UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Medicina Molecolare

CORSO DI DOTTORATO DI RICERCA IN: Medicina Molecolare

CURRICOLO: Biomedicina

CICLO XXX

G-QUADRUPLEXES IN THE HERPES SIMPLEX VIRUS TYPE 1 AND MEASLES VIRUS GENOMES: NEW ANTIVIRAL TARGETS

Coordinatore: Ch.mo Prof. Stefano Piccolo

Supervisore: Prof. ssa Sara Richter

Dottorando : Sara Callegaro

Index

List of Abbreviations

ACV	acyclovir
ADEM	acute disseminated encephalomyelitis
B-19	BRACO-19
bp	base pair
C	cytosine
c-exNDI	core extended naphthalene diimide
CC_{50}	50% cytotoxic concentration
cDNA	complementary DNA
cRNA	complementary RNA
CD	circular dichroism
Cnt/ctrl	control
DMSO	dimetil sulfoxide
dsDNA	double-stranded DNA
E (β)	early gene
G	guanines
G4	G-quadruplex
G4s	G-quadruplexes
h	hour
h.p.i.	hours post infection
HSV-1	herpes- simplex virus type 1
HSV-2	herpes-simplex virus type 2
HIV-1	human immunodeficiency virus type 1
IC_{50}	50% inhibitory concentration
IE (α)	immediate early gene
K^+	potassium
KCl	potassium chloride
kb	kilobase
L (γ)	late gene
LTR	long terminal repeat
MIBE	measles inclusion bodies encephalomyelitis
MOI	multiplicity of infection
MTT	3-(4,5-dimetylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MV	measles virus
MMR	measles- mumps- rubella vaccine
Na^+	sodium
NaCl	sodium chloride
NDI	naphthalene diimide
nt	nucleotide/ nucleotides
PQS	putative quadruplex sequence
PCR	polymerase chain reaction
Pol	polymerase
QGRS	quadruplex forming G-rich sequences
qPCR	quantitative PCR
RQ	relative quantities
SI	selectivity index
SSPE	sub-acute sclerosis panencephalitis
ssRNA	single-stranded RNA
T_m	melting temperature
wt	wild- type

Riassunto

I G-quadruplex sono strutture secondarie a quattro filamenti che possono essere formate all'interno di acidi nucleici ricchi nella composizione di guanine; i G-quadruplex sono implicati in importanti funzioni biologiche: questi agiscono come regolatori epigenetici a livello genomico in cellule umane, procariotiche e anche in virus.

I G-quadruplex agiscono come silenziatori a livello di regioni promotoriali in geni umani e, inoltre, sono stati riportati in letteratura per il loro diretto coinvolgimento nella regolazione dell'espressione genica.

Poiché il virus dell'Herpes Simplex di Tipo 1 (HSV-1) possiede un genoma caratterizzato da un'elevata percentuale di basi guanine (il 68%), e le strutture G-quadruplex identificate sono massivamente presenti durante la replicazione del virus stesso, abbiamo deciso di investigare la possibilità di agire direttamente su tali G-quadruplex attraverso l'utilizzo di un Naftalene Diimmide con core esteso (c-exNDI) per ottenere un'attività antivirale caratterizzata da un innovativo meccanismo d'azione.

Mediante caratterizzazione biofisica e biomolecolare abbiamo dimostrato che tale composto (c-exNDI) è in grado di stabilizzare sequenze G-quadruplex in modo concentrazione-dipendente e, ulteriormente è in grado di inibire il processamento della *Taq* polimerasi, in fase di elongazione del DNA, in corrispondenza delle strutture G-quadruplex.

Abbiamo inoltre dimostrato che il composto testato (c-exNDI) è in grado di riconoscere preferenzialmente i G-quadruplex virali rispetto ai G-quadruplex formati a livello delle regioni telomeriche che, all'interno di una cellula eucariotica, rappresentano i G-quadruplex maggiormente presenti. Il trattamento di cellule infettate con HSV-1 con il composto in analisi, utilizzato a basse concentrazioni, ha causato una significativa inibizione della produzione virale con parallela scarsa citotossicità sulle cellule. Infine, abbiamo dimostrato che il meccanismo d'azione di tale composto risiede nell'inibizione G-quadruplex-mediata della replicazione del DNA virale, con conseguente alterazione a valle della trascrizione genica.

Tali dati suggeriscono che la potente attività antivirale e la scarsa citotossità di tale composto dipendono principalmente da una combinazione di maggiore affinità del composto per i G-quadruplex virali e dalla loro massiccia presenza durante il processo di infezione.

Poiché la maggior parte dei composti leganti G-quadruplex è caratterizzata da un' estesa superficie aromatica planare che li rende scarsamente farmaco-simili, ci siamo focalizzati sullo studio di una libreria di composti, precedentemente creata per agire sui G-quadruplex presenti nei promotori di

oncogeni in cellule tumorali, al fine di valutare la capacità di tali composti di legare e stabilizzare i G-quadruplex presenti in HSV-1 e di esprimere un'attività antivirale diretta contro il medesimo virus. In questa parte di progetto abbiamo dimostrato che tutti i composti della libreria sono in grado di legare e stabilizzare i G-quadruplex virali secondo un meccanismo concentrazione-dipendente e, nuovamente, sono in grado di alterare la processività della polimerasi attraverso un ingombro sterico correlato alla formazione della struttura G-quadruplex.

Il trattamento di cellule infettate da HSV-1 con ciascuno dei composti della libreria a basse concentrazioni, ha mostrato una cospicua diminuzione della produzione virale. Poiché due dei quattro composti sono risultati scarsamente citotossici, per questi sono stati ottenuti promettenti indici di selettività (SI). Il meccanismo di azione di tali composti, come già visto in precedenza, è stato ascritto all'inibizione G-quadruplex-mediata della polimerasi in fase di replicazione del DNA virale.

Poiché tutti i composti della libreria sono caratterizzati da (i) strutture chimiche promettenti (dimensione più contenuta rispetto a quella di composti leganti G-quadruplex precedentemente sviluppati come BRACO-19 (B-19) e le Naftalene Diimmidi) e (ii) mostrano una potente attività contro HSV-1 a basse concentrazioni e una scarsa citotossicità, possono essere considerati composti fortemente adatti per lo sviluppo di nuovi farmaci leganti G-quadruplex contro HSV-1, e contro ceppi di HSV-1 resistenti al trattamento con Acyclovir (ACV), con caratteristiche fortemente farmaco-simili.

Nella terza parte del presente progetto di tesi, mediante un approccio di pull-down, abbiamo cercato di identificare proteine, sia cellulari che virali, in grado di interagire con strutture G-quadruplex presenti in cellule infettate. Risultati preliminari hanno portato all'identificazione di proteine sia cellulari che virali direttamente implicate in meccanismi di replicazione e trascrizione del genoma virale.

Per quanto riguarda le proteine cellulari, abbiamo identificato le proteine nucleari ribonucleoproteina U (hnRNPU), nucleolina, nucleofosmina e le proteine istoniche 11 e 12; alcune di queste proteine (hnRNP, nucleolina e nucleofosmina) erano già state riportate in letterature per la loro capacità di legare i G-quadruplex durante la replicazione della cellula. L'identificazione delle proteine istoniche 11 e 12 suggerisce la possibilità che le strutture G-quadruplex siano direttamente coinvolte nel mantenimento del genoma virale nel processo di latenza che tale virus è in grado di stabilire a livello di cellule del sistema nervoso centrale nell'uomo.

Tra le proteine virali abbiamo identificato il fattore di trascrizione ICP4, la proteina legante il DNA a singola catena ICP8, il fattore di processività della polimerasi PAP1 e la proteina capsidica VP5. L'identificazione delle proteine ICP4, ICP8 e PAP1, direttamente implicate in meccanismi di replicazione e trascrizione, conferma l'ipotesi di un diretto coinvolgimento delle strutture G-

quadruplex nei meccanismi di replicazione e trascrizione del genoma virale. Inoltre, l'identificazione della proteina capsidica VP5 suggerisce la possibilità che le strutture G-quadruplex siano direttamente coinvolte nel processo di incapsidamento delle particelle virali neo-formate come guida per l'incapsidamento del genoma virale.

Infine, attraverso lo studio del genoma RNA a singolo filamento a polarità negativa del virus del Morbillo (MV), abbiamo dimostrato la presenza di sette sequenze caratterizzate da un'elevata probabilità di formare strutture G-quadruplex. Sei delle sette sequenze identificate hanno mostrato effettiva capacità di formare strutture secondarie G-quadruplex, mentre per una è stata dimostrata la capacità di formare una struttura secondaria a forcina. Tutte le sequenze hanno mostrato una stabilizzazione a seguito del legame con il composto B-19 e il trattamento delle cellule infettate ha mostrato una pronunciata riduzione nella sintesi del genoma virale.

In conclusione, dato che (i) le sequenze G-quadruplex sono fortemente conservate e distribuite in molteplici genomi virali (come ad esempio è stato dimostrato nel virus dell'Immunodeficienza Umano HIV-1, in HSV-1 e in MV) a causa della loro implicazione nella regolazione di diversi processi biologici, e (ii) diverse molecole leganti G-quadruplex (B-19, c-exNDI e piccole molecole di una libreria di composti) si sono dimostrate attive contro diversi virus attraverso un meccanismo di inizibione G-quadruplex-mediata della replicazione, possiamo confermare che le strutture G-quadruplex possono essere considerate come un ottimo bersaglio farmacologico per il trattamento di infezioni causate da diversi agenti eziologici e, inoltre, che piccole molecole in grado di legare e stabilizzare i G-quadruplex possono rappresentare innovativi farmaci con un nuovo meccanismo d'azione.

Abstract

G-quadruplexes (G4s) are four-stranded secondary structures formed by guanine-rich nucleic acids; they are implicated in important functions as epigenetic regulators at the genomic level in humans, prokaryotes and viruses. They act as silencers in the promoter regions of human genes and, moreover, they have been proposed to be directly involved in gene regulation at the transcription level.

Since the Herpes Simplex Virus Type 1 (HSV-1) genome is significantly rich in guanines (68%) and its G4 sequences are massively present during viral replication, we aimed at investigating the possibility to target the HSV-1 G4s by a core extended naphthalene diimide (c-exNDI) G4–ligand to obtain anti HSV-1 effects with an innovative mechanism of action.

Biophysical and biomolecular analysis proved that c-exNDI stabilized the HSV-1 G4s in a concentration-dependent manner and was able to inhibit *Taq* polymerase processing at G4 forming sequences. Additionally, we proved that c-exNDI preferentially recognized HSV-1 G4s over cellular telomeric G4s, the most represented G4s within cells. Treatment of HSV-1 infected cells with c-exNDI at low nanomolar concentrations induced significant virus inhibition with low cytotoxicity. The mechanism of action was ascribed to a G4-mediated inhibition of viral DNA replication, with consequent impairment of viral gene transcription.

Our data suggested that the observed potent antiviral activity and low cytotoxicity mainly depend on a combination of c-exNDI affinity for HSV-1 G4s and their massive presence during infection.

Since the vast majority of G4-ligands share the common characteristic of a large aromatic surface which makes them not well druggable, we then aimed at investigating the ability of an in-house library of compounds, previously designed to target promoters in cancer cells, to bind and stabilize HSV-1 G4s and to exert an antiviral activity against HSV-1. We demonstrated that all compounds of the library were able to bind and stabilize HSV-1 G4s in a concentration-dependent manner with different affinities and, again, to induce polymerase stalling because of a steric hindrance linked to the G4 stabilization. Treatment of HSV-1 infected cells with all compounds of the library showed a conspicuous decrease in viral production in the low nanomolar range. Two of the four compounds were also only mildly cytotoxic and thus showed promising selectivity indexes (SI). The mechanism of action, once again, was ascribed to the G4-mediated inhibition of viral DNA replication. Since all compounds of the library possess (i) promising chemical features (i.e their small size, smaller than other well known G4-ligands such as B-19 and NDIs) and (ii) potent activities in the low nanomolar range with no cytotoxicity, this makes them suitable for

future development as novel anti HSV-1 G4-ligands with more drug-like characteristics than previous developed G4-ligands, and as promising compounds for the treatment of ACV-resistant herpetic infections.

In the third part of this study, by using a pull-down approach, we aimed at identifying viral and cellular proteins able to interact specifically with G4 structures within HSV-1 infected cells. We preliminarily identified both cellular and viral proteins directly implicated in replication and transcription processes. Concerning cellular proteins, we identified hnRNPU, nucleolin, nucleophosmin and histone proteins (11 and 12), some of them being already reported for their ability to bind G4s during replication. Identification of histone proteins opens the possibility that G4s are involved in the latent maintenance of the viral genome. Besides viral proteins we identified the transcription factor ICP4, the single-stranded DNA binding protein ICP8, PAP1 polymerase processing factor and VP5. The identification of the two transcription factors and of the polymerase processing factor confirmed the hypothesis of a direct involvement of G4s in the viral genome replication and transcription and the identification of the major capsid protein VP5 suggested a possible role of G4s as a driving force for viral genome encapsidation in the virions assembly process.

Finally, by investigating the negative-stranded RNA genome of Measles Virus (MV) we demonstrated the presence of seven G4-putative folding sequences. Six of these sequences displayed the ability to fold into G4 structures while one showed a hairpin-like conformation. All sequences were bound and stabilized by B-19 G4-ligand and treatment of MV infected cells with B-19 induced a pronounced reduction in the viral genome accumulation.

Since (i) G4s are well conserved and distributed within different viral genomes (as described by our group in HIV-1, HSV-1, and MV) because of their central roles in regulating biological processes and (ii) different G4-ligands have been demonstrated to be active against different viruses (i.e B-19, c-exNDI and G4-ligands of an in-house library in this thesis) with a G4-mediated inhibition of viral DNA replication, we can strongly encourage and confirm the hypothesis using G4s as novel broad range drug targets against different etiological agents and G4-ligands as novel innovative therapeutic molecules.

1. Introduction

1.1. G-quadruplex

In the 1960s it has been observed that DNA molecules can not only associate into the canonical duplex helix (B-DNA) where two complementary strands are held together by Watson- Crick base pairs, (**Figure 1a**) but can also fold in at least ten different non-B-structures such as left-handed Z-DNA, A-motifs, hairpin, triplex, tetraplex as G- quadruplex (G4) and i-motifs [1].

Nucleic acids whether DNA or RNA, that are rich in short run of guanine bases (G-rich), have the ability to self-associate via non-Watson-Crick base pairing to form four-stranded higher-order structures referred to as G-quadruplexes (G4s) (**Figure 1b**) [2,3].

The first evidence of the formation of an alternative conformation was first noted in 1910 from Bang and colleagues by the observation of a gelatinous aggregate formed from G- rich DNA sequences in solution [4,5]. The exact nature of the gelatinous formation, described then as a G4 formed by the self association of guanine bases, was discovered only in 1962 by Davies and co-workers [1].

The existence of G4s was subsequently confirmed first in human telomere and in the genome and then in several organisms such as bacteria, yeasts as well as viruses. Most importantly, their biological functions within mammalian living cells were extensively demonstrated and their stabilization with G4-ligands was proved to interphere with biological mechanisms that include transcription, translation, splicing, DNA recombination and telomere maintenance [6,7]. The strong correlation between G4s and biological functions within cells prospected the use of G4s as novel target for drug design for the treatment of various disorders such as cancer but also viral infections [8].

1.1.1 G-quadruplex structure, polymorphism and topologies

Guanine- rich sequences (both DNA and RNA) possess the ability to fold into non- canonical highly ordered secondary structures called G-quadruplex (G4). Guanine bases that take part into this self- association are linked by alternative hydrogen bonds: Hoogsteen hydrogen bonds that lead the formation of planar G-quartet (G- tetrads) that are the building block of each G4 structure [9]. A G-quadruplex (G4) has at least two contiguous G-tetrads vertically stacked on each other and each G-tetrad is stabilized by Hoogsteen hydrogen bonds between guanine bases at the position N1, N7, O6 and N2 to form the classical planar structure shown in **Figure 1b** [2,10,11].

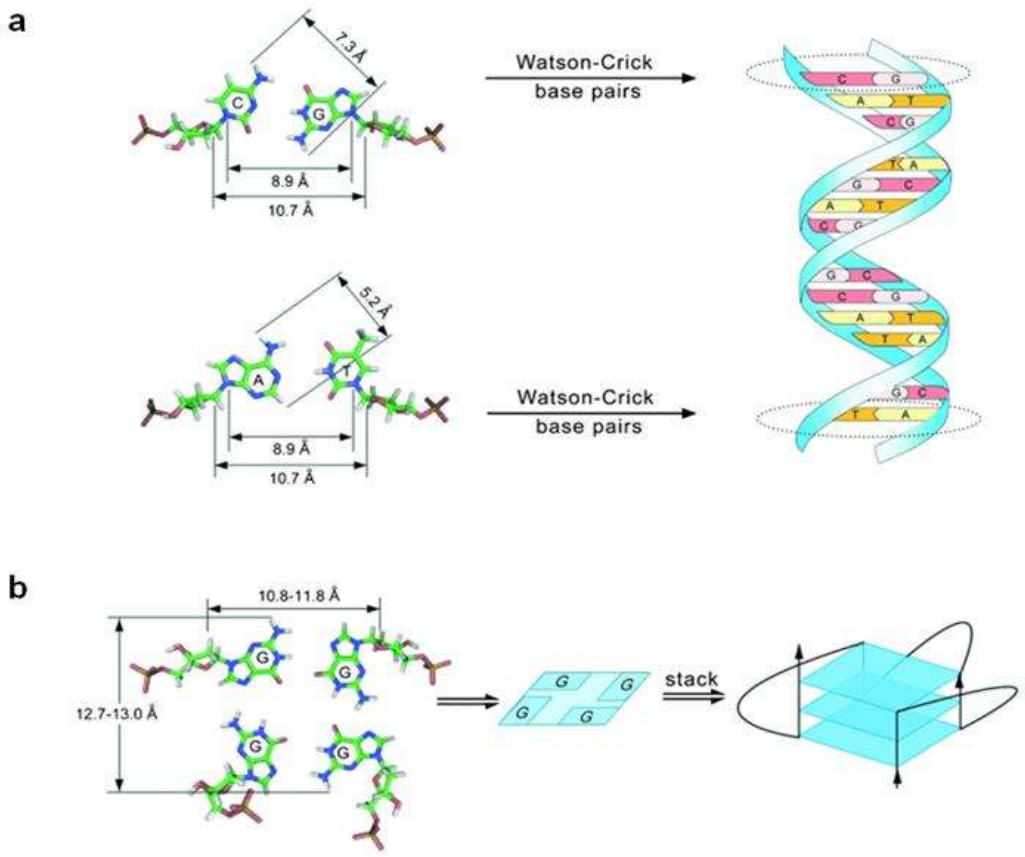


Figure 1: (**a**) Representation of the canonical Watson- Crick B- structure formed by DNA sequences. (**b**) Schematic representation of a G-quadruplex (G4) in terms of disposition of planar guanines (left) forming a G-tetrad (center) and self associated forming a G-quadruplex complete structure (right). Adapted from [12].

Many factors, such as the pH, the presence of alkali cations and G-quadruplex ligands (G4-ligands) within tetrads, positively influence the folding into G4s by promoting their formation and stabilization [11,13,14].

Depending on ion size, monovalent cations, such as K^+, Na^+ and NH_4^+, can localize in the central channel of G-quartets between tetrads by coordinating six negative charges (negative electrostatic potential), formed by the O_6 (oxygen atom) of each guanine (**Figure 2**) [14].

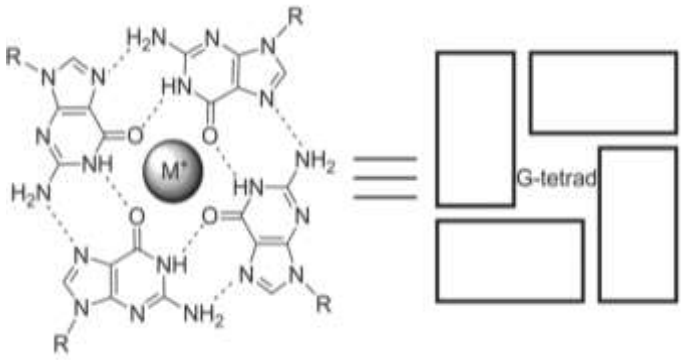


Figure 2: Schematic representation of a G-tetrad in which a monovalent cation (left panel) is coordinated in the center of the G-tetrad. In the right panel a model of a G-tetrad is represented; each coplanar guanine is shown as a rectangle. Adapted from [3].

The presence of a coordinated cation can neutralize the electrostatic repulsion eventually formed in the G4 core [3]. Additionally, depending on the specific dimension, each cation coordinated within the G4 can occupy a different location: Na^+ ions are generally coordinated in the plane of the G4 while K^+ ions are generally tetrad-plane equidistant therefore forming a tetragonal bipyramidal configuration [15].

G4s are characterized by an extensive polymorphism influenced by (i) the number and the strand orientation (referred to as strand polarities), (ii) the number of G-tetrads, (iii) the location and composition of loops, (iv) the dimension of grooves forming the G4 (medium, narrow and wide), and (v) the glycosidic conformation of each guanine torsion angle (*syn* or *anti*) (**Figure 3**) [3].

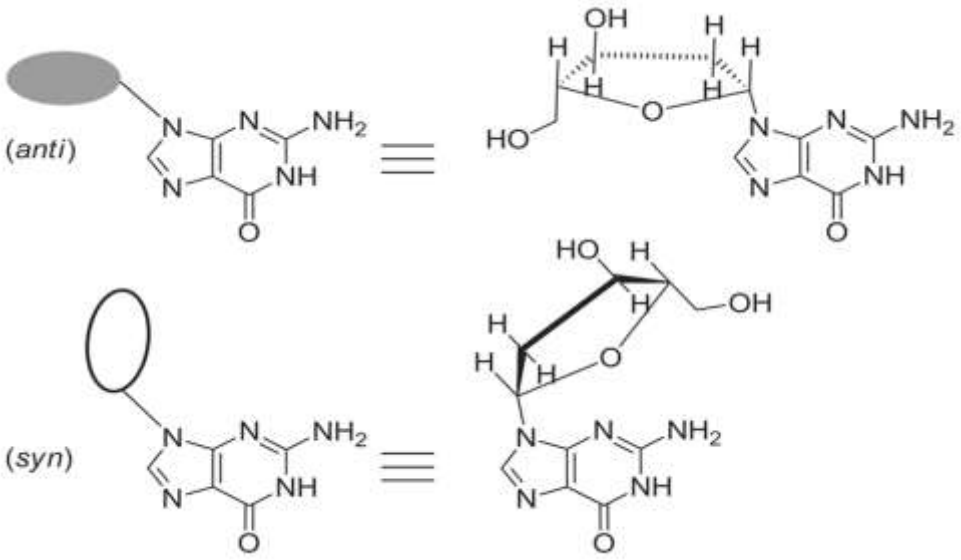


Figure 3: Schematic representation of guanines that are in *anti* and *syn* conformation. Grey and white ellipses represent pentoses in both conformations respectively. Adapted from [3].

G4s, depending on the number of separate strands that take part in their formation, can be structurally classified as (i) monomolecular, (ii) bimolecular, (iii) trimolecular or tetramolecular and have directionality from 5′ to 3′ (**Figure 4**)[13,16,17,18]:

- **Monomolelcular** (unimolecular or intramolecular): G4s are composed by one strand forming four contiguous G-tetrads.
- **Bimolecular:** G4s are formed by two strands that generally possess two G-tracts each one
- **Trimolecular** or **tetramolecular:** are formed by three or four separate strands that associate together. In this case each single strand generally has only one G-tract.

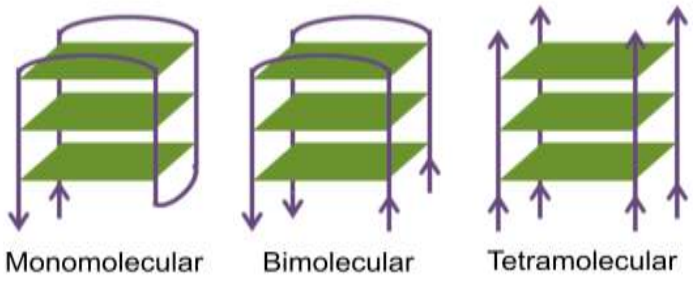


Figure 4: Schematic representation of the polymorphism of G4 structures in terms of separate strands that take part to the structure formation. G4s can be monomolecular, bimolecular, tri- or tetramolecular. Green planes represent G- tetrads, grey lines represent the sugar backbone of the structure while arrows indicate the strand polarity. Adapted from [19].

The general structure of a single strand G4 forming sequence can be recapitulated as:

$$G_{\geq 2}\, N_{L1}\, G_{\geq 2}\, N_{L2}\, G_{\geq 2}\, N_{L3}\, G_{\geq 2}.$$

Concerning G-tracts (also referred to as G4-legs) they can adopt a different conformation and this feature is related to the conformation of each guanine torsion angle. Parallel G4s have all guanines in the *anti* conformation (therefore making all strands parallel to each other) while antiparallel G4s have guanines in both *syn* and *anti* conformation (leading to one strand antiparallel to the others) (**Figure 5a-d**).

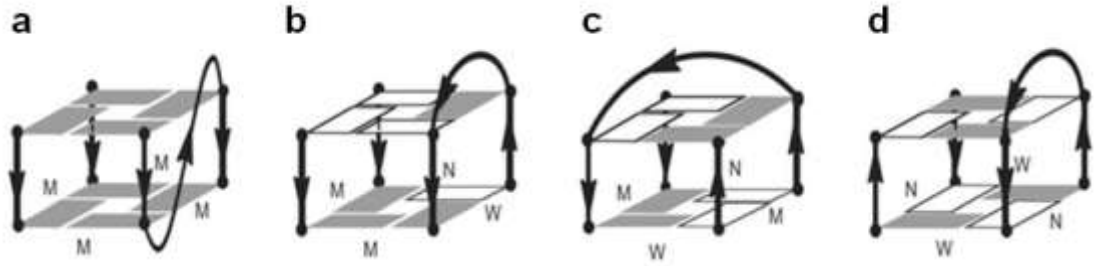


Figure 5: Schematic representation of G-tetrad core types and loops. (**a**) Parallel G-tetrad core with *anti* alignment of all guanines and external loop. (**b**) Hybrid/mixed G-tetrad core (3+1) with *anti-anti-anti-syn/ syn-syn-syn-anti* guanines alignment and lateral loop. (**c**) Type 1 antiparallel G-tetrad core with *syn-syn-anti-anti* guanines alignment and diagonal loop. (**d**) Type 2 antiparallel G-tetrad core with *syn-anti-anti-syn* guanines alignment and lateral loop. Adapted from [3].

The different conformation adopted by G4-tracts (which connect G4-tetrads to each other) can therefore influence the conformation adopted by G4-tetrads [3]:

- Four strands in the same direction and guanines in *anti* conformation form a **parallel** G-tetrad (**Figure 5a**)
- Three strands in the same direction and the fourth in the opposite direction and guanines consequently in *anti- anti- anti- syn* or *syn- syn- syn- anti* conformation form a mixed/ **hybrid-type** G-tetrad (3+1 G-tetrad) (**Figure 5b**)

- Two adjacent strands in the same conformation *syn* or *anti* and the other two in the opposite conformation and guanines in *syn- syn- anti- anti* conformation form an **antiparallel** G-tetrad (**type 1**) (**Figure 5c**)
- Two diagonally opposite strands in the same direction and the opposite two in another same direction and guanines in *anti- syn- anti- syn* form an **antiparallel** G-tetrad (**type 2**) (**Figure 5d**).

Moreover, the G-tetrads are held together by the presence of intervening mixed- sequences called loops (N_{Lx}) that can vary in their length and sequence and that are not usually directly involved in the tetrad themselves [10,15]. Loops are characterized by a wide range of structures classified into three major types [3,10]:

- **Chain-reversal/external** (also called propeller loops:) connecting adjacent parallel strands (**Figure 5a**)
- **Lateral** (also called edgewise loops): connecting antiparallel strands close to each other (**Figure 5b,d**)
- **Diagonal:** connecting diagonally opposite strands in antiparallel conformation (**Figure 5c**)

Another aspect that influences the high polymorphic nature of G4s and differentiates them from B-DNA is the nature of the four grooves formed in each G4-tetrad.

Two adjacent guanines in a G4-tetrad give rise to the formation of grooves that can be (**Figure 6**) [3]:

- Medium: formed by two *anti* of two *syn* guanines
- Narrow: formed by an *anti* or *syn* guanine perpendicular to the *anti* guanine
- Wide: formed by an *anti* guanine perpendicular to the *syn* guanine

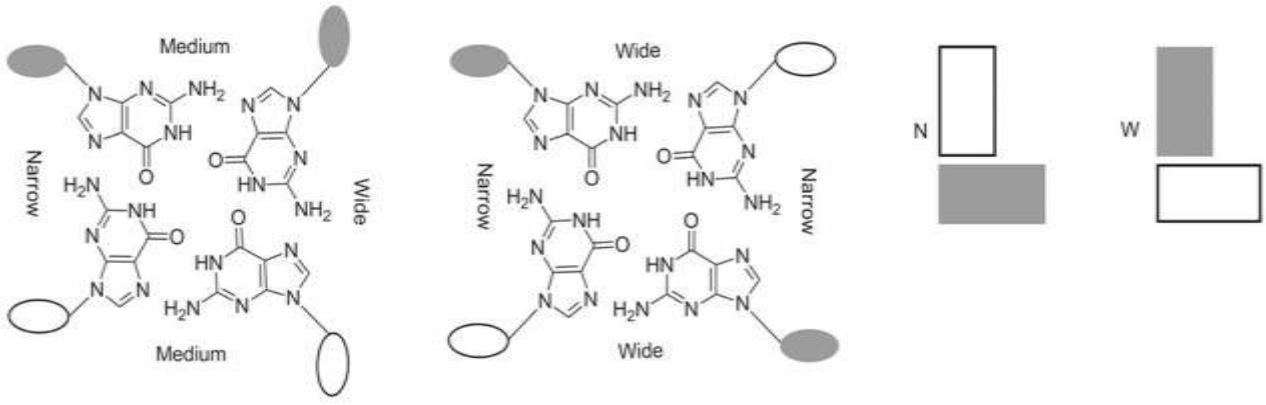


Figure 6: Schematic representation of G-tetrad core grooves. Grooves can be medium, narrow and wide depending on the glucosidic conformation of guanines (*anti* or *syn*) and their spatial position (parallel or perpendicular each other). Adapted from [3].

1.1.2 Distribution and biological role of G-quadruplex

Many efforts have been done in the study of G4s *in vitro* and since late 90s increasing evidences have suggested the possibility of a G4s key role also *in vivo* [6,7].

More than 350000 DNA sequences have the capability of forming G4s and many of them have been observed in telomeres and in promoter regions therefore suggesting their involvement in crucial biological processes [3].

Since G4s are highly dynamic structures they can be potentially formed both from ss-DNA and ds-DNA depending on denaturation mechanisms of the nucleic acid (which create a transient single-stranded DNA intermediate that provide the opportunity for G4 to fold) leading to the possibility of a concrete implication of G4s in the regulation of such mechanisms [20,21].

Since computational analyses demonstrated that G4s are not randomly distributed but are on the contrary located in key regions in the genome, G4s formation has been associated with the regulation of different biological functions such as transcription, translation but also DNA repair, DNA recombination, DNA replication, epigenetic mechanisms and alternative splicing via a *cis*-regulation activity [13,22].

The first evidence of a role of G4s in biological processes was obtained when for the first time G4s were found in the telomeric DNA [23]; these studies have opened the way to the discovery of their presence in many different regions of the human genome such as oncogene promoters, ribosomal DNA, DNA mini-satellites and the heavy chain switch region of immunoglobulines [12].

1.1.2.1 Effect of G-quadruplex during DNA replication

The folding of a G4 structure is a highly dynamic mechanism which requires the presence of a single-stranded nucleic acid or a temporary single-stranded nucleic acid. Of course the separation of a double stranded DNA to a single-stranded DNA occurs during biological processes such as replication or transcription. As a direct consequence of this statements, the presence of a G4 structure within a sequence can easily perturb these two mechanisms [21,24].

When DNA replication starts, replicative helicases unwind supercoiled DNA and separate the two strands allowing the leading and lagging strands synthesis. In this process, as already said before, DNA transiently assumes a single-stranded nature which represents the ideal setting for G4 formation [25].

Since 1990s, helicases have been proposed as necessary actors for the important role of DNA unwinding and some of them resulted active also in G4 unwinding (**Figure 7**) [21].

Deregulation of some helicases, active also on G4s *in vitro,* was correlated with diseases in humans: for example RecQ helicase WRN was associated with premature aging [26], FANCJ (Fanconi anemia

group J protein) and PIF1 deregulation were correlated with an increased incidence of cancer in patients [27,28,29]. PIF1 was one of the first helicases reported for its unwinding activity on G4 structures [27].

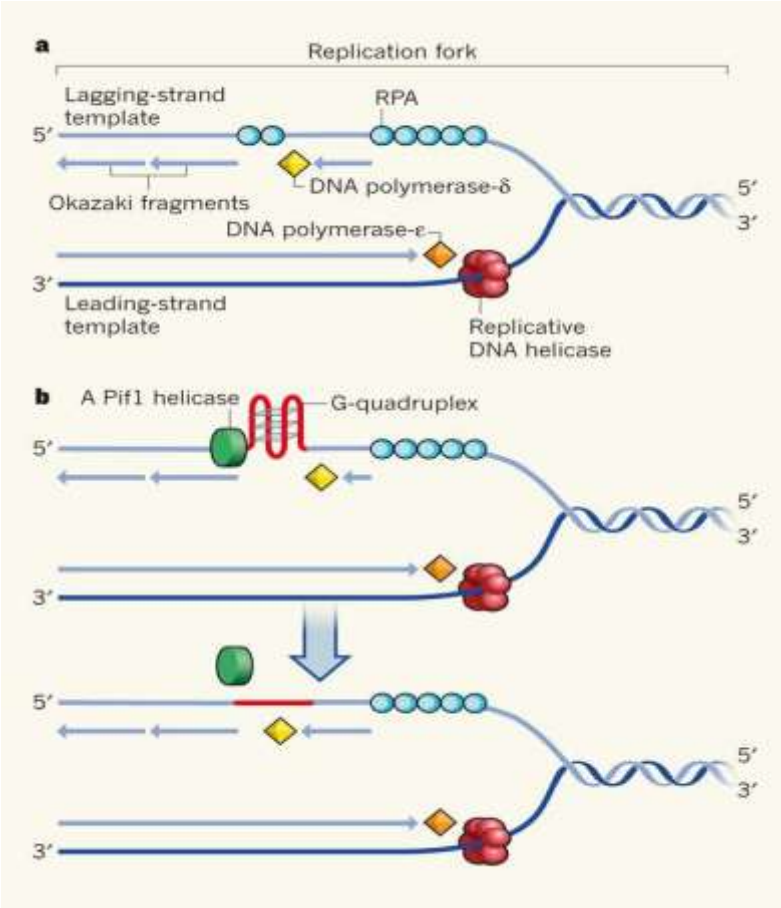


Figure 7: Schematic representation of the DNA replication process. (**a**) During DNA replication, helicases act to unwind the two strands into the leading and the lagging strands to give rise to the replication fork. Protein A (RPA) binds the lagging strand and progress DNA replication by DNA polymerase along the strand. (**b**) If the DNA sequence contains the G4, RPA is no more able to unwind DNA and therefore give the ideal setting for G4 formation. Adapted from [30].

1.1.2.2 Effect of G-quadruplex during transcription

G-quadruplex folded structures present in coding regions can influence transcription both in a positive or a negative way depending on their location. G4s can be located both in the leading and in the lagging strands. When a G4 is located in the leading strand of a coding region it blocks the transcription mechanism, whereas a G4 structure in the lagging strand (non-coding region) promote the formation of the G-loop which is the substrate for polymerase-mediated transcription (**Figure 8**) [31].

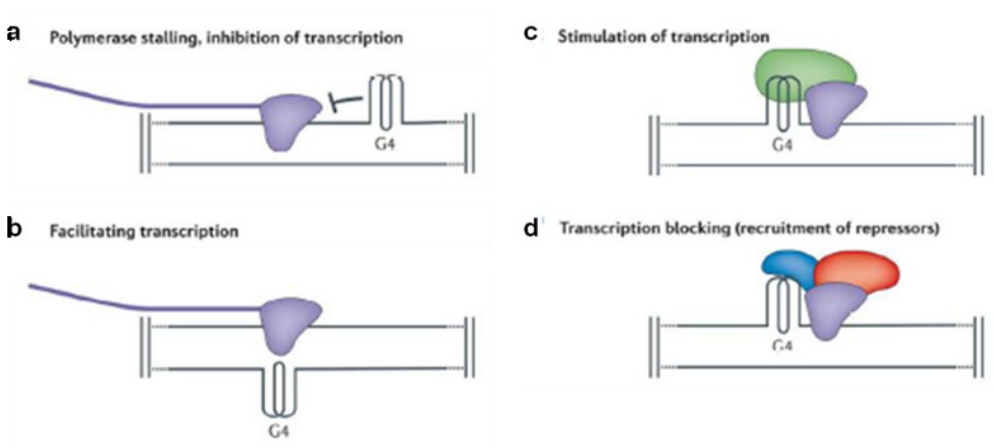


Figure 8: Schematic representation of the effect of G4s on the transcription process. (**a**) When a G4 is folded on the leading strand of a coding region it induces the polymerase stalling because of a steric hindrance which prevents transcription. (**b**) When the G4 is folded on the lagging strand (non-coding sequence) it favors transcription by making disposable for transcription the coding sequence on the leading strand. (**c-d**) By recruiting specific transcription factors, G4s can stimulate or repress transcription in an indirect mode. Adapted from [31].

1.1.2.3 G-quadruplex at telomeric ends

Telomeres, firstly described in 1938 by Hermann Muller and co-workers and defined from Greek "*telos* : end and *meros*: part", are nucleoprotein complexes (DNA and shelterin complex proteins) located at the end of eukaryotic chromosomes (**Figure 9**) [32].

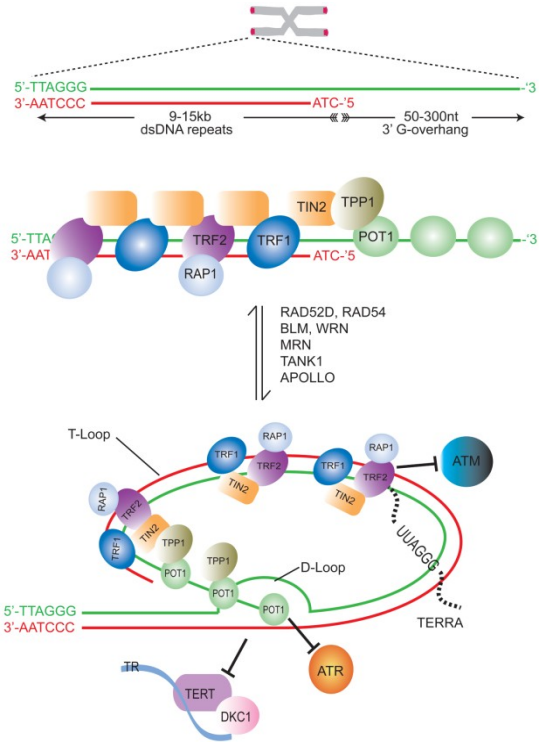


Figure 9: Schematic representation of the structure of human telomeres. Telomeres are formed by 15 kilobases with the tandem repetition of the motif d(TTAGGG) with a 3' G-rich overhang. The leading strand is rich in G- composition while the lagging strand in opposite rich in C- composition. Telomeres are composed of DNA and proteins of the shelterin complex: TRF1 and 2 consisting in the double stranded telomeric repeat binding factors, RAP1 which interacts with TRF2, TIN2 and TPP1 and the telomeric protection factor POT1 that all together cover the double and single-stranded repeats. Taken together all these interacting partners form a protective structure at the chromosome end (T-loop). Adapted from [33].

In human somatic cells telomeres are non-coding double-stranded DNA sequences composed of tandem repeats of the 5′- 3′ oriented motif d(TTAGGG) that can give rise to sequences of 15 kb in length and have a 3′-single-stranded overhang of 300-400 nucleotides (referred to as G-overhang) particularly rich in G-bases composition. Telomeres main function is focused on the preservation of chromosomal stability: they protect chromosomes in their end sequences from degradation and chromosomal end fusion to prevent cellular aging and the acquittal of genetic diseases [34].

The shelterin complex consists of (i) the two double stranded telomeric repeats binding factors TRF1 and TRF2, (ii) RAP1 which is an interacting partner of TRF2, (iii) the two bridging proteins TIN2 and TPP1 and, finally, (iv) the protecting factor POT1. Altogether these interacting factors (i) cover the double and single-stranded repeats at the end of each telomere (T-loop region) to protect telomeric extremities and (ii) transiently interact with other factors that interact with telomeres in a cell-cycle dependent manner [33]. T-loop region protection is necessary for cells survival: T-loop destabilization (which results in the exposure of the 3′ end) is recognized as DNA damage signal which starts a senesce mechanism that therefore results in cell death [35].

Telomeres are essential for chromosome stability and for genomic integrity: they provide sites for recombination events and transcriptional silencing and they play a central role in cellular aging and cancer development [33]. Cancer development is a direct consequence of the "end replication problem": the lagging strand of the chromosome is replicated following a semi-conservative replication process which results in a progressive shortening (loss of 50- 100 bp of telomeric DNA) of the telomeric end at each cell cycle (**Figure 10a,b**).

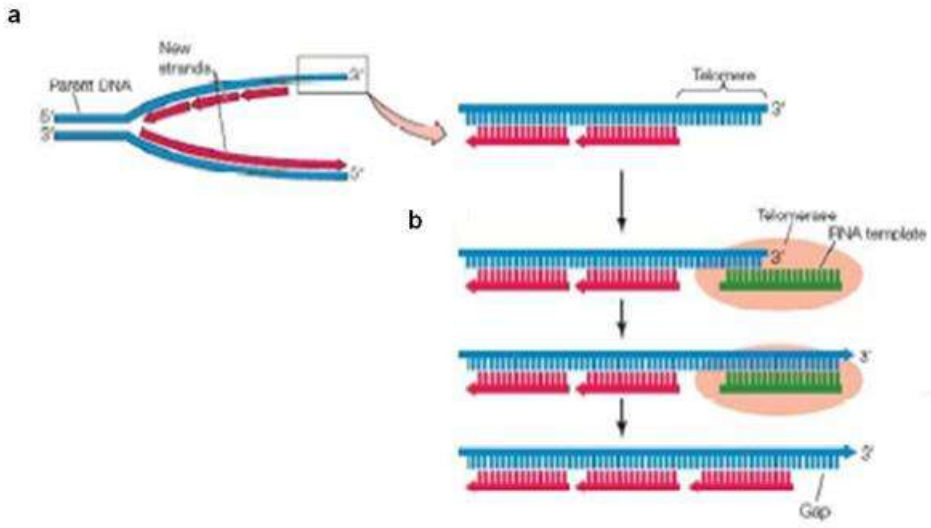


Figure 10: Schematic representation of the mechanism of telomeric ends replication. (a) Semi- conservative replication of the lagging strand which results in a progressive shortening of the end at each cell- cycle. (b) Telomerase drives the synthesis of the G-rich sequence forming the 3′ overhang by using a single- stranded RNA template (hTERC). Adapted from [36].

