Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Medicina Molecolare

CORSO DI DOTTORATO DI RICERCA IN MEDICINA MOLECOLARE
CURRICOLO: BIOMEDICINA
CICLO XXX

G-QUADRUPLEXES IN THE HIV-1 LTR PROMOTER: TARGETING AND BINDING PROTEINS

Tesi redatta con il contributo finanziario dello European Research Council (ERC-2013-CoG 615879)

Coordinatore: Ch.mo Prof. Stefano Piccolo
Supervisore: Prof.ssa Sara Richter

Dottoranda: Martina Tassinari

Contents

List of abbreviations ..... **vii**

Abstract ..... **ix**

Riassunto ..... **xi**

Introduction **1**

1.1 DNA G-quadruplexes 1

1.1.1 General features 1

1.1.2 Polymorphism of G-quadruplex structures 3

1.1.3 G-quadruplexes as target for drug design 7

1.1.3.1 Interaction modes of G-quadruplex ligands 7

1.1.3.2 Classification of G-quadruplex ligands 8

1.1.3.2.1 *In situ* protonated G-quadruplex ligands 8

1.1.3.2.2 *N*-methylated aromatic G4 ligands 10

1.1.3.2.3 Metallo-organic G4 ligands 12

1.1.3.2.4 Non-cationic macrocyclic ligands 12

1.1.4 Biological role of G-quadruplexes 13

1.1.4.1 G-quadruplexes in telomeres 14

1.1.4.2 G-quadruplexes during replication 16

1.1.4.3 G-quadruplexes in gene promoters 17

1.1.4.3.1 G-quadruplexes in the c-myc promoter 18

1.2 G-quadruplexes in pathogens 20

1.3 The Human Immunodeficiency Virus (HIV) 22

1.3.1 Viral structure and genome 24

1.3.2 HIV-1 viral cycle 29

1.3.3 Antiretroviral treatment 34
1.3.4 G-quadruplexes in HIV-1 35
1.4 Peptide nucleic acids 37
1.4.1 Peptide nucleic acids: general features 37
1.4.2 Chemical modifications of peptide nucleic acids 38
1.4.3 Targeting G-quadruplexes with peptide nucleic acids 40
1.4.4 Anti-HIV-1 peptide nucleic acids 42
Aim of the study 43
Materials and methods 45
3.1 Oligonucleotides used in this study 45
3.2 Fluorescence Resonance Energy Transfer analysis 45
3.3 Circular Dichroism analysis 48
3.4 *Taq* polymesare stop assay 49
3.5 Surface Plasmon Resonance 52
3.5.1 SPR analysis of G-quadruplex-hits interactions 52
3.5.2 SPR analysis of G-quadruplex-hnRNP A2/B1 interactions 53
3.6 Cells 54
3.7 Virus stock production 54
3.8 Antiviral assay in HIV-1 infected TZM-bl cells 54
3.9 Luciferase reporter assay 55
3.10 Mass spectrometry analysis 56
3.11 Confocal microscopy 57
3.12 Protein nuclear extraction and pull-down assay 57
3.13 Immunoblot analysis 59
3.14 Silencing and reporter assay 59
Results and discussion 61

4.1 Selective compounds targeting LTR G4s 61
4.1.1 High throughput screening of LTR-IV G4 stabilizing ligands 61
4.1.2 Hits stabilize G4s with a preference for the LTR conformations *vs.* the telomeric sequence 65
4.1.3 The best hits display anti-HIV-1 activity through a G4-mediated mechanism of action 71
4.1.4 Mass spectrometry binding analysis 74
4.1.5 Discussion 77
4.2 NDI-PNA conjugates targeting HIV-1 LTR G4s 78
4.2.1 NDI-PNA selectively targets LTR-IV G4 79
4.2.2 Nuclear localization signal mediates cell entry 88
4.2.3 Discussion 90
4.3 Identification and characterization of HIV-1 LTR G4s interacting proteins 91
4.3.1. The human nuclear ribonucleoprotein (hnRNP) A2/B1 selectively binds the HIV-1 LTR G-quadruplexes 91
4.3.2 HnRNP A2 unfolds the LTR G4s 93
4.3.3 Depletion of hnRNP A2/B1 decreases viral transcription 98
4.3.4 Discussion 100
Conclusions 103
Bibliography 105

List of abbreviations

AIDS	acquired immunodeficiency syndrome
CA	capsid protein
CC_{50}	50% cytotoxic concentration
CD	circular dichroism
c-exNDI	core extended naphtalene diimide
Ds	double strand
FRET	fluorescent resonance energy transfer
G	guanine
G4	G-quadruplex
GBA	glycosidic bond angles
HAART	highly active antiretroviral theraphy
HIV	human immunodeficiency virus
hnRNP	human nuclear ribonucleoprotein
hTERT	human telomerase reverse transcriptase
hTR	human telomerase RNA
IC_{50}	50% inhibitory concentration
IN	integrase
LTR	long terminal repeats
MA	matrix protein
MS	mass spectrometry
MOI	multiplicity of infection
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NC	nucleocapsid protein
NDI	naphthalene diimide
$NHIII_1$	nuclear hypersensitivity element III_1
NMR	nuclear magnetic resonance

NF-kB	nuclear factor kappa B
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
PI	protease inhibitor
PIC	pre-integration complex
PNA	peptide nucleic acid
PR	protease
RT	reverse transcriptase
SI	selectivity index
Sp1	specificity protein 1
SPR	surface plasmon resonance
Ss	single strand
TAR	trans-activating response element
Tat	trans-activator of transcription
T_m	melting temperature
Wt	wild type

Abstract

G-quadruplexes (G4s) are non-canonical nucleic acids secondary structures that may form in G-rich sequences and regulate key biological processes. We have previously identified three mutually exclusive and functionally significant G4s in the unique long terminal repeat (LTR) promoter of the human immunodeficiency virus type 1 (HIV-1): their formation decreases viral transcription, with an effect that is enhanced by the presence of the cellular protein nucleolin and G4 ligands. Given that most available G4 binders display yet insufficient selectivity towards different G4s, we explored the possibility to selectively target LTR G4s, challenge which constitutes the basis for the development of anti-HIV-1 compounds with unprecedented mechanism of action.

By screening of a commercially available library of drug-like small molecules, we found a new class of compounds that displayed a clear-cut selectivity for viral over cellular G4s and showed a promising G4-mediated antiviral activity.

In addition, we proposed a conceptually new approach to selectively target G4s, which was successfully tested on HIV-1 LTR-IV G4. We employed naphthalene diimide-peptide nucleic acid (NDI-PNA) conjugates to combine the stabilizing activity of a G4 ligand, with fingerprint recognition of the nucleotide sequence proximal to the G4 of interest. The conjugate was able to stabilize LTR-IV G4 through the NDI, and, at the same time, to prevent the formation of the most stable G4 (LTR-III) by a PNA specifically designed to bind the 5'-flanking region of LTR-IV G4. The NDI-PNA was also successfully modified with a nuclear localization signal (NLS) to achieve cell entry without affecting activity. This innovative method is fundamental to reach the specific discrimination between LTR-III and LTR-IV G4s, whose balance seems to act as regulatory element of the viral promoter. It could be used to modulate the folding and unfolding of unique G4s within the full-length LTR G-rich sequence in order to better investigate their single role and function.

Overall, these findings clearly highlight the possibility to selectively recognize G4s through both small molecules and NDI-PNA conjugates and pave the way to develop novel antiviral compounds that may complement current clinical AIDS therapies.

Finally we looked for new proteins that specifically recognize LTR G4s and modulate the

folding/unfolding of these structures. We identified and characterized the cellular protein hnRNP A2/B1, the first one shown to unfold G4s in the HIV-1 LTR promoter. We thus present new insights in the regulation mechanisms of HIV-1 transcription and we propose a new possible target for the design of specific inhibitors.

Riassunto

I G-quadruplex sono strutture secondarie non canoniche che gli acidi nucleici possono formare in regioni ricche di guanine e regolano importanti processi biologici. Il nostro gruppo di ricerca ha identificato tre G-quadruplex mutualmente esclusivi e rilevanti dal punto di vista funzionale, localizzati nella regione LTR del promotore del virus dell'immunodeficienza umana di tipo 1 (HIV-1): la loro formazione inibisce la trascrizione virale, effetto che è incrementato dalla proteina cellulare nucleolina e da ligandi specifici. Poiché la maggior parte di composti leganti G-quadruplex conosciuti non sono in grado di riconoscere in modo sufficientemente selettivo G-quadruplex diversi, noi abbiamo indagato la possibilità di riconoscere in modo selettivo i G-quadruplex nella regione LTR di HIV-1, concetto alla base dello sviluppo di nuovi composti anti-HIV-1 caratterizzati da un innovativo meccanismo d'azione. Attraverso lo screening di una libreria commerciale di piccole molecole farmaco-simili, abbiamo identificato una nuova classe di composti con una netta selettività per i G-quadruplex virali rispetto a quelli cellulari e una promettente attività antivirale G-quadruplex mediata. Inoltre, abbiamo proposto un approccio completamente innovativo per il riconoscimento selettivo dei G-quadruplex, il quale è stato testato con successo sui G-quadruplex della regione LTR di HIV-1. A tale scopo, abbiamo utilizzato un coniugato formato da una naphthalene diimide (NDI) e da un PNA, con l'obiettivo di combinare la stabilizzazione di un composto legante G-quadruplex con il riconoscimento specifico della sequenza nucleotidica prossimale al G-quadruplex di interesse. L'utilizzo del coniugato ha permesso di stabilizzare in modo specifico il G-quadruplex LTR-IV attraverso la porzione NDI, allo stesso tempo prevenendo la formazione del G-quadruplex più stabile (dLTR-III), grazie alla presenza di un PNA disegnato per legare la regione fiancheggiante LTR-IV in posizione 5'. Il coniugato è stato inoltre implementato con una sequenza di localizzazione nucleare (NLS) per permettere l'entrata nelle cellule, senza interferire con la sua attività specifica. Questa strategia innovativa è fondamentale per riuscire a discriminare in modo specifico LTR-III e LTR-IV, il cui equilibrio sembra essere un elemento regolatore del promotore virale. Potrà inoltre essere utilizzato per modulare la formazione e la destabilizzazione di specifici singoli G-quadruplex all'interno della sequenza di lunghezza completa ricca in guanine della regione LTR, con lo scopo di investigare meglio il loro singolo ruolo e funzione.
Complessivamente, questi risultati evidenziano la possibilità di riconoscere in modo selettivo i G-

quadruplex attraverso l'utilizzo sia di piccole molecole sia di coniugati NDI-PNA e aprono la strada per lo sviluppo di nuovi composti antivirali che possono affiancare l'attuale terapia clinica per l'AIDS. Infine, abbiamo ricercato nuove proteine che riconoscano in modo specifico i G-quadruplex nella regione LTR di HIV-1 e modulino la loro stabilizzazione/destabilizzazione. Abbiamo identificato e caratterizzato, hnRNP A2/B1, la prima proteina cellulare in grado di destabilizzare i G-quadruplex nella regione LTR di HIV-1. Abbiamo quindi mostrato nuove evidenze nel meccanismo regolatorio della trascrizione di HIV-1, proponendo inoltre un nuovo possibile bersaglio per lo sviluppo di inibitori specifici.

Chapter 1

Introduction

1.1 DNA G-quadruplexes

The DNA mainly exists in its typical double helix (B DNA) conformation proposed by Watson and Crick; however, it can also adopt other secondary structures, including G-quadruplexes (G4s). The knowledge that guanine-rich nucleic acids can self-associate has a long history, pre-dating the double helix itself by almost 50 years. The molecular basis for the association was determined by Davies and co-workers in 1962 [1] by fibre diffraction and, later, by biophysical studies [2] using the concept that the G-tetrad (also termed a G-quartet), formed by Hoogsteen hydrogen-bonded guanines, is the basic structural motif [3].

In the recent years, G4s they have gained importance because of the increasing evidence of their potential to act as regulatory sequences in different biological processes. Indeed, many regions in eukaryotic and prokaryotic genomes display the ability to fold into a G4 conformation. Recently, cell- and virus-cycle dependent G4 formation in mammalian living cells and their stabilization by G4 ligands has been demonstrated [4,5]. Thus, G4 can provide a selective site for small molecules in the treatment of various disorders, for example cancer and viral infections.

Besides DNA, G4s have also been reported in RNA molecules, with many similarities with the DNA G4 [6]. Because of the topic of this thesis, only DNA G4s will be further discussed.

1.1.1 General features

G4s are non-canonical nucleic acids secondary structures adopted by sequences that are rich in guanine (G) bases. Deviating from Watson-Crick hydrogen bonding found in B-form duplex DNA, the building block of G4s is the so-called G-quartet (Figure 1.1, panel B) and it consists of four Gs held together in a square planar arrangement by Hoogsteen hydrogen bonds. Eight hydrogen bonds are present: four N1-H hydrogens are hydrogen-bonded with O6 and four N2-H hydrogens are hydrogen-bonded with N7. O6 and N7 are hydrogen bond acceptors and NH1 and C2NH are hydrogen bond donors. Two or

more quartets stack on top of each other to form the G4 [7]. Comparing the association of G bases in the G-quartet with the Watson-Crick base pairs in the double helix (Figure 1.1), it is evident that the G-quartet's surface reaches bigger dimensions. The surface area of the G-quartet is approximately 135 Å, which is approximately twice as large as the surface area of the canonical Watson-Crick base pair. This peculiarity of G4 in possessing a big aromatic surface makes the basis for the design of selective small molecules [8].

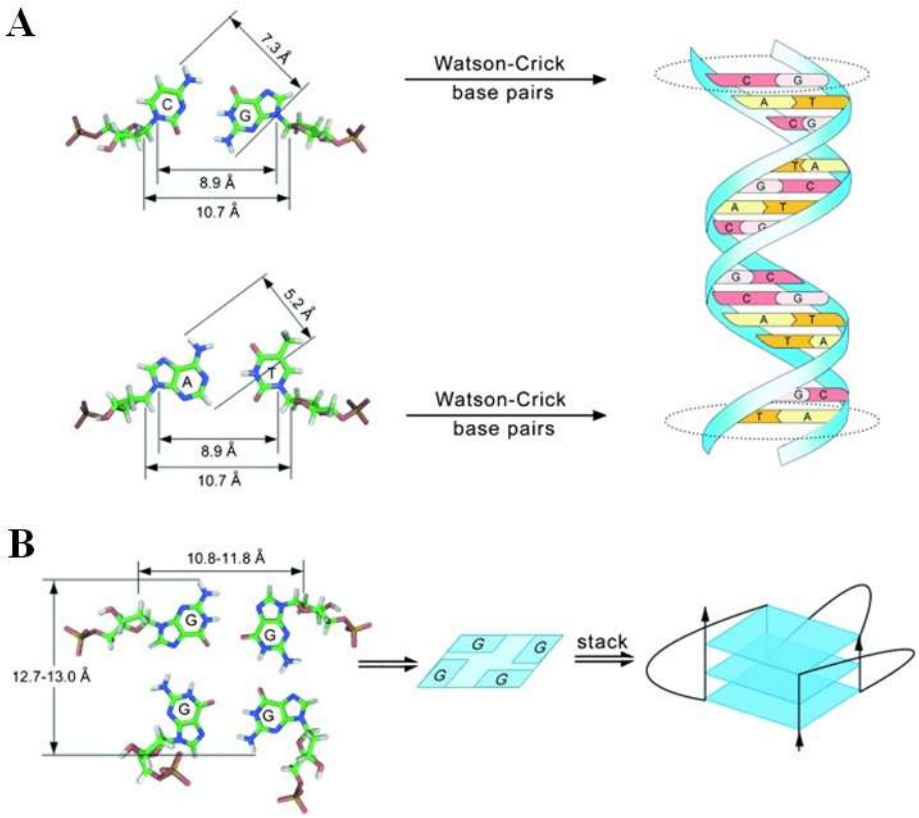


Figure 1.1: Comparison of the dimensions of duplex and G4 DNA structures. A) The double helix and its base pair surface. B) The quadruplex structure and the G-quartet surface. Figure from [8].

G4 formation and stabilization are favored by the presence of monovalent cations, which coordinate with the eight electronegative O6 atoms of the adjacent stacked G-tetrads and thus neutralize electrostatic repulsions [9] (Figure 1.2). Since K^+ and Na^+ are the main cations *in vivo*, G4 formation is favored under physiological conditions. The ions contribute to charge screening, interaction with the loops and the grooves, thus playing an important role in G4 structure formation and in their stability: indeed, molecular dynamics simulation studies suggest that the absence of coordinated cation at the

center of the quartet destabilizes the G4 structure as it is electronically unfavorable [10]. Moreover, a G-rich sequence may adopt different structures in the presence of different cations. The K^+ form is considered to be more biologically relevant due to its higher intracellular concentration (~140 mM) than that of Na^+ (5-15 mM). The precise location of cations between tetrads depends on the nature of the ions: Na^+ ions are positioned mainly in the plane of the G-tetrads, whereas K^+ ions are located between G-tetrad planes [11]. K^+ in general is preferred over Na^+ by G4s as K^+ has a better coordination with eight Gs O6s and a lower dehydration energy [12]. The cation's nature (size/ionic radius and charge) has a considerable effect on the stability of the resultant quadruplex [13].

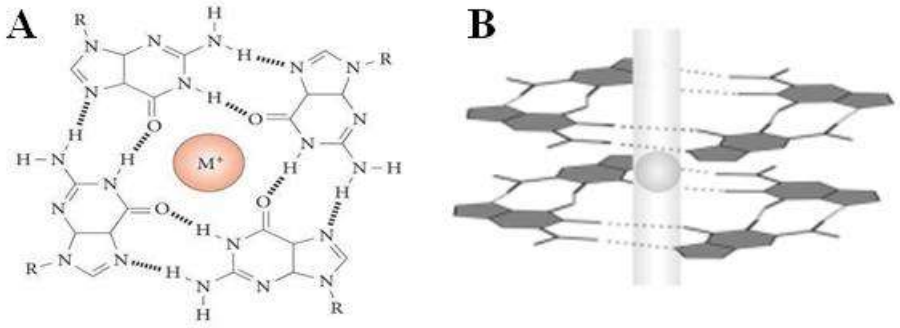


Figure 1.2: G-tetrad structure. A) G-tetrad coordinates monovalent cations thanks to the strong negative electrostatic potential due to carbonyl O6 atoms. B) K^+ in a G4 ion channel. Figure from [8].

1.1.2 Polymorphism of G-quadruplex structures

G4s are highly polymorphic and can adopt different folding topologies depending on several factors, especially G-tetrads number and composition, the orientation of the strands, the syn/anti glycosidic torsion angle of Gs and composition and size of the loops that link the tetrads. Moreover, metal ions in solution, the presence of small molecules and molecular crowding conditions can influence the G4 architecture.

First of all, three main G4 arrangements are possible depending on the number of strands involved:

- tetramolecular: G4 formed by the association of four separated strands, each with at least one G-tract (Figure 1.3, panel A);
- bimolecular: G4 formed by the association of two separated strands, each with two G-tracts (Figure 1.3, panel B);
- monomolecular (or intramolecular): G4 formed within one strand composed of four G-tract connected by loop sequences (Figure 1.3, panel C).

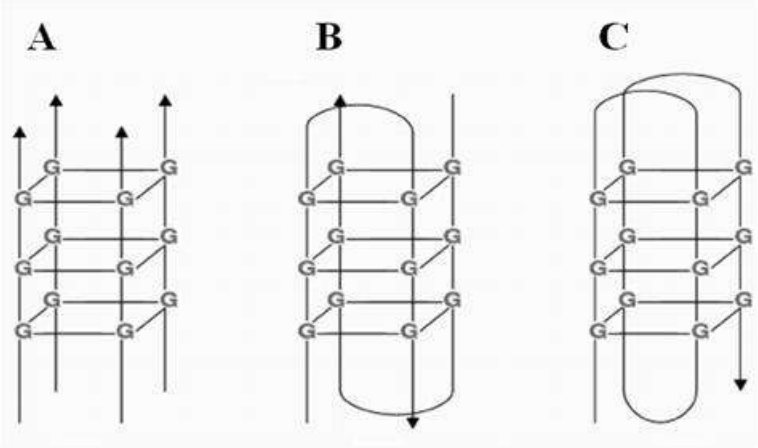


Figure 1.3: Different G4 topologies on the basis of the number of strands involved. A) Tetramolecular. B) Bimolecular. C) Monomolecular. Figure adapted from [14]

The intramolecular G4 forming sequence is composed by at least four runs, containing at least 2 Gs each (G-tracts). This is the so-called "G-4 motif" and can be represented as follow:

G≥2Nx G≥2Ny G≥2Nz G≥2

where N are loops of x, y, z length.

The relative arrangement of strand polarities in a G4 can be classified in parallel (same strand orientation) or antiparallel (opposite strand orientation). Thus, the polarities of the four strands in the G4 can be:

- all parallel;
- three parallel and one antiparallel;
- adjacent parallel;
- alternating parallel.

The relative arrangement of strand polarities gives rise to different G4 conformations:

- antiparallel G4: with adjacent or alternating parallel strands (Figure 1.4, panels A and B);
- 3+1 hybrid (or mixed) G4: three parallel and one antiparallel strands (Figure 1.4, panel C);
- parallel G4: all the four strands are parallel (Figure 1.4, panel D).

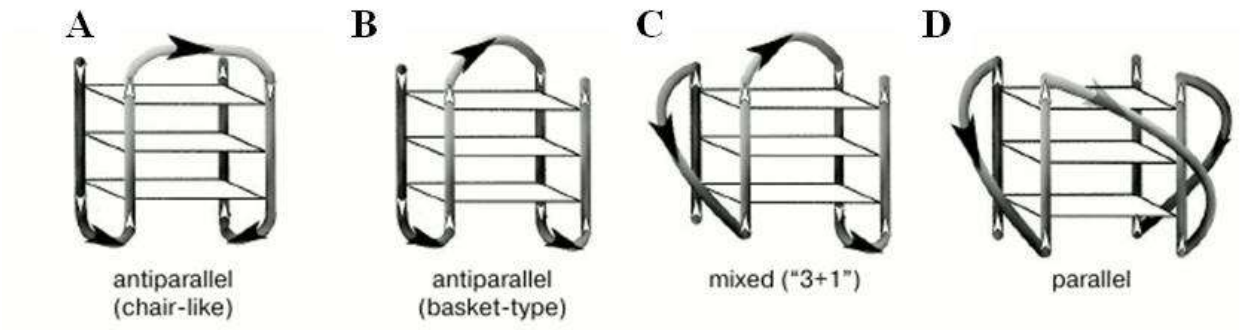


Figure 1.4: **Different G4 topologies on the basis of the relative arrangement of strand polarities.** A) Antiparallel G4 with adjacent parallel strand. B) Antiparallel G4 with alternating parallel strands. C) 3+1 hybrid G4. D) Parallel G4. Figure adapted from [15].

Variations in strand polarities affect also the orientation of connecting loops further increasing the conformational polymorphism. The loops can be classified into four major categories:

- lateral (or edgewise): loop connecting two adjacent antiparallel strands (Figure 1.5, panel A);
- diagonal: loop connecting two opposing antiparallel strands (Figure 1.5, panel B);
- double-chain-reversal (or propeller): loop connecting adjacent parallel strands (Figure 1.5, panel C);
- v-shaped: loop connecting two corners of a G-tetrad core in which one supporting column is lacking (Figure 1.5, panel D).

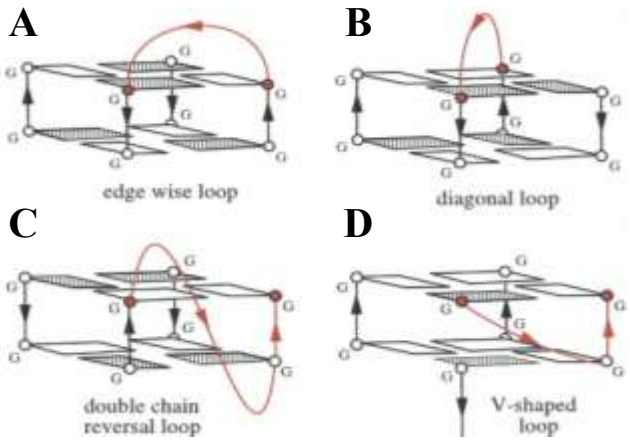


Figure 1.5: Possible orientation of connecting loops in a G4. A) Edge wise loop. B) Diagonal loop. C) Double chain reversal loop. D) V-shaped loop. Loops are in red. Figure adapted from [16].

The parallel intramolecular G4 requires a loop that would connect the bottom G-tetrad with the top one, resulting in the formation of propeller-type. Anti-parallel G4s, in addition to the propeller-type loops, display also diagonal and lateral loops. Two lateral loops can be located both at the same and opposite poles of the molecule, corresponding to the "head-to-head" or "head-to-tail" arrangement in bimolecular complexes [13]. Mixed or (3+1) G4 structures have three strands pointing in one direction, while only one strand is antiparallel. Hybrid-1 and hybrid-2 are subtypes of the mixed topology. Both hybrid structures have two lateral loops that connect the antiparallel strands, while the two parallel strands are connected with the propeller loop.

The G glycosidic bond angles (GBA) are another important parameter that contributes to the G4 high polymorphism. Four Gs within a tetrad can exist in either anti- or syn-conformation (Figure 1.6) with respect to the glycosidic bond, thus providing 16 possible combinations. The mutual orientation of the individual strands within a G4 has an effect on GBAs. For instance, with all four strands oriented in parallel, all GBAs have an anti- conformation. For anti-parallel orientation, the G4 contain both syn- and anti-Gs, regardless of whether the G4 is four- or single-stranded [17].

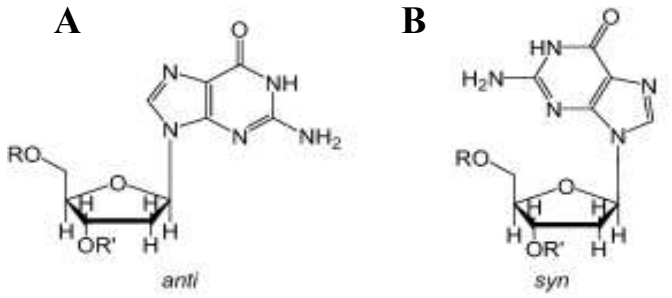


Figure 1.6: Chemical structures of GBAs. A) Anti. B) Syn. Figure from [18].

All G4s have four grooves, unlike the double helix with only two grooves. The grooves are formed by the cavities restricted by sugar-phosphate backbones. The groove's dimensions widely vary depending on the total topology and nature of loops, and on GBAs, as well. In the G4 with loops of only diagonal or lateral type, the grooves are structurally simple, while in the structures with propeller-type loops they possess more complex structural features [13]. Thus, there is a large number of structural variables, which leads to a wide topological and structural variety of G4. Overall, topology and stability of G4 structure is influenced by many factors and, for this reason, could be relatively complex to be predicted and characterized. On the other hand, this great polymorphism constitutes the basis for a drug-targeting

with a high level of selectivity over duplex and non-related G4 structures. Moreover, it could be the basis for the development of ligands specifically designed to selectively recognize and stabilize a specific G4 target structure with respect to other G4s.

1.1.3 G-quadruplexes as target for drug design

1.1.3.1 Interaction modes of G-quadruplex ligands

Since Zahler and co-workers demonstrated in 1991 that K^+-stabilized G4 structures are able to inhibit telomerase activity [19], G4s have emerged as a significant target for telomerase inhibitors. In the recent years, several research groups have searched for small organic molecules that can bind to G4s and hence block telomerase activity or inhibit G4-related oncogene expression. Up to now, many compounds with a wide range of binding affinities have been developed against G4 nucleic acids and several modes of interaction of G4 ligand complexes have been elucidated through computational and chemical-biological approaches. To be efficient, these ligands should display high affinity for the G4s, but more importantly, they should also display higher selectivity for G4s against duplex DNA.

The polymorphism of G4s is thought to allow recognition by G4 ligands trough different binding modes: in fact G4s have several interaction sites, which mainly include the terminal G-tetrads, grooves, loops and the central channel.

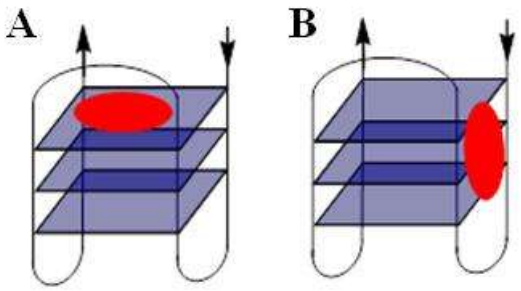


Figure 1.7: Examples of binding mode of small molecule with G4. A) G-quartet stacking. B) Groove binding. Figure from [20].

General features of these G4-recognising ligands include a large flat aromatic surface and cationic charges. The G4 stabilization occurs mainly *via* π–π stacking and electrostatic interactions resulting in the binding of the ligand on the external G-quartet of the structure (Figure 1.7, panel A). This binding mode is called “external stacking” and is typical of flat aromatic molecules. In order to improve the

aromatic-aromatic overlap and provide selectivity, the aromatic surface of a specific G4 ligand should be larger than that of a duplex binder. Electrostatic interactions between the side chains of cationic substituent and the anionic phosphate backbone on one side or more of the G4 also strongly promote stabilization. Moreover, small molecules can interact with the G4 through the groove binding mode (Figure 1.7, panel B), specifically interacting with loops, without significant π stacking with DNA bases [8,21].

Examples of G4 ligands include perylenes, such as PIPER [22], porphyrins, such as TMPyP4 [23], trisubstituted acridines, such as BRACO-19 [24] and natural macrocycles, such as Telomestatin [25]. Some of these compounds have shown encouraging anticancer activity *in vitro* and *in vivo* [26]. Quarfloxin is first-in-class G4 interacting compound in vivo that has reached Phase II clinical trials for the treatment of neuroendocrine/ carcinoid tumors [27].

1.1.3.2 Classification of G-quadruplex ligands

According to the review published by Teulade-Fichou [28], G4 ligands can be classified into four large categories, principally on the basis of their cationic nature:

- *in situ* protonated G4 ligands;
- *N*-methylated aromatic G4 ligands;
- metallo-organic G4 ligands;
- non-cationic macrocyclic ligands.

1.1.3.2.1 *In situ* protonated G-quadruplex ligands

In situ protonated G4 ligands are probably the mostly explored class of G4 binding compounds and are characterized by the presence of protonable sidearms (e.g. amine groups) around the aromatic core.

Large flat aromatic surfaces, essential for stacking interactions with G-tetrads, usually result in high hydrophobic nature. Consequently, protonable side chains are introduced in order to retain sufficient water solubility [28].

BRACO-19 is the best example of this class of compounds. It is an acridine derivative characterized by three protonable substituents (3,6,9-trisubstituted) around the aromatic core (Figure 1.8, panel A), rationally disegned to interact with three G4 grooves [29]. It displays very high selectivity (31-fold) over duplex DNA [30] and greatly stabilizes G4s *in vitro*. It is reported to inhibit telomerase activity and more recently biological investigations revealed its strong anti-proliferative activity on cancer cells.

In situ protonated G4 ligands also include perylene derivatives, such as PIPER. This compound is a perylene diimide (PDI) characterized by a broader hydrophobic core with two external amine chains (Figure 1.8, panel B). It is able to bind G4s with different with stechiometries of 1:2, 1:1 and 2:1 as indicated by NMR studies [31]. PIPER is good telomerase inhibitor, with an IC_{50} (50% inhibitory concentration) in the low nM range and it can induce the duplex-quadruplex transition in the c-myc promoter region [32].

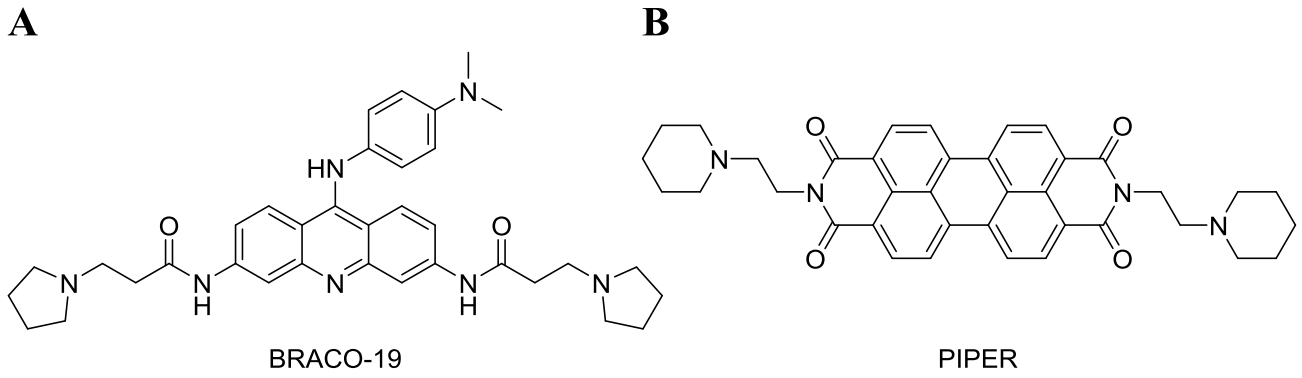


Figure 1.8: Chemical structure of A) BRACO-19 and B) PIPER.

In situ protonated G4 ligands also include naphthalene diimides (NDI) derivates (Figure 1.9). The great advantage of NDIs is that the synthetic route of NDIs allows the introduction to the NDI core of up to four different side chains, characterized by pKa values above 7, which likely make them protonated under physiological conditions. In principle, such substituents can be exploited to produce diverse ligands that may discriminate between different types of G4s.

Among NDI analogues, tri- and tetra-substituted NDIs display very promising G4 binding acitivity. Tri-substituted NDIs chemically engineered to embed an alkylating quinonemethide precursor (QMP) resulted extremely efficient alkylating agents of telomeric G4 DNA, showing a promising anticancer activity *in vitro* [33]. Neidle and co-workers also performed macrocyclization with the aim of improving its G4 binding activity and selectivity over duplex DNA [34].

More recently, a new series of core-extended NDI compounds (Figure 1.9, panel B) have been synthetized. The NDI core was fused to 1,4-dihydropyrazine-2,3-dione, in order to improve the binding to the G4 thanks to the extended aromatic surface of these ligands [35]. This new compounds showed very good G4 binding properties and promising anticancer activity in telomerase-positive cell lines [36]. Besides this advantage, core-extended NDI are also water soluble and display intrinsic fluorescence,

characteristic that make them promising probes with on/off switch of the fluorescence emission in biologically appropriate pH [36]. In fact, interesting results were obtained in testing NDI compounds as selective light-up chemosensors [37,38].

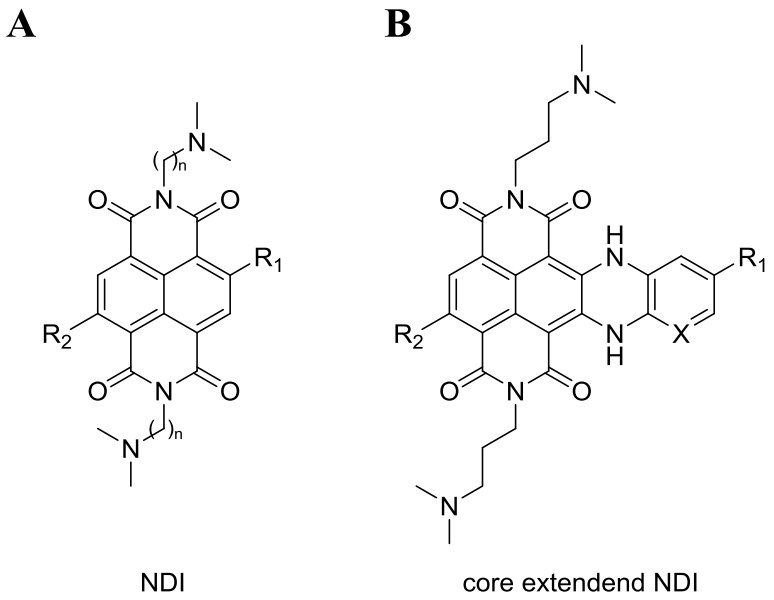


Figure 1.9: General chemical structure of A) NDI and B) core-extended NDI derivatives.

1.1.3.2.2 *N*-methylated aromatic G4 ligands

N-methylated aromatic G4 ligands (quaternized on the aromatic ring nitrogens) exhibit two important advantages: first of all, they conserve water solubility without the need of cationic substituents; secondly, they display a strong π-π stacking ability thanks to the reduction of the electron density of the aromatic part [28].

The lead compound of this category is the tetracationic porphirin TMPyP4 (Figure 1.10, panel C). It has been shown to bind and stabilize both parallel and antiparallel G4s and to efficiently inhibit telomerase [39]. Moreover this compound resulted efficient in downregulating the expression of several oncogenes, such as c-myc [40]. NMR of the complex between TMPyP4 and c-myc pointed out the preferential external tetrad binding and additional electrostatic interactions [41].

Isomers of TMPyP4, named TMPyP2 (Figure 1.10, panel A) and TMPyP3 (Figure 1.10, panel B), have also been tested as G4 ligands. They differ only in the position of *N*-methyl group in the pyridyl ring relative to its connection to the porphyrin core. While TMPyP3 exhibited similar G4 binding properties, TMPyP2 resulted unable to stabilize G4s, probably due to the *orto* position of *N*-methyl

group which hinders the free rotation between the pyridyl ring and the core, resulting in loss of affinity to available accommodation site of the G4. For this reason, TMPyP2 is considered a kind of "negative control" in characterizing G4s and is widely used in parallel with TMPyP4 [42].

Besides porphyrine derivates, a small molecule has to be mentioned in this category: RHPS4 (Figure 1.10, panel D). It is a *N*-methylated pentacyclic acridinium that showed a potent anti-telomerase activity through a selective G4 binding by interaction with the external tetrad [43]. *In vitro* and *in cellulo* studies demonstrated the ability of this ligand to decrease telomeres length and modulate telomere binding proteins [44].

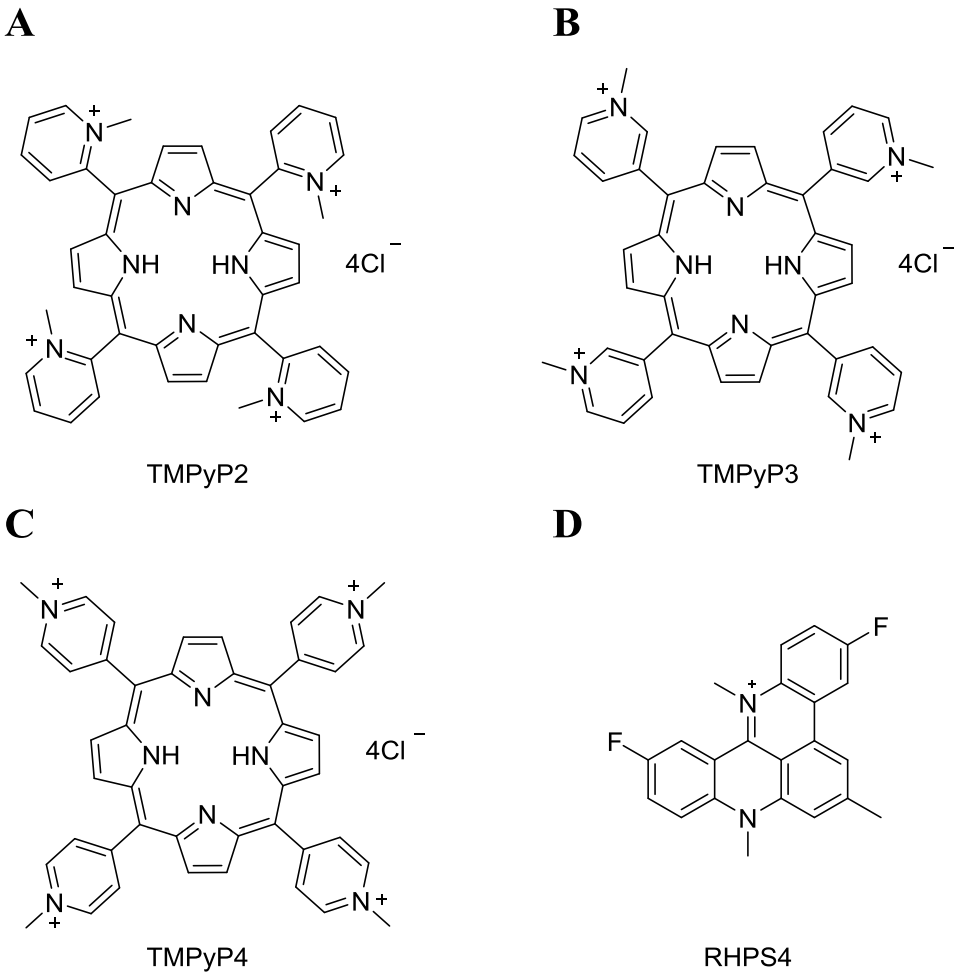


Figure 1.10: Chemical structure of A) CuTMPyP2, B) TMPyP3, C) TMPyP4 and D) RHPS4.

1.1.3.2.3 Metallo-organic G4 ligands

This class of ligands is currently emerging due their interesting G4 binding properties. These are macrocycles, such as previously described porphyrins, coordinating a metal (Cu, Ni, Mn, etc.) in the central cavity [28]. The central metal core of these ligands may be positioned over the cation channel of the G4, thus optimizing the stacking interactions of the surrounding chelating agent with the accessible G-quartet. Moreover, their cationic or highly polarized nature makes the association with the negatively charged DNA more favorable [45].

