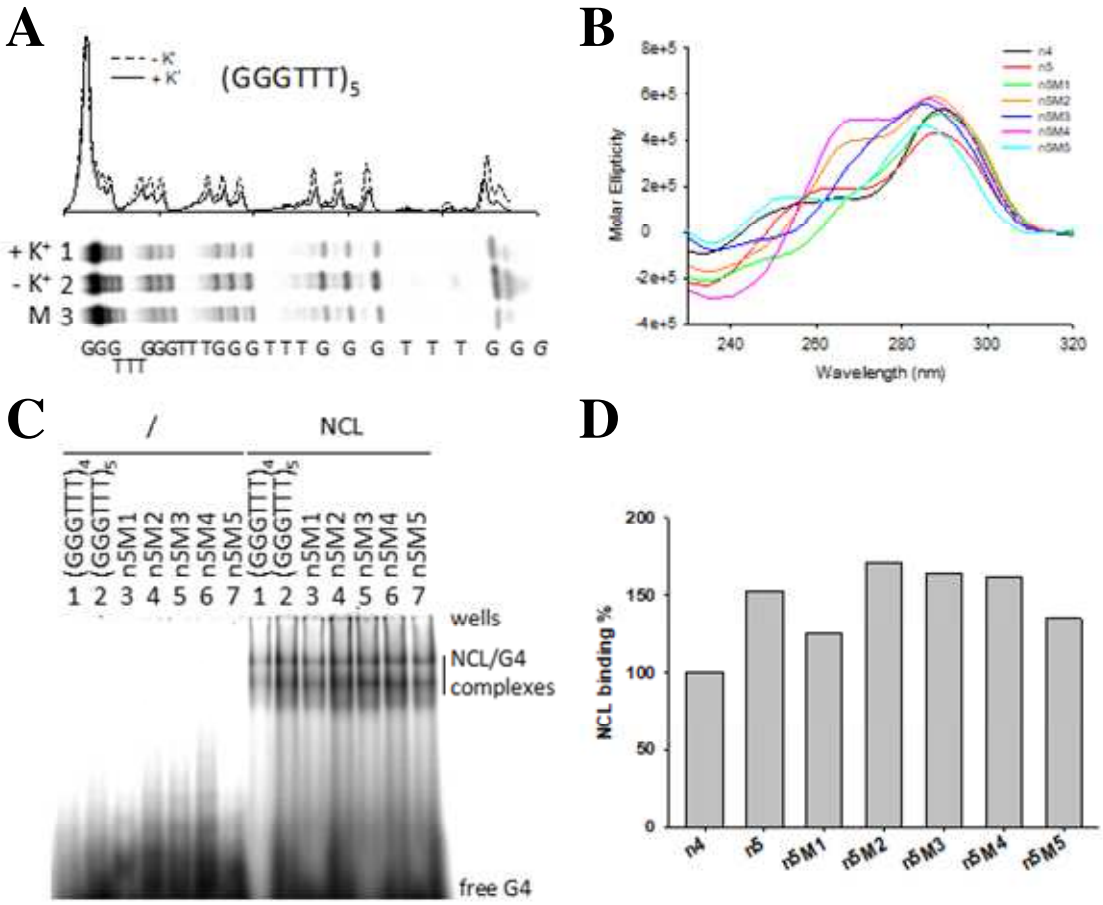

Figure 4-11. ***Analysis of NCL binding to alternative G4 structures formed by the (GGGTTT)$_5$ telomeric sequence***

A) *DMS protection analysis of (GGGTTT)$_5$ folded in the presence or absence of K^+. M is a marker lane obtained with the Maxam and Gilbert sequencing protocol. Oligonucleotide sequence is shown below the gel image. The densitogram above the gel shows quantification of cleaved bands intensity.* ***B)*** *CD spectra of (GGGTTT)$_5$ and mutant derivatives where each G-tract was systematically replaced by Ts.* ***C)*** *EMSA analysis of NCL binding to (GGGTTT)$_4$, (GGGTTT)$_5$ and mutant oligonucleotides. Oligonucleotide identity is shown above each lane in the gel image.* ***D)*** *Quantification of NCL/G4 complex bands obtained in the EMSA in panel C). Bars named n4 and n5 correspond to (GGGTTT)$_4$ and (GGGTTT)$_5$, respectively.*

Table 4-4. ***Stability (T_m values) of oligonucleotides of the (GGGTTT)$_5$ series and mutants.***

Oligonucleotide	T_m (265 nm)	T_m (285 nm)
(GGGTTT)$_4$	66.6 °C	66.1 °C
(GGGTTT)$_5$	58.0 °C	55.7 °C
n5M1	53.4 °C	54.5 °C
n5M2	56.7 °C	56.6 °C

n5M3	52.0 °C	53.2 °C
n5M4	/	60.4 °C
n5M5	65.5 °C	64.5 °C

T_m were measured by CD thermal unfolding and reported at the positive CD peaks, signatures of the parallel (265 nm) and antiparallel (285 nm) G4 conformation contribution

More than two long or extremely short loops destabilize NCL interaction with G4s. To further explore the impact of loop length on NCL binding to G4s, three series of oligonucleotides with different loop-length combinations were tested: in each series, the two lateral loops were kept constant (1, 3 or 6 nts in each series) and the central loop was modified (1, 3 or 6 nts) **(Appendix 7.2)**. All oligonucleotides folded into G4: in this case some sequences displayed a neat parallel topology (111, 131, 161, 313, 616, **Figure 4-12 A**), while other a mixed/hybrid type one (333, 363, 636, 666, **Figure 4-12 A**), in keeping with the notion that G4s with 1-nt loops are constrained to the parallel conformation[13,15]. In general, G4s with shorter loops were more stable than those with longer loops (**Table 4-5)**.

Table 4-5. ***Stability (T_m values) of oligonucleotides of the $(GGGTTT)_4$ series***.

Oligonucleotide	T_m (265 nm)	T_m (285 nm)
111	> 90 °C	/
131	> 90 °C	/
161	72.1 °C	/
313	66.1 °C	/
333	66.6 °C	66.1 °C
363	63. 0°C	64.0 °C
616	52.7 °C	/
636	52.9 °C	53.0 °C
666	45.3 °C	44.4 °C

T_m were measured by CD thermal unfolding and reported at the positive CD peaks, signatures of the parallel (265 nm) and antiparallel (285 nm) G4 conformation contribution.

EMSA binding analysis showed a general increase in NCL binding upon loop extension **(Figure 4-12 B and C)**. This effect was particularly visible in the three G4s where the central loop was modified, while keeping constant the lateral loops, and in the series where the central loop was kept constant, upon the increasing length of the lateral loops. Two notable exceptions were found: when all loops were 1 nt- or 6-nt long, NCL binding was higher or lower, respectively, than expected by the above-described trend (oligonucleotides 111 and 666 in **Figure 4-12 B and C**. The observed behavior was apparently independent of the G4 conformation: oligonucleotides adopted both mixed and parallel-like G4 conformations **(Figure 4-12 A)**, while no straightforward correlation with the intensity of NCL binding could be found **(Figure 4-12 B and C)**. However, when testing G4 folding in EMSA, we observed one main band for all oligonucleotides, except oligonucleotides 111 and 131, which displayed two bands running at different migration rates **(Figure 4-12 D)**.

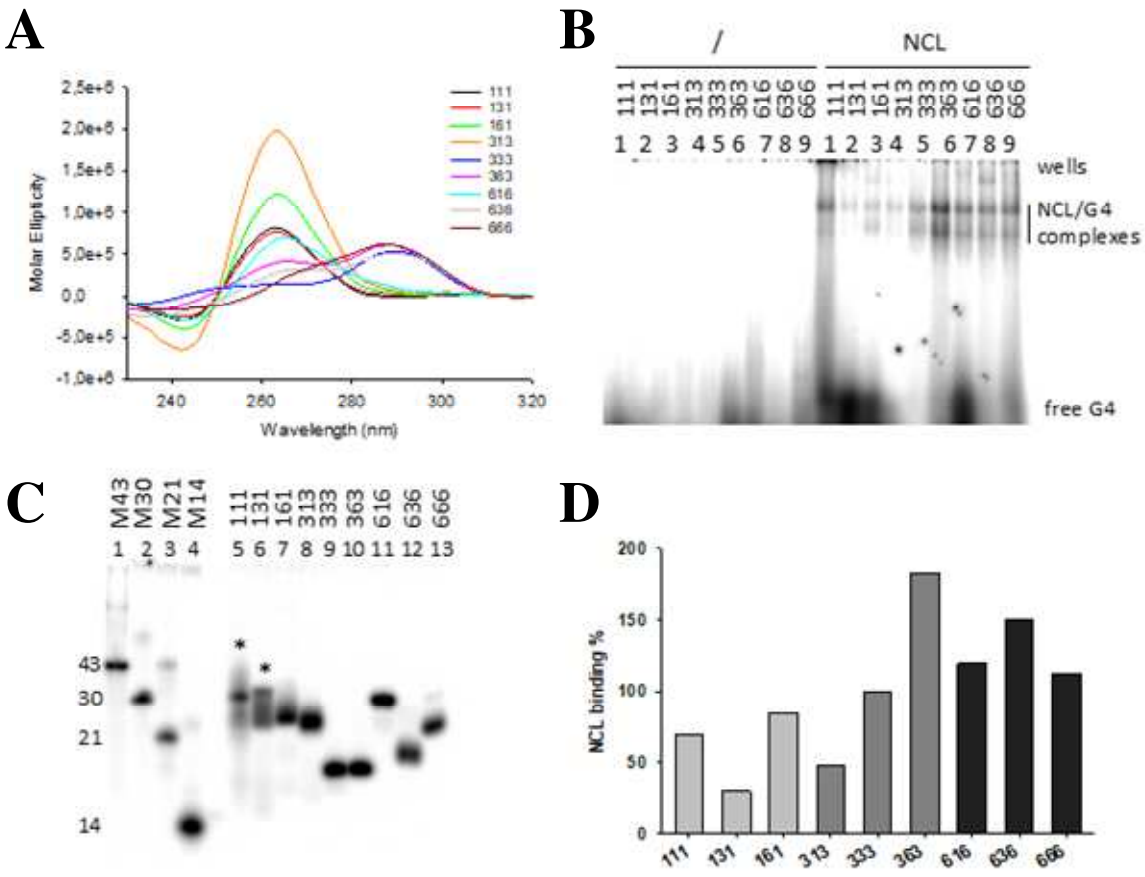

Figure 4-12. ***NCL binding to G4s with systematic elongation of loop length.***

***A)** CD spectra of oligonucleotides with two constant lateral (1, 3 or 6 nts) and variable central loop (1, 3 or 6 nts). **B)** EMSA analysis of NCL binding to each G4 with different loop-length combination. Oligonucleotides identity is shown above each lane. **C)** Quantification of NCL/G4 complex bands obtained in the EMSA in panel (B). **D)** Native gel electrophoresis of G4-folded oligonucleotides. Oligonucleotides identity is shown above each lane. Lanes 1-4 are marker oligonucleotides of known length and unable to fold into secondary structures. The symbol * indicates oligonucleotides showing two bands with different migration rate.*

To test the molecularity of the upper running band, each labelled oligonucleotide was added with increasing amounts of the same unlabelled nucleic acid. As shown in **Figure 4-13 A**, the amount of the upper running band increased with increasing amounts of the unlabelled oligonucleotide, indicating the dimeric nature of the upper band. Since it has been reported that the presence of one T base at the end of a G4 forming oligonucleotide induces a preferential monomeric G4[192,193], we added one T base at each end of the 111, 131 and 161 oligonucleotides (T111T, T131T, T161T, **Appendix 7.2**). The modified sequences maintained indeed the monomeric G4 conformation as shown in figures **Figure 4-13 B and C**. Interestingly, these T-added oligonucleotides bound NCL to a lower extent than those without T, probably due to steric hindrance, however, the binding trend was restored: the increase in the amount of NCL/DNA complex paralleled the lengthening of the loops **(Figure 4-13 D)**.

The above data indicate that in G4-forming sequences formed by GGG tracts and T loops, NCL binding is in general favoured by the presence of one or two long loops (i.e. 6 nt), whereas three long loops are less favoured. The presence of one or two 1-nt long loops in the absence of at least one 6-nt loops disfavours NCL binding.

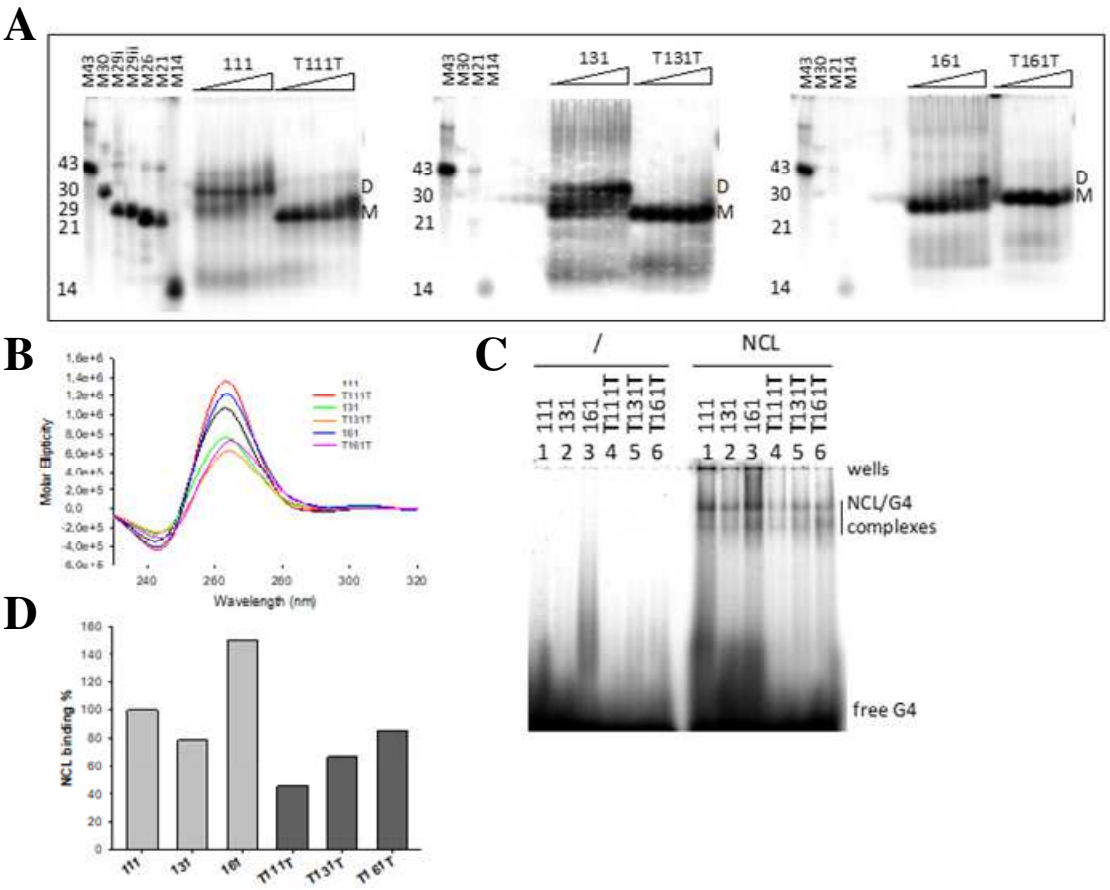

Figure 4-13. ***Analysis of monomeric and dimeric G4 folding of relevant oligonucleotides.***