This mechanism was attributed to the inability of DNA polymerase to completely replicate the ends of chromosomal DNA at the S phase. When a normal cell reach a critical telomere length referred to as Hayflick limit, it exits the cell cycle, enters the M2 phase (mortality stage) and undergoes the cellular senescence: this well-regulated mechanism is the "clock" which determines the human life span [37,38].

Different cell types are correlated with different telomeres lengths: normal somatic cells have telomeres of about 10 kb, stem cells in renewal tissues and fetal tissues have telomeric ends about 12 kb with a reduced shortening rate, cancer cells have an average of telomeric dimension from 2-9 kb. In this case their maintenance and slowed down shortening are linked to telomerases reactivation [38]. For this reason telomerase is thought to be overexpressed in ~85-90% of cancer cells and primary tumors, but not in normal somatic cells [10]. This clearly suggest how telomeres shortening in cancer cells is necessary: telomeres integrity loss, associated with p53 and Rb tumor suppressors activation, is necessary to enter cells in senescence in order to suppress cancer evolution (**Figure 11**) [12,39].

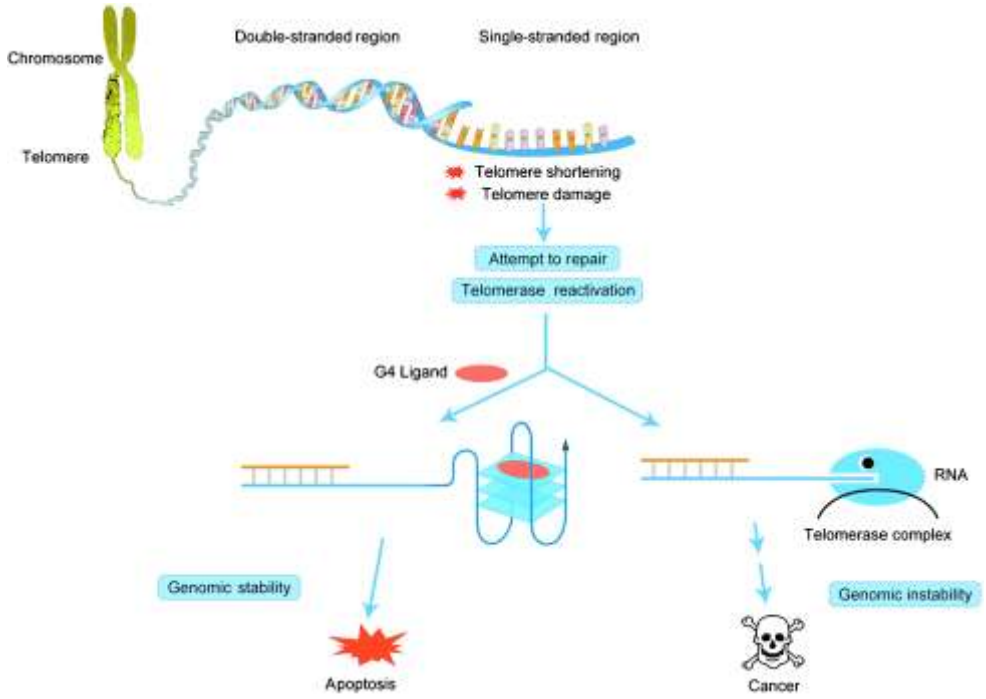


Figure 11: Telomeres shortening or damage can reactivate telomeres enzymes or repair mechanisms, this triggers tumorigenesis. Induction of stabilization of telomeric G4s can therefore block telomerases from binding to telomeric ends. Adapted from [12].

The single-stranded G- rich overhang can fold in a G4 structure in a highly dynamic equilibrium with its T-loop conformation (relaxed). In 1993 it was demonstrated that telomeric intramolecular G4 sequence, d[AGGG(TTAGGG)$_3$], can reach two different stable conformations both in K^+ and in Na^+ [40]. In Na^+ an intramolecular antiparallel G4 is formed with both diagonal and lateral loops

(**Figure 12a**), while in K^+ a parallel G4 with double-chain-reversal loops is formed (**Figure 12b**)[41,42]. Many other different conformations have been addressed to the telomeric G4 in the presence of K^+: (i) two intramolecular mixed–type G4s and (ii) two bimolecular G4s (one parallel with diagonal loops and one antiparallel with lateral loops)(**Figure 12c,d**) [41,43].

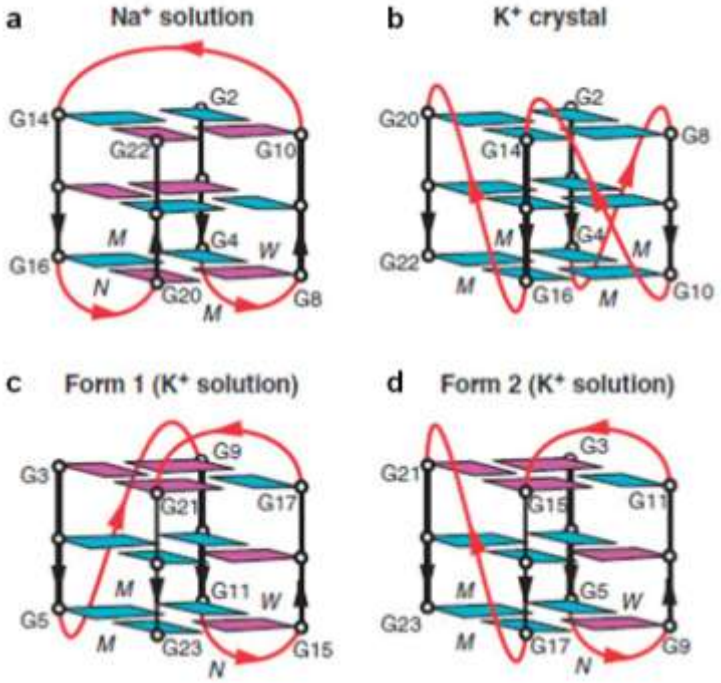


Figure 12: Schematic structures of intramolecular G-quadruplexes formed by the human telomeric sequences. (**a**) G4 telomeric sequence in Na+ solution; (**b**) G4 temoleric sequence in K+ -containing crystal; (**c**) G4 telomeric sequence in K+ solution: natural form 1; (**d**) G4 telomeric sequence in K+ solution: natural form 2. Telomeric G4 is based on the sequence d[AGGG(TTAGGG)$_3$] Adapted from [41].

Docking studies have interestingly demonstrated the possibility to target with different enantiomers different G4 structures: for some Telomestatin enantiomers a different binding mode, depending on the G4 polymorphism, was demonstrated [44].

Recent studies, moreover, demonstrated that naturally occurring mutations within loop sequences of the telomeric G4s can therefore alter the stability and the conformation of the G4 depending on the position of the mutated nucleotide [45]. All together these information obtained for telomeric G4s conformations opened the possibility of a structure-based rational drug design that could be applicable for all known G4s [12,46].

Since telomeric G4s have a major function in protecting G-overhang from degradation by nucleases and in blocking telomerase activity, for this reason telomeric G4s are currently seen as promising candidate for anti-cancer G4-mediated therapy. Their ability to block telomerase activity previously seen in an *in vitro* study [47] is probably linked to the steric hindrance of the G4 which results in the prevention of hTERT annealing to the G-overhang. Moreover, the folding and

stabilization of the telomeric G4 mediated by specific ligands, can act as an alternative protection mechanism to overcome not only genetic instability due to chromosome shortening but also telomeres elongation in cancer cells. Accordingly to this, many different G4-ligands resulted promising in *in vitro* studies: in particular B-19 (an acridine compound) induced a rapid senescence in cancer cells due to (i) the displacement of POT1, the protective protein of the shelterin complex, (ii) the induction of DNA damage but also (iii) the displacement of the hTERT enzyme [48]. Unfortunately, as many other G4-ligands proposed for therapy that are not enough drug-like, this compound did not progress into clinical. This clearly opens the necessity to continue in the research for new G4-ligands applicable in therapy.

1.1.2.4 G-quadruplex in gene promoters

Given the promising opportunity to use telomeric G4s as novel anti-cancer targets, many efforts have been done in the past decade in order to elucidate the presence of G4s in gene promoters [49]. Studies on human gene promoters and correlated bioinformatic analyses have shown the presence of putative G4-forming sequences not only in human genes but also in many other different species (i.e. yeasts, plants and bacteria) therefore suggesting a transversal role of these structures within genomes [50,51,52]. More deeper studies have also raised the over-representation of G4-motifs near binding site for transcription elements, such as Sp1 [53].

In about 50% of human genes G4 structures have been found upstream or downstream the Transcription Start Site (TSS) therefore suggesting an involvement in the regulation of gene expression. Additionally it was observed that tumor-suppression genes possess a low "G4-forming potential (G4P)"of sequences whereas proto-oncogenes have a high G4P: this clearly confirm a direct enrolment of G4s in the up-regulation of cancer-related genes [54]. Further investigations about G4s in oncogenes have elucidated the presence of G4s in promoter regions of several genes correlated with the six hallmarks of cancer (**Figure 13**) [55].

The most representative promoters in which G4s have been identified were *c-myc* [56], *bcl2* [57], *VEGF* [58], *c-kit* [59], *hTERT* [60] and *PDGF-A* [11].

As all other G4s previously identified, also G4s in promotorial regions are subjected to polymorphism which is associated with (i) the sequence composition, (ii) the number of tetrads that form the G4 and (iii) the composition and length of loop sequences. This vast diversity in the folding can be exploited for specific drug targeting [55]. As already said, the great variability between all these secondary structures is seen in loops, in particular in loop length and sequence composition. Loops in these G4s have been seen to vary from 1 nt (the minimum required) up to 26 nt (loop which forms an independent secondary structure in *hTERT* promoter) [55].

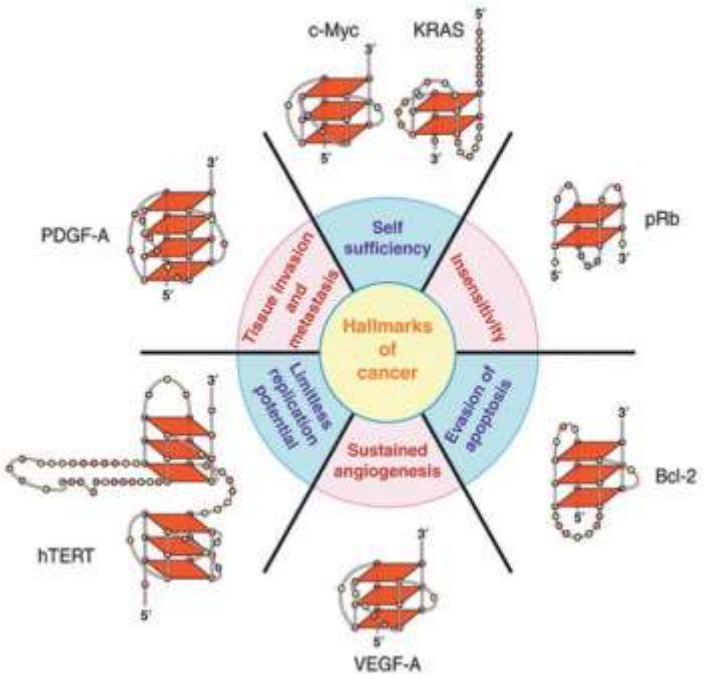


Figure 13: Schematic representation of G4s correlated with the six hallmarks of cancer. Each of these G4 have been found in the promoter region of the corresponding genes. As already described, each G4 is characterized by a different folding pattern which is influenced by (i) the number of tetrads that take part to the G4, (ii) the loop sequence and composition and (iii) the sequence of the G-rich region that folds into a G4. Adapted from [55].

Since G4s in the promoter region of different genes have been extensively studied, many efforts have been done to draw up a classification which took into account the huge polymorphism to which these structures were subjected [55]:

- Class I: made of a single G4, *c-myc*, characterized by loops isomers whose sizes can easily vary (**Figure 14a**). Its presence results in gene silencing.
- Class II: represented by two different G4s separated each other from three turn of DNA. The main sequence of this class is the *c-kit* promoter which biological consequence is gene expression inhibition (**Figure 14b**).
- Class III: in this class a pair of G4s are present and are sufficiently near one another to form a tandem of G4s more stable that each G4 alone. This class is represented by *c-myb* and *hTERT* G4s (**Figure 14c**) that cause inhibition of gene expression.
- Class IV: in this class multiple overlapping G4s are present in a dynamic equilibrium. *Bcl-2* is the main G4 of this class (**Figure 14d**)

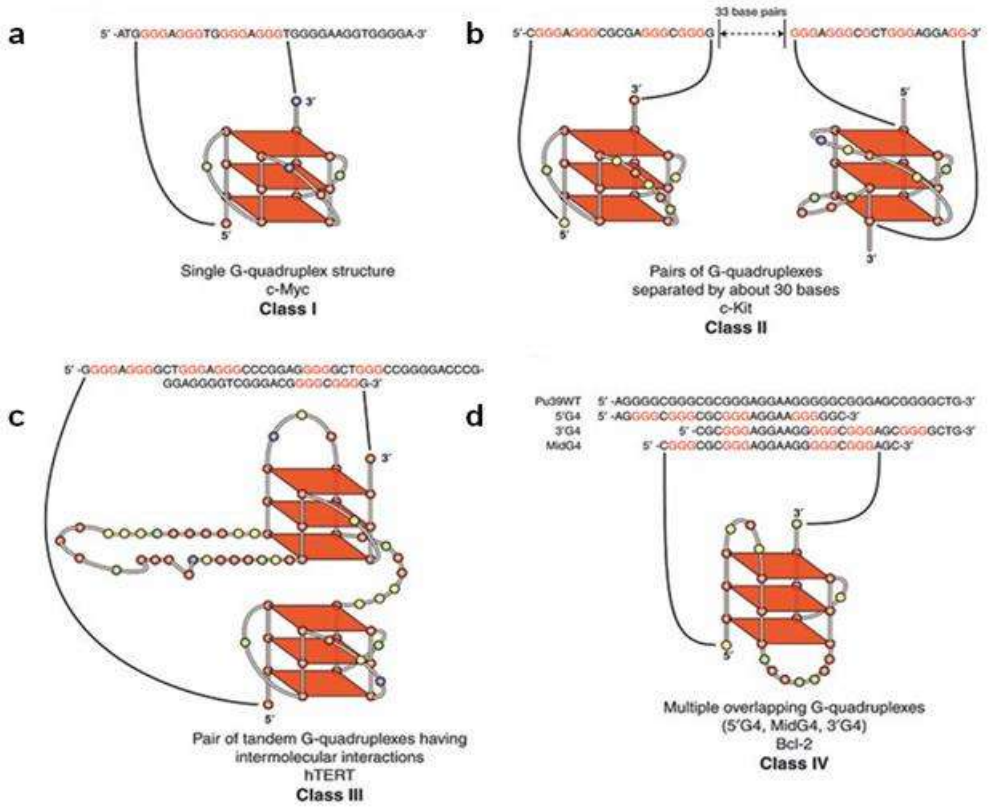


Figure 14: Schematic representation of the classification of G-quadruplexes formed within promoter regions. (**a**) Class I: single G4 structure within the promoter, represented by G4 found in *c-myc* promoter region. (**b**) Class II: pair of different G4s separated by three turns of DNA, represented by G4 folded in the *c-kit* promoter region. (**c**) Class III: tandem of G4s formed in adjacent sequences, represented by *c-myb* and *hTERT* G4s. (**d**) Class IV: multiple overlapping and stable G4s, represented by the G4 found in *bcl-2* promoter region. Adapted from [55].

Increasing evidences support the thesis that G4 structures in promoter regions can regulate repression or activation of gene expression and may have an essential role in gene transcription regulation (**Figure 15**) [10].

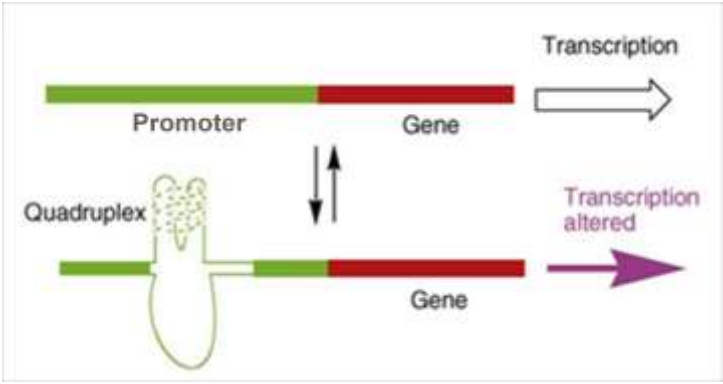


Figure 15: Schematic representation of the promoter- G4 hypothesis. Once a G4 is folded within the promotorial region of a gene, it induces an alteration of the transcription. This hypothesis would suggest that any molecule able to bind and to stabilize a G4 can therefore modulate the transcriptional activity of the associated gene. Adapted from [10].

As an example: the aberrant overexpression of the *c-myc* oncogene induces a potentiation of cellular proliferation and inhibition of differentiation which correlates with a variety of cancer malignancies including breast cancer, carcinoma of the colon, cervix cancer, small-cell lung cancer, osteosarcomas, glioblastomas and myeloid leukemia in which the correlated protein c-MYC is overexpressed (80%) [61]. It has been demonstrated the presence of a G4-forming sequence within the nuclease hypersensitivity element (NHE) III_1 region which is formed when the (NHE) III_1 region is in a single-stranded context and when no proteins are bound. G4 formation favored in these conditions negatively regulates the expression of *c-myc* by preventing its transcription [56]. When a base mutation occurs in the G4-forming sequence and G4 formation is lost, an increase in the expression of *c-myc* and therefore an increase in cancer incidence are seen. It was positively demonstrated that targeting (NHE) III_1 G4 with G4-ligands (i.e. the cationic porphirin TMPyP4 and the analogue of the antineoplastic drug Ellipticine GQC-05) to stabilize the non-canonical secondary structure, resulted in the downregulation of gene expression and a reduction in the transcribed c-MYC mRNA levels [56,62].

Two additional G4s were found in two G-rich sequences within the promoter region of the human *c-kit* protooncogene (*c-kit* native and *c-kit 2*). *C-kit* oncogene encodes for a membrane protein of the thyrosine kinase family of proteins: it consists of a glycoprotein that can stimulate cell proliferation, differentiation, migration and survival. Since these functions are closely related with cell division, regulation of *c-kit* gene overexpression plays a crucial role in human neoplastic malignancies [12]. Eighty-seven base pairs upstream the transcription start site of the *c-kit* gene, in the native sequence, a single intramolecular very stable G4 is formed in K^+ solution [63]. Also for these G4s it has been proposed a G4-ligand mediated therapy to interphere with overexpression of this gene in cancer cells [59].

1.1.2.5 G-quadruplex in coding regions

G-rich sequences were found within other regions than promoters such as minisatellites [64], immunoglobulin heavy chain switch region [65,66] and in human ribosomal DNA (rDNA) [67] and were proposed to be the target of many different binding proteins. Generally G4-folding sequences are mainly found in the leading strand at the 5' end near the UTR region, less frequently they are found in the lagging strand [68]. Recent studies elucidated the presence of G4-forming sequences also in coding regions and on both the leading and the lagging strands of DNA [20]. Very little is known about the mechanism that arises when a G4 structure is encountered by RNA polymerase while processing mRNA (**Figure 16**). The only one evidence concerning the G4 structure interference with transcription was reported for a G4 folded structure occurring in the c-MYB

protooncogene: it was demonstrated that *in vitro* this G4 can block T7 RNA polymerase during transcription therefore maintaining this possibility also *in vivo* [69].

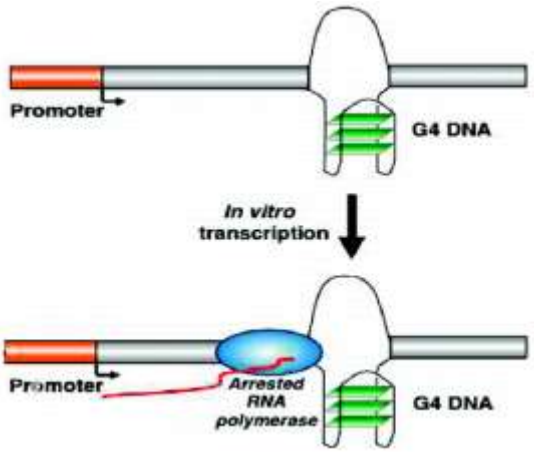


Figure 16: Schematic representation of the mechanism of transcription arrest mediated by the folding of a G4 in a coding region. When a G4 is present the RNA polymerase can not process DNA to mRNA. The first evidence of this mechanism was demonstrated in vitro with a G4 structure folded in the coding region of the c-MYB protooncogene. Adapted from [69].

Studies on G4s folded within coding region of such genes, induces the observation of the presence of a G-loop on the lagging strand at the position where the G4 structure was folded on the leading strand both in vitro and in E. coli [66,70]. These structures were reported for their crucial role: G-loops could maintain accessible the DNA for transcription of other genes by preventing the annealing to the complementary sequence. In this sense the folding of G4 sequences in some coding region can therefore promote transcription of certain genes (**Figure 17**) [71].

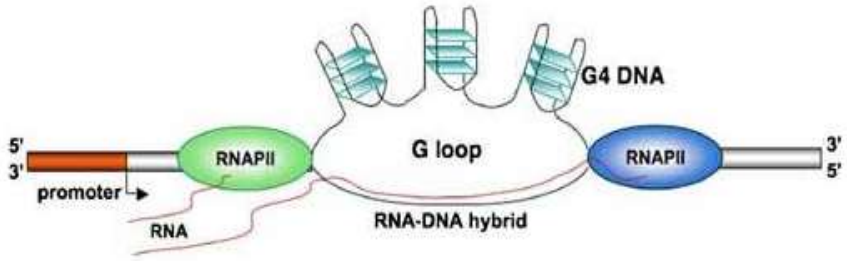


Figure 17: Schematic representation of the induction of transcription mediated by G4 folding on the leading strand. Once the G4 is formed on the leading strand a G-loops is consequently folded on the lagging strand therefore promoting the transcription of coding regions located on this strand. Adapted from [72].

1.1.3 G-quadruplex as new potential therapeutic target: G4-ligands

In recent years G4s have emerged as novel and promising target in the anti-cancer therapy, primarily to block telomerase activity and to inhibit G4s related oncogene expression [12,13].

The very peculiar structure of G4 favored the rational design and development of a diverse array of innovative small molecules (referred to as G4-ligands) with the capability to bind and to stabilize these alternative structures [3].

It was demonstrated that the binding between a G4 and a G4-ligand can happen via different binding modes depending on the structure of both the G4 and the G4-ligand. Many studies have focused their attention on the development of small molecules with a known mechanism of interaction with G4s but the exact interaction mode is still undefined in some cases [12]. Studies on G4/G4-ligands interactions have identified different binding modes expressed by G4-ligands (**Figure 18**) [10,50]:

- $\pi-\pi$ stacking on the terminal G4-tetrads (external stacking) and electrostatic interactions which are typical of flat aromatic molecules
- groove, loops and negatively charged phosphate backbone binding molecules [73]
- combination of two or more binding modes previously described

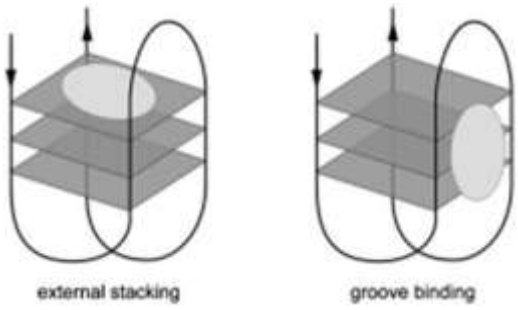


Figure 18: Schematic representation of the mechanism of binding between G4s and G4-ligands. G4-ligands can interact with G4s via (i) an external stacking on the terminal G4-tetrads, or (ii) a groove or loops binding. Adapted from [12].

G4-ligands are mainly characterized by cationic charges and a wide planar aromatic surface that favors the binding on the terminal G4-tetrad rather than intercalation between tetrads. This is due to the fact that G4s are highly structured nucleic acids that would need an extremely high energy to disrupt their structure to allow the intercalation of an exogenous molecule (the G4-ligand). Groove and loops binding are generally mainly mediated by G4-ligands which present substituents as short acyclic chains terminating in a cationic nitrogen-containing group (i.e. diethylamine, pyrrolidine and piperidine), this gives the ability to take contact with negatively charged phosphate groups present in the G4-groove [12].

Finally, some small G4-ligands made of a long positively charged side chain have showed the ability (i) to lie at the central channel of the G-tetrad and (ii) to strong stabilize the G4 structure [3].
The most famous G4-ligands that have shown encouraging anticancer activity *in vitro*, *in vivo* and in clinical trials included the fluoroquinolone-based compound Quarfloxin (CX-3543) [74], perylenes such as PIPER [75], porphyrins such as TMPyP4 [76], trisubstituted acridines such as B-19 [77] and natural macrocycles such as Telomestatin [78].
In general G4-ligands can be classified into different categories depending on their chemical features:

- Anthraquinones
- Acridine analogues
- Porphyrins
- Quindoline and quindoline derivatives
- Telomestatin
- Naphthalene diimides

1.1.3.1 Anthraquinones

This group of G4-ligands interacts with the G-tetrads through a π–π stacking. The first anthraquinone synthesized was BSU-1051, obtained by modification of the anthraquinone main scaffold (**Figure 19a**), was a 2,6-disubstituted aminoalkylamido anthraquinone which displayed a telomerase activity in the micromolar range (IC_{50} 23 μM) (**Figure 19b**) [79].

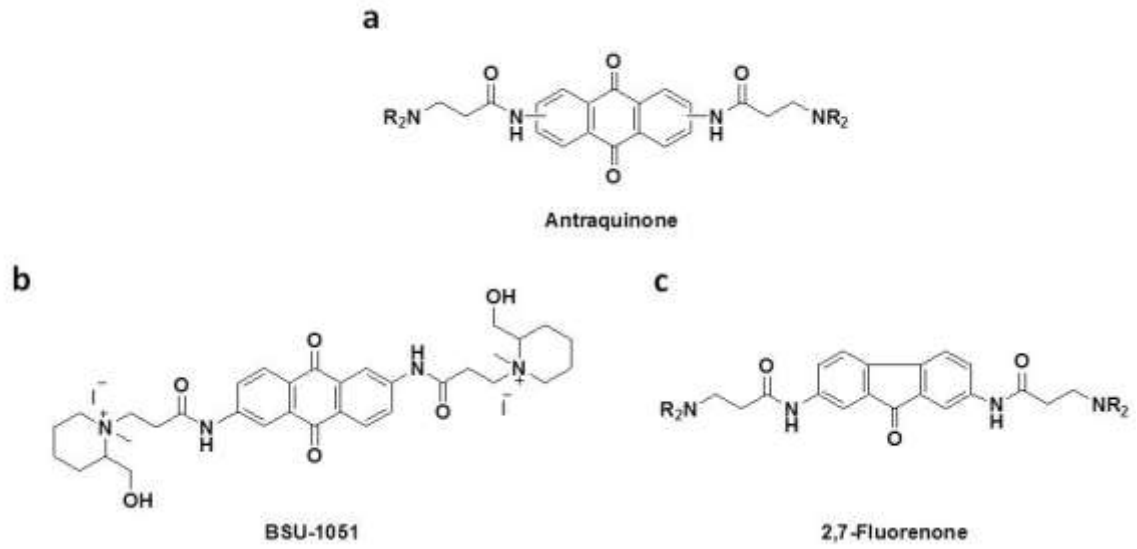


Figure 19: (**a**) Chemical structure of the anthraquinone scaffold. (**b-c**) Chemical structures of the first anthraquinone synthesized, BSU-1051 and of the 2,7-fuorenone used as scaffold for the synthesis of new anthraquinones with increased activity against telomerase and decreased cytotoxicity against cell lines. Adapted from [79].

Then, a huge number of 2,7- fluorenone (2,7-FO) analogues were designed to reduce the cytotoxicity and increase the activity against telomerase (the most potent compound showed an IC_{50} around 8-12 μM against telomerase) with low cytotoxicity on tumor cell lines (**Figure19c**) [12].

1.1.3.2 Acridines

Studies aimed to develop novel G4-ligands with better features than anthraquinones led to the discovery of a new series of chromophores better known as acridines. Differently from anthraquinones, acridines were characterized by the presence of a nitrogen atom in the heterocyclic scaffold which could be protonated at physiological conditions and that exerted an increased interaction with G4 due to the electron deficiency [73]. The first acridine moieties that were synthesized were 3,6-disubstituted then replaced with more potent 3,6,9-trisubstituted acridines.
Effectively the 3,6-disubstituted analogs were poorly able to discriminate between G4s and duple-DNA (same binding constant between the two species) and therefore elicit a low activity against telomerase in comparison with anthraquinones [12]. For these reasons 3,6,9-trisubstituted acridines were synthesized: they showed an encouraging increased inhibitory activity against telomerase in the range of 10-20 nM and, additionally, they were able to selectively bind to G4s species via a π–π external stacking and a groove binding mediated by tertiary amine moieties as side chains [80]. The main G4-ligand representing this class of 3,6,9-trisubstituted acridines was the acridine derivative *N,N'*-(9-((4-(dymethylamino)phenyl)amino)acridine-3,6-diyl)bis)3-(pyrrolidin-1-yl)propan-amide) better known as B-19 (**Figure 20a**) [81]. It has been extensively investigated for its ability to stabilize telomeric G4s and resulted a strong inhibitor of telomerase *in vitro* as demonstrated by TRAP assay [82]. Moreover further studies demonstrated a promising activity of this G4-ligand in the inhibition of cancer cell proliferation *in vitro* and in the arresting of tumor growth *in vivo* with a mechanism telomerase inhibition-mediated (activity in the low nanomolar range) [83]. Other important acridine derivatives which expressed a promising activity against G4s were (i) RHSP4, a pentacyclic acridine 3,11-difluoro-6,8,13-trimethyl(8*H*)-quino[4,3,2-*kl*]acridinium methylsulfate, which showed a potent telomerase inhibition (IC_{50} about 330 nM) (**Figure 20b**) and (ii) BSU-6030 which showed a 70-100 fold greater selectivity than B-19 (**Figure 20c**). Both B-19 and RHSP4 entered in a preclinical development stage [12].

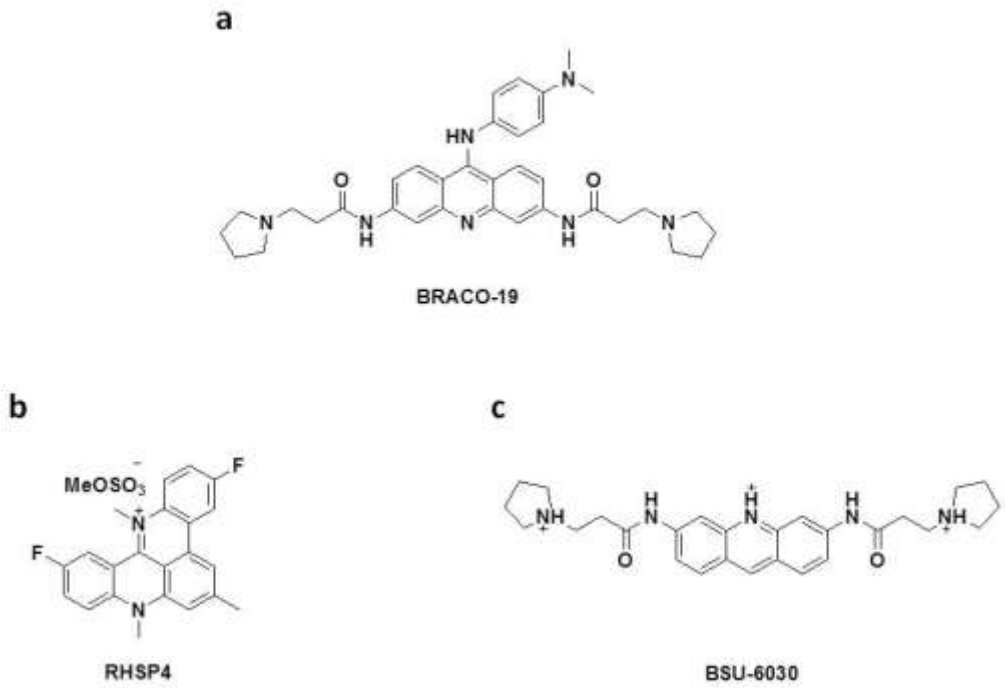


Figure 20: (**a**) Chemical structure of B-19, the main member of the acridine derivative family of G4-ligands. (**b-c**) Chemical structures of other acridine derivatives: RHSP4 and BSU-6030. All compounds showed a promising activity against telomerase *in vitro*. B-19 and RHSP4 entered the preclinical development stage. Adapted from [84].

A possible use of B-19 in a viral environment was first analyzed for the EBV virus in order to elucidate the role and the structure of EBNA1. In this study it was demonstrated that B-19 was able to (i) stabilize RNA G4s, (ii) to reduce the number of viral genome copies produced within an infection cycle and (iii) to reduce the transcription levels of EBNA2 and EBNA3A [85]. Additionally, B-19 was then applied for studies on HIV-1 and HSV-1 G4s and in both cases it showed a promising antiviral activity in the micromolar range. Concerning HIV-1, it was demonstrated that B-19 was able to act both at the reverse transcription step and at the post-integration level during the viral life cycle. Treatment with B-19 induced a decrease in the promoter activity up to 70% while no activity was measured for the G4-ligand when the G4 structure was mutated. The treatment of HIV-1 infected cells induced a strong and significant decrease in the viral titer [86].
Moreover, on HSV-1 it was demonstrated that since this virus contains a high guanine content it is possible to target G4-folding sequences with B-19 in order to inhibit the viral production. Treatment with B-19 of HSV-1 infected cells induced inhibition of viral DNA synthesis and late protein production [87].

1.1.3.3 Porphyrins

Porphyrins are *N*-Methylated ligands characterized by the presence of a central aromatic ring in which four methylated nitrogen groups are located at the four corners of the aromatic planar surface. This family of compounds has been extensively used as G4-ligands thanks to their low electron density that improved water solubility. Moreover, the planar arrangement of the aromatic rings in the porphyrin moiety strongly induced the binding of G4s via an external π-stacking rather than intercalation between G4-tetrads [88].

TMPyP4, the 5,10,15,20-tetrakis-(*N*-methyl-4-pyridyl) porphine (**Figure 21a**), represents the main molecule of this family and was proposed as G4-ligand because of its appropriate physical properties, i.e. molecular size, planar core, positive charges and hydrophobicity. Biophysical analyses demonstrated that TMPyP4, and other molecules of the family, were able to bind and stabilize both parallel and antiparallel G4s with a greater selectivity toward G4 structure over duplex DNA [89].

TMPyP4 was proved for its ability to (i) inhibit telomerase (IC_{50}=6.5 ± 1.4 μM) and (ii) downregulate the proto-oncogene *c-myc* and KRAS promoter expression by a G4-stabilization mechanisms that resulted in an *in vivo* antitumor activity [90]. Our group showed that TMPyP4 was able to stabilize G4s in the HIV-1 *Nef* coding region and to favor and induce their formation within the double-helix conformation [91]. The G4-ligand/G4 binding in the *c-myc* promoter occurred by stacking both in the correspondence of the external TTA loop of the G4 structure or in the 5′ region of the stacked quadruplex [92]. Since the TMPyP4 stacking on the G4 was mediated by the hydrogen bonded base pairs, that were not involved in the formation of the G4, and was not mediated by bases involved in the G-tetrad formation, this specified the lower selectivity of the molecule for quadruplex DNA over duplex DNA [12].

TMPyP4 binding on one external G-tetrad of a parallel G-quadruplex was proved by the complex between TMPyP4 and the parallel G4 formed within the *c-myc* promoter [3].

A conformational isomer obtained from TMPyP4, with a *N*-Methyl group in the sterically hindered 2-position, referred to as TMPyP2 (**Figure 21b**), showed a very weak activity and was generally used as porphyrin with much lower activity in the G4 binding [12]. Moreover a different G4 binding mode, unable to produce a biological effect, was proved for this molecule.

Studies conducted on tumor cell lines demonstrated that TMPyP4 had a greater effect in slowing cells growth, more pronounced than TMPyP2. Additionally it was demonstrated that TMPyP4 induces anaphase chromosomal bridges in sea urchins. This result indicated that G4-ligands might target telomeres directly inside cells [93].

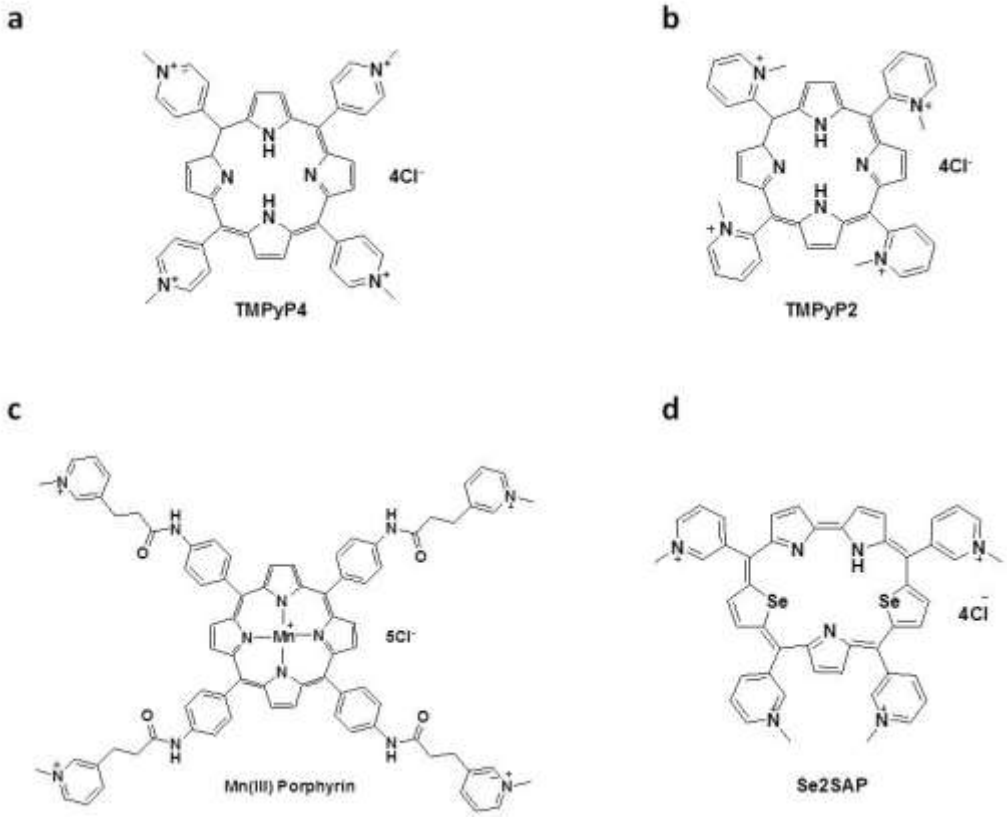


Figure 21: Chemical structures of the cationic porphiryns (**a**) TMPyP4, (**b**) its conformational isomer TMPyP2, (**c**) Mn(III) porphyrin derivative and (**d**) Se2SAP core-extended porphyrin. Adapted from [90].

Mn(III) porphyrin derivatives (**Figure 21c**) were synthesized to overcome the TMPyP4 restricted selectivity on G4 binding and to better increase the differentiation between G4 and duplex DNA. Mn(III) structure was composed of a central porphyrin ring and four flexible arms carrying methylated nitrogen groups and displayed an ability in the discrimination between G4 and duplex DNA of a 4 orders of magnitude than previous porphyrins [94]. The complex between Mn(III) and the human telomeric G4 was resolved and showed that the porphyrin ring was stacked on the terminal G-tetrad and the four lateral and flexible arms interacted with G4 grooves [3]. Then, to further increase the therapeutic selectivity of porphyrins the core-modified expanded porphyrin analogue Se2SAP (5,10,15,20-[tetra(*N*-methyl-3-pyridil)]-28,28-deselenasapphyrin chloride) (**Figure 21d**), was synthesized [12].

1.1.3.4 Quindoline and other molecules

Quindoline molecule (**Figure 22a**), a derivative molecule of the natural product criptolepine has been proposed for its ability to (i) display a higher affinity for G4s than other DNA structures [95], to (ii) bind the intramolecular G4 DNA of the telomere and therefore to (iii) inhibit the human telomerase [96]. Additionally more recent works proposed quindoline as a promising G4-ligand able to stabilize the G4 formed in the *c-myc* promoter and thus able to inhibit the expression of c-MYC mRNAs in the

hepatocellular carcinoma cell line H2p G2 [97]. Recently the NMR solution structure of a 2:1 complex of the small molecule quindoline with the major G4 formed in the *c-myc* promoter was resolved: this was the first representation of the complex structure of a biologically relevant unimolecular promoter with a G4-ligand [98]. As the *c-myc* promoter G4 was a prototype of the family of G4s of promoter regions, the complex structure provided new important insights into the structure-based rational design of drugs able to bind to unimolecular parallel G4s commonly found in promoter elements [9].

Moreover the structure of the complex between *c-myc* G4 and quindoline showed an unexpected drug-induced reorientation of the flanking sequences at both ends of the DNA sequence involved in the G4. The reorientation, referred to as "induced intercalated triad pocket" was previously described for riboswitches and was significantly different from previously proposed binding models for G4-ligands (i.e. binding mode of telomestatin or TMPyP4) which involved the external stack on G-tetrads [99].

Quindoline disubstituted derivatives, such as (2-(4-10*H*-indolo[3,2-*b*]quinolin-11-yl)piperazin-1-yl)-*N,N*-dymethilanamine, GSA-0820, were previously described for their good-cell free profile in the stabilization of G4s (**Figure 22b**). In particular GSA-0820 significantly down-regulated c-MYC mRNA on *c-myc* promoter with a G4-unreleated mechanism [100]; on the contrary GSA-0820 was reported for its stabilizing activity on a competing G4 formed on SMN2 Exon 7 after splicing mechanisms [101]. Actual studies are focusing on the elucidation of G4-related anticancer mechanisms of this compound.

NSC338258 compound, also known as GQC-05 (**Figure 22c**), was a compound derived from the plant alkaloid ellipticine (ellipticine dihydrochloride) that was reported and extensively used for its high affinity and binding activity on *c-myc* G4 able to maintain an "off switch" DNA conformation [102]. In Burkitt's Lymphoma cells, GQC-05 induced a pronounced decrease in c-MYC mRNAs and an altered pattern of protein binding to the NHE III_1 region due to the stabilization of the G4 in that region [62,102]. Additionally, GQC-05, was reported for its dramatic cytotoxicity against myeloma cells mediated by the cell cycle arrest and therefore apoptosis progression [103].

