



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova  
Dipartimento di Medicina Molecolare

SCUOLA DI DOTTORATO DI RICERCA IN BIOMEDICINA  
CICLO XXVI

# **G-QUADRUPLEXES IN THE HIV-1 GENOME AS ANTIVIRAL TARGETS**

**Direttore della Scuola** : Prof. Riccardo Manganelli

**Supervisore** : Prof.ssa Sara Richter

**Dottorando** : Rosalba Perrone



# Index

<b>LIST OF ABBREVIATIONS.....</b>	<b>V</b>
<b>RIASSUNTO .....</b>	<b>VII</b>
<b>1. ABSTRACT .....</b>	<b>1</b>
<b>2. INTRODUCTION.....</b>	<b>3</b>
<b>2.1. DNA G-Quadruplexes.....</b>	<b>3</b>
2.1.1. General features .....	3
2.1.2. Structural polymorphism of G-quadruplex structures .....	5
2.1.3. Biological roles of G-quadruplexes .....	8
2.1.3.1. G-quadruplex at telomeric ends.....	8
2.1.3.2. Effects of G-quadruplex structures during DNA replication.....	13
2.1.3.3. Effects of G-quadruplex structures during transcription .....	15
2.1.3.3.1. G-quadruplexes in promoter regions.....	16
2.1.3.3.1.1. G-quadruplex in <i>c-myc</i> NHE III <sub>1</sub> .....	19
2.1.3.3.2. G-quadruplexes in coding regions .....	20
2.1.4. G-quadruplexes as targets for drug design.....	22
2.1.4.1. Interaction modes of G-quadruplex ligands.....	22
2.1.4.2. Main classes of G-quadruplex ligands .....	23
2.1.4.2.1. <i>In situ</i> protonated G-quadruplex ligands .....	23
2.1.4.2.2. N-methylated aromatic G-quadruplex ligands .....	25
2.1.4.2.3. Metallo-organic G-quadruplex ligands .....	26
2.1.4.2.4. Neutral macrocyclic G-quadruplex ligands .....	27
<b>2.2. G-quadruplex in microorganisms .....</b>	<b>28</b>
<b>2.3. The Human Immunodeficiency Virus (HIV) .....</b>	<b>29</b>
2.3.1. Viral structure and genome .....	30
2.3.2. HIV replication cycle .....	37
2.3.3. Antiretroviral treatment .....	42
2.3.4. G-quadruplexes and HIV-1.....	43
<b>3. AIM OF THE STUDY .....</b>	<b>45</b>
<b>4. MATERIALS AND METHODS.....</b>	<b>47</b>
<b>4.1. Materials and compounds.....</b>	<b>47</b>
<b>4.2. Cells .....</b>	<b>47</b>
<b>4.3. G-quadruplex analysis of the HIV-1 proviral genome .....</b>	<b>48</b>
<b>4.4. Spectroscopic analysis .....</b>	<b>49</b>

4.4.1.	Circular Dichroism analysis .....	49
4.4.2.	UV Spectroscopy analysis .....	51
<b>4.5.</b>	<b><i>Taq</i> polymerase stop assay .....</b>	<b>52</b>
<b>4.6.</b>	<b>Footprinting assays.....</b>	<b>55</b>
<b>4.7.</b>	<b>Reporter assays.....</b>	<b>57</b>
<b>4.8.</b>	<b>Antiviral assays.....</b>	<b>60</b>
4.8.1.	Viruses.....	60
4.8.2.	Antiviral Assay in HIV-1 infected TZM-bl cell line .....	60
4.8.3.	MTT-based antiviral assay in HIV infected MT-4 cell line .....	61
4.8.4.	Antiviral assay in HIV-1 infected MT-4/LTR-eGFP cell line.....	62
4.8.4.1.	Flow Cytometry analysis.....	63
4.8.5.	Antiviral assay in HIV-1 infected PBMCs .....	63
4.8.6.	Antiviral effect of test compounds in persistently HIV-1 infected cells (MT-4/III <sub>B</sub> cell line).....	63
<b>4.9.</b>	<b>Viral binding assay .....</b>	<b>64</b>
<b>4.10.</b>	<b>Virucidal assay.....</b>	<b>64</b>
<b>4.11.</b>	<b>Time-of-Addition experiments .....</b>	<b>64</b>
<b>5.</b>	<b>RESULTS AND DISCUSSION .....</b>	<b>67</b>
<b>5.1.</b>	<b>G-Quadruplex analysis of the HIV-1 Proviral Genome .....</b>	<b>67</b>
5.1.1.	Computational analysis of the HIV-1 proviral genome for the presence of putative G-Quadruplex forming regions .....	67
<b>5.2.</b>	<b>G-Quadruplex structures in the HIV-1 Long Terminal Repeat (LTR) Promoter .....</b>	<b>69</b>
5.2.1.	Analysis of putative G-quadruplex sequences in the HIV-1 LTR promoter region.....	69
5.2.2.	Characterization of putative G-quadruplex forming sequences LTR I, II and III .....	72
5.2.2.1.	Spectroscopic analysis.....	72
5.2.2.2.	Footprinting studies .....	74
5.2.2.3.	Melting studies .....	76
5.2.2.4.	<i>Taq</i> polymerase stop assays .....	79
5.2.3.	Characterization of multiple G-quadruplexes folding in the full-length LTR G-rich region.....	81
5.2.3.1.	Melting experiments .....	81
5.2.3.2.	Footprinting studies .....	82
5.2.3.3.	<i>Taq</i> Polymerase stop assays .....	84
5.2.4.	Characterization of LTR-IV G-quadruplex .....	88
5.2.5.	Biological significance of G-quadruplexes in the LTR region. ....	90
5.2.6.	The G-Quadruplex Ligand, BRACO-19, displays antiviral activity.....	92
<b>5.3.</b>	<b>G-Quadruplex structures in the HIV-1 <i>nef</i> coding region .....</b>	<b>94</b>
5.3.1.	Analysis of putative G-quadruplex sequences in the HIV-1 <i>nef</i> coding region .....	94
5.3.2.	Characterization of putative G-quadruplex forming sequences Nef8528, Nef8547 and Nef 8624.97	
5.3.2.1.	CD spectroscopic analysis.....	97
5.3.2.2.	<i>Taq</i> polymerase stop assay.....	103

5.3.3.	Biological significance of G-quadruplexes in the HIV-1 <i>nef</i> coding region .....	104
5.3.4.	The G-Quadruplex Ligand, TMPyP4, displays antiviral activity.....	105
<b>5.4.</b>	<b>Evaluation of anti HIV-1 activity of G-quadruplex ligands .....</b>	<b>107</b>
5.4.1.	Antiviral activity of BRACO-19 .....	107
5.4.1.1.	HIV-1 infected MT-4 cells .....	107
5.4.1.1.1.	Antiviral assays.....	107
5.4.1.2.	Virucidal assay .....	112
5.4.1.3.	Viral Binding Assay .....	113
5.4.1.4.	HIV-1 infected PBMCs .....	114
5.4.1.5.	HIV-1 persistently infected cells (MT-4/III <sub>B</sub> ) .....	116
5.4.1.6.	Time of Addition (TOA).....	119
5.4.2.	Antiviral screening of G-quadruplex ligands.....	120
5.4.2.1.	Evaluation of antiviral activity of H-SPP-03 .....	127
5.4.2.2.	Evaluation of antiviral activity of PDI-Pip .....	130
<b>6.</b>	<b>CONCLUSIONS .....</b>	<b>133</b>
<b>7.</b>	<b>REFERENCES.....</b>	<b>135</b>
	<b>PUBLICATIONS.....</b>	<b>151</b>
	<b>CONFERENCES.....</b>	<b>151</b>



## List of abbreviations

AIDS	acquired immunodeficiency syndrome
AZT	3'-azido-3'-deoxythymidine
CA	capsid protein
CC <sub>50</sub>	50% cytotoxic concentration
CCID <sub>50</sub>	50% cell culture infectious dose
CD	circular dichroism
CL	clerocidin
CPE	cytopathogenic effect
DMS	dimethylsulfate
DS	dextran sulfate
EC <sub>50</sub>	50% effective concentration
G	guanine
G4P	G-quadruplex forming potential
GBA	glycosidic bond angles
G-N7	N7 position of the guanine
HAART	highly active antiretroviral therapy
HIV	human immunodeficiency virus
hTERT	human Telomerase Reverse Transcriptase
hTR	human telomerase RNA
IC <sub>50</sub>	50% inhibitory concentration
IC <sub>50</sub>	50% inhibitory concentration
IN	integrase
LTR	long terminal repeats
MA	matix protein
MOI	multiplicity of infection
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NC	nucleocapsid protein
NDI	naphtalene diimide
nef	negative factor
NF-kB	nuclear factor kappa B
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
PBMCs	peripheral blood mononuclear cells
PDI	perylene diimide
PI	protease inhibitor
QGRS	quadruplex forming G-rich sequences
rev	regulator of viral expression
RT	reverse transcriptase
SI	selectivity index
Sp1	specificity protein 1
tat	trans-activator of transcription
TDS	thermal difference spectrum

T <sub>m</sub>	melting temperature
vif	virulence factor
vpr	viral protein R
vpu	viral protein U



## Riassunto

Sequenze di acidi nucleici ricche di guanine possiedono la capacità di ripiegarsi in conformazioni secondarie non canoniche dette G-quadruplex. La presenza di sequenze G-quadruplex è stata evidenziata in numerose regioni del genoma umano, ad esempio le estremità telomeriche, i promotori di oncogeni e diverse sequenze codificanti. In particolare, la formazione di strutture G-quadruplex a livello delle regioni promotoriali del genoma umano risulta correlata al silenziamento dell'espressione genica. E' interessante notare come la presenza di sequenze G-quadruplex nei promotori sia stata descritta anche per genomi di altri organismi eucarioti, tra cui lieviti, e procarioti. Queste evidenze supportano fortemente una funzione regolatoria ubiquitaria svolta dalle strutture G-quadruplex a livello promotoriale, dove agiscono per lo più come silenziatori. Inoltre, anche la presenza di strutture G-quadruplex in sequenze codificanti del genoma umano può regolare l'espressione genica, perturbando il processo di trascrizione. Il presente lavoro di tesi ha avuto come scopo principale lo studio della presenza e della formazione di strutture G-quadruplex a livello del genoma provirale di HIV-1 al fine di valutare possibili effetti antivirali mediati dalla stabilizzazione di queste strutture secondarie. A livello della regione LTR di HIV-1, dove è situato il promotore virale, abbiamo dimostrato la presenza di strutture G-quadruplex analoghe a quelle già descritte a livello di promotori eucariotici. I risultati ottenuti evidenziano per la prima volta una regolazione dell'espressione genica virale da parte di queste strutture. La stabilizzazione delle strutture G-quadruplex virali, mediata da un ligando specifico, è risultata sufficiente a reprimere l'attività promotoriale del virus. Lo stesso composto è risultato inoltre efficace nell'inibire la produzione di HIV-1 in cellule infettate. Anche a livello di sequenze codificanti proteine virali, in particolare la proteina accessoria Nef, abbiamo individuato un cluster di sequenze che possono formare strutture G-quadruplex. L'induzione e la stabilizzazione di queste strutture secondarie mediate da un ligando specifico hanno determinato una significativa riduzione dell'espressione della proteina, supportando ulteriormente la possibilità di una regolazione dell'espressione genica mediata dai G-quadruplex. Inoltre, il medesimo composto è risultato efficace nell'inibire la replicazione virale in cellule infettate con un meccanismo d'azione Nef-dipendente. Infine, è stato effettuato uno screening di ligandi G-quadruplex per valutarne il possibile effetto anti-HIV-1. Diversi composti sono risultati promettenti inibitori virali e sono stati oggetto di caratterizzazioni più approfondite per quanto riguarda il meccanismo d'azione antivirale. I

risultati presentati in questa tesi pongono le basi per un'azione anti-HIV-1 mediata da strutture G-quadruplex indotte o stabilizzate da piccole molecole e rappresentano un importante punto di partenza per lo sviluppo di composti antivirali con un meccanismo d'azione innovativo.

## 1. Abstract

G-quadruplexes are non-canonical nucleic acid structures: their presence and functional role have been established in telomeres, oncogene promoters and coding regions of the human chromosome. In particular, they act as silencers in the promoter regions of human genes and putative G-quadruplex forming sequences are also present in promoters of other mammals, yeasts, and prokaryotes. Moreover, they have been proposed to be directly involved in gene regulation at the level of transcription. We investigated G-quadruplex formation in the HIV-1 proviral genome to assess the potential for viral inhibition through G-quadruplex stabilization. Here we show that the HIV-1 LTR promoter exploits G-quadruplex-mediated transcriptional regulation with striking similarities to eukaryotic promoters and that treatment with a G-quadruplex ligand inhibits HIV-1 infectivity. In addition, we found three conserved putative G-quadruplex forming sequences uniquely clustered in the coding region for the accessory protein Nef that were efficiently stabilized or induced by G-quadruplex ligands. Upon incubation with a G-quadruplex ligand, Nef expression was reduced in a reporter gene assay and Nef-dependent enhancement of HIV-1 infectivity was significantly repressed in an antiviral assay. Finally, a comprehensive screening of G-quadruplex ligands against HIV-1 disclosed significant potential of some of them as HIV-1 inhibitors, likely with a G-quadruplex-mediated mechanism of action. These findings open up the possibility of inhibiting the HIV-1 by G-quadruplex-interacting small molecules, providing a new pathway to the development of anti-HIV-1 drugs with unprecedented mechanism of action.



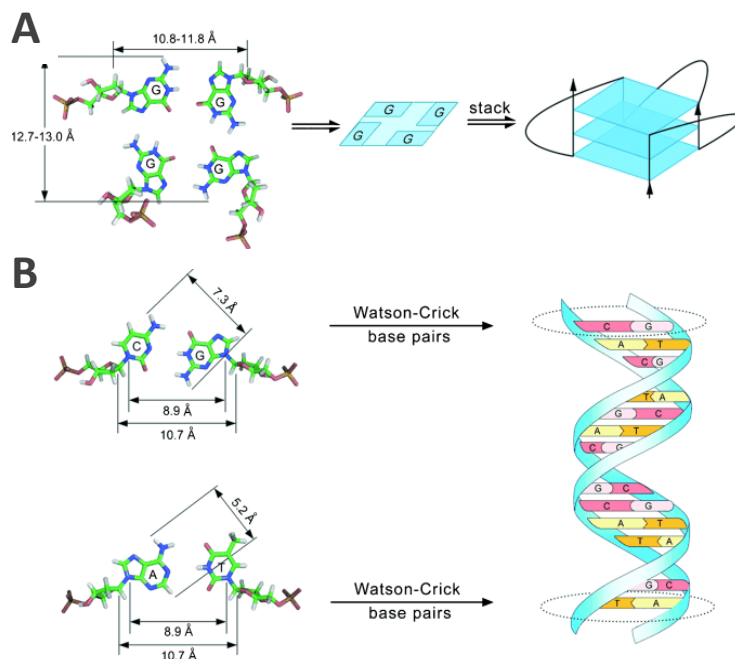
## 2. Introduction

### 2.1. DNA G-Quadruplexes

The DNA mainly exist in its typical double helix (B DNA) conformation proposed by Watson and Crick, however it can also adopt other secondary structures, such as G-quadruplexes. The first proof of the guanine (G) self-association called G-quadruplex was reported by Davies and coworker in 1962 (Gellert et al, 1962), although first evidences were known since 1910 (Bang, 1910). G-quadruplexes are secondary structures formed by single-stranded G-rich sequences. In the recent years, they have gained importance because of the increasing evidence of their potential to act as regulatory sequences in different biological processes. Indeed, many regions in eukaryotic and prokaryotic genomes display the ability to fold into a G-quadruplex conformation. Recently, cell-cycle dependent G-quadruplex formation in mammalian living cells and their stabilization by G-quadruplex ligands has been demonstrated (Biffi et al, 2013). Thus, G-quadruplex can provide a selective site for small molecules in the treatment of various disorders, for example cancer and viral infections. Also RNA molecules can fold into G-quadruplex, with many similarities with the DNA G-quadruplex (Brown & Hurley, 2011). However, because of the topic of this thesis, only DNA-G-quadruplexes will be further discussed.

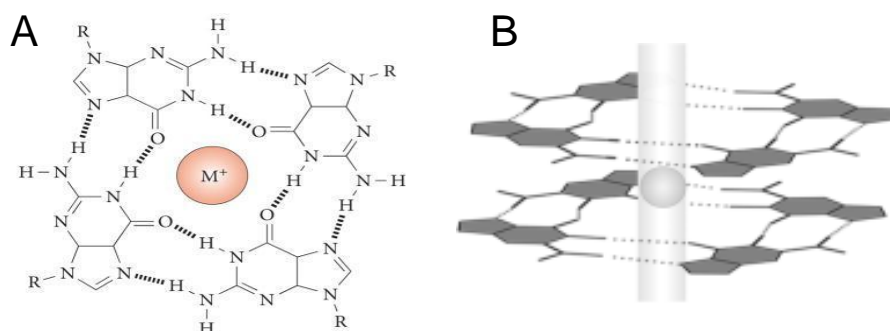
#### 2.1.1. General features

The building blocks of G-quadruplex are the G-quartets (or G-tetrads) that form by the association of 4 Gs in a cyclic Hoogsteen hydrogen bonding planar array. Two or more G-quartets can stack on top of each other to form the tetraplex secondary structure (Figure 2.1a). The sequences located between successive G-tracts serve to link stacked G-quartets and are named loops. Comparing the association of G bases in a G-quartet with Watson-Crick base pairs in the double helix, it is evident that the G-quartet's surface reaches bigger dimensions (Figure 2.1b). This peculiarity of G-quadruplex structure in possessing a big aromatic surface makes the basis for the design of selective small molecules.



**Figure 2.1** Comparison of the dimensions of A) G-quadruplex DNA structures and B) duplex DNA (Ou et al, 2008)

Monovalent cations, such as  $K^+$  and  $Na^+$ , can stabilize and even induce the G-quadruplex folding (Engelhart AE, 2009); the explanation of this behavior is in the G-quartet chemical properties. In fact, the hole between a G-tetrad has a strong negative electrostatic potential due to carbonyl O6 atoms and cations can neutralize electrostatic repulsions (Figure 2.2A). In addition, the cations channel in the hole between stacked tetrads can strongly stabilize the whole G-quadruplex structure (Figure 2.2B). It is important to mention that because of their difference in size, cations can have different position:  $Na^+$  ions are positioned mainly in the plane of the G-tetrads, whereas  $K^+$  ions are positioned between G-tetrad planes (Phan et al, 2006).



**Figure 2.2** G-tetrad structure A) G-tetrad coordinates monovalent cations thanks to the strong negative electrostatic potential due to carbonyl O6 atoms B)  $K^+$  in a G-quadruplex ion channel

### 2.1.2. Structural polymorphism of G-quadruplex structures

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

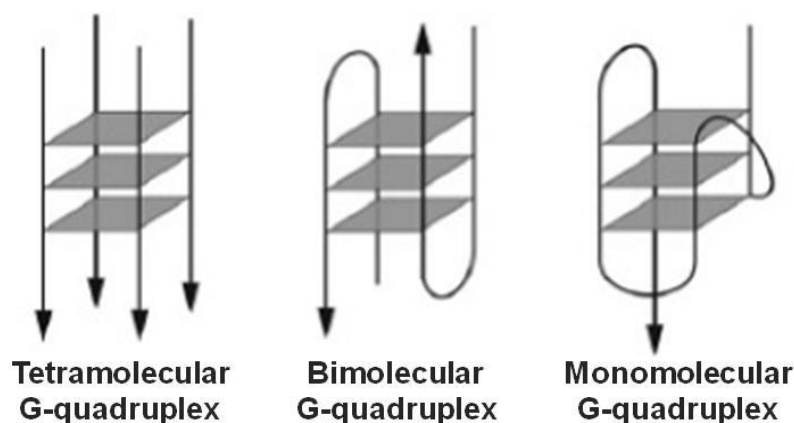
First of all, three main G-quadruplex arrangements are possible (Figure 2.3):

- Tetramolecular: G-quadruplex formed by the association of four separated strands, each with at least one G-tract
- Bimolecular: G-quadruplex formed by the association of two separated strands, each with two G-tracts
- Monomolecular (or intramolecular): G-quadruplex formed within one strand composed of four G-tract connected by loop sequences

The intramolecular G-quadruplex forming sequence is composed by at least four runs of G bases (G-tracts). This is the so called “G-4 motif” and can be represented as follow:



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



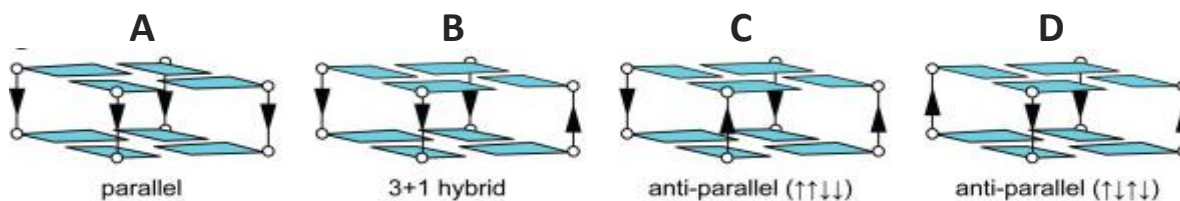
**Figure 2.3** Different G-quadruplex topologies on the basis of the number of strands involved: tetramolecular, bimolecular and monomolecular G-quadruplexes

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

- All parallel
- Three parallel and one antiparallel
- Adjacent parallel
- Alternating parallel

The relative arrangement of strand polarities gives rise to different G-quadruplex conformations:

- Parallel G-quadruplex (Figure 2.4A): all the four strands are parallel
- 3+1 Hybrid (or mixed) G-Quadruplex (Figure 2.4B): three parallel and one antiparallel strands
- Antiparallel G-quadruplex: with adjacent (Figure 2.4C) or alternating parallel strands (Figure 2.4D)

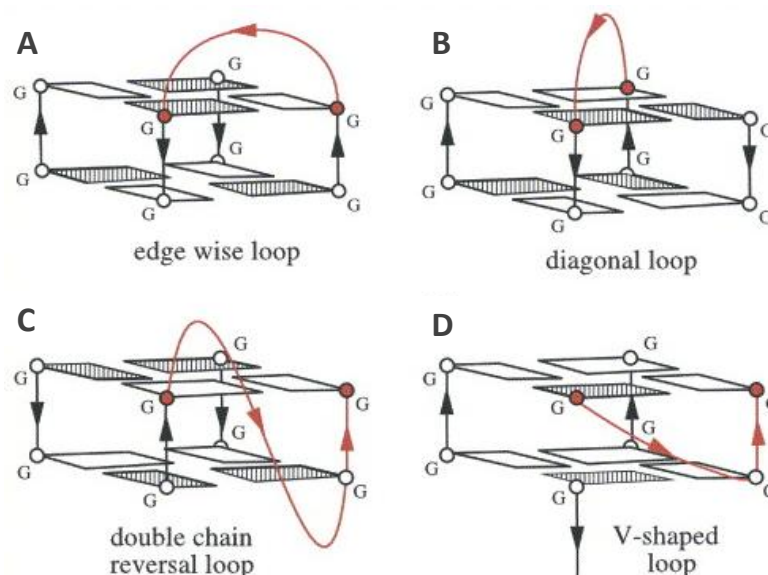


**Figure 2.4 Different G-quadruplex topologies on the basis of the relative arrangement of strand polarities**  
A) Parallel G-quadruplex B) 3+1 hybrid G-quadruplex C) Antiparallel G-quadruplex with adjacent parallel strands D) Antiparallel G-quadruplex with alternating parallel strands

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

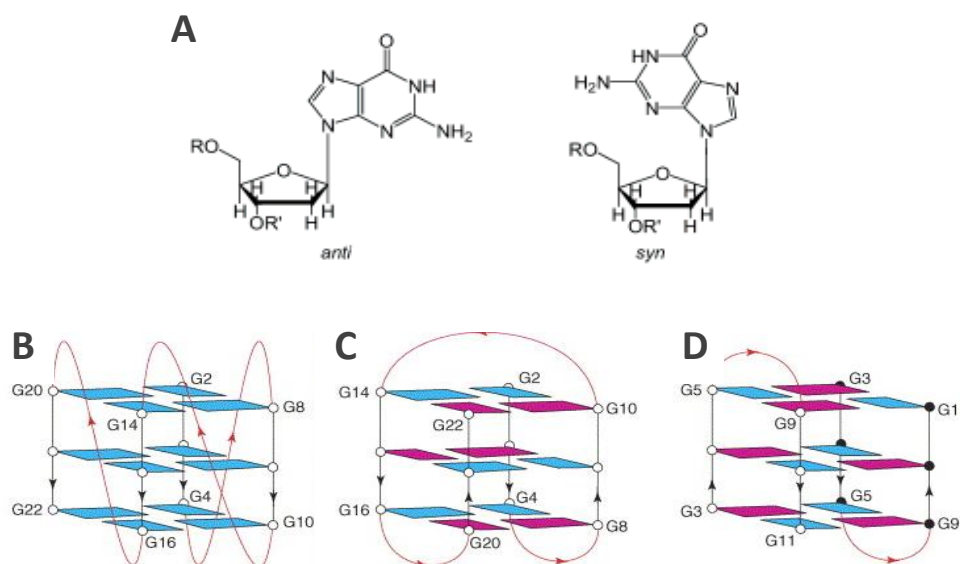
- Lateral (or edgewise): loops connecting two adjacent antiparallel strands (Figure 2.5A)
- Diagonal: loops connecting two opposing antiparallel strands (Figure 2.5B)
- Double-chain-reversal (or propeller): loops connecting adjacent parallel strands (Figure 2.5C)
- V-shaped: loops connecting two corners of a G-tetrad core in which one supporting column is lacking (Figure 2.5D).





**Figure 2.5 Possible orientation of connecting loops in a G-quadruplex structure** A) Edge wise loop (in red) B) Diagonal loop (in red) C) Double chain reversal loop (in red) D) V-shaped loop (in red)

The guanine glycosidic bond angles (GBA) are another important parameter that contributes to the G-quadruplex high polymorphism. In a G-quadruplex structure, stacked G may adopt either an *anti* or a *syn* GBA (Figure 2.6A). All parallel G-quadruplexes have G bases with *anti* GBA (Figure 2.6b), whereas antiparallel G-quadruplexes have G with both *syn* and *anti* GBA (Figure 2.6C-D).



**Figure 2.6 Glycosidic bond angles (GBA) in a G-quadruplex structure** A) *anti* or a *syn* GBA B) Parallel G-quadruplex with *anti* GBA C) antiparallel G-quadruplex with *anti-anti syn-syn* GBA D) antiparallel G-quadruplex with *anti-syn anti-syn* GBA. *anti* GBA are represented as light blue rectangles while *syn* GBA are represented as magenta rectangles

Overall, topology and stability of G-quadruplex structure is influenced by many factors and, for this reason, could be relatively complex to be predicted and characterized. However, this great polymorphism can constitute the basis for a drug-targeting with a high level of selectivity over duplex and theoretically selective for a G-quadruplex structure with unique characteristic over non-related G-quadruplex structures.

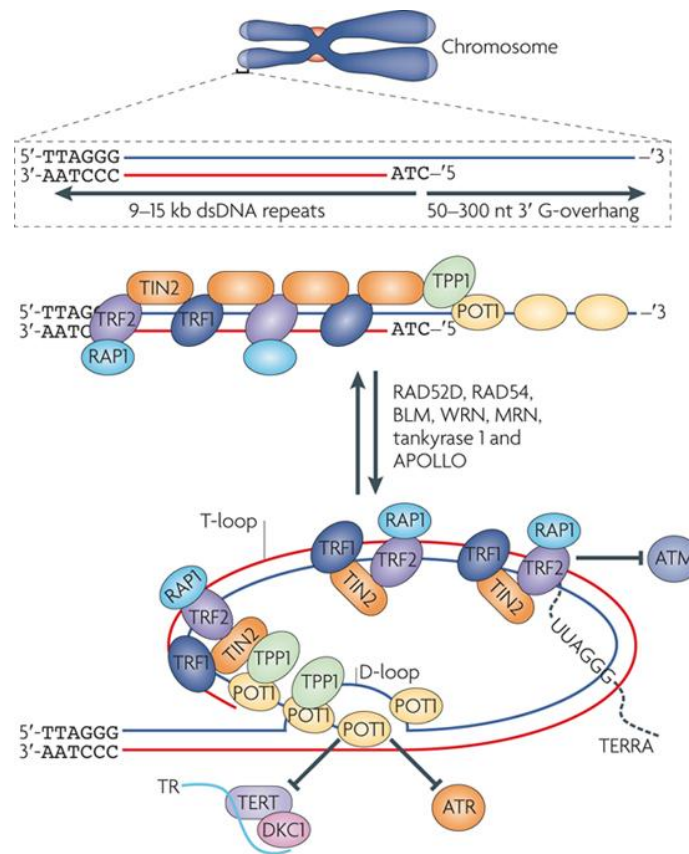
### **2.1.3. Biological roles of G-quadruplexes**

G-quadruplex potential folding sequences occur in different regions of the human genome, such as oncogene promoters, telomeres, ribosomal DNA, mini-satellites and the immunoglobulin heavy chain switch region (Ou et al, 2008). It has been estimated that more than 300,000 sites in the human genome have potential to fold in a G-quadruplex (Huppert & Balasubramanian, 2005), suggesting an important role of these structures in regulating biological processes. Computational analysis showed that “G-4 motif” are not randomly distributed and that are similarly distributed, especially in specific functional region such as promoters (Huppert & Balasubramanian, 2005). G-quadruplexes have been extensively studied *in vitro* and increasing evidences suggested a key biological role of these tetraplex structures *in vivo*.

#### **2.1.3.1. G-quadruplex at telomeric ends**

Telomeres are nucleoprotein complexes located at the ends of eukaryotic chromosomes. Telomeric DNA in human somatic cells is composed of tandem repeats of d(TTAGGG) sequence that can reach 15 kilobases in length. Moreover, telomeric DNA is characterized by a 3' single-stranded overhang of up to 300 bases, named G-overhang, that results particularly G-rich. Since the linear G-overhang can be recognized as DNA double-strand breaks and so degraded by DNA repair mechanisms, a protective structure is formed at this level thanks to proteins of the shelterin complex. The shelterin complex is composed of several binding proteins (telomeric-repeat-binding factor 1 (TRF1), TRF2, TRF1-interacting protein 2 (TIN2), the transcriptional repressor/activator protein RAP1, protection of telomeres 1 (POT1) and the POT1- and TIN2-organizing protein TPP1) that cover both the double and single stranded repeats, generating a protective alternative structure at telomeric ends named T-loop. In detail, the T-loop is generated by invasion of the single-stranded G-overhang into the double-

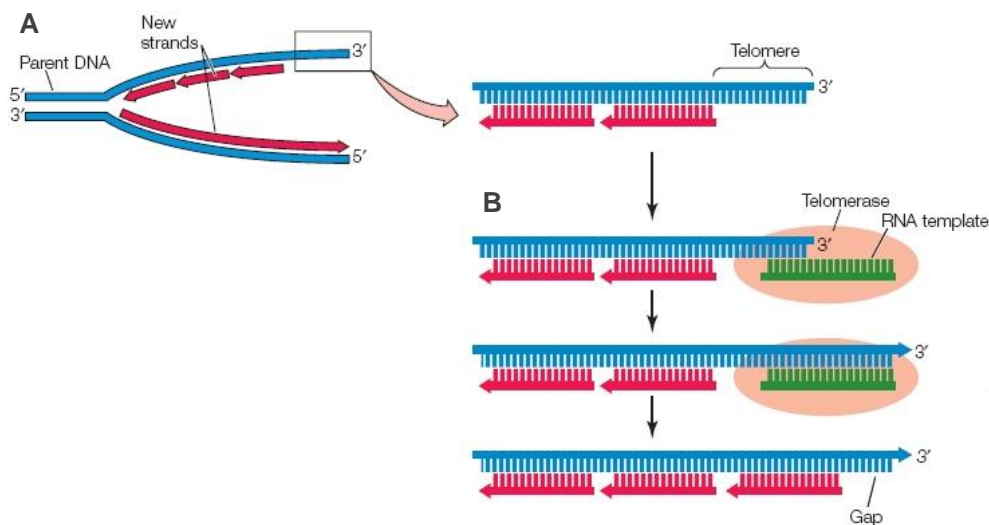
stranded TTAGGG repeats. The looped structure protects telomeres on several levels. Invasion effectively sequesters the G-overhang and allows natural chromosome ends to be distinguished from double-strand breaks (Figure 2.7).



**Figure 2.7 The structure of human telomeres** Human telomeres consist of tandem repeats of TTAGGG sequence with a G-rich strand that extends in the 3' direction, forming the G-OVERHANG. The shelterin complex consists in several proteins (the double stranded telomeric repeat binding factors TRF1 and TRF2, the TRF2 interacting factor RAP1, the bridging molecules TIN2 and TPP1 and the telomeric protection factor POT1), and forms a protective structure at chromosome ends, the T-loop (O'Sullivan & Karlseder, 2010)

Telomeres are thought to be essential for chromosome stability and in general for genomic integrity; they provide sites for recombination events and transcriptional silencing and play a critical role in cellular aging and cancer (O'Sullivan & Karlseder, 2010). This latter aspect is particularly important and refers to the so called “end replication problem”. Briefly, the lagging strand of linear chromosomes is copied in a semi-conservative manner by the replication machinery, causing a progressive shortening of telomeric ends at every

mitotic cycle (Figure 2.8A). As consequence, somatic cells can only undergo a defined number of doublings before telomeres become critically short (Hayflick limit), losing their protective properties and sending cells into a replicative senescence, followed by apoptosis (Hayflick & Moorhead, 1961). Since this mechanism limits the replicative lifespan of individual cells, it is essential not only in controlling cell cycle but also in cellular immortalization and tumorigenesis. In fact, telomere-induced senescence is an important tumor suppressor mechanism that contrasts the infinite replicative potential of cancer cells. This unlimited replicative potential is due to the Telomerase activity that in cancer cells is not repressed. Telomerase is a ribonucleoprotein mainly composed by the enzymatic part hTERT (human Telomerase Reverse Transcriptase) and by the RNA part hTR (human Telomerase RNA). Telomerase drives the synthesis of the G-rich tandem repeats at telomeric G-overhangs by using hTR as template (Figure 2.8B). Then, during the next round of DNA replication, DNA polymerase fill in the other strand. As a consequence, telomerase activity is able to resolve the “end replication problem” mentioned above.

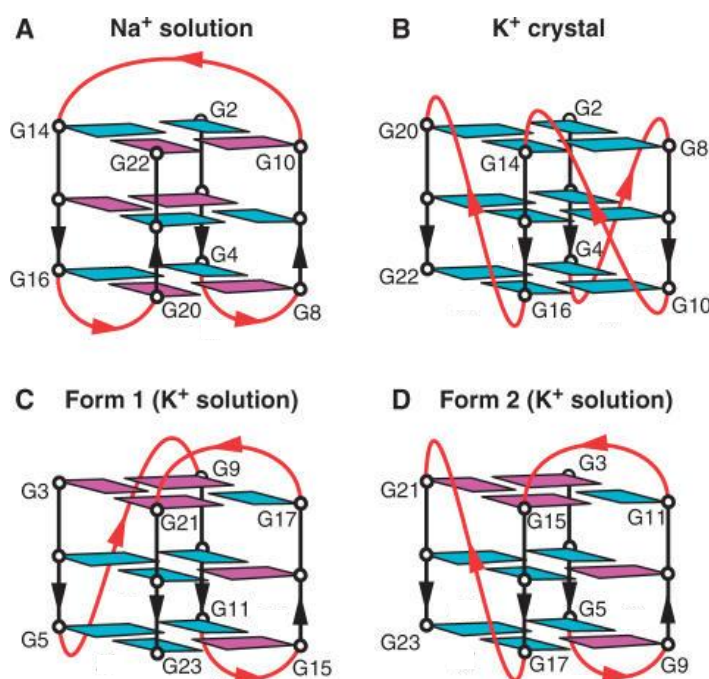


**Figure 2.8 Mechanism of DNA replication of telomeric ends** A) Semi-conservative replication of lagging strand B) Telomerase drives the synthesis of the G-rich tandem repeats at telomeric G-overhangs by using an RNA template (hTR) (O'Connor, 2008)

Telomerase attachment and consequent elongation of telomeres can be inhibited by the sheltering complex that acts as protector of the G-overhang (Figure 2.7). In somatic cells telomerase activity is absent or weak and it is not sufficient to maintain a constant telomeric length. On the contrary, embryonic and germ cells possess telomerase activity that ensures the

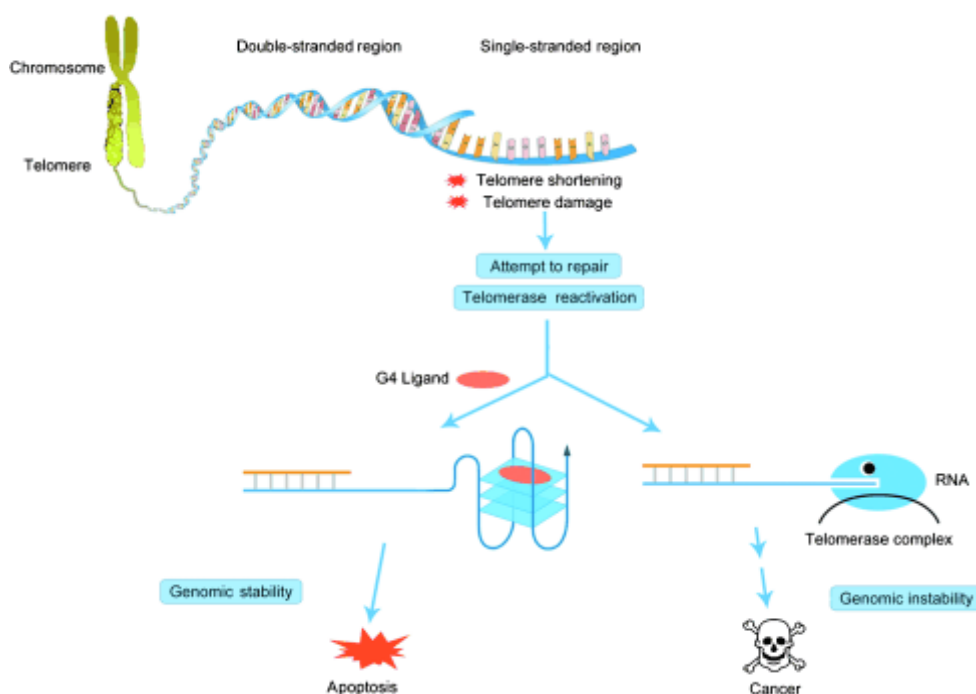
maintenance of telomeric ends in length. The exceptions include highly proliferative cells, such as those in the skin, hematopoietic tissues, and the intestinal epithelium. It has been reported that telomerase is overexpressed in about 85% of cancer cells and primary tumor (Kim et al, 1994), suggesting its importance as tumor promoter element.

The single stranded G-overhang can fold in a G-quadruplex structure. Of course the G-quadruplex folding and the T-loop structure are in a dynamic equilibrium. Several studies on architectures and folding kinetics of telomeric intramolecular G-quadruplex revealed two stable conformation both in  $K^+$  and  $Na^+$  (Ying et al, 2003). In  $Na^+$  solution, an intramolecular antiparallel G-quadruplex can form with both diagonal and lateral loop (Figure 2.9A) (Parkinson et al, 2002; Wang & Patel, 1993). In a  $K^+$  containing crystal a very different G-quadruplex structure can fold: a parallel G-quadruplex conformation with double-chain reversal loops and *anti* GBA (Figure 2.9B) (Parkinson et al, 2002). In  $K^+$  solution several others G-quadruplex structures have been reported and these are in equilibrium with each other. For example, two intramolecular mixed G-quadruplex can form (Figure 2.9C-D) (Phan et al, 2007). Moreover, two bimolecular G-quadruplexes that forms in  $K^+$  solution have been reported: one parallel G-quadruplex with diagonal loops and one antiparallel G-quadruplex with lateral loop (Phan & Patel, 2003).



**Figure 2.9 Telomeric G-quadruplexes** A)  $d[AGGG(TTAGGG)_3]$  in  $Na^+$  solution; B)  $d[AGGG(TTAGGG)_3]$  in  $K^+$  - containing crystal; C-D)  $d[AGGG(TTAGGG)_3]$  in  $K^+$  solution. Loops are red. *Anti* and *syn* G bases are light blue and magenta, respectively.

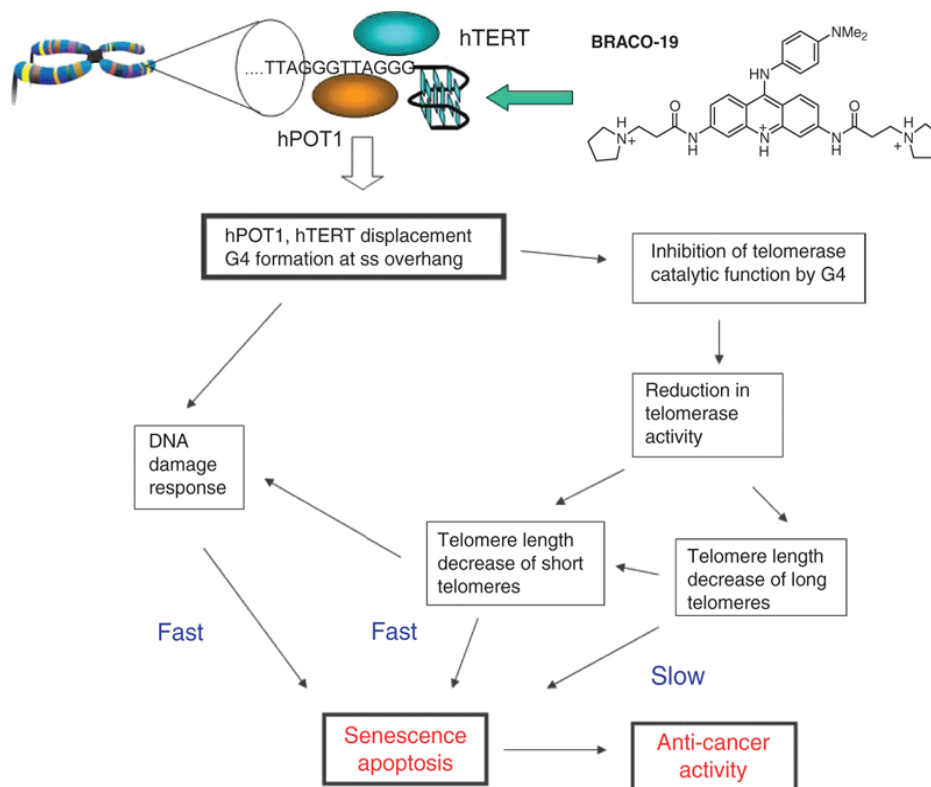
Telomeric G-quadruplexes seem to have two major functions in protecting telomeric 3' overhang from degradation by nucleases and in blocking telomerase activity. In fact, it has been reported that intramolecular antiparallel G-quadruplexes block telomerase activity *in vitro* (Zahler et al, 1991) suggesting a similar mechanism *in vivo*. This effect is probably due to the folding of the secondary structure that prevents the annealing of hTR to the G-overhang. The folding and stabilization of G-quadruplex structures at telomeric ends, especially mediated by specific ligands, can act as an alternative protection mechanism to avoid not only genomic instability and but also telomeres elongation in cancer cells (Figure 2.10).



**Figure 2.10 Biological role of telomeric G-quadruplex** Telomere shortening or damage could reactivate repair mechanisms or telomerase enzyme, triggering tumorigenesis. Induction or stabilization of telomeric G-quadruplex can thus block telomerase from binding to telomeric ends (Ou et al, 2008)

This is the main reason that made telomeric G-quadruplexes promising targets, especially in the anticancer therapy. The interaction of telomeric G-quadruplex with ligands give rise to different effect on telomere functions and it has been extensively studied in the past years. Many of these G-quadruplex ligands resulted promising *in vitro*. One of the most interesting ligands is BRACO-19. This acridine compound induces a rapid senescence in cancer cells due to the displacement of hPOT1, one of the protein of the shelterin complex. The anticancer

effect of BRACO-19 is determined by both an induced DNA damage response and a displacement of hTERT enzyme (Figure 2.11) (Neidle, 2010). Unfortunately, this molecule, as for most of the G-quadruplex ligands, has not progress into clinical trials, probably because it is not drug-like enough. This highlights the necessity of going on with the G-quadruplex research to disclose therapeutically effective ligands.



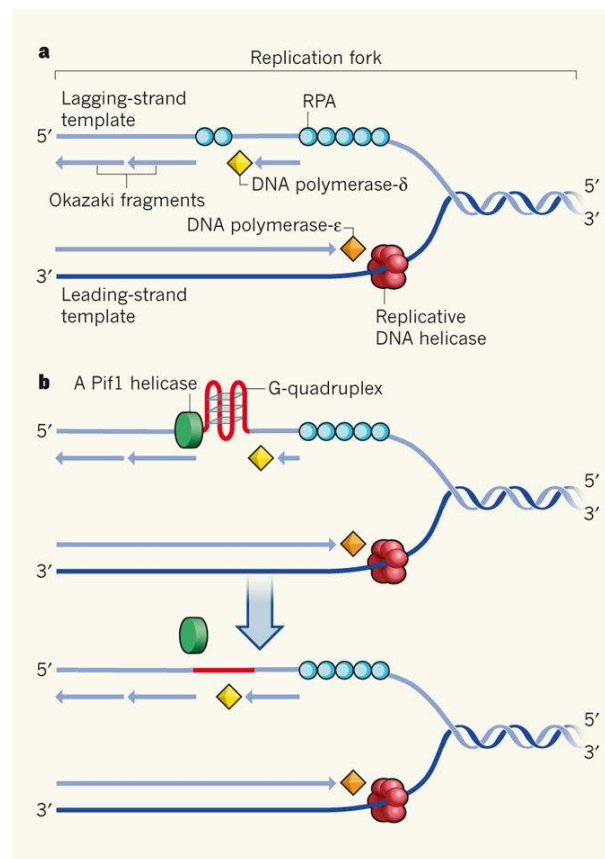
**Figure 2.11** Anticancer effect of BRACO-19 (Neidle, 2009)

### 2.1.3.2. Effects of G-quadruplex structures during DNA replication

Since the G-quadruplex structure is a dynamic structure that can form from a single-stranded sequence, the denaturation of DNA is required in a double-stranded context. The separation of the two complementary strands mainly occurs during physiological processes such as replication or transcription. As a consequence, these two processes can be mainly perturbed by a G-quadruplex folding (Maizels, 2006).

During DNA replication, replicative helicase separates the two strands, allowing the leading and the lagging strand synthesis (Figure 2.12A). During this process, the DNA is transiently single stranded and provides opportunities for G-quadruplex formation. In addition, it has

been reported that the G-quadruplex formation occurs mainly in the lagging strand template that it is replicated discontinuously, triggering fork pausing and instability (Lopes et al, 2011). Of course, G-quadruplex structures have to be solved to allow a correct DNA replication. In this context, helicases have been proposed as necessary to unwind the G-quadruplex structures. Many helicases have been tested for their ability to unwind G-quadruplexes *in vitro* and some of them resulted active. Deregulation of most of the human helicases that unfold G-quadruplexes *in vitro* is associated with diseases correlated with the genomic instability: for example the RecQ helicase WRN is associated with premature ageing (Mohaghegh et al, 2001), whereas FANCF and PIF1 are associated with an increased cancer risk (London et al, 2008; Sanders, 2010). Probably one of the best well-characterized example of G-quadruplex unwinding helicases is PIF1 that has been demonstrated to specifically bind and solve G-quadruplex structures (Figure 2.12B) (Sanders, 2010).

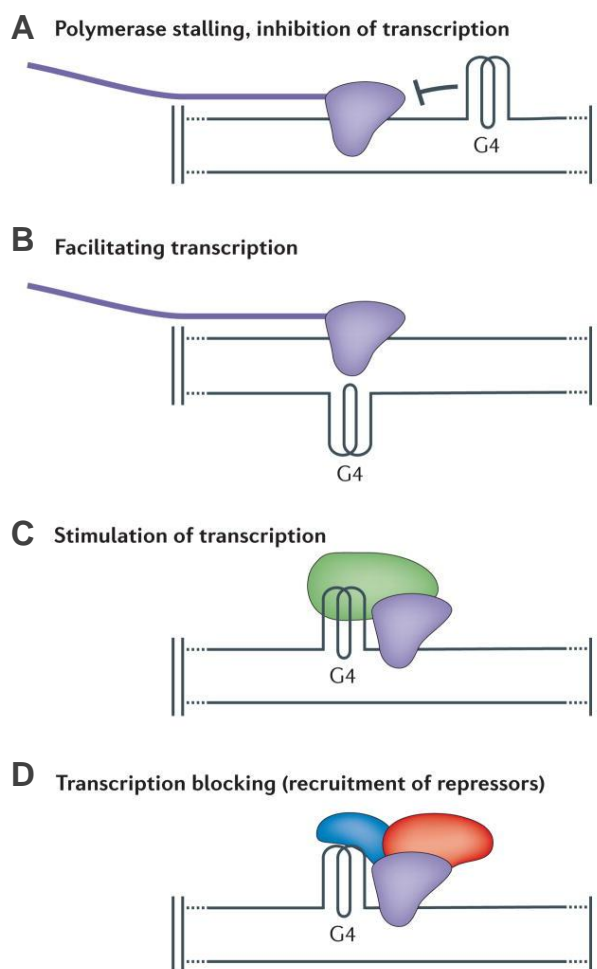


**Figure 2.12 DNA replication process** A) During DNA replication, replicative DNA helicases unwind the two strands into a leading- and a lagging-strand template to form a replication fork. The replication protein A (RPA) binds to the lagging strand and ensures smooth progress of replication along the strand. B) Where this strand contains a G-rich sequence, RPA is inefficient, allowing the strand to fold into a G-quadruplex structure including G-quadruplexes. Consequently, progress of DNA polymerase-δ is blocked, leading to stalling of the. Pif1 helicases unwind G-quadruplexes, allowing replication to progress (Mirkin, 2013)



### 2.1.3.3. Effects of G-quadruplex structures during transcription

G-quadruplex structures could influence transcription both in positive or negative ways, depending on the location of the G-quadruplex sequence. As a matter of fact G-quadruplexes can fold in regulatory (such as promoters) or coding regions. In addition they can form on the template or non-template strand. A G-quadruplex structure in a template strand of a coding region can block the transcription machinery (Figure 2.13a), whereas a G-quadruplex in the non-template strand can promote the transcription, maintaining the single-stranded conformation (Figure 2.13b). G-quadruplexes located in promoters can modulate transcription both preventing and recruiting proteins. Thus, G-quadruplexes can bind proteins (mainly transcriptional activators) that recruit or facilitate polymerase (Figure 2.13c) or can bind repressors that affect transcription (Figure 2.13d).