The most representative examples are Cu(II)-TMPyP4 (Figure 1.11, panel A) and Mn(III)-TMPyP4 (Figure 1.11, panel B). These compounds are characterized by the insertion of a metal in the central cavity of TMPyP4, forming a metallo-complexes. In particular, Mn(III)-TMPyP4 contains four flexible cationic arms in addition to the central aromatic core coordinated by Mn(III). Interestingly, it is one of the most potent G4 ligands, showing a 10000-fold G4 *vs.* duplex selectivity in surface plasmon resonance experiments and good level of telomerase inhibition [46].

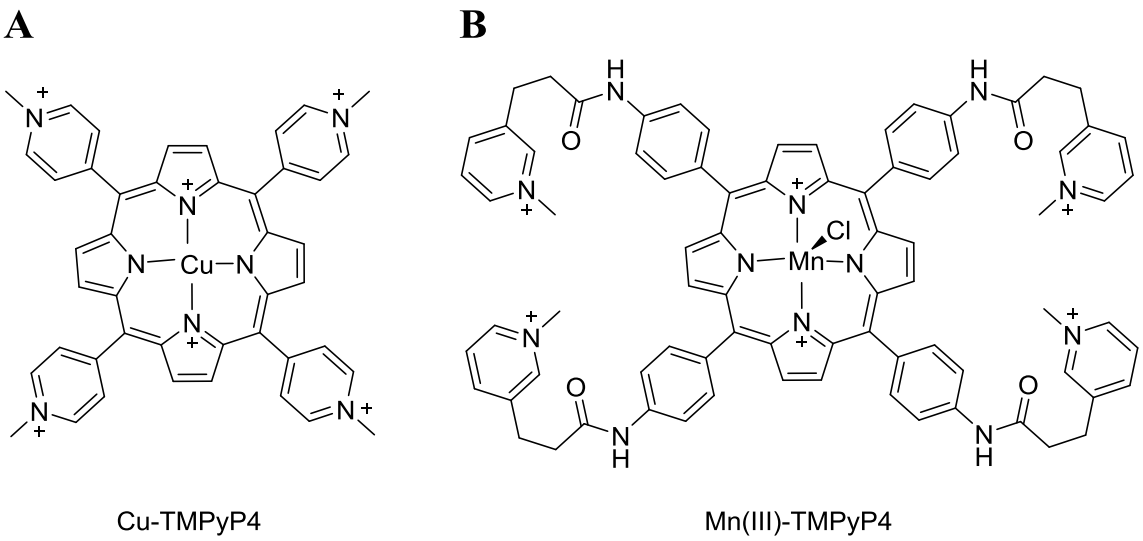
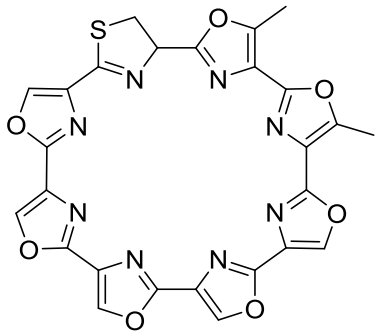


Figure 1.11: Chemical structure of A) CuTMPyP4 and B) Mn(III)TMPyP4 metallo-complex.

1.1.3.2.4 Non-cationic macrocyclic ligands

This group of G4 ligands includes one of the most interesting and better characterized compounds, telomestatin. It is a macrocyclic natural molecule isolated from *Streptomyces annulatus* and consists of seven oxazole rings and one thiazoline ring (Figure 1.12). This G4 ligand completely lacks the affinity towards duplex DNA due to it cyclic shape and neutral character [25]. Perfect shape adaptation to the G-

quartet makes telomestatin highly selective for G4s with great stabilizing and G4 inducing properties even in the absence of cations. Telomestatin is one of the most efficient telomerase inhibitor *in vitro*, with an IC_{50} in the nM range [47,48]. However, its chemical synthesis is really complex and thus hardly compatible with large-scale preparation [49].



Telomestatin

Figure 1.12: Chemical structure of telomestatin.

1.1.4 Biological role of G-quadruplexes

Significantly, it was found that the localization of G4s is non-random: they co-localize with functional regions of the genome and furthermore, are highly conserved between different species indicating a selection pressure to retain such sequences at specific genomic sites [50]. This conservation is highest among mammalian species and decrease in non-mammalian species and lower organisms [51].

The highest abundance of G4s is at telomeres, which in humans consist of tandem repeats of the sequence TTAGGG, typically 5-8 kilobases long, with a single-stranded 3'-overhang of 100-300 bases. They are also highly enriched in gene promoters and immunoglobulin switch regions [52]. Recently, it has emerged that 90% of human DNA replication origins contain G4 motifs, and in higher densities near origins that are used frequently [53–55]. Genome-wide analysis of DNA breakpoints in different cancer types show a significant enrichment in potential G4 forming motifs in the vicinity of somatic copy number alterations [56], as well as telomeres being favored targets of persistent DNA damage response in aging [57]. It also has been found that in about 3000 human genes, G4s are present in the region specifying the 5'-UTR of the encoded mRNAs and may repress translation [58,59].

All this observations of the conserved presence of G4 forming sequences in important genomic regions suggest that they provide a regulatory role through their ability to form G4s [60].

In double stranded DNA, the opportunity for forming G4s arises during DNA replication, transcription and repair when DNA is rendered transiently single stranded through the breaking of Watson-Crick base pairing, which could permit the alternative Hoogsteen base pairing present in G4s to take place [61,62]. In addition, it can be envisaged that G4 formation could be favored by superhelical stress, molecular crowding [63] as well as specific G4 binding proteins [64].

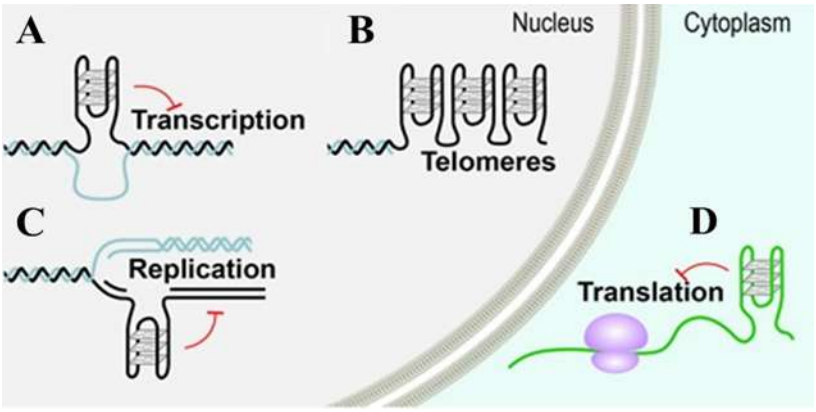


Figure 1.13: Locations of G4s in cells. In the nucleus, G4 formation can occur in double stranded G-rich regions when DNA becomes transiently single stranded, during A) transcription and C) replication and B) at the single stranded telomeric G-rich overhangs. Outside the nucleus, G4s can also form in mRNA and D) are involved in translational control. Red T-bars indicate impediments to transcription, replication and translation. Figure from [60].

1.1.4.1 G-quadruplexes in telomeres

Telomeres are non-coding DNAs located at the termini of linear chromosomes and can form protective structures at these regions, preventing them from being recognized as DNA double-strand breaks and consequently degraded or fused by DNA repair mechanisms [65]. Stringent control of telomere length is important for cell cycle control, cellular immortalization and tumorigenesis [66]. Briefly, telomere shortening is thought to lead to the loss of structural integrity of the telomere nucleoproteins, thus resulting in the activation of p53 and Rb tumor suppressor pathway and cellular senescence, which is an important tumor suppressor mechanism [67]. Telomere-induced senescence is as effective as apoptosis in decreasing cancer incidence [68,69], particularly in preventing oncogene-expressing cells from progressing to malignancy [70]. Normal human somatic cells exhibit weak telomerase activity, which is insufficient to maintain a constant telomere length, whereas more than 90% of human tumor cell populations have high telomerase activity and show various telomere lengths [71,72].

As previously reported, the very distal end of the telomere is a 300 bases single-stranded portion, which forms the T-loop. This loop is analogous to a knot, which stabilizes the telomere, preventing their degradation. It is held together by several proteins collectively referred to as the shelterin complex. In humans, the shelterin complex consists of six elements identified as telomeric-repeat-binding factor 1 (TRF1), telomeric-repeat-binding factor 2 (TRF2), TRF1-interacting protein 2 (TIN2), protection of telomeres 1 (POT1), the transcriptional repressor/activator protein RAP1, and the POT1- and TIN2-organizing protein TPP1. More in detail, the T-loop is generated by invasion of the single-stranded G-overhang into the double-stranded TTAGGG repeats; invasion effectively sequesters the G-overhang and allows natural chromosomes ends to be distinguished from double-strand breaks [73].

The single stranded G-overhang can fold in a G4 structure which is in a dynamic equilibrium with the T-loop structure. Several studies on architectures and folding kinetics of telomeric intramolecular G4s revealed two stable conformations both in K^+ and Na^+ [74]. In Na^+ solution, an intramolecular antiparallel G4 can form with both diagonal and lateral loop [75,76]. In a K^+ containing crystal a very different G-quadruplex structure can fold: a parallel G4 conformation with double-chain reversal loops and anti GBA [75]. In K^+ solution several others G4 structures have been reported and these are in equilibrium with each other. For example, two intramolecular mixed G4 can form [77]. Moreover, two bimolecular G4s that forms in K^+ solution have been reported: one parallel G4 with diagonal loops and one antiparallel G4 with lateral loop [78].

The tandem organization of G-rich telomeric DNA repeats is almost universally conserved in eukaryotes and such sequences are well known for forming G4s *in vitro*. Direct *in vivo* evidence for the presence of G4s at telomers, first came over 10 years ago in studies using specific antibodies direct against G4s [79,80].

The folding and stabilization of G4s in G-rich sequences at the end of telomeres have two major functions: protection of telomeric 3'-overhang from degradation by nucleases and block of telomerase activity. The latter effect is probably due to the folding of the G4 structure which prevents the annealing of human Telomerase RNA (hTR) to the G-overhang. HTR is a component of Telomerase. The latter is a ribonucleoprotein principally composed of the enzymatic part hTERT (human Telomerase Reverse Transcriptase) and by the just mentioned RNA part, hTR. Telomerase drives the synthesis of the G-rich tandem repeats at telomeric G-overhangs by using hTR as template [81]. The effects of stabilization of these structures by small ligands have been extensively studied over the past decade and can act as an alternative protection mechanism to avoid not only genomic instability and but also telomeres elongation in cancer cells (Figure 1.14) [82–85].

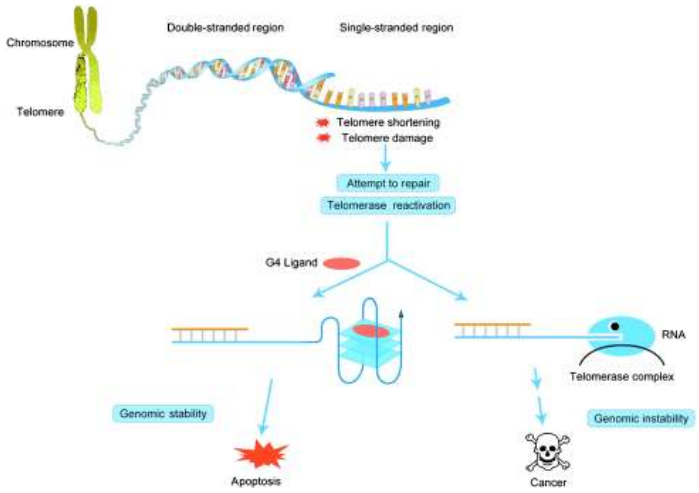


Figure 1.14: Biological role of telomeric G4. Telomere shortening or damage could reactivate repair mechanisms or telomerase enzyme, triggering tumorigenesis. Induction or stabilization of telomeric G4 can thus block telomerase from binding to telomeric ends. Figure from [26].

1.1.4.2 G-quadruplexes during replication

In DNA replication, helicase separates the two strands of dsDNA, allowing the leading and the lagging strand synthesis. During this process, the DNA is thus transiently single stranded and provides a thermodynamically favorable environment for potential G4 forming sequences to fold into G4s. Once formed, G4s may sterically impede the processing of the template by polymerase enzyme, thus inducing the stalling of replication fork progression [86]. In addition, it has been reported that the G4 formation occurs mainly in the lagging strand template that it is replicated discontinuously, triggering fork pausing and instability [87]. To ensure the correct progression of replication process, G4 structures have to be unwound. In this context, many helicases have been shown to solve G4s, indirectly supporting their existence and functions [88].

For example, Human FANCJ helicase have been reported to unwind G4s in ATP-dependent manner in a preferential 5'-3' polarity, thus suggesting the activity on the lagging strand. Its function it is impaired by G4 stabiliziation by telomestatin *in vitro* [89]. Moreover, cell lines expressing mutated FANCJ have been shown to accumulate deletions in sequencing matching with G4 motifs [90].

Another of the best well-characterized example of G4 unwinding helicases is PIF1 which unwinds G4s in the leading strand of replication fork [91,92]. Mutations of Pif slow down the replication fork progression and stabilization of G4s with PhenDC3 inhibits the unwinding activity of Pif1 [90].

As discuss previously, the absence of helicases such as PIF1 or FANCJ leads to G4 instability and G-tract deletions. First, this confirms that G4s are problematic secondary structures that need to be resolved to maintain genome stability. Secondly, it suggests that at least in the absence of unwinding helicases, other enzymes such as nucleases are responsible for processing G4. Several nucleases are known to cleave G4s in vitro. These include, for example, FEN1 and EXO1 [93].

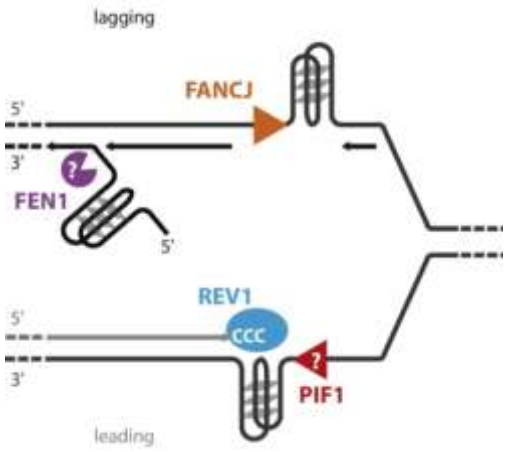


Figure 1.15: G4s and replication. During genomic replication, G4s are unwound by FANCJ on the lagging strand template, and may be cleaved off of a 5′ flap during Okazaki fragment processing by FEN1. G4s that form on the leading strand template may be unwound by PIF1 and destabilized by REV1-dependent synthesis of a run of cytosines opposite the G4. Figure adapted from [93].

1.1.4.3 G-quadruplexes in gene promoters

In a review to characterize the gene ontology of promoters that contain putative G4 forming sequences, Eddy and Maizels discovered that these motifs correlated with the function of human genes, showing, for example, an over-representation in proto-oncogenes and an apparent depletion in tumor suppressor genes [94]. A similar enrichment of G4 motifs in promoters was found in other organisms, including yeast [95], plants [96] and bacteria [97], suggesting a transversal role of G4s among different species. Consistent with this finding, G4 motifs within several oncogene promoters have been shown to transition to stable G4 structures. Importantly, G4 formation was found in promoters of genes that results altered in several diseases and are related to the six hallmarks of cancer (Figure 1.16). In 2000, Hanahan and Weinberg proposed six vital cellular and microenvironmental processes that are aberrantly regulated during oncogenic transformation and malignancy [98]. These include self-sufficiency for growth signals, insensitivity to anti-growth signals, evasion of apoptosis, sustained angiogenesis, limitless replicative

potential, and tissue invasion and metastasis. When each of these categories is examined, a critical protein or proteins can be found with a G4 in the core or proximal promoter. The biological consequence of formation and stabilization of G4 in these promoter elements is mainly gene silencing. All these findings supported G4s associated with oncogenes as a therapeutic potential target for the treatment of human diseases, primarily cancer [99].

The most representative among these genes are c-myc [100], VEGF [101], bcl2 [102], c-kit [103], hTERT [104] and PDGF-A [105]. Promoters in each of these oncogenes are able to form G4s with vast diversity in their folding patterns (e.g. parallel *vs*. mixed parallel/antiparallel), loop sizes and base composition (e.g. one to seven and bases that have specific interactions), number of tetrads (i.e. two, three or four), making them putatively amenable to specific drug targeting [99].

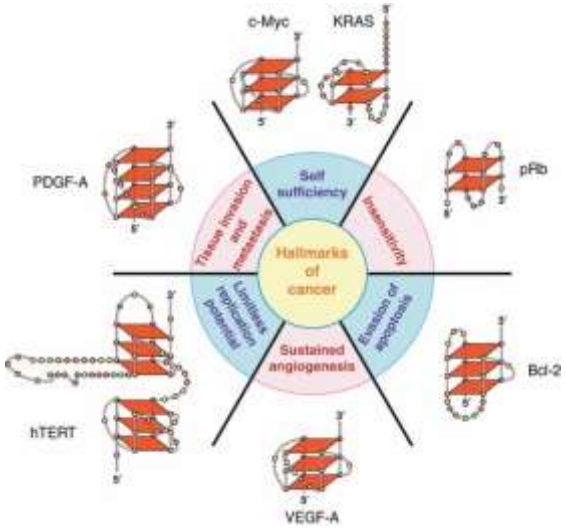


Figure 1.16: The six hallmarks of cancer shown with the associated G4s found in the core or proximal promoters associated genes. Represented structures display a great topological variety. Figure from [99].

1.1.4.3.1 G-quadruplexes in the c-myc promoter

The nuclear hypersensitivity element III_1 ($NHIII_1$) upstream of the P1 promoter of c-myc is the most studied example of G4 forming region. It is a 27-base-pair sequence that contains six G-tracts of unequal length which unsurprisingly has the propensity to adopt complex mixture of G4s in solution. The presence of G4s within this promoter region was initially proposed based on chemical probe studies, gel mobility measurements and fluorescent energy transfer spectroscopy [106,107]. In later studies, the topological structures were determined by circular dichroism, NMR, and mutational experiments [40,108]. Two separate NMR studies on the four central G-tracts have shown that it has an intramolecular

parallel topology, with one two-nucleotide and two single-nucleotide propeller loops, whereas another sequence containing the four central G-tracts similarly forms a parallel G4 with a GGA triad stacked on one end [109–111].

C-myc belongs to the Myc gene family and was one of the first oncogenes identified. It was subsequently linked to a wide range of human cancers. C-myc functions as a gene specific transcription factor through its protein which is thought to regulate 10-15% of all cellular genes and involved in cell cycle regulation, apoptosis, metabolism, cellular differentiation and cell adhesion [112]. As a result, the aberrant overexpression of c-myc is associated with a variety of malignant tumours including those of breast, colon, cervix, small-cell lung, osteosarcomas, gliobastomas and myeloid leukemia [113]. In particular, c-myc has been identified as one of the main activating factors for the human telomerase reverse transcriptase (hTERT) catalytic domain of the telomerase enzyme [114]. The regulation of c-myc promoter activity is extremely complex since it involves several promoters and start sites.

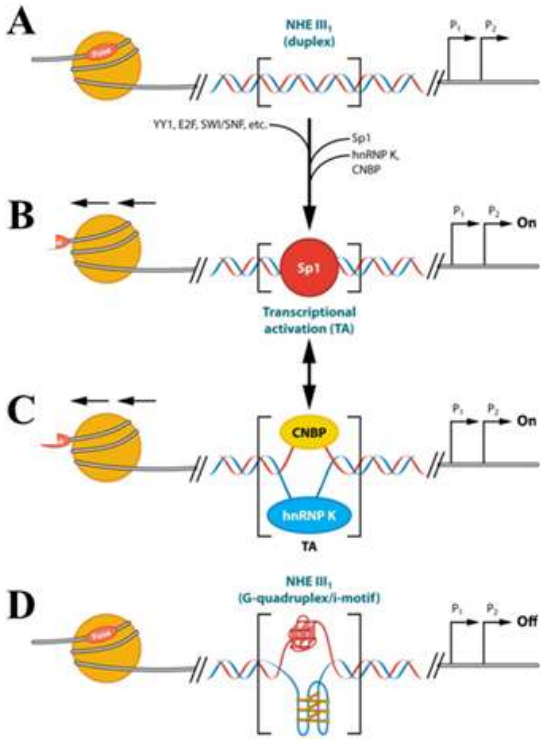


Figure 1.17: Models of different promoter conformations in c-myc NHE III_1. A) Representation of duplex form of NHE III_1 promoter region without bound proteins. The duplex form is a binding site for transcription factor Sp1. B) Binding of Sp1 to the duplex structure, leading to activation of c-myc expression. C) Binding of hnRNP K and CNBP to the single stranded C- and G-rich regions, respectively, leading to activation of c-myc transcription. D) Repression of c-myc transcription when Sp1, hnRNP K and CNBP are not bound, leading to the formation of the G4 and i-motif. Figure from [115].

One important element that controls c-myc expression is the transcriptionally induced negative supercoiling: this effect is due to the RNA polymerase movement during transcription. Since the DNA has to be screwed through the enzyme, this creates an under-twisting behind the transcriptional machinery called negative supercoiling. Some cis-elements in the c-myc promoter, such as the nuclease hypersensitive element 1 ($NHEIII_1$), are dynamically affected by supercoiling. The $NHEIII_1$ region interacts with the transcriptional activator Sp1 in its duplex state. However, the negative superhelical stress facilitates the denaturation of this region and favors the binding of other proteins, heterogeneus ribonucleoprotein K (hnRNP K) and CCHC-zinc finger nucleic acid binding protein (CNBP), to the single strand component, C- and G-rich respectively, again facilitating the transcription process. But the separation of the two complementary strands can allow also the formation of a G4 structure when no proteins are bound. In this case, the G4 prevents c-myc transcription acting as a transcriptional repressor element. This effect is particularly significant since the $NHEIII_1$ controls about 85-90% of c-myc transcription [115]. Stabilization of G4s in this region by small ligands can suppress c-myc transcriptional activation [40,108,116], down-regulate c-myc expression [117,118], inhibit cell proliferation and induce delayed apoptosis in leukemia cells [119].

1.2 G-quadruplexes in pathogens

Besides humans, G4s have been subsequently reported in other different organisms including other mammals [120], yeasts [121] and prokaryotes [97,122–124]. Computational analysis on several organisms showed that, as for the human genome, G4 motifs appear similarly distributed and are over-represented in promoter regions [125].

Evidences of G4 implication in key viral biological processes have been obtained for the *herpesviridae* family, in particular for herpes simplex virus-1 (HSV-1) [126], Epstein-Barr virus (EBV) [127] and Kaposi's sarcoma associated herpesvirus (KSHV) [128]. In 2009, Lieberman and co-workers reported for the first time a G4 structure in the EBV genome. The EBV nuclear antigen 1 (EBNA1) is a virus-encoded protein that is critical for the replication and maintenance of the genome during latency in proliferating cells. It recruits the cellular origin recognition complex (ORC) through a RNA G4 dependent interaction with EBNA1 linking region 1 (LR1) and 2 (LR2). Moreover, G4 RNA-interacting drugs block functions of EBNA1 critical for viral DNA replication and episome maintenance [127]. A few years later, Murat *et al*. demonstrated that destabilization of G4s using antisense oligonucleotides increases

EBNA1 mRNA translation. On the other hand, pretreatment with a G4 stabilizing small molecule, pyridostatin, decreased EBNA1 synthesis, thus highlighting the importance of G4s within virally encoded transcripts as unique regulatory signals for translational control and immune evasion [129]. G4s can also modulate the endogenous presentation of EBNA1-specific $CD8^+$ T cell epitopes which are involved in persistent infections [130]. In addition, Lista and co-workers have very recently reported that the cellular protein nucleolin directly mediates EBV immune evasion through binding to G4s of EBNA1 mRNA [131].

Our group demonstrated that the HSV-1 genome is characterized by 68% GC-content and displays numerous and highly stable G4 sequences that are mainly located in the repeated regions. Their stabilization by G4 ligands, such as BRACO-19 and c-exNDI, resulted in a G4-mediated inhibition of viral DNA replication, with consequent impairment of viral genes transcription [126,132]. In addition, we found that G4 formation and localization within the cells was virus cycle dependent: viral G4s peaked at the time of viral DNA replication in the cell nucleus, moved to the nuclear membrane at the time of virus nuclear egress and were later found in HSV-1 immature virions released from the cell nucleus, thus indicating a key epigenetic role played by these structures during the viral replication cycle [5].

KSHV genome is also rich in G-residues able to form stable G4s, in both strands of a unique region called terminal repeat (TR). Treatment of latently infected cells with G4 stabilizing compounds has been showed to negatively regulate viral replication, leading to a reduction in the KSHV genome copies [128].

Besides the *herpesviridae* family, G-rich regions have also been found in human papillomavirus (HPV) genomes, located in the LCR, L2, E1, and E4 coding sequences, even though G-rich loci have only been found in eight types of HPV [133]. A strong argument for the relevance of G4s in HPV biology is that the viral protein E1 is a helicase that resembles SV40 TAg. Consequently, E1 may also present a G4 unwinding activity. For HPV52 and HPV58, potential G4-forming sequences are located in the long control region (LCR), which is a regulatory sequence, suggesting a potential role in transcription and replication. The presence of G4s in the sequence coding for the L2 protein (HPV57), E1 (HPV32, HPV42) and E4 (HPV3,HPV9, HPV25) suggests that G4 formation may also alter alternative splicing necessary for producing viral proteins from the overlapping Open reading frame (ORFs) [134].

The non-structural protein 3 (Nsp3) of the severe acute respiratory syndrome coronavirus (SARS-CoV) is one component of the viral replicase complex and contains a domain, referred to as SARS unique domain (SUD), which interacts with G4s. Thus, G4s are also relevant for the SARS-CoV and might be involved either in viral replication or host immune evasion [135].

Last year, Wang and co-workers elucidated the existence of two highly conserved RNA G4 sequences in the C gene of the hepatitis C virus (HCV), which can be stabilized by G4 ligands thus reducing RNA replication and inhibiting protein translation [136].

Some G4 forming sequences were also discovered in the Zika virus (ZIKV) RNA genome, of which the most interesting are localized in the unique 3'-UTR region, crucial for initial viral replication of the negative-sense strand [137].

Moreover, a G-rich sequence able to fold into G4 was identified in the negative-sense RNA of Ebola virus (EBOV), in the L gene, and proved to be stabilized by a G4 ligand (TMPyP4), which inhibited gene expression and viral replication [138].

Satkunanathan *et al*. have revealed the presence of a G4 DNA sequence in the adeno-associated viruses (AVV) genome, which interacts with the DNA binding protein nucleophosmin (NPM1) and negatively regulates AVV DNA replication [139].

Finally, Biswas *et al*. have very recently shown the presence of a G-4 motif in the promoter of the preS2/S gene of the hepatitis B virus. This G4 forming sequence is highly conserved in HBV genotype B and forms a hybrid intramolecular G4 which unconventionally positively regulates transcription [140].

Since this thesis focuses on G4 structures in the human immunodeficiency virus (HIV), the main features of this virus and reported evidences for G4 implications in its viral biology will be presented in the following paragraphs.

Understanding the G4 mediated regulation of pathogen virulence has opened the door to innovative therapeutic interventions in the treatment of bacterial or viral infections. Similar to anti-cancer strategies targeting G4s, a pathogen's RNA or DNA could be targeted by G4 specific ligands, and the high topological variety amongst G4 structures suggests that a high level of drug specificity could be achieved [141].

1.3 The Human Immunodeficiency Virus (HIV)

The human immunodeficiency virus (HIV) was first characterized in 1983 when was proposed as the causal agent of the acquired immunodeficiency syndrome (AIDS) [142,143]. The official start of the epidemic occurred in the summer of 1981 when clinical observations of what would later be known as AIDS were reported in the United States and then quickly elsewhere across the world [144]. However, there is substantial evidence that HIV first crossed the simian-human species barrier much earlier,

possibly in Cameroon in West Africa [145]. There is also evidence that HIV found its way to the Caribbean before the 1980s [146].

Since the beginning of the epidemic, more the 70 million people have been infected with HIV and 35 million died. The Joint United Nations Programme on HIV/AIDS (UNAIDS) estimates that there were 36.7 million people living with HIV in 2016, of which 2.1 million were children. In the same year, 1.8 million of new infections and 1 million of AIDS relate deaths have been registered. Sub-Saharan Africa remains the global epicentre of AIDS pandemic with the highest number of people living with HIV infection [147].

HIV selectively infects CD4+ cells. The infection causes the progressive depletion of mainly CD4+ T-helper lymphocyte cells that are a key component of the human immune system. Consequently, the progression of the disease results in a compromised immune system of an individual leading to increase susceptibility to opportunistic infections caused by bacteria, fungi, viruses and parasites.

To date, important progress has been achieved in preventing new HIV infections, which have been steadily declining since the late 1990s and the antiretroviral treatment has reduced the toll of AIDS-related deaths. However, access to therapy is not universal, and the prospects of curative treatments and an effective vaccine are uncertain [148,149]. Thus, AIDS will continue to pose a significant public health threat for decades to come.

HIV belongs to the *Retroviridae* family in the Lentivirus genus and, based on homologies among genomic sequences, two genetically distinct viral types of HIV have been isolated: HIV-type 1 (HIV-1) and HIV-type 2 (HIV-2) [150]. The worldwide main agent of AIDS is HIV-1, while HIV-2 is restricted to some regions of Western and Central Africa [151]. HIV-2 is very similar to HIV-1 in that it has the same tropism for cell of the immune system and causes illness that results from immune deficiency even though it appears less virulent that HIV-1 and infection course takes longer to progress to AIDS [152].

HIV-1 is characterized by a high and complex diversity, mainly due to the error prone reverse transcriptase enzyme, the high virus replication rate and the frequent recombination events. Because of this high diversity in the HIV-1 genome, it has been phylogenetically classified into four groups: group M (Main), group O (Outlier), N (non-M/non-O) and group P. Groups N and P are relatively new groups and together with group O are extremely rare and confined to restricted areas as Gabon and Cameroon. In fact, more than 90% of HIV-1 infections belong to HIV-1 group M which includes 9 subtypes, or "clades", designed as A,B,C,D,F,G,H,J and K, and several circulating recombinant forms (CRFs) [153]. Although subtype B is responsible for only 11% of all infections worldwide, it is the predominant subtype in Europe, besides North America and Australia [154].

1.3.1 Viral structure and genome

HIV-1 virion (Figure 1.18) has a spherical shape (110 nm diameter) and contains two identical positive-sense single-stranded RNA molecules as genetic material. The 9.18 kilobases genome is enclosed by a conical capsid, which is composed of approximately 1500 molecules of nucleocapsid protein (NC), or p24. The capsid also contains the viral enzymes integrase and reverse transcriptase and the four accessory proteins Vif, Vpr, Vpu and Nef. The outer membrane of the virus, called envelope, is composed by a host-cell derived lipid membrane studded by knobs consisting in two glycoproteins: the trimer gp120 surface protein and the gp141 transmembrane protein complex. The envelope is coated in the interior by the matrix proteins (MA) [155].

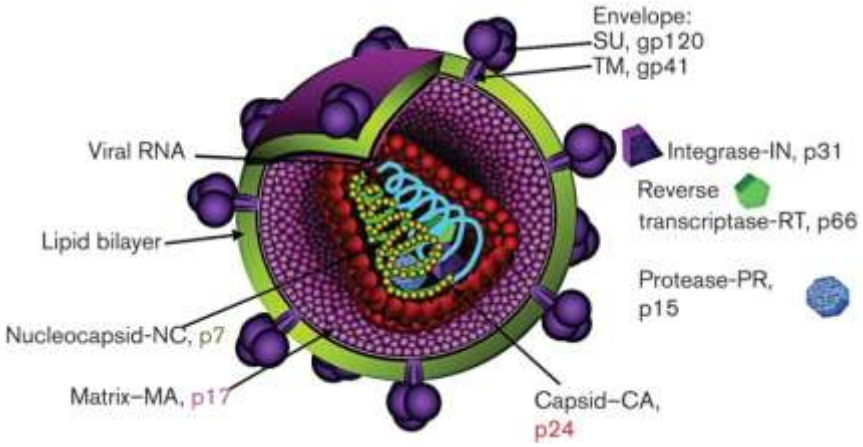


Figure 1.18: HIV virion structure. Schematic representation of a mature HIV-1 virion detailing the localization of viral proteins and the approximate virion structure. The representation is not to scale. Figure from [156].

The HIV-1 RNA genome (Figure 1.19) consists of three primary genes (gag, pol and env), two regulatory genes (tat and rev) and four accessory genes (vif, vpr, vpu, nef). Gag, pol and env encode for polyproteins, which are subsequently proteolyzed into individual proteins common to all retroviruses. The four Gag proteins MA, CA(capsid), NC, and p6, and the two Env proteins, gp120 and gp41, are structural components that make up the core of the virion and outer membrane envelope. Pol encodes for the three viral enzymes protease (PR), reverse transcriptase (RT) and integrase (IN) and is essential for viral replication, as well as gag and env. The two accessory proteins, Tat and Rev, provide essential gene regulatory functions and are also indispensable; on the contrary, the four so-called accessory

proteins Nef, Vif, Vpr and Vpu are not fundamental for viral replication in vitro, but can have consequences on viral life cycle, altering replication or disease progression [157].

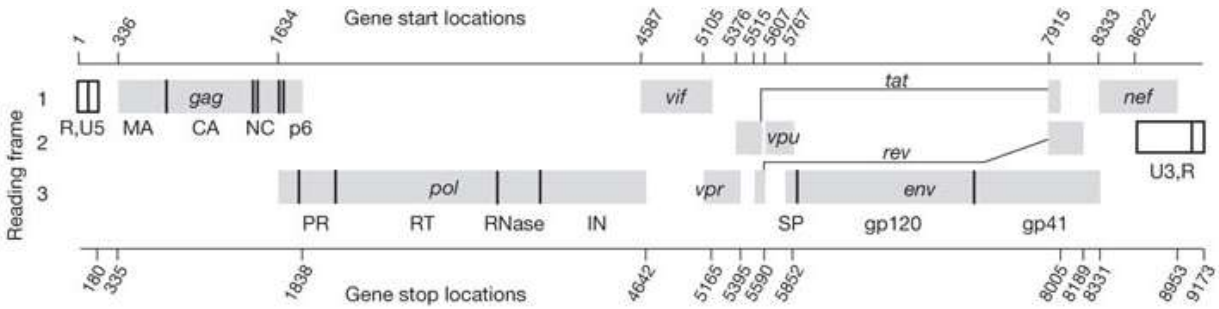


Figure 1.19: HIV-1 RNA genome organization. Protein coding region are shown as grey boxes; polyprotein-domain junctions are depicted as solid vertical lines. Figure from [158].

The main features of the HIV genes and proteins are described below.

- Env

The polyprotein Env (gp160) is expressed from singly spliced mRNA and is synthesized in the endoplasmic reticulum. Then, it undergoes glycosylation, an essential step for viral infectivity [159]. A cellular protease cleaves gp160 into the transmembrane protein gp41 and the surface protein gp120 [160].

- Tat

Tat (trans-activator of transcription) is a transcriptional transactivator essential for HIV-1 replication. The HIV-1 promoter in the 5'-LTR contains sites for cellular transcription factors (NF-kB, Sp1, and TBP) which help control the rate of transcription initiation from the integrated provirus. Despite the importance of these factors, transcription complexes initiated at the HIV-1 promoter are rather inefficient at elongation and require the viral protein Tat to enhance the processivity of transcribing polymerase. Unlike typical transcriptional activators, Tat binds not to a DNA site but rather to an RNA hairpin known as TAR (trans-activating response element), located at the 5'-end of the nascent viral transcripts. An arginine-rich domain of Tat helps mediate binding of a three-nucleotide bulge region of TAR, with one arginine residue being primarily responsible for recognition [161]. The binding of Tat to TAR activates transcription from the HIV promoter at least 1000-fold [162], while in the absence of Tat expression, HIV generates mainly short (>100 nucleotides) transcripts [163].

- Rev

Rev (regulator of viral expression) is a 13-kD sequence-specific RNA binding protein and acts to induce the transition from the early to the late phase of HIV gene expression [164]. When viral mRNAs

are first produced, most are doubly spliced in the nucleus allowing the production of Tat, Rev and Nef proteins. Later, when other viral components are needed to assemble infectious virions, singly spliced and unspliced transcripts are transported in the cytoplasm, where they are translated and where genomic RNAs are packaged. Rev is important in this switch because it overcomes the default pathway in which mRNAs are spliced prior to nuclear export, thus allowing the production of structural proteins and RNA genome. It functions by binding to a RNA secondary structure, called RRE (Rev response element) site located in the env coding region of unspliced and partially spliced mRNA[165].

- Vpu

Vpu (Viral Protein U) is a unique accessory protein of HIV-1 and is expressed from the mRNA that also encodes env. Vpu is translated from this mRNA at levels tenfold lower than that of Env because the Vpu translation initiation codon is not efficient [166]. Vpu is an integral membrane phosphoprotein that is primarily localized in the internal membranes of the host cell. Newly synthesized Env glycoproteins (gp160), which are later cleaved into gp120 and gp41, are sometimes held in the endoplasmic reticulum trough interactions with newly synthesized CD molecules. Vpu promote degradation of CD4 in the complexes, thus allowing Env transport to the cell surface for assembly into viral particles. Besides down-modulation of CD4 antigen, Vpu can also stimulate virion release [167] and it can down-regulate cell surface expression of MHC class I proteins, which may protect infected cells from recognition and killing by T lymphocytes [168].

- Vif

Vif (virulence factor) is a 23 kDa accessory protein that is important for the production of highly infectious mature virions. In fact, virions generated in absence of this protein seem to be about 1000 times less efficient in establishing the infection [169]. Although this protein is not essential for viral replication, the absence of Vif in primary cells results in a defected replication [170].

- Vpr

The 14-kDa accessory protein Vpr (Viral protein R) confers rapid growth advantage to Vpr-expressing viruses. This evidence is more pronounced in macrophages than in primary T-cells [171]. As for the Vif protein, Vpr is incorporated into viral particles, with approximately 100 molecules of Vpr in each virion [172]. Vpr facilitates the nuclear localization of the pre-integration complex (PIC) in non-dividing cells, elucidating its important role at a pre-integration level. In fact, it contains a NLS that direct s transport even in the absence of mitotic nuclear envelope breakdown [173,174]. In addition to its nuclear uptake function, Vpr can influence mutation rates during viral DNA synthesis [175] and can also induce

G2 cell cycle arrest prior to nuclear envelope breakdown and chromosome condensation, and sustained expression can reportedly kill T cells by apoptosis [176].

- Nef

Nef (Negative Factor) is a 27 kDa myristoylated protein expressed early in the HIV-1 life cycle. It was originally characterized as a negative regulator of HIV infection and was thus named as "negative factor". This was then refused by several research groups. Although Nef is not required for HIV-1 replication in vitro, it appears as a fundamental factor for efficient viral replication and pathogenesis in vivo; it also facilitates virus replication and enhances viral infectivity in vitro [177]. Moreover, virus produced from a nef mutated proviral DNA results in a 4 to 40 less infectious virions in single-round infection assay [178]. An essential role for Nef in vivo has been demonstrated in a subset of long-term non progressors, HIV-infected individuals that do not progress to AIDS. Viral isolates from some of these individuals exhibit either a deletion in the nef gene or defective nef alleles [179]. Nef has a positive effect on viral infection and replication by promoting the survival of infected cells by several mechanisms. For example, to promote escape from the immune system and infectivity, it downregulates CD4 and major histocompatibility complex I (MHC I) expression on the cell surface; moreover, to enhance viral replication and infectivity, it activates CD4+ T lymphocytes [180].

- Gag

The gag gene encodes for the 55-kilodalton Gag precursor protein (p55) which is expressed from the unspliced viral mRNA. During translation, the p55 is associates with host cell membranes, recruits two copies of the viral genomic RNA along with other viral and cellular proteins that triggers the budding of the viral particle from the surface of an infected cell. After budding, p55 is cleaved by the virally encoded protease (PR) (a product of the pol gene) during the process of viral maturation into four smaller proteins designated MA (matrix or p17), CA (capsid or p24), NC (nucleocapsid or p9), and p6 [181].

MA is the N-terminal component of the Gag polyprotein and is important for targeting Gag and Gag-pol precursor polyproteins to the plasma membrane prior to viral assembly. In the mature viral particle MA molecules line the inner surface of the virion, stabilizing it [182].

The p24 protein forms the conical core of the virus particle, with ~2000 molecules per virion. Cyclophilin A (a cellular peptidylprolyl isomerase) has been demonstrated to interact with the p24 region of p55 leading to its incorporation into HIV particles [183,184]. The interaction between Gag and cyclophilin A is essential because the disruption of this interaction by cyclosporine A inhibits viral replication [185].

