Università degli Studi di Padova

Head Office: Università degli Studi di Padova

Department of Molecular Medicine

Ph.D. COURSE IN: MOLECULAR MEDICINE
CURRICULUM: BIOMEDICINE
SERIES XXXII

LIGANDS-MEDIATED MODULATION OF G-QUADRUPLEX STRUCTURES WITHIN THE HIV-1 GENOME DURING LYTIC AND LATENT STATE OF INFECTION

Thesis written with the financial contribution of the European Research Council
(ERC-2013-CoG 615879)

Coordinator: Prof. Stefano Piccolo
Supervisor: Prof. Sara Richter

Ph.D. student: Paola Soldà

"*Success is not final, failure is not fatal:*
it is the courage to continue that counts"
Winston Churchill

Index

Abstract ..... *VII*

Sommario ..... *IX*

1. Introduction *1*

1.1 G-quadruplex structures **1**
1.1.1 G-quadruplex structures: beyond the canonical architecture of the double helix 1
1.1.2 DNA and RNA G-quadruplex structures: general features 1
1.1.3 G-quadruplex interaction with ligands 4
1.1.4 G-quadruplex distribution throughout genomes 10

1.2 The human immunodeficiency virus type 1 ..... **19**
1.2.1 The HIV-1 worldwide: epidemiology 19
1.2.2 HIV-1 general features 19
1.2.3 A deepening on HIV-1 latency 25
1.2.4 Currently approved treatments and the hope of eradication 26
1.2.5 G4 structures in HIV-1 28

1.3 The herpes simplex virus type 1 ..... **30**
1.2.1 The HSV-1 worldwide: epidemiology 30
1.2.2 HSV-1 general features 31
1.2.3 Currently approved treatments 34
1.2.4 G4 structures in HSV-1 35

2. Aim of the study *37*

3. Materials and methods *39*

3.1 The HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands **39**
3.1.1 The HIV-1 recombinant nucleocapsid protein 39
3.1.2 Oligonucleotides and compound 39
3.1.3 Electrophoretic mobility shift assay (EMSA) 39
3.1.4 Circular dichroism (CD) analysis 40
3.1.5 Mass spectrometry (MS) analysis 40
3.1.6 Reverse transcriptase (RT) stop assay 41

3.2 The HIV-1 LTR G-quadruplexes fold in latently infected cells and their stabilization by G4 ligands counteract viral reactivation from latency **41**
3.2.1 Cell line 41
3.2.2 Cell treatments 42
3.2.3 BG4 production and purification 42
3.2.4 Chromatin immunoprecipitation (ChIP) assay 42
3.2.5 qPCR 43
3.2.6 Cytotoxicity assay 44
3.2.7 p24 ELISA assay 44

3.3 Screening of a new series of small molecules as antivirals **45**
3.3.1 Cell lines 45
3.3.2 Viruses and viral stocks production 45
3.3.3 Antiviral assays 45
3.3.4 Cytotoxicity assays 46
3.3.5 Oligonucleotides and compounds 46
3.3.6 Circular dichroism (CD) analysis 47
3.3.7 *Taq* Polymerase stop assay 47

3.3.8 Mass spectrometry (MS) competition assay 48
3.3.9 Time of addition (TOA) assay 49
3.3.10 UV-visible and fluorescence spectra 49
3.3.11 Confocal microscopy analysis 49
3.3.12 Immunoblot analysis 50

4. Results and discussion 51

4.1 The HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands 51
4.1.1 NCp7 binds and unfolds the RNA G4s in the U3 region of the HIV-1 RNA genome .. 51
4.1.2 NCp7-mediated unfolding of the RNA G4s promotes reverse transcriptase processivity 56

4.2 The HIV-1 LTR G-quadruplexes fold in latently infected cells and their stabilization by G4 ligands counteracts viral reactivation from latency 59
4.2.1 The HIV-1 LTR G-quadruplexes fold in latently infected cells 60
4.2.2 The HIV-1 LTR G4s stabilization by G4 ligands counteracts viral reactivation from latency 61

4.3 Screening of a new series of small molecules as antivirals 65
4.3.1 Antiviral activities of the Quindoline-derived compounds 66
4.3.2 The Quindoline-derived best candidate greatly stabilizes G4s with a preference for the viral HSV-1 conformations *vs* the telomeric sequence 68
4.3.3 GSA-0932 acts at early events of the viral life cycle 73
4.3.4 GSA-0932 colocalizes with the VP16-GFP viral protein at early stages of infection .. 77
4.3.5 GSA-0932 decreases the essential viral immediate-early ICP4 protein 79

5. Conclusions 83

6. Appendix 85

6.1 List of recurrent abbreviations 85

6.2 Index of Figures 87

6.3 Index of Tables 89

7. Bibliography 91

Abstract

G-quadruplexes (G4s) are non-canonical nucleic acids secondary structures. Their presence has been firmly established in the human genome, as well as in that of many viruses. Our research group has previously identified highly conserved G4s in the U3 region of the HIV-1 genome and the long terminal repeat (LTR) promoter of the proviral genome. In the present work, we addressed how the RNA and DNA HIV-1 G4s can be modulated by various ligands (proteins and chemical compounds), leading to the regulation of important virulence processes. Unveiling so far unknown mechanisms in the HIV-1 infection regulation could help the discovery of new targets for the design of specific inhibitors to be proposed as new antivirals.

Initially, we focused on the identification of new RNA G4 binding proteins and the derived modulation. We proved that the HIV-1 nucleocapsid protein (NCp7) binds and unfolds the HIV-1 RNA U3 G4s. Importantly, the NCp7-mediated RNA G4 structure resolution favoured proceeding of the reverse transcription (RT). Conversely, the G4 ligand BRACO-19 (B19) stabilized the RNA G4 folding, thus hindering RT progression, and also counteracting the destabilizing activity of NCp7. On the one hand, our data pointed out the strength of NCp7 as chaperone protein that is able to process the extremely stable HIV-1 RNA G4s in order to allow viral retro-transcription to occur. On the other, they indicate a new target of G4 ligands that inhibit both RT and NCp7. This information brings out the possibility to develop selective U3 RNA G4 ligands, to be proposed as effective anti-HIV-1 drugs with an innovative mechanism of action.

We then moved to the investigation of the regulatory role of DNA LTR G4s on viral latency. We reported for the first time the actual folding of the viral LTR G4s within the context of HIV-1 latent chromatin, where the viral promoter is in a repressed state. Moreover, LTR G4s stabilization mediated by two G4 ligands down-modulated viral transcription, thus counteracting viral reactivation from latency. We presented here the possibility to selectively target LTR G4s to control viral latency and develop innovative antiretroviral drugs to manage HIV-1 infection. In this context, targeting LTR G4s constitutes the basis for the progress of new antiviral compounds with an unprecedented mechanism of action. However, the vast majority of G4 binders present high molecular weights and protonated side chains, which may cause bioavailability problems when subjected to *in vivo* studies. As a consequence, we here investigated the antiviral activity of a new series of Quindoline-derived compounds presenting more drug-like features. Our data highlight the importance to develop small druggable molecules to target viral G4s in order to handle viral infections.

Sommario

I G-quadruplex (G4s) sono strutture secondarie non canoniche degli acidi nucleici che sono state identificate in regioni regolatorie del genoma umano, così come in quelle di molti virus. Il gruppo di ricerca della Professoressa Richter ha precedentemente identificato strutture G4 altamente conservate nella regione U3 del genoma di HIV-1 e anche in una regione promotoriale del genoma provirale. In questa tesi tratteremo di come i G4 presenti nell'RNA e nel DNA di HIV-1 possano essere modulati da vari ligandi (sia proteine che composti chimici), ottenendo così la modulazione di importanti processi virologici. La scoperta di meccanismi finora sconosciuti nella regolazione dell'infezione da HIV-1 potrebbe portare all'identificazione di nuovi *target* per la progettazione di specifici inibitori da poter proporre come antivirali.

Inizialmente ci siamo concentrati sull'identificazione di nuove proteine in grado di legare gli RNA G4 e sulla loro conseguente modulazione. Abbiamo dimostrato che la proteina virale nucleocapside (NCp7) è in grado di legare e svolgere il G4 nella regione U3 di HIV-1. E' importante sottolineare che la risoluzione della struttura G4 mediata da NCp7 è stata osservata favorire il processo di trascrizione inversa. Al contrario, il ligando BRACO-19 (B19), stabilizzando il ripiegamento dell'RNA a G4, ha ostacolato la progressione della trascrittasi inversa, anche contrastando l'attività destabilizzante di NCp7. Da un lato, i nostri dati hanno sottolineato la forza della proteina NCp7 come *chaperone* che risolve le strutture G4 altamente stabili dell'RNA di HIV-1 al fine di consentire la trascrizione inversa virale. Dall'altro lato, abbiamo portato una nuova conseguenza dell'attività del ligando G4, legata all'inibizione dell'attività della trascrittasi inversa e di NCp7. Queste informazioni hanno messo in evidenza la possibilità di sviluppare ligandi selettivi per il G4 dell'RNA U3 di HIV-1 al fine di proporre terapie anti-HIV-1 più efficaci con un meccanismo d'azione innovativo.

Siamo poi passati ad indagare il ruolo regolatorio che le strutture G4 nel DNA LTR di HIV-1 possono avere sulla latenza. Abbiamo dimostrato per la prima volta il *folding* autentico dei G4 virali nell'LTR in un contesto di cromatina latente, in cui il promotore si trova in uno stato inattivo. Inoltre, la stabilizzazione dei G4 dell'LTR mediata da due ligandi G4 ha inibito la trascrizione virale, contrastando così la riattivazione virale dalla latenza. Abbiamo quindi presentato la possibilità di bersagliare selettivamente i G4 virali nell'LTR di HIV-1 per controllare la latenza virale e sviluppare farmaci antiretrovirali innovativi per gestire l'infezione da HIV-1. In questo contesto, colpire selettivamente i G4 dell'LTR costituisce la base per l'avanzamento di nuovi composti antivirali con un meccanismo d'azione senza precedenti. Tuttavia, la maggior parte dei ligandi G4 presenta pesi molecolari elevati e sostituenti protonati che possono causare problemi di scarsa biodisponibilità quando sottoposti a studi *in vivo*. Di conseguenza, ci siamo focalizzati sullo studio dell'attività antivirale di una nuova serie di composti derivati dalla Quindolina che presentano caratteristiche chimico-fisiche più promettenti. I nostri risultati hanno dimostrato l'importanza di sviluppare piccole molecole per colpire selettivamente i G4 virali al fine di gestire le infezioni virali.

1. Introduction

1.1 G-quadruplex structures

1.1.1 G-quadruplex structures: beyond the canonical architecture of the double helix

The canonical DNA secondary structure is the right-handed double helix (also known as B-DNA form) described by Watson and Crick over 60 years ago.[1] This structure represents the predominant and more energetically favoured form in which DNA can be found.

Despite the traditional view that considers DNA as a relatively stable element, numerous studies report that the DNA is highly dynamic.[2,3] In fact, beyond the canonical DNA form, it can adopt various alternative conformations. More than 10 types of so-called non-B or non-canonical nucleic acids secondary structures have been characterized in the years.[4,5] These include G-quadruplex structures, as well as i-motifs, triplexes, cruciforms, hairpins (Figure 1.1).[6]

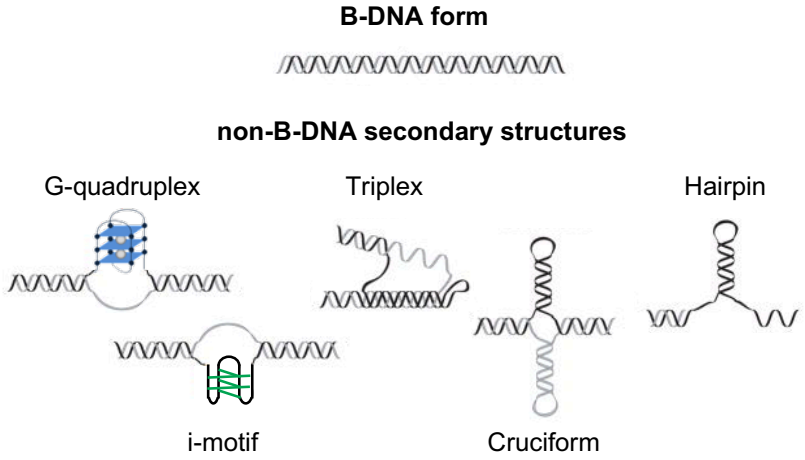


Figure 1.1: **Representation of some non-canonical DNA secondary structures compared to the B-DNA form.** Adapted from Saini et al.[6]

Accumulating evidence suggests that non-B DNA structures may play crucial biological roles in genetic instability, chromatin-remodeling, DNA damage and repair, dysregulation of gene expression, alteration of the replication landscape, thus contributing to the generation of a variety of human diseases, such as cancers.[4,7]

1.1.2 DNA and RNA G-quadruplex structures: general features

The very first scientific evidence that led to the discovery of G-quadruplex structures over the years was reported by the German chemist Ivar Bang in 1910. Almost 40 years before the knowledge of the DNA double helix, he noted that guanylic acid forms a viscous gel at high concentrations.[8] About 50 years later, X-ray diffraction revealed that guanylic acids could

effectively assemble into four-stranded helical structures[9] that account for the gel-like properties of the aqueous solution. Then, the formation and organization of what we now called a G-tetrad or G-quartet were revealed for the first time. The G-tetrad consists of a square planar arrangement in which each of the four molecules of guanylic acid is the donor and acceptor of two hydrogen bonds. Precisely, G-tetrads are held together by eight Hoogsteen hydrogen bonds (Figure 1.2). These bonds differ from those observed in canonical Watson–Crick pairing and involve the interaction of the N7 group from one guanine with the exocyclic amino group from a neighboring base.[10–12]

Interest in nucleic acids intensified over the following decades. Various studies reported that guanosine homo-oligomers could adopt the G-tetrad conformation, both in the ribose and deoxyribose forms.[13] Thus, both DNA and RNA sequences particularly enriched in guanine residues have the potential to assemble into the G-tetrad arrangement. Moreover, two or more G-quartets can stack on top of each other to form the tetraplex secondary structure, now known as G-quadruplex structure (G4).

The formation and stabilization of G4s highly depend on the presence of monovalent and bivalent cations, especially K^+, and to a lesser extent Ca^{2+}, NH_4^+, Na^+, Rb^+, Mg^{2+}, Ni^+, Cs^+.[14,15] The stabilization effect is due to the coordination of the positively charged cations with the electronegative O6 atoms in the center channel of the adjacent stacked G-tetrads.[16] When the ionic radius of the cation is relatively small (e.g., in the case of Na^+), the ions can be located in the middle of a single G-tetrad and coordinate four oxygen atoms. Alternatively, the ion may be embedded between two G-tetrads and coordinate eight oxygen atoms, as for larger cations, like NH_4^+ and K^+.[17] Generally, the nature of the cations has a considerable effect on the stability of the resultant G4s. Since the K^+ induces high thermodynamic stability to the G4 and it is the main cation *in vivo*, G4s can readily form *in vitro* mimicking its physiological intracellular concentration (~140 mM).[18]

From the earliest days of studying G-quadruplexes *in vitro*, extensive structural polymorphism was noted. First of all, G4s can be classified into intramolecular (also unimolecular) or intermolecular, since the G-tracts that participate in the G4s formation can derive from one or multiple (up to four) strands (Figure 1.2).[11,18] Moreover, Gs in each quartet can adopt either a *syn* or an *anti* glycosidic bond angle conformation, which is the torsion angle of the bond between the base and the sugar. Consequently, each of the four G-tracts can run in the same or opposite direction with respect to its two neighbors, forming parallel, anti-parallel, or hybrid core conformations (Figure 1.2). Parallel G4s have all the Gs in the *anti* glycosidic bond angles conformation. Instead, antiparallel topologies derive from the combination of both *anti* and *syn* conformations of the glycosidic bond angle.[11] Depending on these orientations, the G-blocks delimit four negatively charged grooves of different sizes (narrow, medium, or wide).[12] The loops, which are the sequences located between successive G-tracts that serve to link stacked G-quartets, give other contributions to conformational diversity, particularly for intramolecular G4s. The loops can adopt three different conformations: lateral, diagonal, or propeller (Figure 1.2).[18] Specifically, the parallel topology requires a connecting loop to link the bottom G-tetrad with the

top G-tetrad, leading to propeller-type loops. Antiparallel and mixed-type G4s allow greater variety of loop conformations, resulting in diagonal and lateral loops, in addition to the propeller one(Figure 1.2).[11]

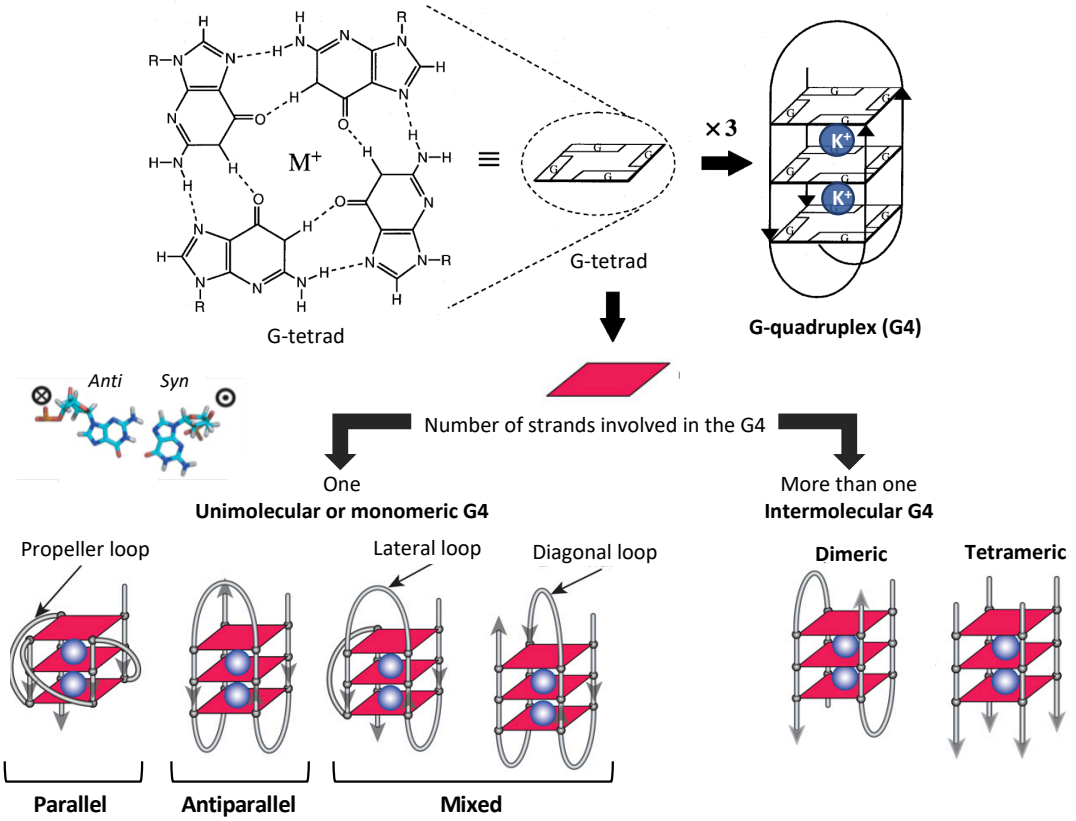


Figure 1.2: **Arrangement of guanosine residues into a G4 structure.** Hoogsteen hydrogen bonds are represented as dash lines. Blue circles represent potassium ions that coordinate the structure. One or multiple (two or more) strands can participate in the G4s formation giving shape to monomeric, dimeric, or tetrameric G4s. Depending on the *syn*/*anti* glycosidic bond angle conformation, the strands can form parallel, anti-parallel, or hybrid G4s. The loops can adopt three different conformations: lateral, diagonal, or propeller. Adapted from Chen et al.[18]

Recently, increasing studies describe the formation of very singular G4 structures. As an example, since the G-blocks can be interrupted by non-G nucleotides, bulges have been observed to protrude from the G4 core, further increasing G4s complexity.[19]

The just explained structural general features reflect the organization of the DNA G4s and also those of RNA G4s, with some variables and topological limitations.[20] Of course, the most relevant differences between RNA and DNA G4s are those between RNA and DNA themselves: the presence of a ribose sugar instead of a deoxyribose sugar and of uracil instead of thymidine.[21] The additional presence of a hydroxyl group in position 2' of the ribose sugar has several consequences. First of all, it makes RNA G4s more stable with respect to the DNA counterpart because of the formation of additional intramolecular hydrogen bonds, the increase in stacking interactions, and the recruitment of more water molecules within the RNA G4 grooves.[21] Moreover, the 2'-hydroxyl group prevents the ribose sugar to adopt the *syn*-type glycosidic angle conformation, strongly favouring the *anti*-conformation. As a consequence, the topology of all the

known naturally occurring RNA G4s is limited only to the parallel conformation, where all the four strands are oriented in the same direction.[20,21]

Loops length and composition highly dictate the stability of both DNA and RNA G4s. Since RNA G4s can adopt only parallel topology, the loops are present only in their propeller-type conformation. Moreover, the substitution of thymidine with uracil in the loop sequences increase the stability of RNA G4s, both augmenting stacking interactions and making them less hydrated.[22]

Other parameters can induce differences between DNA and RNA G4s, such as the cation binding specificity. Even if K^+ greatly stabilizes both DNA and RNA G4s, Na^+, Mg^{2+}, Ca^{2+} seem to have no effect on RNA G4 stability.[21]

Altogether, we can summarize that RNA G4s are generally more compact, less hydrated and often more thermodynamically stable than DNA G4s (Figure 1.3).[21]

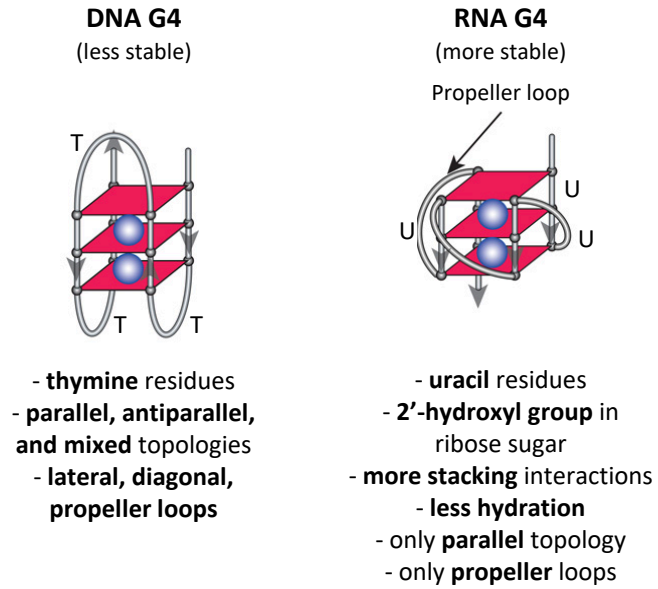


Figure 1.3: **Differences between DNA and RNA G4 structures.** Adapted from [18] and [22].

Overall, DNA and RNA G4s can present a high level of topological variants, depending on multiple factors. Nowadays, advanced technologies could help us to better characterize their precise conformation, spatial arrangement, and stability. The final goal is to exploit the high polymorphism of G4s to identify specific structural determinants in order to allow selective recognition. A special interest comes from the pharmaceutical field. Indeed, ligands could be designed to selectively target a specific G4 over other structures (e.g., duplexes or other G4s), in order to develop new therapeutic drugs against various human diseases.

1.1.3 G-quadruplex interaction with ligands

G4 structures possess structural determinants that can be considered as interaction sites for various ligands. The binding modes between binders and G4s and the specificity of the recognition are important aspects to be considered for the purpose of selectivity. Moreover, since ligands binding can either stabilize or destabilize G4s, the G4 folding and unfolding modulation is a matter of particular interest, especially for the possible biological implications.

Generally, both chemical compounds (natural and synthetic) and proteins can be described by the generic term ligands. They will be discussed separately in the following paragraph.

1.1.3.1 G4s interaction with chemical compounds

The therapeutic possibilities of targeting G4s by interacting compounds were first reported by Sun and co-workers in 1997. They reported that the targeting of the telomeric G4s with small molecules makes telomeric ends inaccessible for the telomerase-mediated extension, thus inhibiting the enzyme activity.[23] From that moment, much effort has been directed toward the identification and design of G4-interactive compounds, in order to propose promising potential therapeutics (especially against cancer).

Since G4s possess various interaction sites, which mainly include terminal G-tetrads, grooves, and loops, several modes of interaction between G4s and compounds have been elucidated (Figure 1.4).

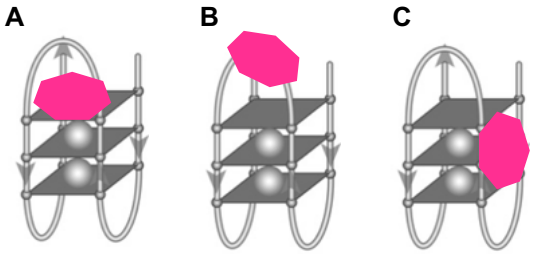


Figure 1.4: **Examples of G4-compound interacting modes.** (**A**) External stacking binding mode. (**B**) Loop binding. (**C**) Groove binding.

Generally, most of the G4-interacting compounds have an aromatic surface for stacking with G-tetrads, a positive charge or basic groups to bind to loops and grooves, and steric bulk to prevent intercalation within double-stranded DNA.[24] Therefore, common G4 binders are flat aromatic molecules with fused-ring systems that can establish $\pi - \pi$ stacking interactions with the external tetrads of the G4.[18,24,25] External stacking binding mode is the preferred way of interaction of the vast majority of the G4 ligands. It does not require the transient unstacking of G-quartets, as instead occurs in the intercalation mode. Intercalation is not energetically favoured because of the G4 rigidity,[25] and it is also not desired because it could lead to duplex targeting as well. For these reasons, the compounds steric bulks are important to prevent intercalation.[26]

The G4 loops and grooves are also accessible for small molecules interactions. Many of the identified ligands contain side chains with amino groups or other substituents that can be positively charged by protonation or methylation. Protonated side chains contribute to better recognition and stronger binding to the loops, grooves, and the negatively charged phosphate backbone of the G4s through electrostatic interactions.[26–28] In contrast to the G4-tetrads, which are a common feature of all the G4s, loops and grooves are strictly dependent on the DNA/RNA sequence involved in the formation of the secondary structure. Then, since they are different in size and shape between one structure and another, they can represent attractive binding sites for the selective targeting of G4s.[26,28]

An important aspect of be considered is the possibility to specifically target RNA G4s versus the DNA counterpart. In fact, as previously explained, RNA G4s possess unique chemical properties, suggesting it is possible to develop selective small molecules against RNA G4.[20] Until now, only one compound exhibits high molecular specificity for RNA over DNA G4s, which is the carboxypyridostatin derivative.[20]

To date, through rational structure-based and *in silico* drug design, a large number of compounds from many chemical classes has been developed against G4 nucleic acids.[29] Specifically, around 1000 chemical compounds targeting both DNA and RNA G4s have been reported in the G-Quadruplex Ligands Database (G4LDB).[24,30] Some of the most studied G4 interacting compounds (natural and synthetic) are: telomestatin,[31] BRACO-19,[32] TMPyP4,[33] PIPER (a perylene diimide),[34] NDIs (naphthalene diimides),[35] PDS (pyridostatin),[36] PhenDC3,[37] Quindoline,[38] Quarfloxin.[39]

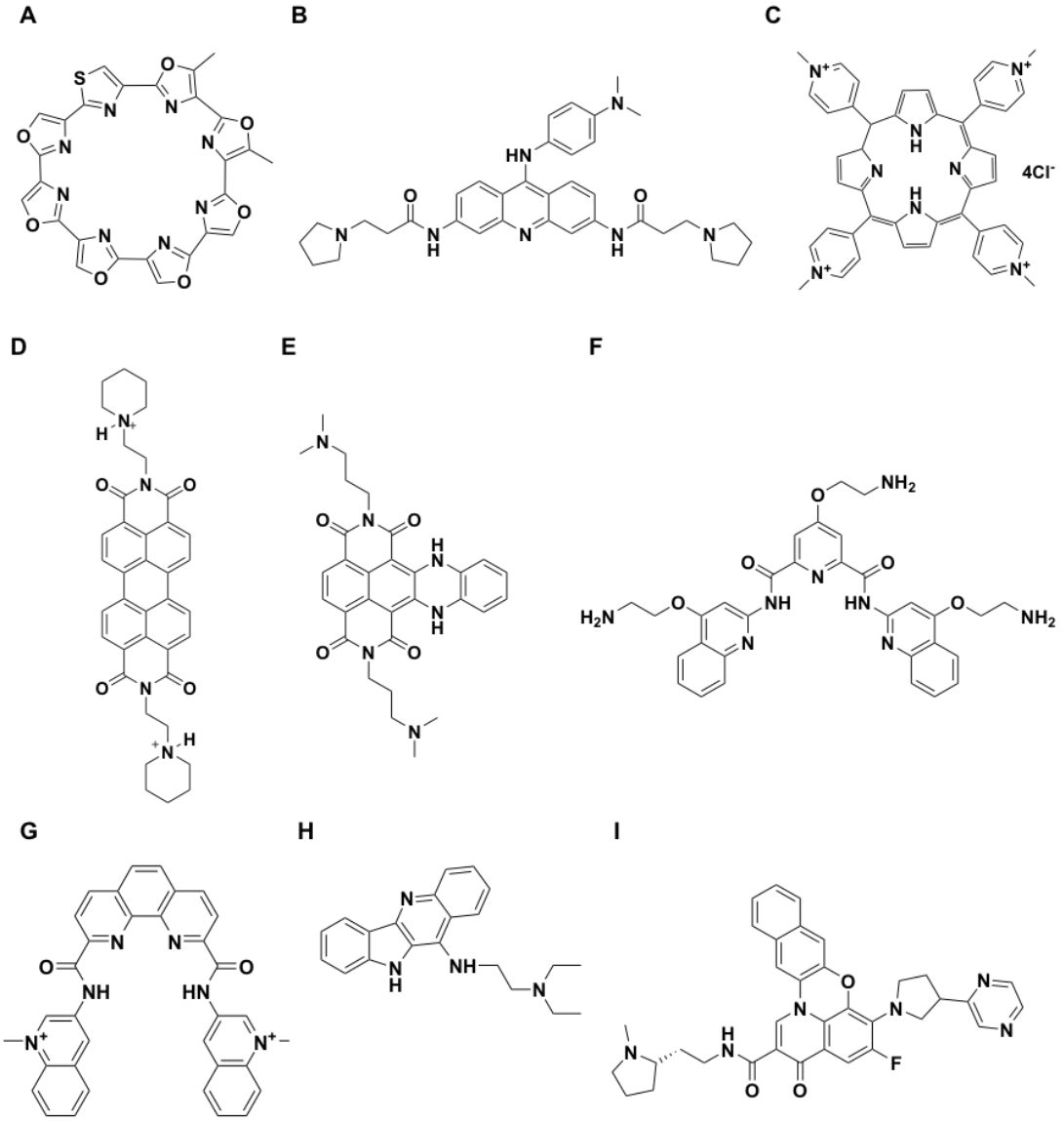


Figure 1.5: **General chemical structures of some of the most relevant G4 ligands.** (**A**) Telomestatin; (**B**) BRACO-19; (**C**) TMPyP4; (**D**) PIPER; (**E**) c-exNDI; (**F**) PDS; (**G**) PhenDC3; (**H**) Quindoline; (**I**) Quarfloxin.

For the purpose of this work, BRACO-19, derivatives of the naphthalene diimides (NDI) family, and Quindoline-like compounds will be focused on.

BRACO-19 (3,6,9-trisubstituted acridine 9-[4-(*N,N*-dimethylamino)phenylamino]-3,6-bis(3-pyrrolodinopropionamido)acridine, or B19) is a trisubstituted acridine and can be considered an "*in situ* protonated G4 ligand".[28] It was rationally designed by computer modeling. Specifically, its central acridine aromatic core has been introduced to allow efficient G-tetrads stacking. Moreover, three tertiary amine substituents were inserted to interact with the grooves, since they are protonated at physiological pH.[32] Overall, B19 has been reported to possess higher affinity for G4s over the duplex DNA, low cytotoxicity, and to efficiently inhibit telomerase activity.[40] Biopharmaceutical studies concluded that its poor permeability across biological barriers restrains its pharmacological application.[41] However, B19 remains one of the most studied G4 interacting compounds, and it is considered a reference molecule in G4 research.[42]

One of the more promising classes of G4 binders is that of the naphthalene diimides (NDI) derivatives.[35] The dimensions of the aromatic central core dictate the recognition of different DNA conformations. In particular, at least four condensed rings are required to bind G4s over the double-stranded DNA efficiently.[43] Moreover, the great advantage of the NDIs is that up to four different side chains can be introduced to the scaffold. These substituents can be protonated at physiological pH, further enhancing G4 affinity and selectivity. The tri- and tetra-substituted NDIs display very promising G4 binding activity with respect to the di-substituted ones.[44]

Thanks to their highly versatile structures, various types of NDIs have been synthesized over the years. Overall, the activity of the NDI compounds was found to be strongly related to the length of the side chains (distance between the central core and nitrogen atom), the entity of substituents, and the number of positive charges. More recently, a new series of NDI has been designed extending the aromatic surface. They have been classified as core-extended NDI derivatives. These c-exNDIs has been obtained fusing the NDI core with various structures, quinone methides,[45] 1,4-dihydropyrazine-2,3-dione,[46] and dihydrobenzophenazine heterocycle,[47] to name a few. Besides increased affinity to the G4 structures, the water solubility and the intrinsic fluorescence of these compounds also make them promising probes for G4 detection.[46]

Some natural alkaloids have been described as G4 ligands, such as GQC-05 and various Quindoline derivatives.[48] GQC-05 (or NSC338258 compound) derived from the plant alkaloid ellipticine (ellipticine dihydrochloride). It was reported to bind with high affinity the *c-myc* G4, stabilizing it and inducing dramatic cytotoxicity against myeloma cells mediated by the cell cycle arrest and apoptosis progression.

Quindoline is a naturally occurring indolo[3,2-*b*]quinoline alkaloid. Some studies highlighted its ability to bind and stabilize the G4 formed in the *c-myc* promoter.[49] NMR studies resolved the 2:1 solution structure complex formed from the small molecules and the biologically relevant *c-myc* G4.[50] The resolution of the interaction showed an unexpected drug-induced reorientation of the flanking sequences at both ends of the DNA sequence involved in the G4. Quindoline derivative compounds have been developed through a rational drug design.[38] In 2000, Neidle et al. first reported that the 2,10-disubstituted quindoline derivative could bind to G-quadruplex structures

in human telomeres, showing inhibitory activity against telomerase with moderate cytotoxicity.[38] They further determined that the 2,7-disubstituted quindoline derivative showed potent activity. Various observations pointed out that electrostatic interactions are important for the G4 selectivity. Therefore it has been expected that introducing an electron donor such as an amino group in the 11-position of quindoline could increase the electron density, increasing the electrostatic interaction between the quindoline derivatives and the negative electrostatic center of the G4.[51] These derivatives showed improved inhibition of the telomerase activity and a marked cessation in cell growth. The GSA-0820 (2-(4-10H-indolo[3,2-b]quiolin-11-yl)piperazin-1-yl)-*N,N*-dymethilanamine) is another derivative of this family. It significantly down-regulated the *c-myc* promoter activity, and it was reported to stabilize a competing G4 formed on SMN2 Exon 7 after splicing mechanisms. Actual studies are focused on the elucidation of the G4-related anticancer mechanisms of this compound.

The group of Professor Laurence Hurley intensively studied Quindoline-like compounds and also designed one of the two G4 binder that entered clinical trials: Quarfloxin. Quarfloxin is a fluoroquinolone derivative compound,[39] originally developed to target the G4 found in the *c-myc* promoter. Quarfloxin is highly selective for G4 over duplex and single-stranded DNA and it reached Phase II clinical trials.[42] Unfortunately, it was withdrawn due to bioavailability related problems.[52]

A major limitation of the so far described G4 interacting compounds is low selectivity profiles and poor drug-like properties. In fact, to improve G4 binding, the aromatic ring count, positive charge, and the number of hydrogen bond donors generally exceed what would be optimal for a small molecule with good pharmacokinetic properties.[24] This is why no G4 ligand has so far advanced beyond Phase II drug discovery pathway. Therefore, research in the G4 binding compounds will need to improve both the selectivity and the drug-likeness of G4 binders, in order for them to be employed in *in vivo* studies in the next future.[42]

1.1.3.2 G4s interaction with proteins

DNA and RNA G4 structures can be specifically recognized by numerous binding proteins.[53] Until now, more than 200 interacting proteins belonging to different species have been identified. The first and free database G4 Interacting Protein DataBase (G4IPDB) contains detailed information about all of them.[54] Among all, 77 are G4 binding proteins of Homo sapiens.

Generally, it has been observed that human G4 binding proteins possess unique features, such as prominent enrichment in glycine (G) and arginine (R) residues.[55] The arginine/glycine-rich region is termed RGG domain and is important in G4-protein interactions. Indeed, studies reported that the RGG motif might enable additional hydrogen bonding and π-stacking interaction with nucleobases, thus strengthening G4 binding affinity.[56]

The expression and the level of expression of proteins can greatly depend on the type and the state of the cell. In addition, they can have variable distribution within the cell, as certain proteins naturally accumulate in distinct compartments (e.g., in the nucleus).

Depending on the preferential binding to DNA or RNA G4s, proteins can be divided into DNA and RNA G4 interactors.[53] Moreover, DNA binders may belong to different classes of proteins, such as telomeres binding proteins, helicases, transcriptional activators, transcriptional repressors, epigenetic regulators.

The group of the telomeric G4 binding proteins mainly includes POT1 (protection of telomere 1), and TEBP (telomere end binding proteins; α-β heterodimer). POT1 is a member of the shelterin complex. It normally binds the single-stranded DNA overhangs at the ends of chromosomes and is essential for chromosome end-protection. It has been demonstrated to bind telomeric G4s.[57]

Helicases are the proteins deputed to separate individual strands of duplex DNA to promote replication and transcription. Pif1, FANCJ, RecQ, BLM (Bloom syndrome protein), and WRN (Werner syndrome protein) are the best characterized G4 helicases.[58] All of them were shown to induce the unfolding of different G4 structures.

Among all the characterized G4 binding protein, many of them are involved in transcriptional regulation. Depending on their effect on transcription, they may be separated into transcriptional activators or repressors. PARP-1 (poly[ADP-ribose]polymerase 1), nucleolin, nucleophosmin, MAZ (myc-associated zinc-finger), and CNBP (cellular nucleic acid binding protein) are the main examples. Soldatenkov and co-workers demonstrated that PARP-1 binds to intramolecular DNA G4 *in vitro*, with a 2:1 stoichiometry.[59] Nucleolin is a nucleolar phosphoprotein particularly expressed in highly proliferating cells. It has been first identified as the binding and stabilizing partner of the G4-structured transcriptionally-inactive form of the *c-myc* promoter.[60] Later it has been demonstrated to bind viral HIV-1 G4s *in vitro,*[61] with a preference for long-looped G4s,[62] and also to recognize the mRNA EBNA1 G4.[63]

Furthermore, various epigenetic and chromatin remodeling enzymes have been reported to bind to G4s. A representative example is that of the DNA methyltransferase 1 (DNMT1) enzyme. It binds to G4s with stronger affinity *in vitro* compared to duplex DNA, and it loses enzymatic activity upon G4 binding.[64]

Due to the structural similarities between RNA and DNA G4s, some of the described proteins have also been reported to bind RNA G4s, with some peculiarities. Mainly hnRNPs (heterogeneous nuclear ribonucleoproteins), ribosomal proteins, and splicing factors were reported to interact with RNA G4s.[65] HnRNPs are a large family of abundant nuclear proteins involved in various aspects of the RNA metabolism (such as alternative splicing, mRNA transport, trafficking, and export), and have been reported to bind G4s through their essential RRG-box region. For instance, the hnRNP A1 has been shown to preferentially bind the telomere RNA G4 both *in vitro* and in live cells, presenting unfolding activity.[66] The same effect has been observed after hnRNP A2/B1 binding to HIV-1 LTR G4s.[67]

Besides human-derived proteins, other proteins have been identified in yeasts, such as Sub1[68] and Slx9.[69] So far, only one viral protein, i.e., EBNA1 of the Epstein-Barr virus, has been shown to bind to folded RNA G4s to promote viral DNA replication.[70]

Although chemical compounds tend to induce a preferential G4 stabilization effect, proteins can both stabilize or destabilize the non-canonical secondary structures, leading to various cellular

effects depending on the G4 location within cells and on the cell cycle step. Approaches that can reveal the composition and dynamics of complexes between G4s and interacting proteins would be needed to improve the knowledge of the G4 interactome. This may in turn present opportunities for small molecule modulations of these interactions.[71]

1.1.3.3 Tools for G4 detection

Until now, both functionalized compounds (i.e., fluorescent probes) and "special proteins" (monoclonal antibodies) have been developed to selectively detect and visualize G4s in cells. Appropriately functionalized compounds, such as radiolabeled and alkyne functionalized coupled with fluorophores G4 ligands, were initially developed. Moreover, intrinsically fluorescent molecules that display different fluorescent emission or excitation maxima, and appropriate fluorescent decay lifetimes upon binding to G4s (e.g., o-BMVC, DAOTA-M2), have also been exploited for cells imaging.[71] The compound o-BMVC is an example of fluorescent biomarker selective for G4 recognition over the duplex.[72]

Antibodies were also employed to visualize G4s. Very few G4-specific antibodies have been prepared: hf2,[73] BG4,[74] and 1H6.[75] The single-chain antibody hf2 detect G4s *in vitro* with at least 100-fold higher affinity with respect to duplexes. However, this tool has been proved to be unsuitable for whole-cell immunofluorescence.[76] 1H6 was obtained immunizing mice with stable G4s. It has been used to visualize G4s within cells, and it does not to detect RNA G4s.[75] The FLAG-tagged single-chain variable fragment (scFv) antibody was generated by phage display and characterized through *in vitro* selection on G4s. It possesses nanomolar affinity towards both DNA and RNA G4s, and it allows detection of G4s in cells by immunofluorescence.[74]

The introduction of G4-selective antibodies allowed the acquisition of substantial evidence of the effective formation of G4 structures in the cell genome.[74] However, questions remain about whether they can be sensitive to the level of a single G-quadruplex or can only detect high local densities of multiple G-quadruplex motifs. Therefore, a map of the precise locations of G4s at the whole genome level would be preferable.

1.1.4 G-quadruplex distribution throughout genomes

For biology, the important question is if, where and when G4 structures are formed *in vivo*. The implementation of predictive techniques (*in silico* algorithms) and technological advancement (i.e., next-generation sequencing methods) enabled the characterization and mapping of a lot of G4 structures in genomes. G4s have been extensively reported in the human genome, as well as in the genome of non-human organisms, including other mammals,[77] bacteria,[78] yeasts,[79] viruses,[12,42,80] and plants.[81] Generally, these structures are highly conserved, indicating a selection pressure to retain such sequences at specific genomic sites.[82] Moreover, this conservation seems to be highest among mammalians with respect to non-mammalian or lower organisms.

The G4 structures that may form in the human genome and the genetic content of the known human viruses are considered as fundamental pillars in the research of Professor Richter's group.

1.1.4.1 G4s in the human genome and their regulatory roles

Bioinformatic predictive tools (e.g., QGRS mapper,[83] QuadBase,[84] G4Hunter,[85] Quadfinder[86]) employ algorithms to determine how many putative G4 sequences (PQSs) a genome would be expected to contain. These computational analyses have predicted over 370000 G4 sequence motifs in the human genome.[87]

Although predictive algorithms are very useful in identifying new G4s, they all present limitations. The main problem is that these techniques cannot ascertain if G4s can effectively be present/folded or not in cells. Moreover, they do not accurately predict all the PQSs in a genome with motifs with large loops, non-guanine bulges, or other variables, which represent another stumbling block.[76,88]

Recently, the advent of next-generation sequencing (NGS) has provided an excellent opportunity to design specific methods to identify G4s at a genome/transcriptome-wide level. G4-seq and BG4-ChIP-seq were then employed to map G4s in the human genome.[89,90]

G4-seq assay offers an *in vitro* reference map of sequences that can form G4s in purified human genomic DNA.[76] The G4-seq technique identified 700000 G4s in the human genome:[89] about twice as much as the number of G4s predicted *in silico* by standard algorithms. This is possibly due to the presence of many non-canonical G4s, with bulges or long loops.

