UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Medicina Molecolare

CORSO DI DOTTORATO DI RICERCA IN BIOMEDICINA

CURRICOLO MEDICINA MOLECOLARE

CICLO XXIX

G-QUADRUPLEXES IN THE HIV-1 GENOME: STRUCTURE AND TARGETING

Tesi redatta con il contributo finanziario di Bill & Melinda Gates Foundation

Coordinatore: Ch.mo Prof. Stefano Piccolo

Supervisore: Prof.ssa Sara Richter

Dottoranda: Elena Butovskaya

Table of contents

Abstract

Nucleic acids may form non-canonical tetraplex secondary structures called G-quadruplexes. G-quadruplexes have been found in eukaryotic and prokaryotic genomes, including viruses. Located in key functional regions of genomes, G-quadruplexes play important regulatory roles in transcription, replication and translation processes. Particularly, G-quadruplex-mediated transcription regulation of oncogene promoters has been widely described.

Previous studies demonstrated that a set of dynamic G-quadruplex structures in the promoter region of the HIV-1 Long Terminal Repeat (LTR) regulates viral transcription acting as repressor elements. G-quadruplex-directed targeting with stabilizing ligands enhances their inhibitory effect, resulting in decrease of viral production and suggesting viral G4 structures as potential antiviral targets.

We aimed at 1) develop antiviral compounds selective for viral G-quadruplexes over cellular structures. We screened a newly designed series of NDI-core based G-quadruplex ligands and pointed out structural features of the compounds that led to improve the selectivity.

2) We identified by nuclear magnetic resonance the deep structural coordinates of the G-quadruplex targets as the preliminary necessary step for rational drug design approach. We described the singular hybrid quadruplex/duplex topology of the major G-quadruplex component within the LTR region, which allows novel implication for selective recognition of viral structures.

3) We also explored the formation of G-quadruplex structures at the HIV-1 RNA genome level, which emerged as a pre-integration target for the antiviral activity of a well-known G-quadruplex ligand. We investigated the formation, stability and involvement of RNA G-quadruplexes in the reverse transcription process and the role of the HIV-1 nucleocapsid protein in controlling folding of these structures.

Riassunto

Acidi nucleici ricchi in guanine possono formare strutture secondarie alternative chiamate G-quadruplex. I G-quadruplex sono stati caratterizzati in diversi tipi di genomi, tra cui genomi virali e umano. Nel genoma umano le strutture G-quadruplex sono prinicipalmente localizzate in importanti regioni funzionali, dove possono assumere ruoli regolatori dei processi come trascrizione, replicazione a traduzione. In particolare, la regolazione della trascrizione dei promotori degli oncogeni mediata dalle strutture G-quadruplex è stata ampiamente descritta, evidenziando che i G-quadruplex promotoriali agiscono principalmente da silenziatori del processo trascrizionale.

I precedenti studi, condotti dal nostro gruppo di ricerca hanno dimostrato che una serie di strutture G-quadruplex al livello del Long Terminal repeat (LTR) del genoma provirale di HIV-1 è coinvolta nella regolazione della trascrizione virale. La stabilizzazione di queste strutture con i ligandi specifici si traduce in repressione dell'attività promotoriale e in inibizione della produzione del virus in cellule infettate, suggerendo che i G-quadruplex nella regione LTR di HIV-1 possono essere dei promettenti target antivirali. Lo scopo principale di questa tesi è stato quello di individuare composti con attività antivirale che mostrano un legame preferenziale verso le strutture G-quadruplex virali. Abbiamo testato una nuova serie di composti leganti G-quadruplex sviluppata a partire dal NDI-core e abbiamo identificato dei componenti strutturali responsabili della maggiore affinità verso le strutture G-quadruplex virali che possono guidare verso ulteriore miglioramento della selettività.

L'obiettivo di utilizzare un approccio razionale per lo sviluppo dei composti selettivi ha richiesto di individuare le coordinate strutturali del target. Abbiamo identificato che il componente G-quadruplex prevalente nella regione LTR considerata è foldato in una topologia molto particolare, descritta come struttura ibrida quadruplex/duplex e presenta interessanti implicazioni per il riconoscimento selettivo da parte delle piccole molecole.

1 Introduction

1.1 DNA G-quadruplexes

The canonical DNA secondary structure is the right-handed double helix, described by Watson and Crick more than 50 years ago, and known as B-DNA form. Driven by different environment conditions and strand composition DNA may adopt alternative secondary conformations, such as A-DNA and Z-DNA forms, triplex and tetraplex forms (Figure 1.1). Non-B-DNA secondary structures participate in biologically important processes and have been associated to genomic instability by inducing genetic expansion and deletions, DNA strand breaks and rearrangements (1,2).

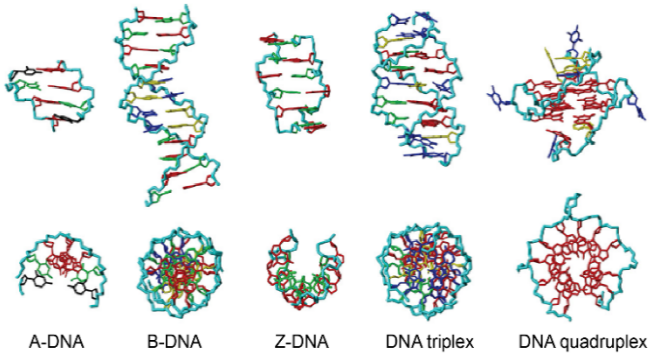


Figure 1.1 Representation of different DNA secondary structures. Figure from (19)

G-quadruplexes (G4s) are formed by single strand guanine rich sequences, where the guanines are interacting through Hoogsteen hydrogen bonds, in contrast to the canonical Watson-Crick base-pairing. First evidences of G4 formation are dated back to 1962, when Gellert and coworkers, proposed that guanylic acids form four-stranded helical structures (3) In early 80s, Sundquist et al. reported the *in vitro* formation of guanine tetrads in telomeric repeats (4). Since then the interest in these alternative secondary structures has kept growing and much evidence, suggesting their important biological roles has been collected.

1.2 DNA G-quadruplexes: general features

As anticipated, guanine-rich nucleic acids may form a non-canonical secondary structure called G4. The building block of a G4 structure is a G-tetrad, where four guanines in a planar arrangement interact through Hoogsteen hydrogen bonds (5,6) (Figure 1.2). Stacking of two or more G-tetrads gives rise to a G4 structure (5,6).

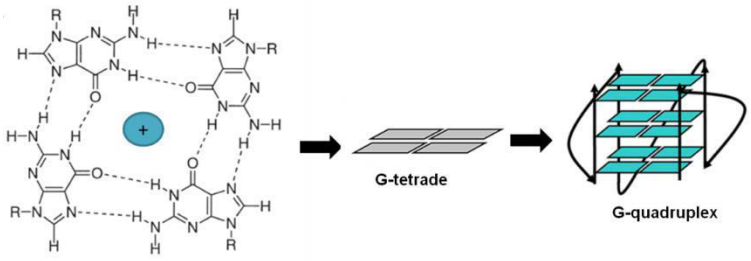


Figure 1.2 Four guanines interacting through Hoogsteen hydrogen bonds. Stacking of three G-tetrads forms a three layered G4 structure.

Cations play an important role in the stabilization of G4 structures. The oxygens (O6) of the guanines define a negatively charged environment in the center of the tetrad creating the channel which can accommodate different types of cations. Molecular dynamics simulation studies suggested that the absence of the coordinating cations highly destabilize the structure by leaving electronically unfavorable environment (7).

NMR studies and crystal structures of G4s revealed the incorporation of positively charged ions, both monovalent and bivalent, such as sodium (8), potassium (9), calcium (10), ammonium (11) and magnesium (12) ions. Particularly, physiological concentrations of potassium (K^+) and sodium (Na^+) were found to greatly stabilize G4s. The positon of the cation in the channel depends on the nature of the cation. More precisely, the atomic radius defines the way of coordination with the oxygens and the location relatively to the tetrads plane. Sodium ions have the atomic radius of 1.02 Å and planar coordination geometry, thus they are placed on the same plane of the tetrads while potassium ions, having larger atomic radius (1.38 Å) and bipiramidal coordination geometry, are distributed between the tetrads coordinating eight oxygens (13) (Figure1.3). The number of coordinating interactions of potassium cations gives to the entire structure increased thermodynamic stability. The presence of a particular cation can also influence the topology of the structure. A study from Smargiasso *et al*. suggested that potassium favors parallel structures, while sodium strongly induces antiparallel conformations (14). This is attributed to the coordination modes described above. The conformational diversity of telomeric G4s in sodium or potassium solutions has been reported (8,15,16).

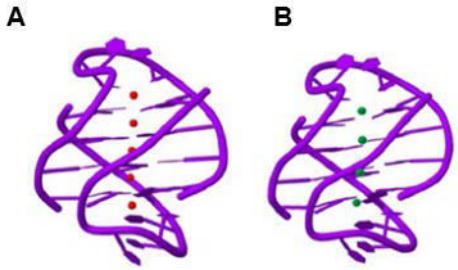


Figure 1.3 Crystal structures of *Oxytricha nova* telomeric DNA in presence of different monovalent cations. Side views of crystal structures evidence the position of **(A)** K^+ between the tetrad planes and **(B)** Na^+ on the same plane of the tetrads. Figure from (7)

1.3 Polymorphism of G-quadruplex structures

G4s are highly polymorphic structures. The conformational differences arise from number of strands involved in the structure formation (maximum 4 strands), strands directionality, loops length and composition, and, in some cases, from the nature of stabilizing cation (5,6). Based on the molecularity, G4s can be classified in two main groups:

- intramolecular G4s, formed from a single strand DNA with at least four G-tracts intercalated by loop sequences of different length and composition
- intermolecular G4s, formed from two or four DNA strands with at least two or one G-tracts respectively.

Four-stranded G4s are the topologically simplest structures as they can be formed by a single repeat of at least two guanines and usually do not present loop connectivity (17). Dimeric structures need at least two runs of guanines to form a G4 and they are topologically more complex because of loop connectivities that may assume different conformation, such as diagonal or lateral (17). Therefore, the single strand G4s display the most abundant variety of topologies and theoretically may adopt 26 different conformations (18,19) (Figure1.4A).
Besides the molecularity, further topological variety is given by the strand polarity, resulting in parallel, antiparallel and mixed structures. This classification is based on the glycosidic bond angle conformation. The glycosidic bond angle is referred to as the torsion angle of the bond between the base and the sugar. In *anti* conformation the pyrimidine ring is spatially far from the deoxyribose sugar while in *syn* conformation it is flipped towards the sugar (Figure1.4B). The combination of the glycosidic torsion angles of the guanines in the tetrad defines different types of folding topologies (18).

The parallel topology results from the *anti* conformation of the glycosidic bond angles of all the guanines participating to the tetrads formation, while the antiparallel topologies are the combination of *anti* and *syn* glycosidic bond angle conformations, such as *syn-anti* or *syn-syn*, *anti-anti* steps within the tetrad (18,20).

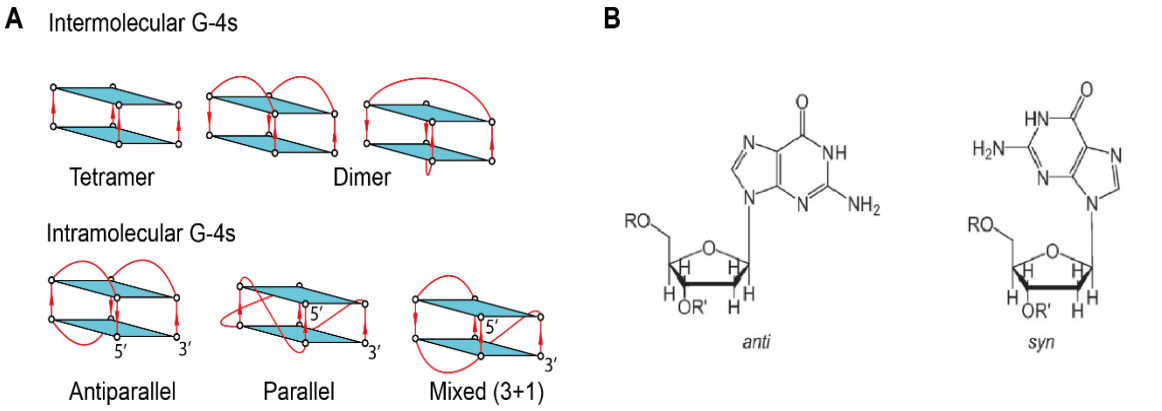


Figure 1.4 (A) Some examples of G4 topologies. The simplest bilayer tetrameric G4 topology is represented on the top. Dimeric structures display examples of lateral and diagonal loops. Intramolecular structures display examples of parallel, antiparallel and mixed topologies with different loop conformation (lateral, propeller and their combination, from the right to the left, respectively). Arrows indicate the strand directionality. **(B)** Chemical structures of *anti* and *syn* guanosines bond angles. Figure **(B)** from (20)

As mentioned above, the loop connectivity gives another contribution to the conformational diversity of parallel and antiparallel structures. Parallel topology requires to connect the bottom G-tetrad to the top G-tetrad resulting in the propeller-type loops. Antiparallel topology allows a greater variety of loop conformations, as adjacent or opposite G-strands can be connected, resulting in lateral or diagonal loops, respectively (17). Combinations of loop progressions through narrow, medium and wide grooves of the structure further define topological varieties (19).

Figure 1.4A displays a summarized representation of topological diversity of G4s structures. Only regular types of G4 topologies are represented here, but there is accumulating structural evidence describing very singular topologies, such as G4s with bulges and with long particularly structured loops (21,22), suggesting the presence of peculiar features that increase the topological variety of G4s .

Conformational polymorphism is a very important feature of G4s since their emerging biological roles make them an attractive therapeutic targets and the selectivity between different structures is a main goal for the design of G4 binding small molecules as will be discussed later.

1.4 Biological role of G-quadruplexes

In the human genome the G/ C content is relatively lower than the A/T content. However, first computational analyses of the human genome revealed the presence of over 300.000 putative G4 forming sequences (PQS) (23). Taken in consideration that the consensus sequence used for analysis ($G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$) is prone to fold into a very stable regular three-tetrad G4s, while structural evidences of peculiar G4s kept accumulating, the underestimation was very likely. Indeed, a recent high-throughput sequencing of human genome performed by Chambers *et al*. identified over 700.000 PQS that include non-canonical G4s with long loop and bulges, previously not considered in computational studies (24). Interestingly, PQS are not distributed randomly within the human genome but clustered in functional regions, such as telomeres, replication origins and gene promoters (24-26). Figure 1.5 represents key cellular mechanisms where G4s may have an important biological role, which will be discussed in details in the following sections.

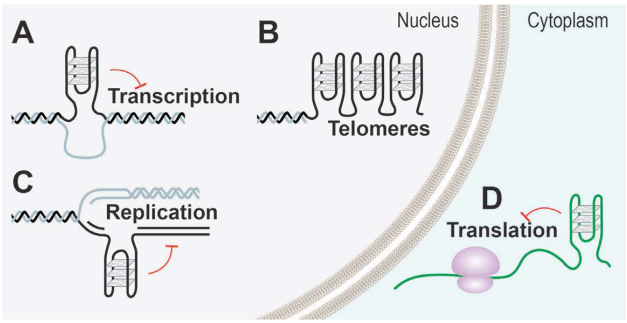


Figure 1.5 Possible distribution of G4s in eukaryotic cells. G4s are not randomly distributed in the human genome, but clustered in functionally important regions as discussed in the main text. In the nucleus DNA G4s mainly occur in the single stranded G-rich 3'-overhangs of telomeres **(B)**, and in the regions transiently becoming single stranded for functional reasons during transcription **(A)** and replication **(C)**. In the cytoplasm, G4s have been described in untranslated regions (UTRs) of mRNA where they are involved in translational regulation. Figure from (26)

1.4.1 G-Quadruplexes in telomeres

Telomeres are nucleoprotein complexes located at the end of eukaryotic chromosomes. In humans, telomeres are composed of tandem repeats of the d(TTAGGG) sequence, partially in the double stranded form (up to 15 kb) that terminates in the single stranded G-rich form at the 3'-overhang (around 200 bases) (27). Telomeres are strongly associated with specific telomere binding proteins, known as shelterin complex. It counts six proteins: TRF1 (telomeric repeat-binding factor 1), TRF2, RAP1 (repressor and activator protein 1), TIN2 (TRF1-interacting nuclear protein 2), POT1 (protection of telomeres 1) and TPP1 (28,29). TRF1, TRF2 and RAP1 bind the telomeric dsDNA portion, while POT1 and TPP1, associated in a ternary complex, bind with high affinity and selectivity the 3' single strand tail (29). The main role of the shelterin complex is to protect the end of chromosomes from DNA repairing

mechanisms as, by their nature, they may be easily recognized as dsDNA damage sites. The binding partnership between these six proteins bridges the formation of the protecting secondary structure, called T-loop, which involves ds and ssDNA portions (Figure 1.6) (30-32).

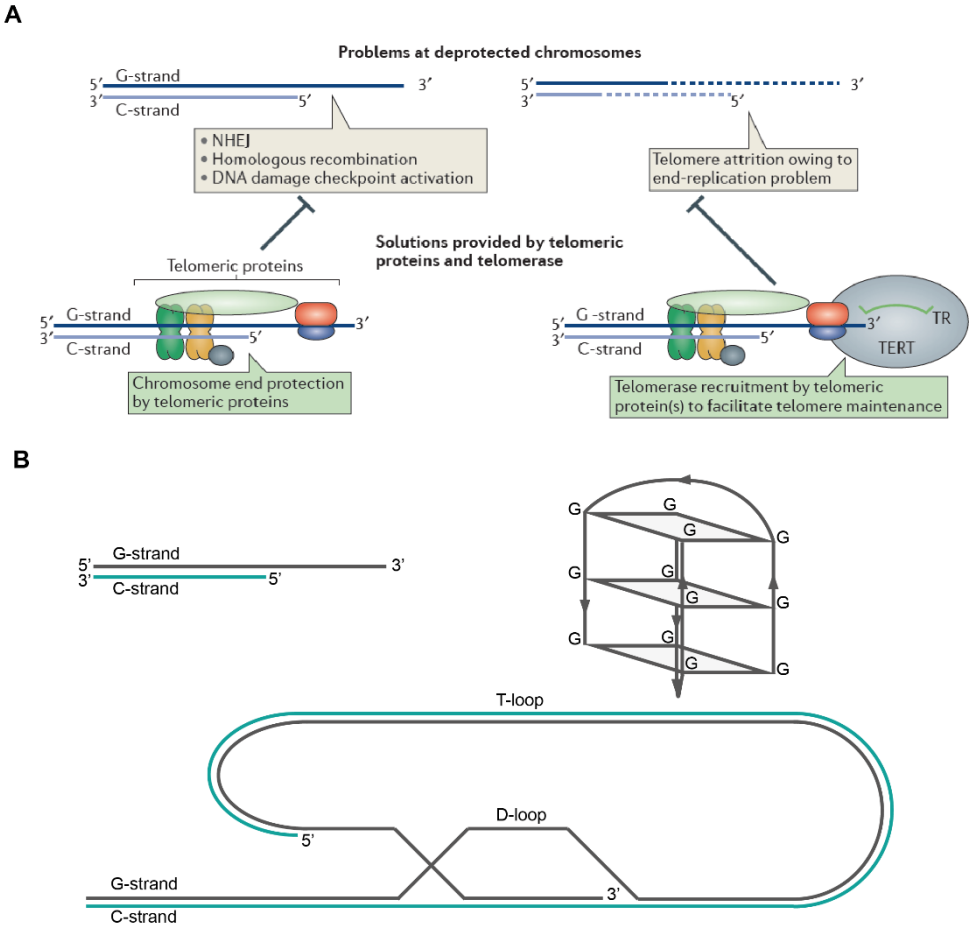


Figure 1.6 Schematic representation of chromosomes end-protection and end-replication problems and solutions provided by shelterin complex and telomerase. **(A)** Due to the single strand nature, chromosomes ends undergo two main problems, known as end-protection problem and end-replication problem. In their linear form, telomeres are recognized as DNA damage points and the activation of repairing mechanisms result in the chromosome end-to-end fusion. Every replication cycle leads to the loss of telomeric sequences because of incomplete replication at the 3'-end by DNA polymerases which are unable to fill in the gap left by the 5' RNA primer. The solutions to these problems are provided by telomeric shelterin complex, capping the end of chromosomes and inducing the formation of protective secondary structures and telomerase recruitment to assure telomere maintenance. **(B)** Secondary structures adopted by telomeric ends. T-loop structure is the result of the single-strand overhang invasion into the double stranded DNA portion, thermodynamically favored by shelterin complex proteins binding. G4 structure formed by tandem d(TTAGGG) repeats in the single-strand overhang. Figure adapted from (29)

During the semi-conservative DNA replication, the polymerase enzyme requires short RNA molecules to act as a primer for DNA synthesis, but at the end of the chromosomes the primer is removed from the template and the gap in the lagging strand cannot be filled leading to the progressive loss of DNA sequences (28). The end-replication problem is the cause of the telomere shortening which leads to the inability of the shelterin complex to bind and

form the protective cap at the end of chromosomes, making them susceptible to the DNA damage responses which are directly correlated to the cell senescence and programmed death (32).

To overcome the progressive shortening of telomeres some human somatic cells, stem cells and cancer cells use the telomerase to elongate telomeres. The holo-enzyme telomerase is a ribonucleoprotein complex composed of a catalytic reverse-transcriptase unit (hTERT) which uses its own RNA component (hTR/hTERC unit) as a template for synthesis of telomeric TTAGGG repeats (33). The activity of telomerase in somatic cells is basically absent and the rate of telomere elongation is not high enough to maintain their length, resulting in gradual shortening of telomeres as consequence of end-replication problem. On the other hand, in 85% of cancer cells the telomerase is overexpressed and up-regulated that permits cancer cells to avoid senescence and apoptosis (34). Therefore, the telomerase activity reactivation has been considered as one of the major hallmarks of cancer (27,29,34).

Besides T-loop, G-rich telomeric sequences may fold into stable G4 structures (Figure1.6B). Folding of the single strand telomeric end into the G4 prevent the telomeric ends from annealing with the telomerase RNA template, arresting the elongation and in such a way inhibiting telomerase catalytic activity (26,35,36).

Telomeric G4s have been widely investigated over the years. First of all a lot of structural work has been conducted to investigate different possibilities of folding depending on the number of tandem repeats in the sequence (from one to four) and nature of stabilizing cations (37). NMR and crystal structures in presence of sodium and potassium ions have been solved and interestingly the nature of cation highly influence the folding topology changing from parallel topologies in presence of potassium to antiparallel in sodium solution. Given the extended amount of tandem repeats, also higher order structures, where more than one G4 structures stack upon each other, have been observed (37).

One of the most direct evidences of G4 formation at telomeric level so far comes from *in vivo* experiments showing that telomere structural proteins, such as TEBPα (Telomere end-binding protein) and TEBPβ in ciliates may promote G4 folding (38), while POT1 has been demonstrated to unwind telomeric G4s *in vitro* (39).

The importance of the biological role of G4 structures at telomeric level has been also supported by recent studies evidencing that human helicases, such as WRN and BLM are involved in telomere maintenance and are able to unwind G4 structures *in vitro* (26,40). Mutations in WRN cause Werner syndrome characterized by premature aging which can be related to the telomere shortening and mutations in BLM cause Blooms syndrome characterized by high susceptibility to cancer diseases (26).

Given the protective role and telomere mediated telomerase inhibition, targeting of G4 structures have been investigated as novel anticancer strategy. Small compounds binding and stabilizing G4s displayed promising telomerase inhibition activity and cell proliferation suppression (41). Different families of tested G4 ligands suggest a general mechanism of POT1 displacement and reactivation of DNA damage responses ultimately leading to apoptosis (41).

1.4.2 G-quadruplexes during replication

In the double strand context, G4s structures are likely to form during a transient opening of the double helix, which is necessary for DNA processing in some mechanisms, such as replication. The naturally occurring strand separation gives a thermodynamically favorable environment for potential G4 forming sequences to fold into G4 structure.

Once formed, these structures may sterically impede the processing of the template by polymerase enzymes. Accumulating evidences suggest that unusual DNA secondary structures, G4s among these, are implicated in the stalling of replication fork progression (42).

To ensure the correct progression of replication process helicases are involved in unwinding of DNA secondary structures. Many helicases have been shown to unwind G4s, indirectly supporting their existence and functions (43). For instance, Pif1 helicases unwind G4s in the leading strand of replication fork and the absence or mutations of Pif1 helicases slow down the replication fork progression (44,45). Moreover, the stabilization of G4s with $PhenDC_3$ inhibits unwinding activity of Pif1 mimicking genome instability induced by Pif1 deletion (46). Human FANCJ helicase has been shown to unwind G4s in ATP-dependent manner in a preferential 5'-3' polarity, suggesting the activity on the lagging strand, and its unwinding

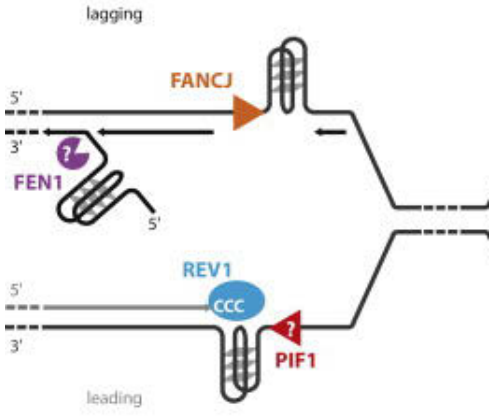


Figure 1.8 G4 involvement in replication process. During the replication fork progression, G4s may form while DNA is transiently in a single strand conformation in both leading and lagging strands. Pif1 helicases have been shown to be associated with the leading strand where they are able to unwind G4s. FEN1 nucleases are able to recognize and cleave G4s in the 5'-flap during the Okazaki fragments processing. Figure adapted from (48)

properties are impaired by G4 stabilizing ligand telomestatin *in vitro* (47). Cell lines encoding mutated FANCJ have been shown to accumulate deletions at sequences matching G4 motifs (47) (Figure 1.8).

The fact that in the absence of helicases, G4 forming regions may undergo deletions suggested both that G4s are problematic sequences to be resolved and that the nucleases may recognize G4 regions for specific cleavage (48). FEN1 and EXO1 nucleases have been shown to bind G4s *in vitro* (48).

The evidences of the role of different helicase families in maintaining the genomic stability through G4 related mechanisms indirectly support the formation of these structures at genomic level and their modulatory role of replication process and genomic integrity (49).

1.4.3 G-quadruplexes in gene promoters

Within gene promoters G4 forming motifs are especially enriched in the oncogene promoters, while being underrepresented in tumor suppressor genes (25), suggesting an evolutionary selection of functionally relevant G4 structures and their therapeutical potential as drug targets (50-53). G4s in oncogene promoters are usually located immediately upstream of transcription start sites (TSS), correlated to binding sites of important transcription factors such as Sp1 and NF-kB and have been proposed as regulatory elements of transcription initiation (25).

Gene promoter G4s are mostly intramolecular structures, formed by a single strand DNA. In contrast to telomeric structures in the single stranded 3'-overhang of human telomeres, G4s in gene promoters, where genomic DNA is in its usual double strand conformation, have to overcome the energetic barrier of the duplex to form (50). Molecular crowding conditions are known to stabilize G4s once they are formed (54,55), while negative superhelicity generated by transcriptional machinery during the transcription may give the driven force to promote the G4 formation (56). Intramolecular structures are more complex and polymorphic as compared to bimolecular o tetramolecular G4s, displaying a great variety of loop conformations, strand directionalities and tetrads arrangement (53). Moreover, G4 forming sequences described in promoter regions often contain more than four G-tracts that potentially allow the sequence to adopt various conformations with involvement of different guanines and thus loop sequences. Different conformations may overlap and G4 structures are in continuous dynamic equilibrium, even if usually there is a single prevalent conformation. This particular feature makes the structural and biophysical characterization more complex. The examples of multiple intramolecular conformations in oncogene promoters are bcl-2 (57-59), PDGFR-β (60), c-kit (12,61,62), while RET (63,64), PDGF-A (65) and VEGF (66-68) represent the class of single conformation G4s.

Significantly, formation of G4s in promoter of oncogenes is highly associated with six hallmarks of cancer (69). Critical proteins associated with self-sufficiency of growth signal (c-Myc, c-KIT, KRAS), insensitivity to anti-growth signals (pRb), evasion of apoptosis (Bcl-2), sustained angiogenesis (VEGF), limitless replication (hTERT) and metastasis (PDGF-A), have been demonstrated to present stable G4s in their gene promoter regions (69).

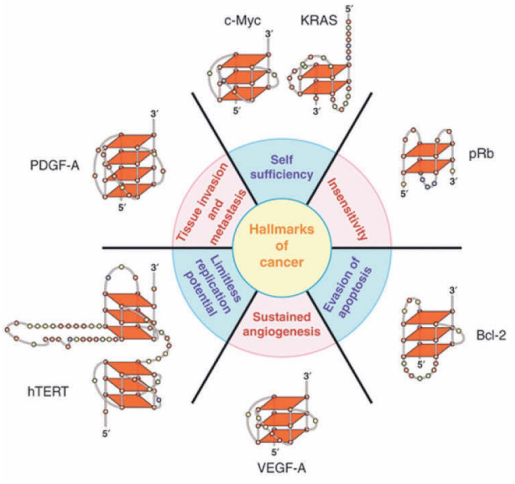


Figure 1.9 Representation of six hallmarks of cancer and G4 structures that have been described in the core or proximal promoters of associated genes. Represented structures display a great topological variety. Figure from (69)

1.4.3.1 G-quadruplexes in the c-myc promoter

G-quadruplex in the promoter region of c-myc proto-oncogene is probably the most representative and widely investigated example of these secondary structures involved in gene transcription regulation.

C-myc plays an essential role in the regulation of apoptosis, protein synthesis and cell adhesion (70). Around 20% of human cancers can be associated with the overexpression of c-Myc which leads to tumorigenesis and sustained tumor growth (70). Therefore, inhibition of c-myc is a therapeutic strategy for human cancer.

Around 80% of transcriptional activity of c-myc is controlled by nuclease hypersensitivity element III_1 (NHE III_1) (71). It is located immediately upstream the promoter and contain a G-rich sequence that has G4 folding potential (71,72). The presence of more than four G-tracts gives rise to a dynamic equilibrium of multiple G4 structures. NMR studies in K^+ solution showed the prevalence of parallel conformations with different loop isomers (73,74).

The transcriptionally active form of c-myc promoter is the duplex conformation and can be activated upon binding of transcription factors, such as Sp1. In fact, NHE III_1 contains high affinity binding sites for Sp1, and cotransfection in mammalian cells indicated that Sp1 transactivates the c-myc promoter (71). It has been proposed that the presence of three

DNA structural populations may regulate the c-myc transcription. Double helix (with or without Sp1 binding) and binding of hnRNPk and CNBP to the single strand component, G- and C-rich respectively, activate transcription, while formation of G4 and i-motif (secondary structure formed in C-rich stand) is the repressor element for transcription (71) (Figure 1.10).

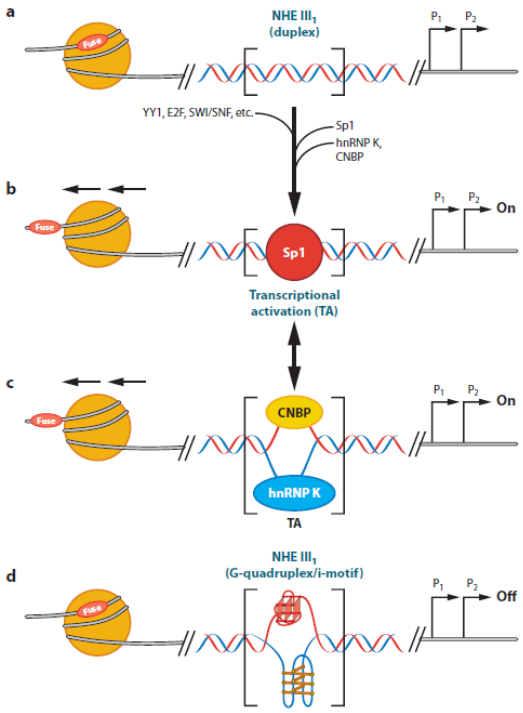


Figure 1.10 Models of different promoter conformations in c-myc NH EIII$_1$. **(A)** Representation of duplex form of NHE III$_1$ promoter region without bound proteins. The duplex form is a binding site for transcription factor Sp1. **(B)** Upon Sp1 binding c-Myc expression is activated. **(C)** Binding of hnRNPk and CNBP to the non-structured G- and C-rich single strands respectively leads to the transcription activation. **(D)** Absence of binding proteins leads to formation of G4 and i-motif structures acting as repressors of c-myc transcription. Figure from (71)

Recently other two proteins have been described as structural modulators of c-myc promoter. It has been proposed that NM23-H2 (human nonmetastatic 23 isoform 2 protein) induces c-myc transcription by stabilizing single strand conformations that can be bound by single-strand transcription factors, such as CNBP, to activate transcription (75). On the other hand, the multifunctional nucleolar protein nucleolin induces formation of G4 structures in the NHE III$_1$ of c-myc leading to transcriptional inhibition (76).

Taken together all these data strongly suggested that G4 structure in the c-myc promoter acts as a silencer element. Thus, the selective stabilization of this structure by G4 binding ligands to repress c-Myc expression has been proposed (75,77). Particularly, a well-known G4 binding compound TMPyP4 was able to lower significantly c-myc transcriptional activa-

tion in cell lines where NHE III_1 was present, while having a small effect on cell lines with deleted NHE III_1 (77).

1.4.4 RNA G-quadruplexes and their role in translation

Beside the DNA G4s, RNA G-rich sequences may also form G4 structures (RNA G4s). The general structural features resemble the DNA G4s with some topological limitations. In fact, so far only parallel RNA G4s have been reported. The major reason is probably the inability of ribose sugar to assume *syn* glycosidic bond angle conformation, which is a requisite for antiparallel structures formation (78). The presence of uraciles instead of thymines in the loop sequences additionally contributes to the limitation of assumed conformations, and to the increased stability of RNA G4s because of increased stacking interactions and release of water molecules (79,80). The hydroxyl group in 2' position of ribose sugar, contributes with additional intramolecular hydrogen bonds making RNA G4s more stable compared to DNA counterparts.

A number of putative RNA G4 forming sequences were found in the 5' untranslated regions (UTR) of human mRNAs including genes of clinical interest. At this level RNA G4s were proposed as regulatory elements of gene expression at the translational level (81,82). The initiation step of the translation is the rate limiting and most regulated step of whole translation process. The recruitment of the ribosomal subunit 40s occurs at the modified cap analog that forms 5' end of mRNA, then the complex scans mRNA until the AUG start codon is reached. Alternatively, the translation starts in the internal ribosome entry site (IRES) which substitutes modified cap analog. Both mechanisms are influenced by the presence of secondary structures, including G4s. Interestingly, during the cap-dependent translation, RNA G4s inhibit translational process, while in the cap-independent translation, secondary structures and G4s among them are required for initiation (Figure 1.11) (81).