NC is the third component of the Gag polyprotein and coats the genomic RNA inside the virion core. The primary function of NC is to bind specifically to the so-called packaging signal (ψ) and deliver full-length viral RNAs into the assembly virion. The packaging signal consists of four stem loop structures located near the 5'-end of the viral RNA and are bounded by NC through interactions mediated by two zinc-finger motifs [186]. NC also binds single-stranded nucleic acids nonspecifically, leading to coating of the genomic RNA that protects it from nucleases and compacts it within the core. NC appears also essential in facilitating the reverse transcription of RNA genome allowing the formation of proviral DNA [187].

Finally, the p6 polypeptide region mediates interactions between p55 Gag and the accessory protein Vpr, leading to the incorporation of Vpr into assembling virions. It also helps mediate efficient particle release thanks to its late domain which is required for the budding from infected cells of Vpr into assembling virions [188]. The p6 region also contains a so-called late domain which is required for the efficient release of budding virions from an infected cell [189].

- Gag-pol precursor

The Gag-Pol precursor (p160) is generated by a ribosomal frame shift [190]. During viral maturation, the virally encoded protease cleaves the Pol polypeptide away from Gag and further digests it to separate the protease (PR or p10), the reverse transcriptase (RT or p50), RNase H (p15), and integrase (INT or p31) [190].

The HIV-1 protease is an aspartyl protease [191] that acts as a dimer. Protease activity is required for cleavage of the Gag and Gag-Pol polyprotein precursors during virion maturation to produce the final MA, p24, NC and p6 proteins from Gag, and PR, RT and IN from Pol.

RT has two enzymatic activities, a DNA polymerase that can copy either a DNA or an RNA template, and an RNase H that cleaves RNA only if the RNA is part of an RNA/DNA duplex. The two enzymatic functions of RT cooperate to convert the RNA into a double-stranded linear DNA. This conversion takes place in the cytoplasm of the infected cell; after DNA synthesis has been completed, the resulting linear double-stranded viral DNA is translocated to the nucleus where the viral DNA is inserted into the host genome by IN [192].

The integrase protein mediates the insertion of the proviral DNA into the genomic DNA of an infected cell. This process is mediated by three distinct functions of IN. In the first step, its exonuclease activity remove two nucleotides from each 3'-end of the linear viral DNA duplex. Then, a double-stranded endonuclease activity cleaves the host DNA at the integration site. Finally, a ligase activity generates a single covalent linkage at each end of the proviral DNA, generating an integrated provirus [193].

1.3.2 HIV-1 viral cycle

The HIV-1 replication cycle in an infected CD4+ T cells takes approximately 24 h [194] but it is subject to a large degree of variation among individual cells, even in relatively homogeneous cell populations [195]. HIV-1 life cycle is schematically shown in Figure 1.20 and can be summarized in six main steps: 1) binding and entry; 2) uncoating; 3) reverse transcription; 4) provirus integration; 5) virus protein synthesis and assembly and 6) budding [196].

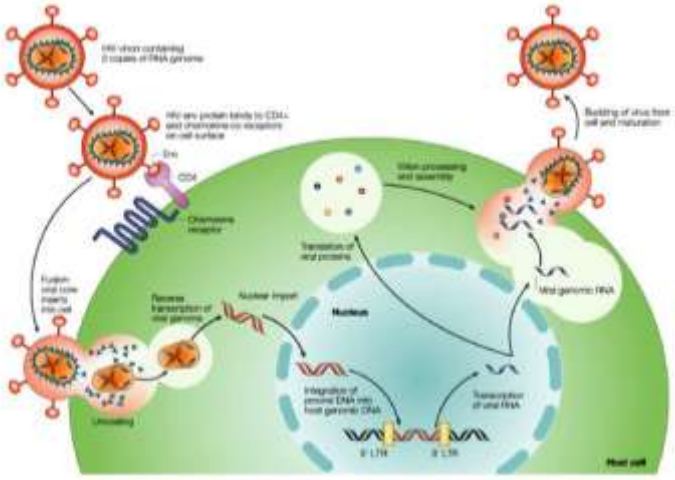


Figure 1.20: HIVreplication cycle. Figure from [197].

The HIV-1 envelope spikes, which comprise trimers of non-covalently linked heterodimers of the surface gp120 and transmembrane gp41 glycoproteins, initiate a cascade of conformational changes that culminates in fusion between the viral and host cell membranes and the release of the viral core into the cytoplasm [198,199].

Indeed, the viral life cycle starts when gp120 binds a 58kDa monomeric glycoprotein, designed as CD4, which is mainly expressed on the cell surface of circulating T-lymphocytes and monocyte/macrophages, but also on T-cells precursors within the bone marrow and thymus, eosinophils, dentritic cells and microglial cells of the central nervous system. However, the binding of the gp120 subunit to CD4 by itself does not trigger membrane fusion. In fact, CD4 binding causes conformational changes in gp120, exposing the binding site for a second cell surface molecule, termed a co-receptor [200,201]. Co-receptor engagement leads to insertion of the fusion peptide located at the N-terminus of gp41 into the cell membrane, which in turn triggers significant rearrangements between trimerized N- and C-terminal heptad repeat sequences within gp41, the formation of a six helical

hairpin structure, and the apposition and fusion of the viral and host cell membranes [202,203]. Most common co-receptors used by HIV are CXCR4 and CCR5, but other potential co-receptors have been described. CXCR4 is expressed on many cells, including T-lymphocytes, whereas CCR5 is present on monocytes/macrophages, dendritic cells and activated T-lymphocytes. The differential recognition of a co-receptor is a major determinant of HIV-1 tropism [204]. In fact, there are strains of HIV-1 preferentially binding the β-chemokine receptor CCR5, conversely, other isolates use preferentially α-chemokine receptor CXCR4. These strains are respectively known as M-tropic (or R5 viruses) and T-tropic (or X4 viruses) [205].

Following membrane fusion, the virus core uncoats into the cytoplasm of the target cell freeing the viral RNA (uncoating) [206]. The conversion of viral RNA into proviral DNA takes place because of the action of the reverse transcriptase and the integrase [207]. The full reverse transcription process is represented in Figure 1.21. Through its ribonucleasi H active site, the reverse transcriptase begins the reverse transcription of viral RNA in the cytoplasm that occurs as a minus-strand polymerization, starting at the primer binding site, until viral RNA is transcribed into a RNA/DNA hybrid double helix. Then, the ribonuclease H site breaks down the RNA strand and the polymerase active site of the reverse transcriptase completes a complementary DNA strand to form a double helix DNA molecule [208]. The full-length double-stranded DNA, viral integrase, MA and Vpr and various cellular proteins form the PIC that is imported into the nucleus [209]. The viral enzyme integrase mediates integration into the human chromosome of the proviral DNA. This protein cleaves nucleotides of each 3'-ends of the double helix DNA creating two sticky ends, transfers the modified provirus DNA into the cell nucleus and facilitates its integration into the host genome [193].

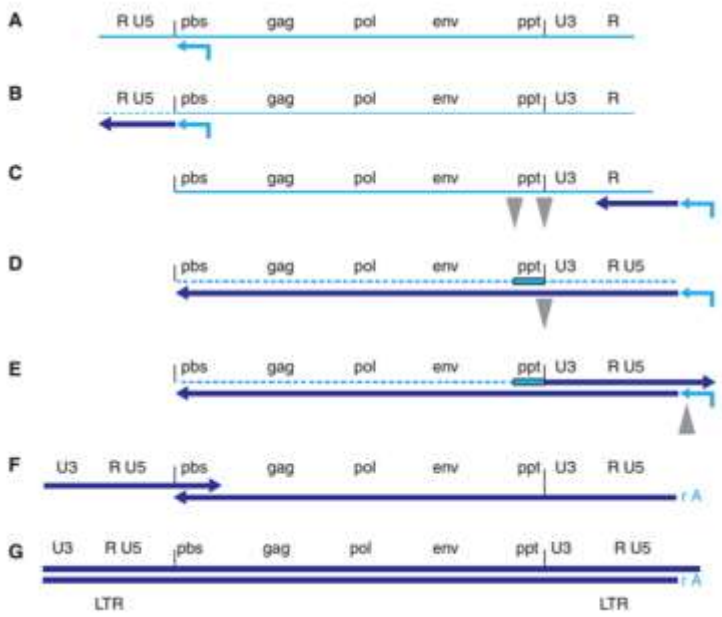


Figure 1.21: Conversion of the single-stranded RNA genome of a retrovirus into double-stranded DNA. A*)* The RNA genome of a retrovirus (light blue) with a tRNA primer base paired near the 5′ end B) RT has initiated reverse transcription, generating minus-strand DNA (dark blue), and the RNase H activity of RT has degraded the RNA template (dashed line). C) Minus-strand transfer has occurred between the R sequences at both ends of the genome, allowing minus-strand DNA synthesis to continue D), accompanied by RNA degradation. A purine-rich sequence (ppt), adjacent to U3, is resistant to RNase H cleavage and serves as the primer for the synthesis of plus-strand DNA. E) Plus-strand synthesis continues until the first 18 nucleotides of the tRNA are copied, allowing RNase H cleavage to remove the tRNA primer. Most retroviruses remove the entire tRNA; the RNase H of HIV-1 RT leaves the rA from the 3′ end of the tRNA attached to minus-strand DNA. Removal of the tRNA primer sets the stage for the second (plus-strand) transfer F) extension of the plus and minus strands leads to the synthesis of the complete double-stranded linear viral DNA G). Figure from [208].

Monocytes/macrophages, microglial cells, and latently infected quiescent CD4+ T-cells contain integrated provirus and are important long-living cellular reservoirs of HIV [210]. In the proviral form the HIV-1 DNA genome (Figure 1.22) contains two identical copies of the long terminal repeat (LTR) at the 5'- and 3'-end of the genome. The LTRs are composed of the segment U3 (derived from a unique sequence located at the 3'-end), R (repeated sequence at both ends) and U5 (derived from a unique sequence located at the 5'-end). The promoter region U3 can be divided in three functional sections: an upstream modulatory element (-454 to -104, with the respect of the first transcribed base) including binding sites for cellular transcription factors, an enhancer (-105 to -79) with two binding sites for the nuclear factor kB (NF-kB), and the core promoter (-78 to -1) composed of three tandem binding sites for specificity protein 1 (Sp1) and a TATA box. The LTR-directed gene expression is regulated in a

cell type and differentiation-dependent manner by the binding of both host and viral protein to the LTR region. The most important host transcription factors are members of the following categories: the specificity protein (Sp) family (e.g. Sp1), nuclear factor kappa B (NF-kB) family, activator protein 1 (AP-1) proteins, nuclear factor of activated T cells (NFAT) and CCAT enhancer binding protein (C/EBP) family. All these proteins bind specific LTR sequences located in the U3 region with a different level of conservation. Genetic variability within LTR binding sites in U3 and TAR regions has been observed in several HIV-1 subtypes [211,212]. However, NF-kB and Sp1 binding sites are remarkably conserved [213]. AP-1 and NFAT bind to the modulatory region, NF-kB binds to 2 sites located in the enhancer region while Sp factors interacts in the core promoter region. Among the Sp proteins, Sp1 plays a crucial role in regulating HIV-1 transcription. In fact, Sp1 is a transcriptional activator that specifically binds 3 sites located in the HIV promoter through its zinc finger binding domain. Interestingly, Sp1 shows affinity and specificity for the GC rich sequence GGGGCGGGGC. The core promoter contains also a TATA box essential for the initiation and regulation of transcription. Also viral protein such as Vpr and Tat bind to the LTR to regulate transcription [214]. The role of these viral proteins has been anticipated in the previous part of the thesis.

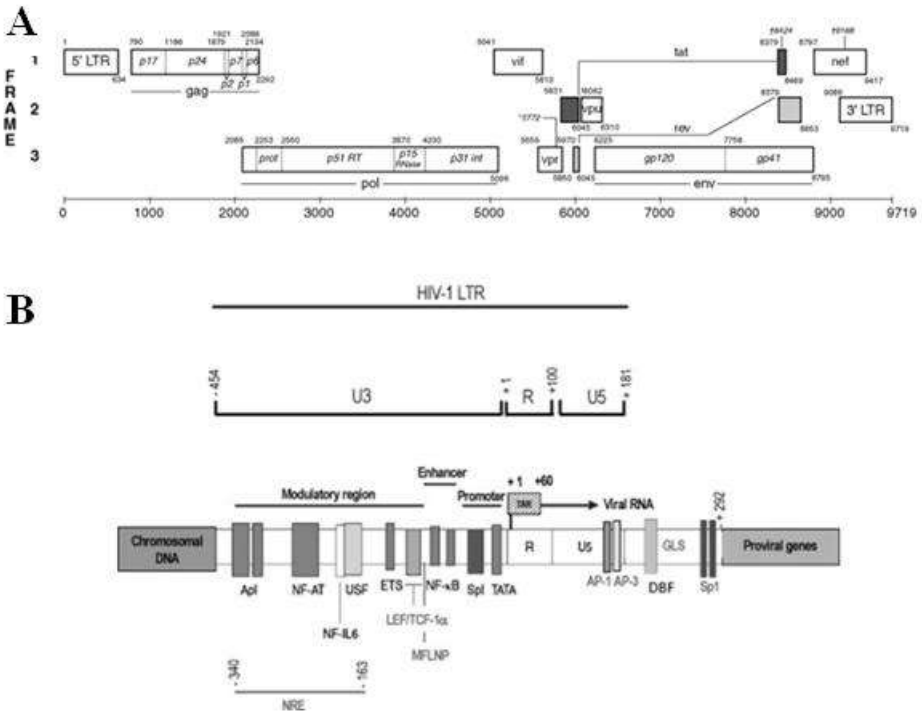


Figure 1.22: HIV-1 proviral genome. A) HIV-1 proviral genome organization. B) HIV Long Terminal Repeat (LTR) organization. Figure from [211].

Upon cell activation, transcription of proviral DNA into a messenger RNA occurs by the host transcription machinery including RNA polymerase II. As anticipated, the viral promoter is located in the U3 region of the 5'-LTR and requires activation by host transcription factors (Figure 1.23). Upstream from the transcription start site, in the core promoter the initiator (Inr), the TATA BOX and three Sp1-binding sites are necessary for the correct position of RNA polymerase II (RNAPII) in the so-called pre-initiation complex. Moreover, also the TATA-binding protein (TBP) and TBP-associated factors (TAFs) contributes to the formation of this complex. After the formation of the pre-initiation complex, the RNAPII clears the promoter, starting the transcription. The initial transcriptional output is however very low, due to the blocked elongation of viral transcripts early in the 5'-portion of the RNAs. The viral transactivator protein Tat, which is made very early from the tiny amounts of successfully terminated mRNAs, is required to achieve normal levels of expression [215]. A peculiar feature of the LTR is the presence of a RNA regulatory element known as the TAR element. As anticipated, Tat protein binds TAR element. Tat and its cellular co-factor P-TEFb (Positive Transcription Elongation factor B) cooperate to bind TAR with high affinity, allowing RNAPII to produce full-length viral transcripts. P-TEFb contains the components cyclin T1 (CYCT1) and cyclin-dependent kinase 9 (CDK9) that mediates a phosphorylation of the carboxy-terminal domain of the RNAPII. This converts the initiating transcription complex to an elongating transcription complex [216].

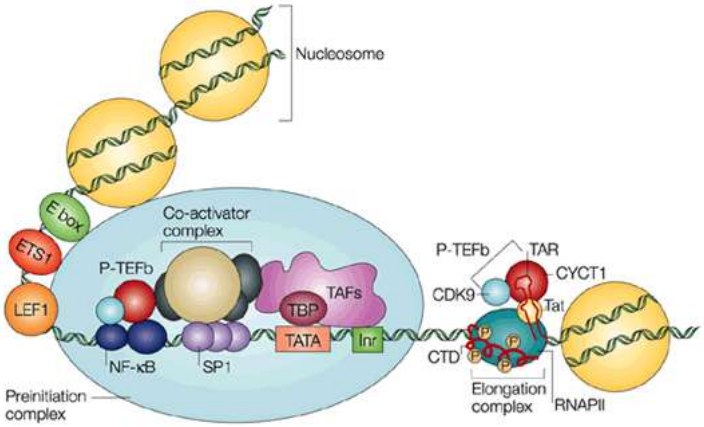


Figure 1.23: The HIV Long Terminal Repeat (LTR) promoter. TBP and TAFs bind the core promoter. The co-activator complex binds SP1 and cooperates to recruit and position RNAPII in the pre-initiation complex. RNAPII then clears the promoter. P-TEFb and Tat bind the TAR element. Thus, the initiating transcription complex is converted into an elongating transcription complex and the transcription efficiently starts. The enhancer binds members of NF-kB, NFAT and ETS families. Figure from [217].

Thereafter, full-length mRNA transcripts are efficiently synthesized. These unspliced transcripts contain multiple splice sites for the generation of over 40 unique viral transcripts for translation of the nine viral proteins. To date, the full-length mRNA serves as the template for the polyproteins Gag (p55), Gag-pol (p160) and Env (gp160), while accessory proteins are translated from spliced mRNA. The mRNA molecules are next transported to the cytoplasm where the translation of viral proteins occurs. Gp160 is then cleaved resulting in the two glycoproteins gp41 and gp120 [218].

The formation of new viral particles is a stepwise process: two viral RNA strands associate together with precursor proteins, while core proteins assemble over them forming the virus capsid. This immature particle migrates towards the cell surface. The large precursor molecules are then cleaved by the HIV-1 protease and the virion starts to bud off from the surface of the cell, thus acquiring a new envelope. During the budding process, the virus lipid membranes may incorporate various host cell proteins and become enriched with phospholipids and cholesterol. Differently from T-lymphocytes, where budding occurs at the cell surface and virions are released into the extracellular space, the budding process in monocytes and macrophages results in the accumulation of virions within intracellular vacuoles which are then released. During or shortly after the release of the virion, the enzyme protease cleaves the Gag and Gag-pol polyproteins to complete the maturation [219].

1.3.3 Antiretroviral treatment

In 1987, the first anti-HIV drug was approved by the U.S. Food and Drug Admisnistration (FDA) for the AIDS treatment. This was the reverse transcriptase inhibitor 3'-azido-3'-deoxythymidine (AZT), also known as Zidovudine. To date, lot of antiretroviral compounds have been approved by FDA. These drugs belong to several classes, based on their mechanism of action and viral target:

- nucleoside reverse transcriptase inhibitors (NRTIs);
- non-nucleoside reverse transcriptase inhibitors (NNRTIs);
- protease inhibitors (PI);
- entry inhibitors;
- integrase inhibitors;
- CCR5 antagonists.

These drug classes target four key steps in the viral cycle: viral entry, reverse transcription, integration and protein processing/maturation [220]. Since the first HIV-1 specific antiviral drugs were given as monotherapy in the early 1990s, the standard of HIV-1 care evolved to include the administration of a combination of antiretroviral agents. The current indicated treatment, the so-called highly active

antiretroviral therapy (HAART), consists of a cocktail containing at least three different drugs against to, at minimum, two distinct molecular targets, usually two NRTIs and one NNRTI or PI. The advent of HAART has dramatically suppressed viral replication and has reduced the plasma HIV-1 viral load to below the limits of detection of the most sensitive clinical assays resulting in a significant reconstitution of the immune system as measured by an increase in circulating CD4+ T-lymphocytes. Consequently, it has been crucial in reducing the morbidity and mortality associated with HIV-1 infection and AIDS, greatly increasing the life expectancy of HIV-1 infected patients [221,222]. However, despite advances in antiretroviral therapy, some treatments still fail. The major cause of treatment failure is the development of drug resistance that arises from the extreme variability of the virus [223,224] which is a consequence of at least three peculiar features: 1) the "error-prone" mechanism of action of RT enzyme, that introduces, on average, one substitution per genome per replication round; 2) the very rapid viral replication, that generates a high number of virions per day (estimated around 10^{10}) in the infected individual and 3) the occurrence of recombination processes between two or more different HIV viruses within the same infected individual. In addition, although HAART is very effective in blocking HIV-1 spread within the body, it does not eradicate the virus which still rests in latently infected CD4+ T-cells. Indeed, viral loads readily rebound when treatment is interrupted making life-long treatment necessary which is often associated with drug toxicities and a risk of viral resistance [225,226]. Therefore, there is an urgent need of therapeutic strategies with an innovative mechanism of action, possibly against highly conserved viral sites, and able to clear the virus from the infected human host in order to reach a complete recovery.

1.3.4 G-quadruplexes in HIV-1

The presence of G4 structures has been highlighted at both RNA and DNA levels in the HIV-1 genome, with implications throughout the viral life cycle [134]. The first evidence of G4 formation in the HIV-1 genome was given in 1993 when Sundquist and Heaphy described a cation dependent dimerization of the two homologous copies of ssRNA genome. They demonstrated that the highly conserved purine-rich region located near the 5'-end of gag gene spontaneously dimerizes in high ionic strength conditions without protein cofactors and that GGGGGAGG sequence at the positions 817-825 is essential for dimerization. They therefore proposed that the dimerization occurs via intramolecular RNA G4 formation [227].

Ten years later, Lyonnais *et al.* discovered a G4 structure in the single-stranded portion (central DNA flap) of the reverse-transcribed pre-integration HIV-1 genome [228] and proposed that G4s have a

nuclease protection function of pre-integration complex, as DNA-flaps are usually recognized by cellular nucleases. In particular, G4s specifically interact with the viral NC protein, promoting it assembly along ssDNA and thus protecting the pre-integrated genome from degradation [229].

More recently, Shen and co-workers presented evidence of a G-rich sequence at the beginning of gag coding region, characterized as preferential site for recombination. G4s in gag recombination hot spot induce RT pausing at G4 forming sites, and those pauses could initiate acceptor invasion-driven transfer events, which highly promotes recombination of the virus at the genomic RNA level [230].

In 2013, our research group identified a G4 forming sequences in the HIV-1 *nef* coding region: three contiguous putative G4 regions are present and that at least two are extremely conserved among most circulating HIV-1 strains. We provided evidence of their G-4 folding and stabilization in the presence of cations and G4 binding compounds, particularly TMPyP, which impaired Nef expression and significantly suppressed Nef-dependent enhancement of HIV-1 infectivity [231].

In the same year, our group identified a putative regulatory G4 forming sequence in the HIV-1 LTR U3 region [232], in a segment corresponding to part of the core promoter and enhancer (positions -105/-48 with respect to the transcription initiation site). Interestingly, this tract was also the binding region of two important cellular transcription factors that stimulate viral transcription, i.e. two binding sites for NF-kB and three tandem binding sites for Sp1. Three mutually exclusive and functionally significant G4s have been found, namely LTR-II, LTR-III and LTR-IV (Figure 1.24).

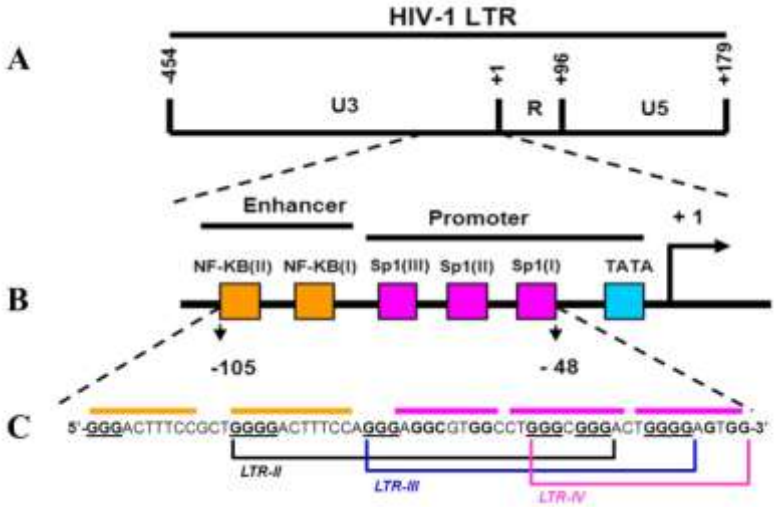


Figure 1.24: The G-rich HIV-1 LTR region organization. A) Schematic representation of HIV-1 LTR, composed of U3, R and U5 regions. B). C) G-rich sequence, -105/-48 with respect to the transcription initiation site, able to fold into three mutually exclusive G4s, termed LTR-II, LTR-III and LTR-IV.

LTR-II and LTR-III G4s spontaneously fold in physiological ionic conditions, while LTR-IV is induced by the presence of G4 stabilizing ligands. The introduction of point mutations disrupting G4 formation stimulated the LTR promoter activity, while G4 stabilization by G4 ligands inhibited it, thus showing a great parallelism between G4s in the LTR and G4s in oncogene promoters of eukaryotic cells. Based on these evidences LTR G4s were proposed as novel antiviral targets [232]. Supporting that, we next demonstrated that BRACO-19, a potent and well-characterized G4 ligand, displayed promising anti-HIV-1 activity through a G4 mediated mechanism of action [233], thus validating LTR G4s as promising anti-viral targets.

Further investigation of LTR G4s involvement in HIV-1 transcriptional process allowed to the identification of the cellular protein nucleolin, which specifically recognize LTR G4s. Nucleolin depletion produced enhancing effects on LTR promoter activity; on the contrary, LTR G4-nucleolin specific interaction resulted in the increase of silencing activity operated by G4s on viral transcription [234]. Nucleolin binding to LTR G4s and its influence on LTR promoter activity indirectly supported the biological relevance of these structures in regulating HIV-1 transcription.

Interestingly, G4s in the 5'-LTR U3 region are highly conserved not only among different HIV-1 strains, but also in primate lentiviruses, thus supporting the role of 5'-LTR G4 region as control centre of viral transcription.

Very recently, G4 stabilizing agents have been also employed in cells infected with latent HIV-1, where their activity resulted in a strong antiviral effect, especially in combination with a DNA repair inhibitor, revealing new aspect of HIV-1 latent infection [235].

1.4 Peptide nucleic acids

1.4.1 Peptide nucleic acids: general features

Peptide nucleic acids (PNAs) are non-natural nucleic acids in which the natural nucleobases are preserved but connected to an uncharged, achiral pseudo-peptide backbone composed of *N*-(2-aminoethyl)glycine units in lieu of the natural sugar phosphate backbone (Figure 1.25) [236].

PNAs bind complementary RNA and DNA with high affinity and specificity. In fact, because the PNAs are uncharged, they tend to exhibit tighter binding to complementary natural nucleic acids than does natural DNA or RNA, due to the lack of poly anionic charge-charge interactions [237]. The stability of PNA-DNA (or RNA) complexes does not required extensive cation shielding of the backbone

phosphate. Thus, in sharp contrast to the behavior of negatively charged nucleic acids, PNA complexes show almost unaltered stability at very low ionic strength. Furthermore, the pseudo-peptide backbone makes them extremely stable in biological fluids: in fact PNAs are resistant to degradation in vivo because they are not recognized by nucleases or proteases. Consequently, PNAs are expected to have sufficient biostability to be used as a drug. All these qualities make them attractive candidates for antigene, antisense, or nucleic acids probes [238].

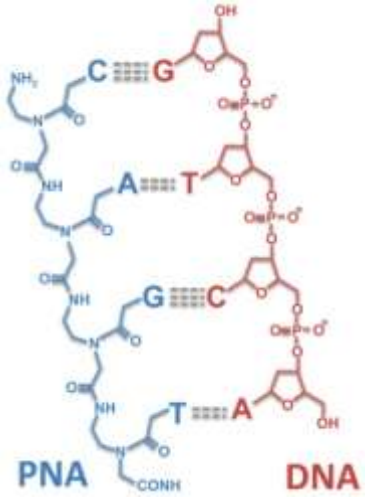


Figure 1.25: PNA-DNA hybrid. PNA (blue) and DNA (red) backbones. Figure from [239].

1.4.2 Chemical modifications of peptide nucleic acids

Since the first reports on PNAs based on *N*-(2-aminoethyl)glycine backbone, a large variety of PNA backbone modifications and substitutions have been synthesized, with the main purpose of improve biological activities and abrogate some of PNAs inherent shortcomings such as solubility, cell permeability, or bioavailability. Despite most of them resulted in PNAs with lower nucleic acid recognition properties, some quite simple substitutions provided PNAs with very useful specifically improved features [239].

Most importantly, in vivo applications of unmodified PNAs are hindered by poor cellular uptake and endosomal entrapment. Because of the neutral backbone, PNAs do not associate with delivery vehicles based on cationic lipids. Therefore, the inefficient crossing of cellular membrane in mammalian cells by unmodified PNAs has been a major problem for their practical *in vivo* applications.

A large number of different conjugation of PNA with cation peptides and cationic backbones modifications have been reported to overcame this critical problem [240]. The most popular approach to

enhance cellular delivery is the conjugation of PNA with cell penetrating peptides (CPPs) that deliver the conjugate through the endocytosis pathway [241]. Conjugates with arginine-rich peptides showed promising activity in HeLa cells in the absence of endosomolytic agents [242]. However, the relatively high concentrations of PNA-CCP, which are required for efficient delivery, may cause off-target binding and toxicity in vivo. Moreover, CPPs are relatively large peptides, which complicate the preparation and use of PNA-CPP conjugates [240].

Another approach to improve PNAs cell entry was developed by Taylor and co-workers who used nanoparticles to deliver PNAs covalently attached to nanoparticles through a biodegradable disulfide linkage. Nanoparticles have a hydrophobic core and a positively charged cross-linked shell and, the latter is highly functionalizable and mediates the cellular delivery through an endocytotic mechanism [243].

Recently, several groups have demonstrated that relatively simple cationic modifications in PNA can substantially improve their cellular uptake and produce effects similar to that of longer and more complex CPPs. The groups of Corey [244] and Gait [242] showed that conjugation of PNA with short oligolysine enabled efficient delivery in fibroblast and various cancer cell lines (T47D, MCF-7, Huh7, and HeLa). As few as four lysine residues achieved similar efficiency as R6-Penetratin, a CPP previously optimized for cellular delivery of PNA [242]. Fabani *et al*. employed PNA-oligolysine conjugates to efficiently block a key inducible miRNA expressed in the hematopoietic system, miR-155, in cultured B cells as well as in mice. Remarkably, miR-155 inhibition by PNA in primary B cells was achieved in the absence of any transfection agent. Most importantly, the high efficiency of the treatment was confirmed in mice by a strong overlap in global gene expression between B cells isolated from anti-miR-155 PNA-treated and miR-155-deficient mice [245].

A very efficient delivery system developed by Wender and co-workers is based on arginine oligomero [246]. Oligoarginines are reported to be internalized into mammalian cells through an endocytotic pathway, mediated by binding to cell surface heparan sulfates linked to proteoglycans [247], followed by back-transport into the Golgi complex and endoplasmic reticulum [248].

All these studies suggest that chemical approaches can be used to tailor cationic modifications that will improve cellular uptake and avoid the problem of endosomal entrapment.

An alternative approach to conjugation of PNA has been direct modification of PNA's backbone. Several groups have explored this kind of modifications. For example, Ly and co-workers introduced guanidine groups at α- [249] and γ-positions [250] of PNA's backbone. The guanidine-modified PNA (GPNA) maintained a good sequence selectivity and was readily taken up by several cell lines

(HCT116, human ES, and HeLa), which was attributed to the cationic guanidine modifications. Moreover GPNA was less toxic to cells than a PNA-polyarginine conjugate. More recently, Manicardi *et al*. used both α- and γ-modified GPNA 15-mers to inhibit microRNA-210 in K562 cells [251].

Another problem related to PNAs is the limited sequence scope of double-stranded nucleic acids that can be recognized by PNAs. To expand the recognition of double-stranded nucleic acids, many strategies have been used, including the one proposed by Ly and co-workers, who showed that γ-methylation pre-organizes PNA into right-handed helix and enhanced its ability to form strand-displacement complex with mixed sequence DNA [252].

The non-charged and less hydrophilic nature of PNAs also affect the solubility in aqueous solutions, which has been demonstrated to be improved through the insertion of one or more lysine residues as a tail into the PNA backbone [253].

Regarding their bioavailability, PNAs are inherently large hydrophilic molecules which are rather quickly excreted from the organism through the kidneys, typically with half-lives of less than 1 h [254]. Thus, to obtain sufficient bioavailability, PNA oligomers must be chemically modified and/or formulated for *in vivo* use. It has been demonstrated that PNA-tetralysine conjugates are active in several tissues in a transgenic mouse expressing a splicing defective GFP gene [255] and PNA-arginine conjugates have shown activity in muscles in the mdx muscular dystrophy mouse model [256]. Despite these very encouraging data, it is quite clear that better and more general delivery methods are eagerly awaited, and these results should be accompanied by comprehensive pharmacokinetics and toxicity studies, of which only very few are available at present.

1.4.3 Targeting G-quadruplexes with peptide nucleic acids

PNAs can provide a powerful instrument to target G-rich regulatory DNA sequences, stabilize or destabilize G4s, and ascertain the effect of G4 formation on the expression of a particular gene. Based on existing experimental data, several modes of gene transcription regulation through G4 targeting via PNAs are possible and many different PNAs were designed to specifically bind G-rich, C-rich, and/or both DNA strands [239].

Initially, PNA probes containing Gs were designed to invade homologous DNA G4-forming sequences and participate in the formation of heterogeneous DNA-PNA G-tetrads. This idea was first put forward by Armitage and co-workers who noticed that G-rich PNA formed hetero-quadruplexes with homologous DNA oligomers. This process occurs because the Hoogsteen base pairing between Gs is bolstered by the lack of negative charge on the PNA backbone. These intermolecular structures are

highly stable, their measured melting temperatures being higher than those of the corresponding intramolecular DNA G4s [257]. DNA sequences with four G-runs can also be targeted in this mode via formation of two consecutive G4s each consisting of two PNA- and two DNA-strands [258]. To increase the efficiency and specificity of G4 invasion versus duplex formation, Lusvarghi *et al.* utilized modified PNAs in which they incorporated abasic sites as well as chiral modifications to the backbone [259].

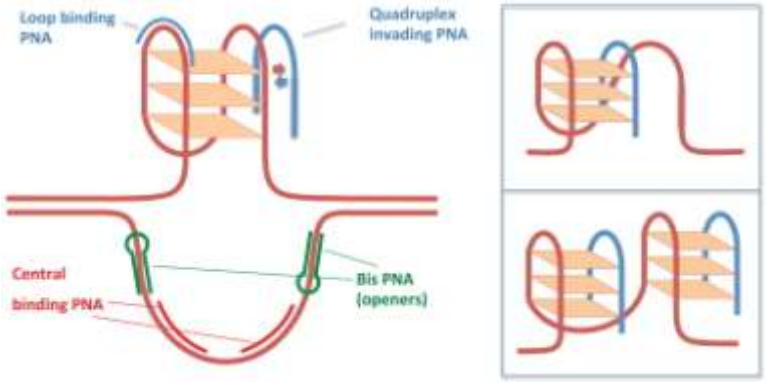


Figure 1.26: Proposed modes of PNA binding to DNA G4 formed in duplex DNA. Quadruplex binding PNA (shown in dashed) replaces one or more DNA strands containing quadruplex-core forming runs of Gs. Figure from [239].

Paul and coauthors proposed another model of G4 targeting: the designed a G-rich PNA probe that combines with three G-runs of a human telomere sequence to form an intramolecular PNA-DNA G4 in a "3+1" mode and the resulting complex mimics the biologically relevant pure DNA telomeric G4. However, the PNA probes described above could not invade the duplex DNA even if the latter contained the motif complementary to the PNA sequence and therefore, further efforts are required to increase the ability of PNA of invade duplex DNA thus making this types of strategies viable [260].

Instead of G4 invasion, an alternative approach was used by Amato *et al.* to target the ssDNA in the exposed loops of G4 structure and exclude disruption of G-tetrads within the G4. They screened a small library of short PNAs complementary to a part of G4-forming DNA sequence that does not contain Gs involves in G4 formation. Depending on PNA length and ionic conditions, PNAs were able to bind to the loops of the G4 and either stabilize or disrupt the G4 [261]. In their earlier studies, Amato *et al.* also used short cytosine-rich PNA probes to show that they could form novel G4-PNA complexes in addition to the expected DNA-PNA heteroduplexes, depending on ionic conditions [262,263].

Besides targeting the G-rich strand, another different approach is based on targeting exclusively the cytosine-rich strand complementary to the G4-forming DNA strand. Onyshchenko *et al*. demonstrated that, in this case, a PNA probe invades the double helix and binds the cytosine-rich strand thus facilitating G4 extrusion in the G-rich strand without interfering with native DNA G4 conformation [264].

1.4.4 Anti-HIV-1 peptide nucleic acids

The success of the anti-HIV activity of PNA depends on two major factors. The first factor is selection of a conserved region on the viral genome; the second is that the target region should be crucial to the HIV life cycle [265].

Various studies have targeted the HIV-1 TAR region, showing encouraging results. Mayhood *et al.* showed that a PNA targeting the TAR sequence of HIV-1 RNA genome was able to prevent Tat-TAR interaction by efficient sequestration of the TAR [266]. A triphenyl-phosphonium-tagged PNA targeted to HIV-1 TAR was found to be efficiently taken up by cells and to block HIV-1 replication in cell culture [267]**.** In one unique study, PNA complementary to the HIV-1 TAR region was conjugated to neamine, a polycationic aminoglycoside. This conjugate was able to enter inside cells and inhibited HIV-1 replication [268]. However, the most efficient cellular uptake of anti-HIV-1 PNA was noted with the use of cell-penetrating peptides, including penetratin, transportan, and HIV-1 Tat-derived peptide as vehicle to deliver anti-HIV-1 PNA into cells [265]. For example, an anti HIV-1 TAR PNA conjugated with transportan peptide was shown to be rapidly taken up by cells, block Tat-mediated trans-activation of HIV-1 transcription, and strongly inhibit HIV-1 production in chronically HIV-1-infected H9 cells [269]. More recently, Das *et al.* employed a PNA-aminosugar conjugate targeting TAR, which showed a high biovability in human cells and strongly inhibited Tat-mediated trans-activation of HIV-1 transcription [270].

Another relatively conserved and crucial region in the HIV genome is the U5 primer binding site (PBS), where the $tRNA^{lys}$ anneals and primes the synthesis of negative-strand strong-stop DNA. A 19-mer PNA targeting HIV-1 PBS blocked *in vitro* reverse transcription by binding to a pre-primed RNA primer duplex [271]. In onother study, a 18-mer PNA targeting the PBS region blocked the initiation of reverse transcription by displacing the $tRNA^{Lys3}$ primer from the viral genome [272]**.**

Another conserved sequence near the 3′ end of the HIV-1 gag-pol trans-frame domain, when challenged by PNA, inhibited the production of virus up to 98.4% in chronically infected H-9 cells [273].

More recently, Parkash *et al*. showed that PNAs can be successfully used to as an antisense to inhibit HIV-1 genome dimerization and template switching process [274].

Chapter 2

Aim of the study

We have shown that the formation of G4s in the HIV-1 LTR results in decreased viral transcription, with an effect that is enhanced by the presence of the cellular protein nucleolin and G4 ligands. In this context, the first aim of this project was to explore the possibility of selectively targeting HIV-1 LTR G4s. Selectivity toward LTR G4s is fundamental in order to avoid side effects due to interaction with cellular G4s. Consequently, this approach constitutes the basis for the development of novel anti-HIV-1 compounds with an innovative mechanism of actions. We faced the selectivity challenge through different approaches: drug-like small molecules and NDI-PNA conjugates. It is worth noting that known G4 ligands are all characterized by large flat scaffolds which make them poorly druggable, which is why it is necessary to discover small molecules able to bind G4s. On the other hand, the employment of PNAs allows conveying a ligand directly to the G4 of interest.

The second part of this thesis was focused on the identification and characterization of new proteins that specifically recognize LTR G4s and modulate the folding/unfolding of these structures. This will provide new insights in the regulation mechanisms of HIV-1 transcription, which may also lead to new possible targets for the design of specific inhibitors.

Chapter 3

Materials and methods

3.1 Oligonucleotides used in this study

Synthetic DNA oligonucleotides used in this study were purchased from Sigma-Aldrich (Milan, Italy) in lyophilized form. They were dissolved in TE buffer (Tris-HCl 10 mm, EDTA 1 mm pH 8.0) at stock concentrations 1 mM or 100 μM and stored at -20 °C.
The oligonucleotide sequences used in this study are reported in tables following each method.

3.2 Fluorescence Resonance Energy Transfer analysis

Fluorescent resonance energy transfer (FRET) based method was developed by Mergny's group [275] and requires oligonucleotides labeled at 3'- and 5'-end with a fluorophore and a quencher, respectively.

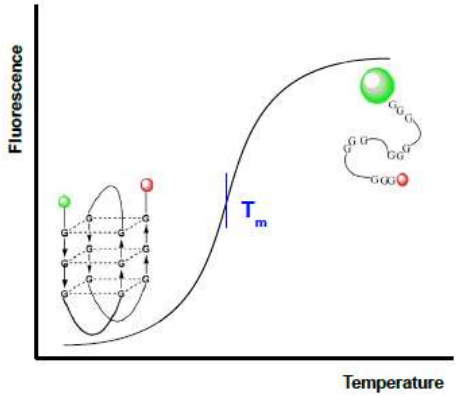


Figure 3.1: Schematic representation of FRET-based melting temperature experiments. When G4 is in the folded form, the fluorescence is quenched. During thermal unfolding the distance between the fluorophore and the quencher increases and the fluorescence is restored.

The principle of the method consists in the fact that the formation of a G4 in a labeled oligonucleotide is accompanied by a decrease in the distance between the donor and the acceptor, thus leading to a more efficient energy transfer from the excited donor fluorophore to the acceptor. Therefore the light emitted from the donor results to be quenched in a distance-dependent extent. When the G4 undergoes thermal denaturation, the quencher moves apart and the fluorescence is restored. The variation of fluorescence emitted by the fluorophore is measured at appropriate wavelength to obtain melting curves (Figure 3.1).