More recently, the BG4 antibody has been used to isolate G4-containing fragments from human chromatin.[90] The analysis yielded about 10000 G4s in the chromatin from human epidermal keratinocytes (NHEKs) and 1000 G4s from spontaneously immortalized HaCaT keratinocytes.[90] This represents only 1% of the sites with the capability to form G4s as observed in G4-Seq.[89] The difference in the number of G4s reported from *in vitro* and *in vivo* techniques (i.e., G4-seq versus G4-ChIP) suggests that the cellular environment may play a central role in affecting the dynamics of G4 formation in cells.[71,76] Specifically, the native chromatin context could mask the majority of G4 epitopes, limiting BG4 sensitivity. The G4 number variability can also depend on the fact that the PQSs can be folded into G4s only in specific cellular conditions. Moreover, many cellular factors, such as binding proteins (e.g., helicases), can likely remodel the DNA G4 landscape *in vivo*.[71,76]

A traditional view considered that since the genomic DNA is primarily double-stranded, the formation of G4s may be favored when DNA is rendered transiently single-stranded. This could occur mainly during DNA replication, transcription, and repair. Telomeres and mRNA are physiologically single-stranded nucleic acids forms, which are expected to be more prone to fold into G4s. However, G4 formation can depend on negative supercoiling, molecular crowding as well as interacting binding proteins. Therefore, these and other observations suggest that G-quadruplex formation may not necessarily require the prior melting of the DNA double helix, but various elements should be considered, making this picture much more complex than initially thought. Then, more analyses are necessary to understand how G4s are distributed in different cell lines and during different phases of the cell cycle, making the G4-related research still an open field.

In silico algorithms and G4-NGS methods allow identifying a particular G4 enrichment in peculiar regions of the human genome. Initially, bioinformatic analysis reported that G4s appeared not randomly distributed, but enriched in functional regions associated with genome regulation, such as telomeres, promoters, and 5′ untranslated regions (5'-UTR) (Figure 1.6).[91] A few years later, the majority of G4s identified by BG4-ChIP-seq were found in nucleosome-depleted chromatin regions corresponding to promoters of highly transcribed genes.[90]

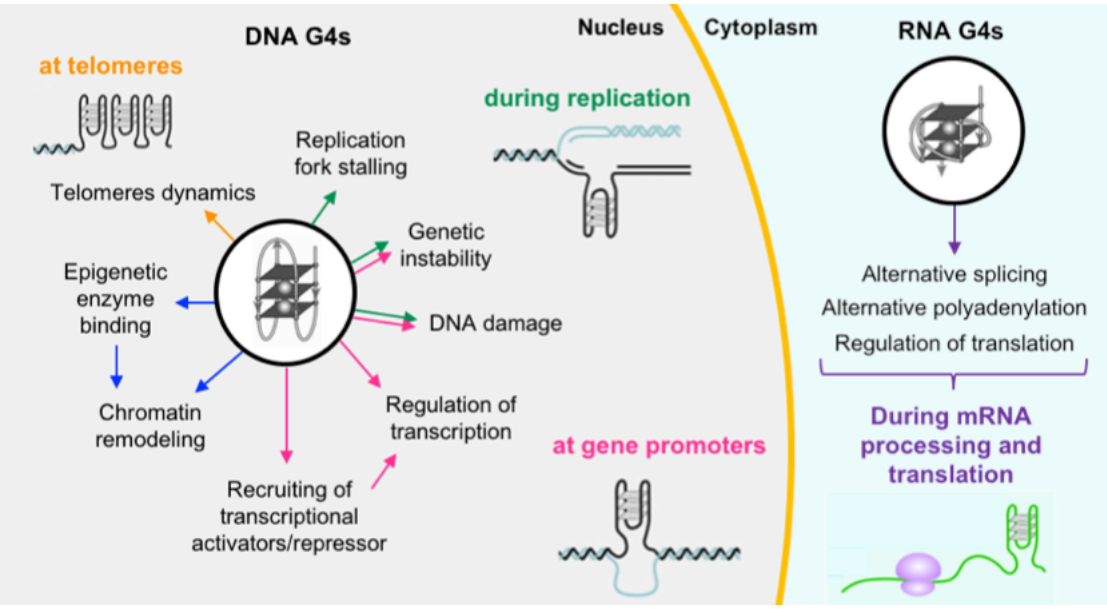


Figure 1.6: A conceptual representation of the regulatory roles of DNA and RNA G4 structures.

1.1.4.1.1 DNA G4s at telomeres

Telomeres, repeated sequences at the end of chromosomes, protect chromosomal DNA from degradation and recombination. They play an important role in genomic stability, aging, and cancer. Approximately 100-200 nucleotides of the telomeric sequence remain unpaired and form single-stranded G-rich regions called telomeric 3' overhang. These ends have strong potential to fold into G4s *in vitro*,[92] as well as *in vivo*.[75] The formation of telomeric G4s had been observed to impair the telomerase function.[93] The telomerase enzyme elongates the telomere sequences in cells with high replication rates. It is overexpressed and activated in 85-90% of human cancer cells.[94] Since the abnormal processing of the telomere 3′ overhang region may confer cellular immortality, the telomere G4 has been thought as a potential target for small molecules. G4-targeting compounds (e.g., telomestatin, BRACO-19) have been shown to inhibit telomerase activity,[40,95,96] to disrupt telomere capping and induce rapid apoptosis.[29] Similarly, the telomeric G4 binding proteins TEBPα and TEBPβ have been reported to promote G4 folding.[97] As an opposite effect, binding of the POT1 protein has been identified to unwind and disrupt the G4 structure, enhancing telomeric ends elongation by the telomerase.[57]

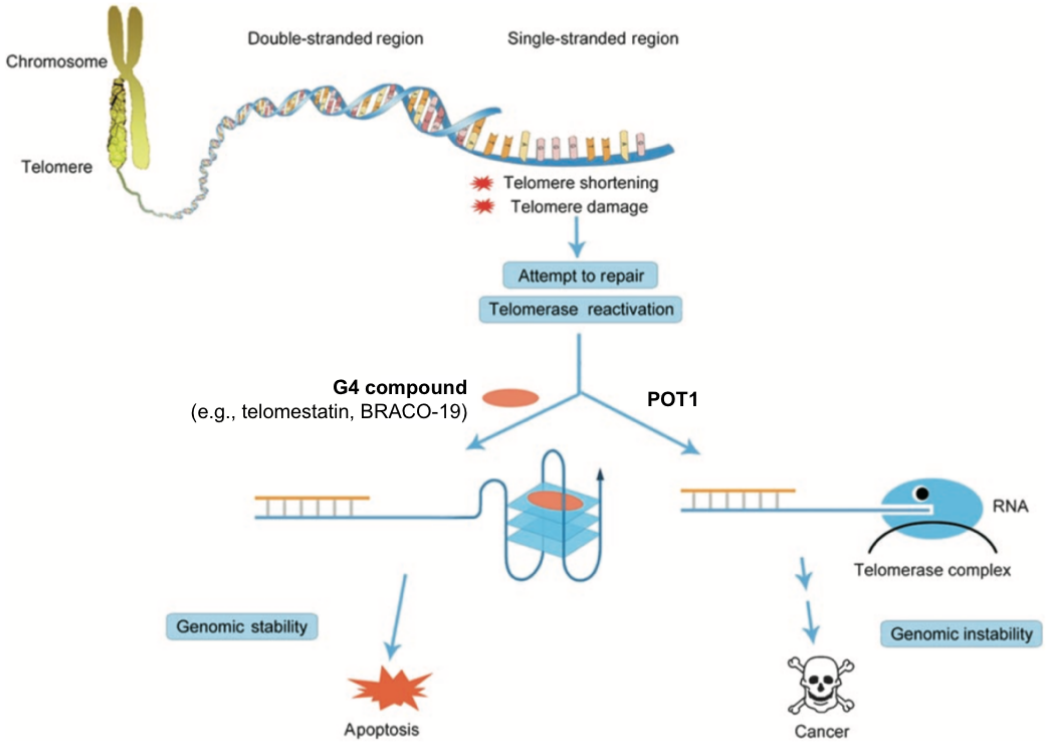


Figure 1.7: **Biological roles of G4s at telomeres**. Excessive repair mechanisms or telomerase activation may induce cellular immortality. The G4 unfolding activity of POT1 could enhance telomeric elongation, leading to tumorigenesis. G4 interacting compounds could induce or stabilize G4s, inhibiting telomerase binding and favouring apoptosis. Figure from [98].

1.1.4.1.2 DNA G4s during replication

During replication, the DNA double strand has to be separated by the helicases activity, allowing the synthesis of the new filaments by the DNA polymerase. In these circumstances, the DNA is single-stranded and provides a favorable environment for G4 formation. Various models have been postulated for the biological role of G4 structures during DNA replication.[99] It has been reported that G4s can form both on the leading or lagging strand during replication. Once formed, G4s may sterically impede template processing by the polymerase enzyme, thus triggering fork pausing and instability.[99,100] This evidence have been further confirmed by the fact that stabilization of G4s by small molecules induced genetic instability. Therefore, to ensure the correct progression of the replication process, G4s have to be denatured. Many helicases have been reported to solve G4s.[58] The best characterized examples are Pif1 and FANCJ. For instance, Pif1 has been proved to unwind a tetramolecular G4 formed by the CEB1 sequence in the leading strand of the replication fork. Moreover, the G4 compound PhenDC3 inhibited Pif1-mediated G4 unwinding.[58] FANCJ unfolded G4s in a preferential 5'-3' polarity, suggesting activity on the lagging strand, and the presence of telomestatin impaired its unwinding activity *in vitro*.[101]

1.1.4.1.3 DNA G4s in gene promoters and their regulation of transcription

Nearly 50% of the human genes have been observed to present G4 motifs within their promoter. Specifically, they have been found in proximity to transcription start sites (TSS).[91,102]

Correlation studies between PQSs and gene functions revealed that G4s are under-represented in the promoters of house-keeping and tumor suppressor genes, whereas proto-oncogenes are over-represented in G4 motifs.[91] These findings opened up further investigations about G4s in promoter regions and strong evidence proved their presence and functions in promoters of several genes that are related to the so-called "six hallmarks of cancer".[103] The most representative G4s in cancer-related genes are: *c-myc*,[104] KRAS,[105] VEGF,[106] Bcl-2,[107] c-kit,[108] hTERT,[109] and PDGF-A[110] (Figure 1.8).

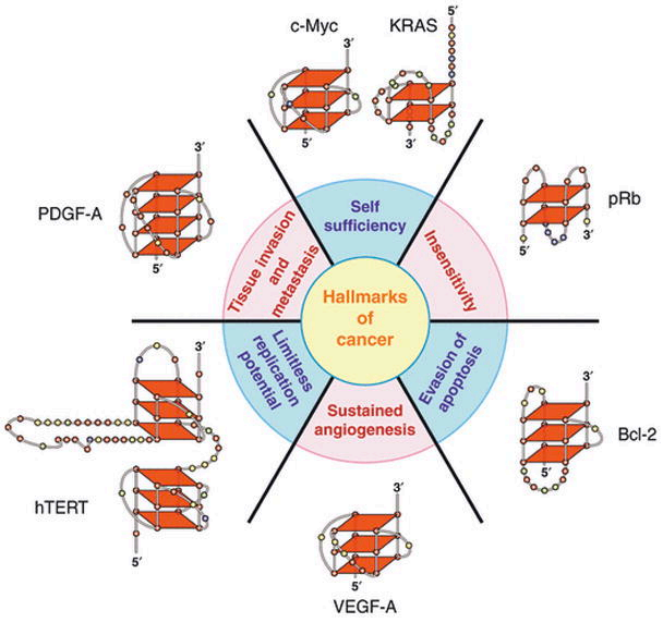


Figure 1.8: **The six hallmarks of cancer and the association of them with the G4 structures found in their promoter regions.** Figure from [103].

G4 structures at promoter regions of oncogenes represent the most deeply investigated G4 sites. Above all, G4 motifs in the nuclease hypersensitive element (NHE)III$_1$ of the *c-myc* promoter are the first and most studied example of the implication of G4 in the fine tuning of transcription.[60,104,111–113]

The high prevalence of G4 motifs in gene promoters suggests that they could modulate repression or activation of gene expression and have an essential role in gene transcription regulation. Generally, G4s near promoter regions may influence transcription in both positive and negative ways, depending on different factors.[114]

First of all, it has to be considered that G4s may or may not be formed under physiological conditions, due to the fact that sufficient kinetic energy is necessary to promote their folding. During transcription, a source of this energy could be represented by the negative superhelicity accumulated behind the RNA polymerase machinery during its progression.[115]

Another important aspect is that G4s can fold in regulatory (such as promoters) or coding regions. In addition, they can form on the template or non-template strand (Figure 1.9, panels A and B). Depending on which DNA strand encodes the G4 motif, the structure could either inhibit transcription (if the motif is on the template strand, blocking the transcription machinery) or

enhance transcription (if the motif is on the non-template strand, maintaining the transcribed strand in a more G4 prone single-stranded conformation).
Moreover, considering that enrichment in G4 motifs correlated with conserved transcription factor binding sites, G4s can modulate transcription both preventing or recruiting interacting proteins (Figure 1.9, panels C and D). Indeed, G4s can bind transcriptional activators to facilitate polymerase progression or can bind repressors that negatively affect transcription.[114] As previously mentioned, PARP-1, nucleolin, nucleophosmin, MAZ, CNBP, and various helicases can be cited as transcription regulatory proteins via G4 interaction.

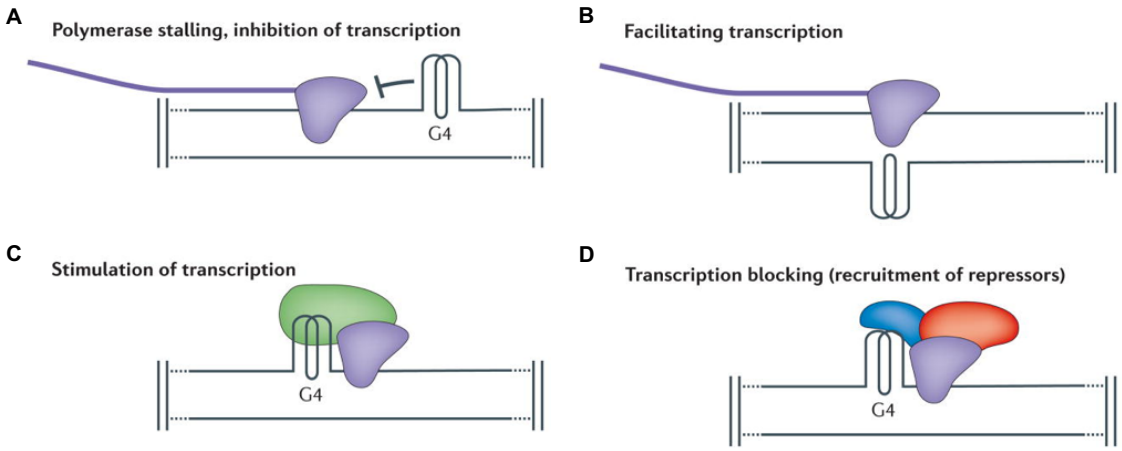


Figure 1.9: Possible functional roles of G4s during the transcription process. Figure from [114].

Another issue that should not be neglected for a comprehensive study of G4 involvement in transcription is epigenetics. Since genes expression depends on cell-types and different phases of the cells life, chromatin compaction, DNA methylation, and histones modifications have to be considered for transcription and gene expression modulation. A correlation between G4s and epigenetic modifications has been reported in several studies.[90,116,117] Specifically, the formation of non-canonical DNA secondary structures is known to affect the positioning of epigenetic marks, such as the cytosine methylation at the CpG islands. An interesting project reported that CpGs with low methylation were enriched in G4 motifs, while high methylation correlated with G4-depleted regions.[118] Various chromatin remodeling enzymes have also been reported to bind to G4s, such as the DNMT1 which loses its methyltransferase activity upon strong G4 interaction.[64] As a consequence, G4 formation can orchestrate CpGs methylation, that is one of the major epigenetic mechanisms for chromatin modelling, thus regulating genes transcription.
Altogether, transcription is a very complex and finely regulated process, where G4s can exert both activator or repressor roles, depending also on the contribution of the whole cellular environment. Further work is needed to unravel the mechanistic and molecular details of how G4s could influence transcription.

1.1.4.1.3 RNA G4s and their role in translation

RNA G4 structures have recently caught the attention thanks to two observations. First, RNA is generally found in cells in a single strand form, so G-rich sequences are more prone to fold into G4s with respect to the DNA duplex form. Second, RNA is located in cells cytoplasm and this is a great advantage because they are more accessible by different types of small molecules.
Computational predictions revealed that RNA G4s are present in the 5′ untranslated region (UTR) of many genes of clinical interest.[119] Some studies reported that RNA G4s in the 5'-UTR inhibited the translational processes, whereas other evidence suggested that G4s and other secondary structures are required for the initiation of translation.[120,121] Moreover, RNA G4s have been discovered in the mRNA 3'-UTR,[122] also downstream of an endo-nucleolytic cleavage site, suggesting their role in mRNA processing.[123] Moreover, it has been proven that RNA G4s may influence the translational process also in the open reading frames (ORFs) of mRNA coding regions, where they can cause frameshift mutations.[124] Altogether, important roles have been attributed to them, such as pre-mRNA splicing and polyadenylation, mRNA translation (Figure 1.6). Therefore, RNA G4s appear to be key regulatory motifs of the transcriptome.

1.1.4.2 G4s in viruses and their potentiality

Besides the human genome, PQSs have been found in the genome of almost all known viruses that can infect humans.[125] Occurrence and location of putative G4s largely varied between each virus class and family. In particular, according to Baltimore's classification, PQSs were highly enriched in the negative strand of retroviruses (ssRNA (RT) viruses), ssDNA viruses and both strands of dsDNA viruses. Conversely, dsRNA and dsDNA (RT) viruses notably lacked the presence of G4 forming motifs.[125] Thus, very different viruses may use G4s to control DNA/RNA dynamics in ways that could be relevant to pathogenic mechanisms and virulence processes.[126] Over the years, the interest in viral G4s has been increasing. Several studies reported the presence of G4s in viruses, together with evidence of their involvement in virus key steps (Figure 1.10).[12,42]

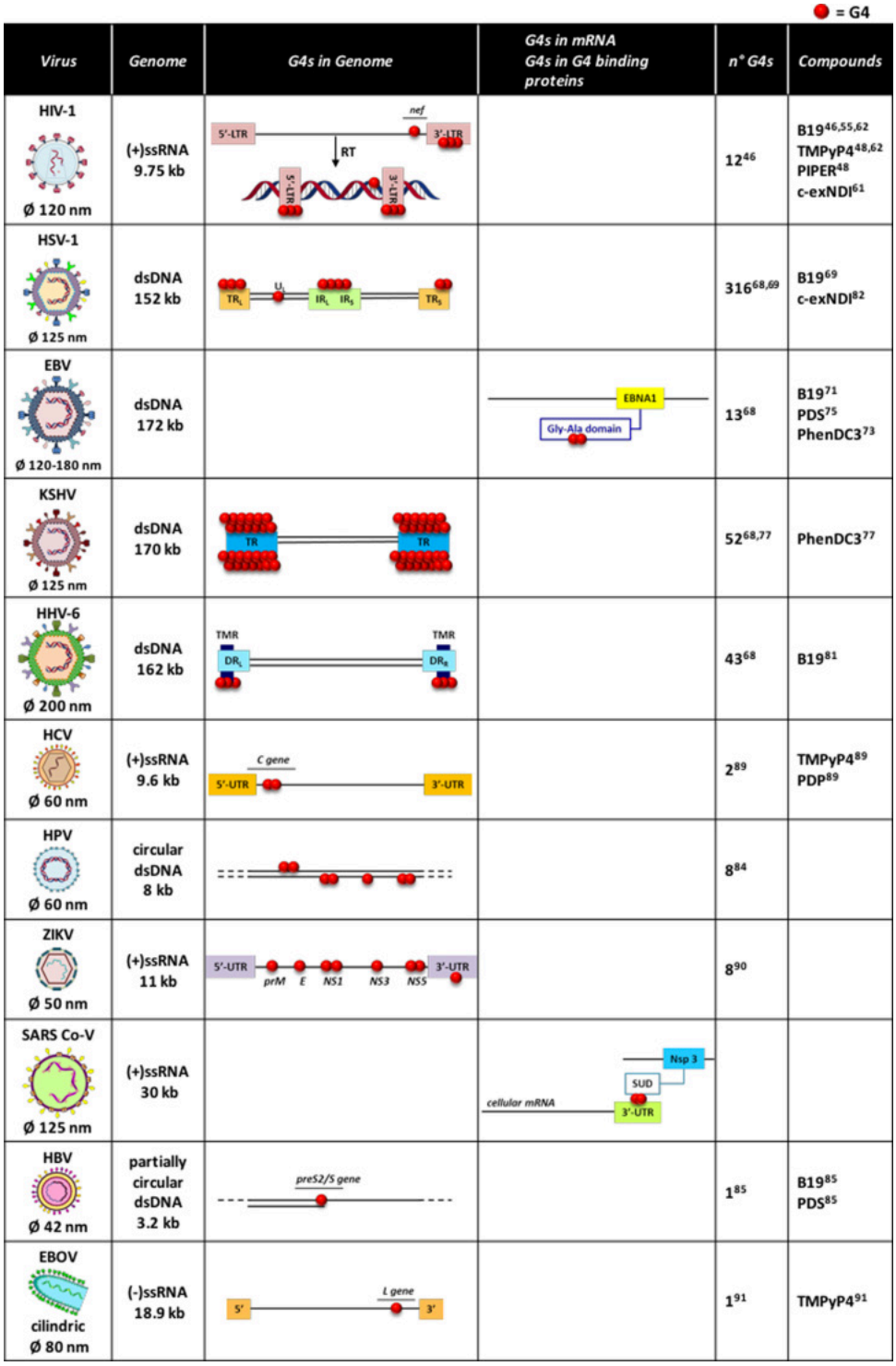


Figure 1.10: **Summary of the so far reported G4s in the viral genomes.** Information regarding virion structure and dimension, genome size and organization, location of G4s, number of assessed G4s, and type of tested G4 ligands are reported for each virus. Figure from [42].

Specifically, in the dsDNA group, G4s have been described in both *Herpesviridae* and *Papillomaviridae* families.[125] Among herpesviruses, G4s have been characterized in herpes simplex virus type 1 (HSV-1),[127] Epstein-Barr virus (EBV),[128] human herpesvirus 6 (HHV-6),[129] and 8 (HHV-8), also known as Kaposi's sarcoma-associated herpesvirus (KHSV).[130]

In 2009, Lieberman and co-workers reported for the first time a G4 structure in the EBV genome.[70] Studies proved that the EBV-encoded nuclear antigen-1 (EBNA1) protein recruits the cellular origin replication complex (ORC) through an interaction with RNA G4s, promoting viral DNA replication. G4 ligand treatment inhibited EBNA1-dependent stimulation of viral DNA replication and EBNA1 synthesis.[131] The EBNA1 mRNA itself can fold into parallel G4s, regulating viral mRNA translation, producing ribosome dissociation. G4s in EBNA1 mRNA have been shown to modulate also the endogenous presentation of EBNA1-specific $CD8^+$ T-cell epitopes, which are involved in persistent infections.[132] The cellular protein nucleolin interacts with EBNA1 mRNA G4s downregulating EBNA1 protein expression and antigen presentation.[63]

The KSHV genome is flanked by G-rich terminal repeats, which are able to form stable G4s, both in the forward and reverse strands.[130]

The HHV-6 genome presents telomeric regions at its termini. Since telomeres can fold into G4s, the stabilization of these structures by a G4 ligand have been observed to inhibit HHV-6 chromosomal integration.[129]

Concerning the *Papillomaviridae* family, G4s have been found in only eight out of 120 identified HPV types.[133] However, these include some of the most high-risk cervical cancer HPV types (e.g., HPV52 and HPV58).[42] In HPV52/58, G4s are located in the long control region (LCR), which is a regulatory sequence, suggesting a potential role in transcription and replication. In the other HPV types, PQSs have been demonstrated in the sequences codifying for the L2, E1, and E4 proteins, suggesting a possible implication of G4s in alternative splicing regulation.[12]

In ssDNA viruses, the presence of G4s was reported in the adeno-associated virus (AAV) genome.[134] In particular, 18 PQSs have been identified in the terminal repeat region. The cellular nucleophosmin protein, which enhances AAV infectivity, has been shown to interact with those G4s directly.[134]

Besides DNA viruses, both positive and negative-sense single-stranded RNA viruses ((+)ssRNA and (-)ssRNA, respectively) were described to contain G4s.[12,125] The severe acute respiratory syndrome coronavirus (SARS-CoV),[135] the Hepatitis C virus (HCV),[136] and the Zika virus (ZIKV)[137] belong to the first group. The SARS unique domain (SUD) of the non-structural protein 3 (Nsp3) has been reported to preferentially bind G4-forming oligonucleotides that may be found in the 3'-UTR of human mRNA. Since these mRNA encodes for proteins involved in apoptosis and signal transduction, it has been proposed that SUD/G4 interaction may be involved in controlling the host cell's response to the viral infection.[135] Wang et al. proved the existence of two highly conserved RNA G4s sequences in the C gene of HCV. Stabilization of these structures by G4 interacting compounds reduces RNA replication and inhibits protein translation.[136] Several PQSs were discovered in the ZIKV genome. The most interesting one localizes in the unique 3'-UTR region and it has appeared crucial for initial viral replication of the negative-sense strand.[137] G4s have also been proved in Ebola virus (EBOV) and Marburg virus (MARV), which belong to the (-)ssRNA class.[42] TMPyP4 treatment has been observed to stabilize the G4 that may form in the L gene of EBOV, thus inhibiting gene expression and viral replication.[138]

A G4 was also detected in the Hepatitis B virus (HBV) genome, the only member of dsDNA viruses with RT activity. The highly conserved G4 motif was identified in the promoter of the preS2/S gene and was observed to regulate transcription positively.[139]

Finally, functionally significant G4s were identified both in the RNA and DNA genomes of the human immunodeficiency virus (HIV)[140–142], a retrovirus, and in the LTR region of lentiviruses in general.[143]

Overall, the presence of G4 in human viruses, their crucial involvement in the virus life cycle, and the possibility to target them in order to obtain an antiviral activity have been illustrated. Future studies on viral G4s will provide a deeper understanding of the G4-mediated regulation of pathogen virulence. Moreover, DNA or RNA viral G4s could be targeted by specific G4 ligands, with the hope of finding new solutions against viral infections.

The research group of Professor S. Richter has been studying G4 structures among virus genomes for years. Since this thesis particularly focused on G4s in the genome of HIV-1 and HSV-1, these viruses' main features and the G4 implications in their biology will be presented in the following paragraphs.

1.2 The human immunodeficiency virus type 1

1.2.1 The HIV-1 worldwide: epidemiology

The Human Immunodeficiency Virus (HIV) was first characterized in 1983.[144] It is a member of the *Retroviridae* family in the Lentivirus genus. HIV is the causal agent of the Acquired Immunodeficiency Syndrome (AIDS), a condition in which the progressive failure of the human immune system allows life-threatening opportunistic infections and cancers to thrive. Nowadays, AIDS remains one of the most significant infectious disease and AIDS-related illnesses are one of the leading causes of death and premature mortality worldwide.[145] The Joint United Nations Programme on HIV/AIDS (UNAIDS) in 2018 estimated that 1.7 million people were newly infected with HIV and 37.9 million people globally were living with HIV. Among them, 23.3 million were accessing antiretroviral therapy. The World Health Organization (WHO) reported also that the Sub-Saharan Africa remained most severely affected, with nearly 1 HIV-positive in every 25 adults, accounting for nearly two-thirds of the people living with HIV worldwide.

Despite advances in our scientific understanding of HIV and its prevention and treatment as well as years of significant effort by the global health community, leading government, and civil society organizations, UNAIDS cautions that the pace of progress in reducing new HIV infections, increasing access to treatment, and ending AIDS-related deaths is slowing down. Therefore, the need to pay global attention to HIV infection is still high.

1.2.2 HIV-1 general features

HIV can be divided into two major types: HIV-1 and HIV-2. HIV-2 is less virulent, has lower transmission rate, and takes longer time to progress to AIDS than HIV-1.[146] Type 1 virus may be

further divided into groups: group M (“major”), O (“outlier”), N (“non-M, non-O”), and a possible group P (“pending the identification of further human cases”). Group M is the most common type of HIV-1, with more than 90% of HIV-1 infections deriving from it. The M group can be further subdivided into clades or subtypes (A, B, C, D, E, F, G, H, I, J, K), but there are also other circulating recombinant forms (CRFs).[147]

HIV-1 can also be grouped into two classes depending on the tropism of the virus to different cell types. Generally, HIV-1 can infects vital cells of the human immune system, such as T lymphocytes (CD4+ T cells, specifically),[148] cells of the monocytes/macrophages lineage (monocytes, macrophages, microglial cells, dendritic cells),[149] as well as T cells precursors. However, the HIV-1 M-tropic strain (or R5 viruses) infects mainly macrophages and monocytes, thanks to its preferentiality to interact with the CCR5 co-receptor present in these cells. In contrast, the T-tropic strain (or X4 viruses) prevalently uses the CXCR4 co-receptor to infect helper T cells. HIV-1 virions has spherical, membrane-enveloped, pleomorphic, and about 120 nm in diameter (Figure 1.11).[150] The envelope (the outer membrane of the virus) contains two glycoproteins: gp120 and gp41, and it is interiorly coated by the matrix proteins (MA). Moreover, each virion encloses a cone-shaped capsid, which contains two identical copies of its positive-sense single-stranded RNA genome (Figure 1.11). The capsid also contains the enzymes reverse transcriptase (RT), integrase (IN) and protease (PR), some other proteins (Vif, Vpr, Nef, nucleocapsid NCp7), and the major core protein p24 (Figure 1.11).

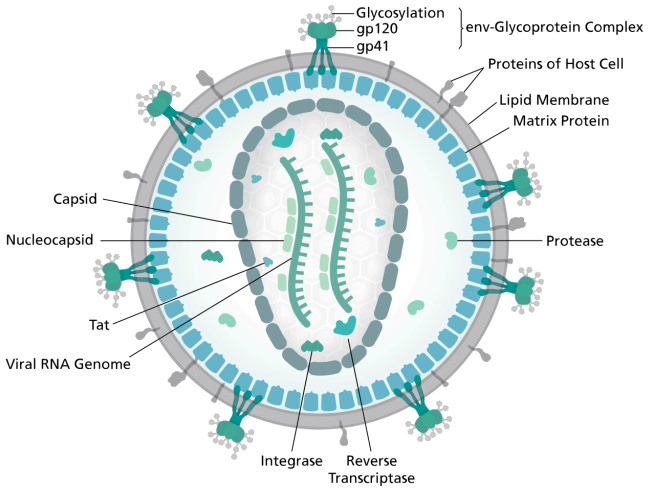


Figure 1.11: **Representation of the structure of a mature HIV-1 virion.** The localization of the RNA genome and of the most important viral proteins is detailed.

The HIV-1 replication cycle consists of seven main steps: binding and entry, uncoating, reverse transcription, provirus integration, replication, virus protein synthesis, and assembly and budding (Figure 1.12).[151]

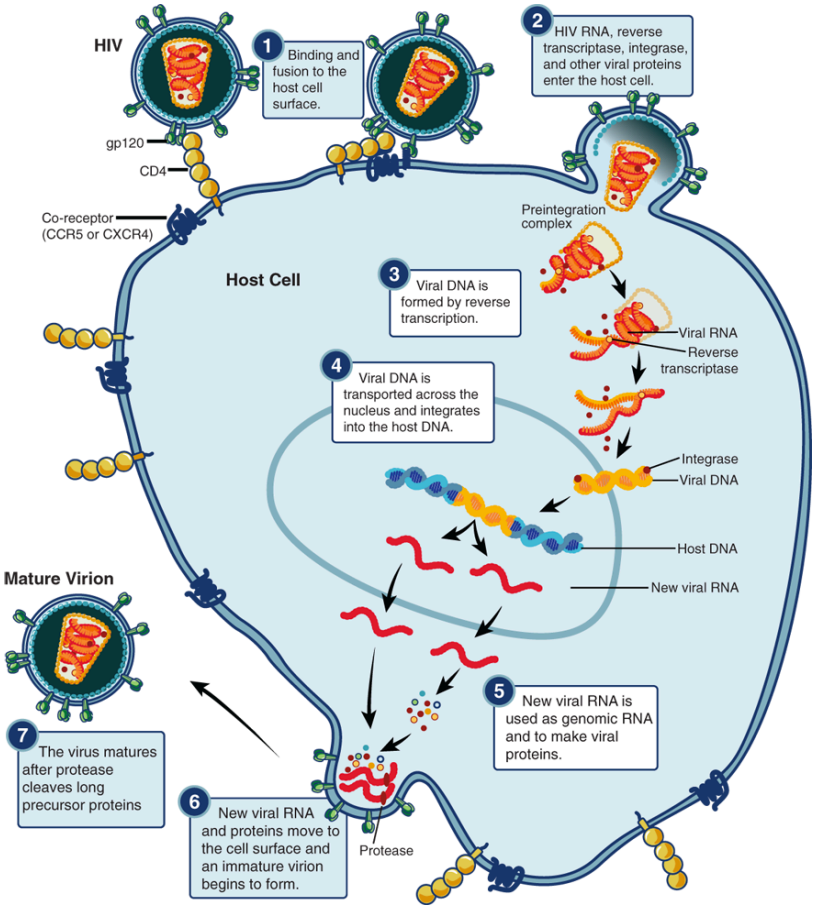


Figure 1.12: **Schematic representation of the seven step of the HIV-1 replicative cycle.** Figure from [152].

The envelope glycoproteins gp120 and gp41 first bind to the host cell binding the CD4 receptor.[153] This interaction causes conformational changes in gp120 that exposes its second binding site for the cell co-receptor. Depending on the virus tropism, the interaction of gp120 with CCR5 or CXCR4 co-receptors induces initiation of the membrane fusion process. After the complete membrane fusion, the virus core is injected into the host cytoplasm. Then, viral uncoating takes place, freeing the two viral RNA genome copies. The HIV-1 genome is relatively small, it is made of about 9000 nucleotides (9.7 kb). Generally, the RNA genome consists of nine genes (gag, pol ,env, tat, rev, nef, vif, vpu, vpr), encoding 16 proteins, and several structural landmarks.[147] In the 5'-3' direction, the RNA genome contains the following structures: R (repeated sequence), U5 (unique sequence located at the 5' end), primer binding site (pbs), the nine genes, a poly-purine rich tract (ppt), U3 (unique sequence located at the 3' end), and R (Figure 1.13).

The two RNA genome copies are reverse transcribed into a linear double-stranded DNA molecule (cDNA) by the viral reverse transcriptase (RT) enzyme (Figure 1.13).[154] The reverse transcription process is one of the key steps in retroviral cycle. Given its importance also for the aim of this dissertation, it will be deeper discussed.

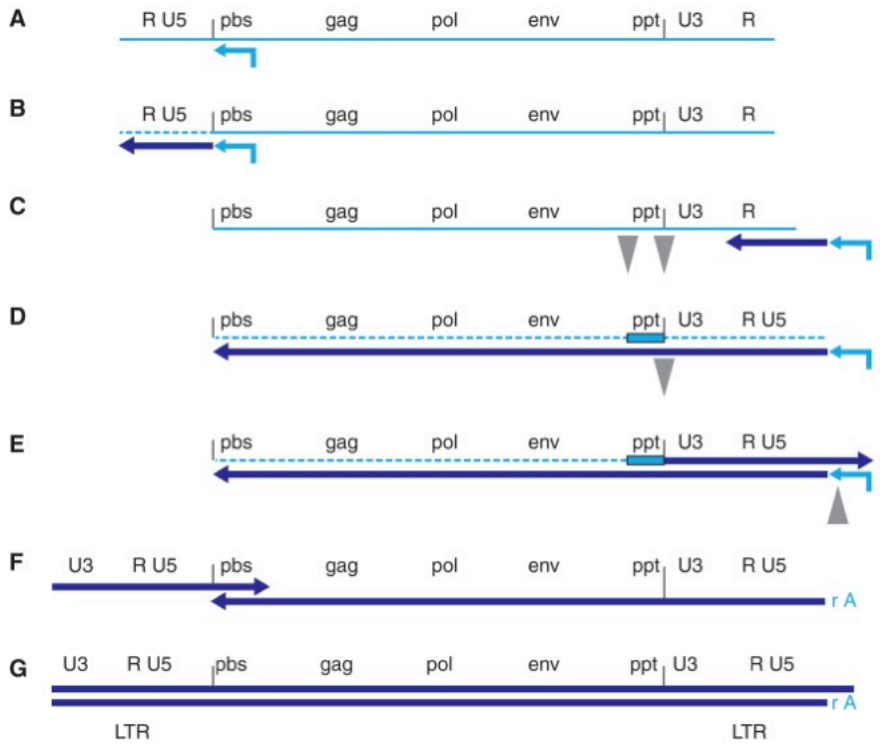


Figure 1.13: Schematic representation of the HIV-1 reverse transcription. The conversion of the single-stranded RNA genome of a retrovirus into double-stranded DNA is a multistep process. (**A**) The tRNA primer base pairs to the pbs site of the RNA genome of the retrovirus (light blue). (**B**) RT generates the minus-strand DNA (dark blue), and the RNase H activity of RT degrades the RNA template (dashed line). (**C**) Minus-strand transfer occurs between the R sequences at both ends of the genome, allowing minus-strand DNA synthesis to continue (**D**), accompanied by RNA degradation. The ppt site serves as the primer for the synthesis of plus-strand DNA (**E**). Plus-strand synthesis continues until the first 18 nucleotides of the tRNA are copied, allowing RNase H cleavage to remove the tRNA primer. The second (plus-strand) transfer occurs (**F**). Extension of the plus and minus strands leads to the synthesis of the complete double-stranded linear viral DNA (**G**). Figure from [154].

The RT enzyme, like other DNA polymerases, needs both a primer and a template. The template is the ssRNA viral genome, while the primer for the synthesis of the first DNA strand (the minus strand) is a host tRNA. Different retroviruses use different host tRNAs; HIV-1 uses the Lys3 one. The 3' end of the tRNA recognizes and binds its complementary sequence (the pbs) near the 5′ end of the viral RNA. Then, the RT allows the DNA synthesis elongating the minus-strand DNA from the pbs site to the 5' end. At this point, the RNA template is substrate for RNase H activity, which degrades the 5′ end of the viral RNA. Then, the R regions act as a bridge that allows the newly synthesized minus-strand DNA to be transferred to the 3′ end of the viral RNA. After this transfer, minus-strand synthesis can continue along the length of the genome. As DNA synthesis proceeds, so does RNase H degradation. However, the ppt of the RNA genome is resistant to RNase H cleavage, so it is maintained and serves as primer for initiation of the second (or plus) strand DNA. The plus DNA synthesis continues and the RNase H cleavage removes the tRNA primer. Removal of the tRNA primer sets the stage for the second strand transfer. Then, the two pbs DNA complementary sequences anneal and the DNA synthesis of both the two strands complete. The reverse transcription process creates a DNA product that is longer than the RNA

genome. In fact, each end of the viral DNA has two identical sequence called the long terminal repeats (LTRs), each one divided in U3, R, and U5 regions.[154]

The newly synthesized proviral dsDNA is then transported across the nucleus membrane as pre-integration complex (PIC). In the nucleus the integrase enzyme catalyzes the integration of the viral DNA into the chromosomes of the host cell.[151] In most cases, integration leads to productive infection, in which viral transcription and gene expression take place. However, in rare instances, little-to-no viral transcripts are detected, due to the fact that the virus undergoes latency instead of productive infection.

When active transcription prevails, the 5' LTR functions as promoters. The U3 region of the LTR is particularly involved in transcription initiation and modulation. U3 can be subdivided into three functional sections: an upstream modulatory element containing important binding sites for cellular transcription factors (e.g., the activator protein 1, AP-1; the nuclear factor of activated T cells, NFAT; the CCAT enhancer binding protein, C/EBP), an enhancer region with two adjacent binding sites for the inducible transcriptional activator nuclear factor kappa B (NF-κB), and a basal core promoter, which contains a specialized initiator element (Inr), a canonical TATA element, and three tandem Sp1 (specificity protein 1) binding sites (Figure 1.14).[155] Genetic variability in proteins binding sites have been observed between different HIV-1 subtypes.[156,157] However, among all of them, the NF-kB and Sp1 binding sites are remarkably conserved, confirming that they are essential for the regulation of viral transcription.

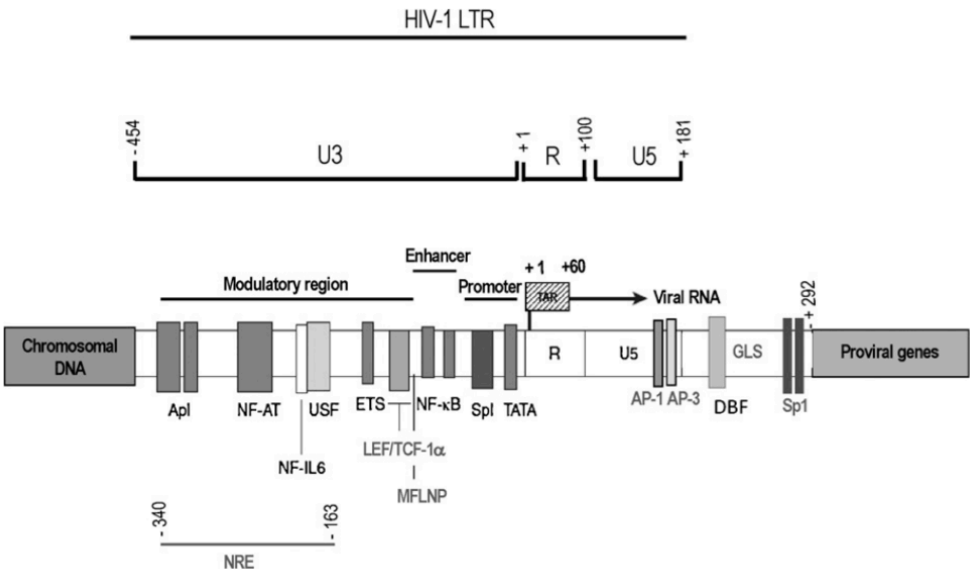


Figure 1.14: HIV-1 LTR organization. The modulatory, enhancer, and promoter elements of the U3 region are underlined. The transcription factors binding sites are illustrated as boxes. Figure from [155].

Viral transcription is operated by the host RNA polymerase II (RNAPII) which is recruited in the PIC based on specific signals. Normally, the TATAA element retrieves the general transcription factor TFIID, a large multiprotein complex consisting of the TATA-box binding protein (TBP) and a number of TBP-associated factors (TAFs). These proteins contributes to PIC formation and to the recruitment of the RNAPII, which initiated transcription. Initially, RNAPII produces only short

mRNA transcripts, due to elongation pausing at the 5' portion of the RNAs. This low level elongation still allows transcription of the two regulatory proteins Tat and Rev, and the TAR element, with the production of the TAR RNA stem-loop, a RNA regulatory element. At this point, the viral trans-activator protein Tat and the P-TEFb (human positive transcription elongation factor b) complex, which in turn is formed by cyclin T1 (CycT1) and cyclin-dependent kinase 9 (CDK9), are recruited to the elongation complex via binding interaction with the TAR RNA. This induces the phosphorylation of the carboxy-terminal domain (CTD) of the RNAPII, which is required to obtain full-length transcripts and normal levels of expression.[158,159]

To produce the full range of mRNAs needed to encode the viral proteins, HIV-1 primary transcripts undergo extensive and complex alternative splicing in the nucleus. HIV-1 can produce more than 40 differently spliced (including unspliced, incompletely spliced, and completely spliced) mRNAs for the translation of viral proteins.[160] Since unspliced and incompletely spliced transcripts from cellular genes are typically degraded in the nucleus, HIV-1 circumvents this problem through the Rev function. In fact, this viral protein allows mRNAs transport from the nucleus to the cytoplasm.[160] Finally, the translation step can take place.