One of the representative examples is RNA G4 motif forming close to the 5' cap of 5'-UTR of the NRAS proto-oncogene mRNA, which repress the gene expression at translational level (83). Moreover, it is a target for small molecules that inhibit translation *in vitro* (84).

RNA G4s mediated translational up-regulation has been reported for cap-independent translation. Formation of G4 structure in the IRES is required for translation initiation of VEGF (85), FGF-2 (86), and TGFβ2 (87).

The discovery of RNA G4 downstream of an endonucleolytic cleavage site in the 3'-UTR of IGF-II mRNA suggesting the role of G4s in mRNA-processing (88). The presence of a stable G4 structure downstream from the p53 pre-mRNA cleavage site is critical for maintaining p53 3′ end processing efficiency. This cis element via interactions with the hnRNP H/F pro-

tein maintains p53 expression and leads to apoptosis following DNA damage (Figure 1.11B) (89).

RNA G4s may influence the translational process not only in the untranslated region but also in the open reading frames (ORFs) of mRNA coding regions, where they cause frameshift mutations leading to the recoding processes (Figure 1.11C)(90).

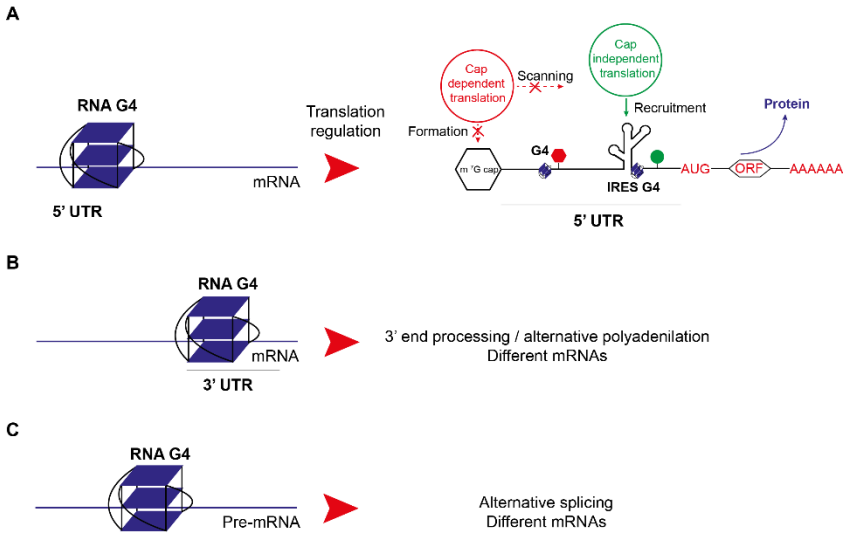


Figure 1.11 Role of RNA G4s in mRNA. **(A)** RNA G4s in 5'-UTRs of mRNA are involved in translational regulation. **(B)** When located in 3'-UTRs of mRNA G4s are involved in 3'-end processing and alternative polyadenylation. **(C)** Alternative splicing and frameshift mutations when located in introns. Model of potential translation regulation operated by RNA G4s in cap-dependent and cap-independent translations.

The increasing interest on these structures is supported by the fact that RNA is generally found in cells in a single strand form, so the G-rich sequences are prone to fold into G4 structure without competing the duplex formation. Moreover RNA is located in cytoplasm of cells and this is a great advantage for targeting of RNA G4s as they are more accessible by different types of small molecules.

1.5 G-quadruplex targeting by small molecules

Since the accumulating evidences of the biological role of G4 structures, they have been proposed as promising targets in anticancer therapy, particularly in terms of telomerase activity inhibition and G4 related transcriptional suppression of some oncogenes. A discrete diversity of small molecules have been synthesized and tested as G4 stabilizing agents.

The very general features of G4 binding compounds resemble the general features of the structure itself in order to target the extended aromatic surface of the tetrads by π-π stacking interactions(17,91-93). Therefore, the G4 binders are flat aromatic molecules that can establish π-π interaction with the external tetrads of the structure and usually present protonated side-chains that contribute to the binding stability by additional electrostatic inter-

actions with negatively charged phosphate backbone. The external stacking binding mode by enlarge aromatic surface of G4 binders ensure the selectivity over duplex form where there are not enough available aromatic interactions.

The external stacking is the preferential way of interacting of the majority of know G4 ligands, although the intercalation, generally typical for duplex DNA targeting, have been observed. The intercalation binding mode is not common for G4 ligands because the G4 structure itself is rigid and the intercalation of chemical compounds is generally not favorable energetically (17,93,94). Moreover the presence of cations in the central channel of the structure hinders the placement of positively charged large aromatic compounds between the tetrads.

G4 loops and grooves are also accessible for small molecules and, importantly, they are very different in size and shape compared to duplex structure which can represent an attractive binding site for selective targeting of G4s (17,93-95).

According to the review published by Teulade-Fichou (95), it is possible to classify the variety of G4 ligands in 4 large groups, based principally on their cationic nature: *in situ* protonated G4 ligands, N-methylated aromatic G4 ligands, metallo-organic G4 ligands, non-cationic macrocyclic ligands.

1.5.1 *In situ* protonated G-quadruplex ligands

In situ protonated G4 ligands are probably the mostly explored type of compounds. Large flat aromatic surfaces, necessary for stacking interactions with G-tetrads result usually in high hydrophobic characteristics. Therefore, in order to maintain the water solubility of these molecules protonable side chains are introduced around the aromatic core.

BRACO-19 (Figure 1.12A) is the best and more extensively studied example of these group of compounds. It has an acridine core with three protonable substituents (3,6,9-

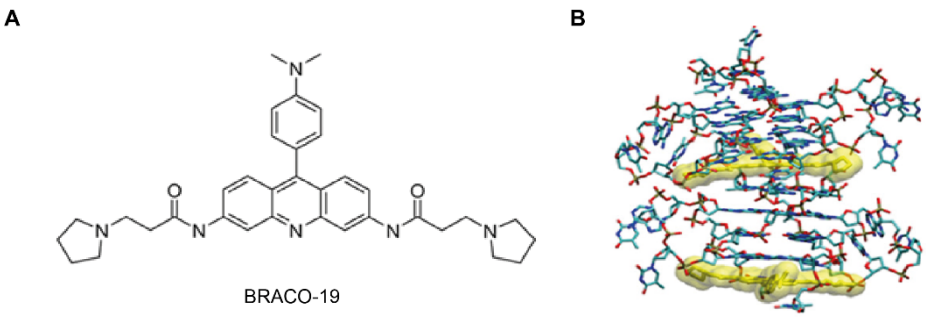


Figure 1.12 (A) Chemical structure of BRACO-19. **(B)** Side view of the crystal structure of BRACO-19/telomeric G4 complex. The ligand is sandwiched between two units of bimolecular three-layered G4s. Figure **(B)** from (98)

trisubstituted) that was rationally designed to interact with three G4 grooves (96). As confirmed by crystal structures of G4 complex with a multiplicity of di-substituted acridines (97,98), and molecular modeling analyses these core appears to be suitable for G4 structure recognition upon optimized protonation of central ring nitrogen (99). A crystal structure of BRACO-19 complexed with telomeric bimolecular G4 revealed interesting interaction features (100). The stacking interactions seems not to be symmetric with only two guanines forming π-π stacking interactions and the positively charged nitrogen atom in the acridine ring is positioned directly on the central axis of the electronegative channel. Apparently, the positively charged nitrogen is able to displace the monovalent cation and compensate its charges (Figure1.12B) (100).

BRACO-19 greatly stabilizes G4 structures *in vitro*, while being able to discriminate between G4 and duplex conformations as confirmed by spectroscopic studies (circular dichroism, FRET, surface plasmon resonance). It strongly inhibits telomerase activity (101), and recent biological investigations revealed that BRACO-19 has a strong antiproliferative effect on cancer cells (102,103).

1.5.1.1 Naphthalene Diimide compounds (NDIs) as G-quadruplex ligands

NDIs are the well-known family of compounds with high affinity to DNA. They nowadays represent a very promising class of G4 stabilizing ligands (Figure 1.13).

Disubstituted NDIs were first described as double stranded DNA intercalators preferentially binding G:C sequences (104). Some of them were shown to display the anti-cancer activity, although with elevated toxicity and limited therapeutic advantages (105,106).

Neidle and coworkers tested NDI compounds for G4 affinity, obtaining initially not very promising results. They then hypothesized that introducing more substituents to the naphthalene diimide core would have result in improved affinity to the G4 motif, rather than extending the planar moiety, proposing a new class of tri- and tetra-substituted NDIs highly selective for telomeric G4 structures (107,108). The high affinity to the G4s was correlated to the cell growth inhibition in cancer cell lines (109,110). Interestingly, the substitution of N-methyl-piperazine group of the side-chain with pyrrolidine, resulted in a significant increase of specificity for a G4 structure forming in c-kit promoter *in vitro*, reflected in inhibition of kit transcription and growth inhibition in primary kit-dependent gastrointestinal tumor cell lines (111).

The availability of G4/NDI crystal structures allowed to expand these series of compounds based on structure-activity relationships (98,112). It has been found that the activity of NDI

compounds is strongly dependent on the length of side chains (distance between the NDI core and nitrogen atom), entity of substituents and number of positive charges.

In order to enhance the biological activity of NDI class of compounds Neidle and coworkers performed also the macrocyclization. The macrocycle compounds are conformationally locked and in this way may establish more interaction with the extended aromatic surface available in G-tetrad arrangement of guanines comparable to the duplex (113).

Recently, a new series of water soluble core-extended (T-shaped) NDI compounds have been engineered (114,115). New NDI scaffolds are suitable for chemical conjugations allowing the improvement of binding affinity, solubility and cellular uptake of the compounds, as suggested by novel carbohydrates-NDI conjugates (116). Besides increased affinity to the G4 structures due to the extended aromatic surface, the water solubility and intrinsic fluo-

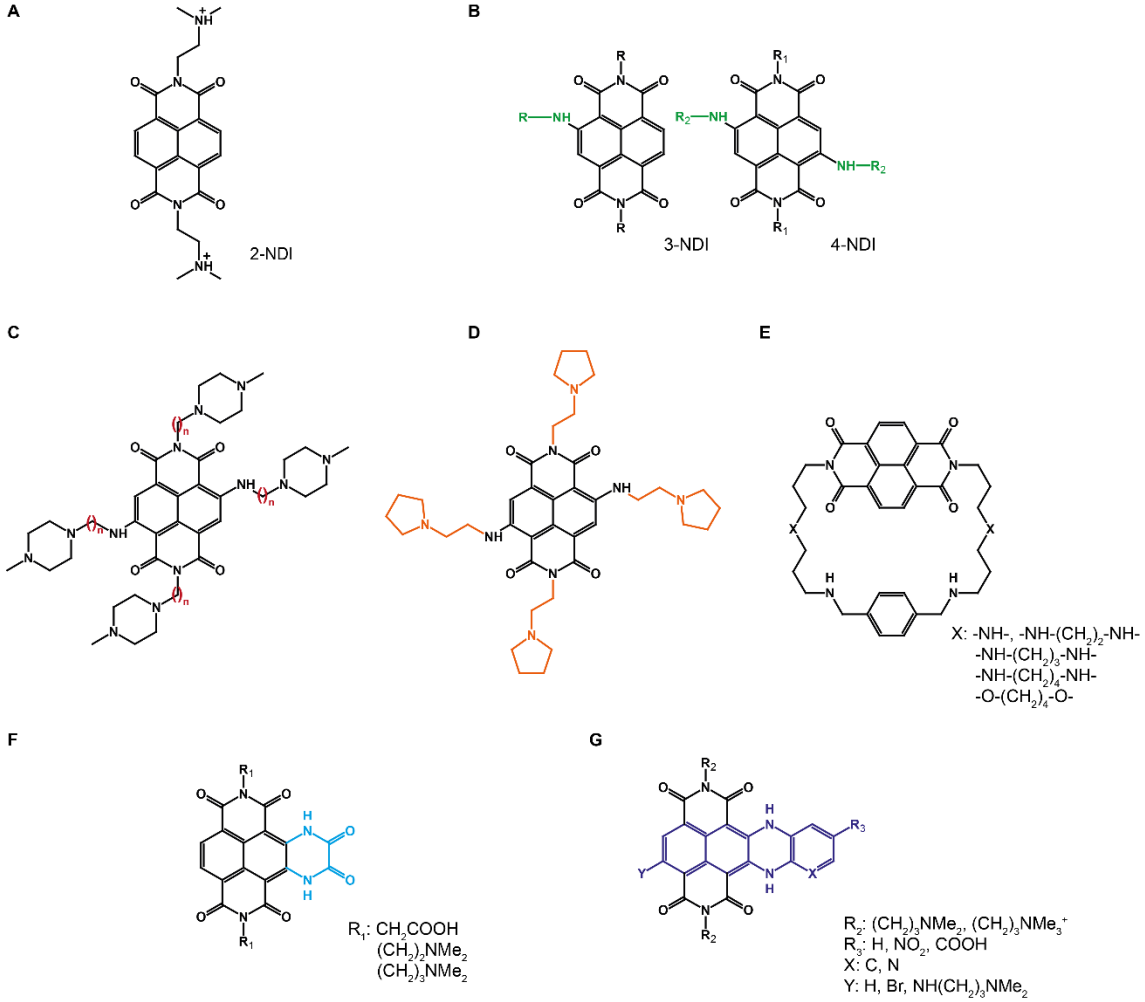


Figure 1.13 Chemical structures of compounds with NDI core. **(A)** Disubstituted NDI core having dsDNA intercalating properties with preferential binding to G/C runs. **(B)** Additional substituents (tri and tetrasubstituted) lend G4 over duplex selectivity properties. **(C)** Tetrasubstituted compounds with long side-chains that contribute to the selectivity towards telomeric G4s. **(D)** Pyrrolidine substitution confers inter-G4 selectivity, having increased affinity towards c-kit G4. **(E)** Conformationally locked marcocyclic NDIs with increased affinity to G4 structures. **(F)** Core-extended NDIs with high affinity towards G4s. **(G)** Core-extended water soluble NDIs.

rescence of these compounds also made them a promising probes with on/off switch of the fluorescence emission in biologically appropriate pH range (5-8.5) (115). In fact, good results have been achieved in testing NDI compounds as G4-selective light-up chemosensors, through exploiting optoelectronic properties of NDI moieties (117,118).

1.5.2 N-Methylated aromatic G-quadruplex ligands

N-Methylated aromatic G-quadruplex ligands (quaternized on the aromatic ring nitrogens) display the double advantage of conserving the water solubility without the need for cationic substituents and increasing the π-π stacking ability thanks to the reduction of the electron density of the aromatic part.

The pivotal compound of these group is tetracationic porphyrin TMPyP4 (Figure 1.14). It has been shown that targeting of c-myc G4 with TMPyP4 leads to the decreasing transcriptional levels of the gene (72). The NMR structure of the complex between TMPyP4 and the modified c-myc sequence has been resolved (119) emphasizing the preferential external tetrad binding and additional electrostatic interactions (Figure 1.14A). Positional isomers of TMPyP4, named TMPyP2 and TMPyP3, which differ only in the position of *N*-methyl group in the pyridyl ring relative to its connection the porphyrin core (Figure 1.14B), have also been testing as G4 ligands (120). Surprisingly, while TMPyP3 displayed similar G4 binding affinity, TMPyP2 does not facilitate, neither bind G4s, probably because the orto position of N-methyl group impedes the free rotation between the pyrididyl ring and the core resulting in loss of affinity to available accommodation sites of the G4 structure (120).

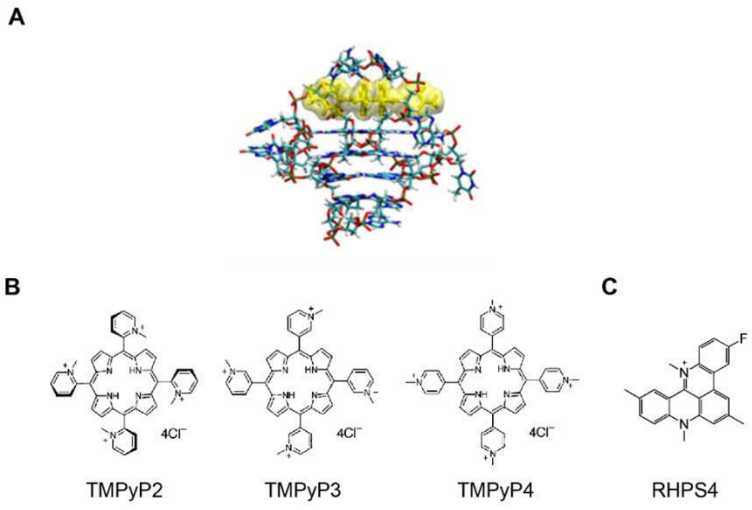


Figure 1.14 (A) NMR structure of TMPyP4/c-myc (Pu24I) G4 complex evidencing the external stacking mode of interaction. **(B)** Chemical structures of porphyrin positional isomers TMPyP2, TMPyP3, TMPyP4 and **(C)** RHPS4. Figure **(A)** from (98)

Among the group of porphyrin derivatives, characterized by enlarged macrocyclic surface, a small molecule RHPS4 deserves to be mentioned (Figure 1.14C). It is a highly condensed aromatic ligand and the NMR solved structure of its complex with tetramolecular G4-DNA showed strong interactions with external tetrads of G4 structures (95,121). *In vitro* and *in cellulo* studies confirmed the ability of RHPS4 to decrease telomere length and modulate telomere binding proteins (122).

1.5.3 Metallo-organic G-quadruplex ligands

This class of ligands is currently emerging because of very promising G4 binding properties. These compounds are usually macrocycles, such as previously described porphyrins, coordinating a metal (Cu, Ni, Mn, etc.) in the central cavity (95) (Figure 1.15). The approach is

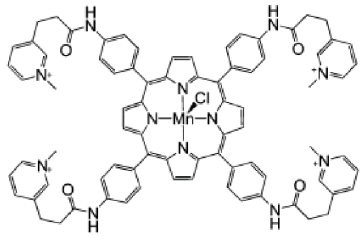


Figure 1.15 Chemical structure of Mn(III)-porphyrin metallo-complex

based on the assumption that the metal in the central cavity may be positioned over the cationic channel of the G4, optimizing stacking interactions of the surrounding molecule with the accessible tetrad (123). First of all, Cu-, Ni-, Mn-, metallo-complexes of TMPyP4 have been tested, giving promising results in terms of G4 stabilization and over duplex selectivity. The most promising metallo-complexes described so far is Mn(III)-porphyrin complex, which displayed a 10000-fold G4 vs. duplex selectivity in surface plasmon resonance experiments and good level of telomerase inhibition (124).

1.5.4 Neutral macrocyclic G-quadruplex ligands

This class includes one of the better characterized G4 ligands telomestatin, a natural macrocyclic molecule isolated from *Streptomyces Annulatus* (Figure 1.16). This macrocyclic com-

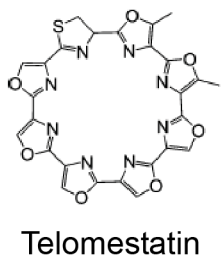


Figure 1.16 Chemical structure of telomestatin

pound completely lack the affinity towards duplex DNA because of cyclic shape and neutral character (125). Perfect shape adaption to the G-quartet arrangement makes this compound highly selective for G4 structures with great stabilizing and G4 inducing properties even in absence of cations. Important telomerase inhibition activity and inhibition of proliferation of telomerase positive cells support the interest in telomestatin (126,127). However, the large-scale chemical synthesis of this natural molecule is very challenging and the investigation of telomestatin-like compounds with comparable activity is currently open (128).

1.6 G-quadruplexes in pathogens

Besides eukaryotic genomes, G-quadruplex forming sequences have been described also in prokariots and viruses (129,130).

The non-coding regulatory region of SV40 polyomavirus contains six (GGGCGG) repeats forming an unusual G4 structure (131). This region is a binding site for transcriptional factor Sp1 and play a regulatory role in the transcription control. HPV genomes contain G4 forming sequences located in the long control regions, responsible of transcription and replication regulation and in the regions coding for L2 protein (late structural protein), E1 and E4 early proteins, suggesting a potential alteration of viral proteins production from overlapping ORFs (open reading frames) mediated by G4 formation (132). Epstein-Barr virus encodes genome maintenance protein EBNA1. Its translation is regulated by G4 secondary structures at mRNA level (133). Stabilization of this G4s by G4-ligand pyridostatin leads to EBV immune evasion due to decrease of EBNA1 synthesis (133). Herpes Simplex Virus-1 dsDNA genome is highly enriched in potential G4 forming sequences, which have been shown to form extremely stable G4 conformations (134). Treatment of HSV-1 infected cells with G4 ligand BRACO-19 lead to inhibition of viral production that has been correlated to the replication step inhibition with G4 mediated mechanism (134).

As well as for human G4s, the listed evidences suggest the importance of these structures as regulatory elements of viral life cycles and targeting of viral G4s is a potential strategy for novel antiviral therapies. This thesis is focused on the G4 structures in the HIV-1 genome and structure, functions and implication for viral cycle will be discussed in details in the next chapters.

1.7 The Human Immunodeficiency Virus (HIV)

HIV is the causative agent of Acquired Immunodeficiency Syndrome (AIDS). The first case of what later would be known as AIDS has been reported in 1981 in United States. In 1983 the first human retrovirus has been isolated from a biopsy of lymph node of a patient with gen-

eralized lymphoadenopathy in Institute Pasteur, in Paris, defined at that time as LAV – lymphoadenopathy-associated virus (135).

Since the beginning of epidemic more than 70 million people have been infected with HIV virus and 35 million died from HIV. Sub-Saharan Africa remains the most affected area with 1 in every 25 adults living with HIV infection (136). HIV is a major contributor to the global burden of disease. In 2010, HIV was the leading cause of disability adjusted life years worldwide for people aged 30–44 years, and the fifth leading cause for all ages (137).

HIV selectively infects CD4+ T-lymphocytes. The infection causes the progressive depletion of CD4 T-cells leading to immune system inefficiency and increase of susceptibility toward opportunistic bacterial and fungal infections.

HIV is a member of the *Retroviridae* family in the Lentivirus genus which counts two types of HIV, HIV-1 and HIV-2. HIV-1 was transmitted from apes and HIV-2 from sooty mangabey monkeys (137). HIV-1 is characterized by high genetic diversity and counts a number of different groups and group subtypes with distinct prevalence in world regions (137). The marked genetic diversity of HIV-1 is a consequence of the error prone function of reverse transcriptase, which results in a high mutation rate. HIV-2 is largely confined to West Africa and causes a similar illness to HIV-1, but immunodeficiency progresses more slowly and HIV-2 is less transmissible (137).

1.7.1 HIV-1 virion organization

HIV-1 is a spherical shape of 100-150 nm diameter dimension virion, coated by viral envelope, which is the bilayer phospholipid membrane originated from the membrane of host cells during viral budding (Figure 1.17). Besides cellular proteins, two important viral envelope glycoproteins (gp) are expressed on the virion membrane and constitute spikes. Spikes

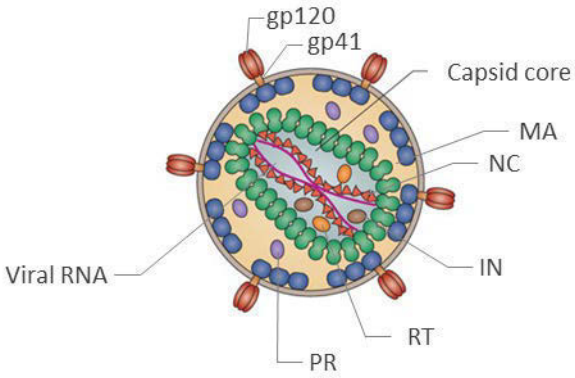


Figure 1.17 HIV-1 mature virion organization. Two copies of viral RNA genome with associated proteins are kept inside the capsid. Matrix proteins are intimately associated with the viral lipid bilayer envelope expressing viral envelope proteins gp120 and gp41. Figure adapted from[1]

[1] Freed, E. O. (2015) HIV-1 assembly, release and maturation. *Nature Reviews Microbiology*, **13**, 484-496

are composed of three units of gp120 and three units of gp41, anchored on the viral membrane. These two glycoproteins are responsible of recognition of target host cells and fusion of the viral envelope with the host cell membrane to allow the release of viral genome and associated proteins essential for the initiation of infection into the cytoplasm.

Inside the virion conical capsid, formed by 2.000 copies of p24 protein, contains two identical copies of positive-sense single-strand RNA genomes. Viral RNA is covered by nucleocapsid protein (p7), which will have a nuclease protective function after virus entry and uncoating inside the host cell. Viral proteins reverse transcriptase (RT), integrase (IN) and protease (PR) are kept inside the capsid. These proteins are major actors of key infectious steps, *i.e*. reverse transcription, integration of HIV-1 proviral genome into the host chromosomes and maturation of newly produced virions, respectively.

1.7.2 Genome structure

HIV-1 RNA genome consists of nine viral genes: three primary structural genes (*gag, pol* and *env*), two regulatory genes (*tat* and *rev*) and four accessory genes (*vif, vpr, vpu* and *nef*) (Figure 1.18).

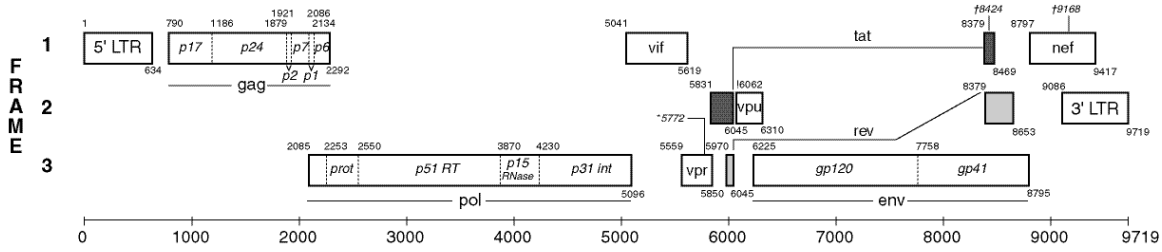


Figure 1.18 HIV-1 genome organisation.

Seven structural landmarks of the genome are described based on their functions:

LTR – long terminal repeats, the DNA sequences flanking the proviral genome. It contains important regulatory regions for transcription initiation and will be discussed in details in next chapters as it is one of the topics of this thesis.

TAR – target sequence for viral transactivation. It is the 45 nucleotides, hairpin structured, mRNA binding site for Tat protein.

RRE – Rev responsive element. 200 nucleotides binding site for regulatory protein Rev.

PE – Psi element. Set of four stem-loop structures recognized by zinc-finger domains of p7 protein.

SLIP – a TTTTTT slippery site, followed by a stem-loop structure, necessary for regulating the ribosomal frameshift out of the Gag reading frame into the Pol reading frame.

1.7.2.1 HIV-1 Structural proteins and enzymes

Gag encodes for structural capsid proteins. It is produced as polyprotein precursor p55, containing capsid protein (p24), matrix protein (p17) and nucleocapsid protein (p7). During the maturation process p55 is cleaved by viral protease to produce mature structural proteins.

Matrix protein is the N-terminal part of Gag precursor and it is associated with the viral lipid envelope in the mature virions. It plays an important role in the viral particle assembly and incorporation of viral glycoproteins into the nascent particles (138).

In the context of Gag precursor capsid protein follows the matrix protein. It has crucial roles in the viral particle assembly and entry into a new target cells (139). In the mature virions capsid protein constitutes the conical shell of the core, coating genomic RNA and viral proteins.

The C-terminal domain of Gag precursor is the small zinc-finger protein called nucleocapsid protein. In the following chapter its structure and functions are discussed in details as it is one the topics of this thesis.

Pol genomic region encodes for viral enzymes RT, IN and PR. As capsid proteins they are also protease cleavage products of the *Gag-pol* precursor polyprotein, produced by ribosome frameshifting. Briefly, RT is a RNA dependent RNA polymerase which synthesizes double stranded DNA proviral genome, subsequently transported into the nucleus and integrated into host chromosomes, from single strand RNA template. The integration step is catalyzed by the integrase, 32 kDa protein, composed of N-terminal zinc-binding domain, catalytic core and C-terminal DNA-binding domain. The 99-aminoacids aspartic protease is a homodimer with the central active site cavity capped by two flexible flap regions. Protease-catalyzed cleavage of viral polyprotein precursors is essential for the formation of mature infective particles.

Viral glycoproteins are encoded by *env* gene and produced as a precursor gp160, processed by the protease into surface membrane gp120 and transmembrane domain gp41. Viral glycoproteins are involved in the recognition of target host cells, attachment and entry processes. Gp120 subunit associates with the CD4 receptor of the host cell and subsequently with chemokine coreceptor (CCR5 or CXCR4 depending on the viral tropism). The coreceptor binding triggers conformational changes in the transmembrane gp41 domain by forcing the fusion peptides into the target cell membrane.

1.7.2.1.1 Nucleocapsid protein

The HIV-1 nucleocapsid protein (NCp7) is a small (55 amino acid residues) basic protein originated from a protease cleavage of Gag polyprotein. The basic N-terminal and C-

terminal domains of the protein are not structured and flexible while the central domain contains two highly conserved CCHC type zinc fingers (140) (Figure 1.19).

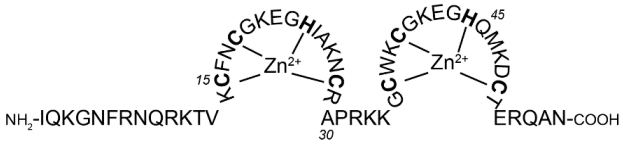


Figure 1.19 Aminoacidic sequence of HIV-1 nucleocapsid protein. Zinc-finger domains participate to the specific binding to nucleic acids by hydrophobic interaction. CCHC aminoacids coordinating zinc atoms are labeled in bold.

NC is a multifunctional nucleic acid binding protein. It acts as a nucleic acid chaperone, favoring the secondary structure to assume thermodynamically most stable conformation (141,142). The highly structured zinc fingers are responsible of sequence specific binding to the nucleic acids (particularly G and UG/TG runs) while the N-terminal basic domain interacts nonspecifically with nucleic acids by electrostatic interactions, giving to the NC also nucleic acids annealing properties (143).

Ncp7 is essentially involved in almost all the steps of HIV-1 replication cycle. First of all, multiple copies of nucleocapsid protein (1.500-2.000) are associated with the RNA dimer genome inside the mature virions, ensuring protection from the RNAse activity once the genome is released into the cytoplasm of host cells (144).

The nucleic acids chaperone properties and annealing activity of nucleocapsid protein contribute to the correct progression of reverse transcription process (145). From the very first step of reverse transcription, i.e. tRNA primer annealing, p7 assists RT enzyme progression. Nucleocapsid protein has been found to partially unfold the tRNA secondary structure to assure the annealing to template RNA giving higher stability to the duplex helix conformation necessary for RT binding and reverse transcription initiation (146,147). Two strand transfers, very delicate and trigging steps, have to occur during the reverse transcription. The first, minus-strand transfer, consists in the translocation of newly synthesized (-)ssDNA at 5' end of the genome to the complementary region R at the 3' end of RNA genome. The R region contains the highly structured trans-activation response element (TAR). 59-nt TAR RNA folds into a very stable hairpin structure. It has been demonstrated *in vitro* that in absence of NCp7 the annealing reaction between TAR RNA and its complementary DNA strand occurs via loop-loop "kissing" intermediate that slowly interconverts to the completely annealed hybrid duplex, while the presence of nucleocapsid protein permits the formation of kinetically favored bimolecular intermediate, facilitating the annealing reaction by around 10^5-folds (148) (Figure 1.20). Besides facilitating the annealing, NCp7 contemporary assures the specificity of minus-strand transfer by competing the nonspecific self-priming re-

action when intramolecular TAR-induced fold-back structures at the 3' end of (-)ssDNA are formed resulting in the abrogation of complete dsDNA synthesis (149).

It is generally accepted that NCp7 assists RT enzyme in the dsDNA synthesis by unfolding complex secondary structures formed by both RNA and DNA templates. Secondary structures, such as hairpins may make difficult the elongation of nucleic acids by polymerase enzymes due to steric impediments and cause pausing sites and incomplete synthesis products. Chaperone activity of NCp7 facilitates RT processing (150).

In the late phase of HIV-1 viral cycle, nucleocapsid protein may have different functions being a part of non-processed Gag polyprotein. Particularly, selection of genomic RNAs and packaging and dimerization are thought to be mediated by Gag NCp7 domain interactions. Genome selection appears to proceed via the direct binding of NCp7 to particularly structured RNA packaging signals, called Ψ-sites, that are generally located near the 5'-end of the viral RNA (5'-UTRs) (151).

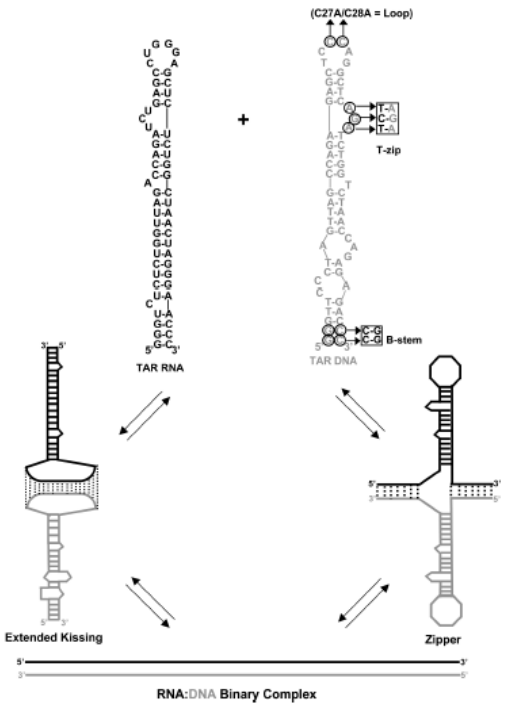


Figure 1.20 RNA TAR/cTAR annealing mechanisms. In absence of NCp7 extended kissing between two hairpin structures mechanism is favored. NCp7 catalyzes the formation of bimolecular intermediate that is fast interconverted in the duplex form. Figure from (148)

Interestingly, NCp7 has been shown to efficiently localize in the nucleus after around 8 hours post-infection (116) and act as an enhancer of basal promoter activity by binding to the promoter regions in the LTR of integrated genome (118).