A) *Native gel electrophoresis at increasing oligonucleotide concentration (20, 45, 120, 415, 1600, 6200 nM). Samples identity is indicated above the gel image. Marker oligonucleotides with known length and unable to fold into secondary structures are loaded.* ***B)*** *CD spectra of G4s showing dimeric folding and their corresponding monomolecular form when one T base was added at each of the two ends.* ***C)*** *EMSA analysis of NCL binding. Oligonucleotides identity is shown above each lane.* ***D)*** *Quantification of NCL/G4 complex bands obtained in the EMSA in panel C).*

NCL binding to G4 oligonucleotides is independent of the G4 conformation. By now, it was not possible to establish in a straightforward manner, whether NCL/G4 binding was affected by the G4 conformation. It has been reported that oligonucleotide G4 folding depends on the presence of the monovalent cation in solution. For instance, the human telomeric sequence adopts a hybrid conformation in K^+ solution[194], while it is antiparallel in Na^+ solution[195]. Moreover, the presence in solution of cosolvents, such as polyethylene glycol (PEG), during G4 folding is commonly used to perturb G4 topology[196]. To investigate the effect of G4 conformation on NCL binding, we folded oligonucleotides previously tested in KCl alone

(Figure 4-12 B), in KCl and NaCl in the presence of PEG 20% (w/v). In these conditions we observed in most cases a clear-cut conformational change **(Figure 4-14 A)**: for instance, oligonucleotide 333 was mixed with a prevalent antiparallel conformation in K^+, while it shifted to a prevalent parallel topology in K^+ + PEG **(compare Figure 4-12 A and Figure 4-14 A)**. Assessing NCL binding on these oligonucleotides folded in PEG, we observed the maintenance of the increased NCL binding to longer loop sequences trend **(Figure 4-14 B and C)**. These data indicate that binding of NCL to telomeric-like G4 oligonucleotides is in general independent of the conformation and that loop length is the major determinant of NCL binding efficiency to telomeric G4s.

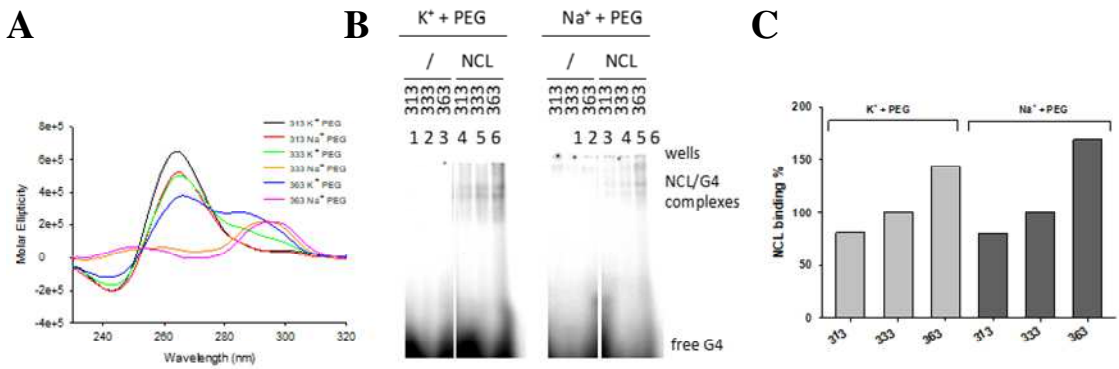

Figure 4-14. ***Analysis of NCL binding towards oligonucleotides with variable G4s conformation.***

A) *CD spectra of 313, 333, 363 oligonucleotides folded in the presence of 20% (w/v) PEG200 and K^+ or Na^+.* ***B)*** *EMSA analysis of NCL binding to 313, 333, 363 oligonucleotides folded in the presence of 20% PEG200 and K^+ or Na^+ as indicated above the gel image.* ***C)*** *Quantification of NCL/G4 complex bands obtained in the EMSA in panel (B). For the K^+ or Na^+ experiment series, binding was normalized to the intensity of the 333 NCL/G4 complex band in the corresponding salt condition.*

Sequence dependence: NCL binding to oncogenes promoter G4s. To verify if NCL binding determinants observed in telomeric $(GGGTTT)_n$ G4s were maintained in G4s with more complex sequences, we tested NCL binding towards four oncogene promoters i.e. *C-MYC, C-KIT1, BCL-2* and *hTERT 1–4*[197–200]. Each of them was shown to form stable G4s whose conformations have been characterized by deep structural analysis. Moreover, all these oncogene promoter G4s display loops with different length, composition and conformation (**Appendix 7.2 Figure 4-15 A**). We observed that NCL bound preferentially to *bcl-2*, followed by *c-myc*; *hTert 1–4* and *c-kit1* displayed similar and lower NCL binding **(Figure 4-15 B and**

C). Among these G4s, *BCL-2* is the only structure exhibiting a long loop (7 nts) and two shorter loops (3 and 1 nts), whereas *C-MYC*, *C-KIT1* and *hTERT 1-4* they all have loops ≤4 nts **(Appendix 7.2)**.

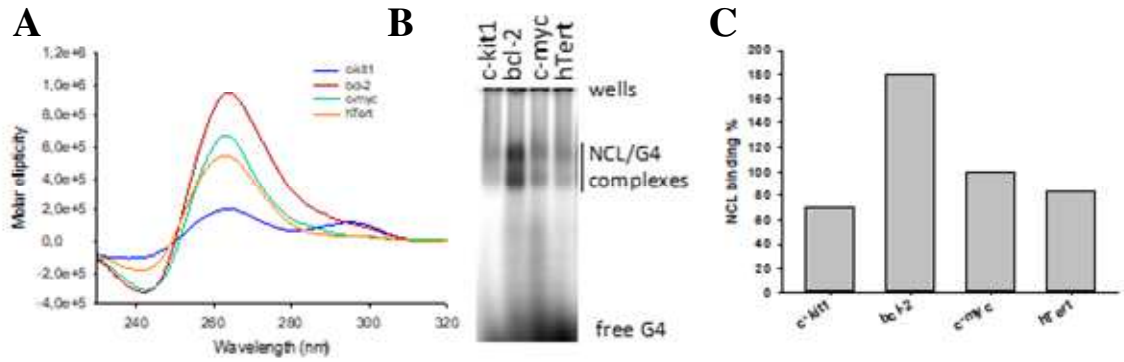

Figure 4-15. ***NCL binding to human oncogene promoter G4s.***

A) *CD spectra of G4s formed in the C-KIT1, BCL-2, C-MYC and hTERT promoters.* ***B)*** *EMSA analysis of NCL binding to oncogene promoter G4s.* ***C)*** *Quantification of NCL/G4 complexes obtained in the EMSA in panel (B).*

It has very recently been shown that the hTert full-length sequence is composed of 12 G-tracts that fold in two adjacent G4s: our tested *hTERT 1-4* comprises the first four G-tracts; the second G4, spanning G-tracts 5–12, displays a peculiar structure where a 26 nt-long loop forms an extended stem-hairpin[201]. We wished to test NCL binding to the two separated *hTERT* G4s *hTERT 1-4* and *hTERT 5-12*) and to the full-length sequence (*hTERT 1-12*). All three *hTERT* G4s have similar CD spectra with a major peak at 260 nm and a shoulder at 295 nm, with molar ellipticity values at 260 nm increasing with the number of G-tracts in the G4-forming oligonucleotides **(Figure 4-16 A)**. NCL binding to *hTERT 1-4*and hTert 5–12 was similar, indicating that the stem-hairpin suppresses the enhancing binding effect of the long loop **(Figure 4-16 B and C)**. Binding of NCL to *hTert* 1–12 was slightly enhanced **(Figure 4-16 B and C)**, retracing the behaviour observed in the $(GGGTTA)_n$ and $(GGGTTT)_n$ oligonucleotides, where NCL preferentially bound to oligonucleotides with longer loops when multiple G-tracts were available for folding **(Figure 4-10 and Figure 4-11)**. It is thus possible that also in the case of hTert 1–12 NCL induces folding of a less stable but favourite long-looped G4. In alternative, the 7 nt spacer between the two G4 s (hTert 1–4 and hTert 5–12) is recognized as a long loop and therefore stimulates NCL binding.

We next tested bcl-2 and c-myc in the presence of PEG 20% (w/v) to rule out an effect of PEG in NCL binding. Both oligonucleotides in K^+ are folded in a parallel G4 conformation, which is not modified in the presence of PEG **(Figure 4-16 D)**. As observed before for the telomeric-like G4s, NCL maintained a similar trend and degree of binding to these G4s in the absence/presence of PEG, confirming that PEG did not significantly alter NCL binding **(Figure 4-16 E)**.

Interestingly, these data fit with NCL binding to HIV-1 LTR promoter G4s, previously reported and obtained by both EMSA and FRET analysis[97]

.]. Among the different G4 species that can independently form in HIV-1 LTR, the best-bound G4 is LTR-III. It displays 1, 11, 3 nt-long loops. The second G4 with better affinity for NCL is LTR-II with 8, 11, 1 nt-long loops. Instead LTR-IV, which presents 1, 4, 1 nt-long loops, was bound to a negligible extent[93,97]. These data indicate that the results collected on $(GGGTTT)_n$ G4s apply in general to G4s with random sequences and that, again, the loop length is the major determinant in driving NCL binding to G4s.

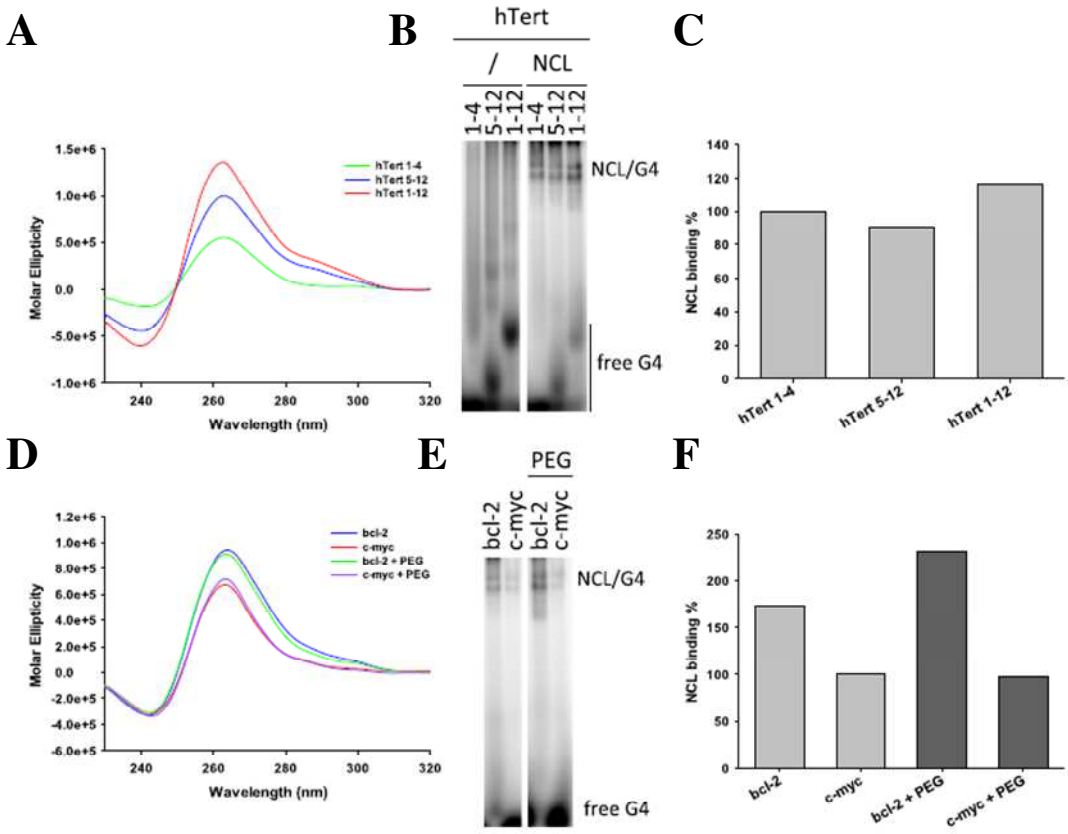

Figure 4-16. ***NCL binding to G4s in special conditions.***

A) *CD spectra of G4s formed by hTert 1–4, hTert 5–12 and hTert 1–12 oligonucleotides. Numbers indicate G4-tracts.* ***B)*** *EMSA analysis of NCL binding to hTert G4s.* ***C)*** *Quantification*

of NCL/G4 complexes obtained in the EMSA in panel B). ***D)*** *CD spectra of G4s formed by c-myc and bcl-2 in the presence/absence of PEG 20% (w/v).* ***E)*** *EMSA analysis of NCL binding to c-myc and bcl-2 G4s in the presence/absence of PEG 20% (w/v).* ***F)*** *Quantification of NCL/G4 complexes obtained in the EMSA in panel E).*

Given the importance of NCL in the regulation of G4-mediated cellular mechanisms, another group before us already attempted to characterize the G4 features that induce NCL binding[44]. In contrast to our work, this group used C-myc as test G4 to prove a preferential binding of NCL to parallel G4s with short loops. This apparent discrepancy may be explained by several reasons: i) we used the full-length human NCL, while Gonzalez et al. used an *E. coli* purified recombinant protein which lacks the N-terminal domain (which in turn was fused to the bacterial maltose-binding protein) and eventual post-translational modifications. These differences may impair the specific recognition of G4s and change the steric occupancy of the protein. In particular because NCL N-terminal domain is known to participate in the reinforcing of the direct binding to nucleic acids mediated by the RGG domain. ii) It has been demonstrated that short loops tend to constrain G4 structures into a parallel conformation, regardless of base composition[13,15], therefore the suggested preferential binding to short looped G4s in the parallel conformation may derive from the analysis of two dependent variables.