All the 16 HIV-1 proteins can be divided into three groups: the fundamental regulatory proteins (Tat and Rev; previously mentioned), the essential structural proteins (Gag, matrix (MA or p17), capsid (CA or p24), nucleocapsid (NCp7), p6, protease, integrase, reverse transcriptase, gp41, and gp120), and the accessory proteins (Nef, Vpr, Vpu, and Vif).[161]

For the intention of this work, the viral nucleocapsid protein (NCp7) requires further insights.

Ncp7 is a small (55 aminoacid residues) basic protein originated from the protease cleavage of the Gag polyprotein. The basic N-terminal and C-terminal domains of the protein are not structured and flexible while the central domain contains two highly conserved CCHC type zinc fingers (Figure 1.15).[162,163]

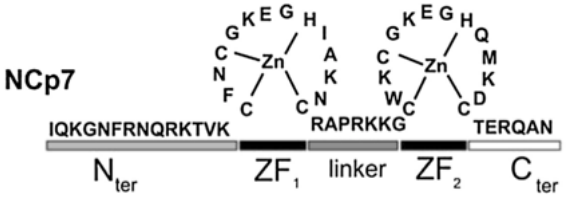


Figure 1.15: **Aminoacid sequence of the HIV-1 nucleocapsid protein (NCp7).** The N-terminal (Nter), C-terminal (Cter), and the two zinc-finger domains (ZF1 and 2) are represented.

The highly structured zinc fingers are responsible for sequence specific binding to the nucleic acids (particularly G and UG/TG runs). NCp7 is essentially involved in almost all the steps of the HIV-1 replication cycle. First of all, multiple copies (1500 – 2000) of the protein are associated with the RNA dimer genome inside the mature virions, ensuring protection from RNase activity once the genome is released into the cytoplasm of host cells. Moreover, NCp7 contributes to the correct progression of the reverse transcription process, thanks both to its chaperone function (it favours the secondary structure to adopt the thermodynamically most stable conformation), and to its annealing activity.[164,165] It is generally accepted that the chaperone activity of the protein

facilitates the RT enzyme in dsDNA synthesis by unfolding complex secondary structures formed by both RNA and DNA templates. In fact secondary structures, such as hairpins, may make difficult the elongation of nucleic acids by polymerase enzymes due to steric impediments that could cause pausing sites and incomplete synthesis products. Given the assistance role of the nucleocapsid protein in key steps of the HIV-1 viral cycle, NCp7 has been considered as a promising target for antiviral therapy. Some approaches are directed to zinc ejection to obtain disruption of zinc finger domains responsible for nucleic acids binding and chaperone activity; other approaches tend to target and stabilize nucleic acids bound by NCp7, such as RNA binders targeting TAR/cTAR.[166–168]

The assembly and budding are the last steps required to complete the HIV-1 replicative cycle. Briefly, the two replicated viral RNA strands associate together with precursor proteins, while core proteins assemble over them forming the virus capsid. This immature particle migrates toward the cell surface. During the budding step, the virus exit the host cell acquiring the host-derived membrane to create the viral envelope, which becomes enriched with phospholipids and cholesterol.[169] Both during and shortly after the release of the virion, the protease cleaves the Gag and gag-pol polyproteins to complete virion maturation.

1.2.3 A deepening on HIV-1 latency

Virus latency is the ability of a pathogenic virus to lie dormant (latent) within a cell.[170] In fact, for certain viruses, after initial infection, proliferation of virus particles ceases, but the viral genome is not fully eradicated and persists within the host indefinitely (as long as the cell lives). From this state the virus can reactivate and begin to produce large amounts of viral progeny without the host being infected by new virus.

Like other retroviruses, HIV-1 establishes a viral latency that is named proviral latency. In this type of post-integration latency that occurs when a provirus fails to effectively express its genome and is reversibly silenced after integration into the host cell genome, the virus persists simply as information (in the form of about 10 kb of integrated HIV-1 DNA). This stably integrated provirus is replication-competent but transcriptionally silent, due to a repressed state of the viral LTR promoter.[159,170,171]

In 1995, the presence of integrated HIV-1 DNA in highly purified populations of resting CD4+ T cells was definitively shown.[172] Many studies reported that all HIV-1 infected patients contain latently infected cells (also known as “latent reservoirs”), mostly memory CD4+ T cells, dendritic cells and cells of the monocyte/macrophage lineage.[173] All of these cells have a long half-life *in vivo* ($t_{1/2}$ = 44 months), thus allowing the latent virus to persist within infected individuals for decades. Moreover, HIV-1 in proviral latency is nearly impossible to target with antiretroviral drugs and it is hidden from immune response because of latently infected cells are indistinguishable from uninfected cells.[170]

The establishment of HIV-1 latency is a complex process and less is known about this.[174] It likely results from the convergence of multiple mechanisms, active at both the transcriptional and the post-transcriptional levels.[159] These mechanisms include: transcriptional interference, insufficient

levels of transcriptional activators (absence of nuclear forms of key host transcription factors; e.g., NFkB and NFAT), the presence of transcriptional repressors, epigenetics (epigenetic changes inhibit HIV-1 gene expression), nucleosome positioning, insufficient Tat activity (or absence of Tat and associated host factors that promote efficient transcriptional elongation), blocks of mRNA splicing or nuclear export, cellular microRNAs (miRNAs), homeostatic proliferation of latently infected cells.[160,175,176] The relative importance of these mechanisms is probably dependent on the physiological state of the cell undergoing infection.

1.2.4 Currently approved treatments and the hope of eradication

To date, important progress has been achieved in preventing new HIV infections and several antiretroviral compounds against HIV-1 have been approved by FDA. These drugs belong to several classes, based on their mechanism of action and viral target: nucleoside reverse transcriptase inhibitors (NRTIs; e.g., azidothymidine, abacavir), non-nucleoside reverse transcriptase inhibitors (NNRTIs; e.g., nevirapine, efavirenz), protease inhibitors (PI; e.g., ritonavir), entry inhibitors (e.g., enfuvirtide), integrase inhibitors (e.g., raltegravir), CCR5 antagonists (e.g., maraviroc). Currently, the indicated treatment consists of a cocktail containing at least three drugs against at minimum two different viral targets. HAART is able to reduce patient viremia to undetectable levels, that is why it has radically changed the course of the HIV-1/AIDS pandemic, transformed it from a deadly disease into a chronic lifelong condition, and saved millions of lives worldwide.[177] However, despite advantages in the antiretroviral therapy, some problems remain. These consist in elevated toxicity (e.g., hepatotoxicity, renal failure, pancreatitis, nausea, insomnia), low adherence to therapy due to lifelong treatments, and the emergence of resistant strains. It has been calculated that about the 10% of adults with HIV-1 developed first-line ART resistance. The development of drug resistance arises from the extreme variability of the virus, which is a consequence of the high evolution rate of the virus due to lack of RT proofreading activity, the high rate of viral replication, and the occurrence of recombination processes.[178,179] In addition, although ART is very effective in blocking HIV-1 spread within the body, it does not eradicate the virus because of the presence of latent reservoirs. Indeed, viral loads readily rebound when treatment is interrupted making life-long treatment necessary which is often associated with drug toxicities and a risk of viral resistance. Therefore, there is an urgent need of therapeutic strategies with an innovative mechanism of action, possibly against highly conserved viral sites, and able to clear the virus from the infected host in order to reach the complete recovery.

Currently, two types of HIV cures are in development: a “sterilizing cure” and a “functional cure.”[180–182] A sterilizing cure refers to the complete elimination of replication-competent proviruses in the body, while a functional cure refers to the long-term control of HIV replication without treatment.[183] Based on these concepts, significant progress has been made in different areas. Effective sterilizing strategies include ‘shock and kill’ and gene editing. The “shock and kill” or “kick and kill” approach uses small molecules called latency-reversing agents (LRAs) to reactivate latent reservoirs, thus making possible for ART and immune response to clear the virus

from the body. Until now, various classes of LRAs have been identified and tested, such as protein kinase C (PKC) activators (e.g., phorbol myristate acetate, PMA; prostratin; ingenol-B), histone deacetylase (HDAC) inhibitors (e.g., valproic acid, vorinostat, SAHA), DNA methyltransferase (DNMT) inhibitors, bromodomain inhibitors, cytokines. Each LRA has a specific efficiency in reactivating the virus from latency.[181] The "shock and kill" approach faces many identified challenges. First of all, the level of LRAs-mediated reactivation of HIV-1 gene expression appears to be suboptimal.[182] Moreover, considering the heterogeneity of the cell reservoirs and the multiplicity of molecular mechanisms which underlie latency, different LRAs can reflect a high diversity in term of the pattern of response.[184] It has been concluded that the mechanisms of latency probably differ from one patient to the other, and even from one cell to another in a single patient, so it seems that HIV-1 latency reactivation by small molecules may be stochastic in nature rather than deterministic.[185]

The case of the "Berlin patient" (Timothy Ray Brown) has raised hope for the possibility of developing an HIV-1 cure. He became HIV-1-negative after receiving hematopoietic stem cell transplantation from a CCR5Δ32 homozygous donor.[186] Individuals who are homozygous for this allele are naturally resistant to HIV-1 infection. However, this approach worked for the "Berlin patient" but not for the "Boston patients", who have experienced viral rebound.[181] Moreover, since CCR5Δ32 is a rare mutation (about 1% of the population), bone marrow transplantation is a risky procedure and not tolerated by most patients, this is not a scalable treatment. Nevertheless, efforts aim to generate this resistant phenotype by disruption or suppression of CCR5 receptor in CD4+ T cells or hematopoietic stem cells derived from bone marrow or peripheral blood by gene editing (for example through zinc finger nuclease, ZFN; transcription activator-like effector nucleases, TALENS; and more recently clustered regularly interspaced short palindromic repeats-associated protein nuclease-9, CRISPR-Cas9).[181] The applicability of this procedure remains doubtful. In fact, aside from the side effects, off-targets, costs of autologous transplantation, another major limitation is potential viral escape from the CCR5-tropism to the X4-tropic virus. Gene editing techniques have been used also to target the provirus in latently infected cells. Generally, ZFN, TALENs, and CRISPR-Cas9 have been employed in T cell lines, primary $CD4^+$ T cells, and patient immune cell engrafts in humanized mice[187] to eradicate proviral HIV-1 DNA. However, neither a complete silencing nor a total eradication was reached so far. Moreover, ethical concerns remain about the application of the CRISPR-Cas9 technique in humans, without fully understanding the possible consequences.

In view of the major hurdles impeding an effective sterilizing cure, more attention is now being placed on functional strategies. Conversely to the "shock and kill" strategy, this "block and lock" approach aimed at fully repressing HIV-1 transcription. Inhibition of HIV-1 transcription could establish a state of deep latency that is refractory to viral reactivation, thereby suppressing the residual viremia that arises from reactivation of latently infected cells or from ongoing viral replication. Some examples are the use of gene-based therapies (RNA interference, RNAi) and the Tat inhibitor didehydro-cortistatin A (dCA).[181,188] RNAi, short interfering RNA (siRNA) and short hairpin RNA (shRNA), can be designed to silence the expression of viral or host mRNA

targets that are necessary for HIV-1 transcription and replication.[188] Indeed, the Tat inhibitor dCA, selectively inhibiting Tat transactivation of the HIV-1 promoter by binding to the TAR-binding domain of Tat, induces a state of latency with a diminished capacity to be reactivated not only *in vitro* but also *ex vivo*. In a primary cell model of HIV-1 latency, dCA-induced inactivation of viral transcription persists even after drug removal, suggesting that the HIV-1 promoter is epigenetically repressed.[181] The advancement of developing a functional HIV cure seems more rapid and effective than that of a sterilizing HIV cure.[180]

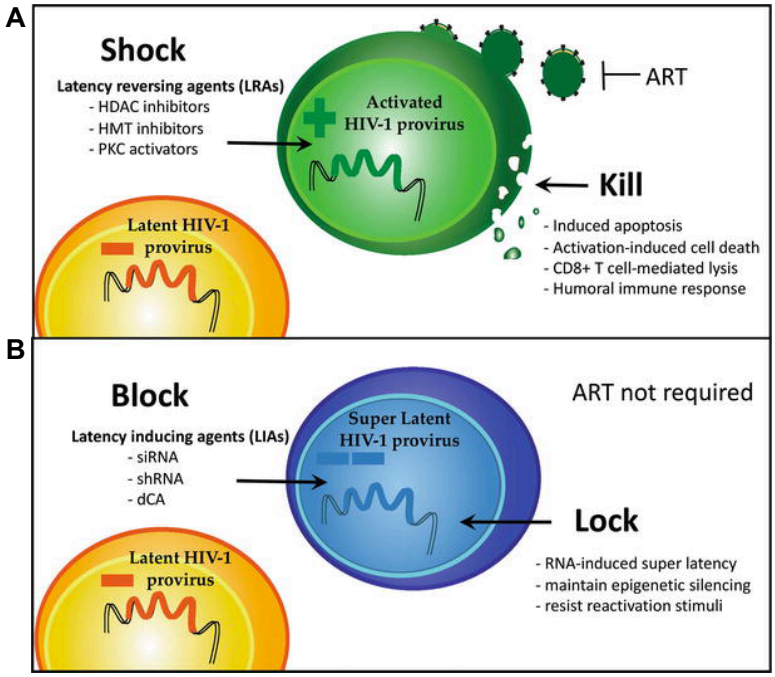


Figure 1.16: **Sterilizing and functional cure in comparison.** The (A) "shock and kill" strategy versus the (B) "block and lock" approach.

1.2.5 G4 structures in HIV-1

The presence of G4 structures has been demonstrated both at the RNA and DNA levels of the HIV-1 genome, together with the evidence of their regulatory role in the viral life cycle.[12,42]

In 2013, our group identified three G4 forming sequences in the HIV-1 nef coding region. At least two of them presented high level of conservation among most circulating HIV-1 strains. The stabilization of the nef G4 folding by cations and G4 binding compounds impaired Nef expression, significantly suppressing HIV-1 infectivity.[189]

The knowledge that the HIV-1 LTR U3 enhancer and promoter regions (positions−105/−48 with respect to the TSS; binding sites for NF-κB and Sp1; see Figure 1.17) was very rich in G bases (50% G and 70% GC) prompted our group to evaluate the presence of G4 folding sequences. Twelve putative sequences were found; four of these, namely LTR-I, LTR-II, LTR-III, and LTR-IV exhibited the highest G-scores and high conservation levels among circulating viral strains. LTR-II and LTR-III demonstrated spontaneous G4 folding in physiological environment, while LTR-IV formation is induced by the presence of stabilizing ligands.[140]

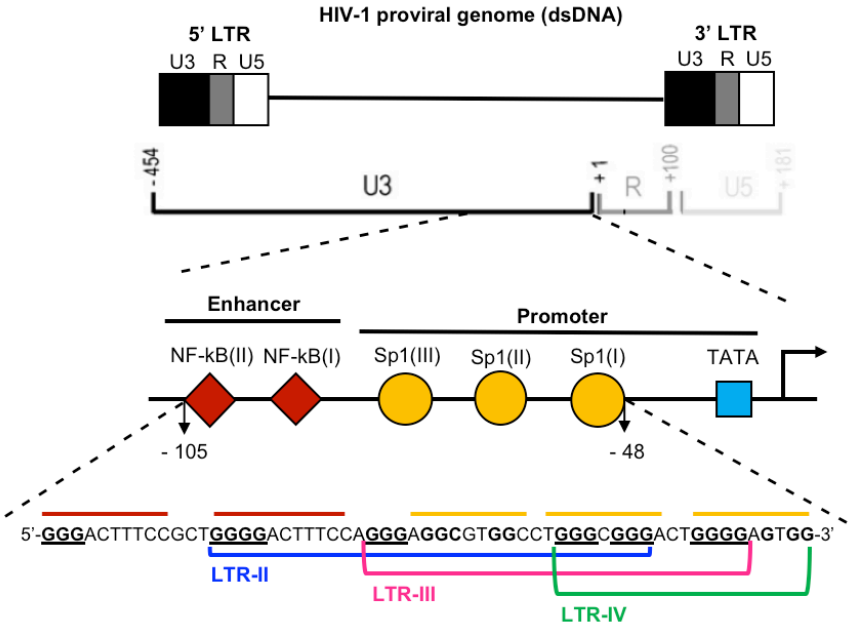


Figure 1.17: **G4 structures in the HIV-1 proviral genome.** From top to bottom: schematic organization of the HIV-1 dsDNA genome; magnification of the LTR region, composed of U3, R, and U5; enlargement of the U3 enhancer and promoter areas, with the indicated binding sites of the two cellular transcription factors (NF-kB and Sp1), the specified -105/-48 sequence, and the highlighted three mutually exclusive G4s (LTR-II, LTR-III, LTR-IV).

The research highlighted also a tightly regulated control of transcription based on G4 folding/unfolding. First of all, single-base mutations (i.e., m4 and m5), which totally prevent G4 formation, produced an increase of the LTR promoter activity, thus proving that G4s act as repressor elements in the transcriptional activation of HIV-1.[140] Other studies reported a fine tuning of transcription regulation by interacting compounds (e.g., BRACO-19 and c-exNDI)[47,140] and cellular proteins (e.g., nucleolin and hnRNP A2/B1). In particular, the compounds and nucleolin induced the stabilization of HIV LTR G4s, suppressing viral transcription.[61] On the contrary, hnRNP A2/B1 was able to efficiently unfold the LTR G4s, acting as a HIV-1 transcriptional activator (Figure 1.18).[67]

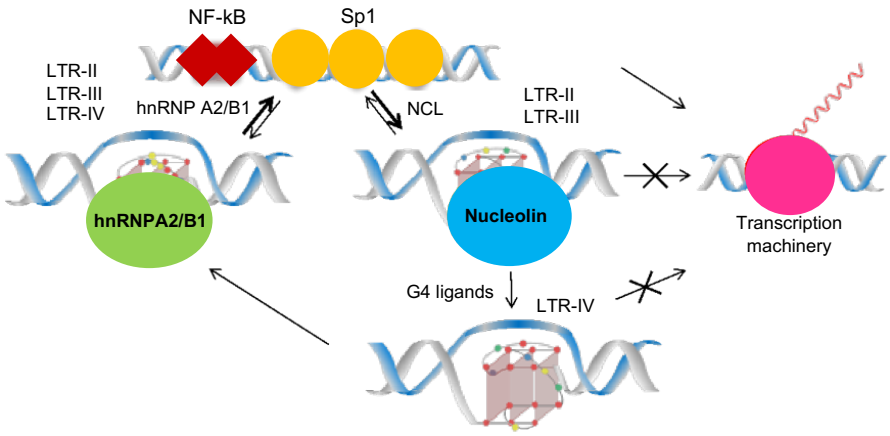


Figure 1.18: **Ligands-mediated HIV-1 LTR G4s modulation and the consequent regulation of transcription.** BRACO-19 and c-exNDIs compounds interact with LTR-II and LTR-III G4s, and induce the

folding of the LTR-IV G4. The interacting compounds and the cellular protein nucleolin stabilize the LTR G4s, thus inhibiting viral transcription. On the contrary, the hnRNP A2/B1 binding promote the unfolding of all the G4s, enhancing the transcription machinery to proceed.

Because the G4 forming LTR DNA region derives from an identical RNA sequence present at the 3′ end of the RNA viral genome (Figure 1.19), the presence of G4 RNA has been tested. Analogously to the DNA region, the RNA counterpart was divided into three regions (namely U3-II, U3-III, and U3-IV) containing four GGG tracts, which are the minimum requirement for the formation of a three-tetrad G4. Biophysical studies proved the G4 folding capabilities of the selected RNA oligonucleotides, that displayed a parallel-like conformation.[190] Moreover, it has been proven that the U3 RNAs exhibited significantly higher stability than that of their LTR DNA counterparts. Incubation of the three U3 RNAs with BRACO-19 greatly stabilized their G-quadruplex conformation. Finally, HIV-1 RT stop assays proved polymerase stalling due to the formation of the U3-III and U3-IV RNA G4s, and when the RNA G4 folding sequence was incubated with increasing concentrations of BRACO-19, HIV-1 RT was inhibited to a much higher extent.[190]

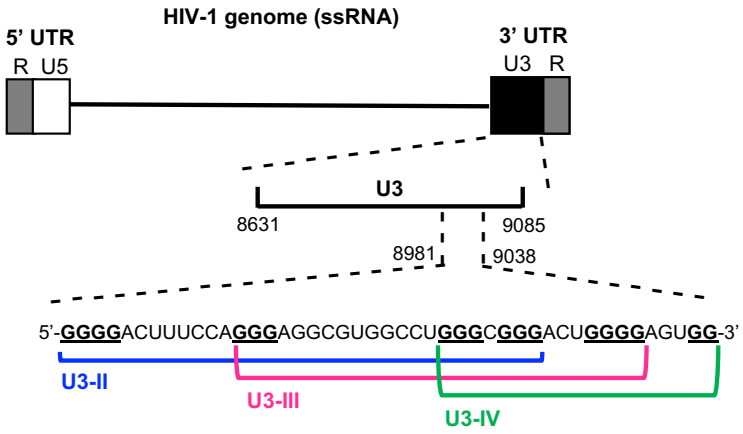


Figure 1.19: **G4 structures in the HIV-1 RNA genome.** From top to bottom: schematic organization of the HIV-1 ssRNA genome; enlargement of the U3 region, with the specification of the sequence between 8981 and 9038 nucleotides, and the highlighted three RNA G4s (U3-II, U3-III, U3-IV).

1.3 The herpes simplex virus type 1

1.2.1 The HSV-1 worldwide: epidemiology

HSV-1 is a highly contagious infection, which is common and endemic throughout the world. Herpetic infections were first documented in ancient Greece, where Greek scholars adopted the word “herpes” (which means “to crawl”) to describe them.

The WHO reported that “the global burden of HSV-1 infection is huge”. The organization estimated that about two-third (67%) of the world population under the age of 50 had HSV-1 infection.[191] The prevalence was highest in Africa, south-east Asia, and western Pacific.

Most HSV-1 infections are acquired during childhood. The transmission takes place through oral-to-skin/oral-to-oral contacts. The greatest risk of transmission is when the skin present active

sores, but it can occur also when the surfaces appear normal and there are no symptoms. In fact, oral infections are mostly asymptomatic. When present, symptoms include painful blisters and ulcers (commonly referred to as “cold sores”) in the skin or mucous membranes of the mouth, lips, nose, and also eyes. Ulcers formation can be preceded by unpleasant sensations of tingling, itching, and burning. HSV-1 can also cause severe complications, such as encephalitis and keratitis especially in immunocompromised patients, and neurologic disability or death in infected infants (“neonatal herpes”).

The HSV-1 is a recurrent and lifelong infection because the virus can establish a latent infection in ganglia. Due to latency, there is still no cure. In the majority of the neurons, the viral genome remains in an episomal state for the entire life of the individual. However, it can reactivate upon several stimuli such as: UV-light exposition, stress, and hormonal imbalance. Studies reported the presence of latent HSV-1 in the brains of elderly people[192] and it has been demonstrated a correlation between HSV-1 and risk of Alzheimer’s disease.[193]

1.2.2 HSV-1 general features

HSV-1 or HHV-1 (human herpesvirus 1) is a member of the *Herpesviridae* family, *Alphaherpesvirinae* subfamily, and *Simplexvirus* genus.[194] The *Simplexvirus* genus include both HSV-1 and HSV-2 viruses, that cause orolabial and genital herpes, respectively.

HIV-1 virions are composed of four main architectural features: envelope, tegument, capsid, and the viral genome. The HSV-1 virion can be represented as a spherical particle with an average diameter of about 200 nm. The external lipid bilayer envelope contain at least 11 viral glycoproteins that protrude from the virus surface and are necessary for viral entry. The internal tegument protect an icosahedral capsid (about 125 nm diameter), which in turn encompasses a linear, double-stranded DNA genome of 152 kb length (Figure 1.20).[195]

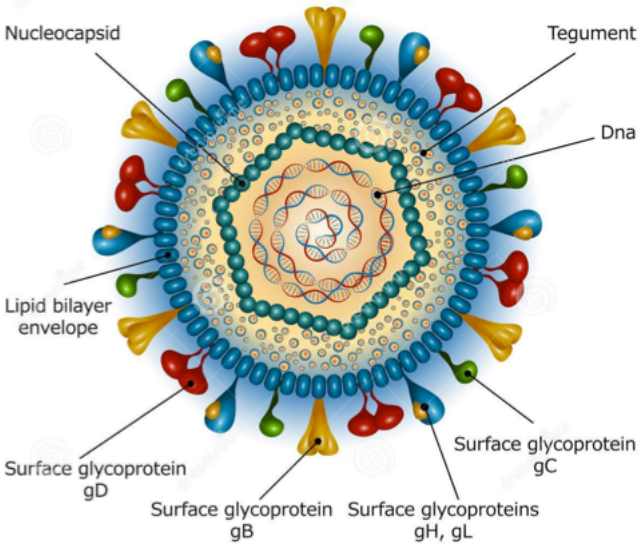


Figure 1.20: **Representation of the structure of an HSV-1 virion.** The dsDNA genome, the icosahedral capsid, the tegument and the envelope with its viral glycoproteins are represented.

The HSV-1 replication cycle consists of six main steps: binding, entry, capsid transport to the nucleus, viral replication and gene expression, capsid assembly, and virion budding (Figure 1.21).[196]

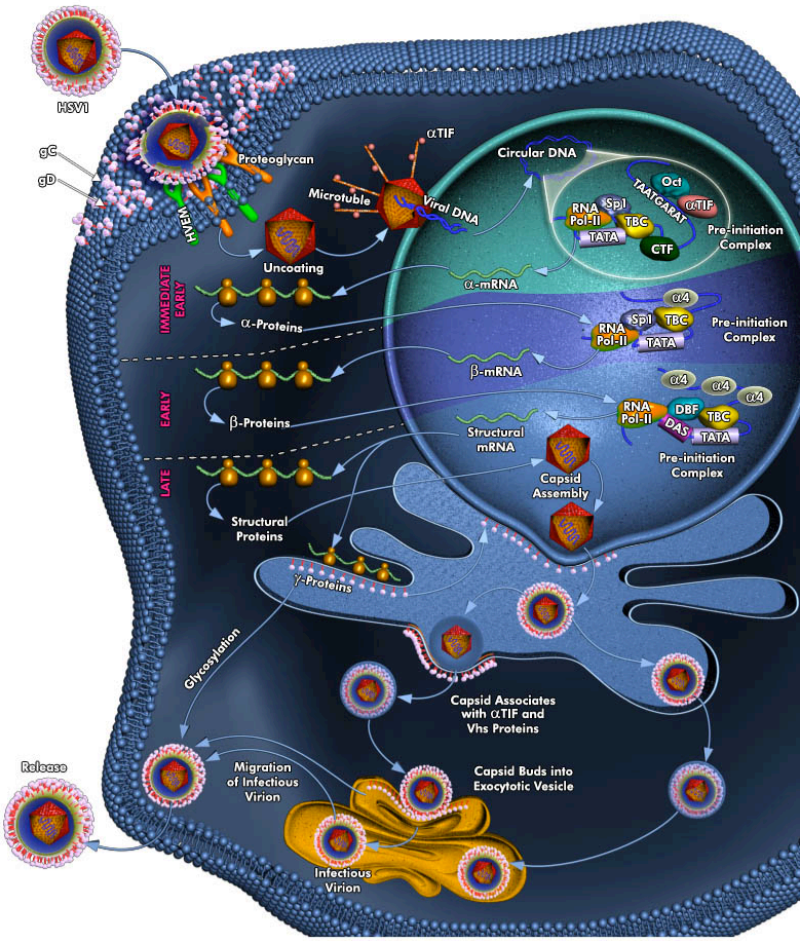


Figure 1.21: **The main steps of the HSV-1 replicative cycle.**

The initial binding of virus to cells is mediated by the association of viral glycoproteins gB plus gC with heparan sulfate proteoglycans (HSPGs) present on the cell surface.[197] These interactions are followed by gD binding to its receptors, such as nectin-1, herpes virus entry mediator (HVEM), and 3-O-sulfated HS (3-OS HS).[197] The gD binding to one of its receptors triggers a conformational change in its protein structure that drives recruitment of a fusion complex (including gB, gH, and gL).[197] After the fusion of membranes, the viral capsid and tegument proteins are internalized in the cytoplasm. Once in the cytoplasm, the viral capsid is transported near the nuclear membrane either by simple diffusion or by cytoskeletal structures, such as microtubules. It has been demonstrated that from 30 minutes to 1 h post-infection, the viral DNA passes through the nuclear pores into the nucleus.[196] Here, the viral DNA adopts its circularized conformation (Figure 1.22).[198] The HSV-1 genome is a dsDNA of 152 kb length that consists of two unique regions, unique long (U_L) and unique short (U_S), flanked by inverted repetitions (TR_L, IR_L, IR_S, and TR_S). The inverted repeat sequences flanking U_L are named ab and b'a', whereas those flanking U_S are a'c' and ca (Figure 1.22).[198]

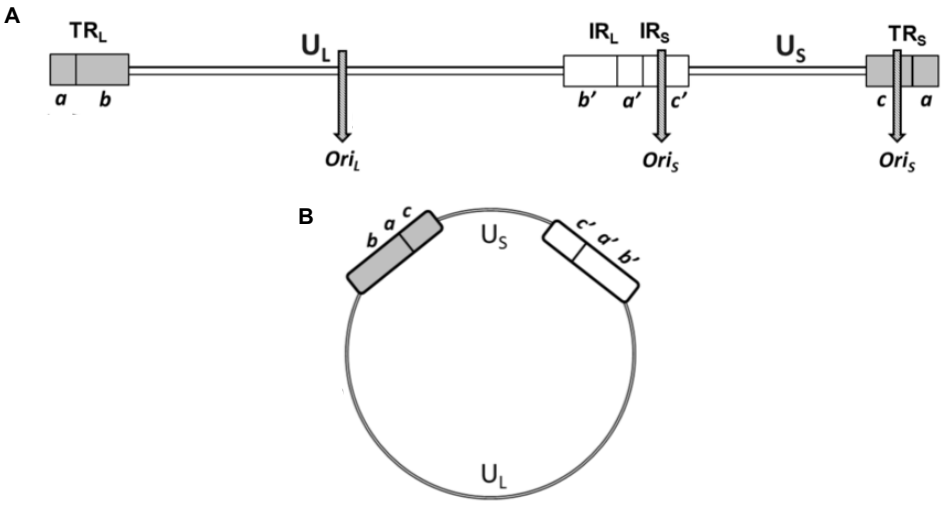


Figure 1.22: **Organization of the HSV-1 genome.** (**A**) The positions of the a, b, and c repeats within the terminal and internal repeats and the positions of the DNA replication origin (Ori_L and Ori_S) are indicated. (**B**) The circularized HSV-1 genome.

A peculiarity of the HSV-1 genome is its high GC content of about 68%, with peak of 87% in simple sequence repeats.[199]

The viral genome contains 84 open reading frames (ORFs) for the expression of the viral proteins. HSV-1 lytic infection involves the temporally regulated expression of three classes of viral genes: immediate early genes (α or IE), early genes (β or E), and late genes (γ or L), which can be further subdivided into early-late (γ_1) or late (γ_2).[200,201]

IE-promoters contain an enhancer element that responds to cellular factors and a TAATGARAT element that is recognized by the virion protein (VP) VP16.[202] VP16, along with the cellular factors Oct1 and host cell factor (HCF), activates transcription of IE genes. They are transcribed between 2 and 4 hours post-infection (h.p.i.).[196] The expression of IE proteins, the infected-cell polypeptide ICP0, ICP4, ICP22, ICP27, and ICP47, does not require prior viral protein synthesis. These firstly transcribed proteins are mainly dedicated at blocking the cellular antiviral host response, such as interferon pathways, and are required for the optimal expression of both E and L gene products.[202] ICP4 is the principal regulatory protein of the virus; it plays a role as a transcriptional repressor or activator.[203,204] Binding to an array of cellular protein, ICP0 acts as a potent transactivator of viral and cellular promoters, providing for efficient viral gene expression and growth *in vitro* and *in vivo*. Moreover, it is required both in lytic viral growth and the efficient reactivation of the virus from the latent state.[205] ICP27 regulates the processing of viral and cellular mRNAs, and it modulates ICP4 and ICP0 activities. ICP27 is required for an optimal viral DNA synthesis; in fact, it increases the early-gene expression levels.[206] Furthermore, it contributes to an efficient late-gene expression, as well as ICP22. ICP22 mediates the formation of a novel phosphorylated form of RNA polymerase II, and it regulates the longevity as well as the splicing pattern of the ICP0 mRNA. ICP47 may help the virus escape immune surveillance on the basis of its ability to block the presentation of antigenic peptides to $CD8^+$ cells.[201]

The synthesis of the IE genes induces the synthesis of the β class of proteins, which reaches peak rates between 5 and 7 h.p.i. and is detectable as early as 3 h.p.i.[196] The β proteins include the enzymes that are required for replication of the viral genome; six of them play "core" replication roles at the replication fork: the single-strand DNA- binding protein ICP8 (or UL29), the two-subunit DNA polymerase (catalytic subunit Pol and processivity subunit UL42), and a three-subunit helicase/primase complex (UL5, UL8, and UL52). The remaining HSV-encoded protein is the origin-binding protein UL9. The viral DNA replication occurs as a rolling circle and begins shortly after the appearance of the β proteins and is detectable as early as 3 h.p.i. and continues up to 15 h.p.i.[198,207]

The temporal program of viral gene expression ends with the appearance of the γ or late proteins. The expression of L genes occurs between 12 and 17 h.p.i. Late proteins are proteins necessary for the DNA cleavage/packaging process and the capsid assembly.[201] At least seven genes (UL6, UL15, UL17, UL25, UL28, UL32, and UL33), together with the terminally redundant a regions of the HSV-1 genome, are essential to accomplish cleavage and packaging of the viral DNA successfully. L structural proteins such as VP5 (major capsid protein), VP19C, VP23, VP24, and VP26 are essential for viral capsid formation. A characteristic of herpesviruses is that capsids are assembled within the nucleus of infected cells.[208] After γ-mRNA translation in the cytoplasm, L capsid proteins enter the nucleus and initiate the capsid assembly. Viral genome packaging and tegument assembly also take place. The capsid maturates, and there is the fusion of the enveloped nuclear capsid with the outer nuclear membrane and release of the capsid into the cytoplasm. Once on the cell surface, viral capsids are endocytosed forming particles of infectious virions. Finally, the virions are secreted into the extracellular medium.

1.2.3 Currently approved treatments

Currently, FDA approved antiviral agents against HSV-1 infection include: acyclovir (ACV), valaciclovir (VCV), and famciclovir (FCV).[209]

ACV, a synthetic acyclic guanosine nucleoside analogue, has become a gold standard for the treatment of HSV-1 infections since its introduction in the 80s. After administration, ACV is selectively phosphorylated to a monophosphate derivative in infected cells by the virus-encoded thymidine kinase (TK) enzyme. The affinity of ACV for the viral TK is about 200 times greater than for the human one. Various cellular kinases convert the monophosphate derivative to di- and triphosphate derivatives. The ACV-triphosphate represent a competitive inhibitor of the viral DNA polymerase. In fact, the lack of the 3' hydroxyl group leads to the termination of the viral DNA replication.[195,210]

VCV and FCV present a better oral bioavailability than ACV and are also first-line drugs for HSV infections. Moreover, the off-label cidofovir and foscarnet (a pyrophosphate analogue) are occasionally used for the treatment of severe HSV infections that have become resistant to other drugs.[209] In fact, drug-resistant HSV-1 strains have emerged over the years. HSV can develop resistance mainly through mutations in the viral gene that encodes TK by the generation of TK-

deficient mutants or by the selection of mutants with a TK unable to phosphorylate the drug. Also, altered DNA polymerase has been detected in some clinical isolates.[195]

A large difference in drug resistance has been observed between immunocompetent and immunocompromised patients. Although low prevalence (0.1-0.6%) of ACV resistance has been reported for "normal" patients, it is more often evident in immunocompromised patients (3.5-10%).[211] This can be explained by the fact that those patients undergo long-term treatments. The resistant strains can cause severe diseases, such as pneumonia and encephalitis, so it has to be considered an important clinical problem for immunocompromised patients (e.g., patients with AIDS).[211]

The emergence of drug resistance has created a barrier for the treatment of HSV infections, especially in immunocompromised patients. Therefore, there is an urgent need to explore new effective agents to circumvent drug resistance to HSV.

1.2.4 G4 structures in HSV-1

As mentioned before (Paragraph 1.1.4.2), through computational analysis, PQSs were found at both the strands of dsDNA viruses. The *Herpesviridae* family was not only the one with the highest PQS content but also the one that displayed significantly more PQSs than expected in both genome strands.[80] In particular, 316 putative G4s have been identified for the HSV-1 (HHV-1) virus;[212] and they have been observed to be highly conserved among various HSV-1 strains.[125] G4 motifs were enriched in the repeat and regulatory regions. Interestingly, more G4-forming motifs have been found in the IE genes compared to the E and L classes.[212] The G4s high number, enrichment in peculiar region and high conservation suggest that these structures may represent a novel type of functional elements in herpesviruses.

Consistently with what has been observed, in 2015 our group has identified and selected nine regions that presented highly repeated and conserved PQSs.[127] Six of them were named gp054 (a-f) because they were mainly found in the gp054 gene, which encodes UL36, the essential viral tegument protein. Two repeats of gp054b were additionally identified in the leading and lagging strand of the gene encoding the ICP0 protein. Three more G4 motifs were found clustered at the terminal and internal repeats (both long and short) of the HSV-1 genome. They were called un1, un2, and un3, and each one was highly repeated (5-18 times). Biophysical techniques proved their ability to fold into physiological conditions, except un1. The stabilization of the selected HSV-1 G4s with the G4 ligand BRACO-19 inhibited the polymerase processing at the G4 forming sites. Moreover, BRACO-19 treatment of HSV-1 infected cells induced inhibition of virus production.[127] The possibility to target HSV-1 G4s in order to obtain an antiviral activity was further corroborated. Indeed, it has been observed that the treatment with the potent c-exNDI induced significant virus inhibition with low cytotoxicity.[213]

The implementation of the G4-specific antibody 1H6 allowed our group to visualize G4s during the HSV-1 life cycle. G4 formation and localization within the cells was found to be viral cycle dependent. In particular, G4 structures were massively present during viral DNA replication.[214] Therefore, the observed antiviral activities of BRACO-19 and c-exNDI may depend on the

combination of their high affinity for viral G4s and the massive prevalence of these structures during infection.

Very recently our group also demonstrated that all IE promoters in the genome of HSV-1 (and of the other *Alphaherpesviruses*) contained fully conserved PQSs. They were located both on the leading and lagging strand. Biophysical and biological analysis proved that all sequences can actually fold into G4s under physiological conditions and can be further stabilized by the G4 ligand BRACO-19, with subsequent impairment of viral IE gene transcription in cells.[215]

Overall, these results help shed light on the control of viral DNA replication and viral transcription, and indicate new viral targets to design specific drugs that impair the early steps of HSV-1 infection.

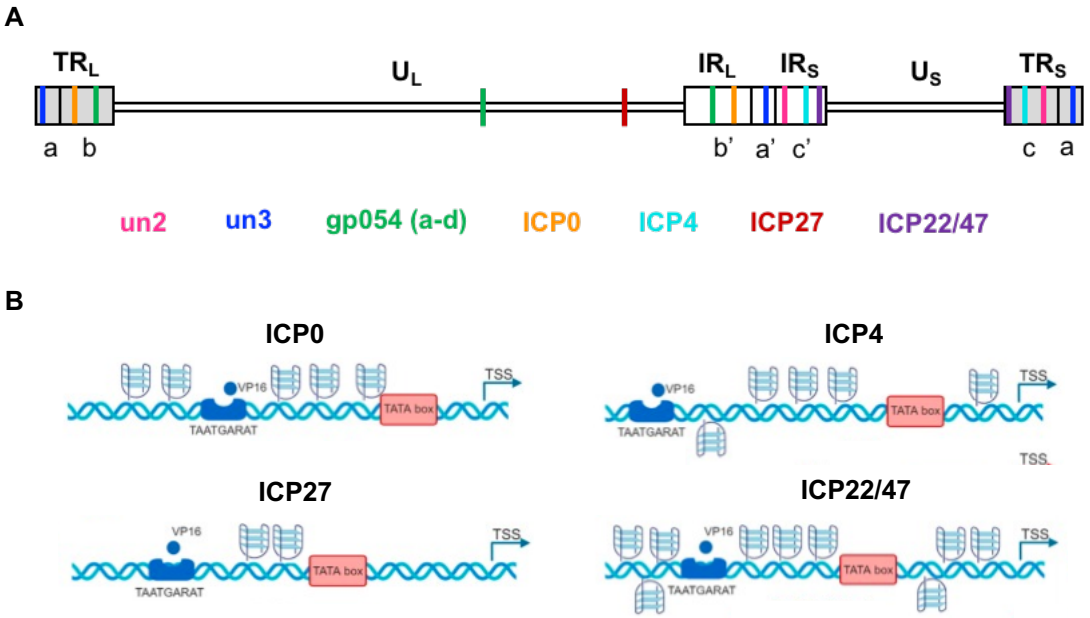


Figure 1.23: **Representation of the most studied HSV-1 G4s.** (**A**) Localization of the HSV-1 G4s within the viral genome. Each G4 is represented with its corresponding color (un2 in pink, un3 in dark blue, gp054 in green, ICP0 in orange, ICP4 in light blue, ICP27 in red, and ICP22/47 in purple). (**B**) G4s in the leading and lagging strands of the IE gene promoters.

2. Aim of the study

The G-quadruplex structures within the HIV-1 genome were the main subject of the projects developed during my Ph.D. The studies can mainly be grouped into three topics.

The first part of the work focused on the identification and characterization of new HIV-1 G4-binding proteins. Particular interest was devoted to the study of the G4 folding/unfolding modulation upon protein binding, as well as to demonstrate its biological relevance. Unveiling new insights in the regulation of HIV-1 mechanisms can lead to the discovery of new possible targets for the design of specific inhibitors.

Secondly, the folding and the regulatory roles of the LTR G4s within the viral latency environment were investigated. The persistence of transcriptionally silent proviruses, invisible both to the immune system and to current therapeutic tools, remains the major obstacle to HIV-1 eradication. As a consequence, the identification of G4-mediated latency regulation mechanisms might help to define a new molecular target for alternative therapeutic approaches.

The third aim of the study was to screen a new series of small drug-like molecules as antivirals. This approach constitutes the basis for the development of novel druggable antiviral compounds with an innovative G4-mediated mechanism of action.

3. Materials and methods

3.1 The HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands

3.1.1 The HIV-1 recombinant nucleocapsid protein

The full-length recombinant nucleocapsid protein was prepared as previously reported[216] in the lab of Professor Fabris (University at Albany, NY, USA). Briefly, an expression vector containing the gene for NCp7 was transformed and expressed in BL21(DE3)-pLysE cells. The protein was subsequently purified and desalted. A UV-Visible Spectrophotometer was utilized to determine the protein concentration in the obtained sample, using an extinction coefficient at 280 nm of 6.41 $M^{-1}cm^{-1}$. The stock solution was stored in aliquots at -80 °C until use.