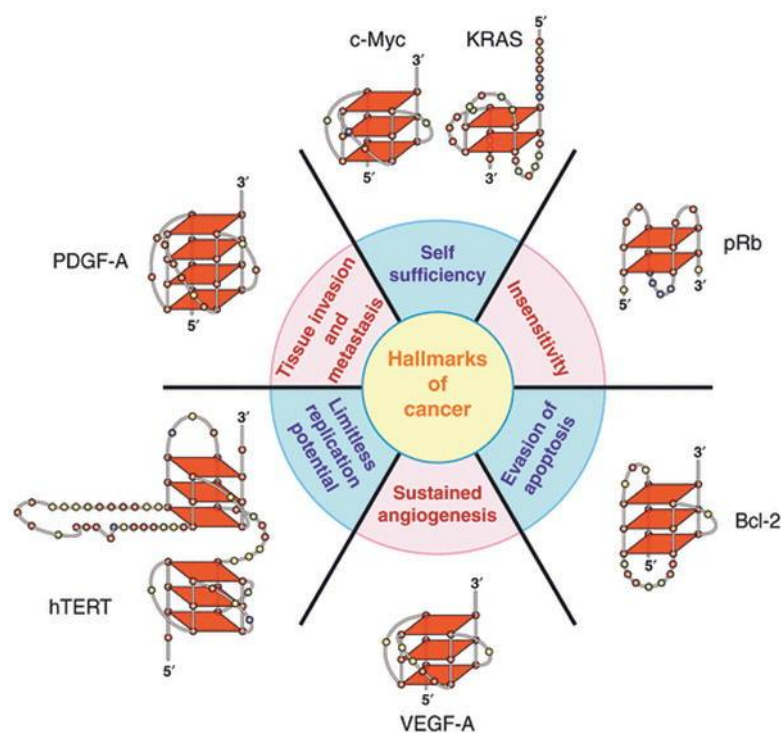


**Figure 2.13** Putative effects of G-quadruplexes during transcription process (Bochman et al, 2012)

### 2.1.3.3.1. G-quadruplexes in promoter regions

Bioinformatic analysis shows that putative G-quadruplex forming sequences are prevalent at promoters in human genome (Huppert & Balasubramanian, 2007). A similar enrichment of “G-4 motifs” in promoters is found in other organisms, including yeast (Capra et al, 2010), plants (Mullen et al, 2010) and bacteria (Rawal et al, 2006), suggesting a transversal role of G-quadruplexes among different species. Interestingly, “G-4 motifs” seem to be over-represented near binding element for transcription factors, such as Sp1 (Todd & Neidle, 2008).

In the human genome, about 50% of human genes present “G-4 motif” upstream or downstream the Transcription Start Site (TSS), suggesting a role of tetraplex structures in regulating gene expression. Further investigations about “G-4 motif” in regulatory regions revealed a correlation between gene function and the so called “G-quadruplex forming potential” (G4P) of sequences. Interestingly, tumor suppressor genes have a very low G4P, whereas proto-oncogenes show an high G4P, suggesting a specific enrolment of G-quadruplexes as transcriptional up-regulators of cancer-related genes (Eddy & Maizels, 2006). These findings opened up further investigations about G-quadruplexes in oncogene promoters and evidences for a G-quadruplex formation was found in promoters of several genes that are related to the six hallmarks of cancer (Figure 2.14) (Brooks et al, 2010). The expression of these genes results altered in cancer diseases supporting the therapeutic potential of targeting G-quadruplex for the treatment of human diseases, primarily cancer. The most representative among these genes are *c-myc* (Siddiqui-Jain et al, 2002), *VEGF* (Sun et al, 2005), *bcl2* (Dexheimer et al, 2006), *c-kit* (Rankin et al, 2005), *hTERT* (Palumbo et al, 2009) and *PDGF-A* (Qin & Hurley, 2008).

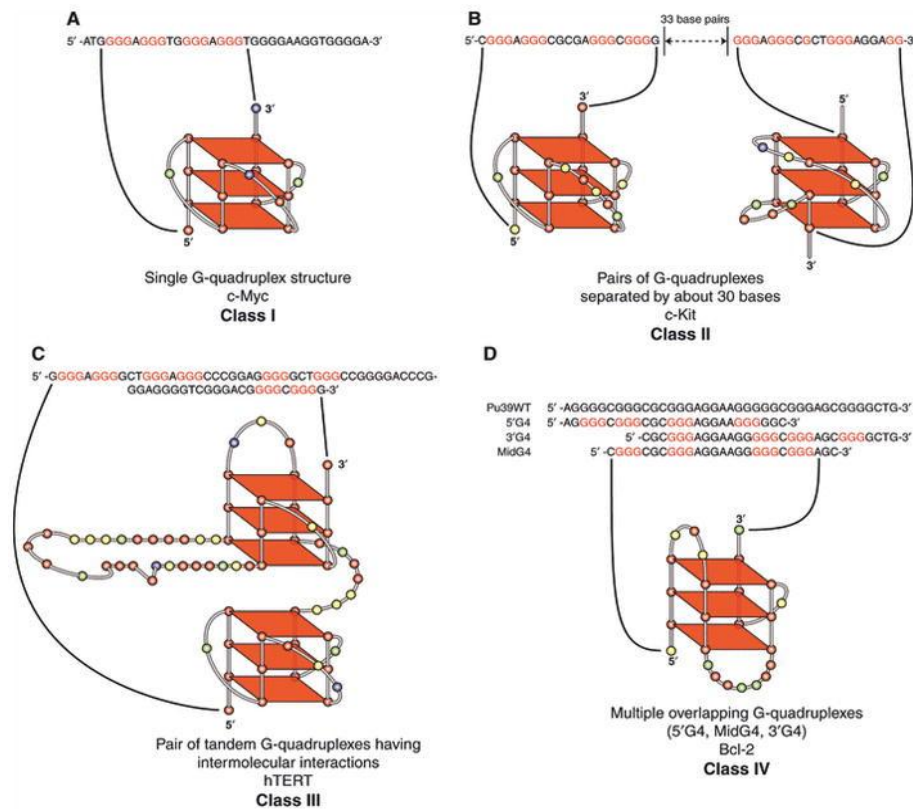


**Figure 2.14** The six hallmarks of cancer and the association with G-quadruplexes found in the promoter region of these genes (Brooks et al, 2010)

Often G-quadruplex forming sequences in promoters contain multiple G-tracts thus multiple tetraplex structures can form. This fact sometimes makes the characterization of the structure complicated. In 2010, Brooks and coworkers tried to categorize promotorial G-quadruplexes in four classes summarized as follow(Brooks et al, 2010).

- Class I (Figure 2.15A): a single G-quadruplex predominates but many loop isomers are possible. The biological consequence of formation and stabilization of G-quadruplex in these promoter elements is gene silencing. One example is G-quadruplex in *c-myc* promoter.
- Class II (Figure 2.15B): two different G-quadruplexes separated by several bases can fold. The unique example of this class is G-quadruplex in *c-kit* promoter. As for Class I, the biological consequence is inhibition of gene expression.
- Class III (Figure 2.15C): two different tandem G-quadruplexes can form. These two structures resulted more stable than the individual structures because of intermolecular interactions. Two examples are G-quadruplexes in *c-myb* and hTERT promoters. For both cases, G-quadruplexe sequences contain binding sites for the transcription factor Sp1. Again, G-quadruplex at this level inhibits gene expression

- Class IV (Figure 2.15D): Multiple overlapping G-quadruplexes species can form in a dynamic equilibrium. The greatest example of this class is *bcl-2*.



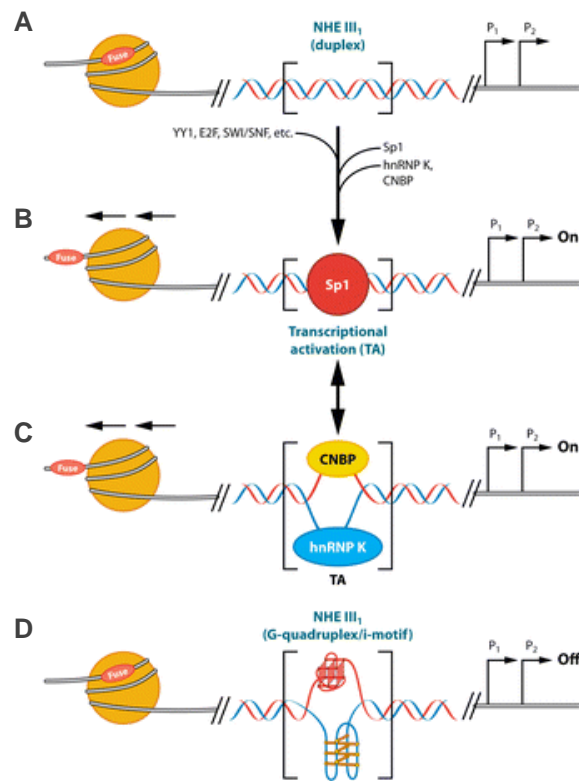
**Figure 2.15** Four classes of unimolecular G-quadruplexes found in eukaryotic promoter regions (Brooks et al, 2010)

It is important to underline that promoter regions are DNA duplex elements that would be unwound prior to G-quadruplex formation. How these tetraplex structures can fold from a double stranded context and whether they are stable once formed remain an open question. However, increasing evidence supports the thesis that G-quadruplex structures in promoter regions could regulate repression or activation of gene expression and may have an essential role in gene transcription regulation. The G-quadruplex folding could be facilitated by formation of single-stranded tract during replication and further stabilized through addition of G-quadruplex-ligands. Moreover, G-quadruplex binding protein could be involved in this dynamic process.

### 2.1.3.3.1.1. G-quadruplex in *c-myc* NHE III<sub>1</sub>

One of the best characterized example of regulatory G-quadruplex structures is located in the NHE III<sub>1</sub> region of the *c-myc* promoter. The reported biological consequence of formation and stabilization of this G-quadruplex structure is the repression of *c-myc* transcription. c-MYC is a protein that plays a central role in several aspect of cancer biology, including proliferation, differentiation, apoptosis, metastases and changing in the tumor microenvironment. The oncogenic role of c-MYC is mainly due to its overexpression in transformed cells which is often an early step in oncogenesis. Since c-MYC appears deregulated in most tumor type (about 80%) and stages, the potential of c-MYC-targeted therapies is really attractive (Gonzalez & Hurley, 2010).

The regulation of *c-myc* promoter activity is complex since it involves several promoters and start sites. One important element that control *c-myc* expression is the transcriptionally induced negative supercoiling: this effect is due to the RNA polymerase movement during transcription. Since the DNA has to be screwed through the enzyme, this creates an under-twisting behind the transcriptional machinery called negative supercoiling. Some *cis*-elements in the *c-myc* promoter, such as the nuclease hypersensitive element 1 (NHE<sub>1</sub>), are dynamically affected by supercoiling. The NHE<sub>1</sub>region interact with the transcriptional activator Sp1 in its duplex state. However, the negative superhelical stress facilitates the denaturation of this regions and favors the binding of other proteins (heterogeneous ribonucleoprotein K (hnRNP K) and CCHC-zinc finger nucleic acid binding protein (CNBP)) in a single-stranded context, again facilitating the transcription process. But the separation of the two complementary strands can allow also the formation of a G-quadruplex structure when no proteins are bound. In this case, the G-quadruplex structure prevent *c-myc* transcription acting as a silencer element (Figure 2.16) (Brooks & Hurley, 2009). This effect is particularly significant since the NHE<sub>1</sub> controls about 80% of *c-myc* transcription. It appears clear that ligands that can stabilize this structure could be used to specifically repress c-MYC expression in cancer cells overexpressing the oncogenic protein. Interestingly, GQC-05, an analogue of the antineoplastic drug Ellipticine binds the G-quadruplex structure in the NHE III<sub>1</sub> region of *c-myc in vitro* with high affinity and selectivity. Moreover, when added to Burkitt's lymphoma cell lines, GQC-05 results in reduced levels of transcribed c-MYC mRNA (Brown et al, 2011).

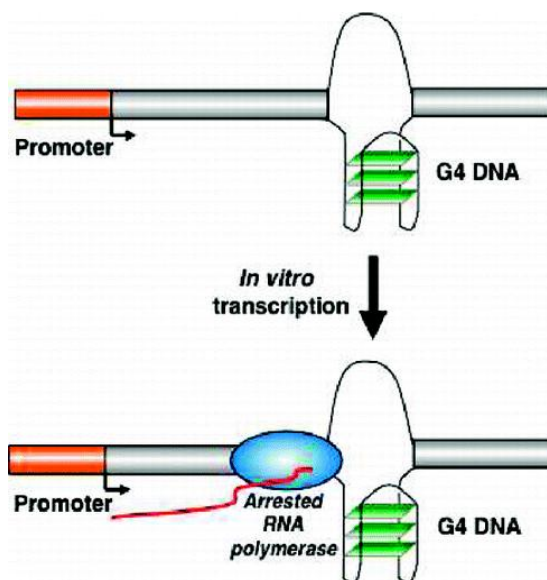


**Figure 2.16. Models of the different promoter forms within the *c-myc* NHE III1** A) Duplex representation of the promoter without any proteins bound. B) Binding of Sp1 to the duplex structure, leading to activation of *c-myc* expression. C) Binding of hnRNP K and CNBP to the single-stranded C- and G-rich regions, respectively, leading to activation of *c-myc* transcription. D) Repression of *c-myc* transcription when Sp1, hnRNP K, and CNBP are not bound, leading to the formation of the G-quadruplex and i-motif (Brooks & Hurley, 2009)

### 2.1.3.3.2. G-quadruplexes in coding regions

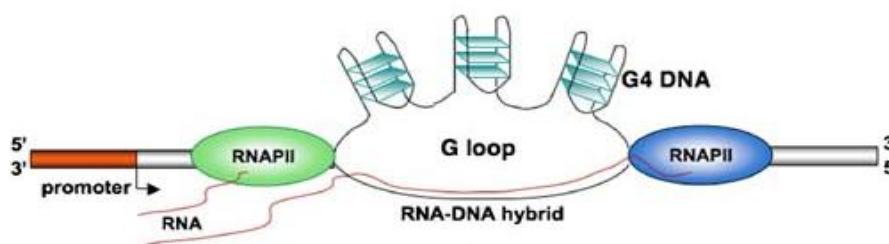
Besides promoter regions, G-rich sequences capable of forming G-quadruplex were also found within other regions of the human genome, including minisatellites (Weitzmann et al, 1997), immunoglobulin heavy chain switch regions (Dunnick et al, 1993) and rDNA (Hanakahi et al, 1999), and were shown to be the target of binding proteins (Law et al, 2010). “G-4 motifs” are less often found in the template strand than in the non-template strand. Normally, those that are located on the template strand tend to cluster at the 5’ and of 5’UTR (Huppert et al, 2008). However, recent work has shown that G-quadruplex can also form within coding regions both on the leading and on the lagging strand. On the leading strand G-quadruplexes arise during replication and promote genetic instability in human and yeast (Lopes et al, 2011; Nambiar et al, 2011). Besides these evidences, very little is known up to now on how G-quadruplex structures are processed when encountered by an elongating RNA polymerase.

Broxson and coworker in 2011 demonstrated that a G-quadruplex forming sequence located in c-Myb proto-oncogene can fold *in vitro* and can cause a T7 RNA Polymerase block during transcription (Figure 2.17). A block of transcription in the dsDNA when the G-quadruplex sequence was located in the template strand was demonstrated, suggesting that c-Myb expression is regulated by G-quadruplex formation *in vivo* (Broxson et al, 2011).



**Figure 2.17** Transcription arrest by a G-quadruplex in a template strand (Broxson et al, 2011)

On the lagging strand G-quadruplex-forming sequences were found to generate G loops (Figure 2.18) during transcription both *in vitro* and in *Escherichia coli* (Belotserkovskii et al, 2010; Duquette et al, 2004). Such structures may help to keep the transcribed template accessible for transcription by preventing its annealing to its complementary strand. In this case, G-quadruplexes could even promote transcription of certain genes.



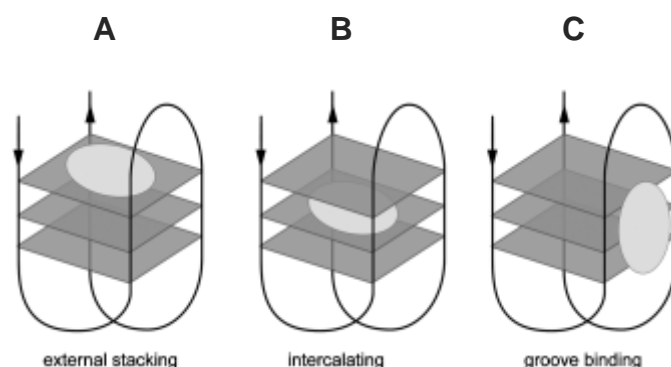
**Figure 2.18** Formation of a G-loop due to G-quadruplex structures in the non-template strand (Tornaletti, 2009)

Further, association between G-quadruplexes and single nucleotide polymorphisms and expression of the corresponding gene in individuals has been proposed (Baral et al, 2012).

## 2.1.4. G-quadruplexes as targets for drug design

### 2.1.4.1. Interaction modes of G-quadruplex ligands

In the recent years, G-quadruplexes have emerged as promising target for the anticancer therapy, primarily to block telomerase activity and to inhibit G-quadruplex related oncogene expression. To date, a diverse array of G-quadruplex stabilizing compounds have been identified. General features of these G-quadruplex recognizing ligands include a large flat aromatic surface and cationic charges. The polymorphism of G-quadruplexes is thought to allow a recognition by G-quadruplex ligands through different binding modes. In fact, these small molecules can both interact with loops and G-quartets of the tetraplex structure. The G-quadruplex stabilization occurs mainly *via*  $\pi$ - $\pi$  stacking and electrostatic interactions resulting in the binding of the ligand on the external G-quartet of the structure. This binding mode is called “external stacking” (Figure 2.19A) and is typical of flat aromatic molecules. A specific G-quadruplex ligand should possess an aromatic surface larger than that of a duplex binder to improve the aromatic–aromatic overlap and provide selectivity. Intercalation of a small molecule between G-tetrads (Figure 2.19B) is theoretically possible but it is thought to be difficult. In fact G-quadruplex structure is a rigid structure, thus the distortion induced by a ligand is energetically unfavorable. However these interactions are much less understood than those occurring with duplex DNA and are under investigation. Finally, small molecules can interact with the tetraplex structure through the groove binding mode (Figure 2.19C), specifically interacting with loops. Electrostatic interactions between positively charged ligands and the G-quadruplex-DNA scaffold also strongly promote stabilization (Ou et al, 2008).



**Figure 2.19 Mode of interaction of G-quadruplex ligands** A) external stacking mode on the surface of the terminal G-quartet B) intercalating mode between stacked G-tetrads C) groove binding mode



To date, numerous ligands have been synthesized but interactions mode resulted really complicated to be elucidated *in vitro*. Examples include perylenes, such as PIPER (Fedoroff et al, 1998), porphyrins, such as TMPyP4 (Izbicka et al, 1999), trisubstituted acridines, such as BRACO-19 (Campbell et al, 2008) and natural macrocycles, such as Telomestatin (Kim et al, 2002). Some of these compounds have shown encouraging anticancer activity *in vitro*, *in vivo* and in clinical trials (Drygin et al, 2009; Ou et al, 2008).

#### 2.1.4.2. Main classes of G-quadruplex ligands

In general, G-quadruplex ligands can be classified into four categories on the basis of their cationic nature:

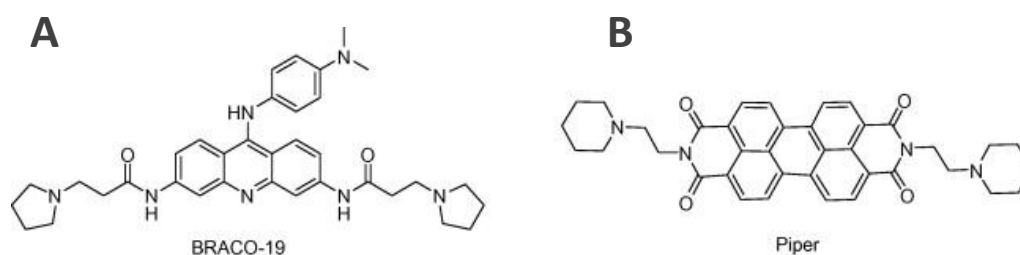
- Cationic compounds upon *in situ* protonation of an amine appendage
- Cationic compounds *via* *N*-methylation of an aza-aromatic moiety
- Cationic compounds due to the presence of a metal centre
- Non-cationic ligands

##### 2.1.4.2.1. *In situ* protonated G-quadruplex ligands

Besides flat aromatic surfaces prone to  $\pi$ -stacking with G-tetrad, G-quadruplex ligands should retain sufficient water solubility. To ensure this, normally protonable sidearms (e.g. amine groups) are introduced around the aromatic core, thus the charges are far from the hydrophobic center.

BRACO-19 is probably one of the best examples of this class of compounds. BRACO-19 is a 3,6,9 trisubstituted acridine derivative (9-[4-(*N,N*-dimethylamino)phenylamino]-3,6-bis (3-pyrrolo-dino-propionamido) acridine  $\times$  3HCl) (Figure 2.20A) that interact with three G-quadruplex grooves thanks to three side-arms (Schultes et al, 2004). The acridine motif seems very efficient in recognizing G-quadruplex structures, with a very high selectivity (31-fold) over duplex DNA (White et al, 2007). BRACO-19 has been intensively investigated for its ability of stabilizing telomeric G-quadruplexes and resulted a strong inhibitor of telomerase firstly by TRAP (telomeric repeat amplification protocol) assay (Kim et al, 1994). Further biological investigations show that BRACO-19 is efficient in inhibiting cancer cell proliferation *in vitro* and in arresting tumor growth *in vivo*, targeting specifically the telomeric G-quadruplex (Burger et al, 2005).

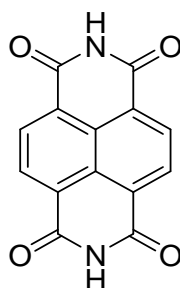
Another group of compounds that belong to the *in situ* protonated G-quadruplex ligands is the perylene derivatives, such as PIPER. This compound is a perylene diimide (PDI) characterized by a broader hydrophobic core with two external amine appendages (Figure 2.20B). NMR studies indicate that PIPER can bind G-quadruplex structures with different topologies with stoichiometries of 1:2, 1:1 and 2:1 (Ou et al, 2008). As for BRACO-19, PIPER resulted a good telomerase inhibitor with an  $IC_{50}$  (50% inhibitory concentration) in the low- $\mu$ M range. Interestingly, it has been reported that PIPER can induce the transition duplex-quadruplex in the *c-myc* promoter region (Rangan et al, 2001).



**Figure 2.20** Chemical structure of A)BRACO-19 and of B) PIPER

The substituted naphthalene diimides represent another interesting class of G-quadruplex ligands. The naphthalene diimide core is reported in Figure 2.21. Among the 1,4,5,8-naphthalene tetracarboxylic diimides (NDIs) analogues, tri- and tetra-substituted NDIs show promising G-quadruplex binding properties (Cuenca et al, 2008). Interestingly, a tetra-substituted NDI is able to strongly bind the human telomeric G-quadruplex DNA and to selectively inhibit the growth of several cancer cell lines at sub- $\mu$ M concentrations (Hampel et al, 2010). The great advantage of NDIs is that the synthetic route of NDIs allows the introduction to the ND core of up to four different side chains, which in principle can be exploited to produce ligand diversity that may discriminate between different types of G-quadruplexes. Tri-substituted NDIs chemically engineered to embed an alkylating quinonemethide precursor (QMP) resulted extremely efficient alkylating agents of telomeric G-quadruplex DNA, showing a promising anticancer activity *in vitro* (Nadai et al, 2011). Moreover, also NDIs with an extended core (e.g. core fused to 1,4-dihydropyrazine-2,3-dione), resulted good G-quadruplex ligands with a promising anticancer activity against different human telomerase-positive cell lines (Doria et al, 2012). NDI core tethered to activatable alkylating moieties by flexible spacers showed a selective alkylation and stabilization of G-quadruplexes, providing the first example of thermally induced non-metal-

based G-quadruplex adduct formation (Di Antonio et al, 2009). Moreover, NDI–oxirane conjugates showed a significant affinity for G-quartets and a selective alkylation of the loop adenines (Doria et al, 2013).



**Figure 2.21** Chemical structure of the naphthalene diimide (NDI) core

Other compounds in this group are the quinoanthroxazine derivatives, such as the fluoroquinolone QQ58 that stacks onto an external G-tetrad as the main binding mode (Duan et al, 2001).

#### **2.1.4.2.2. N-methylated aromatic G-quadruplex ligands**

N-methylated ligands, for example compounds quaternized on the aromatic ring nitrogens, show several advantages such as an increasing water solubility and  $\pi$ -stacking ability.

Porphyrins with their planar arrangement of the aromatic rings have proved to bind G-quadruplexes by stacking with the G-tetrads. One of the most important exponent of this category is the tetracationic porphyrin TMPyP4 (5,10,15,20-tetra-(N-methyl-4-pyridyl)porphyrin) (Figure 2.22A). TMPyP4 have shown to bind and stabilize both parallel and antiparallel G-quadruplex structures and to efficiently inhibit telomerase (Shi et al, 2001). Moreover this compound resulted efficient in downregulating the expression of several oncogenes, such as *c-myc* (Siddiqui-Jain et al, 2002). TMPyP4 showed diverse G-quadruplex binding modes, including intercalation between adjacent G-tetrads and stacking onto the external G-quartet. The TMPyP4 isomer, named TMPyP2 (tetra-(N-methyl-2-pyridyl)porphyrin) (Figure 2.22B) has two N-methyl groups in the sterically hindered 2-position resulted unable to stabilize G-quadruplex structures and showed a weak activity in inhibiting telomerase activity (Rha et al, 2000), supporting the specificity of TMPyP4. For this reason, TMPyP2 is considered a kind of “negative control” in characterizing G-quadruplexes and is widely used in parallel with TMPyP4. However, cationic porphyrins have

only 2-fold greater affinity for quadruplex over duplex, thus they are used only for the preliminary biophysical characterization of G-quadruplex structures.

A small molecule has to be mentioned in this category: RHPS4 (3,11-difluoro-6,8,13-trimethyl(H)-quino[4,3,2-kl] acridinium methylsulfate) is a *N*-methylated pentacyclic acridinium (Figure 2.22C) that showed a potent anti-telomerase activity through a selective G-quadruplex binding. *In vitro* and *in cellulo* studies demonstrated the ability of this ligand to decrease telomeres length and to act in synergy with the anti-cancer agent Taxol (Gowan et al, 2001). RHPS4 is one of the rare ligands whose complex with G-quadruplex-DNA has been solved by NMR (Gavathiotis et al, 2001). As expected, the cationic molecule sandwiches the quadruplex-structure thanks to strong stacking interactions between the ligand and the two external G-quartets of the G-quadruplex.

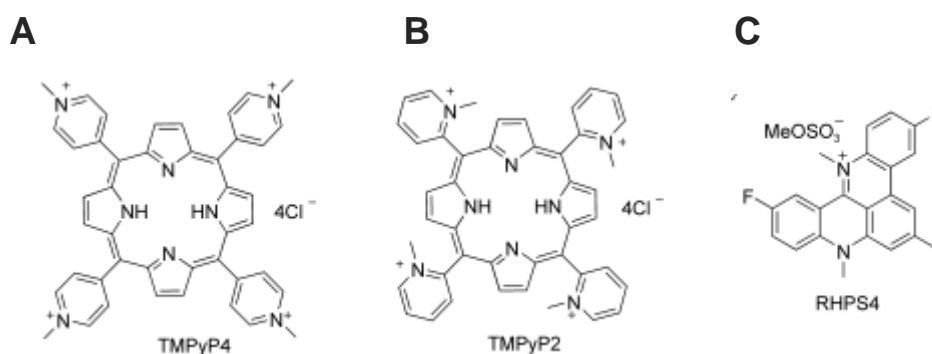


Figure 2.22 Chemical structures of A) TMPyP4 B) TMPyP2 and C) RHPS4

#### 2.1.4.2.3. Metallo-organic G-quadruplex ligands

This class of ligands is interesting thanks to very promising G-quadruplex binding properties. The central metal core of these ligands could be positioned over the cation channel of the G-quadruplex, optimizing the stacking interactions of the surrounding chelating agent with the G-quartet. Moreover, their cationic or highly polarized nature makes the association with the negatively charged G-quadruplex-DNA more favorable.

The most representative examples are Cu(II)-TMPyP4 (Figure 2.23A) and Mn(III)-TMPyP4. These compounds are characterized by the insertion of a metal in the central cavity of TMPyP4, forming a metallo-complexes. In particular, Mn(III)-TMPyP4 showed a 10-fold increased selectivity for quadruplex over duplex (Dixon et al, 2005). The pentacationic manganese(III) porphyrin (Figure 2.23B) is another example of metallo-organic G-quadruplex ligands and contains a central aromatic core with four flexible cationic arms. Interestingly,

this compound is one of the most potent G-quadruplex ligands, showing a 10000-fold quadruplex versus duplex selectivity (Dixon et al, 2005).

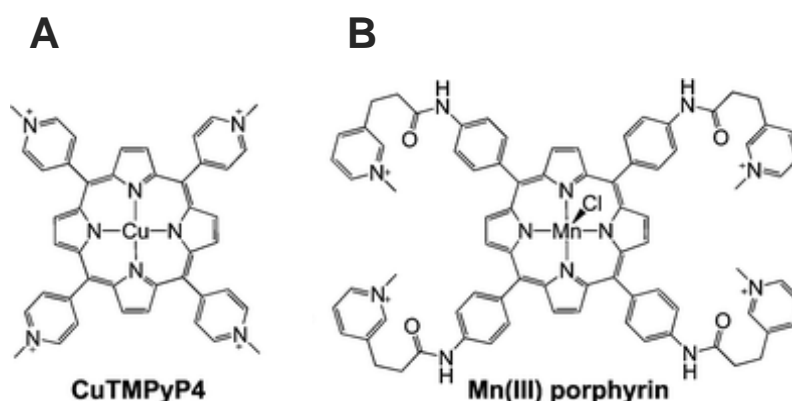


Figure 2.23 Chemical structures of A) CuTMPyP4 and B) Mn(III)porphyrin

#### 2.1.4.2.4. Neutral macrocyclic G-quadruplex ligands

This category includes one of the most interesting G-quadruplex ligands, Telomestatin. This compound is a macrocyclic natural molecule isolated from *Streptomyces annulatus* and consists in seven oxazole rings and one thiazoline ring (Figure 2.24). Telomestatin is actually one of the most efficient telomerase inhibitor *in vitro*, with an  $IC_{50}$  in the nano molar range, and appears as one of the most selective G-quadruplex ligands (70-fold quadruplex over duplex selectivity) (Shin-ya et al, 2001). This efficiency is probably due to a perfect shape adaptation between its structure and the G-quartet. However, Telomestatin has a big disadvantage related to its synthesis that has been recently reported (Doi et al, 2006). In fact, the synthesis is really complex and thus hardly compatible with large-scale preparation.

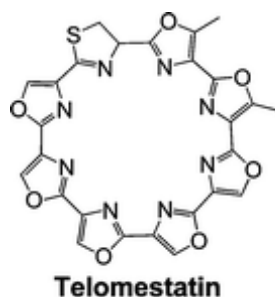


Figure 2.24 Chemical structure of Telomestatin

### **2.2. G-quadruplex in microorganisms**

Besides humans, other mammals (Verma et al, 2008), yeasts (Hershman et al, 2008) and prokaryotic organisms (Beaume et al, 2013; Rawal et al, 2006; Wieland & Hartig, 2009) exhibit putative G-quadruplex forming sequences in key regions. Computational analysis on several organisms showed that, as for the human genome, “G-4 motifs” appear similarly distributed and are over-represented in promoter regions (Huppert & Balasubramanian, 2005). The fact that also prokaryotic genomes possess the potential to adopt a G-quadruplex conformation made G-quadruplexes a new and interesting topic for the scientific community. In fact, G-quadruplex can provide a selective site for small molecules in the treatment of various disorders, for example bacterial or viral infections. To date, very little is known about G-quadruplexes at the viral level. The Epstein Barr virus (EBV) encodes for the EBV nuclear antigen 1 (EBNA1) protein that is critical for replication and maintenance of the genome during latency in proliferating cells. It has been demonstrated that EBNA1 specifically binds to a G-rich RNA sequence that is predicted to form G-quadruplex structures. Interestingly, the G-quadruplex ligand BRACO-19 inhibited EBNA1-dependent stimulation of viral DNA replication. BRACO-19 treatment also disrupted the ability of EBNA1 to tether to metaphase chromosomes, suggesting that maintenance function is also mediated through G-quadruplex recognition (Norseen et al, 2009). However, no more evidences have been reported about G-quadruplex effects on EBV biology. The Human Papillomavirus (HPV) presents highly conserved regions with an high propensity in G-quadruplex folding that could be implicated in regulation of viral processes. These sequences can fold in relatively strong G-quadruplex structures. However, these reported results do not show possible biological consequences of G-quadruplexes for the HPV nor possible implications for medical treatment of viral infection (Tluckova et al, 2013).

Overall, the evidences about G-quadruplex formation and biological roles in viruses remain elusive, suggesting a need to further investigate this promising area of research. Since this thesis focuses on G-quadruplex structures in the Human Immunodeficiency Virus (HIV), the main features of this virus and reported evidences for G-quadruplex implications in its viral biology will be presented in the following paragraphs.

### 2.3. The Human Immunodeficiency Virus (HIV)

The human immunodeficiency virus (HIV) was first characterized in 1983 when it was proposed as the causal agent of the acquired immunodeficiency syndrome (AIDS) (Barre-Sinoussi et al, 1983; Gallo et al, 1983). AIDS was first reported in 1981 by the US (United States) Center for Disease Control and Prevention (CDC). In 1985, the US Food and Drug Administration approved a commercial test to detect the virus. As a result in that year over 17000 cases were reported from 71 countries, clearly showing the AIDS pandemic nature (Merson et al, 2008). The progression of the disease results in a compromised immune system of an individual leading to increased susceptibility to opportunistic infections caused by bacteria, fungi, viruses and parasites. This is mainly due to the depletion of CD4+ T-helper lymphocyte cells that are a key component of the human immune system. Nowadays, AIDS remains one of the most significant infectious diseases and AIDS-related illnesses are one of the leading causes of death and premature mortality worldwide. High mortality is without doubt the most serious outcome of AIDS, but it impairs also other aspects of social life (socio-economic development, poverty) especially in undeveloped countries (e.g. Africa). The Joint United Nations Programme on HIV/AIDS (UNAIDS) estimates that there were 33.3 million people living with HIV at the end of 2009, underlying an increase of about 27% in ten years. Africa remains the global epicentre of AIDS pandemic with the highest number of people living with HIV and AIDS (UNAIDS/WHO, 2012). To date, important progress has been achieved in preventing new HIV infections, which have been steadily declining since the late 1990s, this decrease is offset by the reduction in AIDS-related deaths due to the significant scale up of antiretroviral therapy over the past few years.

HIV is a member of the *Retroviridae* family in the Lentivirus genus which comprises two types of HIV, HIV-1 and HIV-2. HIV-1 is the worldwide predominant and is the main responsible for the AIDS pandemic. HIV-1 is characterized by a high and complex diversity, mainly due to the error-prone reverse transcriptase enzyme, the high virus replication rate and the frequent recombination events. Because of this high diversity in the HIV-1 genome, it has been phylogenetically divided into four groups: group M (major), group O (outlier), group N (new/non M or non O) and group P. Groups N and P are relatively new groups and together with group O are extremely rare. In fact, more than 90% of HIV-1 infections belong to HIV-1 group M. Group M viruses have been classified into 9 main subtypes (A,B,C,D,F,G,H,I and K) and several circulating recombinant forms (CRFs) (Hemelaar, 2012). Although subtype B

is responsible for 11% of all infections worldwide, it represent the predominant subtype in Europe (Hemelaar et al, 2011).

### 2.3.1. Viral structure and genome

HIV-1 virion has a spherical shape (110 nm diameter) and consists of an envelope that surrounds a conical capsid which contains two molecules of the viral single stranded RNA genome. The capsid core is composed of approximately 1500 molecules of the viral p24 capsid protein (CA) and encloses two copies of positive-sense single-stranded RNA genome that is tightly bound to the nucleocapsid protein (NC). Moreover, the capsid contains the viral enzymes integrase and reverse transcriptase and the four accessory proteins Vif, Vpr, Vpu and Nef. The outer membrane of the virus, called envelope, is composed by a host-cell derived lipid membrane and is coated in the interior by the matrix proteins (MA). On the envelope two glycoproteins are anchored: the trimer gp120 surface protein and the gp141 transmembrane protein complex (Figure 2.25).

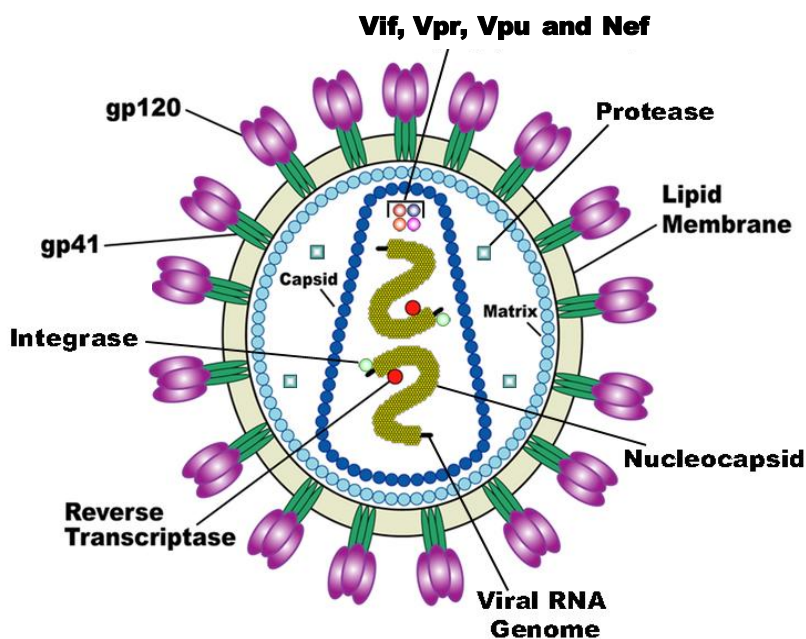
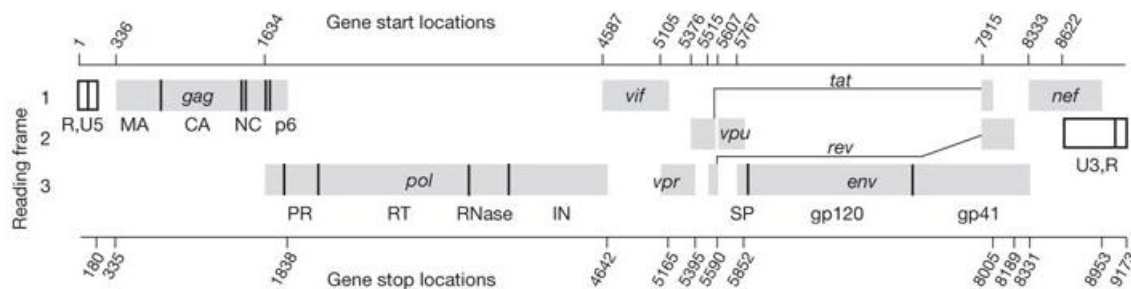


Figure 2.25: Schematic representation of HIV-1 virion

The 9.18 kilobases HIV-1 RNA genome (Figure 2.26) consists in three primary genes (*gag*, *pol* and *env*) and two regulatory genes (*tat* and *rev*) and four accessory genes (*vif*, *vpr*, *vpu*, *nef*). *Gag*, *Pol* and *Env* are the prototypical retroviral proteins: *gag* encodes for capsid proteins (MA, CA, NC and p6), *pol* encodes for the three viral enzymes protease (PR), reverse



transcriptase (RT) and integrase (IN) and *env* encodes for the two viral glycoprotein gp120 and gp141. Gag, Pol and Env result essential for viral replication. Besides these three proteins, also the two regulatory proteins Tat and Rev are essential. On the contrary, the four so-called accessory proteins Nef, Vif, Vpr and Vpu are not essential for viral replication *in vitro*, but can have consequences on viral life cycle, altering replication or disease progression.



**Figure 2.26 HIV-1 RNA genome organization** Protein coding region are shown as grey boxes; polyprotein-domain junctions are depicted as solid vertical lines. CA, capsid; IN, integrase; MA, matrix; NC, nucleocapsid; PR, protease; RT reverse transcriptase; SP, signal peptide (Watts et al, 2009)

The main features of the HIV genes and proteins will follow, with a particular focus on the Nef protein that is one of the topics of this thesis.

- **Gag**

The *gag* gene encodes for the 55-kilodalton Gag precursor protein (p55) which is expressed from the unspliced viral mRNA. During translation, the p55 is associated with host cell membranes, recruits two copies of the viral genomic RNA along with other viral and cellular proteins that triggers the budding of the viral particle from the surface of an infected cell. After budding, p55 is cleaved by the virally encoded protease (PR) (a product of the *pol* gene) during the process of viral maturation into four smaller proteins designated MA (matrix or p17), CA (capsid or p24), NC (nucleocapsid or p9), and p6. The MA molecules are attached to the inner surface of the virion, stabilizing it. Moreover these molecules facilitate the nuclear transport of viral genome (Gallay et al, 1995). The p24 protein forms the conical core of viral particles. It has been reported that the cellular peptidylprolyl isomerase cyclophilin A interacts with the p24 region of p55 leading to its incorporation into HIV particles and it is

essential for viral replication (Franke & Luban, 1996). The NC of p55 is responsible for specifically recognizing the so-called “packaging signal” that mediated the incorporation of RNA genome into HIV-1 virions. The packaging signal consists of four stem loop structures located near the 5' end of the viral RNA and are bounded by NC through interactions mediated by two zinc-finger motifs (Harrison & Lever, 1992). NC appears also essential in facilitating the reverse transcription of RNA genome allowing the formation of proviral DNA (Lapadat-Tapolsky et al, 1993). Finally the p6 polypeptide region mediates interactions between p55 Gag and the accessory protein Vpr, leading to the incorporation of Vpr into assembling virions (Paxton et al, 1993). The p6 region also contains a so-called late domain which is required for the efficient release of budding virions from an infected cell.

- **Gag-Pol precursor**

The Gag-Pol precursor (p160) is generated by a ribosomal frame shift. During viral maturation, the virally encoded protease cleaves the Pol polypeptide away from Gag and further digests it to separate the protease (PR or p10), the reverse transcriptase (RT or p50), RNase H (p15), and integrase (INT or p31). The HIV-1 protease is an aspartyl protease that acts as a dimer. Protease activity is required for cleavage of the Gag and Gag-Pol polyprotein precursors during virion maturation as described previously. The RT enzyme is a DNA polymerase-RNA dependent that synthesizes a double-stranded DNA from the single stranded RNA genome. RNase H removes the original RNA template from the first DNA strand, allowing synthesis of the complementary strand of DNA. Because the polymerase does not contain a proof-reading activity, replication is error-prone and introduces several point mutations into each new copy of the viral genome. The integrase protein mediates the insertion of the proviral DNA into the genomic DNA of an infected cell. This process is mediated by three distinct functions of IN. First, an exonuclease activity trims two nucleotides from each 3' end of the linear viral DNA duplex. Then, a double-stranded endonuclease activity cleaves the host DNA at the integration site. Finally, a ligase activity generates a single covalent linkage at each end of the proviral DNA (Bushman et al, 1990).

- **Env**

The polyprotein Env (gp160) is expressed from singly spliced mRNA and is synthesized in the endoplasmic reticulum. Then, Env undergoes glycosylation, an essential step for viral infectivity (Capon & Ward, 1991). A cellular protease cleaves gp160 into the transmembrane protein gp41 and the surface protein gp120.

- **Tat**

Tat (trans-activator of transcription) is a transcriptional transactivator essential for HIV-1 replication. Tat is an RNA binding protein that interacts with a short-stem loop structure called TAR (transactivation response element) located at the 5' end of viral RNA. The binding of Tat to TAR activates transcription from the HIV promoter LTR (Long Terminal Repeats) at least 1000-fold (Roy et al, 1990). Tat seems to promote primarily the elongation phase of HIV-1 transcription to produce full-length transcripts. In the absence of Tat expression, HIV generates mainly short (>100 nucleotides) transcripts (Feinberg et al, 1991).

- **Rev**

Rev (regulator of viral expression) is a 13-kD sequence-specific RNA binding protein and acts to induce the transition from the early to the late phase of HIV gene expression (Kim et al, 1989). In the absence of the Rev, the host splicing machinery in the nucleus splices the RNA allowing the production of the regulatory proteins Rev and Tat and the accessory protein Nef. In the presence of Rev, RNA is exported from the nucleus before it can be spliced, so that the structural proteins and RNA genome can be produced. Interestingly, Rev binds to a RNA secondary structure, called RRE (Rev response element), facilitating the export of unspliced and incompletely spliced viral RNAs from the nucleus to the cytoplasm. This mechanism allows to overcome the host's splicing machinery system in the nucleus and to produce the structural viral proteins (Strebel, 2003).

- **Vif**

Vif (virulence factor) is a 23 kDa accessory protein and the lack of this protein seems strongly impair the viral infectivity. In fact, virions generated in absence of this protein seem to be about 1000 times less efficient in establishing the infection (Strebel et al, 1987). Although this

protein is not essential for viral replication, the absence of Vif in primary cells results in a defected replication (Fan & Peden, 1992).

- **Vpr**

The 14-kDa accessory protein Vpr (Viral protein R) confers rapid growth advantage to Vpr-expressing viruses. This evidence is more pronounced in macrophages than in primary T-cells (Balliet et al, 1994). As for the Vif protein, Vpr is incorporated into viral particles, with approximately 100 molecules of Vpr in each virion (Cohen et al, 1990). Moreover, Vpr facilitates the nuclear localization of the Pre Integration Complex (PIC, see HIV replication cycle section) in non-dividing cells, elucidating its important role at a preintegration level (Heinzinger et al, 1994).

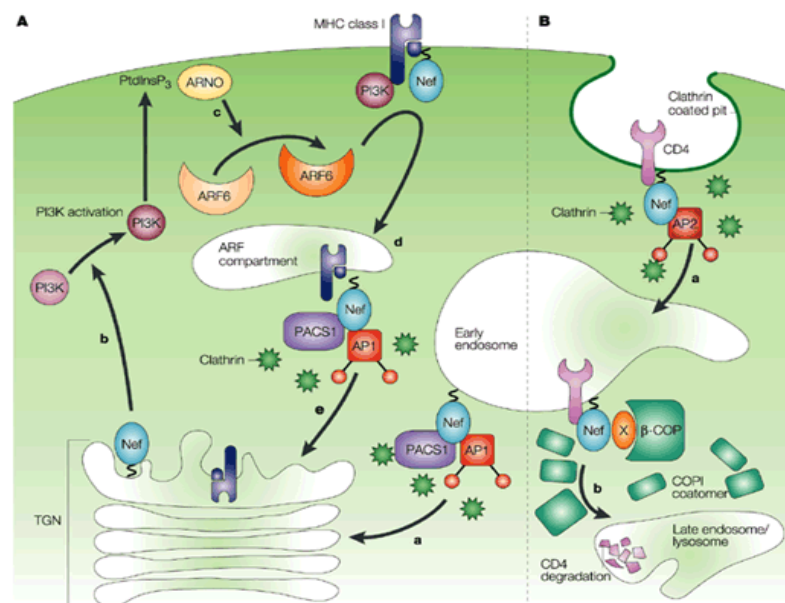
- **Vpu**

Vpu (Viral Protein U) is an unique accessory protein of HIV-1 and is expressed from the mRNA that also encodes *env*. Vpu is translated from this mRNA at levels tenfold lower than that of Env because the Vpu translation initiation codon is not efficient (Schwartz et al, 1990). Vpu is an integral membrane phosphoprotein that is primarily localized in the internal membranes of the host cell. The two main functions of Vpu are the down-modulation of CD4 antigen and the enhancement of virion release (Schubert et al, 1996).