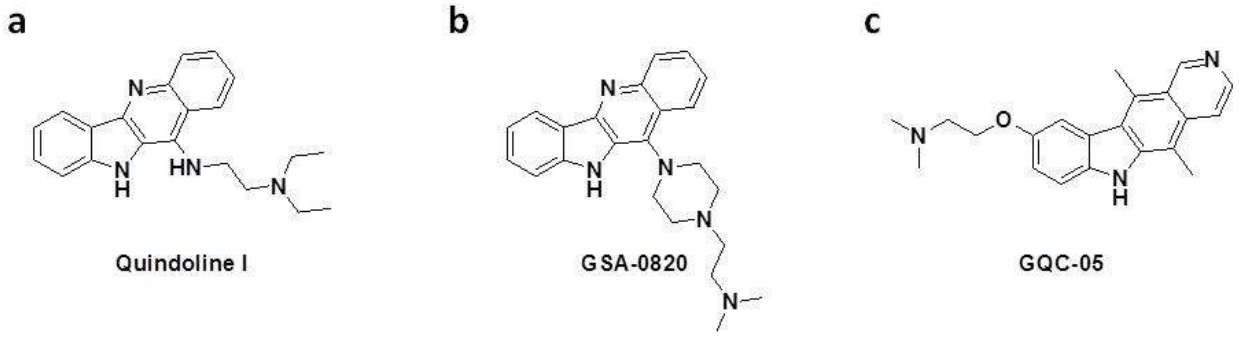


Figure 22: Chemical structures of (a) Quindoline, (b) GSA-0820 quindoline derivative and (c) GQC-05 elliptine analogue molecules. Adapted from [95,100,102].

1.1.3.5 Telomestatin

Another G4-ligand that expressed an interesting activity against telomerase was the natural small molecule Telomestatin (SOT-095). This compound, a neutral macrociclyc G4 ligand, consisted of seven oxazole rings and one thiazoline ring isolated from *Streptomyces annulatus* 3533-SV4 (**Figure 23a**). It was tested against both duplex DNA and G4s and showed a 70-fold stronger binding affinity for G4s over duplex DNA [12,104]. Telomestatin has been for a long time the most efficient telomerase inhibitor *in vitro* with an IC_{50} in the low nanomolar range and a strong selectivity due to a perfect shape which was able to adapt on the G4 structure. Since telomestatin was recently reported for its disadvantageous synthesis hardly compatible with large-scale production, some synthetic derivatives such as macrocyclic hexaoxazole (HXDV) and bistrioxazole acetate have been developed (**Figure 23b,c**) [105].

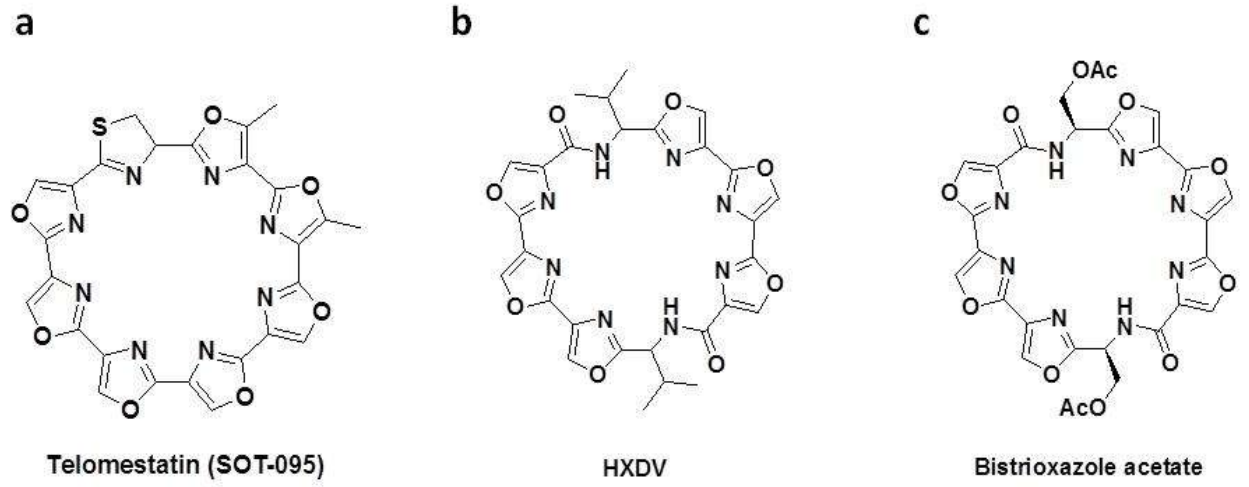


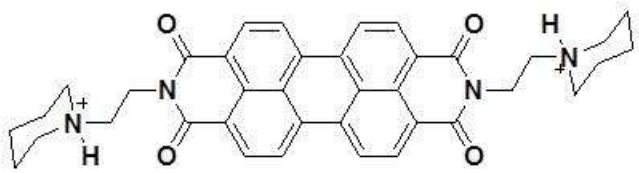
Figure 23: (a) Chemical structure of the natural macrocyclic G4 ligand Telomestatin, isolated from *Strptomyces annulatus*. It has been reported as the most effective compound against telomerase with an IC_{50} in the low nanomolar range. (**b-c**) Chemical structures of telomestatin- analogues HXDV and Bistrioxazole acetate. Adapted from [104].

1.1.3.6 Perylenes

Perylene diimide compounds (PDI) are characterized by a broad hydrophobic core containing large fused aromatic rings and two external amine appendages (**Figure 24**). The main representative compound of this class of G4-ligands was PIPER, *N,N'*-bis[2-(1-piperidino)-ethyl]3,4,9,10-perylenetetracarboxylic diimide, a polycyclic compound derived from perylene scaffold [106]. Since the central core is planar, as for the vast majority of G4-ligands, it was responsible for the binding to G4-tetrads through π–π interactions.

PIPER was reported its ability to bind to a variety of G4 structures with different stoichiometries and its ability to accelerate the assembly of G4 structures *in vitro* with a chaperone-like function [106]. Additionally it was able to induce telomere shortening in treated cells therefore inducing reduction of cell proliferation and tumorigenicity and inducing senescence in cancer cells [105,107].

PIPER resulted, as for B-19, a good telomerase inhibitor with an IC_{50} in the low micromolar range. Moreover, our group demonstrated that this compound was able to induce and stabilize G4 structures in the *Nef* coding region of HIV-1 genome [91].



PIPER

Figure 24: Chemical structure of the perylene diimide compound (PDI) PIPER. It G4 binding is mediated by a π–π stacking on G4-tetrads. It is able to bind to different G4 structures and in particular it favors the formation of G4s by a chaperone-like function. Adapted from [107].

1.1.3.7 Naphthalene Diimide compounds

The biggest challenge that G4-ligands must overcome is a two-level specificity against their target: (i) they must be able to discriminate between G4s and duplex DNA and (ii) they must be able to discriminate between different G4 folding patterns [11]. Another interesting group of promising G4-ligands, designed to satisfy this need and designed to improve physicochemical properties of perylene scaffold, was represented by Naphthalene diimide compounds (NDIs) (**Figure 25a**). NDIs were obtained by a simplification and reduction in the dimension of the central planar core of perylenes and the obtained compounds led to an efficient target of G4s [108]. Each NDI compound was composed of a central core made of condensed rings and four side chains which could modulate the ability to bind G4s depending on their chemical features. Indeed, it was demonstrated that the dimension of the planar central core was able to modulate the ability of this class of compounds to recognize different DNA conformations. Moreover, NDI showed interesting photophysical properties that induced their extensive application as molecular probes with promising properties toward guanine-rich oligonucleotides [109]. Among the 1,4,5,8- tetracarboxylic NDIs analogues, the tri- and tetra- substituted NDIs showed promising activities thanks to their planar core able to accommodate up to four different side chains, which in principle could be exploited to produce ligand diversity that may discriminate between types of G4s [110].

In the late years huge interest has been dedicated to the study of NDIs: in particular many observations have suggested that tri- and tetra- substituted NDIs exhibit high G4 affinity with noteworthy selectivity for the parallel topology [111]. For example, tri-substituted NDIs chemically engineered to embed an alkylating quinomethide precursor (QMP), resulted extremely efficient in the alkylation of the telomeric G4 showing a promising anticancer activity *in vitro* [111]. NDIs were found to inhibit telomerase activity in the low micromolar range and to produce short-term cell

growth inhibition against MCF-7 and A549 cancer cell lines [110]. Additionally some studies reported NDIs ability in the binding and stabilization of the telomeric G4 and of the tyrosine-protein kinase receptor G4 in gastrointestinal tumor cell lines [112]. A tri-substituted NDI compound (H-NDI-NMe_2) was reported not only for its stabilizing activity on the telomeric G4, but also for its high affinity and effective stabilization of *c-myc* and *bcl2* G4s. This binding resulted in a promising reduction of mRNA levels of both genes that could further investigated for anticancer approaches [113].

Neidle and co-workers reported that NDIs promoted the parallel G4 topology therefore forming the 1:1 complex with the telomeric G4. This stoichiometry resulted from the combination of binding site affinity and direct groove interactions that were highly influenced by the protonated moiety in the side chains of compounds able to interact with DNA phosphates in the grooves [114,115]. NDI molecules tethered to activable alkylating moieties by flexible spacers showed a selective alkylation and stabilization of different G4s, therefore providing the first evidence of a thermally induced non-metal-based G4 adduct formation [116]. Additionally NDI-oxirane conjugates showed an increased affinity for G-quartets and a selective alkylation of the loop adenines [117].

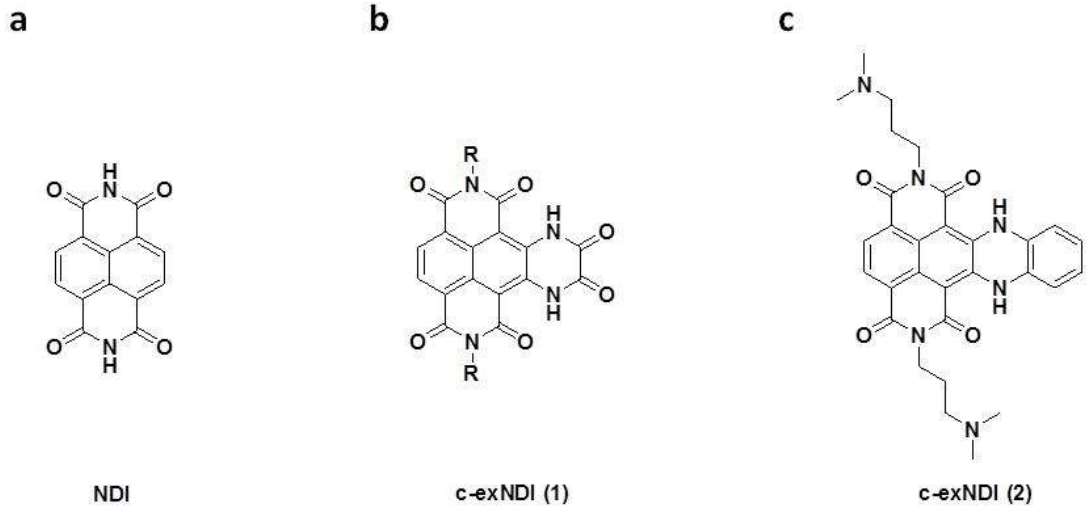


Figure 25: Chemical structure of (**a**) Naphthalene Diimide scaffold (NDI) and (**b-c**) core-extended Naphthalene Diimide (c-exNDI) compounds. In (**b**) R stands for i.e.: $CH_2CH_2N(CH_3)_2$ or $CH_2CH_2CH_2N(CH_3)_2$. The c-exNDI shown in (**c**) was tested against HIV-1 and, in the present work, against HSV-1. Adapted from [118].

In recent years our group investigated a new series of core-extended NDIs where the NDI core was fused with (i) 1,4-dihydropyrazine-2,3-dione (c-exNDI (1); **Figure 25b**) or (ii) 1,4-dihydroquinoxaline heterocycle (c-exNDI (2); **Figure 25c**). Concerning the c-exNDI1 it was proved its anticancer activity against different telomerase-positive cell lines [119]. Biophysical studies on the c-exNDI2 proved that this compound possesses a strong ability to bind the HIV-1 LTR G4s and a higher affinity when compared to the telomeric G4. More importantly the c-exNDIs exhibited a very interesting and promising antiviral activity against two different HIV-1 strains in the low nanomolar range (IC_{50} < 25 nM) with very low cytotoxicity therefore giving a promising

therapeutic window. NDIs mediated anti HIV-1 activity was deeper analyzed and the G4-related mechanism of action was proved by combining time-related and reporter assays [118].

In the present work, the most active c-exNDI against HIV-1 (**Figure 25c**), was analyzed in HSV-1 infection in order to confirm its G4-related mechanism of action.

1.1.4 G-quadruplex as potential therapeutic target: G4 binding proteins

Since G4s are non- randomly located with clusters in the promoter and telomeric regions many proteins with the ability to bind to G4s were identified. The fact that G4s must be unfolded during replication led to the identification of many different G4 binding proteins with different functions at that level (**Table 1**) [19].

Localization/Function	Gene	Protein
Telomere Region		BRCA1
		hnRNP A1
		hnRNP D
		POT1
		RPA
		TEBPs
		TLS/FUS
		Topo I
		TRF2
		UP1
Promoter Regions	*BCL-2*	PARP1
	c-MYC	CNBP
	c-MYC	nucleolin
	c-MYC	nucleophosmin
	Insulin	IGF-2, insulin
	KRAS	hnRNP A1
	KRAS	MAZ
	KRAS	PARP1
	MYB	PARP1
	KIT	PARP1
	VEGF	PARP1
		Mutant p53 protein
		MutSα
		Topo I
RNA Quadruplexes		FMR2
		hnRNP A1 mutant
		hnRNP A2
		nucleolin
		RHAU
		Ribosomal proteins
		SRSF 1 and 9
		TLS
		TRF2
Quadruplex-Resolving Helicases		BLM
		Dna2
		FANCJ
		G4R1/RHAU
		Sgs1
		WRN

Table 1: Proteins involved in G4 recognition, binding and resolving. Adapted from [19].

One of the first evidences of a G4 binding protein corresponded to the identification of proteins able to bind to the telomeric G4-forming sequence. Proteins corresponding to the **Shelterin complex** (6 proteins) expressed the ability to modulate telomerase activity and to protect the human telomeric DNA from being recognized as a chromosomal break.

The shelterin complex is formed by: POT1 (protection of telomeres 1), TRF1 and TRF2 (telomere repeat binding proteins 1 and 2), TPP1, TIN2 (proteins that interact with TRF2) and RAP1 (repressor activation protein 1). Hwang and co-workers firstly depicted the mechanism of interaction of **POT1** and TPP1 with the G4 structure of the *Oxytricha nova* [120]. More recently it was discovered the ability of **BRCA1** protein to bind the telomeric G4 therefore highlighting the possibility of an important role in telomere regulation and in the formation of quadruplexes structure in the genome [19]. As already mentioned, a huge number of different G4s were identified within promoter regions of many important genes associated with oncogenesis, such as *c-myc* oncogene. These findings confirmed a key role of G4s in the regulation of translation and it is therefore not surprising that among G4s binding protein, many of them are involved in transcriptional regulation, in chromatin remodeling and DNA repair [19]. Soldatenkov and colleagues demonstrated that **PARP-1** (a nuclear zing– finger protein) was able to bind to (i) intramolecular *c-kit* DNA G4 *in vitro* with high affinity and with a stoichiometry of two proteins per DNA G4 molecule [121], and (ii) the parallel G4 formed in KRAS promoter region.

More recently some papers described the ability of **Nucleolin**, a nucleolar phosphoprotein highly expressed in proliferating cells, to bind G4s. Nucleolin was reported for its ability, both *in vitro* and *in vivo*, to selectively and strongly bind to *c-myc* G4 structure, therefore facilitating its formation and inhibiting its promoter-driven transcription [122]. Recently, our group demonstrated that Nucleolin recognizes with high affinity and specificity G4 structures present in the LTR promoter region of HIV-1 proviral genome. Additionally we proved that Nucleolin interaction with viral G4s promotes transcription silencing while Nucleolin disruption with both siRNA and a nucleolin binding aptamer greatly increases LTR promoter activity [123].

Nucleophosmin, another multifunctional protein primarily implicated in ribosome maturation and export, centrosome duplication and response to stress stimuli [124] and implicated in the pathogenesis of several human malignancies (i.e. myeloid leukemia) was identified as a G4 binding protein able to bind G4s via its intrinsically unfolded C-terminal region. Its ability to bind to alternative secondary structures was proved for the G4 binding of the *c-myc* promoter region [125]. Mutations in the C-term sequence (i.e. NPM1 mutation) induced a complete loss of the binding ability and, moreover, a cytoplasmic stable translocation [124].

Heterogeneous Nuclear Ribonucleoproteins (**hnRNPs)** were identified for their ability to bind and to disrupt G4 formation thus alleviating the translational block due to G4 folding presence [19]. Our

group recently presented **hnRNP A2/B1** as an alternative binding protein with an efficient unfolding activity toward the HIV-1 LTR G4 in a full-length environment. Moreover, when hnRNP A2/B1 was silenced the LTR promoter region was significantly decreased indicating that the protein acts as a HIV-1 transcription activator [126].

1.1.5 G-quadruplex in viruses

Beside the human genome, many evidences pointed the presence of G4-putative folding sequences in key regions of several microorganisms, such as vertebrates [127], bacteria as *E.coli* and *Mycobacterium Tuberculosis* [52,128], yeasts such as *Saccharomyces cerevisiae* [129] and also in plants [130].

The first evidence of the presence and role of G4s within a viral genome was reported for HIV-1 where G4s were linked to the dimerization of the HIV-1 genome [131].

In recent years the presence of G4-putative folding sequences has been reported in a huge number of viral genomes and this have made G4s a new interesting topic for the scientific community [132].

In Simian virus 40 (SV40), a polyomavirus with a circular 5-kb doube-stranded DNA genome, 6 GC boxes (with consensus sequence GGGCGG) were identified within the ORI of the non-coding regulatory region (NCRR) and were reported by NMR spectra for their ability to form an unusual quadruplex structure containing a C-tetrad stacked between the two G-tetrads [133]. G4s in the SV40 genome were additionally found to be binding motifs for Sp1 transcription factor, therefore playing an important role in early transcription mechanisms. In the study of SV40, perylene diimides compounds (PDI) were extensively used to probe the role of G4s thanks to their ability to bind and stabilize G4 structures [132].

Concerning the human papillomaviruses family (HPV), G-rich regions were found in the 8-kb circular double-stranded DNA genome of 8 types of HPV. In some sub-types of HPV G4-folding sequences were found in the long control region (LCR) strongly suggesting a potential role in the regulation of viral transcription. Additionally the identification of G4s within the coding region of some viral proteins (i.e. L2, E1 and E4) suggested that G4 formation could also alter splicing mechanisms necessary for viral proteins production. Taken together these evidences pointed the possibility to target G4s in HPV for medical treatment [134].

Within the *herpesviridae* family, G4s were moreover identified in the human herpes virus 4 (HHV-4), also known as Epstein-Barr Virus (EBV) able to cause infectious mononucleosis but also cancers (Burkitt's Lymphoma, Hodgkin's Lymphoma and naso-pharyngeal carcinoma) [135]. The nuclear-antigen 1, EBNA1, of Epstein-Barr was reported (i) to bind G-rich sequences, (ii) to recruit the replication complex and (iii) to act in the metaphase chromosome maintenance. Since EBNA1 binding to G-rich sequences induced a stabilization of the secondary structure at the viral

replication origin, this therefore was directly linked to a role in the regulation of EBV replication [136]. Additionally, G4s within EBV were reported for their intrinsic control of EBNA1 mRNA expression regulation: destabilization of the G4-folding sequence in EBNA1 gene by antisense oligonucleotide annealing increased the translation rate of this gene. Since EBNA1 expression is significantly high to maintain EBV replication rate and significantly low to avoid immune recognition by the host-immune system, the activity of antisense oligonucleotides in the unwinding of G4 structures was directly linked to a protective increase in the antigen presentation [136]. The use of pyridostatin to stabilize EBNA1 G4s decreased EBNA1 synthesis and allowed immune evasion [132].

In Kaposi's sarcoma associated herpesvirus (KSHV or HHV-8), a member of the *gamma* subfamily of herpesviruses, a GC-rich DNA element with quadruplex folding capacity was found in the terminal repeat (TR), a region that is bound by the latency associated nuclear antigen (LANA). Stabilization of G4 structures within the TR region with G4-ligands led to the bidirectional pause of replication forks originated in the TR region therefore reducing the DNA replication and the episomal maintenance. These results opened the possibility to use G4-ligands in the treatment of KSHV-related diseases [137].

Among RNA viruses, G4s were identified in Zika virus (ZIKV) and a protein with G4s binding ability was identified in the severe acute respiratory syndrome coronavirus (SARS-CoV).

In ZIKV, member of the *Flaviviridae* family and actually a global concern, > 60 different G4-putative folding sequences were identified in the positive-stranded RNA genome. Restricting analyses identified and proved the formation of a unique parallel G4 structure in a conserved region near the end of the 3'-UTR (3'-untranslated terminal region) of ZIKV. This structure was reported for its ability to bind a G4-ligand and to induce polymerase stalling therefore being a target in the treatment of ZIKV infections [138,139].

Concerning the SARS-CoV enveloped virus, it was demonstrated the ability of the SARS-unique domain of the Nsp3 protein, to bind G4 structures formed both in DNA and in RNA nucleic acids, therefore opening the possibility to interphere with viral replication of immune response in host infected cells [140].

In Hepatitis A virus (HAV) a different approach using G4 structures was used: for this virus it was demonstrated the ability of the 3C protease, originated from a viral polyprotein, to bind the G4 structure formed by the hexanucleotide GGGGGT (G_5T) designed as a potent protease inhibitor. 3C protease binding to G4 resulted in a potent inhibition of the proteolytic activity of the enzyme therefore confirming the possibility to use G4s also as a tool for inhibition of viral infections [140].

A similar approach was investigated and used with Influenza Virus: non-structural protein 1 (NS1), a multifunctional protein, was proved to bind the G4 structure formed by an

oligonucleotide aptamer selected with a SELEX approach. NS1 displayed a high affinity in the G4 binding and, in the cellular context, this restored the IFN production therefore resulting in an antiviral activity and low cytotoxicity [141].

In Adeno-associated viruses (AAV) the presence of G4 folding sequences was demonstrated by a QGRS bioinformatic genetic analysis and their synergistic effect with the DNA binding protein nucleophosmin (NPM1) in enhancing AAV infection and mobilizing AAV capsids was reported [142].

Extensive studies, carried out by Prof. Richter's research group, strongly suggested the presence of important G4-putative folding sequences within the HIV-1 genome in the Long Terminal Repeat region (LTR) and in the *Nef* coding region [91,143], and the possibility to target viral G4s with G4-ligands for antiviral purposes [86]. Our group demonstrated that G4s formed within the HIV-1 LTR region exploit G4-mediated transcriptional regulation with striking similarities to eukaryotic promoters and the treatment with the G4-ligand B-19 strongly inhibited the HIV-1 infectivity. G4-folding sequences were highly conserved within different HIV-1 strains and additionally in Lentiviruses LTR regions therefore confirming the importance of these structures in the replication and transcription of different viruses [144]. Additionally, we have proved the presence of G4 folding sequences within the HSV-1 genome but an extensive overview on these results is reported in the next HSV-1 section.

1.2 Herpes simplex virus type 1 (HSV-1)

1.2.1 General features and structure of HSV-1

The herpes virus family, referred to as *Herpesviridae* family (or *Herpesviridales* order), is composed of three different subfamilies: α, β, and γ (**Table 2**) [145].

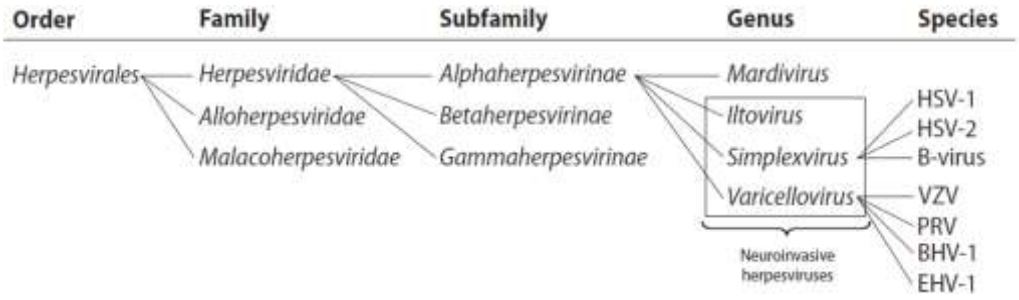


Table 2: Scheme of the taxonomy of neuroinvasive herpesviruses. One of the most important subfamily of herpesviruses is the alphaherpesvirus subfamily containing monkey B-virus, HSV-1 and HSV-2. Adapted from [146].

The *Alphaherpesviridae* subfamily is composed of four different genera (i) Mardivirus, (ii) Iltovirus, (iii) Simplexvirus and (iv) Varicellovirus, the last three being neuroinvasive viruses, able to invade neurons that were not directly exposed to inoculation [146]. The *Betaherpesvirinae* family includes human cytomegalovirus (HCMV or HHV-5) and the human herpes viruses type 6 and 7 (HHV-6 and HHV-7 respectively). The Epstein-Barr virus (EBV or HHV-4) and the Kaposi's sarcoma-associated herpesvirus (KSHV or HHV-8) are grouped in the *Gammaherpesvirinae* subfamily.

Among the *Simplexvirus* genus the most known viruses are the simplex virus type 1 (HSV-1) and the type 2 (HSV-2). The significant phenotypic differences between the phenotype HSV-1 and HSV-2 can be summarized in the fact that HSV-1 strains has been usually isolated from primary and recurrent lesions of the face (epithelial and mucosal cells) and cornea, while HSV-2 has usually been isolated from recurrent infections in the genital area [147]. Additionally HSV-1 has been associated with carcinoma lesions of the lips whereas HSV-2 has been associated with carcinoma of the cervix. Finally, HSV-1 and HSV-2 are similar but not identical in their ability to multiply in cells and in some aspects in their intracellular development [147].

HSV-1 primary infection of epithelial and mucosal cells results in the sores appearance (**Figure 26**). During illness manifestation, which is from two to three weeks, an infected person will develop pain on the affected site (in the case of oral sores) and sores all around the mouth site but also inside the mouth and in the nose. Each single sore can therefore grow to expand to other sores in the same district [148].

Figure 26: Clinical manifestation of the primary infection of HSV-1 referred to as 'cold sores'. HSV-1 primary sites of infection are the epithelial and mucosal cells.

After the primary infection, HSV-1 can establish a life-long latent infection in sensory neurons (resident in the ganglia of the peripheral nervous system) and it can reactivate under several stimuli such as UV-light, hormonal imbalance and emotional stress [146,149]. HSV-1 infection can result in a spectrum of illnesses ranging from asymptomatic, to manifested symptoms such as cold sores or even ocular herpes, up to life-threatening diseases and for the second case a lot of different pathologies have been associated to a latent infection (i.e. cancer, birth defects, neuropathy, cardiovascular disease, chronic inflammation and immunological dysfunctions) [150,151].

All herpesviruses are characterized by a dimension comprised between 800-1200 Å with a diameter at least about 120-200 nm. The viral genome is composed of a linear double-stranded DNA that is located within the core of the virion (**Figure 27**) [145,152].

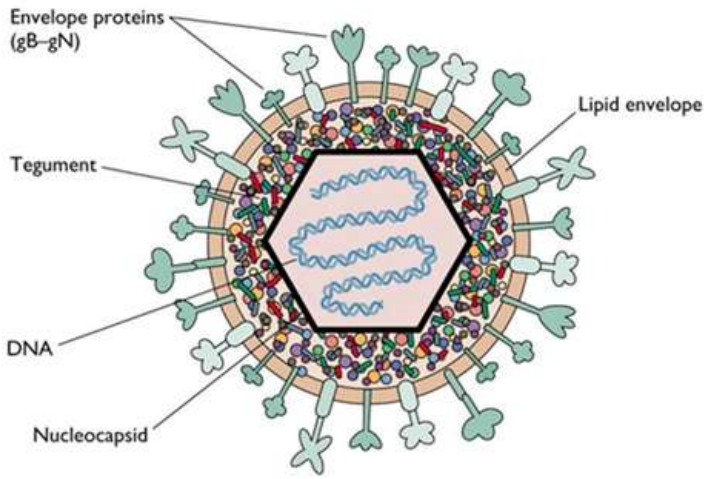


Figure 27: Structure of the HSV-1 virion. The viral particle is composed of a nucleocapsid which contains the double-stranded DNA, capsid, tegument and envelope are all characterized by the presence of important structural glycoproteins.

The region containing the viral DNA is referred to as **nucleocapsid** and shows an icosahedral structure with a diameter of approximately 100-110 nm [149]. This capsidic structure is composed of proteins which mediate the viral DNA entry and exit from the capsid [145]. The nucleocapsid is

enclosed in an external lipids and proteins **envelope** whose proteins mediate the first steps of the viral attachment to the host cells**.** The amorphous region between the nucleocapsid and the external envelope is called **tegument** and is composed of viral protein that may play an important role in the replication of the virus [149].

The HSV-1 DNA genome is estimated to be 152 kb long, is double-stranded and linear while contained in the nucleocapsid of a mature virion. When it is released in the infected cell, the DNA immediately circularizes [145,153]. The genome is divided in two covalently linked unique regions showing a different length: long (U_L) and short (U_S) segments that are flanked by terminal and internal repeated regions (TR_L, IR_L, TR_S and IR_S) referred to as *ab* and *a'b'* for the long segment (U_L) and *a'c'* and *ca* for the short one (U_S), therefore containing packaging sequences required for the cleavage and packaging of the newly synthesized viral genome into virions (**Figure 28**) [87].

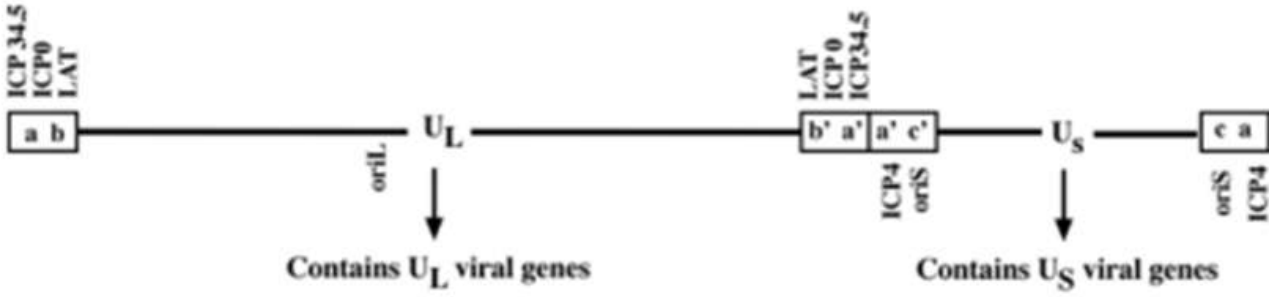


Figure 28: Schematic representation of the HSV-1 genome organization. The linear double-stranded DNA genome of wild-type HSV-1 is 152 kb and encodes for approximately 80-85 different viral genes, which are located in the unique long (U_L) and unique short (U_S) regions. Inverted repeats elements (*ab, a'b', a'c', ca*) flank the unique regions and contain packaging signals required for cleavage and packaging of the replicated viral genome into newly synthesized virions. Three viral origins of replication are located within the genome with 2 of them in the unique short segment (oriS) and one in the unique long segment (oriL) respectively Adapted from [154].

Three origins of replication are present within the HSV-1 genome: origins of replication present in the unique short sequence (U_S), Ori_S, are present in two copies within the *c* and *c'* regions while Ori_L is present as a single copy within the unique long sequence (U_L). *Ori* sequences are strongly related because (i) all contain extensive inverted repeated sequences made of the core palindromic 18 nt- AT bp region and (ii) are located in the promoter-regulatory region of divergent -transcribed genes important for viral replication [155]. *OriL* is located within coding regions for ICP8 (gene encoding a single-stranded DNA-binding protein) and UL30 (gene encoding the catalytic subunit of the polymerase Pol) while *OriS* is located between genes encoding immediate early proteins ICP4, ICP22 and ICP47 [156].

The whole genome contains at least 84 different genes coding for viral proteins (VP) or glycoproteins (gp) and for proteins first identified in infected cells, referred to as infected cells proteins (ICP) (**Figure 29a,b**) [149]. A very peculiar feature of the HSV-1 genome is the high G+C

content (68%) [157] strongly correlated with a viral protection mechanism against retrotransposon insertion [158] and an essential evolutionary driving force selected for the possible implication in different functions [159].

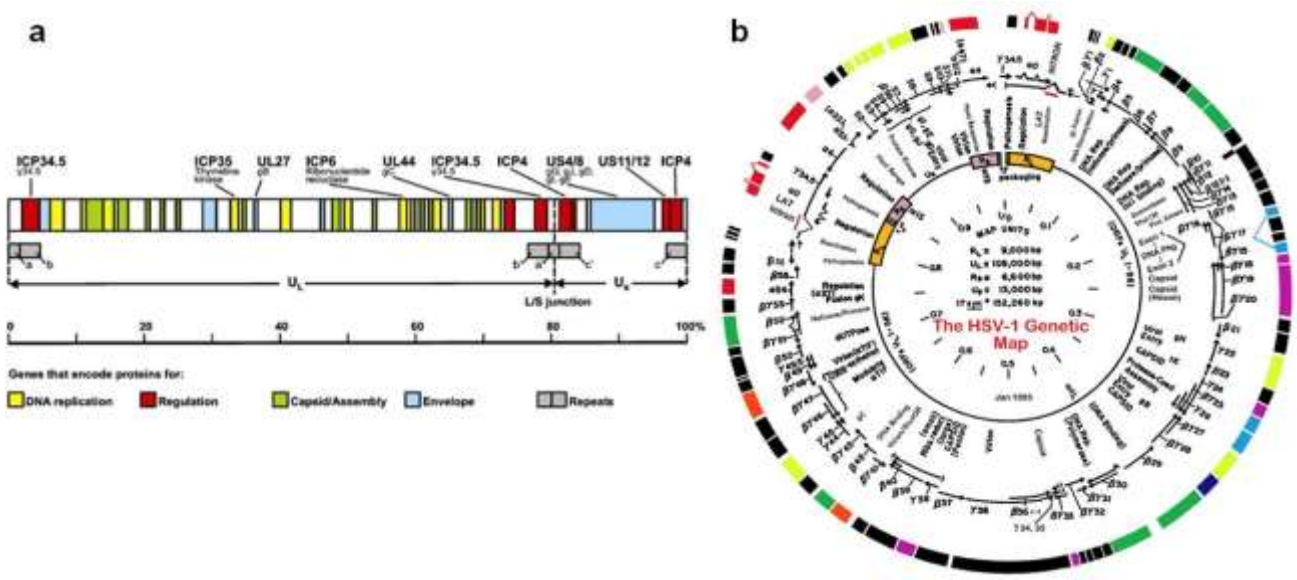


Figure 29: Schematic representation of the HSV-1 genome map. (**a**) Linear form of the genome. (**b**) Circularized form of the genome, episomal form. Adapted from [160].

1.2.2 Viral replication cycle

Glycoproteins of the external envelope are required for the first infection step of HSV-1: the **attachment/adsorption** of the viral particles to the external membrane of host cells which is mediated by heparan sulfate and chondroitin sulfate of the host cell membrane [161]. In particular five viral surface glycoproteins mediate the entry process : gB, gC, gD, gH and gL in addition to a cellular receptor for gD [162]. Attachment and **fusion** of the viral and cellular membranes are mediated by specific co-receptors: the member of the TNF-family of receptors herpes virus entry mediator (HVEM), nectins (1 and/or 2), the paired immunoglobulin-like type 2 receptor alpha (PILRα) and the 3-O heparan sulphate (3-OS HS) [149]. GB and gC interact both with heparan suphate proteoglycans (HSPGs) even if gC is not strictly essential for viral entry. gD is then required for effective viral entry within host cells: it acts via a conformational change that mobilizes a fusion active multi-glycoprotein complex involving all glycoproteins (**Figure 30**) [162].
Once the fusion is completed the tegument proteins that cover the nucleocapsid are released in the host cytoplasm by a pH-independent mechanism [163,164].

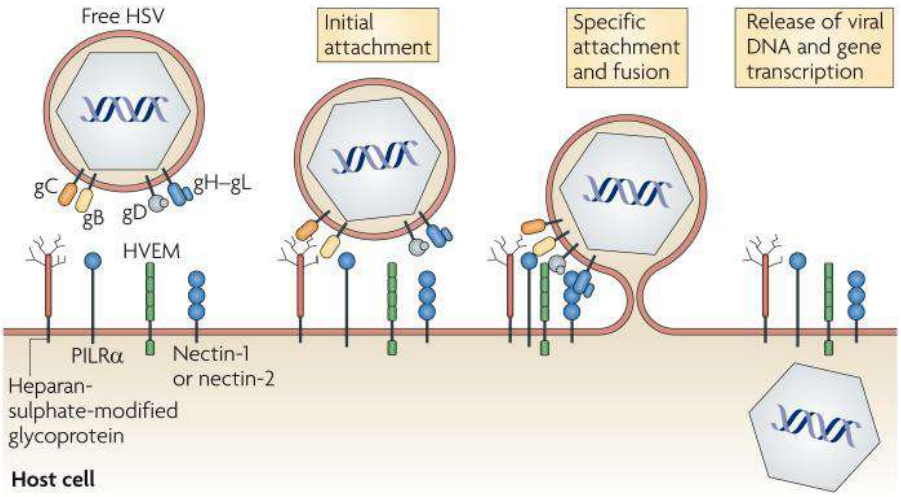


Figure 30: Initial attachment, adsorption and fusion steps of HSV-1 glycoproteins of the envelope to receptors of the external membrane of the host cell. Adapted from [163].

Subsequently, the nucleocapsid dissociates from the tegument and is then transported by the motor protein dynein, along to the nuclear pore via a microtubule-mediated process [149,165]. The vast majority of tegument proteins dissociate from the capsid, although two tegument proteins, VP1/2 and UL37 remain associated for consequent binding to the host nuclear pore complex (NPC) where the viral DNA is released upon the **uncoating** mechanism [165]. Upon arrival to the nucleus pores, the DNA molecule becomes disposable for subsequent **replication** and **transcription** steps, that occurs in a cascade-like fashion in the "lytic infection cycle", by translocating through the pores (**Figure 31**) [149].

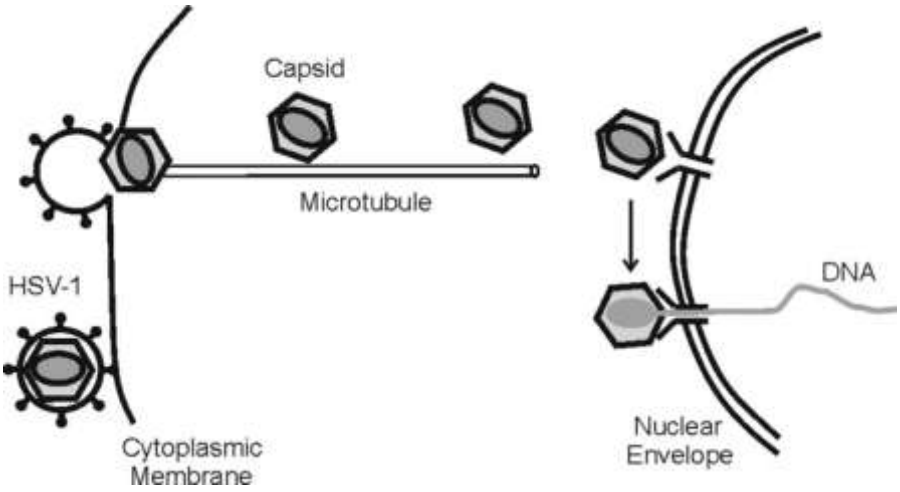


Figure 31: Schematic drawing illustrating events leading to the uncoating of viral DNA before replication step in a host infected cell. HSV-1 entry is mediated by viral-host membranes fusion, then the capsid is transported by dynein through microtubules to the nucleus where the viral DNA is uncoated and then injected within cellular nucleus through viral pores.

Then the lytic replication cycle is assisted by three different classes of viral genes: immediate early (α or IE genes), early (β or E genes) and late genes (γ 1 and 2 or L genes) [149]. As described by their

name, immediate early genes are rapidly transcribed, between 2 and 4 hours post infection (h.p.i), because of their implication in the coding for regulatory proteins that take part to subsequent processes. Immediate early (ICP0, ICP4, ICP22, ICP27 and ICP47) gene expression does not require prior viral synthesis and activates in cascade-fashion early and late genes [166]. Early genes, transcribed between 5 and 7 h.p.i., include enzymes required for viral DNA replication while late genes, whose expression occurs between 12 and 17 h.p.i, are mainly characterized by structural proteins of the viral structure [149].

All immediate early genes contain in their promoter region an enhancer element, characterized by the consensus sequence TAATGARAT, that is bound by VP(16) for the activation of transcription [167].

HSV-1 **DNA synthesis** initiates at the origin of replication where UL9, ICP8, UL30, UL42 and the HP complex upon unwinding of the viral DNA can undergo the DNA replication.

Although the wide variety of tissue tropisms and the different way of interaction with their natural hosts, all herpesviruses, independently from their sub-family of origin, share a common mechanism by which they replicate the genome during the lytic phase of the replication cycle. Additionally, in every herpesvirus family the lytic DNA replication occurs by generating a long head-to-tail concatamers of leading- and lagging- strands of viral genome, that are accumulated in the nucleus and linked to a unique monomer through *cis*-acting packaging signals in the *a* region [168,169].

Once the replication is complete, the **assembly** and **maturation** of progeny virions start to take place through the packaging process [170]. Naked nucleocapsids are then ready to be released from the nucleus to the cytoplasm where the **primary envelopment** (more than 15 tegument proteins are acquired) takes place to give rise to partially mature virions. The first budding event in HSV-1 maturation occurs at the inner nuclear membrane which provides the first primary envelope [170]. Enveloped intravescicular viral particles are ready to be transported to the Golgi apparatus where the **secondary envelopment** takes place. Then mature virions are ready to be transported to the host cell membrane where the **egress** of virus progeny will occur by an exocytosis mechanism of virus-containing vescicles [171,172].

The infection cycle is reported in **Figure 32**.

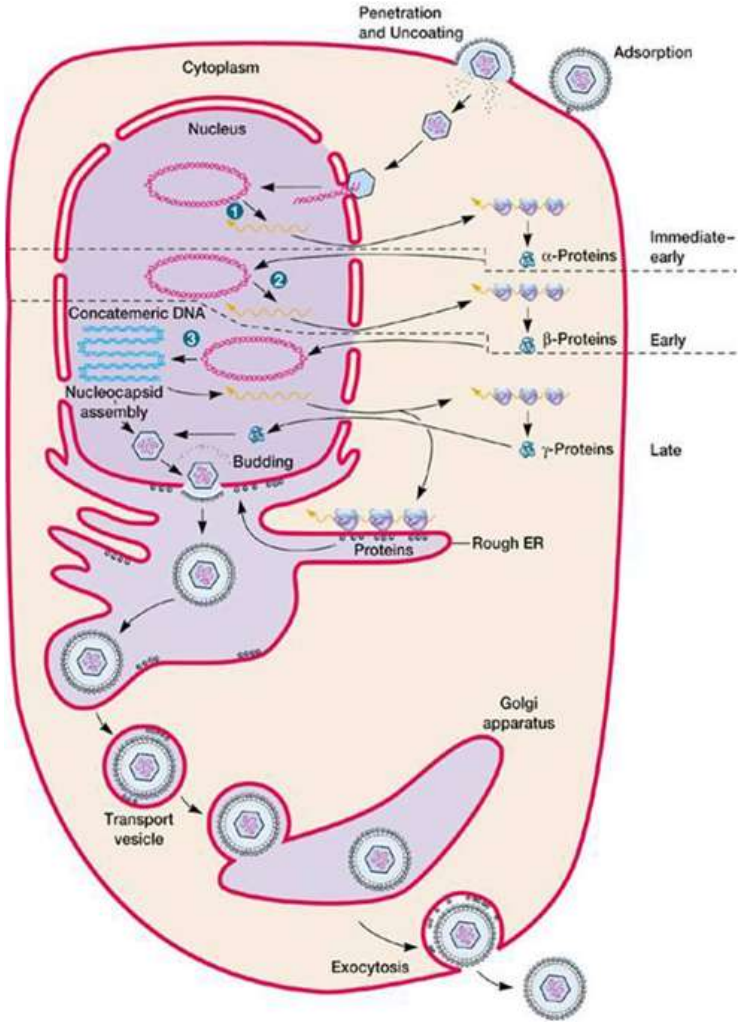


Figure 32: Replication cycle of herpesviruses. After the viral attachment and entry into the host cells, the HSV-1 genome circularizes and the transcription of IE-genes occurs. IE-proteins stimulate the transcription of E-genes. E-proteins promote DNA replication and the formation of a concatemeric DNA, which is processed during cleavage/packaging, and DNA molecules are pre-encapsidated. Nuclear egress and maturation result in mature herpes virions that are spread from the cell by exocytosis mechanism.