3.1.2 Oligonucleotides and compound

Synthetic oligonucleotides were purchased from Sigma-Aldrich (Milan, Italy). RNA oligos were shipped in lyophilized form, then dissolved in diethyl pyrocarbonate (DEPC)-treated water at stock concentrations (1 mM) and stored at -80 °C until use.

Table 3.1 provides specific information about the oligonucleotides used in this study.

The G4 ligand BRACO-19 was obtained from Endotherm GmbH (Saarbrücken, Germany).

Application	Name	Sequence (5'-3')
EMSA, CD, MS analysis	RNA U3-III+IV	GGGAGGCGUGGCCUGGGCGGGACUGGGGAGUGG
RT stop assay	RNA U3-III+IV RT assay	GGGAGGCGUGGCCUGGGCGGGACUGGGGAGUGG CGAGCCCUCAGAUCCUGCAUAUAAGCA
	non-G4 RNA	GUAACCGAUGAGUCUAUGCGAGCCCUCAGAUCCU GCAUAUAAGCA
	Primer RT	TGCTTATATGCAGGATCTGAGG

Table 3.1: Oligonucleotide names, sequences, and their applications in the study of the NCp7-mediated unfolding of RNA G-quadruplex structures.

3.1.3 Electrophoretic mobility shift assay (EMSA)

Native gel shift assay is one of the most sensitive methods to study protein-nucleic acid interactions.[217] For this reason, here it was used to study the HIV-1 NCp7 binding to viral RNA G4s. For EMSAs RNA oligonucleotides, labeled with [γ-^{32}P-ATP] using T4 polynucleotide kinase at 37 °C for 30 min, were annealed by heating at 95 °C for 5 min in lithium cacodylate (10 mM, pH 7.4) and KCl (50 mM) buffer. The annealed oligonucleotides at 15 nM final concentration were added to 20 µl of binding reaction (8% glycerol, 30 mM Tris-HCl, 15 mM $MgCl_2$, 50 µM $ZnCl_2$)

containing appropriate levels of NCp7. For EMSA unfolding assays labeled RNA oligonucleotides were annealed to form G4s, and cold DNA complementary oligonucleotides were added to the binding reactions at equimolar or 2-fold excess strand ratio. Binding reactions were incubated for the indicated time in the presence of appropriate protein concentrations. Mixtures were then loaded on a 12% polyacrylamide native gel and run for 90 min at 90 V. Gels were dried, exposed overnight, and visualized by phosphorimaging (Typhoon FLA 9000, GE Healthcare).

3.1.4 Circular dichroism (CD) analysis

Circular dichroism (CD) is an absorption spectroscopy method considered as a primary tool for the characterization of G4 structures. Indeed, different G4s display unique CD spectral signatures (e.g., 260 nm maximum and 240 nm minimum peaks for parallel G4s, 295 nm maximum and 260 nm minimum peaks for antiparallel G4s, and two positive peaks at 295 nm and 260 nm for mixed G4s).[218,219] Moreover, CD spectroscopy is also useful to study interactions of G4 structures with various ligands (e.g., proteins and compounds). In fact, ligands binding can induce G4 conformational changes and variations in the G4 stability (following the T_m). For the aforementioned reasons, CD spectroscopy was applied in this study. For CD analysis, the RNA oligonucleotides were diluted to 2 μM concentration in 10 mM lithium cacodylate buffer (pH 7.4) supplemented with 50 mM KCl. Samples were annealed by heating at 95 °C for 5 min and gradually cooled to room temperature to allow G4 formation. When the unfolding properties of nucleocapsid protein were analyzed, it was added to the samples at 10-folds NCp7:oligonucleotide ratio and incubated for 3 h before CD analysis. Where specified, B19 was added at 8 μM concentration 4 h after the annealing step, and the samples were placed at 4 °C for 24 h to permit G4 stabilization. CD spectra were recorded on a Chirascan-Plus (Applied Photophysics, Leatherhead, UK) equipped with a Peltier temperature controller using a quartz cell of 5 mm optical path length and an instrument scanning speed of 50 nm/min over a wavelength range of 230−320 nm. The reported spectrum of each sample represents the average of 2 scans, and it is baseline corrected for signal contributions due to the buffer. Observed ellipticities were converted to mean residue ellipticity $(\theta) = deg \times cm^2 \times dmol^{-1}\ (mol\ ellip)$. Unfolding spectra were recorded over a temperature range of 20−90 °C, while 90-20 °C were used for annealing experiments, with a temperature increase rate of 1 °C/min. T_m values were calculated according to the van 't Hoff equation, applied for a two-state transition from a folded to unfolded state, assuming that the heat capacity of the folded and unfolded states are equal.

3.1.5 Mass spectrometry (MS) analysis

Electrospray ionization (ESI) mass spectrometry (MS) represents a powerful technique to investigate both G4 structures and G4/small molecules binding.[220] As a consequence, this technique has been employed to evaluate the NCp7-mediated G4s unfolding. For MS analysis, the RNA oligonucleotides were diluted to 5 μM concentration in a final buffer composition consisting of 0.8 mM KCl, 120 mM trimethylammonium acetate (TMAA) adjusted from pH ~7 to 7.4 with triethylamine (TEA). Samples were annealed by heating at 95 °C for 5 min, gradually

cooled to room temperature, added of 20% of water and incubated overnight at 4 °C. Where appropriate, NCp7 was added to the sample at a 1:1 protein:oligonucleotide ratio and the incubation was performed at 4° C for 3 h. At the time of analysis, a volume of 5 µL of each sample was typically scanned by direct infusion using electrospray ionization (ESI) on a Xevo G2-XS QTOF mass spectrometer (Waters, Manchester, UK). The ESI source settings were: electrospray capillary voltage 1.8 kV; source and desolvation temperatures 45 and 65 °C, respectively; sampling cone voltage 65 V. All these parameters ensured minimal fragmentation of the DNAs complexes. The instrument was calibrated using a 2 mg/mL solution of sodium iodide in 50% of isopropanol (IPA). Additionally, the use of the internal standard LockSpray (a solution of leu-enkephalin 1 µg/mL in acetonitrile/water (50:50, v/v) containing 0.1% of formic acid) provided a typical <5 ppm mass accuracy. This high-resolution system allowed us to visualize the isotopic pattern, identify the charge state, and therefore unambiguously calculate the neutral mass of the detected species.

3.1.6 Reverse transcriptase (RT) stop assay

RT stop assay is used to investigate the formation of RNA G4 structures in a sequence of interest, based on the assumption that a G4 sterically impedes the progression of a reverse transcriptase (RT) enzyme during the elongation of a radiolabeled DNA primer.[221] This assay can also be applied to the study of G4 ligands (e.g., proteins and compounds) interactions. In fact, binding partner-mediated G4 stabilization or unfolding can prevent and encourage, respectively, the progression of the RT enzyme, visible through differences in elongation products. The DNA primer was 5′-labelled with [γ-^{32}P-ATP] using T4 polynucleotide kinase at 37 °C for 30 min. The labeled primer (70 nM) was annealed to the RNA U3-III+IV and non-G4 RNA (control sequence unable to adopt G4 structures) in the presence of 50 mM KCl. The primer extension reaction was performed by recombinant HIV-1 reverse transcriptase (1 U/reaction; Calbiochem) at 44 °C for 1 h. Where specified, samples were incubated overnight with increasing concentrations of B19 (500 nM - 1 μM - 2 μM - 4 μM) at RT before primer extension. When nucleocapsid protein was used, appropriate concentrations (350 nM - 700 nM - 1.4 uM) of it were added immediately before the elongation reaction. Reaction products were treated with NaOH (2 N) at 95 °C for 3 min to permit the alkaline hydrolysis of RNA, and the pH was adjusted with HCl (2 N) to neutrality. Samples were ethanol precipitated, and extension products were separated on 16% denaturing polyacrylamide gel and visualized by phosphorimaging (Typhon FLA9000; GE Healthcare).

3.2 The HIV-1 LTR G-quadruplexes fold in latently infected cells and their stabilization by G4 ligands counteract viral reactivation from latency

3.2.1 Cell line

U1 cells are a subclone of U937 cells that have been chronically infected with HIV-1. U937 is a pro-monocyte obtained from a pleural effusion of a two-year-old Caucasian male with diffuse histiocytic lymphoma. U1 cells are widely used for latency induction experiments. These cells

show minimal constitutive expression of virus, and the surface expression of CD4 is low. Cells were maintained in RPMI 1640 medium (Gibco, Life Technologies, Monza, Italy) supplemented with 10% heat-inactivated FBS (Gibco, Life Technologies, Monza, Italy). All cultures were grown in a humidified incubator maintained at 37 °C with 5% CO_2. Cells were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH.

3.2.2 Cell treatments

Phorbol 12-myristate 13-acetate (PMA), also known as 12-*O*-tetradecanoylphorbol-13-acetate (TPA) (Sigma-Aldrich), were used as HIV-1 latency-reversing agent (LRA). The G4 ligand B19 was obtained from Endotherm GmbH (Saarbrücken, Germany), and the core-extended NDI (c-exNDI) H-NDI-NMe_2-PhAm was synthesized and kindly provided by Professor Filippo Doria and Professor Mauro Freccero (University of Pavia).

3.2.3 BG4 production and purification

BG4-encoding plasmid (kindly provided by Professor Shankar Balasubramanian, University of Cambridge, UK) was transformed into BL21(DE3) competent cells (Stratagene) which were cultured in TY medium (1.6% tryptone peptone, 1% yeast extract, 0.5% NaCl) and 50 µg/ml of kanamycin. Transformed cells were grown at 37 °C 160 rpm to an OD_{600} of 0.7-0.8. BG4 antibody expression was induced with 0.85 mM isopropyl β-D-1-thiogalactopyranoside overnight at RT. Then, cells were pelleted for 25 min at 25000g at 4 °C, resuspended in lysis buffer (20 mM Tris-Cl pH 8.0, 50 mM NaCl, 5% glycerol, 1% Triton X-100 and 100 µM phenylmethanesulfonylfluoride solution), and lysed through 5 cycles of freezing and thawing. After centrifugation at 10000g at 4 °C for 20 min, the supernatant was filtered (0.45 µm) and purified on a Protino Ni-NTA-agarose affinity column (Machery-Nagel, Germany) according to the manufacturer instructions. The column was washed in 20 mM imidazole in 20 mM Tris HCl pH 8.0, 300 mM NaCl, and BG4 antibody 1.5 ml fractions eluted in 250 mM imidazole in 20 mM Tris HCl pH 8.0, and 300 mM NaCl. The eluted fractions were checked on a Coomassie-stained SDS-PAGE and concentrated in Amicon Ultra-3k Centrifugal Filter Unit (Merck Millipore, Milan, Italy). The BG4 concentration was determined using Pierce™ BCA protein assay kit (Thermo Fisher Scientific, Waltham, MA, USA), and the antibody was stored at −20 °C until use.

3.2.4 Chromatin immunoprecipitation (ChIP) assay

The ChIP assay is a powerful and versatile technique to probe protein-DNA interactions within the natural chromatin context of the cell.[222,223] Here it has been employed to study the presence of G4-folded structures in latently infected cells. One million of U1 cells were fixed in RPMI containing 1% (v/v) formaldehyde and 10% (v/v) FBS for 8 min at RT. After 5 min quenching with 125 mM glycine, cells were pelleted and washed twice with phosphate-buffered saline (PBS) 1X (Gibco, Life Technologies, Monza, Italy) containing 10% FBS. The flash-frozen pellets were lysed for 5 min on ice in 100 µl of 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5% SDS and protease inhibitor cocktail. Samples were sonicated using the Covaris E220 to shear chromatin to an

average size of 100-500 bp (2% duty cycle, 200 cycles per burst, at the intensity of 3) for 30 min. Sheared chromatin was diluted 1:5 in IP-buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, and 140 mM NaCl) supplemented with protease inhibitor cocktail. After a centrifugation step of 10 min at 13000g at 4 °C, the supernatant containing soluble chromatin fraction was recovered and incubated with 0.7 mg/ml RNase A (Thermo Fisher Scientific) for 30 min at 37 °C. The quality of the sheared chromatin fragments was checked by Agilent Bioanalyzer using Agilent DNA High Sensitivity Chips (Agilent Technologies, Milan, Italy). For ChIP experiments 10 µl protein-G magnetic beads (Pierce™, Thermo Fisher Scientific) were washed in IP-buffer and incubated with 1 µg Anti-FLAG Ab (Sigma-Aldrich) for 1 h at 4 °C on a rotating wheel. Then, 50 µl of RNA digested chromatin was incubated with 450 ng BG4 Ab (or without for the mock negative control) for 1 h at 16 °C. The anti-FLAG coated beads were washed with IP-buffer and incubated with chromatin-BG4 complex for 3 h at 4 °C on a rotating wheel. Beads were washed 4 times with IP-buffer and once in wash buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA). Elution of immunoprecipitates and chromatin crosslink reversal were performed incubating beads with 70 µl elution buffer (10 mM Tris-HCl pH 8.0, 5 mM EDTA, 300 mM NaCl and 0.5% SDS) containing 0.3 mg/ml RNase A (Thermo Fisher Scientific) for 30 min at 37 °C followed by the addition of 0.5 mg/ml proteinase K for 1 h at 55 °C and 8 h at 65 °C shaking. The supernatant was then recovered and incubated for 1 h at 55 °C in the presence of 0.25 mg/ml proteinase K (Thermo Fisher Scientific). The eluted fragments were purified with the MinElute® PCR purification kit (Qiagen GmbH, Hilden, Germany) and quantified with the Qubit™ 4 fluorometer (Invitrogen™, Thermo Fisher Scientific).

3.2.5 qPCR

Input, immunoprecipitated, and mock-derived eluted fragments were used to quantify G4 enrichment via quantitative PCR, using dual-labeled fluorescent probes (TaqMan probes) in order to increase the specificity and the sensitivity of the qPCR. A standard TaqMan™ universal PCR master mix (Applied Biosystems) was used, and samples were analyzed through an ABI 7900 HT fast real-time PCR system (Applied Biosystems). Cycling conditions were 95 °C for 10 min, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Primer pairs and relative dual-fluorescent probes (5'-FAM and 3'-TAMRA marked probes) (Sigma-Aldrich) are listed in Table 3.2.

They were designed to detect both G4 ChIP positive and negative regions. Relative enrichments were derived with respect to their inputs.

Name	G4	Type	Sequence (5'-3')
HIV-1 LTR	+	Forward primer	GCTGCATCCGGAGTACTACAA
		Reverse primer	GAGAGACCCAGTACAGGCAA
		Probe	FAM-CAGCTGCTTATATGTAGCATCTGAGGGC-TAMRA
Human CDK4	+	Forward primer	CCACCCTCACCATGTGACC
		Reverse primer	CTTACACTCTTCGCCCTCCTC
		Probe	FAM-TGCCAAAGAGGCGCGCGGAAACTG-TAMRA
Human ESR1	-	Forward primer	GAAACAGCCCCAAATCTCAA
		Reverse primer	TTGTAGCCAGCAAGCAAATG
		Probe	FAM-AGTGGCACCCAGACTTGATGGCCGAC-TAMRA

Table 3.2: Names and sequences of primer pairs and probes used in the ChIP-qPCR assay.

3.2.6 Cytotoxicity assay

Cytotoxicity of the tested compounds (H-NDI-NMe_2-PhAm and B19) was determined through the ATPlite™ Luminescence ATP detection assay system (Perkin Elmer, Brussels, Belgium). This assay is based on firefly luciferase and allows quantitative evaluation of proliferation and cytotoxicity of cultured mammalian cells. The adenosine triphosphate (ATP) is present in all metabolically active cells, and its concentration declines very rapidly when the cells undergo necrosis or apoptosis; it is therefore a marker for cell viability. Briefly, $2x10^4$ U1 cells/well were plated in white but clear flat bottom 96 wells plates. Cells were next treated at the appropriate time with serial dilutions of compounds. 72 h after seeding, 100 μL of plated cells were supplemented with 50 μL of mammalian cell lysis solution and shook for 5 min in an orbital shaker at 700 rpm. After that, 50 μL of substrate solution was added to the wells, and the microplate was shaken again for 5 min. Finally, the plate was dark-adapted for 10 min, and the luminescence was read with the microplate luminometer Centro LB 960 (Berthold Technologies GmbH, Germany).

3.2.7 p24 ELISA assay

The enzyme-linked immunosorbent assay (ELISA) is a technique designed for detecting and quantifying substances, such as proteins. The p24 ELISA assay quantifies the amount of the HIV-1 p24 capsid protein in culture supernatants. Given that the level of p24 correlates directly with virus titer, we can use the technique to evaluate virus production. Briefly, U1 cells were seeded into 96-wells plates ($2x10^4$ cells/well) and treated with serial dilutions of compounds at the indicated times after seeding. Cells were also stimulated by PMA treatment (at 10 nM

concentration). At 48 h post-PMA-stimulation, supernatants were collected and subjected to protein titration through the p24 ELISA kit (Perkin Elmer). The obtained absorbance levels were evaluated at 450 nm using a Sunrise Tecan plate reader (Mannendorf, Switzerland). The kit is supplied with a p24 control which is used to generate a standard curve and calibrate the p24 equivalent of the supernatants.

3.3 Screening of a new series of small molecules as antivirals

3.3.1 Cell lines

Human embryonic kidney 293T (HEK 293T, ATCC®) cell line is a highly transfectable derivative of human embryonic kidney 293 cells. Thanks to their features, these type of cells were employed for the production of HIV-1 viral stocks. TZM-bl is a HeLa cell clone that has been engineered to express CD4 and CCR5 receptors and contains integrated reporter genes for both firefly luciferase and E. coli β-galactosidase under the control of the HIV-1 LTR promoter. These cells are highly permissive to infection by most strains of HIV, and they allow sensitive and accurate measurements of infection. The TZM-bl reporter cell line was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH. African green monkey kidney cells (Vero cells, ATCC® CCL-81™) and human bone osteosarcoma cells (U-2 OS, ECACC 92022711) are cell lines highly permissive to infection by various HSV strains. All cell lines used in this study were maintained in Dulbecco's Modified Eagle Medium (DMEM, Gibco, Life Technologies, Monza, Italy) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco, Life Technologies, Monza, Italy), and were maintained in a humidified incubator set at 37 °C with 5% CO_2.

3.3.2 Viruses and viral stocks production

The viral stock of the HIV-1 strain NL4-3 was produced transfecting HEK 293T cells with wild-type (wt) X4 proviral genome (NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH). The CalPhos mammalian transfection kit (Clontech Laboratories, Santa Clara, CA, USA) was utilized following the manufacturer's protocol. Viral particles in supernatants were collected and titrated using the Reed and Muench method.[224] HSV-1 strain F (GenBank: GU734771) was a kind gift of Professor B. Roizman (University of Chicago, Illinois, USA), whereas recombinant HSV-1 expressing VP16-GFP (HSV-1 v41) virus was kindly provided by Peter O'Hare (Imperial College London, UK).[225] For viral stocks propagation, confluent Vero cells were infected with HSV-1 strain F or HSV-1 v41 at a multiplicity of infection (MOI) of 0.01 and maintained from 48 h. Then, supernatants were collected and subsequently titrated through the plaque reduction assay (PRA).[226]

3.3.3 Antiviral assays

The antiviral activity of the tested compounds against HIV-1 was assessed through the measurement of the LTR-driven luciferase expression in the TZM-bl reporter cells. Briefly, TZM-bl cells were seeded in 96-well plates (10000 cells/well) and grown overnight to allow cells

attachment. Cells were next infected with HIV-1 strain NL4-3 at a multiplicity of infection (MOI) of 0.5, treated with serial dilutions of compounds (from 25 µM to 100 nM) and incubated at 37 °C. After 48 h HIV-1 production was quantified using the BriteLite™ plus reporter gene assay system (Perkin Elmer), according to the manufacturer's instructions. Antiviral activities against HSV-1 was investigated by plaque reduction assay. For virus infection, wt HSV-1 (strain F) was added to cells at an MOI of 1 in serum-free medium. After 1 h at 37 °C, the inoculums were replaced with complete medium and compounds were next added at increasing concentrations (25–400 nM). Since a single round of HSV-1 replication takes around 24 h to complete[198], supernatants were collected 24 hours post-infection (h.p.i.) and stored at −80 °C until viral titrations by plaque assay. For plaque reduction assay, U-2 OS cells were seeded in 24-well plates (85000 cells/well) and incubated overnight. Cells were then infected with 250 µl of serially-diluted (10-folds) supernatants for 1 h at 37 °C. After infection, cells were washed with PBS 1X and incubated with 500 µl of DMEM supplemented with 0.6% methylcellulose (Sigma-Aldrich) and 2% FBS. 48 h.p.i., cells were washed with PBS 1X and fixed with formaldehyde 5% in PBS 1X for 20 min at RT, then colored with crystal violet 0.8% (in ethanol 50%). Viral plaques were counted using an optical microscope (Zeiss, Jena, Germany). The 50% inhibitory concentration (IC_{50}), defined as the concentration of compound that inhibits the HIV-1/HSV-1 production by 50%, was determined from the dose-response curve.

3.3.4 Cytotoxicity assays

The cytotoxicity of the tested compounds was evaluated at the same time of antivirals by the MTT assay (Sigma-Aldrich). The MTT assay is a colorimetric assay for assessing cell viability based on the conversion of yellow 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) into purple formazan crystals by mitochondria of living cells. Since the total mitochondrial activity is related to the number of viable cells, this assay is broadly used to measure the *in vitro* cytotoxic effects of drugs of interest on cell lines. TZM-bl (10000 cells/well) and U-2 OS (9000 cells/well) cells were plated in 96-well plates. 24 h after seeding serial dilutions of compounds (from 25 µM to 100 nM) were dispensed. After 48 h and 24 h from TZM-bl and U-2 OS treatments, respectively, cells were supplemented with 10 µl of freshly diluted MTT solution (5 mg/mL in PBS 1X) and incubated for 4 h at 37 °C. MTT crystals were then solubilized by the solubilization solution (10% sodium dodecyl sulfate (SDS) and 0.01 M HCl). After overnight incubation at 37 °C, absorbance was measured by Sunrise Tecan plate reader (Mannendorf, Switzerland) at 540 nm. The concentration of compound required to reduce cell grown by 50% (50% cytotoxic concentration or CC_{50}) was determined from the dose-response curve.

3.3.5 Oligonucleotides and compounds

Synthetic and desalted DNA oligonucleotides were purchased from Sigma-Aldrich. They were dissolved in Tris-EDTA buffer (TE buffer; Tris-HCl 10 mM, EDTA 1 mM pH 8.0) at 1 mM final stock concentration and stored at -20 °C until use. Table 3.3 provides specific information about the oligos used in this third study. The two control compounds acyclovir (ACV) and phosphonoacetic acid (PAA) were purchased from Sigma-Aldrich.

Application	Name	Sequence (5'-3')
CD, MS analysis	un2	GGGGGCGAGGGGCGGGAGGGGGCGAGGGG
	un3	GGGAGGAGCGGGGGGAGGAGCGGG
	gp054a	GGGGTTGGGGCTGGGGTTGGGG
	hTel21	GGGTTAGGGTTAGGGTTAGGG
	c-myc	TGGGGAGGGTGGGGAGGGTGGGGAAGG
Taq pol stop assay	HSV *Taq* primer	GGCAAAAAGCAGCTGCTTATATGCAG
	un2	TTTTTGGGGGCGAGGGGCGGGAGGGGGCGAGG GGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC
	un3	TTTTTGGGAGGAGCGGGGGGAGGAGCGGGTTTT TCTGCATATAAGCAGCTGCTTTTTGCC
	gp054a	TTTTTGGGGTTGGGGCTGGGGTTGGGGTTTTTCT GCATATAAGCAGCTGCTTTTTGCC
	HSV *Taq* non-G4 ct	TTGTCGTTAAAGTCTGACTGCGAGCTCTCAGATCC TGCATATAAGCAGCTGCTTTTTGCC

Table 3.3: Oligonucleotide names, sequences, and their applications in the screening of a new class of G4 ligands.

3.3.6 Circular dichroism (CD) analysis

Circular dichroism experiments have been employed to study G4-small molecules interactions following both G4s conformational changes and G4s stabilization (through T_m variations) induced by the ligand binding upon addition of the compounds.[227] The analysis was performed using a Chirascan-Plus (Applied Photophysics) equipped with a Peltier temperature controller using a quartz cell of 5 mm path length. G4 oligonucleotides were diluted to a final concentration of 4 μM in the absence or presence of different concentrations (2.5, 100) KCl and 10 mM lithium cacodylate buffer. After annealing step (5 min at 95 °C), DNA samples were gradually cooled down and, where specified, compounds were added at a final concentration of 16 μM. Thermal unfolding analyses were recorded from 230 to 320 nm over a temperature range of 20–90 °C (5 °C/min). The reported spectrum of each sample represents the average of 2 scans, and it is baseline corrected for signal contributions due to the buffer. Observed ellipticities were converted to mean residue ellipticity $(\theta) = deg \times cm^2 \times dmol^{-1}\ (mol\ ellip)$. T_m values were calculated according to the van 't Hoff equation, applied for a two-state transition from a folded to unfolded state, assuming that the heat capacity of the folded and unfolded states are equal.

3.3.7 *Taq* Polymerase stop assay

Taq polymerase stop assay is used to investigate the G4 formation in a DNA sequence of interest, based on the assumption that a G4 sterically impedes the progression of a DNA polymerase during the elongation of a ^{32}P-labeled primer complementary to the 3'-end of the G4 template.[221]

When the template containing G4 forming regions is elongated, truncated products are formed and are discriminated in a denaturing polyacrylamide gel. Specific markers (specified above) are used to evaluate whether the stop bands correspond to the Gs involved in the G4 formation. This technique has been employed to validate the G4 stabilizing effect of the best G4 ligand candidate, following enzyme processivity. The DNA primer (Table 3.3) was 5'-end-labeled with [γ-^{32}P-ATP] using T4 polynucleotide kinase (Thermo Fisher Scientific) at 37 °C for 30 min and then purified with Illustra MicroSpin G-25 columns (GE Healthcare). The labeled primer (final concentration 72 nM) was annealed to the templates (final concentration 36 nM) (Table 3.3) in lithium cacodylate buffer (10 mM, pH 7.4) in the presence or absence of the indicated concentration of KCl by heating at 95 °C for 5 min and gradually cooling to RT to allow both primer annealing and G4 folding. Where indicated, samples were incubated with various concentrations (2-8 μM) of the selected compound (GSA-0932) at RT overnight. For primer extension, AmpliTaq Gold DNA polymerase (2 U/reaction, Applied Biosystems) was incubated at the indicated temperature for 30 min. Reactions were stopped by ethanol precipitation, and primer extension products were separated on a 16% denaturing gel, finally visualized by phosphorimaging (Typhoon FLA 9000, GE Healthcare). Markers were prepared based on Maxam and Gilbert sequencing protocol by PCR reaction with ^{32}P-labeled primer. PCR products were treated with formic acid for 5 min at 25 °C and then with piperidine for 30 min at 90 °C.

3.3.8 Mass spectrometry (MS) competition assay

Here, native mass spectrometry was employed to evaluate the binding specificity of the best compound GSA-0932 to HSV-1 G4s instead of various competitors. Oligonucleotides were heat-denatured and folded in 0.4 mM KCl, 120 mM trimethylammonium acetate (TMAA), pH 7.4, and 20% isopropanol (IPA) overnight at 4 °C. The oligonucleotides were diluted to a final concentration of 4 μM and incubated with the tested compound at ratio DNA:compound 1:1.5 overnight at 4 °C. Samples were analyzed by direct infusion electrospray ionization (ESI) on a Xevo G2-XS QTof mass spectrometer (Waters, Manchester, UK). The injection was automatically performed by an Acquity UPLC h-class (Waters) equipped with an autosampler; the carrying buffer was TMAA 120 mM, pH 7.4, 20% IPA. Up to 5 μL samples were typically injected for each analysis. The ESI source settings were the following: electrospray capillary voltage set at 1.8 kV, the source and desolvation temperatures were 45 °C and 65 °C respectively, the sampling cone was at set at 65 V. All these parameters ensured minimal fragmentation of the DNAs complexes. The instrument was calibrated using a 2 mg/mL solution of sodium iodide in 50% of IPA. Binding affinities were calculated for each experiment using the peak intensity for each species calculated by MassLynx V4.1. The binding affinity was calculated with the following formula: [BA = (ΣG4b/(ΣG4f + ΣG4b)) × 100], where BA is the binding affinity, G4b is the intensity of bound G4 DNA, and G4f is the intensity of free G4 DNA.

3.3.9 Time of addition (TOA) assay

The time of addition assay can provide information about the mode of action of newly discovered antiviral agents. Specifically, it determines how long the addition of a compound can be postponed before losing its antiviral activity, and so it establishes what is the last step of the viral life cycle affected by the presence of a compound, as reported by[228]. U-2 OS cells were seeded in 24-well plates at a density of 80000 cells/well and incubated overnight. The day after, cells were infected with HSV-1 strain F at an MOI of 0.5 and treated every two hours (from 0 to 10 h.p.i.) with the tested compound (400 nM) or with ACV (1.6 μM) as reference drug. The supernatants were collected at 30 h.p.i., and then titrated following the plaque reduction assay working protocol described above (Section 3.3.3).

3.3.10 UV-visible and fluorescence spectra

In order to visualize the compound in cells, we have investigated its spectroscopic features. As a consequence, UV-visible and fluorescence emission spectra have been recorded. Spectrophotometric measures were performed using a Lambda 25 UV-Vis spectrophotometer (Perkin Elmer). Absorption spectra and determination of maximum absorption wavelength were carried out in 10 mM lithium cacodylate buffer, pH 7.4 at a compound concentration of 25 μM. The emission spectra were performed on an LS-55 fluorescence spectrophotometer (Perkin Elmer) at a 2.5 μM compound concentration in 10 mM lithium cacodylate buffer, pH 7. All instruments were equipped with Peltier temperature controllers. Quartz cuvettes of 0.4-10 cm path length were used.

3.3.11 Confocal microscopy analysis

Confocal microscopy has been employed both to validate the antiviral effects of the lead compound through the viral GFP intensity reduction and to observe a possible colocalization of the compound with the VP16-GFP virus. U-2 OS cells were seeded at 40000 cells/well in 48-well plates (sterile 48-well glass coverslips) and grown overnight at 37 °C. Cells were next infected with the GFP-expressing recombinant virus v41[225] at MOIs of 1 (for viral replication inhibition) or 3 (for colocalization) for 1 h at 37 °C in serum-free medium. Then, cells were washed with PBS 1X and incubated at 37 °C in complete medium for the indicated time (6, 8, 10 h.p.i.). Cells were treated with the GSA-0932 compound (1-6 μM) for 2 h at 37 °C and eventually with PAA (400 nM), as replication inhibitor control. At the appropriate times post infection, cells were washed 5 times with PBS 1X, fixed with 2% paraformaldehyde (PFA, Sigma-Aldrich) in PBS 1X for 20 min at RT, and subjected to another 5 PBS 1X washes. Mock-infected cells were treated in the same way as infected cells in each type of experiment and staining, except that serum-free medium was added in place of the virus. Coverslips were mounted in Vectashield mounting medium (DBA Italia, Milan, Italy). For viral replication inhibition studies, images were acquired using a Nikon A1Rsi+ Laser Scanning confocal microscope equipped with NIS-Elements Advanced Research software (Nikon Instruments Inc., Melville, USA). GFP fluorescence was evaluated in the 500-550 nm emission filter range, using an excitation wavelength of 488 nm. For overall fluorescence

intensity reduction quantification, cells were counted using ImageJ software, and signals were appropriately normalized. At least three pictures per condition were considered. For colocalization studies, images were captured with a Leica TCS SP5 confocal laser scanning microscope (Leica microsystems, Germany). Laser excitation and emission filters were: diode laser λ_{ex} at 405 nm, λ_{em} 415-460 nm, in order to visualize the GSA-0932 fluorescence; whereas blue argon laser λ_{ex} at 488 nm, λ_{em} 500–550 nm, to detect the GFP fluorophore. Final images included in this work are representative of multiple experiments.

3.3.12 Immunoblot analysis

The western blot (WB) is a widely used analytical technique to detect a specific protein of interest in a sample. In this study, WB assay has been used to quantify the amount of an immediate-early (IE) viral protein expression in treated samples versus untreated controls. U-2 OS cells were seeded at 280000 cells/well in 6-well plates and grown overnight at 37 °C. Cells were next infected with the wt HSV-1 virus at an MOI of 1. 15 min post-infection cells were treated or not with the GSA-0932 compound at increasing concentrations (6-12 μM). At 60 min after infection, complete medium was restored, and 5 h post-infection cells were scraped off and centrifuged at 1200 rpm for 5 min. Pellets were incubated with the radioimmunoprecipitation assay buffer (RIPA, 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EGTA, 1% IGEPAL® CA-630, 1% sodium deoxycholate, 1 mM Na_3VO_4, 1X protease inhibitor cocktail; 25 μl/pellet) for 30 min at -20 °C. After centrifugation at 13000 rpm for 5 min at 4 °C, supernatants were collected, and protein concentrations were quantified by using the Pierce® BCA Protein Assay Kit (Thermo Fisher Scientific), and samples stored at -80 °C. Each sample was electrophoresed on 8% SDS-PAGE and transferred to a nitrocellulose blotting membrane (Amersham™ Protan™, GE Healthcare Life science) by using trans-blot SD semi-dry transfer cell (Bio-Rad Laboratories, Milan, Italy). The membranes were blocked with 2.5% skim milk in PBS 1X. Membranes were then incubated for 2 h with the respective primary antibody directed against ICP4 (mouse monoclonal, H943, Santa Cruz Biotechnology, CA, USA), α-tubulin (mouse monoclonal, Sigma-Aldrich), and β-actin (mouse monoclonal; Sigma-Aldrich). After three washes in PBST (0.05% Tween-20 in PBS 1X), membranes were incubated for 1 h with ECL Plex Goat-α-Mouse IgG-Cy5 (GE Healthcare). Images were captured on the Typhoon FLA 9000 and quantified by ImageQuant TL software.

4. Results and discussion

4.1 The HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands

Our group has previously demonstrated that two RNA G-quadruplexes (i.e., U3-III and U3-IV) can form in the G-rich U3 region (positions 8981–9038, representative strain HXB2 LAI, NC 001802) of the single-stranded RNA genome of HIV-1.[190] The U3 RNA G4s exhibited a remarkable high thermodynamic stability under physiological conditions, significantly higher than that of their LTR DNA counterparts (i.e., T_m = 82.1 °C and 71.2 °C for U3-III and U3-IV, respectively, in 100 mM K^+). Importantly, *in vitro* study of reverse transcriptase progression (RT stop assay) highlighted that RNA G4 folding prevents the enzyme from complete elongation of the DNA strand, thus negatively affecting the reverse transcription process. However, RT is an unavoidable step for successful viral replication, and it is completely carried out during a typical infection. Thus, we envisaged the presence of a protein with G4-unfolding activity that would allow progression of the reverse transcriptase (RT) in infected cells. Since the HIV-1 nucleocapsid protein (NCp7) remains associated with the viral RNA during reverse transcription, displays nucleic acids chaperone activity, and has been proved to bind and unfold DNA G4s,[164,229–232] here we explored the possibility that the NCp7 modulated HIV-1 RNA G4s stability.

4.1.1 NCp7 binds and unfolds the RNA G4s in the U3 region of the HIV-1 RNA genome

We first assessed the ability of NCp7 to bind the G4-folded U3 sequences. To this purpose, the RNA U3-III+IV oligonucleotide, folded into G4 in the presence of increasing amounts of NCp7, was analyzed by electrophoretic mobility shift assay (EMSA) (Figure 4.1).

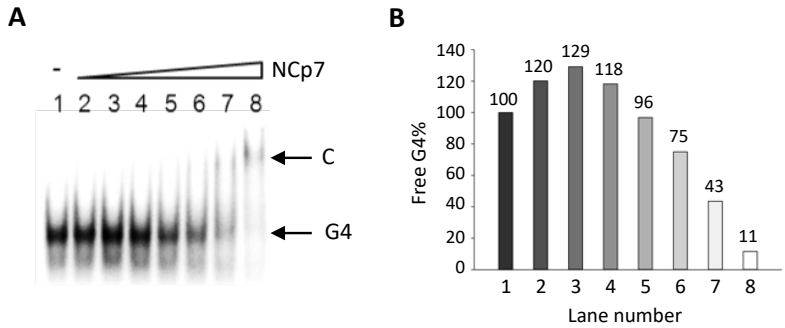


Figure 4.1: **EMSA with the RNA G4 U3-III+IV and NCp7.** (**A**) Increasing concentrations of the protein (7.5 nM - 15 nM - 37.5 nM - 75 nM - 112.5 nM - 150 nM - 300 nM) were incubated with RNA U3-III+IV G4-folded in 50 mM KCl buffer. C indicates the protein/oligonucleotide complex. (**B**) The graph shows the free G4 band intensity quantification for each lane reported in percentage compared to the control.

Complexes with NCp7 were reported to precipitate, and thus the disappearance of the free G4 starting from NCp7 75 nM was deemed indicative of the binding. At NCp7 300 nM (Figure 4.1, lane 8) a faint band corresponding to the G4-NCp7 complex became visible alongside the almost complete disappearance of the free G4, which indicated binding saturation (Figure 4.1). Since during retro-transcription a DNA/RNA intermediate forms, we next evaluated the binding properties of NCp7 to the G4-forming U3-III+IV sequence in the presence of the complementary DNA oligonucleotide (Figure 4.2).

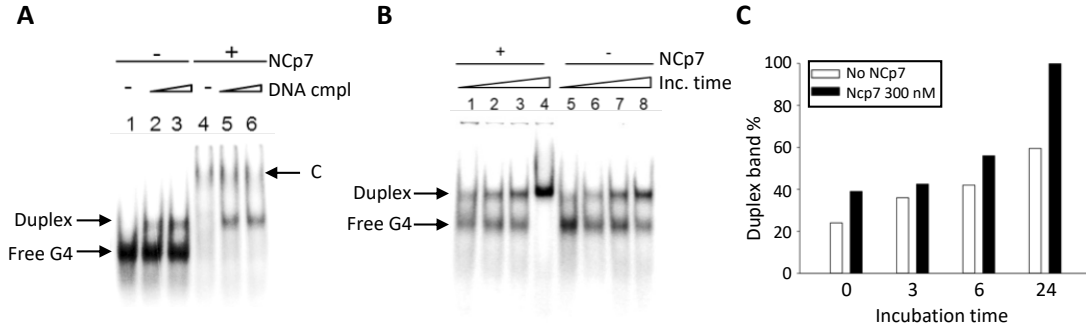


Figure 4.2: **NCp7 binding to the G4 in the presence of cold complementary DNA strand.** (**A**) The ^{32}P-marked RNA U3-III+IV G4 was annealed in 50 mM KCl buffer and incubated with 0.5 (lanes 2 and 5) and equimolar (lanes 3 and 6) ratios of unlabeled complementary DNA for 1 h. Then, NCp7 was added at a final concentration of 300 nM and incubated for 10 min at 37 °C. C indicates the protein/oligonucleotide complex. (**B**) NCp7 binding to the G4 in the presence of cold complementary DNA strand for increasing time. RNA G4 was folded in 50 mM KCl buffer and incubated with a 2-fold excess of cold complementary DNA strand for 0 - 3 - 6 - 24 h in the presence (lanes 1-4) and absence (lanes 4-8) of 300 nM NCp7. Duplex and free G4 species are indicated by arrows. (**C**) Quantification of the duplex from (B).

Addition of the unlabeled (cold) complementary DNA strand to the labeled and G4-folded RNA sequence induced formation of a hybrid RNA-DNA duplex, which had a slower migration rate and thus could be separated from the single-stranded G4-folded species on a native polyacrylamide gel (Figure 4.2 A, lanes 1-3). In the presence of NCp7, the RNA-DNA duplex competed for formation of the RNA G4-NCp7 complex (Figure 4.2 A, lanes 4-6). The free G4 species completely disappeared upon G4-NCp7 complex formation, while the amount of the free duplex was not perturbed. These data indicate the preferential binding of NCp7 to the G4-folded RNA *vs* the DNA/RNA duplex. To form the duplex, the thermodynamically stable RNA G4 has to be unfolded prior to base-pairing with its complementary strand. We thus analyzed duplex formation kinetics: the amount of duplex increased over time and reached 60% at 24 h (Figure 4.2 B, lanes 5-8). Addition of NCp7 highly increased the amount of the duplex species, especially at 24 h when the G4 folded oligonucleotide completely converted to the duplex form (Figure 4.2 B, lanes 1-4). These results indicate that when NCp7 binds to the HIV-1 RNA G4 structures, it stimulates their unfolding.

In order to confirm the observed unfolding properties of the viral protein toward G4s, circular dichroism experiments were performed. Briefly, the G4-folded U3-III+IV RNA was incubated in the absence or presence of NCp7 (10x concentration) and analyzed by CD spectroscopy (Figure 4.3).

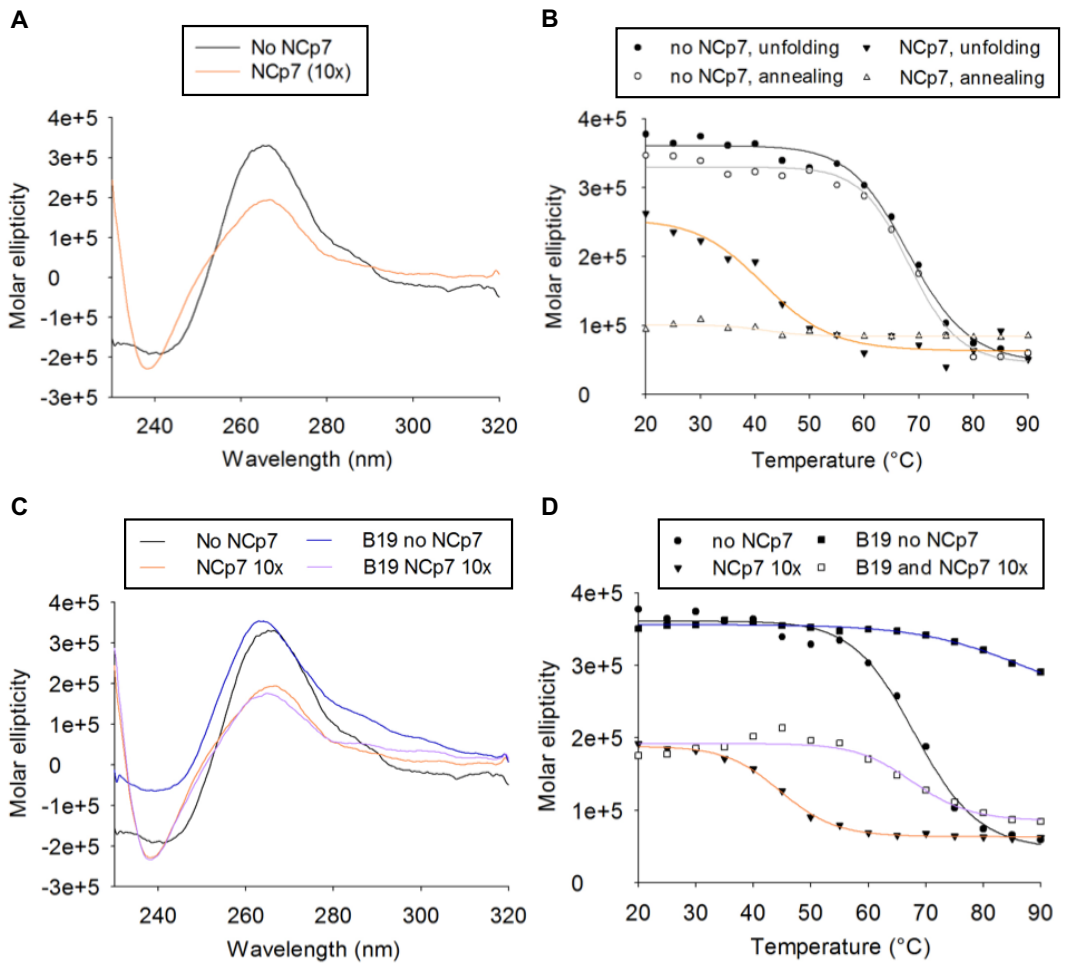


Figure 4.3: **CD analysis of the HIV-1 RNA G4 unfolding by NCp7.** (**A**) CD spectra registered at 20 °C of the G4-folded RNA U3-III+IV (50 mM K^+) in the absence (black line) and presence (orange line) of NCp7. (**B**) CD melting and annealing curves of the RNA G4 U3-III+IV in the absence/presence of NCp7. The molar ellipticity at the peak wavelength (265 nm) is shown as a function of the temperature. (**C**) CD spectra of the RNA G4 U3-III+IV in the absence/presence of both B19 and NCp7. (**D**) CD melting curves of the samples from Figure 4.3 C. The molar ellipticity at the peak wavelength (265 nm) is shown as a function of the temperature.