- **Nef**

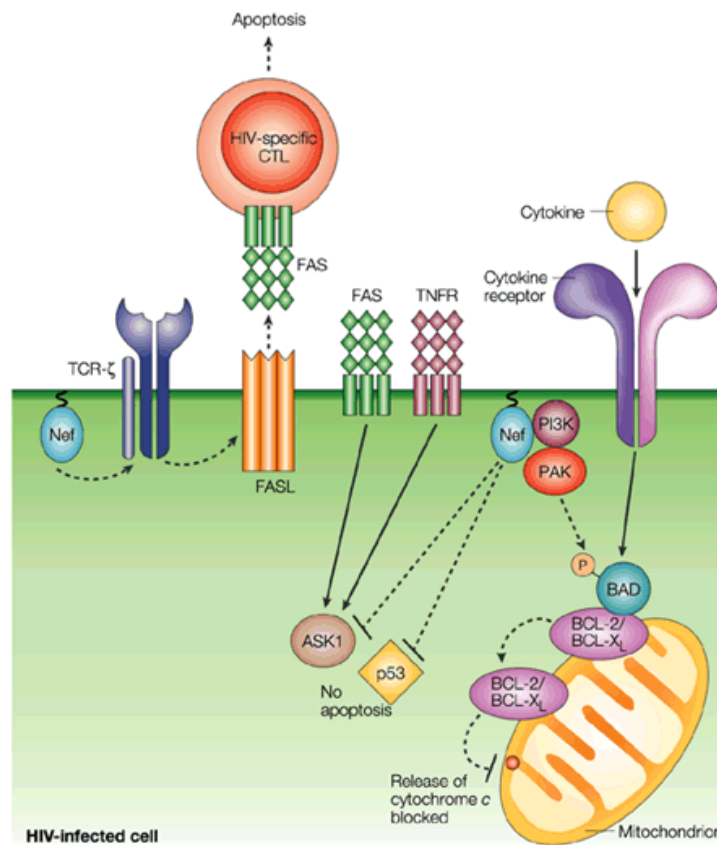
The Nef (Negative Factor) coding region, a 621 bp-long sequence located at the 3'-end of the viral genome, partially overlaps with the 3'-long terminal repeat (LTR) region. Nef was originally characterized as a negative regulator of HIV infection and was thus named as "negative factor". This was then refused by several research groups. Nef is a 27 kDa myristoylated protein expressed early in the HIV-1 life cycle. The *nef* gene is highly conserved in all primate lentivirus e.g. HIV-1, HIV-2 and SIV. Although Nef is not required for HIV-1 replication *in vitro*, it appears as a fundamental factor for efficient viral replication and pathogenesis *in vivo*; it also facilitates virus replication and enhances viral infectivity *in vitro* (Miller et al, 1994). Moreover, virus produced from a *nef* mutated proviral DNA results in a 4 to 40 less infectious virions in single-round infection assay (Das & Jameel, 2005). An essential role for Nef *in vivo* has been demonstrated in a subset of long-term non progressors,

HIV-infected individuals that do not progress to AIDS. Viral isolates from some of these individuals exhibit either a deletion in the *nef* gene or defective *nef* alleles (Salvi et al, 1998). In addition, rhesus macaques infected with an engineered strain of SIV that lacked the functional Nef protein also did not attain high viral loads and did not progress to clinical disease (Kestler et al, 1991). Nef has a positive effect on viral infection and replication by promoting the survival of infected cells by several mechanisms: for example, to promote escape from the immune system and infectivity, it downregulates CD4 and Major Histocompatibility Complex I (MHC I) expression on the cell surface, to enhance viral replication and infectivity it activates CD4<sup>+</sup> T lymphocytes (Richter et al, 2009). The downmodulation of critical cell surface protein, such as CD4 and MHC I, is the most extensively studied function of Nef and is represented in Figure 2.27. Nef accelerates the endocytosis of MHC I molecules through the phosphofurin acidic cluster sorting protein 1 (PACS1)/phosphatidylinositol 3-kinase (PI3K)-dependent activation of ADP ribosylation factor 6 (ARF6) mediated endocytosis (Figure 2.27A). The downregulation of MHC I decreases the efficiency of cytotoxic T cells in killing HIV infected cells. Nef is also responsible for the CD4 downmodulation (Figure 2.27B). At the plasma membrane, Nef connects the cytoplasmic tail of CD4 with clathrin-coated pits through an interaction with adaptor protein 2 (AP2) and the vacuolar ATPase (v-ATPase), triggering rapid endocytosis of the CD4 receptor (Das & Jameel, 2005).



**Figure 2.27** Nef-induced downmodulation of A) MHC class I and B) CD4 molecules (Peterlin & Trono, 2003)

Moreover, Nef induces apoptosis in both infected and uninfected immune effector cells (Figure 2.28). *In vivo* HIV-infection of lymphatic tissue is accompanied by enhanced apoptosis, which affects mainly bystander cells. This effect is in part due to the Nef-induced upregulation of expression of FAS ligand (FASL) on the surface of infected cells. FASL could interact with FAS molecules on neighboring cells, including virus-specific Cytotoxic T-cells (CTLs), thereby triggering their apoptosis (Das & Jameel, 2005).

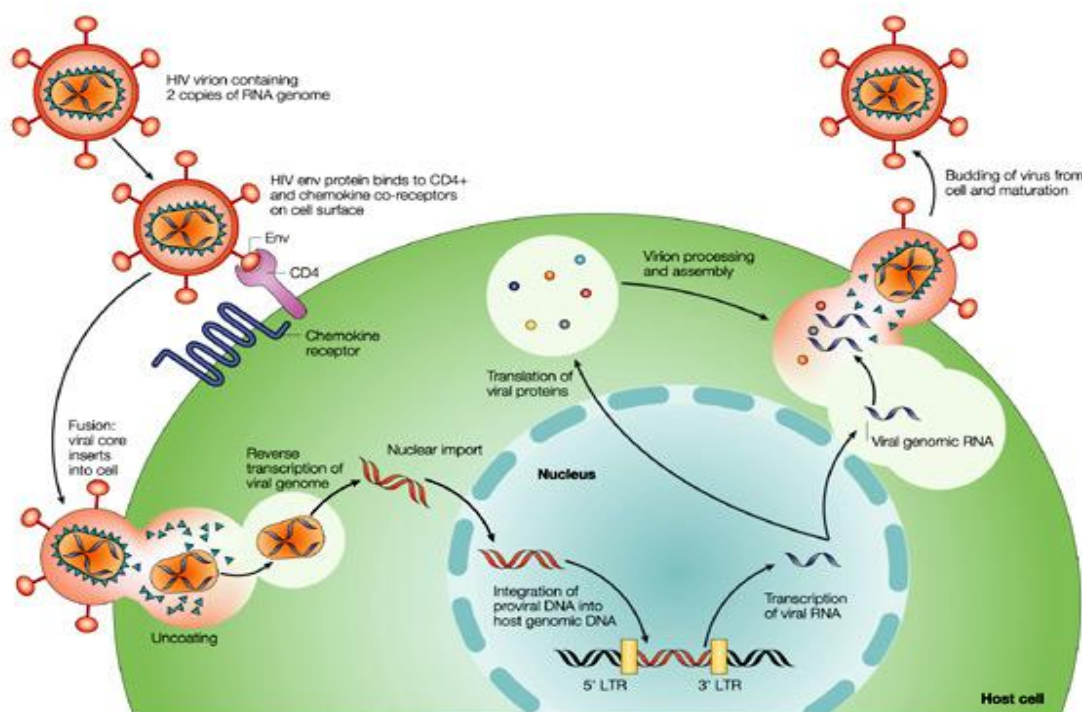


**Figure 2.28 Nef protein induces apoptosis** (Peterlin & Trono, 2003)

Interestingly, Nef protein has been proposed as one attractive target for the anti-HIV therapy (Breuer et al, 2011; Chutiwitoonchai et al, 2011; Narute & Smithgall, 2012). In fact, combinations of the existing drugs are very effective in slowing down progression to AIDS; however, the high mutation rate of HIV gives rise to resistance which ultimately impairs antiretroviral therapy. Therefore, there is an urgent need for new anti-HIV drugs with an innovative mechanism of action, possibly against highly conserved viral sites. Interestingly, Nef represent one of the most conserved viral sites, as previously reported.

### 2.3.2. HIV replication cycle

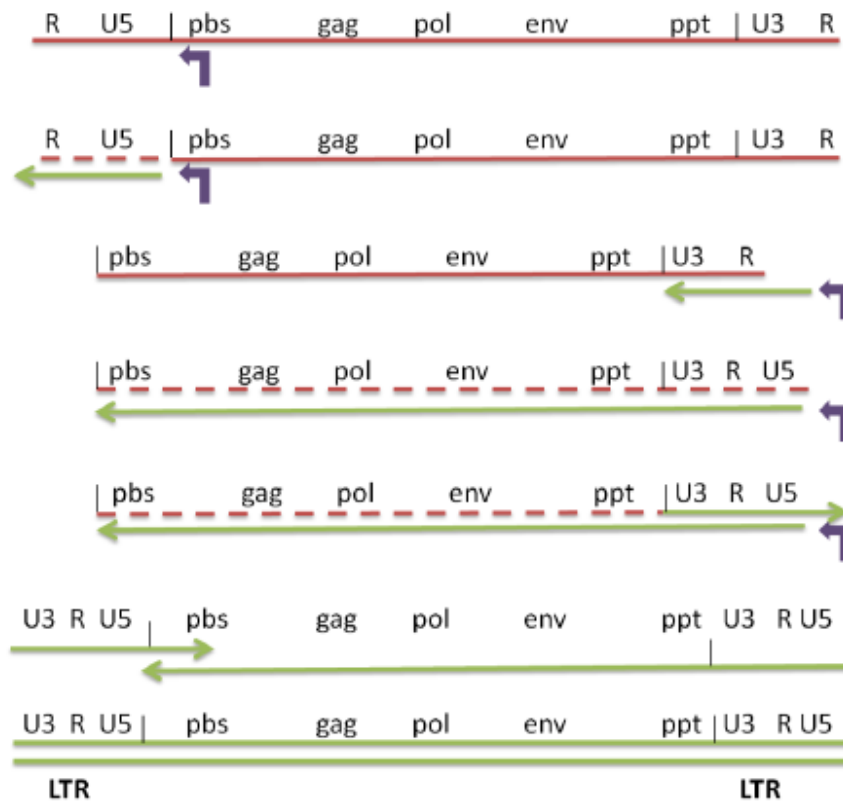
The HIV replication cycle takes about 24 hours and is represented in Figure 2.29.



**Figure 2.29 HIV replication cycle** (Rambaut et al, 2004)

The viral particle recognizes the host cell, mainly CD4+ T lymphocytes and macrophages, through the glycoprotein gp120 that specifically binds to the cellular CD4 receptor. Viral entry requires additional binding to a co-receptor molecule CCR5 or CXCR4 (Doms & Trono, 2000). The differential recognition of a co-receptor is responsible for the viral tropism. In fact, M-tropic strains use the beta-chemokine receptor CCR5 for entry and are thus able to replicate in macrophages and CD4+ T-cells. These strains are now called R5 viruses. T-tropic strains replicate in primary CD4+ T-cells as well as in macrophages and use the alpha-chemokine receptor, CXCR4, for entry. These strains are now called X4 viruses (Clapham & McKnight, 2001). Binding of gp120 to CD4 and the co-receptor triggers a conformational change that exposes the gp41 viral fusion protein which inserts into the host membrane. The resulting fusion of viral and cellular membrane injects the viral core into the host cytoplasm where, after the uncoating process, the two RNA genome copies are reverse transcribed into a

linear double-stranded DNA molecule by the viral reverse transcriptase (Mougel et al, 2009). The full reverse transcription process is represented in Figure 2.30.

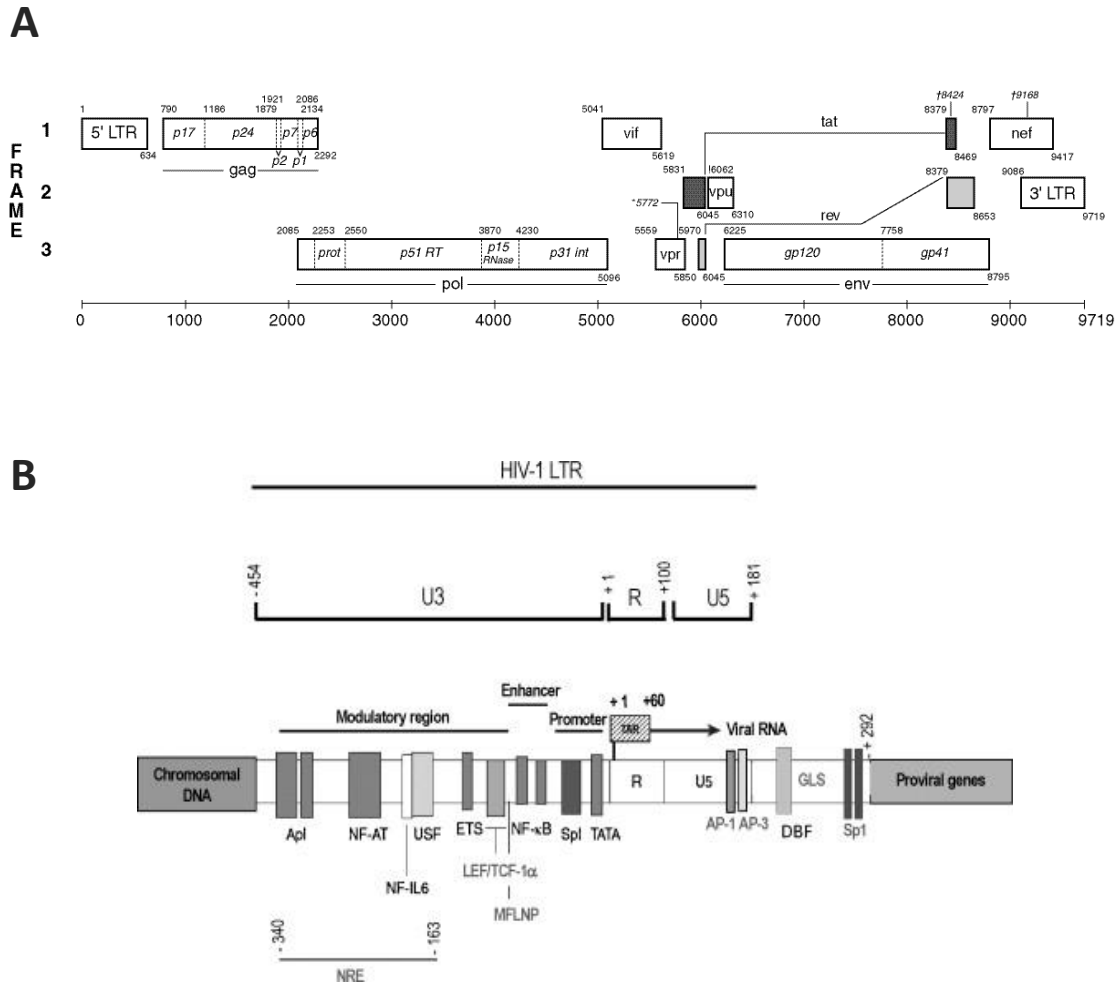


**Figure 2.30 HIV reverse transcription process**

The full-length double-stranded DNA, viral integrase, MA and Vpr and various cellular proteins form the so-called pre-integration complex (PIC) that is imported into the nucleus. The viral enzyme integrase mediates integration into the human chromosome of the proviral DNA. In the proviral form the HIV-1 DNA genome contains two identical copies of the long terminal repeat (LTR) at the 5' and 3' end of the genome (Figure 2.31A). The LTRs are composed of the segment U3 (derived from a unique sequence located at the 3' end), R (repeated sequence at both ends) and U5 (derived from a unique sequence located at the 5' end). The promoter region U3 can be divided in three functional sections: an upstream modulatory element (-454 to -104, with the respect of the first transcribed base) including binding sites for cellular transcription factors, an enhancer (-105 to -79) with two binding sites for the nuclear factor  $\kappa$ B (NF- $\kappa$ B), and the core promoter (-78 to -1) composed of three tandem binding sites for specificity protein 1 (Sp1) and a TATA box (Figure 2.31B) (Luciw,



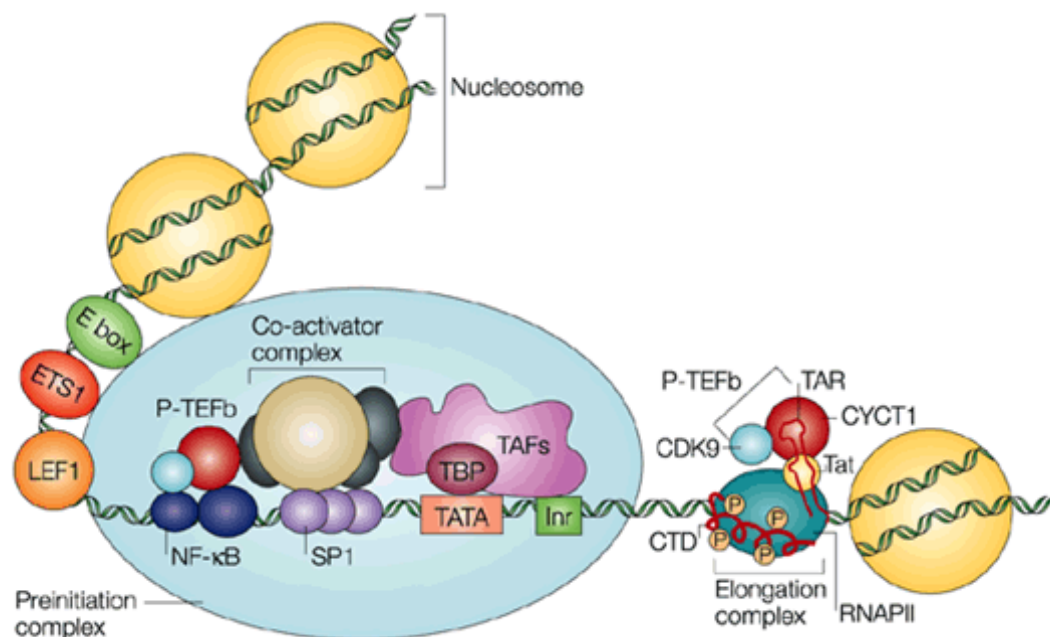
1996). The LTR-directed gene expression is regulated in a cell type and differentiation-dependent manner by the binding of both host and viral protein to the LTR region. The most important host transcription factors are members of the following categories: the Specificity protein (Sp) family (e.g. Sp1), nuclear factor kappa B (NF- $\kappa$ B) family, activator protein 1 (AP-1) proteins, nuclear factor of activated T cells (NFAT) and CCAT enhancer binding protein (C/EBP) family. All these proteins bind specific LTR sequences located in the U3 region with a different level of conservation. Genetic variability within LTR binding sites in U3 and TAR regions has been observed in several HIV-1 subtypes (de Arellano et al, 2010; Michael et al, 1994). However, NF- $\kappa$ B and Sp1 binding sites are remarkably conserved (Jeeninga et al, 2000). AP-1 and NFAT bind to the modulatory region, NF- $\kappa$ B binds to 2 sites located in the enhancer region while Sp factors interacts in the core promoter region. Among the Sp proteins, Sp1 plays a crucial role in regulating HIV-1 transcription. In fact, Sp1 is a transcriptional activator that specifically binds 3 sites located in the HIV promoter through its zinc finger binding domain. Interestingly, Sp1 shows affinity and specificity for the GC rich sequence GGGGCGGGGC. The core promoter contains also a TATA box essential for the initiation and regulation of transcription. Also viral protein such as Vpr and Tat bind to the LTR to regulate transcription. The role of these viral proteins have been anticipated in the previous part of the thesis (Kilareski et al, 2009).



**Figure 2.31 HIV-1 proviral genome** A) HIV-1 proviral genome organization B) HIV Long Terminal Repeat (LTR) organization (de Arellano et al, 2010)

Thus, viral genes are expressed from the stably integrated HIV-1 proviral DNA by the host transcription machinery including RNA polymerase II. As anticipated, the viral promoter is located in the U3 region of the 5' LTR and requires activation by host transcription factors (Figure 2.32). Upstream from the transcription start site, in the core promoter the initiator (Inr), the TATA BOX and three Sp1-binding sites are necessary for the correct position of RNA polymerase II (RNAPII) in the so-called pre-initiation complex. Moreover, also the TATA-binding protein (TBP) and TBP-associated factors (TAFs) contributes to the formation of this complex. After the formation of the pre-initiation complex, the RNAPII clears the promoter, starting the transcription. The initial transcriptional output is however very low, due to the blocked elongation of viral transcripts early in the 5' portion of the RNAs. The viral transactivator protein Tat, which is made very early from the tiny amounts of successfully

terminated mRNAs, is required to achieve normal levels of expression (Coiras et al, 2010). A peculiar feature of the LTR is the presence of a RNA regulatory element known as the TAR element. As anticipated, Tat protein binds TAR element. Tat and its cellular co-factor P-TEFb (Positive Transcription Elongation factor B) cooperate to bind TAR with high affinity, allowing RNAPII to produce full-length viral transcripts. P-TEFb contains the components cyclin T1 (CYCT1) and cyclin-dependent kinase 9 (CDK9) that mediates a phosphorylation of the carboxy-terminal domain of the RNAPII. This converts the initiating transcription complex to an elongating transcription complex (Peterlin & Trono, 2003).



**Figure 2.32 The HIV Long Terminal Repeat (LTR) promoter** TATA-binding protein (TBP) and TBP-associated factors (TAFs) bind the core promoter. The co-activator complex binds SP1 and cooperates to recruit and position RNA polymerase II (RNAPII) in the pre-initiation complex. RNAPII then clears the promoter. Positive transcription elongation factor b (P-TEFb) and transactivator of transcription (Tat) bind the TAR (Transactivation response) element. Thus, the initiating transcription complex is converted into an elongating transcription complex and the transcription efficiently starts. The enhancer binds members of the nuclear factor- $\kappa$ B (NF- $\kappa$ B), nuclear factor of activated T cells (NFAT) and ETS families (Peterlin & Trono, 2003)

Thereafter, full-length mRNA transcripts are efficiently synthesized. These unspliced transcripts contain multiple splice sites for the generation of over 40 unique viral transcripts for translation of the nine viral proteins. To date, the full-length mRNA serve as the template for the polyproteins gag (p55), gag-pol (p160) and env (gp160), while accessory proteins are translated from spliced mRNA. The mRNA molecules are next transported to the cytoplasm where the translation of viral proteins occurs. The env polyprotein (gp160) is then cleaved

resulting in the two glycoproteins gp41 and gp120 (Hallenberger et al, 1992). These envelope glycoproteins are transported to the plasma membrane of the host cell where a immature virion is assembled from the precursors proteins together with and two copies of the viral genomic RNA. After this, the virion starts to bud off from the surface of the cell. After the release, the enzyme protease cleaves the gag and gag-pol polyproteins to complete the maturation of the virion.

### **2.3.3. Antiretroviral treatment**

In 1987, the first anti-HIV drug was approved by the U.S. Food and Drug Administration (FDA) for the AIDS treatment. This was the reverse transcriptase inhibitor 3'-azido-3'-deoxythymidine (AZT), also known as Zidovudine. Until the 2012, 25 more antiretroviral compounds were approved by FDA. These drugs belong to several classes, based on their mechanism of action and viral target:

- 1) Nucleoside Reverse Transcriptase Inhibitors (NRTIs)
- 2) Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)
- 3) Protease Inhibitors (PI)
- 4) Entry Inhibitors
- 5) Integrase inhibitors
- 6) CCR5 antagonists

These drug classes target four key steps in the viral cycle: viral entry, reverse transcription, integration and protein processing/maturation (Palmisano & Vella, 2011). Currently the indicated treatment is a combination of drugs, the so-called highly active antiretroviral therapy (HAART). The HAART consists in a cocktail containing at least three different drugs, usually two NRTIs and one NNRTI or PI. This therapy resulted really effective against AIDS, determining a decline in mortality and mobility. However, the high mutation rate of HIV due to lack of RT proofreading activity gives rise to resistance which, at the end, impairs antiretroviral therapy. In addition, although HAART is very effective in blocking HIV-1 spread within the body, it is not a cure, as viral loads readily rebound when treatment is interrupted (Chun et al, 1999). Moreover, the existence of latently infected CD4<sup>+</sup> T-cells represent a big problem, since HAART do not affect the latent virus resting in this cell population (Chan et al, 2013). Therefore, there is an urgent need of new anti-HIV drugs with

an innovative mechanism of action, possibly against highly conserved viral sites and able to clear the virus from the infected human host in order to reach complete recover. All anti-HIV drugs have focussed so far on inhibition of viral proteins; however, the viral genome, both at the pre-integration (single-stranded RNA) and post-integration (double-stranded DNA) stage could be a very effective and selective target.

#### **2.3.4. G-quadruplexes and HIV-1**

In 1992 Sundquist and Heaphy demonstrated for the first time a role of G-quadruplex structure during the dimerization of HIV-1 genome. As anticipated, mature HIV-1 virions contain 2 homologous copies of their single stranded RNA genome which are stably associated within an RNA-gag protein complex. This RNA-RNA association appears more stable in a site near the 5' end of each strand named dimer linkage structure (DLS). HIV RNA genomes dimerizes spontaneously in the absence of protein cofactors under condition of high ionic strength by forming an interstrand G-quadruplex (Sundquist & Heaphy, 1993). Ten years later, Lyonnais *et al.* discovered a G-quartet structure associated within a single-stranded portion (central DNA flap) of the reverse-transcribed pre-integration HIV-1 genome (Lyonnais et al, 2002). Moreover, this tetraplex structure specifically interacts with the viral nucleocapsid (NC) protein: in particular, G-quadruplexes can promote the NC assembly along ssDNA, thereby protecting the pre-integrated genome from nuclease degradation (Lyonnais et al, 2003). In addition, Kankia *et al.* discovered a peculiar function of NC in unfolding DNA quadruplex (Kankia et al, 2005). Recently, G-quadruplexes have been proposed to promote the genetic recombination in a specific recombinant *gag* hot spot at the genomic RNA level. In fact, when the two co-packaged RNA templates are non-identical, template switching mediated by the RT enzyme results in genetic recombination. Sequences and structures, such as G-quadruplexes, that cause RT pausing seem the main actors of these genetic recombination. (Shen et al, 2009). Together these evidences support the hypothesis of a regulatory role of G-quadruplex structure during HIV-1 viral cycle. However, in the proviral genome no G-quadruplex structures have been reported.



### **3. Aim of the study**

The principal aims of this study were 1) to evaluate the presence of putative G-quadruplex forming sequence in the HIV-1 proviral genome, 2) to dissect their biological/virological significance, 3) to assess the effect of G-quadruplex ligands on the HIV-1 G-quadruplex structures.

In the first part of this study, the LTR promoter region was characterized for the presence of G-quadruplex structures in order to investigate G-quadruplex-mediated alterations in the regulation of transcription. The second part of this thesis focused on G-quadruplex implication in Nef protein expression, evaluating possible consequences for the HIV-1 viral cycle. Finally, the antiviral activity of G-quadruplex ligands was tested to elucidate their modes of action against HIV-1 and to further investigate G-quadruplex as innovative antiviral targets.





## 4. Materials and Methods

### 4.1. Materials and compounds

Oligonucleotides were from Sigma-Aldrich (Milan, Italy). T4 polynucleotide kinase was from Invitrogen (Paisley, UK), [ $\gamma$ - $^{32}$ P]ATP from Perkin-Elmer (MA, USA).

Clerocidin (CL) was a gift of Leo Pharmaceutical Products (Ballerup, Denmark). Dimethylsulfate (DMS) was from Sigma Aldrich, TMPyP4 and PIPER from Calbiochem, (Merck Chemicals, Nottingham, UK), and BRACO-19 from ENDOTHERM, (Saarbruecken, Germany).

Dextran sulfate (DS) was purchased from Sigma (Bornem, Belgium). Nevirapine was obtained from BoehringerIngelheim (Ridgefield, CN). AMD3100 was a gift from Dr. G. Henson (AnorMED, Langley, British Columbia, Canada), Zidovudine (AZT) was synthesized C. Pannecouque (Rega Institute for Medical Research, Leuven, Belgium) according to the method described by Horwitz et al. (Horwitz et al, 1964). Ritonavir was purchased from Molekula (Dorset, United Kingdom)

### 4.2. Cells

HEK 293T (Human Embryonic Kidney 293T) cell line was grown in DMEM medium (Gibco, Life Technologies, Paisley, UK) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Gibco, Life Technologies, Paisley, UK).

TZM-bl is an engineered cell line to express the reporter gene Luciferase from HIV-1 LTR promoter. In this system, infectivity is measured as stimulation of luciferase reporter gene expression driven by the HIV-1 LTR in response to infection with HIV-1, and this effect is enhanced by HIV-1 Nef (Emert-Sedlak et al, 2013). TZM-bl reporter cell line was provided by NIH AIDS Research and Reference Reagent Program to Prof. T.E. Smithgall (University of Pittsburg School of Medicine, Pennsylvania, USA).

MT-4 (human T-lymphotropicvirus type I (HTLV-I)-transformed cells) were provided by Dr. N. Yamamoto (Tokyo Medical School and Dental University School of Medicine, Tokyo, Japan), MT-4/III<sub>B</sub> (HIV-1(III<sub>B</sub>) persistently infected MT-4 cell line) was selected in the lab of Prof. C. Pannecouque (Rega Institute for Medical Research, Leuven, Belgium) and MT-4-LTR-eGFP cell lines were kindly provided by provided by Tibotec. MT-4 and MT-4/III<sub>B</sub> were grown in RPMI 1640 medium (Life Technologies, Merelbeke, Belgium) supplemented with

10% fetal bovine serum (FBS; Sigma-Aldrich, Diegem, Belgium), 2mM L-glutamine(Life Technologies, Merelbeke, Belgium), 0.1% sodium bicarbonate (Life Technologies, Merelbeke, Belgium) and 20 µg/ml gentamicin (Life Technologies, Merelbeke, Belgium). MT-4-LTR-eGFP (MT-4 engineered cell line to express the reporter gene GFP from HIV-1 LTR promoter) cell line were kindly provided by Tibotec and were grown in RPMI 1640 medium supplemented with 10% FBS, 2 mM L-glutamine, 0.1% sodium bicarbonate, 20 µg/ml gentamicin (Life Technologies, Merelbeke, Belgium) and 100µg/ml G418 (Sigma-Aldrich, Diegem, Belgium).

Peripheral blood mononuclear cells (PBMCs) were isolated from HIV-seronegative donor buffy coats using Lymphoprep (Nycomed, Oslo, Norway), and stimulated in RPMI 1640 medium containing 15% FBS (Sigma-Aldrich, Diegem, Belgium), 2 mM L-glutamine (Life Technologies, Merelbeke, Belgium), 0.1% sodium bicarbonate (Life Technologies, Merelbeke, Belgium), 60 µg/ml gentamicin (Life Technologies, Merelbeke, Belgium), 2 µg/ml phytohaemagglutinin (Sigma, Diegem, Belgium) and 5 U/ml human interleukin-2 (Roche Diagnostics, Vilvoorde, Belgium). After 3 days of stimulation the cells are washed and resuspended in RPMI 1640 medium supplemented with 2 mM L-glutamine (Life Technologies, Merelbeke, Belgium), 0.1% sodium bicarbonate (Life Technologies, Merelbeke, Belgium), 15% FBS (Sigma-Aldrich, Diegem, Belgium), 60 µg/ml gentamicin (Life Technologies, Merelbeke, Belgium) and 10 U/ml human interleukin-2 (Roche Diagnostics, Vilvoorde, Belgium).

All the cell lines were incubated at 37°C in a humidified CO<sub>2</sub>-controlled atmosphere.

### **4.3.G-quadruplex analysis of the HIV-1 proviral genome**

The HIV-1 wild type genome (strain HXB2\_LAI; NC\_001802) was analyzed by QGRS Mapper (<http://bioinformatics.ramapo.edu/QGRS/index.php>) for prediction of G-quadruplex forming sequences in both coding and non-coding strands (Kikin et al, 2006). The following restrictions were applied: maximum length 30 nucleotides (nt); minimum G-group size 2 nt; loop size 0–15 nt. The found putative G-quadruplex forming sequences were ranked based on the G-score, which is the likelihood to form a stable G-quadruplex, according to the following principles: a) shorter loops are more common than longer loops; b) G-quadruplexes tend to have loops roughly equal in size; c) the greater the number of G tetrads, the more stable the G-quadruplex. For further investigations, we chose sequences on the basis of assigned G-score and of relevant location in the HIV-1 genome. We focused on putative G-quadruplex

forming sequences in the Long Terminal Repeats (LTR) promoter region and in HIV-1 *nef* coding region.

For the HIV-1 LTR region, conserved bases were evaluated by aligning 953 LTR U3 sequences from the HIV database (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>) using Jalview (<http://www.jalview.org/>).

For the HIV-1 *nef* coding region, three putative G-quadruplex sequences (Nef 8528 and Nef8624 in coding strand, Nef8547 in non-coding strand) were evaluated for their consensus sequence by aligning 3224 Nef sequences (HIV-1, M Group) from the HIV database (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>) using Jalview (<http://www.jalview.org/>). In addition, the conservation grade of the G-pattern necessary for G-quadruplex folding was investigated. Aligned sequences were searched for G-quadruplex patterns expressed as Perl compatible regular expressions using GNU grep command line tool.

## 4.4.Spectroscopic analysis

### 4.4.1. Circular Dichroism analysis

Circular Dichroism (CD) is a polarized light spectroscopy that offers the possibility to discriminate a G-quadruplex conformation from other secondary architectures. CD could give information about topology and thermal stability of a G-quadruplex structure. In fact, since most of the G-quadruplex conformations have a characteristic CD spectrum, it is possible to distinguish them simply analysing the different CD spectra: parallel G-quadruplex has a positive CD peak around 260nm and a negative CD peak around 240 nm; antiparallel G-quadruplex shows a positive peak at 295nm and a negative peak around 260nm while a mixed G-quadruplex conformation exhibits two positive peaks at around 260 nm and 295nm. In addition, following the G-quadruplex CD-spectrum over a range of increasing temperatures, it is possible to study the thermal stability properties of the secondary structure and to have quantitative information in terms of melting temperature. This is particularly useful to investigate the stabilizing effect of G-quadruplex ligands that is proportional to the increasing of melting temperature ( $T_m$ ) of the G-quadruplex. Finally, CD can provide information about G-quadruplex-ligand complexes. Normally, G-quadruplex ligands are non-chiral molecules and so CD-inactive. However, upon interaction with the secondary structure, non-chiral molecules could show an induced CD (ICD) signal due to the chiral environment of the

bounded ligand. Although the ICD spectrum represents a strong and direct evidence of DNA-ligand interaction, the absence of an ICD does not mean the absence of the ligand's binding to the secondary structure.

For CD analysis, all DNA oligonucleotides (Table 4.1) were diluted from stock to final concentration (4  $\mu\text{M}$ ) in lithium cacodylate buffer (10 mM, pH 7.4) and, where appropriate, KCl or NaCl. All samples were annealed by heating at 95°C for 5 min, gradually cooled to room temperature, and measured after 24 h. Compounds at 16  $\mu\text{M}$  final concentration were added after DNA annealing. CD spectra were recorded on a Jasco-810 spectropolarimeter (Jasco, Easton, MD, USA) equipped with a Peltier temperature controller using a quartz cell of 5mm optical path length and an instrument scanning speed of 100 nm/min with a response time of 4 s over a wavelength range of 230–320nm. The reported spectrum of each sample represents the average of 2 scans at 20°C and is baseline-corrected for signal contributions due to the buffer. Observed ellipticities were converted to mean residue ellipticity ( $\theta$ ) = deg  $\times$  cm<sup>2</sup>  $\times$  dmol<sup>-1</sup> (mol.ellip.). For the determination of T<sub>m</sub>, spectra were recorded over a temperature range of 20–95°C, with temperature increase of 5°C/min or 2°C/min, followed by an equilibration step of 1 min. T<sub>m</sub> 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 (Doria et al, 2013; Greenfield, 2006).

Application	Project	Name	Sequence (5'→3')
CD	LTR region characterization	LTR FL	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGG GAGGCGTGGCCTGGGCGGGACTGGGGAGTGG TTTTT
		LTR I	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGG GAGGCGTGGCCTGGGTTTTT
		LTR II	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTG GGCGGGTTTTT
		LTR III	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGG GGTTTTT
		LTR II+III	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTG GGCGGGACTGGGGTTTTT
	<i>nef</i> coding region characterization	Nef8528	GAGGAGGAGGTGGGT
		Nef8547	GGTCTTAAAGGTACCTGAGGTCTGACTGG
		Nef8624	GGGGGGACTGGAAGGG
		Nef8528-Compl	ACCCACCTCCTCCTC
		Nef8547-Compl	CCAGTCAGACCTCAGGTACCTTTAAGACC
		Nef8624-Compl	CCCTTCCAGTCCCCC

**Table 4.1 Oligonucleotides used in HIV-1 LTR region characterization and in *nef* coding region characterization for CD analysis**

#### 4.4.2. UV Spectroscopy analysis

Similarly to CD spectroscopy, UV spectroscopy is used both to discriminate the G-quadruplex from other secondary structures and to further characterize them. For example, the difference in the UV absorbance before and after the melting temperature (respectively 20°C and 95°C) in the region 220-320nm gives rise to the so called Thermal difference spectrum (TDS) that is characteristic for the G-quadruplex structure. This typical TDS shows four distinct positive/negative bands around at 243, 255, 273 and 295 nm and provide further proof of the effective G-quadruplex formation *in vitro*. In addition, it is possible to calculate the TDS factor ( $\Delta A_{240} \text{ nm} / \Delta A_{295} \text{ nm}$ , where  $\Delta A_{\lambda}$  is the difference, at the given wavelength  $\lambda$ , between the absorbance above and below the melting temperature, expressed as absolute value) that could give information about G-quadruplex topology since parallel and mixed-antiparallel G-quadruplex showed a big difference in terms of TDS factor value. Normally parallel G-quadruplex topology gives a TDS factor value around 4, while mixed-antiparallel G-quadruplex shows a TDS factor value around 1.

For thermal difference spectrum (TDS) analysis, all DNA oligonucleotides (Table 4.2) were diluted from stock to final concentration (4  $\mu\text{M}$ ) in lithium cacodylate buffer (10 mM, pH 7.4) and KCl (100mM). All samples were annealed by heating at 95°C for 5 min, gradually cooled to room temperature, and measured after 24 h. UV spectra were recorded on Lambda25 UV/Vis spectrometer (Perkin-Elmer) equipped with a Peltier temperature controller over a temperature range of 20–95°C. TDS spectra were calculated by subtracting the spectrum at 20°C from the spectrum at 95°C. TDS factors were calculated as the absolute values of  $\Delta A_{240} \text{ nm} / \Delta A_{295} \text{ nm}$ , where  $\Delta A_{\lambda}$  is the difference, at the given wavelength  $\lambda$ , between the absorbance above and below the melting.

Application	Project	Name	Sequence (5'→3')
UV	LTR region characterization	LTR FL	TTTTTGGGACTTTCCGCTGGGGACTTTCCAG GGAGGCGTGGCCTGGGCGGGACTGGGGAGT GGTTTTT
		LTR I	TTTTTGGGACTTTCCGCTGGGGACTTTCCAG GGAGGCGTGGCCTGGGTTTTT
		LTR II	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCT GGGCGGGTTTTT
		LTR III	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGG GGTTTTT
		LTR II+III	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCT GGGCGGGACTGGGGTTTTT

**Table 4.2** Oligonucleotides used in HIV-1 LTR region characterization for UV analysis

#### 4.5. *Taq* polymerase stop assay

The formation of G-quadruplex structure in a DNA template can be investigated using a *Taq* Polymerase Stop assay. The basis of this assay is that a G-quadruplex structure could sterically block the *Taq* enzyme progression in elongating a  $^{32}\text{P}$ -labeled primer (Figure 4.1). Products representing the full length and major arrest sites can be evaluated by denaturing electrophoresis. In this assay it is essential to optimally set the G-quadruplex folding conditions (e.g. cations concentration in solution) together with the choice of G-quadruplex ligands concentration to avoid any influences on the enzyme activity. Normally the *Taq* reaction is performed in a single cycle at a temperature below or around the  $T_m$  of the G-quadruplex structure. However, some G-quadruplexes can be stable also in a complete PCR cycle.

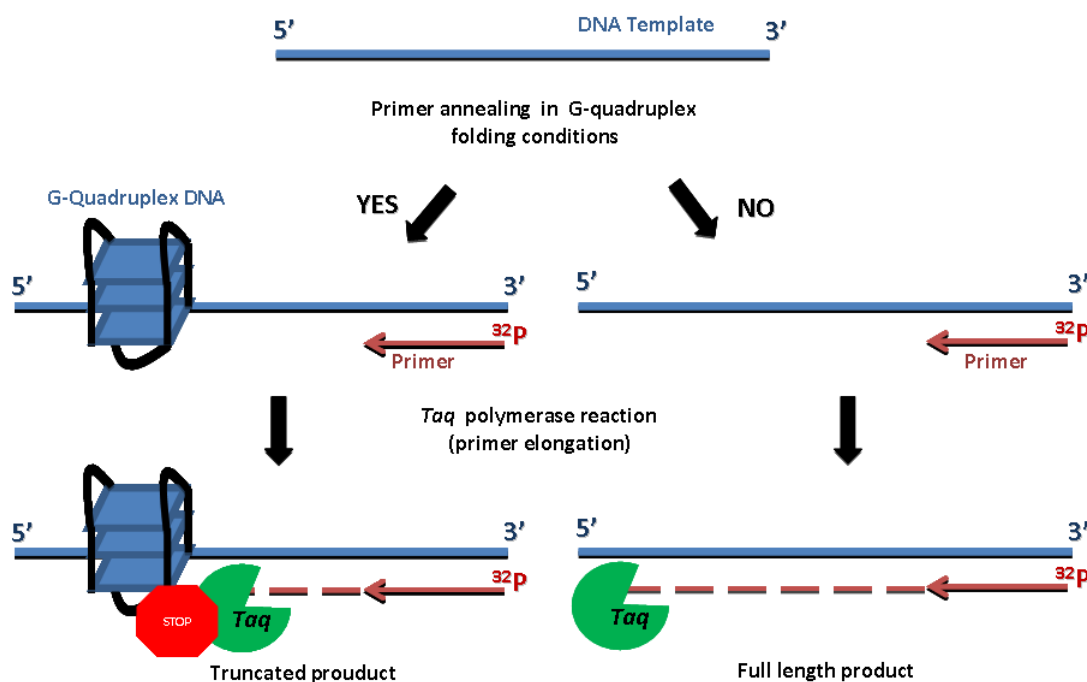


Figure 4.1. *Taq* Polymerase Stop Assay

For the HIV-1 LTR region characterization, primer (LTR G4 *Taq* primer, Table 4.3) was 5'-end labeled with  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  using T4 polynucleotide kinase at  $37^\circ\text{C}$  for 30 min. The labeled primer (72 nM), annealed to the template (36 nM) in lithium cacodylate buffer (10 mM, pH 7.4), was extended with AmpliTaq Gold DNA polymerase (2 U/reaction, Applied Biosystem, California) at  $47^\circ\text{C}$  for 30 min. Sequences of all templates used in this assay were reported in Table 4.3. Where specified, samples were incubated with G-quadruplex ligands and 100 mM

KCl for 20 min at room temperature and primer extension performed as described. Reactions were stopped by ethanol precipitation, extension products were separated on 12% denaturing gel and visualized by phosphorimaging (Typhoon FLA9000, GE Healthcare).

For the HIV-1 *nef* coding region characterization, primers (Table 4.3) were 5'-end-labeled with [ $\gamma$ - $^{32}$ P] ATP using T4 polynucleotide kinase for 30 min at 37°C and purified with Illustra Micro Spin G-25 Column (GE Healthcare, Life Sciences, Milan, Italy). DNA templates (Table 4.3) were diluted from stock to the final concentration (50  $\mu$ M) in lithium cacodylate buffer (10 mM, pH7.4) with 100 mM KCl and then let fold by heating at 95°C for 3 min, gradually cooled to room temperature, and incubated at 4°C overnight. DNA templates were further diluted to a concentration of 1  $\mu$ M and mixed with DNA primer (200 nM), 1X PCR reaction buffer (Applied Biosystems, Carlsbad, California, USA), and 0.1 mM dNTPs. Where appropriate, TMPyP4 was added. AmpliTaq Gold DNA polymerase (1 U/reaction, Applied Biosystem, Carlsbad, California, USA) was then added. Samples were subjected to 30 cycles: 95°C 30 sec, 60°C 30 sec, 72°C 30 sec. Reactions were stopped by ethanol precipitation. Marker lanes were generated on the labelled double stranded PCR product using the Maxam and Gilbert protocol. Briefly, ethanol precipitated PCR products were treated with formic acid (12  $\mu$ l) for 5 min at 20°C. Reactions were stopped by ethanol precipitation. Samples were treated with piperidine 1 M for 30 min at 90°C; reactions were stopped on ice for 5 min. Samples were concentrated in SpeedVac. Markers corresponded to the C-rich complementary strand. The primer extension products and markers were separated on a 12% polyacrylamide denaturing gel and visualized by phosphorimaging (Typhoon FLA 9000, GE Healthcare, Milan, Italy).

Application	Project	Name	Sequence (5'→3')
<i>Taq</i> polymerase stop assay	LTR region characterization	LTR G4 Taq primer	GGCAAAAAGCAGCTGCTTATATGCAG
		LTR G4 LTR I Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGG AGGCGTGGCCTGGGTTTTTCTGCATATAAGCAG CTGCTTTTTGCC
		LTR G4 LTR II Taq	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGG GCGGGTTTTTCTGCATATAAGCAGCTGCTTTTT GCC
		LTR G4 LTR III Taq	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGG GTTTTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL Taq	GCTACAAGGGACTTTCCGCTGGGGACTTTCCA GGGAGGCGTGGCCTGGGCGGGACTGGGGAGTG GCGAGCCCTCAGATCCTGCATATAAGCAGCTG CTTTTTGCC
		LTR G4 FL m1 Taq	TTTTTGTGACTTTCCGCTGGGGACTTTCCAGGG AGGCGTGGCCTGGGCGGGACTGGGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL m2 Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGG AGGCGTTGCCTGGGCGGGACTGGGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL m3 Taq	TTTTTGGGACTTTCCGCTGTGGACTTTCCAGGG AGGCGTGGCCTGGGCGGGACTGGGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL m3' Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGG ATGCGTGGCCTGGGCGGGACTGGGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL m3'' Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGG AGGCGTGGCCTGGGCGGGACTGTGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL m4 Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGG AGGCGTGGCCTGTGCGGGACTGGGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL m5 Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGG AGGCGTGGCCTGGGCGTGACTGGGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL m6 Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGG AGGCGTGGCCTGGGCGGGACTGGGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
		LTR G4 FL m6' Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGTG AGGCGTGGCCTGGGCGGGACTGGGGAGTGGTT TTTCTGCATATAAGCAGCTGCTTTTTGCC
	<i>nef</i> coding region characterization	Nef8528pol	TTGGAGGAGGTGGGTTTTCCAGTCACACACCTC AG
		Nef8624pol	TTGGGGGGACTGGAAGGGTTTTCCAGTCACACA CCTCAG
		Nef8547pol	TTGGTCTTAAAGGTACCTGAGGTCTGACTGGTTT TCGAGACACAGCTCAG
		NefControl1	TTGTCGTCACAGTCTGACTGTTTTCCAGTCACAC ACCTCAG
		NefControl2	TTGTCGTTGAAGATAGCCGTGTAGCTGACGTTTT TCGAGACACAGCTCAG
		DNA primer 1	ATCGATCGCTTCTCGTCTGAGGTGTGTGACTGG
DNA primer 2		ATCGATCGCTTCTCGTCTGAGCTGTGTCTCG	

**Table 4.3** Oligonucleotides used in HIV-1 LTR region characterization and in *nef* coding region characterization for the *Taq* Polymerase Stop Assay



## 4.6. Footprinting assays

The chemical footprinting assay is a useful technique for studying the G-quadruplex structure and in particular to detect guanine bases involved in the G-quadruplex folding. In this assay, a  $^{32}\text{P}$ -labelled template bearing the putative G-quadruplex sequence is incubated in G-quadruplex folding conditions (normally in the presence of cations such as  $\text{K}^+$ ) and subjected to an alkylating agent to alkylate guanine bases. The sample is then treated with piperidine that cleaves the alkylated guanine bases and evaluated by denaturing electrophoresis. The basis of this assay is that guanines involved in the secondary structure's folding are protected from alkylation and create a so called "footprint" in the electrophoretic band pattern (Figure 4.2).

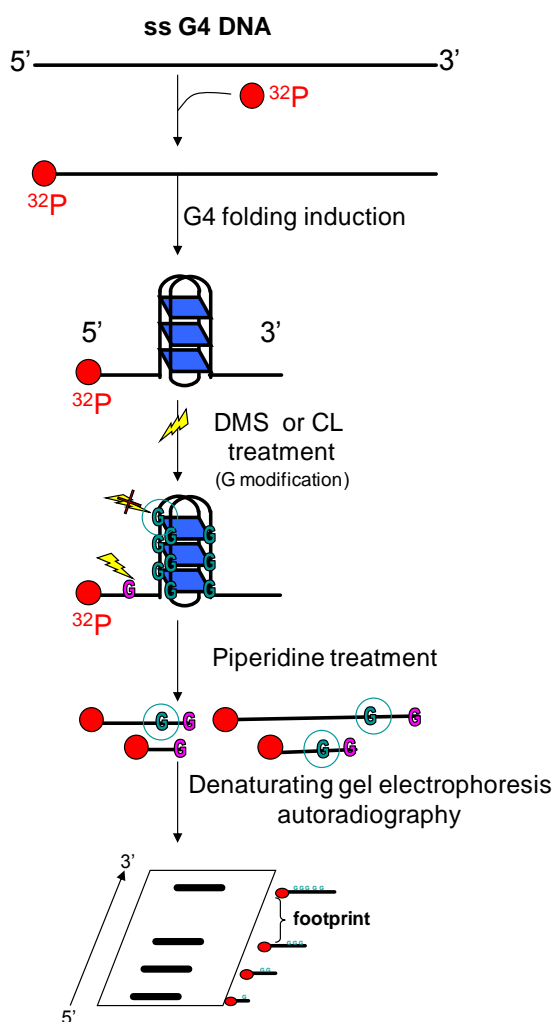


Figure 4.2. Footprinting assay

### 4.6.1. DMS footprinting assay

DMS methylates the N7 position of the guanines (G-N7) both in single-stranded and double-stranded DNA. However, in a G-quadruplex G-N7 is involved in Hoogsteen base pairing and therefore cannot be free for DMS methylation. For this reason, DMS have been widely employed as alkylating agent for characterizing G-quadruplexes in footprinting assay.

All oligonucleotides (Table 4.4) were gel-purified before use and prepared in desalted/lyophilised form. Oligonucleotides were 5'-end-labelled with [ $\gamma$ - $^{32}\text{P}$ ]ATP by T4 polynucleotide kinase and purified by MicroSpin G-25 columns (Amersham Biosciences, Europe). They were next resuspended in lithium cacodylate 10 mM, pH 7.4, with or without KCl 150 mM, heat-denatured and folded. Each  $^{32}\text{P}$ -labeled and annealed DNA was treated with DMS (0.5% in ethanol) for 5 min. Reactions were stopped by adding gel loading buffer containing 10% glycerol and  $\beta$ -mercaptoethanol. Samples were loaded on 16% native polyacrylamide gel and run until the desired resolution was obtained. DNA bands were localized via autoradiography, excised and eluted overnight. The supernatant were recovered, ethanol-precipitated and treated with 1M piperidine at 90°C for 30 min. Samples were dried in a speed vac, washed with water, dried again and resuspended in formamide gel loading buffer. The DMS-treated oligonucleotides were separated on a 20% denaturing gel and visualized by phosphorimaging analysis (Molecular Dynamics Amersham Biosciences).