Tested oligonucleotides were 5'-labeled with 6-carboxyfluorescein (FAM) and 3'-labeled with 6-carboxytetramethylrhodamine (TAMRA). These two molecules work respectively as fluorescence donor and acceptor. For fluorescence melting curves, 20 µL samples of oligonucleotides (0.25 µM) were annealed in lithium cacodylate buffer (10 mM, pH 7.4) supplemented with KCl (100 mM) for 5 min at 95 °C and slowly cooled to room temperature to allow G4 formation. Tested compounds/conjugates at indicated concentration were added to the samples 4 h after annealing. After overnight stabilization at 4 °C, samples were processed by Light Cycler II (Roche, Milan, Italy) or Light Cycler 480 (Roche, Milan, Italy) and the oligonucleotide melting was monitored by following FAM emission in the temperature range of 30-95 °C, with 1 °C/min gradient. Data analysis was performed plotting the normalized fluorescence in function of the temperature. Melting profiles were normalized as previously described [276]. The T_m was determined as the temperature corresponding to the 0.5 fraction of the normalized fluorescence.

To study hnRNP A2/B1:G4 interaction, oligonucleotides were diluted to 0.1 µ M in lithium cacodylate buffer (10 mM, pH 7.4) and different concentration of KCl (indicated in figure legends), heat denatured for 5 min at 95 °C, and folded into G4 at room temperature for 16 h. Samples were incubated alone, with hnRNP A2 (1 µM) or bovine serum albumin (BSA, negative control, 1 µM) and subsequently fluorescence intensity was monitored from 30 °C to 95 °C (1 °C/min) in a LightCycler II (Roche, Milan, Italy) by observing FAM emission. FRET efficiency (E) was calculated as the total fluorescence intensity of the donor in the absence (I_D) and presence of the acceptor (I_{DA}), by using $E = 1 - (I_{DA}/I_D)$. I_D and I_{DA} values were obtained on a LS 55 Fluorescence spectrometer (Perkin Elmer, Waltham, Massachusetts, USA) setting the excitation wavelength at 480 nm and recording the emission from 500 to 650 nm. The dual-labelled oligonucleotide was converted into the corresponding duplex, in which the fluorophores are at a distance (R) for which FRET is ~ 0. R was calculated from FRET efficiency values as $R = R_0 \sqrt[6]{(1/E) - 1}$, where R_0 (Förster distance) is the distance at which energy transfer is

50% of the maximum value. Between FAM and TAMRA fluorophores, R_0 is assumed to be 50 Å [277]. F-test (F) and the probability (P) values were calculated using R statistical environment (v. 3.3.2) [278]. A conventional alpha = 0.05 was considered to evaluate the test significance.

Table 3.1: Oligonucleotides used for FRET

Project	Oligonucleotide	5' → 3' sequence
Screening hnRNP A2/B1	LTR-IV	6FAM-TGGGCGGGACTGGGGAGTGGT-TAMRA
Screening hnRNP A2/B1	LTR-III	6FAM-GGGAGGCGTGGCCTGGGCGGGACTGGGG-TAMRA
NDI-PNA conjugates hnRNP A2/B1	LTR-III+IV	6FAM-TGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGT-TAMRA
hnRNP A2/B1	LTR-II+III+IV	6FAM-TGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGT-TAMRA
Screening NDI-PNA conjugates	dsDNA	6FAM- CTATAGCGCGCTATAG –TAMRA
Screening	hTel	6FAM-AGGGTTAGGGTTAGGGTTAGGG-TAMRA

In FRET-based competition assay 5'-FAM and 3'-TAMRA labeled oligonuclotide and unlabeled oligonucleotides (competitors) were separately folded for 5 min at 95 °C in lithium cacodylate buffer (10 mM, pH 7.4) supplemented with KCl (100 mM). After 4 h at room temperature, labeled oligonucleotide (0.25 μM) was mixed with increasing amounts of competitor (1-8-folds excess) in the presence or absence of NDI-PNA and PNA (1 μM). Fluorescence was measured after overnight stabilization at 4°C by Light Cycler II (Roche, Milan, Italy) at 30 °C for 1 min. The variation of fluorescence (ΔF) was calculated as the difference between the fluorescence of the labeled oligonucleotide in the presence of NDI-PNA or PNA and competitor and the basal fluorescence of the oligonucleotide alone. Data were indicated as percentage (ΔF%) of the control (difference between the fluorescence of the labeled oligonucleotide in the presence of NDI-PNA and the basal fluorescence of the oligonucleotide alone).

Table 3.2: Oligonucleotides used for FRET competition assay

Project	Oligonucleotide	5' → 3' sequence
NDI-PNA conjugates	LTR-III+IV labeled	6FAM-TGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGT-TAMRA
NDI-PNA conjugates	LTR-III+IV unlabeled	TGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGT
NDI-PNA conjugates	hTel unlabeled	AGGGTTAGGGTTAGGGTTAGGG

3.3 Circular Dichroism analysis

Circular dichroism (CD) is a polarized light spectroscopy that is used to discriminate G4s from other secondary structures. In fact, CD spectra generated by G4s display characteristic shapes useful for structure recognition and conformational analysis [279]: parallel G4s show a maximum positive peak at 260 nm and a negative one at 240 nm, whereas antiparallel G4s exhibit a positive peak at 290 nm and a negative one at 260 nm; also mixed conformations are possible and they are characterized by two positive peaks at 290 nm and 260 nm, and a negative one at 240nm. CD can also be employed to evaluate the thermal stability of G4s which is studied following the variation of molar ellipticity corresponding to the maximum peaks as a function of the temperature: this allow calculation of T_m. In addition, CD is useful to study G4-ligand interactions in order to evaluate possible conformational changes induced by the ligand binding and possible G4 stabilization by following the variation of T_m upon addition of compound [280].

For CD analysis, oligonucleotides were diluted to the final indicated concentration in lithium cacodylate buffer (10 mM, pH 7.4) and KCl 100 mM. Samples were annealed by heating at 95 °C for 5 min and gradually cooled to room temperature to allow G4 formation. Where indicated, compounds/conjugates were added at the final reported concentration. CD spectra were recorded on a Chirascan-Plus (Applied Photophysisics, Leatherhead, United Kingdom) equipped with a Peltier temperature controller using a quartz cell of 5 mm optical path length, over a wavelength range of 230–320 nm. The reported spectrum of each sample represents the average of 2 scans and it is baseline corrected for signal contributions due to the buffer. Observed ellipticities were converted to mean residue ellipticity (θ) = deg×cm2×dmol−1 (mol. ellip.). For the determination of T_m, spectra were recorded over a temperature range of 20-90 °C, with temperature increase of 5 °C. T_m values were

calculated according to the van't Hoff equation, applied for a two-state transition from folded to unfolded state, assuming that the heat capacity of the folded and unfolded states are equal [281].

Table 3.3: Oligonucleotides used for CD

Project	Oligonucleotide	5' → 3' sequence
Screening	LTR-IV	CTGGGCGGGACTGGGGAGTGGT
hnRNP A2/B1	LTR-IV	GGGCGGGACTGGGGAGTGG
Screening hnRNP A2/B1	LTR-III	GGGAGGCGTGGCCTGGGCGGGACTGGGG
NDI-PNA conjugates hnRNP A2/B1	LTR-III+IV	GGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGG
Screening NDI-PNA conjugates	hTel	AGGGTTAGGGTTAGGGTTAGGG
hnRNP A2/B1	LTR-II+III+IV	GGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGG

3.4 *Taq* polymesare stop assay

Taq polymerase stop assay is used to investigate the G4 formation in a DNA sequence of interest based on the assumption that a G4 sterically impedes the progression of a DNA polymerase during the elongation of a ^{32}P-labeled primer complementary to the 3'-end of the G4 template [282] (Figure 3.2). When the template containg G4 forming regions is elongated, truncated products are formed and are descriminated in a denaturing polyacrilamide gel. Specific markers are used to evaluate whether the stop bands correspond to the Gs involved in the G4 formation. Markers are prepared based on Maxam & Gilbert sequencing and DNA breakings are produced by treatment of the double stranded sequence firsltly with formic acids and then with piperidin.

The DNA primer was labeled with [γ-^{32}P-ATP] using T4 polyonucleotide kinase (Fermentas, Waltman, Massachusetts, USA) at 37 °C for 30 min. The labeled primer was purified with Illustra MicroSpin G-25 columns (GE Healthcare, Milan, Italy). G4 DNA templates (36 nM) were diluited in lithium cacodylate buffer (10 mM, pH 7.4) in the presence or absence of KCl at the indicated concentrations and ^{32}P-labeled primer (72 nM). The mix was incubated at 95 °C for 5min to allow denaturation, followed by slow cooling to room temperature to allow both primer annealing and G4 folding. When

indicated, compounds/conjugated were added at reported concentrations and incubated overnight. The primer was subsequently extended on the template strand by adding 1X PCR reaction buffer (Applied Biosystems Carlsbad, California, USA), 0.1 mM dNTPs and 2 U/reaction of Ampli*Taq* Gold DNA polymerase (Applied Biosystem, Carlsbad, California, USA) at the indicated temperature for 30 min. The reaction was stopped by EtOH precipitation and the primer extension products were separated on a 16% denaturing gel and visualized by phosphoimaging (Typhoon FLA 9000, GE Healthcare, Milan, Italy). Markers were prepared by PCR reaction with ^{32}P-labeled primer. PCR products were treated with formic acids for 5 min at 25 °C and then with piperidin for 30 min at 90 °C.

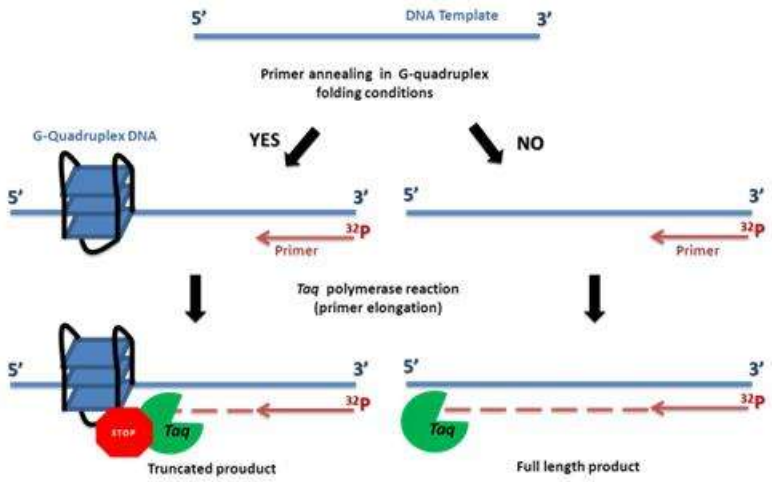


Figure 3.2: Schematics of *Taq* polymerase stop assay. G4s sterically impede the progression of a DNA polymerase during the elongation of a ^{32}P-labeled primer complementary to the 3'-end of the G4 template, thus allowing the formation of truncated products.

Table 3.4: Oligonucleotides used for *Taq* polymerase stop assay

Project	Oligonucleotide	5' → 3' sequence
Screening hnRNP A2/B1	LTR-III	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC
Screening	LTR-IV	TTTTTGGGCGGGACTGGGGAGTGGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC
Screening NDI-PNA conjugates	hTel	TTTTTGGGTTAGGGTTAGGGTTAGGGTTTTTCTGCATATAAGCGCTTTTTGCC
Screening NDI-PNA conjugates	Cnt no G4	TTGTCGTTAAAGTCTGACTGCGAGCTCTCAGATCCTGCATATAAGCAGCTGCTTTTTGCC
NDI-PNA conjugates hnRNP A2/B1	LTR-III+IV	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC
hnRNP A2/B1	LTR-II+III+IV	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC
hnRNP A2/B1	LTR-II+III+IV M4+5	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGTGCGTGACTGGGGAGTGGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC
Screening NDI-PNA conjugates hnRNP A2/B1	Primer	GGCAAAAAGCAGCTGCTTATATGCAG

In competition assay, G4 DNA templates were prepared as previously reported. DNA oligonucleotides designed without the sequence complementary to the primer (competitors) were separately incubated at 95°C for 5 min. After that, the DNA template was mixed with increasing amounts of competitor (1-8X) in presence or absence of NDI-PNA. The primer was next extended at 42°C and products were separated on a 16% denaturing gel and visualized by phosphoimaging as previously described.

Table 3.5: Oligonucleotides used for competition by *Taq* polymerase stop assay

Project	Oligonucleotide	5' → 3' sequence
NDI-PNA conjugates	LTR-III+IV labeled	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGTTTTTTGCATATAAGCAGCTGCTTTTTGCC
NDI-PNA conjugates	LTR-III+IV unlabeled	GGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGG
NDI-PNA conjugates	hTel unlabeled	AGGGTTAGGGTTAGGGTTAGGG

3.5 Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) is an optical sensor based technique allowing the real time measurement of biomolecules interaction. It is a highly sensitive label-free detection method based on a phenomenon in which an incident light photon hits a thin metal surface. The most commonly used metals are silver and gold. In the hit with this surface, part of the energy of the photon is transferred to the metal electrons. If the transferred energy is sufficient to excite electrons to resonance, they start to oscillate, producing a so-called plasmon. The propagation of plasmonic oscillations generates an electric field on the metal surface. This electromagnetic wave results to be strongly concentrated in the dielectric of a biochemical probe layer in contact with the metal surface. For this reason the type of probe (refractive index and mass) highly influences the intensity of reflected light, giving rise to a particular SPR signal. The SPR signal can indeed be simply defined as the change in the critical angle due to the refractive index of a probe in contact with the excited metal surface. By immobilizing one of the binding partners on the sensor chip surface it is possible to investigate binding interactions between G4s and small molecules or proteins [283]. Biotinylated oligonucleotides can be immobilized on the chip surface coated with streptavidin. Affinity and kinetic parameters of interaction with small molecules are measured by injecting increasing concentrations of tested compounds. To investigate G4-protein interactions the preferential choice is to immobilize the proteins on sensor chip surface instead of oligonucleotides, because of the relatively high molecular weight of the proteins compared to the oligonucleotides and because of the advantage in small amounts of proteins required for immobilization. Proteins are immobilized on dextran coated sensor chip by amine coupling.
SPR was performed on the Biacore T100 platform (GE Healthcare, Life Science, Milan, Italy).

3.5.1 SPR analysis of G-quadruplex-hits interactions

For oligonucleotides immobilization we used SA series S sensor chip (Biacore, GE Healthcare, Life Science, Milan, Italy) were used for high affinity capturing of biotinylated oligonucleotides. We docked the SA chip and primed the system for conditioning the surface. As immobilization/running buffer we used HEPES-KCl buffer (HEPES pH 7.4 10 mM, KCl 200 mM, EDTA 3 mM). Folded 5'-biotin labeled oligonucleotides were diluted to 30 nM in HEPES-KCl buffer, allow acquiring the G4 folding at room temperature after heat denaturation (5 min at 95 °C) and injected on the chip to reach the response of around 500 immobilized RU. Flow cell 1 was blank immobilized to allow reference subtraction. Chemical compounds binding analysis was performed at a flow rate of 25 µL/min, with

contact time of 180 s and dissociation time of 240 s in HEPES-KCl buffer. Sensograms were obtained in the concentration range of 39 nM – 1.25 μM. After each compound injection the chip surface was regenerated with Glycine 10 mM pH 2.0 solution (GE Healthcare, Life Science, Milan, Italy). All sensorgrams were corrected by double reference subtraction of blank flow cell response and buffer injection response. Data were fitted to a local 1:1 binding model using BIAevaluation software (GE Healthcare, Life Science, Milan, Italy).

3.5.2 SPR analysis of G-quadruplex-hnRNP A2/B1 interactions

About 1000 RU of human recombinant hnRNP A2 (Origene Technologies, Rockville, USA) were immobilized on Serie S sensor chip CM5 (GE Healthcare, Life Science, Milan, Italy) by amine coupling chemistry. Immobilization was performed in HEPES-NaCl buffer (HEPES pH 7.4 10 mM, NaCl 150 mM, EDTA 3 mM). Flow cell 1 was blank immobilized to permit reference subtraction. Binding analysis of oligonucleotide sequences was performed at a flow rate of 25 μL/min, with contact time and dissociation time of 120 s in HEPES-KCl buffer (HEPES pH 7.4 10 mM, KCl 150 mM, EDTA 3 mM). Oligonucleotides folding into G4s was performed in HEPES-KCl buffer after heat denaturation at 95 °C for 5 min and gradual cooling at room temperature. Sensorgrams were obtained in the concentration range of 31.25 nM –2 μM. After each oligonucleotide injection the chip surface was regenerated with NaCl 2.5 M solution. All sensograms were corrected by double 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, Life Science, Milan, Italy).

Table 3.6: Oligonucleotides used for SPR

Project	Oligonucleotide	5' → 3' sequence
Screening	LTR-III	biot-GGGAGGCGTGGCCTGGGCGGGACTGGGG
Screening	LTR-IV	biot-AGGGTTAGGGTTAGGGTTAGGG
Screening	hTel	biot-TGGGCGGGACTGGGGAGTGGT
hnRNP A2/B1	LTR-II+III+IV	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGTTTTT
hnRNP A2/B1	LTR-II+III+IV M4+5	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGTGCGTGACTGGGGAGTGGTTTTT
hnRNP A2/B1	LTR-II+III+IV random	AAAAACTACTGCACGCTCGCTACGACGACACTGTCGCGCATACAAGCTGCAAAAA

3.6 Cells

TZM-bl is a HeLa cell clone that has been engineered to express CD4 and CCR5 and contains integrated reporter genes for firefly luciferase and E.coli β-galactosidase under control of HIV-1 LTR promoter, permitting sensitive and accurate measurements of infection. The cells are highly permissive to infection by most strains of HIV. TZM-bl reporter cell line was provided by NIH AIDS Research Program.

HEK 293T (Human Embryonic Kidney 293T) cell line is a highly transfectable derivative of human embryonic kidney 293 cells. HEK 293T cell line was provided by ATCC.

Both cell lines were maintained in Dulbecco's modified Eagle Medium (DMEM) (Gibco, Life Technologies, Monza, Italy) supplemented with 10% heat inactivated fetal bovine serum (FBS, Gibco, Life Technologies, Monza, Italy) at 37°C in a humidified CO_2-controlled atmosphere.

3.7 Virus stock production

The HIV-1 strain NL4-3 viral stock was prepared by transfection of HEK 293T cells with wild-type X4 proviral genome (NIH AIDS Research and Reference Reagent Program), using CaIPhos mammalian transfection kit (Clontech Laboratories, Santa Clara, CA, USA) according to the manufacturer's protocol. Viral particles in supernatants were collected and titrated by Reed and Muench method [284].

3.8 Antiviral assay in HIV-1 infected TZM-bl cells

HIV-1 infectivity was measured using the TZM-bl cells, in which the HIV infection drives transcription of the HIV-1 LTR-Luciferase reporter gene construct and thus it is possible to evaluate the antiviral effect of test compounds following the luciferase expression. Cells were seeded in 96-well plates (10,000 cells/well) and grown overnight to permit adherence prior to treatment and viral infection. Cells were next infected with HIV-1 NL4.3 strain (MOI 0.5), treated with serial dilutions of compounds (250.00-0.98 μM) and incubated at 37 °C. After 48 h HIV-1 production was assessed following the LTR-luciferase signal using the britelite™ plus Reporter Gene Assay System (Perkin Elmer, Waltman, Massachusetts, USA) according to the manufacturer's protocol.

Cytotoxicity of test compounds was evaluated in parallel by MTT assay (Sigma-Aldrich, Milan, Italy). The MTT assay is a colorimetric assay for assessing cell viability based on the conversion of yellow MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) into purple formazan crystals by living cells, which determines mitochondrial activity. Since for cells the total mitochondrial activity is related to the number of viable cells, this assay is broadly used to measure the in vitro cytotoxic effects of drugs of interest on cell lines. Absorbance data of formazan are proportional to the number of viable cells. 10 μL of freshly dissolved solution of MTT (5 mg/mL in phosphate buffered saline (PBS)) were added to each well, and after 4 h of incubation, MTT crystals were solubilized in solubilization solution (10% sodium dodecyl sulphate (SDS) and 0.01 M HCl). After overnight incubation at 37 °C, absorbance was read at 540 nm. Data were expressed as mean values of at least two individual experiments conducted in triplicate. The percentage of cell survival was calculated as follows:
cell survival = (Awell − Ablank)/(Acontrol − Ablank) × 100, where blank denotes the medium without cells. The 50% inhibitory concentration (IC_{50}) was defined as the concentration of tested compound that inhibit HIV-1 production by 50%, measured following the LTR-Luciferase signal. The 50% cytotoxic concentration (CC_{50}) was defined as the concentration of compound required to reduce cell grown by 50%. The selectivity index (SI) is the relative effectiveness of the tested compound in inhibiting viral replication compared to inducing cell death (CC_{50} value/IC_{50} value).

3.9 Luciferase reporter assay

This assay is particularly useful to investigate the effect of G4 ligands at the promoter level. The LTR DNA region was amplified by PCR on the HIV-1 genome and then the LTR amplicon was subcloned into a construct (pGL4.10-Luc2) (Promega, Madison, Wisconsi, USA) where it drives the expression of the Firefly luciferase gene. Mutant vector (pGL4.10-Luc2/LTR M4+5) was also generated using QuikChange mutagenesis kit (Stratagene/Agilent Technologies, Santa Clara, California, USA). Vectors pGL4.10Luc2/LTR and mutant pGL4.10Luc2/LTR (300 ng each) were transfected in 350,000 HEK 293T cells per well onto 6-well plates, using TransIT-293 transfection reagent (Mirus, Madison, Wisconsi, USA) following the manufacturer's instructions. After 30 min, cells were treated with serial dilutions of D8 (16-1 μM) and incubated over-night. Expression of firefly luciferase was determined 24 h after transfection using the britelite plus Reporter Gene Assay System (Perkin Elmer, Waltman, Massachusetts, USA) at a Victor X2 multilabel plate reader (Perkin Elmer, Waltman, Massachusetts,

USA). The cell number and the protein content (quantified by BCA assay (Thermo Fisher Scientific, Waltham, Massachusetts, USA)) were also measured 24 h after transfection and were used as a control for transfection efficiency. Both luciferase to protein content and luciferase to cell number output ratios were calculated. Each assay was performed in triplicate and the experiment was repeated at least three times.

The MTT assay was performed to assess cell viability on HEK 293T cells (6,000 cells/well), following the protocol previously reported. Cells were treated with serial dilutions of D8 (up to 200 μM) and cytotoxicity was measured 24 h post-treatment.

3.10 Mass spectrometry analysis

Mass spectrometry (MS) is an analytical technique that ionizes chemical species and separate the ions based on their mass-to-charge ratio. When performed in native conditions it represents a powerful tool to investigate both G4 structure and G4/small molecule binding [285,286]. Native mass spectrometry was employed to evaluate the binding of hits to LTR-IV G4. Samples were analyzed by direct infusion electrospray ionization (ESI) on a Xevo G2-XS QTOF mass spectrometer (Waters, Manchester, United Kingdom). This is a high-resolution instrument that allowed us to visualize the isotopic pattern, identify the charge state, and therefore unambiguously calculate the neutral mass of the detected species. The injection was automatically performed by an Agilent 1290 Infinity HPLC (Agilent Technologies, Santa Clara, California, USA) equipped with an autosampler; the carrying buffer was composed by triethylamine (TEA)/120 mM hexafluoroisopropanol (HFIP) pH 7.4 with 20% isopropanol (IPA). A volume of 5 μL of each sample was typically injected for each analysis. To investigate the binding, the spectra were acquired in negative ion mode. ESI source settings were: electrospray capillary voltage 1.8 kV; source and desolvation temperatures 45 and 65 °C, respectively; sampling cone voltage 65 V. All these parameters ensured minimal fragmentation of the DNAs complexes.

To characterize the hits library and the putative active compounds, the hits were dissolved at the concentration of 50 μM in acetonitrile/water (50:50 v/v) containing 0.1% of formic acid and the spectra were acquire in positive ion mode, ESI source settings were: electrospray capillary voltage 1.5 kV; source and desolvation temperatures 100 °C and 350 °C, respectively; sampling cone voltage 40 V. MS/MS spectra were acquired isolating and fragmenting the ions of interest in the collision cell with a ramp of collision energy between 60 to 80 V.

The instrument was calibrated using a 2 mg/mL solution of sodium iodide in 50% IPA. Additionally, the use of the LockSpray during analysis provided a typical < 5 ppm mass accuracy. The internal standard LockSpray consisted in a solution of leu-enkephalin 1 μg/mL in acetonitrile/water (50:50 v/v) containing 0.1% of formic acid.

Table 3.7: Oligonucleotide used for mass spectrometry

Project	Oligonucleotide	5' → 3' sequence
Screening	LTR-IV	TGGGCGGGACTGGGGAGTGGT

3.11 Confocal microscopy

Confocal microscopy was employed to evaluate cell entry of NDI-PNA conjugates, exploiting the intrinsic fluorescence of NDI. TZM-bl cells were seeded in 24-wells plates (80.000 cells/well) pretreated with polylysine (1μg/μL) for 5 min at room temperature, and grown overnight. Cells were treated with conjugates (50 μM) and, after 3 h, were washed 5 times with PBS to remove the excess of conjugates. Cells were fixed with paraformaldehyde (PFA) 2% for 20 min at room temperature. After 5 washes with PBS, fluorescence was evaluated in the 560-700 nm range, using an excitation wavelength of 543 nm by Leica TCS SP2 confocal microscope (Leica, Wetzlar, Germany).

3.12 Protein nuclear extraction and pull-down assay

Protein nuclear extracts of HEK 293T cells were obtained by using NXTRACT kit (Sigma-Aldrich, Milan, Italy). HEK 293T protein nuclear extracts (150 μg) were incubated with biotinylated LTR-II+III+IV G4, mutated M4+5 and random R oligonucleotides folded (600 nM) in 250 μL of reaction containing Tris-HCl 20 mM, pH 8, KCl 30 mM, MgCl2 1.5 mM, protease inhibitor cocktail (Sigma-Aldrich, Milan, Italy) 1%, NaF 5 mM, Na_3VO_4 (Sigma-Aldrich, Milan, Italy) 1 mM, poly [dI-dC] (Sigma-Aldrich, Milan, Italy) 1.25 ng/μL for 2 h at 37 °C. The binding reaction was followed by incubation (2 h at 37 °C) with 30 μL of streptavidin-agarose beads (Sigma-Aldrich, Milan, Italy). After PBS (0.2M) and NaCl (1M) washes, beads were collected by brief centrifugation, resuspended in 50 μL of Laemmli buffer, and finally incubated at 95 °C for 5 min. Supernatants were separated in 12% SDS-

PAGE and, after coomassie staining, the gel lanes of LTR G4 and LTR M4 +5 were directly in-gel digested and analyzed for protein identification by mass spectrometry. Briefly, gel lanes cut in ~0.5 cm pieces were first washed with 50% CH_3OH and 2.5% acetic acid, dehydrated with CH_3CN, then reduced with 30 µL of dithiothreitol (DTT) (10 mM in 100 mM NH_4HCO_3) for 30 min at room temperature; the excess of DTT was next neutralized by alkylation with 30 µL of iodoacetamide (50 mM in 100 mM NH_4HCO_3) for 30 min at room temperature. Bands were washed with 100 mM NH_4HCO_3, dehydrated with CH_3CN twice, and then digested overnight with 1 µg of MS-grade trypsin (ThermoFisher Scientific, Waltham, Massachusetts, USA) in 50 µL of 50 mM NH_4HCO_3. Peptides were extracted twice with 5% formic acid and twice with 50% CH_3CN/5% formic acid. The peptide mixture was further desalted in a silica nanocolumn (Polymicro Technologies, Phoenix, Arizona, USA) packed in house with pinnacle C18 pack material (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The desalted mixture was finally analyzed by ESI on a LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific Waltham, Massachusetts, USA) utilizing quartz emitters produced in house. A stainless steel wire was inserted through the back-end of the emitter to supply an ionizing voltage that ranged between 0.8 and 1.2 kV. Putative peptides present in wt LTR-II+III+IV (but not in M4+5) samples were submitted to tandem mass spectrometric (MS/MS) analysis. The masses of the 50 most intense fragment ions were employed to perform a Mascot Database Search51 to identify their parent protein. Significant Mascot hits were accepted as positive matches and their ion score reported Supernatants were further analyzed by Western blot with an anti-hnRNP A2/B1 antibody (mouse monoclonal DP3B3; Santa Cruz Biotechnology, Dallas, Texas, USA).

Table 3.8: Oligonucleotides used for pull-down

Project	Oligonucleotide	5' → 3' sequence
hnRNP A2/B1	LTR-II-III+IV	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGG GGAGTGGTTTTT-BtnTg
hnRNP A2/B1	LTR-II-III+IV M4+5	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGTGCGTGACTGGG GAGTGGTTTTT-BtnTg
hnRNP A2/B1	random	AAAAACTACTGCACGCTCGCTACGACGACACTGTCGCGCATAA AGCTGCAAAAA-BtnTg

3.13 Immunoblot analysis

Immunoblot analysis was performed on cell protein extracts obtained in RIPA Buffer (20 mM Tris-HCl pH 7.5; 150 mM NaCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate, 1 mM Na_3VO_4, 1x protease inhibitors). Protein concentrations were quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Each sample was electrophoresed on 10% SDS-PAGE and transferred to a nitrocellulose blotting membrane (Amersham TM Protan TM, GE Healtcare Lifescience, Milan, Italy) by using trans-blot SD semi-dry transfer cell (Bio-Rad Laboratories, Milan, Italy). The membranes were blocked with 2.5% skim milk in PBS-T (0.05% Tween 20 in PBS). Membranes were incubated with the respective primary antibody directed against hnRNP A2/B1 (mouse monoclonal; Santa Cruz Biotechnology, Dallas, Texas, USA), alpha-tubulin (mouse monoclonal; Sigma-Aldrich, Milan, Italy). After three washes in PBS-T, membranes were incubated with ECL Plex Goat-α-Mouse IgG-Cy5 (GE Healthcare Lifesciences, Milan, Italy). Images were captured on the Typhoon FLA 9000 (GE Healthcare, Milan, Italy).

3.14 Silencing and reporter assay

Gene-specific pooled siRNA trilencer targeting human hnRNP A2/B1 and a scrambled negative control duplex were purchased from Origene (human hnRNP A2/B1 Trilencer-27 Human siRNA, OriGene Technologies, Rockville, Maryland, USA). HEK 293T cells or TZM-bl reporter cells were transfected with increasing concentrations of siRNA (1, 2, 4 and 8 nM for HEK 293T and 1, 5, 10 nM for TZM-bl cells) and scrambled siRNA by using Lipofectamine RNAiMAX (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA) following the manufacturer's instructions.

Only for experiments in HEK 293T cells, pLTR luciferase plasmids (pGL4.10-LTRwt or pGL4.10-LTR-M4+5) were transfected 24 h later by using Lipofectamine 3000 (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Luciferase activity was measured using the britelite plus Reporter Gene Assay System (Perkin Elmer, Waltman, Massachusetts, USA) at a Victor X2 multilabel plate reader (Perkin Elmer, Waltman, Massachusetts, USA), according to the manufacturer's instructions. Cells were lysed in 1% Triton-X100-PBS and protein concentration was determined by BCA assay (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Luciferase signals were subsequently normalized to total protein content, according to the manufacturer's protocol. Each assay was

performed in duplicate and each set of experiments was repeated at least three times. The MTT assay was performed to assess cell viability on silenced cells. HEK 293T cells (6,000 cells/well) were plated in 96-well plates and incubated for 24 h. Following siRNA transfection (1-8 nM), cells were incubated for an additional 48 h. Control cells (transfected in absence of siRNA) were treated in the exact same conditions.

Chapter 4

Results and discussion

4.1 Selective compounds targeting LTR G4s

Our group has recently demonstrated the chance to inhibit HIV-I transcription by stabilizing G4s in the LTR promoter [232] and G4 ligands have shown antiviral activity [233] suggesting the possibility to develop new antiviral compounds with G4-mediated mechanisms of action. Given that most available G4 ligands display yet insufficient specificity against different G4s, they cannot be straightforwardly employed as antiviral drugs. In order to avoid side effects due to interactions with cellular G4s and thus develop novel antiviral drugs, it is necessary to generate new G4 binders, selective for LTR G4s over cellular ones, challenge which could take advantage of the high polymorphism of G4s. We faced the selectivity challenge through different approaches. The first one consisted in the screening of the commercially available DIVERSet library from ChemBridge Corporation. The library was composed of 50,000 drug-like small molecules, selected to provide the broadest coverage of pharmacophore space within 50,000 compounds while maintaining structural diversity. Compounds displayed the following physiochemical properties: molecular weight (MW) ≤ 500, partition coefficient for n-octanol/water (clogP) ≤ 5, topological polar surface area (tPSA) ≤ 100, rotatable bonds ≤ 8; hydrogen bond acceptors ≤ 10 and hydrogen bond donors ≤ 5.

4.1.1 High throughput screening of LTR-IV G4 stabilizing ligands

They library was screened by FRET technique that involves a donor fluorophore in an excited state, the excitation energy of which can be transferred to a proximal acceptor chromophore. Given that the major determinant of FRET efficiency is the distance between the acceptor and the donor, fluorescence intensities depict nucleic acids folding state. This method is widely used and reported in the literature and has multiple advantages such as high throughput (use of 384-well plates), small amounts of oligonucleotides and compounds required, and it is relatively easy and very useful for assessing the effects of ligand binding on the structural stability of G4s. In fact, it can not only be applied to study

G4 formation and unfolding, but also stabilization by a G4 ligand can be easily evaluated as increase in the T_m [275].

The G4 target sequence used in the screening was LTR-IV, which is composed of four G runs and is normally unfolded in physiological conditions, but it has been proved that its G4 folding can be induced by G4 ligands [232]. Given that its NMR structure has been recently solved [287] by our group in collaboration with Prof. Anh Tuan Phan's group (Nanyang Technological University, Singapore), a virtual screening based on molecular modeling was also performed by our collaborators, Prof. Stefano Alcaro's group (University "Magna Graecia", Catanzaro, Italy) (not discussed here). By FRET melting analysis we found an interesting family of small molecules composed of the same chemical core, displaying stabilizing effect on LTR-IV G4. The stabilization was measured as increase T_m of the G4 folded LTR-IV in the presence of the compound compared to the oligonucleotide in the absence of compound. Variations in T_m, i.e. ΔT_m, are reported in Table 4.1.

Among the identified compounds, D8 and G8 showed the highest stabilizing effect on LTR-IV G4, with an increase of the T_m of 9.0°C and 8.5°C, respectively. G10 compound was instead chose as negative control because it has a chemical structure similar to that of the other identified compounds but displayed no stabilizing effects on LTR-IV G4. The other hits displayed intermediate level of stabilization.

Table 4.1: FRET analysis.

Compound	$\Delta T_m \pm$ SD (°C)
A8	3.0 ± 0.4
B8	4.5 ± 0.3
C8	2.0 ± 0.3
D8	9.0 ± 0.3
E8	0.5 ± 0.2
F8	1.0 ± 0.1
G8	8.5 ± 0.5
H9	2.5 ± 0.2
G10	0.0 ± 0.1

ΔT_m values of HIV-1 LTR -IV G4-folded sequence (0.25 μM) obtained by FRET upon addition of 16-folds amount of hits. SD stands for standard deviation.

Because oligonucleotides used for FRET analysis display fluorophore-modified ends that may influence the measured stabilization, circular dichroism (CD) was employed as an additional technique that uses label-free oligonucleotides. CD spectra of LTR-IV were recorded in the presence and in the absence of hits and T_m were calculated at the peak wavelength corresponding to the parallel G4 conformation (265 nm). CD analysis relative to the representative compound D8 is shown in Figure 4.1 and ΔT_m values of all hits are reported in Table 4.2.

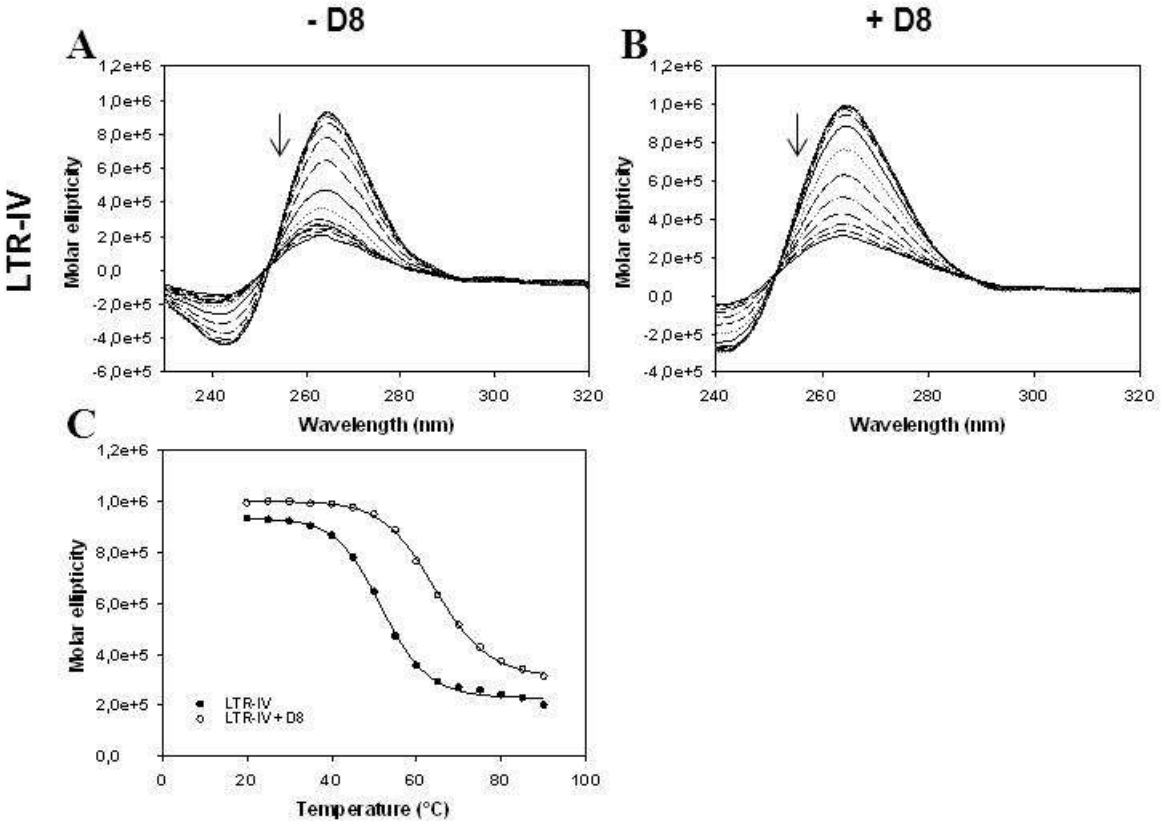


Figure 4.1: CD analysis of LTR-IV G4 oligonucleotide (2 µM) in the absence (panel A) and presence (panel B) of 16-folds excess of compound D8. Panels A and B show CD spectra variation as a function of the wavelength; arrows indicate the spectral change from low to high temperatures. Panel C shows the molar ellipticity at the peak wavelength as a function of the temperature, fitted with the vant' Hoff equation.

As previously observed by FRET, compounds D8 and G8 resulted the most active in stabilizing LTR-IV G4. The other hits displayed intermediate levels of stabilization, while G10 did not show any affect. Therefore, FRET and CD thermal unfolding experiments resulted in accordance. Notably, all compounds with the exception of G10 displayed higher stabilizing effect on the tested oligonucleotides

by CD thermal unfolding than FRET melting analysis, even if the trend of stabilization was similar. This could be due to the fact that oligonucleotides used for FRET display fluorophore-modified ends that may influence the measured stabilization [288].

Table 4.2: CD thermal unfolding analysis.

Compound	$\Delta T_m \pm SD$ (°C)
A8	7.4 ± 0.4
B8	7.2 ± 0.4
C8	8.1 ± 0.5
D8	12.7 ± 0.5
E8	3.5 ± 0.5
F8	5.0 ± 0.4
G8	11.2 ± 0.4
H9	6.4 ± 0.6
G10	0.1 ± 0.1

ΔT_m values of HIV-1 LTR -IV G4-folded sequence (2 μM) obtained by CD upon addition of 16-folds amount of hits. SD stands for standard deviation.

From a structural point of view (Figure 4.2), each identified compound possesses the same chemical core composed of two condensed aromatic rings forming a 1*H*-pyrrolo[2,3-*b*]pyridinic core with a piperidine residue as substituent of the pyridine. This aliphatic moiety appeared to be fundamental for stabilization of LTR-IV G4. Indeed, the absence of the piperidine group in G10 compound seemed to be sufficient to abolish the activity. In addition, the best hits, D8 and G8, were structural isomers and were characterized by the presence of a pyrimidine and a pyrazine ring, respectively, as substituent of the central pyridine; in fact their structures differed only for the position of a nitrogen in the aromatic residue. With respect to the other compounds, the absence of additional substituents to the pyrimidine/pyrazine ring in compounds D8 and G8 appeared to favor LTR-IV G4 stabilization.