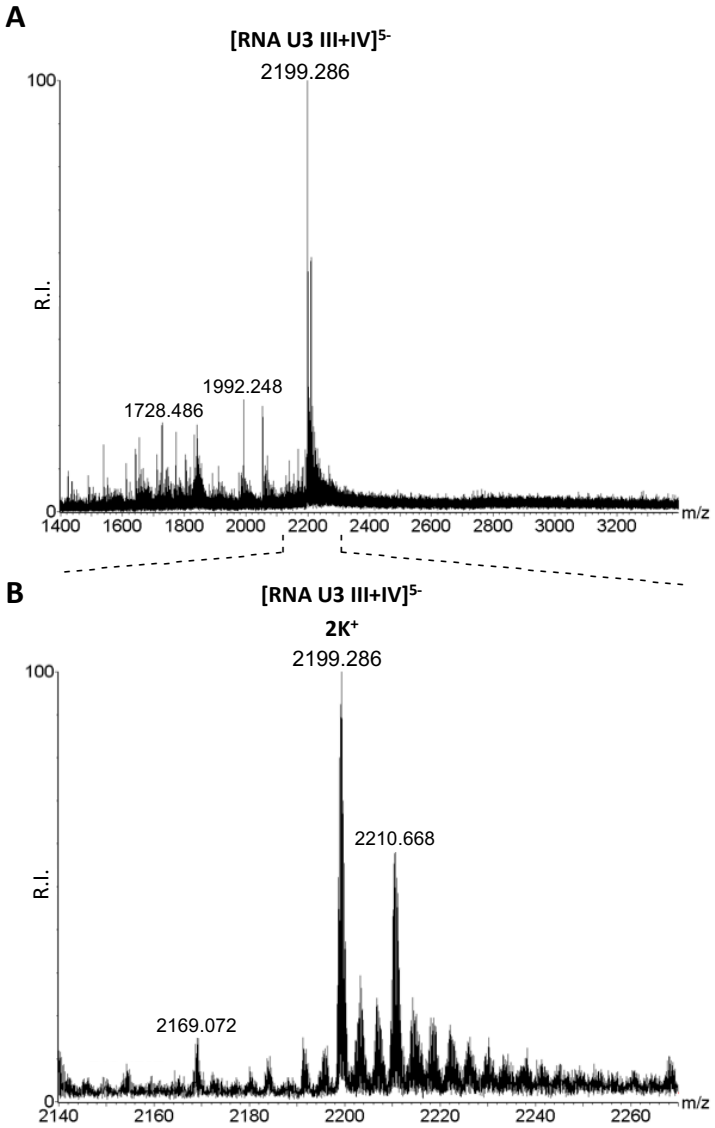
Sample	**T_m ± SD (°C)**
RNA U3-III+IV	68.3 ± 0.9
+ NCp7	44.8 ± 3.6
+ B19	>90
+ NCp7 and B19	67.8 ± 1.8

Table 4.1: **Melting temperatures of the samples from Figure 4.3, panel D.** SD stands for standard deviations.

Upon addition of NCp7, the molar ellipticity of U3-III+IV G4 drastically decreased, while the CD spectrum maintained the G4 signature, with a maximum at 265 nm and a minimum at 238 nm (Figure 4.3 A). Usually, low molar ellipticity indicates low stability of the tested G4s. To check the actual stability of U3-III+IV G4 in the presence/absence of NCp7, CD spectra were recorded at increasing temperature (Figure 4.3 B). The melting temperature (T_m) calculated according to the van 't Hoff equation applied to the molar ellipticity signal at 265 nm *vs* the temperature was

68.3 °C and 44.8 °C in the absence and presence of NCp7, respectively, indicating the effective unfolding of the RNA G4 mediated by NCp7. When the unfolded U3-III+IV G4 was reannealed by steadily decreasing the temperature from 90 °C to 20 °C, in the absence of NCp7 the T_m maintained its value of 68.3 °C whereas, in the presence of the protein, the oligonucleotide was unable to regain the folded G4 structure (Figure 4.3 B), indicating that NCp7 was able to maintain its unfolding properties in this condition. We next investigated if the G4 ligand B19, which has been shown to stabilize the HIV-1 U3 G4s, could inhibit the G4-unfolding activity of NCp7. We incubated U3-III+IV G4 with B19 before addition and further incubation with NCp7 (Figure 4.3 C, D). B19 increased the G4 T_m up to > 90 °C; in the presence of NCp7, this was reduced to 67.8 °C (Figure 4.3 D) indicating that B19 was able to in part suppress NCp7 unfolding activity.

NCp7 unfolding properties were further assessed by electrospray ionization (ESI) mass spectrometry (MS), a powerful technique to investigate both G4 structures and G4/small molecules binding.[220] The number of the coordinated K^+ ions is diagnostic of the number of G-quartets involved in the G4 structures, and therefore of the G4 folded conformation.



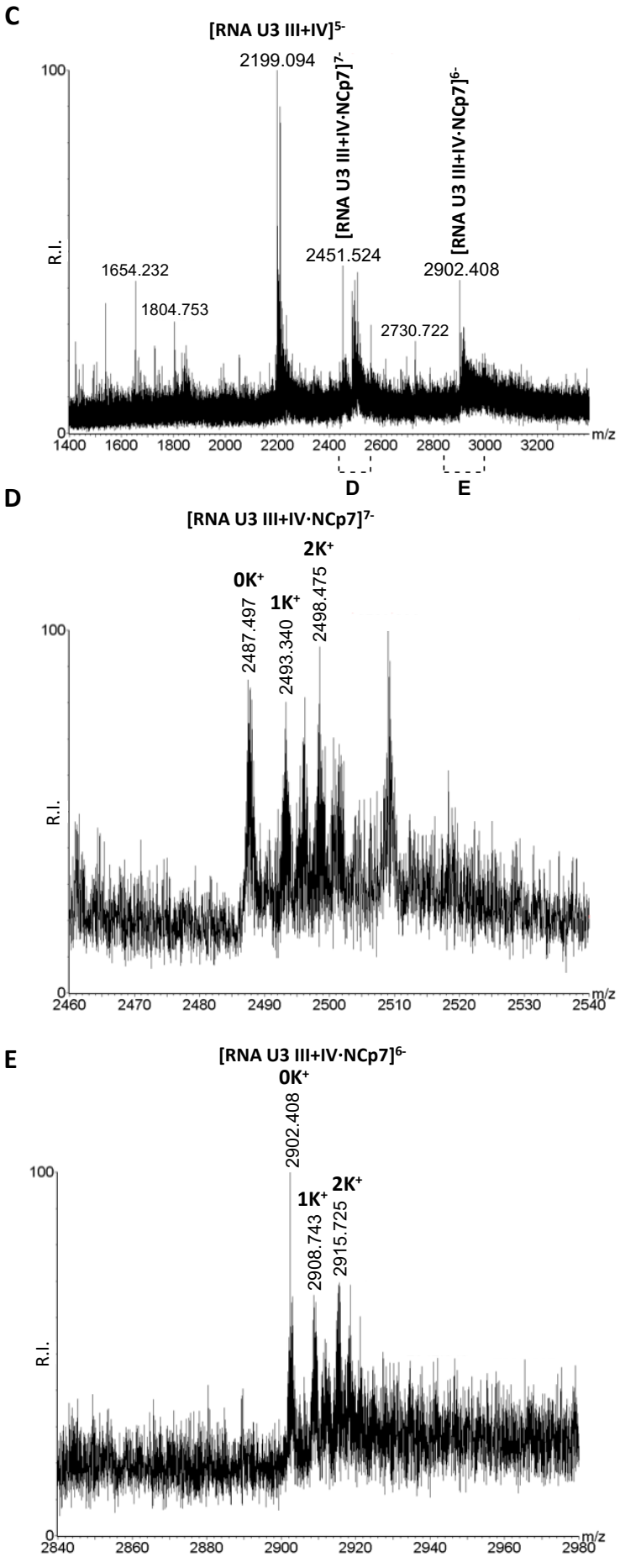

C
[RNA U3 III+IV]5-
2199.094
[RNA U3 III+IV·NCp7]7-
[RNA U3 III+IV·NCp7]6-
2451.524
2902.408
1654.232
1804.753
2730.722
100
R.I.
0
1400 1600 1800 2000 2200 2400 2600 2800 3000 3200
m/z
D
E
D
[RNA U3 III+IV·NCp7]7-
2K+
0K+
1K+
2498.475
2487.497
2493.340
100
R.I.
0
2460 2470 2480 2490 2500 2510 2520 2530 2540
m/z
E
[RNA U3 III+IV·NCp7]6-
0K+
2902.408
1K+
2K+
2908.743
2915.725
100
R.I.
0
2840 2860 2880 2900 2920 2940 2960 2980
m/z

Figure 4.4: **ESI-MS spectra of the RNA U3-III+IV in the absence or the presence of 1 equivalent of NCp7 in 0.8 mM KCl, 120 mM TMAA adjusted to pH 7.4 with TEA.** (**A**) ESI-MS spectra that show the entire sample peak distribution of the RNA U3-III+IV in the absence of NCp7. (**B**) Zoomed-in spectrum corresponding to the RNA U3-III+IV with a 5^- charge state. (**C**) ESI-MS spectra of the RNA U3-III+IV in the presence of 1 equivalent of NCp7. (**D**) Zoomed-in spectrum corresponding to the RNA U3-III+IV complexed with NCp7 with a 7^- charge state. (**E**) Zoomed-in spectrum corresponding to the oligo-protein complex with a 6^- charge state.

The MS spectrum of the RNA U3-III+IV G4 presented a peak corresponding to the oligonucleotide coordinated to two K^+ ions, which indicated the expected three-layered U3-III+IV G4 in this condition (Figure 4.4 A, B). In the presence of NCp7, two additional peaks corresponding to the oligo-protein complex appeared (Figure 4.4 C). In the complex, the species with 0 and 1 K^+ ions were prevalent with respect to the 2 K^+ ion species (Figure 4.4 D, E), indicating the unfolded state of the oligonucleotide in the presence and in complex with NCp7, and thus confirming the unfolding activity of NCp7 towards the HIV-1 RNA G4s.

Using different and complementary techniques (i.e., EMSA, CD, MS), we proved here that NCp7 binds single-stranded RNA G4s, also preferring G4s *vs* the duplex counterpart. Furthermore, the viral NCp7 protein effectively unfolds the G4 structures that form in the HIV-1 RNA genome, decreasing the T_m of the RNA G4 of about 24 °C. On the contrary, the G4 ligand B19 stabilizes the G4 folding, partially counteracting the destabilizing activity of NCp7.

4.1.2 NCp7-mediated unfolding of the RNA G4s promotes reverse transcriptase processivity

To investigate whether NCp7 unfolding properties could abolish the previously observed HIV-1 RNA G4-mediated RT stalling[140], we performed the RT stop assay in the presence of increasing concentrations of NCp7.

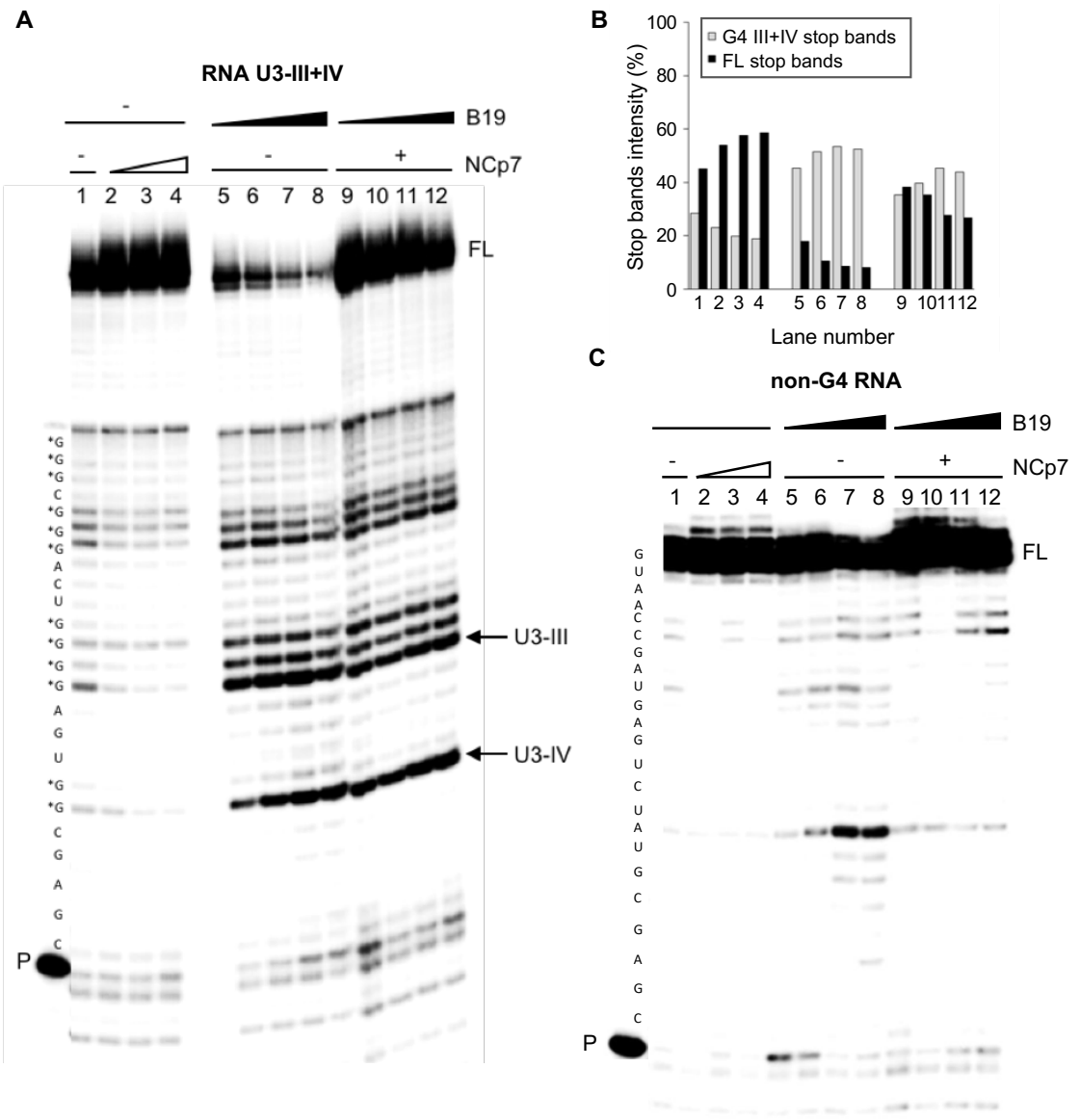


Figure 4.5: **RT stop assays on the RNA G4 U3-III+IV and non-G4 RNA control sequence.** (**A**) Templates were treated with increasing concentrations of NCp7 (350 nM - 700 nM - 1.4 μM; lanes 2-4) prior to elongation. When folded, samples were stabilized by B19 (500 nM - 1μM - 2 μM - 4 μM; lanes 5-8), elongation was performed in the absence (lanes 5-8)/presence (lanes 9-12) of NCp7 at 1.4 μM concentration. Stop bands corresponding to the U3-IV and U3-III G4 species in the RNA U3-III+IV template are indicated by arrows. P and FL indicate primer and full-length product bands, respectively. G bases potentially involved in G4 formation are labeled with asterisks. (**B**) Stop bands intensity quantification relative to the RT stop assay represented in panel (A). (**C**) RT stop assay on a control RNA sequence unable to fold into G4 structures. Each lane/sample was treated according to what previously explained for the RT stop assay in panel (A).

The U3-III+IV sequence in the presence of K^+ induced RT pausing at all G-tracts involved in the formation of the overlapping G4s (i.e., U3-III and U3-IV, Figure 4.5 A, lane 1). Upon addition of NCp7, the stop sites decreased, and the full-length RT product increased (Figure 4.5 A, lanes 2-4, and 4.5 B). When the RNA U3-III+IV G4 template was treated with increasing concentrations of B19, the stop sites significantly increased at the expense of the full-length product (Figure 4.5 A, lanes 5-8, and 4.5 B). When NCp7 was added to the B19-treated samples, the full-length

product was restored, while the stop sites were mainly maintained (Figure 4.5 A, lanes 9-12, and 4.5 B), with a visible B19 concentration-dependent effect on both the full-length product and stop sites (Figure 4.5 A, lanes 9-12, and 4.5 B). When the same reactions were performed on an RNA template unable to form G4s, no significant changes in stop sites and full-length products could be appreciated (Figure 4.5 B). These data indicate that NCp7 unfolds G4 structures that form in the HIV-1 RNA genome, favoring the proceeding of reverse transcription. In the absence of G4 stabilizing compounds, NCp7 is able to destabilize the structures to allow complete synthesis of the DNA, while the G4 ligand B19 can in part counteract this effect.

We hypothesized that the HIV-1 nucleocapsid protein could resolve RNA G4s to allow the correct completion of the RT process *in vivo*. Several considerations brought out this idea. Specifically, hundreds of NCp7 molecules coat and protect the HIV-1 dimeric RNA genome in the virion and later assist several steps of the HIV-1 replication cycle, including reverse transcription.[162,163,165,229,233] Moreover, NCp7 displays nucleic acid chaperone activity that, during reverse transcription, facilitates the rearrangement of nucleic acids into their most thermodynamically stable structures.[164] Additionally, NCp7 has been reported to bind DNA G4s: it was able to unfold a short and synthetic monomeric DNA G4,[234] and to assemble tetramolecular G4 structures.[232] However, NCp7 ability to bind and process RNA G4s has never been presented so far.

Using different and complementary techniques, we proved here that NCp7 was able to bind and unfold the U3 G4s. In addition, we proved that NCp7 preferentially binds the G4 sequence *vs* its duplex counterpart. NCp7 has been reported to bind single-stranded nucleic acid regions preferentially.[233] Here we take this concept further and demonstrate that NCp7 binds to conformationally structured single-stranded regions, such as G4s, and unfolds them. In our case, this activity resulted in the increased formation of duplex RNA/DNA hybrid, a structure that is thermodynamically more stable than the G4, as demonstrated by competition experiments in EMSA. Similar activity has been reported for another single-stranded structured RNA in HIV-1, the TAR hairpin, which gets unfolded by NCp7 to favour annealing to the complementary sequence and thus the formation of the double-stranded molecule.[167,235]

For the first time, a viral protein is reported to unfold RNA G4s. So far, only one other viral protein, i.e., EBNA 1 of the Epstein-Barr virus, has been shown to bind to folded RNA G4s to promote viral DNA replication.[128]

We already demonstrated that the U3 RNA G4s folding inhibits RT progression *in vitro*.[190] Here, the presence of NCp7 stimulated the production of full-length amplification products by RT, as assessed in the RT stop assay. These results indicate that NCp7 is able to resolve the stable RNA G4s, favoring the proceeding of reverse transcription. In a broader perspective, NCp7, thanks to its chaperone activity, is the viral protein that allows viral retro-transcription to occur *in vivo*. Conversely, the G4 ligand B19 stabilized U3 G4s thus inhibiting RT progression at the sites of G formation. This activity resulted in inhibition of the viral life cycle at the pre-integration step,[190]

which we proposed due exclusively to inhibition of RT progression at the G4 site. We proved here by CD and RT stop assay that B19 is also able to counteract the unfolding activity of NCp7.
On the one hand, our data pointed out the strength of NCp7 as a chaperone protein that processes extremely stable G4 structures in order to allow viral retro-transcription to occur. On the other, we reported that G4 ligands possess a dual activity at the U3 RNA level: they sterically hinder RT processing of the structured RNA template, and they inhibit the chaperone activity of NCp7, which in turn assists RT activity (Figure 4.6). Thus, both activities contribute to the final effect of inhibition of the reverse transcription process, and thus of the viral life cycle. As a result, the development of innovative U3 G4-selective compounds may lead to the proposal of new antivirals against HIV-1 infections.

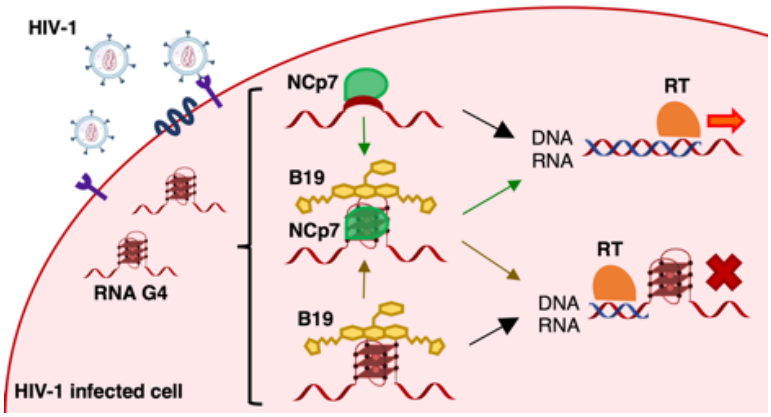


Figure 4.6: **Schematic representation of the unfolding properties of the viral NCp7 vs the stabilizing activity of the G4 ligand B19.** RNA G4 structures that fold in the U3 region of the HIV-1 genome are represented in red. The NCp7 viral protein is reported in green, whereas the G4 ligand B19 in yellow. The reverse transcriptase enzyme is colored in orange.

4.2 The HIV-1 LTR G-quadruplexes fold in latently infected cells and their stabilization by G4 ligands counteracts viral reactivation from latency

Our group has previously reported the first evidence of dynamic DNA G-quadruplex structures in the HIV-1 LTR promoter region.[140] The characterized G4s have been shown to be significant elements both for their high evolutionary conservation[140,143] and, importantly, for their regulatory activity at the promoter level during active transcription. This statement derives from two significant results. First of all, single-base mutations (i.e., m4 and m5) of the sequence of interest, which totally prevent G4 formation, produced a 2-fold increase of the LTR promoter activity in a luciferase reporter assay with respect to the control. This proved that G4s act as repressor elements in the transcriptional activation of HIV-1. Furthermore, G4 ligands (e.g., B19 and c-exNDIs)[47,190] were able to stabilize LTR G4s, thus suppressing viral transcription. Given the observed regulatory activity of the LTR G4s at the HIV-1 promoter level during productive infection, we asked ourselves what role LTR G4s could have in a latency environment.

4.2.1 The HIV-1 LTR G-quadruplexes fold in latently infected cells

In order to investigate the *bona fide* folding of the HIV-1 LTR G4 structures within a latent chromatin context, an antibody-based G4 chromatin immunoprecipitation (BG4-ChIP) assay was set up. A cell system reflecting the state of HIV-1 latency was required. Therefore, the HIV-1 chronically infected U1 cells,[236,237] containing two integrated copies of the HIV-1 proviral genome per cell, were used. Under basal (unstimulated) conditions, cells show undetectable/minimal constitutive expression of the virus because the integrated provirus is transcriptionally silent, due to a repressed state of the viral LTR promoter. U1 chromatin samples were prepared as described in Section 3.2.4, before being subjected to the BG4-mediated ChIP analysis. The FLAG-tagged single-chain variable fragment (scFv) antibody BG4 (developed in [74]) was previously employed to map the genome-wide distribution of G4s in the chromatin of the HaCaT cell line.[90] We adapted the previously published protocol to our system. ChIP assay followed by qPCR was employed to detect and assess the enrichment of LTR G4s with respect to the total sheared chromatin (input). To this aim, a primer pair spanning the LTR region that folds into a dynamic G4[140] *in vitro* was used (see Table 3.2 and Figure 4.7). As quality control, to evaluate the effectiveness of the experimental procedure, both G4-positive and negative regions were taken into account (primer pairs and probes specifications are reported in Table 3.2). Specifically, the human CDK4 (cyclin-dependent kinase 4) gene was used as a positive control because bioinformatic predictive tools reported its ability to fold into a G4 and its enrichment through a BG4-ChIP assay has already been presented.[90] Conversely, the negative control region represented by the human ESR1 (estrogen receptor 1) gene did not contain any sequence with a G4 pattern, was never reported to adopt a G4 structure, and was not detected by previously conducted BG4-ChIP study.[90]

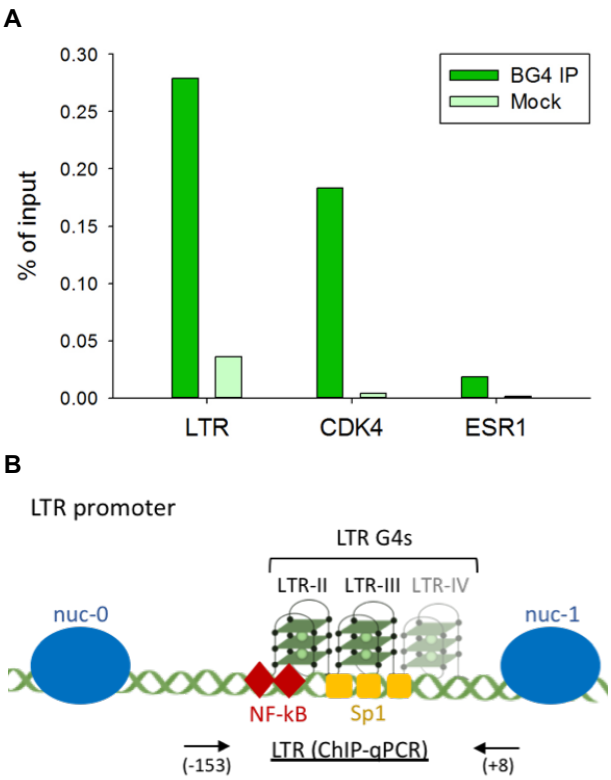


Figure 4.7: **BG4-mediated ChIP-qPCR analysis of G4 structures in latently infected U1 cells chromatin.** (**A**) Image of a typical BG4-ChIP-qPCR analysis. DNA fragments obtained from input (total sheared chromatin), immunoprecipitated (BG4 IP), and mock (control) were subjected to qPCR analysis. Displayed values are calculated from the C_t of the samples over the input, normalized for the input dilution factor and reported as percentages of enrichment. (**B**) Schematic representation of the LTR promoter region. LTR-specific primers amplify a 161 bp length region which includes the previously characterized G4 structures (i.e., LTR-II, LTR-III, and LTR-IV).

First of all, ChIP-qPCR results (Figure 4.7) showed a clear enrichment in the CDK4 G4-positive region, while a negligible signal was detected for the negative control ESR1. These observations validated the goodness of the developed BG4-ChIP protocol. Importantly, the analysis highlighted an evident enrichment in the HIV-1 LTR promoter area comprising the LTR G4 structures. This is the first time the authentic folding of the viral LTR G4 structures has been demonstrated within the context of HIV-1 latent chromatin, where the viral promoter is in a repressed state.

4.2.2 The HIV-1 LTR G4s stabilization by G4 ligands counteracts viral reactivation from latency

Given the proved LTR G4s folding in a transcriptionally inactive state of the HIV-1 promoter and their above-mentioned observed regulatory activity at the promoter level during lytic infection (i.e., negative regulators of transcription), then we wondered if they could modulate viral transcription reactivation from latency.

To this purpose, the chronically infected U1 cells have been employed once again as a model of HIV-1 latency. Different agents[236–238] have been reported to stimulate virus expression in this cell line. Here the potent PKC agonist PMA was selected to be used as latency-reversing agent. We initially tested the HIV-1 promoter inducibility in the U1 cells following PMA stimulation (Figure 4.8). Two typically reported concentrations of the selected LRA in this cell line had been screened.[239] Cells reactivation from latency was determined by the quantification of the Gag-derived p24 HIV protein using a p24 ELISA assay.

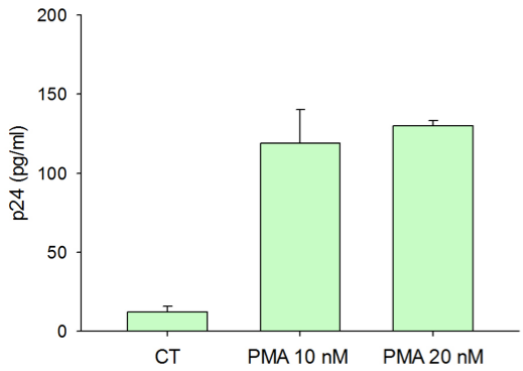


Figure 4.8: **Measurement of HIV-1 production in the U1 cell line after PMA stimulation.** Cells were treated with 10 nM and 20 nM concentrations of PMA. Supernatants were collected 48 h post-stimulation and subjected to the p24 ELISA assay. Data are represented as mean ± standard deviation (n=2).

The unstimulated (DMSO-treated) control reported only minimal constitutive expression of the viral p24 protein. On the contrary, PMA-treated U1 cells showed a clear increase in the levels of

p24 expression, which reflect the HIV LTR transcriptional activation from latency. Since the PMA 10 nM is the minimum concentration of LRA that exert a good increase in the amount of the viral capsid protein in our system, then we decided to use this concentration for further experiments.
In order to evaluate the G4-mediated modulation of the viral transcription reactivation from latency, two previously validated LTR G4 stabilizing compounds (i.e., B19 and H-NDI-NMe_2-PhAm)[47,190] were used. Figure 4.9, panel A graphically represent and simplify the work-flow applied in this study. Briefly, U1 cells were treated with different non-toxic concentrations of both B19 (from 0.39 to 6.25 μM; see Figure 4.9, panel B) and H-NDI-NMe_2-PhAm (from 0.016 to 10 nM, see Figure 4.9, panel C). Treatments were carried out at two different time-points: at 24 h pre-PMA-stimulation (then right after the seeding), and at the same time of the PMA stimulation. Dark green and light green colors were used to distinguish the two times, respectively. PMA stimulation occurred 24 h after the seeding and after the second treatment time-point. In all the experiment, the G4 ligand-treated but unstimulated cells have been checked for their remanence in a latency state, which means undetectable p24. Forty-eight h after stimulation, supernatants were collected and assayed for the amount of the HIV-1 p24 protein. The bar charts represent the percentage of the p24 amount of treated and stimulated samples reported to stimulated, but untreated controls (Figure 4.9, panel B, C). Unstimulated samples were taken as negative control. The line graph was added to represent the cytotoxicity of the G4 ligands measured in 72 h-treated cells.

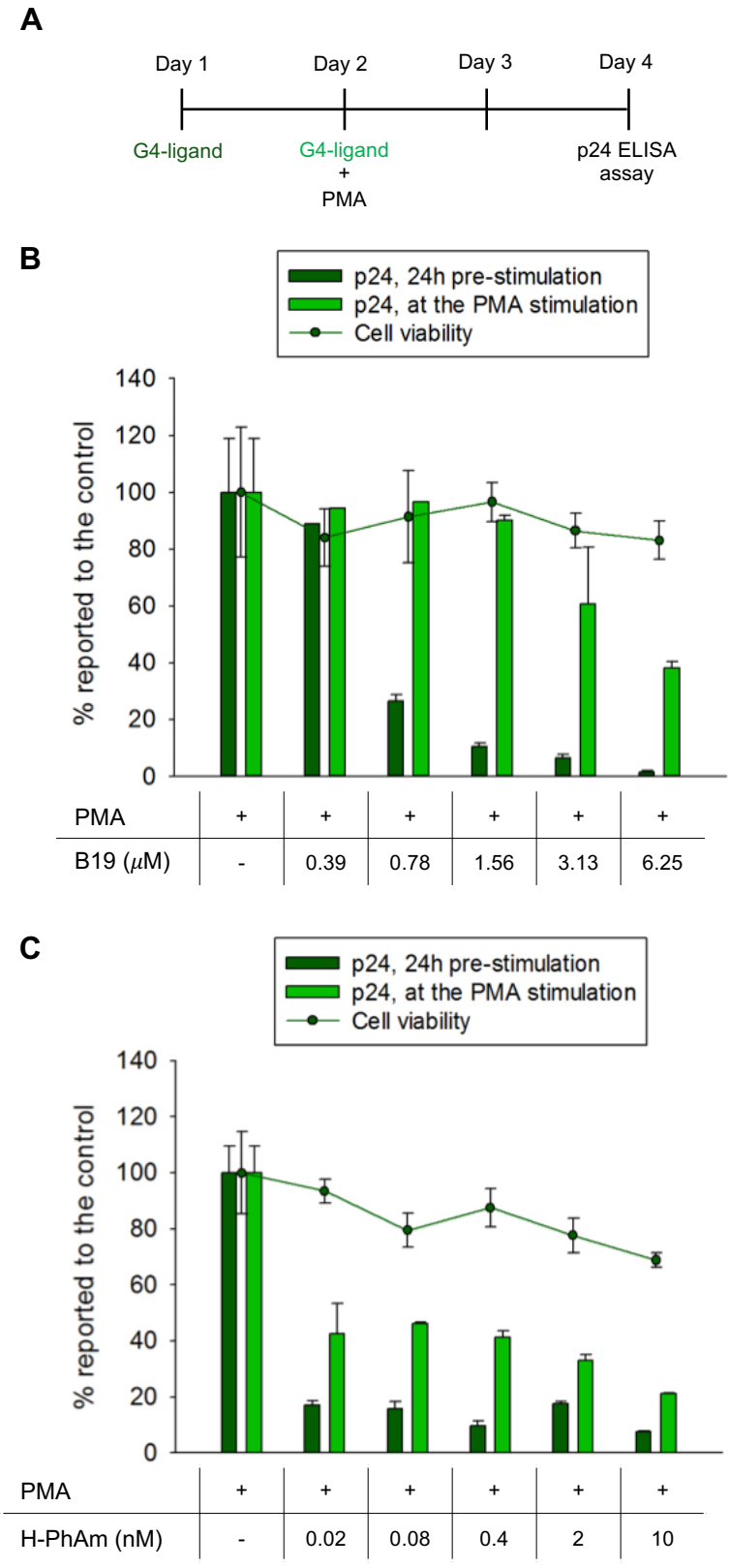


Figure 4.9: **G4 ligands-mediated modulation of viral transcription reactivation from latency.** Panel (**A**) is a schematic representation of the treatment modality applied in this study. (**B**) The bar chart represents the detected amount of the p24 viral protein at the tested B19 concentrations. Data are reported as the percentage of the amount of p24 of the treated and stimulated samples reported to the stimulated but untreated control. Dark green bars refer to the treatment occurred at 24h pre-PMA-stimulation whereas the light green bars to that at the same time of PMA stimulation. The line graph represents the cell cytotoxicity

evaluated 72 h after the B19 treatment using the ATP-lite assay. All the data are represented as mean ± standard deviation (n=2). (**C**) The H-NDI-NMe_2-PhAm mediated effects on the p24 amount and its relative cytotoxicity, represented in the same way as B19.

Increasing non-toxic concentrations of the two HIV-1 LTR stabilizing ligands B19 and H-NDI-NMe_2-PhAm induced a decrease in the percentage of the HIV-1 p24 protein at both processing times, with respect to the controls. Given that the amount of the HIV-1 p24 protein correlates directly with the transcriptional activation of the latent HIV-1 provirus, the G4 ligands promoted a down-modulation of the viral transcriptional activation induced by the PMA. For the first time, these data prove that the LTR G4s stabilization mediated by the G4 ligands counteracts viral reactivation from latency. This evidence confirms that HIV-1 LTR G4 structures act as repressor elements of the transcriptional activation of the viral promoter also in a latency environment as that of the chronically-infected U1 cells.

It is worth noting that in recent years efforts have been made to probe the native G4 landscape *in vivo*.[71,240] In fact, discrepancies were found in the folded state of G4s detected by different techniques (e.g., computational predictions, G4-seq, ChIP-seq).[88–90] The local chromatin context and all the associated proteins likely exert a strong influence on G4 stability and formation in cells:[71] it is therefore essential to consider these factors.
To face the afore issues, we here employed a BG4-ChIP assay focused on the investigation of the *bona fide* folding of the HIV-1 LTR G4 structures within the latent chromatin context of the U1 cells. The analysis highlighted an evident enrichment in the HIV-1 LTR portion. This result represents the first authentic demonstration that a G4 structure can fold within the HIV-1 LTR in the chromatin context of latently infected cells.
Hänsel-Hertsch and coworkers[90] have previously observed that DNA G4 formation is highly dependent on chromatin state and is frequently found in regulatory nucleosome-depleted regions in proximity to genes TSS. In a similar perspective, we proved the folding of G4s within a nucleosome-free, regulatory region near the TSS corresponding to the LTR promoter of HIV-1.
We must keep in mind that it is not clear yet how the G4 folding is influenced by the chromatin state and by the presence of epigenetic modifications.[116] These are two factors that have to be considered in the coming years. Indeed, the inactive/active transcriptional state switch of the promoter has to be evaluated. Since G4s are dependent on chromatin structure and transcriptional status, then changing the chromatin landscape would possibly cause a change in the G4 profile, and consequently affect the transcriptional output. So we are now planning to perform a BG4-ChIP experiment with the chromatin of PMA-induced U1 cells. Activation of transcription at the LTR promoter through stimulation with the latency-reversing agent PMA may or may not change the G4 profile. The other point that could be investigated is the interactions between G4s and epigenetic regulators; also considering that evidence pointed out the important role of G4s in this field.[117] Identifying epigenetic modifiers able to bind to G4s with a different pattern in the active and latent state of HIV-1 infection could lead to new perspectives.

In the second part of this work, we proved that the LTR G4 stabilization induced by the two G4 ligands produced a down-modulation of viral transcription that counteracted viral reactivation from latency. Therefore, our data demonstrate that LTR G4 structures can regulate the transcriptional activity of the HIV-1 promoter also influencing the regulation of virus latency. A recent work already tested and confirmed the B19 antiviral activity in latent HIV-1 infected cells,[241] so we detected the same effect in our system.
All this evidence spur us to further investigate the G4s activity at the LTR promoter level. The understanding of G4-mediated HIV-1 latency regulation mechanisms might help to define a new molecular target for alternative therapeutic approaches to proviral HIV infections.

4.3 Screening of a new series of small molecules as antivirals

Our research group has previously reported the development of new anti-HIV-1 compounds targeting the G4 structures in the viral LTR promoter.[47,190] Most G4 ligands are characterized by high molecular weights and protonated side chains, which may lead to bioavailability problems during biopharmaceutical characterization and clinical trials (e.g., Quarfloxin, B19).[41,42,52] In order to overcome this obstacle, our collaborators (Professor L.H. Hurley group and Reglagene Inc., Arizona) synthesized and provided us four small molecules (i.e., Quindoline-I, GQC-05, GSA-0820, and GSA-0932) (Figure 4.10) characterized by drug-like chemical features. Indeed, unlike the majority of available G4 binders, they meet the Lipinski's rule of five criteria, thus giving hope for their possible success in case of future oral bioavailability studies. Apart from GQC-05 which is an ellipticine analog,[112,242] all the other compounds are derivatives of the quindoline structure (or indolo[3,2-b]quinoline),[50,51] which is, in turn, a derivative of the natural product cryptolepine. While Quindoline-I, GQC-05, and GSA-0820 have already been characterized as G4 ligands, as they were reported to bind and stabilize the *c-myc* G4 (as reported in Section 1.1.3.1),[50,111,112] the GSA-0932 derivative has never been tested before. Here we aimed at testing the effectiveness of the small molecules as antivirals, in order to propose new drug-like compounds with an innovative G4-mediated mechanism of action to be applied in the treatment of HIV-1 infection.

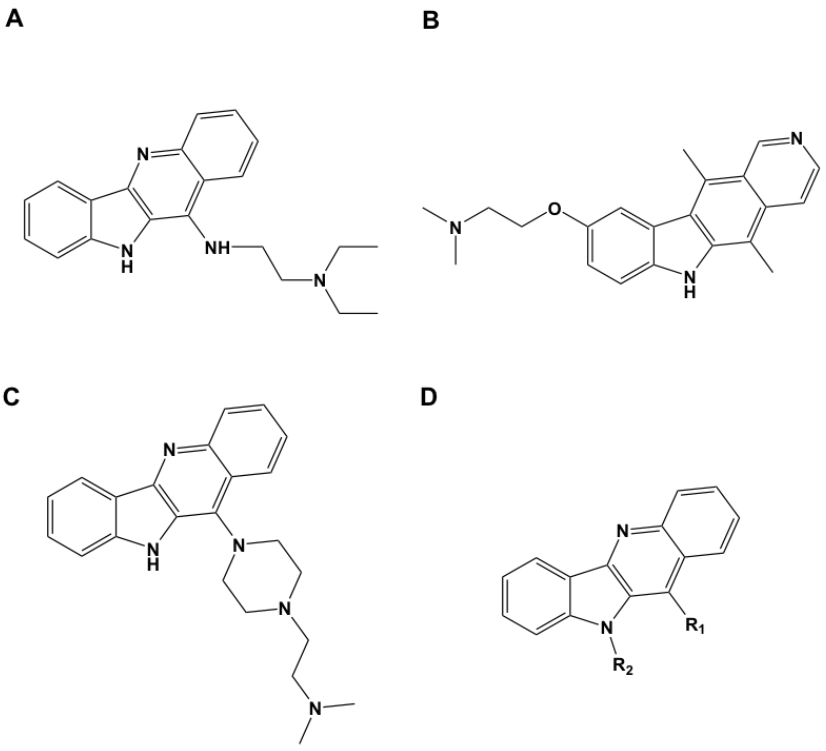


Figure 4.10: **Chemical structures of the small molecules used in this study.** (**A**) Quindoline-I; (**B**) GQC-05; (**C**) GSA-0820; (**D**) GSA-0932: it belongs to a series of Quindoline-derived compounds never published before. R_1 and R_2 indicate the presence of substituents we do not disclose because the intellectual property (IP) is patent-covered (Quindoline compounds and uses thereof, 29-Jan-2019, Serial No.: 62/798,293: Reglagene USA and UNIPD).

4.3.1 Antiviral activities of the Quindoline-derived compounds

Antiviral assays were performed in TZM-bl cells, in which HIV-1 infection drives transcription of an HIV-1 LTR-luciferase reporter gene construct. The wt HIV-1 NL4-3 strain, which is an X4 strain, was used to infect cells. We treated infected cells with increasing concentrations of compounds, and we measured virus production in culture supernatants by quantifying the luciferase signal at 48 hours post-infection (h.p.i.). Concurrently, the cytotoxicity of the small molecules was assessed in the same conditions but on uninfected cells by the MTT assay. Antiviral activities (reported as IC_{50}) and cytotoxicity (CC_{50}) are listed in Table 4.2.

Compound	TZM-bl cell line		
	IC_{50} (nM)	**CC_{50} (nM)**	**SI**
Quindoline-I	0.241	0.236	<1
GQC-05	0.242	0.309	1.28
GSA-0820	1.309	2.519	1.92
GSA-0932	0.635	1.676	2.64

Table 4.2: **Anti-HIV-1 activity and cytotoxicity of the Quindoline-derived compounds.** IC_{50} is the compound concentration required to inhibit 50% of the HIV-1 production; CC_{50} is the compound concentration at which 50% of cell cytotoxicity is observed; SI is the selectivity index calculated as the ratio of CC_{50}/IC_{50}.

The IC_{50} and CC_{50} values obtained for each compound were very similar to each other. The derived selectivity indexes were approximately equal to 1. We concluded that none of the tested compounds showed antiviral effects against HIV-1.