### 4.6.2. Clerocidin footprinting assay

More recently, published data of our research group proposed Clerocidin (CL) as a new tool to assess G-quadruplex conformation in a footprinting assay (Nadai et al, 2012; Nadai et al, 2013). CL is a natural product that has been shown to react at single-stranded DNA regions, with different mechanisms. In particular, CL targets the G-N7 when a guanine base is exposed. The CL protection assay not only can provide a further confirmation of the data obtained with the DMS protection assay but makes also the footprinting protocol simpler and less time-consuming. In addition, at 25°C CL is a stable non-volatile solid and in aqueous solution hydrolysis rates are slow and products are non-toxic, whereas DMS is an unstable, hygroscopic and volatile liquid which degrades fast in solution generating toxic hydrolysis products. Another great advantage of CL is that it does not react with double-stranded DNA, while DMS is equally reactive towards single and double-stranded DNA: this can allow the detection of a G-quadruplex within a double-stranded context, mimicking the DNA folding state in physiological conditions.

All oligonucleotides (Table 4.4) were gel-purified before use and prepared in desalted/lyophilised form. Oligonucleotides were 5'-end-labelled with [ $\gamma$ - $^{32}$ P]ATP by T4 polynucleotide kinase and purified by MicroSpin G-25 columns (Amersham Biosciences, Europe). They were next resuspended in lithium cacodylate 10 mM, pH 7.4, with or without KCl 150 mM, heat-denatured and folded.

CL reactions with the labelled G-quadruplex folded or unfolded oligonucleotides (4 pmol/sample) were performed at 37°C in annealing buffer (50 mM phosphate buffer, pH 7.4) for 24 h. These conditions were selected to maintain the stability of the target structure and to minimize the possible competition of buffer molecules for CL alkylation. Samples were precipitated with ethanol to eliminate non-reacted drug, resuspended and either kept on ice, or treated at 90°C for 30 min with 1 M piperidine. Samples were then lyophilised, resuspended in formamide gel loading buffer, and heated at 95°C for 2 min. Reaction products were analyzed on 20% denaturing polyacrylamide gels and visualized by phosphorimaging analysis (Molecular Dynamics Amersham Biosciences).

Application	Project	Name	Sequence (5'→3')
DMS or CL footprinting assay	LTR region characterization	LTR FL	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGG GAGGCGTGGCCTGGGCGGGACTGGGGAGTGG TTTTT
		LTR I	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGG GAGGCGTGGCCTGGGTTTTT
		LTR II	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTG GGCGGGTTTTT
		LTR III	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGG GGTTTTT
		LTR II+III	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTG GGCGGGACTGGGGTTTTT

**Table 4.4 Oligonucleotides used in HIV-1 LTR region characterization and in nef coding region characterization for the footprinting assay**

#### 4.7. Reporter assays

The reporter assays are useful to investigate the biological function of G-quadruplexes and are usually performed via transient cell-transfection of plasmids carrying the G-quadruplex sequence of interest fused to a reporter gene, such as the enhanced green fluorescent protein (eGFP) gene. Moreover, it is possible to evaluate the efficacy and selectivity of G-quadruplex ligands *in cellulo* comparing and eventually validating the *in vitro* data. To further confirm the G-quadruplex related effect in these assays, it is useful to employ constructs with point mutations that abrogate G-quadruplex folding as control.

#### 4.7.1. Plasmid construction

For the LTR region characterization, the LTR DNA region was amplified by PCR on the HIV-1 genome (AF033819.3) using reported primers (Table 4.5). The LTR amplicon was subcloned into pGL4.10-Luc2 (Promega) within XhoI and HindIII sites. The resulting pGL4.10-Luc2/LTR vector contained the sequenced 464 bp-long LTR-region (corresponding to nts-381/+83 in the HIV-1 genome) fused to the luciferase coding region. Mutant pGL4.10-Luc2/LTR vectors were generated using Quik-Change mutagenesis kit (Stratagene/Agilent Technologies) and primers (Table 4.5).

For the HIV-1 *nef* coding region characterization, plasmid p-Nef-HA EGFP-N1 containing Nef-HA fused to eGFP was obtained by PCR amplification of the Nef-HA coding sequence (strain HXB2/LAI; NC\_001802) in plasmid pNefHABJ5 (kindly donated by Prof. M. Pizzato, Centre for Integrative Biology, University of Trento, Trento, Italy). PCR was performed using primers prFNef and prRNef (Table 4.5), which introduced NheI and EcoRI restriction sites for subsequent insertion in the pEGFP-N1 vector (Clontech, Mountain View, CA, USA). The obtained coding sequence of the fused Nef-HA GFP protein was confirmed by sequencing.

Application	Project	Name	Sequence (5'→3')	Vector
LTR cloning for reporter assays	LTR region characterization	LTR-Xho I	GGGCCCTCGAGCCCTGATTGGCAGAAY TACACACCAGG	pGL4.10-Luc2/LTR-wt
		LTR-Hind III	GGGCCCAAGCTTCCTGCGTCGAGAGAGC TYCTCTGG	
LTR mutants cloning for reporter assays	LTR region characterization	pr m4a	CCAGGGAGGCGTGGCCTGTGCGGGACTG GGGAGTGGCG	pGL4.10-Luc2/LTR-m4
		pr m4b	CGCCACTCCCCAGTCCCGCACAGGCCACG CCTCCCTGG	
		pr m5a	GGGAGGCGTGGCCTGGGCGTACTGGGG AGTGGCGAGC	pGL4.10-Luc2/LTR-m5
		pr m5b	GCTCGCCACTCCCCAGTCACGCCAGGCC ACGCTCCC	
		pr m4+5a	CAGGGAGGCGTGGCCTGTGCGTACTGG GGAGTGGCGAGC	pGL4.10-Luc2/LTR-m4+5
		pr m4+5b	GCTCGCCACTCCCCAGTCACGCACAGGCC ACGCTCCCTG	
		pr m3''a	GACTTTCCAGGGAGGCGTAGCCTGGGCG GGACTGGGG	pGL4.10-Luc2/LTR-m3''
		pr m3''a	CCCCAGTCCCGCCAGGCTACGCCTCCCT GGAAAGTC	
Nef cloning for reporter assays	<i>nef</i> coding region characterization	prFNef	TAAGCTAGCACGCGTCATGGGTGGCAAG TGG	p-Nef-HA-EGFP-N1
		prRNef	ACTGAATTCTAGCGTAATCTGGGACGTC	pNef-HA-EGFP-N1

**Table 4.5 Oligonucleotides used in HIV-1 LTR region characterization and in *nef* coding region characterization for the Reporter Assays**

#### 4.7.2. Dual-Luciferase reporter assay

This assay is particularly useful to investigate the effect of G-quadruplex folding at the promoter level. The promoter bearing the G-quadruplex sequence of interest is cloned in a construct where it drives the expression of the Firefly luciferase gene. The latter construct is co-transfected with a control plasmid containing another promoter (e.g. Herpes simplex virus thymidine kinase promoter) that drives the Renilla luciferase gene expression. Measuring the luminescence signals of the two luciferase, it is possible to normalize the expression of an experimental reporter to the expression of the control reporter and it is essential for differentiating between specific and not-specific cellular responses. Importantly, this normalization can provide information on transfection efficiencies and allow to have more precise data.

Vectors pGL4.10Luc2/LTR and pGL4.74 (200 ng each) were transfected in  $5.5 \times 10^4$  HEK 293T cells per well onto 96-well plates, using TransIT-293 transfection reagent (Mirus, Madison, WI, USA). Plasmid pGL4.74-hRLUC/TK (Promega), containing a Renilla luciferase gene driven by the Herpes Simplex virus thymidine kinase promoter, was used as a control for transfection efficiency. Expression of firefly luciferase, with respect to that of Renilla luciferase, was determined 24 h after transfection using the Dual-Glo luciferase assay system (Promega). Cell lysate (75  $\mu$ L) was mixed with reconstituted Dual-Glo luciferase buffer (75  $\mu$ L) or Dual-Glo Stop&Glo buffer, and light output detected with VICTOR X2 multilabel plate reader (Perkin-Elmer). Luciferase and Renilla output ratio was calculated.

#### 4.7.3. eGFP reporter assays

This assay is useful to investigate G-quadruplexes at different levels, both in promoter and in coding regions, following the eGFP expression by flow cytometry.

For the LTR-GFP reporter assay,  $2 \times 10^5$  of HEK 293T cells in 6-well plates were transfected with pcLTR-EGFP DNA 3.1 (kindly provided by Prof. A. Loregian, University of Padua). After 4 h, cells were treated with BRACO-19 or TMPyP2 and incubated overnight. After trypsinization, cells were washed, resuspended in 500  $\mu$ L of PBS 1X and analyzed by flow cytometry (see flow cytometry analysis below).

For the Nef-eGFP reporter assay,  $8 \times 10^4$  of HEK 293T cells were seeded in a 12-well plate in 1 ml of DMEM/10% FBS medium and incubated for 24 h. Cells were next transfected with pNef- HA EGFP-N1 by *TransIT-293* Transfection Reagent (Mirus, Madison, WI, USA). After 4 h, cells were treated with TMPyP4 or TMPyP2 (10  $\mu$ M) and incubated overnight.

After trypsinization, cells were washed with PBS, resuspended in 500  $\mu$ l of PBS 1X and analyzed by flow cytometry (see flow cytometry analysis below).

### **4.7.3.1. Flow Cytometry analysis**

To evaluate mean of eGFP fluorescence, a total of 30000 events were acquired for each sample with an LRS 2 instrument using FACS DIVA software (BD Bioscience) and analyzed with Flow Jo (Tree Star, OR, USA). Cell debris was excluded from the analysis by gating on forward versus side scatter dot plots

## **4.8. Antiviral assays**

### **4.8.1. Viruses**

The HIV-1 NL4.3 (wild type) and HIV-1  $\Delta$ Nef NL4.3 (Nef defective) were kindly prepared and handled by Dr. J. Poe and Prof. T.E. Smithgall (University of Pittsburg School of Medicine, Pennsylvania, USA).

The HIV-1(III<sub>B</sub>) strain was originally provided by Prof. R.C. Gallo and Dr. M. Popovic (at that time at the NIH, Bethesda, MD, USA). HIV-1(BaL) was originally provided by R.C. Gallo (at that time at the NIH, Bethesda, MD, USA). HIV-2(ROD) was originally obtained from L. Montagnier (at that time at the Pasteur Institute, Paris, France) and SIV(Mac251) from C. Bruck (Smith Kline-RIT, Rixensart, Belgium).

### **4.8.2. Antiviral Assay in HIV-1 infected TZM-bl cell line**

TZM-bl is an engineered cell line which contains the Luciferase reporter gene under the control of the HIV-1 LTR promoter. In this cell line, the HIV infection drives transcription of the HIV-1 LTR-Luciferase reporter gene construct and thus it is possible to evaluate the antiviral effect of test compounds following the Luciferase expression. Since the HIV-1 replication is enhanced by Nef in this cell line (Emert-Sedlak et al, 2013), TZM-bl cells represent a good system to evaluate specifically Nef-related effect during viral replication. For this reason it's essential to work also with a Nef-defective ( $\Delta$ Nef) virus as control. A similar antiviral effect both against wild-type and  $\Delta$ Nef viruses indicate an effect not Nef-related.

For the LTR region characterization, TZM-bl cells were seeded in 96-well plates ( $2 \times 10^4$ ) and grown overnight. BRACO-19 and TMPyP2 were preincubated separately with both cell culture medium (100  $\mu$ L) and wild-type HIV-1 or  $\Delta$ Nef-HIV-1 (100  $\mu$ L) for 4 h prior to infection in a combined final volume of 200  $\mu$ L. After 48 h at 37°C, cells were washed with

PBS and lysedin luciferase lysis buffer (Promega). Lysates (40  $\mu$ L) with 50  $\mu$ L of luciferase reagent (Promega) were read with a delay time of 2 s and an integration period of 10 s. Cytotoxicity of test compounds was assessed in parallel using the Cell Titer Blue reagent (Promega).

For the HIV-1 *nef* coding region characterization, TZM-bl cells were seeded in 96-well plates ( $2 \times 10^4$ ) and grown overnight to permit adherence prior to treatment and viral infection. TMPyP4 was solubilized in DMSO and preincubated separately with both the cell culture medium (100  $\mu$ L) and wild-type HIV-1 or  $\Delta$ Nef HIV-1 (100  $\mu$ L) for 4 h prior to infection in a combined final volume of 200  $\mu$ L. After 48 h at 37C, the cells were washed with PBS and lysed in luciferase lysis buffer (Promega) by rocking for 15 min. Lysates (40  $\mu$ L) were transferred to white 96-well plates, and 50  $\mu$ L luciferase reagent (Promega) was injected into each well. Readings were recorded with a delay time of 2 s and an integration period of 10 s. Cytotoxicity of test compounds was assessed in parallel using the Cell Titer Blue reagent (Promega).

#### **4.8.3. MTT-based antiviral assay in HIV infected MT-4 cell line**

The evaluation of antiviral effect of the compounds against HIV-1 strain III<sub>B</sub>, HIV-2 strain ROD and SIV strain mac251 in MT-4 cells was performed using a tetrazolium-based colorimetric assay (Pannecouque et al, 2008). Briefly, this method is based on HIV-induced cytopathogenic effect (CPE) on MT-4 cells 5 days after infection. The antiviral effects of test compounds is directly correlated to the inhibition of viral induced CPE and can be measured spectrophotometrically following the reduction of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) to a blue formazan product in metabolically active cells. Since this assay covers the complete viral cycle, it is possible to discover direct inhibitors of HIV replication. Mock-infected cells are used in parallel to assess the cytotoxicity of test compounds. Test compounds (25  $\mu$ l/well of 10X final concentration) were added to two series (to evaluate the effect on both HIV and mock-infected cells) of triplicate wells in a 96-well plate and 5-fold diluted using a Biomek 3000 robot (Beckman instrument, Fullertone, CA). For each compound 9 serial 5 fold dilutions were made and untreated HIV and mock samples were also included as controls. Virus stock, 50 $\mu$ l of HIV-1(III<sub>B</sub>), HIV-2(ROD) and SIV(mac251) at 100-300 CCID<sub>50</sub> (50% cell culture infectious doses, corresponding to a multiplicity of infection (MOI) of 0.003) or complete medium were added. Finally, 50  $\mu$ l per well of MT-4 cells ( $6 \times 10^5$  cells/ml) were added. 5 days after infection, the viability of HIV

and mock-infected cells was evaluated using the MTT assay. The absorbances were measured in an eight-channel computer-controlled photometer (Infinite M1000, Tecan, Mechelen, Belgium), at two wavelengths (540 and 690 nm) and the median absorbance value of three wells was calculated. The 50% cytotoxic concentration ( $CC_{50}$ ) was defined as the concentration of test compound that was able to reduce the absorbance of the mock-infected cells by 50%. The 50% effective concentration ( $EC_{50}$ ) was defined as the concentration of test compound that protect 50% of the HIV-infected cells from CPE. HIV-1 production in selected samples was also evaluated by measuring HIV-1 core antigen (p24 antigen) in the supernatant with a p24 antigen enzyme-linked immunosorbent assay (Perkin-Elmer, Brussels, Belgium). The antiviral activity of BRACO-19 against HIV-1(III<sub>B</sub>) in MT-4 cells was tested also 1 day post infection. The antiviral activity of BRACO-19 against HIV-1 was based on the inhibition of HIV-1 p24 antigen production in MT-4 infected cells. BRACO-19 (22  $\mu$ l/well of 10X final concentration) was added to one series of duplicate wells in a 96-well plate and 9 serial 5 fold dilutions were made. MT-4 cells were infected with HIV-1(III<sub>B</sub>) at MOI of 0.5 and incubated at 37°C. After 1h cells were washed 3 times in PBS 1X and seeded (100,000 cells/well) in a 96-well plate. At 31 h post infection, supernatants were collected and p24 antigen was measured with a p24 antigen enzyme-linked immunosorbent assay (Perkin-Elmer, Brussels, Belgium). Cytotoxicity of BRACO-19 on MT-4 cells was tested in parallel using an MTT assay.

#### **4.8.4. Antiviral assay in HIV-1 infected MT-4/LTR-eGFP cell line**

To further confirm the antiviral activity of BRACO-19 at 1 day post infection, another cellular system was used. MT-4-LTR-eGFP is an engineered cell line which contains the eGFP reporter gene under the control of the HIV-1 LTR promoter. In this cell line, the HIV infection drives transcription of the HIV-1 LTR-eGFP reporter gene construct and thus it is possible to evaluate the antiviral effect of test compounds following the eGFP expression. This experiment is useful to have a rapid indication of anti-HIV activity that could eventually be confirmed measuring directly the p24 antigen in the supernatant. BRACO-19 (22  $\mu$ l/well of 10X final concentration) was added to one series of duplicate wells in a 96-well plate and 9 serial 5 fold dilutions were made. MT-4-LTR-eGFP cells were infected with HIV-1(III<sub>B</sub>) at MOI of 0.5 and incubated at 37°C. After 1 h cells were washed 3 times in PBS 1X and seeded (100,000cells/well) in the 96-well plate. 31 h post infection, supernatants were collected and cells were fixed in 3% aqueous paraformaldehyde. HIV-1 production was determined



following the GFP-expression by using flow cytometry (see flow cytometry analysis below). In addition, p24 antigen was measured in the supernatant to further confirm the data (p24 antigen enzyme-linked immunosorbent assay; Perkin-Elmer, Brussels, Belgium). Cytotoxicity of BRACO-19 on MT-4-LTR-eGFP cells was tested in parallel using an MTT assay.

#### **4.8.4.1. Flow Cytometry analysis**

Flow cytometric analysis was performed on a FACS CantoII flow cytometer (Becton Dickinson, San Jose, CA, USA). Before acquisition, cells were pelleted at 1000 rpm for 10 min and fixed in 3% paraformaldehyde solution. Acquisition was stopped when 10,000 events were counted. Data analysis was carried out with FACS Diva Software (Becton Dickinson). Cell debris was excluded from the analysis by gating on forward versus side scatter dot plots

#### **4.8.5. Antiviral assay in HIV-1 infected PBMCs**

The antiviral activity of BRACO-19 was further investigated in stimulated PBMCs against two HIV-1 strains: HIV-1 strain III<sub>B</sub> and HIV-1 strain BaL. These two viruses use two different chemokine-receptors for entry: HIV(III<sub>B</sub>) is a X4 virus that uses the alpha-chemokine receptor CXCR4 while HIV(BaL) uses the beta-chemokine receptor CCR5 for entry.  $2 \times 10^5$  cells were seeded in the presence of serial dilutions of the test compound and were infected with HIV-1(III<sub>B</sub>) or with HIV-1(BaL) at 1000 CCID<sub>50</sub> per ml. At 4 days post infection, 125  $\mu$ l of cell suspension of the infected cultures were removed and replaced with 150  $\mu$ l of fresh medium containing the test compound at the appropriate concentration. After 7 days, supernatants were collected and p24 antigen was detected by enzyme-linked immunosorbent assay (Perkin-Elmer, Brussels, Belgium).

#### **4.8.6. Antiviral effect of test compounds in persistently HIV-1 infected cells (MT-4/III<sub>B</sub> cell line)**

The antiviral activities of test compounds against persistent HIV-1 infection were based on the inhibition of p24 antigen production in persistently-infected MT-4/III<sub>B</sub> cells untreated or pretreated with AZT (10 ng/ml). The pretreatment with AZT, a NRTI, allows to evaluate the effect of the compound only in post integration viral steps. AMD3100 (a CXCR4 Antagonist) and Ritonavir (a PI) were tested in parallel as reference drugs. BRACO-19 (22  $\mu$ l/well of 10X final concentration) was added to one series of triplicate wells in a 96-well plate and 9 serial 5 fold dilutions were made. Finally, untreated or pretreated MT-4/III<sub>B</sub> cells ( $3 \times 10^4$  cells/well)

were added in a final volume of 200  $\mu$ l/well. After 5 days, supernatants were collected and HIV-1 production was determined by measuring p24 antigen with a p24 antigen enzyme-linked immunosorbent assay (Perkin-Elmer, Brussels, Belgium). Cytotoxicity of BRACO-19 on MT-4/III<sub>B</sub> cells was tested in parallel using an MTT assay.

### **4.9. Viral binding assay**

This assay allows to investigate the inhibitory effect of test compounds on viral binding to MT-4 cells. MT-4 cells ( $5 \times 10^5$ /tube in 100  $\mu$ l) were incubated with 4 serial 5-fold dilutions of test compounds (100  $\mu$ l/tube of 4X final concentration) in 5ml tubes. Finally, 200  $\mu$ l of HIV-1(III<sub>B</sub>) dilutions (corresponding to 100 ng of p24) were added to each tube. Dextran Sulfate (DS) and AMD3100 were tested in parallel as reference compounds. After 2 h at 37°C, the cells were washed three times with PBS 1X to remove unbounded viral particles and then lysed with PBS 1X containing 0.5% tergitol NP-40 (Sigma, St-Louis, MO). The amount of p24 antigen was measured by a p24 antigen enzyme-linked immunosorbent assay (Perkin-Elmer, Brussels, Belgium).

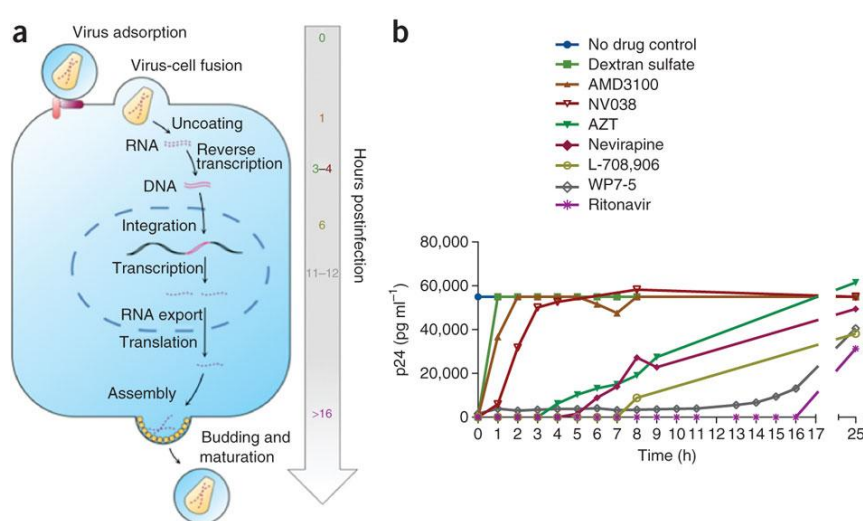
### **4.10. Virucidal assay**

To investigate the antiviral properties of BRACO-19 against HIV-1-virions' structural parts, a virucidal assay was performed. If the compound has virucidal effect against HIV-1 virions, a compound-pretreated and subsequently compound-cleared virus stock has a minor efficacy in infecting MT-4 cells that could be evaluated by titration. Aliquots of HIV-1(III<sub>B</sub>) stock were incubated with various concentrations of test compounds in 100  $\mu$ l of complete RPMI-1640 medium for 1h at 37°C. The samples were next diluted 4000 times with complete medium to reach a concentration of compound so far below the IC<sub>50</sub> and used to infect MT-4 cells. Five days after infection, viral infectivity was quantified by titration (Daelemans et al, 2011).

### **4.11. Time-of-Addition experiments**

The Time-of-Addition (TOA) assay allows to determine how long the addition of a compound can be postponed before it loses its antiviral activity in cells and was used to investigate the target of test compounds as HIV inhibitors (Daelemans et al, 2011). Since the HIV replication cycle proceed in a well-established chronological order (Fig. 4.3A), it is possible to investigate the target of a test compound by comparing its action to that of reference drugs in the time scale (Fig.4.3B). The choice of reference drugs is essential: well-characterized anti-

HIV drugs with a high selectivity are preferred, since 10 to 100-fold their 50% inhibitory concentration ( $IC_{50}$ ) is used. The anti-HIV-1 reference drugs chosen for this assay could cover almost every step of viral cycle: Dextran Sulfate (DS) is an inhibitor of viral adsorption, AMD3100 is a CXCR4 co-receptor antagonist, NV038 is an NC inhibitor, AZT is a NRTI, Nevirapine is a NNRTI, L-708,906 is a strand transfer integrase inhibitor, WP7-5 is a transcription inhibitor and finally Ritonavir acts as PI. For all the compounds mentioned above, HIV-1 replication is blocked up to a time point corresponding to the occurrence of the replication process targeted by the drug.



**Figure 4.3 Time-Of-Addition** A) Chronological representation of the essential steps in the HIV-1 viral replication cycle. B) Example of typical results obtained with the well-characterized anti-HIV drugs

MT-4 cells were infected with HIV-1(III<sub>B</sub>) at the MOI of 0.5. After 1 h, cells were washed 3 times with complete RPMI 1640 medium and seeded into a 96-well plate (100,000 cells/well) and incubated at 37°C. Test compounds and reference compounds were added at different times (0,1,2,3,4,5,6,7,8,24 and 25 h) post infection. The reference compounds DS, AMD3100, AZT, Nevirapine and Ritonavir were added at a concentration of 100-fold their  $IC_{50}$  ( $IC_{50}$  required to reduce by 50% the CPE of HIV-1(III<sub>B</sub>) in MT-4 cells). HIV-1 production was determined 31 h post infection by measuring HIV-1 core antigen (p24 antigen) in the supernatant with a p24 antigen enzyme-linked immunosorbent assay (Perkin-Elmer, Brussels, Belgium).



## 5. Results and Discussion

### 5.1. G-Quadruplex analysis of the HIV-1 Proviral Genome

#### 5.1.1. Computational analysis of the HIV-1 proviral genome for the presence of putative G-Quadruplex forming regions

The HIV-1 proviral genome (strain HXB2/LAI, NC\_001802) was analysed with QGRS Mapper which is an online algorithms-based software program for recognition and mapping of putative Quadruplex forming G-Rich Sequences (QGRS) (Kikin et al, 2006). The found QGRS were ranked based on the G-score, which is the likelihood to form a stable G-quadruplex. Since the single-stranded RNA viral genome is retrotranscribed into double-stranded DNA and inserted into the human genome, analysis for QGRS was performed also on the antisense strand. In fact, as anticipated in the introduction, transcription and/or replication can be perturbed by G-quadruplexes located on both strands. It is important to underline that this QGRS analysis represented only a preliminary screening of the HIV-1 proviral genome, since complex sequences with multiple runs of Gs can form several G-quadruplex structures in equilibrium that are not necessarily predicted by the software.

The distribution of QGRS along the HIV-1 is represented in Figure 5.1. Overall, forty-one QGRS were found: the sequences displayed a G-score from 10 to 33 and appeared uniformly distributed in all the proviral genome. Interestingly, thirty-two sequences were located in the sense strand while only nine sequences were located in the antisense strand. This preliminary analysis showed an asymmetry in QGRS distribution between the two strands, with a strong prevalence of QGRS in the sense strand. Only nine sequences showed a relatively high G-score ( $\geq 20$ ) and were considered for further characterizations, taking into account also the relevance of the location in the HIV-1 genome. Since transcription can be perturbed by G-quadruplexes located both in promoter and in coding regions, we focused on putative G-quadruplex forming sequences in the Long Terminal Repeats (LTR) promoter region and in HIV-1 coding regions. Interestingly, one of the sequences that displayed the highest G-score (=32) was located in the U3 region of LTR, within the viral promoter. Thus, G-quadruplex formation in this region could affect regulation of transcription and consequently the viral cycle. We decided to further investigate this putative G-quadruplex region in the HIV-1 LTR promoter: results are presented in paragraph 5.2. In addition, we noticed that four putative

QGRS were located in the *nef* gene. If these latter sequences were able to fold in G-quadruplex, this region of the HIV-1 proviral genome could perturb polymerase processing. Our results about this putative G-quadruplex forming region in the *nef* coding sequence are described in paragraph 5.3.

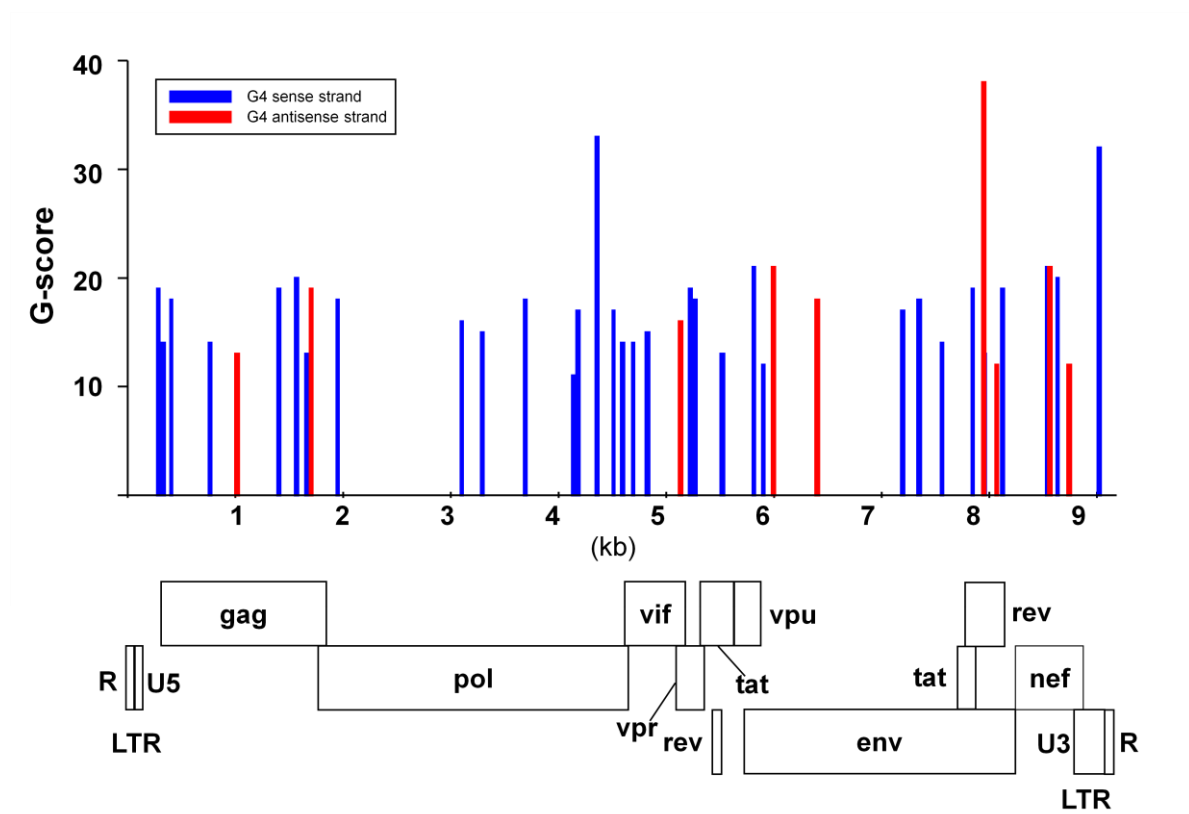
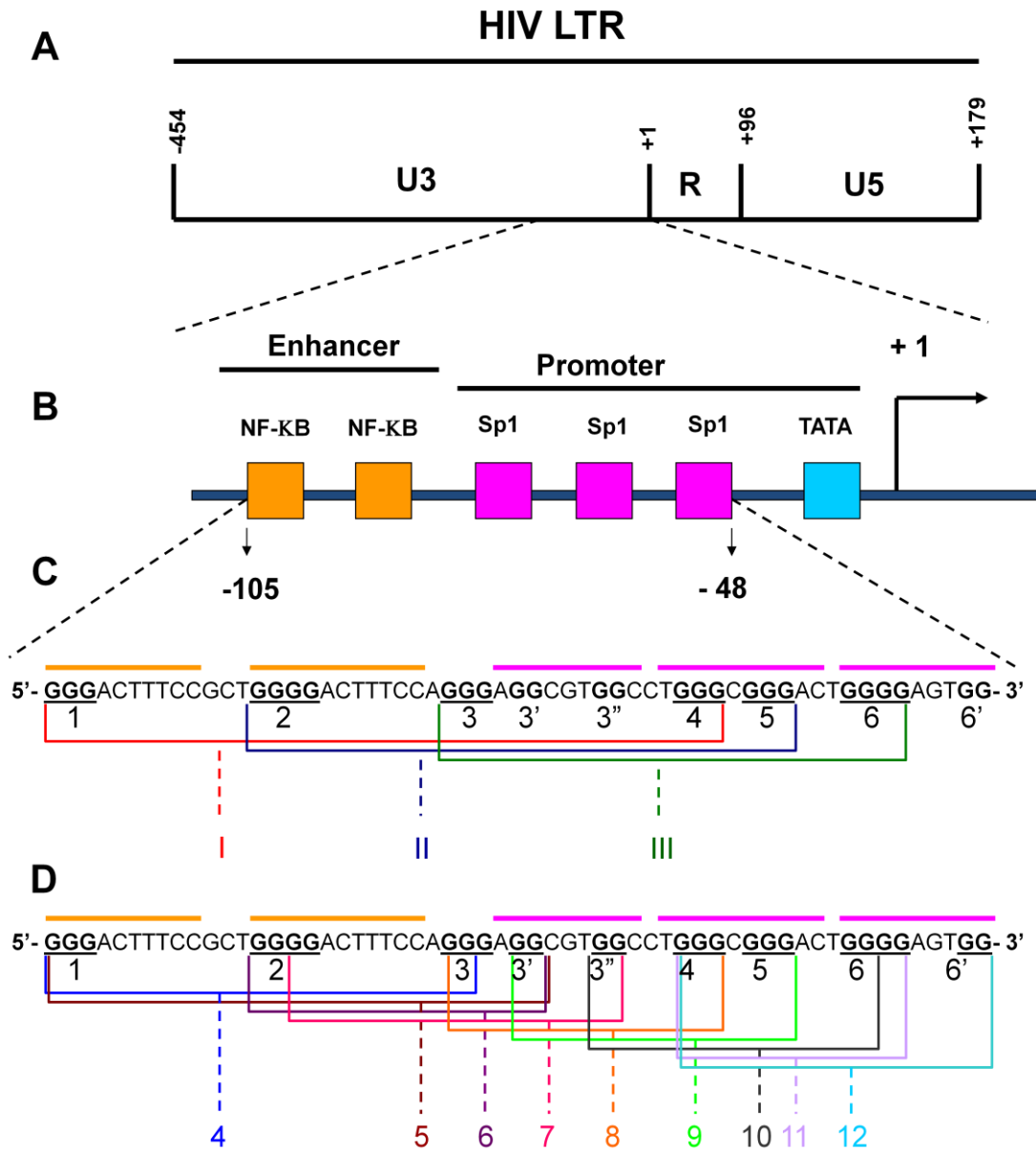


Figure 5.1 Putative Quadruplex forming G-Rich Sequences (QGRS) in the HIV-1 proviral genome

## **5.2.G-Quadruplex structures in the HIV-1 Long Terminal Repeat (LTR) Promoter**

### **5.2.1. Analysis of putative G-quadruplex sequences in the HIV-1 LTR promoter region**

In the proviral HIV-1 DNA, two identical copies of the Long Terminal Repeat (LTR) are located at the 5' and 3' end of the genome. Each LTR is composed of the segment U3, R and U5 (Figure 5.2A). The viral promoter is located in the U3 region of the 5' LTR and requires activation by host transcription factors. The U3 region contains an enhancer (position -105 to -79, with respect to the transcription initiation site) with two binding sites for the nuclear factor  $\kappa$ B (NF- $\kappa$ B), and the core promoter (position -78 to -1, with respect to the transcription initiation site) composed of three tandem binding sites for Specificity protein 1 (Sp1) and a TATA box (Figure 5.2B). This region appears particularly G-rich: the segment corresponding to part of the core promoter and enhancer consists of 50% G and 70% GC. Interestingly, Gs were mainly clustered in groups of 2–4 continuous G bases (Figure 5.2C). Because of these features, the possibility of G-quadruplex folding was further analyzed. Besides the sequence found in the preliminary QGRS analysis (Figure 5.1), a total of twelve putative overlapping sequences were found and were ranked based on G-score (Table 5.1): three sequences, namely LTR-I, -II, and -III, were composed of GGG repeats (numbered 1–6, Figure 5.2C) and could thus generate G-quadruplexes with three stacked G-quartets exhibiting the highest G-scores. The remaining nine sequences (Figure 5.2D) involved two stacked quartets and were ranked with lower G scores. These sequences differed in the length of the loop segments connecting G-tracts (Table 5.1). Interestingly, the G-rich -105/-48 tract was also the binding region of two important cellular transcription factors that stimulate viral transcription: NF- $\kappa$ B, with two binding sites at -105/-96 and -92/-82, and Sp1, three molecules of which bind at -79/-68, -67/-57, and -56/-4817 (Figure 5.2B).



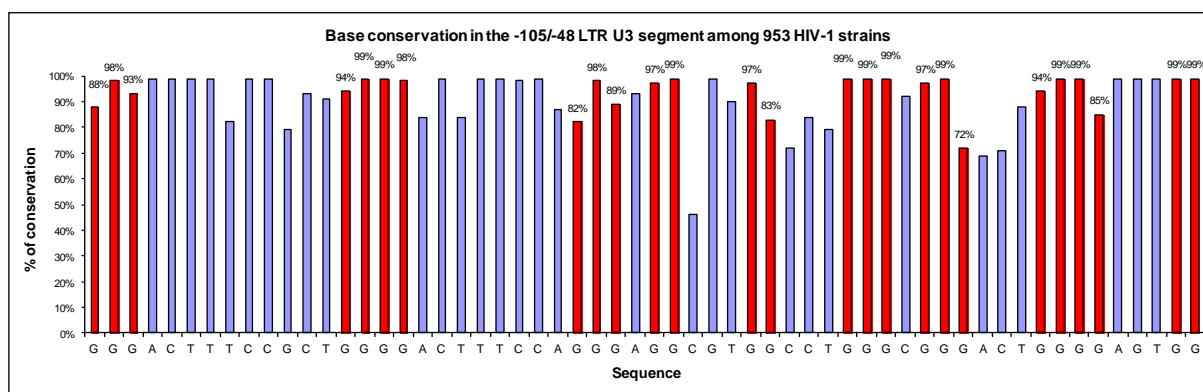
**Figure 5.2 Architecture of the G-rich HIV-1 LTR region spanning -105/-48 nucleotides** A) Scheme representing the U3, R and U5 regions and their positions in the HIV-1 LTR sequence. B) The NF-κB (orange) and Sp1 (magenta) binding sites and the TATA box (blue) are highlighted. C-D) Sequence of the G-rich LTR region spanning -105/-48 nts. The orange and magenta lines above the sequence indicate the bases involved in transcription factor binding. G bases that may be involved in G-quadruplex are shown in bold. G-tracts are consecutively numbered (1-6 for GGG- or GGGG-tracts and 3', 3'', 6' for GG-tracts). The sequences that may form G-quadruplexes with three stacked tetrads are denoted in roman numbers with brackets (C) The sequences that may form G-quadruplexes with two stacked G-quartets are indicated by Arabic numbers and differently colored brackets (D).



Sequence name	# G-quartets	# linker nucleotides (min-max)			G-score
		loop I	loop II	loop III	
LTR-I	3	10	8	11	69-70
LTR-II	3	8-9	11	1	62-68
LTR-III	3	11	1	3-4	62
LTR-4	2	10-11	0	8-9	10-11
LTR-5	2	9-11	7-9	1-2	13
LTR-6	2	0	8-9	12	12-13
LTR-7	2	8-11	1-2	3	15
LTR-8	2	1-2	3	3-4	18-20
LTR-9	2	3	3-4	1-3	18-21
LTR-10	2	3	1-3	3-4	13-21
LTR-11	2	1-3	3-4	0	17-18
LTR-12 (or LTR IV)	2	1-3	3-5	3-4	13-20

**Table 5.1 Properties of G-quadruplex forming sequences in the -105/-48 HIV-1 LTR U3 region**

To establish the importance of these predicted sequences within the viral context, the degree of base conservation among HIV-1 strains was assessed. 953 LTR sequences of different HIV-1 strains were analyzed. In particular, 24 LTR sequences belonged to strains of subtypes A, 485 of subtypes B, 119 of subtype C and 325 of others strains. Among all the analyzed LTR sequences, G bases were highly conserved, reaching a conservation grade up to 99%. Interestingly, also most non-G bases were highly conserved (Figure 5.3). This evidence suggested a key role of this highly conserved viral site.



**Figure 5.3 Base conservation among 953 HIV-1 strains** G-bases are shown in red, non-G-bases are shown in lilac. Percentages of conservation of G-bases are shown above corresponding bars.

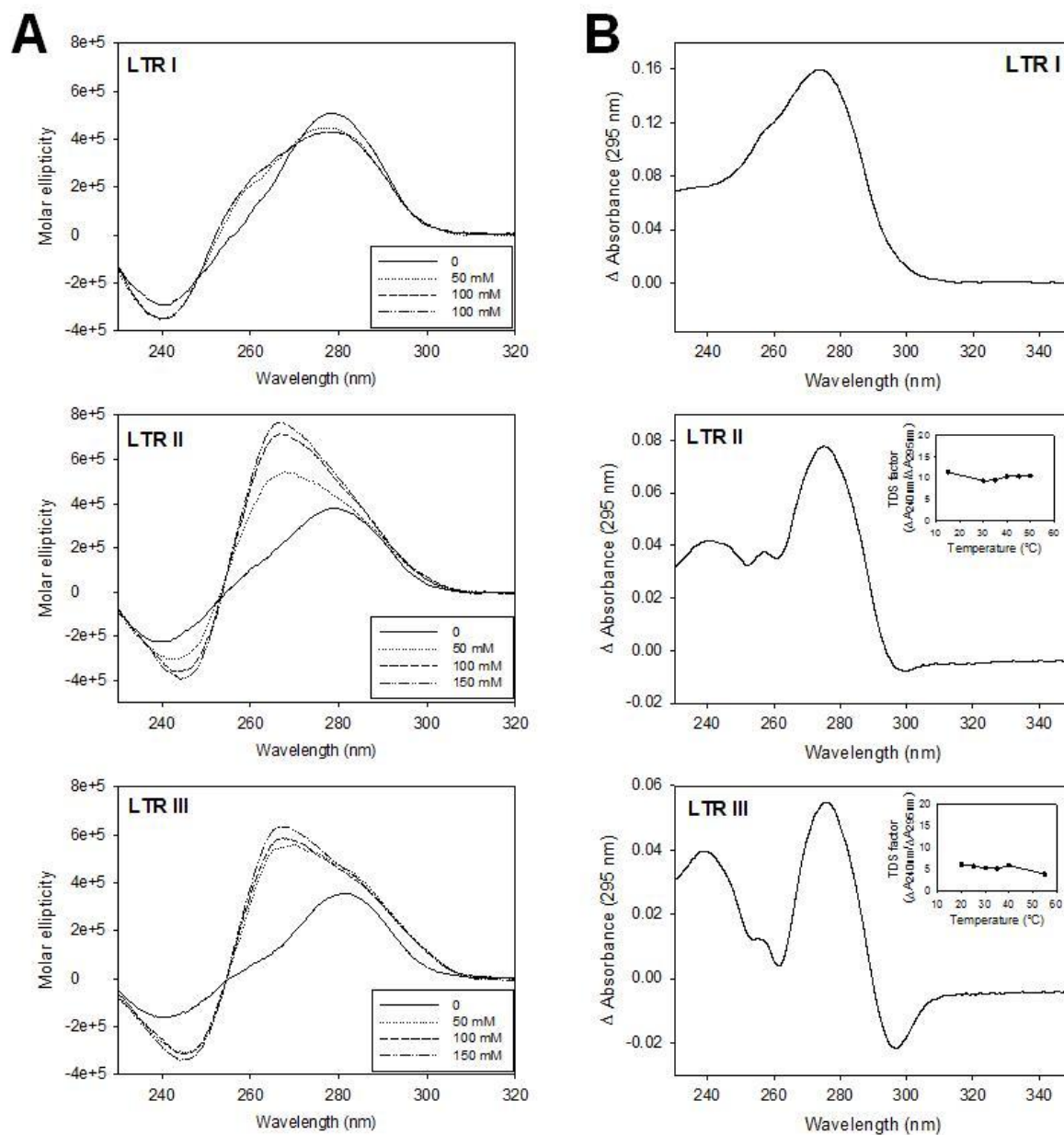
### **5.2.2. Characterization of putative G-quadruplex forming sequences LTR I, II and III**

The three LTR sequences, namely LTR-I, -II, and -III, embedding four GGG tracts each (Figure 5.2C) and displaying higher G-scores (Table 5.1), were selected and analyzed for their ability to form G-quadruplexes.

#### **5.2.2.1. Spectroscopic analysis**

Circular dichroism (CD) spectroscopy was initially performed in the absence or presence of increasing concentrations of  $K^+$  to monitor G-quadruplex formation and its likely topology. LTRII and LTR-III produced CD spectra characteristic of a G-quadruplex structure in a  $K^+$ -dependent manner (Figure 5.4A). Spectra of both oligonucleotides exhibited a positive band at 266 nm and a negative peak at 244 nm, a signature suggesting a parallel-like G-quadruplex. In contrast, LTR-I showed a negative band at 240 nm, a shoulder at 260 nm, and a positive peak at 280 nm, and low  $K^+$ -dependence, features not typical of G-quadruplex structures (Figure 5.4A).

The topology of the selected sequences was further assessed by UV thermal difference spectroscopy (TDS). LTR-II and LTR-III produced TDS signatures with three positive bands at 275, 257 and 239 nm, and three negative bands at 296, 261 and 252 nm, characteristic of the G-quadruplex structure (Figure 5.4B). In addition, the TSD factor, calculated as  $\Delta A_{240nm}/\Delta A_{295nm}$  was above 5, indicating a parallel-like topology (Figure 5.4B; inset). In contrast, the TDS signature of LTR-I was not representative of G-quadruplex (Figure 5.4B). Overall, spectroscopic analysis strongly supported a parallel-like G-quadruplex folding only for LTR II and LTR III sequences. The two independent analysis excluded a G-quadruplex formation for LTR I sequence.



**Figure 5.4 Spectroscopic data of LTR I, LTR II and LTR III putative G-quadruplex forming sequences** A) CD spectra of each oligonucleotide in the presence of increasing concentration of  $K^+$  (0-150 mM). Only LTR II and LTR III present G-quadruplex characteristic spectra. B) TDS spectra of each oligonucleotide. Only LTR II and LTR III exhibit the G-quadruplex TDS signature. Where available, TDS factors graphs are shown in the insets.

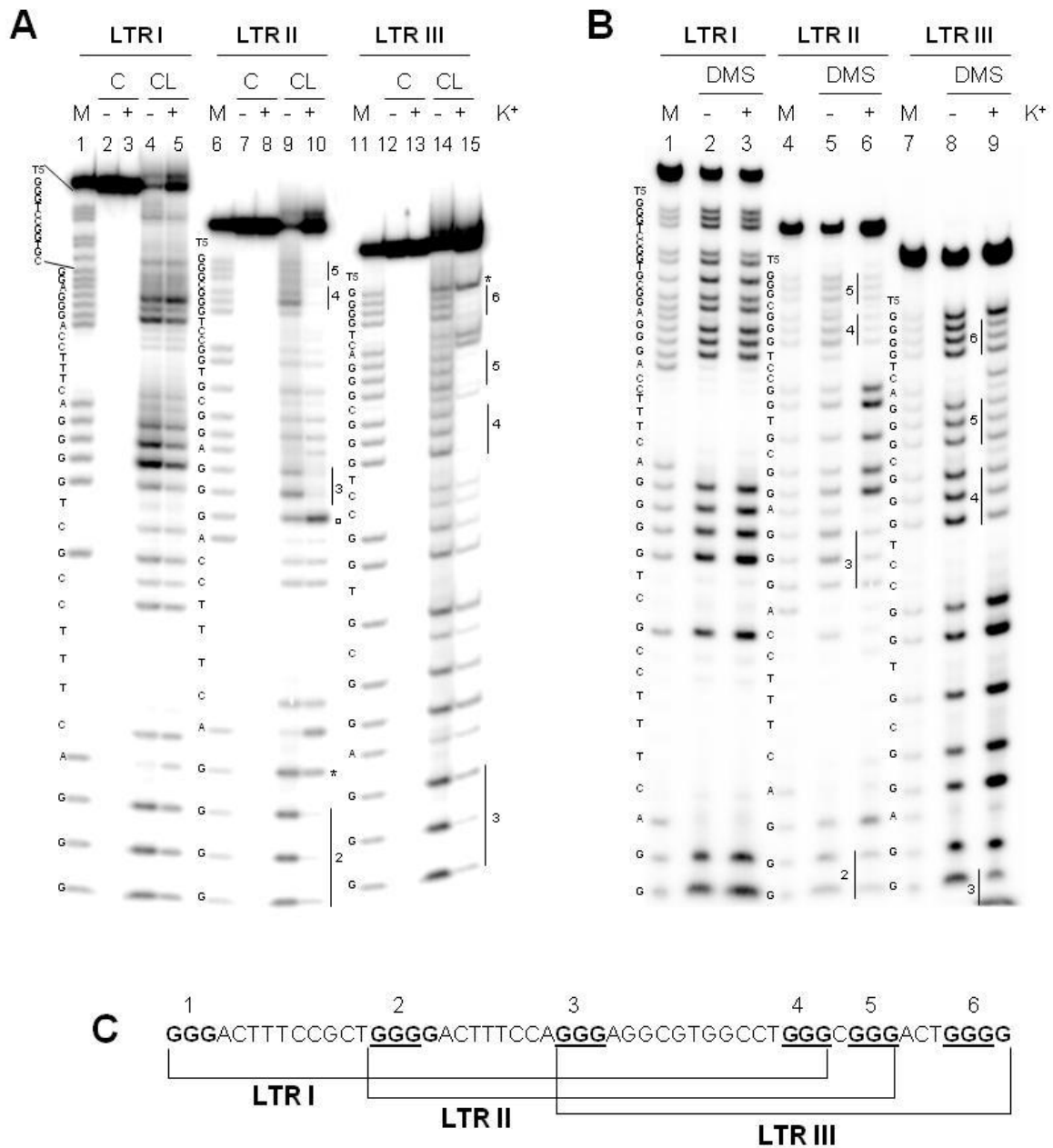
### 5.2.2.2. Footprinting studies

Further support for the formation of G-quadruplexes was obtained by protection assays. Two G-N7 alkylators were separately employed: DMS, the standard reactant to highlight G bases involved in G-quartets, and clerocidin (CL), endowed with finely tunable reactivity.