Given the assistance role of nucleocapsid protein in the key steps of HIV-1 viral cycles, NCp7 has been considered as a promising target for antiviral therapy. The approaches to inhibit the NCp7 activity are directed to the structural unfolding of protein by zinc ejection, which

disrupts highly structured zinc-finger domains responsible of specific nucleic acids binding (117). Another approach is to target and stabilize nucleic acids bound by NCp7. RNA binders targeting TAR/cTAR have been tested (152).
Given high affinity of NCp7 to the guanine rich sequences, it is not surprising that its chaperoning activity may be involved in modulating formation of DNA and RNA G4 structures. In fact, there are evidences of NCp7 mediated assembly of intermolecular, particularly tetrameric, G4s, probably due to its nucleic acids aggregation properties (117,153). In contrast, spectroscopic techniques showed that NCp7 displays an unfolding activity upon binding to the intramolecular G4 formed by thrombin binding aptamer (TBA) sequence (154).

1.7.2.2 Regulatory proteins

Tat and *Rev* are virus-encoded transactivating proteins involved in HIV gene expression. Tat binding to the TAR RNA element activates transcription elongation from the LTR promoter. Rev acts by binding to the RRE element and promotes nuclear export, splicing of structural mRNAs, positively regulating structural genes expression.

1.7.2.3 Accessory proteins

The HIV accessory proteins, Nef, Vif, Vpr and Vpu, represent critical virulence factors *in vivo.* Nef has been shown to have multiple activities, including the downregulation of the cell surface expression of CD4 and stimulation of HIV infectivity. Vif is also responsible of HIV infectivity, probably by affecting reverse transcription step, viral uncoating, or the stability of viral nucleoprotein complexes. Vpr plays a role in the ability of HIV to infect nondividing cells by facilitating the nuclear localization of the preintegration complex (PIC), which contains reverse transcribed proviral DNA, and acting as a nucleocytoplasmatic transport factor. Vpu protein two functions are down-modulation of CD4 and the enhancement of virion release from the surface of infected cells (155).

1.7.3 HIV-1 viral cycle

Figure 1.21 represents the key steps of HIV-1 viral cycle.

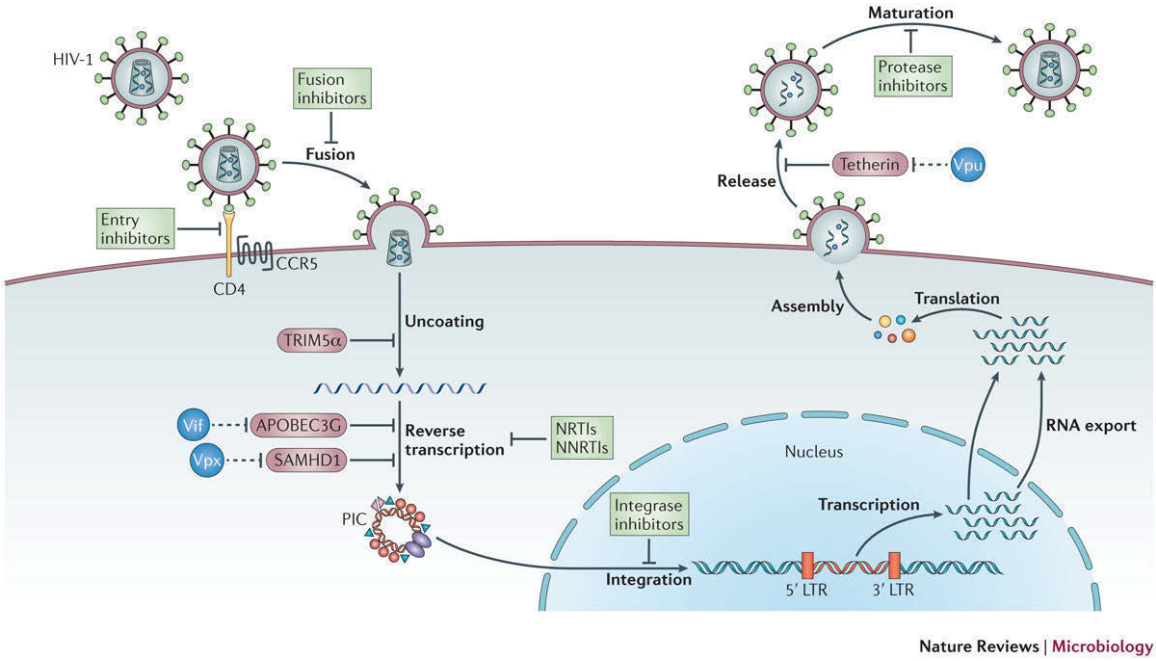


Figure 1.21 Schematic representation of HIV-1 viral cycle

HIV-1 virus infects CD4+ T-cells and macrophages, as the viral glycoprotein gp120 domains exhibit a specific binding for CD4 receptors. After attachment and fusion of viral particle with the host cell membrane, mediated by binding of viral glycoproteins gp120 and gp41 to the cellular receptor CD4, the uncoating of the virus allows the release of genetic information inside the cytoplasm of the cell, where reverse transcription occurs.

1.7.3.1 Reverse transcription mechanism and LTR formation

The reverse transcription process is one of the key steps in retroviral cycle. Viral reverse transcriptase is able to transcribe single strand RNA genome into double strand DNA which will be integrated in the human genome. After virus entry, two copies of genomic RNA and associated proteins are released into the cytoplasm of the host cell. To start the reverse transcription the RT enzyme needs the $tRNA^{lys3}$ cellular primer, which is complementary to the region located at 5'-end of viral RNA, called primer binding site (PBS). The reverse transcription proceeds from PBS to the 5'-end (Figure 1.22). At this point the newly synthesized DNA portion is transferred at the 3'-end of RNA (first strand transfer), while the RNA template is degraded by RNAseH activity of RT. The strand transfer is possible because of the identic R regions situated at both ends of viral RNA, where newly synthesized DNA is complementary to RNA and the cDNA synthesis proceeds across the genome. The synthesis of complementary DNA strand occurs in similar way. In this case the poly-purine tract (PPT),

purine-rich region which is not degraded by RNAseH activity of RT, is used as primer. The

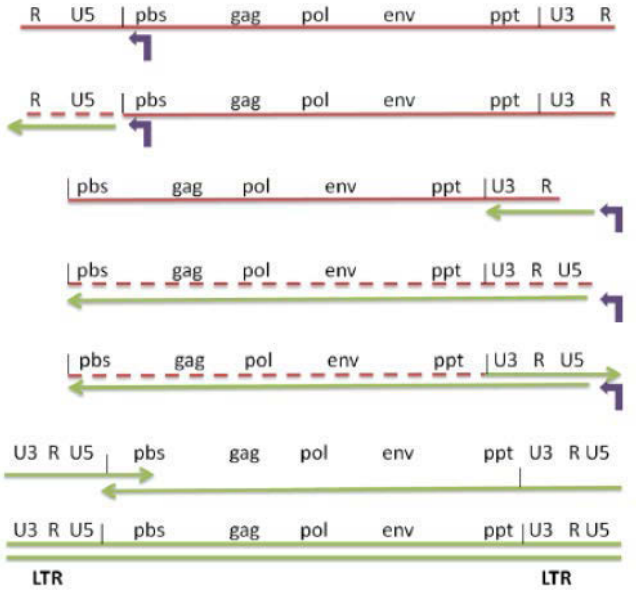


Figure 1.22 Schematic representation of reverse transcription mechanism

transcription proceeds across U3, R and U5 regions and the second strand transfer from 5' to 3' of the genome occurs to complete the synthesis of double stranded proviral genome which is actively transported into the nucleus as pre-integration complex (PIC). After reverse transcription process is completed proviral genome contains U3-R-U5 repeats at both ends (Figure 1.23). These regions are called long terminal repeats (LTR) and have extremely important regulatory functions. Particularly, U3 region contains a promoter region responsible of the transcription of all viral genes, and contains binding sites for transcription factors such as Sp1 and NF-kB. Interaction of transcription factors with binding sites in the U3 region of HIV-1 LTR is critical for transcription activation and viral gene expression and the inhibition of this interaction may bring to negative consequences for the entire viral cycle (Figure 1.23).

The newly synthesized proviral dsDNA is actively transported across the nucleus membrane as Pre-integration complex (PIC) and the integrase catalyzes the integration of proviral genome into the chromosomes of the host cell. At this point virus can undergo latency or the viral transcription can be activated upon binding of viral regulatory protein Tat or of cellular transcription factors. Viral transcription is operated by host RNA polymerase II (RNAPII) which is positioned in the pre-initiation complex guided by the presence of the core promoter initiator (Inr), the TATA BOX and three Sp1 binding sites. RNAPII produces only short transcripts blocked in early 5' portions of mRNAs. At this point, HIV-1-encoded Tat protein is essential for activating transcriptional elongation, but it is not involved in transcriptional initiation and the only Tat-TAR interactions are not enough for transcription stimulation (156). Specific host cellular cofactors are required for this process. Human positive tran-

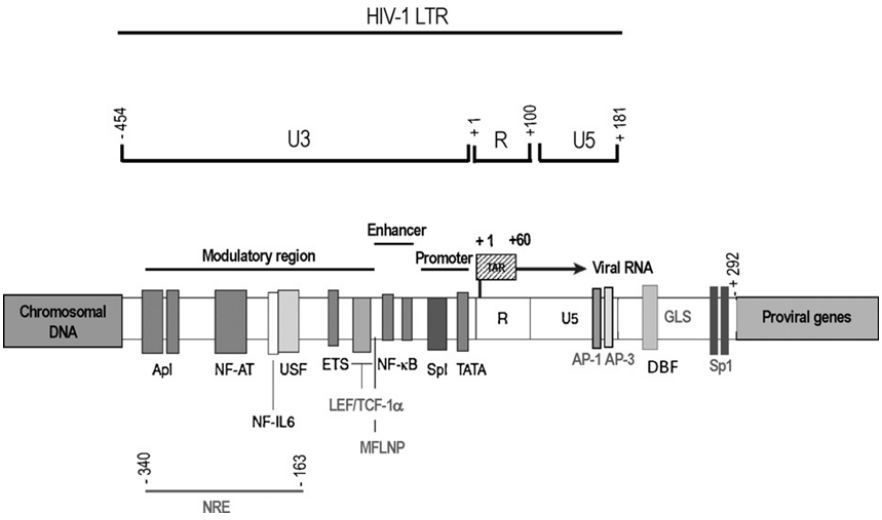


Figure 1.23 HIV-1 LTR organization. LTR is subdivided in three structural regions U3, R, U5 originated from 3' and 5' ends of RNA genome. U3 region is a modulatory region which contains the promoter governing the transcription of all HIV-1 genes. Promoter region is characterized by the presence of three Sp1 and two NF-kB transcription factors binding sites. Figure from [2]

[2]de Arellano, E.R., Alcami, J., Lopez, M., Soriano, V., Holguin A. (2010) Drastic decrease of transcription activity due to hypermutated long terminal repeat (LTR) region in different HIV-1 subtypes and recombinants. *Antiviral research*, **88**, 152-159

scription elongation factor b (P-TEFb) helps Tat protein to ensure high affinity interactions with TAR stem-loop to produce full-length viral transcripts (157).

After full-length mRNAs are produced in the nucleus, they are transported into the cytoplasm where translation and assembly of immature virions occurs. Immature virions buds from the cell and during the maturation process protease activity cleaves viral polyproteins into mature proteins with subsequent virion structure reorganization.

1.7.4 G-quadruplexes in HIV-1

The first evidence of G4 formation in the HIV-1 genome came from Sudquist and Heapy, when they described a cation dependent dimerization of two homologous copies of RNA genomes of HIV-1. They demonstrated that the highly conserved purine-rich region located near the 5'-end of gag gene spontaneously dimerize in high ionic strength conditions without protein cofactors and that GGGGGAGG sequence at the positions 817-825 is essential for the dimerization (158). They therefore proposed that the dimerization occurs via intermolecular RNA G4 formation. Interestingly, based on observations from Marquet *et al*. (159), they also proposed that the genome dimerization may be kinetically favored by the presence of nucleocapsid protein covering the RNA genome.

After the reverse transcription is completed, in the polypurine tract (cPPT) of double stranded HIV-1 proviral genome a so called central DNA-flap is formed. This DNA portion has been demonstrated to fold into G4 structure. Lyonnais *et al*. proposed that they may

have a nuclease protective function of pre-integration complex, as DNA-flaps are usually recognized by cellular nucleases (160).

More recently, a G-rich sequence has been identified at the beginning of gag coding region, characterized as preferential site for recombination. The RT stop products, corresponding to the G-runs involved in G4 formation in RT pausing assays, confirmed the G4 formation. Intermolecular G4s in gag recombination hot spot have been proposed to favor strand transfer of RT by pausing at G4 forming sites, event that highly promotes the genomic recombination of the virus (161). Moreover, NCp7 has been shown to induce dimer structures, amplifying the recombination events (162).

Our group reported the formation of G4 clusters in the HIV-1 *nef* coding region. We showed that three contiguous putative G4 regions are present and that at least two are extremely conserved among most circulating HIV-1 strains. This region have been shown to fold in G4 structures stabilized by presence of cations and G4 binding compounds, particularly TMPyP4 resulting in impaired Nef expression and significant suppression of Nef-dependent enhancement of HIV-1 infectivity (163).

2 Background and aim of the study

Our group has recently described a dynamic G4 forming sequence in the LTR of HIV-1 proviral genome (164). The U3 promoter region of LTR contains regulatory and enhancer elements with binding sites for cellular transcription factors, i.e. two binding sites for NF-kB and three tandem binding sites for Sp1, which cooperate to the viral transcription activation (Figure 2.1). The highly conserved G-rich sequence spanning these five transcription factors binding sites is able to fold into a dynamic set of at least three G4 structures. Spectroscopic studies, supported by *Taq* polymerase Stop assay confirmed formation and thermodynamic stability of LTR-II, LTR-III and LTR-IV G4 structures especially in potassium solution. Binding of G4 ligands, such as BRACO-19 and TMPyP4, greatly increased the stability and steric

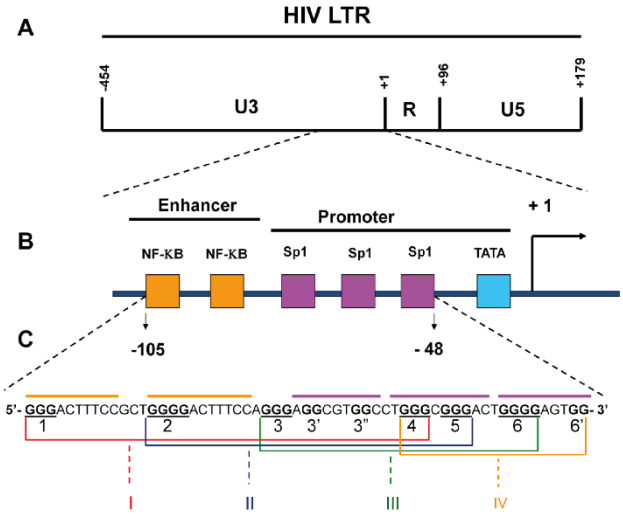


Figure 2.1 The G-rich HIV-1 LTR region organization. **(A)** Schematic representation of HIV LTR, containing U3, R and U5 regions. **(B)** NF-kB (orange) and Sp1 (magenta) binding sites are highlighted in the -105/-48 U3 sequence. **(C)** G-rich sequence able to form different G4 structures, named LTR-I, LTR-II, LTR-III and LTR-IV.

inhibition effect on the processing of the G4 forming template by *Taq* polymerase.

When the wt LTR and selected point mutants were cloned upstream of the firefly luciferase gene, an increase of 50% of promoter activity was observed in presence of LTR point mutants disrupting G4 formation, compared to wt and G4 non-disrupting mutants. In presence of BRACO-19, wt LTR promoter activity decreased, while no effect was observed on G4 abolishing mutants. These data strongly suggested that HIV-1 LTR G4s act as repressor elements of viral transcription activation (164).

Further investigations of LTR G4s involvement in HIV-1 transcriptional process lead our group to analyze potential binding proteins that may control the G4 folding of this region. Electrophoretic mobility shift assay and pull-down approaches combined with mass spectrometry allowed the identification of cellular protein nucleolin, which specifically recognize LTR G4s (165). The LTR G4-nucleolin specific interactions resulted in the increase of silencing activity operated by G4s on viral transcription, as observed with luciferase reporter assay, while nucleolin depletion produced the enhancing effect on LTR promoter activity.

Cellular protein nucleolin binding to the LTR G4s and its influence on LTR promoter activity indirectly supported the biological relevance of these structures for the regulation of HIV-1 transcription, a key viral process. Based on these evidences LTR G4s were proposed as novel antiviral targets. Indeed, BRACO-19 was tested for antiviral activity and displayed promising G4-related viral production inhibition properties, validating LTR G4s as antiviral targets (166). However, BRACO-19 is a well-known G4 ligand able to bind a wide variety of G4 structures and thus, it is not specific for recognition of viral G4 structures and displays a high affinity to the cellular G4s (for instance telomeric structures).

The first focus of this thesis is to explore the possibility of selective targeting of HIV G4s in order to develop new antiviral compounds with G4 mediated mechanism of action. We faced the selectivity challenge through two different approaches. The first approach consists in the screening of small library of NDI-core based G4 ligands modified with new features. The newly synthesized core-extended NDI compounds were tested *in vitro* for selectivity towards viral G4s over cellular (telomeric) structures through spectroscopic analysis combined with surface plasmon resonance analysis and G4-specific electrophoretic techniques. On the other hand, rational drug design approach can be adopted. To develop specific viral G4 ligands the 3D structural coordinates are necessary. To this purpose, we sought to solve the high-resolution structure in potassium solution of the most stable and preferentially targeted G4 component within the LTR G4-forming region by NMR technique.

Interestingly, the investigation of antiviral activity of BRACO-19 revealed two steps targeted by the compound. Besides the post-integration step at proviral genome level, a pre-integrational target has been proposed compatible with the reverse transcription process at

the HIV-1 RNA genome level (166). Thus, the second part of the thesis is focused on the investigation of RNA G4 structures in the U3 region of HIV-1 genome. HIV-1 RNA G4s may likewise represent an attractive antiviral target during the early step of viral cycle. First of all, the *in vitro* characterization of RNA G4s was performed followed by the investigation of the potential involvement of both cellular and viral proteins in controlling the formation of U3 RNA G4 folding.

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, dissolved in Tris-EDTA (TE) buffer at stock concentrations (1 mM or 100 uM) and stored at -20 °C. RNA oligonucleotides were purchased from ThermoFisher Scientific (Waltham, MA, USA) in lyophilized form, dissolved in DEPC-treated water at stock concentrations (1 mM or 100 uM) and stored at –80 °C.

For NMR experiments, unlabeled and site-specific labeled DNA oligonucleotides were prepared by phosphoramidite synthesis using reagents from Glen Research (Sterling, USA). Samples were deprotected in ammonium hydroxide solution at 55 °C for 16 h, purified using Poly-Pak cartridges following Glen Research protocol. Briefly, cartridges were prepared by acetonitrile and 2M TEAA solution treatment to enhance DNA binding. Deprotected DNA was loaded into the cartridge, treated with 2 % TFA and eluted with 20 % acetonitrile solution. Purified DNA was then dialyzed overnight against 20 mM KCl solution to allow G4 formation. The excess of KCl was removed by dialysis against water for 2 hours. Upon lyophilization DNA was obtained in powder form. DNA samples were dissolved in potassium chloride (70 mM) and potassium phosphate (20 mM) buffer. The concentration of the sample was obtained by measuring UV absorbance at 260 nm.

The oligonucleotide sequences used in this study are indicated in the Table 3.1

Application	Oligo name	Sequence (5'-3')
CD analysis	LTR-III	GGGAGGCGTGGCCTGGGCGGGACTGGGG
	LTR-IV	TGGGCGGGACTGGGGAGTGGT
	hTel	GGGTTAGGGTTAGGGTTAGGG
	RNA U3-II	GGGGACUUUCCAGGGAGGCGUGGCCUGGGCGGG
	RNA U3-III	GGGAGGCGUGGCCUGGGCGGGACUGGGG
	RNA U3-IV	GGGCGGGACUGGGGAGUGG
FRET analysis	LTR-III	FAM-TGGGAGGCGTGGCCTGGGCGGGACTGGGGT-TAMRA
	LTR-IV	FAM-TGGGCGGGACTGGGGAGTGGT-TAMRA
	hTel	FAM-GGGTTAGGGTTAGGGTTAGGG-TAMRA
	dsDNA	FAM-CTATAGCGCGCTATAG-TAMRA
***Taq* polymerase Stop Assay**	LTR-III	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGAG-TGGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC
	LTR-IV	TTTTTGGGCGGGACTGGGGAGTGGTTTTTCTGCATA-TAAGCAGCTGCTTTTTGCC
	hTel	TTTTTGGGTTAGGGTTAGGGTTAGGGTTTTTCTGCAT-ATAAGCAGCTGCTTTTTGCC
	No G4	TTGTCGTTAAAGTCTGACTGCGAGCTCTCAGATCCTG-CATATAAGCAGCTGCTTTTTGCC
	Primer	GGCAAAAAGCAGCTGCTTATATGCAG
***RT* Stop Assay**	RNA U3 III+IV	GGGAGGCGUGGCCUGGGCGGGACUGGGGAGUG-GCGAGCCCUCAGAUCCUGCAUAUAAGCA
	RNA no G4	GUAACCGAUGAGUCUAUGCGAGCCCUCAGAUCCUGC AUAUAAGCA
	Primer RT	TGCTTATATGCAGGATCTGAGG
SPR	LTR-III	GGGAGGCGTGGCCTGGGCGGGACTGGGG-biot
	LTR-IV	TGGGCGGGACTGGGGAGTGGT-biot
	hTel	GGGTTAGGGTTAGGGTTAGGG-biot
NMR	LTR-III	GGGAGGCGTGGCCTGGGCGGGACTGGGG
	LTR-III+IV	AGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGGT
	LTR-II+III+IV	GGGGACTTTCCAGGGAGGCGTGGCCTGGGCGG-GACTGGGG
	LTR-FL	GGGACTTTCCGCTGG-GACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGG
	LTR-III G1A	AGGAGGCGTGGCCTGGGCGGGACTGGGG
	LTR-III G2A	GAGAGGCGTGGCCTGGGCGGGACTGGGG
	LTR-III G3T	GGTAGGCGTGGCCTGGGCGGGACTGGGG
	LTR-III ΔG3	GGAGGCGTGGCCTGGGCGGGACTGGGG

	LTR-III G10A	GGGAGGCGTAGCCTGGGCGGGACTGGGG
	LTR-III G11A	GGGAGGCGTGACCTGGGCGGGACTGGGG
	LTR-III G5A-C13T	GGGAGACGTGGCTTGGGCGGGACTGGGG
	LTR-III T14A	GGGAGGCGTGGCCAGGGCGGGACTGGGG
EMSA experiments	RNA U3 III	GGGAGGCGUGGCCUGGGCGGGACUGGGG
	RNA U3 IV	GGGCGGGACUGGGGAGUGG
	RNA U3 III+IV	GGGAGGCGUGGCCUGGGCGGGACUGGGGAGUGG
	RNA U3 III+IV m5	GGGAGGCGUGGCCUGGGCGUGACUGGGGAGUGG
	RNA U3 III+IV m4+5	GGGAGGCGUGGCCUGUGCGUGACUGGGGAGUGG
	RNA scrambled	GAGCGUGCGCGUGCGAGCGUGAGUGAGCGUGGG
	RNA TAR	GGCAGAUCUGAGCCUGGGAGCUCUCUGCC
	LTR-III+IV compl	CCACTCCCCAGTCCCGCCCAGGCCACGCCTCCC

Table 3.1 Oligonucleotide sequences used in this study

3.2 Circular Dichroism (CD) analysis

CD is a polarized light spectroscopy that is used to discriminate G4 structures from other secondary structures. CD spectra generated by G4 structure display characteristic shapes useful for structure formation recognition and conformational analysis (167). Indeed, it is possible to have preliminary indications of G4 formation and adopted folding topology by simply analyzing the CD spectra. Parallel G4s give rise to CD spectra characterized by a maximum peak at 260 nm and a minimum peak at 240 nm, antiparallel structures display one positive and one negative peaks, at 295 nm and 260 nm, respectively, while mixed G4 conformations exhibit two positive peaks at 295 nm and 260 nm. Thermal stability of G4s can be studied following the variations of molar ellipticity corresponding to the maximum peaks with temperature increase. Thermal melting experiments allow the calculation of melting temperature (T_m) values, by plotting molar ellipticity vs temperature. CD spectroscopy is al-

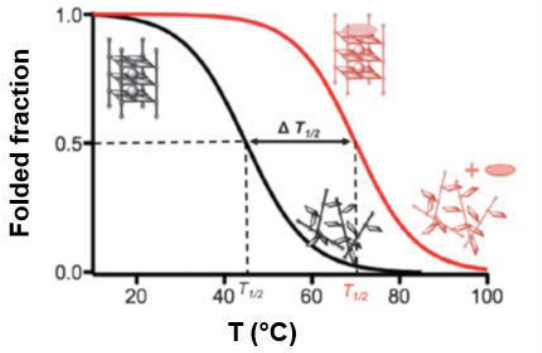


Figure 3.1 Schematic representation of melting curves obtained from CD melting experiments. In black, the melting curve of G4, in red, the melting curve of G4-ligand complex. The T_m corresponds to the 0.5 fraction of folded G4 structure. ΔT_m is the difference between T_m of DNA in presence of stabilizing compound and T_m in absence of G4 ligand. Figure from (168)

so useful for G4-ligand interaction analysis. The simple titration with a G4 ligand may induce changes in the CD spectrum suggesting conformational changes induced by the ligand binding. It is also possible to study the stabilization operated by G4 ligands by following the Tm variation upon addition of different amounts of G4 binders (168)(Figure 3.1).

For CD analysis, DNA and RNA oligonucleotides were diluted to a final concentration of 4 μM in lithium cacodylate buffer (10 mM, pH 7.4) added with appropriate concentrations of KCl (from 0 to 100 mM). After annealing step (95 °C for 5 min), samples were gradually cooled to room temperature to allow G4 formation. When the interaction with G4 ligands was analyzed, compounds were added from stock at final concentration of 16 μM (4:1 ligand/oligonucleotide ratio). Blank samples were prepared by addition of the same volumes of compound solvent (i.e. DMSO).

CD spectra were recorded on a Chirascan-Plus (Applied Photophysisics, Leatherhead, UK) equipped with a Peltier temperature controller using a quartz cell of 5 mm optical path length and an instrument scanning speed of 50 nm/min 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 Tm, spectra were recorded every over a temperature range of 20−95 °C, with temperature increase rate of 5 °C/min, followed by an equilibration step of 1 min before measurement. Tm values were calculated according to the van't Hoff equation, applied for a two-state transition from a folded to unfolded state, assuming that the heat capacity of the folded and unfolded states are equal.

3.3 UV analysis

Similarly to CD spectroscopy, UV analysis is widely used for investigation of G4 formation, thermal stability and interaction with ligands. By UV spectroscopy, it is possible to follow the G4 melting at 295 nm, wavelength where the difference between folded and unfolded states of G4 structure results to be measurable (169). Registering melting curves over 20 °C – 95 °C temperature range allows melting temperatures calculation. The advantage of UV technique in this case relies in the possibility of using high nucleic acids concentration samples. Measuring Tm values at low and high oligonucleotides concentrations allows to assess the molecularity of G4s. Melting temperatures of monomeric structures are independent from strand concentration variation, while melting temperatures of bi- or tetramolecular G4s varies with strand concentration increase.

The difference between the UV absorbance spectra (220-320 nm wavelength range) of folded form (registered at temperatures below T_m) and unfolded form (registered at tempera-

tures above T_m) gives rise to the Thermal Difference Spectrum (TDS) (20). TDS shapes, characteristic of G4 structures display four distinct positive/negative peaks around 243, 255, 273 and 295 nm and further confirm G4 folding.

For UV melting analysis, the same sample preparation protocol as for CD analysis was used. Thermal denaturing experiments were performed on Jasco V-650 UV spectrometer. DNA samples at 5 uM or 100 uM strand concentration were heated to 95°C for 5 minutes and cooled down to 20 °C by a temperature ramp of 0.1 °C per minute, followed by the reverse step of heating from 20 °C to 95°C at the same temperature ramp. UV absorbance at 295 nm were measured every 0.5 °C to follow the G4 folding and unfolding steps. Obtained data were plotted as folded fraction against temperature and the melting temperature was determined as the value corresponding to the 0.5 of the folded fraction.

For TDS analysis, UV spectra over a wavelength range of 220–320 nm were recorded on a Lambda25 UV/Vis spectrometer (Perkin-Elmer) equipped with a Peltier temperature controller. TDS were calculated by subtracting the UV spectrum at 20 °C from the spectrum at 95 °C.

3.4 Fluorescence Resonance Energy Transfer (FRET) analysis

FRET-based method was developed by Mergny group (170). Oligonucleotides, able to fold into intramolecular G4s, are labelled at 3'- and 5'-ends with a fluorophore and a quencher, respectively. Formation of G4 structure brings the fluorophore and the quencher in close proximity and the fluorescence is quenched. 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 the appropriate wavelength to obtain melting curves (Figure 3.2). The 5'-fluorescein and 3'-tetramethylrhodamine fluorophore-quencher couple have been used in our studies.

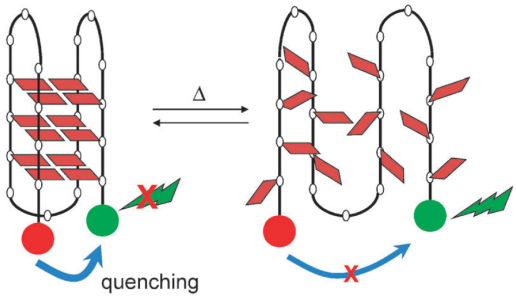


Figure 3.2 Schematic representation of FRET-based melting temperature experiments. When G4 structure is in the folded form, the fluorescence (green) is quenched (the quencher is represented in red). During the structure unfolding the distance between the fluorophore and the quencher increase and the fluorescence is restored. Figure from (168)

The 5'-FAM (6-carboxyfluorescein) and 3'-TAMRA (6-carboxy-tetramethylrhodamine) labeled oligonucleotides used in this study were purchased from Sigma-Aldrich (Milan, Italy), diluted in Tris-EDTA buffer and stored at -20 °C. For fluorescence melting curves 20 ul samples of oligonucleotides at 250 nM concentration were annealed in the lithium cacodylate buffer supplemented with potassium (100 mM) for 5 mins at 95 °C and slowly cooled to room temperature to allow G4 formation. Equimolar amounts of tested compounds were added to the sample 4h after annealing. After overnight stabilization at 4 °C, samples were processed in a LightCycler II LightCycler II (Roche, Milan, Italy) or LightCycler 480 (Roche, Milan, Italy), and the oligonucleotide melting was monitored by observing 6-carboxyfluorescein (6-FAM) emission in the temperature range of 30– 95 °C with 1 °C/min gradient. Melting profiles were normalized as previously described. Tm was defined as the temperature corresponding to the 0.5 fraction of the normalized fluorescence.

3.5 *Taq* polymerase and *RT* stop assays

Taq polymerase stop assay is used to investigate the formation of G4 structures in a DNA sequence based on the assumption that a G4 sterically impedes the progression of a DNA polymerase during the elongation of a ^{32}P-labeled primer (171) (Figure 3.3). When the template containing G4 forming regions is elongated by a labeled primer, truncated synthesis products are formed. These incomplete DNA synthesis products are discriminated in a denaturating polyacrylamide gel. To evaluate whether the stop bands correspond to the guanines involved in the G4 formation, the specific markers are used. The markers are prepared based on Maxam & Gilbert sequencing. Specific chemically induced DNA breakings are pro-

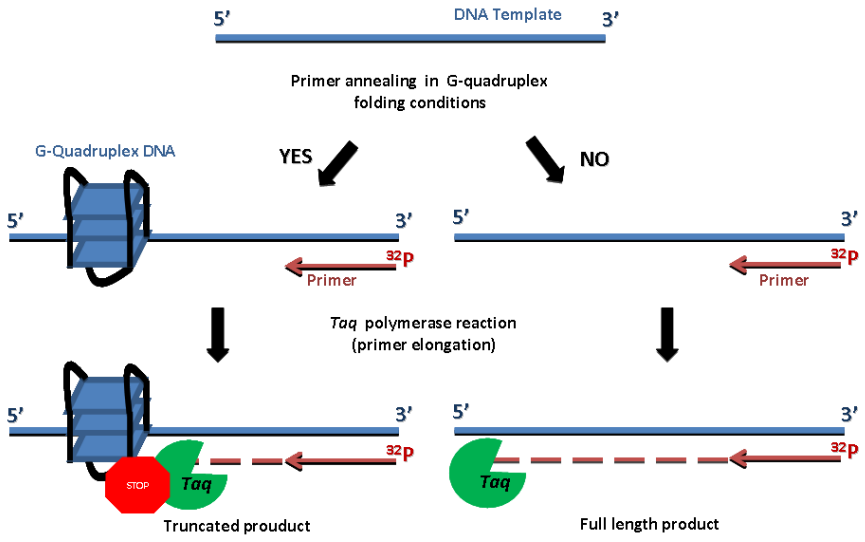


Figure 3.3 Schematics of *Taq* polymerase stop assay, described in the main text.

duced by treatment of the double stranded sequence first with formic acid and then with piperidine. Formic acid produces N7 guanine and N3 adenine methylation making these bases more susceptible for hydrolysis subsequently produced by piperidine treatment.

The DNA primer was 5'-end labeled with [γ-^{32}P-ATP] using T4 polynucleotide kinase (Fermentas) at 37 °C for 30 mins. The labeled primer was purified with Illustra MicroSpin G-25 columns (GE Healthcare, Life Sciences). A reaction mix containing labeled primer (72 nM) and DNA template (36 nM) in 10 mM lithium cacodylate buffer supplemented with appropriate concentration of potassium ions (0 – 100 mM KCl), was heated at 95 °C for 5 mins and gradually cooled to room temperature to allow primer annealing and G4 formation. When the stabilizing effect of G4-binding compounds was tested, an appropriate concentration of the ligand was added to the reaction mix after annealing and the samples kept at 4 °C o/n to allow G4 stabilization. 2U/reaction of Ampli*Taq*Gold DNA polymerase (Fermentas) were added to initiate the single cycle elongation reaction at 47 °C for 30 mins. Reactions were stopped by EtOH precipitation, resuspended in denaturating loading dye (phormamide, EDTA 0.5 M pH 8) and heated at 95 °C for 3 mins prior loading into 16% denaturating polyacrylamide gel. Separated extension products were visualized by phosphorimaging (Typhoon FLA9000, GE Healthcare). Markers were prepared by PCR reaction with ^{32}P-labeled primer. PCR products were treated with formic acid for 5 mins at 25 °C and then with piperidine for 30 mins at 90 °C and loaded into the denaturating gel together with elongation reaction products.