Since Quindoline-derived compounds were already proved to bind G4s,[50,111,112] and considering i) the impressively high number of G4 structures identified in the genome of the HSV-1,[127,215] ii) their massive presence during the viral life cycle,[214] and iii) the possibility to target them to obtain considerable antiviral effects,[127,213] we deemed it worth testing the compounds against HSV-1. Antiviral and cytotoxicity tests were conducted in parallel on U-2 OS infected and uninfected cells, respectively. The HSV-1 strain F (MOI of 1) was used to carry out infections. After that, we treated infected cells with increasing concentrations of compounds and supernatants were collected 24 h.p.i. Viral particles in the supernatants were quantified through the plaque reduction assay. Cytotoxicity was evaluated by the MTT assay. All the compounds displayed a considerable antiviral activity against HSV-1 in the nanomolar range (Table 4.3). GSA-0932 showed a pronounced antiviral activity (IC_{50} = 165.0 nM) together with low cytotoxicity (CC_{50} = 19395.3 nM), thus resulting the most promising compound, with high selectivity index (SI = 117.5). Given that this previously unreported compound showed the most encouraging antiviral effect, seven other derivatives of this new series of compounds were synthesized with different substituents (mainly in position 6 of the Quindoline core) (Figure 4.10, D) to assess the best anti-HSV-1 molecule. The obtained IC_{50}, CC_{50}, and SI are listed in Table 4.3.

Compound	U-2 OS cell line		
	IC_{50} ± SD (nM)	**CC_{50} ± SD (nM)**	**SI**
Quindoline-I	155.9 ± 6.7	2200.2 ± 489.3	14.1
GQC-05	267.4 ± 21.8	1654.6 ± 118.5	6.2
GSA-0820	>400	2984.8 ± 171.2	<7.5
GSA-0932	165.0 ± 49.1	19395.3 ± 738.0	117.5
GSA-0825	356.9 ± 60.9	32709.2 ± 313.8	91.6
GSA-0903	>400	5205.8 ± 492.8	<13.0
GSA-0920	>400	3620.6 ± 92.1	9.1
GSA-1202	248.0 ± 64.4	4852.9 ± 261.3	19.6
GSA-1502	>400	14758.5 ± 1067.4	<36.9
GSA-1504	308.7 ± 82.2	4509.6 ± 516.8	14.6
GSA-1512	255.6 ± 114.3	18089.5 ± 389.8	70.8

Table 4.3: **Anti-HSV-1 activity and cytotoxicity of the Quindoline-derived compounds.** IC_{50} is the compound concentration required to inhibit 50% of the HIV-1 production; CC_{50} is the compound concentration at which 50% of cell cytotoxicity is observed; SI is the selectivity index calculated as the ratio of CC_{50}/IC_{50}. SD stands for standard deviation. An intermediate line separates the two analyzed subgroups.

Above all, GSA-0932 remained the one with the best SI value. A representative antiviral activity profile of it is shown in Figure 4.11.

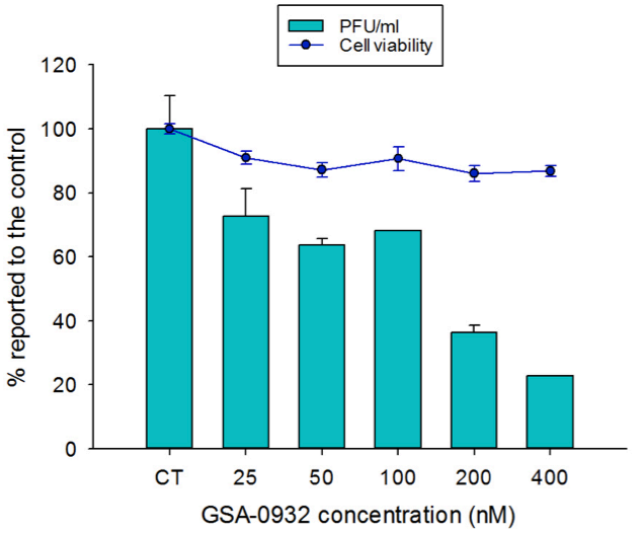


Figure 4.11: **Antiviral activity of the hit compound GSA-0932 on HSV-1-infected U-2 OS cells.** Infected cells were treated with increasing concentrations of the compound. Supernatants were collected 24 h.p.i., and the viral particles were measured with the PRA. Antiviral activity is expressed as the percentage of the plaque forming units per milliliter (PFU/ml). The infected but untreated sample was taken as control. Cell viability was in parallel obtained with an MTT assay on uninfected cells treated in the same conditions. Histograms represent the antiviral activity of the tested compound while the line indicates the cytotoxic effect.

We thus selected this compound GSA-0932 for further investigation.

4.3.2 The Quindoline-derived best candidate greatly stabilizes G4s with a preference for the viral HSV-1 conformations *vs* the telomeric sequence

CD thermal unfolding experiments were performed to investigate the binding and stabilizing activity of GSA-0932 on HSV-1 G4s. Both structural conformational changes and thermal stabilities were analyzed. Three representative and previously characterized G4s were considered: two sequences forming a four-stacked-G-quartet structure (*un2* and *gp054a*) and one forming a three-G-quartet G4 (*un3*).[127,213] Viral G4s were analyzed by CD spectroscopy at a final concentration of 4 µM in the presence of physiological K^+ concentration (100 mM) or lower K^+ concentration (2.5 mM) in order to better appreciate the effect of the compound in the case of extremely stable G4 structures (e.g., *un2* sequence). Oligonucleotides were also analyzed in the absence or presence of 4-folds excess of GSA-0932. Recorded spectra are presented in Figure 4.12.

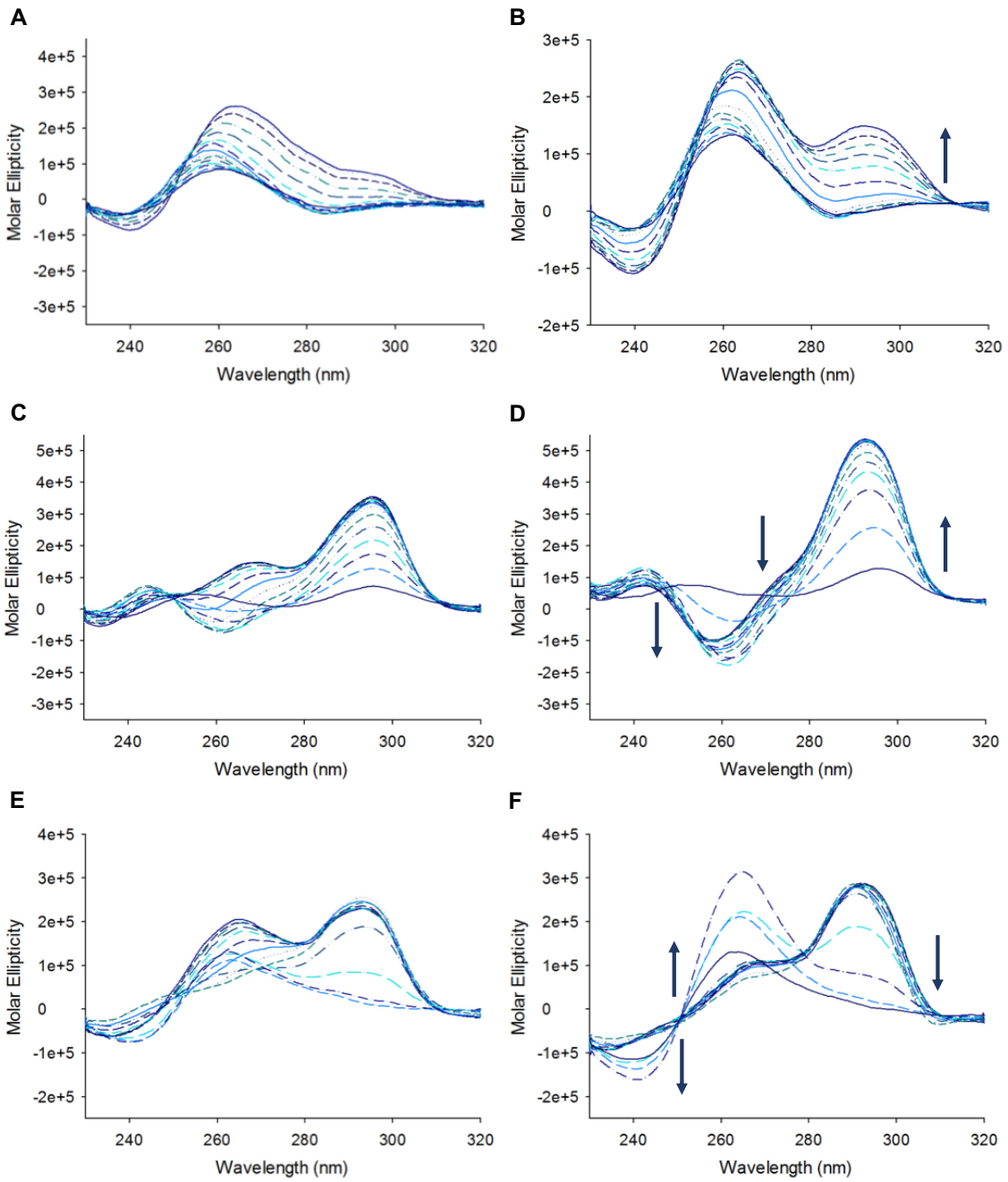


Figure 4.12: **Thermal unfolding experiments on the HSV-1 G4 sequences in the absence or the presence of 16 μM GSA-0932.** (**A**) *un3* G4 (2.5 mM K^+); (**B**) *un3* G4 (2.5 mM K^+) in the presence of GSA-0932; (**C**) *un2* G4 (2.5 mM K^+); (**D**) *un2* G4 (2.5 mM K^+) with the compound; (**E**) *gp054a* G4 (2.5 mM K^+); (**F**) *gp054a* G4 (2.5 mM K^+) with GSA-0932. Spectra were recorded over a temperature range of 20-90 °C. Changes in molar ellipticity and conformation are indicated with arrows.

In 100 and 2.5 mM K^+, the *un3* sequence showed a stable parallel topology with a maximum peak at 260 nm and a negative peak at 240 nm (Figure 4.12 A). An increase in the positive peak (shoulder) at 290 nm was registered upon addition of the compound (Figure 4.12 B). The *un2* sequence, which displays strong stability (above 90 °C at 100 mM K^+),[127] was directly analyzed at 2.5 mM K^+ to better evaluate the effect of the compound. In this condition, the sequence displayed its well-known antiparallel conformation[127,213] and a minor population with hybrid-mixed

conformation (due to the lower stability of the sequence with low concentrations of K^+) (Figure 4.12 C). GSA-0932 induced an increase in the molar ellipticity of the positive peak of *un2* at 290 nm and the negative peak at 240 nm. Additionally, the compound promoted the stabilization of the antiparallel conformation of the oligo, reducing the "shoulder" corresponding to the hybrid-mixed population of G4s (Figure 4.12 D). In 100 and 2.5 mM K^+, *gp054a* showed a typical mixed-type conformation with two positive peaks at 290 and 260 nm (Figure 4.12 E). Upon addition of GSA-0932, a slight change in the conformation of *gp054a* was observed at high temperatures (as shown in Figure 4.12 F); this is in line with the stabilization of an alternative population of G4 with a parallel conformation already seen after c-exNDI treatments.[213] Overall, thermal unfolding experiments proved that addition of the Quindoline-like compound to three peculiar HSV-1 G4s induce an increase in molar ellipticities, together with slight variations in their typical G4 conformations. Usually, high molar ellipticities indicate high stability of the G4s. To prove this point, the melting temperatures (T_m) of each sample were calculated at the peak wavelength (i.e., 264 nm for *un3*, 295.6 nm for *un2*, and 264.2 for *gp054a*). Data are reported in Table 4.4.

	2.5 mM K^+		**100 mM K^+**	
	T_m (°C)	**ΔT_m ± SD (°C)**	**T_m (°C)**	**ΔT_m ± SD (°C)**
un3	37.6±0.4		65.0	
un3+GSA-0932	56.4±1.2	18.8±0.8	66.4	1.4
un2	82.4±0.3		ND	
un2+GSA-0932	>90	>7.5	ND	ND
gp054a	57.3/82.4		58.9/>90	
gp054a+GSA-0932	74.7/>90	17.4/>7.5	74.5/>90	15.6

Table 4.4: **Stabilization (T_m) of HSV-1 G4 sequences (4 µM) at 100 and 2.5 mM K^+ in the absence/presence of GSA-0932 (16 µM).** T_m values (°C) were calculated according to the van 't Hoff equation. SD stands for standard deviation, whereas ND stands for not determined.

Addition of GSA-0932 induced an increase in the melting temperatures of all the tested HSV-1 G4s, proving its ability to effectively bind and stabilize viral G4s.

In order to further confirm the increased stability of the viral G4s in the presence of GSA-0932, a *Taq* polymerase stop assay was next set up (Figure 4.13 A, B).

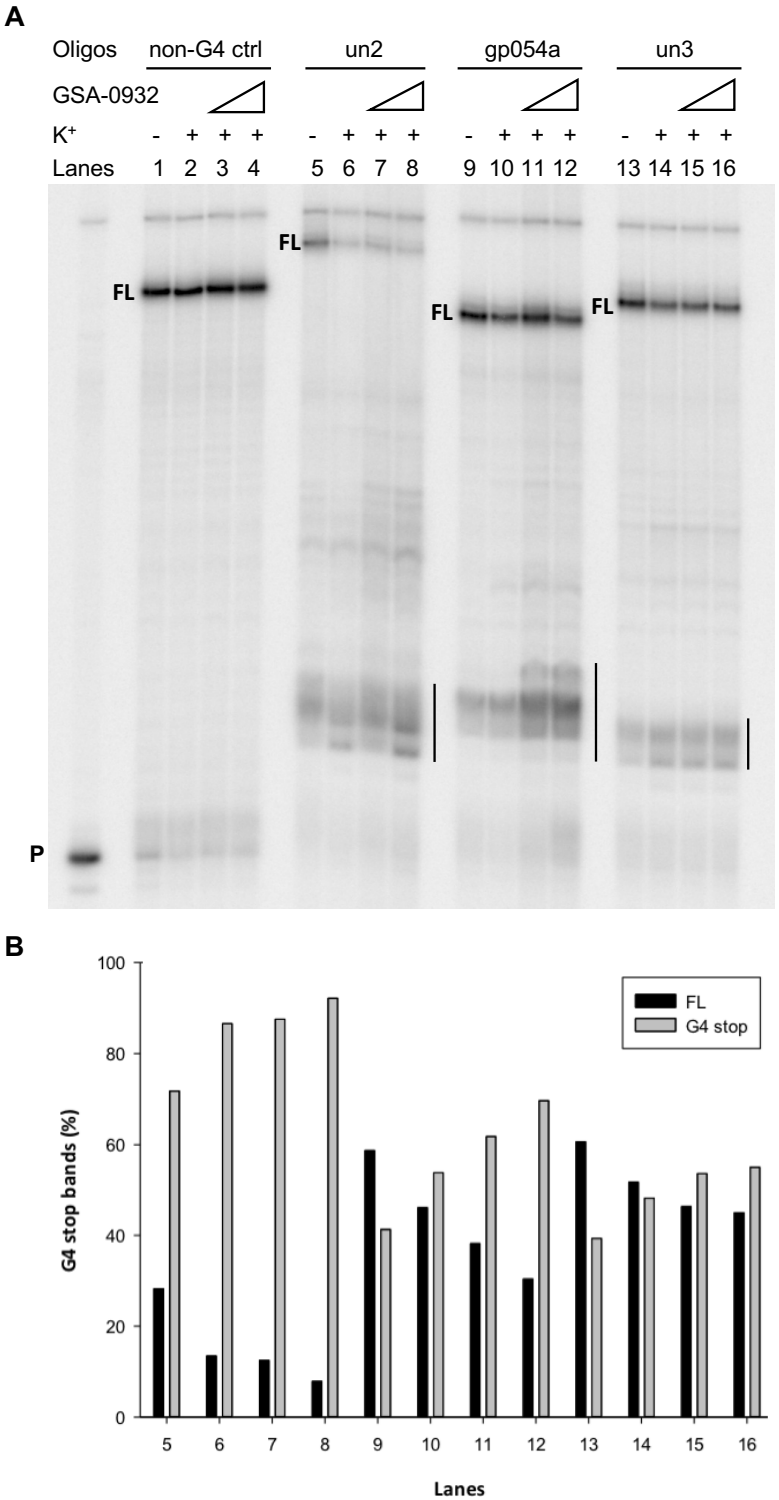


Figure 4.13: **Image of a typical *Taq* polymerase stop assay.** (**A**) The *un2*, *gp054a, un3* templates were amplified by *Taq* polymerase in the absence (lanes 5-9-13) and in the presence of K^+, combined with increasing amounts (2-8 μM) of GSA-0932 (lanes 7-8, 11-12, 15-16) or the same amount of DMSO as that in the ligand (lanes 6-10-14). *Un3* was analyzed at 50 mM of K^+, while the other sequences were investigated at 0.5 mM of K^+. A template (non-G4 ctrl) unable to fold into G4 was also used as control (lanes 1–4). P stands for unreacted labeled primer, FL stands for full-length product. G4-specific *Taq* polymerase stop sites are highlighted by vertical bars. (**B**) Full length and G4 stop bands intensities quantification relative to the *Taq* polymerase stop assay represented in panel (A).

Extended *un2*, *un3* and *gp054a* forming sequences were used, containing additional flanking bases at the 3′-end: a primer annealing sequence and a 5-T linker region to separate the annealing sequence from the first G of the G-tract. In the absence of K^+ (Figure 4.13, lanes 5-9-13), all three G4-forming sequences, especially *un2* (Figure 4.13, lane 5), displayed a marked stop site corresponding to the first G of the most 3′ G-tract, indicating stable G4 folding. In the presence of K^+ (Figure 4.13, lanes 6-10-14) the G4-stop site increased in all templates, indicating that K^+ stimulates G4 folding and thus inhibits of polymerase progression. *Un3* was analyzed at 50 mM of K^+, while the other sequences were investigated at 0.5 mM of K^+ due to their higher stability. Upon addition of increasing amounts (2-8 μM) of GSA-0932, the intensity of the stop bands increased in a concentration-dependent manner in all templates (Figure 4.13, lanes 5-9-13), along with the reduction of the full-length amplicons, thus corroborating the effective stabilization of the G4s by the compound. Quantification of the G4-stop sites and full-length products is reported in Figure 4.13, panel B. In contrast, the compound did not affect a control DNA template unable to fold into G4, indicating that the observed polymerase inhibition was G4-dependent. The above techniques (CD and *Taq* polymerase stop assay) proved the ability of GSA-0932 to efficiently bind and stabilize HSV-1 G4 structures.

In order to check the selectivity of the candidate compound for HSV-1 G4 sequences, competition electrospray ionization mass spectrometry (ESI-MS) experiments were performed. CD analysis previously confirmed that in the buffer conditions used for ESI/MS experiments the topology of the G4-folded oligonucleotides did not significantly modify.[243] Then, the viral *un3*, *un2,* and *gp054a* G4s were analyzed against two selected competitors: *hTel*, the G4-forming sequence of the human telomeric repeat, the most abundant cellular G4, and c-*myc*, one of the most studied G4s present in the *c-myc* oncogene promoter.[244,245] All the oligonucleotides were diluted to a final concentration of 4 μM and incubated with the compound at ratio DNA:compound 1:1.5 overnight at 4 °C. Samples were analyzed by a Xevo G2-XS QTof mass spectrometer. Binding affinities were calculated as specified in Section 3.3.9 for each experiment using the peak intensity for each species measured by MassLynx V4.1. Relative binding affinities are reported in Table 4.5.

Competing G4s	*un3*	*un2*	*gp054a*	Cell G4 (*hTel* or *c-myc*)
un3/*hTel*	50.3±1.2			13.8±0.17
un3/*c-myc*	37.1±0.77			59.5±1.1
un2/*hTel*		38.9±1.70		11.1±0.36
un2/*c-myc*		29.6±1.9		46.2±1.59
gp054a/*hTel*			57.3±0.43	ND
gp054a/*c-myc*			51.8±0.42	45.6±0.75

Table 4.5: **Relative binding affinity analyzed by MS competition assay for *un3*, *un2*, *gp054a*, *hTel*, and *c-myc* G4 oligonucleotides.** ND stands for not determined (due to a technical problem in correctly assigning MS peaks to the respective samples).

The reported analysis clearly showed that our compound GSA-0932 preferentially bound the viral G4s *un2*, *un3*, and *gp054a* over the telomeric G4 sequence (*hTel*). However, the compound preferred the cellular *c-myc* G4 over the viral ones. This is line with previously reported data[50,111,112] that propose the Quindoline-derived compounds as selective for the *c-myc* G4. Overall, these data suggested a general, even though not absolute, selectivity of GSA-0932 toward HSV-1 G4s. What have to be taken into account is the binding selectivity of ligands for G4s with different abundances within cells. Indeed, an infected cell may contain a lot of telomeric (hTel) G4s, many viral G4s (> 300 PQSs identified in the HSV-1 genome),[125,212] and few *c-myc* G4s. Therefore, the antiviral activity of the Quindoline-derived compounds can be ascribed to both the massive presence of viral G4s during HSV-1 infection and the ability of the ligands to preferentially recognize them over the cellular ones.

4.3.3 GSA-0932 acts at early events of the viral life cycle

Time of addition (TOA) assay was performed to assess the temporally last viral step targeted by the most promising compound (Figure 4.14).[228] Since the HSV-1 replication cycle proceeds in a well-established chronological order, it is possible to investigate the last viral process targeted by a compound comparing its action to that of a reference drug in the time frame of replication events. The well-characterized anti-HSV drug acyclovir (ACV) was used as reference drug. HSV-1-infected U2OS cells were treated with 400 nM concentration of GSA-0932, which corresponds to 2.5-folds its IC_{50}. To compare compounds in the same range of IC_{50} folds, ACV was used at 1.6 µM, concentration corresponding to about 2.5-folds its IC_{50}. Infection was carried out at MOI of 0.5 and compounds were added at different times post-infection (corresponding to different viral cycle steps): every 2 h up to 10 h.p.i. Supernatants were then collected at 30 h.p.i. and titrated by PRA.

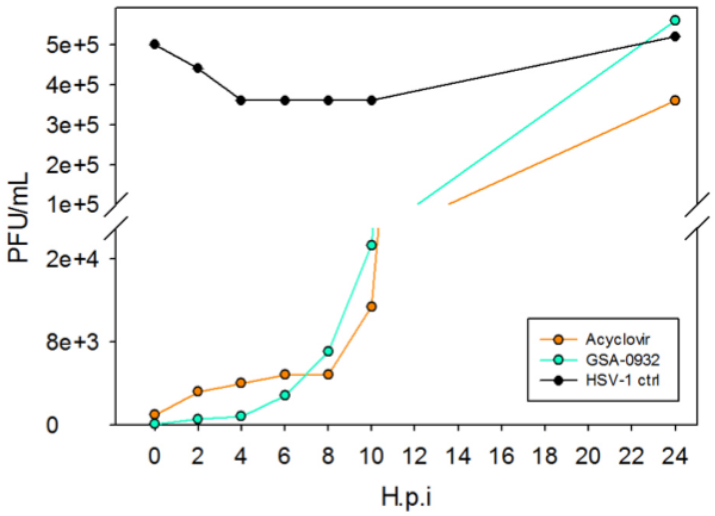
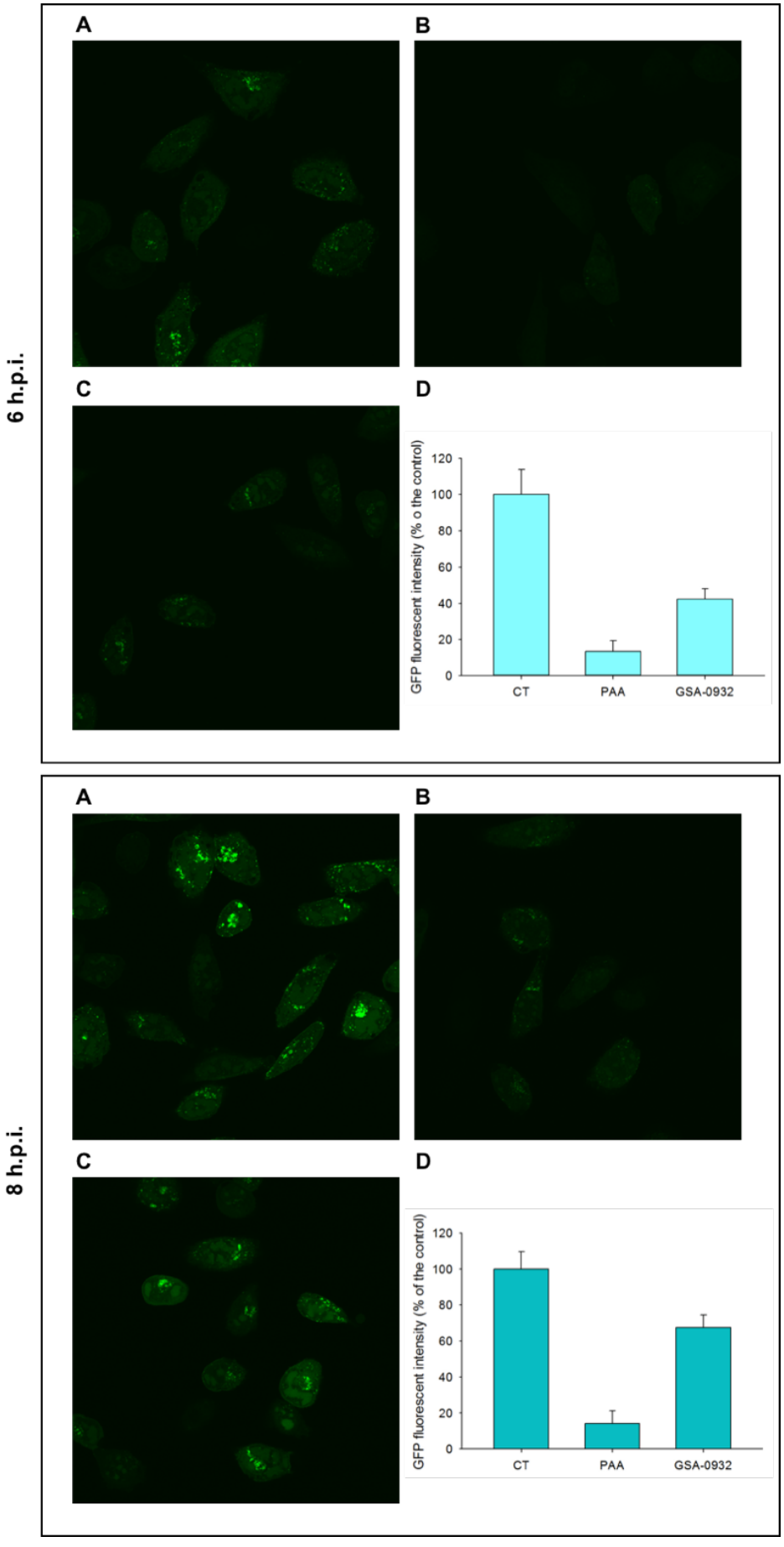


Figure 4.14: **Effect of GSA-0932 on HSV-1 cycle steps evaluated by the time of addition (TOA) assay.** GSA-0932 was tested at 400 nM (light blue line). Acyclovir was used as a reference drug and tested at 1.6 µM (orange line). Data from infected cells treated in the same conditions but without the compounds are reported in black line (HSV-1 ctrl).

Treatment of HSV-1-infected U-2 OS cells with the reference drug showed a pronounced reduction in the viral titer up to 8 h.p.i., whereas increasing levels of PFU/ml were detected after 10 h.p.i. It is worth noting that ACV targets the viral DNA polymerase during the viral DNA replication that occurs between 6-12 h.p.i.[196,207] In a very similar way, GSA-0932 kept infection controlled up to 6-8 h.p.i., while an exponential increase in the viral titer was reached from 10 h.p.i. These results indicate that the compound can target the same viral as ACV (i.e., viral replication).[228] However, one should keep in mind that the TOA allows detection of the last target in the viral cycle, which does not exclude the possibility that earlier targets are present.[228] Therefore, we can conclude that GSA-0932 acts in early events of the viral life cycle (including both pre-replicative and replicative events): it inhibits the viral replication machinery and possibly previous steps.

In order to validate the early effect of the GSA-0932 on HSV-1 infection, we also monitored the time-dependent decrease in virus production through a confocal microscopy-based approach. U-2 OS cells were plated on coverslips and infected with a recombinant HSV-1 expressing the GFP fused to the viral protein VP16 (HSV-1 v41)[225] at an MOI of 1. This mutant virus is characterized by normal replication kinetics and yields.[225] Cells were further incubated in the presence and absence of both the GSA-0932 (1 μM) and the phosphonoacetic acid (PAA; 400 nM), a specific inhibitor of virus DNA synthesis. Previously results reported that the antiviral effect of a viral DNA polymerase inhibitor can be monitored following the decrease in GFP fluorescence.[213] Emission of the GFP fluorescence was evaluated by confocal microscopy, and the reduction in the signal of the treated *vs* the untreated samples was quantified.


6 h.p.i.
A
B
C
D
GFP fluorescent intensity (% o the control)
120
100
80
60
40
20
0
CT
PAA
GSA-0932
8 h.p.i.
A
B
C
D
GFP fluorescent intensity (% of the control)
120
100
80
60
40
20
0
CT
PAA
GSA-0932

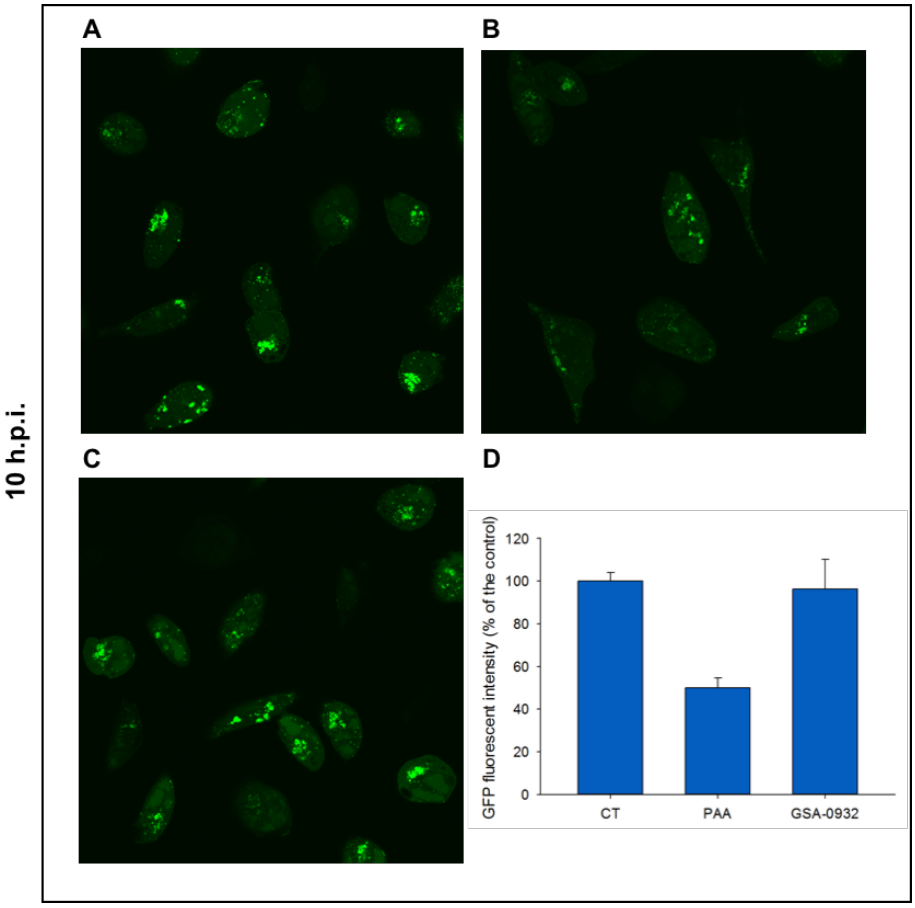


Figure 4.15: **Confocal microscopy-mediated validation of the GSA-0932 antiviral effect from 6 to 10 h.p.i**. Final images presented here are representative of multiple experiments. U-2 OS cells were seeded on coverslips, infected with the recombinant VP16-GFP HSV-1, and treated 1 h.p.i. with the tested compound. The three boxes represent the three chosen time points: 6, 8, and 10 h.p.i. For each box: (**A**) infected but untreated controls; (**B**) PAA-treated samples; (**C**) GSA-0932-treated samples; (**D**) quantification of the GFP fluorescent signal. At least three pictures per sample were quantified. Intensity values were opportunely normalized, and data were expressed as the percentage of the GFP fluorescent intensity reported to the untreated control.

Up to 4 h.p.i. no significant grouping of green particles could be discerned (data not shown). From 6 h.p.i. VP16-GFP began to accumulate in punctate foci still not very bright (Figure 4.15, 6 h.p.i. box, A). In the time between 6-10 h.p.i. (Figure 4.15, 8 and 10 h.p.i. boxes, A) these nuclear foci expanded and coalesced into large globular domains within the nucleus, which likely represented replication compartments. This is in line with what was already reported in [225]. Addition of PAA (Figure 4.15, panel B of each box) induced a very pronounced decreased in GFP fluorescent intensity at 6 and 8 h.p.i. (about 14%), whereas at 10 h.p.i. the green signal was quantified as about 50% with respect to the control. The GSA-0932 promoted a general decrease in the GFP intensity, although not as much as the PAA (Figure 4.15, panel C of each box). The reduction in green intensity was evident at 6 and 8 h.p.i. (42% and 67%, respectively), but at 10 h.p.i. the GFP quantification was almost the same as the untreated control. It has been reported that PAA induced GFP-VP16 to accumulate in a diffuse pattern of the cells, without recruitment into nuclear foci/replication compartments (in line with its inhibitory effect).[225] This was particularly evident at 10 h.p.i. (Figure 4.15, 10 h.p.i. box, B). As to GSA-0932, this could be appreciated in part at 8 but

not at 10 h.p.i. Overall, these data confirmed and expand the data obtained in the TOA: GSA-0932 acts in early events of the viral life cycle, mainly up to 8 h.p.i., including both pre-replicative and replicative events.

4.3.4 GSA-0932 colocalizes with the VP16-GFP viral protein at early stages of infection

We then decided to investigate the optical properties of the Quindoline-derived compounds in order to observe their cellular distribution and eventually their colocalization with the recombinant virus by confocal microscopy. UV-visible spectra of the GSA-0932 (25 µM) were recorded in 10 mM lithium cacodylate buffer, pH 7.4, and the wavelengths at which the compound absorbed UV-visible light at the highest rates were determined (Figure 4.16 A). These corresponded to 419, 346, 332, and 286 nm. GSA-0932 (2.5 µM) fluorescence emissions after excitation at the indicated wavelengths were also detected (Figure 4.16 B).

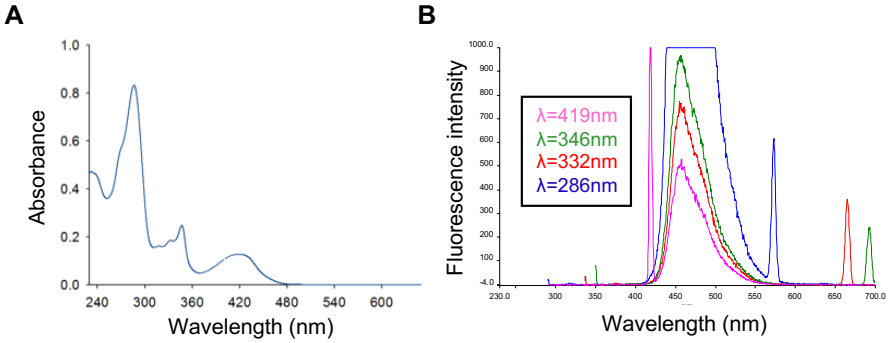


Figure 4.16: **Spectroscopic properties of the GSA-0932.** (**A**) UV-visible spectra were carried out in 10 mM lithium cacodylate buffer, pH 7.4 at a compound concentration of 25 µM using a Lambda 25 UV-Vis spectrophotometer (Perkin Elmer). (**B**) Emission spectra were performed on an LS-55 fluorescence spectrophotometer (Perkin Elmer) at a 2.5 µM compound concentration in 10 mM lithium cacodylate buffer, pH 7.4.

Thanks to its fluorescent properties, the only laser that allowed excitation of the GSA-0932 was the diode laser (λ_{ex} at 405 nm). Emission filters were λ_{em} 415-460 nm. Otherwise, to detect the GFP fluorophore the blue argon laser (λ_{ex} at 488 nm) was used, with emission filters λ_{em} 500–550 nm. Briefly, cells were seeded on coverslips, infected with the recombinant virus at an MOI of 3 for 8 or 10 h, and treated with the GSA-0932 6 µM for 2 h at 37 °C. Fixed and appropriately prepared cells were subjected to analysis. Images were captured with a Leica TCS SP5 confocal laser-scanning microscope (Leica microsystems, Germany). Final images reported here are representative of multiple experiments.

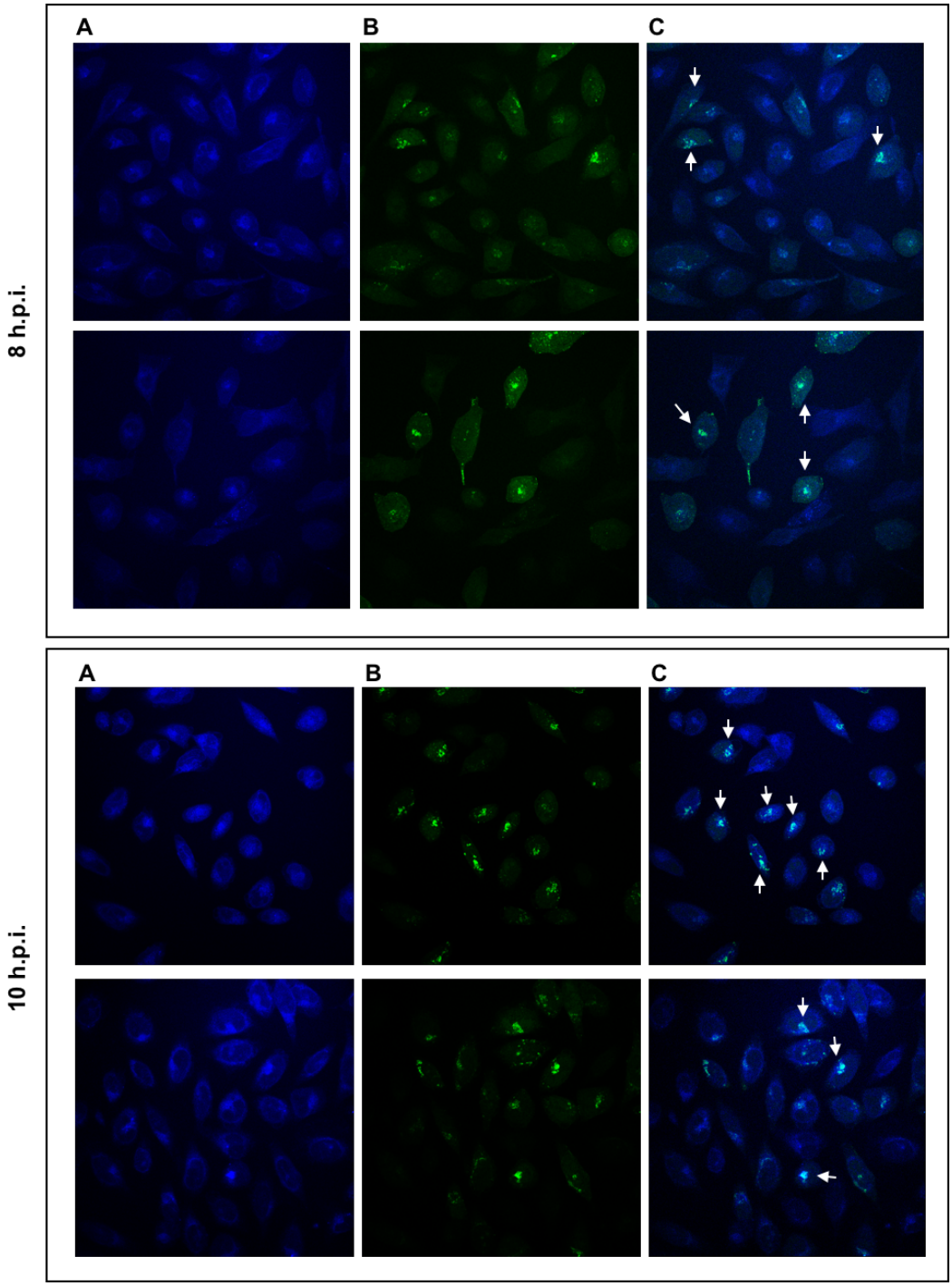


Figure 4.17: **Colocalization between GSA-0932 and the recombinant VP16-GFP virus.** Final images presented here are representative of multiple experiments. U-2 OS cells were seeded on coverslips, infected with the recombinant HSV-1, and treated 2 h with the tested compound. The two boxes represent the 8 and 10 h.p.i. time points. For each box, two captured images are shown. (**A**) GSA-0932 (6 µM) was represented in blue; (**B**) VP16-GFP HSV-1 had the green signal; (**C**) merge. Arrows indicated the colocalization areas.

Both at 8 and 10 h.p.i. we appreciated green nuclear foci which likely represented replication compartments (Figure 4.17 B), as was already seen in Figure 4.15. From pictures A of Figure 4.17, we observed that the GSA-0932 tended to accumulate in perinuclear and nuclear regions. From the merge of the two time points (Figure 4.17 C), we observed colocalization between the compound and the viral VP16-GFP virus. More precisely, the GSA-0932 accumulated in the cell

(as shown by the cell blue appearance) in the areas where we also detected the green signal corresponding to the HSV-1 replication compartments. As a consequence, the GSA-0932 significantly localized in the replication compartments at times appointed to this process. Therefore, it is possible to conclude that Quindoline-derived compounds could exert their antiviral activity targeting the viral DNA replication step. This also matches with the previous observations.

4.3.5 GSA-0932 decreases the essential viral immediate-early ICP4 protein

We reasoned that since the compound exerts its antiviral activity very early/early in infection, and considering the recently identified G4s in the IE gene promoters,[215] the GSA-0932-mediated stabilization of these G4s could inhibit IE gene expression. In order to test this hypothesis, an immunoblot assay was set up. We decided to evaluate the GSA-0932-mediated viral IE proteins expression inhibition following changes in the amount of the IE essential protein ICP4 between untreated and treated samples. ICP4 is a DNA binding protein considered as the major viral transcription factor thanks to its regulatory role in the viral gene expression.[246,247]

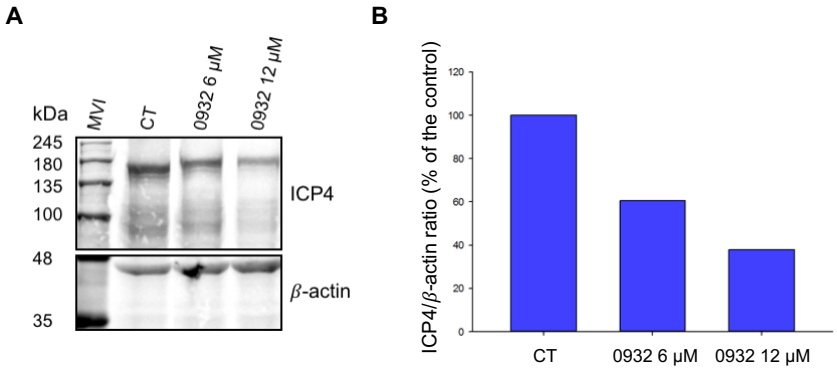


Figure 4.18: **Image of a typical immunoblot assay with its relative quantification plot.** (**A**) U-2 OS-infected cells were treated or not with increasing concentrations of GSA-0932 (6-12 µM). Specific primary antibodies were used to detect two proteins: the viral IE ICP4 protein and the human housekeeping β-actin. (**B**) Quantification of the bands from panel (A). ICP4 contents were normalized to the housekeeping. Data are presented as the percentage with respect to the untreated control.