In the CL protection assay, tracts 2, 3, 4, and 5 in the LTR-II sequence were all protected (Figure 5.5A). In particular, in G-tract 2, the three 5'-Gs were involved in G-quadruplex, while the 3'-G was excluded (symbol \*, lane 10). Surprisingly, in tract 3, the 5'-G was overexposed to cleavage (symbol  $\varpi$ , lane 10). On the basis of CL discrimination between protected and stretched bases, the cleaved G, adjacent to the 8-nt-long linker region, is likely strained and therefore exposed to CL alkylation. In the LTR-III sequence, tracts 3, 4, 5, and 6 were clearly protected, while the 3'-G of tract 6 was not (symbol \*, lane 15). Conversely, no protection was observed in the LTR-I sequence in the CL protection assay (lane 2-3, Figure 5.5 A).

Similar results were obtained using the G-N7 alkylator DMS (Figure 5.5B): LTR-I seemed not to form any G-quadruplex since no protection sites were detectable (lane 2-3, Figure 5.5B). Again, LTR-II showed a protection of G-tracts from 2 to 5 (lane 5-6, Figure 5.5B). In LTR-III sequence, G-tracts from 3 to 6 were protected and again in G-tract 2, the 3'-G appeared excluded from G-quadruplex structure (lane 8-9, Figure 5.5B).

A summary scheme of the protected G-bases involved in the G-quadruplex folding is reported in Figure 5.5C.



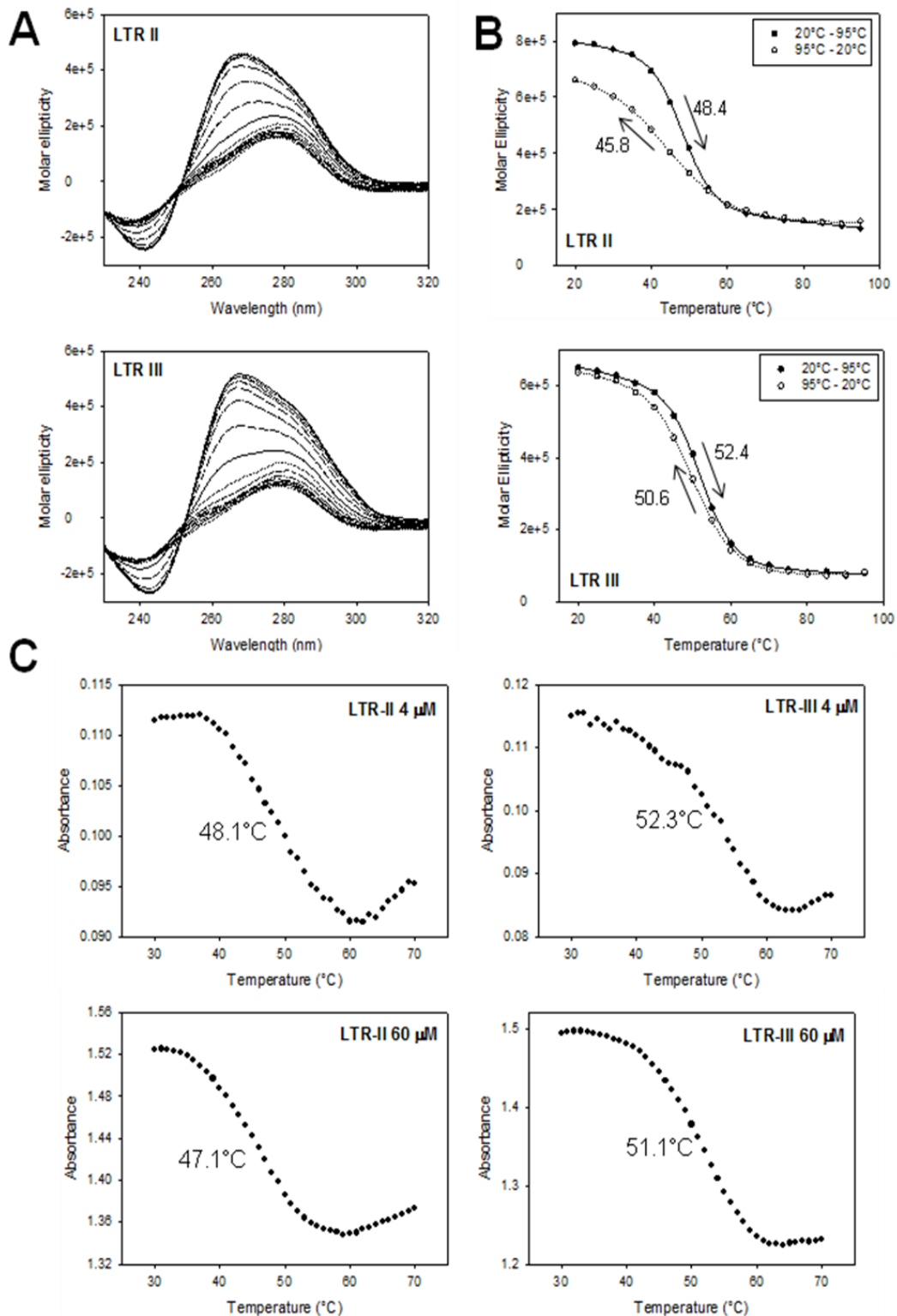
**Figure 5.5 Protection assays of LTR I, LTR II and LTR III** A) LTR sequences were heat denatured, cooled down in the presence or absence of K<sup>+</sup>, and treated with DMS before cleavage induction at the G alkylated sites with hot piperidine. B) LTR sequences were heat denatured, cooled down in the presence or absence of K<sup>+</sup>, and treated with clerocidin followed by hot piperidine to induce cleavage at the G alkylated sites (CL lanes) or just treated with hot piperidine (C lanes). Base sequences are provided on the left of each oligonucleotide. Protected G-tracts are indicated with vertical lines and corresponding numbers. The  $\boxtimes$  symbol indicate overexposed bases and the \* symbol indicate non-protected bases within a protected G-tract. M indicate marker lanes obtained with the Maxam and Gilbert sequencing protocol. C) Overlapping sequences of LTR I, LTR II and LTR III. Relevant G-tracts are shown in bold and numbered (1-6). Protected G bases are underlines.

### 5.2.2.3. Melting studies

Preliminary spectroscopic analysis revealed a G-quadruplex formation in LTR-II and LTR-III sequences, thus only these two sequences were selected for further characterizations.

Stability of LTR-II and LTR-III G-quadruplexes in the absence/presence of 100 mM  $K^+$  was assessed by thermal unfolding experiments monitored by CD and melting temperatures ( $T_m$ ) were calculated. In the presence of  $K^+$ , both LTR-II and LTR-III showed an increase of  $T_m$  of 9.3 and 13.9°C, respectively, suggesting a  $K^+$ -dependent G-quadruplex stabilization. Moreover, LTR-III appeared slightly more stable than LTR-II with a  $T_m$  of 51.9°C (Table 5.2).

Next, we asked whether the G-quadruplexes in LTR-II and LTR-III are prevalently inter or intramolecular structures.  $T_m$  values of LTR-II and LTR-III unfolding were 2.6 and 1.8°C (respectively) higher than  $T_m$  values gained during the refolding process (Figure 5.6B). The small hysteresis was indicative of reversible and intramolecular G-quadruplex formation, with folding kinetics faster for LTR-III than for LTR-II. These data indicate that the presence of two long loops in LTR-II (9, 11, 1 nt-loops) versus only one long loop in LTR-III (11, 1, 3 nt loops) moderately affected both the thermodynamic stability and the kinetics of G-quadruplex formation. To further confirm these data, UV melting experiments at 4 and 60  $\mu$ M oligomer concentrations were made. In fact, an higher oligomer concentration promotes the intermolecular structure with a difference in  $T_m$  values. UV melting experiments showed a superimposable  $T_m$  values for both concentrations, confirming intramolecular G-quadruplex formation for LTR-II and LTR-III (Figure 5.6C).



**Figure 5.6 Thermal stability of LTR-II and LTR-III** A) LTR-II and LTR-III oligonucleotides (3-6  $\mu\text{M}$ ) were treated at increasing/decreasing temperature (20-95°C 2°C/min) and CD spectra recorded. B) Molar ellipticity values recorded at 265 nm were plotted against temperature and fitted with the van't Hoff equation to extrapolate  $T_m$  values. Down- and up-pointing arrows indicate denaturation and renaturation experiments, respectively.  $T_m$  values are indicated aside the corresponding melting curve. C) LTR-II and LTR-III oligonucleotides at low (4  $\mu\text{M}$ ) and high concentration (60  $\mu\text{M}$ ) were treated at increasing temperature (20-95°C) and UV absorbance spectra recorded. Absorbance values recorded at 295 nm were plotted against temperature to extrapolate  $T_m$  values, which are indicated aside the corresponding melting curve.

The commercially available G-quadruplex ligands TMPyP4, BRACO-19, and PIPER were next tested by CD for binding to the novel LTR G-quadruplex structures in the absence/presence of 100 mM K<sup>+</sup>. In the presence of K<sup>+</sup>, the average  $\Delta T_m$  was above 30°C, indicating both LTR sequences were significantly stabilized by the compounds. BRACO-19 was most efficient determining an increase of T<sub>m</sub> up to 43.3°C for LTR-III. TMPyP4 resulted in an increase of T<sub>m</sub> of about 30°C for both LTR sequences. PIPER was less effective, especially with LTR-III with an increase of T<sub>m</sub> of 15.3°C. Interestingly, even in the absence of K<sup>+</sup>, compounds induced a G-quadruplex conformation with average  $\Delta T_m$  around 15°C (Table 5.2).

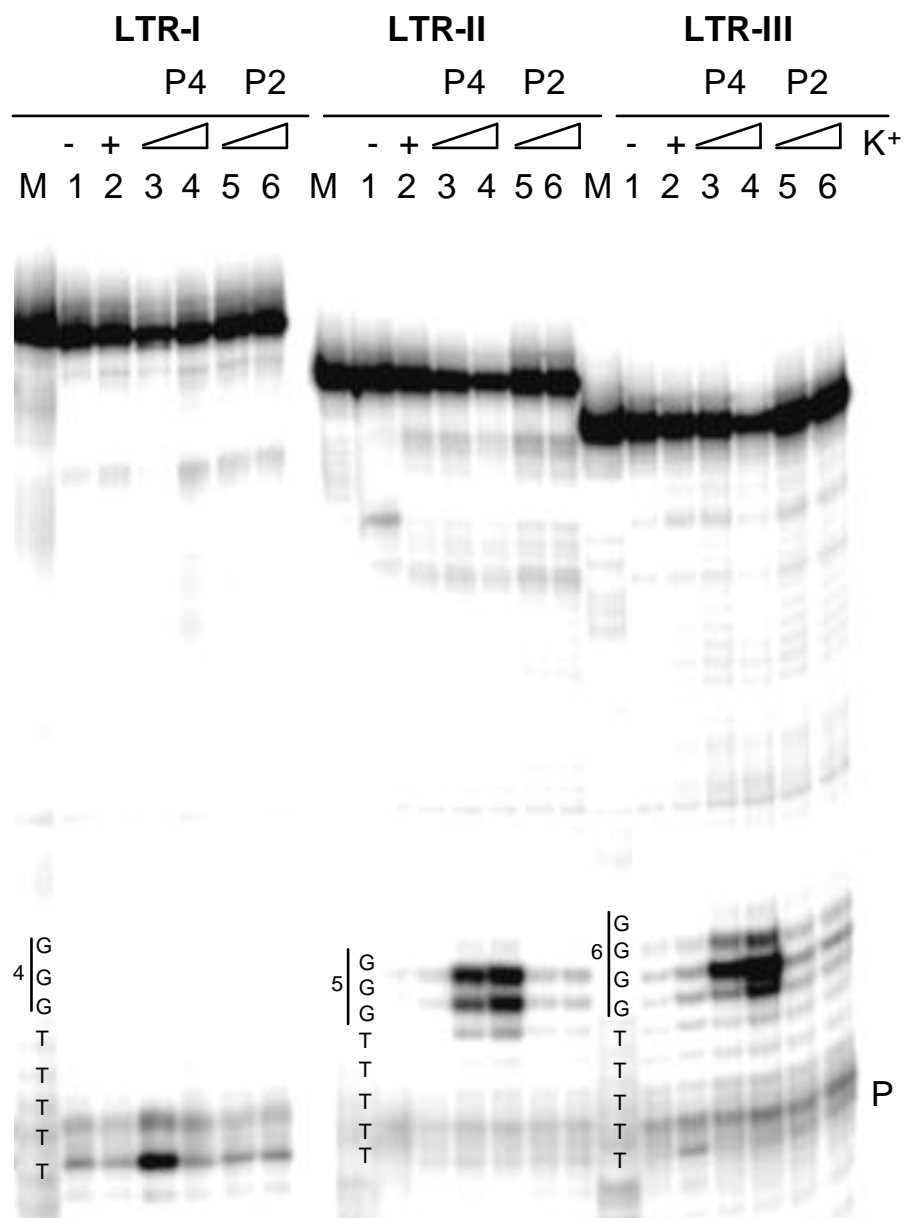
G-quadruplex sequence	K <sup>+</sup> (mM)	Drug added	T <sub>m</sub> (°C)	$\Delta T_m$ (°C) (T <sub>m</sub> <sub>K<sup>+</sup>[100]</sub> - T <sub>m</sub> <sub>K<sup>+</sup>[0]</sub> )	$\Delta T_m$ (°C) (T <sub>m</sub> <sub>K<sup>+</sup>[100] drug</sub> - T <sub>m</sub> <sub>K<sup>+</sup>[100]</sub> )
LTR-II	0	-	39.7 ± 1.2	-	-
	0	TMPyP4	52.1 ± 1.8	-	12.4
	0	BRACO-19	57.3 ± 1.7	-	17.6
	0	PIPER	53.5 ± 4.0	-	13.8
	100	-	49.0 ± 0.2	9.3	-
	100	TMPyP4	82.4 ± 1.5	30.3	33.4
	100	BRACO-19	88.9 ± 0.6	31.6	39.9
	100	PIPER	80.1 ± 3.3	26.6	31.1
LTR-III	0	-	38.0 ± 1.5	-	-
	0	TMPyP4	60.9 ± 2.5	-	22.9
	0	BRACO-19	50.9 ± 1.1	-	12.9
	0	PIPER	53.0 ± 1.2	-	15
	100	-	51.9 ± 0.2	13.9	-
	100	TMPyP4	89.9 ± 0.7	29	38
	100	BRACO-19	94.2 ± 2.0	43.3	42.3
	100	PIPER	68.3 ± 1.3	15.3	16.4

**Table 5.2 Stabilization (T<sub>m</sub>) of LTR II and III Sequences (4 μM) in the absence and presence of 100 mM K<sup>+</sup> and G-quadruplex ligands (16 μM)**



#### 5.2.2.4. *Taq* polymerase stop assays

Further evidence of the stability of the LTR G-quadruplexes was provided by *Taq* polymerase stop assay (Figure 5.7). Samples were incubated in the absence/presence of 100 mM K<sup>+</sup> (Figure 5.7, lanes 1 and 2) and with increasing concentrations (50–100nM) of TMPyP4 (Figure 5.7, lanes 3 and 4). TMPyP2, a non-G-quadruplex binding porphyrin (Han et al, 2001), was used as a negative control (Figure 5.7, lanes 5–6). *Taq* polymerase activity was tested at 47°C against LTR-I, LTR-II, and LTR-III DNA templates. Full length products were obtained in the absence of K<sup>+</sup> in the three sequences. However, in the presence of K<sup>+</sup>, a premature stop site occurred in LTR-II and LTR-III (Figure 5.7, lanes 2) at the first two 3'-G bases involved in G-quadruplex (G-tract 5 in LTR-II and 6 in LTR-III). The stop became clearer upon incubation with TMPyP4 (Figure 5.7, lanes 3–4). The stop observed in the presence of TMPyP2 was comparable to the stop with K<sup>+</sup> alone (Figure 5.7, compare lanes 5–6 with 2), confirming that the compound did not affect G-quadruplex formation. In addition, no stop was observed in LTR-I confirming that no G-quadruplex formation occurred and showing that polymerase activity was not directly inhibited by the compounds (LTR-I, lanes 2–6).



**Figure 5.7 *Taq* polymerase stop assay** Oligonucleotides were folded in the presence or absence of K<sup>+</sup>. K<sup>+</sup>-treated samples were further incubated with either TMPyP4 (P4) or the control compound TMPyP2 (P2). Oligonucleotides were used as templates in a *Taq* polymerase reaction at 47°C. Bases at the 3'-end and the corresponding 3'-G-tract are indicated for each sequence. P indicates the band of the labeled primer. M is a marker lane obtained with the Maxam and Gilbert sequencing protocol.

### 5.2.3. Characterization of multiple G-quadruplexes folding in the full-length LTR G-rich region

Considering that LTR-II and LTR-III sequences have a portion in common in the full length LTR G-rich region, we investigated the dynamic equilibrium that regulates the G-quadruplex folding in the full-length region. In fact, the full-length LTR G-rich sequence comprises four GGG-tracts (1, 3–5), two GGGG-tracts (2, 6) and three additional GG-tracts (3', 3'', 6') which could be used for multiple G-quadruplex formation (Figure 5.2C-D). Moreover, we examined a number of subsequences to dissect this issue, corresponding to LTR-II, LTR-III, and their combination of LTR-(II+III).

#### 5.2.3.1. Melting experiments

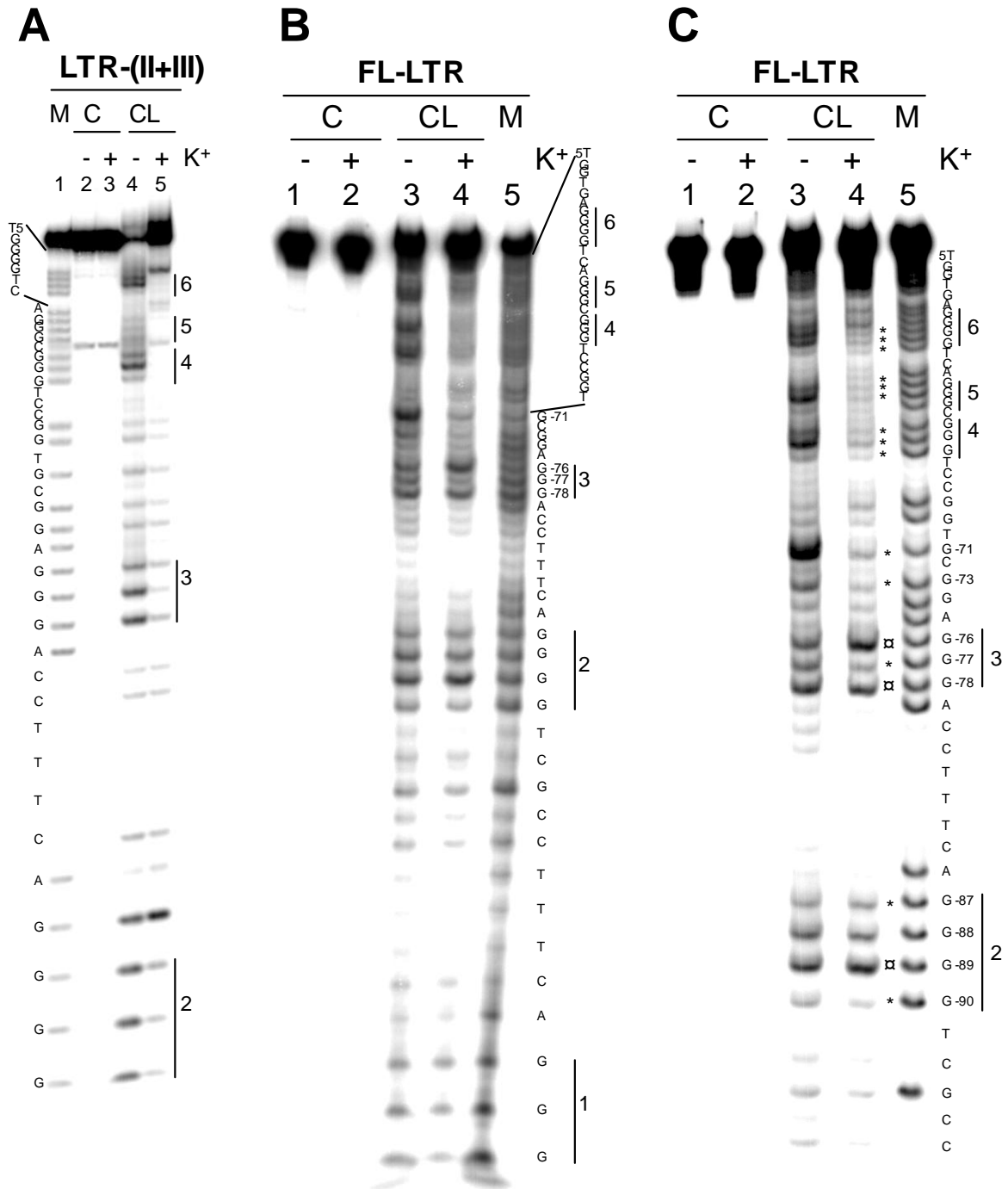
CD-monitored melting experiments showed that the  $T_m$ s of LTR-II and LTR-III increased by 10–13°C in the presence of  $K^+$  (Table 5.2). Also LTR-(II+III) and the full length LTR (FL-LTR) were greatly stabilized by the metal ion, with an average  $\Delta T_m$  around 22°C. The two  $T_m$ s resulted very close to each other (55.8°C for LTR-(II+III) and 61.1°C for FL-LTR) and 4–6°C higher than for LTR-II and LTR-III. The G-quadruplex ligands TMPyP4, BRACO-19, and PIPER were again tested by CD for stabilizing the G-quadruplex structures in the FL-LTR sequence. In the presence of  $K^+$ , G-quadruplex ligands further stabilized these conformations with an average  $\Delta T_m$  from 4.5 to 8.8°C. Interestingly, PIPER resulted the most efficient ligand in stabilizing (Table 5.3).

G-quadruplex sequence	$K^+$ (mM)	Drug added	$T_m$ (°C)	$\Delta T_m$ (°C) ( $T_{m_{K^+[100]}} - T_{m_{K^+[0]}}$ )	$\Delta T_m$ (°C) ( $T_{m_{K^+[100] \text{ drug}}} - T_{m_{K^+[100]}}$ )
LTR-(II+III)	0	-	34.4 ± 2.0	-	-
	100	-	55.8 ± 0.3	21.4	-
FL-LTR	0	-	33.8 ± 5.1	-	-
	100	-	56.6 ± 0.8	22.8	-
	100	TMPyP4	61.1 ± 0.5	-	4.5
	100	BRACO-19	62.2 ± 0.5	-	5.6
	100	PIPER	65.4 ± 0.4	-	8.8

**Table 5.3 Stabilization ( $T_m$ ) of LTR-(II+III) and FL-LTR sequences (4  $\mu$ M) in the absence and presence of 100 mM  $K^+$  and G-quadruplex ligands (16  $\mu$ M)**

### 5.2.3.2. Footprinting studies

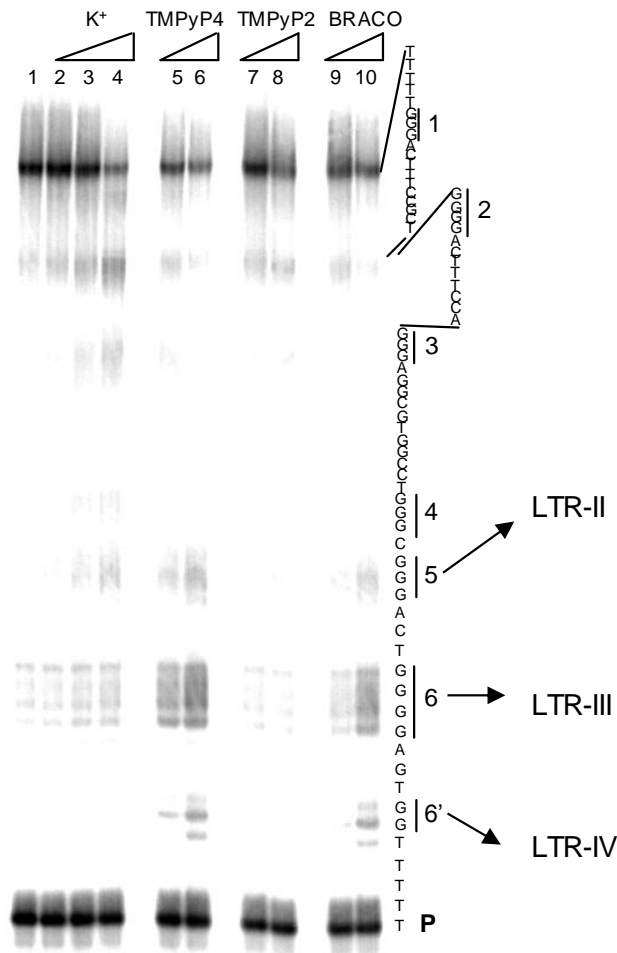
LTR-(II+III) and FL-LTR were next examined in a CL-protection assay (Figure 5.8). In the LTR-(II+III) sequence, all G-tracts present (2–6) were protected, indicating the coexistence of LTR-II and LTR-III structures (compare lanes 5 and 4, Figure 5.8A). In the FL-LTR, G-tract 1 was not protected (compare lanes 4 and 3, Figure 5.8B), showing its exclusion from G-quadruplex folding. In the 3'-region of FL-LTR, magnified in Figure 5.8C, G-tracts 4, 5, and 6 were clearly protected (compare lanes 4 and 3, Figure 5.8C). Tracts 3 and 2 were partially protected (G -77, G -90, and G -87) and partially overexposed (G -76, G -78, and G -89) (symbols \* and  $\alpha$ , respectively, lane 4, Figure 5.8C), indicating involvement of these tracts in the G-quadruplex, with buried and stretched bases. In addition, G-71 was protected, suggesting participation in the G-quadruplex conformation or burial within the long linker region possibly folded on the quadruplex core. These data imply that both LTR-II and LTR-III conformations form in the FL-LTR. Note that similar dynamic G-quadruplexes have been previously proposed (Qin et al, 2010). In addition, both LTR-II and LTR-III display a 11-bp long loop that could form four Watson–Crick base pairs and a GTG hairpin. Similar folding in the loop of the hTERT promoter G-quadruplex structure has been shown to promote cooperative binding (Palumbo et al, 2009; Yu et al, 2012). However, on the basis of the footprinting data, such a folding does not likely occur in the LTR G-quadruplexes.



**Figure 5.8 Characterization of G-quadruplex structures in extended regions of the LTR G-rich sequence**  
 A) The LTR-(II+III) oligonucleotide and (B) the FL-LTR sequence were folded in the presence or absence of K<sup>+</sup> and treated with CL followed by hot piperidine (CL lanes) or just treated with piperidine (C lanes). Relevant G-tracts are highlighted by vertical lines and corresponding numbers. M indicates the marker lane. Base sequences are shown aside each gel image. C) Magnification of the 3'-end of the FL-LTR oligonucleotide. Samples were treated as described above but they were run for shorter time in the denaturing gel. The \* and α symbols indicate protected and overexposed nucleotides, respectively.

### 5.2.3.3. *Taq* Polymerase stop assays

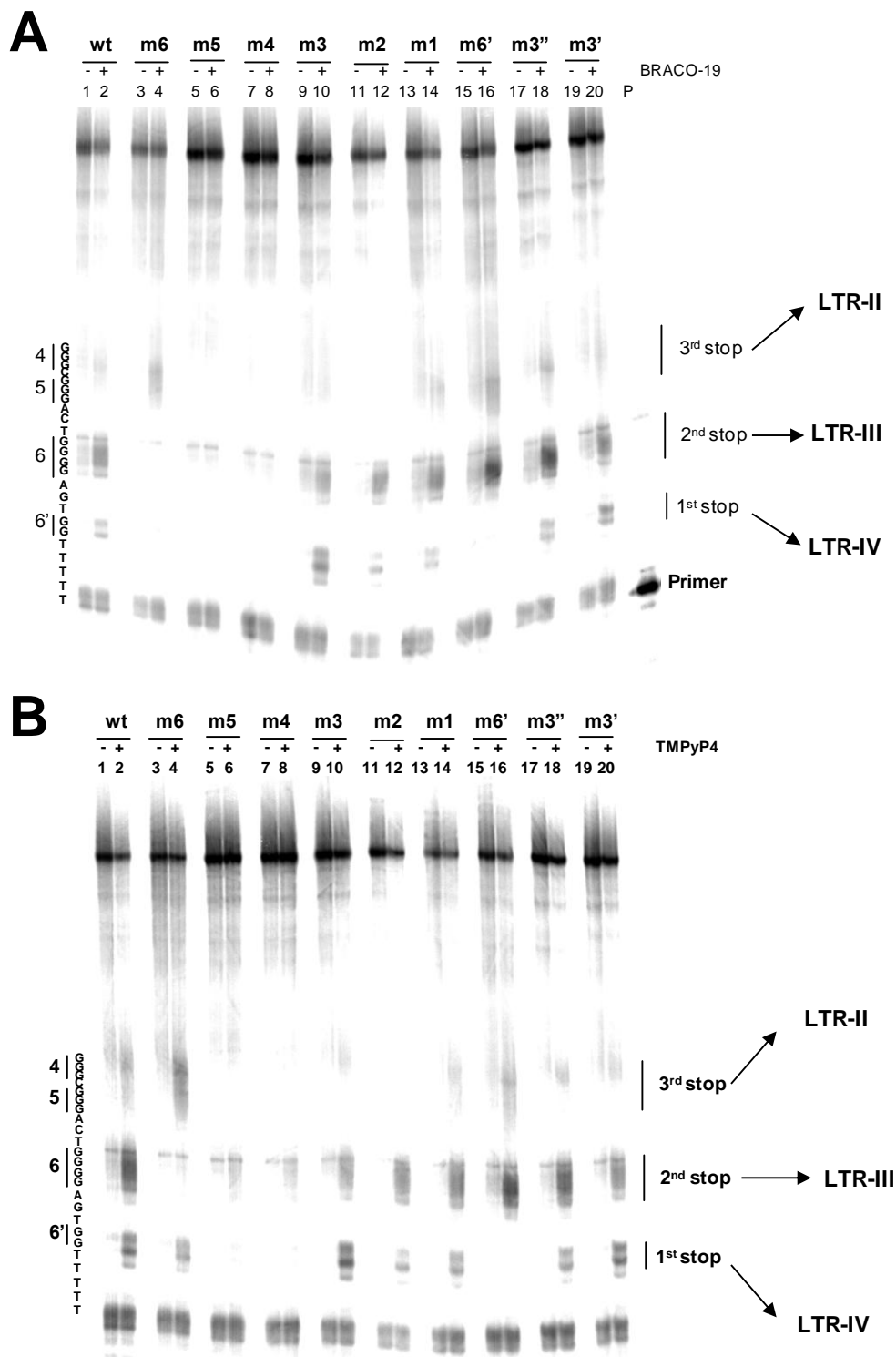
In *Taq* polymerase stop assays, at increasing concentrations of  $K^+$ , pausing sites were observed at all G-quadruplex relevant G-tracts, i.e., 6, 5, 4, 3, and 2 (lanes 2–4, Figure 5.9), confirming the coexistence of LTR-II and LTR-III structures. With TMPyP4, two major stops emerged at G-tracts 6 and 5, while pausing at G-tracts 4–2 was no longer observed (lanes 5–6, Figure 5.9). Using BRACO-19, an identical behavior was obtained (lanes 9–10, Figure 5.9), whereas the control compound TMPyP2 maintained a stop site pattern similar to that observed in  $K^+$  alone (lanes 7–8, Figure 5.9). The strong stabilization imparted by the ligands (see Table 5.2 and 5.3) apparently caused the polymerase to stop at the first G repeat involved in a G-quadruplex, without allowing the enzyme to further proceed on the DNA template. We concluded that BRACO-19 and TMPyP4 mainly stimulated formation of an LTR-III-like structure and in part of an LTR-II-like structure (3'-stops at 6 and 5, respectively, Figure 5.9). Interestingly, a new pausing site arose at G-tract 6', which suggests the formation a new LTR structure, namely LTR-IV (Figure 5.9). The LTR-IV is a sequence of 19 bp that may form a G-quadruplex with two stacked G-quartets (Table 5.1).



**Figure 5.9 G-quadruplexes forming in the FL-LTR sequence** *Taq* polymerase stop assay on the FL-LTR template in the presence of increasing concentration of K<sup>+</sup> (0–150 mM) (lanes 1–4) and of 100 mM K<sup>+</sup> and G-quadruplex ligands TMPyP4 (lanes 5–6), TMPyP2 (lanes 7–8) and BRACO-19 (lanes 9–10). The base sequence is shown on the right. Vertical lines indicate stop sites observed in the presence of K<sup>+</sup>. Arrows point to the structures of G-quadruplexes stabilized in the presence of G-quadruplex ligands.

Single-base mutations were next introduced in the FL-LTR to assess the role of each G-tract on G-quadruplex formation. Mutants were evaluated in the absence/presence of BRACO-19 and TMPyP4. In the presence of BRACO-19, deletion of one G base in G-tract 6 (m6) resulted in maintenance of the third stop site only at G-tracts 5/4 (formation of only LTR-II, Figure 5.10A). Mutations in G-tracts 5 and 4 (m5 and m4) produced the harshest effects with all stops removed (no LTR G-quadruplex formed). Mutation in G-tract 3 (m3) increased the efficiency of the newly proposed first stop site (LTR-IV), while removal of one G in G-tract 6' (m6') effectively deleted the first stop (only LTR-II and LTR-III formed). Mutation in G-tract 2 (m2) hampered formation of the third stop (no LTR-II), while mutations in G-tracts 1 (m1), 3' (m3'), and 3'' (m3'') produced no effect. Deletion of G-tracts m6', m6, m5, and m4

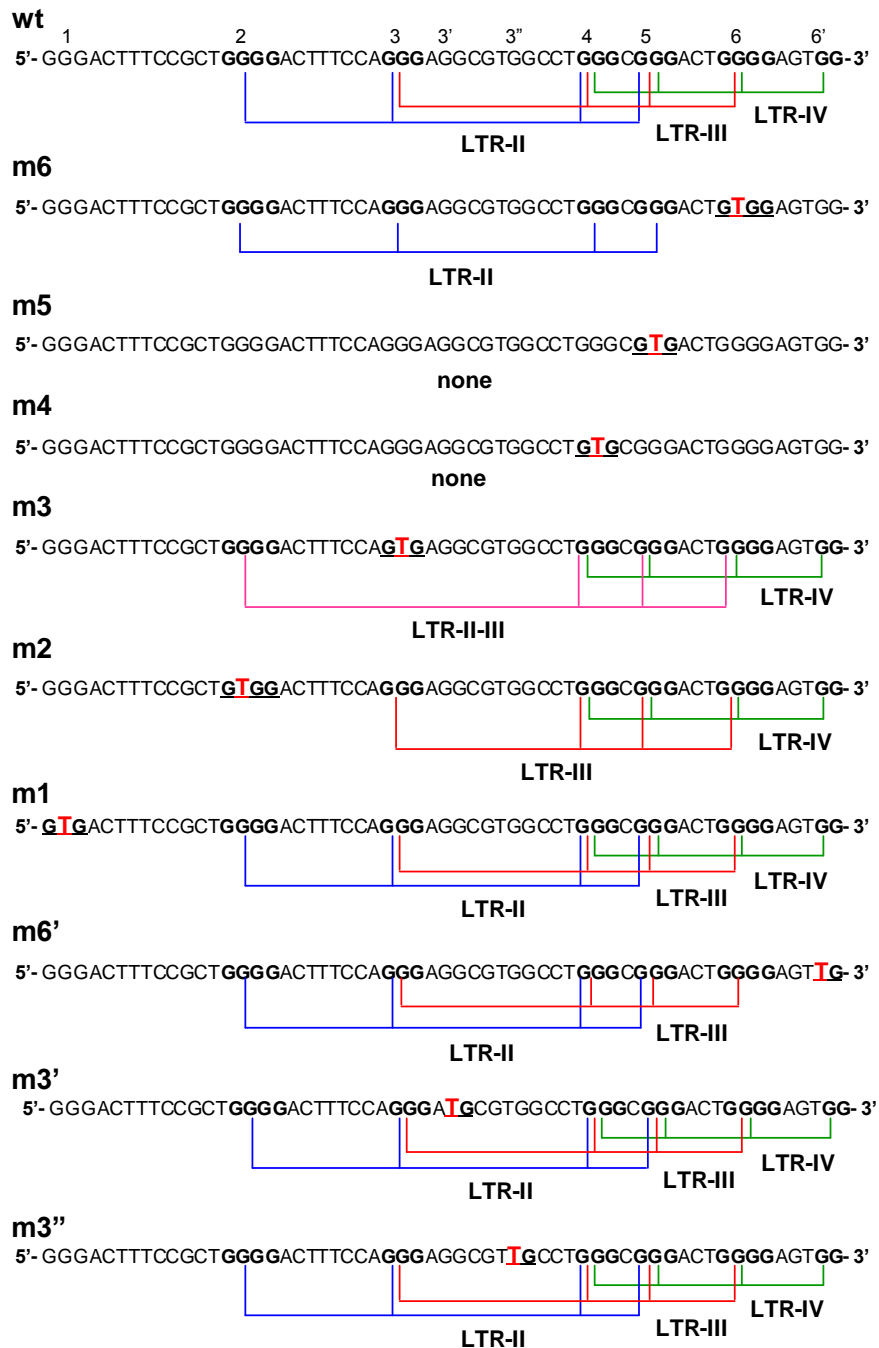
inhibited formation of LTR-IV, indicating that these G-tracts must all be involved in its structure. Similar results were obtained with TMPyP4 (Figure 5.10B).



**Figure 5.10 G-quadruplexes forming in the FL-LTR sequence** *Taq* polymerase stop assay on the wild-type and mutants FL-LTR templates in the presence of 100 mM K<sup>+</sup> and (A) 100 nM BRACO-19 or (B) 100nM TMPyP4. The base sequence is shown on the right. Vertical lines indicate stop sites observed in the presence of K<sup>+</sup>. Arrows point to the structures of G-quadruplexes stabilized in the presence of G-quadruplex ligands.



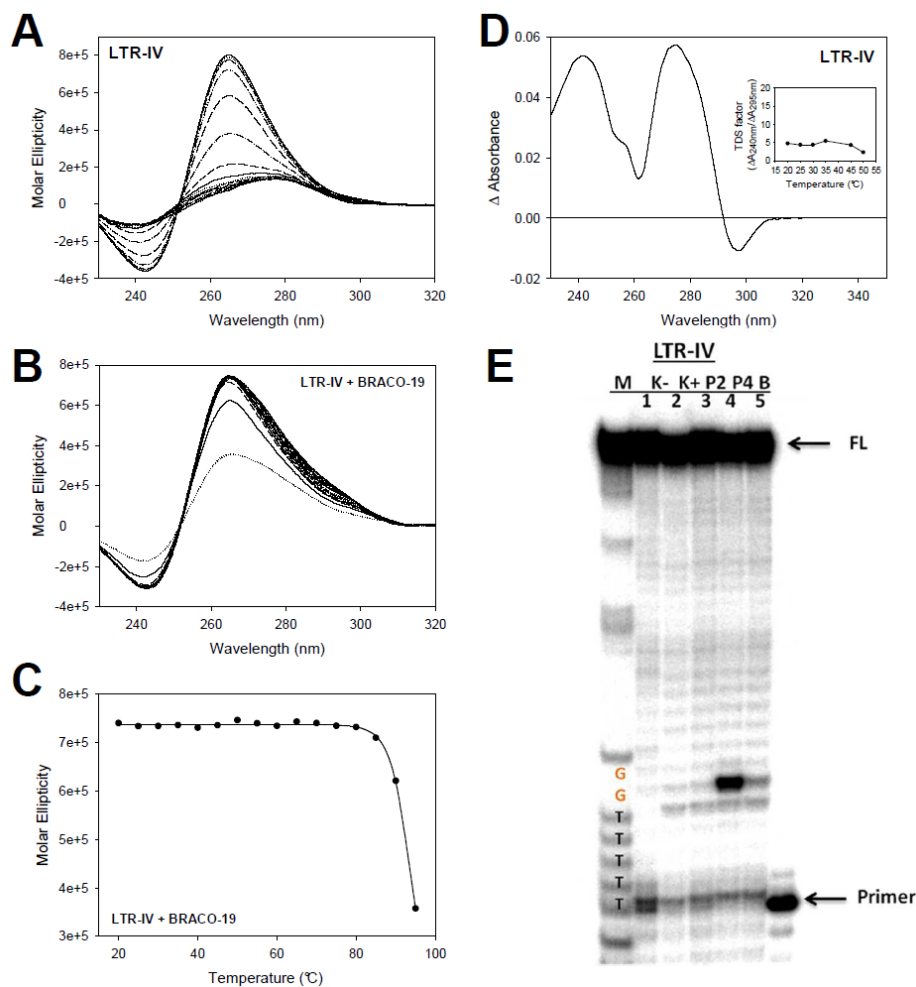
A summary of single-base mutation analysis is reported in Figure 5.11. In the WT sequence, in the presence of ligands, three G-quadruplexes corresponding to LTR-II, LTR-III, and LTR-IV are present. The core of the LTR G-quadruplex architecture rests on G-tracts 4 and 5 whose mutations completely abolish quadruplex building.



**Figure 5.11 G-quadruplexes forming in each of the examined mutant FL-LTRs** Brackets indicate G-tracts involved in the relevant G-quadruplex. Blue, red and green brackets indicate LTR-II, LTR-III, LTR-IV G-quadruplexes, respectively. Mutated bases are shown in red. LTR-II-III is an alternative structure that forms involving G-tracts 2,4,5 and 6 when G-tract 3 is unavailable.

### 5.2.4. Characterization of LTR-IV G-quadruplex

The G-quadruplex analysis on FL-LTR revealed a putative new G-quadruplex structure namely LTR-IV. Thus, we decided to further characterize this putative G-quadruplex forming sequence. The 19 bp LTR-IV sequence including these G-tracts revealed CD and UV features characteristic of a parallel G-quadruplex (Figure 5.12A and D) showing increased stability upon  $K^+$  addition (Table 5.4). Although in the presence of 100 mM  $K^+$  LTR-IV was less stable than LTR-II, LTR-III, LTR-(II+III), and FL-LTR, upon addition of G-quadruplex ligands, LTR-IV stability significantly increased; in particular,  $T_m$  values of LTR-IV in the presence of ligands were consistently higher than  $T_m$  of LTR-III and LTR-II (Table 5.4). These data corroborate the results presented in Figure 5.9 and 5.10 and demonstrate that G-quadruplex ligands both stabilize naturally occurring G-quadruplex conformations and induce (and greatly stabilize) a novel G-quadruplex structure which is not present under physiological conditions. Interestingly, LTR-IV could form a G-quadruplex with a single bulge, a type of structure recently reported in G-quadruplex conformations (Mukundan & Phan, 2013).

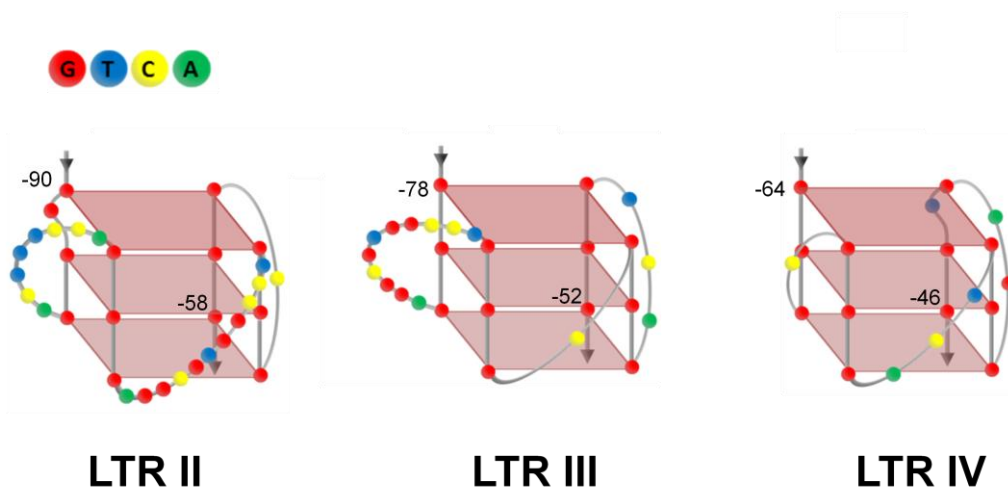


**Figure 5.12 Thermal stability of LTR-IV** in A) the absence and B) presence of BRACO-19. LTR-IV oligonucleotide was treated at increasing temperature (20–95°C) and CD spectra recorded. C) Molar ellipticity values recorded at 265nm were plotted against temperature and fitted with the van't Hoff equation to extrapolate  $T_m$  values. D) TDS spectrum of LTR IV. LTR IV exhibit the G-quadruplex TDS signature. TDS factor graph is shown in the inset. E) *Taq* polymerase stop assay performed at 47°C in the absence (K-) and presence (K+) of  $K^+$ , and in presence of TMPyP2 (P2), TMPyP4 (P4) or BRACO-19 (B). FL indicates full-length polymerase product. The 3' end sequence is shown on the left.

G-quadruplex sequence	K <sup>+</sup> (mM)	Drug added	T <sub>m</sub> (°C)	ΔT <sub>m</sub> (°C) (T <sub>m</sub> <sub>K<sup>+</sup>[100]</sub> - T <sub>m</sub> <sub>K<sup>+</sup>[0]</sub> )	ΔT <sub>m</sub> (°C) (T <sub>m</sub> <sub>K<sup>+</sup>[100] drug</sub> - T <sub>m</sub> <sub>K<sup>+</sup>[100]</sub> )
LTR IV	0	-	26.2 ± 0.5	-	-
	100	-	43.4 ± 0.2	17.2	-
	100	TMPyP4	87.3 ± 3.4	-	43.9
	100	BRACO-19	97.4 ± 3.2	-	54.0
	100	PIPER	99.2 ± 6.5	-	55.8

**Table 5.4 Stabilization (T<sub>m</sub>) of LTR-IV sequence (4 μM) in the absence and presence of 100 mM K<sup>+</sup> and G-quadruplex ligands (16 μM)**

On the basis of these data, involvement of G-tracts and G-quadruplex types could be extrapolated as shown in Figure 5.13.



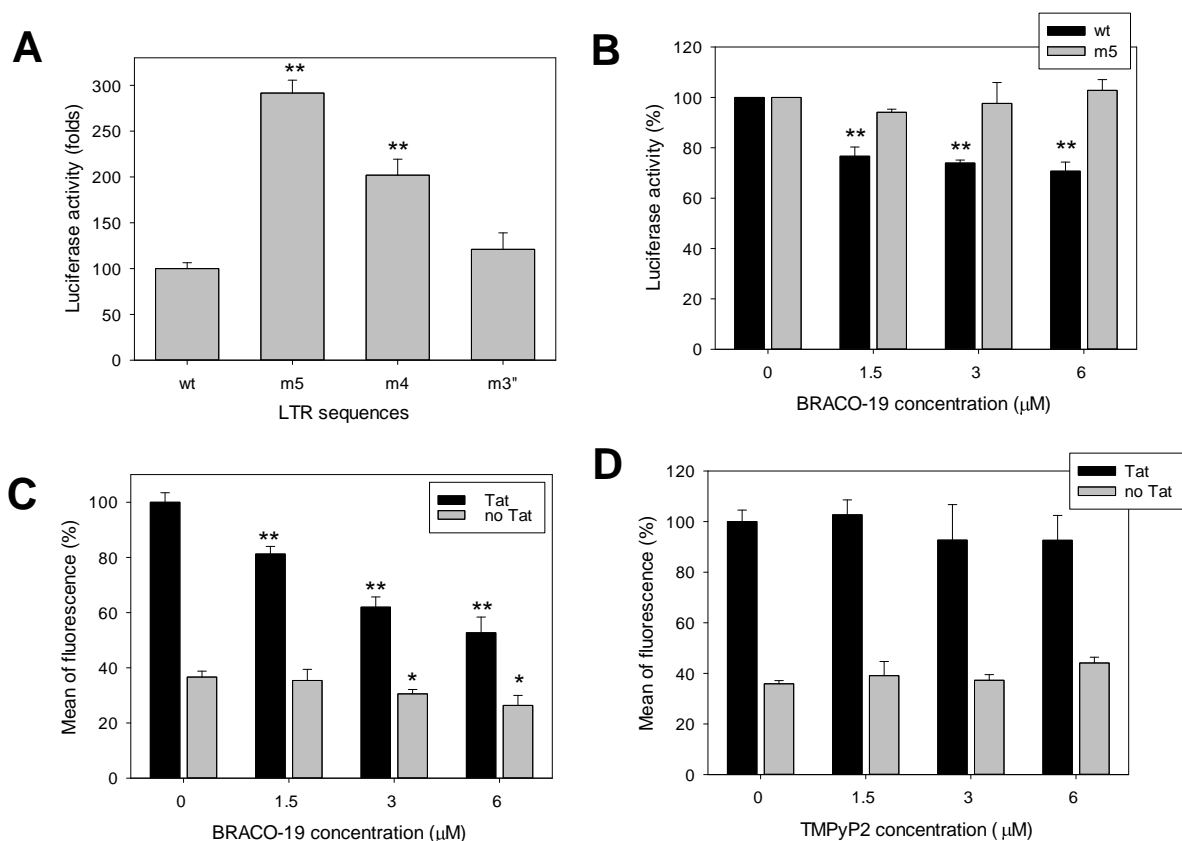
**Figure 5.13 Models of characterized LTR G-quadruplex structures.** G, T, C, and A bases are shown in red, blue, yellow, and green, respectively. Numbers in the structures indicate nucleotide position within the HIV-1 integrated genome.