The principles of RT stop assay are the same as for *Taq* polymerase stop assay with the difference that the RNA template is elongated by a reverse transcriptase enzyme from a ^{32}P-labeled DNA primer. It is worth to notice that RNA sequences exhibit a higher potential of forming different secondary structures, for instance hairpins, which may as well sterically impede reverse transcriptase progression. To correlate the truncated products to the RNA G4 formation, markers are used. Moreover, performing the elongation reactions at higher temperature may help to avoid secondary structure formation without compromising G4 stability.

DNA primer was 5′-labelled with [γ-^{32}P-ATP] using T4 polynucleotide kinase at 37 °C for 30 mins. The labelled primer (100 nM) was annealed to the appropriate template (70 nM) in the absence or presence of various concentrations of KCl. The primer extension reaction was performed by Recombinant HIV-1 Reverse Transcriptase (1 U/reaction; Calbiochem) at 44 °C for 1 h. Where specified, samples were incubated with increasing concentrations of BRACO-19 (125 nM–8 mM) for 1 h at room temperature before primer extension. When nucleocapsid protein was used, appropriate concentrations of the protein were added immediately prior to the elongation reaction. Reaction products were treated with NaOH (2 N) at

95 °C for 3 min to permit the alkaline hydrolysis of RNA and the pH was adjusted with HCl (2 N) to neutrality. Samples were ethanol precipitated and extension products were separated on 16% denaturating gel and visualized by phosphorimaging (Typhon FLA9000; GE Healthcare).

3.6 Surface Plasmon Resonance (SPR) analysis

The surface plasmon resonance instrument detects a decrease in intensity of polarized light reflected from the surface of a glass prism coated with a thin metal film (172). The minimum in reflected light intensity occurs at a specific angle at which surface plasmons of the metal are excited. The evanescent electromagnetic field extends a short (200 nm) distance from the reverse surface of the film into the surrounding medium, with the consequence that the angle at which resonance occurs is function of the refractive index of the medium in the vicinity of the surface. The refractive index is a function of the composition of the interface, and hence binding interactions which bring about concentration of a species at the surface of the metal film can be detected. 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 (173). The biotin-functionalization is performed to immobilize the oligonucleotides 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 nucleotides, because of the relatively high molecular weight of the proteins compared to the oligonucleotides and because of the advantages in a small amounts of the protein necessary for immobilization. Proteins are easily immobilized on the dextran sensor chip surface by amine coupling.

3.6.1 SPR analysis of G-quadruplex – c-exNDI compounds interaction

SPR was performed on the Biacore T100 platform (GE Healthcare). 5′-Biotinylated LTR-III, LTR-IV, and hTel oligonucleotides were heated at 95 °C for 5 min and cooled down at room temperature to allow G4 folding. Immobilization was performed in HEPES-KCl running buffer (0.01 M HEPES pH 7.4, 0.2 M KCl, 3 mM EDTA) on a steptavidin coated surface (SA sensor chip, Biacore). Oligonucleotides were diluted in HEPES-KCl running buffer to a concentration of 30 nM and injected to reach the response of around 500 RU. Flow cell 1 was left empty to allow reference subtraction. C-exNDIs binding analysis was performed at a flow rate of 30 μL/min, with contact time of 280 s and dissociation time of 360 s in HEPES-KCl buffer. Sensorgrams were obtained in the concentration range of 25–200 nM. After each compound injection the chip surface was regenerated with glycine 10 mM pH 2.0 solution

(GE Healthcare). All sensorgrams were corrected by buffer injection response. Data were fitted to a 1:1 binding model with Rmax initial parameter set to theoretical calculated Rmax using BIAevaluation software (GE Healthcare). All experiments were performed independently at least twice, and in each instance χ^2 values, which indicate the reliability of the fitting, were below 0.2.

3.7 NMR spectroscopy

NMR spectroscopy is widely used for structural studies of G4 forming nucleic acids. This technique allows the determination of high-resolution structures of G4s, their dynamics, stability and intermolecular interactions (174-176). The high resolution structures provide structural coordinates which could permit the identification of ligands specific for the given G4 by putting in evidence particular features to target and improve selectivity of G4 binders. A bench of biologically relevant G4 structures have been studied with NMR technique e their high-resolution structures in salt solutions have been successfully solved. However, NMR analysis of interaction with G4 binding ligands is more challenging, as it requires particular conditions and crystallographic analysis results more suitable to this purpose. In fact, only a few ligand-G4 complexes structures have been solved by NMR technique.

3.7.1 One-dimensional (1D) ^{1}H NMR analysis

NMR analysis can be applied at several levels. In ideal circumstances, an oligonucleotide that forms a single kinetically stable structure in buffer solution is suitable for a detailed structural analysis. Limited information about the structural state of the G4 is available when more than one G4 form is present in solution or the original G4 is kinetically unstable. A simple 1D ^{1}H NMR spectrum of a G4 forming sequence provides initial structural information. It is possible to identify distinct regions of the spectrum where distinct proton groups resonate (176). Imino protons (H1) chemical shifts span between 10-14 ppm range and can be further discriminated within these region, as imino protons involved in Watson-Crick base pairs resonate above 12.5 ppm, while resonances of imino protons interacting through Hoogsteen hydrogen bonds are usually in the chemical shift range between 10 – 12 ppm. Aromatic protons (H8 of guanines, H8 and H2 of adenines, H6 of thymines and cytosines) resonate in the 7 – 8 ppm region (Figure 3.4).

The quality of the ^{1}H spectrum of the exchangeable proton resonances (H1 of guanines) is a good indication of how suitable is the G4 for structural analysis. The presence of countable, well-resolved peaks in the Hoogsteen hydrogen bonds resonances region (10.5 – 12 ppm) and in the aromatic protons region is usually the first indication of single conformation structure in solution suitable for NMR structural analysis.

The number of resonances in 10.5 – 12 ppm region is indicative of number of H1 protons involved in Hoogsteen hydrogen bonds formation and thus, of number of G-tetrads forming the structure.

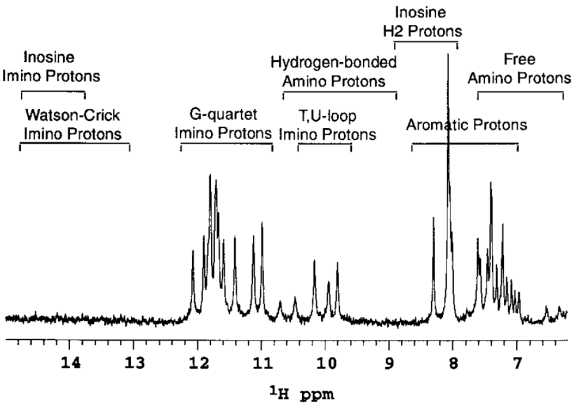


Figure 3.4 1D ^{1}H NMR spectrum of $d(G_3T_4G_3)_2$ in aqueous buffer. Overlaid on the spectrum are frequency ranges of the resonances of various protons observed in G4 structures. Figure from (176)

3.7.2 Unambiguous spectral assignment and site-specific labeling techniques

The elucidation of G4 structure starts from unambiguous spectral assignment, employing site-specific low enrichment isotope labeling approach (174). Imino protons are usually assigned by site specific 2 – 4% ^{15}N-enrichment (177), while aromatic protons can be assigned by site-specific ^{2}H labeling (Figure 3.5). Through-bond correlations (^{1}H-^{13}C)-HMBC (coupling of H1 and H8 protons from the same guanine via $^{13}C5$) can be used to confirm the assignments of imino and aromatic protons (178).

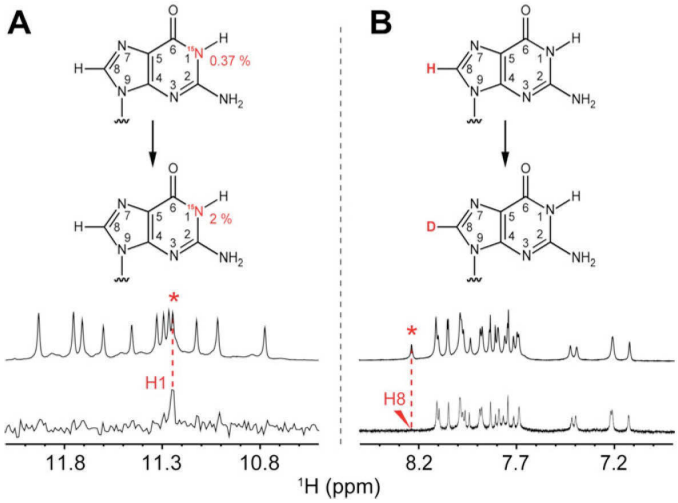


Figure 3.5 Techniques for unambiguous spectral assignment. **(A)** Imino protons assignment by site-specific low-enrichment 15N labeling. **(B)** Aromatic protons assigned by site-specific deuterium labeling. Replacement of H8 of guanines with deuterium results in disappearing of corresponding peak from the spectrum. Figure from (174)

3.7.3 G-quadruplex folding determination

Specific NOE patterns are used for G4 folding determination. Particularly, within a G-tetrad cyclic connectivities can be established between the imino (H1) proton of a guanine and the aromatic H8 proton of the adiacent guanine. Tracing this type of connectivities in a water NOESY spectrum allows to define the composition and alignment of the tetrads. Several H1/H1 short distances can be detected between stacked G-tetrads and be useful for tetrads alignment determination. Solvent exchange experiments independently confirm the G4 folding topology. When the sample is dissolved in deuterium solution, imino protons exchange with the solvent and disappear from the spectrum. The internal tetrad remains protected from the exchange for a certain period of time allowing the detection of the resonances belonging to the guanines participating to the central tetrad formation.

NMR experiments were performed on 600 MHz Bruker NMR spectrometer. All the DNA samples were processed in the same buffer conditions, i.e. 20 mM potassium phosphate – 70 mM potassium chloride buffer pH 7. Reported 1D NMR spectra were recorded at 25 °C, whether other temperature conditions are specified.

Spectral assignment: H1 resonances of guanines were assigned by site-specific low enrichment ^{15}N labeling, with the exception of G11, which H1 was indirectly assigned by NOESY at 10 °C, mixing time 75 ms. H8 resonances of guanosine participating to the G-tetrads were assigned by heteronuclear multibond correlations (^{1}H-^{13}C)-HMBC at 35 °C. Site specific ^{2}H-labeling of the guanines confirmed through-bond assignments. Assignments were supported by TOCSY and ^{13}C-HSQC experiments. NOESY experiments in H_2O solution were performed at 35 °C, mixing time 200 ms and at 10 °C, mixing time 75 ms. NOESY experiments in $^{2}H_2O$ solution were performed at 35 °C and two different mixing times, 100 ms and 300 ms.

3.8 Electrophoretic Mobility Shift Assay (EMSA)

Native gel electrophoresis is routinely used for discrimination of unfolded nucleic acid sequences from G4 and of different G4 species providing a simple tool for G4 structures characterization. For instance, multimolecular and monomolecular species can be detected by native PAGE separation based on their different electrophoretic mobility. This technique is also useful for G4-small molecules and G4-protein interactions, again based on different electrophoretic mobility of complexes and free species. In this case, EMSA assay was used with both purposes.

3.8.1 G4 molecularity definition

For LTR-III molecularity investigation, DNA folded samples at 100 μM strand concentration in potassium-phosphate buffer were loaded on 15% native polyacrylamide gel containing

10 mM of KCl. The gel was run at 90V for 30 minutes at room temperature in TBE – KCl buffer and DNA bands were visualized by UV shadowing. Control sequences with known molecularity, dimeric T95 and monomeric T95-2T sequences, were used to ensure the meaningful comparison.

3.8.2 RNA G4-cellular proteins interactions

For EMSA experiments RNA oligonucleotides, labeled with [γ-^{32}P-ATP] using T4 polynucleotide kinase at 37°C for 30 min, were annealed by heating at 95 °C for 5 min in lithium cacodylate (10 mM, pH 7.4) and KCl (100 mM) buffer and gradually cooled to room temperature to allow G-quadruplex formation. The annealed oligonucleotides at 15 nM final concentration were added to 20ul of binding reaction containing 15 ug of appropriate fraction of cell extract for 2 h at 37 °C. For competition expertiments the appropriate concentrations (15 nM, 150 nM, 750 nM, 1,5 uM and 3 uM) of not labeled competitor oligonucleotides were added to the binding reaction. Reactions were loaded on a 6% polyacrilamide native gel and run for 19 h and 30 min at 27V. Gels were dryed, exposed overnight and visualized by prosphorimaging (Typhoon FLA 9000, GE Healthcare).

3.8.2.1 Cells and protein extraction

HEK293T cells were cultured in DMEM containing 10% fetal bovine serum. Cells were grown at 37 °Cin humidified atmosphere with 5% of CO_2.
For the preparation of the fractioned (cytoplasmatic and nuclear) cell extracts CelLytic NuCLEAR extraction kit (Sigma-Aldrich) was used according to manufacturer's instructions. The concentration of extracted proteins was quantified by BCA Protein Assay kit (Thermo Scientific Pierce).

3.8.3 RNA G4 – nucleocapsid protein interactions

For EMSA experiments RNA oligonucleotides, labeled with [γ-^{32}P-ATP] using T4 polynucleotide kinase at 37°C for 30 min, were annealed by heating at 95 °C for 5 min in lithium cacodylate (10 mM, pH 7.4) and KCl (100 mM) buffer and gradually cooled to room temperature to allow G-quadruplex formation. The annealed oligonucleotides at 15 nM final concentration were added to 20 µl of binding reaction (8% Glycerol, 30 mM Tris-HCl, 15mM $MgCl_2$, 50 µM $ZnCl_2$) containing appropriate concentrations of nucleocapsid protein. For EMSA unfolding assays, labeled RNA oligonucleotides were annealed to form G4s and cold DNA complementary oligonucleotides were added to the binding reactions at equimolar or 2-fold excess strand ratio. Binding reactions were incubated for different period of time in the presence of

appropriate nucleocapsid protein concentrations. Control sample in absence of the protein were incubated for the same time range.

Reactions were loaded on a 12% polyacrilamide native gel and run for 90 mins at 90V. Gels were dryed, exposed overnight and visualized by prosphorimaging (Typhoon FLA 9000, GE Healthcare).

3.9 Clerocidin (CL) footprinting assays

Chemical footprinting assays are commonly used for G4 structures characterization. Particularly, DMS (Dimethyl sulfate) footprinting allows to determine which guanine repeats in the G4 forming sequence are involved in structure folding (171). To this purpose, the G4 forming DNA, labeled with ^{32}P on its 5'-end is annealed in the appropriate conditions to form G4. The resulting structure is treated with DMS to methylate N7 of the guanine residues in the oligonucleotides. The guanine bases involved in the Hoogsteen hydrogen bonds to form G-tetrads are not accessible for N7-methylation. The subsequent treatment with piperidine produces specific DNA strand breakage at methylated guanine residues, resolvable on denaturating PAGE gel. Guanines participating to the tetrads can be identified based on protection patterns of the sequence (171).

Clerocidin (CL) (Figure 3.6A) is a natural product that has been shown to react at ssDNA regions, with different mechanisms depending on the exposed nucleotide. In particular, CL electrophilic groups target the nucleophilic N7 of guanine inducing spontaneous depurination and DNA strand cleavage and the NH_2 and N3 of cytosine (C) with formation of a stable condensed ring system, which is degraded to induce DNA cleavage only after hot alkali treatment and, finally, the NH_2 and N1 of adenine to generate an adduct that degrades upon alkali but does not result in DNA strand scission. Based on these characteristics, when a DNA subjected to alkylation by CL is loaded on a sequencing gel, differential reactivity at DNA bases is manifested (179) (Figure 3.6).

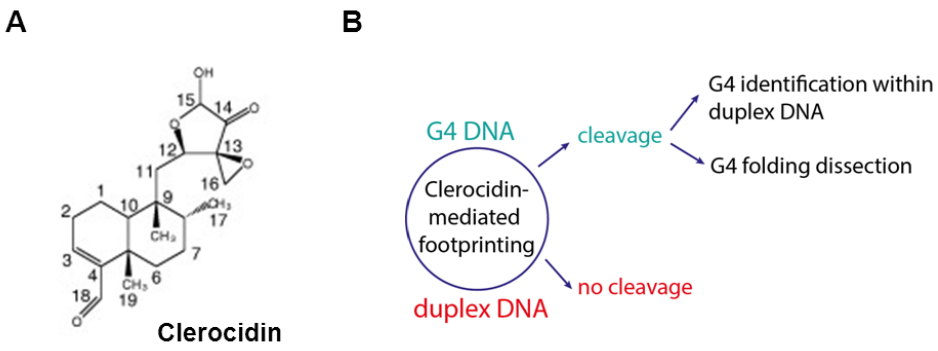


Figure 3.6 (A) Chemical structure of Clerocidin. **(B)** Schematic representation of clerocidin mediated footprinting assay.

All oligonucleotides were gel-purified before use and prepared in desalted/lyophilized form. Oligonucleotides were 5′-end-labeled with γ- ^{32}P-ATP by T4 polynucleotide kinase and purified by MicroSpin G-25 columns (GE Healthcare, Europe). They were next resuspended in lithium cacodylate 10 mM,pH 7.4, and KCl 100 mM, heat-denatured, and folded. Reactions of the labeled G-quadruplex folded oligonucleotides (4 pmol/sample) with increasing amounts of compound 2 (4−100 nM) were performed at 20 °C for 24 h in resuspension buffer. Samples were then reacted with CL (100 μM) at 37 °C for 24 h. Reactions were stopped by ethanol precipitation, resuspended, and either kept on ice or treated at 90 °C for 30 min with 1 M piperidine to complete strand scission according to the Maxam and Gilbert protocol. Samples were then lyophilized, resuspended in formamide gel loading buffer, and heated at 95 °C for 3 min. Reaction products were analyzed on 20% denaturing polyacrylamide gels and visualized by Typhoon FLA 9000 phosphorimaging analysis (GE Healthcare).

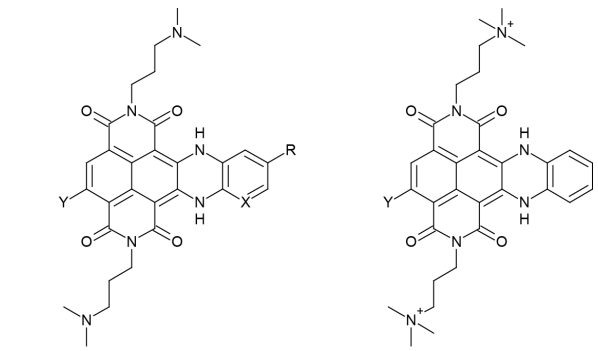
4 Results and Discussion

Part 1. HIV-1 LTR G-quadruplexes: structure and targeting with small molecules

4.1 Potent core-extended Naphthalene Diimides targeting the HIV-1 LTR G-quadruplexes.

The emerging role of G4 structures in human diseases, especially in tumorigenesis, encouraged the development of a wide spectrum of G4 binders. Our group demonstrated recently the possibility to inhibit the HIV-1 transcription by stabilizing the G4 structures in the LTR promoter and G4 binding small molecules have shown antiviral activity suggesting the possibility to develop antiviral compounds with G4-mediated mechanism of action (166).

Among G4 binders, the Naphtalene Diimide (NDI) family of compounds was extensively studied and displayed promising G4 binding properties with good selectivity towards G4 structures vs duplex conformation (105,107,110). Our collaborators, Prof. Freccero group, synthesized a novel series of water soluble core-extended NDIs (c-exNDIs) with NDI core merged to the dehydrobenzophenazine polynuclear heterocycle and carrying a variety of substituents (Figure 4.1). These compounds are characterized by a larger flat core than the NDIs, which ensures improved p-stacking interaction by extensive overlap to the quartets of G4 structure. Compounds with H and Br substituents in Y position were synthesized to evaluate the effect of bulky substituent in this position (Figure 4.1).



X = C, Y = H, R = H **2**
Y = Br, R = H **3**
Y = H, R = NO_2 **4**
Y = Br, R = NO_2 **5**
Y = H, R = COOH **6**
Y = Br, R = COOH **7**
X = N, Y = Br, R=H **8**

R = H, Y = $NH(CH_2)_3N(CH_3)_2$ **9**

Y = H **10**
Y = Br **11**

Figure 4.1 Chemical structures of c-exNDIs **2-11** used in this study

4.1.1 C-exNDIs greatly stabilize G4s with a preference towards HIV-1 LTR conformations vs the telomeric sequence.

To assess the ability of the newly synthesized c-exNDI compounds to bind and stabilize HIV-1 LTR G4s two sequences composed of four G-tracts and able to fold in different G4 conformations within G4 forming sequence of HIV-1 LTR promoter region have been chosen for the analysis, named LTR-III and LTR-IV (Figure 4.2).

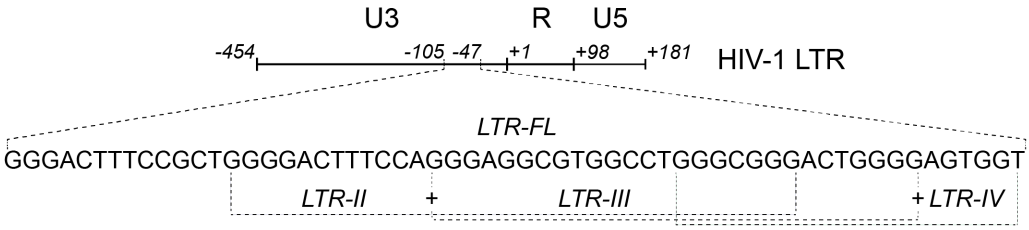


Figure 4.2 Schematic representation of HIV-1 LTR G4 forming sequence. Subsequences forming different G4 conformation are marked and named LTR-II, LTR-III and LTR-IV.

LTR-III 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 studies. LTR-IV is slightly less stable and in full-length context folds only when induced and stabilized by G4 ligands. To evaluate the binding affinity and stabilization performed by core-exNDIs towards cellular G4s, a highly represented intracellular telomeric (hTel) G4 sequences composed of four TTAGGG repeats has been used. A dsDNA has been employed as control sequence to confirm the binding specificity towards G4 vs duplex conformation. The

melting temperatures of LTR-III, LTR-IV, hTel and dsDNA were 66.9 °C, 61.9 °C, 66.9 °C and 66.5 °C respectively, sufficiently similar to allow a meaningful comparison.

First of all, the whole series of c-exNDI compounds has been tested by FRET melting assay in order to evaluate the degree of stabilization operated by these compounds on viral G4s compared to hTel sequence and dsDNA control. We first tested the 4-fold molar excess of the compound, typically used for this kind of assay, to assess their stabilizing activity. The results obtained in these experimental conditions suggested that c-exNDIs have a great stabilization effect on G4-forming sequences LTR-III, LTR-IV and hTel with an increase of melting temperatures of above 25 °C, which is the maximum increase measurable in these settings (Table 4.1). The melting temperature increase registered for non-G4 dsDNA control was in the range 2 – 7 °C for compounds **2 – 9** suggesting the preferential binding toward G4 forming sequences, while compounds **10** and **11** stabilized dsDNA up to 80 °C (Table 4.1). These data suggested that c-exNDI compounds have a high propensity to selectively bind G4 forming sequences vs duplex conformation.

c-exNDI	Tm (°C) LTR-III/c-exNDI 1 / 4	Tm (°C) LTR-IV/c-exNDI 1 / 4	Tm (°C) hTel/c-exNDI 1 / 4	Tm (°C) dsDNA/c-exNDI 1 / 4
2	>95	>95	>95	72.9
3	86.9	83.2	>95	72.4
4	>95	>95	>95	73.9
5	>95	83.4	>95	69.9
6	78.4	76.9	86.4	72.5
7	77.6	76.9	83.4	69.9
8	79.2	77.2	78.5	69.5
9	73.5	73.9	77.9	68.5
10	>95	>95	>95	79.9
11	>95	>95	>95	80.9

Table 4.1 Melting temperatures of LTR-III, LTR-IV, hTel and dsDNA at 0.25 μM concentration in presence of 4-fold molar excess of c-exNDIs (1 μM) obtained from FRET analysis

Because the oligonucleotides used for FRET analysis display fluorophore-modified ends that may influence the measured stabilization, additional techniques using label-free oligonucleotides were employed. Thus, the ability of c-exNDIs was also evaluated by circular dichroism technique. CD analysis relative to the representative compound **2** is displayed in Figure 4.3. Compound **2** was assayed on LTR-IV, LTR-III and hTel G4s. At 4:1 c-exNDI:DNA ratio (at 16 μM c-exNDI conc.), a stabilization higher than 27°C was obtained on both LTR-III and LTR-IV, while hTel G4s was stabilized by only 19.9°C, confirming excellent G4 stabilizing properties of the compound (Figure 4.3). The peak at 290 nm, which is diagnostic of the an-

tiparallel G4 conformation, increased in both LTR-III and hTel sequences upon addition of the c-exNDI, indicating the ability of the compound to drive the folding of both sequences from an initial mixed-type towards an antiparallel-like topology (Figure 4.3). In contrast, LTR-IV maintained a parallel conformation upon addition of compound **2**.

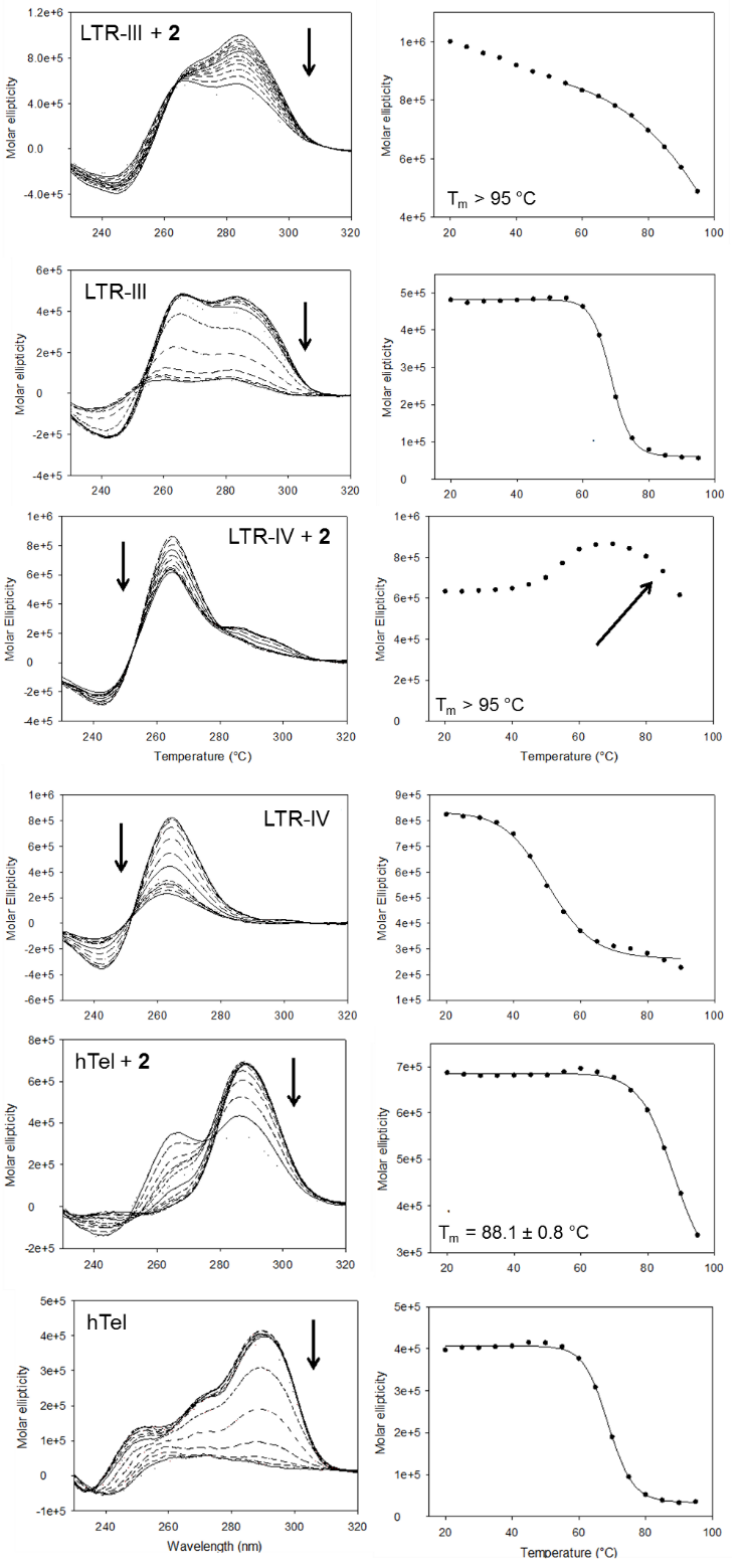


Figure 4.3 CD analysis of LTR-III, LTR-IV and hTel G4 oligonucleotides in absence and in presence of 4-fold excess of compound c-exNDI **2**. On the left panels CD spectra variation in function of the temperature are represented; arrows indicate the spectral change from low to high temperatures. The right panels display melting curves extrapolated from the CD spectra at the maximum peak. Corresponding molar ellipticities are plotted in function of temperature.

To better differentiate the stabilization efficiency of c-exNDIs we performed FRET melting assays in presence of equimolar amounts of target oligonucleotide and compound. Among the tested c-exNDIs, **2**, **3**, **4**, **10**, and **11** were the most efficient ligands, with a stabilization on LTR-IV G4 of 10−16°C at equimolar amounts. The other c-exNDIs showed intermediate (4 °C ≤ΔTm < 7 °C, **5**, **6**, and **7**), or low (ΔTm < 4 °C, **8** and **9**) level of stabilization on LTR-IV G4. Compounds **2** and **10** stabilized LTR-III G4 above 10 °C (Figure 4.4 and Table 4.2). Among the G4 structures, hTel G4 was the least efficiently stabilized, whereas stabilization on dsDNA was negligible. Only compounds **10** and **11** displayed a significant stabilization on dsDNA, indicating that the presence of two permanent charges (the quaternary ammonium moieties) favors an unspecific binding to the negatively charged nucleic acid. In general, the H series of all derivatives (Y = H, **2**, **4**, **6**, **10**; Figure 4.1) was more active and more selective for the viral G4s compared to the Br substituted ones (Y = Br, **3**, **5**, **7**, **8**; Figure 4.1). Compounds **2**, **3**, and **10** exhibited a difference in stabilization of LTR-IV G4 vs hTel G4 equal or above 6 °C, which roughly corresponds to 40% of the overall stabilization efficiency (Figure 4.3 and Table 4.2).

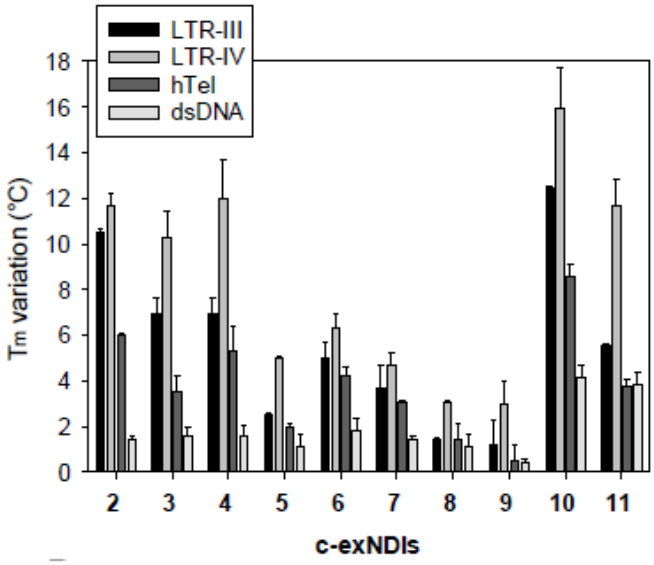


Figure 4.4 ΔTm values of HIV-1 LTR G4-folded sequences (LTR-III and LTR-IV), the G4-folded telomeric sequence (hTel) and a control double stranded sequence dsDNA registered upon addition of equimolar amounts of core-exNDIs **2-11** (0.25 μM).

c-exNDI	ΔTm (°C) LTR-III/c-exNDI 1 / 1	ΔTm (°C) LTR-IV/c-exNDI 1 / 1	ΔTm (°C) hTel/c-exNDI 1 / 1	ΔTm (°C) dsDNA/c-exNDI 1 / 1
2	10.5 ± 0.1	11.7 ± 0.6	6.0 ± 0.1	1.5 ± 0.1
3	7.0 ± 0.7	10.3 ± 1.1	3.5 ± 0.7	1.6 ± 0.4
4	7.0 ± 0.7	12.0 ± 1.7	5.3 ± 1.1	1.6 ± 0.5
5	2.5 ± 0.1	5.0 ± 0.1	2.0 ± 0.1	1.1 ± 0.6
6	5.0 ± 0.7	6.3 ± 0.6	4.2 ± 0.3	1.8 ± 0.6
7	3.7 ± 1.0	4.7 ± 0.6	3.0 ± 0.1	1.5 ± 0.1
8	1.5 ± 0.1	3.0 ± 0.1	1.5 ± 0.7	1.1 ± 0.6
9	1.2 ± 1.1	3.0 ± 1.1	0.5 ± 0.7	0.5 ± 0.1
10	12.4 ± 0.1	16.0 ± 1.7	8.5 ± 0.6	4.5 ± 0.6
11	5.5 ± 0.1	11.7 ± 1.2	3.8 ± 0.3	4.1 ± 0.6

Table 4.2 Melting temperatures of LTR-III, LTR-IV, hTel and dsDNA at 0.25 μM concentration in presence of equimolar amounts of c-exNDIs (0.25 μM) obtained from FRET analysis.

A *Taq* polymerase stop assay was performed on the LTR-III, LTR-IV, and hTel templates to investigate wherever binding of c-exNDI compound **2**, which displayed best viral G4 binding properties in spectroscopic analysis, was able to selectively impair the elongation of viral G4 templates by polymerase enzyme. A control template unable to fold into G4 was used to exclude unspecific inhibition of the polymerase enzyme by the c-exNDI. Compound **2** induced an intense stop site at the most 3′ G tract in each G4-forming template (Figure 4.5A), indicating effective stabilization of the G4 conformation by the compound. Quantification of pausing bands corresponding to the G4 structure formation revealed a considerable increase in the intensity of stop bands corresponding to LTR-III and LTR-IV G4s, which was particularly sensitive to the lowest concentrations of **2**. The same concentrations of the compound did not produce comparable relevant effect on the telomeric template, confirming again a preferential stabilization of the viral G4s. (Figure 4.5B).