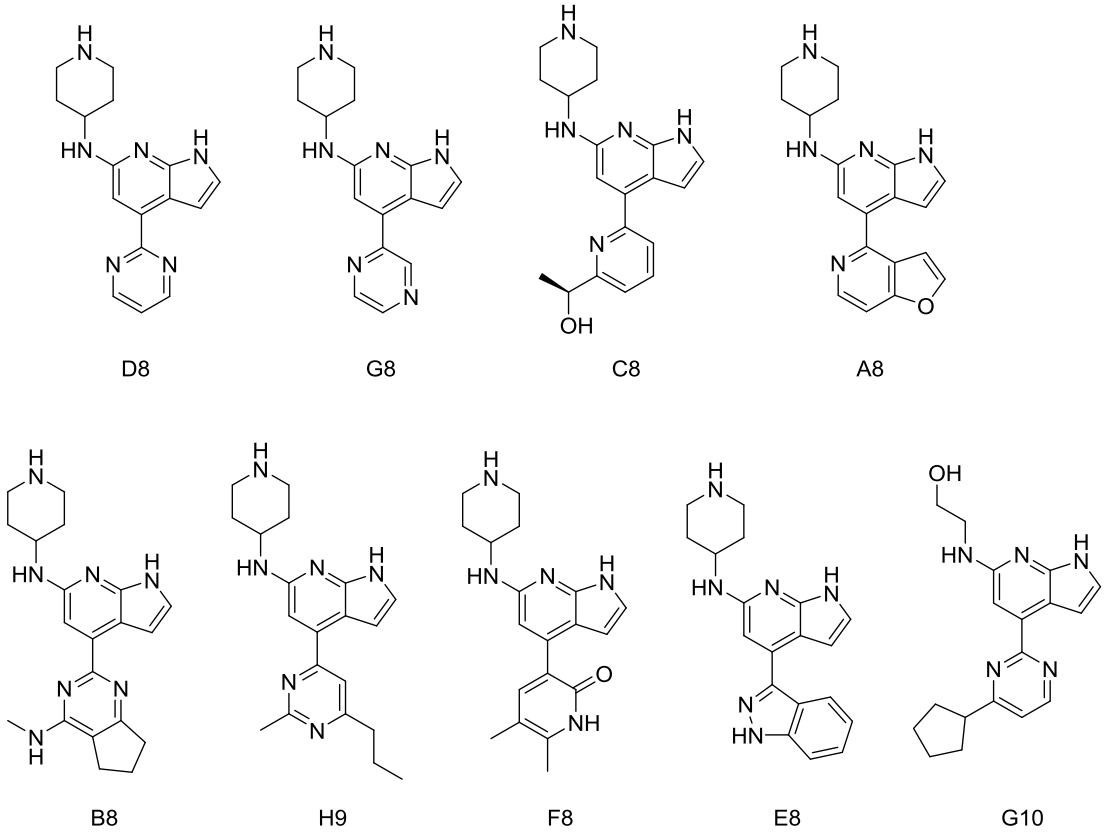


Figure 4.2: Chemical structure of the hits.

4.1.2 Hits stabilize G4s with a preference for the LTR conformations *vs.* the telomeric sequence

To assess the ability of the hits to selectively bind and stabilize HIV-1 LTR G4s, we chose another HIV-1 LTR G4, i.e. LTR-III G4s, which is composed of four G-tracts and is the most stable G4 component within the full-length LTR G4 forming sequence, as demonstrated by spectroscopic analysis and *Taq* polymerase stop assay in our previous study [232]. The minimal intramolecular G4-forming human telomeric sequence (hTel) was also employed to evaluate the activity of the hits on one of the most highly represented and accessible cellular G4: in fact, the TTAGGG repeat is present as 2000–3000 double-stranded (ds) and 50–200 single-stranded sequences at the human telomeres [72], and therefore is representative of cellular G4s. A double-stranded (ds) DNA, which is the main form into which DNA is present in cells, was also added to address specificity toward the G4 conformation. The T_m values of LTR-IV, LTR-III, hTel and dsDNA oligonucleotides were 59.1 ± 0.1 °C, 62.1 ± 0.1 °C, 64.2 ± 0.1 °C and 65.1 ± 0.1 °C respectively, sufficiently similar to allow a meaningful comparison.

First of all, the whole series of compounds was tested by FRET melting assay in order to investigate the degree of stabilization operated by these compounds on viral G4s compared to hTel sequence and dsDNA control. In general FRET melting experiment showed very promising results: the best hits stabilized both LTR G4s while hTel G4 was less efficiently stabilized. Importantly, the stabilization on dsDNA control was negligible for all the tested compounds, indicating the specificity of the hits toward the G4 conformation. More in details, among the tested hits, compounds D8 and G8 were the most efficient ligands with a stabilization on LTR-IV of 9°C and 8.5°C and a stabilization on LTR-III of 4.0°C and 2.5°C, respectively. Their effect on hTel was less significant (2.5°C and 1.5°C, respectively) thus indicating a preferential activity on LTR G4s with respect to the cellular one. The other hits showed intermediate (A8 and B8), or low (C8, E8, F8, H9) level of stabilization of LTR G4s. G10 compound confirmed its consistence as negative control having a non-significant stabilization on each tested oligonucleotide. ΔT_m values are reported in Table 4.3

Table 4.3: FRET analysis.

	$\Delta T_m \pm$ SD (°C)		
Compound	**LTR-III**	**hTel**	**dsDNA**
A8	3.0 ± 0.1	2.0 ± 0.2	0.5 ± 0.2
B8	3.0 ± 0.2	2.9 ± 0.3	0.0 ± 0.1
C8	1.0 ± 0.1	2.0 ± 0.1	0.0 ± 0.1
D8	4.0 ± 0.3	2.5 ± 0.2	0.5 ± 0.1
E8	2.0 ± 0.3	1.5 ± 0.1	1.0 ± 0.2
F8	0.0 ± 0.1	2.0 ± 0.1	0.0 ± 0.1
G8	2.5 ± 0.2	1.5 ± 0.1	0.0 ± 0.1
H9	0.0 ± 0.1	1.5 ± 0.2	0.0 ± 0.1
G10	0.0 ± 0.1	0.5 ± 0.1	0.0 ± 01

ΔT_m values of HIV-1 LTR-III, hTel and dsDNA (0.25 μM) obtained by FRET upon addition of 16-folds amount of hits. SD stands for standard deviation.

As previously reported for LTR-IV G4, we also performed CD thermal unfolding experiments for LTR-III and hTel G4s to confirm the FRET data. CD analysis relative to the representative compound D8 is shown in Figure 4.3. Differently from LTR-IV, that showed a parallel G4 conformation, LTR-III and hTel adopted a mixed conformation with two positive peaks around 265 nm and 290 nm and a

negative one at 240 nm. T_m values were calculated at the peak wavelength corresponding to the prevalent G4 conformation.

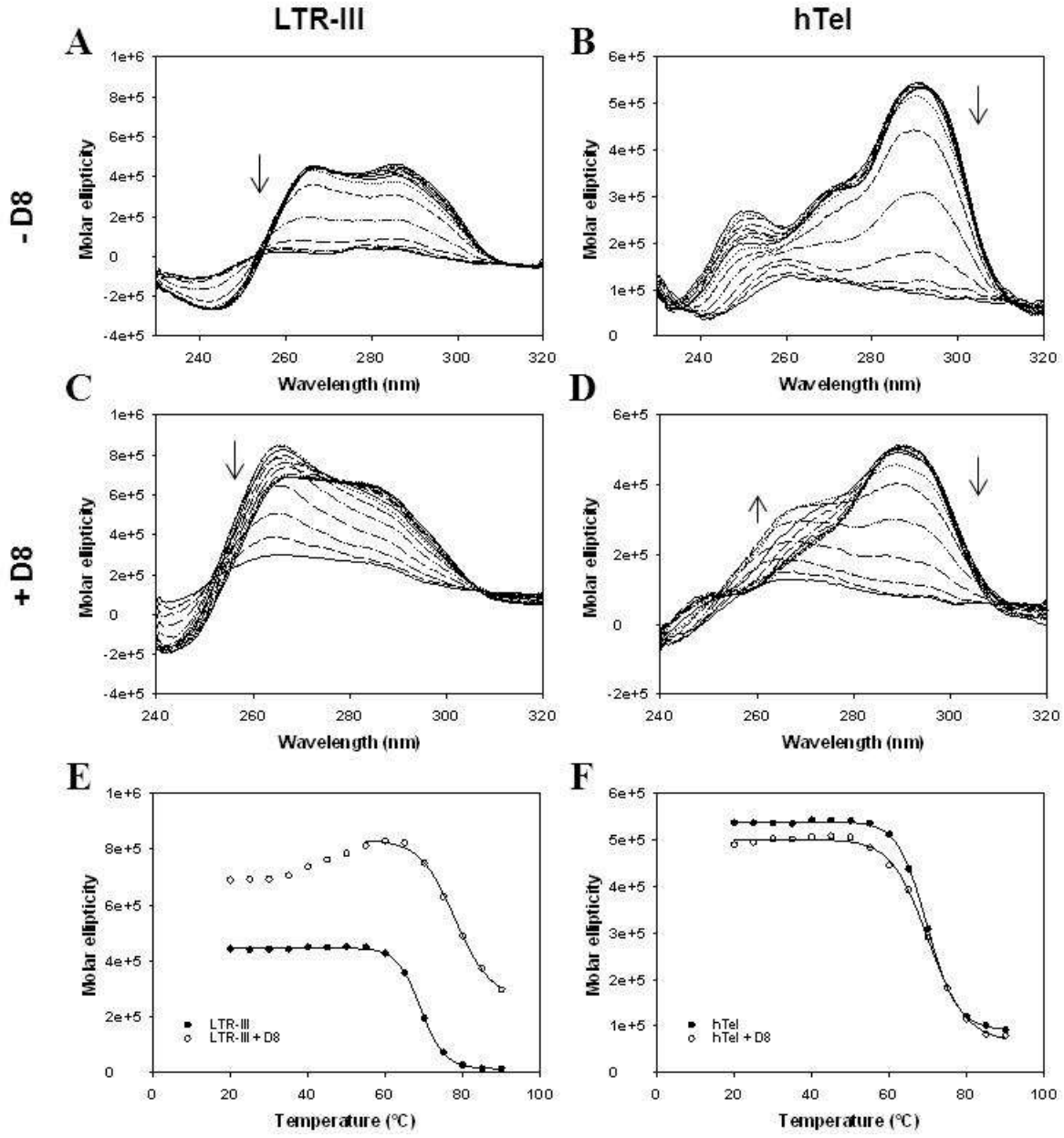


Figure 4.3: CD analysis of LTR-III and hTel G4 oligonucleotides (2 µM) in the absence (panels A and B) and presence (panels C and D) of 16-folds excess of compound D8. Panels A, B, C and D show CD spectra variation as a function of the wavelength; arrows indicate the spectral change from low to high temperatures. Panels E and F show the molar ellipticity at the peak wavelength as a function of the temperature, fitted with the vant' Hoff equation, where possible.

At 16:1 compound:DNA ratio, a stabilization corresponding to 9.0 °C and 7.9 °C was obtain on LTR-III G4 in the presence of compounds D8 and G8 respectively, confirming good stabilizing properties of these compounds. Moreover, both compounds displayed a negligible stabilizing effect on hTel, indicating their selective activity for viral G4s over the cellular G4. The peak at 265 nm, which is diagnostic for the parallel conformation increased in LTR-III sequence upon addition of D8, in contrast, hTel maintained a prevalent antiparallel conformation even if an increment of the peak corresponding to the parallel conformation was observed at temperatures higher than 60°C.

The other compounds displayed good stabilizing properties on LTR-III G4, with the exception of G10 which had no effect on LTR G4s and displayed a weak activity on hTel (ΔT_m= 2.3 °C) (Table 4.4). Notably, compounds C8 and F9 could not be further investigated because they were out of stock by ChemBridge Corporation.

Table 4.4: CD thermal unfolding analysis.

	$\Delta T_m \pm$ SD (°C)	
Compound	**LTR-III**	**hTel**
A8	9.9 ± 1.3	0.1 ± 0.4
B8	10.7 ± 0.8	0.1 ± 0.1
D8	9.0 ± 0.6	0.3 ± 0.2
E8	7.7 ± 0.8	1.4 ± 0.4
G8	7.9 ± 0.8	0.2 ± 0.2
H9	7.8 ± 1.3	-0.8 ± 0.4
G10	0.6 ± 0.3	2.3 ± 0.4

ΔT_m values of HIV-1 LTR-III, hTel and dsDNA (2 μM) obtained by CD upon addition of 16-folds amount of hits. SD stands for standard deviation.

The effect of compound D8, a representative of the best ligands, was next validated by *Taq* polymerase stop assay performed on LTR-III, LTR-IV and hTel templates. All tested oligonucleotides were designed in order to contain additional flanking bases at both the 5'- and 3'-ends; in particular, an additional sequence at their 3'-end was used as primer annealing region and 5-T linker regions were added both at the 3'-end, to separate the primer annealing region and the first G of the G4 portion, and at the 5'-end.

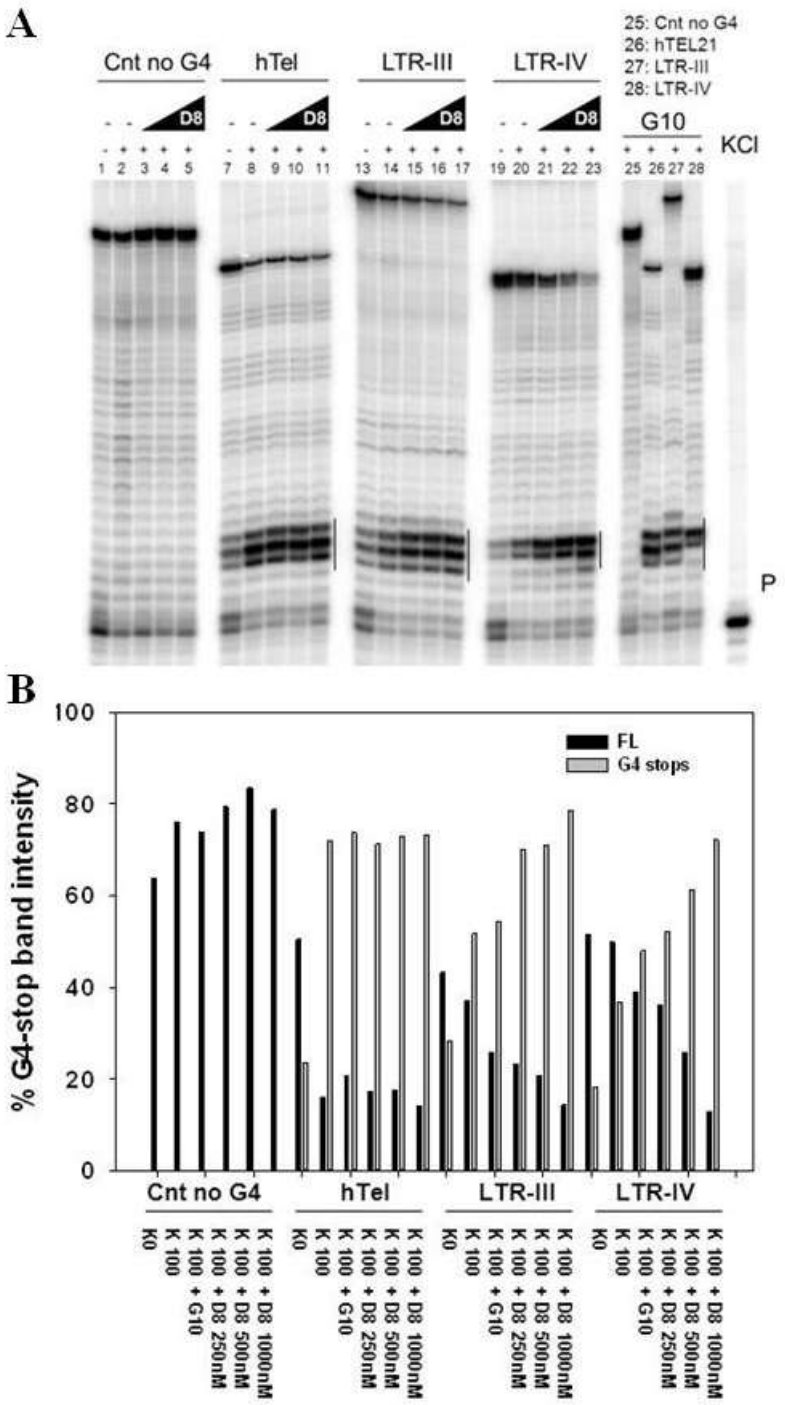


Figure 4.4: *Taq* polymerase stop assay in the presence of D8. A) Htel, LTR-III and LTR-IV templates were amplified by the *Taq* polymerase in the absence and presence of 100 mM K^+, alone or combined with increasing amounts of D8 (0.25-1 µM). G10 (1µM) was tested as negative control. Elongation was performed at 37 °C. A template (cnt no G4) made of a scramble sequence unable to fold into G4 was also used to exclude unspecific stalling of the polymerase. G4 stop regions are highlighted by vertical bars. Lane P: unreacted labeled primer. B) Quantification of stop bands corresponding to formation of hTel, LTR-III and LTR-IV G4s and of the full-length amplification product (FL) is shown.

Specifically, all G4 forming sequences stopped the polymerase at the first 3'-G-rich region encountered by the polymerase. Upon addition of increasing amounts (0.25-1 μM) of D8, the intensity of the stop bands greatly increased in LTR G4 templates (Figure 4.4, panel A, lanes 15-17 and 21-23) accompanied by a considerable reduction of the full-length amplicons, thus corroborating an effective stabilization of G4s by the compound. As indicated in stop bands quantification (Figure 4.4, panel B), this effect was significantly more relevant on the viral sequences *vs.* the telomeric template, confirming again a preferential stabilization of the viral G4s. Notably, G10 resulted a proper negative control since no significant stabilization was observed in all tested oligonucleotide (Figure 4.4, panel A, lanes 25-28). Moreover, a control sequence devoid of G-tracts did not show any stop site in the presence of K^+ and D8 (Figure 4.4, panel A, lanes 1-5) indicating that the observed polymerase inhibition was G4-dependent and specific.

The above techniques evaluated stabilization of the G4 conformation as an indicator of the efficiency of the interaction of the ligands with the tested G4s. Binding affinity of the best hits was next directly assessed by SPR. Affinity of the best hits, D8 and G8, was measured on the three G4-folded sequences, LTR-IV, LTR-III and hTel. Biotinylated G4-folded oligonucleotides were immobilized on a streptavidin coated SPR sensor chip and K_D values, determined from sensorgram local fitting, were calculated. This analysis corroborated that both D8 and G8 possess higher affinity for LTR G4s with respect to hTel G4. In particular, in the presence of D8, LTR-III displayed the highest affinity (K_D = 128.9 ± 31.1nM) (Figure 4.5, panel D), followed by LTR-IV G4 (K_D = 254.1 ± 53.1nM) (Figure 4.5, panel A). The same trend obtained for D8 was also observed in the presence of G8 even though K_D values were slightly higher (LTR-III K_D = 138.00 ± 58.60 nM; LTR-IV K_D = 592.40 ± 278.00 nM) (Figure 4.5, panels B and E), indicating a slightly less affinity of this compound for LTR G4s. The K_D values of hTel were not determined for both D8 and G8 compounds due to too low affinity interaction (Figure 4.5, panels G and H), thus confirming the preferential interaction of the hits for viral over the cellular G4s. Binding of the negative control G10 to all tested oligonucleotides was almost undetectable and it was not possible to obtain a meaningful K_D values (Figure 4.5, panels C, F and I).

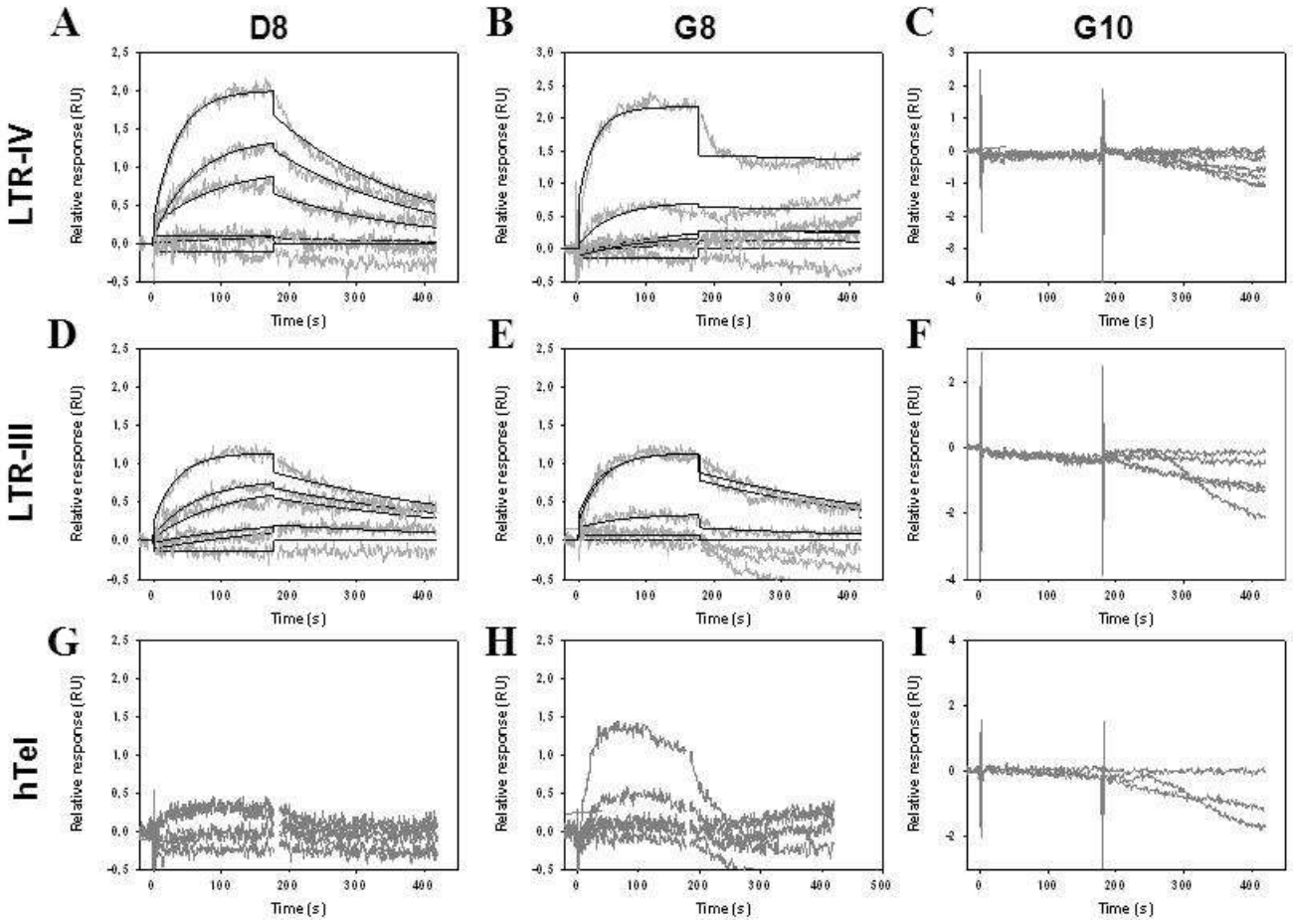


Figure 4.5: Binding affinity of D8, G8 and G10 for the tested oligonucleotides by SPR analysis. Sensograms were obtained in the compounds concentration range of 0.039–1.25 μM. SPR sensograms of A) LTR-IV + D8 with ka $1.87\times 10^{4} \pm 0.08 \times 10^{4}\ Ms^{-1}$ and kd $4.74\times 10^{-3} \pm 0.08 \times 10^{-3}\ s^{-1}$; B) LTR-IV + G8 with ka $3.28\times 10^{4} \pm 0.18 \times 10^{4}\ Ms^{-1}$ and kd $0.19\times 10^{-3} \pm 0.08 \times 10^{-3}\ s^{-1}$; C) LTR-IV + G10; D) LTR-III + D8 with ka $2.08\times 10^{4} \pm 0.36 \times 10^{4}\ Ms^{-1}$ and kd $2.68 \times 10^{-3} \pm 0.08 \times 10^{-3}\ s^{-1}$; E) LTR-III + G8 with ka $2.08 \times 10^{4} \pm 0.36 \times 10^{4}\ Ms^{-1}$ and kd $2.87 \times 10^{-3} \pm 0.25 \times 10^{-3}\ s^{-1}$; F) LTR-III + G10; G) hTel + D8; H) hTel + G8; I) hTel+ G10. Where not indicated, the ka and kd values were not reliably measurable. Sensograms are shown as grey lines and their respective fits as black lines.

4.1.3 The best hits display anti-HIV-1 activity through a G4-mediated mechanism of action

We have previously shown that the HIV-1 viral promoter folds into dynamic G4s [232] and that stabilization of these G4 conformations by the G4 ligand BRACO-19 represses viral transcription and virus production in infected cells [233]. Given the hits binding/stabilization selectivity observed in vitro for the viral G4s *vs.* the cellular G4, we set out to investigate the compounds for their antiviral properties.

Antiviral assays were performed on TZM-bl cells, in which HIV infection drives transcription of an HIV-1 LTR-luciferase reporter gene construct. We treated infected cells with increasing amounts of compounds and measured virus production in the supernatant by quantifying the luciferase signal at 48 hours post-infection (h.p.i.). All compounds were tested against the HIV-1 NL4-3 strain, which is a X4 strain. Concurrently, the cytotoxicity of the hits in the same conditions but on uninfected cells was measured by MTT assay. Compounds B8, D8 and G8 markedly reduced virus production with IC_{50} values in the μM range. CC_{50} values were higher than the highest tested concentration (250 μM) and therefore the calculated selectivity indexes (SI) were generally promising, up to 7.4. On the contrary, A8, E8, H9 and G10 did not show any antiviral effect. Interestingly, the hits displaying anti-HIV-1 activity also showed the highest and more selective stabilizing effect on LTR G4s, thus suggesting a correlation between antiviral activity and LTR G4s stabilization. CC_{50}, IC_{50} and S.I. values are indicated in Table 4.5. A representative antiviral assay is shown in Figure 4.6.

Table 4.5: Anti-HIV-1 activity and cytotoxicity of hits.

Compound	CC_{50} ± SD (μM)	IC_{50} ± SD (μM)	SI
A8	35.9 ± 1.7	72.1 ± 3.7	0.5
B8	> 250.0	34.5 ± 3.0	> 7.4
D8	> 250.0	40.8 ± 2.9	> 6.2
E8	30.3 ± 1.8	47.3 ± 1.9	0.6
G8	> 250.0	42.4 ± 3.3	> 6.0
H9	> 250.0	68.0 ± 7.8	3.7
G10	220.0 ± 5.7	165.5 ± 9.2	1.3

IC_{50} is the compound concentration required to inhibit 50% of HIV-1 production; CC_{50} is the compound concentration at which 50% of cell toxicity is observed; SI is the selectivity index calculated as the ratio of CC_{50} / IC_{50}. SD stands for standard deviation.

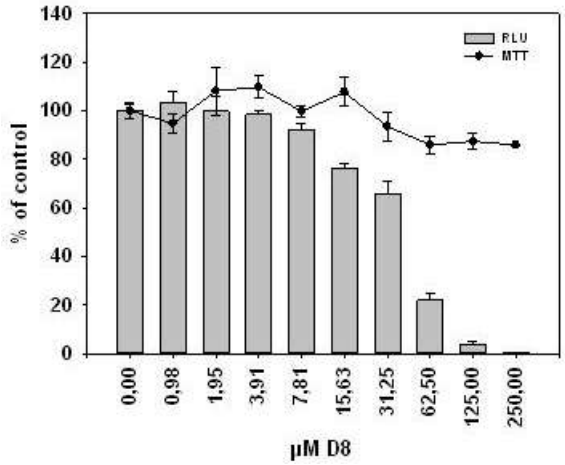


Figure 4.6: Antiviral activity of compound D8. Antiviral activity of compound D8 on infected TZM-bl cells is expressed as relative luciferase units (RLU), while cell viability was in parallel obtained with a MTT assay on uninfected cells treated in the same conditions. Histograms represent the antiviral activity of the tested compound while the line indicates the cytotoxic effect.

In order to evaluate if the antiviral activity of the best hit, compound D8, was related to a direct effect on HIV-1 LTR G4s, two point mutations that abolish G4 folding were introduced in the LTR sequence (M4+5 LTR) [232].

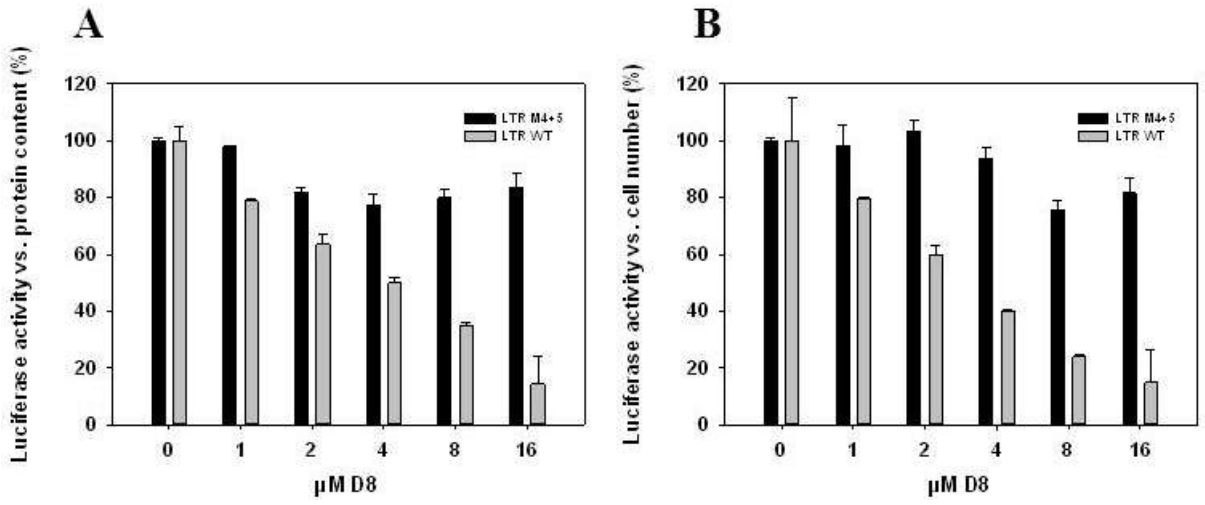


Figure 4.7: Effect of D8 on the HIV-1 LTR promoter in HEK 293T cells. Analysis of the luciferase activity of the WT and M4 + 5 LTR promoters in HEK 293T cells treated with increasing amounts of D8, normalized to A) protein content and B) cell number.

The wt LTR and M4+5 LTR luciferase reporter plasmids were transiently transfected into HEK 293T cells either alone or in the presence of increasing amounts (1-16 μM) of compound D8, and a luciferase reporter assay was performed. Luciferase signals were normalized to total protein content (Figure 4.7, panel A) and to cell number (Figure 4.7, panel B). Wt LTR promoter activity decreased in a concentration dependent manner up to ~70% with respect to the untreated control, whereas no effect was detected on the non-G4-forming mutant LTR sequence, supporting a G4 mediated inhibition, in both normalized experimental conditions. Cytotoxicity of compound D8 was previously evaluated in HEK 293T cell line, and no effect was observed up to the highest tested concentration (200 μM).

4.1.4 Mass spectrometry binding analysis

Native mass spectrometry was employed to investigate the binding of hits to the LTR-IV G4. Surprisingly, when LTR-IV was mixed with D8 at ratio 1:10, no adducts having a mass shift consistent with D8 molecular weight were detected. The spectrum in Figure 4.8 showed the formation of two adducts having a mass shift of 890.5 Da and 1156.7 Da, respectively.

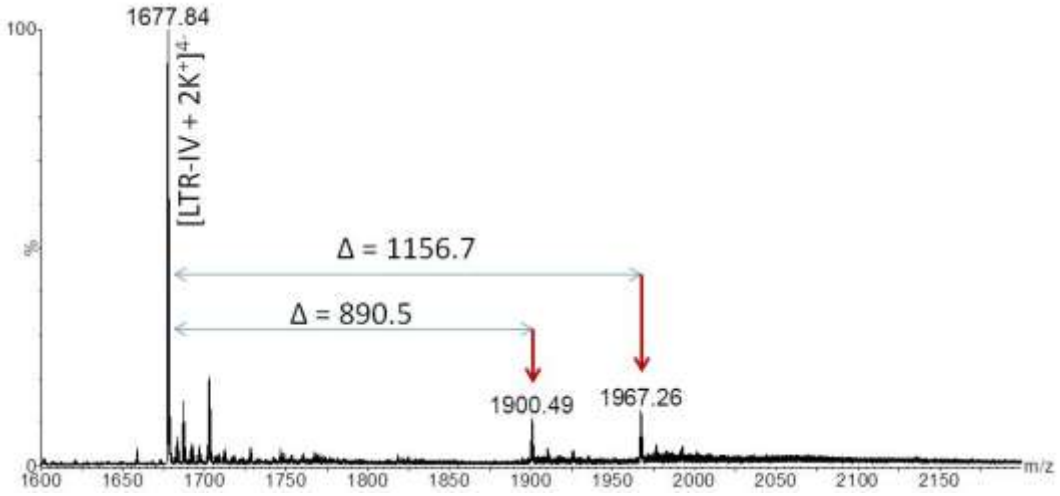


Figure 4.8: Mass spectrum of LTR-IV G4 in the presence of D8. LTR-IV was mixed with D8 at ratio 1:10.

Because the detected mass shifts were not consistent with the binding of one or more D8 molecules (Table 4.6), we suspected that the active compound was a D8 polymer that rearranged in solution. Mass spectrometry analysis of the D8 solution in the absence of the oligonucleotide next revealed that it was a complex mixture of species and the component having a molecular weight corresponding to D8

(294.2 Da) did not represent the main one (Figure 4.9). Intense peaks corresponding to compounds having neutral mass of 890.5 Da and 1156.7 Da were also detected (Figure 4.9).

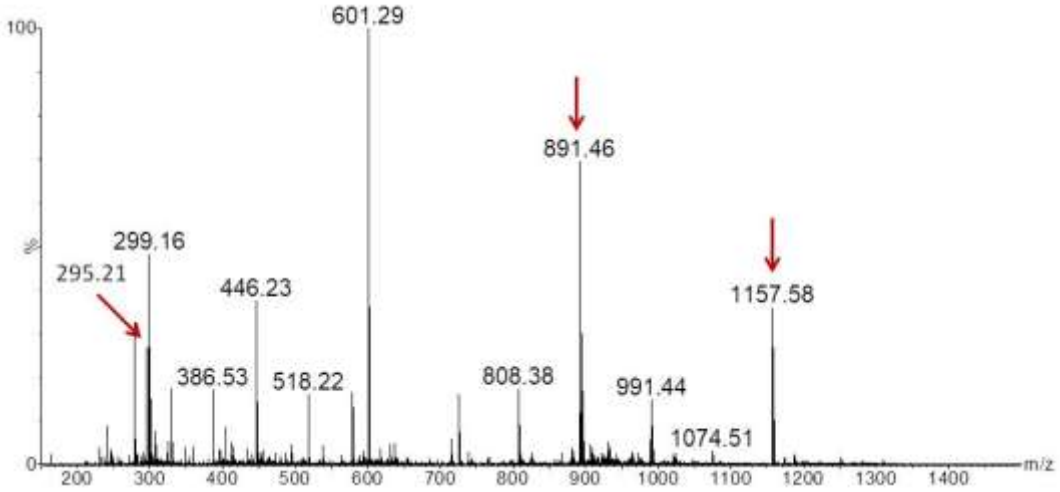


Figure 4.9: Mass spectrum of 50 µM of D8 solution in 50:50 acetonitrile/water. Species having molecular weight corresponding to 294.2 Da, 890.5 Da and 1156.7 Da were identified.

Structural analysis by MS/MS of species having a molecular weight of 890.5 Da and 1156.7 Da revealed that they contained some portions of D8 (i.e. multiple piperidinic rings) (Figures 4.10 and 4.11).

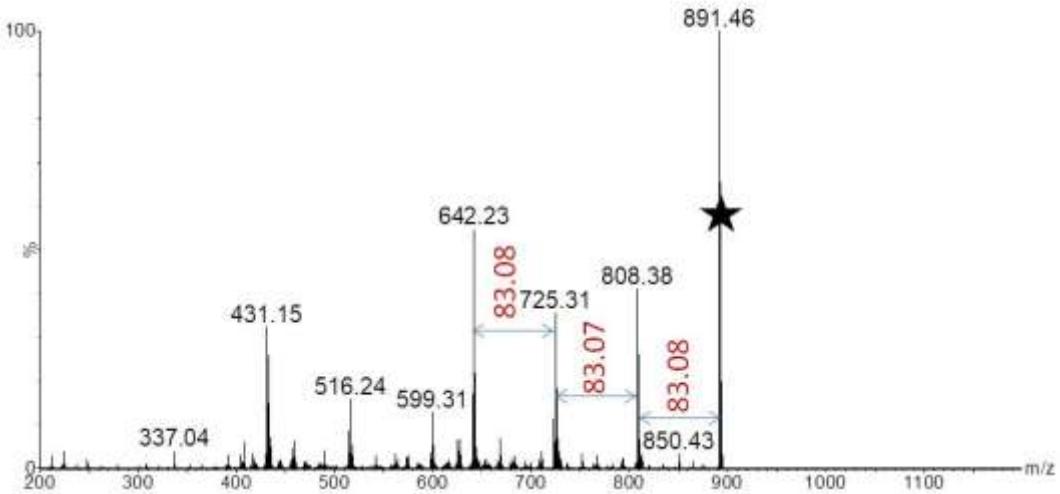


Figure 4.10: MS/MS spectra of 890.5 Da (M^+ = 891.46) molecular weight species. Fragments generated by the lost of piperidinic ring (mass shift = 83.07 Da) were detected.

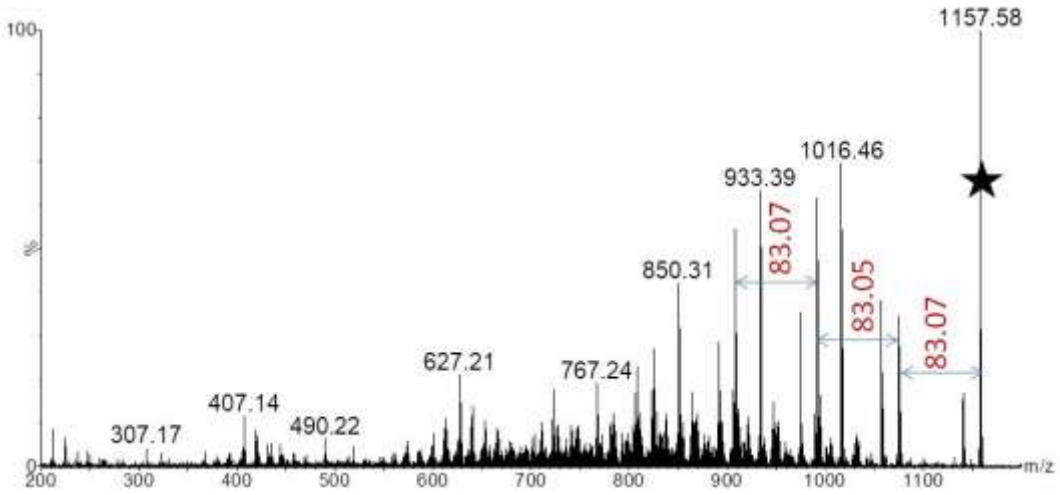


Figure 4.11: MS/MS spectra of 1156.7 Da (M^+ = 1157.58) molecular weight species. Fragments generated by the loss of piperidinic ring (mass shift = 83.07 Da) were detected.

Based on these data, we suspected that such compounds were formed by a polymerization process of D8. This assumption was corroborated by the evidence that also additional hits were able to form polymerization products that were subjected to the same rearrangement pathway (Table 4.6).

Table 4.6: Mass spectrometry binding analysis of hits to LTR-IV G4

Compound	MW (Da)	Δ1 (Da)	Match	BA (%)	Δ1 (Da)	Match	BA (%)
A8	333.159	1007.6	3A8 + 8.009	5.9	1312.7	4A8 – 19.896	10.6
B8	363.217	1097.7	3B8 + 8.065	3.3	/	/	/
C8	337.190	1019.7	3C8 + 8.098	9.0	/	/	/
D8	294.159	890.5	3D8 + 8.055	11.7	1156.7	4D8 – 19.940	14.0
E8	332.175	/	/	/	1308.9	4E8 – 19.844	8.4
F8	337.190	1014.8	3F8 + 3.243	8.1	/	/	/
G8	294.159	890.6	3G8 + 8.127	11.7	1156.8	4G8 – 19.868	11.6
G10	350.222	323.3	1G10	25.4	646.5	2G10	9.4
H9	350.222	1058.8	3H9 + 8.784	12.3	/	/	/

MW stands for the molecular weight of compound calculated by raw formula. Δ stands for the mass shift observed in G4-compound adducts. Match represents the number of molecules bound to G4. BA is the ratio between bound and unbound G4 expressed in percentage.