### 5.2.5. Biological significance of G-quadruplexes in the LTR region.

To evaluate the biological significance of G-quadruplex structures within the LTR sequence, the WT LTR sequence (position -381/+83) and selected point mutants were cloned upstream of the firefly luciferase gene in a promoterless plasmid. A control vector with Renilla luciferase under the HSV-1 TK promoter was used to normalize transfection efficiency. Mutants corresponding to m4 and m5, which totally prevent G-quadruplex formation as described above, were assayed along with WT FL-LTR and mutant m3" which served as controls, respectively, for the original G-quadruplex-forming sequence and for a mutated sequence that does not disrupt G-quadruplex. LTR promoter activity was tested in HEK 293T

cells: m4 and m5 LTR promoter activities were about twice as high as that of the WT LTR (Figure 5.14A). This increment is in line with that observed in eukaryotic promoters, i.e., human c-MYC, KRAS, and thymidine kinase 1 (Basundra et al, 2010; Membrino et al, 2011; Siddiqui-Jain et al, 2002). In contrast, when G-quadruplex formation was unharmed (m3''), LTR promoter activity was comparable to wild-type. These data suggest that G-quadruplexes act as repressor elements in the transcriptional activation of HIV-1. The promoter activities of WT and m5 LTRs were next tested in the presence of increasing concentrations of BRACO-19. As shown in Figure 5.14B, WT LTR promoter activity decreased to around 70% of the untreated control while displaying no effect on m5 LTR activity, supporting a G-quadruplex-mediated inhibition. We also observed concentration-dependent inhibition of the control Renilla luciferase reporter gene by BRACO-19 (less intense than inhibition of the LTR-driven firefly luciferase) attributable to the toxic effects produced by ligand binding to eukaryotic G-quadruplex structures (data not shown). A similar effect has been previously observed (Siddiqui-Jain et al, 2002).

Therefore, the effects of G-quadruplex ligands were next evaluated in a eGFP-reporter system in the presence/absence of the transcriptional activator Tat, using flow cytometry to gate only GFP<sup>+</sup> cells with viable morphology. In the presence of Tat, BRACO-19 was able to impair LTR promoter activity in a dose-dependent fashion (around 50% inhibition at 6  $\mu$ M, Figure 5.14C). Moreover, in the absence of Tat, inhibition was 30%, confirming the results obtained in the luciferase assay. In contrast, the control ligand TMPyP2 did not show significant inhibition under these conditions (Figure 5.14D). Because Tat-mediated trans-activation boosts viral transcription acting on early mRNA transcripts originating by LTR-mediated transcription, when the latter is blocked, a magnified inhibition may be observed in the presence of Tat.

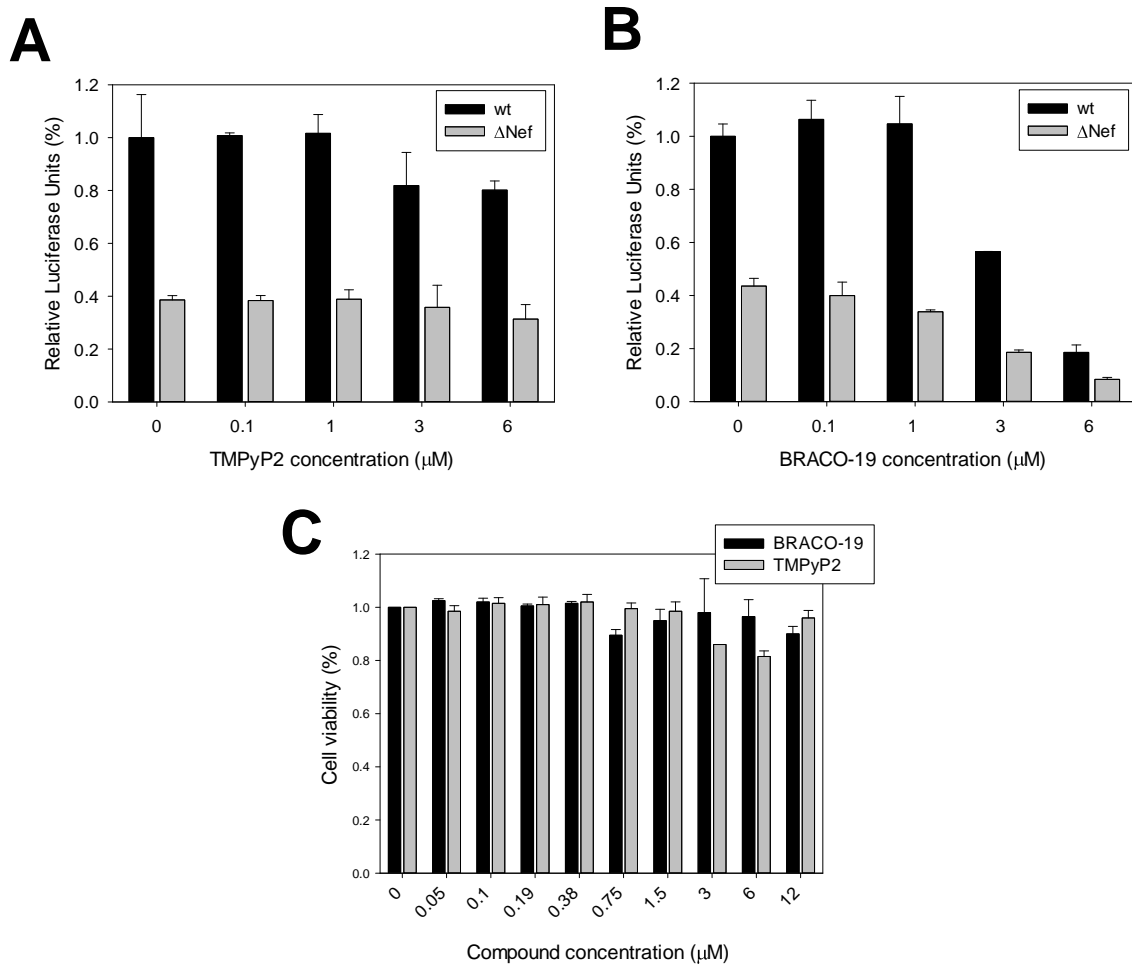


**Figure 5.14 Biological evaluation of LTR G-quadruplexes and treatment with G-quadruplex ligands** (A) Luciferase expression of the wild-type and mutant LTRs normalized to the Renilla luciferase expression and to the wild-type sequence in HEK 293T cells. (B) Normalized luciferase expression of the WT and m5 LTRs in the presence of BRACO-19 (1.5–6.0  $\mu$ M). (C) EGFP mean of fluorescence of cells transfected with the WT LTR-GFP plasmid and treated with increasing concentration (1.5–6.0  $\mu$ M) of BRACO-19 in the absence or presence of Tat. (D) EGFP mean of fluorescence of cells transfected with the WT LTR-GFP plasmid and treated with increasing concentration (1.5–6.0  $\mu$ M) of TMPyP2 in the absence or presence of Tat. In all data sets: n = 3, mean  $\pm$  SD, Student's t-test, \*P < 0.05, \*\*P < 0.01

### 5.2.6. The G-Quadruplex Ligand, BRACO-19, displays antiviral activity.

Having observed a reduction in HIV LTR-driven promoter activity, we next assayed the most active compound, BRACO-19, for antiviral activity. For these experiments, we used infected TZM-bl reporter cell line, in which HIV infection drives transcription of an HIV-1 LTR-luciferase reporter gene construct (Derdeyn et al, 2000; Platt et al, 1998). We observed a significant inhibition of HIV-dependent gene expression upon treatment with the G-quadruplex ligand (Figure 5.15A). HIV LTR activity in TZM-bl cells has been previously reported to be sensitive to the HIV accessory protein Nef (Derdeyn et al, 2000; Emert-Sedlak et al, 2013). To rule out a role for Nef in BRACO-19-mediated inhibition, the same experiment was repeated with Nef-defective HIV ( $\Delta$ Nef). A similar inhibitory effect was observed indicating that the effect of BRACO-19 is Nef-independent. In contrast, the control

ligand TMPyP2 did not affect LTR activity in response to either virus (Figure 5.15B). To date, tested ligands' concentrations did not affect cell viability (Figure 5.15C). These data indicate that G-quadruplex ligands significantly inhibit HIV-1 and that this effect may depend on stabilization of G-quadruplex structures within the HIV-1 LTR promoter region.



**Figure 5.15. Antiviral effects of LTR G-quadruplexes and treatment with G-quadruplex ligands** TZM-bl cells were infected with wild-type (black bars) and  $\Delta$ Nef (gray bars) HIV NL4-3 in the presence of the (A) G-4 stabilizing ligand, BRACO-19 or (B) the negative control G-4 ligand TMPyP2. After 48 h, levels of gene expression were assessed as relative luciferase activity in infected cells. Results are shown relative to the wild-type control cells incubated with only the carrier solvent (DMSO)  $\pm$  SEM (n = 3). Significant inhibition ( $P < 0.05$ ) was observed at 3 and 6  $\mu$ M for both wild-type and  $\Delta$ Nef infected cells. C) Cytotoxicity of BRACO-19 and TMPyP2 on TZM-bl cells was assessed via the Cell-Titer Blue assay (Promega). TZM-bl cells were incubated with the indicated concentrations of compounds for 48 h and cell viability was assessed via the Cell-Titer Blue assay relative to control cells incubated with carrier solvent. Assays were done in triplicate.

### 5.3.G-Quadruplex structures in the HIV-1 *nef* coding region

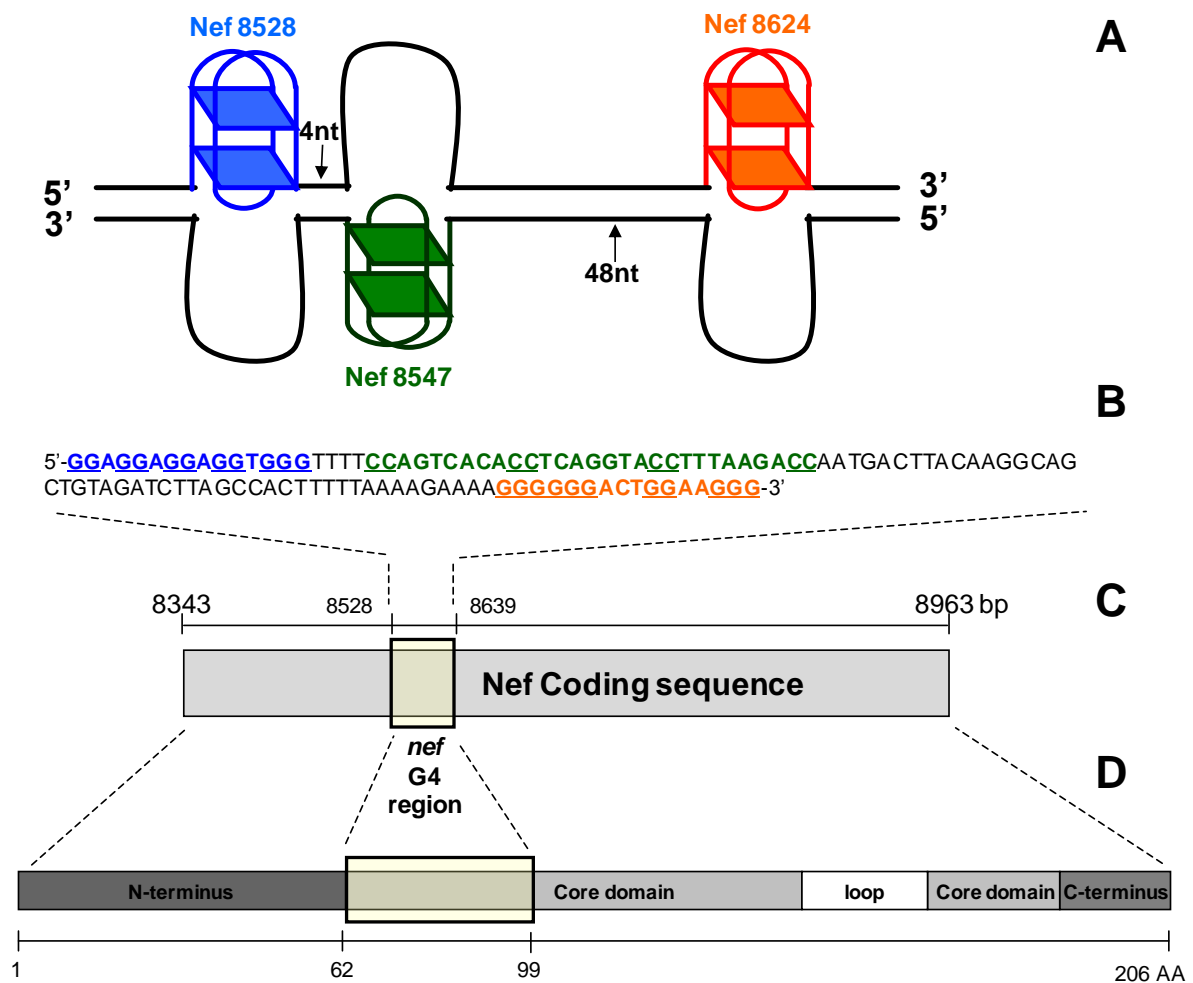
#### 5.3.1. Analysis of putative G-quadruplex sequences in the HIV-1 *nef* coding region

As anticipated, the QGRS analysis of the HIV-1 proviral genome pointed out four putative QGRS located in the *nef* gene (Table 5.5): two were located in the forward strand (positions 8528 and 8624, where +1 is the first base of the HIV-1 genome as reported in GenBank, NC\_001802) and two on the reverse strand (positions 8547 and 8727). In particular, the three that displayed the highest G-score ( $\geq 20$ ) were also adjacent to one another: the first on the forward strand (Nef8528) and the second on the reverse strand (Nef8547) were separated by just 4 nucleotides (nts); the second from the third sequence on the forward strand (Nef8624) by 48 nts (Figure 5.16A-B). The identified G-quadruplex forming sequences comprised at least four tracts of two consecutive Gs, with the possibility to form G-quadruplexes with two stacked tetrads. Although in principle two G-tetrads would confer less stability than more extended G-tetrads, the existence and biological role of G-quadruplexes with two stacked tetrads has been reported in several cases (Raiber et al, 2012; Xu & Sugiyama, 2006). Interestingly, in certain instances, two-tetrads G-quadruplexes showed higher stability than three-tetrads G-quadruplexes (Hu et al, 2009; Phan, 2010). If these three sequences were able to fold in G-quadruplex, this region of the HIV-1 proviral genome could constitute an important cluster of non-canonical DNA structures with possible effects on polymerase processing (and therefore impact on replication and transcription events). In addition, in a supercoiled environment, I-motif conformations may arise in the C-rich complementary sequences independently of G-quadruplex formation (Brooks et al, 2010). We thus foresaw the possibility of creating a structured environment upon induction of non-canonical nucleic acid conformations, with the possibility of blocking enzymes involved in Nef protein expression. I-motif formation in this region will not be discussed in this thesis. It is interesting to note that all selected G-quadruplex putative sequences code for amino acids of the Nef core, which is the most conserved region of the protein and is essential for interaction with cellular proteins to mediate key viral functions (Cheng et al, 1999; Grzesiek et al, 1996; Lee et al, 1996) (Figure 5.16C-D).



Position in the F strand	Length	Putative G-quadruplex nef sequence	G-score
8528	15	<u><b>GGAGGAGGAGGTGGG</b></u>	17-21
8547	29	<u><b>CCAGTCACACCTCAGGTACCTTTAAGACC</b></u>	21
8624	16	<u><b>GGGGGACTGGAAGGG</b></u>	16-20
8727	30	<u><b>CCAGGGCCAGGGGTCAGATATCCACTGACC</b></u>	12

**Table 5.5 Putative G-quadruplex forming sequences within the HIV-1 nef gene (HIV-1 strain HXB2/LAI, NC\_001802) in the forward and reverse strand** Putative bases involved in G-quadruplex formation are in bold and underlined



**Figure 5.16 Putative G-forming regions in the HIV-1 nef coding region** A) Scheme of G-quadruplex formation within the double-stranded DNA of the nef region: Nef8528, Nef8547, Nef8624 G-quadruplex structures are shown in blue, green and red, respectively. The numbers of nts separating each G-quadruplex structure are indicated. The scheme indicates the possibility of formation of a cluster of non-canonical DNA structures within a small portion (112 nts) of the HIV-1 genome. B) Nucleotide sequence of the nef coding region where three putative G-quadruplex sequences were identified. Nef8528 is shown in blue and Nef8624 in red. Nef8547 was identified on the non-coding strand, thus the reverse complementary sequence is shown on the upper strand (in green). C) Scheme of the HIV-1 nef coding sequence with numbering referring to the HIV-1 strain HXB2/LAI, NC\_001802. D) Scheme of the aminoacidic sequence of the Nef protein indicating reported structural domains (Geyer et al, 2001). The protein moiety coded by the G-quadruplex rich nucleotide region is highlighted by the rectangular yellow shape, indicating involvement of the conserved N-terminal Nef core region. Note that the first three nts of the Nef8528 sequence exactly code for the first amino acid of the protein core region

To establish the importance of the identified sequences from a virus standpoint, the degree of conservation in terms of sequence and G-quadruplex formation among HIV-1 strains was assessed. Initially, the presence of the exact sequences identified in the HIV-1 HXB2/LAI strain was analysed in 3224 *nef* sequences of the HIV-1 M group reported in the HIV database (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>). Among these, 1538 sequences belonged to clade B, 612 to clade C, 486 to clade A, and 588 to other clades. As shown in Table 5.6, Nef8528, Nef8547 and Nef8624 were fairly well conserved in the M group, especially in the B subtype, where conservation was higher with respect to the other clades of the M group. Next, the possibility of G-quadruplex formation was statistically analysed by maintaining the number and size of G repeats, while varying loop regions. Two cases were considered: in instance i) loops could diverge in base composition while maintaining a constant length; in instance ii) both loop composition and length were allowed to vary. As shown in Table 5.6, Nef8547 and Nef8624 reached a degree of conservation higher than 95% in both cases across all considered HIV subtypes. Nef8528 was conserved to a significant extent in clades A and B (up to 66.3%); its presence was negligible only in clade C.

G-quadruplex name	G-quadruplex sequence or pattern	Conservation grade (%)			
		Group M	Clade A	Clade B	Clade C
Nef8528	<b>GGAGGAGGTGGG</b>	13.9%	0.6%	27.2%	0.3%
A)	G <sub>2</sub> X <sub>1</sub> G <sub>2</sub> X <sub>1</sub> G <sub>2</sub> X <sub>1-2</sub> G <sub>2</sub>	15.7%	2.4%	29.9%	0.5%
B)	G <sub>2</sub> X <sub>0-7</sub> G <sub>2</sub> X <sub>0-7</sub> G <sub>2</sub> X <sub>0-7</sub> G <sub>2</sub>	46.8%	57.6%	66.3%	9.3%
Nef8547	<b>CCAGTCAGACCTCAGGTACCTTTAAGACC</b>	24.1%	0.0%	39.2%	11.1%
A)	C <sub>2</sub> X <sub>7</sub> C <sub>2</sub> X <sub>7</sub> C <sub>2</sub> X <sub>7</sub> C <sub>2</sub>	98.6%	97.6%	99.1%	98.5%
B)	C <sub>2</sub> X <sub>1-10</sub> C <sub>2</sub> X <sub>1-10</sub> C <sub>2</sub> X <sub>1-10</sub> C <sub>2</sub>	99.0%	99.4%	99.1%	98.9%
Nef8624	<b>GGGGGACTGGAAGGG</b>	66.4%	13.3%	86.5%	58.8%
A)	G <sub>2</sub> X <sub>2</sub> G <sub>2</sub> X <sub>3</sub> G <sub>2</sub> X <sub>2-3</sub> G <sub>2</sub>	98.0%	95.2%	97.9%	98.7%
B)	G <sub>2</sub> X <sub>0-7</sub> G <sub>2</sub> X <sub>0-7</sub> G <sub>2</sub> X <sub>0-7</sub> G <sub>2</sub>	99.9%	99.4%	99.9%	100.0%

**Table 5.6. Statistical analysis of the conservation grade of the G-quadruplex *nef* sequences or their G-quadruplex patterns.** In bold bases possibly involved in G-quadruplex folding

Consensus sequences and base conservation in each position are reported in Table 5.7.

<b>Nef8528</b>	G	G	A	G	G	A	G	G	A	G	G	T	G	G	G														
Consensus seq	G	G	A	G	G	A	G	G	A	G	G	T	A	G	G														
% consensus	76	96	90	35	94	87	65	99	95	50	99	99	51	99	99														
<b>Nef8547</b>	C	C	A	G	T	C	A	C	A	C	C	T	C	A	G	G	T	A	C	C	T	T	T	A	A	G	A	C	C
Consensus seq	C	C	A	G	T	C	A	C	A	C	C	T	C	A	G	G	T	A	C	C	T	T	T	A	A	G	A	C	C
% consensus	99	99	97	99	99	91	98	85	80	99	99	75	99	99	86	99	99	77	99	99	86	79	98	90	99	99	91	99	99
<b>Nef8624</b>	G	G	G	G	G	G	A	C	T	G	G	A	A	G	G	G													
Consensus seq	G	G	G	G	G	A	C	T	G	G	A	A	G	G	G														
% consensus	98	99	99	99	99	99	99	99	99	99	99	99	68	99	99	99													

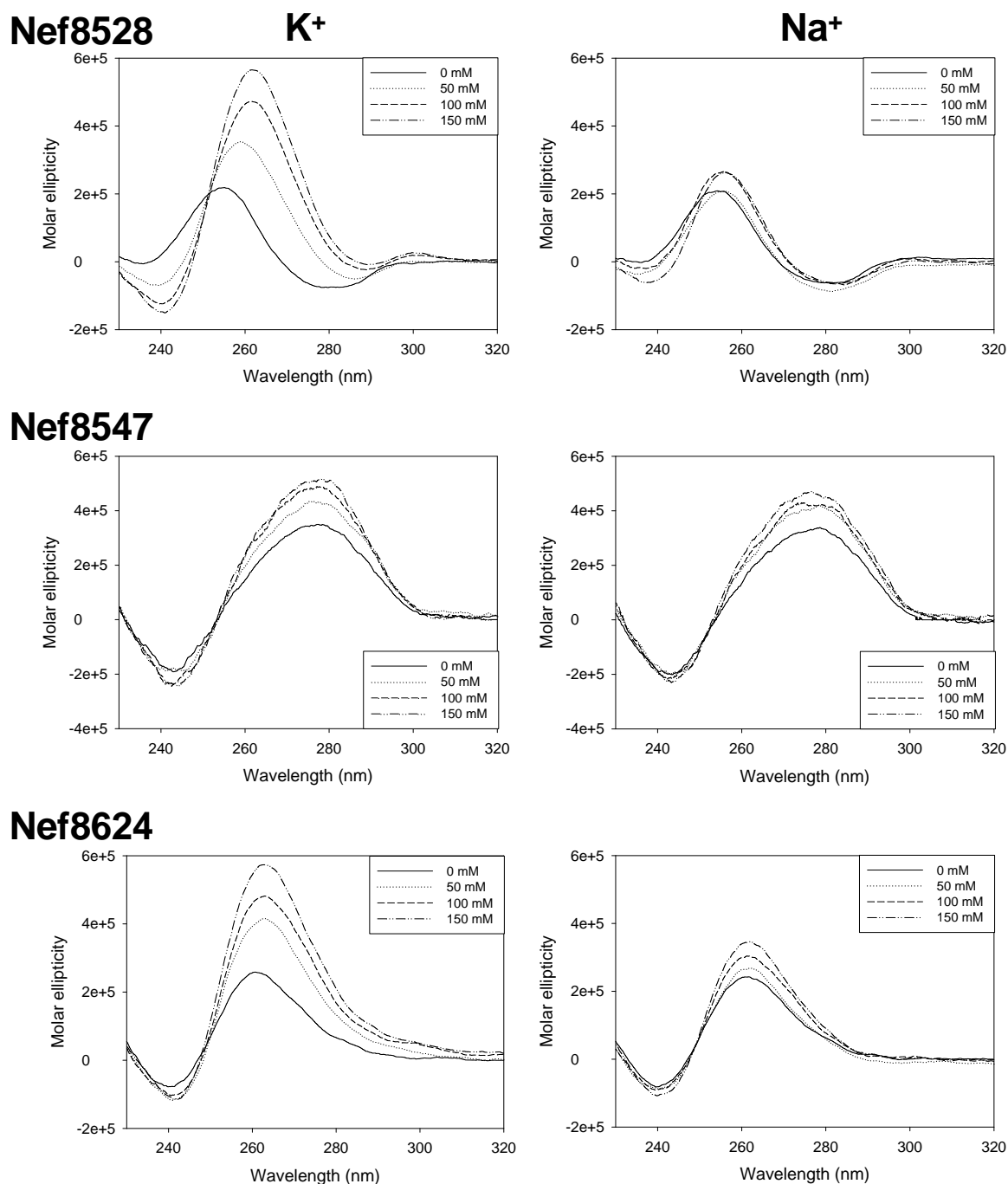
**Table 5.7** Consensus sequences and percentages of base conservation at each position

Overall these data indicate that the G-quadruplex pattern, therefore the possibility of G-quadruplex folding, in the selected G-rich sequences in the Nef coding region is extremely conserved among circulating HIV-1 strains, at least for Nef8547 and Nef8624.

### 5.3.2. Characterization of putative G-quadruplex forming sequences Nef8528, Nef8547 and Nef 8624

#### 5.3.2.1. CD spectroscopic analysis

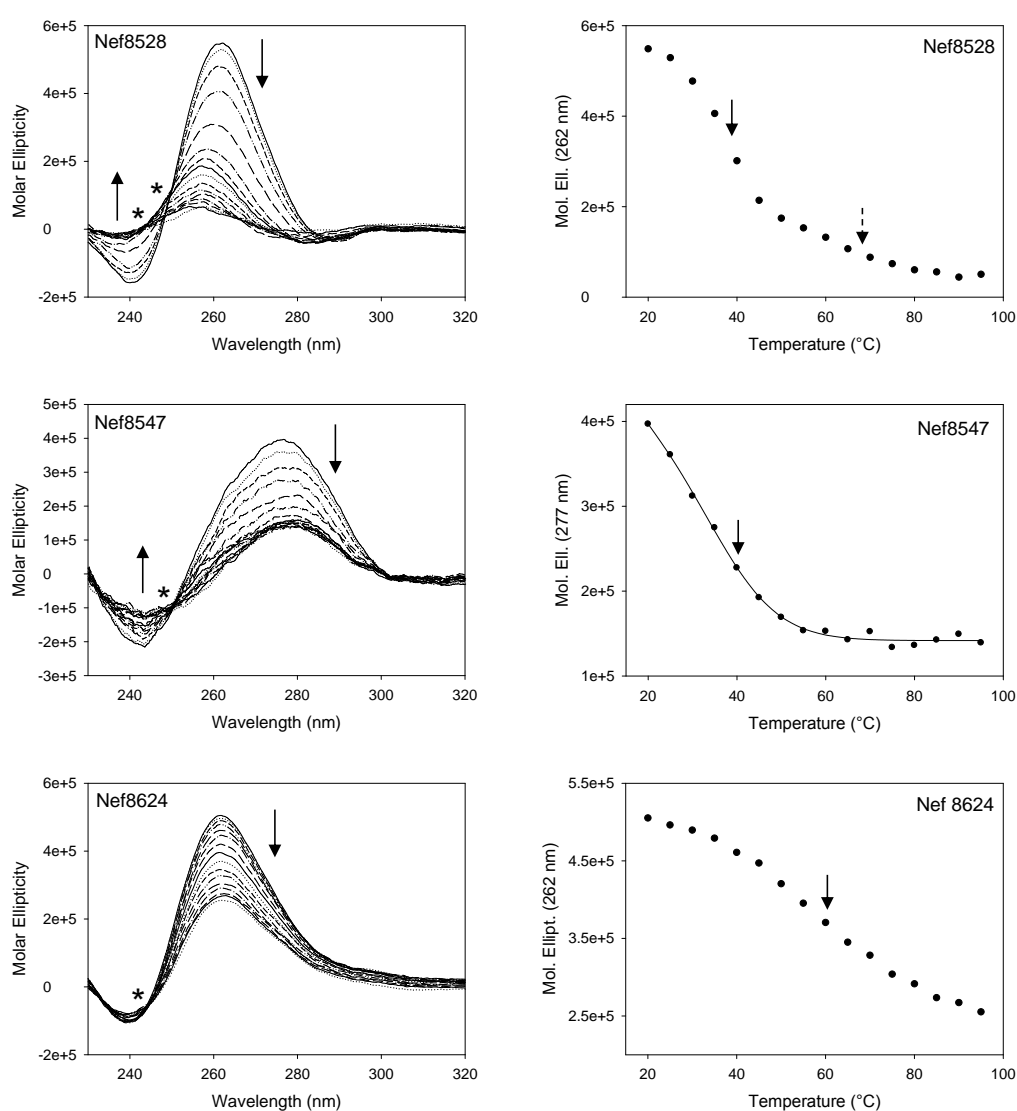
The actual folding of the selected nef sequences in a G-quadruplex conformation was initially assessed by circular dichroism (CD) spectroscopy. In the case of Nef8528, which presents five GG repeats, we used the minimal sequence that could fold into G-quadruplex. The three nef sequences were incubated in the presence of increasing concentrations of K<sup>+</sup> or Na<sup>+</sup>. Both cations increased the CD signal, with K<sup>+</sup> exhibiting a remarkably higher effect than Na<sup>+</sup>. Nef8528 and Nef8624 displayed a clear parallel-type conformation upon addition of K<sup>+</sup> and Na<sup>+</sup>, with a maximum at 260 nm and a minimum at 240 nm (Figure 5.17). Conversely, Nef8547 presented a maximum at around 275 nm and a negative peak at 240 nm, with low K<sup>+</sup>/Na<sup>+</sup> dependence. Therefore, the 7-nt-long loop Nef8547 apparently did not naturally fold into a canonical G-quadruplex conformation: to note, however, that similar CD spectra have been reported for G-quadruplex forming oligonucleotides with long loops (i.e. at least two loops  $\geq 5$  nts) (Guedin et al, 2010).



**Figure 5.17** CD spectra of the putative G-quadruplex forming oligonucleotides in the nef region For each oligonucleotide, CD spectra were measured in the absence or presence of increasing concentrations (50-150 mM) of  $K^+$  or  $Na^+$  cations

Stability of the G-quadruplex structures was next evaluated by CD thermal unfolding and  $T_m$  were calculated as the first derivative of the melting profiles (Figure 5.18 and Table 5.8). In all cases the CD signal decreased with increasing temperature. For Nef8547 and Nef8624 a single transition between 25°C and 95°C was appreciable leading to discrete  $T_m$  values (35°C and 59°C, respectively, Table 5.8). For Nef8528 two transitions were present: a first structural

variation at 39°C and a second at  $T_m > 50^\circ\text{C}$ , where a clear inflection point was not observed. This behaviour could also be evinced by spectra overlapping, where two isosbestic points (asterisks in Figure 5.18) were detected, indicating the presence of at least three spectroscopically distinct species: the initial G-quadruplex structure, a second folded form, likely a more flexible G-quadruplex conformation, and the unfolded random coiled structure. In addition,  $T_m$  values measured by UV thermal unfolding at oligonucleotide concentration of 40  $\mu\text{M}$  were very similar to those obtained at 4  $\mu\text{M}$  (Table 5.8), indicating a prevalent intramolecular G-quadruplex folding.



**Figure 5.18** CD thermal unfolding of the G-quadruplex nef oligonucleotides CD spectra measured at increasing temperatures (25-95°C) are shown on the left. Arrows indicate spectral trends at the corresponding wavelengths. Asterisks indicate isosbestic points. Plots of molar ellipticity values (black circles) measured at the indicated wavelength (corresponding to positive peaks) as a function of temperature are reported on the right. Arrows indicate  $T_m$  points

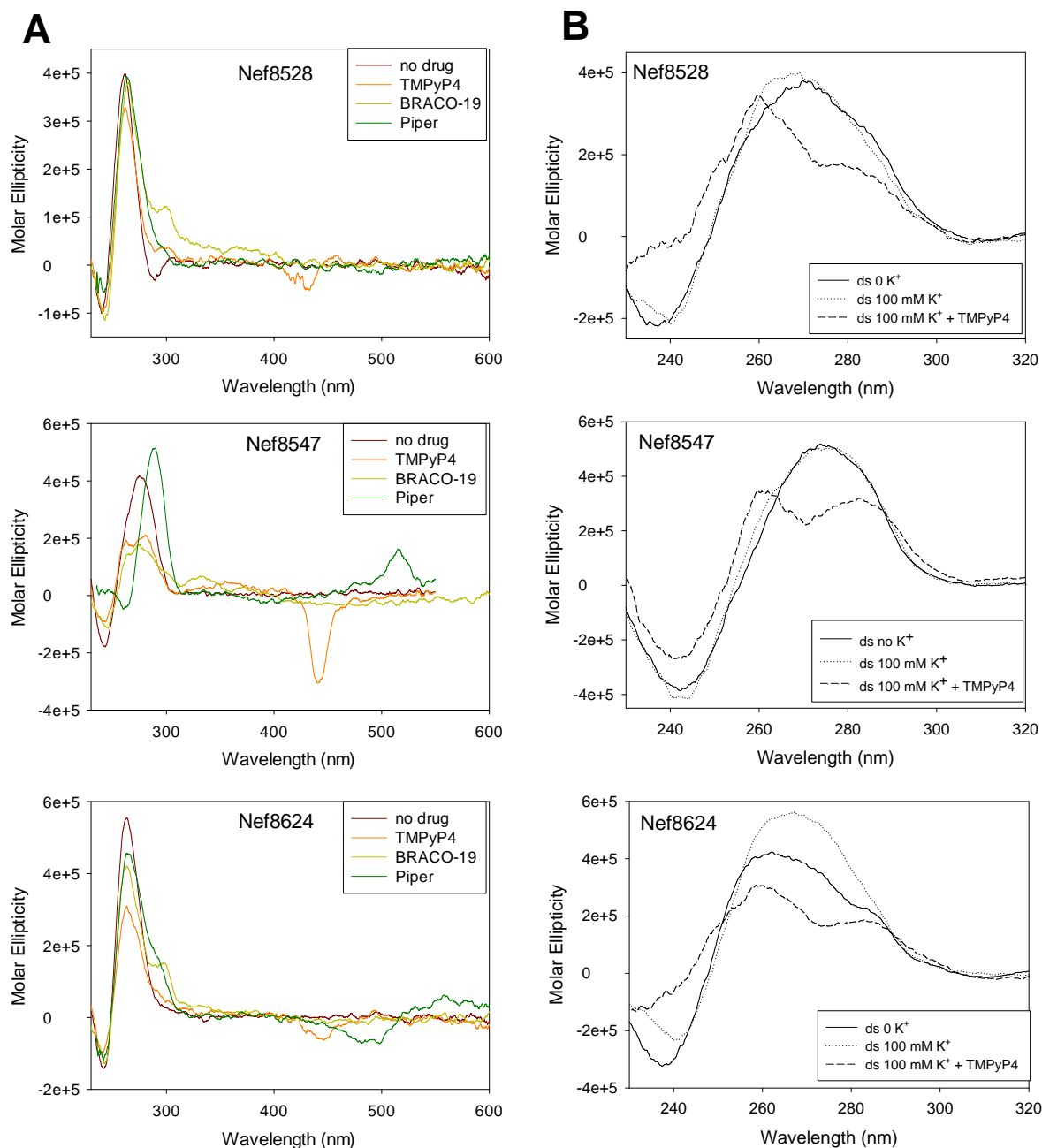
Three G-quadruplex ligands with different central cores (i.e. porphyrin for TMPyP4, acridine for BRACO-19, and perylene for PIPER) were incubated in the presence of each of the three nef G-rich oligonucleotides to check for the compound ability to induce/stabilize the G-quadruplex topology. CD thermal unfolding analysis was employed to check stabilization of the G-quadruplex conformation imposed by the G-quadruplex ligands. G-quadruplex ligands highly stabilized the G-quadruplex conformations of Nef8528 and Nef8547 (Table 5.8). Nef8624 in general was less efficiently stabilized, likely due to the higher innate stability of this latter oligonucleotide. In cases where two transitions were observed, T<sub>m</sub> values for each transition were reported (Table 5.8).

G-quadruplex DNA	Drug added	T <sub>m</sub> (°C) 4μM *	ΔT <sub>m</sub> (°C) 4μM **	T <sub>m</sub> (°C) 40μM *
Nef8528	-	39 / > 50	-	41
	TMPyP4	57	18	-
	BRACO	69	30	-
	Piper	80	41	-
Nef8547	-	35	-	37
	TMPyP4	56	21	-
	BRACO	48	13	-
	Piper	56	21	-
Nef8624	-	59	-	58
	TMPyP4	72	13	-
	BRACO	64 / 75	5 / 16	-
	Piper	61 / >100	2 / > 40	-

**Table 5.8 T<sub>m</sub> and ΔT<sub>m</sub> of the three G-quadruplex nef sequences (4 μM and 40 μM) in the absence and presence of G-quadruplex ligands (16 μM) measured by CD and UV spectroscopy.**\*Average standard deviation was 0.3. \*\*Average standard deviation was 0.4.

Moreover, CD spectra of all three oligonucleotides in the presence of G-quadruplex ligands were evaluated and resulted characteristic of G-quadruplex conformations. In particular, Nef8528 and Nef8624 maintained the initial parallel-like topology with TMPyP4 and PIPER; in the presence of BRACO-19 an additional positive peak appeared at 290 nm, which is characteristic of a G-quadruplex topology and might depict a shifting towards a hybrid topology. In the case of Nef8547, PIPER induced an antiparallel-like spectrum, while TMPyP4 and BRACO-19 stabilized hybrid-type G-quadruplex structures (Figure 5.19A), indicating that G-quadruplex ligands are able to drive Nef8547 folding into a G-quadruplex conformation. To note that in most cases an induced CD spectrum was observed in the UV/Vis absorption region of the ligands, further confirming oligonucleotide/compound interaction.

Since the G-rich nef sequences in the proviral genome are normally embedded in a DNA double-helix, TMPyP4 was incubated with each nef oligonucleotide in the presence of its complementary counterpart in order to assess the ability of a G-quadruplex ligand to promote G-quadruplex folding from a double-stranded (ds) DNA. CD spectra of the ds oligonucleotides in the absence or in the presence of K<sup>+</sup> were very similar and characteristic of a B-DNA (Kypr et al, 2009). However, when TMPyP4 was added, the CD spectra clearly shifted, presenting two maxima at 290 and 260 nm, which are indicative of a G-quadruplex conformation (Figure 5.19B). Since TMPyP4 absorbance below 300 nm is extremely low (Morris et al, 2012), the observed molar ellipticity variation must be due to changes in the absorbance of the nucleic acid. These data altogether confirm the ability of the nef sequences to fold in G-quadruplex: G-quadruplex ligands can bind, induce and stabilize their G-quadruplex conformations. Moreover, our results show that nef sequences are normally present in a B-DNA conformation within the double-helix, and that a G-quadruplex binder is able to induce their folding in G-quadruplex.



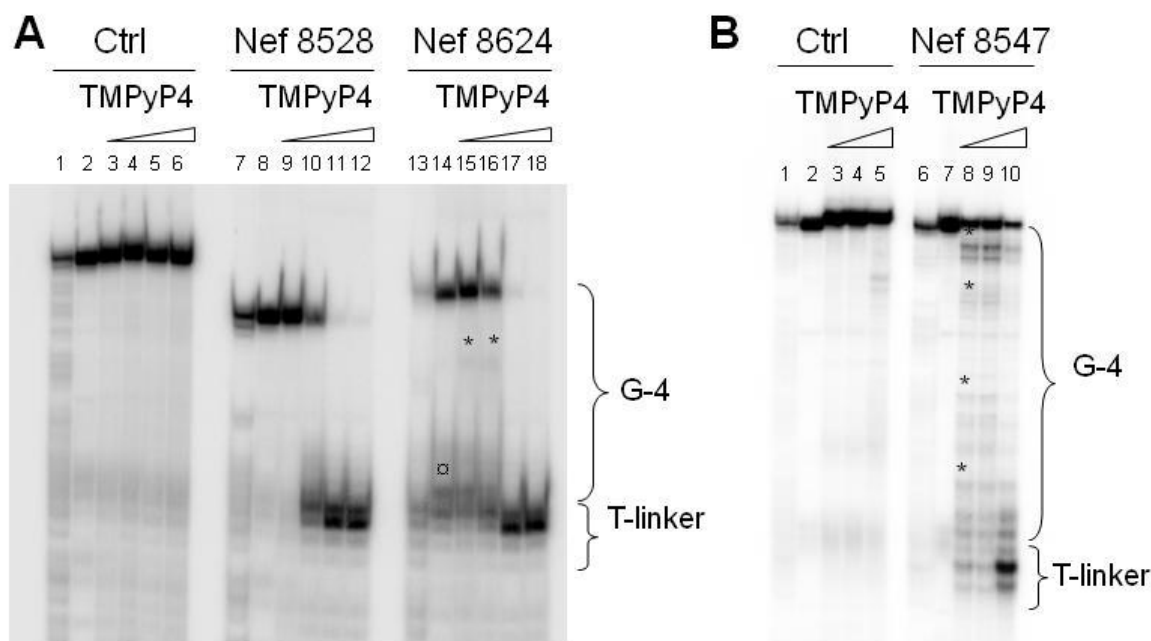
**Figure 5.19** CD spectra of G-quadruplex *nef* single-stranded or double-stranded oligonucleotides in the presence of G-quadruplex ligands A) CD spectra of G-quadruplex *nef* single-stranded oligonucleotides in the presence of TMPyP4, BRACO-19 or PIPER. Addition of ligands stabilized G-quadruplex conformations and generated ICD bands in the UV/Vis absorption regions of G-quadruplex ligands. B) CD spectra of G-quadruplex *nef* doublestranded oligonucleotides in the presence or absence of K<sup>+</sup> and TMPyP4. Addition of the G-quadruplex binding compound induced shifting from double-stranded DNA spectra to mixed type G-quadruplex signatures in all three cases



### 5.3.2.2. *Taq* polymerase stop assay

G-quadruplex folding in the selected sequences was additionally proved by the *Taq* polymerase stop assay. The three G-quadruplex *nef* oligonucleotides were designed in order to contain additional flanking bases at both the 5'- and 3'-ends (see Materials and Methods Section); in particular, an additional sequence at their 3'-end was used as primer annealing region. A 4-T linker region was added to separate the 3'-end of the primer and the first G of the G-quadruplex portion. An additional oligonucleotide lacking the possibility to fold in G-quadruplex was designed and used as negative control. Primer annealing and G-quadruplex folding were obtained by incubating the template G-quadruplex forming oligonucleotides and the primer in  $K^+$  buffer at 95°C and slowly cooling down to room temperature. *Taq* polymerase was incubated with the different template/primer combinations in the presence of increasing amount of TMPyP4. As shown in Figure 5.20A, in the presence of both Nef8528 and Nef8624 templates, increasing drug concentrations induced arrest of the DNA polymerase processing at the T linker region, just before the G-quadruplex folded region (lanes 10-12 and 17-18, Figure 5.20A), while no effect was detected in the negative control (lanes 2-6, Figure 5.20A). In the case of the Nef8624 template, at low amounts of TMPyP4, a polymerase pausing site was observed corresponding to the second G-tract of the Nef8624 oligonucleotide (\* symbol, lanes 15-16, Figure 5.20A). In addition the polymerase was partially inhibited also in the absence of TMPyP4 (α symbol, lane 14, Figure 5.20A). We ascribed this behavior to the fact that the G-quadruplex conformation of Nef8624 was inherently very stable in the presence of  $K^+$  ( $T_m$  59.8°C, see Table 5.8) and thus could affect polymerase activity even without G-quadruplex ligands, similarly to other reported G-quadruplex structures (Palumbo et al, 2009). In the case of the Nef8547 template, several stop sites were observed in the G-quadruplex forming template, while the polymerase was not inhibited in the control (compare lanes 2-5 and 7-10, Figure 5.20B). In particular, stop sites were clustered at G bases at low TMPyP4 concentrations (\* symbols, lane 8, Figure 5.20B) and at the T-linker at higher G-ligand amounts (lane 10, Figure 5.20B).

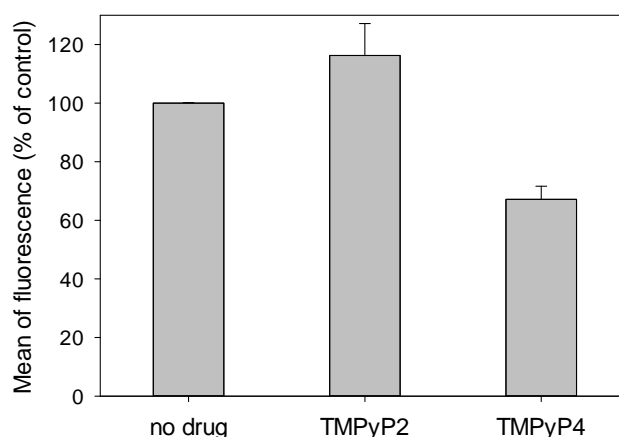
Overall these data indicate that a G-quadruplex binder can induce and stabilize the G-quadruplex conformations at the *nef* DNA level.



**Figure 5.20 *Taq* polymerase stop assay** A) and B) Templates containing the G-quadruplex nef sequences Nef8528, Nef8624 and Nef8547, a 4-T linker and a primer annealing region were allowed to fold and anneal to the P32-5'-end labelled primer in  $K^+$  100 mM, treated with increasing concentrations of TMPyP4 (0-2  $\mu$ M) and subjected to *Taq* polymerase extension. The control template contained a sequence unable to fold in G-quadruplex, and the same 4-T linker and primer annealing region as the nef templates. A) The \* symbol indicates pausing sites in the G-quadruplex region of nef templates. The ◻ symbol indicates a polymerase stop site obtained prior to addition of TMPyP4 in Nef8624. Lanes 1, 7 and 13 (A), and lanes 1 and 6 (B) were Maxam and Gilbert marker lanes performed on the double stranded PCR amplified region. Markers indicate the C-rich complementary strand.

### 5.3.3. Biological significance of G-quadruplexes in the HIV-1 *nef* coding region

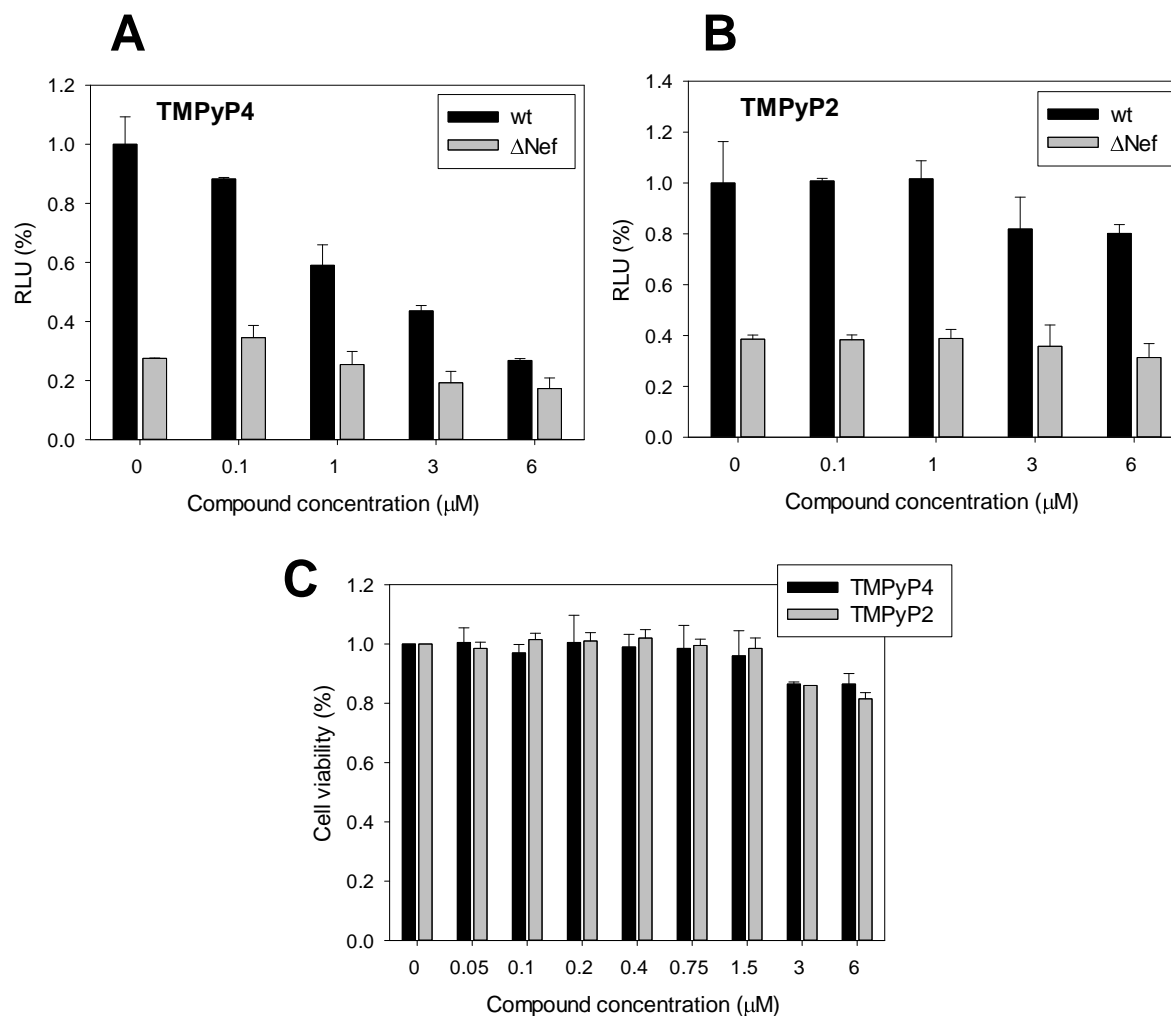
To check the effect of G-quadruplex stabilizing ligands on gene expression, a eGFP-based reporter gene assay was set up. The *nef* gene was cloned upstream of a eGFP coding region in a plasmid optimized for protein expression in mammalian cells. In principle, G-quadruplex folding in the *nef* sequence should impair the polymerase activity on the DNA template, therefore reducing expression of the fused Nef-eGFP protein. Transfected cells were treated with TMPyP4 (10  $\mu$ M) or the control compound, TMPyP2, which is not able to bind G-quadruplexes; both compounds did not show toxicity on HEK293T cells up to 100  $\mu$ M (data not shown). eGFP fluorescence was quantified by flow cytometry-based analysis. As shown in Figure 5.21, mean of fluorescence consistently decreased in the presence of TMPyP4, while slightly increased upon treatment with the control TMPyP2 indicating a selective impairment of gene expression mediated by the interaction of the G-quadruplex ligand with the *nef* sequence.



**Figure 5.21. Effects of the stabilization of the nef G-quadruplexes by TMPyP4 on gene expression** Effect of TMPyP4 and TMPyP2 on Nef- GFP expression measured by flow-cytometry. HEK 293T cells were transfected with a Nef-GFP encoding plasmid and treated with TMPyP4 or TMPyP2 (10  $\mu$ M) for 24 h. Results are shown as percent mean of fluorescence relative to the control cells incubated  $\pm$  SD (n = 4). Statistical difference was observed for TMPyP4 ( $p < 0.05$ ), but not for TMPyP2.