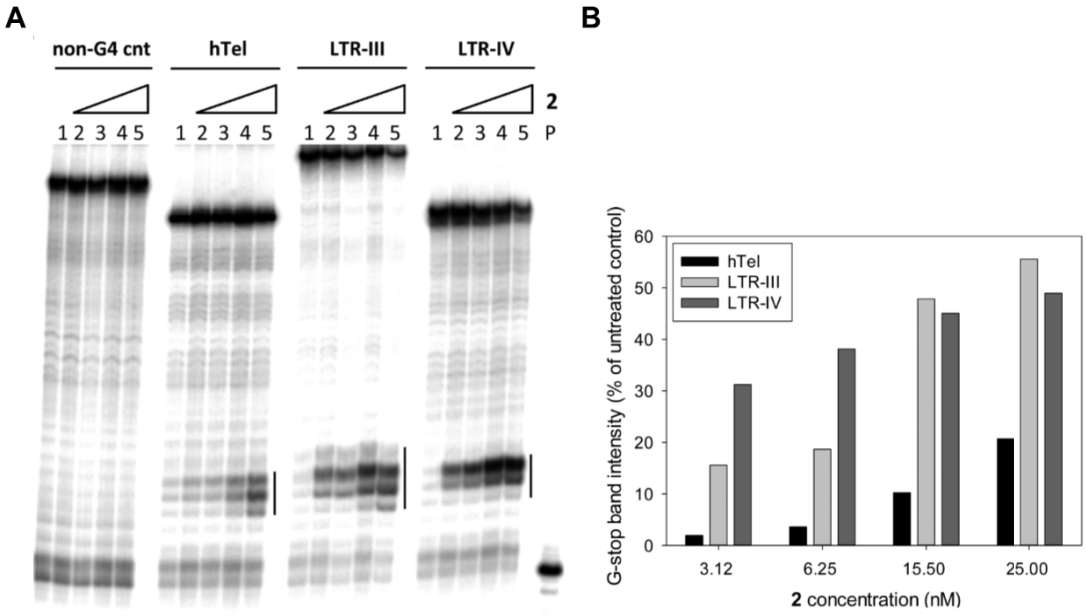


Figure 4.5 *Taq* polymerase stop assay. **(A)** *Taq* polymerization was performed in the presence of K+(lanes 1) and increasing concentrations of compound **2** (lanes 3 – 5), as indicated, on the hTel, LTR-III, and LTR-IV templates, which were previously folded into G4. A control template unable to fold into G4 was also used (non-G4 cnt). Stop regions are highlighted by vertical bars. **(B)** Quantification of the intensity of the stop bands obtained in the *Taq* polymerase stop assay. Stop band intensity of the treated samples were normalized vs the untreated sample.

Next, absolute affinity of **2** was measured by surface plasmon resonance (SPR) toward the three G4-folded sequences. LTR-III and LTR-IV G4s displayed higher association and lower dissociation rates, as represented by sensograms in Figure 4.6. Kinetic analysis confirmed that LTR-III G4 displays the highest binding affinity (K_D 3.2 nM), followed by LTR-IV G4 (K_D 6.1 nM); hTel G4 exhibited the lowest affinity (K_D 13.1 nM) (Figure 4.6). The binding affinity of NDI derivatives with each of the substituents in position R against LTR-IV was next as-

sessed. Compounds **3** and **5** had K_D values similar to that of **2** (K_D 6.6 and 5.9 nM, respectively), **10** displayed an increased binding affinity (K_D 1.7 nM), whereas **7** and **11** showed the

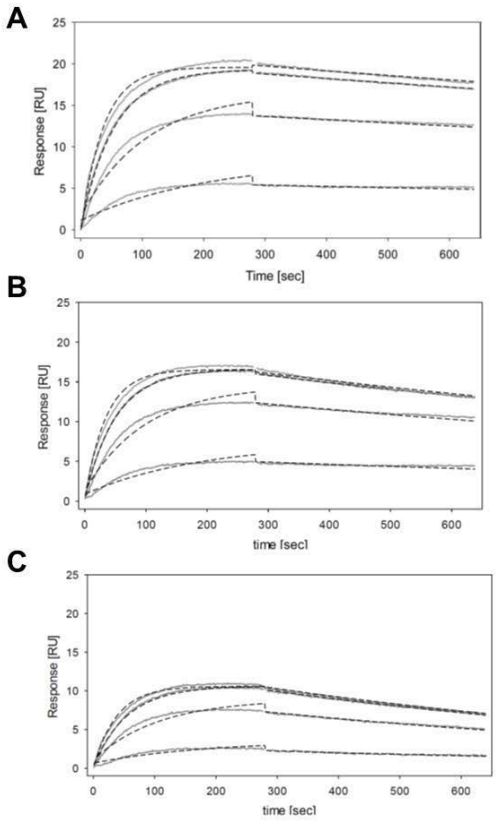


Figure 4.6 SPR analysis of compound **2** binding towards **(A)** LTR-III, **(B)** LTR-IV, **(C)** hTel G4s. The compound **2** was injected at increasing concentrations (50-100-200-300 nM). Registered sensorgrams are shown in grey solid lines. Dotted black lines represent 1:1 binding fits.

lowest affinity (K_D 21.6 and 12.9 nM, respectively). These data are in line with the stabilization properties observed for these derivatives by FRET and CD analysis and confirm the improved and reduced recognition properties of the c-exNDIs exhibiting permanent dicationic (**10**) and neutral or monocationic charge states (**7**), respectively, under physiological conditions. They also confirm that a bulky substituent in position Y of the NDI core (**11**) highly reduces binding, whereas a NO_2 group in position R maintains excellent binding properties (**5**).

In summary, FRET and CD melting data combined to *Taq* polymerase stop assay confirmed the preferential binding of representative of the best ligands compound **2** towards LTR G4s *vs* telomeric G4, with a slight further preference towards LTR-IV G4. On the other hand, SPR analysis and mass spectroscopy (not discussed here) suggested the increased binding towards LTR-III G4. This may rely on the fact that the increased binding affinity of the compound towards LTR-III compared to LTR-IV is not translated in relatively increased stabilizing properties due to higher dissociation rate.

We reasoned that the observed biased affinity for the LTR sequences could derive by specific binding of the c-exNDIs to structural/sequence features shared by the LTR G4s. To this end, we evaluated the protection exerted by compound **2** over residues of the LTR G4s against clerocidin (CL). This molecule has been shown to alkylate DNA single-stranded G, C, and A bases and to be able to predict G4 structural folding (180). Labeled oligonucleotides

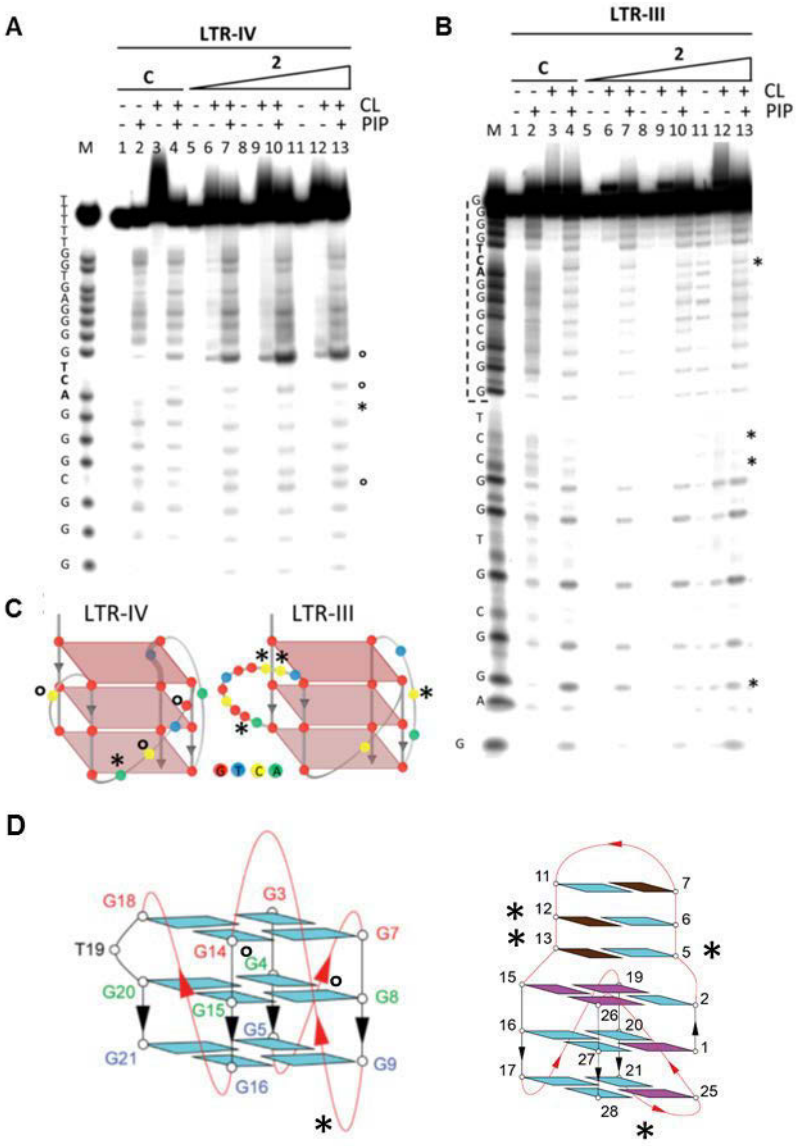


Figure 4.7 CL-mediated footprinting. LTR-IV **(A)** and LTR-III **(B)** oligonucleotides wee G4 folded, incubated with increasing amounts of 2 (lanes 5-13), alkylated with CL and subjected to hot piperidine treatment to induce cleavage of the phosphodiester bond at the alkylated bases. Lanes 1-4 correspond to samples not treated with c-exNDI **2**. Stars and circles indicate base protection and exposure, respectively. M is a marker lane evidencing G and A bases. The vertical dotted bar indicates the region of the LTR-III sequence shared with LTR-IV. **(C)** Schematics of the LTR-IV and LTR-III G4s topologies detected by low-resolution structural data. **(D)** Schematics of the LTR-IV and LTR-III G4s topologies obtained from NMR analysis. LTR-IV topology from (184)

were folded, treated with increasing amounts of **2**, exposed to CL, and further treated with piperidine to highlight CL-alkylated sites. In these settings, we expected to visualize only loop bases bound by the c-exNDI because base-paired Gs involved in G-quartet formation would not be available to CL alkylation, independently of the presence of the compound. The NMR structure of LTR-IV has been solved (PDB code 2N4Y), and therefore we were able to assess that nucleotides affected by the presence of the c-exNDI were in two loops of the structure: in particular, in the ACTG loop the A base was protected, whereas C and G were exposed; in the 1-base-loop, the C base was exposed (Figure 4.7). This protection and exposure pattern may suggest the preferential and more stable stacking of the compound **2** on the bottom tetrad of the LTR-IV G4 (Figure 4.7D).

Based on previous data, we hypothesized that LTR-III G4 can fold into different conformations, the main one of which was likely the one depicted in Figure 4.7C, detected by low-resolution structural data. In this case, we observed protection of the C base in the ACTG loop (Figure 4.7B), which is shared with the LTR-IV sequence (Figure 4.7); three additional residues (C, C, and G) were protected in the 11-base-loop (Figure 4.7A and B). Protection and exposure of residues indicated specific interaction of **2** at the affected LTR sites. Interestingly, unique LTR loops and in particular the ACTG loop shared by both LTR-III and LTR-IV, and absent in the telomeric sequence, were involved, indicating that these are likely the moieties that induce selectivity. We cannot exclude that other regions are bound by the c-exNDI, but probably the interaction at other sites is less stable or more dynamic so that a clear protection/exposure could not be observed.

We later obtained the LTR-III topology (Figure 4.7D), which will be described in details. Interestingly, we found that the 11-nt diagonal loop spanning through the top tetrad is structured in a 3-basepairing hairipin and two bases involved in the basepairing resulted to be protected in CL-footprinting assay (Figure 4.7D). Moreover, also C in the 3-nt loop facing the bottom tetrad resulted protected. This data suggest a possible binding of c-exNDI **2** on the external tetrads establishing additional stacking interaction with paired bases and providing further details for selectivity.

4.1.2 Antiviral activity of c-exNDI derivatives

The anti-HIV-1 activity of the new series of c-exNDI derivatives was next tested. Compounds **2**, **3**, **4** and **5** displayed inhibitory concentrations below 25 nM, with selectivity indexes (SI) around 20 for compounds **3**, **4**, **5** and 40 for compound **2**, best candidates also from spectroscopic analysis able to discriminate viral G4s from telomeric G4. Luciferase reporter assay confirmed that the c-exNDI activity at the LTR promoter was specific for the G4 conformation. An important inhibition of LTR activity was observed on the wt LTR in presence of

increasing amounts of compound **2**, while no effect was detected on the non-G4-forming mutant LTR sequence.

4.1.3 Discussion

We were able to assess the G4 binding properties, and antiviral activity of a new series of NDI derivatives with an extended aromatic core. First, most of the new compounds displayed improved stabilization on G4 structures, with substantial increments in T_m. This feature is an indication that the expansion of the aromatic surface allows a more effective recognition of the G4 structures as compared to previously reported NDI derivatives (108). This may be due to a more extensive overlap to the quartet of the G4 structure. Second, these compounds exhibited a very promising antiviral activity. The best antiviral **2–5** showed promising IC_{50} values in low nM range. This feature may derive from the selectivity observed *in vitro* for the LTR G4s over the telomeric G4 structures. The viral and telomeric G4s are likely the two most abundant DNA G4 species in the cell during infection, therefore a preferential effect on the viral G4s would result in higher antiviral activity and lower cytotoxicity, as observed. The c-exNDIs displayed both selective binding affinity and stabilization for the viral G4s. Between the two LTR G4s, the LTR-III sequence was most efficiently bound by **2**, as measured by SPR. Even though LTR-III (28 nts) and LTR-IV (19 nts) display an identical number of G quartets, LTR-III offers longer structured loops that may accommodate additional molecules and therefore may display increased binding affinity. Indeed, two loops in both LTR-III and LTR-IV G4s were shown to be involved in c-exNDI interaction by footprinting analyses. We suggest that the compounds interact with the edge G-quartets through the core NDI planar aromatic surface, while the extended moiety and side chains interact with the side loops, affording selectivity toward the LTR G4s.

Analysis of the binding activity of the series of c-exNDIs led us to the following structure–activity relationship (SAR) conclusions: (i) the non-extended side of the c-exNDI core position Y (Figure 4.1), lacking substituents as the introduction of an aminoalkyl side chain, highly hinders stabilization of the G4 structures (i.e., **9**); (ii) Br is allowed as Y substituent, but it lowers binding and selectivity toward the viral sequences. The Br group may both hinder specific interactions (hence lower selectivity) and partially distort the aromatic core from planarity (hence lower G4 binding); (iii) the pyridine ring in place of the benzene in the extended core is strongly detrimental (compound **8**). The fair antiviral activity is paralleled by moderate cytotoxicity, allowing no therapeutic window (TW), possibly due to off-target activity; (iv) the NO_2 moiety is allowed (compounds **4** and **5**) as compound 4 is the most powerful and selective ligand toward LTR-III; (v) the negatively charged carboxylate group decrease the electrostatic component of the binding to the DNA G4s (compounds **6**

and **7**); (vi) on the contrary, the permanently positively charged ammonium groups extensively increase binding to the target G4s by adding electrostatic interaction (compounds **10** and **11**). However, compounds with permanent charges are not suitable for cell entry and hence cannot be exploited for antiviral purposes.

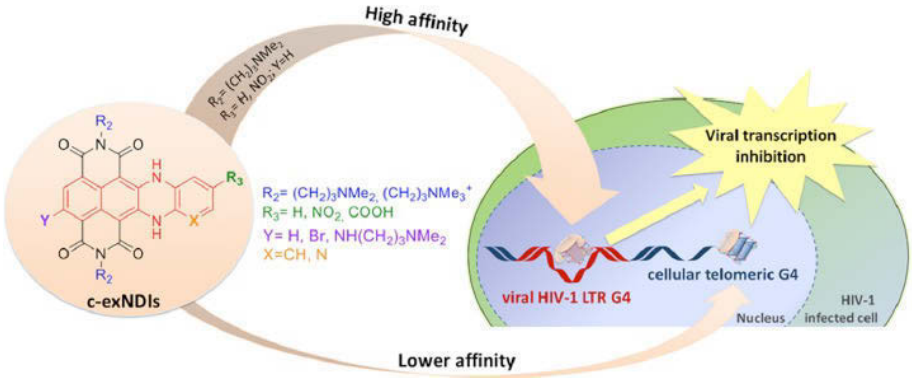


Figure 4.8 Schematic representation of c-exNDIs as novel antiviral compounds.

4.2 The major G-quadruplex form in the LTR of HIV-1 proviral genome reveals a (3+1) folding topology containing a stem-loop.

4.2.1 LTR-III sequence forms stable monomeric G-quadruplex structure

As discussed above, the propensity of LTR-III sequence to fold into a stable G4 structure was previously demonstrated by our group.

Within the full-length LTR G-rich sequence, formation of multiple G4 conformations involving different G-tracts is possible. This sequence can be divided in three main G4 forming components, named LTR-II, LTR-III and LTR-IV in Figure 4.9A. In previous studies, LTR-III conformation showed the highest thermal stability in circular dichroism and FRET melting experiments (180). Moreover, it has been assessed by *Taq* Polymerase Stop Assay that binding of G4 stabilizing small molecules, such as BRACO-19, prevalently occurs on the site corresponding to the formation of LTR-III G4 conformation in the full-length sequence context (Figure 4.9B).

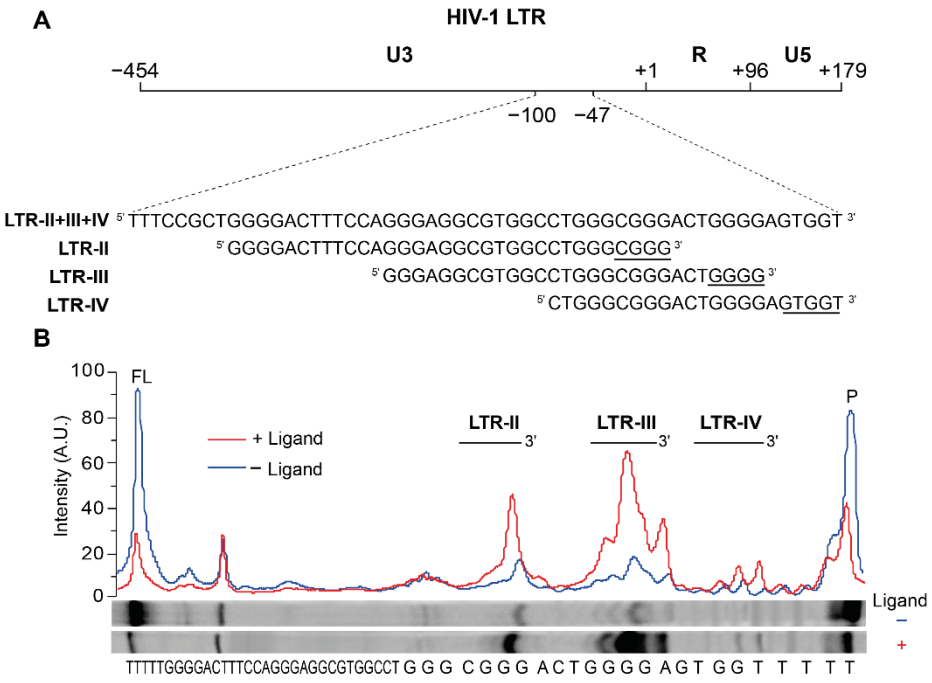


Figure 4.9 (A) LTR G-rich sequence in the U3 promoter region of HIV-1 proviral genome and the associated subsequences LTR-II, LTR-III and LTR-IV. **(B)** Gel quantification of a Taq polymerase stop assay (100 mM of potassium) on the LTR-II+III+IV template in the absence (blue) or presence of an excess of BRACO-19 (250 nM) (red). Bands corresponding to full-length products and primer are labeled with FL and P, respectively.

The 28-nt LTR-III sequence d(GGGAGGCGTGGCCTGGGCGGGACTGGGG) contains six tracts of two Gs or more. We investigated, using different spectroscopic techniques, the G4 formation of LTR-III sequence in K+ solution. First of all, the NMR spectrum of LTR-III showed twelve well-resolved peaks between 11 to 12.2 ppm, and three broad peaks around 12.6 to 13 ppm (Figure 4.10). The presence of well-resolved peaks in this region is characteristic of G4 formation and the number of peaks corresponds to the number of guanines participating to the structure formation. Therefore, in this case twelve resonances are characteristic of a three-layered G4 structure with a single prevalent conformation in solution. CD spectrum of LTR-III showed a maximum peak at 260 nm and a mild shoulder peak around 285 nm, suggesting the formation of a single conformation non-parallel G4 topology (Figure 4.10B).

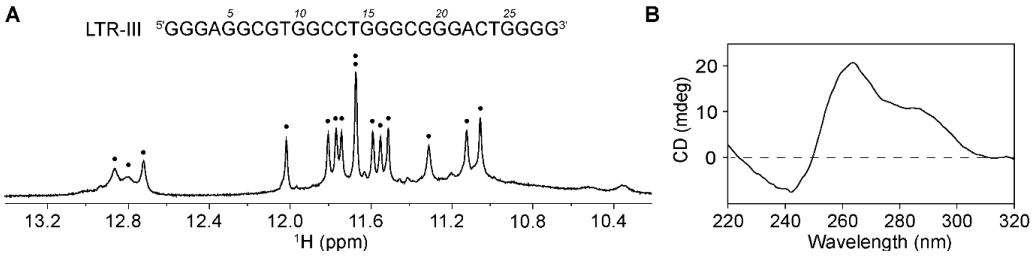


Figure 4.10 Characterization of LTR-III sequence in K⁺ solution **(A)** NMR imino spectrum of LTR-III sequence. Imino proton are labeled with black dots. **(B)** CD spectrum of LTR-III G4 registered at 20°C.

Thermal stability of LTR-III G4 in the conditions used for NMR analysis have been assessed by UV melting experiments (Figure 4.11A). A melting temperature of 65.5 °C determined using UV spectroscopy, was independent of the strand concentration (5 to 100 μM), indicating the formation of a monomeric G4 structure. In addition, on a native gel, LTR-III migration was similar to a monomeric G-quadruplex reference structure (Figure 4.11B). In summary, those data strongly suggest the formation of a non-parallel monomeric G-quadruplex structure.

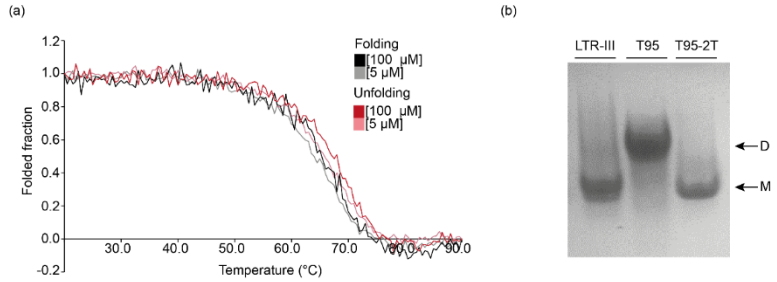


Figure 4.11 (A) UV melting curves of LTR-III G4 at different strand concentrations, 5 uM and 100 uM, in presence of 90 mM of potassium. Black curves represent annealing steps, red curves represent unfolding steps. **(B)** Electrophoretic mobility of monomeric and dimeric G4 forms on a native gel. T95 and T95-2T sequences were used as a control for dimeric and monomeric species respectively. D indicates dimer, M indicates monomer.

We then investigated the LTR-III imino spectrum variation in the 5 – 35°C temperature range (Figure 4.12). As expected, the imino spectrum was extremely well-resolved at higher temperatures, displaying 12 sharp and distinct peaks in the 11 – 12 ppm region. Further structural investigation were generally performed at 30°C, as the aromatic proton region spectrum was also more suitable for structural data collection.

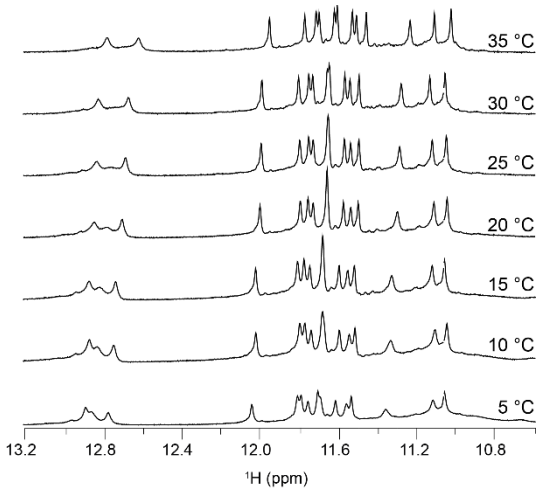


Figure 4.12 NMR imino spectra of the LTR-III sequence over the range of 5 – 35°C temperature

4.2.2 LTR-III G-quadruplex adopts (3+1) folding topology with three G-C base pairs diagonal loop

In order to elucidate the folding topology of LTR-III G4, NMR assignment was performed using site-specific approaches combined with through bond and space NMR experiments. Imino protons (H1) peaks involved in base-pairing formation were assigned using site-specific low-enrichment ^{15}N-labeling (Figure 4.13A). All the guanines in the sequence were site-specifically labeled and the imino protons of guanines in positions 1-2-15-16-17-19-20-21-25-26-27-28 were assigned to the resonances in the 12-11 ppm region suggesting their involvement in the Hoogsteen hydrogen bonds formation and, thus, participation to three G-tetrads of the structure. Interestingly, two resonances in the Watson-Crick base-pairing region were assigned to the H1 of guanines in positions 5 and 6, while the third broaden resonance assignment was challenging with this technique and G11-H1 was assigned using NOEs connectivity observed at low temperature (10 °C) (Figure 4.13C). Subsequently, imino protons were correlated to their corresponding aromatics (H8) using through bond JR-HMBC experiment (Figure 4.13D) and confirmed using site specific ^{2}H labeling (Figure 4.13B).

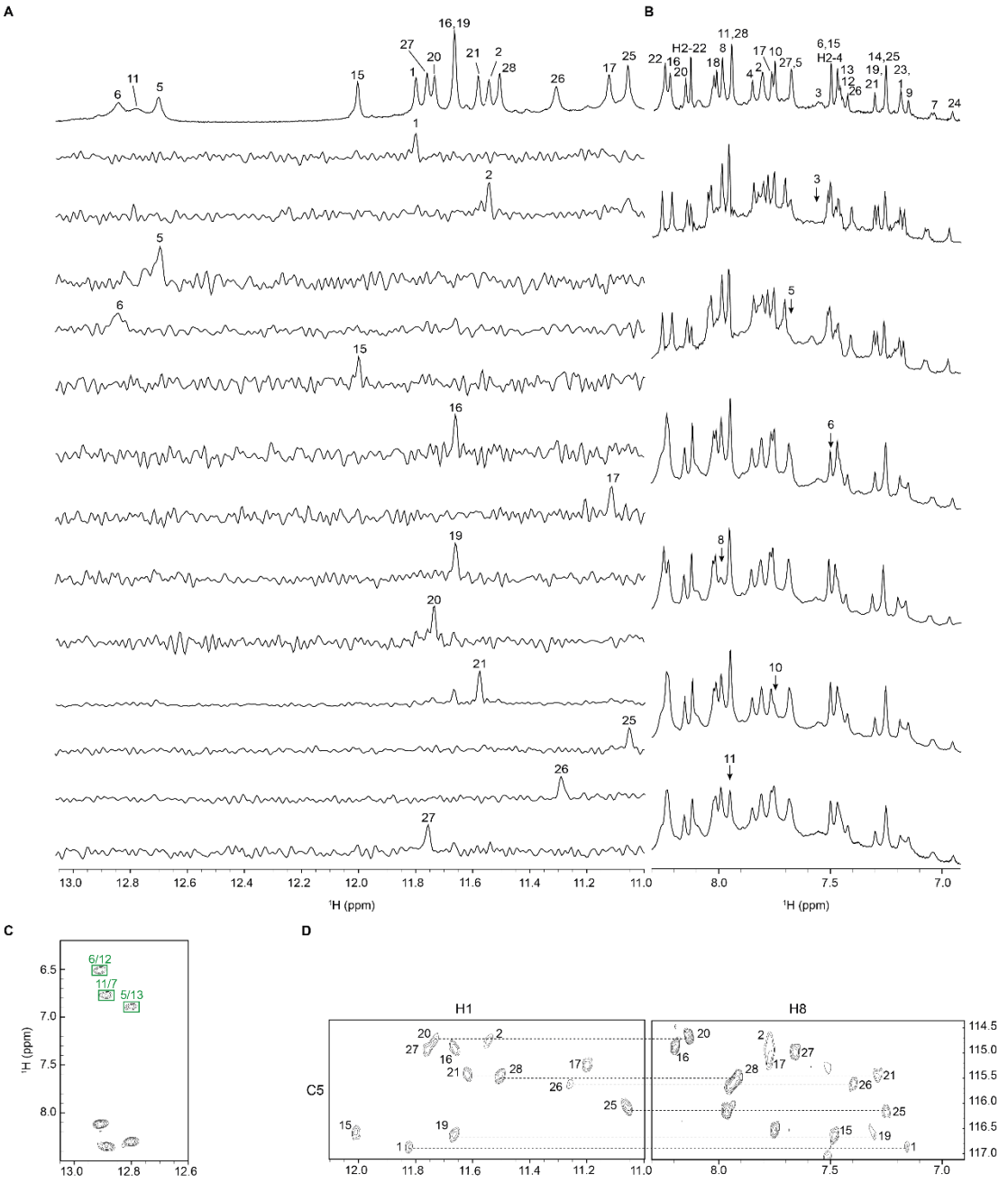


Figure 4.13 LTR-III resonances assignment by NMR. **(A)** Assignment of H1 protons from ^{15}N-filtered spectra of samples containing 4% of ^{15}N-enriched isotope at the indicated position, reference spectrum is shown on top. **(B)** Amino proton (H8) resonances assignment by site-specific ^{2}H-labeling. Only H8 of guanines not participating in the Hoogsteen hydrogen bonding are represented here. All the resonances are labeled with the number corresponding to the position in the sequence on the reference spectrum on the top. Resonance of adenine amino protons are labeled with H2. **(C)** NOESY spectrum of LTR-III acquired at 10 °C, 75 ms mixing time. Cross-peaks between guanines H1 protons and cytosine amino protons involved in Watson-Crick base pair are framed and labeled in green. **(D)** Assignment of H8 protons using H1-H8 through-bond correlation in JR-HMBC experiment.

The exchangeable protons (H1 of guanines) resonances are visible in the 1D NMR spectrum only in water solution. Once the sample is dissolved in deuterium, these protons will rapidly exchange with deuterium and disappear from the spectrum. However, the guanines in the central tetrad of G4 structure are protected from the exchange for a certain period of time. Thus, by deuterium exchange experiment it was possible to preliminary define the composition of the central tetrad of LTR-III G4 resulting in G1•G27•G16•G20 (Figure 4.14).

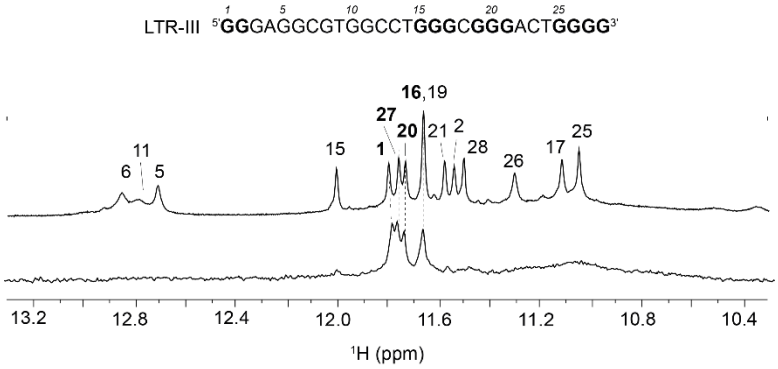


Figure 4.14 Imino proton region of 1D NMR spectra of LTR-III sequence in water (top) and after 1h in deuterium solvent (bottom). Four peaks corresponding to guanines belonging to the central tetrad are marked in bold numbers.

NOESY spectra provide sequential NOE connectivities between residues *n* and *n+1* H8/6(*n*)-H1'(*n*)-H8/6(*n+1*), that can be traced from 5' end to 3' end of a G4 (174). In early '90s, when site-specific labeling techniques allowing unambiguous spectral assignments were not refined, proton spectral assignments were done using NOESY approach. Intra-residue protons assignment are observed by strong NOE cross-peaks between sugar H1' and H4' or H2'/H2'' protons and between sugar H2' and H2'' or H3' protons. In this case, the independent proof of the assignment came from $^{2}H_2O$ NOESY spectrum at 300 ms mixing time. Sequential NOE connectivities were traced for H1' (Figure 4.15A), H2'/H2'', H3', H4' and H5'/H5'' spin systems.

The folding topology of a G4 structure can be determined from specific NOE patterns. Within a G-tetrad, cyclic connectivities between imino proton of a guanine and aromatic proton of the adjacent guanine can be established and used for tetrad composition determination (Figure 4.15B). Short distances between stacked G-tetrads can be detected and suggest same-polarity or opposite-polarity stacking mode.

H1-H8 NOEs connectivity (Figure 4.15A) allowed us to establish the formation of a three layered G-quadruplex core composed of G2•G26•G15•G19, G1•G27•G16•G20 and G25•G28•G17•G21, the former being in reverse polarity. NOEs between H1 to cytosine amino proton established three Watson-Crick base pairs G5•C13, G6•C12, and C7•G11. The topology found in this study is consistent with imino protons of guanines (G1, G27, G16 and

G20) in the central layer found to be the most protected in a deuterium exchange experiment (Figure 4.14).

According to the NOESY data at short mixing time (100 ms), guanosines in positions 1, 15, 19 and 25 are clearly in a *syn* conformation, while the guanosine in the position 26 does not display all the typical features of the *syn* conformation because of the particular sugar distortion in the V-shaped loop formed by adjacent residues G25 and G26, as reported previously (181).