1.2.2.1 Immediate-early proteins

Infected-cell polypeptide 4 (ICP4) is the main regulatory protein of the virus, it plays important roles in transcription mechanisms (both by repressing and activating this mechanism) and is necessary for the transition of transcription from IE expression to E expression. ICP4 stimulates initiation of transcription from the gD promoter region therefore facilitating the formation of transcription complexes [173].

ICP0 nuclear phosphoprotein, on the contrary, acts as a potent transactivator by binding an array of cellular proteins (i.e. transcription factors) and nonspecifically DNA. It is required for lytic viral infection and for viral reactivation from the latent state [174].

ICP27 regulates the viral and cellular processing of mRNAs and therefore modulates ICP4 and ICP0 activities. This protein is required for an optimal viral DNA synthesis since it increases the early-gene expression levels. In collaboration with ICP22 it contributes also to an efficient late-gene

expression. ICP22 has many functions in viral proliferation: it mediates the formation of a novel phosphorylated form of the RNA polymerase II (pol II), and it regulates splicing pattern of ICP0 mRNAs [167]. Since ICP47 inhibits TAP (the transporter associated with antigen presentation protein) and blocks the antigen presentation process in CD8+ cells therefore causing inhibition of host immune response, it helps the virus escape from immune surveillance [175].ù

1.2.2.2 Early proteins

HSV-1 early proteins (β class of proteins) are crucial replication proteins. Some proteins of this family play essential "core" replication roles. Infected cell polypeptide protein 8 (ICP8 or UL29) of HSV-1 is one of the main representative proteins of this class. ICP8 is encoded by the UL29 gene and is an abundant HSV-1-induced protein of about 130 kD. It acts as a helix-destabilizing protein and binds more tightly single-stranded DNA molecules than double-stranded DNA in a sequence-independent manner [176]. Since ICP8 prefers the ss-DNA, its binding at the replication fork site favors the synthesis of the DNA template by the DNA polymerase.

HSV-1 DNA polymerase, main actor of the replication process, is a two-subunits heterodimeric protein made of the catalytic subunit Pol or UL30 (140 kD) and the processivity subunit UL42 (50 kD). UL42 was proved for being an accessory subunit of the catalytic subunit Pol that increases the processivity of the polymerization process [177]. UL30 possess an intrinsic 3'-5' exonuclease activity and a 5'-3' exonuclease/RNase H activity: the first activity cooperates with the maintenance of the DNA synthesis fidelity while the second cooperates with UL12 (alkaline exonuclease) in the removal of primers from the Okazaki fragments, produced by the semi-discontinuous synthesis of the DNA [156]. UL5, UL8, and UL52 are three-subunit helicase primase complex proteins (HP) [178] while UL9 is the origin-binding protein able to unwind short partially duplex DNA and enhanced in its activity by the presence of UL42 protein [179].

1.2.2.3 Late proteins

Late proteins are mainly involved in the packaging process and some of them constitute the vast majority of HSV-1 structural proteins. In particular, UL6, UL15, UL17, UL25, UL28, UL32 and UL33 genes encode proteins that are required for the DNA cleavage/packaging final processes of the replication cycle. Late gene promoters are recognized to be differently organized from IE and E promoters: they lack *cis*-acting sequences and are more easily organized [180]. UL15 and UL28 are recognized as the putative terminase of the packaging process [169]. UL36, UL37 and UL19 are late proteins that take part to the secondary envelopment of newly formed virions, the tegumentation process. UL36 is the HSV-1 largest gene product with more than 2000 amino acids and is well conserved between herpesviruses subfamilies. The absence of the UL36 protein induces a complete

block of the virions morphogenesis. Additionally, many other tegument-capsid proteins and glycoproteins take part in the well-defined protein-protein interaction-mediated encapsidation process but proteins functions are still largely undefined [170,171].

1.2.3 HSV-1 latency

Concerning latency the precise mechanism is currently not well known. What scientists were able to prove is that after the first lytic infection at the primary site, where the virus must come in contact with mucosal surfaces or abraded skin for infection initiation process, the progeny virus is spread within the sensory neurons in the trigeminal ganglia (TG), innervating lips, gingiva and eyes of the host, to establish a life-long latent infection (**Figure 33**) [181,182].

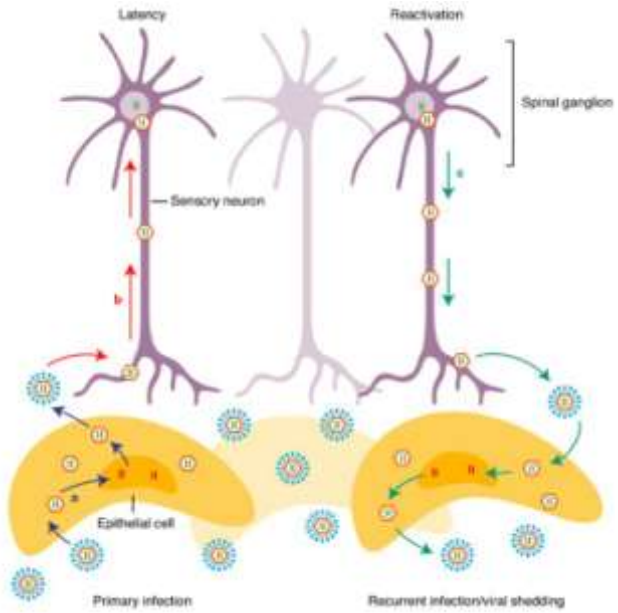


Figure 33: Schematic representation of the HSV-1 life cycle. HSV-1 infects epithelial cells of the skin where it establishes the lytic infection cycle. Some virus enters the sensory neurons terminal and travel retrogradely to the nucleus where it establishes latency. Adapted from [183].

The retrograde transport along the neurons which innervate the primary site of infection, mediated by dynein motor protein, let HSV-1 to arrive in the nuclei of the trigeminal sensory ganglia and to establish a long-life threatening latent infection [181]. In latently infected neurons HSV-1 genome is in a quiescent state with a detectable non-replicating chromatin-associated state (episomal state) and viral gene expression is largely repressed.

The only exception in viral transcription is made for a non-coding RNA known as the latency-associated transcript (LAT) which is splices to yield a stable 2 kb intron that accumulates in neurons and is commonly believed for maintaining HSV-1 latency. The exact mechanism of LAT activity is currently under investigation [184].

What is known is that reactivation from a latent state to a lytic one, influenced by different systemic or local stimuli, requires an immediate early protein, ICP0, which is also required for an efficient initiation of the lytic infection [168]. All herpesviruses have developed specific mechanisms to bypass the host immune response in order to persist in it [185].

1.2.4 HSV-1 epidemiology

Viruses of the HSV family are the most ubiquitous human pathogens currently known. They have the ability to use human skin epithelium and mucous membranes as their port of entry into the human body. The frequency of HSV infection (both the labial HSV-1 and the genital HSV-2) is generally measured by the use of type-specific immunoassays to test the presence of specific antibodies directed toward these pathogens [186]. Actually the global incidence of HSV-1 is ~90%, with a prevalence of 65% in the USA and from 52 to 67% in northern Europe [187]. In recent years a decreased trend is seropositivity for both HSV viruses especially in adolescent and young adults was measured, therefore providing important evidences about the changing epidemiology [186].
In developing countries HSV-1 is actually still universal and is usually acquired in early childhood upon intimate, personal contacts by seronegative individuals with someone excreting HSV [161]. Then the HSV-1 seroprevalence rate increases minimally with age and in a similar way between men and women. In the vast majority of cases HSV-1 infections have an orolabial localization in vesicular lesions and most of them still remain asymptomatic [188].
HSV-1 has been linked to severe encephalitis, keratitis and, in 1979, has also been related to Alzheimer's diseases [189,190].

1.2.5 Antiherpetic current therapy

Over 200 different viruses are already known at the moment and each virus has its own way to infect cells. However all human viruses share general common steps that are targeted by currently available antiviral drugs [191].
To date, the main group of antiviral drugs for the HSV-1 treatment is composed of nucleoside analogues that target and inhibit the DNA replication machinery by mimicking nucleotides incorporated by the DNA polymerase and their insertion give rise to an abortive replication [192,193].
Acyclovir (Zovirax®) (**Figure 34a**) is the most famous antiviral drug used in the treatment of HSV infections. It is generally used topically, intravenously or orally, even if in this way it shows a low absorption [193]. Its nucleoside structure was designed in 1977, in the 80s was introduced in the market and since then several nucleoside analogues have been developed [193].

Valacyclovir (Valtrex®) (**Figure 34b**), a L-valyl-ester prodrug of acyclovir shows a better bioavailability than acyclovir and is administred orally [194]. Penciclovir (**Figure 34c**), has an antiviral potency comparable to acyclovir but as acyclovir is poorly absorbed orally: farmcyclovir, oral diacetylated prodrug of penciclovir, has been evolved to overcome the low absorption of penciclovir (**Figure 34d**) [195].

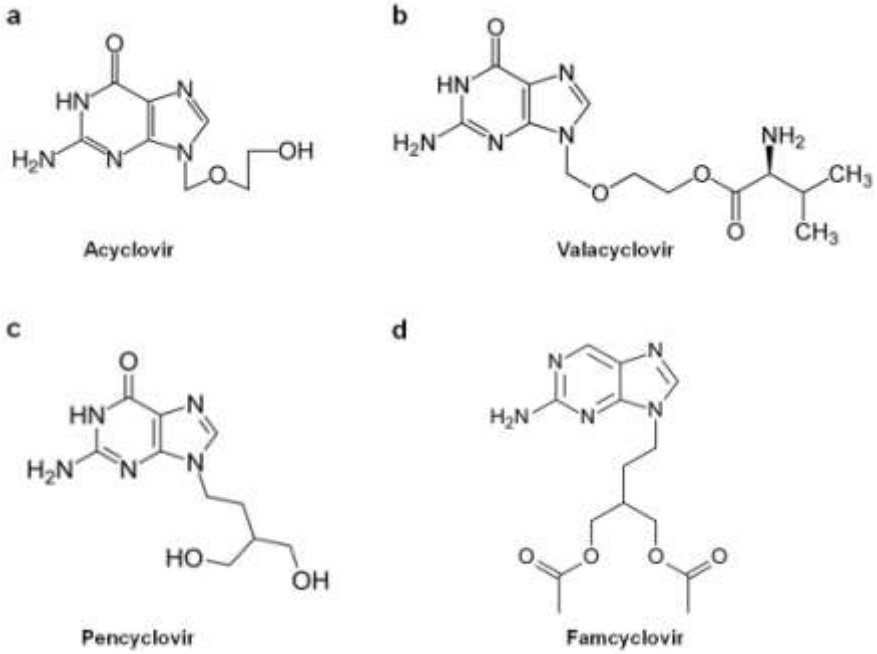


Figure 34: Chemical structures of the nucleoside analogues compounds (**a**) Acyclovir , (**c**) Penciclovir and their respective prodrug structures : (**b**) Valacyclovir and (**d**) Farmcyclovir. In both cases prodrugs showed a better bioavailability.

However the emergence of many acyclovir-resistant HSV strains, also resistant to the analog penciclovir, has prompted the need of the development of novel and alternative antiherpetic drug. As described above, nucleoside analogues constitute a treatment against herpesvirus symptoms but not a cure. Given the urgent need (urgent need because of HSV-1 resistance in AIDS patients) of new small molecules, capable of completely eradicating these viruses from the host, efforts in Richter's group are mainly addressed to the study of the G4 conformation as an innovative antiherpetic target as well as the discovery of new effective compounds. In particular, extensive knowledge about Naphthalene diimide scaffolds in terms of structure, general behavior and activity on cellular G4s has launched the hypothesis and the interest in their study and use as novel potential lead with antiherpetic activity [111,118,196].

1.2.6 HSV-1 and G-quadruplex

Exciting reports revealed important roles of G4s in virulence processes of different human pathogens such as bacteria, yeasts and viruses such as HIV, EBV, HPV [18,143].

The remarkably high guanine content of the HSV-1 genome prompted the Richter's research group to investigate the presence of G4-folding sequences and the possibility to target them with G4-ligands to obtain an anti-herpetic effect with a novel mechanism of action [87].

By biophysical assays our group showed the presence of multiple clusters of repeated sequences that form very stable G4s. These 9 G4 forming sequences are mainly located in the inverted repeats of the HSV-1 genome (**Figure 35**): six of these regions were found in the leading strand of the *gp054* gene which encodes for the largest viral protein of the tegument UL36, and four others were observed at the terminal and internal repeated sequences.

B-19 was able to inhibit *Taq* polymerase processing at the G4 forming sequences within the HSV-1 genome and decreased viral intracellular DNA in infected cells. Treatment of HSV-1 infected cells with B-19 induced inhibition of virus production with a mechanism of action that could be addressed to impairment of viral DNA replication [87].

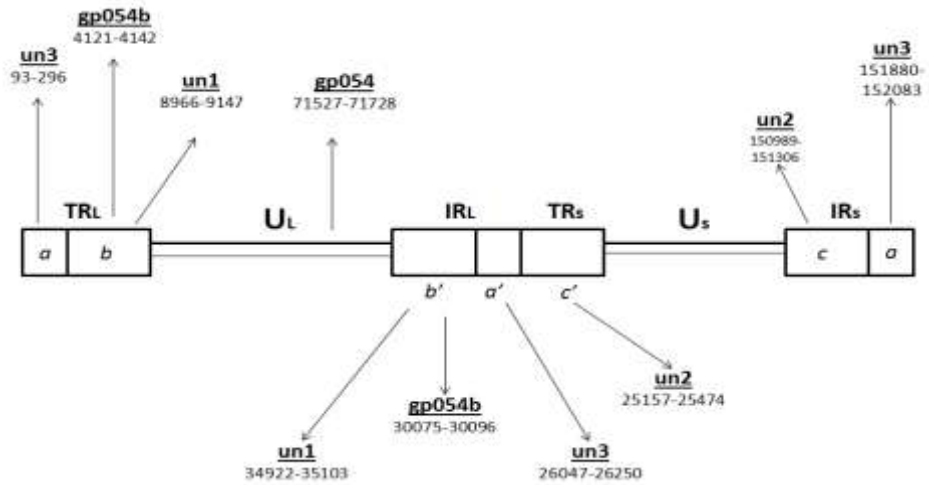


Figure 35: HSV-1 putative G4 forming sequences and their position within the viral genome. Terminal and internal repeats (TRs and IRs) are represented as rectangles while the leading and lagging strands are thick and thin line respectively. Each G4 forming sequence is presented with its name and position within the genome. Adapted from [87].

In this work our group presented the first evidence of extended G4 sites in key regions of the HSV-1 genome therefore indicating the possibility to block viral DNA replication by G4-ligands. Moreover, our group demonstrated the G4 formation and localization within the infected cells was virus cycle dependent: in fact, viral G4s peaked at the time of viral DNA replication in the cell nucleus, moved to the nuclear membrane at the time of virus nuclear egress and were later found in immature virions released from the cell nucleus. The notable increase in G4s upon HSV-1

infection strongly suggested a key role of these structures in the HSV-1 biology to control the lytic infection. This work provided, again, a proof of concept for the use of G4 ligands as a new anti-herpetic therapeutic option with a mechanism of action that targets viral DNA replication [197].

1.3 Measles Virus (MV)

1.3.1 General features and structure of MV

Measles virus (MV) is the prototype RNA virus member of the *Morbillivirus* genus, family of *Paramyxoviridae*, and the subfamily of *Paramyxovirinae* [198,199].

As for the other viruses of the morbillivirus genus (i.e. canine distemper virus of dogs and rinderpest virus of cattle), all morbilliviruses are strictly species-specific: no evidence has been reported for measles animal to human and viceversa transmission notable for epidemiologic reasons [198].

Measles virus consists of a spheric enveloped particle with a diameter around 200 nm [199]. It has a non-segmented negative-sensed RNA molecule which is packaged in a central core composed of nucleoprotein (N), large polymerase protein (L) and polymerase-associated protein or phosphoprotein (P). The core complex is externally surrounded by a layer of matrix protein (M) and an external membrane-derived lipid layer (**Figure 36**). The glycoproteins H and F are located on the envelope surface where they exert their function for virus attachment on host cells: trans-membrane hemagglutinin H mediates the attachment to specific receptors on the host-cell surface (CD150, CD46 and Nectin-4) while the fusion protein F mediates the fusion between the two membranes (viral and host-cell membranes) [198,200,201].

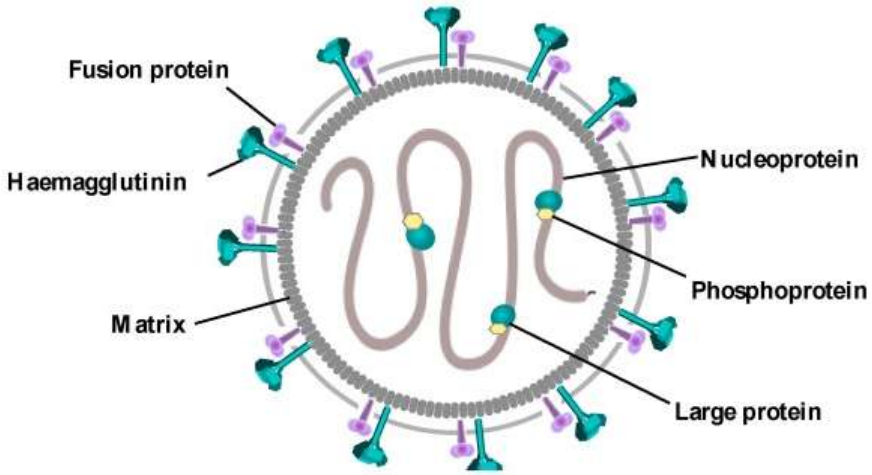


Figure 36: Structure of the Measles virus (MV) virion. The viral particle is composed of a central core which contains the single-stranded negative-sensed RNA genome associated to the nucleoprotein (N), the large polymerase protein (L) and the polymerase-associated protein of phosphoprotein (P). Externally a matrix layer and an external membrane-derived lipid layer form the viral envelope in which hemagglutinin (H) and fusion protein (F) exert their function mediating viral adsorption and entry to host cells. Adapted from [202].

Measles genome is composed of 16000 nucleotides and encodes for six structural proteins (**Figure 37**): three viral envelope proteins and three ribonucleoproteins. The three proteins involved in the

viral envelope are the matrix protein (encoded from the M gene), the hemagglutinin protein (encoded from H gene) and the fusion protein (encoded from the F gene). The three proteins involved in the central core structure (associated with RNA) are the nucleoprotein (encoded from N gene), the large polymerase (encoded from L gene) and the polymerase cofactor phosphoprotein (encoded from the P gene). This last gene, P gene encoding the phosphoprotein P, also encodes for two non-structural proteins: C and V proteins which suppress host interferon production therefore facilitating virus replication [201].

C and V proteins are produced by two alternative mechanisms: the C protein is encoded by the presence of an internal translation initiation site in the P gene, while the V protein is encoded from a frameshift mutation in the P gene (insertion of a non-template G nucleotide in position 751 during transcription) [203].

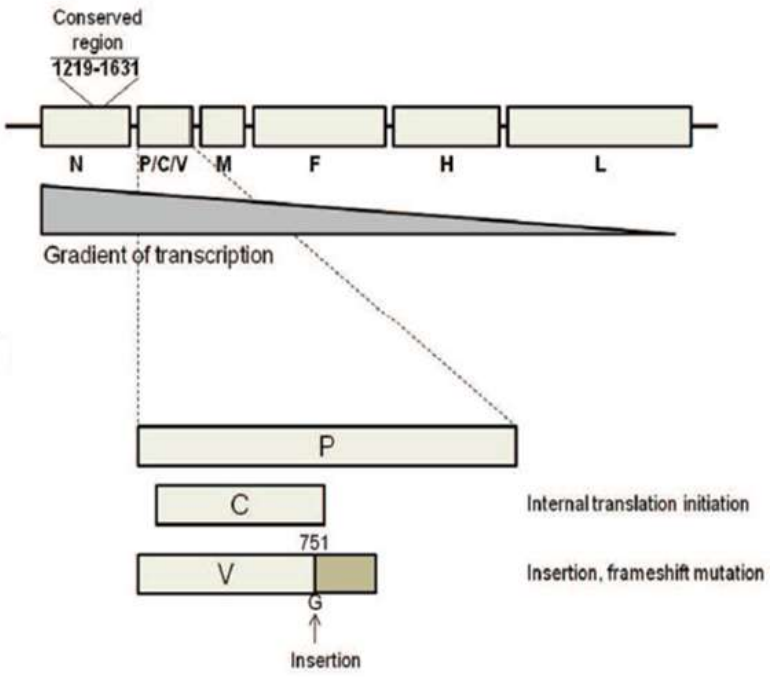


Figure 37: Schematic diagram of measles RNA genome organization and various genes encoded in the genome. N: nucleoprotein, P: phosphoprotein (with additional C and V proteins), M: matrix protein, F: fusion protein, H: hemagglutinin and L: large polymerase protein. The gradient of transcription is reported below the diagram. The variable region in the C- Term of the N region is used in sequencing for epidemiological studies. Adapted from [203].

By sequencing the minimum amount of sequence recommended by WHO for genotyping (450 nucleotides of the variable region of the N gene coding for the 150 amino acids C-term region of the protein), scientists identified 24 genotypes of MV then compiled in eight clades (from A to H) [198,204]. This genetic characterization have opened to the possibility to characterize circulating measles in order to effectively document the transmission pathways to discriminate between endemic and imported strains (transmission pathways) [205]. Many of the 24 identified genotypes are

no longer circulating, suggesting that effective vaccination programs are sorting positive effects on worldwide distributed infection dynamics [201].

Current laboratory diagnosis of measles infection is based on complementary serological and virological methods: identification of virus-specific immunoglobulin M (IgM) in serum of oral fluids samples is the gold standard for serological diagnosis while research for MV-RNA by applying a RT- PCR is the main method used in molecular biology [204]. These two methods are used in parallel for laboratory confirmation of suspected cases because each variation in day of specimen collection can alter the sensitivity and therefore the diagnosis and surveillance [205]. Recently it was reported that detection of high-avidity measles IgG in serum samples provides laboratory evidence of a past immunologic response to measles from natural infection of immunization [206]. For laboratory virus isolation Vero/SLAM and B95a cells are currently used. Vero/SLAM are Vero cells (African green monkey kidney epithelial cells) transfected with a plasmid encoding for the human signaling lymphocyte activation molecule (SLAM) while B95a are persistently infected with Epstein-Barr virus [205].

1.3.2 Viral replication cycle

As already mentioned above, in the initial process of **attachment/adsorption**, MV attaches to the host cell membrane through the interaction of viral protein H with cellular receptors such as hSLAM (CD150), CD46 or Nectin-4 [207].

Once this proteins complex is formed, the **fusion** of membranes occurs by the activity of the F protein at a neutral pH; in particular the hydrophobic core of this viral glycoprotein is inserted within the host cell membrane in order to introduce the negative-stranded RNA-RNP complex within the cell cytoplasm [207]. Negative-stranded RNA is then available as a template for both (i) **primary transcription** to produce mRNAs for subsequent protein synthesis, and for (ii) **secondary transcription** or **replication**, which produces a positive-stranded antigenomic RNA that will be used as a template for subsequent full-length negative-stranded RNA synthesis [207]. During this replication step the gene ends and the intergenic trinucleotide regions and the gene starts are ignored by the polymerase to produce the full-length nucleic acid.

mRNAs are directly synthesised from the negative-stranded RNA by the viral RNA-dependent RNA-polymerase which recognized the gene start (GS) for the initiation and the gene end for termination. mRNAs are directly capped, polyadenilated and terminated at the gene signals by the polymerase and this rapid transcription termination explains the rarity of viral recombination events [208]. Both negative-stranded and positive-stranded RNA are directly encapsidated to produce RNPs (**Figure 38a**) [209]. Primary transcription, concomitantly, produces all viral proteins

and structural proteins, such as F and H proteins and the M protein, which migrate to the host cell membrane: F and H proteins insert to the plasma membrane while M protein is aligned inside the plasma membrane to attach the cytoplasmic tails of the glycoproteins of the RNP complex. Once the virion attachment to the host cell membrane is completed the virion buds (budding event) from the plasma membrane to complete the replication cycle (**Figure 38b**) [207]. Infected cells that express viral glycoproteins on their surface can fuse with adjacent uninfected cells to produce multinucleated giant cells (also known as syncythia) followed by cell death [209].

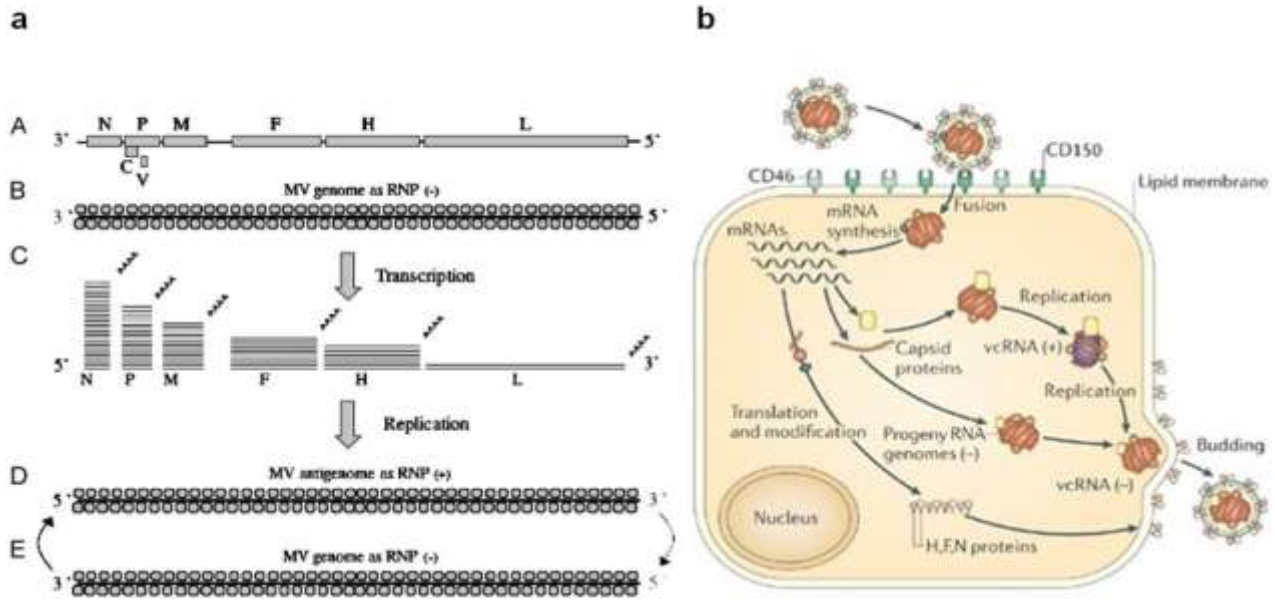


Figure 38: Schematic representation of MV infection and replication cycle. (**a**) Schematic representation of the transcription and replication of MV genome. The single-stranded negative-sensed RNA of MV is used as a template for transcription into mRNAs and subsequent translation into viral proteins, but also is used as a template for replication of a single-stranded positive-sensed RNA again used as a template for single-stranded negative-sensed RNA synthesis. Adapted from [209]. (**b**) Schematic representation of MV replication cycle: H proteins interacts with MV receptors on the host cell membrane to induce F-mediated fusion of the two membranes. Replication occurs in the cytoplasm of the host infected cell and once virions are assembled they bud from the host plasma membrane. Adapted from [210].

The first 12 hours of replication are generally characterized by a mild accumulation of viral mRNAs, genome and antigenome molecules, while from 12 to 30 h.p.i. a general exponential accumulation is observed for all products [208]. In details, from 0 to 5 h.p.i, newly synthesized RNA-dependent RNA-polymerase initiates the primary transcription for each gene with no detectable lag phase, suggesting that every mRNA is accumulated linearly; in this phase no genomes and antigenomes are produced. Then, from 5 to 12 h.p.i., mRNAs are accumulated exponentially indicating a regular accumulation of transcriptases. From 12 to 24 h.p.i, a switch of the RNA-polymerase to its replicase activity induces genomes and antigenomes exponential accumulation while no boost in mRNAs accumulation is observed. This slow down in mRNAs accumulation is moreover observed from 24 to 30 h.p.i where a sufficient number of viral proteins that can be used

for virions production is accumulated. From 30 to 48 h.p.i a plateau in mRNAs and genomes is reached (**Figure 39**) [208].

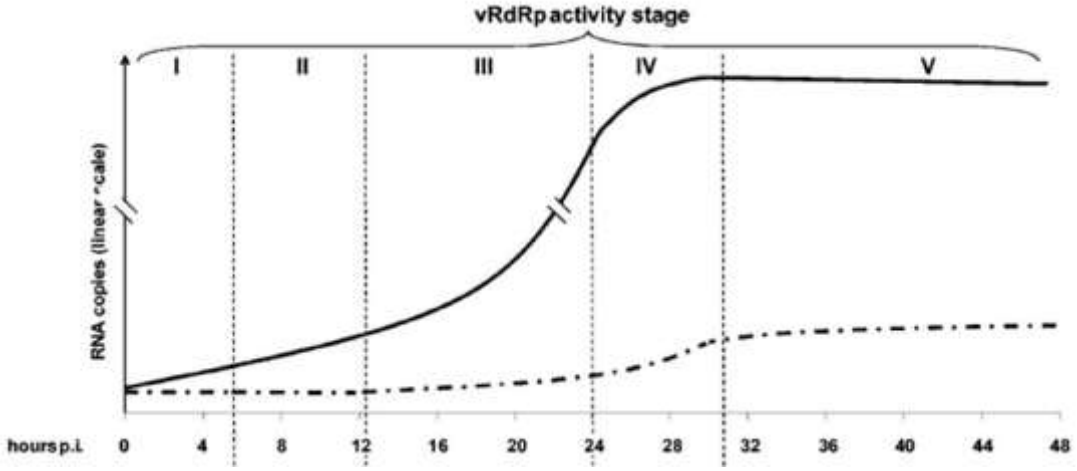


Figure 39: Model of the RNA-dependent RNA-polymerase activities during MV infection. From 0 to 5 h.p.i only mRNAs are linearly accumulated. From 5 to 12 h.p.i mRNAs accumulate exponentially while genomes and antigenomes are poorly produced. From 12 to 24 some RNA polymerases switch to the replication mode to produce genomes and antigenomes then exponentially accumulated. Plateau is reach from ~ 30 h.p.i where a slowdown in mRNAs production is counterbalanced by the formation of mature virions. mRNAs are indicated by the solid line while genomes and antigenomes are indicated by the dotted line. Adapted from [208].

1.3.2.1 MV infection cycle

Signaling lymphocyte activation molecule family member 1 (SLAM1, also known as CD150), is expressed on the surface of subsets of tymocites, on dendritic cells (DCs), on hematopoietic stem cells (HSCs), on macrophages and B- and T- cells and has been identified as the predominant cellular receptor for MV entry and infection *in vivo*. More recently Nectin cell adhesion molecule 4 (Nectin-4) has been identified as another alternative cellular receptor for MV that is expressed by epithelial cells [199]. Nectin-4 is a part of the adherent junction complex located in the basolateral side of epithelium underneath the tight junction and, in association with CD150 molecules, plays crucial roles in the pathogenesis of measles [211,212].

Measles infection is initially established in the mucosal cells of the **respiratory tract** where the virus replicates in the tracheal and bronchial epithelial cells, in alveolar macrophages, in dendritic cells and in monocytes. In some cases another alternative route of MV entry has been described: MV infection of myeloid and lymphoid cells in the conjunctiva [199]. Then, from the lower respiratory tract the virus spreads to **primary** (bone marrow and thymus), **secondary** (spleen, tonsils, and lymph nodes) and **tertiary** (bronchus-associated lymphoid tissue, BALT) local **lymphatic tissues** (in particular **to subsets of B and** memory CD4+ and CD8+ **T cells,** rich in CD150) where the infection massively runs, in most cases without any important complication, for 2 weeks [213]. Infection of lymph nodes, which are prominent sites of virus replication, results in a

rapid increase in the viremia and a potent spread of measles through the **blood** to systemically infect (trough infected circulating CD150+ immune cells) many different organs such as the skin, conjunctivae, lung, kidney, the gastrointestinal tract, genital mucosa, liver and again the port of entry, the respiratory mucosa (i.e endothelial cells, epithelial of the upper respiratory tract (**Figure 40**) [198,213]. Viral dissemination is predominantly mediated by cell-to-cell transmission.

After a primary infection with MV, the lifelong immunity that follows is mediated by neutralizing antibodies (IgG) directed toward the hemagglutinin protein in order to block its assembly to host cell receptors on the membrane surface [201].

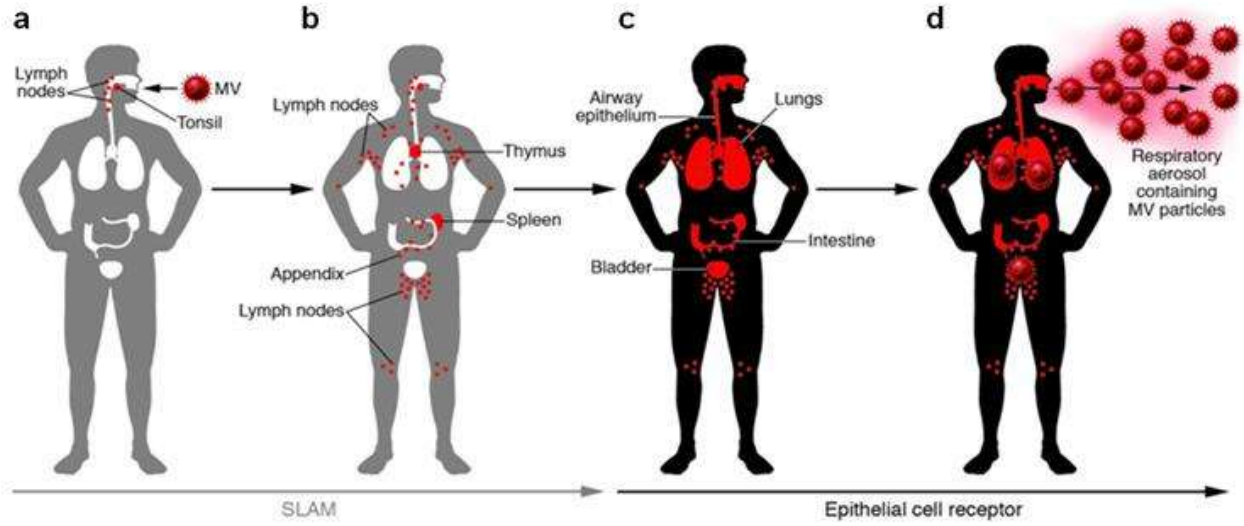


Figure 40: The time course of MV infection through the human body of a susceptible host. (**a-b**) MV, in the form of respiratory droplets, enters the human body through the upper respiratory tract where it encounters CD150+ (SLAM) cells. From the lower respiratory tract the virus is then spread into lymph nodes of different lymphatic tissues. (**c-d**) Infected lymphocytes enter the bloodstream and MV propagates in different organs throughout the body such as the skin, conjunctivae, kidney, liver, spleen and again the respiratory tract via nectin-4 receptor. MV is then replicated in epithelial cells and progeny viruses are released in respiratory aerosol that can infect other patients. Adapted from [214].

1.3.3 MV related diseases

MV is a highly contagious respiratory pathogen that causes a systemic disease [199,213]. MV is air-borne transmitted mainly by (i) nasopharyngeal large droplets, (ii) aerosolized particles or (iii) by direct contact [198]. Respiratory droplets usually cover short distances in the measles transmission and aerosolized particles remain suspended in the air for up to 2 hours [201]. In general almost every infection from MV becomes clinically visible in infected people [198]. MV has an incubation period of 8-12 days and people infected are contagious around 2-4 days before the appearance of the epithelial rash and from 2 to 5 days after the rash appears [213].

Symptoms correlated with an acute MV infection in humans, referred to as prodromal stage, consist of fever, cought, coryza, dry, conjunctivitis, photophobia and the characteristic, and utile for diagnosis, grey-white spots on the mucosal surface of the mouth referred to as Koplik's spots

(**Figure 41a**) [198,213,215]. In the subsequent days (12 days post infection) patients usually develop the main visible effect of measles, the typical erythematous maculopapular rash that usually starts behind the ears and then, with the progression of the infection, spreads to the face, trunk and extremities of the body. The typical rash, following the infection progression increases not only in terms of number (appearance in other districts of the body) but also increases in the dimension (sometimes lesions become confluent) (**Figure 41b**) [216]. This typical rash can be potentially explained by infection of dermal endothelial cells and keratinocytes that are subsequently cleared by the virus-specific host immune T-cells [199].

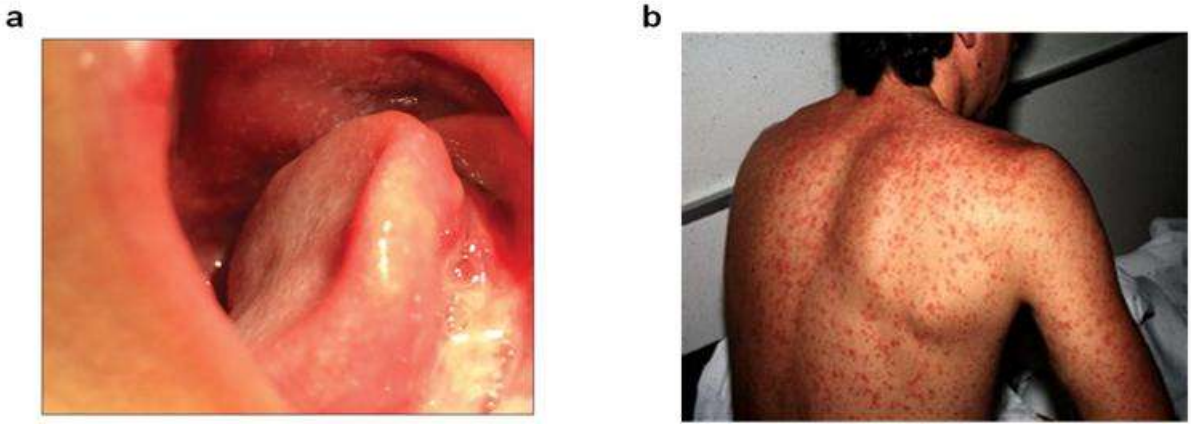


Figure 41: (**a**) Photograph showing Koplik Spots on the buccal mucosa of a 10- months- old girl confirmed to be exposed to measles infection. Adapted from [217]. (**b**) Adult male with a measles maculopapular rash extended to the trunk (back) and arms. Adapted from [218].

1.3.3.1 MV complications

Since MV infects immune cells, the host immune system is inevitably activated to contrast the progression of the infection and the viral replication: proinflammatory cytokines such as Interferon (IFN)-γ and Interleukin (IL)-2, in fact, are up-regulated [213].

MV infection is able to cause long-lasting impairment of memory B and T-cells (lymphopenia both in circulating cells and in tissue homing) therefore predisposing people to subsequent opportunistic infections even several years after recovery [198,219]. This represents the main unresolved "measles paradox": prolonged increased susceptibility to infectious disease due to immune suppression and at the same time induction of strong MV-specific immune responses (i.e. production of measles antibodies) [199].

Complications may include, in the presence of bacterial co-infections, to the insurgence of diarroea (8%), pneumonia (1- 6%), laryngotracheobronchitis, otitis (7- 9%) and stomatitis [198,213].

In around 0.1% of MV acute infection severe complications can occur after months or years from the first infection within the central nervous system (CNS): acute disseminated encephalomyelitis

(ADEM), subacute sclerosis panencephalitis (SSPE) and measles inclusion bodies encephalitis (MIBE) in immunocompromised patients and are the most known severe complication [220,221]. These encephalic and fatal diseases show a different pattern of virological and immunological features than that seen in the classical acute infection: measles persistently infects brain cells such as neurons, glial cells and endothelial cells but the classical formation of giant cells and the budding of virions particles are virtually absent [213]: this mechanism is virtually called "white" measles [198].

Acute disseminated encephalomyelitis disease (ADEM), which is immune-mediated and has a higher incidence than other complications (~1:1000), is generally hallmarked by demyelination of CNS cells therefore resulting in ataxia, motor and sensory neurons loss, mental status changes and in some cases can also results in death [222].

Sub-acute sclerosis panencephalitis (SSPE), a very rare measles complication, can be developed after 1-10 years after the first infection and the incidence of this recurrence is higher in infants infected in the first year of age and in these cases the prognosis of SSPE is lethal [223]. Symptoms usually start with a decline of school performance and a slight change in the behaviour, progressively myoclonic seizures, ataxia and death appear. This neurologic complication is linked only with infections with wild-type MV and no studies have reported its development after genotype A measles vaccination [199].

Concerning MIBE, this neurologic complication has an increased rate of manifestation in young infants or immunocompromised individuals unable to completely clear the infection. The symptoms are in part shared with ADEM and comprise mental status changes, focal seizures and occasionally visual or hearing loss. This disease progresses rapidly with coma and death in the majority of patients [224].

It has been reported that (i) complication rates are higher in patients <2 and >20 years old, (ii) bacterial complications are more common in patients <2 years old while encephalitis are more common in older children and adults. All these complication are increased in their incidence in the presence of an immune-deficiency disorder, vitamin A deficiency, lack of previous measles vaccination and intense and continuous exposure to measles [219]. All complications resulting from measles infection are reported in **Figure 42**.

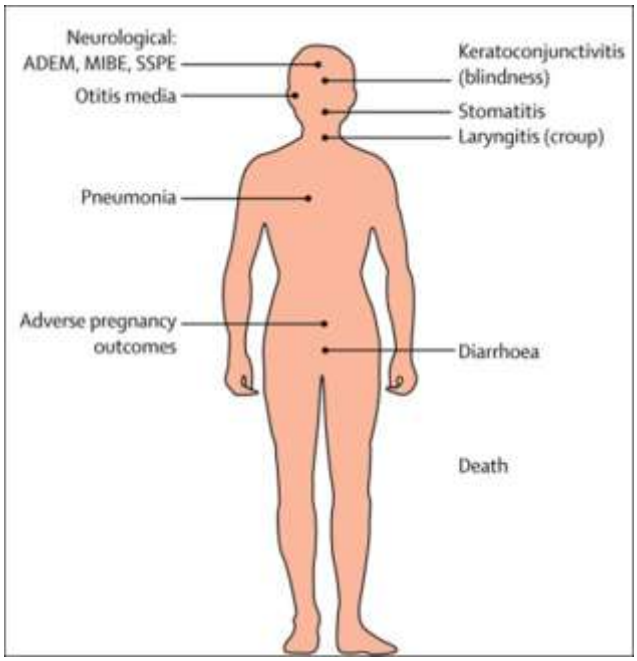


Figure 42: Measles complications generally occurring in patients in different organ systems. Measles infection can cause conjunctivitis, stomatitis, laryngitis, otitis media, pneumonia and diarrhoea when caused by subsequent bacterial and viral infections due to the transient impairment of the immune system. Neurological complications, such as ADEM, MIBE and SSPE are caused by infection of neurons from measles ADEM: Acute disseminated encephalomyelitis disease, MIBE: measles inclusion bodies encephalitis and SSPE: Sub- acute sclerosis panencephalitis. Adapted from [201].