In summary, our results were obtained through a systematic analysis of loop sequence and oligonucleotide conformation. We unambiguously showed that NCL prefers G4s with at least one long loop (≥3 nts), while three short loops (1 nt) or three long loops (6 nts) are disfavoured. The analysis on the simplified sequences allowed us to modify the conformation of the oligonucleotides working in different salt (K^+ or Na^+) and crowding conditions (PEG 20% w/v)[192,196,202]. to show that NCL binding is in general independent of the oligonucleotide conformation and that loop lengths is the driving force that enhances NCL binding to G4s.

By extending the results from the telomeric sequences to cellular oncogene and HIV-1 promoter G4s, we were able to show that loop length is the major determinant in the enhancement of NCL binding to G4s even in more variable and complex sequences. This observation is in line with the reported recognition of G4s by the cellular protein EWS: EWS shares with NCL the RGG domain which mediates recognition of G4s[203]. and binds to G4s independently of base composition in long loops[162]. Finally, our results may shed some light on NCL recognition of HIV-1 LTR G4s: it is known that regions containing multiple G-tracts can fold into several conformations involving the alternative use of different G-tracts. We here indicate that NCL binding stimulates folding of these regions to G4s containing long loops (i.e.

some G-tracts may be included to the loop moiety) therefore modifying the overall 3D structure and steric hindrance of these nucleic acid repeated sequences. Interestingly, indeed, NCL preferentially stabilized G4s that were less intrinsically stable, likely because of the possibility in this condition to fold them in the best protein-fitting conformation. This also explains the strong affinity of NCL to the LTR G4s, which is a dynamic sequence allowing more equilibria (at least two of which comprising long loops) and potentially allowing the modulation of its structure by the interaction with specific proteins.

4.3.2 The G4 aptamer AS1411 as anti HIV-1 agent

The role of NCL in the HIV-1 life cycle is not restricted to the regulation of provirus transcription, but it is also involved in virion attachment to permissive $CD4^+$ cells. This phase is normally mediated by the main co-receptors CXCR4 and CCR5, but can also occur through heparan sulfate proteoglycans and cell-surface-expressed NCL[204]. Immediately after HIV-1 entry into susceptible cells, cell-surface NCL is down-regulated as a consequence of cytoplasmic translocation[205]. Due to the important involvement of cell-surface NCL in the viral initial phases of infection, NCL has been proposed as an anti-HIV-1 target. Here we tested the G4 forming aptamer AS1411 for targeting NCL-mediated HIV-1 cellular uptake[117]. AS1411 already completed two phase II clinical trials as an anticancer agent displaying very low toxicity. We demonstrated through antiviral assay in TZM-bl and MT-4 cells, as well as in chronically and latently infected cells, that AS1411 has the ability to inhibit viral replication. A time of addition and viral binding assays confirmed our hypothesis that AS1411 acts at the level of viral entry. Other anti-HIV-1 G4-forming aptamers have been reported to block viral attachment by binding to the gp120 envelope glycoprotein. We used SPR optical biosensor to test AS1411 binding to gp120 and NCL. As a control we also measured binding affinity of another HIV-1 G4 of the LTR region (LTR-III), as well as a scrambled non G-rich oligonucleotide (SCRA) and the C-rich sequence (CRO26) **(Appendix 7.2)**. AS1411 showed a fast association and high affinity (34.2 nM) for NCL, while its binding to gp120 was too low to assign a meaningful binding constant **(Figure 4-17 A and B)**. LTR-III had a similar behavior, but its affinity for NCL (76.3 nM) was lower with respect to that of AS1411 and showed a slower association rate **(Figure 4-17 C and D)**. SCRA and CRO26 in contrast, exhibited negligible interaction both with NCL and gp120 **(Figure 4-17 E-H)**.

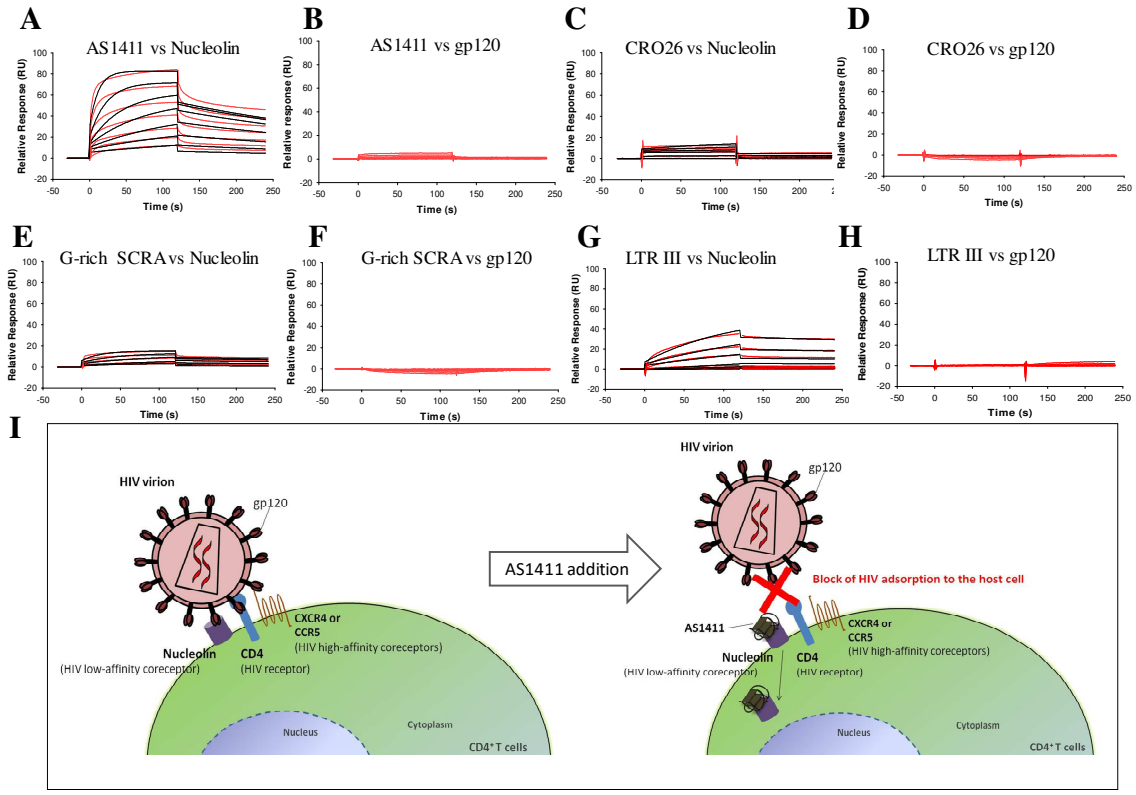

Figure 4-17. ***Binding affinity of NCL and gp120 measured by SPR.***

Sensorgrams were obtained by injection over immobilized NCL and gp120 respectively of the oligonucleitdes ***A-B)*** *AS1411,* ***C-D)*** *CRO26,* ***E-F)*** *SCRA,* ***G-H)*** *LTR-III. Oligonucleotides were tested in the concentration range 15.6 nM – 1 uM. Experimental curves are shown in gray, while 1:1 global fittings are in black.* ***I)*** *AS1411 mechanism of action in HIV-1 entry*[117].

These data strongly support that binding to NCL is the main mechanism of action of AS1411. In particular, we demonstrated through SPR technology that AS1411 effectively occupies the NCL binding site required for viral attachment, thus inhibiting its entry into cells. Moreover, the dimeric nature of this aptamer, its high stability and natural propensity to fold into G4 may account for the improved antiviral activity compared with other G4-forming oligonucleotides. Having AS1411 already passed two phase II clinical trials, it represents a really promising antiviral compounds with high safety and tolerance and specific effect.

Together our results concerning G4s in HIV-1, at genomic level and as tools for inhibiting viral entry and infection, highlight the importance of NCL-G4 complex as antiviral target.

Besides HIV-1, G4s investigation in the genome of the herpesvirus HSV-1 was addressed in the present work, as described in the next chapter.

4.4 HSV-1

The HSV-1 genome is of particular interest for the study of G4s given its incredible richness in G-rich sequences. Recently we provided evidence for the presence of very stable G4-forming regions located in the HSV-1 inverted repeats. Multiple conserved and extended clusters of G4 forming sequences were observed, covering about 2 kbp of the 152 kbp-viral genome[101]. As previously discussed, the direct visualization and study of G4s *in vivo* is currently one of the main challenges in the dissection of their function. Beside BG4, a monoclonal antibody (1H6) was recently developed in mice immunized with a G4-carrying antigen. 1H6 displayed the ability to recognize with high affinity tetramolecular and monomolecular G4s. Immunofluorescence microscopy exploiting 1H6 resulted in strong nuclear staining in most human cells, which was suppressed by the addition of soluble G4 DNA, with prior treatment with DNase and in cells deficient for a G4-specific helicase (FANCJ). In contrast, cell treatment with G4 stabilizing agents increased the intensity and extension of staining[119].

Given the extraordinary extension of G4 forming regions in the HSV-1 genome, we aimed at visualizing G4s in eukaryotic cells infected with HSV-1 by employing 1H6. To reach this goal, we had to first determine the ability of the antibody to specifically interact with HSV-1 G4s. We therefore employed a SPR based method for the measurement of 1H6 affinity to HSV-1 G4s with respect to other cellular G4s.

4.4.1 Development of a new SPR immobilization strategy for 1H6 anti-G4 antibody

SPR is a fast, label-free and highly sensitive technique allowing the real-time measurement of binding kinetics and affinity between two interacting partners. Given the advantages of this technology, we exploited it to screen 1H6 interaction with some selected G4 oligonucleotides: the tetramolecular Oxytrichia telomeric G4 (*Oxy2*) which is the antigen against which 1H6 was developed[119]. a representative G4 of HSV-1 (*un3*), a human G4 (*Bcl-2*), and three negative controls unable to form G4, i.e. a double stranded (*dsDNA*), a single stranded (*ssDNA*) and a hairpin DNA (*hp*) as reported in **Appendix 7.2**.

As a first attempt we directly immobilized 1H6 (ligand) on the sensor chip surface through amine coupling chemistry which is the most common strategy providing a very broad efficacy[206]. We successfully immobilized about 1000 RU of Ab, satisfying the conditions of

low-density surface and good signal-to-noise-ratio[207]. According to *Equation (1)*, in this condition the G4 binding molecules (analytes) were expected to reach a response ranging between 50-130 RU. However, when the G4 and non-G4 analytes were flowed on the immobilized 1H6, no binding was observed **(Figure 4-18 A)**. The reason of the observed negative result is likely the inactivation of the ligand during the immobilization procedure. This could be due to: i) 1H6 denaturation caused by its dilution in low pH buffer; ii) presence of basic residues in the G4 binding site of 1H6 that have reacted with the activated dextran matrix impairing 1H6 ability to bind G4s[119] **(Figure 4-19 E)**; iii) reduction of 1H6 "breathing" and steric hindrance that interferes with G4 recognition[208]; iv) a combination of all the above.

To avoid the mentioned issues, we shifted to a capturing immobilization strategy, in which 1H6 was non-covalently bound to the chip surface by mean of an anti-mouse IgG, recognizing the species-specific portion (Fc region or exposed constant chains) of the ligand Ab of interest. We set up the capturing conditions to reach about 1000 RU of bound 1H6 and performed the binding analysis in the same conditions used for the standard amine coupling. A positive binding response of all the G4 oligonucleotides was now detected, while no interaction was visible for the non-G4 negative controls **(Figure 4-18 B)**. Despite the improvement provided by the capturing mediated strategy, the obtained response was not sufficient to reliably determine the kinetic constants of all the G4 analytes [**Table 4-6**]. Less than 15% of the expected maximal response was indeed obtained for *un3* and *bcl-2*, while 45% of surface activity was reached for the best analyte (*Oxy2*) **(Figure 4-19)**.

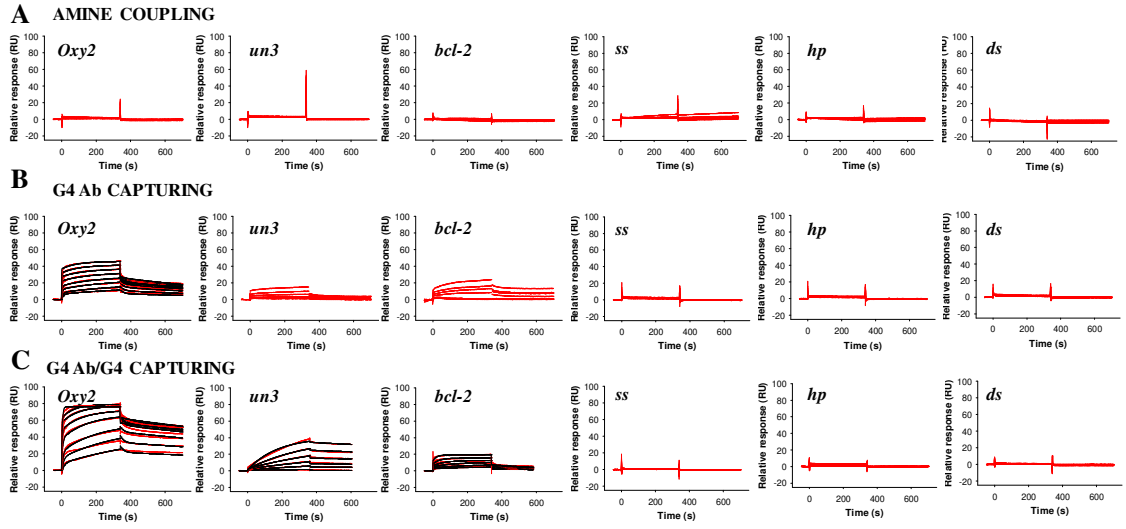

Figure 4-18. ***Binding analysis of 1H6-G4 and control nucleic acids interaction after.***

***A)** 1H6 direct amine coupling, **B)** anti-mouse mediated 1H6 capturing, **C)** anti-mouse mediated capturing of the 1H6-G4 complex. Oxy2 G4 analyte was tested in the concentration range 62.50 nM – 8.00 μM, while for un3 and bcl-2 the 125 nM – 8.00 μM range was sufficient to cover the kinetic spectrum. Recorded sensorgrams are shown in red, while fitting curves are in black. Fitting curves were not reported when they did not fit or the kinetic constants were outside the sensitivity of the instrument.*

Table 4-6. ***Kinetic parameters for 1H6-G4 interaction.***

	AMINE COUPLING			**G4 Ab CAPTURING**			**G4 Ab/G4 CAPTURING**			
	ka (Ms^{-1})	**kd (s^{-1})**	**KD (nM)**	**ka (10^4 Ms^{-1})**	**kd (10^{-4} s^{-1})**	**KD (nM)**	**ka (10^4 Ms^{-1})**	**kd (10^{-4} s^{-1})**	**KD (nM)**	
Oxy2	nd	nd	nd	29.50 ± 0.92	6.99 ± 1.73	26.13 ± 0.51	5.11 ± 0.15	4.73 ± 1.54	9.16 ± 2.81	tetra
				110.60 ± 56.40	2385.26 ± 1774.46	74.45 ± 3.85	106.87 ± 37.22	395.73 ± 106.10	41.73 ± 13.15	bi
un3	nd	nd	nd	nd	nd	nd	0.07 ± 0.01	3.30 ± 0.10	479.65 ± 2.75	
bcl-2	nd	nd	nd	nd	nd	nd	0.58 ± 0.01	30.10 ± 11,0	518.05 ± 20.20	

nd = non detectable; Uncertainties of the reported values are calculated as the standard deviation of at least two experimental replicates obtained with independent ligand immobilizations.