These data suggested the formation of hybrid duplex/quadruplex structure (Figure 4.15C). The quadruplex core forms a (3+1) folding topology, characterized by three parallel G-strands and one antiparallel, defining two medium, one wide and one narrow grooves (Figure 4.15C). Four loops connect the tetrads: 11-nt diagonal loop forming stem-loop structure, 1-nt propeller loop, 3-nt lateral loop and V-shaped loop (Figure 4.15C). G25-G26 residues belonging one to the bottom tetrad and one to the top tetrad form V-shaped loop. This ar-

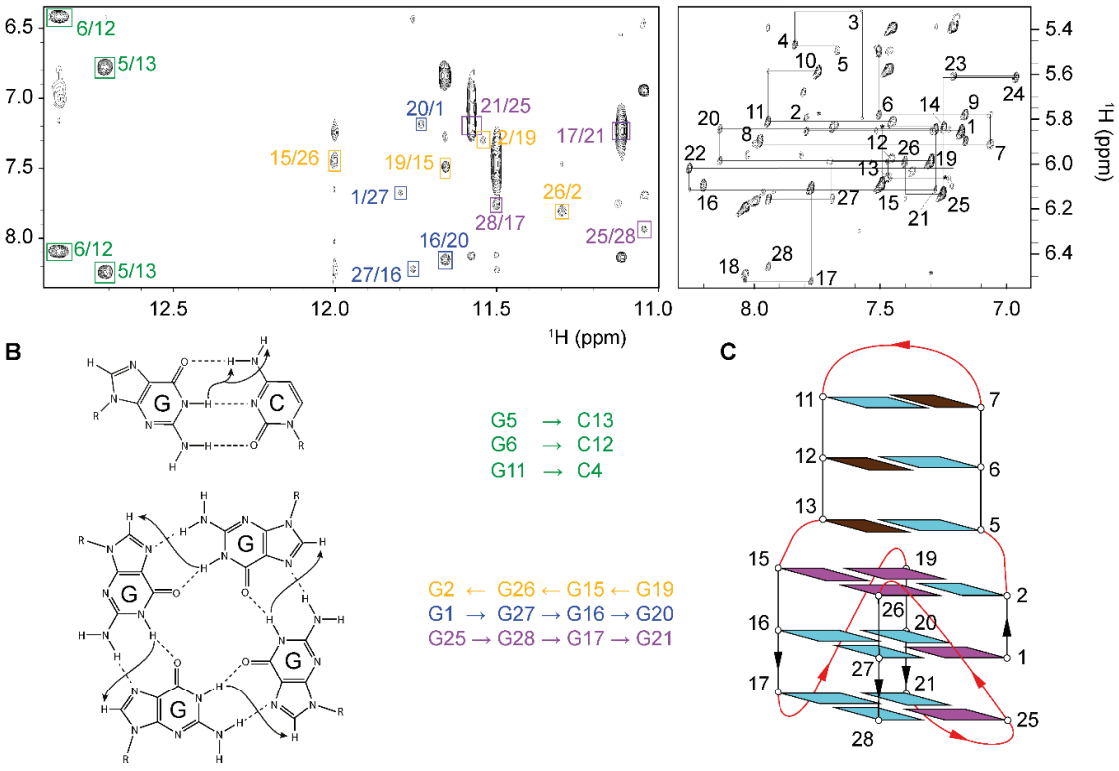


Figure 4.15 LTR-III G4 folding topology determination by NMR. **(A)** Right panel, H8/H6-H1' NOE sequential connectivities in ^2H_2O NOESY spectrum (mixing time, 300 ms at 35 °C). Intraresidue cross-peaks are labeled with residue number. Cross-peaks marked with an asterisks are seen in lower threshold. Left panel, H1-H8 NOEs cyclical connectivities in H_2O NOESY spectrum (mixing time, 200 ms at 25 °C). Cross-peaks used for G-tetrads determination are framed and colored based on the G-tetrad participation. Cross-peaks between guanines H1 protons and cytosine amino protons involved in Watson-Crick base pair are framed and labeled in green. **(B)** Schematic representation of H1-H8 cyclic NOE connectivities within the G-tetrad and resonances between amino protons of cytosines and H1 of guanines in Watson-Crick basepairing **(C)** Folding topology of LTR-III. Guanines in anti and syn conformations are colored in cyan and magenta respectively, cytosines in brown.

rangement creates a phosphate backbone extension and breaking connection in the anti-parallel G-strand, where the first guanine of the sequence G1 is positioned in the central tetrad, in contrast to the majority of intramolecular G-quadrpulex topologies published so far. Within the long 11-nt diagonal loop, six nt are interacting by Watson-Crick hydrogen bonds to form a hairpin (or stem-loop) structure with GTG loop (Figure 4.15C). The helix and G-core are in the coaxial orientation. The hypothetic additional A4-T14 basepairing at the junction is not favored because the distance between the diagonal corners of the tetrads (20 Å) does not allow the hydrogen bond formation (182).

4.2.3 LTR-III sequence mutational analysis

Mutational analysis was performed in order to confirm the investigated LTR-III topology. We considered different point mutations in the LTR-III sequence to assess which positions are critical for the G-quadruplex/duplex hybrid structure formation. To this purpose, we virtually subdivided the structure in four parts: tetrads (residues 1-2-15-16-17-19-20-21-25-26-27-28), junction (residues 3-4-14), stem (residues 5-6-7-11-12-13) and loop (residues 8-9-10) (Figure 4.16).

The base-pairing region of the long loop in the wild type sequence does not display three well-resolved peaks in the 1D NMR spectrum, specifically the resonance corresponding to C7-G11 basepairing is particularly broaden. Recent studies on hybrid constructs showed that the GTA loop in the stem-loop structure favors the base pairing (183). In fact, mutation of GTG to GTA loop allowed to clearly identify three base pairs in the stem (Figure 4.16). The stem formation is the very important feature of the hybrid G-quadruplex/duplex structure. Complementary bases mutations led to the disruption of the basepairing: thus, we effectively confirmed the formation of the C7-G11 basepairing by G11 to A mutation and G5-

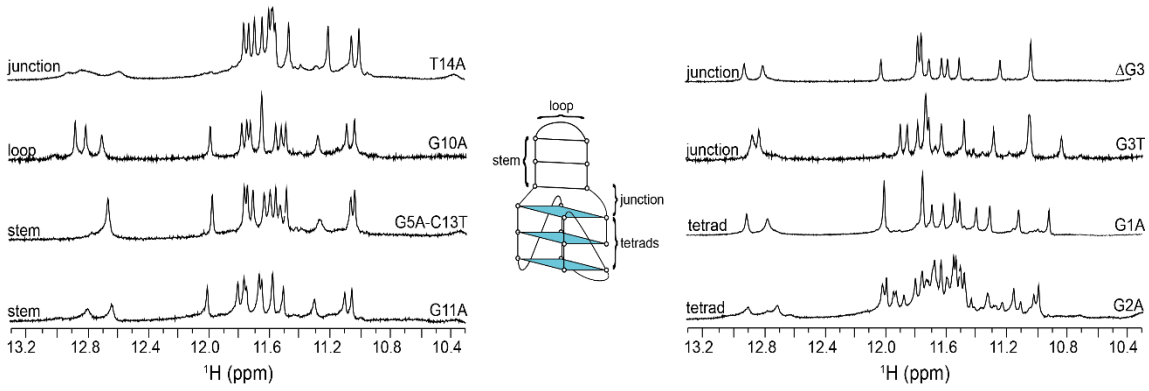


Figure 4.16 LTR-III sequence mutational analysis. NMR imino spectra of LTR-III mutated sequences. Point mutations corresponding to each spectrum are indicated. In the central panel schematic representation of LTR-III topology.

C13 by G5 to A mutation (Figure 4.16). We also asked ourselves if the G5A-C13T double mutation would restore the basepairing by A-T interaction. This was not possible probably due to the incompatibility of A-T basepairing with the dimension of the tetrad the diagonal loop is spanning through. Worth noting that all the stem-loop mutations did not affect the G-core formation. The junction between the stem-loop and the tetrads is very important for the formation of the base pairing. Mutation of G3 from the first run of guanines to T still allows G-tetrads formation while the base pairing is affected in a negative way and only two peaks are present around 12.8 ppm (Figure 4.16). T residue in position 14 seems to be important for baseparing formation since T14 to A mutation leads to the broadening of corresponding resonances, probably because the interaction of A4 and mutated A14 are not favorable (Figure 14.6). Since G3 does not participate to the tetrad formation and the first G-strand of the structure present the strand breakage, we investigated the influence of G1, G2 and G3 on the tetrads formation. The substitution of G2 with A destabilizes the tetrads leading to the formation of a multiple conformations. As mentioned above, G3 to T substitution does not affect the G-core formation, but only the stem loop formation, confirming that this nucleotide is not involved in the tetrad. Moreover, the same effect was produced by the deletion of the G3. Interestingly, G1 to A mutation did not affect the formation of G-core, but only of the basepairing region mimicking the G3 deletion effect, suggesting the possibility that in the absence of G1, G2 and G3 participate to the first G-strand (Figure 4.16). In contrast, G2 to A mutation was more disruptive for the core as the imino protons region of the spectrum presented the number of resonances corresponding to multiple G4 conformations formation.

4.2.4 The sequence containing LTR-III and LTR-IV G-quadruplexes displays the same topology of LTR-III G-quadruplex.

We demonstrate here that the LTR-III sequence forms a unique hybrid G-quadruplex/duplex structure. Previously, the structure of LTR-IV component of LTR G4-forming region has been solved (183). Therefore, we investigated whether in the longer sequence a single conformation prevailed. The 1D spectrum of the LTR sequence able to form both LTR-III and LTR-IV G4 structures displayed 12 well-resolved peaks in the imino proton region, which shared many similarities with the 1D spectrum of the LTR-III sequence (Figure 4.17). Particularly, we found that three peaks in the Watson-Crick base-pairing region of the spectrum are still present, suggesting that the unique feature of LTR-III structure, the three base-pairing stem-loop, is still conserved in the longer sequence. By site-specific ^{15}N-labeling approach we demonstrated that the twelve peaks in the imino protons region of the LTR-III+IV G4 belong to the same guanines which participate to the formation of LTR-III sequence, while the guanines responsible of the LTR-IV G4 structure are not involved in the Hoogsteen hydrogen bond formation (Figure 4.17A and C). Moreover, the deuterium exchange experiments showed that the guanines involved in the central tetrad formation of

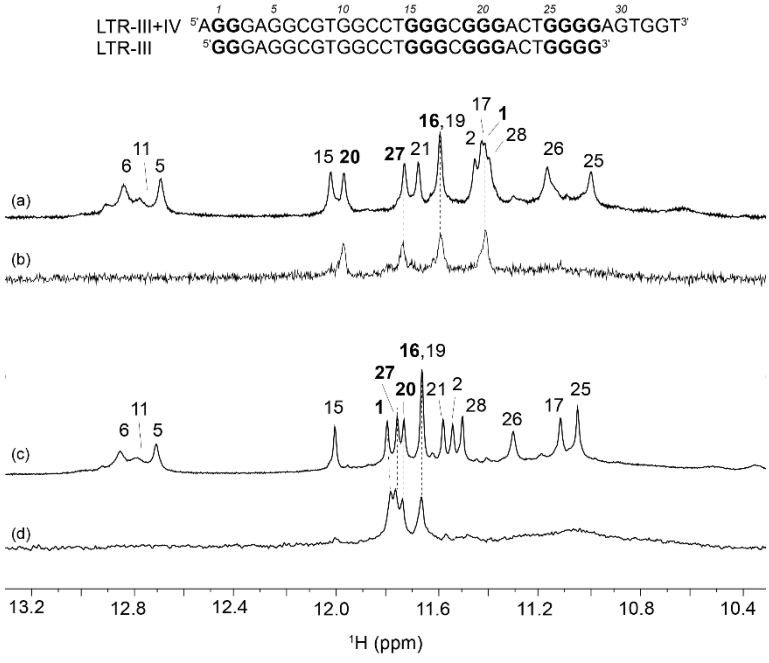


Figure 4.17 Comparison between LTR-III G4 forming sequence and LTR-III+IV. 1D NMR spectra of LTR-III+IV **(A)** and LTR-III **(C)** G4s with the assignment of guanines participating to the G-tetrads and stem-loop formation. Corresponding 1D NMR spectra of LTR-III+IV **(B)** and LTR-III **(D)** G4s after 1h in D_2O. Four peaks, numbered in bold in spectra **(A)** and **(C)**, belong to the central tetrad of G4 structures.

the LTR-III+IV G-quadruplex (1, 16, 20, 27) exactly correspond to the guanines in the same position of LTR-III sequence (Figure 4.17B and D). Full-length G-quadruplex forming region of HIV-1 LTR and LTR-II+III+IV (see Figure 4.8) were also analyzed to assess the presence of a prevalent conformation. The imino proton region of 1D NMR spectra displayed the presence of multiple conformations resulting in a very broad spectra, probably due to the presence of additional LTR-II G4 component (Figure 4.18).

According to these data, it is clear that the longer sequence is prone to fold in a favored single conformation of LTR-III G4, conserving its unique structural features. This fact suggests that the LTR-III G4, previously demonstrated to be the major and most stable form of the considered region, is also prevalent in the longer and more dynamic context.

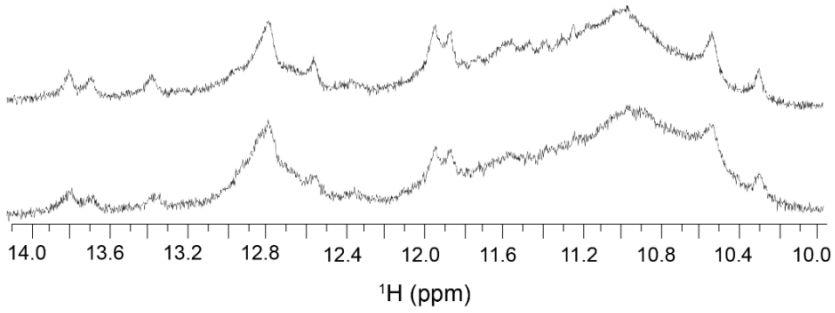


Figure 4.18 1D NMR spectra of LTR-II+III+IV (top spectrum) and LTR-FL (bottom spectrum) G4s.

4.2.5 Discussion

In this work, we demonstrate that LTR-III sequence folds in a very peculiar hybrid duplex/G-quadruplex conformation with three-layered G-tetrad core arranged in a (3+1) topology and long 11-nt loop forming a hairpin structure. NMR analysis of the sequence able to form both LTR-III and LTR-IV components, named LTR-III+IV, showed that the folding topology of LTR-III is still conserved suggesting the preferential folding of LTR-III G-quadruplex within the dynamic context of multiple conformations. Hybrid duplex/G-quadruplex structures have been described previously as artificial constructs with different relative duplex/quadruplex orientations, exploring junction and connection varieties (182,184). Even if computational studies performed on the human genome showed the potential of around 20.000 sequences prone to fold in such a structures (185), so far only one naturally occurring and biologically relevant hybrid duplex/quadruplex topology has been reported (186).

The guanine content in the HIV-1 G-quadruplex forming region is highly conserved (ref LTR). Mutations in the loop sequence, giving rise to the stem-loop formation, may disrupt Watson-Crick base pairing, removing the duplex component of the structure and leading to the loss of the stability of entire structure. Therefore, the conservation of the nucleotides

participating to the stem-loop formation has also been assessed, revealing high percentages of conservation (70% - 99%) for all the nucleotides with the exception of cytosine in position 7, which displayed around 50 % of probability for thymine mutation.

Multiple G-quadruplex structures in the HIV-1 LTR promoter region have been proposed as regulatory elements of viral transcription and therefore as promising targets for viral cycle inhibition. Stabilization of viral G-quadruplexes by the well-known G-quadruplex ligand BRACO-19 resulted in the inhibition of viral production (164,166). Recently, a newly synthesized family of ex-core NDI compounds were found to act as antiviral agents with G-quadruplex related mechanism, targeting selectively viral over telomeric G-quadruplexes (180). Interestingly, we found that mutations in LTR-IV G-quadruplex component do not abolish the inhibitory effect on viral transcription probably due to the stable presence of LTR-III conformation. Therefore, selective targeting of this major LTR G-quadruplex component may result in promising strategy for viral transcription inhibition.

Such a singular structure of LTR-III G-quadruplex opens the possibility of improving the selectivity by targeting G-quadruplex/duplex junction. Examples of the compounds targeting this feature may come, for instance, from DHFBI fluorogens intercalating on the junction between the G-quadruplex and hairpin of RNA light-up aptamers (187).

Part 2. HIV-1 U3 RNA G-quadruplexes: targeting and interactions with cellular and viral proteins.

4.3 The G-quadruplex ligand BRACO-19 displays antiviral activity targeting two steps of the HIV-1 viral cycle.

Our group has recently reported the formation of DNA G4s in the integrated HIV-1 DNA genome. Precisely, we described the dynamic set of G4 structures in the LTR promoter region, involved in viral transcription inhibition, and a cluster of G4s in the leading and lagging strands of *nef*, impairing its expression (164,165). Viral DNA G4s were stabilized by G4 ligands, such as acridine BRACO-19 and porphyrin TMPyP4, resulting in virus inhibition with different molecular targets. While TMPyP4 activity was mainly correlated to the targeting of *nef* coding sequence, BRACO-19 activity was *nef* independent (163). A detailed study on the antiviral activity of BRACO-19 has been performed. It has been shown previously that BRACO-19 inhibits HIV-1 NL4-3 strain in the TZM-bl reporter cell line, which contains a luciferase reporter under the control of the HIV-1 LTR promoter (167). To expand the understanding of antiviral activity, BRACO-19 was tested against a different viral strain (HIV-1$(III)_B$) in a different reporter cell line (MT-4-LTR-eGFP). At 8.4 uM, BRACO-19 caused a decrease in eGFP amounts of 58% (IC_{50}=6.2 μM), while displaying low cytotoxicity (CC_{50}=92.9 μM) and therefore a good SI (SI=15).

To determine the viral stage at which BRACO-19 was active, first of all the direct effect on the viral particle prior to infection was excluded by virucidal assay. Then, a Time of Addition (TOA) experiment was set up to identify the viral step targeted by the compound. This experiment determines how long the addition of a compound can be postponed before its antiviral activity is lost in a single replication cycle. The antiviral effect of the test compound is compared to reference compounds with established modes of action in the time frame of replication events. Interestingly, the BRACO-19 behavior was similar to that of reference RT inhibitors zidovudine and nevirapine, which block the HIV-1 replicative cycle in the reverse transcription time frame, if administered up to 3 – 5 h post infection. TOA experiment pointed out a preintegration target for BRACO-19, prior of coinciding with the reverse transcription process, not excluding other targets during post-integration events. We thus investigated whether G4 structures could also form at the pre-integration level.

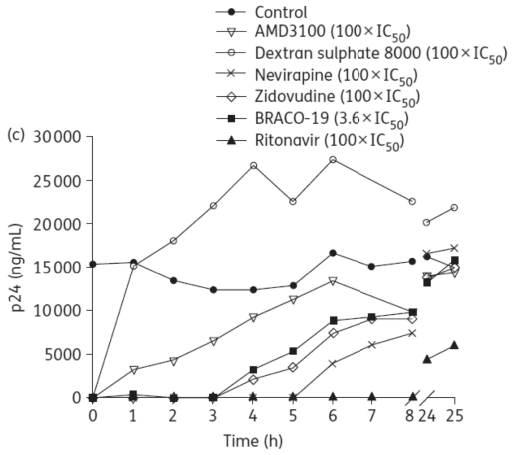


Figure 4.19 TOA experiment. The activity of BRACO-19 was compared with that of the negative control and reference drugs for different steps of HIV-1 infection. Figure from (167)

4.3.1 Characterization of RNA G-quadruplex structure in the U3 region of HIV-1 ssRNA genome

We have previously reported that the G-rich region of the LTR promoter in the proviral genome can fold into G4 and that this folding is stabilized by G4 ligands, such as BRACO-19, resulting in inhibition of viral transcription. Because the G4-forming LTR DNA region derives from an identical RNA sequence present at the 3′-end of the RNA viral genome (positions 8981–9038), we tested if this RNA sequence could also fold into G4. Analogously to the DNA region, the RNA counterpart was divided into three regions (namely U3-II RNA, U3-III RNA and U3-IV RNA) containing four GGG tracts, which are the minimum requirement for the formation of a three-tetrad G4 structure (Figure 4.20).

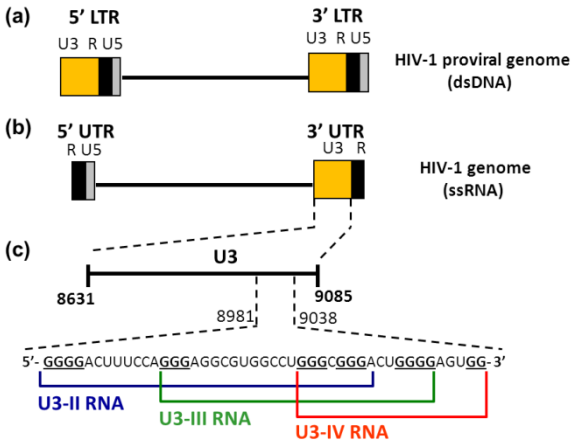


Figure 4.20 Scheme of the HIV-1 proviral DNA genome **(A)** and of the viral RNA genome **(B-C)**. U3 sequences embedding the G-quadruplex forming region are represented by the orange box. G bases involved in G-quadruplex are in bold and underlined. Brackets indicate sequences that form G-quadruplexes with 3 stacked tetrads.

We initially performed the spectroscopic characterization of U3 RNA in order to elucidate G4 formation. Thus, the formation of G4 in the selected RNA sequences was investigated by circular dichroism (CD) technique. In the presence of K^+, the resulting oligonucleotides showed CD signatures characteristic of RNA G4s, i.e. a positive peak at 260 nm and a negative peak at 240 nm (Figure 4.21A), which depict a parallel-like G4 conformation. All RNA G4s so far reported fold into a parallel topology. To confirm CD data, TDS signatures were acquired by UV spectroscopy. Typical shapes of thermal difference spectra shown G4 folding (Figure 4.21B).

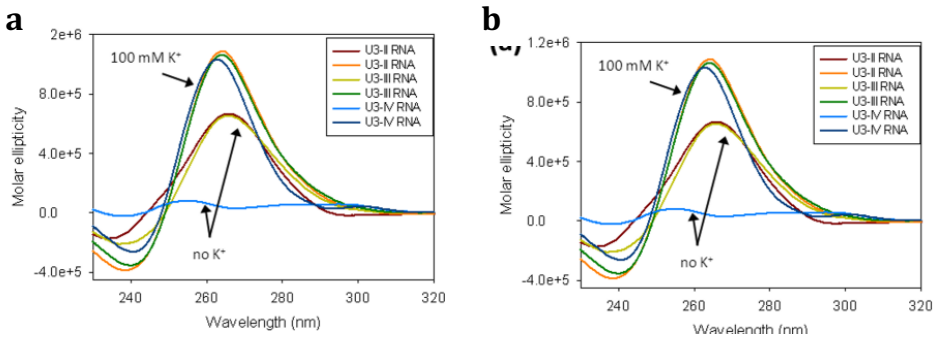


Figure 4.21 CD and UV spectra of U3 RNA sequences**. (A)** Dichroic spectra of U3 RNA sequences in absence and presence of 100 mM KCl. **(B)** UV Thermal difference spectra of U3 RNA in presence of 100 mM KCl.

CD thermal unfolding experiments were performed to investigate thermal stability of U3 RNA G4s. As shown in Figure 4.22 and Table 4.3, U3-II RNA and U3-III RNA displayed T_m of ~50°C and >70°C without and with 100 mM K^+, respectively. U3-IV RNA was less stable without K^+ (also see Figure 4.21), while in the presence of the cation it reached T_m values >70°C. The addition of potassium ions greatly stabilized RNA G4 structures increasing T_m values up to 30 °C. As expected, the U3 RNAs exhibited significantly higher stability (10–30°C) than that of their LTR DNA counterparts (Table 4.3).

Thermal stability experiments on the selected oligonucleotides were also performed in the presence of G4 ligand BRACO-19. Incubation with 16uM BRACO-19 o/n further greatly stabilized G4 conformations, which could not be unfolded up to 95 °C (Figure 4.22, Table 4.3).

Oligonucleotide	T_m (°C), no K^+	T_m (°C), K^+ 100 mM	T_m (°C), K^+ 100 mM + BRACO-19
U3-II RNA	56.5 ± 0.5	73 ± 1.1	> 95
U3-III RNA	47.2 ± 0.7	82.1 ± 2.6	> 95
U3-IV RNA	35.5 ± 2.6	71.2 ± 0.7	> 95
LTR-II DNA	39.7 ± 1.2	49.0 ± 0.29	88.9 ± 0.6
LTR-III DNA	38.0 ± 1.5	51.9 ± 0.12	94.2 ± 2.0
LTR-IV DNA	26.2 ± 0.5	43.4 ± 0.21	97.4 ± 3.2

Table 4.3 T_m of U3 RNA sequences and their DNA counterparts in different experimental conditions.

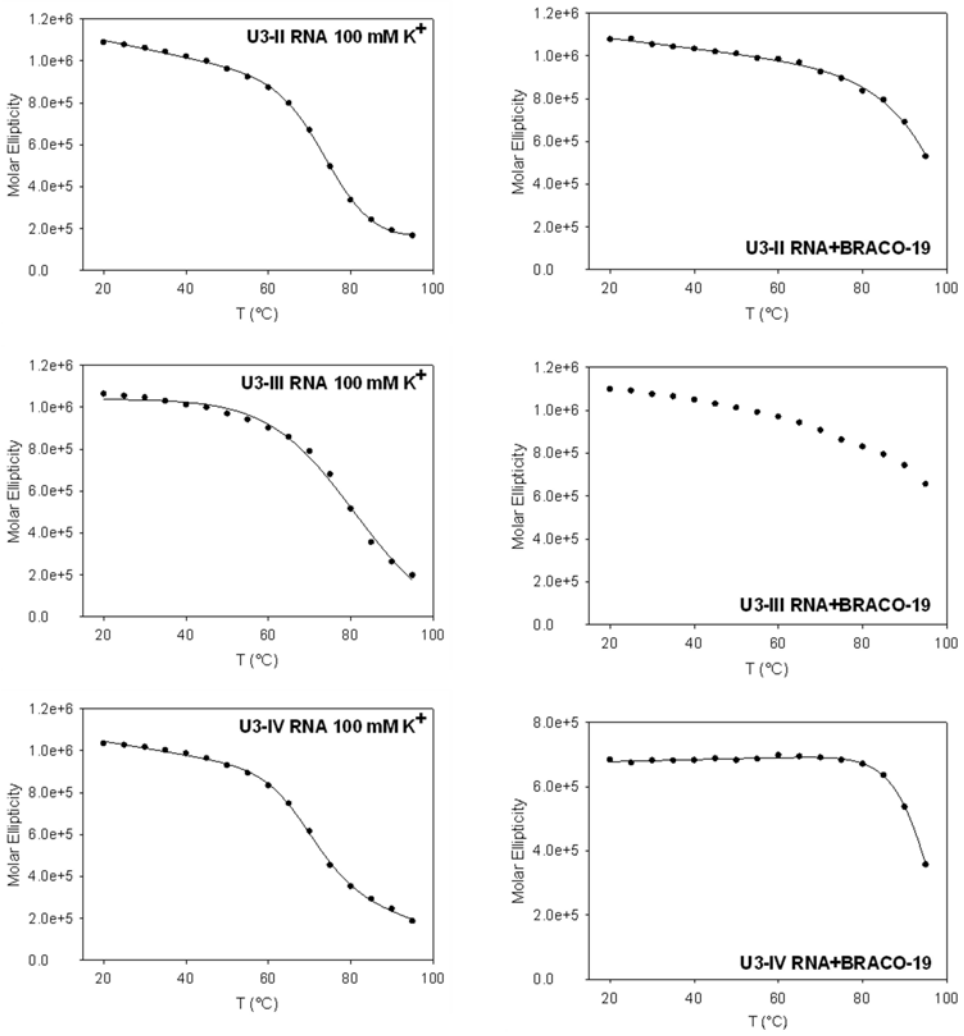


Figure 4.22. CD thermal melting curves. On the left, CD melting curves of U3 RNA sequences in presence of 100 mM KCl. On the right, corresponding CD melting curves of U3 RNA sequences in presence of 100 mM KCl and stabilized by BRACO-19 (16uM).

Based on the above information, we tested whether G4 formation in the G-rich region of U3 RNA could influence RT processing. To this end, we performed an HIV-1 RT stop assay on U3-III+IV RNA template, which was elongated to include a primer annealing region. We used this template because we had previously shown that in the full-length DNA template, which also comprised the U3-II tract, mainly U3-IV and U3-III G4s induced *Taq* polymerase pausing. U3-III+IV RNA template was incubated in the presence of K^+ to promote G4 folding. As shown in Figure 4.23, full-length product was obtained in the absence of K^+ (Figure 4.23A, lane 1). In the presence of K^+, two stops corresponding to the formation of U3-IV and U3-III G4s were visible (Figure 4.23A, lane 2), indicating the coexistence of both structures.

When the U3-III+IV RNA template was incubated with increasing concentrations of BRACO-19, a known G4 ligand, HIV-1 RT was inhibited to a much higher extent (Figure 4.23, lanes

3–9). In particular, inhibition of the viral polymerase was observed from the lowest concentration (125 nM) of the compound: at 8 μM the stop band reached 77% in the U3-III+IV template with respect to the full-extended product (Figure 4.23A and C). Furthermore, a stop corresponding to an additional G-tract in U3-III+IV G4 formation was evident (Figure 4.23A, broken arrow). In contrast, only a modest nonspecific inhibition was induced at the highest concentrations (4–8 μM) in a scrambled RNA template (Figure 4.23B and C). Reverse Transcriptase Stop assay also confirmed that BRACO-19 does not affect the processing of the enzyme by direct interaction with reverse transcriptase, as all the elongation reactions on different templates display full-length products.

All these data confirmed the folding of a very stable parallel G4 structures in the U3 region of HIV-1 RNA genome. The formation of this structure can affect the processing of RNA by reverse transcriptase enzyme already in the presence of potassium, but especially if stabilized by G4 ligands such as BRACO-19.

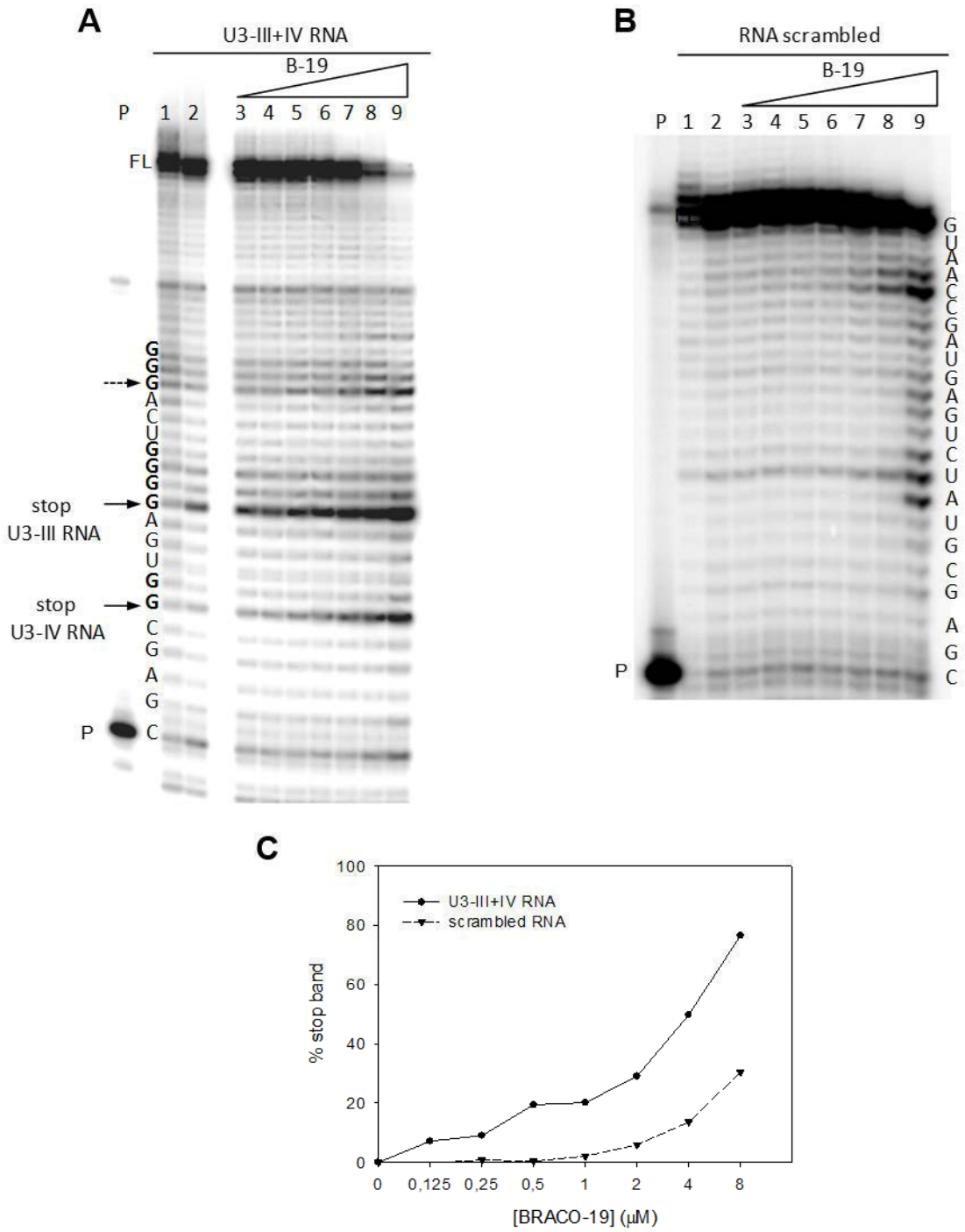


Figure 4.23 *RT* stop assay. **(A)** Lanes 1-2, RT stop assay on U3-III+IV RNA template in the absence and in the presence of 100 mM K^+. Lanes 3-9, RT stop assay on U3-III+IV RNA template in the presence of increasing concentrations of BRACO-19 (125 nM, 250 nM, 500 nM, 1μM, 2 μM, 4 μM, 8 μM). **(B)** RT stop assay on a scrambled RNA template in the absence and presence of K^+ 100 mM (lanes 1-2, respectively) and in the presence of increasing concentrations of BRACO-19 (125 nM, 250 nM, 500 nM, 1 μM, 2 μM, 4 μM, 8 μM) (lanes 3-9). P indicates primer. FL indicates the full-length products. RT pausing sites are indicated by arrows. Sequences are shown aside each gel image. **(C)** Quantification of stop bands obtained in panel (A) and (B). The sum of stop bands obtained by treatment with BRACO-19 was calculated with respect to the full-length product obtained in the absence of the compound.

4.3.2 Discussion

We had shown previously that BRACO-19 interacted and stabilized the DNA G4s structures formed by HIV-1 sequences in the proviral DNA genome. We demonstrated its mechanism of action as an anti-HIV-1 compound. Interestingly, a dual mode of action was observed for BRACO-19: the antiviral activity occurred both at the post- and pre-integration steps. This behavior has been rationalized by the discovery presented her that U3 G4s form both at the DNA and the RNA genomic level. It is possible that the prevalent effect of BRACO-19 observed at the RNA level is ascribable to the reported preferential binding of the compound towards parallel G4 conformations (refs). In fact, parallel folding of RNA U3 G4s was significantly more definite.