#### 5.3.4. The G-Quadruplex Ligand, TMPyP4, displays antiviral activity

Having observed an impairment of Nef protein expression levels, the effect of G-quadruplex stabilizing ligands was next assessed in HIV-1 infected cells. As Nef has been previously shown to enhance HIV-1 infectivity (Aiken & Trono, 1995; Chowers et al, 1994; Spina et al, 1994; Vermeire et al, 2011), we evaluated the impact of TMPyP4 on infectivity using the TZM-bl reporter cell line. As anticipated, TZM-bl cells support HIV-1 replication in a Nef-dependent manner and contain a luciferase reporter driven by the HIV-1 LTR in response to infection with HIV-1 (Derdeyn et al, 2000; Emert-Sedlak et al, 2013). As shown in Figure 5.22A, TMPyP4 impaired the enhancement of Nef-mediated viral infectivity in a dose-dependent manner, while the negative control (TMPyP2) had no effect (Figure 5.22B). In addition, a Nef-deleted virus ( $\Delta$ Nef) showed no significant effects by treatment with either TMPyP4 or TMPyP2 at the highest concentrations (Figure 5.22A-B). Cytotoxicity assays further confirmed the observed impairment of viral infectivity as minimal toxicity was observed over the range of concentrations tested (0.1-6  $\mu$ M) (Figure 5.22C). Taken together, these results provide support for a G-quadruplex mediated, Nef-directed anti-retroviral mechanism of action. Since the Nef protein has a fundamental function *in vivo* and its lack of activity prevents progression to the clinical development of AIDS (Salvi et al, 1998), its G-quadruplex mediated depletion could have critical antiviral effects and open a new avenue in the development of anti-Nef compounds with an unprecedented mechanism of action.



**Figure 5.22. Effects of the stabilization of the nef G-quadruplexes by TMPyP4 on viral infectivity in Nef sensitive cells** A) and B) TZM-bl cells were infected with wild-type (black bars) and  $\Delta\text{Nef}$  (grey bars) HIV NL4-3 in the presence of either the G-quadruplex ligand, TMPyP4 (A), or the negative control compound, TMPyP2 (B). After 48 h, infectivity was assessed as relative luciferase activity in infected cells. Results are shown as percent infectivity relative to the control cells incubated with carrier solvent (DMSO)  $\pm$  SEM ( $n = 3$ ). In A), no statistical difference was observed across  $\Delta\text{Nef}$  infected cells, even at the highest concentration ( $p > 0.15$ ). In B), no statistical difference was observed for the wild-type virus at 3  $\mu\text{M}$  and 6  $\mu\text{M}$  ( $p > 0.345$  and  $> 0.325$ , respectively) relative to the untreated control. The negative control compound, TMPyP2, further had no impact on the  $\Delta\text{Nef}$  virus at any concentration tested ( $p > 0.29$  at 6  $\mu\text{M}$ ). C) TZM-bl cell viability in the presence of compounds was assessed via the Cell-Titer Blue assay (Promega). TZM-bl cells were incubated with the indicated concentrations of compounds for 48 h and cell viability was assessed via the Cell-Titer Blue assay relative to control cells incubated with carrier solvent. Assays were done in triplicate.

## 5.4. Evaluation of anti HIV-1 activity of G-quadruplex ligands

### 5.4.1. Antiviral activity of BRACO-19

We discovered several putative G-quadruplex forming sequences similarly distributed along all the HIV-1 proviral genome (Figure 5.1) and thus we tried to assess biological the relevance of these G-quadruplexes in the viral context. Interestingly, we demonstrated that G-quadruplexes formation and stabilization in LTR promoter (see paragraph 5.2) and in the *nef* coding region (see paragraph 5.3) perturbed viral processes. In addition, the commercially available G-quadruplex ligand BRACO-19 (Table 5.9) displayed antiviral activity against HIV-1 in infected TZM-bl cells and a significant inhibition of HIV-dependent gene expression (Figure 5.15). This promising antiviral activity of BRACO-19 prompted us to carry out further studies on its mechanism of action.

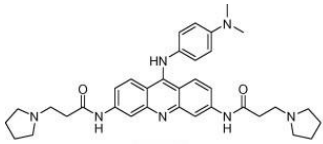
Class of compound	Name	Molecular Weight	Chemical structure
Trisubstituted Acridine	BRACO-19	593.77	

Table 5.9 The G-quadruplex ligand BRACO-19

#### 5.4.1.1. HIV-1 infected MT-4 cells

##### 5.4.1.1.1. Antiviral assays

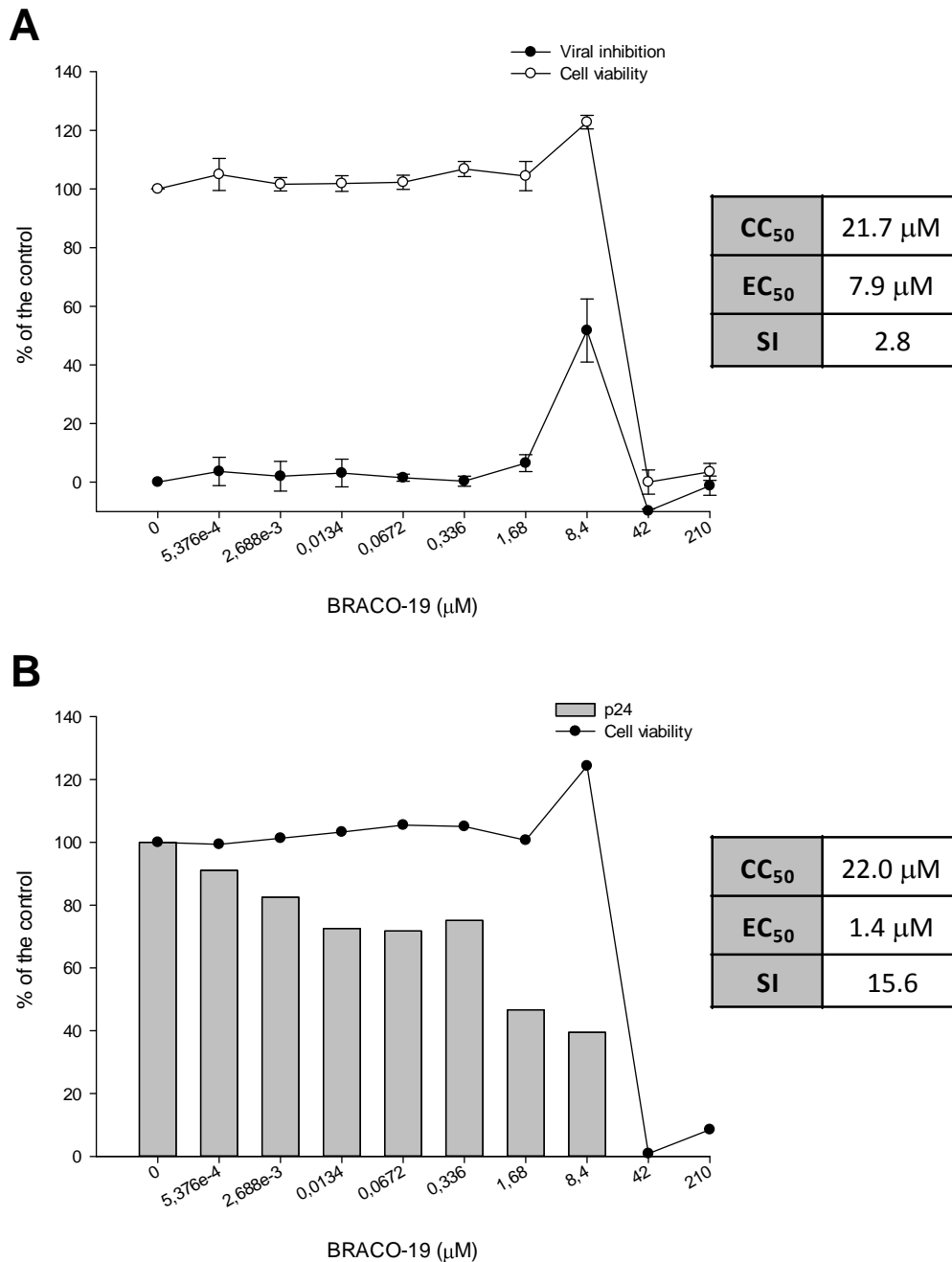
To further confirm the antiretroviral activity of BRACO-19, we investigated the effect of this compound in other HIV-1-infected cell lines.

We started to evaluate BRACO-19 for its antiretroviral activity against HIV-1 strain III<sub>B</sub> (HIV-1(III<sub>B</sub>)). The inhibition of viral replication was first monitored by evaluating the protection from HIV-induced cytopathogenic effect (CPE) in MT-4 cells 5 days after infection. This MTT-based assay is relatively simple to perform and allows to screen large numbers of compounds evaluating in parallel the cytotoxicity in mock-infected cells. Thus, it is possible to obtain both the 50% cytotoxic concentration (CC<sub>50</sub>) and the 50% effective concentration (EC<sub>50</sub>). The CC<sub>50</sub> was defined as the concentration of test compound that

induced cytotoxicity in 50% of mock-infected while the  $EC_{50}$  was defined as the concentration of the test compound that protect the 50% of the HIV-infected cells from CPE. In addition, it is possible to evaluate the maximum percentage of protection of infected cells from HIV-induced CPE. Having the  $CC_{50}$  and the  $EC_{50}$  values, it is possible to calculate the selectivity index (SI) that is the relative effectiveness of the tested product in inhibiting viral replication compared to inducing cell death ( $CC_{50}$  value/ $EC_{50}$  value). Obviously, it is desirable to have a high SI giving maximum antiviral activity with minimal cytotoxicity.

BRACO-19 yielded  $EC_{50}$  of 7.9  $\mu$ M showing a maximum of protection from CPE of around 51% with average SI around 3 (Figure 5.23A). Interestingly, we noticed that BRACO-19 induced an increase in cell viability in mock-infected cells and this effect seemed correlated with the observed antiviral effect. In fact, cell viability reached 120% with respect to the control at 8.4  $\mu$ M, in correspondence to the peak of antiviral effect (Figure 5.23A). This effect observed in MT-4 cells was completely absent in TMZ-bl cells (Figure 5.15C) and suggested a dual effect of BRACO-19, both at cellular and viral level. This unexpected effect of BRACO-19 in promoting cell growth and viability has never been reported. In fact, it has been demonstrated that BRACO-19 targets telomeric G-quadruplex inducing cellular senescence after 15 days in uterus carcinoma cell line UXF1138L (Burger et al, 2005). A similar effect was observed in DU145 prostate cancer cell lines where BRACO-19 blocked cell growth and induced senescence after 7 days of treatment (Incles et al, 2004). However, in our case we observed the opposite effect of BRACO-19 on MT-4 cells: this aspect will be further investigated in a different work. Overall these data confirmed that BRACO-19 retained its antiviral activity also 5 days after infection in HIV-1 infected MT-4 cells and was able to partially reduce the viral induced CPE up to 50 %. Apart from the evaluation of viral induced CPE that is an indirect measure, it is possible also to measure inhibition of HIV-1 p24 antigen production. The measurement of p24 protein in the supernatant of HIV-1 infected MT-4 cells provides a direct and precise information about the antiviral effect of a test compound. Thus, we decided to evaluate the p24 amount of MT-4 infected cells and treated with increasing concentrations of BRACO-19. We observed a progressive dose-dependent decrease of p24 amount in presence of BRACO-19 that resulted in an  $EC_{50}$  of 1.4  $\mu$ M and in an increased SI of 15.6 (Figure 5.23B). This evidence suggested that the antiviral effect of BRACO-19 was not fully detectable with the MTT-based assay evaluating the inhibition of CPE. Moreover, BRACO-19 seemed to produce a decrease in p24 starting from very low concentration: in fact a decrease of about 30% in p24 amount was observed in presence of

BRACO-19 0.013  $\mu\text{M}$ . This decrease was more pronounced at 1.68  $\mu\text{M}$  reaching 47% (Figure 5.23B).

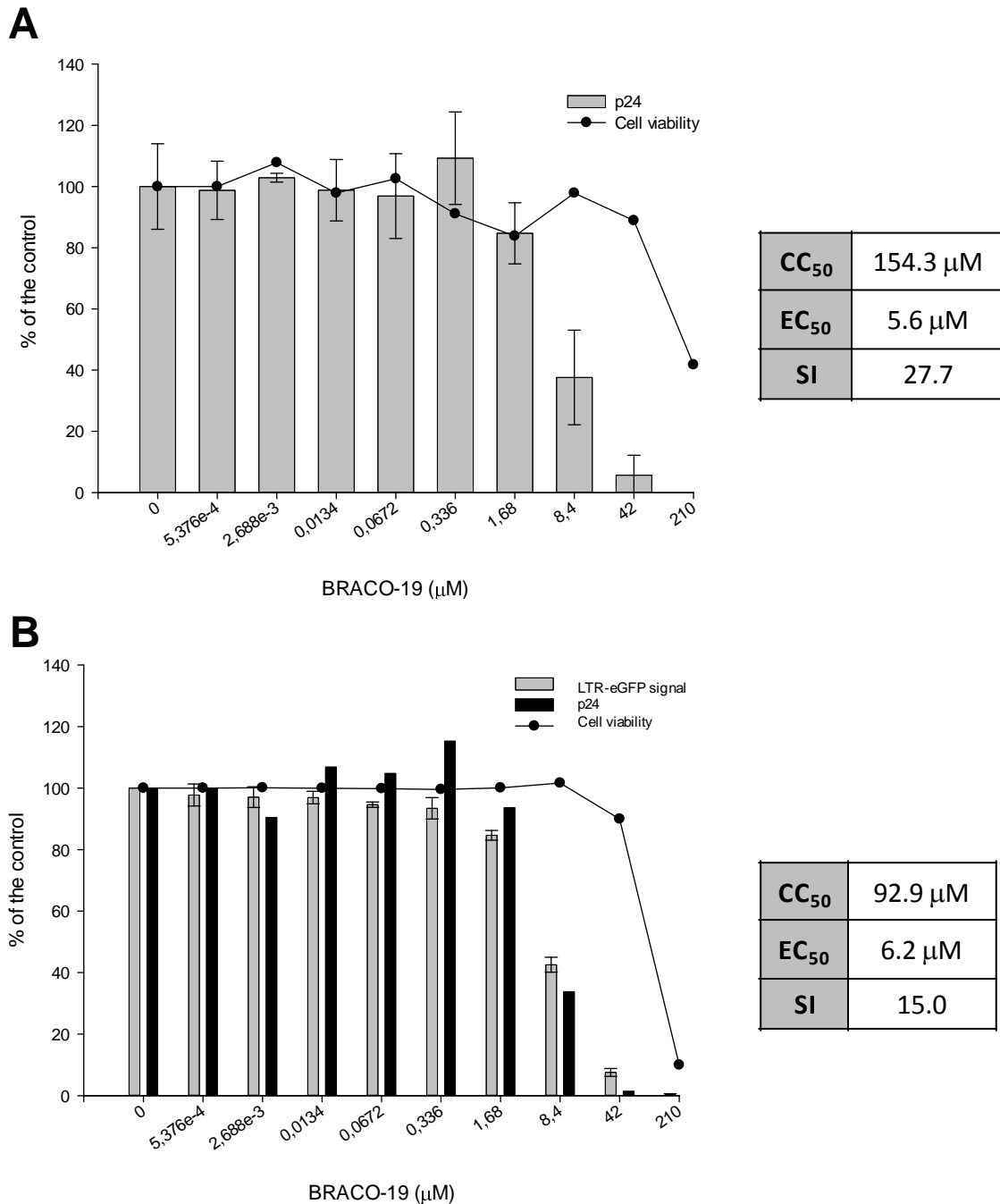


**Figure 5.23 Antiviral effect of BRACO-19 on MT-4 infected with HIV-1(III<sub>B</sub>) 5 days post infection** MT-4 cells were infected with HIV-1(III<sub>B</sub>) in the presence of the G-4 stabilizing ligand, BRACO-19. A) After 5 days, the viral inhibition was assessed as % of protection against viral induced CPE using an MTT-based assay. Cytotoxicity of BRACO-19 on MT-4 cells was assessed via MTT-assay in parallel. Results are presented as mean $\pm$ std dev from 2 independent experiments each in triplicate. The CC<sub>50</sub>, EC<sub>50</sub> and SI values are shown in the inset table. B) After 5 days, the viral inhibition in one series of samples treated with increasing concentrations of BRACO-19 was determined as % of p24 antigen present in the supernatant compared to the control. Cytotoxicity of BRACO-19 on MT-4 cells was assessed via MTT-assay in parallel. Results are presented as values of p24 measured in a single series of samples. The CC<sub>50</sub>, EC<sub>50</sub> and SI values are shown in the inset table.

The antiviral activity of BRACO-19 against HIV-1(III<sub>B</sub>) in MT-4 cells was further tested 1 day post infection. This allowed to evaluate the effect of the test compound in a single replication cycle that takes about 24 h. In this assay the antiviral effect was based on the inhibition of HIV-1 p24 antigen production in MT-4 infected cells, since no evaluation of CPE is possible in these conditions. Again, BRACO-19 displayed a significant antiviral activity with an EC<sub>50</sub> in the micro molar range (5.6 μM). The SI was 28. The increased SI was due to the fact that BRACO-19 was less cytotoxic 24 h post infection (CC<sub>50</sub> 154 μM) suggesting that the cytotoxicity is the main limiting factor for using this compound in this cellular system (Figure 5.24A).

To further confirm the antiviral activity of BRACO-19 at 1 day post infection, the MT-4-LTR-eGFP cellular system was used. In this cell line, the HIV infection drives transcription of the HIV-1 LTR-eGFP reporter gene construct and thus it is possible to evaluate the antiviral effect of test compounds following the GFP expression. Moreover, it is possible to measure directly the p24 antigen in the supernatant to have a further confirmation of the results. At 8,4 μM BRACO-19 caused a decrease in GFP signal of about 58% and in p24 amount of about 67%, resulting in an EC<sub>50</sub> of 6.2 μM. Overall the dose-dependent decrease in eGFP signals appeared in line with the progressive decrease in p24 amounts with increasing concentrations of BRACO-19. Again the SI value of 15 confirmed a significant effect of BRACO-19 in inhibiting HIV-1(III<sub>B</sub>) virus production (Figure 5.24B). Interestingly, BRACO-19 did not induce any increase in cell viability in mock-infected cells 1 day post infection, both in MT-4 and in MT-4-LTR-eGFP cells. This evidence suggested that the antiviral effect of BRACO-19 was not correlated with the cellular effect on cell viability previously observed 5 days post infection (Figure 5.23).



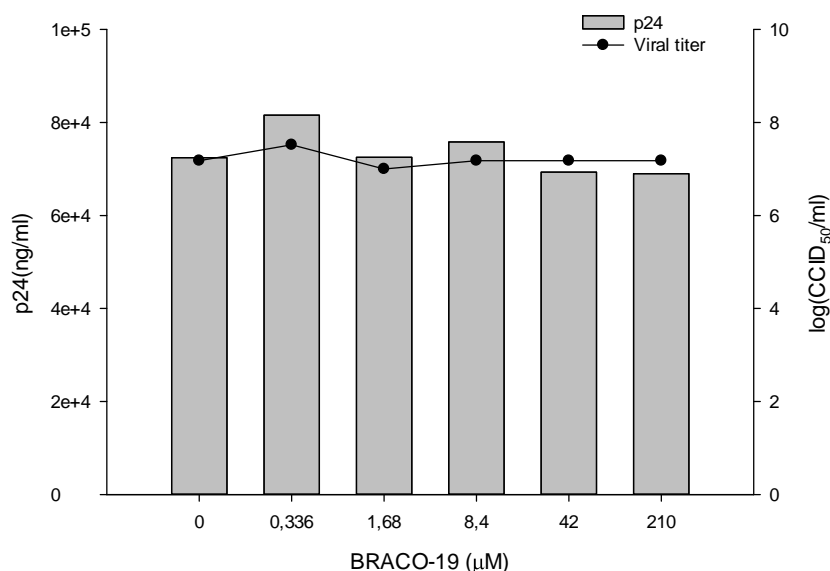


**Figure 5.24 Antiviral effect of BRACO-19 on MT-4 and MT-4-LTR-eGFP infected with HIV-1(III<sub>B</sub>) 1 day post infection** A) MT-4 cells were infected with HIV-1(III<sub>B</sub>) in the presence of the G-4 stabilizing ligand, BRACO-19. After 1 day, viral inhibition was assessed as % of p24 antigen present in the supernatant compared to the untreated control. Cytotoxicity of BRACO-19 on MT-4 cells was assessed via MTT-assay in parallel. Results are presented as mean±std dev from 3 independent experiments each in triplicate. The CC<sub>50</sub>, EC<sub>50</sub> and SI values are shown in the inset table. B) MT-4-LTR-eGFP cells were infected with HIV-1(III<sub>B</sub>) in the presence of the G-4 stabilizing ligand, BRACO-19. After 1 day, the viral inhibition was assessed as % of eGFP signal measured by Flow Cytometry compared to the untreated control. Results are presented as mean±std dev from 2 independent experiments. Viral inhibition in one series of samples treated with increasing concentrations of BRACO-19 was evaluated as % of p24 antigen present in the supernatant compared to the control cytotoxicity of BRACO-19 on MT-4-LTR-eGFP cells was assessed via MTT-assay in parallel. CC<sub>50</sub>, EC<sub>50</sub> and SI values are shown in the inset table.

Therefore, we demonstrated a pronounced antiviral effect of BRACO-19 both 1 and 5 days post infection in HIV-1 infected MT-4 cells. The maximum SI was observed in MT-4 cells infected with HIV-1(III<sub>B</sub>) 1 day post infection (Figure 5.24A).

#### 5.4.1.2. Virucidal assay

Since BRACO-19 is a G-quadruplex ligand, we hypothesized an intracellular viral target responsible for its antiretroviral effect. However, BRACO-19 could be active against the infectious viral particle preventing the virus entry process. To exclude that the observed activity of BRACO-19 was due to an effect on the viral particle prior infection, an assessment of the virucidal effect was performed. In fact, if the compound has a virucidal effect on HIV-1 particles, a compound-pretreated and subsequently compound-cleared virus stock should have a diminished infectivity in MT-4 cells. The virus infectivity can be then evaluated by titration 5 days post infection and expressed as the 50% cell culture infective dose (CCID<sub>50</sub>). As reported in Figure 5.25, BRACO-19 had no effect on the infectivity of HIV-1(III<sub>B</sub>) virions. In addition, treatment with BRACO-19 did not affect the amount of virus associated p24 core protein in the viral stock prior infection (Figure 5.25). These evidences suggested that BRACO-19 did not interfere with the HIV-1 virions.



**Figure 5.25 Virucidal assay for BRACO-19 against HIV-1(III<sub>B</sub>) viral particles** Aliquots of HIV-1(III<sub>B</sub>) stock were pretreated with increasing concentration of BRACO-19 and subsequently compound-cleared. These samples were then diluted and used to infect MT-4 cells. 5 days after infection, viral infectivity was quantified by titration and expressed as logCCID<sub>50</sub>/ml. A part of pretreated viral aliquots were tested in parallel for virus associated p24.

#### 5.4.1.3. Viral Binding Assay

As anticipated in the introduction of this thesis, BRACO-19 is a trisubstituted acridine with protonable amine-containing sidearms (Table 5.9), ensuring a good water solubility of the compound. Depending on the protonation state, BRACO-19 can possess two positive charges becoming a cationic ligand. It has been reported that compounds with multiple positive charges can specifically block viral entry. In particular, CXCR4-specific inhibitors are characterized by several positive charges that specifically interact with acidic residues in the extracellular loops of the coreceptor. These loops are not conserved in the coreceptor CCR5 that seems less affected by the presence of cationic groups in its antagonist (Wilkinson et al, 2011). Since in the performed antiviral assays we used HIV-1(III<sub>B</sub>) that is a X4 strain (a virus that needs the CXCR4 coreceptor for its replication), we wanted to exclude an effect of BRACO-19 on viral adsorption. Thus we performed a viral binding assay that allows to investigate the inhibitory effect of test compounds on HIV-1(III<sub>B</sub>) binding to MT-4 cells. BRACO-19 proved to be ineffective in inhibiting viral binding to MT-4 cells at concentrations up to 125 µg/ml (corresponding to 210 µM). As expected, the reference entry inhibitors DS and AMD3100 were able to block entry process up to 60 and 50%, respectively. Surprisingly, BRACO-19 seemed to slightly stimulate the viral adsorption with an increase in binding affinity of about 20% at the highest concentrations (25-125 µg/ml corresponding to 42 and 210 µM) (Figure 5.26). Importantly, these concentrations were highly far from the EC<sub>50</sub> calculated in the antiviral assays on MT-4 cells both 1 and 5 days post infection. In fact, we observed EC<sub>50</sub> values from 1.4 to 7.9 µM (Figure 5.23 and 5.24). We concluded that the antiviral effect observed in infected MT-4 cells was most probably not due to an inhibition of viral binding.

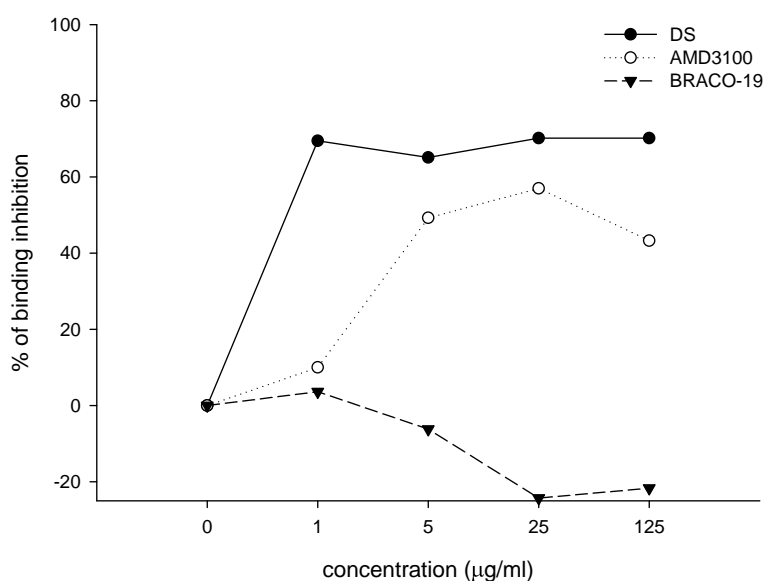


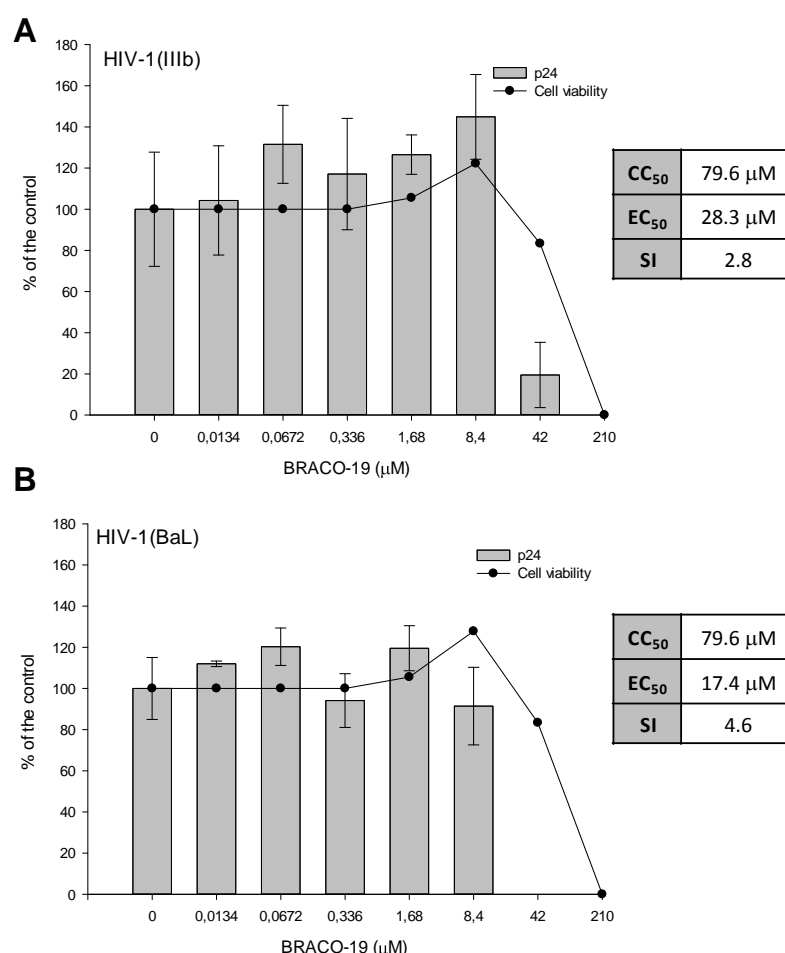
Figure 5.26 Effect of BRACO-19 on HIV-1(IIIB) binding to MT-4 cells

#### 5.4.1.4. HIV-1 infected PBMCs

We also investigated the antiretroviral properties of BRACO-19 in freshly isolated Peripheral Blood Mononuclear Cells (PBMCs) from an healthy donor that were subsequently infected with HIV-1(III<sub>B</sub>). This experiment is particularly significant since PBMCs are primary cells and represents the cells that are naturally infected by HIV *in vivo*. Again, to exclude an antiviral effect of BRACO-19 on viral entry due to its positive charges, the experiment was carried out infecting PBMCs with two viruses characterized by a different tropism. HIV(III<sub>B</sub>) is a X4 virus that uses the alpha-chemokine receptor CXCR4 while HIV(BaL) uses the beta-chemokine receptor CCR5 for entry. This allowed to exclude any effect of BRACO-19 on the CXCR4 co-receptor in PBMCs, as previously observed in MT-4 infected cells (Figure 5.26).

We demonstrated that BRACO-19 was active against both strains 7 days after infection of PBMCs, confirming that the effect of the compound was not correlated to the nature of the co-receptor involved in the entry process (Figure 5.27). In PBMCs infected with HIV-1(III<sub>B</sub>), the treatment with BRACO-19 caused a peculiar dose-dependent relation. Up to 8.4 µM, BRACO-19 induced an increase in cell viability that is directly correlated with a slightly increase in p24 production up to 140%. This latter effect probably reflected an effect of BRACO-19 at cellular level. However, in presence of BRACO-19 42 µM, viral production dramatically decreased by approximately 80% compared to the control. The observed

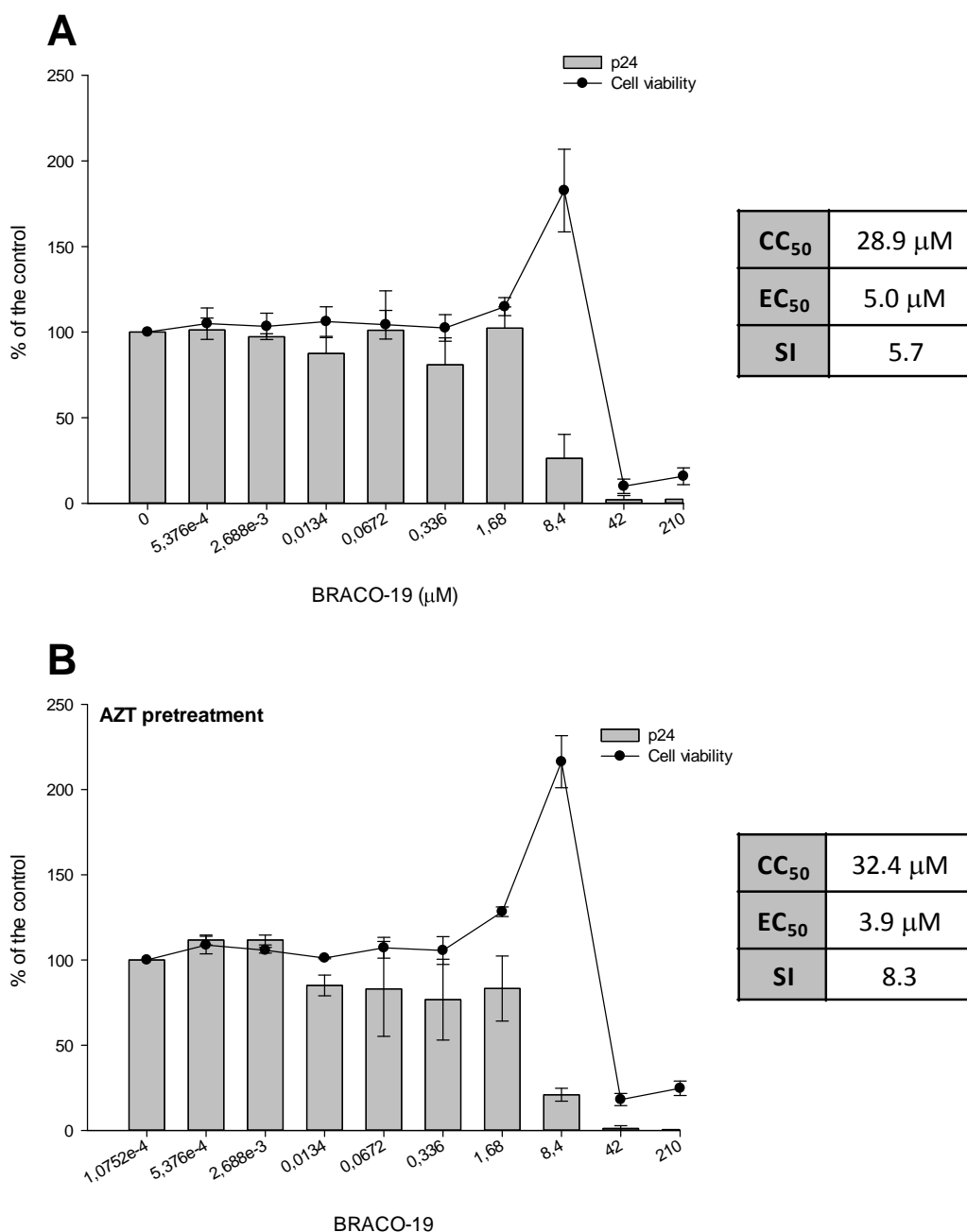
antiviral activity of BRACO-19 did not correlate with cytotoxicity, since at 42  $\mu\text{M}$  cell viability was around 80% (Figure 5.27A). This evidence suggested a different target, probably at viral level, of BRACO-19. In PBMCs, the antiviral effect of BRACO-19 against HIV-1(III<sub>B</sub>) resulted in an EC<sub>50</sub> of 28.3  $\mu\text{M}$  with an SI of 3. Similar results were obtained in PBMCs infected with HIV-1 (BaL). Again, the compound seemed to promote an increase of about 30% in cell viability in presence of BRACO-19 up to 8,4  $\mu\text{M}$ . However, in this case the amount of p24 did not reflect this increasing trend. The antiviral effect of BRACO-19 against HIV-1 (BaL) resulted more pronounced with an EC<sub>50</sub> of 17.4  $\mu\text{M}$  and a with a SI of 4.6 (Figure 5.27B). Besides the fact that infected PBMCs appeared less susceptible to the antiviral action of BRACO-19 7 days after infection, the anti-HIV-1 effect of the G-quadruplex ligand was confirmed.



**Figure 5.27 Antiviral effect of BRACO-19 on PBMCs infected with HIV-1(III<sub>B</sub>) or HIV-1(BaL) 7 days post infection** PBMCs were infected with (A) HIV-1(III<sub>B</sub>) or (B) HIV-1(BaL) in the presence of the G-4 stabilizing ligand, BRACO-19. After 7 days, the viral inhibition was assessed as % of p24 antigen present in the supernatant compared to the untreated control. Cytotoxicity of BRACO-19 on untreated PBMCs was assessed in parallel via cell counting. Results are presented as mean $\pm$ std dev from 1 experiment made in triplicate. The CC<sub>50</sub>, EC<sub>50</sub> and SI values are shown in the inset tables.

### 5.4.1.5. HIV-1 persistently infected cells (MT-4/III<sub>B</sub>)

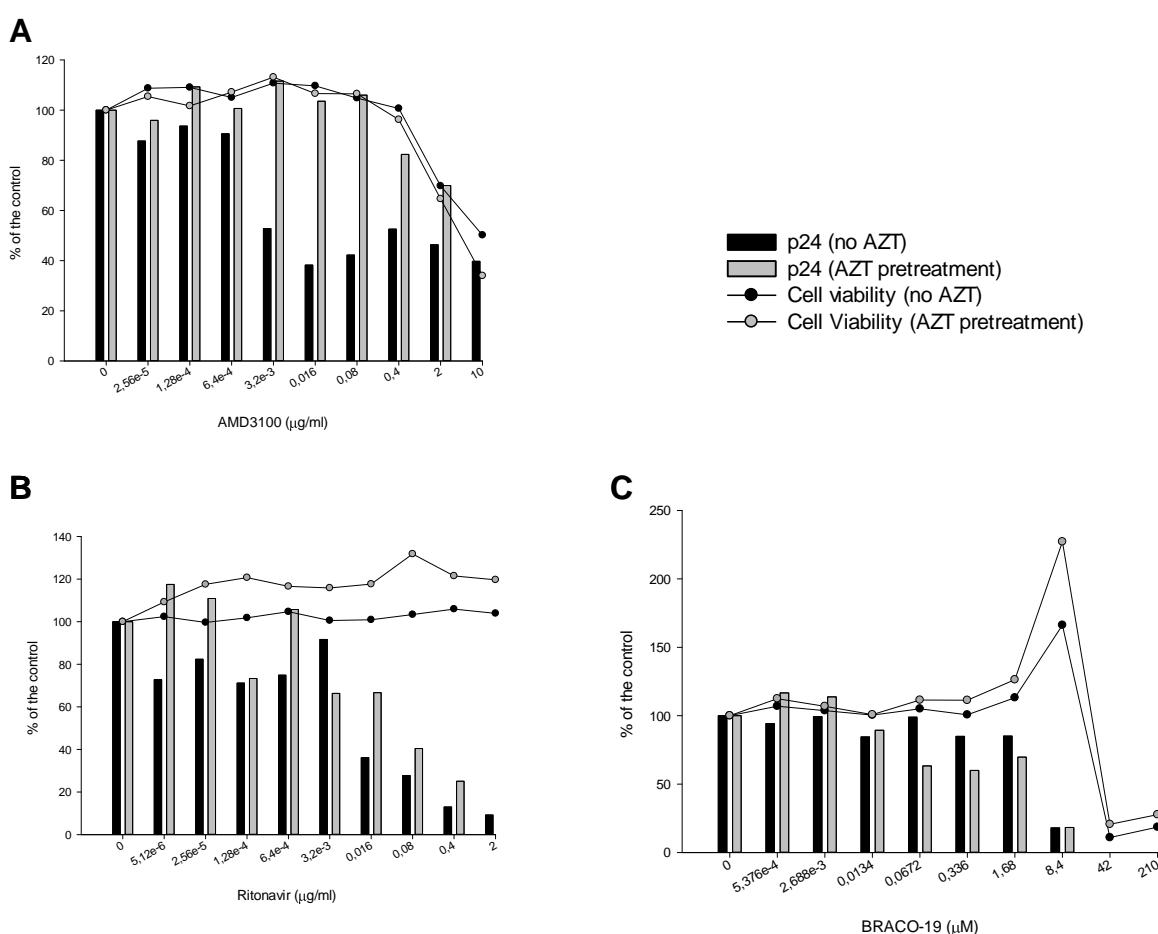
We decided to further investigate the antiviral activity of BRACO-19 in persistently HIV-1 infected cells using the MT-4/III<sub>B</sub> cell line. These cells are persistently infected by HIV-1(III<sub>B</sub>) and resistant to the viral induced CPE, stably producing infectious viral particles. MT-4/III<sub>B</sub> treated with increasing concentration of BRACO-19 for 5 days showed a progressive decrease in p24 production that reached the 26% of the control. The G-quadruplex ligand determined a pronounced increase in cell viability that corresponded to the maximum antiviral effect of the compound. Overall, BRACO-19 is effective with an EC<sub>50</sub> of 5 μM with a SI of around 6 (Figure 5.28A). Overall the SI was lower as compared to that observed in an acute infection model using HIV-1 (III<sub>B</sub>) in MT-4 cells (SI around 15, see Figure 5.23 B). This cellular system provided all viral steps, from entry to budding of viral particles, thus it was not possible to evaluate which viral target could be perturbed by BRACO-19. Since we excluded an effect of BRACO-19 on HIV-1(III<sub>B</sub>) absorption to MT-4 cells (Figure 5.26), we decided to evaluate a possible implication of this test compound in post-integration steps and to test BRACO-19 in persistently-infected cells pretreated with Zidovudine (AZT, 10 ng/ml). The pretreatment with AZT allows to evaluate the effect of the compound only at post integration level, since this compound inhibits the viral cycle at the reverse transcription step. Interestingly, the results obtained after the pretreatment with AZT were similar. Again, BRACO-19 caused an increase in cell viability and an inhibition of about 80% of viral production showing an EC<sub>50</sub> of 3.9 μM. The SI showed a slightly increase reaching the value of 8.3 (Figure 5.28B).



**Figure 5.28 Antiviral effect of BRACO-19 on MT-4/III<sub>B</sub> persistently infected cells** MT-4/III<sub>B</sub> cells (A) untreated or (B) pretreated with AZT were treated with increasing concentration of the G-4 stabilizing ligand, BRACO-19. After 5 days, the viral inhibition was assessed as % of p24 antigen present in the supernatant compared to the untreated control. Cytotoxicity of BRACO-19 on MT-4/III<sub>B</sub> cells was assessed via MTT-assay. Results are presented as mean±std dev from 2 independent experiments each performed in triplicate. The CC<sub>50</sub>, EC<sub>50</sub> and SI values are shown in the inset tables.

We proposed that MT-4/III<sub>B</sub> pretreated with AZT could provide a good cellular system to investigate possible implication of our test compound after viral integration in the host genome. To further validate this system, we decided to test in parallel two reference drugs: the

entry inhibitor AMD3100 and the protease inhibitor Ritonavir. In fact, if the AZT-pretreatment resulted efficient AMD3100 would lose its antiviral effect while Ritonavir would retain its antiretroviral properties. As expected, AMD3100 was active in MT-4/III<sub>B</sub> acting pre-integration and blocking viral entry while completely lost its activity after the pretreatment with AZT (Figure 5.29A). On the contrary, Ritonavir showed similar antiviral activity in untreated and AZT-pretreated MT-4/III<sub>B</sub> since it acted post integration as protease inhibitor (Figure 5.29B). As for Ritonavir, BRACO-19 resulted active in inhibiting HIV-1(III<sub>B</sub>) persistent infection both in AZT-pretreated and untreated MT-4/III<sub>B</sub> (Figure 5.29C). These data suggested that BRACO-19 could perturb viral cycle acting at post integration steps, such as transcription process.



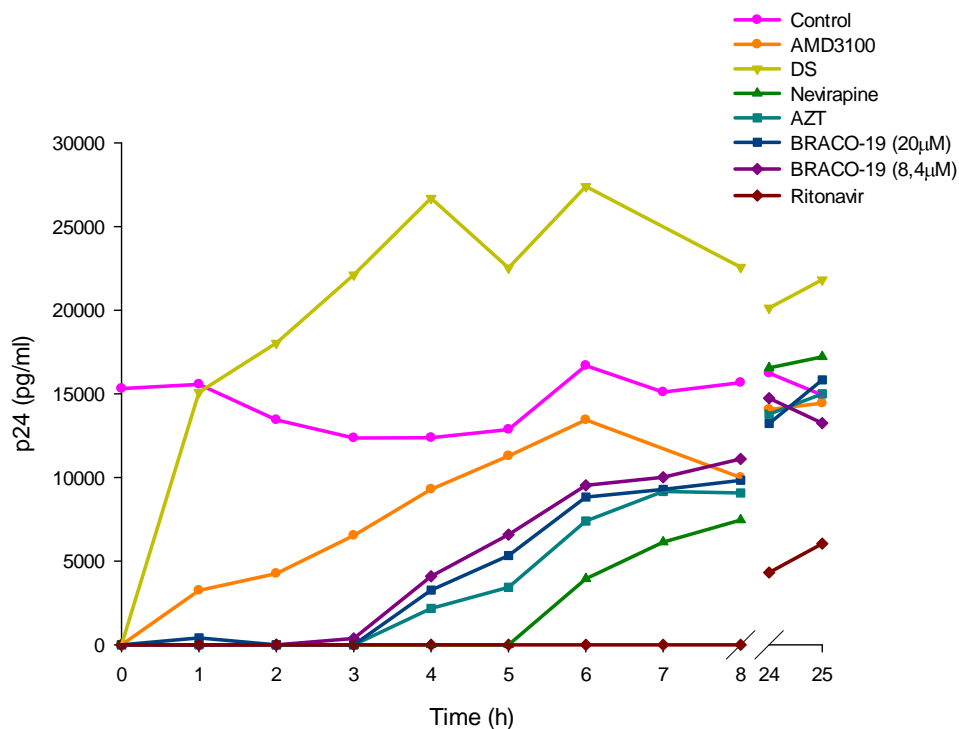
**Figure 5.29** Antiviral effect of AMD3100, Ritonavir and BRACO-19 on MT-4/III<sub>B</sub> persistently infected cells untreated and pretreated with AZT MT-4/III<sub>B</sub> cells untreated or pretreated with AZT were treated with increasing concentration of (A)AMD3100 or (B)Ritonavir or (C) BRACO-19. After 5 days, the viral inhibition was assessed as % of p24 antigen present in the supernatant compared to the untreated control. Cytotoxicity of test compounds on MT-4/III<sub>B</sub> cells was assessed via MTT-assay. Results are presented as value of one single experiment.



#### 5.4.1.6. Time of Addition (TOA)

To further investigate the target of BRACO-19 as HIV-1 inhibitor, a time-of-addition (TOA) assay was performed. This experiment allows to determine how long the addition of a test compound can be postponed before it loses its antiviral activity. HIV-1 replication is blocked up to a time point corresponding to the occurrence of the last replication process targeted by the drug. Thus, the last inhibited viral step can be revealed comparing the antiviral effect of test compound to that of reference drugs in the time frame of replication events. The antiviral activities of test compounds were based on the assessment of the inhibition of p24 antigen production (Daelemans et al, 2011).


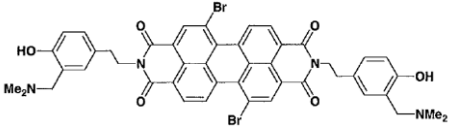
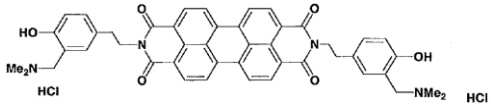
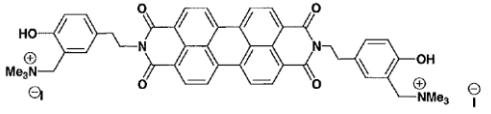
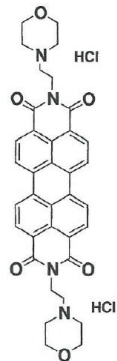
Surprisingly, the final viral step inhibited by BRACO-19 at the two concentration tested (8.4 and 20  $\mu\text{M}$ ) was the reverse transcription process that normally occurs from 3 to 5 hours post infection. As for the reference drug AZT that is a NRTI, BRACO-19 was able to fully block the HIV-1 cycle if administered up to 3 post infection. It is interesting to note that addition of BRACO-19 from 4 to 8 h post infection resulted in a slight inhibition of HIV-1, since p24 levels did not reach the control values (Figure 5.30). Overall, BRACO-19 seemed act prevalently at a pre integration level blocking the reverse transcription process as the final target process. Apparently, these results are in contrast with the evidence presented in Figure 5.29 where BRACO-19 showed a nice antiviral effect with a probable post integration target of action. A possible explanation is that it was not possible to use 100-fold compound's  $\text{EC}_{50}$  as required for the TOA experiment, since at that concentration BRACO-19 resulted cytotoxic. In fact, the  $\text{EC}_{50}$  of BRACO-19 against HIV-1(III<sub>B</sub>) 1 day post infection was 5.6  $\mu\text{M}$  (Figure 5.24) while the  $\text{CC}_{50}$  was 154.6  $\mu\text{M}$ , thus making it impossible to test the compound at 100-fold its  $\text{EC}_{50}$ . This represented the main limitation in characterizing BRACO-19 target of action with TOA. Besides the TOA experiment suggested an implication of BRACO-19 in perturbing reverse transcription process, this fact does not exclude another post-integration target of the test compound as previously described. Further experiments have to be carried out to further investigate the viral targets of BRACO-19.

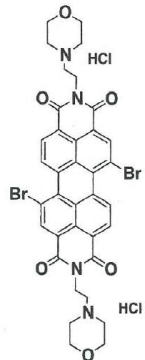
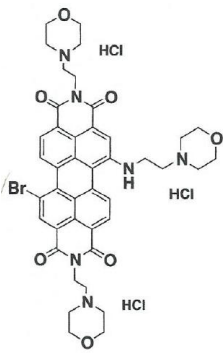
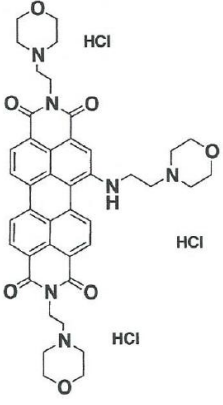


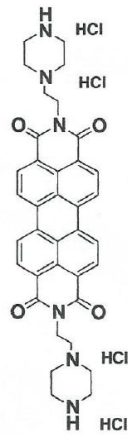

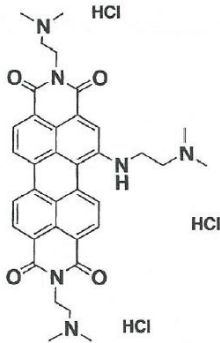
**Figure 5.30 Time-of-addition experiment with the test compound BRACO-19** MT-4 cells were infected with HIV-1(III<sub>B</sub>) and test compounds were added at different time points after infection. Virus associated p24 antigen was measured 31 h post-infection.