We reasoned that such a reduced binding capacity of the immobilized ligand could be due again to heterogeneous orientation of 1H6 G4 binding site. Being the capturing anti-mouse antibody a polyclonal immunoglobulin, there are several portions of the 1H6 Fc region that can be recognized. Differently oriented 1H6 molecules on a crowded surface can impair or hide the availability of the G4 binding site (**Figure 4-19 E**).

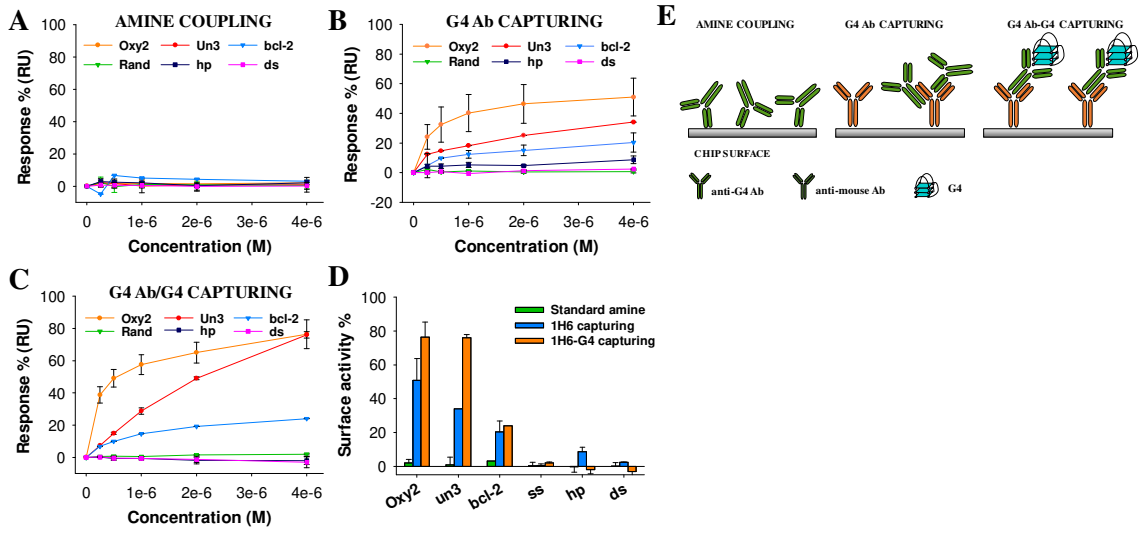

Figure 4-19. ***Sensor chip surface activity.***

1H6 surface activity was measured as percentage of the theoretically expected R_{max} obtained for each tested analyte (0 nM – 4.00 µM) after ***A)*** *direct amine coupling of 1H6,* ***B)*** *anti-mouse mediated 1H6 capturing and* ***C)*** *anti-mouse mediated capturing of previously formed 1H6-G4 complex. Panel* ***D)*** *shows the comparison of surface activity determined at analyte saturating concentration (4.00 µM) for the three immobilization strategies as indicated.* ***E)*** *Representation of 1H6 antibody orientation on the chip surface according to the different immobilization strategies: Direct amine coupling, Anti-mouse antibody mediated capturing and anti-mouse antibody mediated capturing following the incubation of 1H6 with a G4 partner.*

To further improve 1H6 immobilization we modified the capturing step by adding a second orientating factor, with the aim to limit the availability of 1H6 portions that can be recognized by the anti-mouse Ab and at the same time assuring the exposure of the G4 binding site. We achieved these conditions by incubating 1H6 with a high affinity G4 partner (*Oxy2*) at saturating concentration. The 1H6/G4 complex was then injected on the covalently immobilized anti-mouse site. Being 1H6-anti-mouse Ab interaction much stronger than the 1H6-G4 one, once captured, the complexed G4 was washed through a flow of high ionic strength buffer (KCl 1 M), leaving 1H6 antigen binding site free for successive binding analysis. The successful dissociation of the orienting G4 was confirmed by the mild drop in response (about 100 RU) observed after the wash at the end of the complex capturing, corresponding to the expected amount of analyte necessary to saturate the Ab **(Figure 4-20 B)**. Comparing the capturing curves in the presence or absence of the G4 partner we observed different kinetics of association **(Figure 4-20 A and B)**. In particular, the association of 1H6-G4 complex was slower than 1H6 alone. This is likely due to the presence of the G4, which

reduces the availability of some 1H6 binding sites recognized by the anti-mouse Ab. About 1000 RU of 1H6 remained captured on the anti-mouse Ab after the wash, thus a kinetic analysis could be performed and the results could be compared to the previous strategies. As in the previous capturing procedure, the doubly oriented method that we developed allowed to record a binding interaction between 1H6 and all the tested G4 analytes, while no binding was detected for the non -G4 controls. The improvements gained by the new strategy also allowed to increase the surface activity (i.e. the amount of 1H6 molecules that are functionally captured and can be recognized by the flowing analytes). In fact, *Oxy2* and *un3* reached now about 65% of the expected response and *bcl-2* reached 27% **(Figure 4-19 C and D)** allowing a reliable determination of the kinetic constants [**Table 4-6**]. Based on these results, we found that *Oxy2* is the partner with the best affinity (KD_{tetra} 9.16 ± 2.81 nM, KD_{bi} 41.73 ± 13.15 nM), followed by *bcl-2* (375.50 ± 1.46 nM) and *un3* (479.65 ± 2.75 nM). Analyzing in a closer way the binding curves of *Oxy2*, a complex behavior can be noticed, characterized by a slow rise of the association phase and a fast dissociation followed by a slower decrease in the response. This biphasic dissociation can be explained by the presence of two different G4 species formed in solution, as demonstrated by EMSA experiment (**Figure 4-20 C**).

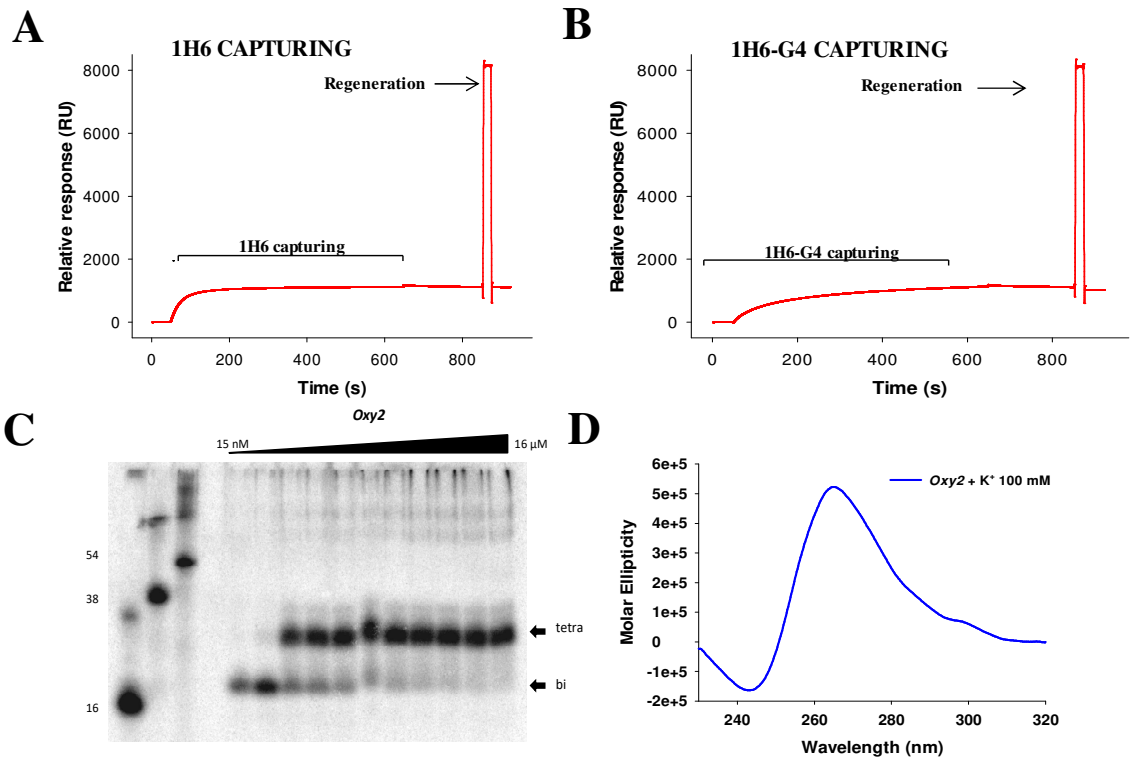

Figure 4-20. ***Association curve of 1H6 capturing procedures and Oxy2 G4 forms.***

A) *1H6 alone, or* ***B)*** *1H6-G4 complex were captured on the sensor chips surface previously immobilized anti-mouse antibody.* ***C)*** *EMSA of Oxy2 G4 folded at different concentrations in*

the same range used for SPR analysis, showing the bi- and tetramolecular form folding in a concentration dependent manner. ***D)*** *CD Spectroscopy: The spectrum was recorded at 20 °C after folding Oxy2 (4 μM) overnight in the presence of 100 mM KCl. The displayed positive peak at 265 nm and a negative peak at 240 nm are typical of a G4 parallel topology.*

Two different species were visible in the gel and were ascribed to the bi- and tetramolecular forms of *Oxy2* G4. The parallel topology recorded by CD spectroscopy for *Oxy2*, folded at a concentration corresponding to the presence of the sole upper band in EMSA native gel, supported the predicted stoichiometry. It has indeed been previously demonstrated that the tetramolecular form of *Oxy2* adopts a parallel conformation, while the dimer folds as antiparallel G4[209–213].

The effective concentration of the two *Oxy2* species was determined by quantification of the corresponding EMSA bands [**Table 4-7**] and used to fit the SPR sensorgrams to a heterogeneous model, taking into account the different species and therefore being suitable for a precise fitting of the experimental curves.

Table 4-7. ***Percentage abundance and estimated concentration of the bi- and tetramolecular G4 species formed by Oxy2 in the presence of 100 mM K^+.***

% abundance		**Concentration (nM)**		
tetra	***bi***	***tetra***	***bi***	**Total**
	100		15.12	15.12
11.22	88.78	3.51	27.74	31.25
59.52	40.48	37.20	25.30	62.50
64.89	35.11	81.11	43.89	125
71.54	28.46	178.85	71.15	250
76.55	23.45	382.75	117.25	500
77.85	22.15	778.50	221.50	1000
80.74	19.26	1614.80	385.20	2000

86.28	13.72	3451.20	548.80	4000
89.16	10.84	7132.80	867.20	8000
92.74	7.26	14838.40	1161.60	16000

The reported values were obtained from the quantification of shift bands obtained by native EMSA experiment **(Figure 4-20 *C*)**.

Altogether, we described a new SPR strategy for the functional immobilization of the anti-G4 Ab 1H6. It consists in a doubly oriented capturing procedure providing several advantages with respect to the standard covalent coupling: i) it allows to work at physiological pH, therefore preserving the integrity and functionality of the ligand; ii) the pH scouting procedure is not necessary allowing to spare time and sample; iii) the captured ligand can be completely regenerated and re-immobilized, so the same flow cell can be used several times; iv) lower ligand volume is required to obtain the same immobilized level. Moreover, the SPR immobilization of a ligand mediated by previous incubation with an analyte can be applied to proteins in general and to molecules that require correct folding. It can also be applied to position on the chip surface every type of molecules so that the binding site is correctly exposed and available for analyte binding.

We employed the described method for the accurate and sensitive characterization of the binding between 1H6 and some selected G4 antigens, demonstrating the suitability of SPR technology to verify the real specificity and affinity of the Ab to various G4 targets. The latter is a fundamental issue for the further exploitation of the Ab in *in vivo* experiments for the monitoring of G4 structures in a cellular context and should be extended to a wider range of targets.

4.4.2 Screening of 1H6 affinity for HSV-1 and human G4s

With the aim of employing 1H6 Ab to visualize HSV-1 G4s in cells, we used the above described capturing method to test 1H6 affinity for three selected HSV-1 G4s over a series of controls. We have previously shown that at least 4 G-rich repeated sequences in the HSV-1 genome can fold into G4 structures *in vitro*: *un2* (antiparallel), *un3* (parallel), *gp054a* (mixed-type) and *un1*, which is however unstable in solution and therefore could not be tested[100]. As

controls we employed: the tetramolecular Oxytrichia telomeric G4 (*Oxy2*), the monomolecular Oxytrichia G4 (*OxyTel*), three different-length human telomeric G4s (*hTel21*, *hTel22* and *hTel54*), two G-rich oligonucleotides corresponding to the scrambled sequences of the human telomeric and HIV-1 LTR promoter G4s (*hTelScra and LTRScra*), one random ssDNA (*ssDNA*), two poly Ts (*T9* and *T30*), one hairpin oligonucleotide (*hp*) and a dsDNA (*dsDNA*).

The absolute binding affinity at the thermodynamic equilibrium could be obtained only for few of these sequences. In particular, *Oxy2* was confirmed to have the best affinity (K_D 62.0 ± 4.2 nM); *hTel54* had a 10-fold lower affinity (K_D 654.5 ± 70.7 nM), slightly lower than *un2* and *un3* (K_D 535.3 ± 20.5 and 479.7 ± 14.1 nM, respectively); *gp054* was 3 times a lesser good binder than *un2* (K_D 1.67 ± 0.46 μM). The other sequences displayed either no or non-specific interaction with 1H6 **(Figure 4-21)**. K_D values represent the binding affinity at the equilibrium. However, other kinetic parameters can be obtained by SPR, such as the binding stability. Stability is a fundamental aspect to be considered when investigating binding partners that should interact in the cellular environment. Even slight variations in the stability of an interaction can be detrimental to target binding. Therefore, to compare all the tested sequences, we measured the stability of 1H6/oligonucleotide right before the end of the dissociation phase, which can be roughly compared to the final wash step in an ELISA assay. In these conditions, the best stabilization was again obtained with *Oxy2*, *un2* and *un3* sequences **(Figure 4-21)**; *gp054*, *hTel54*, *hTel21* and *OxyTel* (all G4 forming sequences) also showed stabilization though to a lower extent; *hTel22* and all non-G4-forming oligonucleotides displayed no stability **(Figure 4-21)**.