No evidences on how the virus enters the CNS are yet available but recent groups proposed the possibility that (i) blood-brain-barrier allows the entry of viral particles or (ii) lymphatic vessels of the brain allow the transportation of viral particle from infected circulating lymphocites [199].

The exact mechanism of measles persistence is far from being exactly understood but recent evidences show that MV classical spread is defective in CNS tissues [225]. It has been proposed a mechanism of spread from a neuron to the adjacent one without the spread of viral particles both *in vitro* and *in vivo*. This mechanism may be due to the membrane fusion between an infected neuron with the uninfected adjacent one that leads a trans-synaptic transmission of the minimal unit of infection: the RNA genome associated with viral nucleoprotein and the viral polymerase [199].

1.3.4 MV epidemiology

MV is one of the most contagious pathogenic virus for humans [198], and has been recognized for a long time as one of the "inevitable traumas of childhood" [226].

Even if historical data is lacking, epidemiological evidences show that measles became a human disease 5000-10000 years ago when the first agrarian civilization led a sufficient number of people in villages able to maintain virus transmission [201].

Before the introduction of measles vaccine in the 1960s, measles was responsible for more than 2 million deaths annually and was the main cause of child morbidity and mortality.

In recent years the vast majority of reported cases of MV infections were from Africa (40%), Western Pacific (27%) and South East Asia (12%). Importantly, it was also observed that 11% of the reported global measles cases were from the European region [201,227].

1.3.5 Current therapy against MV

MV is characterized by one of the highest Reproductive Rate (R_0) between all human viruses: 12-18. This number gives the idea on how infective a virus can be: a single person with measles infects an average of 12 to 18 people in a population fully susceptible to this infection (**Figure 43**) [198,228].

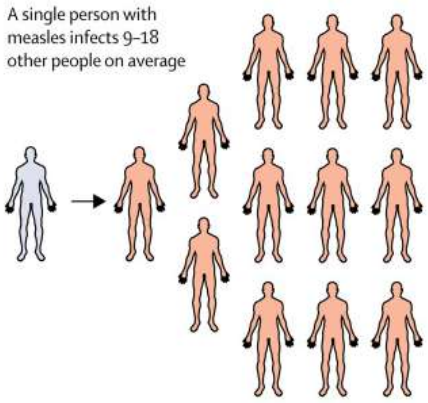


Figure 43: MV basic Reproductive Number (R_0). This number is the average number of secondary cases resulting from the introduction of an infectious individual into a completely fully susceptible population. MV estimated R_0 is about 12- 18. Adapted from [201].

Since this virus is easily transmitted between people it is necessary to manage it with potent antiviral drugs. No specific antiviral drugs are available now for the treatment of acute measles, although ribavirin and IFN-α have been used in the past to treat severe measles in immunocompromised patients [229,230,231]. The actual management of infected patients consists of a "supportive" therapy because no anti-measles drugs are available. In particular the supportive therapy is directed toward the correction and the protection against dehydration and nutritional deficiencies with particular attention to under-developed countries and their malnourished inhabitants [201]. Since in many cases vitamin A deficiency in malnourished children was correlated with increased fatality-rates of measles infections, provision of vitamin A is strongly recommended by the World health Organization (WHO). Actually antibiotics are indicated for people with measles infection and the correlated bacterial infection [219].

1.3.6 MV vaccination program

In 2001 the WHO launched an important program for the complete elimination of measles by a worldwide preventive vaccination strategy [198].

Vaccination became available from 1963 when MV was isolated and propagated for the first time in human or monkey cell lines. Initially the first available vaccine was made of an inactivated MV but unfortunately it did not provide a good and long-lived protection (antibodies were not protective against wild-type MV) because of the denaturation of some relevant regions of the F protein during the inactivation process. Subsequently, to overcome this important problem, this vaccine has been replaced by an attenuated live-based MV vaccine. The latter was made of the A genotype derived from the Edmonston strain isolated in 1954 [232].

In recent years several other MVs vaccine made with "Schwarz" and "Moraten" viruses have been developed by serial passages in human kidney, human amnion and embryonic chicken cell cultures. All these vaccines are actually combined with mumps and rubella vaccines (MMR) and millions of doses are administered since decades. MMR vaccines are well tolerated, are very safe and give a long-lasting protection to vaccinated patients with about 95% of protective antibodies after the first dose and up to 99% protective antibodies after the second dose [233].

Nowadays, following the WHO measles eradication program, vaccination for MV has been introduced in the vast majority of countries in child routine immunization and it is administered in two doses. The first dose (MCV1) is administered at the age of 9-12 months in order to give protection to the children as soon as maternal antibodies have vanished. The second dose (MCV2), which must be administered after a minimum interval of 4 week from the first dose, is usually administered at the age of 15-18 months [234]. During an outbreak, where the risk of exposure is significantly increased, the administration of the first dose of MMR should be done between 6-9 months of age [198]. It has been calculated that since the introduction of the vaccination program, 35 million cases of measles have been prevented.

Many efforts in the MV eradication have been done up to recent years and many positive data have shown the reduction in the incidence of measles: from 2005 to 2015 measles reduced by 70% from 853479 to 254928 cases worldwide [201].

Even if MV incidence and mortality have been contained due to the increasingly widespread use of the attenuated MMR vaccine, (i.e. in Europe MV infections have decreased by >90%) measles still remains an important vaccine-preventable cause of morbidity and mortality and is still responsible for more than 100000 death annually all over the world [219].

Since MV is an human-specific morbillivirus that does not have an animal reservoir, the vaccination strategy has the potential to help the complete elimination of the virus in many countries [198].

Unfortunately, the decrease in the population confidence for MV vaccination program has caused the re-emergence of MV cases that were strongly contrasted in the last century.

Since the beginning of 2017, 2719 new cases of MV were registered in Italy with 89 % people being unvaccinated; in the Veneto region 160 cases were confirmed with 62% of people being unvaccinated. Only in the department of Microbiology and Virology of the Padua Hospital 113 new cases were reported since the beginning of the year [235,236]. This re-insurgence of MV infections is of a social-concern: circulation of MV and consequent infections that cause a long-lasting immune suppression could be fatal for people with a personal history of immune suppression due to other pathologies or medical interventions (i.e. AIDS patients, cancer patients or immune-compromised transplanted patients).

2. Aim of the study

The presence of G-quadruplex (G4) in key regions of the human genome, such as telomeres and oncogene promoters, and in viral genomes of several microorganisms and in particular viruses (i.e. HIV, HSV, EBV, HPV, ZV), strongly suggests their potential as innovative drug targets.
Additionally, stabilization of G4 structures by rationally-designed small molecules that interact with these secondary structures indicates the possibility to interphere with crucial biological mechanisms like replication and translation. The possibility to target G4s and thus interfere with biological processes has ben provided for both tumors and viruses.
Our group has studied the presence of G4s in the genomes of several viruses and the data collected indicate that viral G4s can be exploited as new drug targets [86,87,118,143].
From this point of view the herpes simplex virus type 1 (HSV-1) is of particular interest because it showed the presence of a large number of G4s in its genome, especially during viral DNA replication [197], and an increasing number of viral strains is becoming resistant to current therapies, mainly limiting effective treatments.
On these bases, the aim of this thesis was:

1. Analyze the activity against HSV-1 G4s of the G4-ligand c-exNDI, previously reported for its promising activity against HIV-1. In particular, study c-exNDI binding and stabilization of the HSV-1 G4 forming sequences and evaluate its antiviral activity and mechanism of action to confirm its potential use as innovative anti-herpetic drug [87].
2. Screen the anti HSV-1 activity of an in-house library of G4-ligands with promising drug-like characteristics. The library was made of compounds previously proposed for their anticancer activity [62,99,100,101].
3. Investigate the role of G4 structures in HSV-1 replication cycle by identifying proteins that specifically interacts with viral G4s to better understand the biological role of G4s and their use as drug targets.

Among viral genomes, G4 structures were identified not only in DNA but also in RNA genomes.
Since in recent years a strong decline in the confidence for viral vaccination programs has caused the re-emergence of infections that were previously eradicated, we decided to focus our attention in the main viral infection re-emerged and registered this year in the Department of Microbiology and Virology of the Padua Hospital [235], the Measles virus (MV) infection in order to investigate the presence, conservation and the possible role of G4-putative folding sequences within the RNA

genome of MV to further propose G4 structures as drug targets against a broad range of viral infections and to further develop innovative G4-ligands for the treatment of measles-related infections that lack a specific therapeutic approach.

3. Materials and methods

3.1 HSV-1 STUDY: C-exNDI and screening of an in-house library of compounds

3.1.1 Cells and viruses

Vero cells (African green monkey kidney cells from Sigma-Aldrich, Saint Louis; USA), used in the study of c-exNDI, and U2OS cells (Homo sapiens bone osteosarcoma cells from ATCC® HTB-96™), used in the in-house library of compounds screening were maintained and propagated 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 1X PenStrep antibiotic (Gibco, Life Technologies, Monza, Italy) in humidified incubator maintained at 37°C, 5% CO_2.

HSV-1 strain F was a kind gift of B. Roizman (University of Chicago, Illinois, USA). Recombinant HSV-1 expressing VP16-GFP (HSV-1 [V41]) was kindly provided by Peter O' Hare (Imperial College, London, UK).

For virus infection, wild- type (wt) HSV-1 or mutant virus were applied to cells at different multiplicities of infection (MOI) in serum-free medium. After 1 h of incubation at 37°C, the inoculums was replaced with complete medium.

3.1.2 Oligonucleotides and compounds

All oligonucleotides used in the study of c-exNDI and in the screening of the library of compounds were purchased from Sigma- Aldrich (Milan, Italy) and are reported in **Table 3**.

G-quadruplex ligand c-exNDI2 [118], herein referred to as c-exNDI, was synthesized and provided from Prof. Freccero (University of Pavia, Italy). Then diluted at 10 mM in DMSO and stored at -20°C. C-exNDI chemical structure is reported in **Figure 25c** (*section 1.1.3.7 "Naphthalene Diimide compounds", Introduction*).

The in-house library of compounds was composed of GQC-05, GSA-0932, GSA-0820 and QuindolineI. All compounds were diluted in DMSO at the final concentration of 20 mM and stored at -20°C. For three compounds (GQC-05, GSA-0820 and QuindolineI) the chemical structures are known while for one (GSA-0932) it is not available since the molecule has not been published yet. However, moleculer weights fpr alla the four compounds are available, so that we were able to

prepare the solution at the expected concentration. Available chemical structures are reported in **Figure 22a-c** (*section 1.1.3.4 "Quindoline and other molecules", Introduction*).
Control compound Acyclovir (ACV) was purchased from Sigma-Aldrich (Milan, Italy) and diluted at the final concentration of 5 mM in sterile milliQ water as provided by the manufacturer's instructions.

Technique	Oligo Name	Oligo sequence (5′- 3′)
CD/ MS	*gp054a*	GGGGTTGGGGCTGGGGTTGGGG
	un2	GGGGGCGAGGGGCGGGAGGGGGCGAGGGG
	un3	GGGAGGAGCGGGGGGAGGAGCGGG
	hTel21	GGGTTAGGGTTAGGGTTAGGG
	c-myc	TGGGGAGGGTGGGGAGGGTGGGGAAGG
	c-kit2	CGGGCGGGCGCGAGGGAGGGG
	LTR-III	TGGGAGGCGTGGCCTGGGCGGGACTGGGGT

Table 3: Oligonucleotide sequences used in the study of c-exNDI and in the screening of the library of compounds for circular dichroism (CD) and mass spectrometry competition assay (MS).

3.1.3 Circular Dichroism (CD)

Circular dichroism (CD) experiments were performed using a Chirascan- Plus (Applied Photophysics, Leatherhead, UK) equipped with a Peltier temperature controller using a quartz cell of 5mm path length. In each experiment, G4 folding oligonucleotides were diluted at a final concentration of 4 μM in KCl and 10 mM Lithium Cacodilate pH 7.4. C-exNDI was studied at KCl concentration of 100 mM, 20 mM, 5 mM and 0 while the screening of the in-house library of compounds was performed at KCl 100 mM and 2.5 mM. After annealing step (5 min at 95°C), DNA samples were gradually cooled down and after 4h the G4 ligand (c-exNDI or compounds of the in-house library) was eventually added from stock at a final concentration of 16 μM (molar ratio between oligonucleotide and compound 1:4). Thermal unfolding analyses were recorded from 230 to 320 nm over a temperature range of 20- 90°C with temperature increase of 5°C/min.
The reported spectrum of each sample represents the average of 2 scans and is baseline- corrected for signal contributions due to the buffer. Observed ellipticities were converted to mean residue ellipticity (θ) = deg × cm^2 × $dmol^{-1}$ (mol ellip). T_m values were calculated according to the Van't

Hoff equation, applied for a two- state transition from a folded to an unfolded state, assuming that the heat capacity of the folded and unfolded states are equal.

3.1.4 *Taq* Polymerase Stop Assay

The 5′-end labeled primer (HSV Taq primer, **Table 4**), was annealed to the specific template (**Table 4**) in Lithium Cacodylate buffer. In the c-exNDI study where KCl should be present, samples were incubated with KCl 10mM in the presence or absence of c-exNDI (0, 37.5 nM, 75 nM and 150 nM) at room temperature. Elongation was carried out at 60°C for 30 min for all the tested sequences (*gp054a*, *un3* and *un2*). In the library of compounds screening samples were incubated with KCl 2.5 mM in the presence or absence of the compounds GQC-05, GSA-0932, GSA- 0820 and QuindolineI (250 nM, 500 nM, 1 μM, 2 μM) at room temperature. Elongation was carried out at 42°C for 30 min for *un3* sequence and at 60°C for 30 min for *gp054a* and *un2* sequences. In both studies elongation was carried out using 2 U of Ampli*Taq* Gold DNA polymerase (Applied Biosystem, Carlsbad, California, USA). All reactions were stopped by ethanol precipitation and primers elongation products were resolved on a 15% polyacrylamide denaturing gel. Gel visualization was performed with phosphoimaging (Typhoon FLA 9000) while gel quantification was performed using ImageQuant TL Software (GE Healthcare, Milan, Italy).

Technique	Oligo Name	Oligo sequence (5′- 3′)
Taq Polymerase Stop Assay	HSV Taq primer	GGCAAAAAGCAGCTGCTTATATGCAG
	HSV Taq no G4 cnt	TTGTCGTTAAAGTCTGACTGCGAGCTCTCA GATCCTGCATATAAGCAGCTGCTTTTTGCC
	gp054a	TTTTTGGGGTTGGGGCTGGGGTTGGGGTTTTTCTG CATATAAGCAGCTGCTTTTTGCC
	un2	TTTTTGGGGGCGAGGGGCGGGAGGGGGCGAGGGG TTTTTCTGCATATAAGCAGCTGCTTTTTGCC
	un3	TTTTTGGGAGGAGCGGGGGGAGGAGCGGGTTTTT CTGCATATAAGCAGCTGCTTTTTGCC

Table 4: Oligonucleotide sequences used in the study of c-exNDI and in-house library screening for *Taq* polymerase stop assay.

3.1.5 Mass Spectrometric (MS) competition assay

For mass spectrometric (MS) competition assay oligonucleotides were heat-denatured and folded in 0.8 mM KCl, 120 mM trimethylammonium acetate (TMAA), pH 7.0, and 20 % isopropanol (IPA) overnight at 4 °C. The oligonucleotides where diluted to a final concentration of 4 µM and incubated with c-exNDI at ratio DNA:compound 1:1 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 Agilent 1290 Infinity HPLC (Agilent Technologies, Santa Clara, CA,US) equipped with an auto sampler; the carrying buffer was TMAA 80 mM, 20% IPA. The absence of potassium in the carrying buffer allows the dilution of the salts before MS analysis decreasing the signal suppression. Up to 5 µL samples were 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 set at 65 V. All these parameters ensured minimal fragmentation of the DNA complexes. The instrument was calibrated using a 2 mg/mL solution of sodium iodide in 50% of IPA. Additionally, the use of the LockSpray during the analysis provided a typical <2 ppm mass accuracy. The internal standard LockSpray consisted of a solution of leu-enkephalin 1 µg/mL in acetonitrile/water (50:50, v/v) containing 0.1% of formic acid. Binding affinities were calculated for each experiment using the reconstructed-ion chromatogram area for each species calculated by MassLynx V4.1: this analysis was made possible by the experimental design that used an HPLC system to inject the samples in the mass spectrometer. The binding affinity was calculated with the following formula: [BA = (ΣG4b/(ΣG4f + ΣG4b)) × 100], where BA is the binding affinity, G4b is chromatogram area of bound G4 DNA, and G4f is the chromatogram area of free G4 DNA. G4b comprises DNA with one or two bound c-exNDI molecules, where present. Signals with charge states , 5− and 4− were used for the free and bound hTel21 (hTel) DNA and ckit2 (kit). Signals with charge states 6−, 5− and 4− for the free and bound c-myc (myc), *gp054a* and *un2*; signals with charge states 5− and 4− for *un3* and signals with charge states 5- and 6- for LTR-III G4.

3.1.6 Cytotoxicity test

The cytotoxic effect of c-exNDI and the in-house library of compounds, respectively on Vero and U2OS cells was determined by using the colorimetric 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay (Sigma-Aldrich, Milan, Italy) according to the manufacturer's instructions.

Vero and U2OS cells were seeded onto 96-well culture plates at cell density of $1x10^4$ cells per well. Serial dilutions of c-exNDI were prepared in DMEM in a final volume of 100 µl (at concentrations from 100 nM to 0.7 nM, dilutions 1:2). Control samples were treated with DMSO 0.02%.
The plates were incubated at 37°C with 5% CO_2 for 48 hours.
Serial dilutions (from 40 µM to 10 nM) of library compounds were prepared and dispensed to 24h plated U2OS cells. Control samples were treated with DMSO 0.02%. The plates were incubated at 37°C with 5% CO_2 for 24 hours.
After incubation, 10 µL of freshly diluted 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) solution 5mg/mL (in PBS 1X) were added into each well and incubated for 4 hours at 37°C with 5% CO_2.
Insoluble formazan crystals formed by viable cells were then solubilized with solubilizing solution (10 % SDS, 36% HCl) O/N at 37°C.
Absorbance was measured using the Sunrise Tecan plate reader (Mannendorf, Switzerland) at 620 nm. The 50% cytotoxic concentration (CC_{50}) was determined from the dose-response curve and lower non-toxic concentrations were used in subsequent experiments.

3.1.7 Antiviral assays

The antiviral activities of c-exNDI and of the in-house library of compounds against HSV-1 were investigated by plaque assay. C-exNDI was used from 0.7 nM to 100 nM, 1:2 dilutions while compounds from the in-house library were used from 10 nM to 400 nM, dilutions 1:5 and 1:2 to cover a broad range of concentrations.
Vero and U2OS cells infection was performed as described above on plated cells at a density of $8*10^4$ cells/ well at a MOI of 1. After infection, each compound was added at the chosen concentrations. Because a single round of HSV-1 replication takes around 24 h to complete and at this time p.i. the production of virus reaches a plateau [237,238], supernatants were collected at 24 hours post infection (h.p.i) and then stored at -80°C until viral titrations. Untreated infected cells were used as negative controls while infected controls were treated with DMSO (0.02%), solvent used to dissolve compounds at their final stock concentrations.
For viral titrations Vero or U2OS cells were seeded at a density of $1x10^5$/ well in 24-well culture plates and incubated overnight at 37°C with 5% CO_2. Cells were then infected with 250 µl of serially-diluted (10-folds) supernatants for 1 hour at 37°C. Each dilution was tested in triplicate. After the infection cells were washed with PBS 1X and incubated with 500 µl of DMEM, supplemented with 0.6% methylcellulose and 2% FBS, for 48 hours at 37°C with 5% CO_2.

At 48 h.p.i cells were washed with PBS 1X and fixed with formaldehyde 5% in PBS 1X for 10 min at room temperature, then colored with crystal violet 0.8% (in ethanol 50%). Viral plaques were counted using an optical microscope (Zeiss, Jena, Germany).

3.1.8 Time of addition assay (TOA)

Time of addition assay was performed to establish which was the last step of the whole viral life cycle affected by the presence of a compound as reported by [239]. Vero or U2OS cells were seeded in 24-well plates at a density of $8*10^4$ cells and incubated at 37°C with 5% CO_2 overnight. As described above cells were infected with HSV-1 at a MOI of 0.5 (as suggested in [239]) and treated every two hours (from 0 to 12 h.p.i for c-exNDI and from 0 to 10 h.p.i for compounds of the library) with c-exNDI (100 nM) or compounds from the in-house library (400 nM) and Acyclovir (3 μM and 1.6 μM depending on the compounds studied) as reference drug. To allow estimation of the compounds effects at all viral steps, supernatatns were collected at 30 h.p.i and then titrated following plaque assay working protocol as described above.

3.1.9 Flow cytometry

Flow cytometry was used to compare antiviral activites of c-exNDI and ACV.
Vero cells were seeded into 12-wells plates at a density of $1*10^5$ cells/ well. Infection of Vero cells was carried out with mutant HSV-1 [V41] expressing VP16- GFP at a MOI of 0.3 using the standard procedure usually used for infections. MOI of 0.3 was used because at hgher MOIs the GFP signal reached saturation levels. After infection Vero cells were incubated for 24 h with complete medium (DMEM supplemented with 10% FBS) or with complete medium containing c-exNDI (100 nM) or ACV (3 μM). At 24 h.p.i GFP-fluorescence was acquired for 30.000 events in each sample using FACS Cytofluorometer BD LSR II (BD Bioscences, New Jersey, USA). Fluorescence acquisitions were analyzed with FACS DIVA software (BD Bioscences, New Jersey, USA) and FlowJo software (Tree Star, Oregon, USA). Activities of c-exNDI and Acyclovir against HSV-1 [V41] were plotted as mean of GFP Fluorescence in function of the time.

3.1.10 Quantitative PCR (qPCR)

For quantitative PCR (qPCR) cells were plated ($1.8*10^5$/ well) in 6-wells plates, then mock- or HSV-1 (F)-infected (MOI of 1) and treated with c-exNDI 100 nM. At 4 or 24 h.p.i. cells were collected and total DNA was extracted using NucleoSpin Tissue (Macherey- Nagel, Duren, Germany) according to the manufacturer's instruction. Isolated intracellular DNA was analyzed

by realtime PCR using TaqMan chemistry with 5'-[FAM] and 3'-[TAMRA]-end labeled probes (**Table 5**) in a final volume of 25 µL. qPCR reaction was composed of FAST Master Mix 2X (Applied Biosystem, Monza, Italy), 500 nM of forward/reverse primer mixture (US1 gene, **Table 5**), 200 nM of TaqMan probe, sterile water and 10 µL of isolated DNA. Experiments were performed using ABI 7900 HT – FAST RealTime PCR System under the following conditions: 95°C for 5 min followed by 45 cycles of 5 seconds at 95°C and 10 seconds at 60°C. Each sample was analyzed in duplicate.

Technique	Gene name	Gene class	Primer/ Probe	Sequence (5'- 3')	Viral protein	Amplicon lenght (bp)
qPCR/ Real- time PCR	US1	α	Fw Rev Pr	GGCCCGGAGTGTGATCTTAG GGTGGCATCGGAGATTTCAT [FAM]AGATTCATCTCAGCGCGACAAGCGA[TAMRA]	ICP22	70
Real- time PCR	US12	α	Fw Rev Pr	CAACGGGTTACCGGATTACG TTGGGTGTGGCACATCGA [FAM]ACTGTCGGTCACGGTCCCGCC[TAMRA]	ICP47	68
	UL30	β	Fw Rev Pr	TTCGACTTTGCCAGCCTGTA CAGGGAGAGCGTGCTGAAG [FAM]AGCATCATCCAGGCCCACAACCTG[TAMRA]	DNA polymerase	69
	UL36	γ	Fw Rev Pr	AGGGAGGATGCCCACGAA TCCGCGTCTTCCACAAATC [FAM]ACACCCCAGACAAGGAGCTGCCC[TAMRA]	Ubiquitin specific protease	68
	b-actin	HK	Fw Rev Pr	TCACTGAGCGCGGCTACA CCTTAATGTCACGCACGATTTC [FAM]TCACCACCACGGCCGAGCG[TAMRA]	b-actin	69

Table 5: Primers and probe sequences used to perform qPCR and Real-time PCR. Fw stands for forward primer, Pr stands for probe and Rev stands for reverse primer.

3.1.11 Real-time PCR

For total RNA extraction Vero cells were seeded into 6- wells plates at a density of $1.8*10^5$ in 6-wells plated and then mock infected or infected with HSV-1 (F) at a MOI of 3 and treated with c-exNDI 100 nM. At 4 or 24 h.p.i total RNA was isolated using TRIzol reagent (Life Technologies, Monza, Italy), according to the manufacturer's instructions. RNA-se free DNase I treatment (Ambion Turbo DNA free, Life Technologies, Monza, Italy) was performed on the eluted RNA to avoid residual DNA contamination. One microgram (1 µg) of the extracted RNA was subjected to reverse transcription by 1.5 U/µL MuLv. Each 50 µL of retrotranscription mixture contained 1X PCR buffer II, 3 mM $MgCl_2$, 0.5 mM dNTPs, 2 µM random hexamers, 0.5 U/µL RNase inhibitor and DEPC water (Life Technologies, Monza, Italy). Reverse transcription was carried out under

the following conditions: 10 min at 25°C, 60 min at 48°C and 5 min at 95°C using the Thermo Cycler Verity 96 (Applied Biosystem, Monza, Italy). Forward/ reverse primers used to perform real-time PCR were designed within conserved HSV-1 gene sequences using Primer Express 3 (Applied Biosystem, Monza, Italy) (**Table 5**). All primers were purchased from Sigma- Aldrich (Sigma- Aldrich, Milan, Italy) and before their use were tested for potential cross-reactivity with unreleated viral or other sequences using the BLAST alignment tool (http://blast.ncbi.nlm.nih.gov/). Real-time PCR was performed using TaqMan chemistry with 5'-[FAM] and 3'-[TAMRA] end labeled probes designed to anneal between the two primers. Real-time reaction (final volume of 15 μL), was composed of FAST Master Mix 2X (Applied Biosystem, Monza, Italy), 900nM of forward/reverse primer mixture, 200nM of TaqMan probe, sterile water and 3μL of cDNA. Experiments were performed using ABI 7900 HT – FAST RealTime PCR System under the following conditions: 95°C for 10 min followed by 40 cycles of 30 seconds at 95°C and 30 seconds at 58°C, 1 minute at 72°C. Messenger RNA transcription levels were standardized against the housekeeping gene β-actin. Non- treated retrotranscribed HSV-1 RNA was used as mRNA expressed control. Each sample was analyzed in duplicate.

3.2 HSV-1 STUDY: G-quadruplex binding protein identification

3.2.1 Pull-down assay

For pull-down assay U2OS (Homo sapiens bone osteosarcoma cells from ATCC® HTB-96) cells were used thanks to their permissiveness to HSV-1 infection. U2OS cells were grown in 150- cm^2 flasks at a density of $5.5*10^5$ cells/flask and were infected with HSV-1 at MOI of 3 using the same infection procedure described above. At 8 h.p.i cells were collected and nuclear proteins were extracted using the CelLytic™ NuCLEAR Extraction kit (Sigma Aldrich, Milan, Italy) following the manufacturer's procedure.

Nuclear extracted proteins were quantified using the Pierce® BCA Protein Assay kit (Thermo Scientific, Rockford, IL, USA) according to the manufacturer's protocol.

Nuclear protein extraction was monitored by western blot analysis by using ICP4 and α-tubulin presence as extraction marker as described below.

Briefly, each sample was electrophoresed on 10% SDS-PAGE for 1h at 90V and transferred (24 min at 19V) to a nitrocellulose blotting membrane (Amersham TM Protan TM, GE Healthcare Life science, Milan, Italy) by using trans- blot SD semi- dry transfer cell (Bio-Rad Laboratories, Milan, Italy).

The membrane was then blocked with 5% skim milk in PBST (0.05% Tween20 in PBS 1X). Membranes were then incubated O/N at 4° with either mouse monoclonal anti HSV-1-ICP4 antibody (ab6514; Abcam, Cambridge, UK) or mouse α-tubulin (T9026, Sigma- Aldrich, Milan, Italy) used at a dilution of respectively 1:500 and 1:5000 in 1% milk- PBST (0.05% Tween20 in PBS 1X). After incubation, membranes were washes in PBST, then incubated for 1 h at room temperature in the dark with ECL Plex-Goat-α- Mouse IgG-Cy5 (Ge Healthcare, Milan, Italy) diluted in 1% milk- PBST.

Images were acquired with Typhoon FLA 9000, and bands were quantified using ImageQuant TL software.

HSV-1 5'- biotinilated G4 forming sequence (*un3long3*) and a negative control for G4 formation (LTR II+ III+ IV scrambled) (**Table 6**) were previously annealed (5 min at 95°C) in buffer A (20 mM Phosphate buffer pH 7.4, KCl 100 mM) and gradually cooled down at room temperature before being subjected to beads attachment and protein pull-down.

300 pmol (6 μM) of folded 5'-biotinilated DNA sequences (un3long3 and LTR II+ III+ IV scrambled) were attached to 50 μL of Dynabeads® M-280 Streptavidin (Invitrogen, Life Technologies, Carlsbad, USA) at 4°C O/N under agitation. Oligonucleotides concentration was monitored before and after beads attachment. Then 25 μg of nuclear protein extract, diluted 1:2 in

milliQ water, were incubated with magnetic beads for 2 h at 4°C under agitation. After 3*100 µL buffer A washes, proteins were eluted in 20 µL of 2 M NaCl buffer for 5 min in ice and then were secondly eluted with 20 µL of 2X Laemmli buffer (4% SDS, 10% glycerol, 10 mM DTT, 125 mM Tris- HCl pH 6.8). Supernatants were stained for SDS-PAGE separation with 2X Laemli buffer, denatured for 5 min at 95°C and then electrophoretically separated (1 h at 120 V). Bands visualization was performed with Colloidal Coomassie solution (0.02% w/v coomassie brilliant blue CBB G-250, 5% w/v ammonium sulphate-(14-18)-hydrate, 10% v/v ethanol (96%), 2% v/v ortophosphoric acid (85%) O/N under agitation. Gel de-staining was preceded by two milliQ water washes and was then performed with de- staining solution (10% v/v ethanol (96%), 2% ortophosphoric acid (85%) for 10 min .

Technique	Oligo Name	Oligo sequence (5'- 3')
Pull-Down	*un3long3*	[Btn]GGGAGGAGCGGGGGGAGGAGCGGGGGGAGGAGCGGGGGG AGGAGCGGGGGGAGGAGCGGGGGGAGGAGCGGG
	LTR II+ III+ IV scrambled	[Btn]TTTTTGGAGCGTGTGTGTGCGCGAGAGCGTGCGCGCGTGG CGAGCGTGGAGTGGTTTTT

Table 6: HSV-1 *Un3long3* and HIV-1 LTR II+ III+ IV srambled G4 sequences used for pull-down assay.

3.2.2 Mass spectrometric (MS) protein identification

Bands that represent candidate proteins able to bind to G4 were cut from the gel and were treated according to established in-gel proteins digestion protocols.

Bands were first destained with 500 µl of destaining solution (50% CH_3OH, 2.5% acetic acid) for 2 h at RT and then for another 1 h with fresh destaining solution. Then bands were dehydrated with CH_3CN twice and then reduced with 50 µl of DTT (10 mM in 100 mM NH_4HCO_3) for 30 min at room temperature.

The excess of DTT was eliminated before treating the bands with iodoacetamide (50 mM in 100 mM NH_4HCO_3) to alkylate cisteine residues.

Bands were then washed with 100 mM NH_4HCO_3, dehydrated twice with CH_3CN and digested 30 min on ice with 1µg of MS-grade trypsin (Thermo Fisher Scientific, Waltham, Massachusetts, USA) diluted in 50 µL of 50 mM NH_4HCO_3. The excess of trypsin was eliminated and bands were incubated with 30 µl of 50 mM NH_4HCO_3 O/N at 37°C. Peptides were first extracted twice with 5% formic acid and then extracted two more times with 50% CH_3CN and 5% formic acid. The peptide mixture collected in all these steps was further separated in a LC chromatography silica column, packed with C18 pack material (Thermo Fisher Scientific, Waltham, MA, USA), and then analyzed by direct infusion electrospray ionization (ESI) on a Xevo G2 mass spectrometer Quadrupole-Timo

of flight (Q-TOF) (Waters, Milford, Massachusetts, USA). All analyses were performed in nanoflow mode and up to10 µl were loaded automatically. Empty gel bands were used as negative controls. Putative peptides that were not present in blank samples were subjected to MS/MS analysis. Peptides charged 2+ and 3+ were employed to perform a Mascot Database search to identify their parent protein.

3.3 MV STUDY: Investigation of G-quadruplex in MV genome

3.3.1 G-quadruplex analysis in MV genome

Measles virus (MV) genome from the NCBI nucleotide database (NC_001498.1 from https://www.ncbi.nlm.nih.gov/nuccore) was analyzed by QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/index.php) for prediction of G4-putative forming sequences. The following restrictions were applied to the analysis: (i) maximum oligonucleotide length of 45 nucleotides, (ii) minimum G-group size of 3 nt and (iii) loop size from 0-12 nt. This analysis was applied for both the deposited reference sequence (cRNA with positive polarity 5'-3') and its reverse complement sequence, automatically obtained from the website, which corresponds to the negative stranded RNA MV genome (3'-5'). G4-putative folding sequences obtained from both analyses were then used for subsequent studies.

3.3.2 Analysis of sequence conservation

The conservation of each G4-putative folding sequence was determined by looking at different aspects of the patterns and distinguishing between the G-runs and their connecting loops. The following categories were defined for subsequent analyses:

- G4_islands: the basic structural units of a G4, which included only the G-runs;
- G4_islands_loops: the complete sequence of a G4-putative folding sequence, which included both G-runs and their connecting loops

Then, different indexes were established to measure the nucleotide sequence conservation of each measles strain genome and relative G4s:

- G4_scaffold_conservation_index (G4_SCI): referred to the G4_islands and calculated as the percentage of independent genomes with the corresponding G4_island (calculated in each viral genome and for each detected G4-pattern). **$G4\_SCI = N_{G4\_islands} / N_{tot} * 100$** where $N_{G4\_islands}$ is the number of sequences possessing the N_{G4_island} in a certain genome position and N_{tot} is the total number of sequences available for a virus.
- *Delta conservation*: referred to the difference between G4_SCI and the local conservation of viral genome sequence. ***Delta conservation* = G4_SCI-LSC** where LSC (Local Sequence Conservation) is the average sequence conservation of all viral sequences in a multiple alignment calculated with a sliding window of 20 and shift of 10. LCS was calculated by averaging the values of all windows overlapping a G4. For LSC calculation, the highest conservation value for each positionin in the genome was extracted from multiple sequence alignments using Jalview [240].

The results of these analyses were plotted in bar charts using Plotly (https://plot.ly/).

3.3.3 Analysis of base conservation

The degree of base conservation was applied for all G4-putative folding sequences previously identified. Fifty-five deposited sequences of MV cRNA were aligned using the FastAlign.pl software. The degree of conservation was expressed as % of base conservation and was automatically calculated by the WebLogo software (http://weblogo.berkeley.edu/logo.cgi) at the moment of the consensus sequence extrapolation. Each consensus sequence was calculated on the alignment previously described. The LOGO representation was expressed as conservation of the G-pattern for the G4-putative folding sequence found in the deposited sequence (RNA +) while for sequences found in its reverse complement (RNA -) was expressed as conservation of the C-pattern.

3.3.4 Oligonucleotides and compounds

RNA G4-putative folding sequences obtained from the QGRS Mapper software analysis were purchased from Sigma-Aldrich (Milan, Italy) and are reported in **Table 12** in *section 4.4.1, Results and discussion*. Oligonucleotides were resuspended to a final concentration of 100 μM with DEPC Water (Life Technologies, Monza, Italy). Concentrations were checked by using Nanodrop 1000 Spectrophotometer (Thermo Scientific, Rockford, IL, USA).

G-quadruplex ligand B-19 was purchased from Saarbruecken (Germany) and diluted in sterile milliQ water at the stock concentration of 5 mM. The compound was stored at -20°C. Ribavirin, reference drug in the treatment of MV, was purchased from Sigma-Aldrich (Milan, Italy) and diluted at the final concentration of 5 mM in sterile milliQ water.

3.3.5 Cell line and virus

Vero cells (African green monkey kidney cells from Sigma-Aldrich, Saint Louis; USA) were maintained and propagated 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 1X of Fungizone, Ciprofloxacin, Ampicillin and Gentamicin (Gibco, Life Technologies, Monza, Italy) in humidified incubator maintained at 37°C, 5% CO_2.

Measles virus (MV) was isolated from the trivalent vaccine MMR (Measles, Mumps and Rubella Vaccine, Merck, Darmstadt, Germany). The virus corresponds to the A isolate from Edmonston-

Schwarz Measles Virus strain. It was a kind gift of Dott. M. Pacenti (Department of Microbiology and Virology, Padua Hospital, Italy).
For virus infection, Measles virus were applied to plated cells ($7*10^4$ cells/well in 24- wells plates) at a multiplicity of infection (MOI) of 0.1 in serum- free medium. After 1 h of incubation at 37°C, the inoculums were replaced with complete medium.

3.3.6 Circular Dichroism (CD)

Circular dichroism (CD) experiments were performed using a Chirascan- Plus (Applied Photophysics, Leatherhead, UK) equipped with a Peltier temperature controller using a quartz cell of 5mm path length. In each experiment, G4 folding oligonucleotides were diluted at a final concentration of 4 μM in KCl (100 mM), 10 mM Lithium Cacodilate pH 7.4. After annealing step (5 min at 95°C), RNA samples were gradually cooled down and after 4 h G4-ligand B-19 was eventually added from stock at a final concentration of 16 μM (molar ratio between oligonucleotide and compound 1:4). All samples were measured with Nanodrop 1000 Spectrophotometer (Thermo Scientific, Rockford, IL, USA). CD spectra were performed using the parameters described above.
The reported spectrum of each sample represents the average of 2 scans and is baseline- corrected for signal contributions due to the buffer. Observed ellipticities were converted to mean residue ellipticity (θ) = deg × cm^2 × $dmol^{-1}$ (mol ellip). T_m values were calculated according to the Van't Hoff equation, applied for a two- state transition from a folded to an unfolded state, assuming that the heat capacity of the folded and unfolded states are equal.

3.3.7 Antiviral activity

The antiviral activity of B-19 against measles virus (MV) was investigated by real-time PCR and espressed as % of genome copies/mL. The compound was used at 100 μM, since in previous tests we already observed the absence of cytotoxicity. Antiviral activity of B-19 was compared with the activity of Ribavirin (RBV), a broad spectrum antiviral nucleoside analogue also used in the treatment of measles *in vitro* [241,242]; RBV was used at 10 and 50 μM, concentrations in the range of IC_{50} against MV in Vero cells [241,242].
Vero cells infection was performed as described above. After infection, compounds were added at the chosen concentrations. Supernatants were collected at 24 and 48 h.p.i and stored at -80°C until viral titrations. Untreated infected cells were used as negative controls while infected controls were treated with milliQ sterile water.

3.3.8 Real-time PCR

Genomic RNA of MV was extracted directly from supernatants of the antiviral assay at 24 and 48 h.p.i. 100 µL of each supernatant were incubated at 4°C with 100 µL NucliSENS easyMAG Lysis buffer (BioMerieux, Marci-l'Etoile, France) until extraction. Extraction was carried out using the Pathogen Universal 200 program from Magna Pure 96 Extractor (Roche, Basel, France). Samples containing 3µL of 900 nM primers/ 200 nM probe mixture, 1.67 µL of Detection Enhancer (4388513, Ambion, Thermo Fischer, Waltham, Massachusetts, USA), 12.5 µL of FAST Master Mix 2X buffer, 1 µL of 25X RT-PCR Enzyme (Ambion, Thermo Fischer, Waltham, Massachusetts, USA), sterile milliQ water and 5 µL of extracted RNA were subjected to reverse transcription and real-time PCR. Forward/ reverse primers used to perform real-time PCR were designed to amplify the a 72 bp selected region of the N-gene conserved region. All primers were kindly provided from Dott. E. Franchin (Department of Microbiology and Virology, Padua Hospital, Italy). TaqMan chemistry was used for probe design. Probe sequence was 5'-[FAM] and 3'-[TAMRA] end labeled and was designed to anneal between the two primers in the N-gene conserved region. Primers, probe and N-gene region sequences are reported in **Table 7**. RNA reverse transcription and real-time PCR were coupled in an unique protocol and performed using ABI 7900 HT – FAST RealTime PCR System under the following conditions: 50°C for 30 min for the reverse transcription followed by 10 minutes at 95°C and 45 cycles of 15 seconds at 95°C and 1 minute at 60°C. For quantification a serial dilution of measles cDNA of the 72 bp selected region of the N-gene cloned in the pGEM® -T Easy plasmid (Promega, Madison, Wisconsin, USA) was used. Each sample was analyzed in duplicate and quantification was expressed as genome copies/mL.

Primer/ Probe	Sequence (5'-3')
N-gene region	CGGCAGTGATACCGAGTTCGGATGCCAATGTGGAACTGACCTTTCCAGCTGACCTCCTCACCATCTCTTGCC
Fw	CGGCAGTGATACCGAGTTC
Rev	GGCAAGAGATGGTGAGGAGG
Pr	AATGTGGAACTGACCTTTCC

Table 7: Primers, probe and N-gene region sequences used to perform Real-time PCR for MV genome copies/ mL quantification. Fw stands for forward primer, Pr stands for probe and Rev stands for reverse primer.

4. Results and Discussion

4.1 HSV-1 STUDY: C-exNDI

4.1.1 The G4-ligand c-exNDI highly stabilizes the HSV-1 G4s

As described by Artusi S. *et al* [87], nine regions characterized by G-rich repeated tracts were identified within the genome of HSV-1. All the identified sequences were analyzed by CD and their ability to fold into G4s was demonstrated [87].

Starting from the observation of the ability of these sequences to fold into G4s, we focused our attention on the study of the interaction between G4s forming in the terminal repeat of HSV-1 and a promising G4 ligand, a c-exNDI derivative, previously developed against HIV-1 G4s [118]. It was demonstrated its ability to display a fair selectivity towards the HIV-1 G4s formed in the LTR promoter of the proviral genome versus cellular G4s (i.e. the human telomeric G4 and the G4s formed in the promoters of some oncogenes) [118].