Therefore, ligands that display biased affinity towards the viral DNA and RNA G4 structures might alternately drive the antiviral effect towards inhibition of the reverse or direct transcription process. A second possibility is that the G4 structures formed in the viral RNA genome are more available to the drug with respect to those in the proviral DNA genome. First of all, the RNA genome is in the cell cytoplasm, while the DNA genome is in the nucleus; further, the viral genome is present in a single-stranded conformation, which facilitates tetraplex folding, while the proviral genome rests in a double-stranded helix, which can be unfolded during specific events, such as polymerase processing. The fact that activity has also been observed at the post-integration level indicates that the viral DNA genome is still accessible, even if to a lower extent than the RNA genome. An additional interesting aspect is the natural propensity of the U3 sequence in the viral RNA genome to fold into G4s at physiological conditions. The higher stability observed for these structures, with respect to that of their DNA counterparts, is ascribable to the differences in hydration and increased intra-

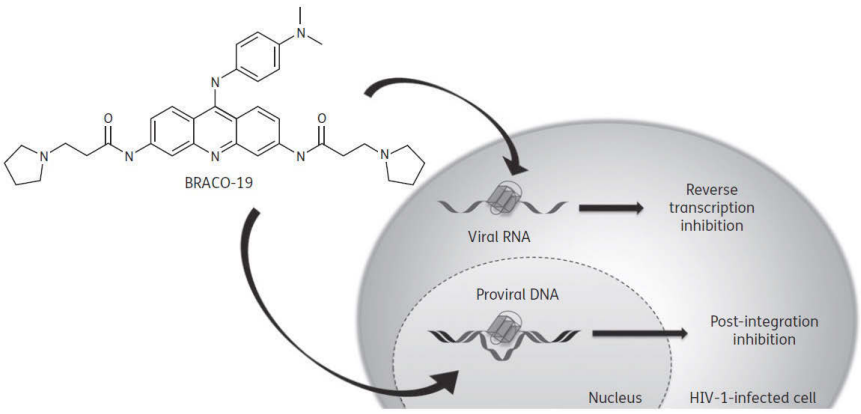


Figure 4.24 Proposed model of BRACO-19 antiviral activity through double mechanism of action: at the reverse transcription step, targeting U3 RNA G4s, and during viral transcription inhibition on post-integration level targeting DNA G4s.

molecular hydrogen bonding owing to the 2′-OH group of ribose sugar. Considering the higher stability and the less hindered conformation of the linear sequence, it is likely that the U3 RNA G4s naturally form in vivo and exert a precise biological function, such as that observed for the LTR DNA G4s (i.e. inhibition of transcription at the promoter level). Therefore, BRACO-19, by binding both RNA and DNA G4s, inhibits viral replication and may decrease production of variant HIV-1 strains (Figure 4.24).

We have shown that antiviral activity of BRACO-19 can be ascribed to a G4 mediated mechanism of action. Selectivity towards viral G4 structures has to be improved to avoid targeting of cellular G4s (for instance telomeric sequences) leading to cell toxicity and increased cell proliferation. A demonstration that a G4 ligand displays a G4 related antiviral activity opens interesting perspectives for development of new selective compounds with novel mechanism of action.

4.4 HIV-1 U3 RNA G-quadruplexes – protein interactions

As discussed in the introduction, in the last years RNA G-quadruplexes attracted a great interest, especially as translational regulators at the mRNA level. So far, a few proteins that bind RNA G4 structures have been discovered, but this is an open field, currently under investigation.

It has been previously described that the G-rich sequence, located in the U3 region of single stranded RNA genome of HIV-1 can adopt a G4 conformation. Extensive spectroscopic studies have demonstrated an exceptional thermodynamic stability of these structures in the presence of physiological concentrations of potassium cations (100 mM). The presence of the folded conformation stabilized by 100 mM of KCl in the RNA template induced strong stop bands during the elongation by reverse transcriptase in the *RT* stop assay.

In physiological context the reverse transcription is not affected by the G4 folding and it is successfully completed so the infection can proceed. We, thus, hypothesized that binding of viral proteins negatively controls RNA G4 folding to allow successful progression of infection. The HIV-1 nucleocapsid protein has been shown to assist reverse transcriptase during the viral DNA synthesis as a chaperone that drives nucleic acids to the thermodynamically more stable conformation. We here investigate the possible role of the nucleocapsid protein in the unfolding of U3 RNA G4s.

4.4.1 HIV-1 nucleocapsid protein interaction with viral RNA G-quadruplexes

To assess the involvement of nucleocapsid protein in G4 structures modulation, we first confirmed the binding of this protein to RNA sequences folded in G4s. To this purpose we performed the EMSA assay on the RNA U3 III+IV sequence folded in 50 mM KCl buffer in the presence of increasing concentration of nucleocapsid protein (Figure 4.25). At intermediate

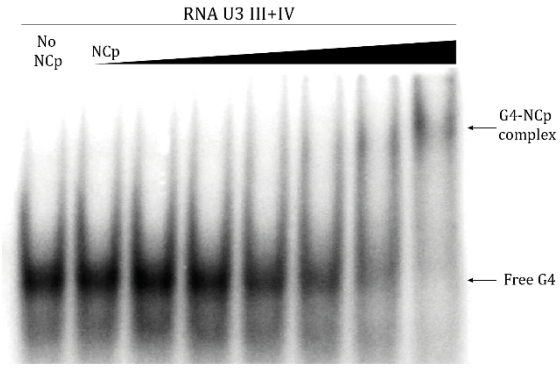


Figure 4.25 EMSA on RNA U3 III+IV sequence in presence of increasing concentrations of NCp (7.5 nM – 15 nM – 37.5 nM – 75 nM – 150 nM – 300 nM). Arrows indicate free G4 species and G4-NCp complexes.

concentrations G4-NCp binding complexes were observed (75nM – 300 nM), while at 300 nM NCp concentration binding saturation was reached with complete disappearance of free G4 species (Figure 4.25).

In order to evaluate the unfolding properties of NCp we set up an EMSA assay in presence of the complementary DNA strand. Addition of cold complementary DNA strand to the labeled RNA sequence produces the formation of hybrid RNA-DNA duplex, which can been separated from the single strand species on a native polyacrylamide gel. In the presence of thermodynamically stable RNA G4 structure the formation of the duplex is slowed down, as the G4 structure has to be unfolded prior to the basepairing. The G4 unfolding properties of nucleocapsid protein can be evaluated by the increased rates of duplex formation in the presence of the protein. We first assessed the hybrid duplex formation and binding of NCp to both species, G4 and duplex (Figure 4.26). RNA U3 III+IV sequence was folded in potassium buffer and then incubated with the complementary DNA strand for 1h. Formation of duplex form has been observed at 7.5 nM and 15 nM complementary strand concentrations (Figure 4.26, lanes 2 – 3). Upon addition of nucleocapsid protein, formation of the binding complex was observed with complete disappearance of free G4 species, while free duplex form was still present (Figure 4.26, lanes 5 – 6). This data confirmed the preferential binding of the protein to the single strand RNA conformations.

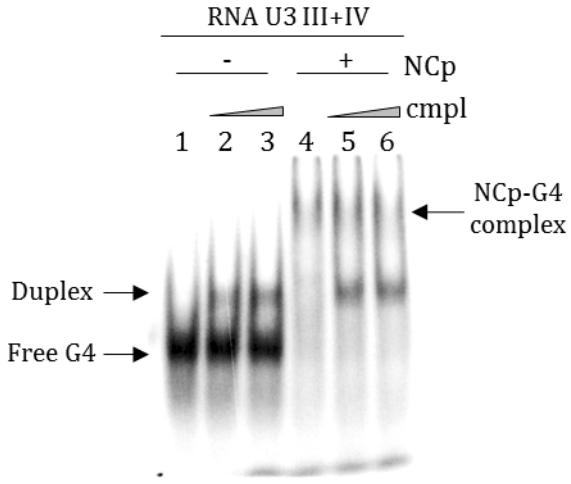


Figure 4.26 EMSA assay on the RNA U3 III+IV sequence folded in 50 mM potassium buffer in presence of cold complementary (cmpl) DNA strand and NCp protein. Complementary strand was added at 0.5 and 1 RNA/DNA molar ratio. Duplex, free G4 and NCp-G4 complex species are indicated with arrows. NCp concentration was 300 nM

We next investigated the unfolding properties of nucleocapsid protein following its ability to favor hybrid RNA/DNA duplex. RNA U3 III+IV G4 was incubated for 0 – 6 h with equimolar amount of complementary DNA strand in the absence or presence of nucleocapsid protein. Duplex and quadruplex species were separated on a native gel (Figure 4.27A). The

quantification of corresponding bands put in evidence that in presence of NCp the rate of duplex formation was higher of about 10% compared to the samples incubated in absence of the protein (Figure 4.27B).

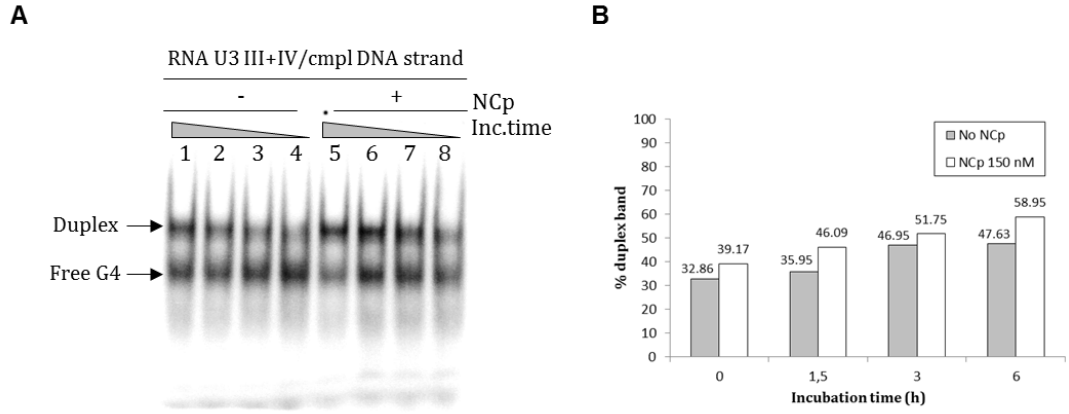


Figure 4.27 EMSA assay on RNA U3 III+IV G4 forming sequence **(A)** RNA G4 was folded in 50 mM KCl buffer and incubated with 2-fold excess of cold complementary DNA strand for 6h – 3h – 1.5h – 0h in absence (lanes 1-4) and presence (lanes 4-8) of 150 nM NCp. Duplex and free G4 species are indicated with arrows. **(B)** Quantification of duplex form bands.

This effect was greatly enhanced in the prolonged time scale. After the incubation for 24 h with 300 nM of nucleocapsid protein, the free G4 species disappeared completely, while the samples incubated only with two molar excess of the complementary strand still presented 40% of free RNA G4 (Figure 4.28). Again, incubation in the 0 – 6 h time range showed the increase of duplex species of around 10% in presence of nucleocapsid protein compared to the untreated samples (Figure 4.28B).

These preliminary data suggested that, in presence of nucleocapsid protein the G4 conformation is less stable and more prone to be unfolded.

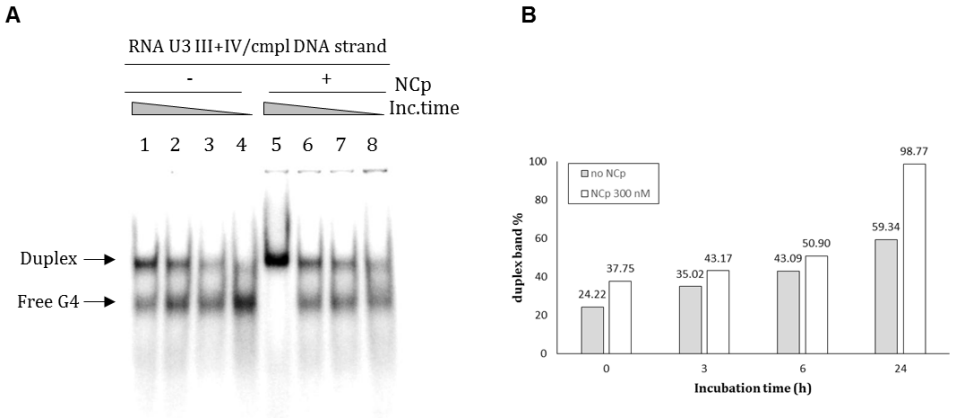


Figure 4.28 EMSA assay on RNA U3 III+IV G4 forming sequence **(A)** RNA G4 was folded in 50 mM KCl buffer and incubated with 2-fold excess of cold complementary DNA strand for 24h – 6h – 3h – 0h in absence (lanes 1-4) and presence (lanes 4-8) of 300 nM NCp. Duplex and free G4 species are indicated with arrows. **(B)** Quantification of duplex form bands.

We then investigated if these unfolding properties may abolish the stalling of reverse transcriptase enzyme induced by RNA G4s formation during the cDNA synthesis. We performed the RT stop assay in the presence of increasing concentrations of nucleocapsid protein (Figure 4.29). We were able to assess a direct unfolding effect on the G4s, as the bands corresponding to G4s decreased in presence of the protein (Figure 4.29A, lanes 3-5). In addition, we observed an augmented rate of full-length products, when the templates were elongated in the presence of NCp (Figure 4.29A, lanes 3-5, and B). When RNA U3 III+IV G4 templates were treated with increasing concentrations of BRACO-19, the same full-length increment effect was observed compared to templates not treated with the nucleocapsid protein (Figure 4.29A, lanes 6-13, and B). However, the full-length products enhancement decreased in BRACO-19 concentration dependent manner (Figure 4.29A, lanes 10-13). In fact, in the presence of NCp, BRACO-19 effectively stabilized G4s in the template, suggesting that the

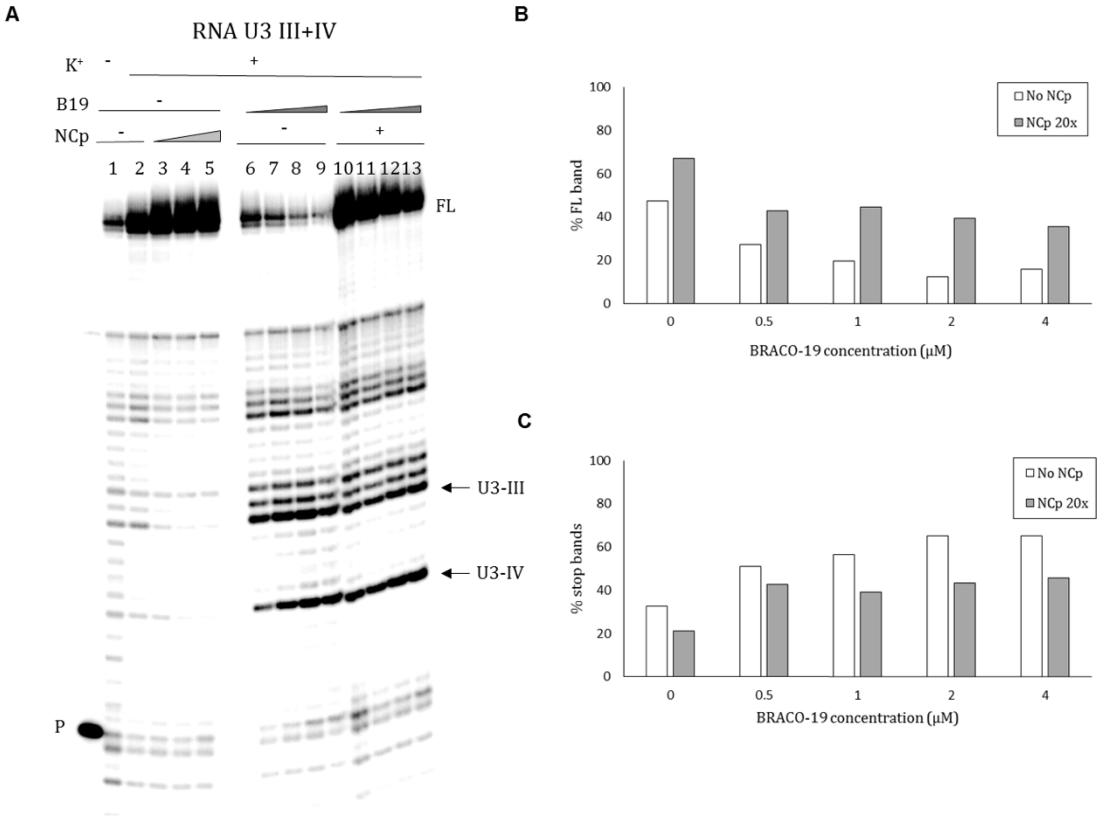


Figure 4.29 RT Stop Assay on RNA U3 III+IV G4 forming template. **(A)** Lane 1, sample folded in absence of K+, lanes 2 – 13, samples folded in 50 mM KCl solution. The presence/absence of BRACO-19 (B19) and NCp is indicated by +/- symbols respectively. Increasing concentrations of the compound and protein are indicated by triangles. BRACO-19 concentrations used were 500 nM – 1 μM – 2 μM – 4 μM, NCp concentrations were 75 nM – 150 nM – 300 nM. Primer is labeled with P, full-length elongation product with FL. Arrows indicate stop bands corresponding to the RNA U3-III and U3-IV G4 species formation. **(B)** FL product bands quantification. **(C)** U3-III stop bands quantification.

unfolding activity of nucleocapsid protein is less pronounced when these structures are stabilized by a G4 ligand (Figure 4.29A, lanes 10-13, and C).

Taken together these preliminary data suggests that the nucleocapsid protein unfolds G4 structures that form in HIV-1 RNA genome and that would impair reverse transcription. In the absence of G4 stabilizing compounds the nucleocapsid protein is able to destabilize the structures to allow complete synthesis of DNA, while higher concentrations of the ligand diminish this effect, further confirming that introducing a G4 ligand in the reverse transcription may result in inhibition of the process with consequent effect on the viral cycle.

4.4.2 U3 RNA G-quadruplexes – cellular proteins interactions

Our group recently reported that the cellular protein nucleolin specifically recognizes and stabilizes HIV-1 LTR G4s, increasing LTR promoter activity. We have also acquired indications that the viral nucleocapsid protein negatively controls the G4 formation at the RNA genome level to ensure reverse transcription progression. At this point, we asked ourselves whether cellular proteins exist that modulate viral RNA G4s folding with consequent effects on the reverse transcription process.

To investigate if the RNA U3 III+IV G4 can interact with cellular proteins, EMSA experiments was performed. The binding of RNA to proteins slows down the electrophoretic mobility of the labeled nucleic acids resulting in shifted bands on a native polyacrilamide gel. The HEK293T cell extracts, divided in cytoplasmic and nuclear fractions, were used to investigate potential protein binding partners. The first EMSA experiments showed different elec-

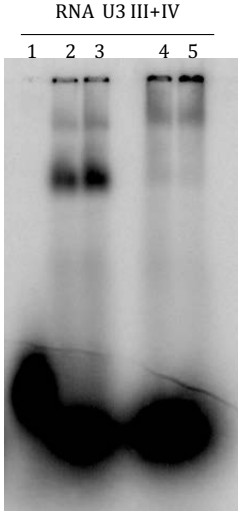


Figure 4.30. EMSA on RNA U3 III+IV oligonucleotide in presence of cell extracts. Lane 1 – oligonucleotide folded in 100 mM KCl buffer, lanes 2, 3 – folded oligonucleotide with cytoplasmic fraction of cell extracts (15 ug and 25 ug respectively), lanes 4, 5 - folded oligonucleotide with nuclear fraction of cell extracts (15 ug and 25 ug respectively).

trophoretic mobility of folded RNA oligonucleotide especially in the presence of cytoplasmic fraction of extracts (Figure 4.30, lanes 2-3). The nuclear fraction caused shifted bands of lower intensity, showing only weak binding to the G4 motif (Figure 4.30, lanes 4-5).

We mainly focused our attention on the cytoplasmic fraction of protein extracts, because HIV-1 RNA G4s are prevalently formed in cell cytoplasm prior to or during reverse transcription. To investigate if the binding of cell extracts could be specifically ascribed to the formation of the G4 structures, mutated sequences were considered. In previous studies, performed with DNA G4 counterparts, the substitution of one guanine in the repeats 4 and/or 5 with a thymine successfully disrupted DNA G4 structure. Therefore, the same point mutation was done to abolish the G4 folding of RNA oligonucleotides (Table 4.4). In addition a two points mutant (m4+5) and a scrambled RNA template that has the same base composition but no ability to form a G4 structure were considered (Table 4.4).

oligonucleotide	Sequence (5'-3')	Length
RNA III	GGGAGGCGUGGCCUGGGCGGGACUGGGG	28
RNA IV	GGGCGGGACUGGGGAGUGG	19
RNA III+IVwt	GGGAGGCGUGGCCUGGGCGGGACUGGGGAGUGG	32
RNA III+IV m5	GGGAGGCGUGGCCUGGGCGUGACUGGGGAGUGG	32
RNA III+IV m4+5	GGGAGGCGUGGCCUGUGCGUGACUGGGGAGUGG	32
RNA scrambled	GAGCGUGCGCGUGCGAGCGUGAGUGAGCGUGGG	32
TAR	GGCAGAUCUGAGCCUGGGAGCUCUCUGCC	29

Table 4.4. RNA oligonucleotides used for EMSA experiments

To confirm that mutated RNA sequences do not fold into G4 conformation, CD spectra in the absence and in the presence of K^+ were registered. Unfortunately, in the presence of 100 mM of potassium RNA U3 III+IV m5 and RNA U3 III+IV m4+5 displayed a typical CD spectra indicating the G4 formation was not completely abolished by single or double point mutation. In fact, RNA sequences have a major potential in forming secondary structures compared to their DNA counterparts. As shown in Figure 4.31, both oligonucleotides generated dichroic signal with a maximum at 260 nm and minimum at 240 nm. In contrast the RNA scrambled template showed a completely different spectrum in the presence of potassium that confirmed its inability to form G4.

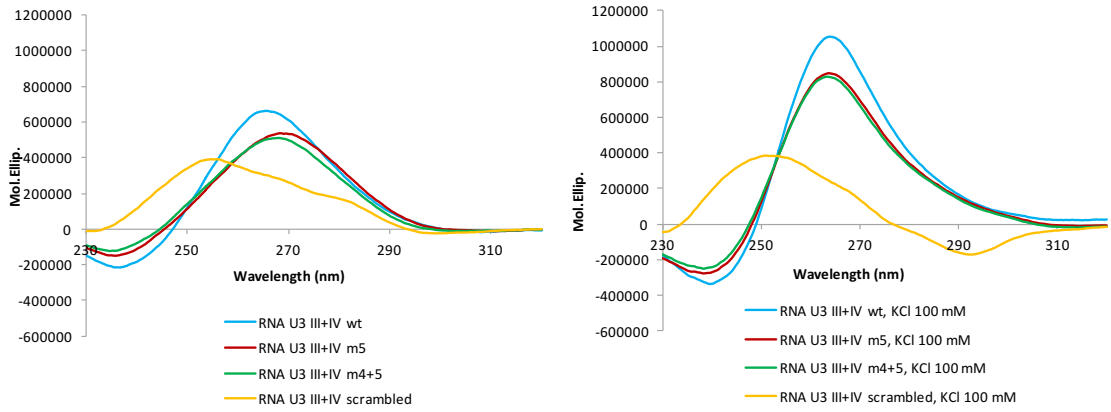


Figure 4.31. CD spectra of wildtype, mutant and scrambled RNA U3 III+IV sequences in absence (left) and presence (right) of 100 mM KCl.

Thus, wild type, mutant and scrambled sequences were incubated with 15 ug of cytoplasmic and nuclear extracts in presence of 100 mM of potassium to check their affinity toward the components of the extracts generating shift bands. The mutant oligonucleotides RNA U3 III+IVm5 and RNA U3 III+IVm4+5 displayed the same shift as the wild type sequence on a native polyacrilamide gel, while the scrambled template did not display the same binding to the cell extract components (Figure 4.32).

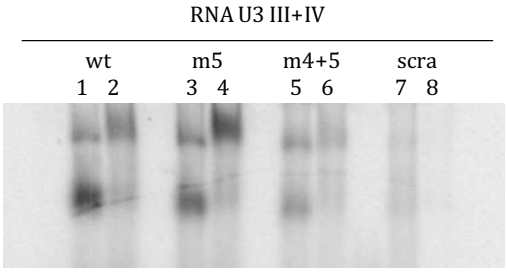


Figure 4.32. EMSA of wt and mutant RNA U3 III+IV sequences with cell extracts. Odd lanes – RNA U3 III+IV sequences with 15 ug cytoplasmatic fraction of cell extract. Even lanes RNA U3 III+IV sequences with 15 ug nuclear fraction of cell extract.

Based on the CD data, cell extracts are specific for the G4 conformation. For further analysis the scrambled template was used to confirm the specificity of cell extract toward RNA G4 conformations.

To further investigate the specificity of the binding for selected RNA G4 structures an EMSA competition assay was set up. Increasing concentrations of unlabeled competitors were added to the constant concentration of the labeled G4 oligonulceotide. When the unlabeled oligonucleotide competes with the labeled sequence for the binding with a protein the intensity of the band decreases in a competitor's concentration-dependent manner.

In Figure 4.32 the competition of the scrambled template is shown. At low scrambled oligonucleotide concentration, this sequence did not compete for the protein binding with a wildtype G4 forming sequence (Figure 4.32, lanes 7-8), while at high concentrations the competition was comparable to the wt sequence self-competition (Figure 4.32, lanes 4-6, 9-11). These data indicated a mild and not absolute specificity of the cell extracts to the G4 conformation.

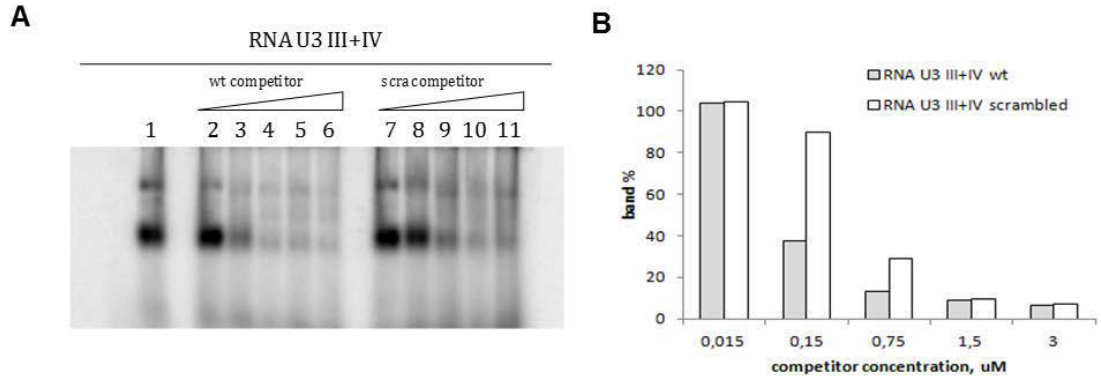


Figure 4.32. EMSA competition assay on RNA U3 III+IV wt sequence**. (A)** Lane 1 – RNA U3 III+IV wt (15 nM) incubated with 15 ug of cytoplasmatic fraction of cell extract. Lanes 2-6, RNA U3 III+IV wt (15 nM) incubated with 15ug of cytoplasmatic fraction of cell extract in presence of increasing concentrations of not labeled wt sequence (15 nM, 150 nM, 750 nM, 1,5uM and 3 uM respectively). Lanes 7-11, RNA U3 III+IV wt (15 nM) incubated with 15ug of cytoplasmatic fraction of cell extract in presence of increasing concentrations of not labeled scrambled sequence (15 nM, 150 nM, 750 nM, 1,5uM and 3 uM respectively). **(B)** Bands intensity quantification.

It is worth noting that the guanine content of the scrambled sequence is the same of the wild type G4 forming RNA sequence, rearranged in different positions in order to abolish intramolecular G4 folding. However, at high strand concentrations formation of intramolecular species in such guanine enriched sequences is not excluded.

We, thus, wanted to evaluate the binding behavior employing an RNA sequence with lower guanine content but able to form other secondary structures. At this point TAR sequence (Table 4.4) was chosen to evaluate the specificity of the binding toward. It is well known that TAR sequence is not able to form G4, but it folds into a hairpin structure. TAR did not compete for binding in the competition assay even at 200-fold molar excess (Figure 4.33), indicating the specificity of the cellular protein for G-rich, possibly G4-folded sequences.

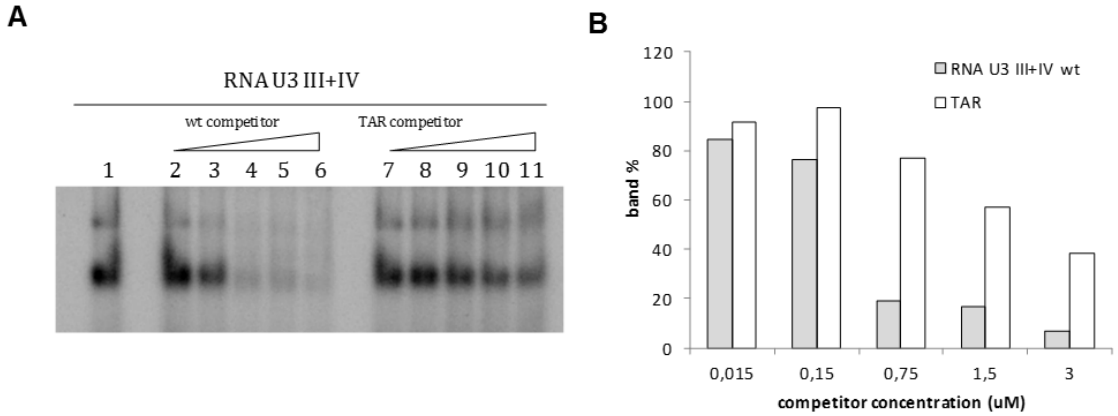


Figure 4.33. EMSA competition assay on RNA U3 III+IV wt sequence**. (A)** Lane 1 – RNA U3 III+IV wt (15 nM) incubated with 15 ug of cytoplasmatic fraction of cell extract. Lanes 2-6, RNA U3 III+IV wt (15 nM) incubated with 15ug of cytoplasmatic fraction of cell extract in presence of increasing concentrations of not labeled wt sequence (15 nM, 150 nM, 750 nM, 1,5uM and 3 uM respectively). Lanes 7-11, RNA U3 III+IV wt (15 nM) incubated with 15ug of cytoplasmatic fraction of cell extract in presence of increasing concentrations of not labeled TAR sequence (15 nM, 150 nM, 750 nM, 1,5uM and 3 uM respectively). **(B)** Bands intensity quantification

These preliminary data suggested the possibility that protein components from the cytoplasmic fraction of cell extracts may exhibit specific binding towards RNA G4s. Further investigations are necessary to clearly assess the specificity of the binding, identify the binding partner and its role in the modutation of RNA G4 folding.

5 Conclusions

The first part of this thesis was focused on the investigation of DNA G4 structures in the HIV-1 LTR promoter region, previously characterized as inhibitory elements of viral transcription, as antiviral targets. We tested a newly synthesized series of c-exNDI compounds for selective binding towards viral G4s. Spectroscopic analysis combined with polymerase enzyme inhibition assays confirmed preferential binding of these series of compounds to viral LTR G4s, compared to one of the most abundant cellular G4s, i.e. telomeric G4. The design of the side-chain substituents of c-exNDI family of compounds allowed us to evaluate several structure-activity relationships which the selectivity was relied on. Particularly, NO_2 moieties located in the side-chains of extended NDI core are well tolerated, as they lend the best selectivity and activity properties, while carboxylate groups decreased G4 affinity. Bulky substituents, directly attached to the aromatic core decrease binding specificity, probably because of planarity distortion. The selectivity and high affinity toward LTR G4s of the best c-exNDIs were translated in low nM range antiviral activity with good selectivity indexes, defining promising therapeutic windows. These data strongly support the possibility to develop selective antiviral molecules with G4-mediated mechanism of action.

In this direction, a rational approach for the design of G4 binding ligands may significantly improve the selectivity toward viral G4s. We provided here a starting point to the rational drug design approach by defining solution structure of LTR-III G4, the major component within the LTR G4-forming motif. LTR-III G4 revealed a particular (3+1) folding topology with singular loops conformations, i.e. abasic V-shaped loop and long 11-bases loop structured as a duplex stem-loop. Interestingly, the unique features of LTR-III conformation are conserved in the dynamic expanded context of the LTR G-rich sequence. The unique hybrid duplex/quadruplex topology is an attractive motif for selective drug recognition of this viral G4 structure represented particularly by the pocket available on the junction between G-

tetrad core and stem-loop. Refining the selectivity of G4 binding ligands towards particular features of the viral G4 structure by rational drug design approach may be a strategy to selectively impair the viral cycle targeting G4 related mechanisms.

In the second part of the thesis the major focus was on the RNA G4 structures that form in the U3 region of HIV-1 ssRNA genome. Investigations on the antiviral activity of G4 ligand BRACO-19 revealed that reverse transcription process is a pre-integration target of the compound. Thus, we analyzed the RNA sequence corresponding to the previously studied U3 LTR G4 forming region. We confirmed formation of three extremely stable parallel RNA G4 structures, which were able to block reverse transcriptase processing, causing enzyme stalling at the guanine bases involved in G4 formation. BRACO-19 strongly enhanced this inhibitory effect by targeting RNA G4s. We also pointed out that HIV-1 nucleocapsid protein can be implicated in controlling the formation of RNA G4s by unfolding these structures and facilitate reverse transcriptase progression.

Taken together, the results of this thesis strongly support a role of the G4s that form in the viral and proviral genome in controlling the HIV-1 life cycle. As a consequence, targeting of HIV-1 G4 may provide an invaluable new tool to inhibit the virus. In particular, targeting viral structures in the proviral genome would allow the unprecedented possibility to attack both the actively replicating and latent virus. The inability of the current anti-HIV-1 armamentarium to target the latter is the main cause of the failure to cure the infection; the possibility to effectively address both the latent and active viral forms could in principle allow viral eradication. Rational drug design combined to the high-throughput screenings may provide effective tools for development of selective compounds. Given the fact that the majority of G4 binding ligands tested so far display structural features directed to target G-tetrads prevalently by stacking interaction and the clear selectivity is achieved only over duplex conformation, our findings open new perspectives to the possibility of discriminating between different G4 conformations. The future approach should be directed to the development of small molecules with structural features compatible with unique loop sequences and arrangements.