These data indicate that two of the three tested HSV-1 G4s are bound at least 30 times better than ssDNA. Interestingly, *un2* and *un3* are the most abundant repeats in all HSV-1 strains. The SPR data also indicate the excellent G4 selectivity of 1H6, which recognizes G4s with different efficiency, depending on the conformation, and does not bind unstructured oligonucleotides[102].

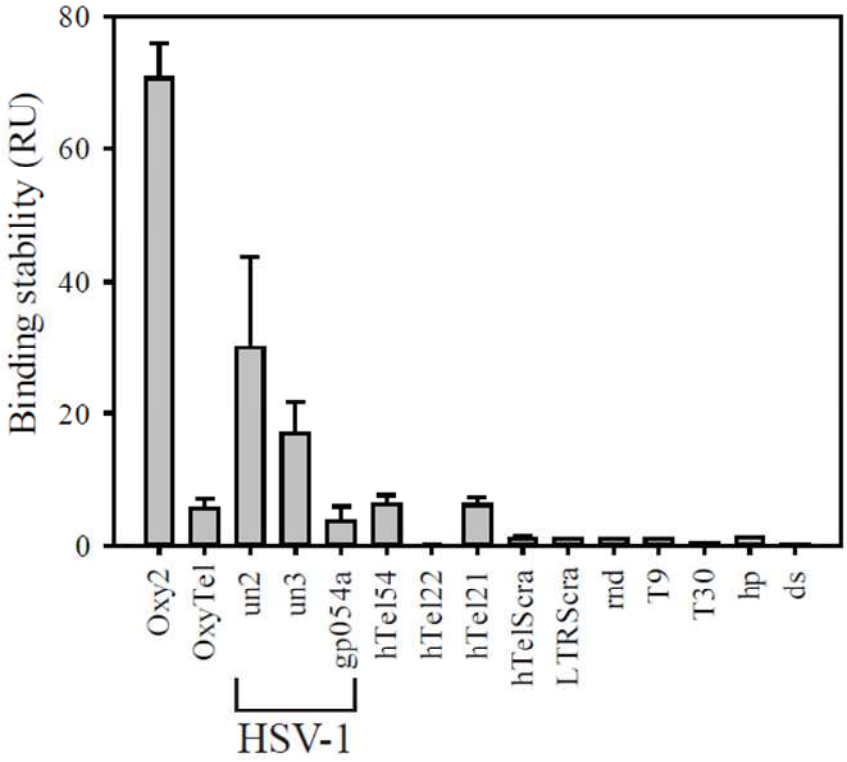

Figure 4-21. ***SPR analysis of 1H6 antibody binding affinity and stability to HSV-1 G4-forming oligonucleotides and control sequences.***

Binding stability of HSV-1 G4s and control oligonucleotides was measured for each sample in the late dissociation phase at the concentration 4 μM . Results are shown as RU (response units) values ± SEM (n = 3). After 1H6 natural antigen Oxy2, the HSV-1 G4s display the strongest binding.

Thanks to the development of a tailored SPR strategy and based on the promising obtained data, showing that 1H6 Ab binds with the best efficiency to HSV-1 G4s with respect to others cellular G4s and non-structured DNA sequences possibly present in a mammalian cell, 1H6 was exploited to visualize HSV-1 G4s in infected cells during different phases of viral infection. Viral G4s appeared to form in a dependent way from the step of the viral life cycle, with a peak during viral DNA replication (6 h.p.i.). At this time point G4s were accumulated in the cellular nucleus, while with the proceeding of virions maturation they migrate to the nuclear membrane[102].

5 CONCLUSIONS

In the present work different aspects of G4 structures' biology were analyzed. In the first part we investigated the presence and role of G4s in the human LPS cancer. As model for this type of cancer we exploited immortalized WD-LPS cells from retroperitoneal origin. WD-LPS is the most common form of human soft tissue sarcoma and it is thought to be the precursor of the more aggressive form of dedifferentiated liposarcoma (DD-LPS)[136]. Therefore, targeting WD-LPS will not only represent a cure for this specific type of LPS, but also allow the prevention of its malignant progression and metastasis. Through a computational prediction algorithm and then biophysical techniques, we identified and characterized a highly stable antiparallel G4 in the inducible promoter of *MDM2* proto-oncogene. *MDM2* is the principal inhibitor of p53 tumor suppressor gene and it is contained in the 12q13-15 region, which is amplified in both WD and DD-LPSs. The consequent overexpression of *MDM2* is the main cause of LPS carcinogenesis[76]. Moreover, *MDM2* transcripts originating from the inducible P2 promoter were reported to have a faster and tumor-associated translational rate[73]. The presence of a stable G4 in close proximity to transcription factor binding sites (e.g. ETS), suggests an involvement of this structure in the regulation of the P2-mediated transcription. To add biological significance to this hypothesis we identified, through a pull-down assay, several proteins with helicase function that specifically recognize and bind to the *MDM2* G4. This result suggests the importance in maintaining the unfolded state of *MDM2* G4 in WD-LPS cells. With this premise, the stabilization of *MDM2* G4 has the potential to impair transcription initiation form the P2 promoter, restoring the p53 apoptotic stimulus and counteracting cancer progression. The antiparallel topology and the four stacked G-quartets that we observed in *MDM2* are two rare conformational features that makes *MDM2* G4 a unique structure. *MDM2* G4 is therefore an ideal target for anticancer therapy.

In addition to the focused G4 characterization in the *MDM2* gene, we also expanded the identification of LPS G4s at the genome-wide level. In this regard we addressed one of the main challenges in the study of G4 structures: the *in vivo* evidence of their formation and function. Employing the scFv anti-G4 antibody BG4[55] in a ChIP-seq experiment, we mapped about 4800 G4s in the genome of WD-LPS cells. The non-random distribution and overrepresentation of G4s in gene promoters supports their involvement in the cell-type specific regulation of transcription. A huge amount of *in vitro* data exploring G4 function in this process is available. However, a clear picture is still missing and contrasting data can be found. Here

we observed a strong, genome-wide correlation between the presence of G4s in gene promoters and active, high level transcription. This is the first time that G4s are proposed as positive marks for active transcription. Instead, despite generally considered spatial obstacles for RNA polymerase processivity, several mechanisms through which G4s can facilitate transcription have been already hypothesized: i) their formation helps to keep the partially single stranded transcriptional bubble open, and ii) when downstream from the TSS, G4s can facilitate the promoter escape process[188], destabilizing the early transcriptional complex.

In support to the G4 role at the transcriptome level, the comparison of the LPS cell line with prooncogenic keratinocytes. Looking at the differentially expressed genes, we found that generally the same gene in the two cell lines possess a folded G4 structure when it has a high expression level, while the G4 is not folded when the gene is transcriptionally repressed. This result is the first indication of the cell-specific modulation of the G4 state as a mean to regulate the transcriptional output. How G4 folding is related to the chromatin state is still unclear and further experiments will thus be necessary.

Moreover, we observed that the stabilization of G4s through the potent G4 ligand H-NDI-NMe_2PhAm led to a massive alteration of the LPS transcriptome, driving to cell death. The main observed effect on G4 containing genes was their downregulation, in accordance with the majority of *in vitro* published data. This effect can be attributed to the ability of G4s to compensate for the negative superhelicity accumulated behind the transcriptional machinery: blocking the G4s in their conformation causes a reduction in the DNA intrinsic flexibility and impairs the polymerase progression. In summary, these data represent one of the firsts attempts to study G4s in an *in vivo* system and give a genome-wide overview of how G4s correlates with gene transcription in WD-LPS. Together with the discovery and characterization of *MDM2* G4, these data increase the understanding of the low explored biology of LPS cancer and open the possibility for testing new therapeutic targets.

In the second part of this work we focused on the study of G4s in the genome of two worldwide-spread viruses: HIV-1 and HSV-1. A cure for the infectious diseases that they cause is still lacking and the research for new therapeutic approaches is needed. Our group recently discovered a dynamic G4 cluster in the LTR promoter of HIV-1[93]. The high affinity interaction with the cellular protein NCL to this G4 was able to suppress viral transcription[97]. We here analyzed the different G4 structural features (loop number, sequence and length, G4 conformation and stability) that make NCL interaction specific for some G4s. We found that

G4 loop length is the main determinant for NCL binding and that more flexible G4s are favored. This result perfectly explains the high affinity between NCL and the LTR G4s. Among the mutually exclusive G4s that can fold in the HIV-1 LTR, NCL binds with higher affinity to LTR-III which possess a long loop[97]. Moreover, the dynamic equilibrium of the different G4 species in the LTR makes it a flexible structure that can be adapted to the spatial occupancy of the protein DNA-binding domain. This result is of importance also from a therapeutic point of view: the understanding of the G4 features necessary for NCL interaction can be exploited for the targeted design of chemical compounds mimicking NCL action and repressing the HIV-1 transcription through the stabilization of the LTR-G4. This approach represents an innovative antiviral strategy that directly acts on the proviral genome and, in combination with HAART, can improve the antiviral efficacy and reduce the development of resistant strains.

NCL is a key protein in the HIV-1 life cycle, not only at the proviral level but also for the entry step[204]. NCL is indeed a coreceptor for viral attachment to the membrane of permissive cells. We here demonstrated that the G4-forming aptamer AS1411 is able to sequester the membrane NCL, impairing HIV-1 entry. AS1411 already reached phase II clinical trial as antitumoral agent[112], its efficacy and high tolerability makes the G4 aptamer a promising alternative to the current entry inhibitors.

Apart from HIV-1, our group also found a huge number of G4-forming sequences in the G-rich genome of HSV-1[101]. We exploited this unique feature of HSV-1 aiming at visualizing the modulation of the viral G4s in infected cells. Through a newly developed immobilization strategy, we here determined that the anti-G4 monoclonal antibody 1H6[102] displays higher affinity to the most represented HSV-1 G4s than to the cellular ones. We therefore reasoned that 1H6 would have been the optimal mean to visualize viral G4s by IF in infected cells. We observed a viral infection step-dependent formation of G4s, which peak during replication and move from the nucleoplasm to the nuclear membrane together with the maturation and nuclear egress of the virion[102]. These results represent an important step forward in the understanding of the biological modulation of G4s in viral genomes. Moreover, the developed SPR protocol represent a valuable alternative immobilization strategy that can be theoretically applied to all ligands that are inactivated by the standard covalent couplings, allowing to specifically orient the ligand from two sides, spare material, time and keep delicate molecules in a more physiologic condition[214].

6 ACKNOWLEDGEMENTS

Reaching the end of a Ph.D. program is a tortuous path that seems too long and too fast at the same time. It makes you grow from a scientific, but also, and most importantly from a personal point of view. You are faced every day with different kinds of problems and you have to deal with may persons and be a group with people that you didn't chose.

If you chose this path is probably because you have a great passion for science, and therefore you are willing to invest all your energies for give your best. But you'll learn that sometimes this isn't sufficient. You must accept that sometimes things do not work as you wanted, and you must start again several times before to find the correct solution. You must learn to don't get angry when is Friday evening, everyone but you is out with friends and you are still working for an experiment that won't work. You will learn that science is 90 % (to be optimistic) unsuccess.

A Ph.D *iter* will reveal all the worst of you in the attempt to make you best at work and in everyday life. Sometimes you'll feel alone and lost, some others you'll feel part of a family that always supports and understand you.

For these reasons I want first of all thank all the persons that crossed the big family of Prof. Sara Richter's group: Elena Tosoni, who followed me during the Master thesis period, never leaving me alone and teaching me to be rigorous an focused; Matteo Nadai, the wise and mysterious Delphi's oracle who is always there when you really need help in science and in life; Ilaria Frasson who, as a mother, was able to bring as up in one of the most difficult period of the Richter's group; Martina Tassinari, for being my second problematic half; Emanuela Ruggiero (Manù) for being always able and secure in finding a positive way to afford obstacles and for being a person able to listen and advise; Paola Soldà for her diplomacy and availability, Matteo Scalabrin for his crazy ideas; and all the ones that changed their way: Sara Callegaro, Rosalba Perrone, Elena Butovskaya, Giulia Zorzan, Elena Trallori and finally, of course, Prof. Sara Richter who believed in me and pushed me to start the Ph.D. with her group.

The people that had a role in making me what I am today, after these three years are however many and I hope not to forget anyone. I want to thank Prof. Francesca Demichelis, Alessandro Romanel and Davide Prandi for introducing me to the scientific approach. I realized only later how useful it has been and how much I learnt during those few months in their group.

I also want to thank Prof. Laura Cendron for supporting me in the difficult choice of starting a Ph.D.. Several times during these three years I found again her words in my mind.

I want to thank Prof. Gunnar Schotta, Mana Kazerani, Helena Moreno, Sophia Groh, Lisa Marinelli, Filippo Cernilogar, Zeyang Wang, Christine Klement, Irina Schcerbakova and the BMC coffee team for hosting me, teaching me lot of things and making me really feel one of them. Thanks to Prof. Peter Becker because I really admire how he leads his department, and to the administration office for the incredible and unexpected availability.

I would like to thank Prof.Stefano Moro, for his extreme kindness and for what he tried to for my happiness.

Thanks to Prof. Stefano Mancuso for the great passion for science that he invests and transmit.

Thanks to Monica Rossetto, Paola Vanzani, Maria Luisa Di Paolo and Lucio Zennaro for supporting me when doing SPR at Vallisneri. In this regard also I want to thank Jordy, who was always able to make me laugh and sharing the desperation when working with Biacore diabolic machine. Thanks to Néstor Santiago Gonzaléz Biacore specialist for the support and advises.

Besides, I also must thank my friends, because doing a PhD is more than a job, is a way of living and it influences your life more than you know and more than you can imagine. You became always busy, always stressed and only those friend that really care about you will remain by your side until the end. Therefore, I must thank Ilaria Cagol, for being my best friend, for never abandon me after all these years (almost 17) even if I live far, even if I don't give notices of me very often and even if I am always reserved and don't like to tell her about my problems and thoughts, which can be frustrating for a best friend. So thank you Illy for accepting me for what I am. Then, I have to thank Elisa Mazzucco e Giorgia Cantisani for having helped me in one of the worst moments in my life. I would like to thank Luca Degasper (Deguz) because we see each other few times every year, but is always fun and always like no time has passed. Thanks to Marco Moretti (you'll always be Mr Metal for me) because I understand I am not the best friend, but you are one of the few who are still there for me. I would like to thank Martin Schiavon, Matteo dell'Osso, Tommaso Caldognetto and Mirko Bez because you helped me even if I am not a key person in your life.