We observed a reduction in the amount of the viral ICP4 protein in GSA-0932-treated samples (Figure 4.18 A), whereas the housekeeping protein remained essentially constant. The relative quantification (Figure 4.18 B) indicated a reduction of ICP4 to about 36% in the 12 µM-treated sample, with respect to the untreated control. These results confirmed that the GSA-0932 treatment induces inhibition of the IE proteins expression through its binding and stabilizing activities toward the G4 structures in the IE genes promoters.

The possibility to target G4 structures in order to obtain new antivirals with an innovative mechanism of action has been proposed for several viruses.[47,127,190,213] However, most of the already presented G4 ligands are characterized by large flat scaffolds that make them poorly druggable,[52,248] which is why it is necessary to discover small molecules able to bind G4s. The drug-likeness of the Quindoline-derived compounds has given us hope in their potential.

All the compounds displayed considerable anti-HSV-1 (but not anti-HIV-1) activity in the nanomolar range with low cytotoxicity, resulting in promising selectivity indexes.

Through biophysical techniques (CD and *Taq* polymerase stop assay), we definitely proved the ability of the tested compounds to efficiently bind and stabilize HSV-1 G4 structures. These observations confirmed what was already reported for some of the compounds, that is their ability to potently stack on G4-tetrads.[50]

In addition to the drugability, another common problem with G4 ligands is their insufficient specificity against different G4s.[248,249] In order to avoid side effects, it would be appropriate to generate G4 binders selective for viral G4s over cellular ones. That is why subsequent studies of ESI-MS binding affinity were conducted, and they revealed a general selectivity of the Quindoline derivatives toward HSV-1 G4s *versus* the telomeric G4s (the most abundant within cells). Their lack of optimal recognition by the compound is likely the main reason for the observed low cellular toxicity at effective antiviral doses. Therefore, we could ascribe the compounds antiviral activities to the already reported massive presence of HSV-1 G4s in the nucleus,[214] which represents the main key factor for the promising use of G4 ligands against HSV-1 infections, and to the high affinity of this class of compounds for the viral G4s.

TOA and confocal microscopy analysis were next set up to investigate the mechanism of action of the Quindoline-derived compounds thoroughly. We unambiguously proved that GSA-0932 acted in early events of the viral infection, precisely in a time window restricted up to 8 h.p.i. Given that GSA-0932 overlapped the TOA profile of ACV and it also significantly localized in the replication compartments at times appointed to this process, it is reasonable to conclude that it targets the viral replication step. Generally, this includes both direct inhibition of the replication machinery and the reduction of function in *trans*, such as the activity of promoters/the expression of genes that are involved in virus DNA replication. In fact, in the analyzed time window, G4 structures could control both viral DNA replication but also transcription events.[196] Moreover, stable G4 structures are also distributed throughout the viral genome and embedded in the promoters and coding regions of fundamental genes (especially IE genes),[127,215] and their stabilization by a G4 ligand could result in dramatic consequences. Then, in order to investigate a possible GSA-0932-mediated inhibition of IE proteins expression, an immunoblot assay was performed. We observed a reduction in the essential viral IE ICP4 protein. Therefore, the Quindoline-derived compounds could also inhibit IE genes expression by binding and stabilizing the DNA G4s in the IE genes promoters during transcription.

In a broader perspective, since HSV-1 IE proteins promoted the expression of E genes, which in turn are involved in the viral DNA replication,[207,247,250] the antiviral effects mediated by the compounds could derive from this repressive mechanism. Overall, for this new class of compounds, we proposed an antiviral mechanism of action mediated both by the inhibition of gene expression and of viral DNA replication; or at least the first target also affects the replication step (Figure 4.19). Generally, as it has been postulated for the c-exNDI,[213] we could say that the potent effect of the Quindoline-like compounds on the HSV-1 genome is likely sensed at a more extended level with respect to that of the B19.

Since all compounds of the family possess promising chemical features and potent antiviral activities, this makes them suitable for future development as novel anti-HSV-1 G4 ligands with more drug-like characteristics. For what has been said so far, the Quindoline-derived compounds have been recently subjected to U.S. patent coverage (Quindoline compounds and uses thereof, 29-Jan-2019, Serial No.: 62/798,293: Reglagene USA and UNIPD).

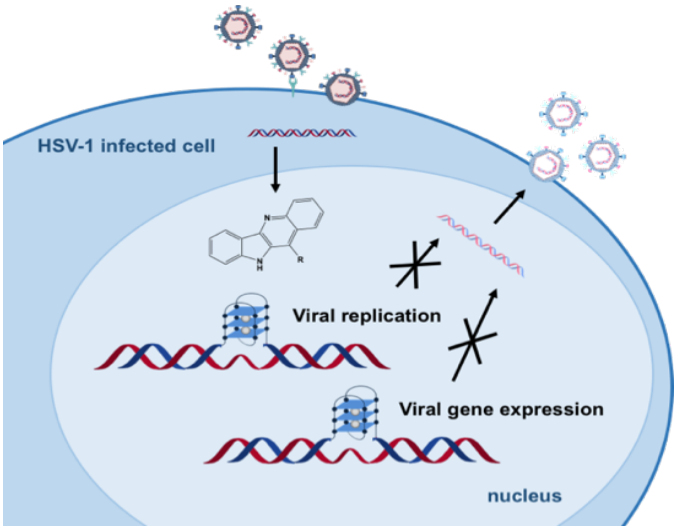


Figure 4.19: **Schematic representation of the proposed anti-HSV-1 mechanism of action of the small molecules Quindoline-derived compounds.**

5. Conclusions

The first part of this thesis focused on the identification of innovative HIV-1 RNA G4-binding proteins. For the first time, we reported the ability of the HIV-1 NCp7 protein to interact with the viral RNA U3 G4s. This evidence not only advances our knowledge on the RNA G4 interactome but also allows us to hypothesize a biological relevance linked to this interaction. Indeed, we proved that the NCp7 binding induced the unfolding of the extremely stable RNA G4s, thus favouring the completion of the reverse transcription process. In the opposite way, the G4 ligand B19-mediated stabilization of the RNA U3 G4s inhibited RT progression also counteracting the unfolding activity of NCp7. Inhibition of NCp7 is an additional and previously unknown activity of G4 ligands. Then, increasing the stability of the RNA G4 structures could be a valuable strategy to decrease the unfolding capacity of NCp7 and thus further inhibit RT progression. Accordingly, we can highlight the need for G4 ligands with improved U3 G4 selectivity and stabilizing activity that will likely allow exploiting inhibition of both NCp7 and RT to the fullest extent. This may lead to the development of anti-HIV-1 drugs with new targets and mechanism of action.

In the second part of the work, the folding and regulatory roles of the HIV-1 LTR G4s within the viral latency environment were investigated. Through a BG4-mediated ChIP-qPCR assay, we reported the authentic demonstration of LTR G4 folding in the chromatin context of latently infected cells. It is worth noting that the formation of G4s within the LTR promoter during virus latency has never been reported so far. This evidence spurred us to in-depth investigate the role of HIV-1 LTR G4s in the regulation of viral latency. Then, we proved that the LTR G4s stabilization induced by two G4 ligands (B19 and the c-exNDI) produced a down-modulation of the viral transcription that counteracted viral reactivation from latency. Targeting G4s to control viral latency had been a possibility so far only hypothesized. Now, although further studies are needed to unravel the correlation between G4 and latency, the design of selective HIV-1 LTR G4 ligands could allow the development of innovative antiretroviral drugs. The resultant inhibition of HIV-1 transcription could establish a state of deep latency that is refractory to viral reactivation, thereby suppressing the residual viremia that arises from reactivation of latently infected cells or ongoing viral replication. As a result, LTR G4s stabilizing agents could become part of the so-called “block and lock” approach to managing HIV-1 infections.

Aiming at the identification of new G4 ligands to be proposed as antivirals, we had to circumvent one of the major obstacle of the so far identified compounds, namely their poor physicochemical properties that make them inadequate for *in vivo* studies. As a result, we focused on the screening of a series of small drug-like Quindoline-derived molecules. Even if none of the Quindoline-like compounds displayed antiviral activity against HIV-1, they exert a nanomolar range antiviral activities against HSV-1, with negligible cytotoxicity. Biophysical assays proved their ability to selectively bind and stabilize HSV-1 G4s over the telomeric most abundant G4s in cells. This

evidence suggests that the antiviral activity of the Quindoline-derived compounds can be ascribed to both the massive presence of viral G4s within infected cells and the ability of the ligands to preferentially recognize them over the cellular ones. Time of addition (TOA) assay and confocal microscopy analysis proved that the compounds acted in the early events of the viral life cycle (up to 8 h.p.i.), and immunoblot assay also demonstrated inhibition of the expression of an essential IE protein. Therefore, we propose an antiviral mechanism of action mediated both by the inhibition of gene expression and of viral DNA replication; or at least the first target also affects the replication step. Our data firstly confirmed the central role of HSV-1 G4s in the regulation of viral processes. Moreover, the results highlighted the importance to target them to develop new antiviral therapies with innovative mechanisms of action for the treatment of ACV resistant herpetic infections. In fact, since the emergence of ACV resistant strains may be extremely dangerous, especially in transplant and immunocompromised patients, new antivirals with a different mechanism of action are highly wished for. Considering the antiviral activities at nanomolar concentrations, the hopeful binding selectivity, the ability to hinder the viral cycle at early stages, and the promising drug-like profiles of the Quindoline-derived compounds, we proposed them suitable for future development as new G4 ligands with anti-herpetic activity.

Altogether, the results presented in this thesis not only improve our understanding of the G4-mediated regulation of virulence processes and pathogenic mechanisms but may also constitute progress from a therapeutic point of view.

6. Appendix

6.1 List of recurrent abbreviations

G4	G-quadruplex
HIV-1	Human immunodeficiency virus type 1
LTR	Long terminal repeat
NCp7	Nucleocapsid protein
RT	Reverse transcription
B19	BRACO-19
c-exNDI	Core extended naphthalene diimide
HSV-1	Herpes simplex virus type 1
ACV	Acyclovir
G4IPDB	G4 interacting protein database
NGS	Next generation sequencing
UTR	Untranslated region
TSS	Transcription start site
PIC	Pre-integration complex
AIDS	Acquired immunodeficiency syndrome
ART	Antiretroviral therapy
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
PI	Protease inhibitor
VCV	Valaciclovir
FCV	Famciclovir
PAA	Phosphonoacetic acid
PBS	Phosphate-buffered saline
FBS	Fetal bovine serum
EMSA	Electrophoretic mobility shift assay
CD	Circular dichroism
T_m	Melting temperature
ESI	Electrospray ionization
MS	Mass spectrometry
RT	Reverse transcriptase
ChIP	Chromatin immunoprecipitation
PMA	Phorbol-12-myristate-13-acetate
LRA	Latency reversing agent

TOA	Time of addition
H.p.i.	Hours post-infection
MOI	Multiplicity of infection
GFP	Green fluorescent protein
IE	Immediate early
E	Early
L	Late
ICP4	Infected cell polypeptide 4

6.2 Index of Figures

Figure 1.1: Representation of some non-canonical DNA secondary structures compared to the B-DNA form. 1
Figure 1.2: Arrangement of guanosine residues into a G4 structure. 3
Figure 1.3: Differences between DNA and RNA G4 structures. 4
Figure 1.4: Examples of G4-compound interacting modes 5
Figure 1.5: General chemical structures of some of the most relevant G4 ligands 6
Figure 1.6: A conceptual representation of the regulatory roles of DNA and RNA G4 structures. 12
Figure 1.7: Biological roles of G4s at telomeres. 13
Figure 1.8: The six hallmarks of cancer and the association of them with the G4 structures found in their promoter regions 14
Figure 1.9: Possible functional roles of G4s during the transcription process. 15
Figure 1.10: Summary of the so far reported G4s in the viral genomes. 17
Figure 1.11: Representation of the structure of a mature HIV-1 virion. 20
Figure 1.12: Schematic representation of the seven step of the HIV-1 replicative cycle 21
Figure 1.13: Schematic representation of the HIV-1 reverse transcription. 22
Figure 1.14: HIV-1 LTR organization. 23
Figure 1.15: Aminoacid sequence of the HIV-1 nucleocapsid protein (NCp7) 24
Figure 1.16: Sterilizing and functional cure in comparison 28
Figure 1.17: G4 structures in the HIV-1 proviral genome 29
Figure 1.18: Ligands-mediated HIV-1 LTR G4s modulation and the consequent regulation of transcription 29
Figure 1.19: G4 structures in the HIV-1 RNA genome 30
Figure 1.20: Representation of the structure of an HSV-1 virion. 31
Figure 1.21: The main steps of the HSV-1 replicative cycle. 32
Figure 1.22: Organization of the HSV-1 genome 33
Figure 1.23: Representation of the most studied HSV-1 G4s 36
Figure 4.1: EMSA with the RNA G4 U3-III+IV and NCp7 51
Figure 4.2: NCp7 binding to the G4 in the presence of cold complementary DNA strand 52
Figure 4.3: CD analysis of the HIV-1 RNA G4 unfolding by NCp7 53
Figure 4.4: ESI-MS spectra of the RNA U3-III+IV in the absence or the presence of 1 equivalent of NCp7 in 0.8 mM KCl, 120 mM TMAA adjusted to pH 7.4 with TEA 56
Figure 4.5: RT stop assays on the RNA G4 U3-III+IV and non-G4 RNA control sequence 57
Figure 4.6: Schematic representation of the unfolding properties of the viral NCp7 vs the stabilizing activity of the G4 ligand B19 59
Figure 4.7: BG4-mediated ChIP-qPCR analysis of G4 structures in latently infected U1 cells chromatin 61
Figure 4.8: Measurement of HIV-1 production in the U1 cell line after PMA stimulation. 61
Figure 4.9: G4 ligands-mediated modulation of viral transcription reactivation from latency 63

Figure 4.10: Chemical structures of the small molecules used in this study 66
Figure 4.11: Antiviral activity of the hit compound GSA-0932 on HSV-1-infected U-2 OS cells. 68
Figure 4.12: Thermal unfolding experiments on the HSV-1 G4 sequences in the absence or the presence of 16 μM GSA-0932 69
Figure 4.13: Image of a typical Taq polymerase stop assay 71
Figure 4.14: Effect of GSA-0932 on HSV-1 cycle steps evaluated by the time of addition (TOA) assay 73
Figure 4.15: Confocal microscopy-mediated validation of the GSA-0932 antiviral effect from 6 to 10 h.p.i. 76
Figure 4.16: Spectroscopic properties of the GSA-0932 77
Figure 4.17: Colocalization between GSA-0932 and the recombinant VP16-GFP virus 78
Figure 4.18: Image of a typical immunoblot assay with its relative quantification plot 79
Figure 4.19: Schematic representation of the proposed anti-HSV-1 mechanism of action of the small molecules Quindoline-derived compounds 81

6.3 Index of Tables

Table 3.1: Oligonucleotide names, sequences, and their applications in the study of the NCp7-mediated unfolding of RNA G-quadruplex structures 39
Table 3.2: Names and sequences of primer pairs and probes used in the ChIP-qPCR assay .. 44
Table 3.3: Oligonucleotide names, sequences, and their applications in the screening of a new class of G4 ligands 47
Table 4.1: Melting temperatures of the samples from Figure 4.3, panel D..... 53
Table 4.2: Anti-HIV-1 activity and cytotoxicity of the Quindoline-derived compounds..... 66
Table 4.3: Anti-HSV-1 activity and cytotoxicity of the Quindoline-derived compounds 67
Table 4.4: Stabilization (T_m) of HSV-1 G4 sequences (4 µM) at 100 and 2.5 mM K^+ in the absence/presence of GSA-0932 (16 µM) 70
Table 4.5: Relative binding affinity analyzed by MS competition assay for un3, un2, gp054a, hTel, and c-myc G4 oligonucleotides 72

7. Bibliography

(1) WATSON, J. D.; CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738. https://doi.org/10.1038/171737a0.

(2) Duzdevich, D.; Redding, S.; Greene, E. C. DNA Dynamics and Single-Molecule Biology. *Chem. Rev.* **2014**, *114* (6), 3072. https://doi.org/10.1021/CR4004117.

(3) Zhou, H.; Kimsey, I. J.; Nikolova, E. N.; Sathyamoorthy, B.; Grazioli, G.; McSally, J.; Bai, T.; Wunderlich, C. H.; Kreutz, C.; Andricioaei, I.; et al. M1A and M1G Disrupt A-RNA Structure through the Intrinsic Instability of Hoogsteen Base Pairs. *Nat. Struct. Mol. Biol.* **2016**, *23* (9), 803–810. https://doi.org/10.1038/nsmb.3270.

(4) Wang, G.; Vasquez, K. M. Impact of Alternative DNA Structures on DNA Damage, DNA Repair, and Genetic Instability. *DNA Repair (Amst).* **2014**, *19*, 143–151. https://doi.org/10.1016/j.dnarep.2014.03.017.

(5) Choi, J.; Majima, T. Conformational Changes of Non-B DNA. *Chem. Soc. Rev.* **2011**, *40* (12), 5893. https://doi.org/10.1039/c1cs15153c.

(6) Saini, N.; Zhang, Y.; Usdin, K.; Lobachev, K. S. When Secondary Comes First – the Importance of Non-Canonical DNA Structures. *Biochimie* **2013**, *95* (2), 117. https://doi.org/10.1016/J.BIOCHI.2012.10.005.

(7) Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K. M. Non-B DNA Structure-Induced Genetic Instability and Evolution. *Cell. Mol. Life Sci.* **2010**, *67* (1), 43–62. https://doi.org/10.1007/s00018-009-0131-2.

(8) Bang, I. Untersuchungen Über Die Guanylsäure. *Biochem. Z.* **1910**, *26*, 293–311.

(9) GELLERT, M.; LIPSETT, M. N.; DAVIES, D. R. *Helix Formation by Guanylic Acid*; **1962**; Vol. 48.

(10) Sen, D.; Gilbert, W. Formation of Parallel Four-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis. *Nature* **1988**, *334* (6180), 364–366. https://doi.org/10.1038/334364a0.

(11) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402. https://doi.org/10.1093/NAR/GKL655.

(12) Métifiot, M.; Amrane, S.; Litvak, S.; Andreola, M.-L. G-Quadruplexes in Viruses: Function and Potential Therapeutic Applications. *Nucleic Acids Res.* **2014**, *42* (20), 12352–12366. https://doi.org/10.1093/nar/gku999.

(13) Bryan, T. M.; Baumann, P. G-Quadruplexes: From Guanine Gels to Chemotherapeutics. *Mol. Biotechnol.* **2011**, *49* (2), 198–208. https://doi.org/10.1007/s12033-011-9395-5.

(14) Sen, D.; Gilbert, W. A Sodium-Potassium Switch in the Formation of Four-Stranded G4-

DNA. *Nature* **1990**, *344* (6265), 410–414. https://doi.org/10.1038/344410a0.

(15) Largy, E.; Mergny, J.-L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. In *Metal ions in life sciences*; **2016**; Vol. 16, pp 203–258. https://doi.org/10.1007/978-3-319-21756-7_7.

(16) van Mourik, T.; Dingley, A. J. Characterization of the Monovalent Ion Position and Hydrogen-Bond Network in Guanine Quartets by DFT Calculations of NMR Parameters. *Chem. - A Eur. J.* **2005**, *11* (20), 6064–6079. https://doi.org/10.1002/chem.200500198.

(17) Campbell, N. H.; Neidle, S. G-Quadruplexes and Metal Ions. In *Metal ions in life sciences*; **2012**; Vol. 10, pp 119–134. https://doi.org/10.1007/978-94-007-2172-2_4.

(18) Chen, Y.; Yang, D. Sequence, Stability, and Structure of G-Quadruplexes and Their Interactions with Drugs. *Curr. Protoc. nucleic acid Chem.* **2012**, *Chapter 17*, Unit17.5. https://doi.org/10.1002/0471142700.nc1705s50.

(19) Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135* (13), 5017–5028. https://doi.org/10.1021/ja310251r.

(20) Malgowska, M.; Czajczynska, K.; Gudanis, D.; Tworak, A.; Gdaniec, Z. Overview of the RNA G-Quadruplex Structures. *Acta Biochim. Pol.* **2016**. https://doi.org/10.18388/abp.2016_1335.

(21) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, *429* (14), 2127–2147. https://doi.org/10.1016/j.jmb.2017.05.017.

(22) Agarwala, P.; Pandey, S.; Maiti, S. The Tale of RNA G-Quadruplex. *Org. Biomol. Chem.* **2015**, *13* (20), 5570–5585. https://doi.org/10.1039/C4OB02681K.

(23) Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *J. Med. Chem.* **1997**, *40* (14), 2113–2116. https://doi.org/10.1021/jm970199z.

(24) Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2019**, *xx*. https://doi.org/10.1016/j.trechm.2019.07.002.

(25) Neidle, S. The Structures of Quadruplex Nucleic Acids and Their Drug Complexes. *Curr. Opin. Struct. Biol.* **2009**, *19* (3), 239–250. https://doi.org/10.1016/j.sbi.2009.04.001.

(26) Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016**, *59* (13), 5987–6011. https://doi.org/10.1021/acs.jmedchem.5b01835.

(27) Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2019**. https://doi.org/10.1016/j.trechm.2019.07.002.

(28) Monchaud, D.; Teulade-Fichou, M.-P. A Hitchhiker's Guide to G-Quadruplex Ligands. *Org. Biomol. Chem.* **2008**, *6* (4), 627–636. https://doi.org/10.1039/b714772b.

(29) Yang, D.; Okamoto, K. Structural Insights into G-Quadruplexes: Towards New Anticancer Drugs. *Future Med. Chem.* **2010**, *2* (4), 619–646.

(30) Li, Q.; Xiang, J.-F.; Yang, Q.-F.; Sun, H.-X.; Guan, A.-J.; Tang, Y.-L. G4LDB: A Database

for Discovering and Studying G-Quadruplex Ligands. *Nucleic Acids Res.* **2013**, *41* (Database issue), D1115-23. https://doi.org/10.1093/nar/gks1101.

(31) Shin-ya, K.; Wierzba, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H. Telomestatin, a Novel Telomerase Inhibitor from Streptomyces Anulatus. *J. Am. Chem. Soc.* **2001**, *123* (6), 1262–1263. https://doi.org/10.1021/ja005780q.

(32) Read, M.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. Structure-Based Design of Selective and Potent G Quadruplex-Mediated Telomerase Inhibitors. *Proc. Natl. Acad. Sci.* **2001**, *98* (9), 4844–4849. https://doi.org/10.1073/pnas.081560598.

(33) Wheelhouse, R. T.; Sun, D.; Han, H.; Han, F. X.; Hurley, L. H. Cationic Porphyrins as Telomerase Inhibitors: The Interaction of Tetra- (N-Methyl-4-Pyridyl)Porphine with Quadruplex DNA [1]. *J. Am. Chem. Soc.* **1998**, *120* (13), 3261–3262. https://doi.org/10.1021/ja973792e.

(34) Fedoroff, O. Y.; Salazar, M.; Han, H.; Chemeris, V. V.; Kerwin, S. M.; Hurley, L. H. NMR-Based Model of a Telomerase-Inhibiting Compound Bound to G-Quadruplex DNA †. *Biochemistry* **1998**, *37* (36), 12367–12374. https://doi.org/10.1021/bi981330n.

(35) Collie, G. W.; Promontorio, R.; Hampel, S. M.; Micco, M.; Neidle, S.; Parkinson, G. N. Structural Basis for Telomeric G-Quadruplex Targeting by Naphthalene Diimide Ligands. *J. Am. Chem. Soc.* **2012**, *134* (5), 2723–2731. https://doi.org/10.1021/ja2102423.

(36) Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *J. Am. Chem. Soc.* **2008**, *130* (47), 15758–15759. https://doi.org/10.1021/ja805615w.

(37) De Cian, A.; DeLemos, E.; Mergny, J.-L.; Teulade-Fichou, M.-P.; Monchaud, D. Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J. Am. Chem. Soc.* **2007**, *129* (7), 1856–1857. https://doi.org/10.1021/ja067352b.

(38) Caprio, V.; Guyen, B.; Opoku-Boahen, Y.; Mann, J.; Gowan, S. M.; Kelland, L. M.; Read, M. A.; Neidle, S. A Novel Inhibitor of Human Telomerase Derived from 10H-Indolo[3,2-b]Quinoline. *Bioorg. Med. Chem. Lett.* **2000**, *10* (18), 2063–2066. https://doi.org/10.1016/s0960-894x(00)00378-4.

(39) Kim, M.-Y.; Duan, W.; Gleason-Guzman, M.; Hurley, L. H. Design, Synthesis, and Biological Evaluation of a Series of Fluoroquinoanthroxazines with Contrasting Dual Mechanisms of Action against Topoisomerase II and G-Quadruplexes. *J. Med. Chem.* **2003**, *46* (4), 571–583. https://doi.org/10.1021/jm0203377.

(40) Gowan, S. M.; Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. A G-Quadruplex-Interactive Potent Small-Molecule Inhibitor of Telomerase Exhibiting in Vitro and in Vivo Antitumor Activity. *Mol. Pharmacol.* **2002**, *61* (5), 1154–1162. https://doi.org/10.1124/mol.61.5.1154.

(41) Taetz, S.; Baldes, C.; Mürdter, T. E.; Kleideiter, E.; Piotrowska, K.; Bock, U.; Haltner-Ukomadu, E.; Mueller, J.; Huwer, H.; Schaefer, U. F.; et al. Biopharmaceutical

Characterization of the Telomerase Inhibitor BRACO19. *Pharm. Res.* **2006**, *23* (5), 1031–1037. https://doi.org/10.1007/s11095-006-0026-y.

(42) Ruggiero, E.; Richter, S. N. G-Quadruplexes and G-Quadruplex Ligands: Targets and Tools in Antiviral Therapy. *Nucleic Acids Res.* **2018**, *46* (7), 3270–3283. https://doi.org/10.1093/nar/gky187.

(43) Sissi, C.; Lucatello, L.; Paul Krapcho, A.; Maloney, D. J.; Boxer, M. B.; Camarasa, M. V.; Pezzoni, G.; Menta, E.; Palumbo, M. Tri-, Tetra- and Heptacyclic Perylene Analogues as New Potential Antineoplastic Agents Based on DNA Telomerase Inhibition. *Bioorg. Med. Chem.* **2007**, *15* (1), 555–562. https://doi.org/10.1016/j.bmc.2006.09.029.

(44) Cuenca, F.; Greciano, O.; Gunaratnam, M.; Haider, S.; Munnur, D.; Nanjunda, R.; Wilson, W. D.; Neidle, S. Tri- and Tetra-Substituted Naphthalene Diimides as Potent G-Quadruplex Ligands. *Bioorg. Med. Chem. Lett.* **2008**, *18* (5), 1668–1673. https://doi.org/10.1016/j.bmcl.2008.01.050.

(45) Di Antonio, M.; Doria, F.; Richter, S. N.; Bertipaglia, C.; Mella, M.; Sissi, C.; Palumbo, M.; Freccero, M. Quinone Methides Tethered to Naphthalene Diimides as Selective G-Quadruplex Alkylating Agents. *J. Am. Chem. Soc.* **2009**, *131* (36), 13132–13141. https://doi.org/10.1021/ja904876q.

(46) Doria, F.; Nadai, M.; Sattin, G.; Pasotti, L.; Richter, S. N.; Freccero, M. Water Soluble Extended Naphthalene Diimides as PH Fluorescent Sensors and G-Quadruplex Ligands. *Org. Biomol. Chem.* **2012**, *10* (19), 3830. https://doi.org/10.1039/c2ob07006e.

(47) Perrone, R.; Doria, F.; Butovskaya, E.; Frasson, I.; Botti, S.; Scalabrin, M.; Lago, S.; Grande, V.; Nadai, M.; Freccero, M.; et al. Synthesis, Binding and Antiviral Properties of Potent Core-Extended Naphthalene Diimides Targeting the HIV-1 Long Terminal Repeat Promoter G-Quadruplexes. *J. Med. Chem.* **2015**, *58* (24), 9639–9652. https://doi.org/10.1021/acs.jmedchem.5b01283.

(48) Che, T.; Wang, Y. Q.; Huang, Z. L.; Tan, J. H.; Huang, Z. S.; Chen, S. Bin. Natural Alkaloids and Heterocycles as G-Quadruplex Ligands and Potential Anticancer Agents. *Molecules* **2018**, *23* (2). https://doi.org/10.3390/molecules23020493.

(49) Ou, T. M.; Lu, Y. J.; Zhang, C.; Huang, Z. S.; Wang, X. D.; Tan, J. H.; Chen, Y.; Ma, D. L.; Wong, K. Y.; Tang, J. C. O.; et al. Stabilization of G-Quadruplex DNA and down-Regulation of Oncogene c-Myc by Quindoline Derivatives. *J. Med. Chem.* **2007**, *50* (7), 1465–1474. https://doi.org/10.1021/jm0610088.

(50) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline–c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **2011**, *133* (44), 17673–17680. https://doi.org/10.1021/ja205646q.

(51) Zhou, J.-L.; Lu, Y.-J.; Ou, T.-M.; Zhou, J.-M.; Huang, Z.-S.; Zhu, X.-F.; Du, C.-J.; Bu, X.-Z.; Ma, L.; Gu, L.-Q.; et al. Synthesis and Evaluation of Quindoline Derivatives as G-Quadruplex Inducing and Stabilizing Ligands and Potential Inhibitors of Telomerase. *J. Med. Chem.* **2005**, *48* (23), 7315–7321. https://doi.org/10.1021/jm050041b.

(52) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene

Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discov.* **2011**, *10* (4), 261–275. https://doi.org/10.1038/nrd3428.

(53) Brázda, V.; Hároníková, L.; Liao, J.; Fojta, M. DNA and RNA Quadruplex-Binding Proteins. *Int. J. Mol. Sci.* **2014**, *15* (10), 17493–17517. https://doi.org/10.3390/ijms151017493.

(54) Mishra, S. K.; Tawani, A.; Mishra, A.; Kumar, A. G4IPDB: A Database for G-Quadruplex Structure Forming Nucleic Acid Interacting Proteins. *Sci. Rep.* **2016**, *6* (1), 38144. https://doi.org/10.1038/srep38144.

(55) Brázda, V.; Jiří˘jiříčervě N, J.; Bartas, M.; Mikysková, N.; Coufal, J.; Pečinka, P. The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* **2018**. https://doi.org/10.3390/molecules23092341.

(56) Huang, Z.-L.; Dai, J.; Luo, W.-H.; Wang, X.-G.; Tan, J.-H.; Chen, S.-B.; Huang, Z.-S. Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *J. Am. Chem. Soc.* **2018**, *140* (51), 17945–17955. https://doi.org/10.1021/jacs.8b09329.

(57) Zaug, A. J.; Podell, E. R.; Cech, T. R. Human POT1 Disrupts Telomeric G-Quadruplexes Allowing Telomerase Extension in Vitro. *Proc. Natl. Acad. Sci.* **2005**, *102* (31), 10864–10869. https://doi.org/10.1073/pnas.0504744102.

(58) Mendoza, O.; Bourdoncle, A.; Boulé, J.-B.; Brosh, R. M.; Mergny, J.-L.; Mergny, J.-L. G-Quadruplexes and Helicases. *Nucleic Acids Res.* **2016**, *44* (5), 1989–2006. https://doi.org/10.1093/nar/gkw079.

(59) Soldatenkov, V. A.; Vetcher, A. A.; Duka, T.; Ladame, S. First Evidence of a Functional Interaction between DNA Quadruplexes and Poly(ADP-Ribose) Polymerase-1. *ACS Chem. Biol.* **2008**, *3* (4), 214–219. https://doi.org/10.1021/cb700234f.

(60) González, V.; Guo, K.; Hurley, L.; Sun, D. Identification and Characterization of Nucleolin as a C-Myc G-Quadruplex-Binding Protein. *J. Biol. Chem.* **2009**, *284* (35), 23622–23635. https://doi.org/10.1074/jbc.M109.018028.

(61) Tosoni, E.; Frasson, I.; Scalabrin, M.; Perrone, R.; Butovskaya, E.; Nadai, M.; Palù, G.; Fabris, D.; Richter, S. N. Nucleolin Stabilizes G-Quadruplex Structures Folded by the LTR Promoter and Silences HIV-1 Viral Transcription. *Nucleic Acids Res.* **2015**, *43* (18), 8884–8897. https://doi.org/10.1093/nar/gkv897.

(62) Lago, S.; Tosoni, E.; Nadai, M.; Palumbo, M.; Richter, S. N. The Cellular Protein Nucleolin Preferentially Binds Long-Looped G-Quadruplex Nucleic Acids. *Biochim. Biophys. acta. Gen. Subj.* **2017**, *1861* (5 Pt B), 1371–1381. https://doi.org/10.1016/j.bbagen.2016.11.036.

(63) Lista, M. J.; Martins, R. P.; Billant, O.; Contesse, M.-A.; Findakly, S.; Pochard, P.; Daskalogianni, C.; Beauvineau, C.; Guetta, C.; Jamin, C.; et al. Nucleolin Directly Mediates Epstein-Barr Virus Immune Evasion through Binding to G-Quadruplexes of EBNA1 MRNA. *Nat. Commun.* **2017**, *8*, 16043. https://doi.org/10.1038/ncomms16043.

(64) Cree, S. L.; Fredericks, R.; Miller, A.; Pearce, F. G.; Filichev, V.; Fee, C.; Kennedy, M. A.

DNA G-Quadruplexes Show Strong Interaction with DNA Methyltransferases in Vitro. *FEBS Lett.* **2016**, *590* (17), 2870–2883. https://doi.org/10.1002/1873-3468.12331.

(65) Hacht, A. von; Seifert, O.; Menger, M.; Schütze, T.; Arora, A.; Konthur, Z.; Neubauer, P.; Wagner, A.; Weise, C.; Kurreck, J. Identification and Characterization of RNA Guanine-Quadruplex Binding Proteins. *Nucleic Acids Res.* **2014**, *42* (10), 6630–6644. https://doi.org/10.1093/nar/gku290.

(66) Ghosh, M.; Singh, M. RGG-Box in HnRNPA1 Specifically Recognizes the Telomere G-Quadruplex DNA and Enhances the G-Quadruplex Unfolding Ability of UP1 Domain. *Nucleic Acids Res.* **2018**, *46* (19), 10246–10261. https://doi.org/10.1093/nar/gky854.

(67) Scalabrin, M.; Frasson, I.; Ruggiero, E.; Perrone, R.; Tosoni, E.; Lago, S.; Tassinari, M.; Palù, G.; Richter, S. N. The Cellular Protein HnRNP A2/B1 Enhances HIV-1 Transcription by Unfolding LTR Promoter G-Quadruplexes. *Sci. Rep.* **2017**, *7*, 45244. https://doi.org/10.1038/srep45244.

(68) Lopez, C. R.; Singh, S.; Hambarde, S.; Griffin, W. C.; Gao, J.; Chib, S.; Yu, Y.; Ira, G.; Raney, K. D.; Kim, N. Yeast Sub1 and Human PC4 Are G-Quadruplex Binding Proteins That Suppress Genome Instability at Co-Transcriptionally Formed G4 DNA. *Nucleic Acids Res.* **2017**, *45* (10), 5850–5862. https://doi.org/10.1093/nar/gkx201.

(69) Götz, S.; Pandey, S.; Bartsch, S.; Juranek, S.; Paeschke, K. A Novel G-Quadruplex Binding Protein in Yeast-Slx9. *Molecules* **2019**, *24* (9). https://doi.org/10.3390/molecules24091774.

(70) Norseen, J.; Johnson, F. B.; Lieberman, P. M. Role for G-Quadruplex RNA Binding by Epstein-Barr Virus Nuclear Antigen 1 in DNA Replication and Metaphase Chromosome Attachment. *J. Virol.* **2009**, *83* (20), 10336–10346. https://doi.org/10.1128/JVI.00747-09.

(71) Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2019**. https://doi.org/10.1016/J.TRECHM.2019.07.002.

(72) Tseng, T.-Y.; Chien, C.-H.; Chu, J.-F.; Huang, W.-C.; Lin, M.-Y.; Chang, C.-C.; Chang, T.-C. Fluorescent Probe for Visualizing Guanine-Quadruplex DNA by Fluorescence Lifetime Imaging Microscopy. *J. Biomed. Opt.* **2013**, *18* (10), 101309. https://doi.org/10.1117/1.JBO.18.10.101309.

(73) Fernando, H.; Rodriguez, R.; Balasubramanian, S. Selective Recognition of a DNA G-Quadruplex by an Engineered Antibody. *Biochemistry* **2008**, *47* (36), 9365–9371. https://doi.org/10.1021/bi800983u.

(74) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5* (3), 182–186. https://doi.org/10.1038/nchem.1548.

(75) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M.; Sen, D.; Lansdorp, P. M.; Lansdorp, P. M. Detection of G-Quadruplex DNA in Mammalian Cells. *Nucleic Acids Res.* **2014**, *42* (2), 860–869. https://doi.org/10.1093/nar/gkt957.

(76) Kwok, C. K.; Merrick, C. J.; Kwok, C. K.; Uk, C. M. A.; Merrick, C. J. G-Quadruplexes:

Prediction, Characterization, and Biological Application. *Trends Biotechnol.* **2017**, *35* (10). https://doi.org/10.1016/j.tibtech.2017.06.012.

(77) Verma, A.; Halder, K.; Halder, R.; Yadav, V. K.; Rawal, P.; Thakur, R. K.; Mohd, F.; Sharma, A.; Chowdhury, S. Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved *Cis* -Regulatory Elements in Human and Related Species. *J. Med. Chem.* **2008**, *51* (18), 5641–5649. https://doi.org/10.1021/jm800448a.

(78) Bartas, M.; Čutová, M.; Brázda, V.; Kaura, P.; Šťastný, J.; Kolomazník, J.; Coufal, J.; Goswami, P.; Červeň, J.; Pečinka, P. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. *Molecules* **2019**, *24* (9). https://doi.org/10.3390/molecules24091711.

(79) Hershman, S. G.; Chen, Q.; Lee, J. Y.; Kozak, M. L.; Yue, P.; Wang, L.-S.; Johnson, F. B. Genomic Distribution and Functional Analyses of Potential G-Quadruplex-Forming Sequences in Saccharomyces Cerevisiae. *Nucleic Acids Res.* **2008**, *36* (1), 144–156. https://doi.org/10.1093/nar/gkm986.

(80) Lavezzo, E.; Berselli, M.; Frasson, I.; Perrone, R.; Palù, G.; Brazzale, A. R.; Richter, S. N.; Toppo, S. G-Quadruplex Forming Sequences in the Genome of All Known Human Viruses: A Comprehensive Guide. *PLoS Comput. Biol.* **2018**, *14* (12), e1006675. https://doi.org/10.1371/journal.pcbi.1006675.

(81) Griffin, B. D.; Bass, H. W. Plant G-Quadruplex (G4) Motifs in DNA and RNA. *Plant Sci.* **2018**, *269*, 143–147. https://doi.org/10.1016/j.plantsci.2018.01.011.

(82) Frees, S.; Menendez, C.; Crum, M.; Bagga, P. S. QGRS-Conserve: A Computational Method for Discovering Evolutionarily Conserved G-Quadruplex Motifs. *Hum. Genomics* **2014**, *8* (1), 8. https://doi.org/10.1186/1479-7364-8-8.

(83) Kikin, O.; D'Antonio, L.; Bagga, P. S. QGRS Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Res.* **2006**, *34* (Web Server issue), W676-82. https://doi.org/10.1093/nar/gkl253.

(84) Yadav, V. K.; Abraham, J. K.; Mani, P.; Kulshrestha, R.; Chowdhury, S. QuadBase: Genome-Wide Database of G4 DNA--Occurrence and Conservation in Human, Chimpanzee, Mouse and Rat Promoters and 146 Microbes. *Nucleic Acids Res.* **2008**, *36* (Database issue), D381-5. https://doi.org/10.1093/nar/gkm781.

(85) Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44* (4), 1746–1759. https://doi.org/10.1093/nar/gkw006.

(86) Scaria, V.; Hariharan, M.; Arora, A.; Maiti, S. Quadfinder: Server for Identification and Analysis of Quadruplex-Forming Motifs in Nucleotide Sequences. *Nucleic Acids Res.* **2006**, *34* (Web Server issue), W683-5. https://doi.org/10.1093/nar/gkl299.

(87) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* **2005**, *33* (9), 2908–2916. https://doi.org/10.1093/nar/gki609.

(88) Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity with

G4Hunter. *Nucleic Acids Res.* **2016**, *44* (4), 1746–1759. https://doi.org/10.1093/nar/gkw006.

(89) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* **2015**, *33* (8), 877–881. https://doi.org/10.1038/nbt.3295.

(90) Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-Quadruplex Structures Mark Human Regulatory Chromatin. *Nat. Genet.* **2016**, *48* (10), 1267–1272. https://doi.org/10.1038/ng.3662.

(91) Huppert, J. L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35* (2), 406–413. https://doi.org/10.1093/nar/gkl1057.

(92) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417* (6891), 876–880. https://doi.org/10.1038/nature755.

(93) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. Inhibition of Telomerase by G-Quartet DMA Structures. *Nature* **1991**, *350* (6320), 718–720. https://doi.org/10.1038/350718a0.

(94) Kim, N.; Piatyszek, M.; Prowse, K.; Harley, C.; West, M.; Ho, P.; Coviello, G.; Wright, W.; Weinrich, S.; Shay, J. Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science (80-.).* **1994**, *266* (5193), 2011–2015. https://doi.org/10.1126/science.7605428.

(95) Kim, M. Y.; Vankayalapati, H.; Shin-Ya, K.; Wierzba, K.; Hurley, L. H. Telomestatin, a Potent Telomerase Inhibitor That Interacts Quite Specifically with the Human Telomeric Intramolecular G-Quadruplex. *J. Am. Chem. Soc.* **2002**, *124* (10), 2098–2099. https://doi.org/10.1021/ja017308q.

(96) Burger, A. M.; Dai, F.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S. The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* **2005**, *65* (4), 1489–1496. https://doi.org/10.1158/0008-5472.CAN-04-2910.

(97) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. Telomere End-Binding Proteins Control the Formation of G-Quadruplex DNA Structures in Vivo. *Nat. Struct. Mol. Biol.* **2005**, *12* (10), 847–854. https://doi.org/10.1038/nsmb982.

(98) Ou, T.; Lu, Y.; Tan, J.; Huang, Z.; Wong, K.-Y.; Gu, L. G-Quadruplexes: Targets in Anticancer Drug Design. *ChemMedChem* **2008**, *3* (5), 690–713. https://doi.org/10.1002/cmdc.200700300.

(99) Lerner, L. K.; Sale, J. E. Replication of G Quadruplex DNA. *Genes (Basel).* **2019**, *10* (2). https://doi.org/10.3390/genes10020095.

(100) Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M.-P.; Foiani, M.; Nicolas, A. G-Quadruplex-Induced Instability during Leading-Strand

Replication. *EMBO J.* **2011**, *30* (19), 4033–4046. https://doi.org/10.1038/emboj.2011.316.

(101) London, T. B. C.; Barber, L. J.; Mosedale, G.; Kelly, G. P.; Balasubramanian, S.; Hickson, I. D.; Boulton, S. J.; Hiom, K. FANCJ Is a Structure-Specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts. *J. Biol. Chem.* **2008**, *283* (52), 36132–36139. https://doi.org/10.1074/jbc.M808152200.

(102) Kim, N. The Interplay between G-Quadruplex and Transcription. *Curr. Med. Chem.* **2019**, *26* (16), 2898–2917. https://doi.org/10.2174/0929867325666171229132619.

(103) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters. *FEBS J.* **2010**, *277* (17), 3459–3469. https://doi.org/10.1111/j.1742-4658.2010.07759.x.

(104) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11593–11598. https://doi.org/10.1073/pnas.182256799.