These polymerization products were also detected when analyzed by mass spectrometry in the absence of the oligonucleotide; additionally MS/MS structural investigation revealed the same fragmentation pathway (data not shown). The library supplier (Chembridge Corporation) also confirmed that this class of compounds underwent degradation and, for this reason, they are currently out of production. Our collaborators, Prof. Mauro Freccero's group (University of Pavia, Italy), are now re-synthesizing the positive hits in order to produce, purify and characterize the active fractions.

4.1.5 Discussion

We were able to identify a new class of G4 ligands through a screening approach. In particular by FRET melting analysis we found a family of small molecules composed of the same chemical core, showing selective stabilization on LTR G4s with respect to the telomeric G4. The preferential stabilization was also confirmed by CD thermal unfolding, *Taq* polymerase stop assay along with SPR sensorgrams that corroborated the preferential affinity. Interestingly, the best LTR G4-stabilizing compounds displayed a modest but promising antiviral activity in HIV-1 infected TZM-bl cells, which was ascribed to a G4-mediated mechanism of action by luciferase reporter assay. On one hand, our findings open the possibility to develop novel anti-HIV compounds with a new mechanism of action that may complement current clinical AIDS therapies. In this context, it is worth nothing that G4s within the LTR are localized in an extremely conserved region: the degree of base conservation among HIV-1 strains have previously been assessed and 953 LTR sequences have been analyzed revealing that most Gs and also non-G bases are highly conserved [232]. This represents an advantage to avoid drug resistance that normally arises from the extreme variability and the high evolution rate of the virus and is the major cause of treatment failure. Moreover, because the LTR promoter has been suggested to be the region where viral latency is regulated [289,290], the G4 switch may play a role not only in activation of effective viral transcription, but also in its shift to latency. Currently, only the actively transcribed virus is targeted by antiviral drugs, while eradication of the HIV-1 infection has been made impossible by the existence of reservoirs of the latent virus [291]. Therefore, our findings not only advance our understanding on the mechanism of viral transcription but may also constitute a progress from a therapeutic point of view. On the other hand, our work clearly highlights the possibility to selectively recognize G4 structures with small molecules, prompting the search and development of G4 ligands specific for G4s implicated in several important human diseases. To this end and basing on our promising results, it is fundamental to elucidate the structure of active fractions of hits in order to better characterize their anti-HIV-1 activity.

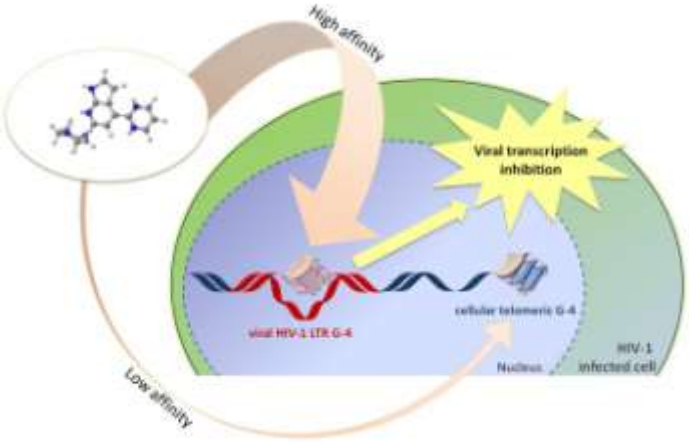


Figure 4.12: Schematic representation of hits as novel antiviral compounds.

4.2 NDI-PNA conjugates targeting HIV-1 LTR G4s

In order to selectively target LTR G4s, we proposed a conceptually new approach, combining the stabilizing activity of a G4 ligand with fingerprint recognition of the nucleotide sequence proximal to the G4 of interest. We chose a tetrasubstituted naphthalene diimide (NDI), a well-characterized G4 ligand with good G4 binding properties, functionalized for conjugation. It has been reported to effectively stabilize the terminal G-quartet of a G4 by stacking interactions on the external tetrads [292] and thus, because it targets a common feature of G4s, it displays insufficient selectivity towards different G4s. For the fingerprint recognition we selected peptide nucleic acid (PNA), a synthetic analogue of nucleic acids equipped with a peptide backbone. This makes interaction with complementary strands more stable, due to lack of charge repulsion, and enables PNA to escape DNAse and protease degradation [236].

Our collaborators, Prof. Mauro Freccero's group (University of Pavia, Italy), have synthesized conjugates functionalized with PNA complementary to the flanking region of a selected LTR G4, so as to force the system to fold into the structure of choice. As starting point, we decided to selectively target LTR-IV G4. Therefore, the conjugate was projected to stabilize LTR-IV G4 through the NDI, and, at the same time, to prevent the formation of the most stable G4 (LTR-III) by a PNA specifically designed to bind the 5'-flanking region of LTR-IV G4. We chose to use 13 membered PNA to balance

sequence selectivity and hydrophobicity, and lysines were attached at both ends of the sequence to improve water-solubility. Chemical structure of the NDI-PNA conjugate is reported in Figure 4.13.

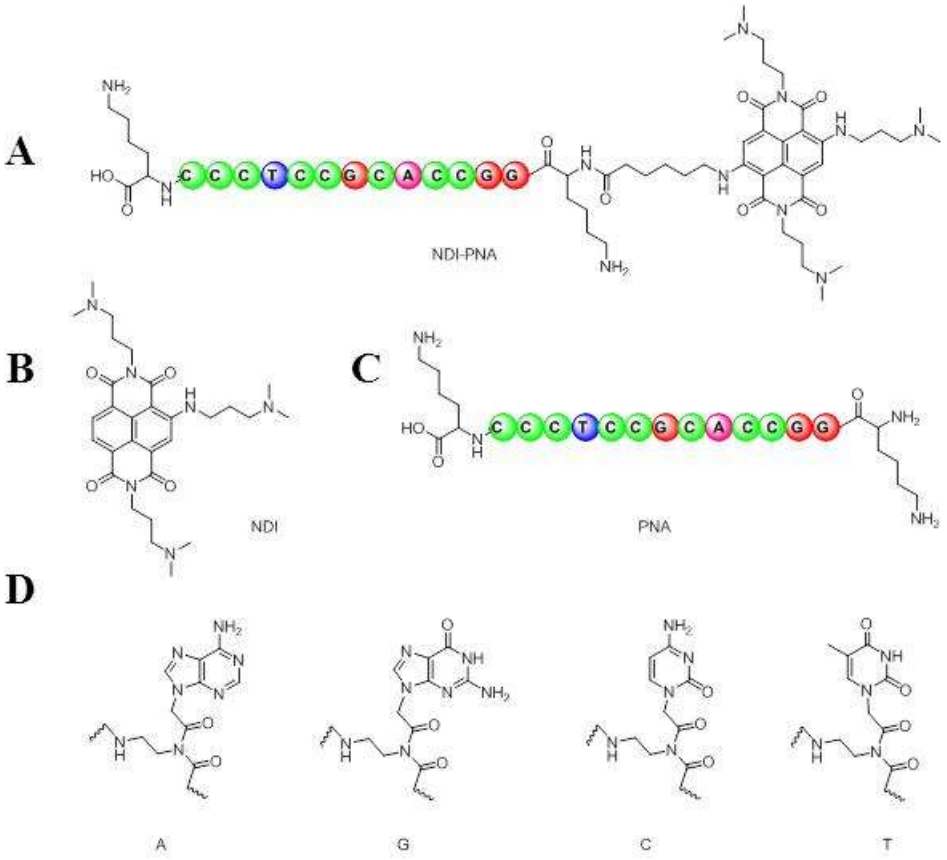


Figure 4.13: Chemical structure of A) NDI-PNA; B) NDI; C) PNA; D) nucleobases bounded to pseudo-peptide backbone.

4.2.1 NDI-PNA selectively targets LTR-IV G4

FRET melting experiments were initially performed to study the interaction between the NDI-PNA and its target sequence, i.e. LTR-III+IV. NDI and PNA were also investigated as controls to evaluate their single contribution. Moreover, dsDNA, which is the main form into which DNA is present in cells, was tested to address specificity toward the G4 conformation. ΔT_m were calculated as previously indicated.

In the presence of both PNA and NDI-PNA, the fluorescence of LTR-III+IV at 30°C was about 4-folds higher than the control (Figure 4.14, panel A): this was be ascribed to the fact that the PNA sequence was complementary to the 5'-end of the labeled oligonucleotide and, consequently, the binding of the PNA to the oligonucleotide was accompanied by an increase in the distance between the donor and the acceptor, thus leading to a less efficient energy transfer that resulted in an higher detected emission

from the donor. This behaviour was considered an indirect evidence of the binding of the PNA to its complementary DNA sequence. Because these fluorescence curves did not have a sigmoidal shape, they could not be normalized and thus ΔT_m values could not be calculated. As expected, NDI greatly stabilized LTR-III+IV with a $\Delta T_m = 21.5 \pm 0.1$ °C (Figure 4.14, panel A). A mild level of stabilization ($\Delta T_m = 3.4 \pm 0.5$ °C) was also observed on dsDNA, suggesting that NDI could interact with dsDNA but to a lesser extent with respect to G4s (Figure 4.14, panel B). Interestingly, both NDI-PNA and PNA did not stabilize dsDNA (ΔT_m values were respectively 0.6 ± 0.5 °C and 0.3 ± 0.3 °C), supporting a sequence-specific mechanism of binding. Figure 4.14, panel B).

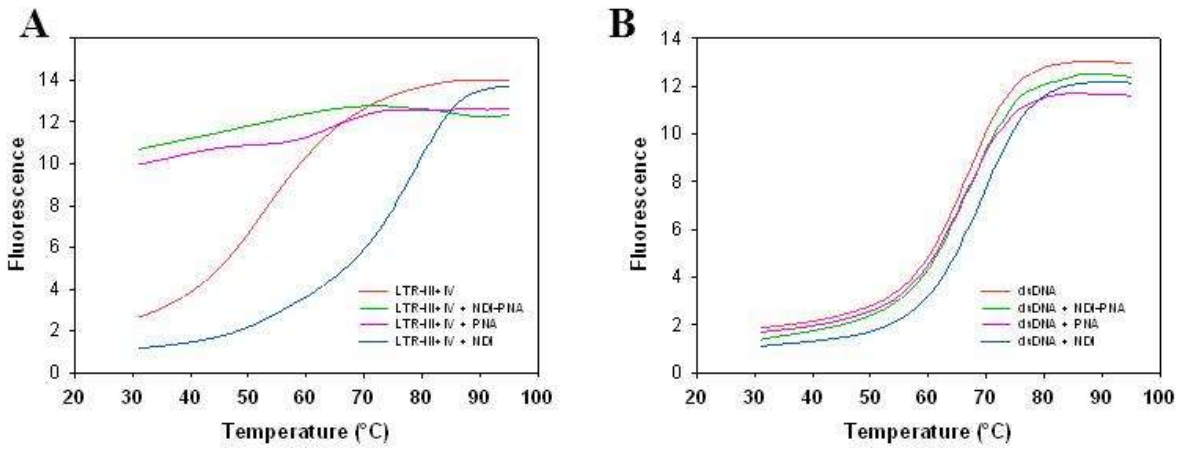


Figure 4.14: FRET analysis of the NDI-PNA. FRET-melting curves of A) LTR-III+IV and B) dsDNA oligonucleotides (0.25 μM) in the absence or presence of 4-fold excess of NDI-PNA, NDI and PNA.

In order to evaluate G4 topology of LTR-III+IV in the presence of NDI-PNA, CD spectra were next recorded at 20°C. Htel was also analysed as representative of cellular G4s. At 100 mM K^+, LTR-III+IV G4 showed a mixed conformation with two positive peaks at 290 nm and 265 nm and a negative one at 240 nm (Figure 4.15, panel A). No significant conformational changes were observed upon addition of NDI (Figure 4.15, panel A). On the contrary, the presence of PNA led to a prevalent parallel structure with a positive peak at 265 nm and a negative peak at 240 nm while the NDI-PNA clearly determined a parallel G4 characterized by a higher molar ellipticity. These data are consisted with the fact that the PNA sequence was specifically designed to target the flanking region of LTR-IV: PNA binding to its complementary DNA region prevented the folding of LTR-III because three runs of Gs were involved in the interaction with the PNA. Therefore a parallel G4, i.e. LTR-IV [232,287], was expected. Htel G4 showed a hybrid "3+1" structure that was not significantly modified by PNA and NDI-PNA, while

NDI induced a prevalent anti-parallel conformation with a positive peak at 290 nm (Figure 4.15, panel B).

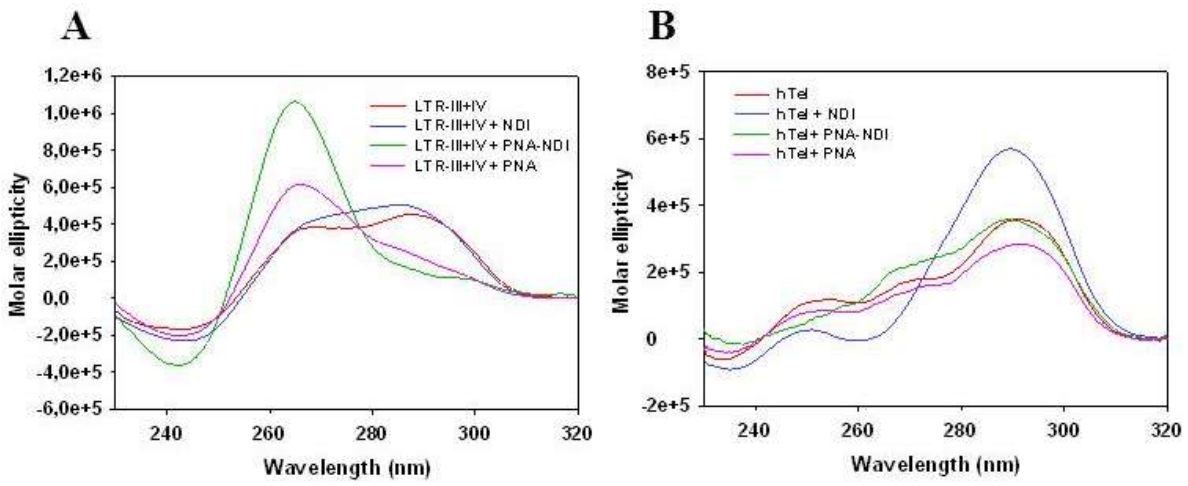


Figure 4.15: Representative CD spectra of G4s formed by A) LTR-III+IV and B) hTel oligonucleotides (2 µM). Spectra were recorded in the absence or presence of 4-fold excess of NDI-PNA, NDI and PNA.

CD thermal unfolding experiments were next performed to better investigate the activity of NDI-PNA, PNA and NDI on LTR-III+IV and hTel G4s, as in terms of thermal stability as well as structural conformational changes. The T_m of LTR-III+IV was calculated at the wavelengths corresponding to the parallel conformation and the anti-parallel one, and were respectively 64.9 ± 0.5 °C and 64.6 ± 0.1 °C (Figure 4.16, panel A). NDI stabilized LTR-III+IV anti-parallel conformation with a ΔT_m = 9.8 ± 1.3 °C while the parallel conformation was stabilized only at temperature higher than 65°C (Figure 4.16, panel D). The presence of PNA led to a prevalent parallel structure formed already at 30°C but no significant effects were detected in terms of thermal stability (ΔT_m = 0.5 ± 0.5 °C) (Figure 4.16, panel F). Interestingly, NDI-PNA greatly stabilized the parallel conformation of LTR-III+IV with a T_m > 90°C and ΔT_m > 25°C (which corresponds to the maximum calculable variation in these settings) (Figure 4.16, panel B). HTel G4 displayed a T_m = 68.7 ± 0.1 °C (Figure 4.17, panel A) and NDI mainly stabilized its anti-parallel conformation with a ΔT_m = 7.3 ± 0.8 °C even if a mild stabilization was detected on the parallel conformation at temperatures higher than 70°C (Figure 4.17, panel D). No significant variations in the structural conformation and in the thermal stability of hTel were detected upon addition of PNA (ΔT_m = 1.2 ± 0.4 °C) (Figure 4.17, panel F). On the contrary, NDI-PNA showed a destabilizing activity on the anti-parallel conformation (ΔT_m = -8.40 ± 1.2 °C) while a mild stabilizing effect was observed on the parallel conformation (Figure 4.17, panel B).

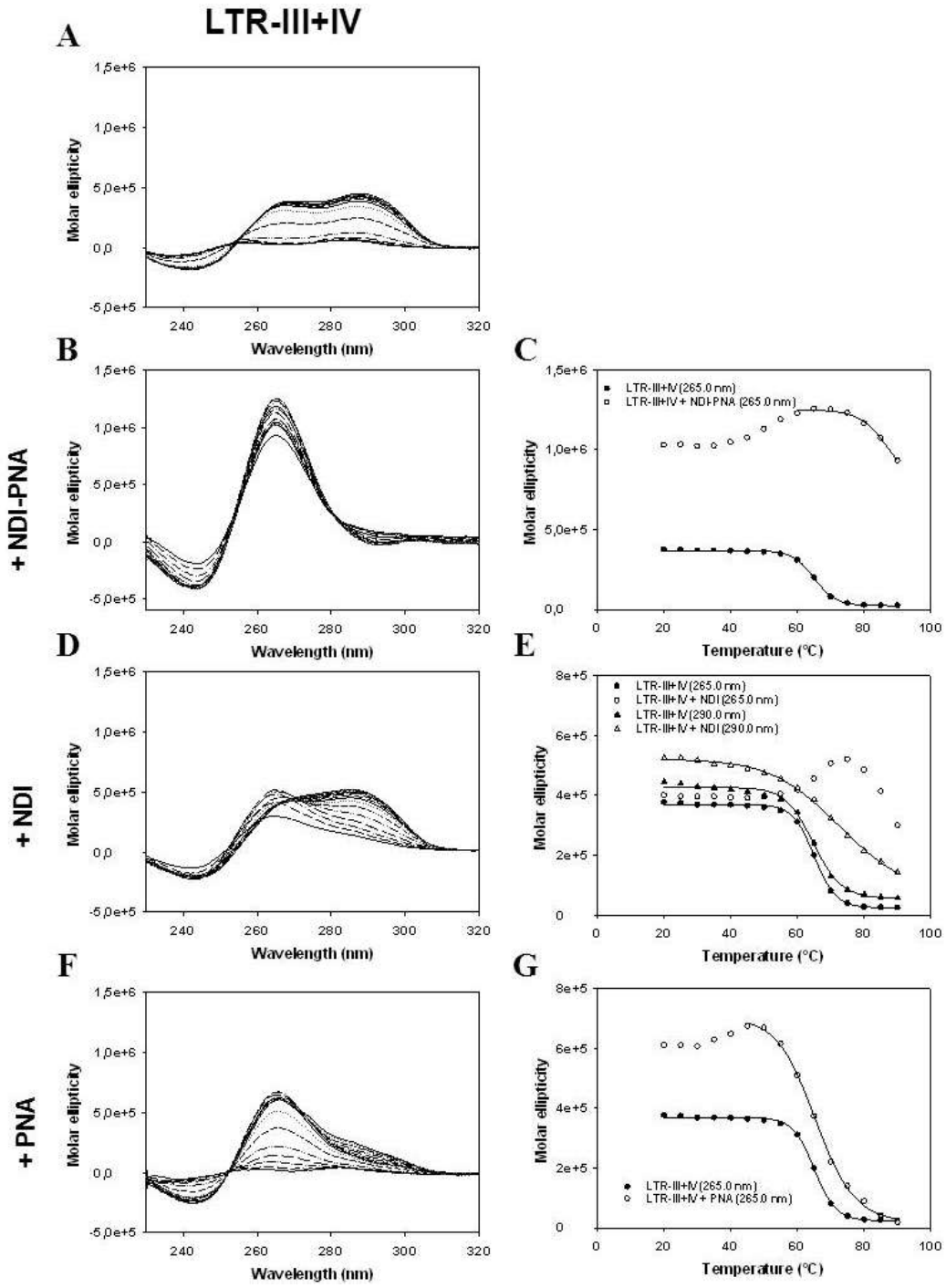


Figure 4.16: CD analysis of LTR-III+IV (2 µM) alone (A) or in the presence of 4-fold excess of B) NDI-PNA, D) NDI, F) PNA. Panels A, B, D, F show CD spectra variation as a function of the wavelength; arrows indicate the spectral change from low to high temperatures. Panels C, E, G show the molar ellipticity at the peak wavelengths as a function of the temperature, fitted with the vant' Hoff equation, where possible. Circles and triangles indicate the parallel (265 nm) and antiparallel (290 nm) G4 conformation contribution, respectively, in the absence (black) and presence (white) of conjugate.

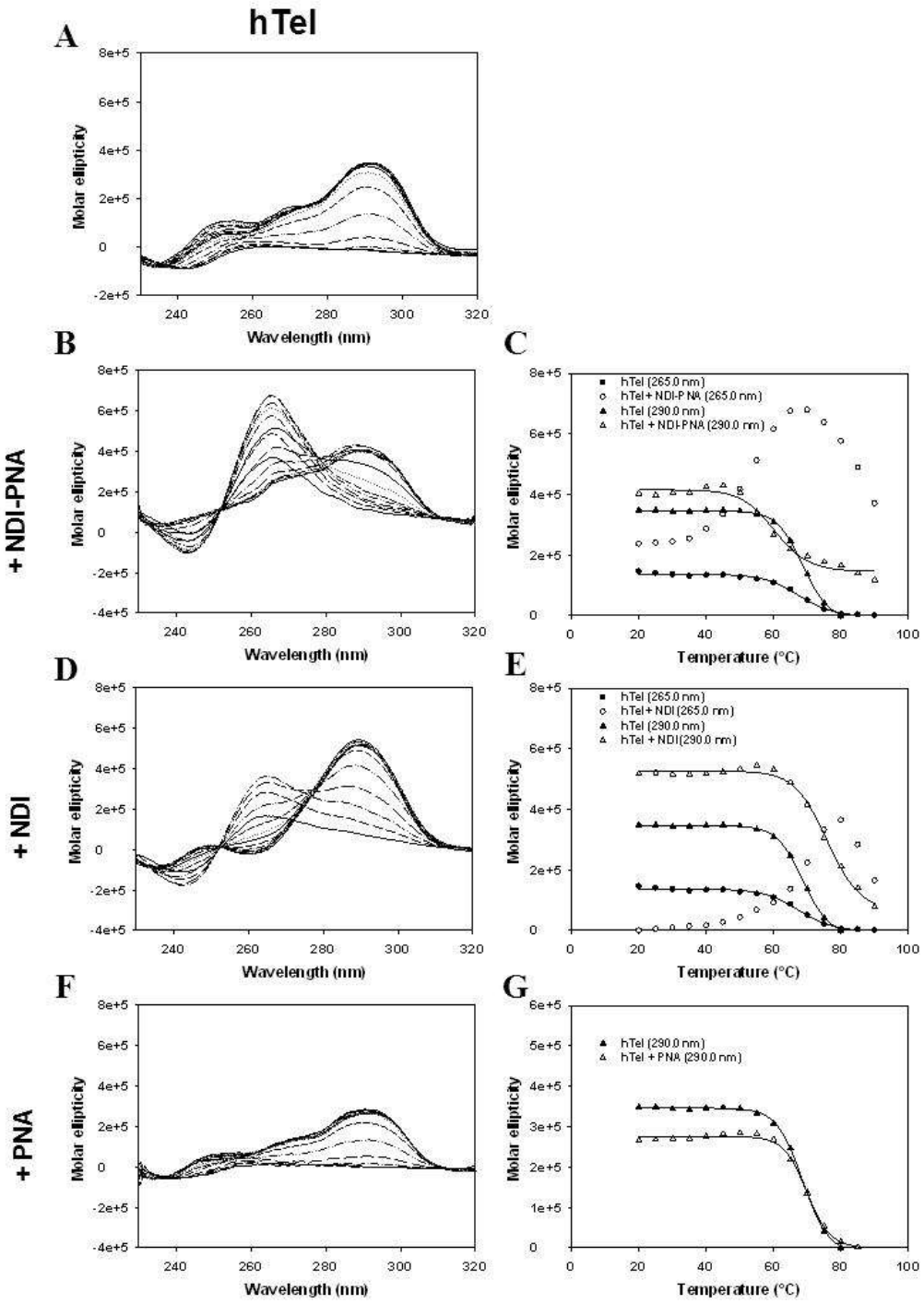


Figure 4.17: CD analysis of hTel (2 µM) alone (A) or in the presence of 4-fold excess of B) NDI-PNA, D) NDI, F) PNA. Panels A, B, D, F show CD spectra variation as a function of the wavelength; arrows indicate the spectral change from low to high temperatures. Panels C, E, G show the molar ellipticity at the peak wavelengths as a function of the temperature, fitted with the vant' Hoff equation, where possible. Circles and triangles indicate the parallel (265 nm) and antiparallel (290 nm) G4 conformation contribution, respectively, in the absence (black) and presence (white) of conjugate.

CD analysis revealed that NDI-PNA induced an extremely stable parallel G4 structure on LTR-III+IV template, which was expected to be LTR-IV. We set out to validate it by *Taq* polymerase stop assay.

The assay was performed on the LTR-III+IV and hTel templates; a control template unable to fold into G4 was used to exclude unspecific inhibition of the polymerase enzyme by the conjugate. Samples were incubated in the absence/presence of 100 mM K^+ and increasing amounts of NDI-PNA (100-400 nM). PNA and NDI were only tested at the highest concentration (400 nM).

As shown in Figure 4.18, LTR-III formed the most stable G4 when embedded in the LTR-III+IV (Figure 4.18, panel A, lane 21) and it was further greatly stabilized by NDI (Figure 4.18, panel A, lane 26). PNA induced the formation of an intense stop site corresponding to the LTR-IV, while a decrease in the LTR-III stop site was observed (Figure 4.18, panel A, lane 25). This data was in accordance to the fact that the PNA binding to its DNA complementary region, prevented the formation of LTR-III and therefore, a block of the *Taq* Polymerase was detectable upstream. Upon addition of increasing amounts of NDI-PNA, a decrease in the stop sites corresponding to the LTR-III G4 was detected while a great concentration dependent increase was observed at the level of LTR-IV G4 (Figure 4.18, panel A, lanes 22-24). Comparing PNA and NDI-PNA (Figure 4.18, panel A, compare lanes 24 and 25), the latter induced a more intense stop, as determined by stop bands quantification (Figure 4.18, panel B). Moreover, G4 stop induced by NDI-PNA was interestingly higher with respect to only NDI (Figure 4.18, panel A, compare lanes 24 and 26).

Regarding the hTel G4, CD data showed that NDI-PNA destabilized its anti-parallel conformation while stabilized the parallel one. However, on the whole, it was not possible to understand if its effect on hTel G4 was lower than NDI. To answer this question, *Taq* polymerase stop assay was also performed on the hTel template. NDI induced a great intense stop at the 3'-G-tract of hTel indicating a strong stabilization of the G4 by the compound (Figure 4.18, panel A, lane 17), while a lower stabilization was observed in the presence of NDI-PNA (Figure 4.18, panel A, compare lanes 15 and 17).

Data collected by *Taq* polymerase stop assay clearly and directly evidence that the conjugation of NDI with PNA resulted in many advantages. Firstly, it was possible to switch the NDI stabilizing activity on another G4 (LTR-IV) with respect to the naturally mainly stabilized G4 (LTR-III). Secondly, the NDI-PNA displayed a higher stabilizing activity on LTR G4s and a lower one on hTel than NDI alone, thus suggesting that the conjugate was more selective for its G4 target.

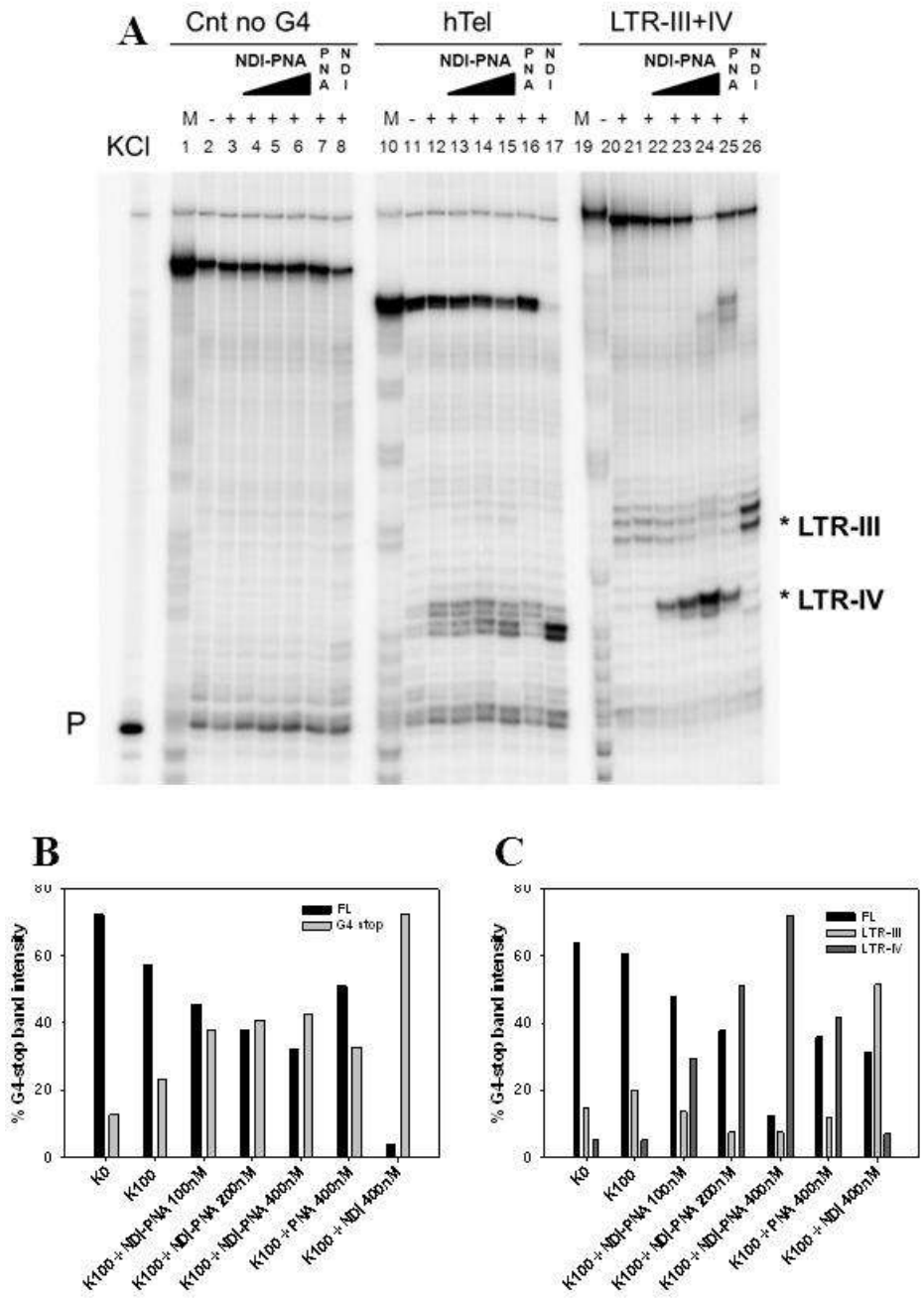


Figure 4.18: *Taq* polymerase stop assay in the presence of NDI-PNA. A) hTel, LTR-III+IV templates were amplified by the *Taq* polymerase in the absence and presence of 100 mM K^+, alone or combined with increasing amounts of NDI-PNA (100-400 nM). PNA and NDI (400 nM) alone were used as controls. Elongation was performed at 42 °C. A template (cnt no G4) made of a scramble sequence unable to fold into G4 was also used to exclude unspecific stalling of the polymerase. Lanes M: ladder of markers obtained by the Maxam and Gilbert sequencing carried out on the amplified strand complementary to the template strand. Lane P: unreacted labeled primer. B) Quantification of stop bands corresponding to formation of hTel G4 and of the full-length amplification product (FL) is shown. C) Quantification of stop bands corresponding to formation of LTR-III, LTR-IV G4s and of the full-length amplification product (FL) is shown.

To further investigate the selectivity of NDI-PNA for LTR-II+IV G4 over hTel G4, a FRET based competition assay was next performed. 5'-FAM and 3'-TAMRA labeled LTR-III+IV was mixed with increasing concentrations of unlabeled competitor, LTR-III+IV or hTel, and a constant amount of NDI-PNA that could compete for labeled *vs*. unlabeled oligonucleotide. Unlabeled LTR-III+IV and hTel G4 sequences, used as competitors, displayed a T_m of 64.9 ± 0.5 °C and 68.7 ± 0.1 °C respectively, sufficiently similar to allow a meaningful comparison. Given that, in FRET melting experiments the NDI-PNA led to an increase in labeled LTR-III+IV fluorescence at 30°C, the fluorescence of the labeled oligonucleotide was monitored at 30°C and the variation of fluorescence (ΔF) was calculated as the difference between the fluorescence of the labeled oligonucleotide in the presence of NDI-PNA and competitor and the basal fluorescence of the oligonucleotide alone. Data were indicated as percentage of the control (difference between the fluorescence of the labeled oligonucleotide in the presence of NDI-PNA and the basal fluorescence of the oligonucleotide alone) and are reported in Figure 4.19.

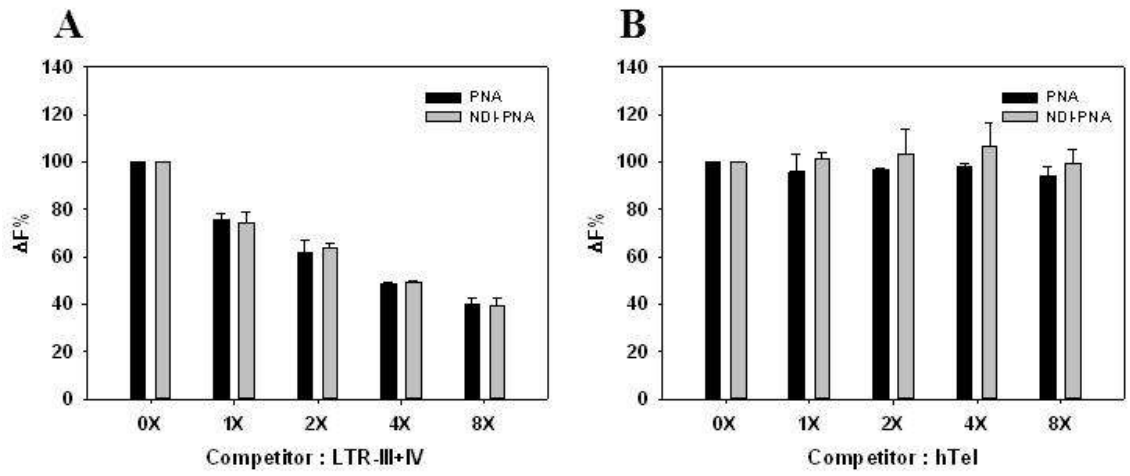


Figure 4.19: Competition analysis by FRET assay. 5'-FAM and 3'-TAMRA labeled LTR-III+IV (0.25 μM) was mixed with increasing concentrations (1-8X) of unlabeled competitor, A) LTR-III+IV or B) hTel, and a constant amount of NDI-PNA (1 μM) that could compete for labeled *vs.* unlabeled oligonucleotide. ΔF is as the difference between the fluorescence of the labeled oligonucleotide in the presence of NDI-PNA or PNA and competitor and the basal fluorescence of the oligonucleotide alone.

The behaviour of PNA and NDI-PNA was almost identical: a dose dependent decrease of ΔF% was detected in the presence of LTR-III+IV competitor suggesting that NDI-PNA and PNA bound both labeled and unlabeled oligonucleotides (Figure 4.19, panel A). On the contrary, no significant variation of ΔF% was observed with hTel, indicating that the latter did not compete with the labeled LTR-III+IV

for the binding of NDI-PNA and PNA (Figure 4.19, panel B). This data further indicate that PNA and NDI-PNA preferentially bound LTR-III+IV over hTel.

To confirm the data obtained through FRET-based competition assay, we finally performed competition by *Taq* polymerase stop assay. The assay was performed on the LTR-III+IV template and samples were incubated in the presence of 100 mM K^+, 400 nM of NDI-PNA and increasing amounts (1-8X) of competitor (LTR-III+IV or hTel).

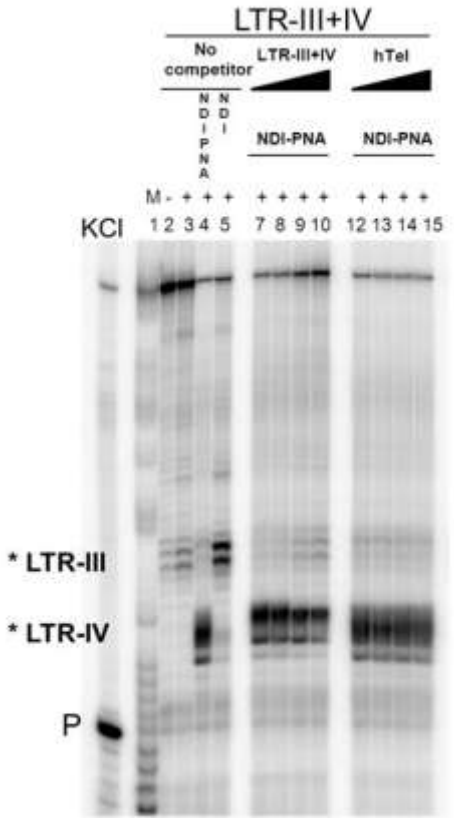


Figure 4.20: Competition analysis performed by *Taq* Polymerase stop assay. LTR-III+IV template was amplified by the *Taq* polymerase in the absence and in the presence of 100 mM K^+, alone or combined with a competitor (LTR-III+IV or hTel) (1-8X) in the presence of NDI-PNA 400 nM. Elongation was performed at 42 °C. Lanes M: ladder of markers obtained by the Maxam and Gilbert sequencing carried out on the amplified strand complementary to the template strand. Lane P: unreacted labeled primer.

As shown in Figure 4.20, in presence of increasing amounts of LTR-III+IV competitor, the intense stop site induced by NDI-PNA at the first 3'-G-tract corresponding to LTR-IV decreased in a concentration dependent manner while an increase of the stop site corresponding to LTR-III and the full-length product were observed. On the contrary, a mild effect was detected in presence of hTel competitor,

once again indicating that the NDI-PNA preferentially bound and stabilized LTR-III+IV with respect to hTel.

4.2.2 Nuclear localization signal mediates cell entry

In vivo applications of unmodified PNAs are hindered by poor cellular uptake and endosomal entrapment. In fact, due to the neutral backbone, PNAs do not associate with delivery vehicles based on cationic lipids [236]. Because the inefficient crossing of cellular membrane in mammalian cells by unmodified PNAs has been a major problem for their practical *in vivo* applications, we investigated NDI-PNA cell entry by confocal microscopy, exploiting the intrinsic fluorescence of NDI. TZM-bl cells were treated with 50 μM of conjugate and fluorescence was evaluated in the 560-700 nm range, using an excitation wavelength of 543 nm.

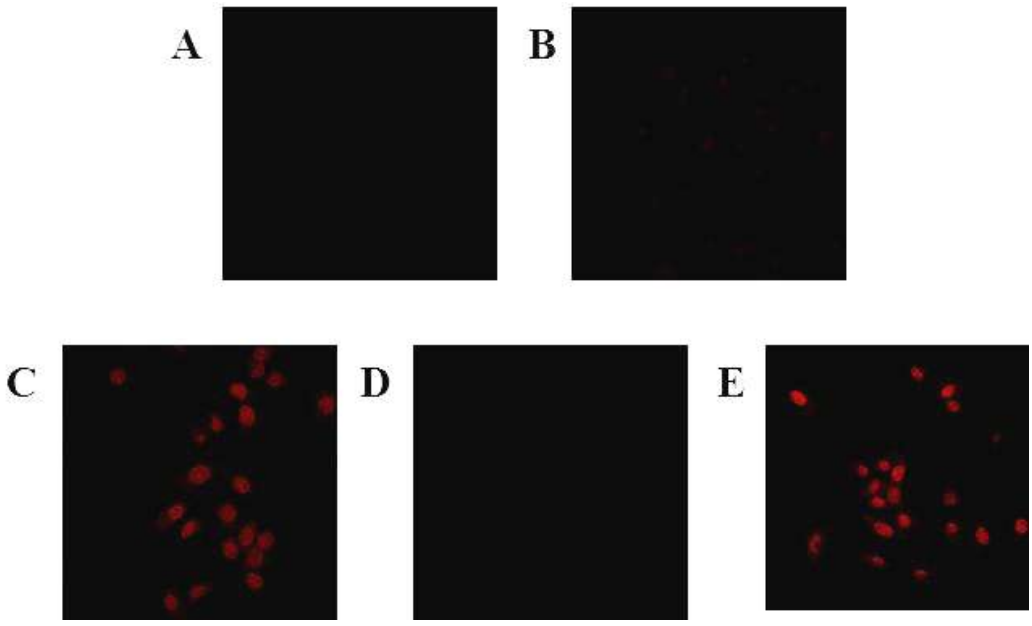


Figure 4.21: Evaluation of cell entry by confocal microscopy. TZM-bl cells were treated with 50 μM of conjugate for 3 h, washed with PBS to remove the excess of compounds and then fixed. Fluorescence was evaluated in the 560-700 nm range, using an excitation wavelength of 543 nm. A) no drug control; B) PNA; C) NDI; D) NDI-PNA; E) NLS NDI-PNA.