In our work three sequences were considered: two sequences forming a four-stacked-G-quartet (*un2* and *gp054a*) and one forming a three-G-quartet G4 (*un3*) [87]. HSV-1 G4s sequences were analyzed by CD spectroscopy in the absence or presence of 100 mM, 20 mM, 5 mM K^+ and in the absence or presence of c-exNDI in order to evaluate (i) an eventual binding of the compound and subsequent (ii) change in the sequences conformation and/or (iii) a stabilization of the conformation due to the G4 ligand.

In 100 mM K^+, u*n3* sequence showed (i) a parallel topology with a maximum peak at 260 nm and a negative peak at 240 nm (**Figure 44a**) and (ii) a pronounced stability (Tm = 65.1 ± 3.03 °C) increased above 90°C upon the addition of c-exNDI (**Table 8**).

Melting temperature						
G4 sequence	KCl 0 (°C)		ΔT_m (°C)	KCl 100 mM (°C)		ΔT_m (°C)
c-exNDI	-	+		-	+	
gp054a	nd	nd	-	64.3 ± 0.84	79.5 ± 0.61	15.2
un2	50.6 ± 0.53	56.8 ± 0.62	6.2	>90	>90	-
un3	nd	nd	-	65.1 ± 3.03	>90	>24.9

Table 8: Melting temperatures (T_m, °C) of HSV-1 G4 folding sequences measured by CD in the absence or presence (100mM) of K^+. Each sequence (4 μM) was analyzed in the absence or presence of c-exNDI (16 μM). nd stands for "not detected". Adapted from [243].

The molar ellipticity was not subjected to a change due to the presence of the compound while the conformation had a slight change at 20°C at 290 nm where a "shoulder", corresponding to an additional conformational population, appeared (**Figure 44b,c**). Since in the absence of K^+ the *un3* sequence does not adopt a G4 conformation, no parameters were detected in that condition.

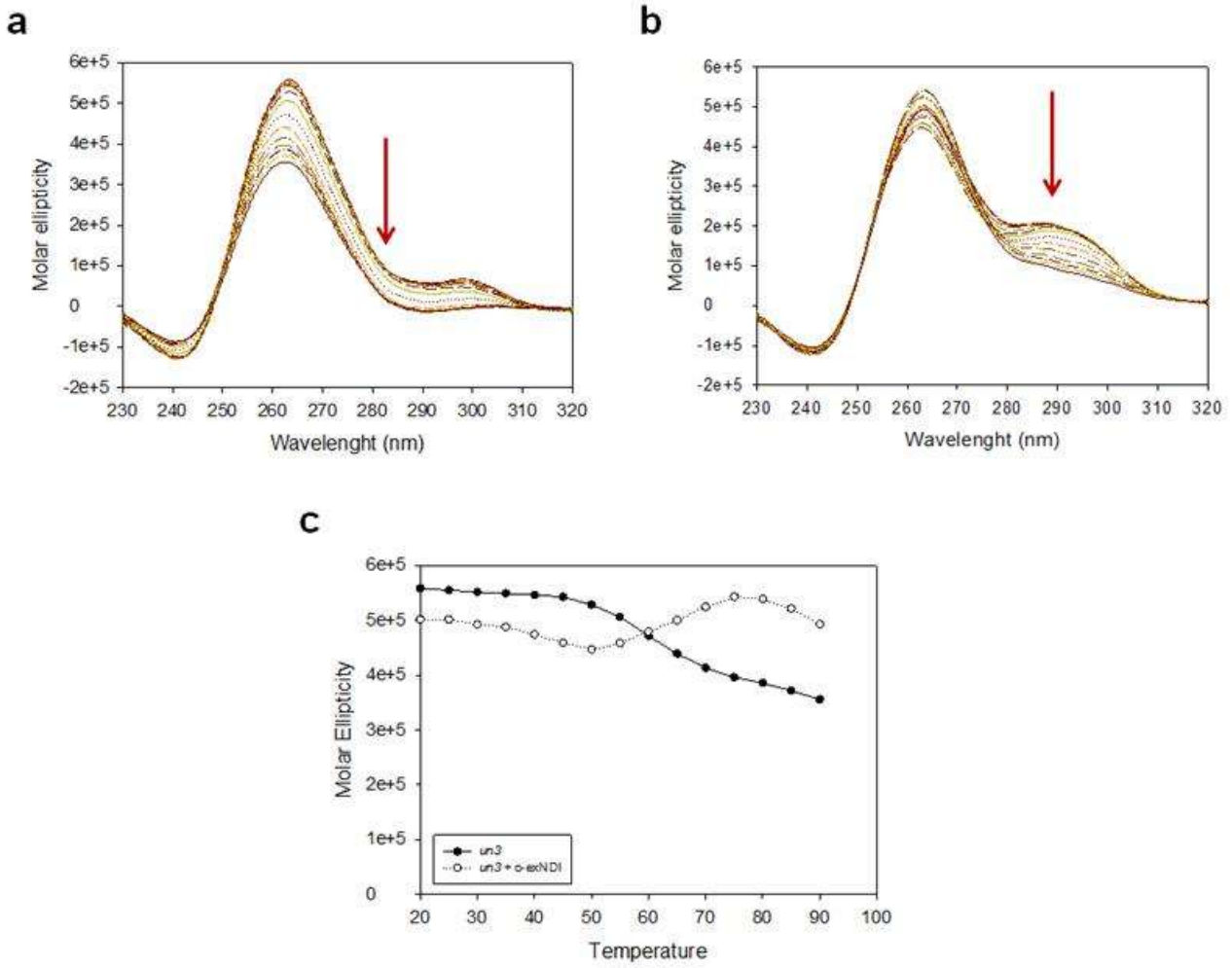


Figure 44: Thermal unfolding experiments on the *un3* parallel sequence in 100 mM K^+ (**a**) in the absence and (**b**) in the presence of 16 μM c-exNDI. Thermal unfolding analyses were recorded over a temperature range of 20 °C- 90°C with temperature increase of 5°C. (**c**) Melting temperatures calculated at 263.2 nm in the presence of 100mM K^+. Adapted from [243].

Concerning *un2*, in presence of 100 mM K^+ we observed a significant stability even in the absence of c-exNDI (**Figure 45a**) where the melting temperature calculated was above 90°C, therefore making impossible to appreciate a variation in the ΔT_m (**Figure 45b,c**). In these conditions only a mild variation in the antiparallel CD spectrum was observed upon addition of c-exNDI as shown in **Figure 45b**. To circumvent this problem, *un2* was measured at lower K^+ concentration (20 and 5 mM); however only in the absence of K^+ an appreciable ΔT_m was obtained (6.2°C) upon addition of c-exNDI (**Table 8**) as shown in **Figure 45d**.

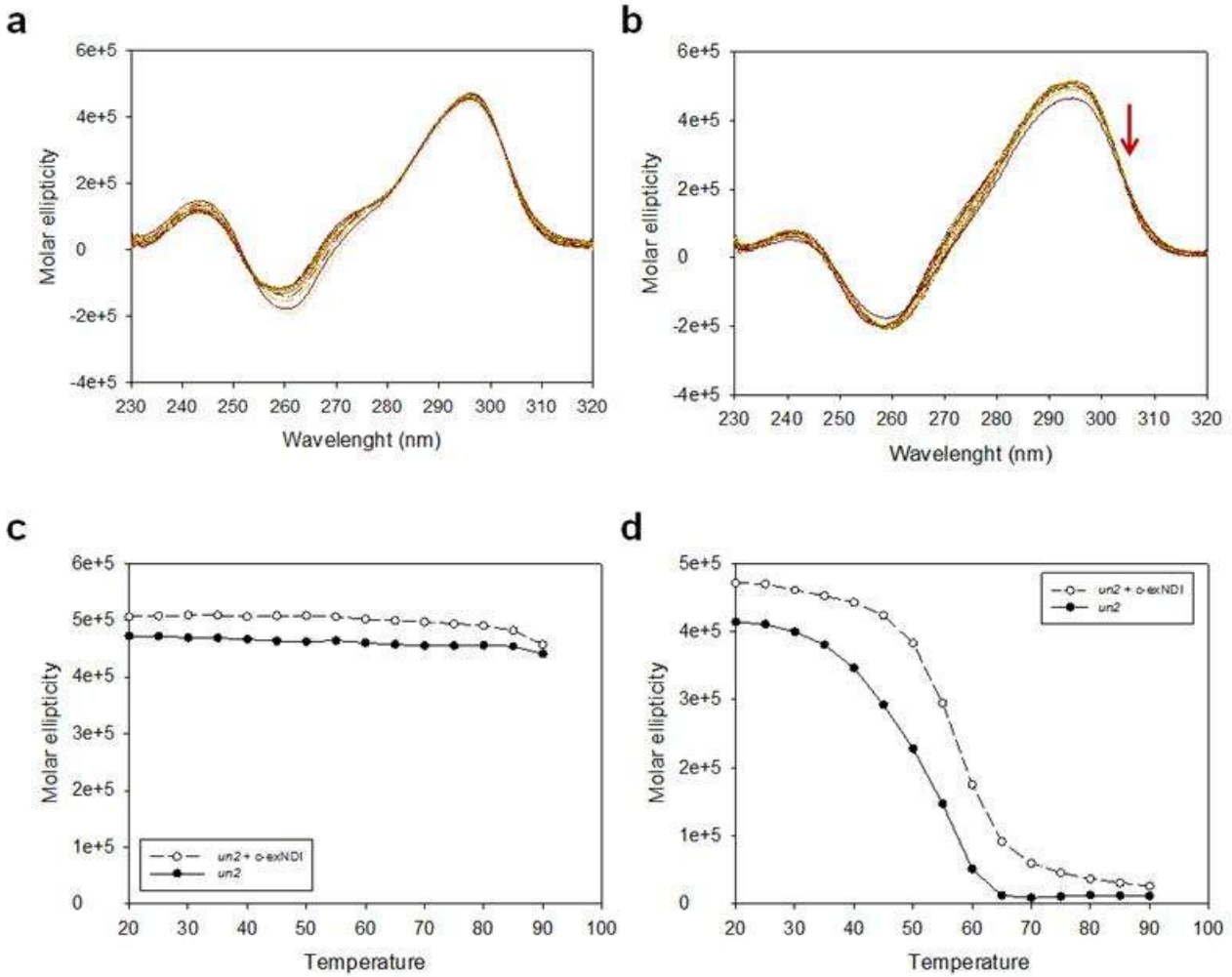


Figure 45: Thermal unfolding experiments on the *un2* antiparallel sequence in 100 mM K^+ (**a**) in the absence and (**b**) in the presence of 16 μM c-exNDI. Thermal unfolding analyses were recorded over a temperature range of 20 °C- 90°C with temperature increase of 5°C. Melting temperatures calculated at 296.4 nm (**c**) in the presence of 100mM K^+ and (**d**) in the absence of K^+. Adapted from [243].

Gp054a sequence in the presence of 100 mM K^+ showed a typical mixed-type conformation with two positive peaks at 290 nm and 260 nm (**Figure 46a**). Addition of c-exNDI stabilized the sequence melting temperature by 15.2 °C (from 64.3 ± 0.84 to 79.5 ± 0.61 °C, **Table 8**) and significantly increased the molar ellipticity of the sequence. Interestingly, during thermal unfolding at high temperatures, the presence of c-exNDI showed protection from denaturation (alternatively seen in the absence of c-exNDI at high temperature) in favor to an increased stabilization of the alternative parallel- like conformation (**Figure 46b,c**) therefore confirming the noteworthy selectivity of c-exNDI for the G4 parallel conformation already observed and described in [111].

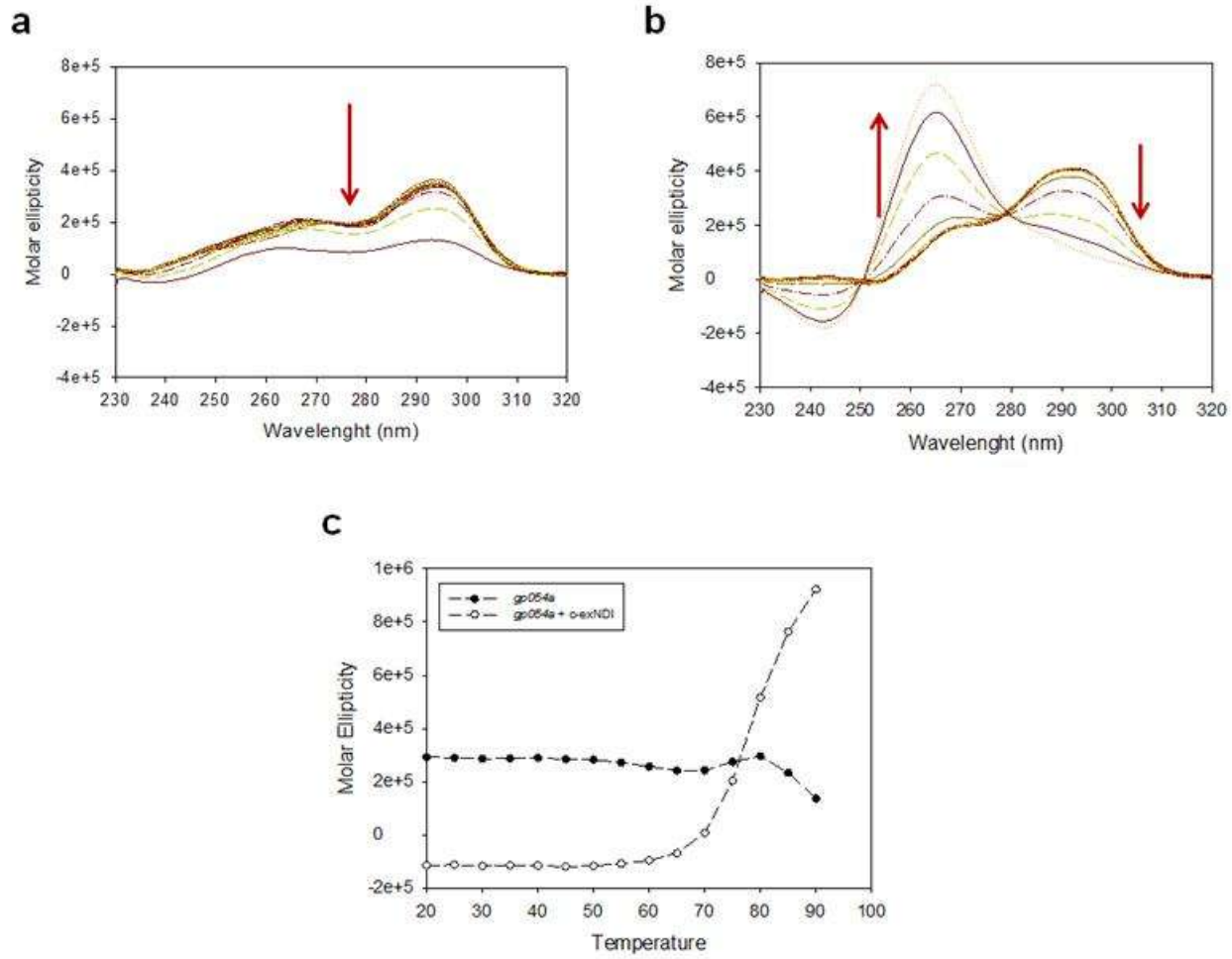


Figure 46: Thermal unfolding experiments on the *gp054a* mixed- type sequence in 100 mM K^+ (**a**) in the absence and (**b**) in the presence of 16 μM c-exNDI. Thermal unfolding analyses were recorded over a temperature range of 20 °C- 90°C with temperature increase of 5°C. (**c**) Melting temperatures calculated at 265.2 nm in the presence of 100mM K^+. Adapted from [243].

Taken together, these results suggested that c-exNDI is able to effectively bind and stabilize all G4 HSV-1 forming sequences.

To confirm the increased stability of the viral G4s in the presence of the compound, a *Taq* polymerase stop assay was set up.

Templates corresponding to *un2*, *un3* and *gp054a* sequences (**Table 4**) were annealed to a primer and incubated with *Taq* polymerase for 30 min at 60°C, a temperature lower than the T_m of the tested sequences. A sequence unable to fold into G4 was alongside assayed as negative control. As already observed with CD analyses in the absence of K^+, *un2* and *gp054a* displayed a marked propensity to fold into G4 and a marked stop site corresponding to the first G of the most 3′ G-tract of the sequence (**Figure 47a, lanes1 *un2* and *gp054a***), indicating a stable G4 folding. The stop site increased upon addition of K^+ and c-exNDI (**Figure 47a, lanes 2-5 *un2* and *gp054a***); moreover, in the presence of the compound the full-length amplified product sharply decreased, indicating

effective stabilization of *un2* and *gp054a* G4s and thus inhibition of polymerase progression by steric hindrance due to the G4 structure (**Figure 47a, lanes 3-5 *un2* and *gp054a***).

In the absence of K^+ no stop site was observed in the *un3* sequence, which thus did not form G4 in these conditions (**Figure 47a, lane 1 *un3***); upon addition of K^+ and c-exNDI, a marked stop became visible at the most 3′ G-tract (**Figure 47a, lanes 2-5 *un3***), indicating effective stabilization also of this template. Quantification of the stop sites indicated a similar degree of polymerase stalling induced by the compound in the three G4- forming sequences (**Figure 47b**).

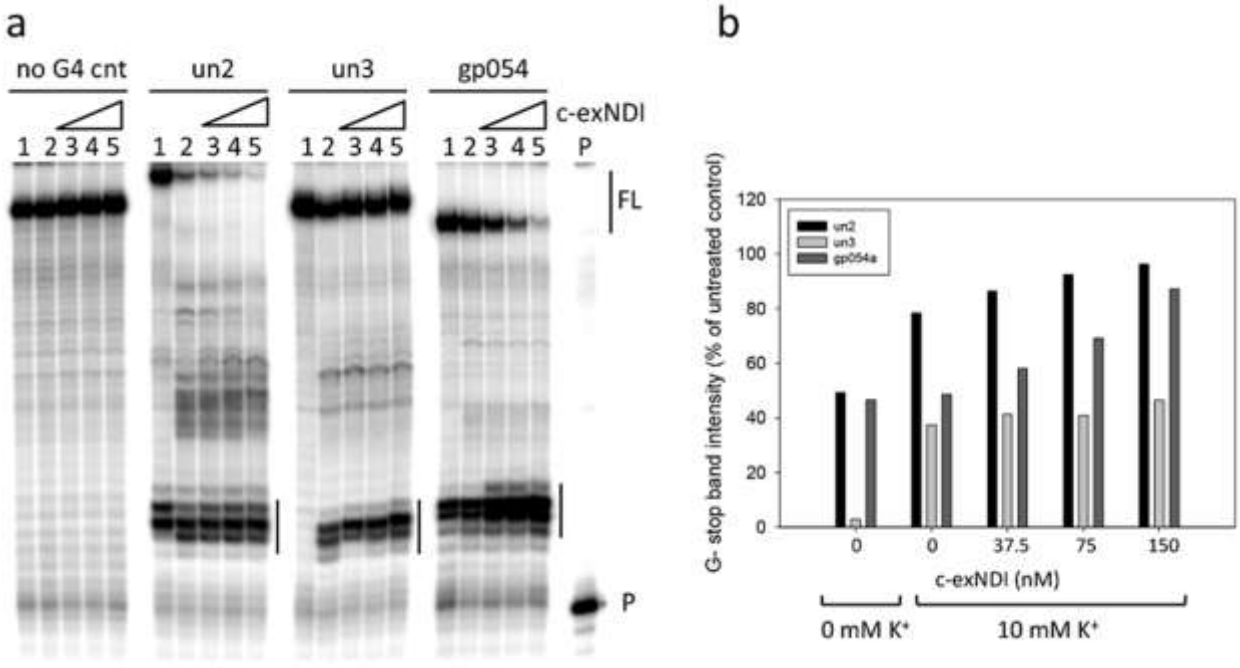


Figure 47: Taq polymerase stop assay on HSV-1 G4- folding sequences in the presence of c-exNDI. (**a**) *un2*, *un3* and *gp054a* were analyzed in the absence (lane 1) or presence of 10mM K^+ (lanes 2-5). C-exNDI was used at concentrations 37.5, 75 and 100 nM (lanes 3-5). Elongation was performed at 60°C. A non- G4 folding sequence was used as negative control (no G4 cnt). G4-related stops were indicated by vertical bars. FL stands for "full- length" amplified template. P indicates the primer lane. (**b**) Quantification of c-exNDI- induced stops sites observed in (**a**). Adapted from [243].

To check the selectivity of c-exNDI for HSV-1 G4s, mass spectrometry (MS) competition assays were performed. Selected competitor G4s were hTel, c-myc and c-kit2 G4s, which correspond to the G4- forming sequences of the human telomeric repeat, the promoter of the *c-myc* oncogene and *c-kit* protooncogene [244,245], and HIV-1 LTR-III G4, the HIV-1 G4 previously shown to be selectively bound with high affinity by c-exNDI (**Table 9**) [118].

Binding affinity c-exNDI [118]					
Competing G4s	*un2*	*un3*	*gp054a*	LTR-III	Cell (hTel/myc/kit)
un2/LTR-III	20			58	
un2/kit	24				46
un2/myc	15				68
un2/hTel	34				30
un3/LTR-III		43		61	
un3/kit		52			31
un3/myc		36			66
un3/hTel		72			21
gp054a/LTR-III			3	75	
gp054a/kit			7		64
gp054a/myc			4		78
gp054a/hTel			12		39

Table 9: Relative binding affinity, analyzed by MS competition assay for *un2*, *un3*, *gp054a*, hTel21, c-kit2 (kit), c- myc (myc) and LTR-III G4- folded oligonucleotides. Adapted from [243].

In each case, HIV-1 LTR-III G4s was preferred over HSV-1 G4s, data that confirm c-exNDI selectivity for HIV-1 G4s [118].

C- myc G4 was also always preferred and c-kit 2 G4 was preferred *vs* HSV-1 *un2* and *gp054a*. In contrast, HSV-1 *un2* and *un3* were preferred over hTel. *Gp054a* was generally less preferentially bound over all the other analysed sequences therefore suggesting a possible role of parallel (*un3*) and antiparallel (*un2*) conformations in the binding, in agreement with the preferential binding of c-exNDI to parallel conformations already demonstrated in [118]. The measured selectivity did not depend on the starting stability of the competing G4s: in fact, in 100 mM K^+, preferred G4s such as LTR-III (T_m 52.9 ± 0.2°C) and c-myc (T_m >90°C) had T_m values both lower and higher, respectively, than HSV-1 G4s (**Table 9**); in contrast, non-preferred G4s had similar T_m values (hTel 68.6 ± 0.2; c-kit2 72.9 ± 1.1°C) (**Table 9**).

We confirmed a general, even though not absolute, selectivity toward the *un* sequences therefore giving rise to the possibility that c-exNDI is able to discriminate between G4s within cells, therefore preferring HSV-1 G4s, which are also more numerous than cellular ones.

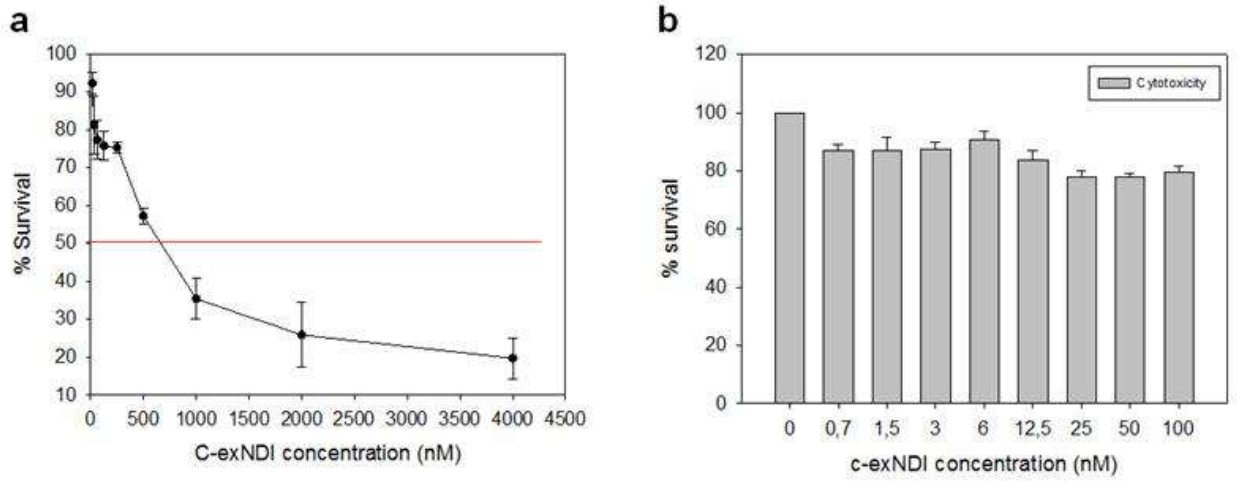
4.1.2 C-exNDI displays potent anti-HSV-1 activity

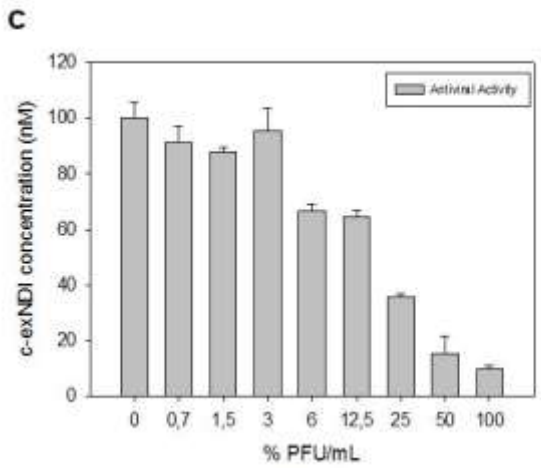
Since our data indicate that even if HSV-1 G4s are not totally preferred, nonetheless they are preferred over telomeric G4s, the most abundant cellular G4 structures in infected cells, we tested the ability of c-exNDI to inhibit HSV-1.

The 50% cytotoxic concentration (CC_{50}) of c-exNDI was first calculated on Vero cells by MTT assay to determine in which range of concentrations the compound could be used without conspicuous cytotoxicity.

The compound concentration able to kill 50% of the cells, as measured by MTT assay, was 624.8 ± 1.2 nM (**Figure 48a**). Anti HSV-1 activity of c-exNDI was assessed by plaque assay. Increasing concentrations of c-exNDI (previously tested for their cytotoxic effect on Vero cells, **Figure 48b**) were added to Vero cells infected with HSV-1 (strain F) at a MOI of 1. Supernatants were collected at 24 h.p.i and then titrated as described in the *section 3.1.7 "Antiviral assays", Materials and Methods).*

C-exNDI showed an inhibitory effect already at 6 nM (viral load reduced of about 35%) and a remarkable antiviral activity up to 90% was observed at 100 nM. The 50% inhibition of HSV-1 production (IC_{50}) was 18.3 ± 1.4 nM. Since the compound induced a pronounced inhibition of HSV-1 production with low cytotoxicity at the concentrations used, this resulted in a remarkable selectivity index (SI = 34.3 ± 2.7) (**Figure 48c,d**) strongly suggesting that this compound has an interesting therapeutic window.





d

Viral Strain	CC_{50}	IC_{50}	SI (CC_{50}/IC_{50})
HSV-1	624.8 36 ± 1.2 nM	18.34 nM ± 1.41	34.35 ± 2.64

Figure 48: Cytotoxicity and anti-HSV-1 activity of c-exNDI. (**a**) Vero cells were treated with serial dilution of c-exNDI (15.6 nM- 4 μM) to calculate the CC_{50} of the compound on this cell line. (**b**) Vero cells treated with the concentrations of c-exNDI (0.7 nM- 100nM) used in the antiviral assay. Cytotoxicity was evaluated by MTT assay. (**c**) Plaque assay: Vero cells were infected with HSV-1 strain F (MOI 1, as previously reported [87] and treated with increasing concentrations of c-exNDI (0.7 nM- 100 nM). Supernatants were collected 24 h.p.i. and the number of plaque forming units was determined. (**d**) Recapitulative table of c-exNDI IC_{50}, CC_{50} and SI calculated in this work. Adapted from [243].

To confirm these data, a recombinant HSV-1 expressing GFP fused to the viral protein VP16 (HSV-1 [V41]) was used to infect cells; this mutant virus is characterized by normal replication kinetics and yields [246]. Analysis was performed by flow cytometry. Cells were treated with c-exNDI and Acyclovir (ACV) as the antiviral drug of choice for the treatment of HSV-1 infections with a known mechanism of action [247]. Both compounds were used at low/non-toxic concentration (i.e 100 nM c-exNDi and 3 μM ACV) corresponding to 5 times their IC_{50} values [248].

Cells infected with HSV-1 [V41] and treated with c-exNDI or ACV at 24 h.p.i were monitored for their GFP fluorescence. C-exNDI (**Figure 49a,b**) induced an almost complete reduction of GFP fluorescence (8.6%, expressed as mean of GFP fluorescence), whereas 26% of cells remained fluorescent upon treatment with ACV (**Figure 49a,b**).

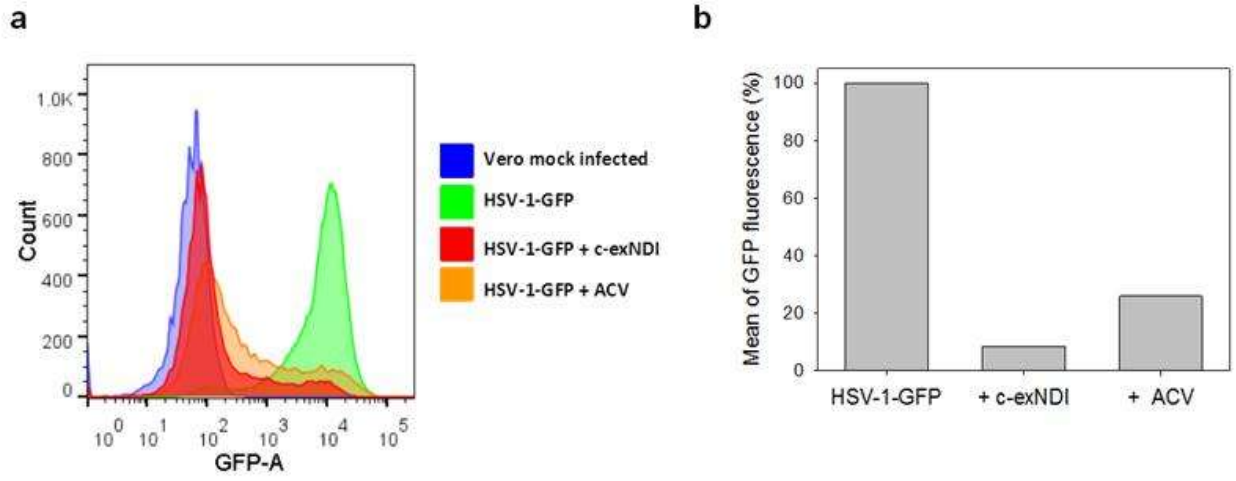


Figure 49: Flow-cytometry of HSV-1 [V41]-infected cells treated with c-exNDI and ACV. HSV-1 [V-41]- infected cells were used as positive control and their GFP fluorescence set to 100%. (**a**) C-exNDI (red curve) and ACV (orange curve)-treated infected cells were compared to HSV-1 [V41]-infected cells (green curve) to monitor compound inhibition of GFP fluorescence. (**b**) Quantification of the mean of GFP fluorescence upon treatment of HSV-1 [V41]-infected cells with c-exNDI and ACV observed in (**a**). Adapted from [243].

To assess the main and temporally last viral step targeted by c-exNDI, we performed a time of addition (TOA) assay where the maximal low/mild-cytotoxic concentration of the compound was added at different times post-infection, corresponding to different viral cycle steps. Also in TOA ACV was used as reference drug. Both compounds (c-exNDI and ACV) were added to infected cells every 2 hours from 0 up to 12 h.p.i and supernatants were collected at 30 h.p.i.

C-exNDI maintained its inhibitory activity when added up to 8 h.p.i (**Figure 50**). This is the time when viral replication occurs; indeed, ACV, a known inhibitor of the viral DNA polymerase [192], showed a remarkably overlapping TOA profile. These data indicate that c-exNDI inhibits viral production mainly by targeting viral DNA replication.

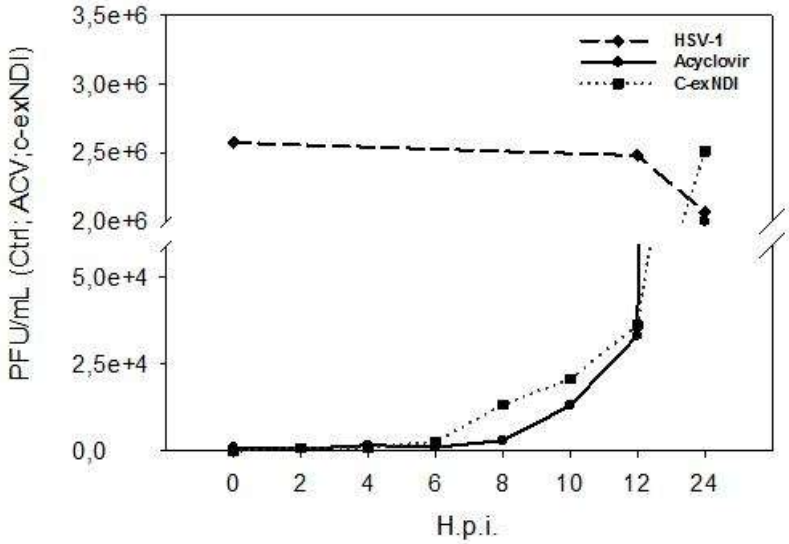


Figure 50: Effect of c-exNDI on HSV-1 cycle steps evaluated by time of addition assay (TOA). C-exNDI was tested at 100 nM (dashed line/squares), Acyclovir (ACV) was used as a reference drug and tested at 3 μM (solid line/circles). Data from infected cells treated in the same conditions but without the compounds are reported as dished line/triangles. The left and right Y-axis refer to ACV and c-exNDI data, respectively. Adapted from [243].

Inhibition of viral DNA replication was further confirmed by qPCR analysis. At 4 h.p.i no reduction in viral DNA replication and production was observed, while at 24 h.p.i. viral DNA was reduced of about 40% (**Figure 51**), similarly to what observed with another G4-ligand (B-19) targeting HSV-1 DNA [87].

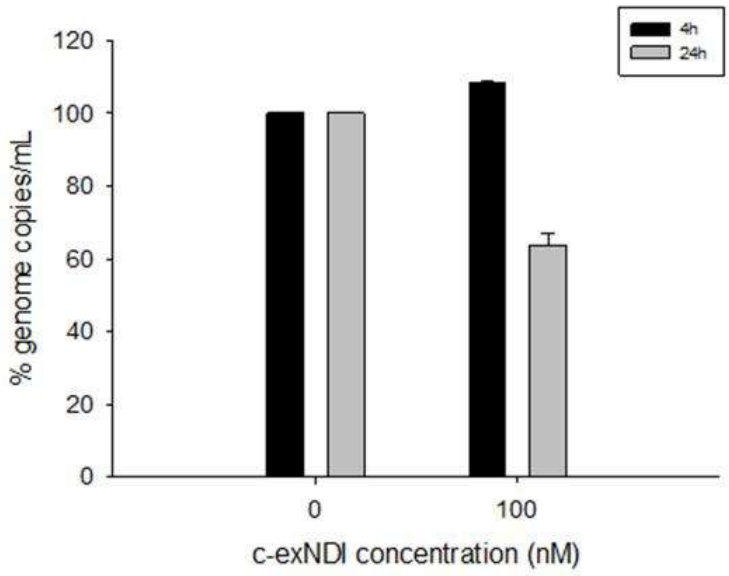


Figure 51: Effect of c-exNDI on HSV-1 DNA replication. Infected cells were treated with c-exNDI (100 nM). Four and 24 h.p.i total DNA was isolated and quantified by qPCR (US1 gene). Each sample was analyzed in duplicate. Adapted from [243].

To further investigate the exact mechanism of action of c-exNDI on HSV-1 and to prove the inhibition of viral DNA replication hypothesized with TOA, mRNA levels of representative viral genes were measured at 4 and 24 h.p.i. upon treatment with c-exNDI. In particular, genes corresponding to immediate-early (IE: ICP22, ICP47), early (E: UL30) and late (L: UL36) proteins were considered.

At 4 h.p.i. no significant inhibition of viral mRNAs production was observed, whereas at 24 h.p.i all mRNAs were reduced to a similar extent (28-45%) (**Figure 52**).

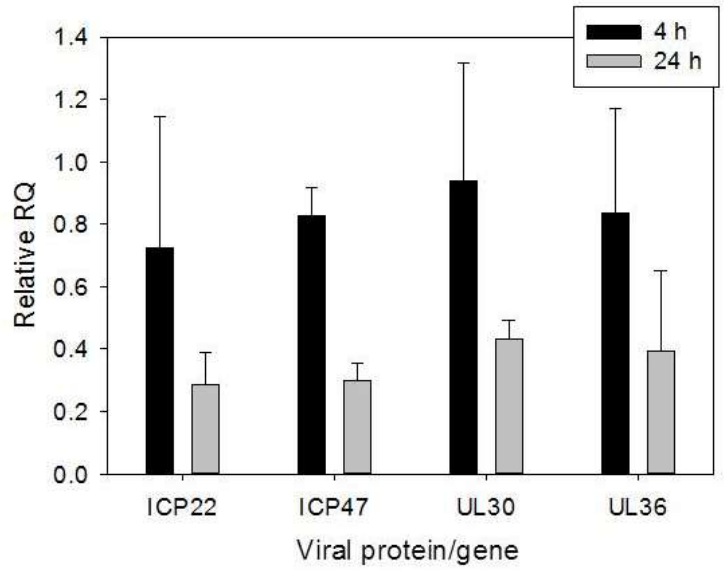


Figure 52: Antiviral effect of c-exNDI on mRNA levels of immediate- early (IE), early (E) and late (L) proteins of HSV-1. Infected cells were treated with c-exNDI (100 nM); 4 or 24 h.p.i total RNA was isolated, retrotranscribed into cDNA and expression of specific genes was determined by RT-PCR. RQ are Relative Quantities. Each gene was analyzed in duplicate. In all data sets: $n \geq 2$, mean ± s.d., student's t-test, $p \leq 0.01$. Adapted from [243].

A similar trend has been reported for ACV [249]. These results strongly corroborate our hypothesis of activity of c-exNDI on viral DNA replication. By RT-PCR we therefore confirmed that treatment with c-exNDI, at the time in which viral DNA replication occurs, has the potential to block, as a cascade of events, viral genes expression, while on the contrary this effect is absent when we quantify the activity of c-exNDI on viral genes when cells are treated before the time of viral DNA replication.

4.2 HSV-1 STUDY: Screening of an in-house library of compounds

4.2.1 Compounds of the in-house library of compounds highly stabilize HSV-1 G4s

Since the majority of G4-ligands, and additionally c-exNDI previously proposed against HSV-1, are characterized by a large planar surface which makes them not easily druggable, we decided to focus our attention on the study of an in-house library of molecules characterized by drug like characteristics.

We applied the same approach used for c-exNDI to study the ability of these compounds to bind and stabilize HSV-1 G4s *in vitro*.

GQC-05 (NSC338258), an ellipticine analogue, was previously reported for its high affinity and selectivity toward *c-myc* G4 [62]. In cells it induced cytotoxicity with a corresponding decrease of *c-myc* correlated mRNAs and altered the binding of DNA binding proteins to the (NHE) III_1 region in agreement with the stabilization of the G4 formed in this region [62].

QuindolineI, a derivative of the natural cryptolepine, as observed for GQC-05 compound, was previously reported for its ability to selectively bind *c-myc* G4 and for its ability to lower *c-myc* transcription in hepatocellular carcinoma cell lines H2p. Additionally, QuindolineI was the first G4-ligand to be resolved by NMR spectroscopy in complex with a G4 structure (*c-myc*), and NMR spectra showed a mode of binding of the compound to the G4 which was radically different from binding models previously proposed: the drug induced a reorientation of the flanking sequences at both ends of the DNA sequence and this observation were the proof of concept of the importance of the drug shape as well as G4 flanking bases in determining drug-binding specificity [99].

Quindoline disubstituted derivatives, such as GSA-0820, were previously described for their good-cell free profile in the stabilization of G4s. In particular GSA-0820 significantly down-regulated c-MYC mRNA on *c-myc* promoter with a G4-unreleated mechanism [100]; on the contrary GSA-0820 was reported for its stabilizing activity on a competing G4 formed on SMN2 Exon 7 after splicing mechanisms [101]. In this part of the work, HSV-1 G4 sequences were analyzed by CD spectroscopy in the absence or presence of 100 mM and 2.5 mM K^+ and in the absence or presence of library compounds. In this case analyses in 2.5 mM K^+ were done to overcome the high stability of the *un2* sequence G4.

In the presence of 100 mM K^+, *un3* sequence showed its well known parallel topology and the pronounced stability previously described. Upon addition of each compound of the library, *un3* melting temperature (T_m) was subjected to a mild increase: GQC-05 was the compound that

induced the most pronounced increase with a ΔT_m of about 7.4°C and QuindolineI induced an increase of 3°C (**Figure 53a**).

a

K^+	100 mM		2.5 mM	
sample	T_m	ΔT_m	T_m	ΔT_m
un3	65.01		37.6±0.4	
+GQC-05	72.41	7.4	64.5±0.3	27±0.02
+GSA-0932	66.40	1.39	56.4±1.2	18.8±0.8
+GSA-0820	65.46	0.45	52.2±0.4	14
+QuindolineI	68	3	57.5±0.6	20±1.06

b

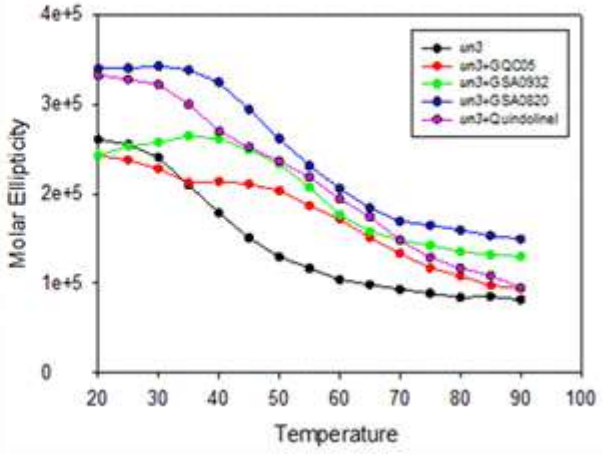
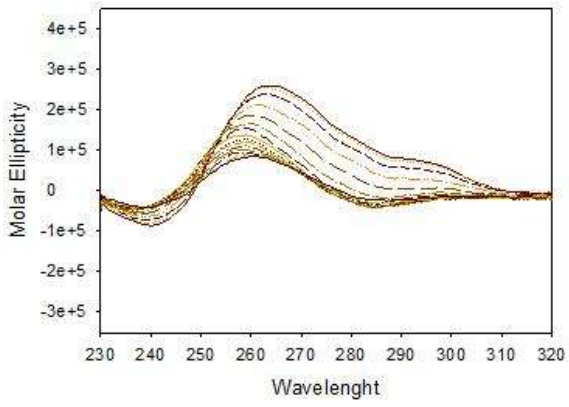


Figure 53: (**a**) Melting temperatures (T_m, °C) of HSV-1 *un3* G4 folding sequences measured by CD in the absence or presence (100 mM, 2.5 mM) of K^+. *Un3* sequence (4 μM) was analyzed in the absence or presence of the library of compounds (16 μM) respectively. (**b**) Melting curves calculated accordingly to Van't Hoff Equation applied for a two-state transition at 264 nm.