(105) Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Res.* **2006**, *34* (9), 2536. https://doi.org/10.1093/NAR/GKL286.

(106) Sun, D.; Liu, W.-J.; Guo, K.; Rusche, J. J.; Ebbinghaus, S.; Gokhale, V.; Hurley, L. H. The Proximal Promoter Region of the Human Vascular Endothelial Growth Factor Gene Has a G-Quadruplex Structure That Can Be Targeted by G-Quadruplex-Interactive Agents. *Mol. Cancer Ther.* **2008**, *7* (4), 880–889. https://doi.org/10.1158/1535-7163.MCT-07-2119.

(107) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution. *J. Am. Chem. Soc.* **2006**, *128* (4), 1096–1098. https://doi.org/10.1021/ja055636a.

(108) Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human C-Kit Oncogene. *Biochemistry* **2006**, *45* (25), 7854–7860. https://doi.org/10.1021/bi0601510.

(109) Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H. Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the HTERT Core Promoter with Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands. *J. Am. Chem. Soc.* **2009**, *131* (31), 10878–10891. https://doi.org/10.1021/ja902281d.

(110) Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H. Characterization of the G-Quadruplexes in the Duplex Nuclease Hypersensitive Element of the PDGF-A Promoter and Modulation of PDGF-A Promoter Activity by TMPyP4. *Nucleic Acids Res.* **2007**, *35* (22), 7698–7713. https://doi.org/10.1093/nar/gkm538.

(111) Brooks, T. A.; Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **2010**, *1* (6), 641–649. https://doi.org/10.1177/1947601910377493.

(112) Brown, R. V.; Danford, F. L.; Gokhale, V.; Hurley, L. H.; Brooks, T. A. Demonstration That Drug-Targeted Down-Regulation of MYC in Non-Hodgkins Lymphoma Is Directly Mediated through the Promoter G-Quadruplex. *J. Biol. Chem.* **2011**, *286* (47), 41018–41027. https://doi.org/10.1074/jbc.M111.274720.

(113) Boddupally, P. V. L.; Hahn, S.; Beman, C.; De, B.; Brooks, T. A.; Gokhale, V.; Hurley, L. H. Anticancer Activity and Cellular Repression of C-MYC by the G-Quadruplex-Stabilizing 11-Piperazinylquindoline Is Not Dependent on Direct Targeting of the G-Quadruplex in the c-MYC Promoter. *J. Med. Chem.* **2012**, *55* (13), 6076–6086. https://doi.org/10.1021/jm300282c.

(114) Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA Secondary Structures: Stability and Function of G-Quadruplex Structures. *Nat. Rev. Genet.* **2012**, *13* (11), 770–780. https://doi.org/10.1038/nrg3296.

(115) Sun, D.; Hurley, L. H. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* **2009**, *52* (9), 2863. https://doi.org/10.1021/JM900055S.

(116) Mukherjee, A. K.; Sharma, S.; Chowdhury, S. Non-Duplex G-Quadruplex Structures Emerge as Mediators of Epigenetic Modifications. *Trends Genet.* **2019**, *35* (2), 129–144. https://doi.org/10.1016/j.tig.2018.11.001.

(117) Sengupta, A.; Ganguly, A.; Chowdhury, S. Promise of G-Quadruplex Structure Binding Ligands as Epigenetic Modifiers with Anti-Cancer Effects. *Molecules* **2019**, *24* (3). https://doi.org/10.3390/molecules24030582.

(118) Halder, R.; Halder, K.; Sharma, P.; Garg, G.; Sengupta, S.; Chowdhury, S. Guanine Quadruplex DNA Structure Restricts Methylation of CpG Dinucleotides Genome-Wide. *Mol. Biosyst.* **2010**, *6* (12), 2439. https://doi.org/10.1039/c0mb00009d.

(119) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. An RNA G-Quadruplex in the 5′ UTR of the NRAS Proto-Oncogene Modulates Translation. *Nat. Chem. Biol.* **2007**, *3* (4), 218–221. https://doi.org/10.1038/nchembio864.

(120) Bugaut, A.; Balasubramanian, S. 5'-UTR RNA G-Quadruplexes: Translation Regulation and Targeting. *Nucleic Acids Res.* **2012**, *40* (11), 4727–4741. https://doi.org/10.1093/nar/gks068.

(121) Millevoi, S.; Moine, H.; Vagner, S. G-Quadruplexes in RNA Biology. *Wiley Interdiscip. Rev. RNA* **2012**, *3* (4), 495–507. https://doi.org/10.1002/wrna.1113.

(122) Arora, A.; Suess, B. An RNA G-Quadruplex in the 3' UTR of the Proto-Oncogene PIM1 Represses Translation. *RNA Biol.* **2011**, *8* (5), 802–805. https://doi.org/10.4161/rna.8.5.16038.

(123) Christiansen, J.; Kofod, M.; Nielsen, F. C. A Guanosine Quadruplex and Two Stable Hairpins Flank a Major Cleavage Site in Insulin-like Growth Factor II MRNA. *Nucleic Acids Res.* **1994**, *22* (25), 5709–5716. https://doi.org/10.1093/nar/22.25.5709.

(124) Endoh, T.; Kawasaki, Y.; Sugimoto, N. Stability of RNA Quadruplex in Open Reading

Frame Determines Proteolysis of Human Estrogen Receptor α. *Nucleic Acids Res.* **2013**, *41* (12), 6222–6231. https://doi.org/10.1093/nar/gkt286.

(125) Lavezzo, E.; Berselli, M.; Frasson, I.; Perrone, R.; Palù, G.; Brazzale, A. R.; Richter, S. N.; Toppo, S. G-Quadruplex Forming Sequences in the Genome of All Known Human Viruses: A Comprehensive Guide. *PLOS Comput. Biol.* **2018**, *14* (12), e1006675. https://doi.org/10.1371/journal.pcbi.1006675.

(126) Harris, L. M.; Merrick, C. J. G-Quadruplexes in Pathogens: A Common Route to Virulence Control? *PLOS Pathog.* **2015**, *11* (2), e1004562. https://doi.org/10.1371/journal.ppat.1004562.

(127) Artusi, S.; Nadai, M.; Perrone, R.; Biasolo, M. A.; Palù, G.; Flamand, L.; Calistri, A.; Richter, S. N. The Herpes Simplex Virus-1 Genome Contains Multiple Clusters of Repeated G-Quadruplex: Implications for the Antiviral Activity of a G-Quadruplex Ligand. *Antiviral Res.* **2015**, *118*, 123–131. https://doi.org/10.1016/j.antiviral.2015.03.016.

(128) Norseen, J.; Johnson, F. B.; Lieberman, P. M. Role for G-Quadruplex RNA Binding by Epstein-Barr Virus Nuclear Antigen 1 in DNA Replication and Metaphase Chromosome Attachment. *J. Virol.* **2009**, *83* (20), 10336. https://doi.org/10.1128/JVI.00747-09.

(129) Gilbert-Girard, S.; Gravel, A.; Artusi, S.; Richter, S. N.; Wallaschek, N.; Kaufer, B. B.; Flamand, L. Stabilization of Telomere G-Quadruplexes Interferes with Human Herpesvirus 6A Chromosomal Integration. *J. Virol.* **2017**, *91* (14). https://doi.org/10.1128/JVI.00402-17.

(130) Madireddy, A.; Purushothaman, P.; Loosbroock, C. P.; Robertson, E. S.; Schildkraut, C. L.; Verma, S. C. G-Quadruplex-Interacting Compounds Alter Latent DNA Replication and Episomal Persistence of KSHV. *Nucleic Acids Res.* **2016**, *44* (8), 3675–3694. https://doi.org/10.1093/nar/gkw038.

(131) Murat, P.; Zhong, J.; Lekieffre, L.; Cowieson, N. P.; Clancy, J. L.; Preiss, T.; Balasubramanian, S.; Khanna, R.; Tellam, J. G-Quadruplexes Regulate Epstein-Barr Virus–Encoded Nuclear Antigen 1 MRNA Translation. *Nat. Chem. Biol.* **2014**, *10* (5), 358–364. https://doi.org/10.1038/nchembio.1479.

(132) Tellam, J. T.; Zhong, J.; Lekieffre, L.; Bhat, P.; Martinez, M.; Croft, N. P.; Kaplan, W.; Tellam, R. L.; Khanna, R. MRNA Structural Constraints on EBNA1 Synthesis Impact on In Vivo Antigen Presentation and Early Priming of CD8+ T Cells. *PLoS Pathog.* **2014**, *10* (10), e1004423. https://doi.org/10.1371/journal.ppat.1004423.

(133) Tlučková, K.; Marušič, M.; Tóthová, P.; Bauer, L.; Šket, P.; Plavec, J.; Viglasky, V. Human Papillomavirus G-Quadruplexes. *Biochemistry* **2013**, *52* (41), 7207–7216. https://doi.org/10.1021/bi400897g.

(134) Satkunanathan, S.; Thorpe, R.; Zhao, Y. The Function of DNA Binding Protein Nucleophosmin in AAV Replication. *Virology* **2017**, *510*, 46–54. https://doi.org/10.1016/j.virol.2017.07.007.

(135) Tan, J.; Vonrhein, C.; Smart, O. S.; Bricogne, G.; Bollati, M.; Kusov, Y.; Hansen, G.; Mesters, J. R.; Schmidt, C. L.; Hilgenfeld, R. The SARS-Unique Domain (SUD) of SARS

Coronavirus Contains Two Macrodomains That Bind G-Quadruplexes. *PLoS Pathog.* **2009**, *5* (5), e1000428. https://doi.org/10.1371/journal.ppat.1000428.

(136) Bian, W.-X.; Xie, Y.; Wang, X.-N.; Xu, G.-H.; Fu, B.-S.; Li, S.; Long, G.; Zhou, X.; Zhang, X.-L. Binding of Cellular Nucleolin with the Viral Core RNA G-Quadruplex Structure Suppresses HCV Replication. *Nucleic Acids Res.* **2019**, *47* (1), 56–68. https://doi.org/10.1093/nar/gky1177.

(137) Fleming, A. M.; Ding, Y.; Alenko, A.; Burrows, C. J. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS Infect. Dis.* **2016**, *2* (10), 674–681. https://doi.org/10.1021/acsinfecdis.6b00109.

(138) Wang, S.-R.; Zhang, Q.-Y.; Wang, J.-Q.; Ge, X.-Y.; Song, Y.-Y.; Wang, Y.-F.; Li, X.-D.; Fu, B.-S.; Xu, G.-H.; Shu, B.; et al. Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell Chem. Biol.* **2016**, *23* (9), 1113–1122. https://doi.org/10.1016/j.chembiol.2016.07.019.

(139) Biswas, B.; Kandpal, M.; Vivekanandan, P. A G-Quadruplex Motif in an Envelope Gene Promoter Regulates Transcription and Virion Secretion in HBV Genotype B. *Nucleic Acids Res.* **2017**, *45* (19), 11268–11280. https://doi.org/10.1093/nar/gkx823.

(140) Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palù, G.; Richter, S. N. A Dynamic G-Quadruplex Region Regulates the HIV-1 Long Terminal Repeat Promoter. *J. Med. Chem.* **2013**, *56* (16), 6521–6530. https://doi.org/10.1021/jm400914r.

(141) Piekna-Przybylska, D.; Sullivan, M. A.; Sharma, G.; Bambara, R. A. U3 Region in the HIV-1 Genome Adopts a G-Quadruplex Structure in Its RNA and DNA Sequence. *Biochemistry* **2014**, *53* (16), 2581–2593. https://doi.org/10.1021/bi4016692.

(142) Amrane, S.; Kerkour, A.; Bedrat, A.; Vialet, B.; Andreola, M.-L.; Mergny, J.-L. Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development. *J. Am. Chem. Soc.* **2014**, *136* (14), 5249–5252. https://doi.org/10.1021/ja501500c.

(143) Perrone, R.; Lavezzo, E.; Palù, G.; Richter, S. N. Conserved Presence of G-Quadruplex Forming Sequences in the Long Terminal Repeat Promoter of Lentiviruses. *Sci. Rep.* **2017**, *7* (1), 2018. https://doi.org/10.1038/s41598-017-02291-1.

(144) Barre-Sinoussi, F.; Chermann, J.; Rey, F.; Nugeyre, M.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; et al. Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science (80-.).* **1983**, *220* (4599), 868–871. https://doi.org/10.1126/science.6189183.

(145) Danforth, K.; Granich, R.; Wiedeman, D.; Baxi, S.; Padian, N. *Global Mortality and Morbidity of HIV/AIDS*; The International Bank for Reconstruction and Development / The World Bank, **2017**. https://doi.org/10.1596/978-1-4648-0524-0/CH2.

(146) Nyamweya, S.; Hegedus, A.; Jaye, A.; Rowland-Jones, S.; Flanagan, K. L.; Macallan, D. C. Comparing HIV-1 and HIV-2 Infection: Lessons for Viral Immunopathogenesis. *Rev.*

Med. Virol. **2013**, *23* (4), 221–240. https://doi.org/10.1002/rmv.1739.

(147) German Advisory Committee Blood (Arbeitskreis Blut), Subgroup 'Assessment of Pathogens Transmissible by Blood,' G. A. C. B. (Arbeitskreis; Blood', S. 'Assessment of P. T. by. Human Immunodeficiency Virus (HIV). *Transfus. Med. Hemother.* **2016**, *43* (3), 203–222. https://doi.org/10.1159/000445852.

(148) Perelson, A. S.; Kirschner, D. E.; De Boer, R. Dynamics of HIV Infection of CD4+ T Cells. *Math. Biosci.* **1993**, *114* (1), 81–125. https://doi.org/10.1016/0025-5564(93)90043-A.

(149) Wong, M. E.; Jaworowski, A.; Hearps, A. C. The HIV Reservoir in Monocytes and Macrophages. *Front. Immunol.* **2019**, *10*, 1435. https://doi.org/10.3389/fimmu.2019.01435.

(150) Ganser-Pornillos, B. K.; Yeager, M.; Pornillos, O. Assembly and Architecture of HIV. In *Advances in experimental medicine and biology*; **2012**; Vol. 726, pp 441–465. https://doi.org/10.1007/978-1-4614-0980-9_20.

(151) Engelman, A.; Cherepanov, P. The Structural Biology of HIV-1: Mechanistic and Therapeutic Insights. *Nat. Rev. Microbiol.* **2012**, *10* (4), 279–290. https://doi.org/10.1038/nrmicro2747.

(152) Jameson, J.L.; Fauci, A.S.; Kasper, D.L.; Hauser, S.L.; Longo, D.L.; Loscalzo, J. Human Immunodeficiency Virus Disease: AIDS and Related Disorders, Harrison's Principles of Internal Medicine.". In *Harrisons Principles of Internal Medicine*; **2005**; pp 1076–1139.

(153) Wilen, C. B.; Tilton, J. C.; Doms, R. W. HIV: Cell Binding and Entry. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (8). https://doi.org/10.1101/cshperspect.a006866.

(154) Hu, W.-S.; Hughes, S. H. HIV-1 Reverse Transcription. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (10). https://doi.org/10.1101/cshperspect.a006882.

(155) de Arellano, E. R.; Alcamí, J.; López, M.; Soriano, V.; Holguín, Á. Drastic Decrease of Transcription Activity Due to Hypermutated Long Terminal Repeat (LTR) Region in Different HIV-1 Subtypes and Recombinants. *Antiviral Res.* **2010**, *88* (2), 152–159. https://doi.org/10.1016/J.ANTIVIRAL.2010.08.007.

(156) Pereira, L. A.; Bentley, K.; Peeters, A.; Churchill, M. J.; Deacon, N. J. A Compilation of Cellular Transcription Factor Interactions with the HIV-1 LTR Promoter. *Nucleic Acids Res.* **2000**, *28* (3), 663–668. https://doi.org/10.1093/nar/28.3.663.

(157) Mbondji-wonje, C.; Dong, M.; Wang, X.; Zhao, J.; Ragupathy, V.; Sanchez, A. M.; Denny, T. N.; Hewlett, I. Distinctive Variation in the U3R Region of the 5' Long Terminal Repeat from Diverse HIV-1 Strains. *PLoS One* **2018**, *13* (4), e0195661. https://doi.org/10.1371/JOURNAL.PONE.0195661.

(158) Sims, R. J.; Belotserkovskaya, R.; Reinberg, D. Elongation by RNA Polymerase II: The Short and Long of It. *Genes Dev.* **2004**, *18* (20), 2437–2468. https://doi.org/10.1101/gad.1235904.

(159) Liu, R.; Wu, J.; Shao, R.; Xue, Y. Mechanism and Factors That Control HIV-1 Transcription and Latency Activation. *J. Zhejiang Univ. Sci. B* **2014**, *15* (5), 455–465. https://doi.org/10.1631/jzus.B1400059.

(160) Karn, J.; Stoltzfus, C. M. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (2), a006916. https://doi.org/10.1101/cshperspect.a006916.

(161) Frankel, A. D.; Young, J. A. T. HIV-1: Fifteen Proteins and an RNA. *Annu. Rev. Biochem.* **1998**, *67* (1), 1–25. https://doi.org/10.1146/annurev.biochem.67.1.1.

(162) Darlix, J.-L.; Godet, J.; Ivanyi-Nagy, R.; Fossé, P.; Mauffret, O.; Mély, Y. Flexible Nature and Specific Functions of the HIV-1 Nucleocapsid Protein. *J. Mol. Biol.* **2011**, *410* (4), 565–581. https://doi.org/10.1016/j.jmb.2011.03.037.

(163) Darlix, J.-L.; de Rocquigny, H.; Mauffret, O.; Mély, Y. Retrospective on the All-in-One Retroviral Nucleocapsid Protein. *Virus Res.* **2014**, *193*, 2–15. https://doi.org/10.1016/J.VIRUSRES.2014.05.011.

(164) Williams, M. C.; Gorelick, R. J.; Musier-Forsyth, K. Specific Zinc-Finger Architecture Required for HIV-1 Nucleocapsid Protein's Nucleic Acid Chaperone Function. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (13), 8614–8619. https://doi.org/10.1073/pnas.132128999.

(165) Bampi, C.; Jacquenet, S.; Lener, D.; Décimo, D.; Darlix, J.-L. The Chaperoning and Assistance Roles of the HIV-1 Nucleocapsid Protein in Proviral DNA Synthesis and Maintenance. *Int. J. Biochem. Cell Biol.* **2004**, *36* (9), 1668–1686. https://doi.org/10.1016/j.biocel.2004.02.024.

(166) Darlix, J.; Garrido, J. L.; Morellet, N.; Mély, Y.; Rocquigny, H. de. Properties, Functions, and Drug Targeting of the Multifunctional Nucleocapsid Protein of the Human Immunodeficiency Virus. *Adv. Pharmacol.* **2007**, *55*, 299–346. https://doi.org/10.1016/S1054-3589(07)55009-X.

(167) Kanevsky, I.; Chaminade, F.; Chen, Y.; Godet, J.; René, B.; Darlix, J.-L.; Mély, Y.; Mauffret, O.; Fossé, P. Structural Determinants of TAR RNA-DNA Annealing in the Absence and Presence of HIV-1 Nucleocapsid Protein. *Nucleic Acids Res.* **2011**, *39* (18), 8148–8162. https://doi.org/10.1093/nar/gkr526.

(168) Sosic, A.; Saccone, I.; Carraro, C.; Kenderdine, T.; Gamba, E.; Caliendo, G.; Corvino, A.; Di Vaio, P.; Fiorino, F.; Magli, E.; et al. Non-Natural Linker Configuration in 2,6-Dipeptidyl-Anthraquinones Enhances the Inhibition of TAR RNA Binding/Annealing Activities by HIV-1 NC and Tat Proteins. *Bioconjug. Chem.* **2018**, *29* (7), 2195–2207. https://doi.org/10.1021/acs.bioconjchem.8b00104.

(169) Sundquist, W. I.; Kräusslich, H.-G. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (7), a006924. https://doi.org/10.1101/cshperspect.a006924.

(170) Siliciano, R. F.; Greene, W. C. HIV Latency. *Cold Spring Harb. Perspect. Med.* **2011**, *1* (1), a007096. https://doi.org/10.1101/cshperspect.a007096.

(171) Bosque, A.; Planelles, V. Induction of HIV-1 Latency and Reactivation in Primary Memory CD4+ T Cells. *Blood* **2009**, *113* (1), 58–65. https://doi.org/10.1182/blood-2008-07-168393.

(172) Chun, T.-W.; Finzi, D.; Margolick, J.; Chadwick, K.; Schwartz, D.; Siliciano, R. F. In Vivo Fate of HIV-1-Infected T Cells: Quantitative Analysis of the Transition to Stable Latency. *Nat. Med.* **1995**, *1* (12), 1284–1290. https://doi.org/10.1038/nm1295-1284.

(173) Alexaki, A.; Liu, Y.; Wigdahl, B. Cellular Reservoirs of HIV-1 and Their Role in Viral Persistence. *Curr. HIV Res.* **2008**, *6* (5), 388–400.

(174) Vanhamel, J.; Bruggemans, A.; Debyser, Z. Establishment of Latent HIV-1 Reservoirs: What Do We Really Know? *J. virus Erad.* **2019**, *5* (1), 3–9.

(175) Mbonye, U.; Karn, J. Transcriptional Control of HIV Latency: Cellular Signaling Pathways, Epigenetics, Happenstance and the Hope for a Cure. *Virology* **2014**, *454–455*, 328–339. https://doi.org/10.1016/j.virol.2014.02.008.

(176) Cary, D. C.; Fujinaga, K.; Peterlin, B. M. Molecular Mechanisms of HIV Latency. *J. Clin. Invest.* **2016**, *126* (2), 448. https://doi.org/10.1172/JCI80565.

(177) Chavez, L.; Calvanese, V.; Verdin, E. HIV Latency Is Established Directly and Early in Both Resting and Activated Primary CD4 T Cells. *PLOS Pathog.* **2015**, *11* (6), e1004955. https://doi.org/10.1371/journal.ppat.1004955.

(178) Clavel, F.; Hance, A. J. HIV Drug Resistance. *N. Engl. J. Med.* **2004**, *350* (10), 1023–1035. https://doi.org/10.1056/NEJMra025195.

(179) Santoro, M. M.; Perno, C. F. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiol.* **2013**, *2013*, 481314. https://doi.org/10.1155/2013/481314.

(180) Xu, W.; Li, H.; Wang, Q.; Hua, C.; Zhang, H.; Li, W.; Jiang, S.; Lu, L. Advancements in Developing Strategies for Sterilizing and Functional HIV Cures. *Biomed Res. Int.* **2017**, *2017*, 1–12. https://doi.org/10.1155/2017/6096134.

(181) Pham, H. T.; Mesplède, T. The Latest Evidence for Possible HIV-1 Curative Strategies. *Drugs Context* **2018**, *7*, 212522. https://doi.org/10.7573/dic.212522.

(182) Darcis, G.; Van Driessche, B.; Van Lint, C. HIV Latency: Should We Shock or Lock? *Trends Immunol.* **2017**, *38* (3), 217–228. https://doi.org/10.1016/j.it.2016.12.003.

(183) Xu, W.; Li, H.; Wang, Q.; Hua, C.; Zhang, H.; Li, W.; Jiang, S.; Lu, L. Advancements in Developing Strategies for Sterilizing and Functional HIV Cures. *Biomed Res. Int.* **2017**, *2017*, 1–12. https://doi.org/10.1155/2017/6096134.

(184) Spina, C. A.; Anderson, J.; Archin, N. M.; Bosque, A.; Chan, J.; Famiglietti, M.; Greene, W. C.; Kashuba, A.; Lewin, S. R.; Margolis, D. M.; et al. An In-Depth Comparison of Latent HIV-1 Reactivation in Multiple Cell Model Systems and Resting CD4+ T Cells from Aviremic Patients. *PLoS Pathog.* **2013**, *9* (12), e1003834. https://doi.org/10.1371/journal.ppat.1003834.

(185) Dahabieh, M. S.; Battivelli, E.; Verdin, E. Understanding HIV Latency: The Road to an HIV Cure. *Annu. Rev. Med.* **2015**, *66*, 407–421. https://doi.org/10.1146/annurev-med-092112-152941.

(186) Hütter, G.; Nowak, D.; Mossner, M.; Ganepola, S.; Müßig, A.; Allers, K.; Schneider, T.; Hofmann, J.; Kücherer, C.; Blau, O.; et al. Long-Term Control of HIV by *CCR5* Delta32/Delta32 Stem-Cell Transplantation. *N. Engl. J. Med.* **2009**, *360* (7), 692–698. https://doi.org/10.1056/NEJMoa0802905.

(187) Kaminski, R.; Bella, R.; Yin, C.; Otte, J.; Ferrante, P.; Gendelman, H. E.; Li, H.; Booze, R.; Gordon, J.; Hu, W.; et al. Excision of HIV-1 DNA by Gene Editing: A Proof-of-Concept in

Vivo Study. *Gene Ther.* **2016**, *23* (8–9), 690–695. https://doi.org/10.1038/gt.2016.41.

(188) Bobbin, M. L.; Burnett, J. C.; Rossi, J. J. RNA Interference Approaches for Treatment of HIV-1 Infection. *Genome Med.* **2015**, *7* (1), 50. https://doi.org/10.1186/s13073-015-0174-y.

(189) Perrone, R.; Nadai, M.; Poe, J. A.; Frasson, I.; Palumbo, M.; Palù, G.; Smithgall, T. E.; Richter, S. N. Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 Nef Coding Region: Implications for Antiviral Activity. *PLoS One* **2013**, *8* (8), e73121. https://doi.org/10.1371/journal.pone.0073121.

(190) Perrone, R.; Butovskaya, E.; Daelemans, D.; Palu, G.; Pannecouque, C.; Richter, S. N. Anti-HIV-1 Activity of the G-Quadruplex Ligand BRACO-19. *J. Antimicrob. Chemother.* **2014**, *69* (12), 3248–3258. https://doi.org/10.1093/jac/dku280.

(191) Looker, K. J.; Magaret, A. S.; May, M. T.; Turner, K. M. E.; Vickerman, P.; Gottlieb, S. L.; Newman, L. M. Global and Regional Estimates of Prevalent and Incident Herpes Simplex Virus Type 1 Infections in 2012. *PLoS One* **2015**, *10* (10), e0140765. https://doi.org/10.1371/journal.pone.0140765.

(192) Jamieson, G. A.; Maitland, N. J.; Wilcock, G. K.; Craske, J.; Itzhaki, R. F. Latent Herpes Simplex Virus Type 1 in Normal and Alzheimer's Disease Brains. *J. Med. Virol.* **1991**, *33* (4), 224–227. https://doi.org/10.1002/jmv.1890330403.

(193) Itzhaki, R. F.; Lin, W. R.; Shang, D.; Wilcock, G. K.; Faragher, B.; Jamieson, G. A. Herpes Simplex Virus Type 1 in Brain and Risk of Alzheimer's Disease. *Lancet (London, England)* **1997**, *349* (9047), 241–244. https://doi.org/10.1016/S0140-6736(96)10149-5.

(194) Davison, A. J. *Overview of Classification*; Cambridge University Press, **2007**.

(195) Whitley, R. J.; Roizman, B. Herpes Simplex Virus Infections. *Lancet (London, England)* **2001**, *357* (9267), 1513–1518. https://doi.org/10.1016/S0140-6736(00)04638-9.

(196) Ibáñez, F. J.; Farías, M. A.; Gonzalez-Troncoso, M. P.; Corrales, N.; Duarte, L. F.; Retamal-Díaz, A.; González, P. A. Experimental Dissection of the Lytic Replication Cycles of Herpes Simplex Viruses in Vitro. *Front. Microbiol.* **2018**, *9*, 2406. https://doi.org/10.3389/fmicb.2018.02406.

(197) Agelidis, A. M.; Shukla, D. Cell Entry Mechanisms of HSV: What We Have Learned in Recent Years. *Future Virol.* **2015**, *10* (10), 1145–1154. https://doi.org/10.2217/fvl.15.85.

(198) Lehman, I. R.; Boehmer, P. E. Replication of Herpes Simplex Virus DNA. *J. Biol. Chem.* **1999**, *274* (40), 28059–28062. https://doi.org/10.1074/jbc.274.40.28059.

(199) Ouyang, Q.; Zhao, X.; Feng, H.; Tian, Y.; Li, D.; Li, M.; Tan, Z. High GC Content of Simple Sequence Repeats in Herpes Simplex Virus Type 1 Genome. *Gene* **2012**, *499* (1), 37–40. https://doi.org/10.1016/j.gene.2012.02.049.

(200) Honess, R. W.; Roizman, B. Regulation of Herpesvirus Macromolecular Synthesis. I. Cascade Regulation of the Synthesis of Three Groups of Viral Proteins. *J. Virol.* **1974**, *14* (1), 8–19.

(201) Roizman, B.; Zhou, G. The 3 Facets of Regulation of Herpes Simplex Virus Gene Expression: A Critical Inquiry. *Virology* **2015**, *479–480*, 562–567.

https://doi.org/10.1016/j.virol.2015.02.036.

(202) Everett, R. D. Trans Activation of Transcription by Herpes Virus Products: Requirement for Two HSV-1 Immediate-Early Polypeptides for Maximum Activity. *EMBO J.* **1984**, *3* (13), 3135–3141.

(203) Imbalzano, A. N.; Coen, D. M.; DeLuca, N. A. Herpes Simplex Virus Transactivator ICP4 Operationally Substitutes for the Cellular Transcription Factor Sp1 for Efficient Expression of the Viral Thymidine Kinase Gene. *J. Virol.* **1991**, *65* (2), 565–574.

(204) Grondin, B.; DeLuca, N.; DeLuca, N. A.; Gu, B.; DeLuca, N. A. Herpes Simplex Virus Type 1 ICP4 Promotes Transcription Preinitiation Complex Formation by Enhancing the Binding of TFIID to DNA. *J. Virol.* **2000**, *74* (24), 11504–11510. https://doi.org/10.1128/jvi.74.24.11504-11510.2000.

(205) Everett, R. D. ICP0, a Regulator of Herpes Simplex Virus during Lytic and Latent Infection. *BioEssays* **2000**, *22* (8), 761–770. https://doi.org/10.1002/1521-1878(200008)22:8<761::AID-BIES10>3.0.CO;2-A.

(206) Sandri-Goldin, R. M. ICP27 Mediates HSV RNA Export by Shuttling through a Leucine-Rich Nuclear Export Signal and Binding Viral Intronless RNAs through an RGG Motif. *Genes Dev.* **1998**, *12* (6), 868–879. https://doi.org/10.1101/gad.12.6.868.

(207) Weller, S. K.; Coen, D. M. Herpes Simplex Viruses: Mechanisms of DNA Replication. *Cold Spring Harb. Perspect. Biol.* **2012**, *4* (9), a013011. https://doi.org/10.1101/cshperspect.a013011.

(208) Heming, J. D.; Conway, J. F.; Homa, F. L. Herpesvirus Capsid Assembly and DNA Packaging. *Adv. Anat. Embryol. Cell Biol.* **2017**, *223*, 119–142. https://doi.org/10.1007/978-3-319-53168-7_6.

(209) De Clercq, E. Selective Anti-Herpesvirus Agents. *Antivir. Chem. Chemother.* **2013**, *23*, 93–101.

(210) Superti, F.; Ammendolia, M.; Marchetti, M. New Advances in Anti-HSV Chemotherapy. *Curr. Med. Chem.* **2008**, *15* (9), 900–911. https://doi.org/10.2174/092986708783955419.

(211) Jiang, Y.-C.; Feng, H.; Lin, Y.-C.; Guo, X.-R. New Strategies against Drug Resistance to Herpes Simplex Virus. *Int. J. Oral Sci.* **2016**, *8* (1), 1–6. https://doi.org/10.1038/ijos.2016.3.

(212) Biswas, B.; Kandpal, M.; Jauhari, U. K.; Vivekanandan, P. Genome-Wide Analysis of G-Quadruplexes in Herpesvirus Genomes. *BMC Genomics* **2016**, *17* (1), 949. https://doi.org/10.1186/s12864-016-3282-1.

(213) Callegaro, S.; Perrone, R.; Scalabrin, M.; Doria, F.; Palù, G.; Richter, S. N. A Core Extended Naphtalene Diimide G-Quadruplex Ligand Potently Inhibits Herpes Simplex Virus 1 Replication. *Sci. Rep.* **2017**, *7* (1), 2341. https://doi.org/10.1038/s41598-017-02667-3.

(214) Artusi, S.; Perrone, R.; Lago, S.; Raffa, P.; Di Iorio, E.; Palù, G.; Richter, S. N. Visualization of DNA G-Quadruplexes in Herpes Simplex Virus 1-Infected Cells. *Nucleic Acids Res.* **2016**, *44* (21), gkw968. https://doi.org/10.1093/nar/gkw968.

(215) Frasson, I.; Nadai, M.; Richter, S. N. Conserved G-Quadruplexes Regulate the Immediate

Early Promoters of Human Alphaherpesviruses. *Molecules* **2019**, *24* (13), 2375. https://doi.org/10.3390/molecules24132375.

(216) Hagan, N.; Fabris, D. Direct Mass Spectrometric Determination of the Stoichiometry and Binding Affinity of the Complexes between Nucleocapsid Protein and RNA Stem−Loop Hairpins of the HIV-1 Ψ-Recognition Element †. *Biochemistry* **2003**, *42* (36), 10736–10745. https://doi.org/10.1021/bi0348922.

(217) Hellman, L. M.; Fried, M. G. Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein-Nucleic Acid Interactions. *Nat. Protoc.* **2007**, *2* (8), 1849. https://doi.org/10.1038/NPROT.2007.249.

(218) Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. Circular Dichroism and Guanine Quadruplexes. *Methods* **2012**, *57* (1), 64–75. https://doi.org/10.1016/j.ymeth.2012.03.011.

(219) Randazzo, A.; Spada, G. P.; da Silva, M. W. Circular Dichroism of Quadruplex Structures. In *Topics in current chemistry*; **2012**; Vol. 330, pp 67–86. https://doi.org/10.1007/128_2012_331.

(220) Marchand, A.; Gabelica, V. Native Electrospray Mass Spectrometry of DNA G-Quadruplexes in Potassium Solution. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (7), 1146–1154. https://doi.org/10.1007/s13361-014-0890-3.

(221) Sun, D.; Hurley, L. H. Biochemical Techniques for the Characterization of G-Quadruplex Structures: EMSA, DMS Footprinting, and DNA Polymerase Stop Assay. *Methods Mol. Biol.* **2010**, *608*, 65–79. https://doi.org/10.1007/978-1-59745-363-9_5.

(222) Orlando, V. Mapping Chromosomal Proteins in Vivo by Formaldehyde-Crosslinked-Chromatin Immunoprecipitation. *Trends Biochem. Sci.* **2000**, *25* (3), 99–104.

(223) Kuo, M.-H.; Allis, C. D. In Vivo Cross-Linking and Immunoprecipitation for Studying Dynamic Protein:DNA Associations in a Chromatin Environment. *Methods* **1999**, *19* (3), 425–433. https://doi.org/10.1006/meth.1999.0879.

(224) Reed, L. J.; Muench, H. A Simple Method of Estimating Fifty per Cent Endpoints. *Am. J. Epidemiol.* **1938**, *27* (3), 493–497. https://doi.org/10.1093/oxfordjournals.aje.a118408.

(225) La Boissière, S.; Izeta, A.; Malcomber, S.; O'Hare, P. Compartmentalization of VP16 in Cells Infected with Recombinant Herpes Simplex Virus Expressing VP16-Green Fluorescent Protein Fusion Proteins. *J. Virol.* **2004**, *78* (15), 8002–8014. https://doi.org/10.1128/JVI.78.15.8002-8014.2004.

(226) Blaho, J. A.; Morton, E. R.; Yedowitz, J. C. Herpes Simplex Virus: Propagation, Quantification, and Storage. In *Current Protocols in Microbiology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, **2006**; Vol. Chapter 14, p 14E.1.1-14E.1.23. https://doi.org/10.1002/9780471729259.mc14e01s00.

(227) Murat, P.; Singh, Y.; Defrancq, E. Methods for Investigating G-Quadruplex DNA/Ligand Interactions. *Chem. Soc. Rev.* **2011**, *40* (11), 5293. https://doi.org/10.1039/c1cs15117g.

(228) Daelemans, D.; Pauwels, R.; De Clercq, E.; Pannecouque, C. A Time-of–Drug Addition Approach to Target Identification of Antiviral Compounds. *Nat. Protoc.* **2011**, *6* (6), 925–

933. https://doi.org/10.1038/nprot.2011.330.

(229) Levin, J. G.; Mitra, M.; Mascarenhas, A.; Musier-Forsyth, K. Role of HIV-1 Nucleocapsid Protein in HIV-1 Reverse Transcription. *RNA Biol.* **2010**, *7* (6), 754–774. https://doi.org/10.4161/RNA.7.6.14115.

(230) Iwatani, Y.; Rosen, A. E.; Guo, J.; Musier-Forsyth, K.; Levin, J. G. Efficient Initiation of HIV-1 Reverse Transcription in Vitro: Requirement for RNA Sequences Downstream of the Primer Binding Site Abrogated by Nucleocapsid Protein-Dependent Primer-Template Interactions. *J. Biol. Chem.* **2003**, *278* (16), 14185–14195. https://doi.org/10.1074/jbc.M211618200.

(231) Kankia, B. I.; Barany, G.; Musier-Forsyth, K. Unfolding of DNA Quadruplexes Induced by HIV-1 Nucleocapsid Protein. *Nucleic Acids Res.* **2005**, *33* (14), 4395–4403. https://doi.org/10.1093/nar/gki741.

(232) Rajendran, A.; Endo, M.; Hidaka, K.; Tran, P. L. T.; Mergny, J.-L.; Gorelick, R. J.; Sugiyama, H. HIV-1 Nucleocapsid Proteins as Molecular Chaperones for Tetramolecular Antiparallel G-Quadruplex Formation. *J. Am. Chem. Soc.* **2013**, *135* (49), 18575–18585. https://doi.org/10.1021/ja409085j.

(233) Levin, J. G.; Guo, J.; Rouzina, I.; Musier-Forsyth, K. Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Critical Role in Reverse Transcription and Molecular Mechanism. *Prog. Nucleic Acid Res. Mol. Biol.* **2005**, *80*, 217–286. https://doi.org/10.1016/S0079-6603(05)80006-6.

(234) Kankia, B. I.; Barany, G.; Musier-Forsyth, K. Unfolding of DNA Quadruplexes Induced by HIV-1 Nucleocapsid Protein. *Nucleic Acids Res.* **2005**, *33* (14), 4395–4403. https://doi.org/10.1093/nar/gki741.

(235) Bernacchi, S.; Stoylov, S.; Piémont, E.; Ficheux, D.; Roques, B. P.; Darlix, J. L.; Mély, Y. HIV-1 Nucleocapsid Protein Activates Transient Melting of Least Stable Parts of the Secondary Structure of TAR and Its Complementary Sequence. *J. Mol. Biol.* **2002**, *317* (3), 385–399. https://doi.org/10.1006/jmbi.2002.5429.

(236) Folks, T.; Justement, J.; Kinter, A.; Dinarello, C.; Fauci, A. Cytokine-Induced Expression of HIV-1 in a Chronically Infected Promonocyte Cell Line. *Science (80-.).* **1987**, *238* (4828), 800–802. https://doi.org/10.1126/science.3313729.

(237) Folks, T. M.; Justement, J.; Kinter, A.; Schnittman, S.; Orenstein, J.; Poli, G.; Fauci, A. S. Characterization of a Promonocyte Clone Chronically Infected with HIV and Inducible by 13-Phorbol-12-Myristate Acetate. *J. Immunol.* **1988**, *140* (4), 1117–1122.

(238) Biswas, P. Interferon Gamma Induces the Expression of Human Immunodeficiency Virus in Persistently Infected Promonocytic Cells (U1) and Redirects the Production of Virions to Intracytoplasmic Vacuoles in Phorbol Myristate Acetate-Differentiated U1 Cells. *J. Exp. Med.* **1992**, *176* (3), 739–750. https://doi.org/10.1084/jem.176.3.739.

(239) Sen, S.; Kaminiski, R.; Deshmane, S.; Langford, D.; Khalili, K.; Amini, S.; Datta, P. K. Role of Hexokinase-1 in the Survival of Hiv-1-Infected Macrophages. *Cell Cycle* **2015**, *14* (7), 980–989. https://doi.org/10.1080/15384101.2015.1006971.

(240) Renard, I.; Grandmougin, M.; Roux, A.; Yang, S. Y.; Lejault, P.; Pirrotta, M.; Wong, J. M. Y.; Monchaud, D. Small-Molecule Affinity Capture of DNA/RNA Quadruplexes and Their Identification in Vitro and in Vivo through the G4RP Protocol. *Nucleic Acids Res.* **2019**, *47* (11), 5502–5510. https://doi.org/10.1093/nar/gkz215.

(241) Piekna-Przybylska, D.; Sharma, G.; Maggirwar, S. B.; Bambara, R. A. Deficiency in DNA Damage Response, a New Characteristic of Cells Infected with Latent HIV-1. *Cell Cycle* **2017**, *16* (10), 968–978. https://doi.org/10.1080/15384101.2017.1312225.

(242) Weldon, C.; Dacanay, J. G.; Gokhale, V.; Boddupally, P. V. L.; Behm-Ansmant, I.; Burley, G. A.; Branlant, C.; Hurley, L. H.; Dominguez, C.; Eperon, I. C. Specific G-Quadruplex Ligands Modulate the Alternative Splicing of Bcl-X. *Nucleic Acids Res.* **2018**, *46* (2), 886–896. https://doi.org/10.1093/nar/gkx1122.

(243) Scalabrin, M.; Palumbo, M.; Richter, S. N. Highly Improved Electrospray Ionization-Mass Spectrometry Detection of G-Quadruplex-Folded Oligonucleotides and Their Complexes with Small Molecules. *Anal. Chem.* **2017**, *89* (17), 8632–8637. https://doi.org/10.1021/acs.analchem.7b01282.

(244) Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. C-MYC Promoter G-Quadruplex Formed at the 5'-End of NHE III1 Element: Insights into Biological Relevance and Parallel-Stranded G-Quadruplex Stability. *Nucleic Acids Res.* **2011**, *39* (20), 9023–9033. https://doi.org/10.1093/nar/gkr612.

(245) Stump, S.; Mou, T.-C.; Sprang, S. R.; Natale, N. R.; Beall, H. D. Crystal Structure of the Major Quadruplex Formed in the Promoter Region of the Human C-MYC Oncogene. *PLoS One* **2018**, *13* (10), e0205584. https://doi.org/10.1371/journal.pone.0205584.

(246) Smith, C. A.; Bates, P.; Rivera-Gonzalez, R.; Gu, B.; DeLuca, N. A. ICP4, the Major Transcriptional Regulatory Protein of Herpes Simplex Virus Type 1, Forms a Tripartite Complex with TATA-Binding Protein and TFIIB. *J. Virol.* **1993**, *67* (8), 4676–4687.

(247) Wagner, L. M.; DeLuca, N. A. Temporal Association of Herpes Simplex Virus ICP4 with Cellular Complexes Functioning at Multiple Steps in PolII Transcription. *PLoS One* **2013**, *8* (10), e78242. https://doi.org/10.1371/journal.pone.0078242.

(248) Ruggiero, E.; Richter, S. N. Survey and Summary G-Quadruplexes and G-Quadruplex Ligands: Targets and Tools in Antiviral Therapy. *Nucleic Acids Res.* **2018**, *46* (7), 3270–3283. https://doi.org/10.1093/nar/gky187.

(249) Asamitsu, S.; Obata, S.; Yu, Z.; Bando, T.; Sugiyama, H. Recent Progress of Targeted G-Quadruplex-Preferred Ligands Toward Cancer Therapy. *Molecules* **2019**, *24* (3). https://doi.org/10.3390/molecules24030429.

(250) Dremel, S. E.; DeLuca, N. A. Genome Replication Affects Transcription Factor Binding Mediating the Cascade of Herpes Simplex Virus Transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (9), 3734–3739. https://doi.org/10.1073/pnas.1818463116.