#### 5.4.2. Antiviral screening of G-quadruplex ligands

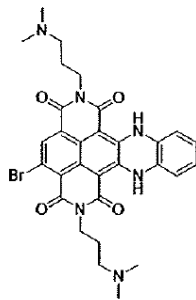
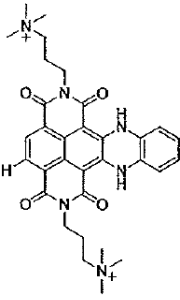
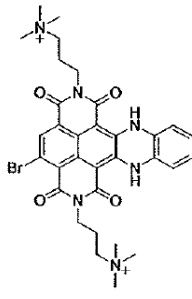
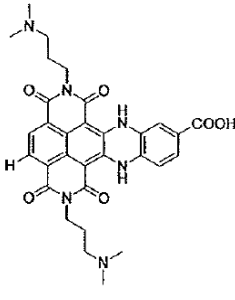
The possible antiviral effect of a variety of newly synthesized G-quadruplex ligands provided by Prof. M. Freccero (University of Pavia, Italy) was subsequently investigated (Table 5.10). These compounds belonged to two of the major G-quadruplex binder classes such as perylene diimides (PDIs) and naphthalene diimides (NDIs). As mentioned in the introduction of this thesis, PDIs and NDIs are *in situ* protonated G-quadruplex ligands that are prone to  $\pi$ -stacking with G-tetrad thanks to their aromatic core. Originally, these compounds were synthesized as ligands against telomeric G-quadruplex or oncogene promoter G-quadruplexes.

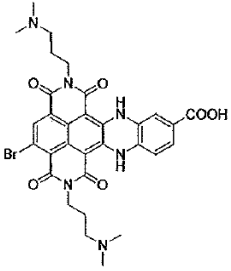
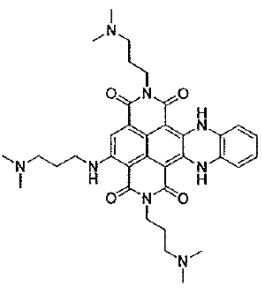
Class of compound	Name	Molecular Weight	Chemical structure
PDI <sub>s</sub>	Br-GBP-02	872.08	
	Br-SPP-02	900.12	
	H-SPP-02	816.25	
	H-SPP-03	1028.15	
	PDI-Morph	688.19	

	Br-PDI-Morph	844.01	
	Br-Tri-PDI-Morph	930.17	
	Tri-PDI-Morph	852.26	

	PDI-Pip	758.17	
	Br-PDI-Pip	913.99	
	Tri-PDI-NMe <sub>2</sub>	726.23	

	H-Tri-prop-NMe <sub>2</sub>	644.24	
NDIs	Br-tri-prop-NMe <sub>2</sub>	722.15	
	Tetra-prop-NMe <sub>2</sub>	780.32	
	H-NMe <sub>2</sub> -PhAm	540.61	

NDIs Extended Core	Br-NMe <sub>2</sub> -PhAm	619.51	
	H-NMe <sub>3</sub> -PhAm	570.68	
	Br-NMe <sub>3</sub> -PhAm	649.58	
	H-NMe <sub>2</sub> -PhAmCOOH	584.62	

	Br-NMe <sub>2</sub> -PhAmCOOH	663.52	
	Tri-NMe <sub>2</sub> -PhAm	640.78	

**Table 5.10 G-quadruplex ligands tested in the antiviral screening**

We started to screen selected compounds for their antiretroviral activity against HIV-1 strain III<sub>B</sub> (HIV-1(III<sub>B</sub>)). Inhibition of viral replication was first monitored by evaluating the protection from HIV-induced CPE on MT-4 cells 5 days after infection. Among the twenty-one tested compounds, only two showed a significant antiviral activity with an EC<sub>50</sub> in the micromolar range (Table 5.11). These two compounds belonged to the PDI class. The perylenes H-SPP-03 and PDI-Pip reached a high percentage of protection from HIV-induced CPE of 96,7% and 73%, respectively. The estimated EC<sub>50</sub> values were 10.41 μM for H-SPP-03 and 3.89 μM for PDI-Pip. The average SI for these active compounds was not so high since H-SPP-03 and PDI-Pip showed a SI of 9 and 5, respectively. Other PDIs compounds appeared slightly active in inhibiting HIV-1: PDI-Morph and Br-PDI-Pip produced a percentage protection against CPE up to 11.8% and 17.7%, respectively. PDI-Morph is characterized by a perylene core with two morpholine groups in the sidearms while Br-PDI-Pip has two piperazine groups at the sidearms and two bromines linked to the aromatic core. Since these latter two compounds did not reach the protection in 50% of HIV-1-infected cells, it was not possible to calculate their EC<sub>50</sub>. Overall, PDIs compounds resulted more effective as antiretroviral agents as compared to NDIs. This preliminary anti-HIV-1 screening



highlighted at least two G-4 ligands, H-SPP-03 and PDI-Pip with a promising antiviral activity and prompted us to carry out a more detailed study of their mechanism of action.

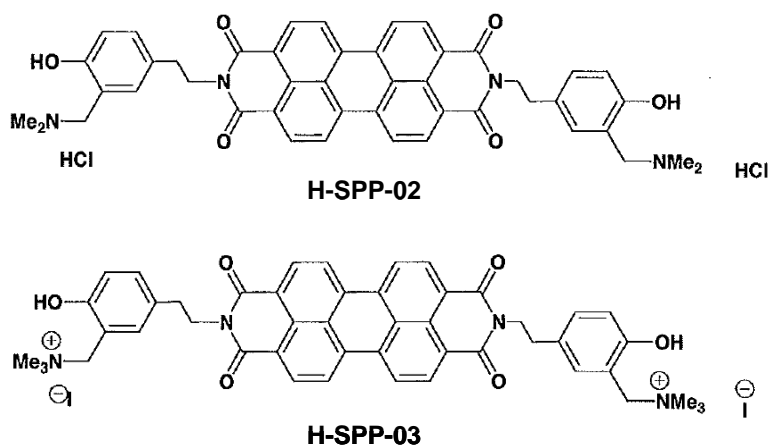
Compound	EC <sub>50</sub> $\mu$ M	CC <sub>50</sub> $\mu$ M	SI	Max protection (%)
Br-GBP-O2	-	2.60	<1	-
Br-SPP-O2	-	2.75	<1	-
H-SPP-O2	-	14.33	<1	-
H-SPP-03	10.41	88.21	9	97
PDI-Morph	-	>36.32	-	12
Br-PDI-Morph	-	0.11	<1	-
Br-Tri-PDI-Morph	-	3.13	<1	-
Tri-PDI-Morph	-	10.16	<1	-
PDI-Pip	3.89	18.99	5	73
Br-PDI-Pip	-	2.65	<1	18
Tri-PDI-NMe2	-	3.44	<1	-
H-Tri Prop-Nme2	-	0.04	<1	-
Br-Tri-Prop Nme2	-	0.03	<1	-
Tetra Prop-Nme2	-	0.60	<1	-
H-Nme2 PhAm	-	0.007	<1	-
Br-NMe2 PhAm	-	0.006	<1	-
H-NMe3 PhAm	-	>43.80	-	-
Br-Nme3 PhAm	-	8.65	<1	-
H-NMe2 PhAm COOH	-	0.93	<1	-
Br-NMe2 PhAm COOH	-	0.76	<1	-
Tri-Nme2 PhAm	-	0.033	<1	-

**Table 5.11** Antiviral activity of test compounds against HIV-1(III<sub>B</sub>) on MT-4 cells 5 days post infection

#### 5.4.2.1. Evaluation of antiviral activity of H-SPP-03

H-SPP-03 (Figure 5.31) is a perylene and it has been synthesized by Prof. M Freccero (University of Pavia) and proposed as G-quadruplex ligand. In fact, the chemical structure of this compound resulted similar to the well-known G-quadruplex binder PIPER (See Introduction). The big aromatic core of this compound ensures  $\pi$ -stacking interactions with G-tetrads, moreover the two stable positive charges at the sidearms increase the affinity for tetraplex structures. To date, persistent positive charges may have a key role in the antiviral activity of this compound. In fact, it has been reported that several CXCR4-specific inhibitors

possesses several positive charges that specifically interact with the extracellular loops of the co-receptor (Wilkinson et al, 2011). Moreover, a series of quaternary ammonium salts (QAS) are highly potent and selective inhibitors of HIV-1 replication *in vitro*, perturbing viral adsorption and/or virus-cell fusion process (Pannecouque, 2000). To elucidate the role of positive charges in H-SPP-03, the analogue H-SPP-02 was used in comparison. H-SPP-02 (Figure 5.31) has the same chemical structure as H-SPP-03 but is protonable instead of being permanently charged. Since previous results of our lab indicated that the positive charges of H-SPP-03 prevented the entry of the compound in cells (data not shown), we hypothesized an action of the compound at cell membrane level probably in preventing viral attachment or entry. An additional indication that probably H-SPP-03 did not enter the cells was reflected by its high  $CC_{50}$  (88 $\mu$ M) compared to the lower  $CC_{50}$  of H-SPP-02 (14 $\mu$ M).



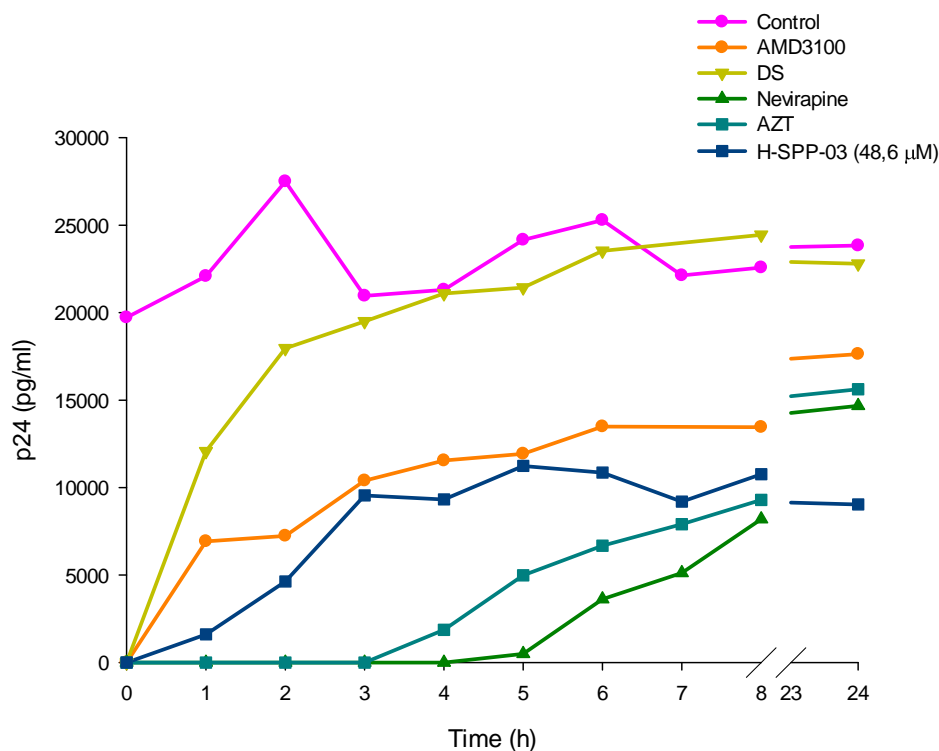
**Figure 5.31** Chemical structures of H-SPP-02 and H-SPP-03

As anticipated in the previous paragraph, H-SPP-03 is active in blocking HIV-1 cycle producing a protection from viral induced CPE up to 96%. The antiviral activity of H-SPP-03 resulted in an  $EC_{50}$  of 10.4  $\mu$ M. Interestingly, H-SPP-02 did not show any antiviral activity against HIV-1(III<sub>B</sub>) (Table 5.12), supporting the hypothesis that the permanent positive charges of H-SPP-03 are required for antiviral activity. We decided to further test both compounds against HIV-2 strain ROD (HIV-2(ROD)) to try to assess the specificity. We found that both H-SPP-03 and H-SPP-02 did not inhibit viral cycle and did not show any antiviral effect in infected MT-4 cells 5 days post infection. These data suggested that H-SPP-03 displayed a specific antiviral activity against HIV-1. This effect against HIV-1 is probably due to the positive charges present in the compound.

Compound	HIV-1				HIV-2			
	EC <sub>50</sub> $\mu$ M	CC <sub>50</sub> $\mu$ M	SI	Max protection (%)	EC <sub>50</sub> $\mu$ M	CC <sub>50</sub> $\mu$ M	SI	Max protection (%)
H-SPP-O2	-	14.33	-	-	-	14.82	-	-
H-SPP-O3	10.41	88.21	9	96.7	-	88.30	-	-

**Table 5.12** Antiviral activity of H-SPP-03 and H-SPP-02 against HIV-1(III<sub>B</sub>) and HIV-2(ROD) on MT-4 cells 5 days post infection

To further investigate the target of H-SPP-03 as HIV-1 inhibitor, a time-of-addition (TOA) assay was performed. The final viral step inhibited by the compound at 48 $\mu$ M was the viral entry process that normally occurs within 1 hour post infection (Figure 5.32).



**Figure 5.32** Time-of-addition experiment with the test compound H-SPP-03 MT-4 cells were infected with HIV-1(III<sub>B</sub>) and test compounds were added at different time after infection. Virus associated p24 antigen was measured 31h post infection.

Even if H-SPP-03 did not result a specific G-quadruplex ligand acting at intracellular level, we were able to demonstrate a significant antiviral activity of this compound against HIV-1(III<sub>B</sub>).

### 5.4.2.2. Evaluation of antiviral activity of PDI-Pip

PDI-Pip is a compound characterized by a perylene aromatic core with piperazine groups at the two sidearms. As for H-SPP-03, the core is similar to that of PIPER. Since we found PDI-Pip active against HIV-1(III<sub>B</sub>) with a SI of 5 (Table 5.11), we decided to further investigate its antiretroviral properties. Surprisingly, we observed that the antiretroviral activity of the compound against HIV-1(III<sub>B</sub>) in infected MT-4 cells was not so reproducible. An example of two independent experiments each in triplicate is reported in Table 5.13. The compound resulted in a consistent decreased antiviral activity in the second experiment, showing a maximum protection from CPE of 16% instead of 73%. These results suggested a mode of action similar to that observed for QAS in inhibiting viral entry (Pannecouque, 2000). The explanation of this peculiar behavior relies in the composition of cell membrane of cells and in particular in the polysaccharide heparan sulfate (HS) role. It has been demonstrated that cell-surface HS participates to the HIV-1 attachment and entry processes in MT-4 cells (Patel et al, 1993). In particular, the interaction between HIV-1 gp41 and HS on T-lymphocyte cell membrane seems to promote the viral adsorption (Cladera et al, 2001). The HIV-1 strain III<sub>B</sub> used to infect MT-4 cells in these antiviral experiments consisted in mixed subpopulations and only some of them strictly required HS involvement in entry process. The initial prevalent subpopulation in the infectious stock and the expression of HS on the cell surface could select viral subpopulations that are differently dependent on HS for efficient infection. If a viral subpopulation fully dependent on HS is selected, compounds that interfere with HS are extremely efficient in blocking HIV-1 cycle. On the contrary, if some selection occurs, HIV-1 infectious population could be partially insensitive to the inhibitory effect of test compound directed against HS. Thus, this could determine different results in antiviral assays with HIV-1(III<sub>B</sub>)-MT-4 infected cells. These evidences have been reported for QAS (Pannecouque, 2000). It is possible to test this hypothesis implying the use of Simian Immunodeficiency Virus (SIV) to infect MT-4 cells. In fact, SIV strain mac251 (SIV(mac251)) depends on the presence of HS to infect MT-4 cells and it has been reported that QAS are able to completely block infection at this level. Thus, if an antiviral compound act interfering with HS, it should be extremely active against SIV(mac251) and the obtained results should be extremely reproducible. Moreover, HS seems not to be required in HIV-2 infection, compound that interfere with HS should be completely inactive against HIV-2.

We thus decided so to test PDI-Pip against SIV(mac251) and HIV-2(ROD). The inhibition of viral replication was monitored by evaluating the protection from induced CPE on MT-4 cells 5 days after infection. Exactly as for QAS, PDI-Pip resulted extremely efficient in blocking SIV(mac251) infection in a reproducible way while no antiviral activity was observed against HIV-2(ROD) (Table 5.13).

Virus	PDI-Pip				
	# exp	EC <sub>50</sub> $\mu$ M	CC <sub>50</sub> $\mu$ M	SI	Max protection (%)
HIV-1 (III <sub>B</sub> )	I	3.9	18.9	5	73
	II	-	16.3	-	16
SIV(Mac251)	I	2.5	17.3	7	100
	II	2.2	18.9	8	100
HIV-2(ROD)	I	-	19.1	-	-

**Table 5.13 Antiviral activity of PDI-Pip against HIV-1(III<sub>B</sub>) , SIV(ROD) and HIV-2(ROD) on MT-4 cells 5 days post infection.**

Therefore, even if also PDI-Pip did not show a G-quadruplex-mediated antiviral activity, we argued evidences for its antiretroviral mode of action by inhibiting virus adsorption and/or virus-cell fusion processes.



---

## 6. Conclusions

This thesis focused on the evaluation of the presence of conserved putative G-quadruplex forming regions in the HIV-1 proviral genome and on possible implications of G-quadruplex structures as target for antiviral therapy. We checked the effect of several G-quadruplex ligands as HIV-1 inhibitors. We presented the first reported evidence of conserved G-rich sequences that can fold in the G-quadruplex conformation within the HIV-1 proviral genome. In particular, we highlighted two G-quadruplex-forming regions located in the HIV-1 LTR promoter and in the *nef* coding region. The U3 region of the HIV-1 LTR had the potential to form multiple G-quadruplexes (i.e., LTR-III, LTR-II, and LTRIV) that we demonstrated could actually fold in a G-quadruplex structure. We investigated structural features of these G-quadruplexes, observing a shared parallel-like topology for all the evaluated structures. Moreover, we were able to characterize the G bases effectively involved in G-quadruplex folding, discovering a specific core of the LTR G-quadruplex architecture whose mutations completely abolished quadruplex building. Interestingly, these single-base mutations were able to increase by 2–3-folds the LTR promoter activity in a luciferase reporter assay, suggesting that G-quadruplexes act as repressor elements in the transcriptional activation of HIV-1. Moreover, the G-quadruplex ligand BRACO-19 inhibited LTR promoter activity only when the LTR promoter sequence maintained its ability to fold into G-quadruplex, whereas resulted inactive in the presence of mutations that abolished G-quadruplex folding. These features point out the biological relevance of the HIV-1 promoter G-quadruplex forming region and open up the possibility of inhibiting the HIV-1 LTR promoter activity by G-quadruplex interacting small molecules.

In the HIV-1 *nef* coding region three G-rich/G-quadruplex prone tracts were closely clustered and located on both the leading and lagging strands. These G-quadruplex forming sequences, and in particular the G pattern necessary for G-quadruplex folding, resulted extremely conserved within the HIV-1 M group and in its subtypes largely distributed worldwide. The identified sequences could form G-quadruplexes with two stacked tetrads that could be strongly stabilized by G-quadruplex ligands. Interestingly, the G-quadruplex ligand TMPyP4 resulted efficient in inhibiting Nef expression in a eGFP-reporter assay. Most importantly, the same G-quadruplex ligand was sufficient to impair the Nef-mediated enhancement of HIV-1 infectivity. In fact, a significant antiviral effect of TMPyP4 against the wt HIV-1 in Nef sensitive cells was observed, while the effect on the  $\Delta$ Nef HIV-1 resulted negligible, further

indicating a specific anti Nef effect. Thus, G-quadruplex stabilization within the *nef* coding region of the viral genome may impair not only Nef expression through inhibition of transcription that directly generates mRNAs, but also overall transcription for production of new copies of the RNA genome to be assembled in the viral progeny.

Here we have shown that the commercially available G-quadruplex ligands BRACO-19 and TMPyP4 were effective towards HIV-1 with minimal effect on cell viability. Therefore a therapeutic window can be envisaged that would consent the employment of known G-quadruplex binders as anti-HIV compounds. Several G-quadruplex ligands were tested for their anti-HIV-1 effect. We confirmed a promising antiviral activity of BRACO-19 in several cell lines and primary cells, with an encouraging Selectivity Index up to 28. BRACO-19 resulted effective both against *ex novo* and persistent HIV-1 infection. We excluded any HIV-1-virions' structural parts as possible target for BRACO-19 activity. Moreover, since BRACO-19 was not able to block viral adsorption/entry, we concluded that the ligands had intracellular viral targets. Preliminary evidences suggested that BRACO-19 interfered with a pre-integration viral step, probably reverse transcription, together with a post-integration step. Finally, we discovered two perylenes G-quadruplex ligands that showed an interesting anti HIV-1 activity, probably direct against viral entry. Thus, for these two compounds we tended to exclude a viral G-quadruplex target. However, our antiviral characterization of BRACO-19 resulted compatible with a direct viral G-quadruplex target. It is important to consider that available G-quadruplex ligands were principally developed against the telomeric/eukaryotic G-quadruplexes and that they interact with G-quadruplex features that are common to all G-quadruplex conformations, therefore displaying no specificity against different G-quadruplex structures. In progress detailed structural characterization of the HIV-1 G-quadruplexes will allow rational design of small molecules for improved selectivity towards the viral G-quadruplexes versus the eukaryotic structures. This work has thus paved the way for the development of anti-HIV-1 drugs with an unprecedented mechanism of action.



---

## 7. References

Aiken C, Trono D (1995) Nef stimulates human immunodeficiency virus type 1 proviral DNA synthesis. *Journal of Virology* **69**: 5048-5056

Balliet JW, Kolson DL, Eiger G, Kim FM, McGann KA, Srinivasan A, Collman R (1994) Distinct effects in primary macrophages and lymphocytes of the human immunodeficiency virus type 1 accessory genes vpr, vpu, and nef: mutational analysis of a primary HIV-1 isolate. *Virology* **200**: 623-631

Bang I (1910) Untersuchungen über die Guanylsäure. *Biochemische Zeitschrift* **26**: 293–311

Baral A, Kumar P, Halder R, Mani P, Yadav VK, Singh A, Das SK, Chowdhury S (2012) Quadruplex-single nucleotide polymorphisms (Quad-SNP) influence gene expression difference among individuals. *Nucleic Acids Research* **40**: 3800-3811

Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**: 868-871

Basundra R, Kumar A, Amrane S, Verma A, Phan AT, Chowdhury S (2010) A novel G-quadruplex motif modulates promoter activity of human thymidine kinase 1. *The FEBS Journal* **277**: 4254-4264

Beaume N, Pathak R, Yadav VK, Kota S, Misra HS, Gautam HK, Chowdhury S (2013) Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of *D. radiodurans* involves G4 DNA-mediated regulation. *Nucleic Acids Research* **41**: 76-89

Belotserkovskii BP, Liu R, Tornaletti S, Krasilnikova MM, Mirkin SM, Hanawalt PC (2010) Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 12816-12821

Biffi G, Tannahill D, McCafferty J, Balasubramanian S (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry* **5**: 182-186

Bochman ML, Paeschke K, Zakian VA (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics* **13**: 770-780

## References

---

- Breuer S, Schievink SI, Schulte A, Blankenfeldt W, Fackler OT, Geyer M (2011) Molecular design, functional characterization and structural basis of a protein inhibitor against the HIV-1 pathogenicity factor Nef. *PloS One* **6**: e20033
- Brooks TA, Hurley LH (2009) The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nature Reviews Cancer* **9**: 849-861
- Brooks TA, Kendrick S, Hurley L (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *The FEBS Journal* **277**: 3459-3469
- Brown RV, Danford FL, Gokhale V, Hurley LH, Brooks TA (2011) Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *The Journal of Biological Chemistry* **286**: 41018-41027
- Brown RV, Hurley LH (2011) DNA acting like RNA. *Biochemical Society Transactions* **39**: 635-640
- Broxson C, Beckett J, Tornaletti S (2011) Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myc gene. *Biochemistry* **50**: 4162-4172
- Burger AM, Dai F, Schultes CM, Reszka AP, Moore MJ, Double JA, 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
- Bushman FD, Fujiwara T, Craigie R (1990) Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249**: 1555-1558
- Campbell NH, Parkinson GN, Reszka AP, Neidle S (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *Journal of the American Chemical Society* **130**: 6722-6724
- Capon DJ, Ward RH (1991) The CD4-gp120 interaction and AIDS pathogenesis. *Annual Review of Immunology* **9**: 649-678
- Capra JA, Paeschke K, Singh M, Zakian VA (2010) G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Computational Biology* **6**: e1000861
- Chan CN, Dietrich I, Hosie MJ, Willett BJ (2013) Recent developments in human immunodeficiency virus-1 latency research. *The Journal of General Virology* **94**: 917-932

- Cheng H, Hoxie JP, Parks WP (1999) The conserved core of human immunodeficiency virus type 1 Nef is essential for association with Lck and for enhanced viral replication in T-lymphocytes. *Virology* **264**: 5-15
- Chowers MY, Spina CA, Kwoh TJ, Fitch NJ, Richman DD, Guatelli JC (1994) Optimal infectivity in vitro of human immunodeficiency virus type 1 requires an intact nef gene. *Journal of Virology* **68**: 2906-2914
- Chun TW, Davey RT, Jr., Engel D, Lane HC, Fauci AS (1999) Re-emergence of HIV after stopping therapy. *Nature* **401**: 874-875
- Chutiwitoonchai N, Hiyoshi M, Mwimanzi P, Ueno T, Adachi A, Ode H, Sato H, Fackler OT, Okada S, Suzu S (2011) The identification of a small molecule compound that reduces HIV-1 Nef-mediated viral infectivity enhancement. *PloS One* **6**: e27696
- Cladera J, Martin I, O'Shea P (2001) The fusion domain of HIV gp41 interacts specifically with heparan sulfate on the T-lymphocyte cell surface. *The EMBO Journal* **20**: 19-26
- Clapham PR, McKnight A (2001) HIV-1 receptors and cell tropism. *British Medical Bulletin* **58**: 43-59
- Cohen EA, Dehni G, Sodroski JG, Haseltine WA (1990) Human immunodeficiency virus vpr product is a virion-associated regulatory protein. *Journal of Virology* **64**: 3097-3099
- Coiras M, Lopez-Huertas MR, Sanchez del Cojo M, Mateos E, Alcamí J (2010) Dual role of host cell factors in HIV-1 replication: restriction and enhancement of the viral cycle. *AIDS Reviews* **12**: 103-112
- Cuenca F, Greciano O, Gunaratnam M, Haider S, Munnur D, Nanjunda R, Wilson WD, Neidle S (2008) Tri- and tetra-substituted naphthalene diimides as potent G-quadruplex ligands. *Bioorganic & Medicinal Chemistry Letters* **18**: 1668-1673
- Daelemans D, Pauwels R, De Clercq E, Pannecouque C (2011) A time-of-drug addition approach to target identification of antiviral compounds. *Nature Protocols* **6**: 925-933
- Das SR, Jameel S (2005) Biology of the HIV Nef protein. *The Indian Journal of Medical Research* **121**: 315-332
- de Arellano ER, Alcamí 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

## References

---

Derdeyn CA, Decker JM, Sfakianos JN, Wu X, O'Brien WA, Ratner L, Kappes JC, Shaw GM, Hunter E (2000) Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *Journal of Virology* **74**: 8358-8367

Dexheimer TS, Sun D, Hurley LH (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

Di Antonio M, Doria F, Richter SN, Bertipaglia C, Mella M, Sissi C, Palumbo M, Freccero M (2009) Quinone methides tethered to naphthalene diimides as selective G-quadruplex alkylating agents. *Journal of the American Chemical Society* **131**: 13132-13141

Dixon IM, Lopez F, Esteve JP, Tejera AM, Blasco MA, Pratviel G, Meunier B (2005) Porphyrin derivatives for telomere binding and telomerase inhibition. *Chembiochem : a European Journal of Chemical Biology* **6**: 123-132

Doi T, Yoshida M, Shin-ya K, Takahashi T (2006) Total synthesis of (R)-telomestatin. *Organic Letters* **8**: 4165-4167

Doms RW, Trono D (2000) The plasma membrane as a combat zone in the HIV battlefield. *Genes & Development* **14**: 2677-2688

Doria F, Nadai M, Folini M, Scalabrin M, Germani L, Sattin G, Mella M, Palumbo M, Zaffaroni N, Fabris D, Freccero M, Richter SN (2013) Targeting loop adenines in G-quadruplex by a selective oxirane. *Chemistry* **19**: 78-81

Doria F, Nadai M, Sattin G, Pasotti L, Richter SN, Freccero M (2012) Water soluble extended naphthalene diimides as pH fluorescent sensors and G-quadruplex ligands. *Organic & Biomolecular Chemistry* **10**: 3830-3840

Drygin D, Siddiqui-Jain A, O'Brien S, Schwaebe M, Lin A, Bliesath J, Ho CB, Proffitt C, Trent K, Whitten JP, Lim JK, Von Hoff D, Anderes K, Rice WG (2009) Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Research* **69**: 7653-7661

Duan W, Rangan A, Vankayalapati H, Kim MY, Zeng Q, Sun D, Han H, Fedoroff OY, Nishioka D, Rha SY, Izbicka E, Von Hoff DD, Hurley LH (2001) Design and synthesis of fluoroquinophenoxazines that interact with human telomeric G-quadruplexes and their biological effects. *Molecular Cancer Therapeutics* **1**: 103-120

- Dunnick W, Hertz GZ, Scappino L, Gritzmacher C (1993) DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Research* **21**: 365-372
- Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes & Development* **18**: 1618-1629
- Eddy J, Maizels N (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Research* **34**: 3887-3896
- Emert-Sedlak LA, Narute P, Shu ST, Poe JA, Shi H, Yanamala N, Alvarado JJ, Lazo JS, Yeh JI, Johnston PA, Smithgall TE (2013) Effector kinase coupling enables high-throughput screens for direct HIV-1 Nef antagonists with antiretroviral activity. *Chemistry & Biology* **20**: 82-91
- Engelhart A.E. PJ, Persil O. and Hud N.V. (2009) Metal Ion Interactions with G-Quadruplex Structures. In *Nucleic Acid-Metal Ion Interactions*, N.V. H (ed), Chapter 4, pp 118-153. RSC Publishing
- Fan L, Peden K (1992) Cell-free transmission of Vif mutants of HIV-1. *Virology* **190**: 19-29
- Fedoroff OY, Salazar M, Han H, Chemeris VV, Kerwin SM, Hurley LH (1998) NMR-Based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry* **37**: 12367-12374
- Feinberg MB, Baltimore D, Frankel AD (1991) The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 4045-4049
- Franke EK, Luban J (1996) Inhibition of HIV-1 replication by cyclosporine A or related compounds correlates with the ability to disrupt the Gag-cyclophilin A interaction. *Virology* **222**: 279-282
- Gallay P, Swingler S, Song J, Bushman F, Trono D (1995) HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell* **83**: 569-576
- Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S, Leibowitch J, Popovic M (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220**: 865-867

## References

---

Gavathiotis E, Heald RA, Stevens MF, Searle MS (2001) Recognition and Stabilization of Quadruplex DNA by a Potent New Telomerase Inhibitor: NMR Studies of the 2:1 Complex of a Pentacyclic Methylacridinium Cation with d(TTAGGGT)<sub>4</sub>. We thank the EPSRC of the UK and AstraZeneca for financial support to E.G. M.F.G.S. and R.A.H. are supported by the Cancer Research Campaign of the UK. *Angewandte Chemie* **40**: 4749-4751

Gellert M, Lipsett MN, Davies DR (1962) Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America* **48**: 2013-2018

Geyer M, Fackler OT, Peterlin BM (2001) Structure--function relationships in HIV-1 Nef. *EMBO Reports* **2**: 580-585

Gonzalez V, Hurley LH (2010) The c-MYC NHE III(1): function and regulation. *Annual Review of Pharmacology and Toxicology* **50**: 111-129

Gowan SM, Heald R, Stevens MF, Kelland LR (2001) Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with G-quadruplexes. *Molecular Pharmacology* **60**: 981-988

Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols* **1**: 2527-2535

Grzesiek S, Bax A, Clore GM, Gronenborn AM, Hu JS, Kaufman J, Palmer I, Stahl SJ, Wingfield PT (1996) The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. *Nature Structural Biology* **3**: 340-345

Guedin A, Gros J, Alberti P, Mergny JL (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Research* **38**: 7858-7868

Hallenberger S, Bosch V, Angliker H, Shaw E, Klenk HD, Garten W (1992) Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* **360**: 358-361

Hampel SM, Sidibe A, Gunaratnam M, Riou JF, Neidle S (2010) Tetrasubstituted naphthalene diimide ligands with selectivity for telomeric G-quadruplexes and cancer cells. *Bioorganic & Medicinal Chemistry Letters* **20**: 6459-6463

Han H, Langley DR, Rangan A, Hurley LH (2001) Selective interactions of cationic porphyrins with G-quadruplex structures. *Journal of the American Chemical Society* **123**: 8902-8913

- Hanakahi LA, Sun H, Maizels N (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *The Journal of Biological Chemistry* **274**: 15908-15912
- Harrison GP, Lever AM (1992) The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *Journal of Virology* **66**: 4144-4153
- Hayflick L, Moorhead PS (1961) The serial cultivation of human diploid cell strains. *Experimental Cell Research* **25**: 585-621
- Heinzinger NK, Bukinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, Gendelman HE, Ratner L, Stevenson M, Emerman M (1994) The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 7311-7315
- Hemelaar J (2012) The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine* **18**: 182-192
- Hemelaar J, Gouws E, Ghys PD, Osmanov S, Isolation W-UNfH, Characterisation (2011) Global trends in molecular epidemiology of HIV-1 during 2000-2007. *Aids* **25**: 679-689
- Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang LS, Johnson FB (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **36**: 144-156
- Horwitz JP, Chua J, Noel M, Darooge MA (1964) Nucleosides. Iv. 1-(2-Deoxy-Beta-D-Lyxofuranosyl)-5-Iodouracil. *Journal of Medicinal Chemistry* **7**: 385-386
- Hu L, Lim KW, Bouaziz S, Phan AT (2009) Giardia telomeric sequence d(TAGGG)<sub>4</sub> forms two intramolecular G-quadruplexes in K<sup>+</sup> solution: effect of loop length and sequence on the folding topology. *Journal of the American Chemical Society* **131**: 16824-16831
- Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Research* **33**: 2908-2916
- Huppert JL, Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research* **35**: 406-413

## References

---

- Huppert JL, Bugaut A, Kumari S, Balasubramanian S (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Research* **36**: 6260-6268
- Incles CM, Schultes CM, Kempfski H, Koehler H, Kelland LR, Neidle S (2004) A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. *Molecular Cancer Therapeutics* **3**: 1201-1206
- Izbicka E, Wheelhouse RT, Raymond E, Davidson KK, Lawrence RA, Sun D, Windle BE, Hurley LH, Von Hoff DD (1999) Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Research* **59**: 639-644
- Jeeninga RE, Hoogenkamp M, Armand-Ugon M, de Baar M, Verhoef K, Berkhout B (2000) Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *Journal of Virology* **74**: 3740-3751
- Kankia BI, Barany G, Musier-Forsyth K (2005) Unfolding of DNA quadruplexes induced by HIV-1 nucleocapsid protein. *Nucleic Acids Research* **33**: 4395-4403
- Kestler HW, 3rd, Ringler DJ, Mori K, Panicali DL, Sehgal PK, Daniel MD, Desrosiers RC (1991) Importance of the nef gene for maintenance of high virus loads and for development of AIDS. *Cell* **65**: 651-662
- Kikin O, D'Antonio L, Bagga PS (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research* **34**: W676-682
- Kilareski EM, Shah S, Nonnemacher MR, Wigdahl B (2009) Regulation of HIV-1 transcription in cells of the monocyte-macrophage lineage. *Retrovirology* **6**: 118
- Kim MY, Vankayalapati H, Shin-Ya K, Wierzba K, Hurley LH (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
- Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**: 2011-2015
- Kim SY, Byrn R, Groopman J, Baltimore D (1989) Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *Journal of Virology* **63**: 3708-3713



---

Kypr J, Kejnovska I, Renciuik D, Vorlickova M (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research* **37**: 1713-1725

Lapadat-Tapolsky M, De Rocquigny H, Van Gent D, Roques B, Plasterk R, Darlix JL (1993) Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Research* **21**: 831-839

Law MJ, Lower KM, Voon HP, Hughes JR, Garrick D, Viprakasit V, Mitson M, De Gobbi M, Marra M, Morris A, Abbott A, Wilder SP, Taylor S, Santos GM, Cross J, Ayyub H, Jones S, Ragoussis J, Rhodes D, Dunham I, Higgs DR, Gibbons RJ (2010) ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143**: 367-378

Lee CH, Saksela K, Mirza UA, Chait BT, Kuriyan J (1996) Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. *Cell* **85**: 931-942

London TB, Barber LJ, Mosedale G, Kelly GP, Balasubramanian S, Hickson ID, Boulton SJ, Hiom K (2008) FANCI is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *The Journal of Biological Chemistry* **283**: 36132-36139

Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, Teulade-Fichou MP, Foiani M, Nicolas A (2011) G-quadruplex-induced instability during leading-strand replication. *The EMBO journal* **30**: 4033-4046

Luciw PA (1996) Human Immunodeficiency Viruses and their replication. In *Fields Virology*, third edn, 27, pp 845-918.

Lyonnais S, Gorelick RJ, Mergny JL, Le Cam E, Mirambeau G (2003) G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA. *Nucleic Acids Research* **31**: 5754-5763

Lyonnais S, Hounsou C, Teulade-Fichou MP, Jeusset J, Le Cam E, 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

Maizels N (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature Structural & Molecular Biology* **13**: 1055-1059

Membrino A, Cogoi S, Pedersen EB, Xodo LE (2011) G4-DNA formation in the HRAS promoter and rational design of decoy oligonucleotides for cancer therapy. *PloS One* **6**: e24421

## References

---

- Merson MH, O'Malley J, Serwadda D, Apisuk C (2008) The history and challenge of HIV prevention. *Lancet* **372**: 475-488
- Michael NL, D'Arcy L, Ehrenberg PK, Redfield RR (1994) Naturally occurring genotypes of the human immunodeficiency virus type 1 long terminal repeat display a wide range of basal and Tat-induced transcriptional activities. *Journal of Virology* **68**: 3163-3174
- Miller MD, Warmerdam MT, Gaston I, Greene WC, Feinberg MB (1994) The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. *The Journal of Experimental Medicine* **179**: 101-113
- Mirkin SM (2013) DNA replication: driving past four-stranded snags. *Nature* **497**: 449-450
- Mohaghegh P, Karow JK, Brosh RM, Jr., Bohr VA, Hickson ID (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Research* **29**: 2843-2849
- Morris MJ, Wingate KL, Silwal J, Leeper TC, Basu S (2012) The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Research* **40**: 4137-4145
- Mougel M, Houzet L, Darlix JL (2009) When is it time for reverse transcription to start and go? *Retrovirology* **6**: 24
- Mukundan VT, Phan AT (2013) Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *Journal of the American Chemical Society*
- Mullen MA, Olson KJ, Dallaire P, Major F, Assmann SM, Bevilacqua PC (2010) RNA G-Quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. *Nucleic Acids Research* **38**: 8149-8163
- Nadai M, Doria F, Di Antonio M, Sattin G, Germani L, Percivalle C, Palumbo M, Richter SN, Freccero M (2011) Naphthalene diimide scaffolds with dual reversible and covalent interaction properties towards G-quadruplex. *Biochimie* **93**: 1328-1340
- Nadai M, Palu G, Palumbo M, Richter SN (2012) Differential targeting of unpaired bases within duplex DNA by the natural compound clerocidin: a valuable tool to dissect DNA secondary structure. *PloS One* **7**: e52994

- Nadai M, Sattin G, Palu G, Palumbo M, Richter SN (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
- Nambiar M, Goldsmith G, Moorthy BT, Lieber MR, Joshi MV, Choudhary B, Hosur RV, Raghavan SC (2011) Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. *Nucleic Acids Research* **39**: 936-948
- Narute PS, Smithgall TE (2012) Nef alleles from all major HIV-1 clades activate Src-family kinases and enhance HIV-1 replication in an inhibitor-sensitive manner. *PloS One* **7**: e32561
- Neidle S (2009) The structures of quadruplex nucleic acids and their drug complexes. *Current Opinion in Structural Biology* **19**: 239-250
- 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
- Norseen J, Johnson FB, Lieberman PM (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
- O'Connor C (2008) Telomeres of human chromosomes. *Nature Education* **1**: 166
- O'Sullivan RJ, Karlseder J (2010) Telomeres: protecting chromosomes against genome instability. *Nature Reviews Molecular Cell Biology* **11**: 171-181
- Ou TM, Lu YJ, Tan JH, Huang ZS, Wong KY, Gu LQ (2008) G-quadruplexes: targets in anticancer drug design. *ChemMedChem* **3**: 690-713
- Palmisano L, Vella S (2011) A brief history of antiretroviral therapy of HIV infection: success and challenges. *Annali dell'Istituto superiore di sanità* **47**: 44-48
- Palumbo SL, Ebbinghaus SW, Hurley LH (2009) Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *Journal of the American Chemical Society* **131**: 10878-10891
- Pannecouque C, Daelemans D, De Clercq E (2008) Tetrazolium-based colorimetric assay for the detection of HIV replication inhibitors: revisited 20 years later. *Nature Protocols* **3**: 427-434

## References

---

- Pannecouque C, Fikkert, V., Schols, D., De Clercq, E., Witvrouw, M. (2000) Quaternary Ammonium Salts inhibiting HIV replication in vitro. *13th International Conference on Antiviral Research*
- Parkinson GN, Lee MP, Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**: 876-880
- Patel M, Yanagishita M, Roderiquez G, Bou-Habib DC, Oravec T, Hascall VC, Norcross MA (1993) Cell-surface heparan sulfate proteoglycan mediates HIV-1 infection of T-cell lines. *AIDS Research and Human Retroviruses* **9**: 167-174
- Paxton W, Connor RI, Landau NR (1993) Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis. *Journal of Virology* **67**: 7229-7237
- Peterlin BM, Trono D (2003) Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nature reviews Immunology* **3**: 97-107
- Phan AT (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *The FEBS Journal* **277**: 1107-1117
- Phan AT, Kuryavyi V, Luu KN, Patel DJ (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K<sup>+</sup> solution. *Nucleic Acids Research* **35**: 6517-6525
- Phan AT, Kuryavyi V, Patel DJ (2006) DNA architecture: from G to Z. *Current Opinion in Structural Biology* **16**: 288-298
- Phan AT, Patel DJ (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *Journal of the American Chemical Society* **125**: 15021-15027
- Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D (1998) Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *Journal of Virology* **72**: 2855-2864
- Qin Y, Fortin JS, Tye D, Gleason-Guzman M, Brooks TA, Hurley LH (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

- Qin Y, Hurley LH (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90**: 1149-1171
- Raiber EA, Kranaster R, Lam E, Nikan M, Balasubramanian S (2012) A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Research* **40**: 1499-1508
- Rambaut A, Posada D, Crandall KA, Holmes EC (2004) The causes and consequences of HIV evolution. *Nature Reviews Genetics* **5**: 52-61
- Rangan A, Fedoroff OY, Hurley LH (2001) Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule. *The Journal of Biological Chemistry* **276**: 4640-4646
- Rankin S, Reszka AP, Huppert J, Zloh M, Parkinson GN, Todd AK, Ladame S, Balasubramanian S, Neidle S (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *Journal of the American Chemical Society* **127**: 10584-10589
- Rawal P, Kummarasetti VB, Ravindran J, Kumar N, Halder K, Sharma R, Mukerji M, Das SK, Chowdhury S (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. *Genome Research* **16**: 644-655
- Rha SY, Izbicka E, Lawrence R, Davidson K, Sun D, Moyer MP, Roodman GD, Hurley L, Von Hoff D (2000) Effect of telomere and telomerase interactive agents on human tumor and normal cell lines. *Clinical Cancer Research : an official journal of the American Association for Cancer Research* **6**: 987-993
- Richter SN, Frasson I, Palu G (2009) Strategies for inhibiting function of HIV-1 accessory proteins: a necessary route to AIDS therapy? *Current Medicinal Chemistry* **16**: 267-286
- Roy S, Delling U, Chen CH, Rosen CA, Sonenberg N (1990) A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes & Development* **4**: 1365-1373
- Salvi R, Garbuglia AR, Di Caro A, Pulciani S, Montella F, Benedetto A (1998) Grossly defective nef gene sequences in a human immunodeficiency virus type 1-seropositive long-term nonprogressor. *Journal of Virology* **72**: 3646-3657
- Sanders CM (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *The Biochemical Journal* **430**: 119-128

## References

---

Schubert U, Bour S, Ferrer-Montiel AV, Montal M, Maldarell F, Strebel K (1996) The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains. *Journal of Virology* **70**: 809-819

Schultes CM, Guyen B, Cuesta J, Neidle S (2004) Synthesis, biophysical and biological evaluation of 3,6-bis-amidoacridines with extended 9-anilino substituents as potent G-quadruplex-binding telomerase inhibitors. *Bioorganic & Medicinal Chemistry Letters* **14**: 4347-4351

Schwartz S, Felber BK, Fenyo EM, Pavlakis GN (1990) Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *Journal of Virology* **64**: 5448-5456

Shen W, Gao L, Balakrishnan M, Bambara RA (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

Shi DF, Wheelhouse RT, Sun D, Hurley LH (2001) Quadruplex-interactive agents as telomerase inhibitors: synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. *Journal of Medicinal Chemistry* **44**: 4509-4523

Shin-ya K, Wierzba K, Matsuo K, Ohtani T, Yamada Y, Furihata K, Hayakawa Y, Seto H (2001) Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*. *Journal of the American Chemical Society* **123**: 1262-1263

Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH (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

Spina CA, Kwoh TJ, Chowes MY, Guatelli JC, Richman DD (1994) The importance of nef in the induction of human immunodeficiency virus type 1 replication from primary quiescent CD4 lymphocytes. *The Journal of Experimental Medicine* **179**: 115-123

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

Strebel K, Daugherty D, Clouse K, Cohen D, Folks T, Martin MA (1987) The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328**: 728-730

- Sun D, Guo K, Rusche JJ, Hurley LH (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
- Sundquist WI, 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
- Tluckova K, Marusic M, Tothova P, Bauer L, Sket P, Plavec J, Viglasky V (2013) Human papillomavirus g-quadruplexes. *Biochemistry* **52**: 7207-7216
- Todd AK, Neidle S (2008) The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Research* **36**: 2700-2704
- Tornaletti S (2009) Transcriptional processing of G4 DNA. *Molecular Carcinogenesis* **48**: 326-335
- UNAIDS/WHO, (2012) The strategic use of antiretrovirals to help end the HIV epidemic. WHO. Available at: [www.who.int/hiv/pub/arv/strategic\\_use/en/index.html](http://www.who.int/hiv/pub/arv/strategic_use/en/index.html). Accessed: November 1 2013.
- Verma A, Halder K, Halder R, Yadav VK, Rawal P, Thakur RK, Mohd F, Sharma A, Chowdhury S (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *Journal of Medicinal Chemistry* **51**: 5641-5649
- Vermeire J, Vanbillemont G, Witkowski W, Verhasselt B (2011) The Nef-infectivity enigma: mechanisms of enhanced lentiviral infection. *Current HIV research* **9**: 474-489
- Wang Y, Patel DJ (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1**: 263-282
- Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**: 711-716
- Weitzmann MN, Woodford KJ, Usdin K (1997) DNA secondary structures and the evolution of hypervariable tandem arrays. *The Journal of Biological Chemistry* **272**: 9517-9523

## References

---

White EW, Tanious F, Ismail MA, Reszka AP, Neidle S, Boykin DW, Wilson WD (2007) Structure-specific recognition of quadruplex DNA by organic cations: influence of shape, substituents and charge. *Biophysical Chemistry* **126**: 140-153

Wieland M, Hartig JS (2009) Investigation of mRNA quadruplex formation in *Escherichia coli*. *Nature Protocols* **4**: 1632-1640

Wilkinson RA, Pincus SH, Shepard JB, Walton SK, Bergin EP, Labib M, Teintze M (2011) Novel compounds containing multiple guanide groups that bind the HIV coreceptor CXCR4. *Antimicrobial Agents and Chemotherapy* **55**: 255-263

Xu Y, Sugiyama H (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Research* **34**: 949-954

Ying L, Green JJ, Li H, Klenerman D, Balasubramanian S (2003) Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 14629-14634

Yu Z, Gaerig V, Cui Y, Kang H, Gokhale V, Zhao Y, Hurley LH, Mao H (2012) Tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. *Journal of the American Chemical Society* **134**: 5157-5164

Zahler AM, Williamson JR, Cech TR, Prescott DM (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**: 718-720



---

## Publications

- Perrone R, Nadai M, Frasson I, Poe JA, Butovskaya E, Smithgall TE, Palumbo M, Palù G, Richter SN (2013) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *Journal of Medicinal Chemistry* **56**: 6521-6530
- Perrone R, Nadai M, Poe JA, Frasson I, Palumbo M, Palù G, Smithgall TE, Richter SN (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

## Conferences

- Perrone R, Frasson I, Nadai M, Freccero M, Palù G, Richter SN (2011) Investigation of nucleic acid secondary structures in the HIV-1 integrated genome *10<sup>th</sup> National Congress of the Italian Society for Virology*. Orvieto (Italy), September 12-14 **Poster Presentation**
- Perrone R, Nadai M, Del Vecchio C, Frasson I, Palumbo M, Palù G, Parolin C, Richter SN (2012) Formation of a unique cluster of G-quadruplex structures in the HIV-1 *nef* coding region: implications for antiviral activity *11<sup>th</sup> National Congress of the Italian Society for Virology*. Orvieto (Italy), September 17-19 **Poster Presentation**
- Perrone R, Nadai M, Del Vecchio C, Frasson I, Palumbo M, Parolin C, Palù G, Richter SN (2012) Formation of a unique cluster of G-quadruplex structures in the HIV-1 *nef* coding region: implications for antiviral activity *COST Action MP0802 Final Conference*. Sitges (Spain), October 6-8 **Poster Presentation**
- Perrone R, Nadai M, Frasson I, Butovskaya E, Poe JA, Palumbo M, Palù G, Smithgall TE, Richter SN (2013) Targeting G-quadruplexes in the HIV-1 integrated genome: a novel antiviral strategy? *4<sup>th</sup> International Meeting on G-Quadruplex Nucleic Acids from Structure to Chemistry and Biology*. Singapore, July 1-4 **Poster Presentation**
- Perrone R, Nadai M, Frasson I, Butovskaya E, Palumbo M, Palù G, Richter SN (2013) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter *5<sup>th</sup> European Congress of Virology*. Lyon (France), September 11-14 **Oral Presentation**