These data were confirmed, with a more pronounced increase in the T_m calculated, when *un3* sequence was analyzed in the presence of 2.5 mM K^+ (**Figure 53a,b**). In this condition, where the sequence itself is less stable but displays its parallel topology (**Figure 54a**), an increase in the T_m of 27 ± 0.02 and 20 ± 1.06 was registered upon addition of GQC-05 and QuindolineI, respectively. No significant changes in molar ellipticity were notable upon addition of all library compounds at both 100 mM and 2.5 mM K^+, except for the increase in the negative peak at 240 nm registered upon addition of compounds GQC-05 (**Figure 54b**) and QuindolineI (**Figure 54e**) and for the increase in the positive peak at 290 nm upon addition of GSA-0932 (**Figure 54c**) and GSA-0820 (**Figure 54d**) in the presence of 2.5 mM K^+, suggesting a stabilization of other alternative conformations in solution.

a



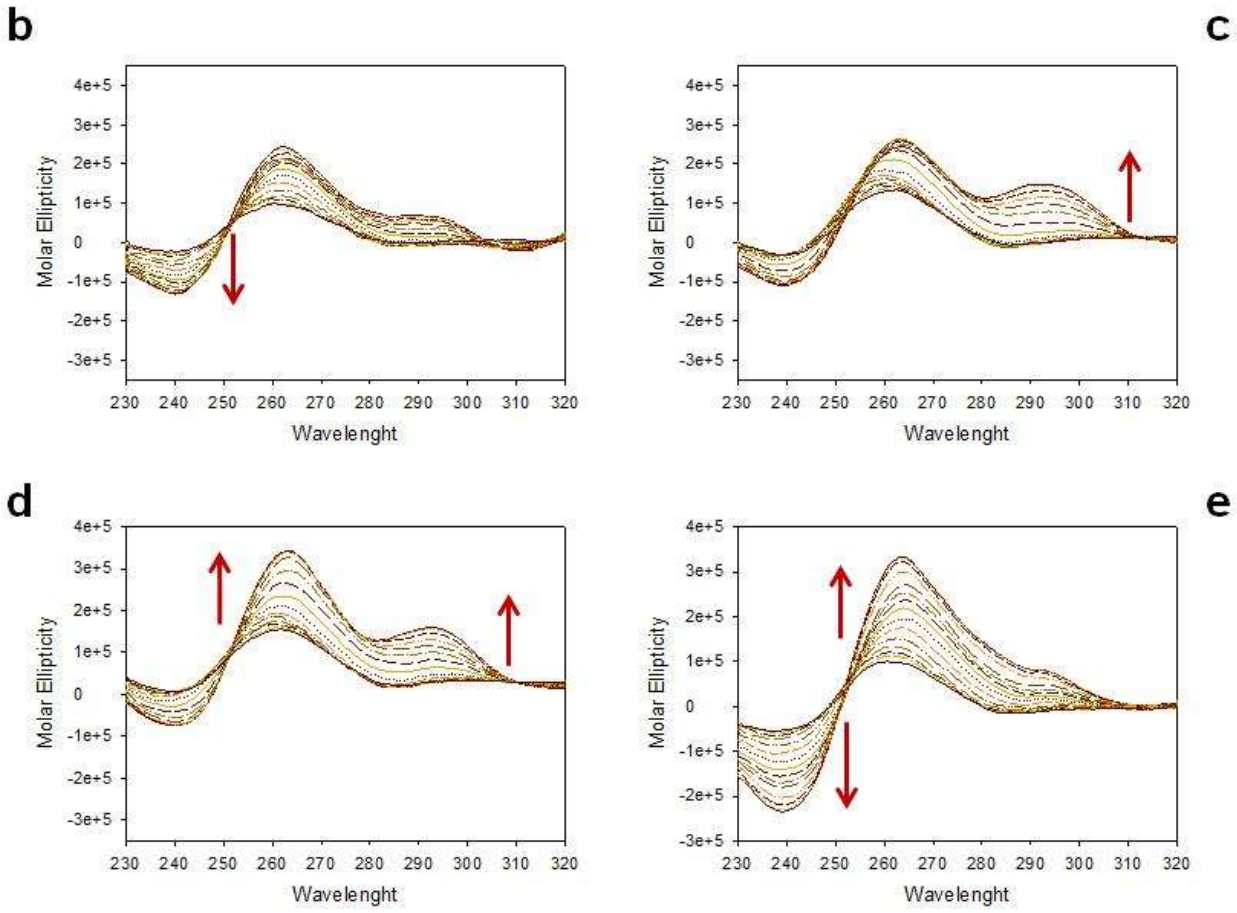


Figure 54: Thermal unfolding experiments on the *un3* parallel sequence in 2.5 mM K^+ (**a**) in the absence of compounds and in the presence of 16 μM (**b**) GQC-05, (**c**) GSA-0932, (**d**) GSA-0820 and (**e**) QuindolineI. Thermal unfolding analyses were recorded over a temperature range of 20°C- 90°C with temperature increase of 5°C. Changes in molar ellipticity and in the conformation are indicated with red arrows.

Un2 sequence, which displays a strong stability (above 90°C at 100 mM K^+), was directly analyzed at 2.5 mM K^+ to better evaluate the effect of each compound. In this condition, the sequence displayed (i) principally its well known antiparallel conformation (**Figure 56a**), (ii) a minor population with hybrid-mixed conformation (due to the lower stability of the sequence with low concentrations of K^+) and (iii) a pronounced stability even without compounds (T_m of 82.4 ± 0.32°C) (**Figure 55a,b**). Upon addition of each compound of the library, the *un2* T_m was increased: in particular, the major increase in the ΔT_m was registered upon addition of GSA-0932 in solution where the T_m increased above 90°C. GQC-05 and GSA-0820 induced a comparable variation in the T_m (ΔT_m of 4.5 ± 0.36°C and 4.25 ± 1.13°C respectively) while QuindolineI did not significantly stabilize the sequence.

a

K^+	100 mM		2.5 mM	
sample	T_m	ΔT_m	T_m	ΔT_m
un2	>90		82.4±0.32	
+GQC-05	n.d.	n.d.	87±0.03	4.5±0.36
+GSA-0932	n.d.	n.d.	>90	>7.5
+GSA-0820	n.d.	n.d.	86.7±0.9	4.25±1.13
+QuindolineI	n.d.	n.d.	83.4±0.1	0.95±0.4

b

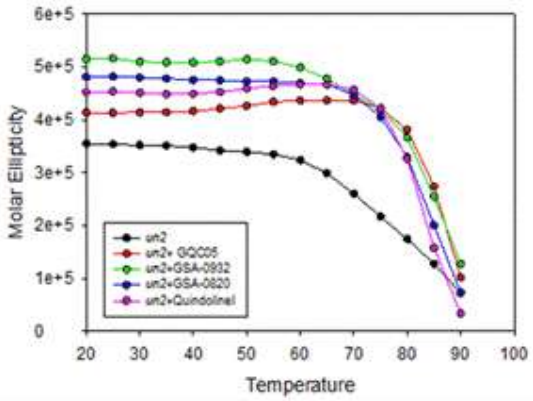
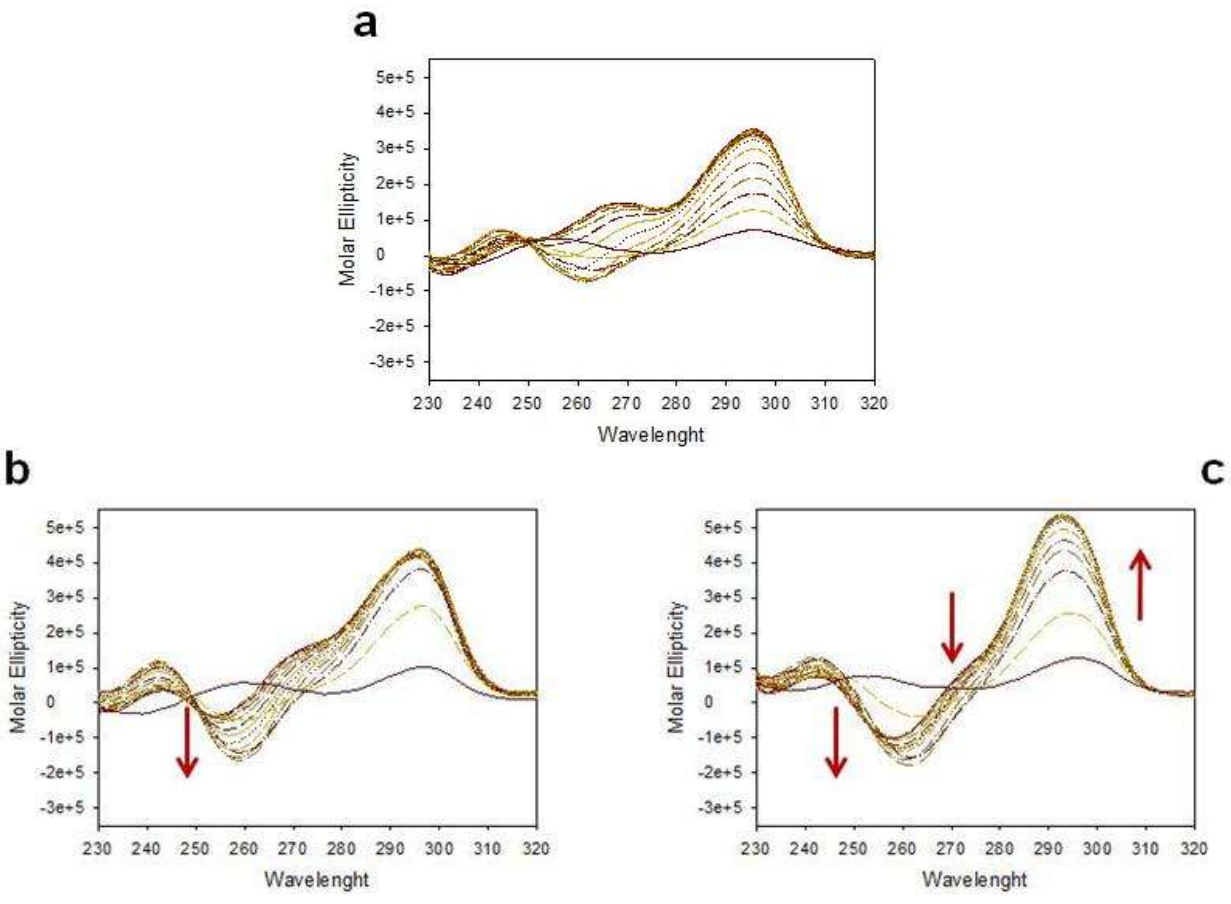


Figure 55: (**a**) Melting temperatures (T_m, °C) of HSV-1 *un2* G4 folding sequences measured by CD in the absence or presence (100 mM, 2.5 mM) of K^+. *Un2* sequence (4 μM) was analyzed in the absence or presence of the library of compounds (16 μM) respectively. (**b**) Melting curves calculated accordingly to Van't Hoff Equation applied for a two-state transition at 295.6 nm.

A significant increase in the molar ellipticity was obtained upon addition of all library compounds except GQC-05: GSA-0932, GSA-0820 and QuindolineI induced an increase in the molar ellipticity both in the positive peak al 290 nm and in the negative peak at 240 nm (**Figure 56c,d,e**), while GQC-05 induced only an increase in the negative peak (**Figure 56b**). Additionally, all compounds except GQC-05 induced stabilization of the antiparallel conformation (reduction in the "shoulder" corresponding to the hybrid-mixed population of G4s).



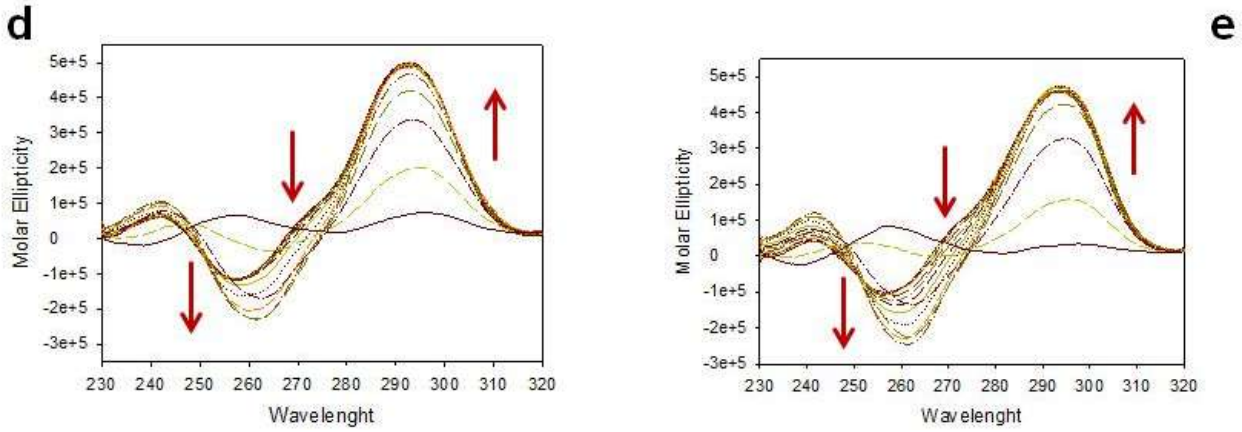


Figure 56: Thermal unfolding experiments on the *un2* antiparallel sequence in 2.5 mM K^+ (**a**) in the absence of compounds and in the presence of 16 μM (**b**) GQC-05, (**c**) GSA-0932, (**d**) GSA-0820 and (**e**) QuindolineI. Thermal unfolding analyses were recorded over a temperature range of 20°C- 90°C with temperature increase of 5°C. Changes in molar ellipticity and in the conformation are indicated with red arrows.

Gp054a was analyzed both in the presence of 100 mM and 2.5 mM K^+. In both conditions, the sequence in the absence of the library of compounds displayed its typical mixed-type conformation with two positive peaks at both 290 nm and at 260 nm (**Figure 58a**). In the presence of 100 mM K^+ all compounds of the library induced a similar stabilization with ΔT_m of 17-18°C and an additional stabilization of the sequence above 90°C (**Figure 57a**).

Moreover, *gp054a* analyzed in the presence of 2.5 mM K^+, showed a pronounced stabilization reached upon addition of GSA-0932 and QuindolineI: ΔT_m of 15.6 ± 1.1 and 22.2 ± 0.6, respectively (**Figure 57a,b**).

a

K^+	100 mM		2.5 mM	
sample	T_m (°C)	ΔT_m (°C)	T_m (°C)	ΔT_m (°C)
gp054a	57.3/ 82.4		58.9/ >90	
+ GQC-05	75.1/ >90	17.8/ >7.5	63	4.1±0.2
+ GSA-0932	74.7/ >90	17.4/ >7.5	74.5/ > 90	15.6±1.1
+ GSA-0820	74/ >90	16.7/ >7.5	72.8/ >90	14±0.4
+ QuindolineI	75.1/ >90	17.8/ >7.5	81/ > 90	22.2±0.6

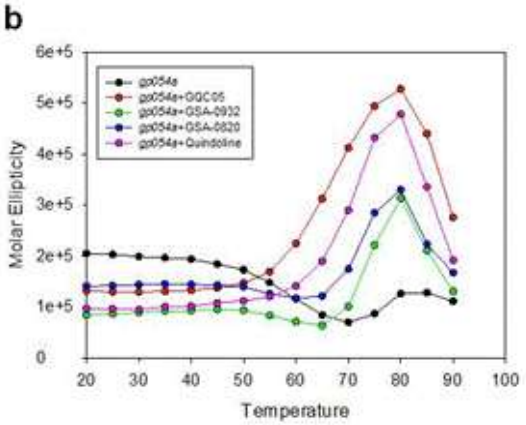


Figure 57: (**a**) Melting temperatures (T_m, °C) of HSV-1 *gp054a* G4 folding sequences measured by CD in the absence and presence (100 mM, 2.5 mM) of K^+. *Gp054a* sequence (4 μM) was analyzed in the absence or presence of the library of compounds (16 μM) respectively. (**b**) Melting curves calculated accordingly to Van't Hoff Equation applied for a two-state transition at 264.2nm.

Upon addition of all the library compounds, both in the presence of 100 mM and 2.5 mM K^+, a slight change in the *gp054a* conformation was registered at 290 nm at high temperatures (**Figure**

58b,c,d,e): this is in accord to the stabilization of an alternative population of G4 with a parallel conformation already seen upon addition of c-exNDI [243].

The major increase in molar ellipticity corresponding to the stabilization of the parallel population of G4s in solution was measured upond addition of GQC-05 and QuindolineI and was accompained by decrease in molar ellipticity at 290 nm and at 240 nm (**Figure 58b,e**). Addition of GSA-0932 and GSA-0820 induced a less pronounced stabilization of the alternative parallel conformation (**Figure 58c,d**).

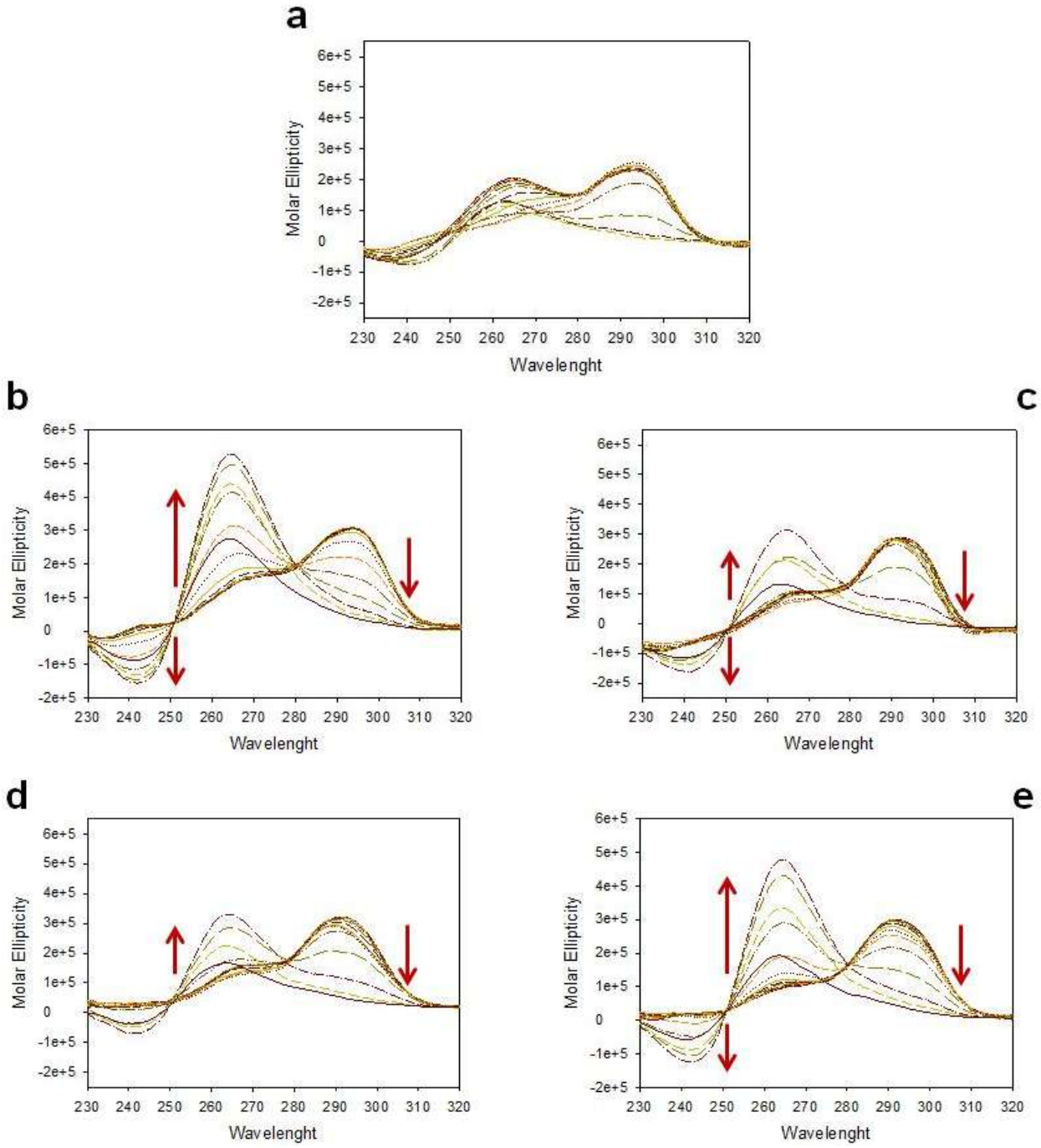
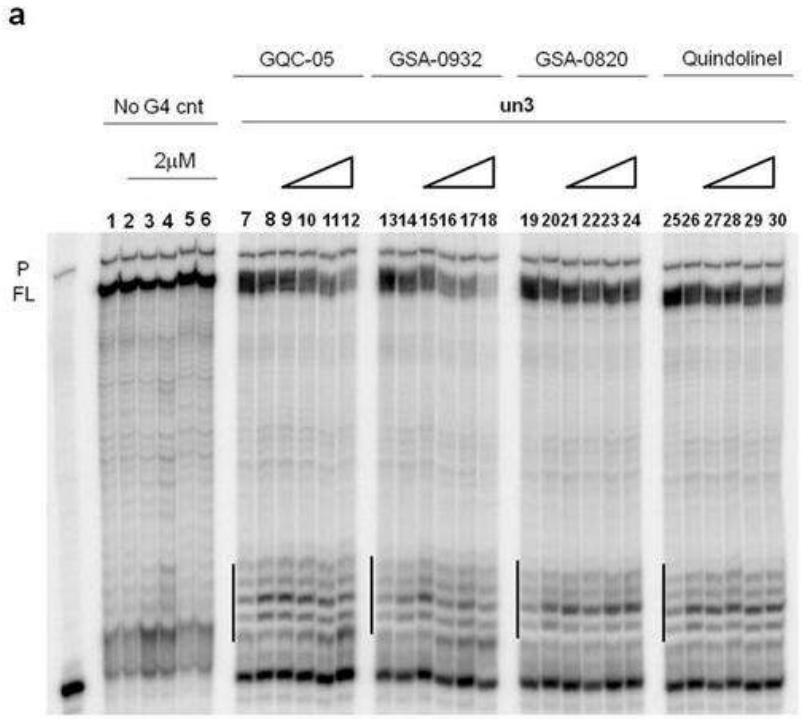


Figure 58: Thermal unfolding experiments on the *gp054a* mixed-type sequence in 2.5 mM K^+ (**a**) in the absence of compounds and in the presence of 16 μM (**b**) GQC-05, (**c**) GSA-0932, (**d**) GSA-0820 and (**e**) QuindolineI. Thermal unfolding analyses were recorded over a temperature range of 20°C- 90°C with temperature increase of 5°C. Changes in molar ellipticity and in the conformation are indicated with red arrows.

Further evidences of the binding and stabilization of the in-house library of compounds on HSV-1 G4s were provided by *Taq* polymerase stop assay. Each HSV-1 G4 forming sequence (*un3*, *un2* and *gp054a* templates) were incubated in the absence or in the presence of 2.5 mM K^+ and with increasing concentrations (250 nM- 2 μM) of each G4-ligand of the screened library. A control sequence unable to fold into G4 structure was used as negative control for G4 formation and was treated with the highest tested dose of each compound.

Full-length products were quantified both in the absence and presence of K^+ and compounds.

Un3 sequence was elongated at 42°C both in the absence and presence of 2.5 mM K^+. In the absence of K^+, the bands corresponding to G4 formation appeared very low while on the contrary the full-length band was very intense therefore suggesting that the polymerase was able to elongate the full template. Upon addition of K^+ a more pronounced G-stop band appeared at the first 3' G4-tract involved in the tetrads. Addition of each compound of the in-house library induced an increase in the intensity of the G4 stop bands (**Figure 59a, lanes 9-12, 15-18, 21-24, 27-30**). This was also confirmed by the decrease in the bands intensity of the full-length template. Quantification of G4 stop bands showed a pronounced decrease in the full-length *un3* sequence amplification (**Figure 59b**) and a pronounced stabilization of the *un3* sequence (~ 20-30% of additional G4 formation) upon addition of GQC-05 and GSA-0932, respectively (**Figure 59c**).



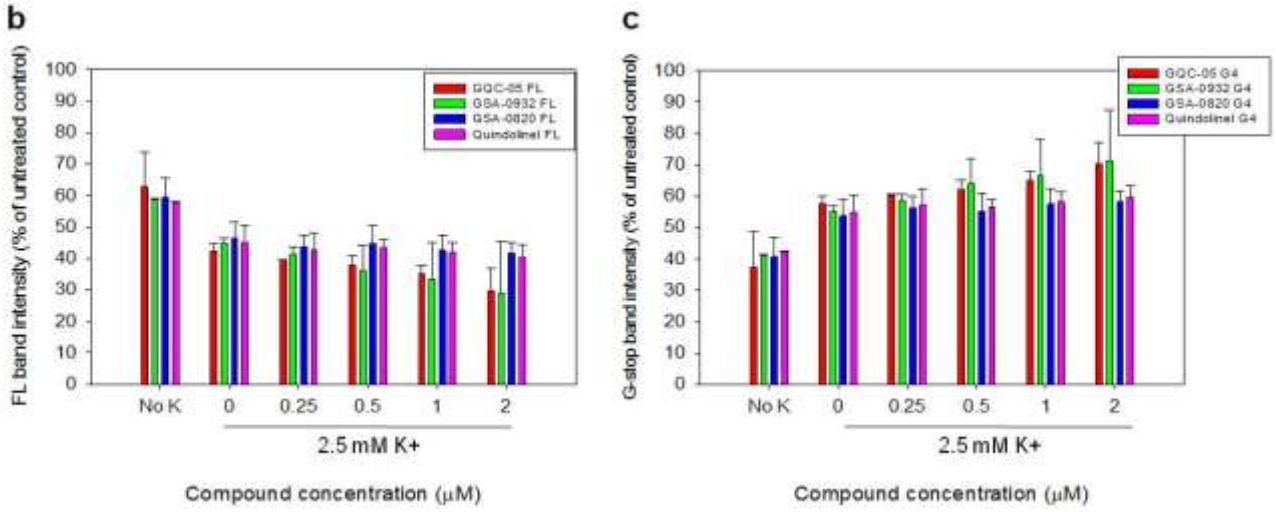
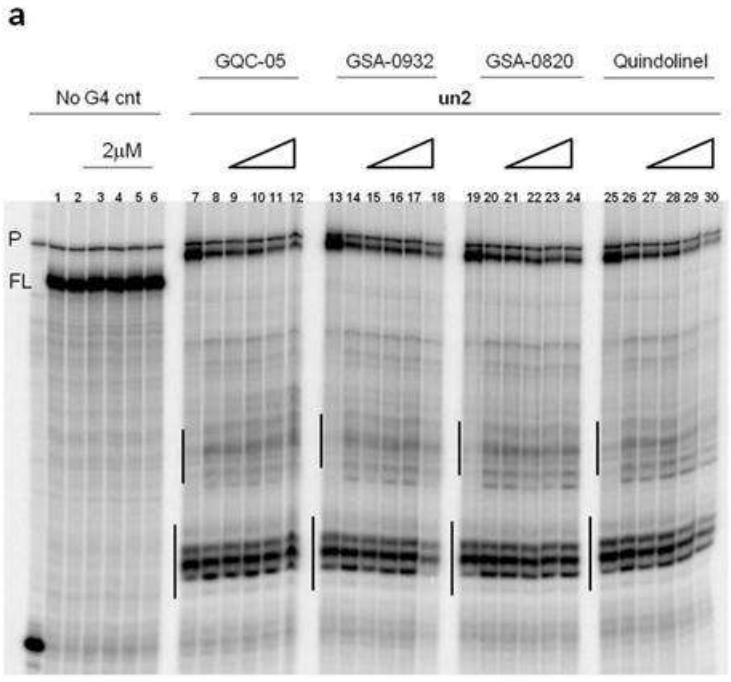


Figure 59: (**a**) *Taq* polymerase stop assay on HSV-1 *un3* G4-folding sequence in the presence of the in-house library of compounds. (**a**) *un3* was analyzed in the absence (lane 7-13-19-25) or presence of 2.5 mM K^+ (lanes 8-12, 14-18, 20-24, 26-30). GQC-05, GSA-0932, GSA-0820 and Quindolinel were used at concentrations 0.2, 0.5, 1, 2 μM (lanes 9-12, 15-18, 21-24, 27-30). Elongation was performed at 42°C. A non-G4 folding sequence was used as negative control (no G4 cnt). G4-related stops are indicated by vertical bars. FL stands for "full- length" amplified template. P indicates the primer lane. (**b-c**) Quantification of the compound-induced reduction in full-length synthesis and in the compound-induced stop sites observed in (**a**).

Concerning HSV-1 *un2* sequence, a premature stop at the first 3′ G-tract involved in the G4 was seen already in the absence of K^+ (**Figure 60a, lanes 7,13,19,25**), accordingly to its pronounced stability already measured by CD in these conditions. However, a slight stabilization was calculated upon addition of each compound of the library **(Figure 60a**, **lanes 9-12, 15-18, 21-24, 27-30).** Interestingly, the stabilization of the *un2* G4 forming sequence by each G4-ligand (**Figure 60c**) was accompanied by a conspicuous reduction in the intensity of the bands corresponding to full-length amplicon (**Figure 60b**), therefore confirming that the presence of a G4 structure directly inhibits polymerase activity.



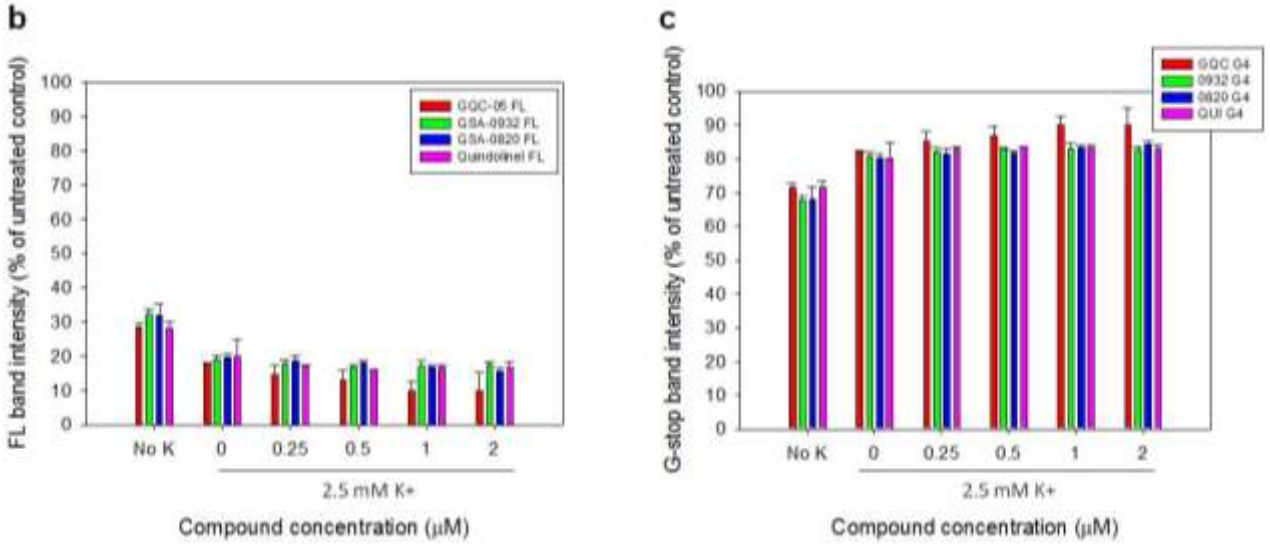


Figure 60: *Taq* polymerase stop assay on HSV-1 *un2* G4-folding sequences in the presence of the in-house library of compounds. (**a**) *un2* was analyzed in the absence (lane 7-13-19-25) or presence of 2.5 mM K^+ (lanes 8-12, 14-18, 20-24, 26-30). GQC-05, GSA-0932, GSA-0820 and QuindolineI were used at concentrations 0.2, 0.5, 1, 2 μM (lanes 9-12, 15-18, 21-24, 27-30). Elongation was performed at 60°C. A non- G4 folding sequence was used as negative control (no G4 cnt). G4-related stops are indicated by vertical bars. FL stands for "full- length" amplified template. P indicates the primer lane. (**b-c**) Quantification of the compound-induced reduction in full-length synthesis and in the compound-induced stop sites observed in (**a**).

Gp054a G4 forming sequence was incubated in the presence of 2.5 mM K^+ and increasing concentrations of each compound of the library as already described (0.2, 0.5, 1, 2 μM) (**Figure 61a**). *Gp054a* was elongated at 60°C and already in the absence of K^+ displayed a 50% of the sequence folded in a G4 structure (**Figure 61c**).

Addition of GSA-0932, GSA-0820 and QuindolineI induced a poor stabilization of the G4 structure (**Figure 61a, lanes 15-18, 21-24, 27-30**) with only a 5% of additional G4 stabilization (**Figure 61c**) while, addition of GQC-05 induced a more pronounced effect (**Figure 61a, lanes 9-12**) with ~ 25% of additional G4 formation (**Figure 61c**). Moreover, treatment with GQC-05 induced a more visible effect on the decrease of signal intensity of bands corresponding to full-length formation (**Figure 61b**). These observations are partially in agreement with what calculated by CD on *gp054a* sequence: GQC-05, as observed in T_m curves, was the compound that induced the major increase in molar ellipticity (**Figure 57b**).

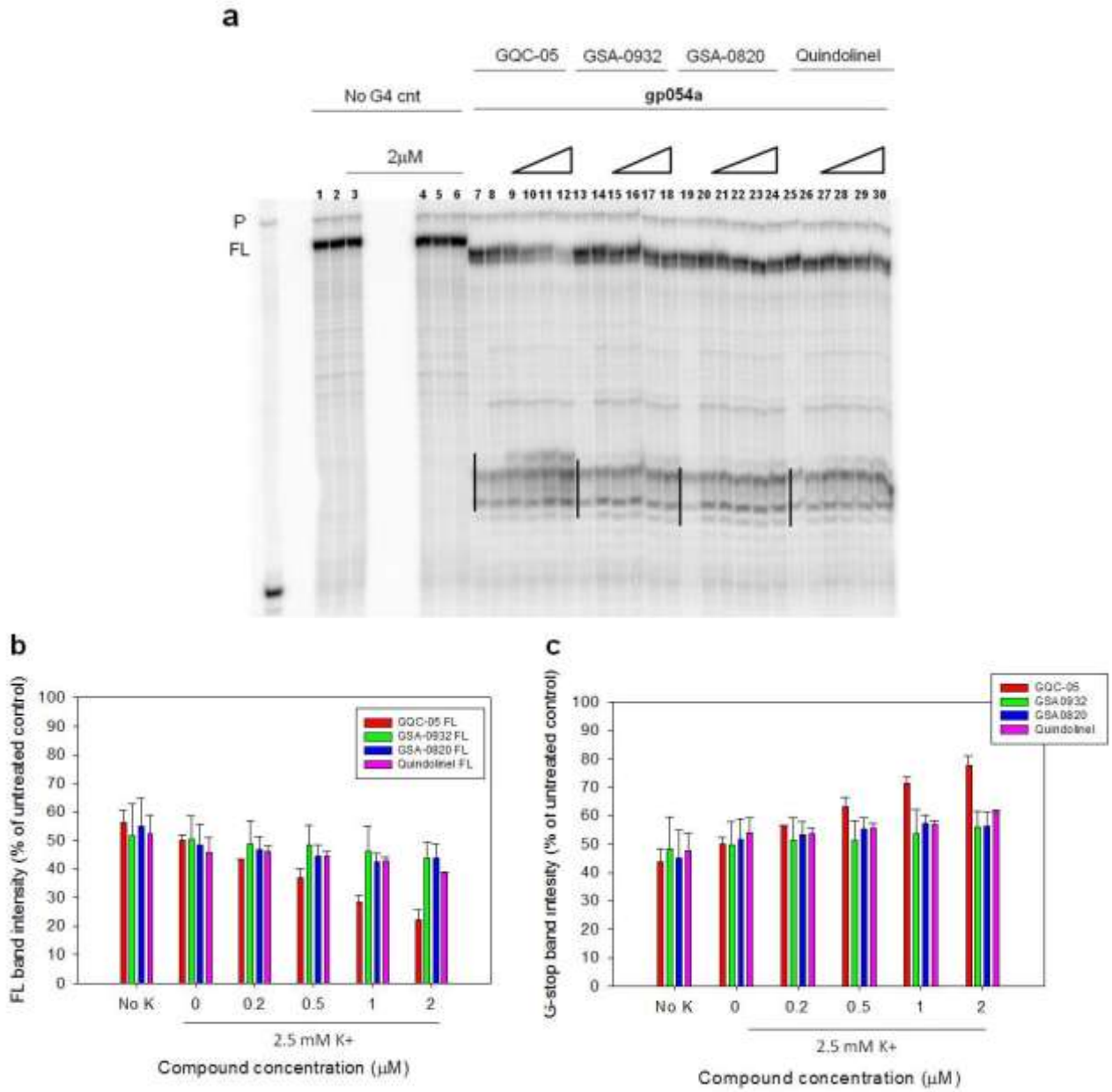


Figure 61: *Taq* polymerase stop assay on HSV-1 *gp054a* G4-folding sequences in the presence of the in-house library of compounds. (**a**) *gp054a* was analyzed in the absence (lane 7-13-19-25) or presence of 2.5 mM K^+ (lanes 8-12, 14-18, 20-24, 26-30). GQC-05, GSA-0932, GSA-0820 and Quindolinel were used at concentrations 0.2, 0.5, 1, 2 μM (lanes 9-12, 15-18, 21-24, 27-30). Elongation was performed at 60°C. A non- G4 folding sequence was used as negative control (no G4 cnt). G4-related stops are indicated by vertical bars. FL stands for "full- length" amplified template. P indicates the primer lane. (**b-c**) Quantification of the compound-induced reduction in full-length synthesis and in the compound-induced stop sites observed in (**a**).

4.2.2 Compounds of the in-house library display anti HSV-1 activity

Given the ability of all the tested compounds to bind and stabilize HSV-1 G4s, we tested the possibility that they displayed antiviral activity against HSV-1.

Cytotoxicity of these compounds was assessed on U2OS cells, a human bone osteosarcoma cell line able to sustain HSV-1 infection (**Figure 62a-d**).

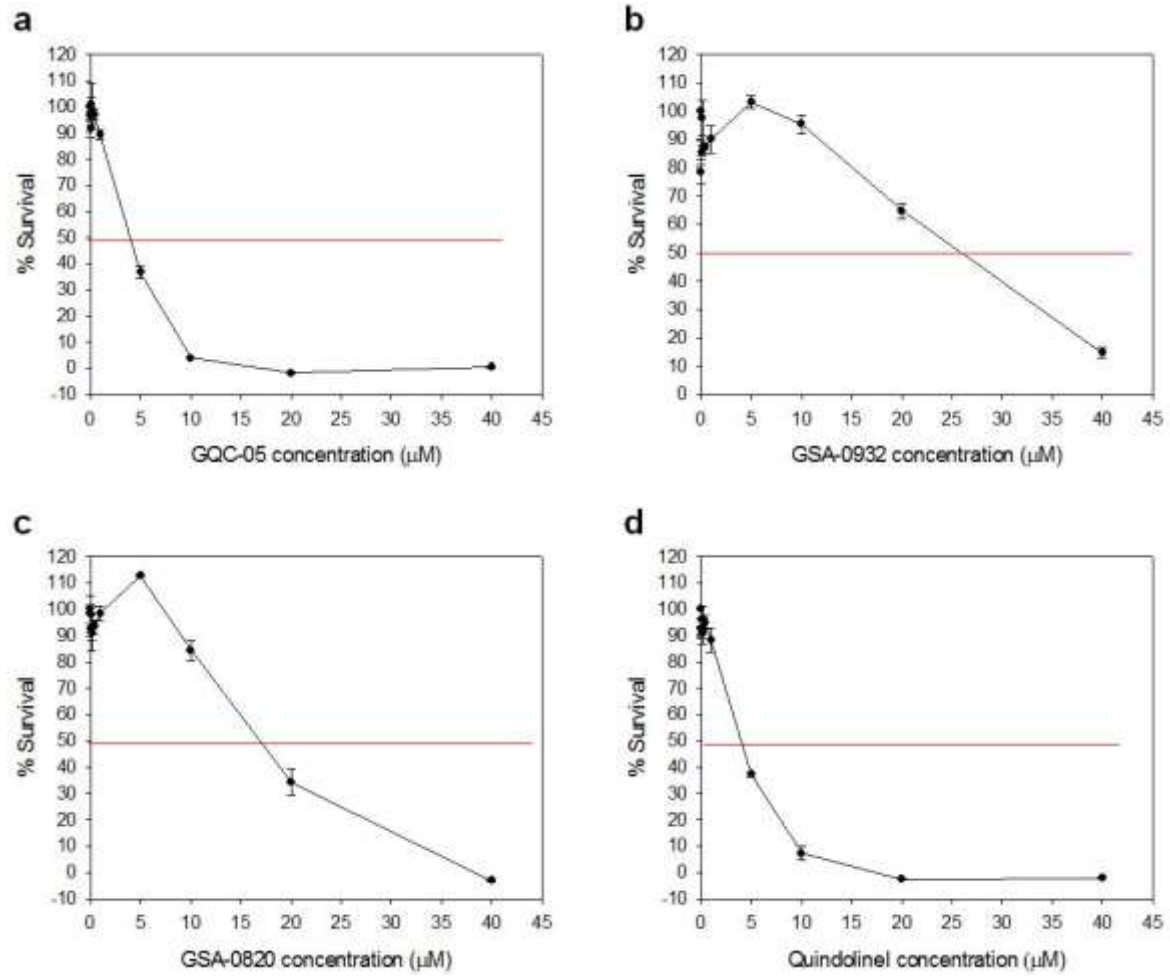


Figure 62: Cytotoxicity of the in-house library of compounds screened in U2OS cells (human bone osteosarcoma). Cytotoxicity was assessed by MTT assay at 24 h. Each compound was analyzed from 40 μM to 10 nM following serial dilutions (1:2 and 1:5). (**a**) GQC-05, (**b**) GSA-0932, (**c**) GSA-0820, (**d**) QuindolineI.

GQC-05 and QuindolineI displayed a CC_{50} in the low micromolar range (CC_{50} 2.9 ± 0.6 and 3.7 ± 0.4 μM, respectively) (**Figure 62a,d**) while GSA-0932 and GSA-0820 displayed an almost 10-fold lower cytotoxicity (CC_{50} 21.8 ± 3.9 and 17 ± 1.2 μM, respectively) (**Figure 62b,c**).

Increasing concentrations of compounds, in the low nanomolar range where they did not display cytotoxicity, were added to U2OS cells infected with HSV-1 strain F (MOI 1), and supernatants were collected at 24 h.p.i. Infectious viral particles in the supernatants were quantified by plaque assay.

As shown in **Figure 63a-d**, all compounds displayed a significant antiviral activity against HSV-1 in the low nanomolar range. GQC-05 and QuindolineI at 400 nM induced a reduction in the HSV-1 production up to ~ 65% with no cytotoxicity (**Figure 63a,d**); GSA-0932 and GSA-0820 at 400 nM induced a reduction in viral production of about 60%, always in the absence of cytotoxicity (**Figure 63b,c**).