I would like to thank Andrea Vantini, Matteo Mongardini and Walter Dughera to educate me (everyone in a different way) to run with all my energies, to run in a rational way and to resist until the last step, or better still until one step after the finish line.

A special thank is for Francesco, for his patience, for his calmness, for helping me with everything, but most of all for putting me back on the right way, for making me grow, for making me realize that I don't need anyone to make decisions, to carry my life forward and imposing me (in the worst way possible) to be strong, always, and to face difficulties with determination and courage.

And finally, I want to thank my parents for always being happy to see me, even if many times I don't think I deserve all what they do for me. They always love me, despite I am often unpleasant, stressed and focused on my plans. I hope this work can be enough to justify some of my behaviours and for always being away from my hometown and to make you proud of me, at least a bit.

Thank you.

7 APPENDIX

7.1 List of recurrent abbreviations

G4	G-quadruplex
UTR	Untranslated Region
IRES	Internal Ribosome Entry Site
TSS	Transcription Start Site
TTS	Transcription Termination Site
NCL	Nucleolin
DNMTs	DNA methyltransferases
HIV-1	Human Immunodeficiency Virus 1
HSV-1	Herpes Simplex Virus 1
AIDS	Acquired Immunodeficiency Syndrome
LTR	Long Terminal Repeat
G	Guanine
c-exNDIs	core-extended Naphtalene Diimides
HAART	Highly Active Anti-Retroviral Therapy
LPS	Liposarcoma
WD-LPS	Well-differentiated Liposarcoma
DD-LPD	Dediffentiated Liposarcoma

CD	Circular Dichroism
DMS	Dimethylsulphate
Tm	Melting Temperature
MS	Mass Spectroscopy
Ab	Antibody
RT	Room Temperature
TPM	Transcript Per Million
IF	Immunofluorescence
FC	Fold Change

7.2 List of used oligonucleotides

Name	Sequence (5'-3')
CD, DMS footprinting and protein pull-down – MDM2	
Mdm2-G4	GTTCGTGGCTG GGGGCTCGGGGCGCGGGGCGCGGGGCATGGGGC
Mdm2-biot	BtnTg-TGGGGGCTCGGGGCGCGGGGCGCGGGGCATGGGGC
G-ss-biot	Btn-TTTTTGGAGTCGTGTCGCGTGTCGAGCGTGTGTAGTGGTTTTT
C-ss-biot	Btn-AAAAACCCCAGTCCCGCCCAGGCCACGCCTCCCAAAAA
ChIP-qPCR primers	
GAPDH	FW: GCTACTAGCGGTTTTACGGGCG – RV: TGCGGCTGACTGTCGAACAGG
EIF4A	FW: CCGGAGCGACTAGGAACTAAC – RV: GCCTTTCTTACCGGGAATCCT
MDM2	FW: GGATTTCGGACGGCTCTCG - RV: CGTTCACACTAGTGACCCGA
CDK4	FW: CCACCCTCACCATGTGACC - RV: CTTACACTCTTCGCCCTCCTC
TMCC2	FW: CCAGACACTTTGGGTGACCT – RV: AACACCTGCTCTGCCAACTT
MAP3K13	FW: GACATAGGAACGGGCAAAGA – RV: CCCATGCTGTATGTGGTCTG
ESR1	FW: GAAACAGCCCCAAATCTCAA – RV: TTGTAGCCAGCAAGCAAATG
LRRN4	FW: GAGGCTGGGATCTCAGTGTTCGG – RV: TACTCTCTGAACCAAGGGGCACT
EMSA and CD - NCL	
(GGGTTA)$_4$	GGGTTAGGGTTAGGGTTAGGG
(GGGTTA)$_6$	GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
(GGGTTA)$_8$	TCGAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAC
(GGGTTT)$_4$	GGGTTTGGGTTTGGGTTTGGG
(GGGTTT)$_6$	GGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTT
(GGGTTT)$_8$	GGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTT
(GGGTTT)$_5$	GGGTTTGGGTTTGGGTTTGGGTTTGGG
n5M1	TTTTTTGGGTTTGGGTTTGGGTTTGGG
n5M2	GGGTTTTTTTTTGGGTTTGGGTTTGGG
n5M3	GGGTTTGGGTTTTTTTTTGGGTTTGGG
n5M4	GGGTTTGGGTTTGGGTTTTTTTTTGGG

n5M5	GGGTTTGGGTTTGGGTTTGGGTTTTTT
M43	TTTTTGGAGTCGTGTCGCGTGTCGAGCGTGTGTAGTGGTTTTT
M30	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
M29i	AAAACAGACAGCTCACTGCGCTCACAAAA
M29ii	TTTTGTGAGCGCGGTGAGGTGTGTGTTTT
M26	CCTCCTCCTCCTTCTCCTCCTCCTCC
M21	TTTTTTTTTTTTTTTTTTTTT
M14	CGGCATGCGTGGCT
111	GGGTGGGTGGGTGGG
131	GGGTGGGTTTGGGTGGG
161	GGGTGGGTTTTTTGGGTGGG
313	GGGTTTGGGTGGGTTTGGG
333	GGGTTTGGGTTTGGGTTTGGG
363	GGGTTTGGGTTTTTTGGGTTTGGG
616	GGGTTTTTTGGGTGGGTTTTTTGGG
636	GGGTTTTTTGGGTTTGGGTTTTTTGGG
666	GGGTTTTTTGGGTTTTTTGGGTTTTTTGGG
T111T	TGGGTGGGTGGGTGGGT
T131T	TGGGTGGGTTTGGGTGGGT
T161T	TGGGTGGGTTTTTTGGGTGGGT
c-kit1	AGGGAGGGCGCTGGGAGGAGG
bcl-2	AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG
c-myc	TGGGGAGGGTGGGGAGGGTGGGGAAGG
hTert 1–4	AGGGGAGGGGCTGGGAGGGC
hTert 5–12	AGGGGGCTGGGCCGGGGACCCGGGAGGGGTCGGGACGGGGCGGGGT
hTert 1–12	AGGGGAGGGGCTGGGAGGGCCCGGAGGGGGCTGGGCCGGGGACCCGGGAGGGG TCGGGACGGGGCGGGGT
Surface Plasmon Resonance – AS1411	
AS1411	GGTGGTGGTGGTTGTGGTGGTGGTGG
CRO26	CCTCCTCCTCCTTCTCCTCCTCCTCC
LTR-III	GGGAGGCGTGGCCTGGGCGGGACTGGGG
SCRA	TTTTTGGAGCGTGTGTGCGCGAGAGCGTGCGCGTGGCGAGCGTGGAGTGGTTTTT
Surface Plasmon Resonance - 1H6	

bcl-2	AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG
gp054a	GGGGTTGGGGCTGGGGTTGGGG
un2	GGGGGCGAGGGGCGGGAGGGGGCGAGGGG
un3	GGGAGGAGCGGGGGGAGGAGCGGG
Oxy2	TTTTGGGGTTTTGGGG
OxyTel	GGGGTTTTGGGGTTTTGGGGTTTTGGGG
hTel21	GGGTTAGGGTTAGGGTTAGGG
hTel22	AGGGTTAGGGTTAGGGTTAGGG
hTel54	TCGAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAC
hTelScra	GGATGTGAGTGTGAGTGTGAGG
LTRScra	TTTTTGGAGCGTGTGTGCGCGAGAGCGTGCGCGTGGCGAGCGTTGAGTGGTTTTT
ssDNA	AAAAACTACTGCACGCTCGCTACGACGACACTGTCGCGCATACAAGCTGCAAAAA
T9	TTTTTTTTT
T30	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
hp	CGCAGCGTGGCTTTGTTTGCCACGCTGCG
dsDNA	hTelScra+complementary sequence

7.3 Index of figures and tables

Figure 1-1. *Alternative forms of DNA.* 5

Figure 1-2. *G-quadruplex polymorphism.* 7

Figure 1-3. *Role of G4 structures in the genome.* 9

Figure 1-4. *G4 mediated regulation of transcription.* 11

Figure 1-5. *G-quadruplex structures role in regulatory chromatin.* 13

Figure 1-6. *G-quadruplexes in the hallmarks of cancer and LPS.* 15

Figure 1-7. *HIV-1 life cycle, genome organization and G4s.* 18

Figure 1-8. *HSV-1 life cycle, genome organization and G4s.* 20

Figure 1-9. *Chemical structures of some G4 ligands.* 22

Figure 4-1. *Position and score of the putative G4s predicted by QGRS-Mapper.* 40

Figure 4-2. *CD characterization of Mdm2-G4 folding and thermal stability.* 41

Figure 4-3. *DMS protection assay.* 42

Figure 4-4. *BG4 chromatin immunoprecipitation of G4 structures.* 48

Figure 4-5. *LPS G4s position and annotation.* 49

Figure 4-6. *Transcriptional state of G4-containing genes:* 51

Figure 4-7. *Transcriptional effect of G4 position.* 52

Figure 4-8. *G4 comparison in 93T449 vs. HaCaT cell lines.* 54

Figure 4-9. *Figure. Analysis of gene expression after treatment with H-NDI-NMe$_2$Pham G4 ligand.* 57

Figure 4-10. *NCL binding to G4s with increasing number of G-tracts.* 60

Figure 4-11. *Analysis of NCL binding to alternative G4 structures formed by the (GGGTTT)$_5$ telomeric sequence* 62

Figure 4-12. *NCL binding to G4s with systematic elongation of loop length.* 64

Figure 4-13. *Analysis of monomeric and dimeric G4 folding of relevant oligonucleotides.* 66

Figure 4-14. *Analysis of NCL binding towards oligonucleotides with variable G4s conformation.* 67

Figure 4-15. *NCL binding to human oncogene promoter G4s.* 68

Figure 4-16. *NCL binding to G4s in special conditions.* 69

Figure 4-17. *Binding affinity of NCL and gp120 measured by SPR.* 72

Figure 4-18. *Binding analysis of 1H6-G4 and control nucleic acids interaction after.* 74

Figure 4-19. *Sensor chip surface activity.* 76

Figure 4-20. *Association curve of 1H6 capturing procedures and Oxy2 G4 forms.* 77

Figure 4-21. *SPR analysis of 1H6 antibody binding affinity and stability to HSV-1 G4-forming oligonucleotides and control sequences.* 81

Table 4-1. *Putative G4 sequences identified by QGRS-Mapper.* 39

Table 4-2. *CD topological and thermal analysis of Mdm2-G4 oligonucleotide.* 42

Table 4-3. *MS identification of specific Mdm2-G4 binding proteins.* 45

Table 4-4. *Stability (T_m values) of oligonucleotides of the $(GGGTTT)_5$ series and mutants.* 62

Table 4-5. *Stability (T_m values) of oligonucleotides of the $(GGGTTT)_4$ series.* 63

Table 4-6. *Kinetic parameters for 1H6-G4 interaction.* 75

Table 4-7. *Percentage abundance and estimated concentration of the bi- and tetramolecular G4 species formed by Oxy2 in the presence of 100 mM K^+.* 78

8 BIBLIOGRAPHY

1. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 737–738 (1953).

2. Franklin, R. E. & Goslino, X. R. G. *The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content*. *Acta Cryst* **6,** (1953).

3. Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **16,** 907–916 (1963).

4. Harvey, S. C., Luo, J. & Lavery, R. DNA stem-loop structures in oligopurine-oligopyrimidine triplexes. *Nucleic Acids Res.* **16,** 11795–11809 (1988).

5. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. Crystal structure of four-stranded Oxytricha telomeric DNA. *Nature* **356,** 126–131 (1992).

6. Lilley, D. M. J. *The inverted repeat as a recognizable structural feature in supercoiled DNA molecules (SI single-strand-specific nuclease/ColEl, pBR322, and 4X174 replicative form DNA conformation/nucleic acid protein recognition). Proc. Natl. Acad. Sci.* **77,** (1980).

7. Day, H. A., Pavlou, P. & Waller, Z. A. E. i-Motif DNA: Structure, stability and targeting with ligands. *Bioorg. Med. Chem.* **22,** 4407–4418 (2014).

8. Zhou, H. *et al.* New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic Acids Res.* **43,** 3420–3433 (2015).

9. Brázda, V., Coufal, J., Brázda, V. & Coufal, J. Recognition of Local DNA Structures by p53 Protein. *Int. J. Mol. Sci.* **18,** 375 (2017).

10. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* **48,** 2013–8 (1962).

11. Tassinari, M. *et al.* A Fragment-Based Approach for the Development of G-Quadruplex Ligands: Role of the Amidoxime Moiety. *Molecules* **23,** 1874 (2018).

12. Ruggiero, E. & Richter, S. N. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.* **46,** 3270–3283 (2018).

13. Tippana, R., Xiao, W. & Myong, S. G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **42,** 8106–8114 (2014).

14. Pascale Hazel, Julian Huppert, Shankar Balasubramanian, and & Stephen Neidle*. Loop-Length-Dependent Folding of G-Quadruplexes. (2004). doi:10.1021/JA045154J

15. Bugaut, A. & Balasubramanian, S. A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes †. *Biochemistry* **47,** 689–697 (2008).

16. Guédin, A., De Cian, A., Gros, J., Lacroix, L. & Mergny, J.-L. Sequence effects in single-base loops for quadruplexes. *Biochimie* **90,** 686–96 (2008).

17. Wong, H. M., Stegle, O., Rodgers, S. & Huppert, J. L. A toolbox for predicting g-quadruplex formation and stability. *J. Nucleic Acids* **2010,** (2010).

18. Hershman, S. G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic Acids Res.* **36,** 144–156 (2008).

19. Cheung, I., Schertzer, M., Rose, A. & Lansdorp, P. M. Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. *Nat. Genet.* **31,** 405–409 (2002).

20. He, Q. *et al.* G-quadruplex-mediated regulation of telomere binding protein POT1 gene expression. *Biochim. Biophys. Acta - Gen. Subj.* **1840,** 2222–2233 (2014).

21. Yadav, V., Hemansi, Kim, N., Tuteja, N. & Yadav, P. G Quadruplex in Plants: A Ubiquitous Regulatory Element and Its Biological Relevance. *Front. Plant Sci.* **8,** 1163 (2017).

22. Thakur, R. S. *et al. Mycobacterium tuberculosis* DinG Is a Structure-specific Helicase That Unwinds G4 DNA. *J. Biol. Chem.* **289,** 25112–25136 (2014).

23. Wieland, M. & Hartig, J. S. Investigation of mRNA quadruplex formation in Escherichia coli. *Nat. Protoc.* **4,** 1632–1640 (2009).

24. Ruggiero, E. & Richter, S. N. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.* **46,** 3270–3283 (2018).

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

26. Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **277,** 1107–1117 (2010).