While NDI (Figure 4.21, panel C) was visualized in cell nuclei, NDI-PNA was not able to enter in cells (Figure 4.21, panel D). To overcome this critical problem a large number of different conjugation of PNA with cation peptides and cationic backbones modifications have been reported in literature [239]. As first approach, NDI-PNA was modified with a basic nuclear localization signal (NLS) (Pro-Lys-Lys-

Lys-Arg-Lys-Val) [293]. As reported in Figure 4.21, panel E, the modification was sufficient to mediate the transfer across cell membrane, delivering the NDI-PNA in the cell nucleus.

To investigate the effect of NLS on NDI-PNA activity, a *Taq* polymerase stop assay was performed on LTR-III+IV and hTel templates. Samples were incubated in the absence/presence of 100 mM K^+ and increasing amounts (100-400 nM) of NLS NDI-PNA. NDI-PNA, PNA and NDI were also tested at 400 nM. As indicated in Figure 4.22 NLS NDI-PNA displayed a stabilizing activity of hTel and LTR-III+IV very similar to NDI-PNA (compare FL products of lane 16 to lane 17 for hTel and FL products of lane 26 to lane 27 for LTR-III+IV), suggesting that NLS did not significantly modify the effect of NDI-PNA.

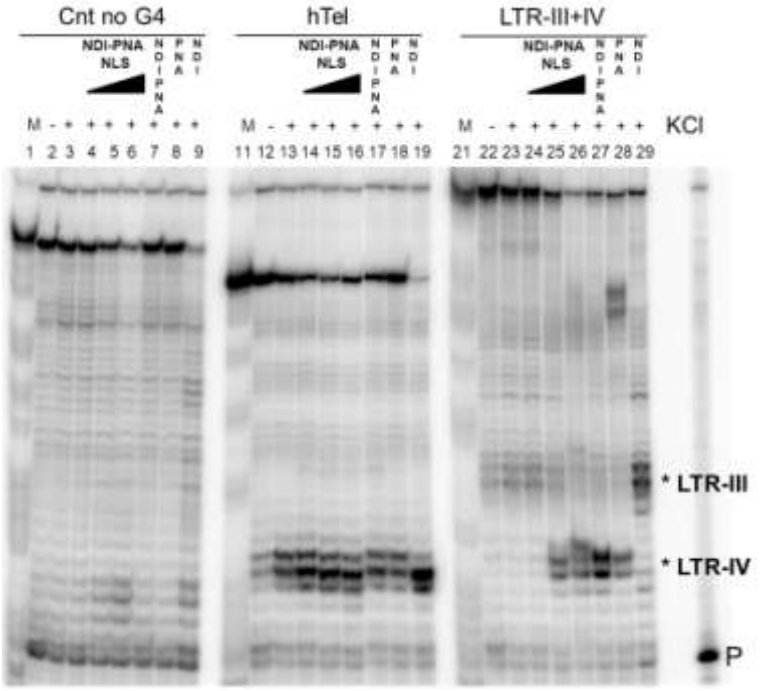


Figure 4.22: *Taq* polymerase stop assay in the presence of NLS NDI-PNA. A) hTel, LTR-III+IV templates were amplified by the *Taq* polymerase in the absence and presence of 100 mM K^+, alone or combined with increasing amounts of NLS NDI-PNA (100-400 nM). NDI-PNA, PNA and NDI (400 nM) alone were used as controls. Elongation was performed at 37 °C. A template (cnt no G4) made of a scramble sequence unable to fold into G4 was also usedto exclude unspecific stalling of the polymerase. Lanes M: ladder of markers obtained by the Maxam and Gilbert sequencing carried out on the amplified strand complementary to the template strand. Lane P: unreacted labeled primer.

4.2.3 Discussion

By synthesizing NDI-PNA conjugates, we were able to set a completely new method to selectively target G4s, which was successfully tested on HIV-1 LTR G4s. In particular, this first generation of "sequence-structure" specific NDI-PNA conjugates was able to induce the formation of LTR-IV G4, destabilizing the folding of the most stable G4 in the full-length G-rich sequence of LTR promoter, i.e. LTR-III. Their efficient selectivity was driven by a combination of π-staking interactions between the NDI core and an external tetrad of the G4 and by PNA/nucleic acid sequence specific hybridization. To overcome the lack of cellular uptake, the conjugated was successfully modified with NLS, without affecting activity.

This innovative approach is fundamental to reach the specific discrimination between LTR-III and LTR-IV G4s, whose balance seems to act as regulator element of the viral promoter [287]. Therefore, NDI-PNA conjugates could be used to modulate the folding and unfolding of unique G4s within the full-length G-rich sequence of the LTR promoter and better investigate their single role and function. Given NDI-PNA selectivity over the cellular G4, they represent a key point for the future development of new anti-HIV-1 drugs with an innovative mechanism of action that may complement current clinical AIDS therapies. Our new approach also highlights the possibility to selectively recognize G4 structures with NDI-PNA conjugates, prompting the development of G4 binders specific for G4s associated to several important diseases.

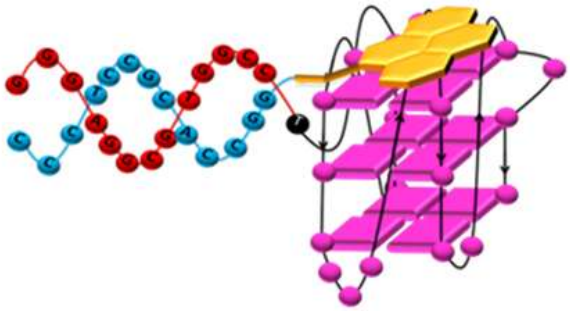


Figure 4.23: New proposal mechanism of targeting LTR-IV G4 with NDI-PNA conjugate. G4-folded LTR-IV is shown in magenta, LTR-IV 5'-flanking region is shown in red, PNA sequence is shown in light blue, NDI is shown in yellow.

4.3 Identification and characterization of HIV-1 LTR G4s interacting proteins

Regulatory mechanisms controlled by G4s involve the binding of protein factors that modulate G4 conformation and/or serve as a bridge to recruit additional protein regulators. G4 binding proteins can be classified into three functional groups: telomere-related proteins, such as the shelterin complex; proteins that unfold the G4s, such as the helicase and heterogeneous nuclear ribonucleoprotein families; proteins that stabilize G4s, a large group which includes nucleolin, MAZ and nucleophosmin. G4s and their cognate proteins are key players in numerous essential processes in eukaryotic cells and in recent years, new studies have contributed to increase our knowledge of the biological significance of G4s in prokaryotes and viruses. In this direction, the significance of G4s as focal points of interactions with host and viral factors is supported also by the observation that G4-folded sequences are specifically recognized by various viral proteins, such as the EBNA1 and the SARS-CoV unique domain, which occurs exclusively in highly pathogenic strains. For this reason, we have decided to pursue the investigation of putative cellular/viral proteins that may be involved in the regulation of the G4 LTR promoter activity in HIV [64,234]. We have previously shown that nucleolin, the most abundant nucleolar protein, can bind, induce and stabilize the LTR G4s, thus silencing viral transcriptional process [234]. Therefore, we reasoned that other proteins may exist that regulate the G4/ds equilibrium at the LTR.

4.3.1. The human nuclear ribonucleoprotein (hnRNP) A2/B1 selectively binds the HIV-1 LTR G-quadruplexes

We looked for additional proteins that recognize and modulate LTR G4s by pull-down assay of HEK 293T nuclear cell extracts against the whole region in the HIV-1 LTR that can fold into G4, i.e. LTR-II+III+IV, which can alternatively fold into three G4s, i.e. LTR-II, LTR-III and LTR-IV. We employed two different oligonucleotides as controls: a two-point mutation LTR-II+III+IV oligonucleotide, i.e. LTR-II+III+IV M4+5, which has been previously shown to be unable to fold into G4 [232], and a random oligonucleotide of the same length, i.e. LTR-II+III+IV random. After washing at increasing ionic strength to destabilize weak protein-G4 interactions, the final eluted sample was run on SDS-PAGE and subjected to MS analysis. A series of heterogeneous nuclear ribonucleoproteins (hnRNPs) were found to be present in the G4-folded oligonucleotide; among them, only hnRNP A2/B1 was obtained in all three independent tests. The identity of hnRNP A2/B1 was confirmed by pull-down assay followed by western blot analysis with an anti-hnRNP A2/B1 antibody. The hnRNP A2/B1 protein was confirmed to bind mainly to the G4-folded sequence, even if in these settings it displayed mild

recognition also of the G-rich non-G4-folding oligonucleotide. No binding was detected in the presence of the random oligonucleotide (Figure 4.24, panel B).

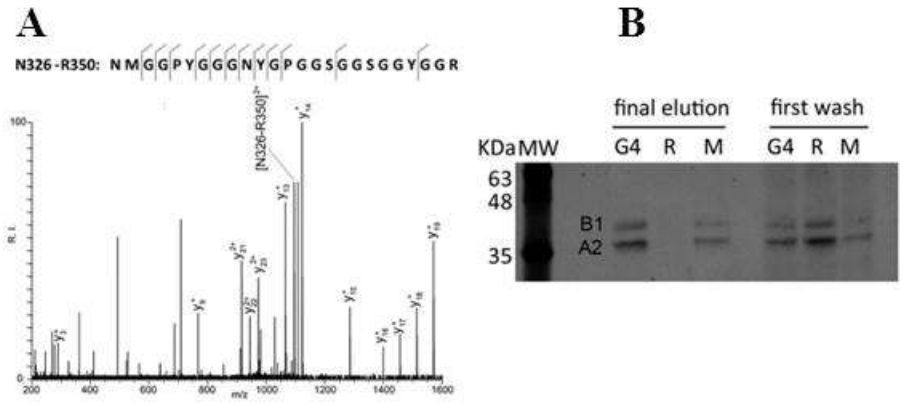


Figure 4.24: A) MS/MS spectrum of the precursor ion observed at m/z 1095.46 in the sample mixture from digestion of LTR-II + III + IV G4 sample. Only characteristic y ions are indicated [294]. The data match the sequence of peptide N326-R350 of hnRNP A2/B1, which is reported on top with the observed fragments. B) Pull-down assay of nuclear extract proteins with wt, mutant G4 LTR-II + III + IV (M) and random (R) sequences, immobilized on agarose beads. Shown is the western blot analysis with an anti-hnRNP A2/B1 antibody.

To confirm selectivity toward the G4 LTR oligonucleotide, SPR analysis was performed. We employed the recombinant purified hnRNP A2, which lacks 12 amino acids in the amino-terminal region and is the main isoform, accounting for ~ 90% of the protein in most tissues [295]. It was immobilized on a SPR chip and binding affinity of the unlabeled oligonucleotides was measured. LTR-II+III+IV displayed the highest affinity (K_D = 19.5 ± 1.5 nM) (Figure 4.25, panel A), while the mutant LTR-II+III+IV M4+5 showed a higher K_D (35.5 ± 3.0 nM) (Figure 4.25, panel B). Chi^2 and U-values were < 10 and < 15, respectively, indicating optimal data fitting [296]. In contrast, binding to a random oligonucleotide of the same length as LTR-II+III+IV was so low that in these conditions it was not possible to obtain a meaningful K_D value (Figure 4.25, panel C).

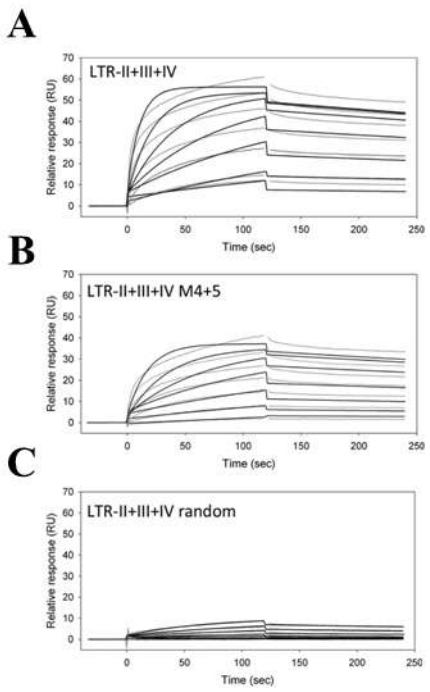


Figure 4.25: Binding affinity of hnRNP A2 for the tested oligonucleotides by SPR analysis. The human recombinant hnRNP A2 was immobilized on the SPR chip by amine coupling chemistry and the unlabeled oligonucleotides were bound at a flow rate of 25 µL/min, with contact time and dissociation time of 120 s in HEPES-KCl buffer. Sensograms were obtained in the oligonucleotide concentration range of 31.25 nM–2 µM. SPR sensograms of A) LTR-II + III + IV with ka $4.7 \times 10^4 \pm 5.2 \times 10^2$ Ms^{-1} and kd $9.2 \times 10^{-4} \pm 7.5 \times 10^{-5}$ s^{-1}, B) LTR-II + III + IV M4-5 with ka $2.8 \times 10^4 \pm 3.4 \times 10^2$ Ms^{-1} and kd $10.0 \times 10^{-4} \pm 8.4 \times 10^{-5}$ s^{-1} and C) LTR-II + III + IV Random, the ka and kd values of which are not reliably measurable. Sensograms are shown as grey lines and their respective fits as black lines.

4.3.2 HnRNP A2 unfolds the LTR G4s

We next set out to investigate the effect of hnRNP A2 binding to the LTR G4s. First of all, a *Taq* polymerase stop assay was performed on the LTR-II+III+IV template. In the presence of 100 mM K^+, stop sites corresponding to formation of LTR-III and LTR-II were visible (Figure 4.26, panel A, lane 2). No stop band corresponding to LTR-IV was detected, as expected, since LTR-IV has been previously reported to form only in the presence of G4 ligands. Upon addition of hnRNP A2, both LTR-III and LTR-II stop sites showed a slight decrease (Figure 4.26, panel A, compare lanes 3 and 2). No effect was induced by addition of K^+ and hnRNP A2 in the control LTR-II+III+IV M4+5 template lacking the possibility to form G4 (Figure 4.26, panel A, lanes 5-7). Quantification of stop bands corresponding to formation of LTR-II and LTR-III G4s and of the full-length amplification product (FL) in lanes 1-3 are reported in Figure 4.26, panel B.

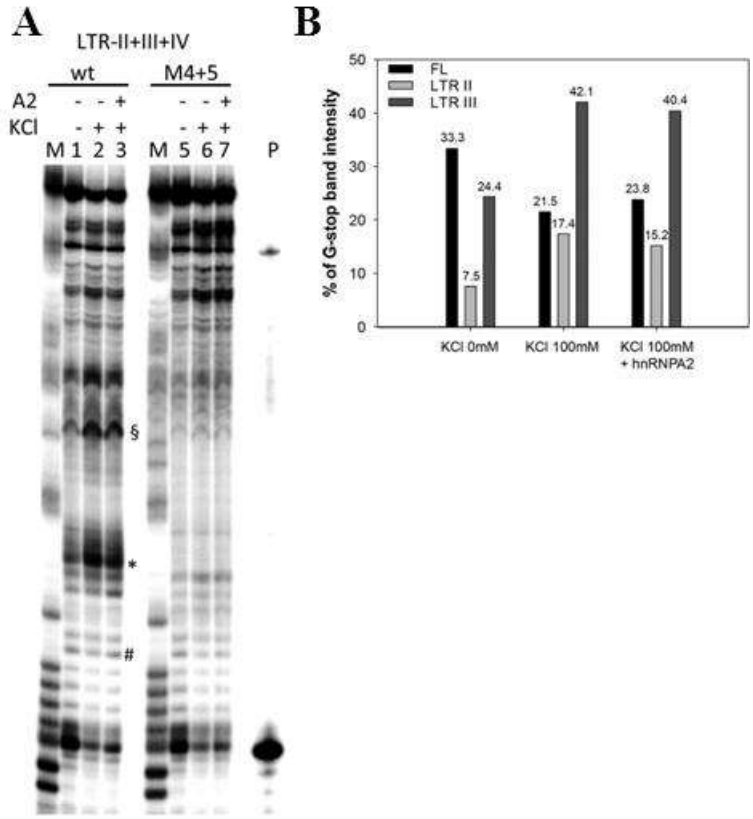


Figure 4.26: *Taq* polymerase stop assay in the presence of hnRNP A2. A) wt and mutant M4+5 LTR-II+III+IV templates were amplified by the *Taq* polymerase in the absence (lanes 1 and 5) and presence of 100 mM K^+, alone (lanes 2 and 6) or combined with hnRNP A2 (lanes 3 and 7). Elongation was performed at 37 °C. Stop regions corresponding to LTR-III are indicated by the * symbol, to LTR-II by § and to LTR-IV by#. Lanes M: ladder of markers obtained by the Maxam and Gilbert sequencing carried out on the amplified strand complementary to the template strand. Lane P: unreacted labeled primer. B) Quantification of stop bands corresponding to formation of LTR-II and LTR-III G4s and of the full-length amplification product (FL) is shown.

As suggested by Xodo [293], the apparent lack of effect in the presence of an unfolding protein could be due to the protein binding to the template sequence which would stimulate polymerase stop and mask G4 release at the same binding site. We thus switched to FRET, a dual-labeled system where oligonucleotide folding could be monitored by changes in fluorescence. Given that the major determinant of FRET efficiency is the distance between the acceptor and the donor, fluorescence intensities depict nucleic acids folding states. Consequently, when annealed to its complementary sequence to form a double-stranded structure, the tested oligonucleotide would yield the maximum fluorescence intensity, while the G4 folded conformation would be the least fluorescent. In these

conditions the measured fluorescence intensity allows to calculate the energy transfer (E) and the end-to-end distance (R) between the two fluorophores, and therefore the unfolding degree. The G4 folded LTR-II+III+IV sequence was characterized by R of ~38 Å, with E ~0.84 (Table 4.7). When the G4 structure was converted into the duplex conformation by addition of the complementary C-rich strand, the fluorophores were separated by ~150 Å, and FRET was approximately 0. Treatment of LTR-II+III+IV G4 with hnRNP A2 increased fluorescence by 3-folds with respect to the free G4, and E decreased to 0.51 (Figure 4.27). Considering that the complete unfolding of the G4 structure required $\Delta E = 0.72$, the $\Delta E = 0.30$ induced by hnRNP A2 reflected a 42% unfolding (Figure 4.27). The negative control protein BSA showed negligible unfolding activity in these conditions (Figure 4.27).

Table 4.7: Energy Transfer (E), Energy Transfer Difference (ΔE) and Radius (Å) of LTR-II+III+IV G4 (0.1 μM) in the absence and presence of hnRNPA2 or BSA, at different K^+ concentrations.

K^+(mM)	100			50			25		
	E ± SD	**ΔE***	**R (Å)**	**E ± SD**	**ΔE***	**R (Å)**	**E ± SD**	**ΔE***	**R (Å)**
G4	0.8 4 ± 0.01	0.727	38.1	0.81 ± 0.01	0.716	38.6	0.83 ± 0.01	0.705	39.0
A2	0.53 ± 0.02	0.309	49.1	0.22 ± 0.01	0.564	59.4	0.26 ± 0.02	0.592	61.6
BSA	0.82 ± 0.01	0.018	38.9	0.81 ± 0.01	0.024	39.6	0.80 ± 0.01	0.006	39.3

*For free G4, ΔE was calculated between the single-stranded oligonucleotide and the corresponding duplex. SD stands for standard deviation.

To test the statistical significance of the effect observed in the presence of the protein, we applied a model comparison approach, which allows testing the statistical role of an independent variable (in our case the hnRNP A2 protein) through the comparison of two models differing only in that variable. The test yielded $F(2, 189) = 36.27$ and $P < 0.05$, indicating that the difference observed in the presence and absence of the protein was statistically significant.

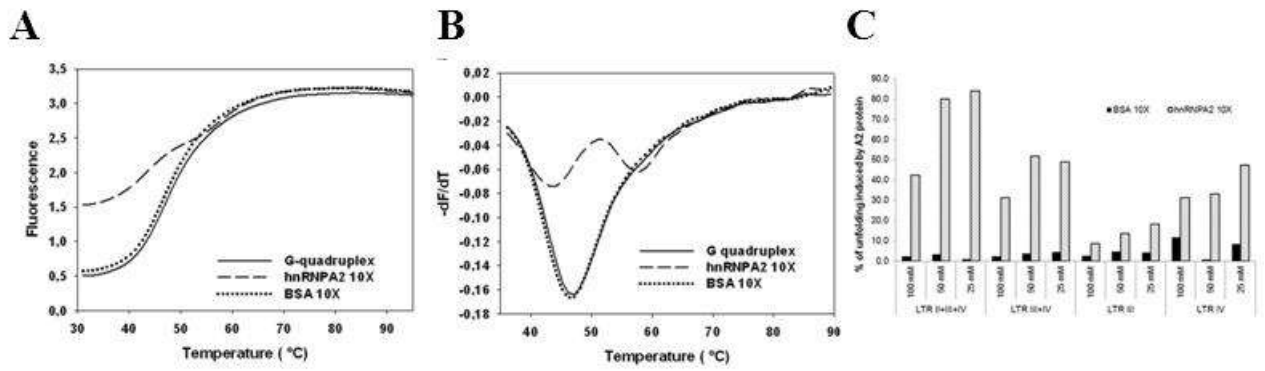


Figure 4.27: FRET analysis of hnRNP A2 unfolding. A) FRET-melting curves of LTR F-II + III + IV-T, in the absence and presence of hnRNP A2 1:10 or of BSA as control. B) Corresponding first derivative curves, -dF/dT versus T. Proteins were incubated with the LTR G4s for 30 min at 37 °C prior to analysis. C) Unfolding % of all tested sequences at 100, 50, 25 mM K^+, incubated in the presence of hnRNP A2 or BSA at 37 °C for 30 min. Unfolding values are referred to unfolding in the absence of proteins (100%).

Analysis of the activity of hnRNP A2 was next extended to shorter LTR sequences, i.e. LTR-III+IV, LTR-III and LTR-IV G4s, folded in 100 mM K^+. The unfolding of LTR-III+IV and LTR-IV measured by FRET was similar but lower than that on the full-length sequence (31%); unfolding of LTR-III was very low (8.5%) (Figure 4.27, panel C). The effect of hnRNP A2 was also investigated in G4-forming sequences folded in lower K^+ concentrations, i.e. 50 and 25 mM. In these conditions, a distinct enhancement of unfolding was observed: in particular, unfolding up to 83% was obtained in the LTR-II+III+IV oligonucleotide (Figure 4.27, panel C).

Because G4s are less stable at lower K^+ concentrations, we proposed that the diverse unfolding efficiency towards the different-length LTR G4s depended on the stability of the LTR G4s. To test this hypothesis, T_m values of the dual-labeled oligonucleotides were thus measured by FRET-monitored thermal unfolding (Table 4.8). The most stable sequence was LTR-III (T_m = 62.1 ± 0.1°C), followed by LTR-IV and LTR-III+IV (T_m = 59.1 ± 0.1°C and T_m = 55.2 ± 0.1°C, respectively); LTR-II+III+IV was the least stable sequence (T_m = 49.1 ± 0.1), confirming that the observed unfolding scale depended on the oligonucleotide stability in the FRET system. To note that in the unlabeled oligonucleotides measured by circular dichroism (CD), LTR-IV was the least stable sequence (Table 4.8), bias which likely derives by the absence of fluorophores at the oligonucleotides' 5′- and 3′-end, which affect T_m values, as previously noted [288].

In addition, the longest sequences, i.e. LTR-II+III+IV and LTR-III+IV, where multiple G4s may form, display T_m values that represent the average stability, which depends on the effective formation of the possible G4s.

Table 4.8: Tm of LTR G4s measured on dual-labeled oligonucleotides by FRET and on unlabeled oligonucleotides by CD.

LTR	K^+ (mM)	Tm ± SD (°C)	
		FRET	CD
II + III + IV	100	49.1 ± 0.1	61.1 ± 0.2
	50	45.1 ± 0.1	nd
	25	43.2 ± 0.1	nd
III + IV	100	55.2 ± 0.1	64.6 ± 0.1
	50	49.1 ± 0.1	nd
	25	46.2 ± 0.1	nd
III	100	62.1 ± 0.1	68.7 ± 0.5
	50	56.0 ± 0.1	nd
	25	52.1 ± 0.1	nd
IV	100	59.1 ± 0.1	50.6 ± 0.8
	50	57.1 ± 0.1	nd
	25	51.1 ± 0.1	nd

T_m values are reported with standard deviation. Nd stands for "not-determined". SD stands for standard deviation.

We thus performed a *Taq* polymerase stop assay to assess the G4 mutual formation and stability at increasing K^+ concentration in the full-length LTR. LTR-III was used as a control sequence where only one G4 could form. We observed that in the full-length LTR-II+III+IV sequence, both LTR-II and LTR-III G4s formed upon addition of K^+ (Figure 4.28, panel A, lanes 2-3); among them, LTR-III G4 induced a slightly more intense stop than LTR-II, suggesting a preferred formation of LTR-III in the full-length sequence (Figure 4.28, panel A, lanes 2-3). LTR-III was also the only G4 forming in the LTR-III+IV template (Figure 4.28, panel A, lanes 5-7). Interestingly, the LTR-III that formed in the full-length sequence was less intense than LTR-III forming in the shortest oligonucleotides (Figure 4.28, panel A, compare lanes 2-3, with 6–7 and 9-10). These data indicate that i) LTR-III is the most prominent G4 among the mutually exclusive LTR G4s and ii) when LTR-III forms in the full-length LTR promoter, its stability maintains a level that is fully susceptible to hnRNP A2 processing.

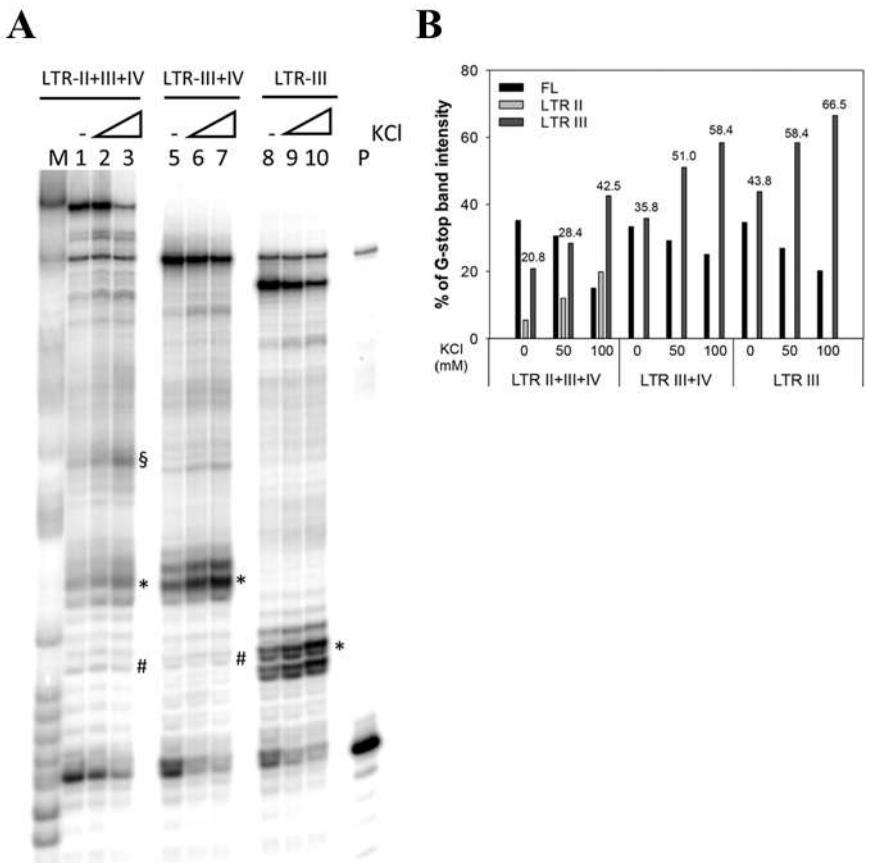


Figure 4.28: *Taq* polymerase stop assay at increasing K^+ concentrations. A) LTR-II+III+IV, LTR-III+IV and LTR-III templates were amplified by the *Taq* polymerase in the absence or presence of K^+ (50-100 mM). Elongation was performed at 37 °C. Stop regions corresponding to LTR-III are indicated by the * symbol, to LTR-II by § and to LTR-IV by#. Lanes M: ladder of markers obtained by the Maxam and Gilbert sequencing carried out on the amplified strand complementary to the template strand. Lane P: unreacted labeled primer. B) Quantification of stop bands corresponding to formation of LTR-II and LTR-III G4s and of the full-length amplification product (FL) is shown.

4.3.3 Depletion of hnRNP A2/B1 decreases viral transcription

We have shown that LTR G4 folding inhibits viral transcription, whereas when point mutations that disrupt G4s are introduced in the LTR promoter, an increase in promoter activity is evidenced [232]. Silencing of the LTR G4 folding/stabilizing protein nucleolin significantly increased promoter activity, indicating the inhibitory effect of nucleolin on LTR-driven transcription [234].

hnRNP A2/B1 silencing followed by luciferase reporter assay was established to explore the downstream biological effects of hnRNP A2/B1 binding/unfolding to the full-length LTR G4s. HEK 293T cells were treated with increasing amounts of anti-hnRNP A2/B1 siRNAs, and then transfected with wt or mutant M4+5 LTR luciferase reporter plasmids. Analysis of hnRNP A2/B1 content showed

that the protein was effectively depleted (Figure 4.29, panel A). In these conditions, LTR-driven transcription decreased by 45% of the untreated control at the highest siRNA concentration in the wt LTR sequence (Figure 4.29, panel B). In contrast, the LTR sequence with two point mutations that disrupt G4 folding was only marginally affected by hnRNP A2/B1 depletion (Figure 4.29, panel B), indicating that the effect was G4-specific. In these conditions, hnRNP A2/B1-silenced HEK 293T cells showed no reduction in cell viability at 48 h with respect to the controls, thus confirming the consistency of the inhibitory effects on LTR promoter activity.

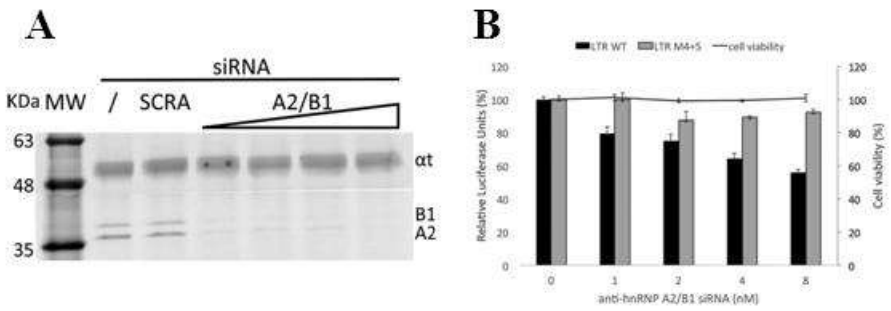


Figure 4.29: Effect of hnRNP A2/B1 on the HIV-1 LTR promoter in HEK 293T cells. A) hnRNP A2/B1 depletion in HEK 293T cells by siRNAs analysed by western blot with anti hnRNP A2/B1 antibody. Scra indicates scrambled siRNAs. Detection of α-tubulin (αt) was used as control. The symbol "/" indicates untreated cells. B) Analysis of the luciferase activity of the wt and M4+5 LTR promoters in HEK 293T cells treated with hnRNP A2/B1 siRNAs, normalized to protein content.

Silencing of hnRNP A2/B1 was also performed in the TZM-bl reporter cell line, which contains stably integrated copies of the luciferase gene under control of the LTR wt promoter and thus closely resembles the integrated provirus. The LTR-driven reporter transcription decreased up to 33% in hnRNP A2/B1-silenced cells (Figure 4.30). These data further confirm the unfolding activity of hnRNP A2/B1 on the LTR G4s.

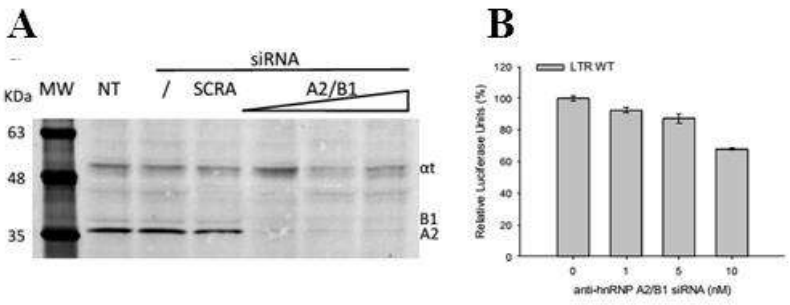


Figure 4.30: Activity of hnRNP A2/B1 on the HIV-1 LTR promoter in TZM-bl cells. A) hnRNP A2/B1 depletion in TZM-bl cells by siRNAs analyzed by western blot with anti hnRNP A2/B1 antibody. Scra indicates scrambled siRNAs. Detection of α-tubulin (αt) was used as control. NT indicates untreated cells. The symbol "/" indicates cells treated only with Lipofectamine RNAimax. B) Analysis of the luciferase activity of the wt LTR promoter in TZM-bl cells treated with hnRNP A2/B1 siRNAs, normalized to protein content.

4.3.4 Discussion

We identified and characterized G4-selectivity and function of the cellular protein hnRNP A2/B1, the first one shown to unfold G4s in the HIV-1 LTR promoter. HnRNPs are RNA/protein complexes which bind newly synthesized RNA (pre-mRNA) in the cell nucleus during gene transcription and post-transcriptional modifications [297]. However, two of the most abundant of them, hnRNP A1 and hnRNP A2/B1, have been reported to bind also DNA G4s and unfold them. In particular, hnRNP A2 has been shown to unfold the fragile X repeats [298] and the shorter hnRNP A2 splice variant to promote telomere extension in mammalian cells [299]. In this context, our finding that hnRNP A2 binds and unfolds the G4s in the promoter of HIV-1 not only validates the DNA G4 unfolding activity of this protein, but also points out for the first time that this type of activity is exploited by a virus.

In HIV-1, hnRNP A2/B1 has been reported to play with Rev an important accessory role in promoting nuclear viral RNA retention and nucleocytoplasmic viral RNA transport [300]. In this context, the lower virus production observed upon siRNA-mediated depletion of hnRNP A2/B1 in infected cells was ascribed to accumulation of viral genomic RNA in cytoplasmic compartments or in the nucleus [301]. Our results showing decreased transcriptional activity upon depletion of hnRNP A2/B1 indicate a possible additional mechanism of inhibition of virus production mediated by increased G4 folding in the HIV-1 LTR promoter in the absence of hnRNP A2/B1. Surprisingly, overexpression of this protein and other hnRNP members induced similar effects, i.e. inhibition of transcription and reduction of virus production [302]. We suggest that upon stimuli that boost viral transcription (i.e. integral release of LTR

G4s upon hnRNP A2/B1 massive overexpression) the virus (or the cell) activates counteracting mechanisms to avoid excessive exploitation of the cell by the virus that would lead to fast cell death, in the end impairing virus production. Therefore, the reported transcription inhibition might be mediated by other factors triggered by hnRNP A2/B1 overexpression.

The activity of hnRNP A2/B1 is fit to unfold the G4s that actually form in the HIV-1 LTR region. In fact, G4s folded in short oligonucleotides, such as LTR-III and LTR-IV, which displayed higher thermal stability, were only partially unfolded by the protein. In contrast, the same G4s folding in a longer context (LTR-II+III+IV) were less stable and were effectively processed by hnRNP A2/B1. Obviously, the longer oligonucleotide better mimes the actual G4 condition in the full-length LTR, as also proved by transcription inhibition obtained by hnRNP A2/B1 depletion in cells that contained the entire HIV-1 LTR.

As in eukaryotic G4-modulated promoters, the HIV-1 LTR promoter is thus processed by G4 stabilizing (nucleolin) [234] and destabilizing proteins (hnRNP A2/B1). Because the G4 switch plays an important role in activation of effective viral transcription, our findings not only advance our understanding on the mechanism of viral transcription but may also constitute a progress from a therapeutic point of view.

Chapter 5
Conclusions

In the first part of this thesis we explored the possibility of selectively target LTR G4s. Through a screening approach we identified a new family of small molecules displaying clear-cut selectivity for viral G4s over the telomeric G4, which is representative of cellular G4. Spectroscopic analysis combined with polymerase enzyme inhibition assay confirmed the preferential binding of these compounds to viral G4s, along with SPR sensorgrams which corroborated the preferential affinity. The selectivity toward LTR G4s of the best hits was translated to a promising G4-mediated anti-HIV-1 activity. It is worth noting that no drugs acting on the integrated provirus have ever been reported so far; consequently, these compounds are encouraging backbones for the future development of new anti-HIV-1 drugs with an innovative mechanism of action which could be employed to implement the current antiviral treatment. In this context, it is crucial to figure out the structure of active fractions of hits in order to better investigate and characterize their antiviral activity. Considering that known G4 ligands are all characterized by large flat scaffolds which make them poorly druggable, our findings strongly pave the way to selective recognize G4s with small molecules, prompting the development of G4 ligands specific for G4s not only implicated in HIV-1 transcription, but to a wider extension in other several important human diseases.

We also faced the selectivity challenge through naphthalene diimide-peptide nucleic acid (NDI-PNA) conjugates to combine the stabilizing activity of a G4 ligand, with fingerprint recognition of the nucleotide sequence proximal to the G4 of interest. Spectroscopic techniques along with *Taq* polymerase stop assay revealed that the NDI-PNA conjugate was effectively able to selectively target the G4 of interest, the LTR-IV G4, preventing the formation of the most stable G4, the LTR-III G4. NDI-PNA was also successfully modified with NLS to achieve cell entry without affecting activity. It is worth noting that while NDI stabilises both the telomeric and LTR G4s (mostly LTR-III), the NDI-PNA conjugate showed a higher stabilizing activity for viral G4s over the cellular one: this result is necessary to avoid side effects due to interaction with cellular G4. Moreover, it is fundamental to reach

the specific discrimination between LTR-III and LTR-IV G4s, as their balance seems to act as regulator element of the viral promoter. This innovative approach allows conveying the stabilizing activity of a ligand directly to the G4 of interest and it could be more generally applied to selectively target G4s involved in different key biological processes dissect the activity of each investigated G4.

Finally, through pull-down assay combined with mass spectrometry analysis we identified the cellular protein hnRNP A2/B1, which selectively binds LTR G4s. It is the first protein able to unfold LTR G4s, as demonstrated by FRET analysis and proved by LTR transcription inhibition obtained in hnRNP A2/B1 depleted cells. This finding, not only confirmed the reported unfolding activity of this protein, but also indicates for the first time that this kind of effect is exploited by a virus. Because the LTR promoter has been suggested to be the region where viral latency is regulated, the G4 switch may play a role not only in activation of effective viral transcription, but also in its shift to latency. Therefore, our data not only advance our understanding on the mechanism of viral transcription but may also constitute a progress from a therapeutic point of view.

Chapter 6

Bibliography

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

2. Howard, F. B., Frazier, J. & Miles, H. T. Stable and metastable forms of poly(G). *Biopolymers* **16,** 791–809 (1977).

3. Davis, J. T. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.* **43,** 668–98 (2004).

4. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5,** 182–6 (2013).

5. Artusi, S. *et al.* Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. *Nucleic Acids Res.* **44,** 10343–10353 (2016).

6. Brown, R. V & Hurley, L. H. DNA acting like RNA. *Biochem. Soc. Trans.* **39,** 635–40 (2011).

7. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334,** 364–6 (1988).

8. Ou, T. *et al.* G-quadruplexes: targets in anticancer drug design. *ChemMedChem* **3,** 690–713 (2008).

9. Sen, D. & Gilbert, W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **344,** 410–4 (1990).

10. Chowdhury, S. & Bansal, M. A Nanosecond Molecular Dynamics Study of Antiparallel d(G) $_7$ Quadruplex Structures: Effect of the Coordinated Cations. *J. Biomol. Struct. Dyn.* **18,** 647–669 (2001).

11. Phan, A. T., Kuryavyi, V., Luu, K. N. & Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Res.* **35,** 6517–25 (2007).

12. Yang, D. & Okamoto, K. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.* **2,** 619–46 (2010).

13. Reshetnikov, R. V, Kopylov, A. M. & Golovin, A. V. Classification of g-quadruplex DNA on the basis of the quadruplex twist angle and planarity of g-quartets. *Acta Naturae* **2,** 72–81

(2010).

14. Borbone, N. *et al.* Synthesis and characterization of tetra-end linked oligonucleotides capable of forming monomolecular G-quadruplexes. *Nucleosides. Nucleotides Nucleic Acids* **26,** 1231–6 (2007).

15. Ogloblina, A. M. *et al.* Parallel G-Quadruplexes Formed by Guanine-Rich Microsatellite Repeats Inhibit Human Topoisomerase I. *Biochemistry. (Mosc).* **80,** 1026–38 (2015).

16. Patel, D. J., Phan, A. T. & Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **35,** 7429–7455 (2007).

17. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34,** 5402–15 (2006).

18. Karsisiotis, A. I. *et al.* Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed. Engl.* **50,** 10645–8 (2011).

19. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **350,** 718–20 (1991).

20. Chen, B.-J., Wu, Y.-L., Tanaka, Y. & Zhang, W. Small Molecules Targeting c-Myc Oncogene: Promising Anti-Cancer Therapeutics. *Int. J. Biol. Sci.* **10,** 1084–1096 (2014).