6 References

1. Choi, J. and Majima, T. (2011) Conformational changes of non-B DNA. *Chemical Society reviews*, **40**, 5893-5909.
2. Wang, G. and Vasquez, K.M. (2014) Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA repair*, **19**, 143-151.
3. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America*, **48**, 2013-2018.
4. Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825-829.
5. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic acids research*, **34**, 5402-5415.
6. Ou, T.M., Lu, Y.J., Tan, J.H., Huang, Z.S., Wong, K.Y. and Gu, L.Q. (2008) G-quadruplexes: targets in anticancer drug design. *ChemMedChem*, **3**, 690-713.
7. Bhattacharyya, D., Mirihana Arachchilage, G. and Basu, S. (2016) Metal Cations in G-Quadruplex Folding and Stability. *Frontiers in chemistry*, **4**, 38.
8. Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876-880.
9. Haider, S., Parkinson, G.N. and Neidle, S. (2002) Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. *Journal of molecular biology*, **320**, 189-200.
10. Lee, M.P., Parkinson, G.N., Hazel, P. and Neidle, S. (2007) Observation of the coexistence of sodium and calcium ions in a DNA G-quadruplex ion channel. *Journal of the American Chemical Society*, **129**, 10106-10107.
11. Podbevsek, P., Hud, N.V. and Plavec, J. (2007) NMR evaluation of ammonium ion movement within a unimolecular G-quadruplex in solution. *Nucleic acids research*, **35**, 2554-2563.
12. Wei, D., Parkinson, G.N., Reszka, A.P. and Neidle, S. (2012) Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic acids research*, **40**, 4691-4700.
13. Campbell, N.H. and Neidle, S. (2012) G-quadruplexes and metal ions. *Metal ions in life sciences*, **10**, 119-134.
14. Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E.S., Bowers, M.T., De Pauw, E. and Gabelica, V. (2008) G-quadruplex DNA assemblies: loop length, cation identity,

and multimer formation. *Journal of the American Chemical Society*, **130**, 10208-10216.
15. Sen, D. and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*, **344**, 410-414.
16. Gaynutdinov, T.I., Neumann, R.D. and Panyutin, I.G. (2008) Structural polymorphism of intramolecular quadruplex of human telomeric DNA: effect of cations, quadruplex-binding drugs and flanking sequences. *Nucleic acids research*, **36**, 4079-4087.
17. Zhang, S., Wu, Y. and Zhang, W. (2014) G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem*, **9**, 899-911.
18. Karsisiotis, A.I., O'Kane, C. and Webba da Silva, M. (2013) DNA quadruplex folding formalism--a tutorial on quadruplex topologies. *Methods*, **64**, 28-35.
19. Webba da Silva, M. (2007) Geometric formalism for DNA quadruplex folding. *Chemistry*, **13**, 9738-9745.
20. Karsisiotis, A.I., Hessari, N.M., Novellino, E., Spada, G.P., Randazzo, A. and Webba da Silva, M. (2011) Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angewandte Chemie*, **50**, 10645-10648.
21. Heddi, B., Martin-Pintado, N., Serimbetov, Z., Kari, T.M. and Phan, A.T. (2016) G-quadruplexes with (4n - 1) guanines in the G-tetrad core: formation of a G-triad.water complex and implication for small-molecule binding. *Nucleic acids research*, **44**, 910-916.
22. Mukundan, V.T. and Phan, A.T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *Journal of the American Chemical Society*, **135**, 5017-5028.
23. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic acids research*, **33**, 2908-2916.
24. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P. and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature biotechnology*, **33**, 877-881.
25. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic acids research*, **35**, 406-413.
26. Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic acids research*, **43**, 8627-8637.
27. Lu, W., Zhang, Y., Liu, D., Songyang, Z. and Wan, M. (2013) Telomeres-structure, function, and regulation. *Experimental cell research*, **319**, 133-141.
28. De Boeck, G., Forsyth, R.G., Praet, M. and Hogendoorn, P.C. (2009) Telomere-associated proteins: cross-talk between telomere maintenance and telomere-lengthening mechanisms. *The Journal of pathology*, **217**, 327-344.
29. Nandakumar, J. and Cech, T.R. (2013) Finding the end: recruitment of telomerase to telomeres. *Nature reviews. Molecular cell biology*, **14**, 69-82.
30. Greider, C.W. (1999) Telomeres do D-loop-T-loop. *Cell*, **97**, 419-422.
31. Smogorzewska, A., van Steensel, B., Bianchi, A., Oelmann, S., Schaefer, M.R., Schnapp, G. and de Lange, T. (2000) Control of human telomere length by TRF1 and TRF2. *Molecular and cellular biology*, **20**, 1659-1668.
32. de Lange, T. (2002) Protection of mammalian telomeres. *Oncogene*, **21**, 532-540.
33. Martinez, P. and Blasco, M.A. (2015) Replicating through telomeres: a means to an end. *Trends in biochemical sciences*, **40**, 504-515.
34. Collins, K. (2006) The biogenesis and regulation of telomerase holoenzymes. *Nature reviews. Molecular cell biology*, **7**, 484-494.
35. Neidle, S. (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *The FEBS journal*, **277**, 1118-1125.

36. Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews. Genetics*, **13**, 770-780.
37. Phan, A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *The FEBS journal*, **277**, 1107-1117.
38. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. and Lipps, H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nature structural & molecular biology*, **12**, 847-854.
39. Zaug, A.J., Podell, E.R. and Cech, T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 10864-10869.
40. Lipps, H.J. and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends in cell biology*, **19**, 414-422.
41. De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.F. and Mergny, J.L. (2008) Targeting telomeres and telomerase. *Biochimie*, **90**, 131-155.
42. Wickramasinghe, C.M., Arzouk, H., Frey, A., Maiter, A. and Sale, J.E. (2015) Contributions of the specialised DNA polymerases to replication of structured DNA. *DNA repair*, **29**, 83-90.
43. Mendoza, O., Bourdoncle, A., Boule, J.B., Brosh, R.M., Jr. and Mergny, J.L. (2016) G-quadruplexes and helicases. *Nucleic acids research*, **44**, 1989-2006.
44. Paeschke, K., Bochman, M.L., Garcia, P.D., Cejka, P., Friedman, K.L., Kowalczykowski, S.C. and Zakian, V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458-462.
45. Paeschke, K., Capra, J.A. and Zakian, V.A. (2011) DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*, **145**, 678-691.
46. Piazza, A., Boule, J.B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M.P. and Nicolas, A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic acids research*, **38**, 4337-4348.
47. London, T.B., Barber, L.J., Mosedale, G., Kelly, G.P., Balasubramanian, S., Hickson, I.D., Boulton, S.J. and Hiom, K. (2008) FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *The Journal of biological chemistry*, **283**, 36132-36139.
48. Leon-Ortiz, A.M., Svendsen, J. and Boulton, S.J. (2014) Metabolism of DNA secondary structures at the eukaryotic replication fork. *DNA repair*, **19**, 152-162.
49. Murat, P. and Balasubramanian, S. (2014) Existence and consequences of G-quadruplex structures in DNA. *Current opinion in genetics & development*, **25**, 22-29.
50. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature reviews. Drug discovery*, **10**, 261-275.
51. Huppert, J.L. (2007) Four-stranded DNA: cancer, gene regulation and drug development. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **365**, 2969-2984.
52. Huppert, J.L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chemical Society reviews*, **37**, 1375-1384.
53. Qin, Y. and Hurley, L.H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **90**, 1149-1171.
54. Verdian Doghaei, A., Housaindokht, M.R. and Bozorgmehr, M.R. (2015) Molecular crowding effects on conformation and stability of G-quadruplex DNA structure: insights from molecular dynamics simulation. *Journal of theoretical biology*, **364**, 103-112.

55. Zhou, J., Tateishi-Karimata, H., Mergny, J.L., Cheng, M., Feng, Z., Miyoshi, D., Sugimoto, N. and Li, C. (2016) Reevaluation of the stability of G-quadruplex structures under crowding conditions. *Biochimie*, **121**, 204-208.
56. Sun, D. and Hurley, L.H. (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *Journal of medicinal chemistry*, **52**, 2863-2874.
57. Dai, J., Chen, D., Jones, R.A., Hurley, L.H. and Yang, D. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic acids research*, **34**, 5133-5144.
58. Dai, J., Dexheimer, T.S., Chen, D., Carver, M., Ambrus, A., Jones, R.A. and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *Journal of the American Chemical Society*, **128**, 1096-1098.
59. Dexheimer, T.S., Sun, D. and Hurley, L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *Journal of the American Chemical Society*, **128**, 5404-5415.
60. Qin, Y., Fortin, J.S., Tye, D., Gleason-Guzman, M., Brooks, T.A. and Hurley, L.H. (2010) Molecular cloning of the human platelet-derived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-beta expression. *Biochemistry*, **49**, 4208-4219.
61. Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S. and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry*, **45**, 7854-7860.
62. Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *Journal of the American Chemical Society*, **127**, 10584-10589.
63. Guo, K., Gokhale, V., Hurley, L.H. and Sun, D. (2008) Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic acids research*, **36**, 4598-4608.
64. Tong, X., Lan, W., Zhang, X., Wu, H., Liu, M. and Cao, C. (2011) Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence. *Nucleic acids research*, **39**, 6753-6763.
65. Qin, Y., Rezler, E.M., Gokhale, V., Sun, D. and Hurley, L.H. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic acids research*, **35**, 7698-7713.
66. Guo, K., Pourpak, A., Beetz-Rogers, K., Gokhale, V., Sun, D. and Hurley, L.H. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *Journal of the American Chemical Society*, **129**, 10220-10228.
67. Sun, D., Guo, K., Rusche, J.J. and Hurley, L.H. (2005) 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 research*, **33**, 6070-6080.
68. Sun, D., Liu, W.J., Guo, K., Rusche, J.J., Ebbinghaus, S., Gokhale, V. and Hurley, L.H. (2008) The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Molecular cancer therapeutics*, **7**, 880-889.
69. Brooks, T.A., Kendrick, S. and Hurley, L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *The FEBS journal*, **277**, 3459-3469.

70. Chen, B.J., Wu, Y.L., Tanaka, Y. and Zhang, W. (2014) Small molecules targeting c-Myc oncogene: promising anti-cancer therapeutics. *International journal of biological sciences*, **10**, 1084-1096.
71. Gonzalez, V. and Hurley, L.H. (2010) The c-MYC NHE III(1): function and regulation. *Annual review of pharmacology and toxicology*, **50**, 111-129.
72. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11593-11598.
73. Yang, D. and Hurley, L.H. (2006) Structure of the biologically relevant G-quadruplex in the c-MYC promoter. *Nucleosides, nucleotides & nucleic acids*, **25**, 951-968.
74. Ambrus, A., Chen, D., Dai, J., Jones, R.A. and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048-2058.
75. Brooks, T.A. and Hurley, L.H. (2010) Targeting MYC Expression through G-Quadruplexes. *Genes & cancer*, **1**, 641-649.
76. Gonzalez, V., Guo, K., Hurley, L. and Sun, D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *The Journal of biological chemistry*, **284**, 23622-23635.
77. Hurley, L.H., Von Hoff, D.D., Siddiqui-Jain, A. and Yang, D. (2006) Drug targeting of the c-MYC promoter to repress gene expression via a G-quadruplex silencer element. *Seminars in oncology*, **33**, 498-512.
78. Agarwala, P., Pandey, S. and Maiti, S. (2015) The tale of RNA G-quadruplex. *Organic & biomolecular chemistry*, **13**, 5570-5585.
79. Agarwal, T., Jayaraj, G., Pandey, S.P., Agarwala, P. and Maiti, S. (2012) RNA G-quadruplexes: G-quadruplexes with "U" turns. *Current pharmaceutical design*, **18**, 2102-2111.
80. Pandey, S., Agarwala, P. and Maiti, S. (2013) Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *The journal of physical chemistry. B*, **117**, 6896-6905.
81. Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic acids research*, **40**, 4727-4741.
82. Millevoi, S., Moine, H. and Vagner, S. (2012) G-quadruplexes in RNA biology. *Wiley interdisciplinary reviews. RNA*, **3**, 495-507.
83. Kumari, S., Bugaut, A., Huppert, J.L. and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature chemical biology*, **3**, 218-221.
84. Bugaut, A., Rodriguez, R., Kumari, S., Hsu, S.T. and Balasubramanian, S. (2010) Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Organic & biomolecular chemistry*, **8**, 2771-2776.
85. Morris, M.J., Negishi, Y., Pazsint, C., Schonhoft, J.D. and Basu, S. (2010) An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *Journal of the American Chemical Society*, **132**, 17831-17839.
86. Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A.C. and Vagner, S. (2003) A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *The Journal of biological chemistry*, **278**, 39330-39336.
87. Agarwala, P., Pandey, S., Mapa, K. and Maiti, S. (2013) The G-quadruplex augments translation in the 5' untranslated region of transforming growth factor beta2. *Biochemistry*, **52**, 1528-1538.
88. Christiansen, J., Kofod, M. and Nielsen, F.C. (1994) A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. *Nucleic acids research*, **22**, 5709-5716.

89. Decorsiere, A., Cayrel, A., Vagner, S. and Millevoi, S. (2011) Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes & development*, **25**, 220-225.
90. Endoh, T., Kawasaki, Y. and Sugimoto, N. (2013) Translational halt during elongation caused by G-quadruplex formed by mRNA. *Methods*, **64**, 73-78.
91. Chen, Y. and Yang, D. (2012) Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. *Current protocols in nucleic acid chemistry*, **Chapter 17**, Unit17 15.
92. Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Current opinion in structural biology*, **19**, 239-250.
93. Neidle, S. and Read, M.A. (2000) G-quadruplexes as therapeutic targets. *Biopolymers*, **56**, 195-208.
94. Hurley, L.H., Wheelhouse, R.T., Sun, D., Kerwin, S.M., Salazar, M., Fedoroff, O.Y., Han, F.X., Han, H., Izbicka, E. and Von Hoff, D.D. (2000) G-quadruplexes as targets for drug design. *Pharmacology & therapeutics*, **85**, 141-158.
95. Monchaud, D. and Teulade-Fichou, M.P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Organic & biomolecular chemistry*, **6**, 627-636.
96. Read, M., Harrison, R.J., Romagnoli, B., Tanious, F.A., Gowan, S.H., Reszka, A.P., Wilson, W.D., Kelland, L.R. and Neidle, S. (2001) Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4844-4849.
97. Haider, S.M., Parkinson, G.N. and Neidle, S. (2003) Structure of a G-quadruplex-ligand complex. *Journal of molecular biology*, **326**, 117-125.
98. Haider, S.M., Neidle, S. and Parkinson, G.N. (2011) A structural analysis of G-quadruplex/ligand interactions. *Biochimie*, **93**, 1239-1251.
99. Read, M.A., Wood, A.A., Harrison, J.R., Gowan, S.M., Kelland, L.R., Dosanjh, H.S. and Neidle, S. (1999) Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: structure-activity relationships. *Journal of medicinal chemistry*, **42**, 4538-4546.
100. Campbell, N.H., Parkinson, G.N., Reszka, A.P. and Neidle, S. (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *Journal of the American Chemical Society*, **130**, 6722-6724.
101. Cuesta, J., Read, M.A. and Neidle, S. (2003) The design of G-quadruplex ligands as telomerase inhibitors. *Mini reviews in medicinal chemistry*, **3**, 11-21.
102. Gowan, S.M., Harrison, J.R., Patterson, L., Valenti, M., Read, M.A., Neidle, S. and Kelland, L.R. (2002) A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity. *Molecular pharmacology*, **61**, 1154-1162.
103. Burger, A.M., Dai, F., Schultes, C.M., Reszka, A.P., Moore, M.J., Double, J.A. and Neidle, S. (2005) The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer research*, **65**, 1489-1496.
104. Tanious, F.A., Yen, S.F. and Wilson, W.D. (1991) Kinetic and equilibrium analysis of a threading intercalation mode: DNA sequence and ion effects. *Biochemistry*, **30**, 1813-1819.
105. Milelli, A., Tumiatti, V., Micco, M., Rosini, M., Zuccari, G., Raffaghello, L., Bianchi, G., Pistoia, V., Fernando Diaz, J., Pera, B. *et al.* (2012) Structure-activity relationships of novel substituted naphthalene diimides as anticancer agents. *European journal of medicinal chemistry*, **57**, 417-428.
106. Tumiatti, V., Milelli, A., Minarini, A., Micco, M., Gasperi Campani, A., Roncuzzi, L., Baiocchi, D., Marinello, J., Capranico, G., Zini, M. *et al.* (2009) Design, synthesis, and biological evaluation of substituted naphthalene imides and diimides as anticancer agent. *Journal of medicinal chemistry*, **52**, 7873-7877.

107. Cuenca, F., Greciano, O., Gunaratnam, M., Haider, S., Munnur, D., Nanjunda, R., Wilson, W.D. and Neidle, S. (2008) Tri- and tetra-substituted naphthalene diimides as potent G-quadruplex ligands. *Bioorganic & medicinal chemistry letters*, **18**, 1668-1673.
108. Collie, G.W., Promontorio, R., Hampel, S.M., Micco, M., Neidle, S. and Parkinson, G.N. (2012) Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. *Journal of the American Chemical Society*, **134**, 2723-2731.
109. Hampel, S.M., Pepe, A., Greulich-Bode, K.M., Malhotra, S.V., Reszka, A.P., Veith, S., Boukamp, P. and Neidle, S. (2013) Mechanism of the antiproliferative activity of some naphthalene diimide G-quadruplex ligands. *Molecular pharmacology*, **83**, 470-480.
110. Hampel, S.M., Sidibe, A., Gunaratnam, M., Riou, J.F. and Neidle, S. (2010) Tetrasubstituted naphthalene diimide ligands with selectivity for telomeric G-quadruplexes and cancer cells. *Bioorganic & medicinal chemistry letters*, **20**, 6459-6463.
111. Gunaratnam, M., Swank, S., Haider, S.M., Galesa, K., Reszka, A.P., Beltran, M., Cuenca, F., Fletcher, J.A. and Neidle, S. (2009) Targeting human gastrointestinal stromal tumor cells with a quadruplex-binding small molecule. *Journal of medicinal chemistry*, **52**, 3774-3783.
112. Parkinson, G.N., Cuenca, F. and Neidle, S. (2008) Topology conservation and loop flexibility in quadruplex-drug recognition: crystal structures of inter- and intramolecular telomeric DNA quadruplex-drug complexes. *Journal of molecular biology*, **381**, 1145-1156.
113. Marchetti, C., Minarini, A., Tumiatti, V., Moraca, F., Parrotta, L., Alcaro, S., Rigo, R., Sissi, C., Gunaratnam, M., Ohnmacht, S.A. *et al.* (2015) Macrocyclic naphthalene diimides as G-quadruplex binders. *Bioorganic & medicinal chemistry*, **23**, 3819-3830.
114. Doria, F., Manet, I., Grande, V., Monti, S. and Freccero, M. (2013) Water-soluble naphthalene diimides as singlet oxygen sensitizers. *The Journal of organic chemistry*, **78**, 8065-8073.
115. Doria, F., Nadai, M., Sattin, G., Pasotti, L., Richter, S.N. and Freccero, M. (2012) Water soluble extended naphthalene diimides as pH fluorescent sensors and G-quadruplex ligands. *Organic & biomolecular chemistry*, **10**, 3830-3840.
116. Arevalo, M., Doria, F., Belmonte-Reche, E., De Rache, A., Campos-Salinas, J., Lucas, R., Falomir, E., Carda, M., Perez-Victoria, J.M., Mergny, J.L. *et al.* (2016) Synthesis, binding properties and differences on cell uptake of G-quadruplex ligands based on carbohydrate naphthalene diimide conjugates. *Chemistry*.
117. Zuffo, M., Doria, F., Spalluto, V., Ladame, S. and Freccero, M. (2015) Red/NIR G-Quadruplex Sensing, Harvesting Blue Light by a Coumarin-Naphthalene Diimide Dyad. *Chemistry*, **21**, 17596-17600.
118. Doria, F., Oppi, A., Manoli, F., Botti, S., Kandoth, N., Grande, V., Manet, I. and Freccero, M. (2015) A naphthalene diimide dyad for fluorescence switch-on detection of G-quadruplexes. *Chemical communications (Cambridge, England)*, **51**, 9105-9108.
119. Phan, A.T., Kuryavyi, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nature chemical biology*, **1**, 167-173.
120. Han, H., Langley, D.R., Rangan, A. and Hurley, L.H. (2001) Selective interactions of cationic porphyrins with G-quadruplex structures. *Journal of the American Chemical Society*, **123**, 8902-8913.
121. Gavathiotis, E., Heald, R.A., Stevens, M.F. and Searle, M.S. (2003) Drug recognition and stabilisation of the parallel-stranded DNA quadruplex d(TTAGGGT)4 containing the human telomeric repeat. *Journal of molecular biology*, **334**, 25-36.
122. Salvati, E., Leonetti, C., Rizzo, A., Scarsella, M., Mottolese, M., Galati, R., Sperduti, I., Stevens, M.F., D'Incalci, M., Blasco, M. *et al.* (2007) Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *The Journal of clinical investigation*, **117**, 3236-3247.

123. Jiang, Y.L. and Liu, Z.P. (2010) Metallo-organic G-quadruplex ligands in anticancer drug design. *Mini reviews in medicinal chemistry*, **10**, 726-736.
124. Dixon, I.M., Lopez, F., Esteve, J.P., Tejera, A.M., Blasco, M.A., Pratviel, G. and Meunier, B. (2005) Porphyrin derivatives for telomere binding and telomerase inhibition. *Chembiochem : a European journal of chemical biology*, **6**, 123-132.
125. Kim, M.Y., Vankayalapati, H., Shin-Ya, K., Wierzba, K. and Hurley, L.H. (2002) Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *Journal of the American Chemical Society*, **124**, 2098-2099.
126. Gomez, D., Paterski, R., Lemarteleur, T., Shin-Ya, K., Mergny, J.L. and Riou, J.F. (2004) Interaction of telomestatin with the telomeric single-strand overhang. *The Journal of biological chemistry*, **279**, 41487-41494.
127. Gomez, D., O'Donohue, M.F., Wenner, T., Douarre, C., Macadre, J., Koebel, P., Giraud-Panis, M.J., Kaplan, H., Kolkes, A., Shin-ya, K. *et al.* (2006) The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences in vitro and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer research*, **66**, 6908-6912.
128. Iida, K. and Nagasawa, K. (2013) Macrocyclic polyoxazoles as G-quadruplex ligands. *Chemical record (New York, N.Y.)*, **13**, 539-548.
129. Harris, L.M. and Merrick, C.J. (2015) G-quadruplexes in pathogens: a common route to virulence control? *PLoS pathogens*, **11**, e1004562.
130. Metifiot, M., Amrane, S., Litvak, S. and Andreola, M.L. (2014) G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic acids research*, **42**, 12352-12366.
131. Tuesuwan, B., Kern, J.T., Thomas, P.W., Rodriguez, M., Li, J., David, W.M. and Kerwin, S.M. (2008) Simian virus 40 large T-antigen G-quadruplex DNA helicase inhibition by G-quadruplex DNA-interactive agents. *Biochemistry*, **47**, 1896-1909.
132. Tluckova, K., Marusic, M., Tothova, P., Bauer, L., Sket, P., Plavec, J. and Viglasky, V. (2013) Human papillomavirus G-quadruplexes. *Biochemistry*, **52**, 7207-7216.
133. Norseen, J., Johnson, F.B. and Lieberman, P.M. (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *Journal of virology*, **83**, 10336-10346.
134. Artusi, S., Nadai, M., Perrone, R., Biasolo, M.A., Palu, G., Flamand, L., Calistri, A. and Richter, S.N. (2015) The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: Implications for the antiviral activity of a G-quadruplex ligand. *Antiviral research*, **118**, 123-131.
135. Barre-Sinoussi, F., Ross, A.L. and Delfraissy, J.F. (2013) Past, present and future: 30 years of HIV research. *Nature reviews. Microbiology*, **11**, 877-883.
136. UNAIDS. (2016). UNAIDS.
137. Maartens, G., Celum, C. and Lewin, S.R. (2014) HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet*, **384**, 258-271.
138. Bell, N.M. and Lever, A.M. (2013) HIV Gag polyprotein: processing and early viral particle assembly. *Trends in microbiology*, **21**, 136-144.
139. Fassati, A. (2012) Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus research*, **170**, 15-24.
140. Darlix, J.L., Godet, J., Ivanyi-Nagy, R., Fosse, P., Mauffret, O. and Mely, Y. (2011) Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *Journal of molecular biology*, **410**, 565-581.
141. Williams, M.C., Gorelick, R.J. and Musier-Forsyth, K. (2002) Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 8614-8619.
142. Darlix, J.L., de Rocquigny, H., Mauffret, O. and Mely, Y. (2014) Retrospective on the all-in-one retroviral nucleocapsid protein. *Virus research*, **193**, 2-15.

143. Urbaneja, M.A., Kane, B.P., Johnson, D.G., Gorelick, R.J., Henderson, L.E. and Casas-Finet, J.R. (1999) Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function. *Journal of molecular biology*, **287**, 59-75.
144. Tanchou, V., Gabus, C., Rogemond, V. and Darlix, J.L. (1995) Formation of stable and functional HIV-1 nucleoprotein complexes in vitro. *Journal of molecular biology*, **252**, 563-571.
145. Lapadat-Tapolsky, M., Pernelle, C., Borie, C. and Darlix, J.L. (1995) Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. *Nucleic acids research*, **23**, 2434-2441.
146. Iwatani, Y., Rosen, A.E., Guo, J., Musier-Forsyth, K. and Levin, J.G. (2003) Efficient initiation of HIV-1 reverse transcription in vitro. Requirement for RNA sequences downstream of the primer binding site abrogated by nucleocapsid protein-dependent primer-template interactions. *The Journal of biological chemistry*, **278**, 14185-14195.
147. Sleiman, D., Goldschmidt, V., Barraud, P., Marquet, R., Paillart, J.C. and Tisne, C. (2012) Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus research*, **169**, 324-339.
148. Vo, M.N., Barany, G., Rouzina, I. and Musier-Forsyth, K. (2009) HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *Journal of molecular biology*, **386**, 789-801.
149. Beltz, H., Clauss, C., Piemont, E., Ficheux, D., Gorelick, R.J., Roques, B., Gabus, C., Darlix, J.L., de Rocquigny, H. and Mely, Y. (2005) Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *Journal of molecular biology*, **348**, 1113-1126.
150. Rodriguez-Rodriguez, L., Tsuchihashi, Z., Fuentes, G.M., Bambara, R.A. and Fay, P.J. (1995) Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro. *The Journal of biological chemistry*, **270**, 15005-15011.
151. Lu, K., Heng, X. and Summers, M.F. (2011) Structural determinants and mechanism of HIV-1 genome packaging. *Journal of molecular biology*, **410**, 609-633.
152. Micco, M., Collie, G.W., Dale, A.G., Ohnmacht, S.A., Pazitna, I., Gunaratnam, M., Reszka, A.P. and Neidle, S. (2013) Structure-based design and evaluation of naphthalene diimide G-quadruplex ligands as telomere targeting agents in pancreatic cancer cells. *Journal of medicinal chemistry*, **56**, 2959-2974.
153. Rajendran, A., Endo, M., Hidaka, K., Tran, P.L., Mergny, J.L., Gorelick, R.J. and Sugiyama, H. (2013) HIV-1 nucleocapsid proteins as molecular chaperones for tetramolecular antiparallel G-quadruplex formation. *Journal of the American Chemical Society*, **135**, 18575-18585.
154. Kankia, B.I., Barany, G. and Musier-Forsyth, K. (2005) Unfolding of DNA quadruplexes induced by HIV-1 nucleocapsid protein. *Nucleic acids research*, **33**, 4395-4403.
155. Malim, M.H. and Emerman, M. (2008) HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell host & microbe*, **3**, 388-398.
156. Berkhout, B., Silverman, R.H. and Jeang, K.T. (1989) Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell*, **59**, 273-282.
157. Karn, J. (1999) Tackling Tat. *Journal of molecular biology*, **293**, 235-254.
158. Sundquist, W.I. and Heaphy, S. (1993) Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 3393-3397.

159. Marquet, R., Baudin, F., Gabus, C., Darlix, J.L., Mougel, M., Ehresmann, C. and Ehresmann, B. (1991) Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic acids research*, **19**, 2349-2357.
160. Lyonnais, S., Hounsou, C., Teulade-Fichou, M.P., Jeusset, J., Le Cam, E. and Mirambeau, G. (2002) G-quartets assembly within a G-rich DNA flap. A possible event at the center of the HIV-1 genome. *Nucleic acids research*, **30**, 5276-5283.
161. Shen, W., Gao, L., Balakrishnan, M. and Bambara, R.A. (2009) A recombination hot spot in HIV-1 contains guanosine runs that can form a G-quartet structure and promote strand transfer in vitro. *The Journal of biological chemistry*, **284**, 33883-33893.
162. Shen, W., Gorelick, R.J. and Bambara, R.A. (2011) HIV-1 nucleocapsid protein increases strand transfer recombination by promoting dimeric G-quartet formation. *The Journal of biological chemistry*, **286**, 29838-29847.
163. Perrone, R., Nadai, M., Poe, J.A., Frasson, I., Palumbo, M., Palu, G., Smithgall, T.E. and Richter, S.N. (2013) Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity. *PloS one*, **8**, e73121.
164. Perrone, R., Nadai, M., Frasson, I., Poe, J.A., Butovskaya, E., Smithgall, T.E., Palumbo, M., Palu, G. and Richter, S.N. (2013) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *Journal of medicinal chemistry*, **56**, 6521-6530.
165. Tosoni, E., Frasson, I., Scalabrin, M., Perrone, R., Butovskaya, E., Nadai, M., Palu, G., Fabris, D. and Richter, S.N. (2015) Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic acids research*, **43**, 8884-8897.
166. Perrone, R., Butovskaya, E., Daelemans, D., Palu, G., Pannecouque, C. and Richter, S.N. (2014) Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *The Journal of antimicrobial chemotherapy*, **69**, 3248-3258.
167. Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A. and Spada, G.P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Organic & biomolecular chemistry*, **8**, 2683-2692.
168. Murat, P., Singh, Y. and Defrancq, E. (2011) Methods for investigating G-quadruplex DNA/ligand interactions. *Chemical Society reviews*, **40**, 5293-5307.
169. Mergny, J.L. and Lacroix, L. (2009) UV Melting of G-Quadruplexes. *Current protocols in nucleic acid chemistry*, **Chapter 17**, Unit 17 11.
170. De Cian, A., Guittat, L., Kaiser, M., Sacca, B., Amrane, S., Bourdoncle, A., Alberti, P., Teulade-Fichou, M.P., Lacroix, L. and Mergny, J.L. (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods*, **42**, 183-195.
171. Sun, D. and Hurley, L.H. (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods in molecular biology (Clifton, N.J.)*, **608**, 65-79.
172. Nguyen, B., Tanious, F.A. and Wilson, W.D. (2007) Biosensor-surface plasmon resonance: quantitative analysis of small molecule-nucleic acid interactions. *Methods*, **42**, 150-161.
173. Redman, J.E. (2007) Surface plasmon resonance for probing quadruplex folding and interactions with proteins and small molecules. *Methods*, **43**, 302-312.
174. Adrian, M., Heddi, B. and Phan, A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods*, **57**, 11-24.
175. Webba da Silva, M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods*, **43**, 264-277.
176. Keniry, M.A. (2000) Quadruplex structures in nucleic acids. *Biopolymers*, **56**, 123-146.
177. Phan, A.T. and Patel, D.J. (2002) A site-specific low-enrichment (15)N,(13)C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *Journal of the American Chemical Society*, **124**, 1160-1161.

178. Phan, A.T. (2000) Long-range imino proton-13C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *Journal of biomolecular NMR*, **16**, 175-178.
179. Nadai, M., Sattin, G., Palu, G., Palumbo, M. and Richter, S.N. (2013) Clerocidin-mediated DNA footprinting discriminates among different G-quadruplex conformations and detects tetraplex folding in a duplex environment. *Biochimica et biophysica acta*, **1830**, 4660-4668.
180. Perrone, R., Doria, F., Butovskaya, E., Frasson, I., Botti, S., Scalabrin, M., Lago, S., Grande, V., Nadai, M., Freccero, M. *et al.* (2015) Synthesis, Binding and Antiviral Properties of Potent Core-Extended Naphthalene Diimides Targeting the HIV-1 Long Terminal Repeat Promoter G-Quadruplexes. *Journal of medicinal chemistry*, **58**, 9639-9652.
181. Kuryavyi, V. and Patel, D.J. (2010) Solution structure of a unique G-quadruplex scaffold adopted by a guanosine-rich human intronic sequence. *Structure (London, England : 1993)*, **18**, 73-82.
182. Lim, K.W. and Phan, A.T. (2013) Structural basis of DNA quadruplex-duplex junction formation. *Angewandte Chemie*, **52**, 8566-8569.
183. De Nicola, B., Lech, C.J., Heddi, B., Regmi, S., Frasson, I., Perrone, R., Richter, S.N. and Phan, A.T. (2016) Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic acids research*, **44**, 6442-6451.
184. Lim, K.W., Nguyen, T.Q. and Phan, A.T. (2014) Joining of multiple duplex stems at a single quadruplex loop. *Journal of the American Chemical Society*, **136**, 17969-17973.
185. Lim, K.W., Jenjaroenpun, P., Low, Z.J., Khong, Z.J., Ng, Y.S., Kuznetsov, V.A. and Phan, A.T. (2015) Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic acids research*, **43**, 5630-5646.
186. Onel, B., Carver, M., Wu, G., Timonina, D., Kalarn, S., Larriva, M. and Yang, D. (2016) A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription. *Journal of the American Chemical Society*, **138**, 2563-2570.
187. Ouellet, J. (2016) RNA Fluorescence with Light-Up Aptamers. *Frontiers in chemistry*, **4**, 29.

7 List of publications

Perrone, R., Butovskaya, E., Lago, S., Garzino-Demo, A., Pannecouque, C., Palù, G., Richter, S.N. (2016) The G-quadruplex-forming aptamer AS1411 potently inhibits HIV-1 attachment to the host cell. *Int J Antimicrob Agents,* **4**, 311-316.

Perrone, R., Doria, F., Butovskaya, E., Frasson, I., Botti, S., Scalabrin, M., Lago, S., Grande, V., Nadai, M., Freccero, M., Richter, S.N. (2015) Synthesis, Binding and Antiviral Properties of Potent Core-Extended Naphthalene Diimides Targeting the HIV-1 Long Terminal Repeat Promoter G-Quadruplexes. *J Med Chem,* **24**, 9639-9652.

Tosoni, E., Frasson, I., Scalabrin, M., Perrone, R., Butovskaya, E., Nadai, M., Palù, G., Fabris, D., Richter, S.N. (2015) Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Researh*, **18**, 8884-8897

Perrone, R., Butovskaya, E., Daelemans, D., Palu, G., Pannecouque, C. and Richter, S.N. (2014) Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *The Journal of antimicrobial chemotherapy*, **69**, 3248-3258.

Perrone, R., Nadai, M., Frasson, I., Poe, J.A., Butovskaya, E., Smithgall, T.E., Palumbo, M., Palu, G. and Richter, S.N. (2013) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *Journal of medicinal chemistry*, **56**, 6521-6530.