27. Lim, K. W., Ng, V. C. M., Martín-Pintado, N., Heddi, B. & Phan, A. T. Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.* **41,** 10556–10562 (2013).

28. Heddi, B. & Phan, A. T. Structure of Human Telomeric DNA in Crowded Solution. *J. Am. Chem. Soc.* **133,** 9824–9833 (2011).

29. Smith, J. S. *et al.* Rudimentary G-quadruplex–based telomere capping in Saccharomyces cerevisiae. *Nat. Struct. Mol. Biol.* **18,** 478–485 (2011).

30. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Mol. Cell* **35,** 403–413 (2009).

31. Schaeffer, C. *et al.* The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **20,** 4803–4813 (2001).

32. Arora, A. *et al.* Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA* **14,** 1290–1296 (2008).

33. Bonnal, S. *et al.* 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. *J. Biol. Chem.* **278,** 39330–39336 (2003).

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

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

36. Valton, A.-L. *et al.* G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* **33,** 732–746 (2014).

37. Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22,** 2399–408 (2012).

38. López Castel, A., Cleary, J. D. & Pearson, C. E. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* **11,** 165–170 (2010).

39. Larson, E. D., Duquette, M. L., Cummings, W. J., Streiff, R. J. & Maizels, N. MutSα Binds to and Promotes Synapsis of Transcriptionally Activated Immunoglobulin Switch Regions. *Curr. Biol.* **15,** 470–474 (2005).

40. Métifiot, M., Amrane, S., Litvak, S. & Andreola, M.-L. G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res.* **42,** 12352–12366 (2014).

41. Rigo, R., Palumbo, M. & Sissi, C. G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim. Biophys. Acta - Gen. Subj.* **1861,** 1399–1413 (2017).

42. González, V. & Hurley, L. H. The c-MYC NHE III(1): function and regulation. *Annu. Rev. Pharmacol. Toxicol.* **50,** 111–29 (2010).

43. González, V. & Hurley, L. H. The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry* **49,** 9706–14 (2010).

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

45. Uribe, D. J., Guo, K., Shin, Y.-J. & Sun, D. Heterogeneous Nuclear Ribonucleoprotein K and Nucleolin as Transcriptional Activators of the Vascular Endothelial Growth Factor Promoter through Interaction with Secondary DNA Structures. *Biochemistry* **50,** 3796–3806 (2011).

46. Wei, P.-C. *et al.* A cis-element with mixed G-quadruplex structure of NPGPx promoter is essential for nucleolin-mediated transactivation on non-targeting siRNA stress. *Nucleic Acids Res.* **41,** 1533–1543 (2013).

47. Baranello, L., Levens, D., Gupta, A. & Kouzine, F. The importance of being supercoiled: How DNA mechanics regulate dynamic processes. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819,** 632–638 (2012).

48. Brooks, T. A. & Hurley, L. H. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer* **9,** 849–861 (2009).

49. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and Mechanism of Helicases and Nucleic Acid Translocases. *Annu. Rev. Biochem.* **76,** 23–50 (2007).

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

51. Agarwal, T., Roy, S., Kumar, S., Chakraborty, T. K. & Maiti, S. In the Sense of Transcription Regulation by G-Quadruplexes: Asymmetric Effects in Sense and Antisense Strands. *Biochemistry* **53,** 3711–3718 (2014).

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

53. Nguyen, G. H. *et al.* Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proc. Natl. Acad. Sci.* **111,** 9905–9910 (2014).

54. Armas, P. & Calcaterra, N. B. G-quadruplex in animal development: Contribution to gene expression and genomic heterogeneity. *Mech. Dev.* (2018). doi:10.1016/j.mod.2018.05.004

55. Hänsel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48,** 1267–1272 (2016).

56. Halder, R. *et al.* Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. *Mol. Biosyst.* **6,** 2439 (2010).

57. Baral, A., Kumar, P., Pathak, R. & Chowdhury, S. Emerging trends in G-quadruplex biology – role in epigenetic and evolutionary events. *Mol. Biosyst.* **9,** 1568 (2013).

58. Schiavone, D. *et al.* Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J.* **33,** 2507–2520 (2014).

59. Guilbaud, G. *et al.* Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat. Chem.* **9,** 1110–1117 (2017).

60. Maizels, N. G4-associated human diseases. *EMBO Rep.* **16,** 910–922 (2015).

61. Rossi, M. L., Ghosh, A. K. & Bohr, V. A. Roles of Werner syndrome protein in protection of genome integrity. *DNA Repair (Amst).* **9,** 331–344 (2010).

62. Vannier, J.-B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I. R., Ding, H. & Boulton, S. J. RTEL1 Dismantles T Loops and Counteracts Telomeric G4-DNA to Maintain Telomere Integrity. *Cell* **149,** 795–806 (2012).

63. Lin, W. *et al.* Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *EMBO J.* **32,** 1425–39 (2013).

64. Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.* **273,** 27587–92 (1998).

65. Castillo Bosch, P. *et al.* FANCJ promotes DNA synthesis through G-quadruplex structures. *EMBO J.* **33,** 2521–33 (2014).

66. Wickramasinghe, C. M., Arzouk, H., Frey, A., Maiter, A. & Sale, J. E. Contributions of the specialised DNA polymerases to replication of structured DNA. *DNA Repair (Amst).* **29,** 83–90 (2015).

67. Fratta, P. *et al.* C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.* **2,** 1016 (2012).

68. Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507,** 195–200 (2014).

69. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).

70. Takahama, K., Kino, K., Arai, S., Kurokawa, R. & Oyoshi, T. Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS J.* **278,** 988–998 (2011).

71. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **277,** 3459–3469 (2010).

72. Nguyen, D., Liao, W., Zeng, S. X. & Lu, H. Reviving the guardian of the genome: Small molecule activators of p53. *Pharmacol. Ther.* **178,** 92–108 (2017).

73. Jin, X., Turcott, E., Englehardt, S., Mize, G. J. & Morris, D. R. The two upstream open reading frames of oncogene mdm2 have different translational regulatory properties. *J. Biol. Chem.* **278,** 25716–21 (2003).

74. Wade, M., Li, Y.-C. & Wahl, G. M. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nat. Rev. Cancer* **13,** 83–96 (2013).

75. Stiller, C. A. *et al.* Descriptive epidemiology of sarcomas in Europe: Report from the RARECARE project. *Eur. J. Cancer* **49,** 684–695 (2013).

76. Patel, R. B. *et al.* Recent translational research into targeted therapy for liposarcoma. *Stem Cell Investig.* **4,** 21–21 (2017).

77. Oliner, J. D., Saiki, A. Y. & Caenepeel, S. The Role of MDM2 Amplification and Overexpression in Tumorigenesis. *Cold Spring Harb. Perspect. Med.* **6,** a026336 (2016).

78. Barak, Y., Gottlieb, E., Juven-Gershon, T. & Oren, M. Regulation of mdm2 expression by p53: alternative promoters produce transcripts with nonidentical translation potential. *Genes Dev.* **8,** 1739–49 (1994).

79. Phelps, M., Darley, M., Primrose, J. N. & Blaydes, J. P. p53-independent activation of the hdm2-P2 promoter through multiple transcription factor response elements results in elevated hdm2 expression in estrogen receptor alpha-positive breast cancer cells. *Cancer Res.* **63,** 2616–23 (2003).

80. Ding, Q. *et al.* Discovery of RG7388, a Potent and Selective p53–MDM2 Inhibitor in Clinical Development. *J. Med. Chem.* **56,** 5979–5983 (2013).

81. Norseen, J., Johnson, F. B. & Lieberman, P. M. Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.* **83,** 10336–46 (2009).

82. Tellam, J. T. *et al.* mRNA Structural Constraints on EBNA1 Synthesis Impact on In Vivo Antigen Presentation and Early Priming of CD8+ T Cells. *PLoS Pathog.* **10,** e1004423 (2014).

83. Madireddy, A. *et al.* G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. *Nucleic Acids Res.* **44,** 3675–3694 (2016).

84. Gilbert-Girard, S. *et al.* Stabilization of Telomere G-Quadruplexes Interferes with Human Herpesvirus 6A Chromosomal Integration. *J. Virol.* **91,** (2017).

85. Barré-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220,** 868–71 (1983).

86. METLAS, R. & VELJKOVIC, V. HIV-1 gp120 AND IMMUNE NETWORK. *Int. Rev. Immunol.* **23,** 413–422 (2004).

87. Freed, E. O. HIV-1 Gag Proteins: Diverse Functions in the Virus Life Cycle. *Virology* **251,** 1–15 (1998).

88. Han, Y., Wind-Rotolo, M., Yang, H.-C., Siliciano, J. D. & Siliciano, R. F. Experimental approaches to the study of HIV-1 latency. *Nat. Rev. Microbiol.* **5,** 95–106 (2007).

89. Simonetti, F. R. & Kearney, M. F. Review. *Curr. Opin. HIV AIDS* **10,** 49–54 (2015).

90. Marquet, R., Paillart, J. C., Skripkin, E., Ehresmann, C. & Ehresmann, B. Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res.* **22,** 145–51 (1994).

91. Shen, W., Gorelick, R. J. & Bambara, R. A. HIV-1 nucleocapsid protein increases strand transfer recombination by promoting dimeric G-quartet formation. *J. Biol. Chem.* **286,** 29838–47 (2011).

92. Perrone, R. *et al.* Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 nef Coding Region: Implications for Antiviral Activity. *PLoS One* **8,** e73121 (2013).

93. Perrone, R. *et al.* A Dynamic G-Quadruplex Region Regulates the HIV-1 Long Terminal Repeat Promoter. *J. Med. Chem.* **56,** 6521–6530 (2013).

94. Amrane, S. *et al.* Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development. *J. Am. Chem. Soc.* **136,** 5249–5252 (2014).

95. Perrone, R. *et al.* Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J. Antimicrob. Chemother.* **69,** 3248–3258 (2014).

96. De Nicola, B. *et al.* Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic Acids Res.* **44,** 6442–51 (2016).

97. Tosoni, E. *et al.* Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* **43,** 8884–8897 (2015).

98. Scalabrin, M. *et al.* The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Sci. Rep.* **7,** 45244 (2017).

99. Kukhanova, M. K., Korovina, A. N. & Kochetkov, S. N. Human herpes simplex virus: Life cycle and development of inhibitors. *Biochem.* **79,** 1635–1652 (2014).

100. Artusi, S. *et al.* The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: Implications for the antiviral activity of a G-quadruplex ligand. *Antiviral Res.* **118,** 123–131 (2015).

101. Artusi, S. *et al.* The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: Implications for the antiviral activity of a G-quadruplex ligand. *Antiviral Res.* **118,** 123–131 (2015).

102. Artusi, S. *et al.* Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. *Nucleic Acids Res.* **44,** gkw968 (2016).

103. Drygin, D. *et al.* Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Res.* **69,** 7653–7661 (2009).

104. Boddupally, P. V. L. *et al.* Anticancer Activity and Cellular Repression of c-MYC by the G-Quadruplex-Stabilizing 11-Piperazinylquindoline Is Not Dependent on Direct Targeting of the G-Quadruplex in the c-MYC Promoter. *J. Med. Chem.* **55,** 6076–6086 (2012).

105. Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H. & Brooks, T. A. Demonstration that Drug-targeted Down-regulation of MYC in Non-Hodgkins Lymphoma Is Directly Mediated through the Promoter G-quadruplex. *J. Biol. Chem.* **286,** 41018–41027 (2011).

106. Wang, X.-D. *et al.* Turning off Transcription of the *bcl-2* Gene by Stabilizing the *bcl-2* Promoter Quadruplex with Quindoline Derivatives. *J. Med. Chem.* **53,** 4390–4398 (2010).

107. McLuckie, K. I. E. *et al.* G-Quadruplex-Binding Benzo[*a*]phenoxazines Down-Regulate *c-KIT* Expression in Human Gastric Carcinoma Cells. *J. Am. Chem. Soc.* **133,** 2658–2663 (2011).

108. Neidle, S. Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat. Rev.*

Chem. **1,** 0041 (2017).

109. Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **59,** 5987–6011 (2016).

110. Burger, A. M. *et al.* The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* **65,** 1489–1496 (2005).

111. Callegaro, S. *et al.* A core extended naphtalene diimide G-quadruplex ligand potently inhibits herpes simplex virus 1 replication. *Sci. Rep.* **7,** 2341 (2017).

112. Girvan, A. C. *et al.* AGRO100 inhibits activation of nuclear factor- B (NF- B) by forming a complex with NF- B essential modulator (NEMO) and nucleolin. *Mol. Cancer Ther.* **5,** 1790–1799 (2006).

113. Ecker, D. J., Vickers, T. A., Hanecak, R., Driver, V. & Anderson, K. Rational screening of oligonucleotide combinatorial libraries for drug discovery. *Nucleic Acids Res.* **21,** 1853–6 (1993).

114. Buckheit, R. W. *et al.* Potent and specific inhibition of HIV envelope-mediated cell fusion and virus binding by G quartet-forming oligonucleotide (ISIS 5320). *AIDS Res. Hum. Retroviruses* **10,** 1497–506 (1994).

115. Abhijit Mazumder, ‡ *et al.* Inhibition of the Human Immunodeficiency Virus Type 1 Integrase by Guanosine Quartet Structures†. (1996). doi:10.1021/BI960541U

116. Esté, J. A. *et al.* Human immunodeficiency virus glycoprotein gp120 as the primary target for the antiviral action of AR177 (Zintevir). *Mol. Pharmacol.* **53,** 340–5 (1998).

117. Perrone, R. *et al.* The G-quadruplex-forming aptamer AS1411 potently inhibits HIV-1 attachment to the host cell. *Int. J. Antimicrob. Agents* **47,** 311–6 (2016).

118. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18,** 279–284 (2017).

119. Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42,** 860–869 (2014).

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

121. Rodriguez, R. *et al.* Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8,** 301–310 (2012).

122. Huang, W.-C. *et al.* Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acids Res.* **43,** gkv1061 (2015).

123. Shivalingam, A. *et al.* The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.* **6,** 8178 (2015).

124. Kikin, O., D'Antonio, L. & Bagga, P. S. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34,** W676-82 (2006).

125. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–9 (2009).

126. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *bioRxiv* 303255 (2018). doi:10.1101/303255

127. Sinha-Datta, U., Khan, S. & Wadgaonkar, D. Label-free interaction analysis as a tool to demonstrate biosimilarity of therapeutic monoclonal antibodies. *Biosimilars* **5,** 83 (2015).

128. Zhang, X. & Oglesbee, M. Use of surface plasmon resonance for the measurement of low affinity binding interactions between HSP72 and measles virus nucleocapsid protein. *Biol. Proced. Online* **5,** 170–181 (2003).