21. Zhang, S., Wu, Y. & Zhang, W. G-Quadruplex Structures and Their Interaction Diversity with Ligands. *ChemMedChem* **9,** 899–911 (2014).

22. Fedoroff, O. Y. *et al.* NMR-Based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry* **37,** 12367–74 (1998).

23. Izbicka, E. *et al.* Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res.* **59,** 639–44 (1999).

24. 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,** 6722–4 (2008).

25. Kim, M.-Y., Vankayalapati, H., Shin-ya, K., Wierzba, K. & Hurley, L. H. Telomestatin, a Potent Telomerase Inhibitor That Interacts Quite Specifically with the Human Telomeric Intramolecular G-Quadruplex. *J. Am. Chem. Soc.* **124,** 2098–2099 (2002).

26. Ou, T. *et al.* G-Quadruplexes: Targets in Anticancer Drug Design. *ChemMedChem* **3,** 690–713 (2008).

27. Duan, W. *et al.* Design and synthesis of fluoroquinophenoxazines that interact with human telomeric G-quadruplexes and their biological effects. *Mol. Cancer Ther.* **1,** 103–20 (2001).

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

29. Schultes, C. M., Guyen, B., Cuesta, J. & Neidle, S. Synthesis, biophysical and biological evaluation of 3,6-bis-amidoacridines with extended 9-anilino substituents as potent G-

quadruplex-binding telomerase inhibitors. *Bioorg. Med. Chem. Lett.* **14,** 4347–51 (2004).

30. White, E. W. *et al.* Structure-specific recognition of quadruplex DNA by organic cations: influence of shape, substituents and charge. *Biophys. Chem.* **126,** 140–53 (2007).

31. Ou, T. *et al.* G-Quadruplexes: Targets in Anticancer Drug Design. *ChemMedChem* **3,** 690–713 (2008).

32. Rangan, A., Fedoroff, O. Y. & Hurley, L. H. Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule. *J. Biol. Chem.* **276,** 4640–6 (2001).

33. Nadai, M. *et al.* Naphthalene diimide scaffolds with dual reversible and covalent interaction properties towards G-quadruplex. *Biochimie* **93,** 1328–40 (2011).

34. Marchetti, C. *et al.* Macrocyclic naphthalene diimides as G-quadruplex binders. *Bioorg. Med. Chem.* **23,** 3819–30 (2015).

35. Doria, F., Manet, I., Grande, V., Monti, S. & Freccero, M. Water-soluble naphthalene diimides as singlet oxygen sensitizers. *J. Org. Chem.* **78,** 8065–73 (2013).

36. Doria, F. *et al.* Water soluble extended naphthalene diimides as pH fluorescent sensors and G-quadruplex ligands. *Org. Biomol. Chem.* **10,** 3830–40 (2012).

37. Zuffo, M., Doria, F., Spalluto, V., Ladame, S. & Freccero, M. Red/NIR G-Quadruplex Sensing, Harvesting Blue Light by a Coumarin-Naphthalene Diimide Dyad. *Chemistry* **21,** 17596–600 (2015).

38. Doria, F. *et al.* A naphthalene diimide dyad for fluorescence switch-on detection of G-quadruplexes. *Chem. Commun.* **51,** 9105–9108 (2015).

39. Shi, D.-F., Wheelhouse, R. T., Sun, D. & Hurley, L. H. Quadruplex-Interactive Agents as Telomerase Inhibitors: Synthesis of Porphyrins and Structure−Activity Relationship for the Inhibition of Telomerase. *J. Med. Chem.* **44,** 4509–4523 (2001).

40. 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).

41. Phan, A. T., Kuryavyi, V., Gaw, H. Y. & Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* **1,** 167–73 (2005).

42. Han, H., Langley, D. R., Rangan, A. & Hurley, L. H. Selective Interactions of Cationic Porphyrins with G-Quadruplex Structures. *J. Am. Chem. Soc.* **123,** 8902–8913 (2001).

43. Gavathiotis, E., Heald, R. A., Stevens, M. F. G. & Searle, M. S. Drug recognition and stabilisation of the parallel-stranded DNA quadruplex d(TTAGGGT)4 containing the human telomeric repeat. *J. Mol. Biol.* **334,** 25–36 (2003).

44. Salvati, E. *et al.* Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J. Clin. Invest.* **117,** 3236–3247 (2007).

45. Jiang, Y.-L. & Liu, Z.-P. Metallo-organic G-quadruplex ligands in anticancer drug design. *Mini Rev. Med. Chem.* **10,** 726–36 (2010).

46. Dixon, I. M. *et al.* Porphyrin Derivatives for Telomere Binding and Telomerase Inhibition. *ChemBioChem* **6,** 123–132 (2005).

47. Gomez, D. *et al.* The G-quadruplex Ligand Telomestatin Inhibits POT1 Binding to Telomeric Sequences *In vitro* and Induces GFP-POT1 Dissociation from Telomeres in Human Cells. *Cancer Res.* **66,** 6908–6912 (2006).

48. Gomez, D. *et al.* Interaction of Telomestatin with the Telomeric Single-strand Overhang. *J. Biol. Chem.* **279,** 41487–41494 (2004).

49. Iida, K. & Nagasawa, K. Macrocyclic Polyoxazoles as G-Quadruplex Ligands. *Chem. Rec.* **13,** 539–548 (2013).

50. König, S. L. B., Evans, A. C. & Huppert, J. L. Seven essential questions on G-quadruplexes. *Biomol. Concepts* **1,** 197–213 (2010).

51. Frees, S., Menendez, C., Crum, M. & Bagga, P. S. QGRS-Conserve: a computational method for discovering evolutionarily conserved G-quadruplex motifs. *Hum. Genomics* **8,** 8 (2014).

52. Maizels, N. & Gray, L. T. The G4 genome. *PLoS Genet.* **9,** e1003468 (2013).

53. Cayrou, C., Grégoire, D., Coulombe, P., Danis, E. & Méchali, M. Genome-scale identification of active DNA replication origins. *Methods* **57,** 158–64 (2012).

54. Cayrou, C. *et al.* Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* **21,** 1438–49 (2011).

55. 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).

56. De, S. & Michor, F. DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat. Struct. Mol. Biol.* **18,** 950–955 (2011).

57. Hewitt, G. *et al.* Telomeres are favoured targets of a persistent DNA damage response in ageing and stress-induced senescence. *Nat. Commun.* **3,** 708 (2012).

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

59. Beaudoin, J.-D. & Perreault, J.-P. Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* **41,** 5898–5911 (2013).

60. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43,** 8627–37 (2015).

61. Hardin, C. C., Watson, T., Corregan, M. & Bailey, C. Cation-dependent transition between the

quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry* **31,** 833–41 (1992).

62. Li, W., Wu, P., Ohmichi, T. & Sugimoto, N. Characterization and thermodynamic properties of quadruplex/duplex competition. *FEBS Lett.* **526,** 77–81 (2002).

63. Miyoshi, D., Karimata, H. & Sugimoto, N. Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions. *J. Am. Chem. Soc.* **128,** 7957–7963 (2006).

64. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* **19,** 414–422 (2009).

65. Blackburn, E. H. Switching and signaling at the telomere. *Cell* **106,** 661–73 (2001).

66. Stewart, S. A. & Weinberg, R. A. Telomeres: Cancer to Human Aging. *Annu. Rev. Cell Dev. Biol.* **22,** 531–557 (2006).

67. Chai, W., Shay, J. W. & Wright, W. E. Human Telomeres Maintain Their Overhang Length at Senescence. *Mol. Cell. Biol.* **25,** 2158–2168 (2005).

68. Sedivy, J. M. Telomeres Limit Cancer Growth by Inducing Senescence: Long-Sought In Vivo Evidence Obtained. *Cancer Cell* **11,** 389–391 (2007).

69. Feldser, D. M. & Greider, C. W. Short telomeres limit tumor progression in vivo by inducing senescence. *Cancer Cell* **11,** 461–9 (2007).

70. Cosme-Blanco, W. *et al.* Telomere dysfunction suppresses spontaneous tumorigenesis in vivo by initiating p53-dependent cellular senescence. *EMBO Rep.* **8,** 497–503 (2007).

71. Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* **33,** 787–791 (1997).

72. Hahn, W. C. Telomere and telomerase dynamics in human cells. *Curr. Mol. Med.* **5,** 227–31 (2005).

73. Martínez, P. & Blasco, M. A. Role of shelterin in cancer and aging. *Aging Cell* **9,** 653–666 (2010).

74. Ying, L., Green, J. J., Li, H., Klenerman, D. & Balasubramanian, S. Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 14629–34 (2003).

75. Parkinson, G. N., Lee, M. P. H. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417,** 876–80 (2002).

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

77. Phan, A. T., Kuryavyi, V., Luu, K. N. & Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Res.* **35,** 6517–25 (2007).

78. Phan, A. T. & Patel, D. J. Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.* **125,** 15021–7 (2003).

79. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **12,** 847–54 (2005).

80. Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8572–7 (2001).

81. Cong, Y.-S., Wright, W. E. & Shay, J. W. Human telomerase and its regulation. *Microbiol. Mol. Biol. Rev.* **66,** 407–25, table of contents (2002).

82. Cuesta, J., Read, M. A. & Neidle, S. The design of G-quadruplex ligands as telomerase inhibitors. *Mini Rev. Med. Chem.* **3,** 11–21 (2003).

83. Guittat, L. *et al.* Targeting human telomerase for cancer therapeutics. *Cytotechnology* **45,** 75–90 (2004).

84. Kerwin, S. M. G-Quadruplex DNA as a target for drug design. *Curr. Pharm. Des.* **6,** 441–78 (2000).

85. Pendino, F. *et al.* Telomeres and telomerase: Pharmacological targets for new anticancer strategies? *Curr. Cancer Drug Targets* **6,** 147–80 (2006).

86. 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).

87. Lopes, J. *et al.* G-quadruplex-induced instability during leading-strand replication. *EMBO J.* **30,** 4033–46 (2011).

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

89. London, T. B. C. *et al.* FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.* **283,** 36132–9 (2008).

90. Piazza, A. *et al.* Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acids Res.* **38,** 4337–48 (2010).

91. Paeschke, K. *et al.* Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497,** 458–62 (2013).

92. Paeschke, K., Capra, J. A. & Zakian, V. A. DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces cerevisiae Pif1 DNA Helicase. *Cell* **145,** 678–691 (2011).

93. León-Ortiz, A. M., Svendsen, J. & Boulton, S. J. Metabolism of DNA secondary structures at the

eukaryotic replication fork. *DNA Repair (Amst).* **19,** 152–162 (2014).

94. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34,** 3887–3896 (2006).

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

96. Mullen, M. A. *et al.* RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles. *Nucleic Acids Res.* **38,** 8149–63 (2010).

97. Rawal, P. *et al.* Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Res.* **16,** 644–655 (2006).

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

99. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **277,** 3459–69 (2010).

100. 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).

101. Sun, D., Guo, K., Rusche, J. J. & Hurley, L. H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **33,** 6070–80 (2005).

102. 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,** 5404–15 (2006).

103. Rankin, S. *et al.* Putative DNA Quadruplex Formation within the Human *c-kit* Oncogene. *J. Am. Chem. Soc.* **127,** 10584–10589 (2005).

104. Palumbo, S. L., Ebbinghaus, S. W. & Hurley, L. H. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc.* **131,** 10878–91 (2009).

105. Qin, Y. & Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90,** 1149–71 (2008).

106. Simonsson, T., Pecinka, P. & Kubista, M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* **26,** 1167–72 (1998).

107. Simonsson, T. & Sjöback, R. DNA tetraplex formation studied with fluorescence resonance energy transfer. *J. Biol. Chem.* **274,** 17379–83 (1999).

108. Seenisamy, J. *et al.* The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* **126,** 8702–9 (2004).

109. Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44,** 2048–58 (2005).

110. Phan, A. T., Modi, Y. S. & Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human *c-myc* Promoter. *J. Am. Chem. Soc.* **126,** 8710–8716 (2004).

111. Kettani, A. *et al.* A dimeric DNA interface stabilized by stacked A.(G.G.G.G).A hexads and coordinated monovalent cations. *J. Mol. Biol.* **297,** 627–44 (2000).

112. Dang, C. V. *et al.* The c-Myc target gene network. *Semin. Cancer Biol.* **16,** 253–264 (2006).

113. Dang, C. V *et al.* Function of the c-Myc oncogenic transcription factor. *Exp. Cell Res.* **253,** 63–77 (1999).

114. Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E. & Tulinsky, A. The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* **268,** 17651–4 (1993).

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

116. Grand, C. L. *et al.* Retraction for Grand et al., Mutations in the G-quadruplex silencer element and their relationship to c-MYC overexpression, NM23 repression, and therapeutic rescue, PNAS 2004 101:6140-6145. *Proc. Natl. Acad. Sci.* **102,** 516–516 (2005).

117. Ou, T.-M. *et al.* Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. *J. Med. Chem.* **50,** 1465–74 (2007).

118. Yin, F., Liu, J., Deng, X. & Wang, J. Effects of Triethylene Tetraamine on the G-quadruplex Structure in the Human c-myc Promoter. *J. Biochem.* **141,** 669–674 (2007).

119. Liu, J.-N. *et al.* Inhibition of myc promoter and telomerase activity and induction of delayed apoptosis by SYUIQ-5, a novel G-quadruplex interactive agent in leukemia cells. *Leukemia* **21,** 1300–2 (2007).

120. Verma, A. *et al.* Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved *cis* -Regulatory Elements in Human and Related Species. *J. Med. Chem.* **51,** 5641–5649 (2008).

121. 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).

122. Beaume, N. *et al.* Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of D. radiodurans involves G4 DNA-mediated regulation. *Nucleic Acids Res.* **41,** 76–89 (2013).

123. Wieland, M. & Hartig, J. S. Investigation of mRNA quadruplex formation in Escherichia coli. *Nat. Protoc.* **4,** 1632–40 (2009).

124. Perrone, R. *et al.* Mapping and characterization of G-quadruplexes in Mycobacterium

tuberculosis gene promoter regions. *Sci. Rep.* **7,** 5743 (2017).

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

126. 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,** 123–31 (2015).

127. 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).

128. Madireddy, A. *et al.* G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. *Nucleic Acids Res.* **44,** 3675–94 (2016).

129. Murat, P. *et al.* G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* **10,** 358–64 (2014).

130. 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).

131. Lista, M. J. *et al.* Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. *Nat. Commun.* **8,** 16043 (2017).

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

133. Tlučková, K. *et al.* Human papillomavirus G-quadruplexes. *Biochemistry* **52,** 7207–16 (2013).

134. Métifiot, M., Amrane, S., Litvak, S. & Andreola, M.-L. G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res.* **42,** 12352–66 (2014).

135. Tan, J. *et al.* The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.* **5,** e1000428 (2009).

136. Wang, S.-R. *et al.* A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci. Adv.* **2,** e1501535 (2016).

137. Fleming, A. M., Ding, Y., Alenko, A. & Burrows, C. J. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS Infect. Dis.* **2,** 674–681 (2016).

138. Wang, S.-R. *et al.* Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell Chem. Biol.* **23,** 1113–22 (2016).

139. Satkunanathan, S., Thorpe, R. & Zhao, Y. The function of DNA binding protein nucleophosmin in AAV replication. *Virology* **510,** 46–54 (2017).

140. Biswas, B., Kandpal, M. & Vivekanandan, P. A G-quadruplex motif in an envelope gene promoter regulates transcription and virion secretion in HBV genotype B. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx823

141. Balasubramanian, S. & Neidle, S. G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **13,** 345–53 (2009).

142. 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).

143. Gallo, R. C. *et al.* Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220,** 865–7 (1983).

144. Barré-Sinoussi, F., Ross, A. L. & Delfraissy, J.-F. Past, present and future: 30 years of HIV research. *Nat. Rev. Microbiol.* **11,** 877–883 (2013).

145. Keele, B. F. *et al.* Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1. *Science (80-.).* **313,** 523–526 (2006).

146. de Medina, M. *et al.* Serological evidence for HIV infection in Cuban immigrants in 1980. *Lancet (London, England)* **2,** 166 (1987).

147. UNAIDS. Global AIDS uptade, techincal report. (2017).

148. Richman, D. D. *et al.* The Challenge of Finding a Cure for HIV Infection. *Science (80-.).* **323,** 1304–1307 (2009).

149. Barouch, D. H. Challenges in the development of an HIV-1 vaccine. *Nature* **455,** 613–619 (2008).

150. Butler, I. F., Pandrea, I., Marx, P. A. & Apetrei, C. HIV genetic diversity: biological and public health consequences. *Curr. HIV Res.* **5,** 23–45 (2007).

151. de Silva, T. I., Cotten, M. & Rowland-Jones, S. L. HIV-2: the forgotten AIDS virus. *Trends Microbiol.* **16,** 588–595 (2008).

152. Whittle, H. *et al.* HIV-2-infected patients survive longer than HIV-1-infected patients. *AIDS* **8,** 1617–20 (1994).

153. Buonaguro, L., Tornesello, M. L. & Buonaguro, F. M. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J. Virol.* **81,** 10209–19 (2007).

154. Kantor, R. *et al.* Impact of HIV-1 Subtype and Antiretroviral Therapy on Protease and Reverse Transcriptase Genotype: Results of a Global Collaboration. *PLoS Med.* **2,** e112 (2005).

155. Briggs, J. A. G. & Kräusslich, H.-G. The Molecular Architecture of HIV. *J. Mol. Biol.* **410,** 491–500 (2011).

156. Steckbeck, J. D., Kuhlmann, A.-S. & Montelaro, R. C. C-terminal tail of human immunodeficiency virus gp41: functionally rich and structurally enigmatic. *J. Gen. Virol.* **94,** 1–19 (2013).

157. Frankel, A. D. & Young, J. A. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67,** 1–25 (1998).

158. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460,** 711–716 (2009).

159. Capon, D. J. & Ward, R. H. R. The CD4-gpl20 Interaction and Aids Pathogenesis. *Annu. Rev. Immunol.* **9,** 649–678 (1991).

160. Checkley, M. A., Luttge, B. G. & Freed, E. O. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J. Mol. Biol.* **410,** 582–608 (2011).

161. Jones, K. A. HIV trans-activation and transcription control mechanisms. *New Biol.* **1,** 127–35 (1989).

162. Roy, S., Delling, U., Chen, C. H., Rosen, C. A. & Sonenberg, N. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* **4,** 1365–73 (1990).

163. Feinberg, M. B., Baltimore, D. & Frankel, A. D. The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 4045–9 (1991).

164. Kim, S. Y., Byrn, R., Groopman, J. & Baltimore, D. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J. Virol.* **63,** 3708–13 (1989).

165. Strebel, K. Virus-host interactions: role of HIV proteins Vif, Tat, and Rev. *AIDS* **17 Suppl 4,** S25-34 (2003).

166. Schwartz, S., Felber, B. K., Fenyö, E. M. & Pavlakis, G. N. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J. Virol.* **64,** 5448–56 (1990).

167. Schubert, U. *et al.* The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains. *J. Virol.* **70,** 809–19 (1996).

168. Kerkau, T. *et al.* The human immunodeficiency virus type 1 (HIV-1) Vpu protein interferes with an early step in the biosynthesis of major histocompatibility complex (MHC) class I molecules. *J. Exp. Med.* **185,** 1295–305 (1997).

169. Strebel, K. *et al.* The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328,** 728–30 (1987).

170. Fan, L. & Peden, K. Cell-free transmission of Vif mutants of HIV-1. *Virology* **190,** 19–29 (1992).

171. Balliet, J. W. *et al.* Distinct effects in primary macrophages and lymphocytes of the human immunodeficiency virus type 1 accessory genes vpr, vpu, and nef: mutational analysis of a primary HIV-1 isolate. *Virology* **200,** 623–31 (1994).

172. Cohen, E. A., Dehni, G., Sodroski, J. G. & Haseltine, W. A. Human immunodeficiency virus vpr product is a virion-associated regulatory protein. *J. Virol.* **64,** 3097–9 (1990).

173. Heinzinger, N. K. *et al.* The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 7311–5 (1994).

174. Cohen, E. A., Subbramanian, R. A. & Göttlinger, H. G. Role of auxiliary proteins in retroviral morphogenesis. *Curr. Top. Microbiol. Immunol.* **214,** 219–35 (1996).

175. Mansky, L. M. The Mutation Rate of Human Immunodeficiency Virus Type 1 Is Influenced by thevprGene. *Virology* **222,** 391–400 (1996).

176. Emerman, M. HIV-1, Vpr and the cell cycle. *Curr. Biol.* **6,** 1096–103 (1996).

177. Miller, M. D., Warmerdam, M. T., Gaston, I., Greene, W. C. & Feinberg, M. B. The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. *J. Exp. Med.* **179,** 101–13 (1994).

178. Das, S. R. & Jameel, S. Biology of the HIV Nef protein. *Indian J. Med. Res.* **121,** 315–32 (2005).

179. Salvi, R. *et al.* Grossly defective nef gene sequences in a human immunodeficiency virus type 1-seropositive long-term nonprogressor. *J. Virol.* **72,** 3646–57 (1998).

180. Richter, S. N., Frasson, I. & Palù, G. Strategies for inhibiting function of HIV-1 accessory proteins: a necessary route to AIDS therapy? *Curr. Med. Chem.* **16,** 267–86 (2009).

181. Göttlinger, H. G., Sodroski, J. G. & Haseltine, W. A. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 5781–5 (1989).

182. Gallay, P., Swingler, S., Song, J., Bushman, F. & Trono, D. HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell* **83,** 569–76 (1995).

183. Franke, E. K., Yuan, H. E. H. & Luban, J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **372,** 359–362 (1994).

184. Thali, M. *et al.* Functional association of cyclophilin A with HIV-1 virions. *Nature* **372,** 363–5 (1994).

185. Franke, E. K. & Luban, J. Inhibition of HIV-1 replication by cyclosporine A or related compounds correlates with the ability to disrupt the Gag-cyclophilin A interaction. *Virology* **222,** 279–82 (1996).

186. Harrison, G. P. & Lever, A. M. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.* **66,** 4144–53 (1992).

187. Lapadat-Tapolsky, M. *et al.* Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res.* **21,** 831–9 (1993).

188. Paxton, W., Connor, R. I. & Landau, N. R. Incorporation of Vpr into human immunodeficiency

virus type 1 virions: requirement for the p6 region of gag and mutational analysis. *J. Virol.* **67,** 7229–37 (1993).

189. Demirov, D. G., Orenstein, J. M. & Freed, E. O. The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner. *J. Virol.* **76,** 105–17 (2002).

190. Parkin, N. T., Chamorro, M. & Varmus, H. E. Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J. Virol.* **66,** 5147–51 (1992).

191. Ashorn, P. *et al.* An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 7472–6 (1990).

192. Sarafianos, S. G. *et al.* Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition. *J. Mol. Biol.* **385,** 693–713 (2009).

193. Bushman, F. D., Fujiwara, T. & Craigie, R. Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249,** 1555–8 (1990).

194. Mohammadi, P. *et al.* 24 Hours in the Life of HIV-1 in a T Cell Line. *PLoS Pathog.* **9,** e1003161 (2013).

195. Holmes, M., Zhang, F. & Bieniasz, P. D. Single-Cell and Single-Cycle Analysis of HIV-1 Replication. *PLOS Pathog.* **11,** e1004961 (2015).

196. Fanales-Belasio, E., Raimondo, M., Suligoi, B. & Buttò, S. HIV virology and pathogenetic mechanisms of infection: a brief overview. *Ann. Ist. Super. Sanita* **46,** 5–14 (2010).

197. Rambaut, A., Posada, D., Crandall, K. A. & Holmes, E. C. The causes and consequences of HIV evolution. *Nat. Rev. Genet.* **5,** 52–61 (2004).

198. Zhu, P. *et al.* Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441,** 847–52 (2006).

199. Liu, J., Bartesaghi, A., Borgnia, M. J., Sapiro, G. & Subramaniam, S. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455,** 109–113 (2008).

200. Rizzuto, C. D. *et al.* A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* **280,** 1949–53 (1998).

201. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393,** 648–59 (1998).

202. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89,** 263–73 (1997).

203. Buzon, V. *et al.* Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog.* **6,** e1000880 (2010).

204. Alkhatib, G. & Berger, E. A. HIV coreceptors: from discovery and designation to new

paradigms and promise. *Eur. J. Med. Res.* **12,** 375–84 (2007).

205. Coakley, E., Petropoulos, C. J. & Whitcomb, J. M. Assessing chemokine co-receptor usage in HIV. *Curr. Opin. Infect. Dis.* **18,** 9–15 (2005).

206. Arhel, N. Revisiting HIV-1 uncoating. *Retrovirology* **7,** 96 (2010).

207. Mougel, M., Houzet, L. & Darlix, J.-L. When is it time for reverse transcription to start and go? *Retrovirology* **6,** 24 (2009).

208. Hu, W.-S. & Hughes, S. H. HIV-1 Reverse Transcription. *Cold Spring Harb. Perspect. Med.* **2,** a006882–a006882 (2012).

209. Piller, S. C., Caly, L. & Jans, D. A. Nuclear import of the pre-integration complex (PIC): the Achilles heel of HIV? *Curr. Drug Targets* **4,** 409–29 (2003).

210. Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387,** 183–8 (1997).

211. de Arellano, E. R., Alcamí, J., López, M., Soriano, V. & Holguín, A. Drastic decrease of transcription activity due to hypermutated long terminal repeat (LTR) region in different HIV-1 subtypes and recombinants. *Antiviral Res.* **88,** 152–9 (2010).

212. Michael, N. L., D'Arcy, L., Ehrenberg, P. K. & Redfield, R. R. Naturally occurring genotypes of the human immunodeficiency virus type 1 long terminal repeat display a wide range of basal and Tat-induced transcriptional activities. *J. Virol.* **68,** 3163–74 (1994).

213. Jeeninga, R. E. *et al.* Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *J. Virol.* **74,** 3740–51 (2000).

214. Kilareski, E. M., Shah, S., Nonnemacher, M. R. & Wigdahl, B. Regulation of HIV-1 transcription in cells of the monocyte-macrophage lineage. *Retrovirology* **6,** 118 (2009).

215. Coiras, M., López-Huertas, M. R., Sánchez del Cojo, M., Mateos, E. & Alcamí, J. Dual role of host cell factors in HIV-1 replication: restriction and enhancement of the viral cycle. *AIDS Rev.* **12,** 103–12

216. Imai, K. *et al.* Cyclin T1 stabilizes expression levels of HIV-1 Tat in cells. *FEBS J.* **276,** 7124–33 (2009).

217. Peterlin, B. M. & Trono, D. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nat. Rev. Immunol.* **3,** 97–107 (2003).

218. Hallenberger, S. *et al.* Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* **360,** 358–61 (1992).

219. Fanales-Belasio, E., Raimondo, M., Suligoi, B. & Buttò, S. HIV virology and pathogenetic mechanisms of infection: a brief overview. *Ann. Ist. Super. Sanita* **46,** 5–14 (2010).

220. Palmisano, L. & Vella, S. A brief history of antiretroviral therapy of HIV infection: success and challenges. *doi.org* **47,** 44–8 (2011).

221. Autran, B. *et al.* Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. *Science* **277,** 112–6 (1997).

222. Arts, E. J. & Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2,** a007161 (2012).

223. Clavel, F. & Hance, A. J. HIV Drug Resistance. *N. Engl. J. Med.* **350,** 1023–1035 (2004).

224. Santoro, M. M. & Perno, C. F. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiol.* **2013,** 481314 (2013).

225. Chun, T. W. & Fauci, A. S. Latent reservoirs of HIV: obstacles to the eradication of virus. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 10958–61 (1999).

226. Chun, T. W., Davey, R. T., Engel, D., Lane, H. C. & Fauci, A. S. Re-emergence of HIV after stopping therapy. *Nature* **401,** 874–5 (1999).

227. Sundquist, W. I. & Heaphy, S. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 3393–7 (1993).

228. Lyonnais, S. *et al.* G-quartets assembly within a G-rich DNA flap. A possible event at the center of the HIV-1 genome. *Nucleic Acids Res.* **30,** 5276–83 (2002).

229. Lyonnais, S., Gorelick, R. J., Mergny, J.-L., Le Cam, E. & Mirambeau, G. G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA. *Nucleic Acids Res.* **31,** 5754–63 (2003).

230. Shen, W., Gao, L., Balakrishnan, M. & Bambara, R. A. A recombination hot spot in HIV-1 contains guanosine runs that can form a G-quartet structure and promote strand transfer in vitro. *J. Biol. Chem.* **284,** 33883–93 (2009).

231. 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).

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

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

234. 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).

235. Piekna-Przybylska, D., Sharma, G., Maggirwar, S. B. & Bambara, R. A. Deficiency in DNA damage response, a new characteristic of cells infected with latent HIV-1. *Cell Cycle* **16,** 968–978 (2017).

236. Nielsen, P. E. & Egholm, M. An introduction to peptide nucleic acid. *Curr. Issues Mol. Biol.* **1,** 89–104 (1999).

237. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. Sequence specific inhibition of DNA

restriction enzyme cleavage by PNA. *Nucleic Acids Res.* **21,** 197–200 (1993).

238. Demidov, V. V *et al.* Stability of peptide nucleic acids in human serum and cellular extracts. *Biochem. Pharmacol.* **48,** 1310–3 (1994).

239. Panyutin, I. G., Onyshchenko, M. I., Englund, E. A., Appella, D. H. & Neumann, R. D. Targeting DNA G-quadruplex structures with peptide nucleic acids. *Curr. Pharm. Des.* **18,** 1984–91 (2012).

240. Rozners, E. Recent advances in chemical modification of Peptide nucleic acids. *J. Nucleic Acids* **2012,** 518162 (2012).

241. Said Hassane, F., Saleh, A. F., Abes, R., Gait, M. J. & Lebleu, B. Cell penetrating peptides: overview and applications to the delivery of oligonucleotides. *Cell. Mol. Life Sci.* **67,** 715–26 (2010).

242. Abes, S. *et al.* Efficient splicing correction by PNA conjugation to an R6-Penetratin delivery peptide. *Nucleic Acids Res.* **35,** 4495–502 (2007).

243. Fang, H., Zhang, K., Shen, G., Wooley, K. L. & Taylor, J.-S. A. Cationic shell-cross-linked knedel-like (cSCK) nanoparticles for highly efficient PNA delivery. *Mol. Pharm.* **6,** 615–26 (2009).

244. Hu, J. & Corey, D. R. Inhibiting gene expression with peptide nucleic acid (PNA)--peptide conjugates that target chromosomal DNA. *Biochemistry* **46,** 7581–9 (2007).

245. Fabani, M. M. *et al.* Efficient inhibition of miR-155 function in vivo by peptide nucleic acids. *Nucleic Acids Res.* **38,** 4466–75 (2010).

246. Mitchell, D. J., Kim, D. T., Steinman, L., Fathman, C. G. & Rothbard, J. B. Polyarginine enters cells more efficiently than other polycationic homopolymers. *J. Pept. Res.* **56,** 318–25 (2000).

247. Fuchs, S. M. & Raines, R. T. Pathway for polyarginine entry into mammalian cells. *Biochemistry* **43,** 2438–44 (2004).

248. Fischer, R., Köhler, K., Fotin-Mleczek, M. & Brock, R. A stepwise dissection of the intracellular fate of cationic cell-penetrating peptides. *J. Biol. Chem.* **279,** 12625–35 (2004).

249. Zhou, P. *et al.* Novel binding and efficient cellular uptake of guanidine-based peptide nucleic acids (GPNA). *J. Am. Chem. Soc.* **125,** 6878–9 (2003).

250. Sahu, B. *et al.* Synthesis of conformationally preorganized and cell-permeable guanidine-based gamma-peptide nucleic acids (gammaGPNAs). *J. Org. Chem.* **74,** 1509–16 (2009).

251. Manicardi, A. *et al.* Cellular uptakes, biostabilities and anti-miR-210 activities of chiral arginine-PNAs in leukaemic K562 cells. *Chembiochem* **13,** 1327–37 (2012).

252. Dragulescu-Andrasi, A. *et al.* A simple gamma-backbone modification preorganizes peptide nucleic acid into a helical structure. *J. Am. Chem. Soc.* **128,** 10258–67 (2006).

253. Zanardi, C. *et al.* Peptide nucleic acids tagged with four lysine residues for amperometric genosensors. *Artif. DNA. PNA XNA* **3,** 80–7 (2012).

254. Hamzavi, R., Dolle, F., Tavitian, B., Dahl, O. & Nielsen, P. E. Modulation of the pharmacokinetic properties of PNA: preparation of galactosyl, mannosyl, fucosyl, N-acetylgalactosaminyl, and N-acetylglucosaminyl derivatives of aminoethylglycine peptide nucleic acid monomers and their incorporation into PNA oligomers. *Bioconjug. Chem.* **14,** 941–54 (2003).

255. Sazani, P. *et al.* Systemically delivered antisense oligomers upregulate gene expression in mouse tissues. *Nat. Biotechnol.* **20,** 1228–33 (2002).

256. Yin, H., Lu, Q. & Wood, M. Effective exon skipping and restoration of dystrophin expression by peptide nucleic acid antisense oligonucleotides in mdx mice. *Mol. Ther.* **16,** 38–45 (2008).

257. Datta, B. & Armitage, B. A. Hybridization of PNA to structured DNA targets: quadruplex invasion and the overhang effect. *J. Am. Chem. Soc.* **123,** 9612–9 (2001).

258. Roy, S., Tanious, F. A., Wilson, W. D., Ly, D. H. & Armitage, B. A. High-affinity homologous peptide nucleic acid probes for targeting a quadruplex-forming sequence from a MYC promoter element. *Biochemistry* **46,** 10433–43 (2007).

259. Lusvarghi, S. *et al.* Loop and backbone modifications of peptide nucleic acid improve g-quadruplex binding selectivity. *J. Am. Chem. Soc.* **131,** 18415–24 (2009).

260. Paul, A., Sengupta, P., Krishnan, Y. & Ladame, S. Combining G-quadruplex targeting motifs on a single peptide nucleic acid scaffold: a hybrid (3+1) PNA-DNA bimolecular quadruplex. *Chemistry* **14,** 8682–9 (2008).

261. Amato, J. *et al.* Targeting G-quadruplex structure in the human c-Kit promoter with short PNA sequences. *Bioconjug. Chem.* **22,** 654–63 (2011).

262. Amato, J. *et al.* A short C-rich PNA fragment capable to form novel G-quadruplex-PNA complexes. *Nucleic Acids Symp. Ser.* **52,** 167–168 (2008).

263. Amato, J., Oliviero, G., De Pauw, E. & Gabelica, V. Hybridization of short complementary PNAs to G-quadruplex forming oligonucleotides: An electrospray mass spectrometry study. *Biopolymers* **91,** 244–255 (2009).

264. Onyshchenko, M. I. *et al.* Stabilization of G-quadruplex in the BCL2 promoter region in double-stranded DNA by invading short PNAs. *Nucleic Acids Res.* **37,** 7570–80 (2009).

265. Pandey, V. N., Upadhyay, A. & Chaubey, B. Prospects for antisense peptide nucleic acid (PNA) therapies for HIV. *Expert Opin. Biol. Ther.* **9,** 975–89 (2009).

266. Mayhood, T. *et al.* Inhibition of Tat-Mediated Transactivation of HIV-1 LTR Transcription by Polyamide Nucleic Acid Targeted to TAR Hairpin Element †. *Biochemistry* **39,** 11532–11539 (2000).

267. Mehiri, M. *et al.* An efficient biodelivery system for antisense polyamide nucleic acid (PNA). *Oligonucleotides* **18,** 245–56 (2008).

268. Riguet, E. *et al.* A peptide nucleic acid-neamine conjugate that targets and cleaves HIV-1 TAR RNA inhibits viral replication. *J. Med. Chem.* **47,** 4806–9 (2004).

269. Kaushik, N., Basu, A., Palumbo, P., Myers, R. L. & Pandey, V. N. Anti-TAR polyamide nucleotide analog conjugated with a membrane-permeating peptide inhibits human immunodeficiency virus type 1 production. *J. Virol.* **76,** 3881–91 (2002).

270. Das, I. *et al.* A peptide nucleic acid-aminosugar conjugate targeting transactivation response element of HIV-1 RNA genome shows a high bioavailability in human cells and strongly inhibits tat-mediated transactivation of HIV-1 transcription. *J. Med. Chem.* **55,** 6021–32 (2012).

271. Lee, R., Kaushik, N., Modak, M. J., Vinayak, R. & Pandey, V. N. Polyamide nucleic acid targeted to the primer binding site of the HIV-1 RNA genome blocks in vitro HIV-1 reverse transcription. *Biochemistry* **37,** 900–10 (1998).

272. Kaushik, N. & Pandey, V. N. PNA targeting the PBS and A-loop sequences of HIV-1 genome destabilizes packaged tRNA3(Lys) in the virions and inhibits HIV-1 replication. *Virology* **303,** 297–308 (2002).

273. Sei, S. *et al.* Identification of a key target sequence to block human immunodeficiency virus type 1 replication within the gag-pol transframe domain. *J. Virol.* **74,** 4621–33 (2000).

274. Parkash, B. *et al.* Inhibition of 5'-UTR RNA conformational switching in HIV-1 using antisense PNAs. *PLoS One* **7,** e49310 (2012).

275. De Cian, A. *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods* **42,** 183–95 (2007).

276. Rachwal, P. A. & Fox, K. R. Quadruplex melting. *Methods* **43,** 291–301 (2007).

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

278. Team, R. C. *ed Vienna R Foundation for Statistical Computing, Austria*. (2016).

279. Masiero, S. *et al.* A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org. Biomol. Chem.* **8,** 2683–92 (2010).

280. Murat, P., Singh, Y. & Defrancq, E. Methods for investigating G-quadruplex DNA/ligand interactions. *Chem. Soc. Rev.* **40,** 5293–307 (2011).

281. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1,** 2527–2535 (2007).

282. Sun, D. & Hurley, L. H. Biochemical Techniques for the Characterization of G-Quadruplex Structures: EMSA, DMS Footprinting, and DNA Polymerase Stop Assay. in *Methods in molecular biology (Clifton, N.J.)* **608,** 65–79 (2010).

283. Redman, J. E. Surface plasmon resonance for probing quadruplex folding and interactions with proteins and small molecules. *Methods* **43,** 302–312 (2007).

284. Reed, L. J. & Muench, H. A simple method of estimating fifty percent endpoints. *Am. J. Hyg.*

27, 493–497 (1938).

285. Scalabrin, M., Palumbo, M. & Richter, S. N. Highly Improved Electrospray Ionization-Mass Spectrometry Detection of G-Quadruplex-Folded Oligonucleotides and Their Complexes with Small Molecules. *Anal. Chem.* **89,** 8632–8637 (2017).

286. Marchand, A. & Gabelica, V. Native electrospray mass spectrometry of DNA G-quadruplexes in potassium solution. *J. Am. Soc. Mass Spectrom.* **25,** 1146–54 (2014).

287. 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).

288. Gray, R. D. & Chaires, J. B. Analysis of multidimensional G-quadruplex melting curves. *Curr. Protoc. nucleic acid Chem.* **Chapter 17,** Unit17.4 (2011).

289. Jeeninga, R. E., Westerhout, E. M., van Gerven, M. L. & Berkhout, B. HIV-1 latency in actively dividing human T cell lines. *Retrovirology* **5,** 37 (2008).

290. Duverger, A. *et al.* An AP-1 binding site in the enhancer/core element of the HIV-1 promoter controls the ability of HIV-1 to establish latent infection. *J. Virol.* **87,** 2264–77 (2013).

291. Finzi, D. *et al.* Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278,** 1295–300 (1997).

292. Doria, F., Manet, I., Grande, V., Monti, S. & Freccero, M. Water-Soluble Naphthalene Diimides as Singlet Oxygen Sensitizers. *J. Org. Chem.* **78,** 8065–8073 (2013).

293. Cogoi, S. *et al.* Transcription inhibition of oncogenic KRAS by a mutation-selective peptide nucleic acid conjugated to the PKKKRKV nuclear localization signal peptide. *Biochemistry* **44,** 10510–9 (2005).

294. Roepstorff, P. & Fohlman, J. Letter to the editors. *Biol. Mass Spectrom.* **11,** 601–601 (1984).

295. He, Y. & Smith, R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell. Mol. Life Sci.* **66,** 1239–1256 (2009).

296. 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).

297. Mayeda, A., Munroe, S. H., Cáceres, J. F. & Krainer, A. R. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.* **13,** 5483–95 (1994).

298. Khateb, S., Weisman-Shomer, P., Hershco, I., Loeb, L. A. & Fry, M. Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.* **32,** 4145–4154 (2004).

299. Wang, F. *et al.* Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 20413–8 (2012).

300. Gordon, H. *et al.* Depletion of hnRNP A2/B1 overrides the nuclear retention of the HIV-1

genomic RNA. *RNA Biol.* **10,** 1714–25 (2013).

301. Lévesque, K. *et al.* Trafficking of HIV-1 RNA is mediated by heterogeneous nuclear ribonucleoprotein A2 expression and impacts on viral assembly. *Traffic* **7,** 1177–93 (2006).

302. Jablonski, J. A. & Caputi, M. Role of Cellular RNA Processing Factors in Human Immunodeficiency Virus Type 1 mRNA Metabolism, Replication, and Infectivity. *J. Virol.* **83,** 981–992 (2009).