129. Di Primo, C. & Lebars, I. Determination of refractive index increment ratios for protein–nucleic acid complexes by surface plasmon resonance. *Anal. Biochem.* **368,** 148–155 (2007).

130. Frenzel, D. & Willbold, D. Kinetic Titration Series with Biolayer Interferometry. *PLoS One* **9,** e106882 (2014).

131. *Biacore™ Assay Handbook*. (2012).

132. Lynch, H. E., Stewart, S. M., Kepler, T. B., Sempowski, G. D. & Alam, S. M. Surface plasmon resonance measurements of plasma antibody avidity during primary and secondary responses to anthrax protective antigen. *J. Immunol. Methods* **404,** 1–12 (2014).

133. Wei, D., Husby, J. & Neidle, S. Flexibility and structural conservation in a c-KIT G-quadruplex. *Nucleic Acids Res.* **43,** 629–644 (2015).

134. Dexheimer, T. S., Sun, D. & Hurley, L. H. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the *bcl-2* P1 Promoter. *J. Am. Chem. Soc.* **128,** 5404–5415 (2006).

135. Lee, S. *et al.* CDK4 Amplification Predicts Recurrence of Well-Differentiated Liposarcoma of the Abdomen. *PLoS One* **9,** e99452 (2014).

136. Crago, A. M. & Singer, S. Clinical and molecular approaches to well differentiated and dedifferentiated liposarcoma. *Curr. Opin. Oncol.* **23,** 373–378 (2011).

137. Massari, M. E. & Murre, C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* **20,** 429–40 (2000).

138. Knappskog, S. *et al.* The MDM2 Promoter SNP285C/309G Haplotype Diminishes Sp1 Transcription Factor Binding and Reduces Risk for Breast and Ovarian Cancer in Caucasians. *Cancer Cell* **19,** 273–282 (2011).

139. Bond, G. L. *et al.* A Single Nucleotide Polymorphism in the MDM2 Promoter Attenuates the p53 Tumor Suppressor Pathway and Accelerates Tumor Formation in Humans. *Cell* **119,** 591–602 (2004).

140. Bouaziz, S., Kettani, A. & Patel, D. J. A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats 1 1Edited by I. Tinoco. *J. Mol. Biol.* **282,** 637–652 (1998).

141. Campbell, N. H., Parkinson, G. N., Reszka, A. P. & Neidle, S. Structural Basis of DNA Quadruplex Recognition by an Acridine Drug. *J. Am. Chem. Soc.* **130,** 6722–6724 (2008).

142. Venczel, E. A. & Sen, D. Synapsable DNA. *J. Mol. Biol.* **257,** 219–224 (1996).

143. Mendoza, O., Bourdoncle, A., Boulé, J.-B., Brosh, R. M. & Mergny, J.-L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44,** 1989–2006 (2016).

144. Scalabrin, M. *et al.* The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Sci. Rep.* **7,** 45244 (2017).

145. Paramasivam, M. *et al.* Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription. *Nucleic Acids Res.* **37,** 2841–2853 (2009).

146. Huang, P.-R., Hung, S.-C. & Wang, T.-C. V. Telomeric DNA-binding activities of heterogeneous nuclear ribonucleoprotein A3 in vitro and in vivo. *Biochim. Biophys. Acta - Mol. Cell Res.* **1803,** 1164–1174 (2010).

147. Mori, K. *et al.* hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol.* **125,** 413–423 (2013).

148. Fry, M. *Tetraplex DNA and its interacting proteins*. *Frontiers in Bioscience* **12,** (2007).

149. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum. Genet.* **135,** 851–867 (2016).

150. Dominguez, C., Fisette, J.-F., Chabot, B. & Allain, F. H.-T. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.* **17,** 853–861 (2010).

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

152. Huang, H., Zhang, J., Harvey, S. E., Hu, X. & Cheng, C. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev.* **31,** 2296–2309 (2017).

153. Willis, D. M. *et al.* Regulation of *Osteocalcin* Gene Expression by a Novel Ku Antigen Transcription Factor Complex. *J. Biol. Chem.* **277,** 37280–37291 (2002).

154. Roberts, S. A. *et al.* Ku is a 5′-dRP/AP lyase that excises nucleotide damage near broken

ends. *Nature* **464,** 1214–1217 (2010).

155. Tuteja, N. *et al.* Human DNA helicase II: a novel DNA unwinding enzyme identified as the Ku autoantigen. *EMBO J.* **13,** 4991–5001 (1994).

156. Chung, U. *et al.* The interaction between Ku antigen and REF1 protein mediates negative gene regulation by extracellular calcium. *J. Biol. Chem.* **271,** 8593–8 (1996).

157. Zhang, T., Zhang, H., Wang, Y. & McGown, L. B. Capture and identification of proteins that bind to a GGA-rich sequence from the ERBB2 gene promoter region. *Anal. Bioanal. Chem.* **404,** 1867–1876 (2012).

158. Cogoi, S., Paramasivam, M., Spolaore, B. & Xodo, L. E. Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Res.* **36,** 3765–3780 (2008).

159. Pagano, B. *et al.* Identification of novel interactors of human telomeric G-quadruplex DNA. *Chem. Commun.* **51,** 2964–2967 (2015).

160. Dardenne, E. *et al.* RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep.* **7,** 1900–13 (2014).

161. Takahama, K., Kino, K., Arai, S., Kurokawa, R. & Oyoshi, T. Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS J.* **278,** 988–998 (2011).

162. Takahama, K., Sugimoto, C., Arai, S., Kurokawa, R. & Oyoshi, T. Loop Lengths of G-Quadruplex Structures Affect the G-Quadruplex DNA Binding Selectivity of the RGG Motif in Ewing's Sarcoma. *Biochemistry* **50,** 5369–5378 (2011).

163. Seri, M. *et al.* Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Heggllin/Fechtner Syndrome Consortium. *Nat. Genet.* **26,** 103–5 (2000).

164. Riley, M. F. & Lozano, G. The Many Faces of MDM2 Binding Partners. *Genes Cancer* **3,** 226–239 (2012).

165. Conlon, E. G. *et al.* The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife* **5,** (2016).

166. Hacht, A. von *et al.* Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res.* **42,** 6630–6644 (2014).

167. Zamiri, B., Reddy, K., Macgregor, R. B. & Pearson, C. E. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J. Biol. Chem.* **289,** 4653–9 (2014).

168. Decorsière, A., Cayrel, A., Vagner, S. & Millevoi, S. 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 Dev.* **25,** 220–5 (2011).

169. Garneau, D., Revil, T., Fisette, J.-F. & Chabot, B. Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J. Biol. Chem.* **280,** 22641–50 (2005).

170. Williams, K. R. *et al.* hnRNP-Q1 represses nascent axon growth in cortical neurons by inhibiting *Gap-43* mRNA translation. *Mol. Biol. Cell* **27,** 518–534 (2016).

171. Serikawa, T. *et al.* Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5′ UTR of tumor-associated mRNAs. *Biochimie* **144,** 169–184 (2018).

172. Bedrat, A., Lacroix, L. & Mergny, J.-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44,** 1746–1759 (2016).

173. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33,** 877–881 (2015).

174. Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci.* **98,** 8572–8577 (2001).

175. Laguerre, A., Wong, J. M. Y. & Monchaud, D. Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic method. *Sci. Rep.* **6,** 32141 (2016).

176. Doria, F. *et al.* A red-NIR fluorescent dye detecting nuclear DNA G-quadruplexes: in vitro analysis and cell imaging. *Chem. Commun.* **53,** 2268–2271 (2017).

177. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **13,** 551–564 (2018).

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

179. Zhou, W. *et al.* Possible Regulatory Roles of Promoter G-Quadruplexes in Cardiac Function-Related Genes – Human TnIc as a Model. *PLoS One* **8,** e53137 (2013).

180. Zhang, L., Tan, W., Zhou, J., Xu, M. & Yuan, G. Investigation of G-quadruplex formation in the FGFR2 promoter region and its transcriptional regulation by liensinine. *Biochim. Biophys. Acta - Gen. Subj.* **1861,** 884–891 (2017).

181. Lopergolo, A. *et al.* Targeting of <i>RET</i> oncogene by naphthalene diimide-mediated gene promoter G-quadruplex stabilization exerts anti-tumor activity in oncogene-addicted human medullary thyroid cancer. *Oncotarget* **7,** 49649–49663 (2016).

182. Brito, H. *et al.* Targeting KRAS Oncogene in Colon Cancer Cells with 7-Carboxylate Indolo[3,2-b]quinoline Tri-Alkylamine Derivatives. *PLoS One* **10,** e0126891 (2015).

183. Lavrado, J. *et al.* KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines. *Sci. Rep.* **5,** 9696 (2015).

184. Bay, D. H. *et al.* Identification of G-quadruplex structures that possess transcriptional regulating functions in the Dele and Cdc6 CpG islands. *BMC Mol. Biol.* **18,** 17 (2017).

185. Barros, P., Boán, F., Blanco, M. G. & Gómez-Márquez, J. Effect of monovalent cations and G-quadruplex structures on the outcome of intramolecular homologous recombination. *FEBS J.* **276,** 2983–2993 (2009).

186. Ribeiro, M. M. *et al.* G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. *Hum. Genet.* **134,** 37–44 (2015).

187. Araujo, P. R. *et al.* Before It Gets Started: Regulating Translation at the 5′ UTR. *Comp. Funct. Genomics* **2012,** 1–8 (2012).

188. Dvir, A. Promoter escape by RNA polymerase II. *Biochim. Biophys. Acta* **1577,** 208–223 (2002).

189. Perrone, R. *et al.* Synthesis, Binding and Antiviral Properties of Potent Core-Extended Naphthalene Diimides Targeting the HIV-1 Long Terminal Repeat Promoter G-Quadruplexes. *J. Med. Chem.* **58,** 9639–9652 (2015).

190. Campbell, N. H., Parkinson, G. N., Reszka, A. P. & Neidle, S. Structural Basis of DNA Quadruplex Recognition by an Acridine Drug. *J. Am. Chem. Soc.* **130,** 6722–6724 (2008).

191. McCormick-Graham, M. & Romero, D. P. A single telomerase RNA is sufficient for the synthesis of variable telomeric DNA repeats in ciliates of the genus Paramecium. *Mol. Cell. Biol.* **16,** 1871–9 (1996).

192. Largy, E., Marchand, A., Amrane, S., Gabelica, V. & Mergny, J.-L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *J. Am. Chem. Soc.* **138,** 2780–2792 (2016).

193. Smargiasso, N. *et al.* G-Quadruplex DNA Assemblies: Loop Length, Cation Identity, and Multimer Formation †. *J. Am. Chem. Soc.* **130,** 10208–10216 (2008).

194. Ambrus, A. *et al.* Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **34,** 2723–2735 (2006).

195. Wang, Y. & Patel, D. J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1,** 263–82 (1993).

196. Buscaglia, R. *et al.* Polyethylene glycol binding alters human telomere G-quadruplex structure by conformational selection. *Nucleic Acids Res.* **41,** 7934–7946 (2013).

197. Dai, J., Chen, D., Jones, R. A., Hurley, L. H. & Yang, D. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **34,** 5133–5144 (2006).

198. Lim, K. W. *et al.* Coexistence of Two Distinct G-Quadruplex Conformations in the hTERT Promoter. *J. Am. Chem. Soc.* **132,** 12331–12342 (2010).

199. Rankin, S. *et al.* Putative DNA Quadruplex Formation within the Human *c-kit*

Oncogene. *J. Am. Chem. Soc.* **127,** 10584–10589 (2005).

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

201. Kang, H.-J. *et al.* A Pharmacological Chaperone Molecule Induces Cancer Cell Death by Restoring Tertiary DNA Structures in Mutant hTERT Promoters. *J. Am. Chem. Soc.* **138,** 13673–13692 (2016).

202. Bončina, M., Vesnaver, G., Chaires, J. B. & Lah, J. Unraveling the Thermodynamics of the Folding and Interconversion of Human Telomere G-Quadruplexes. *Angew. Chemie Int. Ed.* **55,** 10340–10344 (2016).

203. Thandapani, P., O'Connor, T. R., Bailey, T. L. & Richard, S. Defining the RGG/RG Motif. *Mol. Cell* **50,** 613–623 (2013).

204. Nisole, S. *et al.* The Anti-HIV Pentameric Pseudopeptide HB-19 Binds the C-terminal End of Nucleolin and Prevents Anchorage of Virus Particles in the Plasma Membrane of Target Cells. *J. Biol. Chem.* **277,** 20877–20886 (2002).

205. Galati, D. *et al.* Specific Changes in the Posttranslational Regulation of Nucleolin in Lymphocytes from Patients Infected with Human Immunodeficiency Virus. *J. Infect. Dis.* **188,** 1483–1491 (2003).

206. Fischer, M. J. E. Amine Coupling Through EDC/NHS: A Practical Approach. in *Methods in molecular biology (Clifton, N.J.)* **627,** 55–73 (2010).

207. Tanious, F. A., Nguyen, B. & Wilson, W. D. Biosensor-Surface Plasmon Resonance Methods for Quantitative Analysis of Biomolecular Interactions. in *Methods in cell biology* **84,** 53–77 (2008).

208. Jung, Y., Jeong, J. Y. & Chung, B. H. Recent advances in immobilization methods of antibodies on solid supports. *Analyst* **133,** 697 (2008).

209. Čeru, S., Šket, P., Prislan, I., Lah, J. & Plavec, J. A New Pathway of DNA G-Quadruplex Formation. *Angew. Chemie Int. Ed.* **53,** 4881–4884 (2014).

210. Miura, T., Benevides, J. M. & Thomas, G. J. A phase diagram for sodium and potassium ion control of polymorphism in telomeric DNA. *J. Mol. Biol.* **248,** 233–8 (1995).

211. Oganesian, L., Moon, I. K., Bryan, T. M. & Jarstfer, M. B. Extension of G-quadruplex DNA by ciliate telomerase. *EMBO J.* **25,** 1148–59 (2006).

212. Haider, S., Parkinson, G. N. & Neidle, S. Crystal Structure of the Potassium Form of an Oxytricha nova G-quadruplex. *J. Mol. Biol.* **320,** 189–200 (2002).

213. Schultze, P., Hud, N. V, Smith, F. W. & Feigon, J. The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G(4)T(4)G(4)). *Nucleic Acids Res.* **27,** 3018–28 (1999).

214. Lago, S., Nadai, M., Rossetto, M. & Richter, S. N. Surface Plasmon Resonance kinetic

analysis of the interaction between G-quadruplex nucleic acids and an anti-G-quadruplex monoclonal antibody. *Biochim. Biophys. Acta - Gen. Subj.* **1862,** 1276–1282 (2018).