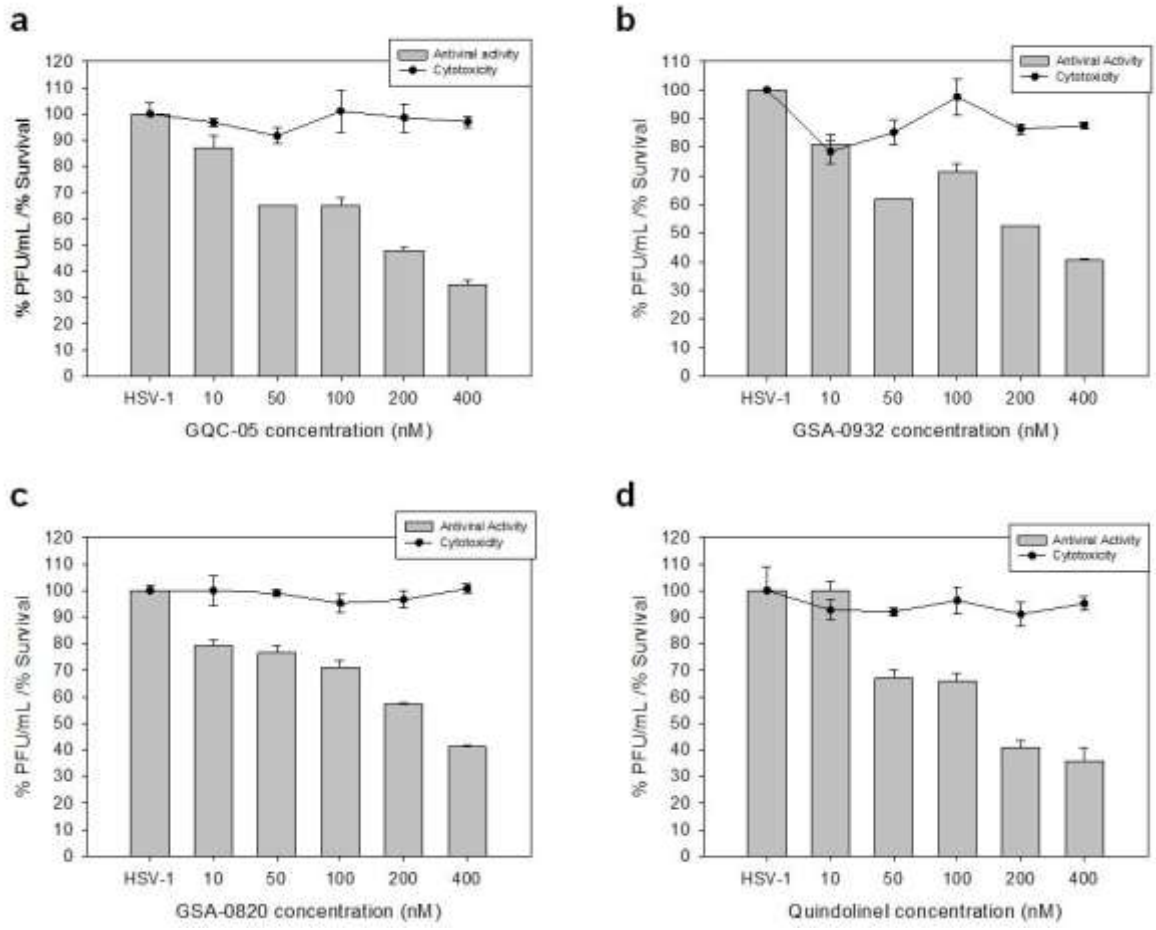


Figure 63: Antiviral activity of the in-house library of compounds on HSV-1 strain F U2OS infected cells. Antiviral activity was measured by plaque assay: U2OS cells were infected with HSV-1 strain F (MOI 1, as previously reported [87,243] and treated with increasing concentrations of each compound (10 nM-400 nM), (**a**) GQC-05, (**b**) GSA-0932, (**c**) GSA-0820 and (**d**) QuindolineI. Supernatants were collected 24 h.p.i. and the number of plaque forming units was determined.

All compounds displayed IC_{50} values in the low nanomolar range. Selectivity indexes (SI) were also measured and resulted to be promising for GSA-0932 and GSA-0820 due to their lower cell cytotoxicity (~3-4 μM) (**Table 10**).

Compound	IC_{50}	CC_{50}	SI
GQC-05	187.4 ± 5.6	2.9 ± 0.6	**15.5**
GSA-0932	214.85 ± 21	21.8 ± 3.9	**101.5**
GSA-0820	199.15 ± 1.2	17 ± 1.2	**85.4**
QuindolineI	153.9 ± 3.4	3.7 ± 0.4	**24**

Table 10: Summary table of IC_{50}, CC_{50} and SI calculated for each compound of the in-house library screened in U2OS cells infected with HSV-1 strain F.

To investigate the target of the G4-ligand with the most promising SI, i.e. GSA-0932, the TOA experiment was set up [243].

As already described, this experiment determines how long the addition of a compound can be postponed in the viral cycle before its antiviral activity is lost. HSV-1 replication is blocked up to a time point corresponding to the occurrence of the last process targeted by the drug and the identification of the last inhibited viral step can be revealed by comparing the antiviral activity of the tested compound to the antiviral activity of the reference drug ACV as previously reported [243].

To this purpose, all compounds of the in-house library were tested on HSV-1 U2OS infected cells at 400 nM, the highest concentration used in antiviral assays, which corresponds to 1.8-2.6-folds their IC_{50}. Compounds activities were compared with that of ACV which was used at 1.6 μM, concentration corresponding to ~2.6-folds its IC_{50}, to compare compounds in the same range of IC_{50}-folds. Infection was carried out at MOI 0.5 and compounds were added every 2 h up to 10 h.p.i; supernatants were collected at 30 h.p.i and then titrated by plaque assay.

Treatment of HSV-1 U2OS infected cells with GSA-0932 kept infection controlled up to 4 h.p.i, while from 6 h.p.i a slight increase in the viral titer was measured and, finally, from 8 to 10 h.p.i an exponential increase in the viral titer was reached.

Treatment of infected cells with ACV showed a perfectly similar increase in viral titer from 8 to 10 h.p.i. ACV, a well-known reference drug, acts by inhibiting viral DNA polymerase during viral DNA replication (that occurs between 8-10 h.p.i): this therefore confirms that GSA-0932 acts in early events of the viral life cycle, including a possible direct inhibition of the replication machinery that causes a pronounced reduction in viral particle production (**Figure 64**).

These results led us to speculate that the replication inhibition mechanism of GSA-0932, as already described also for c-exNDI compound, could also be shared by other compounds of the library.

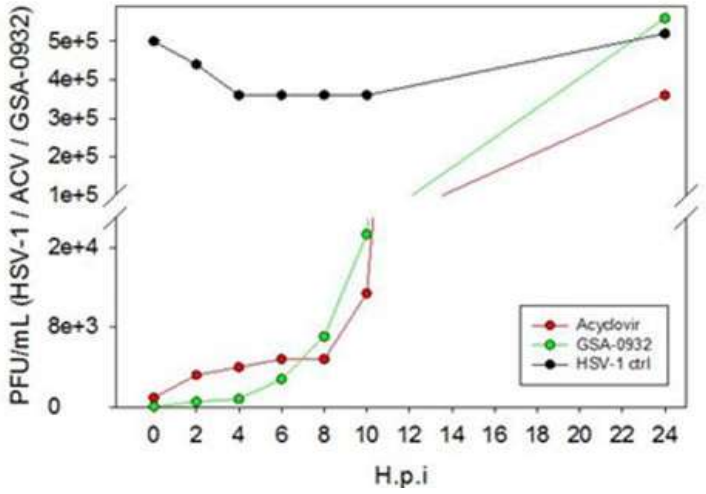


Figure 64: Effect of GSA-0932 on HSV-1 cycle steps evaluated by time of addition assay (TOA). GSA-0932 was tested at 400 nM (green line) which corresponds to 1.8-folds its IC_{50}. Acyclovir (ACV) was used as a reference drug and tested at 1.6 μM (red line) which corresponds to 2.6-folds its IC_{50}. Data from infected cells treated in the same conditions but without the compounds are reported in black line (HSV-1 ctrl).

In this second part of the work we have characterized the G-quadruplex-mediated mechanism of action of in-house library of compounds with drug-like characteristics.

We have demonstrated that all compounds, GQC-05, GSA-0932, GSA-0820 and QuindolineI, (i) potently stabilize G4s in the HSV-1 genome and cause polymerase stalling during DNA elongation therefore confirming their ability to potently stack on G4-tetrads already reported for some of these compounds [102].

Additionally, all compounds displayed potent HSV-1 antiviral activities in the low nanomolar range (concentrations sufficient to reach a reduction in viral production about 60-65%) with low cytotoxicity, resulting in promising selectivity indexes.

We could ascribe the compounds antiviral activities, as previously demonstrated for c-exNDI [243], to the massive presence of G4s in the nucleus during the HSV-1 viral replication which represents the main key factor for the promising use of G4-ligands against HSV-1 infections. During the replication step, G4s control both viral DNA replication but also transcription events and their stabilization by a G4-ligand result in dramatic consequences. Since HSV-1 contains multiple clusters of G4-forming sequences, all G4-ligands possess the theorical possibility to decrease viral production. The GSA-0932 compound targets the viral replication step, strongly suggesting that its anti-HSV-1 activity is G4-mediated.

Since all compounds of the library possess (i) promising chemical features (i.e their small size, smaller than other well known G4-ligands, i.e. B-19 and NDIs) and (ii) potent activities in the low nanomolar range with no cytotoxicity, these characteristics make them suitable for future development as novel anti HSV-1 G4-ligands with more drug-like characteristics than previous developed G4-ligands and as promising compounds for the treatment of ACV-resistant herpetic infections.

4.3 HSV-1 STUDY: G-quadruplex binding protein identification

Starting from our findings on G4-forming sequences described in the previous sections and their importance, we further analyzed their properties to elucidate their potential role within the viral life cycle.

The main interest in performing these analyses was the identification of DNA interacting proteins able to bind to G4s and to express a specific activity once in the nuclear compartment.

The first approach used to identify proteins was a pull-down assay in the presence of 5' biotinylated *un3long3* G4 sequence (*un3* G4 minimal sequence flanked by the natural sequence present in the genome), a negative control for G4 formation (LTR II+III+IV scrambled) and the nuclear protein extract of U2-OS cells infected with HSV-1. Nuclear proteins were extracted at 8 h.p.i. in order to reach the maximum level in the expression of proteins implicated in the replication.

Preliminarily, oligonucleotide sequences were analyzed by CD for their ability/inability to fold into G4 (data not shown).

From the gel-excised and trypsin digested bands, analyzed by a MS/MS experiment and database searching, we identified both cellular and viral proteins, some of them being already identified as G4 binding proteins (**Table 11**).

Protein	Score	Cellular/Viral
Human Ribonucleoprotein U	19	Cellular
Nucleolin	167	Cellular
Nucleophosmin	63	Cellular
Histone 11	145	Cellular
Histone 12	301	Cellular
Major capsid protein (VP5)	165	Viral
Major viral transcription factor (ICP4)	90	Viral
DNA binding protein (ICP8)	51	Viral
DNA polymerase processivity factor (PAP1)	29	Viral

Table 11: Cellular and viral proteins identified with a pull- down assay coupled with MS. Results obtained by analyzing the binding between U2-OS nuclear proteins extract and HSV-1 *un3long3* G4 folded sequence.

Interestingly both cellular and viral proteins were identified. In particular, concerning cellular proteins, the human ribonuclear protein U with the lowest but significant score was identified. This is in agreement with what previously observed by [19]: ribonucleoproteins are involved in the

replication machinery and in particular they are involved in G4 unfolding required for polymerase processivity at that level. Additionally, nucleolin and nucleophosmin were identified as already observed by [123,125]: nucleolin was already reported to be able to bind G4 structures within the LTR promoter region of HIV-1, therefore expressing the ability to shut down viral transcription mechanisms, while nucleophosmin was reported to be able to bind to *c-myc* promoter, therefore altering its functions. Our result, obtained in the presence of a G4 folding sequence naturally contained within the HSV genome in the promoter region of ICP34.5 gene, clearly confirms that G4s contained in promoter regions can be regulated in their function by the binding of specific protein partners. Finally Histone 11 and 12 were identified and showed high interesting scores. Interaction of G4s with engineered RGGF domains were correlated with histone methylation inhibition [250], but no evidence has ever been observed for a direct binding between G4s and histone proteins. Further pull-down assays will be performed to highlight the significance of this result and the possibility that these binding proteins could have implications for viral DNA maintenance within latency mechanisms. Concerning viral proteins, interestingly we identified (i) two infected cells proteins, ICP4 and ICP8, (ii) a structural protein, the major capsid protein VP5 and (iii) the viral polymerase processivity factor, PAP1.

Regarding ICP4, the major transcriptional regulatory protein of HSV-1, its ability to bind to ssDNA, in particular in consensus sequences present upstream of different viral genes, has been demonstrated by Resnik and co-workers [251]. The fact that ICP4 is localized at the replication and transcription forks [238], where G4s can be implicated in the regulation of replication and transcription, strongly support the relevance of this preliminary interaction observed in our research. In the same way, we consider very interesting visualization of ICP8 and the processivity factor of the viral DNA polymerase (PAP-UL42) at the G4 structure level. The major DNA binding protein ICP8 was already reported by Gao and Knipe [252] to be able to bind to ssDNA and to exert nuclear function at the replication level. We need to further confirm these potential binding but, as for ICP4, the localization of these proteins at the replication compartment, where G4s exert their function, positively confirm that their role can be modulated by protein partners.

The last viral protein identified was the major capsid protein VP5. There are no evidences in literature that show a direct binding of VP5 to viral DNA but it was previously reported that VP19, a VP5 protein partner, is able to bind to DNA in the genome condensation mechanism within virions production [253]. We have to further confirm this binding to evaluate if it is direct or protein-mediated, but it is possible that this binding to G4 can have implications in the process of virions formation.

4.4 MV STUDY: Investigation of G-quadruplex in MV genome

4.4.1 Putative G4-forming sequences are present in MV genome

Since many reports described the presence of G4s in different organisms (i.e bacteria and viruses[254]) and not only in DNA genomes but also in RNA genomes [255][132], and taking into account the strong evidence of a G4-mediated regulatory mechanism both in HSV-1 [87][243] and HIV-1 [143], we decided to investigate the theoretical possibility of the presence of G4 forming sequences within the RNA genome of MV.

We initially investigated the possible presence of G4s within the MV genome by analyzing the NCBI Reference Sequence NC_001498.1 (https://www.ncbi.nlm.nih.gov/nucleotide/) using an algorithm-based software for recognition and mapping of putative Quadruplex forming G-Rich Sequences available on-line (QGRS) (http://bioinformatics.ramapo.edu/QGRS/index.php). The analysis was conducted on both the reference sequence (cRNA with positive polarity 5′- 3′) and its reverse complement sequence, which corresponds to the negative-stranded (3′- 5′) RNA genome of MV.

The search for PQS was limited to G4s with this composition: GGG≥3N0-12GGG≥3N0-12GGG≥3N0-12GGG≥3, where N corresponds to loop with length from 0-12 nt. This initial analysis allowed us to identify the most stable putative G4-forming sequences (PQS); putative sequences with G tracts = GG were additionally analyzed but were excluded from further analyses because of their known less stability.

One sequence that presented highly interesting QGRS score (herein termed MV1423) was found in the reference sequence (cRNA, positive polarity) while other six sequences (referred to as MV4500, MV10603, MV10904, MV10953, MV11114, MV11414) were found in the reverse complement strand (which corresponds to the negative-stranded RNA main genome sequence of MV). MV G4s-putative folding sequences are reported in **Table 12.**

Name	QGRS Position	Length	QGRS Sequence	G-score*
MV1423	**1423 (+)**	30	GGGTCAAACAGGGTCGGGGAGAAGCCAGGG	67
MV4500	**4500 (-)**	36	GGGACATGGGCGTCAAGGTCGGGGGGGCAATGAGGG	66
MV10603	**10603 (-)**	39	GGGGGTCCCCCAGGGATCGCTCTTGGGTGCTGGTTCGGG	72
MV10904 (MV10902)	**10904 (-)**	32	GGGCACCGGGGGGGCCGGCGGGTTTGGTTGGG	71
MV10953	**10953 (-)**	38	GGGGGGTGGGGGCGCGGTGCCGCGGGCCGGGGTCTGGG	67
MV11114	**11114 (-)**	32	GGGGTTCTTGCGTTGGGGGAGGGGGGGAGGGG	101
MV11414	**11414 (-)**	27	GGGGGGTCGTTTTCGGGTATTGCTGGG	64

Table 12: QGRS analysis of Measles genome G4s-putative folding sequences. GGG tracts are represented in red and bold. *G-score indicates the propensity of a sequence to fold into monomolecular G-quadruplex assigned by the QGRS software. GGG tracts are shown in red and bold.

Each identified G4-putative folding sequence was then analyzed for its effective location within the negative stranded RNA genome of MV (**Table 12**). In one case, MV 10904, the sequence was slightly changed from the QGRS sequence output in order to include two additional 3'- flanking Gs omitted from the algorithm that could take part in the G4 formation. This sequence is herein referred to as MV10902. MV sequence 10603, MV10902 and MV10953 are located in the F gene, which encodes for the fusion protein, MV11114 and MV11414 are located in the intergenic region comprised between the M gene (encoding for the matrix protein) and the F gene. MV4500 sequence is located in the coding region of the L gene (encoding for the large polymerase). MV1423, the only sequence identified in the positive-stranded RNA (cRNA with positive polarity) is located in the coding region of N gene (encoding for the nucleocapsid protein). All putative G4-forming sequences identified by QGRS algorithm were characterized by high G-scores, from 61 to 101, strongly suggesting their theorical ability to fold into monomolecular G4 structures (**Table 12**).

4.4.2 Base conservation analysis of putative G4-forming sequences

To establish the importance of the predicted sequences within the viral context, and to check if G4 formation in MV is a conserved feature of all strains, we assessed the degree of base conservation among different MV strains. Fifty-five deposited sequences of MV cRNA were aligned using FASTAALIGN0 in order to generate a complete local alignment of all genomes for the future estimation of the degree of conservation for each identified G4. Conservation analyses were carried out taking into account (i) the conservation of G base composition of each G4 and (ii) the degree of conservation of the G4 with its connecting loops (which are generally less conserved than the G-pattern and confer the specificity of each structure). The degree of conservation of the G4 scaffold was calculated for each sequence and the higher was the score, the stronger was the conservation of each G4 structure. Each identified G4 was plotted as a vertical bar, with height representing the degree of conservation and with the position on the x-axis corresponding to the position in the viral genome.

These data give a clear visualization method to identify (i) the presence, position and conservation of each G4 structure within the genome and (ii) the overall local conservation of the genomic context.

The GGG-pattern of G4 structure previously identified by QGRS was the only pattern analyzed for the MV G4 conservation; GG-patterns were excluded from the analyses because of their less pronounced stability.

As shown in **Figure 65**, G4s identified in the reverse complement strand were generally well conserved (the vast majority of sequences reached a conservation rate of more than 80%) and were

embedded in the most conserved regions of the genome suggesting a key role of this highly conserved viral site.

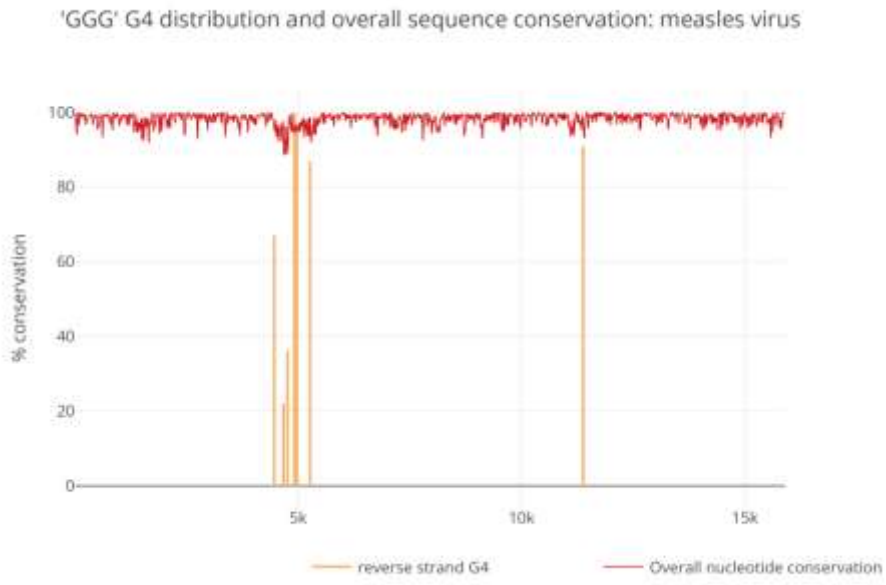


Figure 65: Conservation of MV G4s and viral genome. G4s in the reverse strand (**orange vertical bars**) and with a GGG-pattern in MV genome are shown. The height of bar indicates the degree of conservation of G bases of each G4. The **red broken line** indicates the local sequence conservation of the viral genome. The X-axis indicates the genome position, the y-axis indicates the % of conservation.

Consensus sequences of each MV G4 identified were calculated on the alignment previously described and the LOGO representation was obtained by the WebLogo software (http://weblogo.berkeley.edu/logo.cgi). MV1423 G4 sequence was directly found in the cRNA resulting from the alignment, while all other sequences (located in the reverse complement strand of the cRNA deposited) were obtained by searching for the reverse complement sequence of the QGRS G4 sequences. Consensus sequences and their base conservation, in fact, are reported for the corresponding C-rich sequence of each G4 (**Table 13**).

Base conservation was reported as % of base conservation within aligned genomes. As reported in **Table 13,** the vast majority of G bases (for MV1423) and C bases (for all the other identified sequences) were generally extremely conserved with occurrence ranging between 75 and 100%. The only exception was observed for MV11114 sequence where a lower conservation was measured (~ 60% for the central C-stretch): this indicates that a lower number of MV strains shared that sequence portion. However, since the MV11114 sequence is significantly rich in Gs, it is possible that anyone of the multiple Gs alternatively takes part to G4 formation, therefore lowering the requirement of a high conservation rate of few specific Gs.

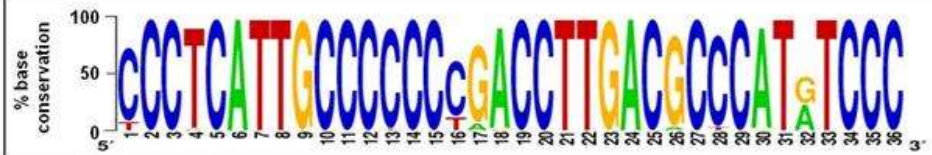
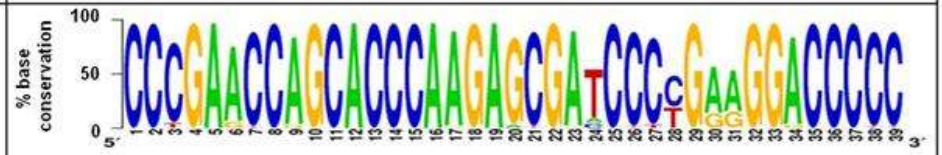
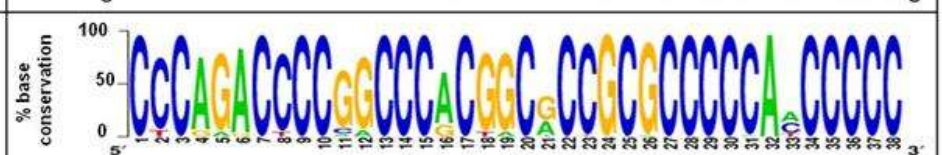
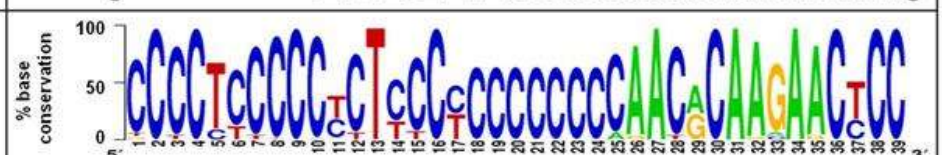
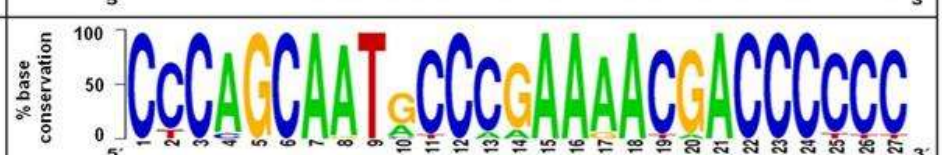
Sequence name	Consensus sequence
MV1423	
MV4500	
MV10603	
MV10902	
MV10953	
MV11114	
MV11414	

Table 13: Base conservation of G4 sequences among 55 measles strains. Consensus sequences were obtained by the local alignment of each cRNA deposited sequence. Consensus logos were obtained using WebLogo software (http://weblogo.berkeley.edu/logo.cgi).

4.4.3 Putative G4-forming sequences can fold into G4 structures

The actual ability of putative G4-folding sequences and their thermal stability were next investigated by CD spectroscopy.

The parallel G4 conformation usually shows a strong positive peak at 260 nm and a minimum peak around 240 nm; CD spectra with a positive peak around 295 nm and a negative peak around 260 nm are usually assigned to the antiparallel conformation, while CD spectra with two positive peaks at 260 and 295 nm are characteristic of the mixed- type conformation [256].

All identified putative G4-folding sequences were analyzed in the presence of physiological K^+ concentrations (100 mM K^+) to monitor their formation and their likely topology. All tested oligonucleotides (**Figure 66, solid lines**), except for MV10603 (**Figure 66, dotted line**), displayed

CD spectra typical of the parallel topology, with a positive peak around 265 nm and a negative peak around 240 nm, therefore confirming the fact that RNA G4s almost exclusively adopt a parallel conformation in which the 4 strands have the same directionality. This feature is totally addressed by the fact that the 2'-hydroxil group of the ribose locks the RNA in an anti-conformation, which therefore favors the parallel topology [257].

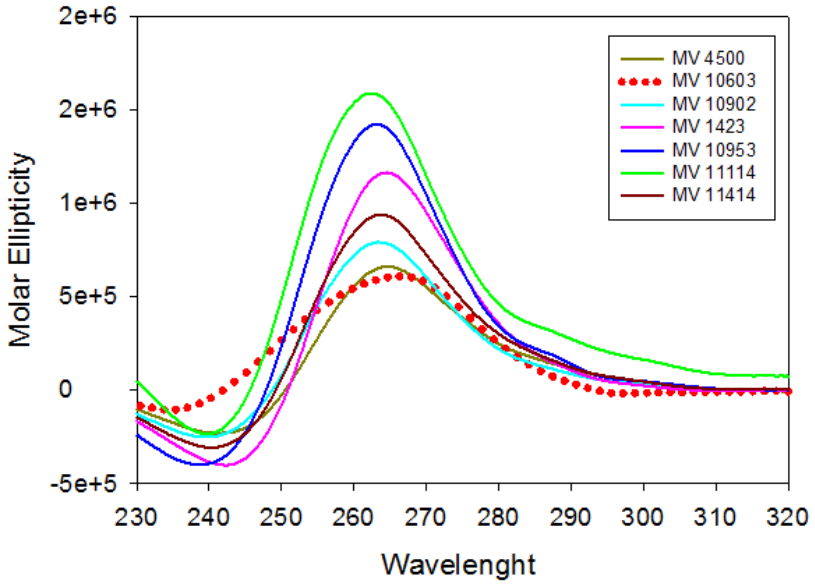


Figure 66: CD analysis of MV G-quadruplex forming sequences identified by QGRS Mapper. Each sequence was analyzed in the presence of physiological concentrations of K^+ (100 mM). All sequences (solid lines) displayed a CD signature typical of parallel G4s except for MV10603 sequence (dotted line), which displayed an hairpin-like secondary conformation.

MV10603 did not display the typical G4 topology: the CD spectra showed a broad and weak positive peak at 265 nm, which resembled the peak of an hairpin as described by [258].

Each sequence was then subjected to thermal unfolding analyses to monitor the thermal stability in 100 mM K^+ (**Table 14**).

T_m (°C) 100 mM K^+				
Sequence	**-Braco-19**	**+Braco-19**	**DT_m**	**Structure**
MV1423	62.2 ± 1.7	>90	>27.7	G4
MV4500	68.9 ± 0.5	76.7 ± 0.1	7.7	G4
MV10603	57.8 ± 2.2 / >90	79.4 ± 7	21.6	Hairpin
MV10902	>90	>90	n.d.	G4
MV10953	>90	>90	n.d.	G4
MV11114	>90	>90	n.d.	G4
MV11414	53.6 ± 0.8	89.2 ± 0.4	35.6	G4

Table 14: Stabilization (T_m) of MV G4 forming sequences identified with QGRS Mapper. Thermal stability was assessed by CD spectroscopy in the presence of 100 mM K^+ and in the presence or absence of commercial G4 ligand B-19. N.d. stands for not detected.

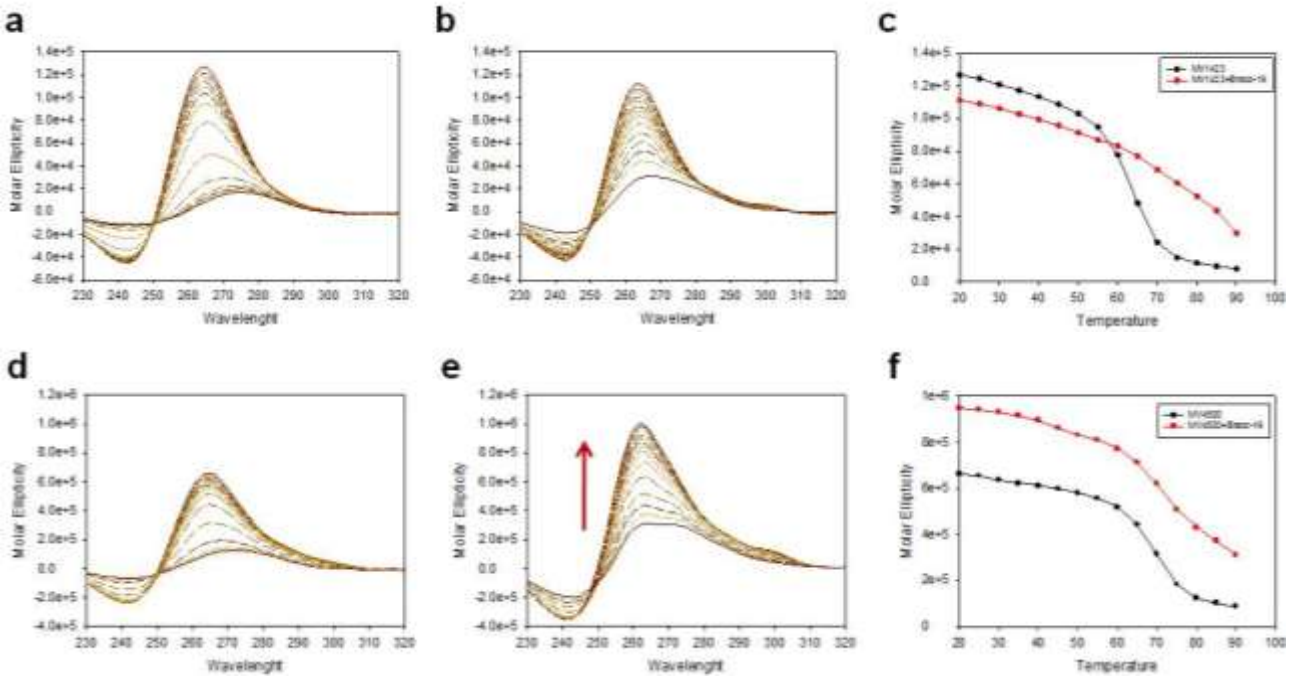
Three of the seven sequence identified, MV10902, MV10953 and MV 11114, showed a thermal stability above 90°C while MV1423, MV4500, MV10603 and MV11414 displayed a less pronounced stability with T_m about 62.2 ± 1.7, 68.9 ± 0.5, 57.8 ± 2.2 / >90 and 53.6 ± 0.8 °C respectively (**Table 14**).

Next we asked whether the G4s identified within the genome of MV could be subjected to binding and stabilization from the commercial available G4-ligand B-19, in the presence of 100 mM K^+.

Sequences that already in the absence of the compound displayed a significant stability, MV 10902, MV10953 and MV11114, were additionally stabilized by the binding of the compound as described by T_m curves shift to higher temperatures (**Figure 67n,q,t**). Moreover for MV10902 sequence a significant increase in the molar ellipticity was obtained upon addition of B-19 (**Figure 67m**).

MV1423, MV10603 and MV 11414 were subjected to an increase in the T_m of >27.7, 21.6 and 35.6 °C, respectively, while MV4500 was the less stabilized sequence with a ΔT_m of 7.7 °C (**Table 14**). Both MV4500 and MV10603 were subjected to a conspicuous increase in molar ellipticity upon addition of B-19 (**Figure 67e,h**) and, moreover, MV10603 adopted a G4 signature upon addition of the compound (**Figure 67h**).

Additionally for MV1423, MV10953, MV11114 and MV11414 a decrease in molar ellipticity in thermal spectra was measured upon addition of B-19, but this was accompanied by a shift of the T_m curves to higher temperatures (**Figure 67b,p,s,v**).



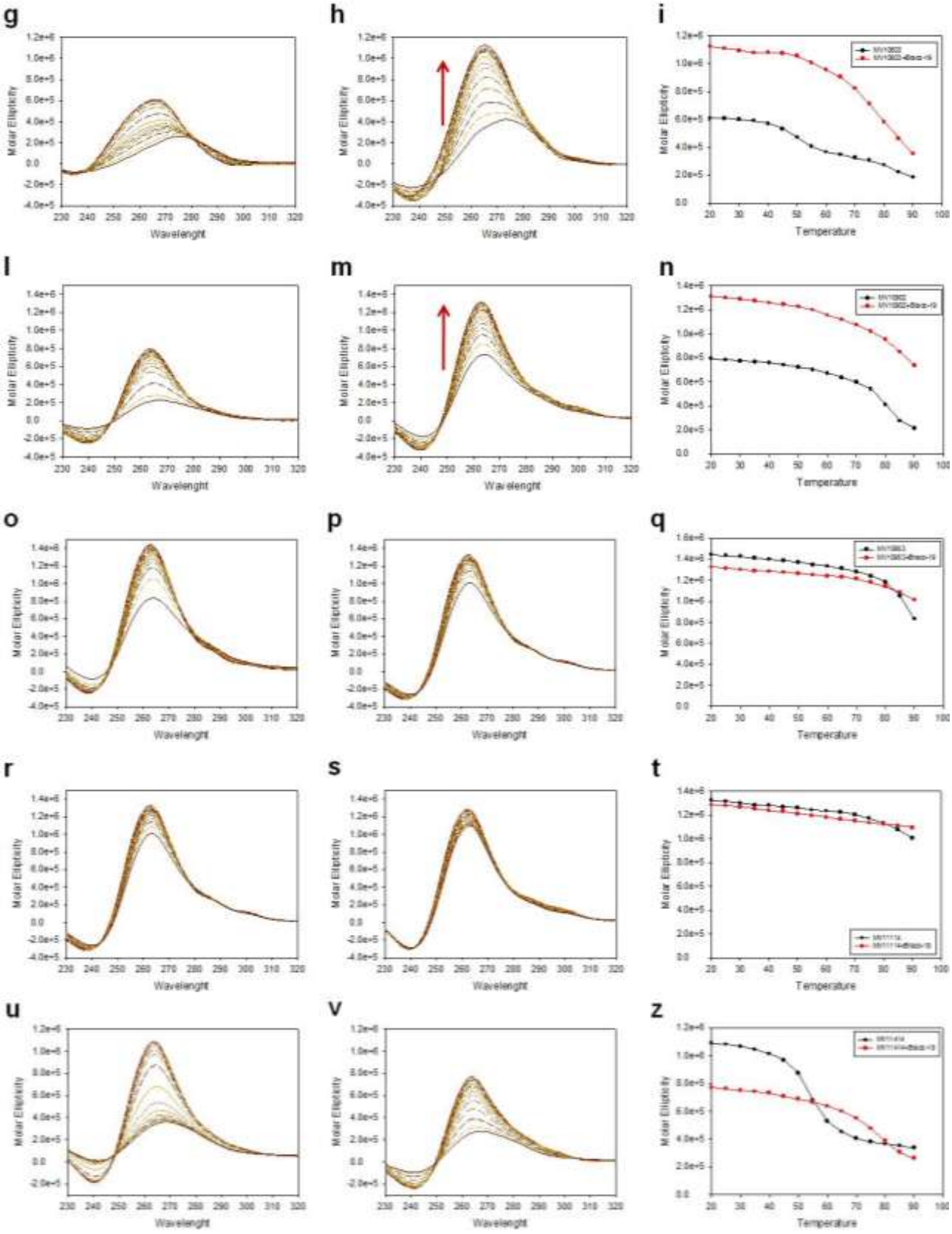


Figure 67: Thermal unfolding experiments on the G4-putative folding sequences identified within measles virus (MV) genome. Each sequence was analyzed in the presence of 100 mM K^+ and in the absence of presence of the commercial available G4-ligand B-19 (B-19). (**Left panel**) thermal unfolding analyses performed in the absence of B-19. (**Central panel**) thermal unfolding analyses performed in the presence of B-19. (Right panel) Melting curves calculated around 260 nm for each sequence. (**a-b-c**) MV1423. (**d-e-f**) MV4500. (**g-h-i**) MV10603. (**l-m-n**) MV10902. (**o,p,q**) MV10953. (**r-s-t**) MV11114. (**u-v-z**) MV11414. Thermal unfolding analyses were recorded over a temperature range of 20 °C- 90°C with temperature increase of 5°C. Increase in the molar ellipticity are indicated with red arrows.

4.4.4 BRACO-19 displays antiviral activity against MV

Infected cells were preliminary treated with B-19 to investigate if this G4-ligand is active against MV.

Since B-19 displays very mild cytotoxicity, it was used at the concentration of 100 μM (highest concentration tested in previous experiments with different viruses). B-19 activity was compared with that of Ribavirin (RBV), a well-known antiviral compound used in therapy and as a reference drug. RBV was used at concentration of 10 μM and 50 μM (the first one being in the IC_{50} range on Vero cells) [241].

Cells were infected with wild-type MV for 1 h at 37°C in DMEM without serum and then incubated with each compound for 48 h. Since viral products accumulation (mRNA, genomes and antigenomes) is very slow in the first 24 h of infection [208], we monitored the progression at 24 h.p.i with optical observation. Supernatants were collected at 48 h.p.i and further processed for viral genome quantification.

Preliminarily, as described in **Figure 68**, quantitative real-time PCR showed that B-19 possesses the ability to reduce viral genome accumulation of about 60%, while a more pronounced antiviral activity was observed for RBV at both concentrations used (between 80 and 100% of antiviral activity).

We may hypothesize that the limited effect of B-19 is due to a limited number of G4s within the MV genome that reduces the probability of the G4/G4-ligand binding in the infected cell environment. On the contrary, since RBV acts as a nucleoside analog that induces errors in the replication and translation of MV genome, its activity is directed toward all genome molecules present during viral replication.

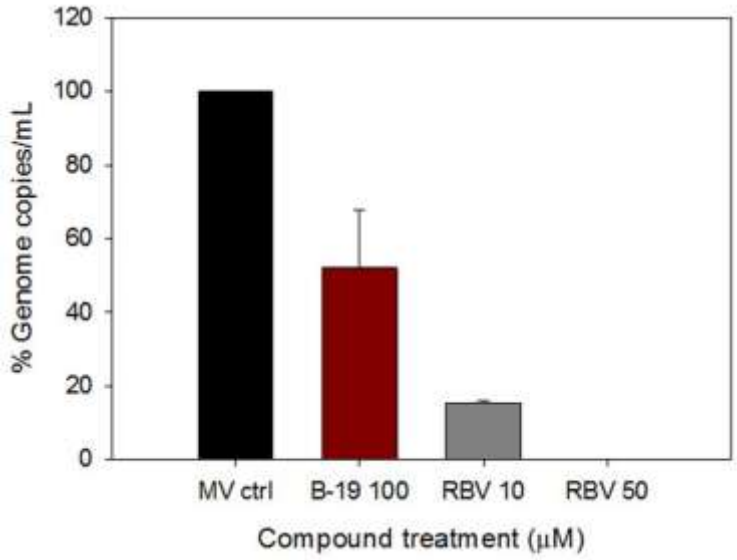


Figure 68: Effect of B-19 and RBV on MV replication. Infected cells were treated with B-19 (100 μM) or RBV (10 or 50 μM). Forty-eight h.p.i total RNA was isolated, retrotranscribed and quantified by quantitative real-time PCR (N gene). Each sample was analyzed in duplicate.

Considerable evidence has been reported on the existence of G4s structures within viral and other microorganisms genomes and their possible use as novel antiviral targets has been extensively investigate by our group [52,144,197,243,259].

In recent years the decline of the confidence for vaccination programs has caused the re-emergence of infection previously eradicated. This is the case of MV infections that in recent years have caused an increase in hospitalized patients with severe adverse effects [201]. Since G4 structures are strongly correlated with control mechanisms of viral biological processes and are strongly conserved within different genomes we decided to focus our studies in the investigation of the presence of RNA G4s within MV, principal actor of an outbreak in Padua hospital this year.

By analyzing the viral genome, we have identified 7 sequences with not only a putative possibility to fold into G4s but with the effective ability to adopt a secondary structure. Each sequence was stabilized by the binding of the G4-ligand B-19 and this clearly opens the possibility of an antiviral activity mediated by G4-ligands. Moreover, the non-selective G4-ligand B-19 demonstrated a preliminary promising antiviral activity that reduced viral production up to 60%. Thorough investigations must be done to evaluate B-19 mechanism of action, but taken together all these preliminary results strongly support our hypothesis of a potential use of G4s as broad range drug targets in infections caused by different etiologic agents.

5. Conclusions

This thesis focused on the evaluation of G4-ligands activities on HSV-1 G4s previously presented by our research group.

Since HSV-1 genome was reported for the high guanine content (68%) [158] and for the high GC content in repeated sequences [157], our group extensively investigated the presence of G4-putative folding sequences and the possibility to target them with G4-ligands.

A previous work from our group demonstrated that HSV-1 presents multiple extended repeated clusters of G4-forming sequences, characterized by a high stability.

The use of the non selective G4-ligand B-19, a tri-substituted acridine characterized by excellent G4 binding properties [81,83], showed that (i) all identified sequences were strongly stabilized upon the binding, (ii) viral DNA processing was arrested *in vitro* when G4s were folded and (iii) viral DNA synthesis was inhibited *in vivo* in infected cells. Finally, it was demonstrated that B-19 acts on viral DNA polymerase by stabilizing the extended clusters of G4s present in repeated regions [87].

This first evidence, the presence of G4-folding sequences within HSV-1 genome and the impairment of DNA synthesis when G4s were folded, was the starting point for subsequent analyses.

Naphthalene Diimides (NDI) and c-exNDIs started to acquire importance as innovative families of compounds with promising activity toward cellular G4s [109,110,115] and promising selectivity toward viral G4s [118] and for this reason, we employed the most promising c-exNDI, previously tested against HIV-1, to investigate its ability to act toward HSV-1 G4s.

We showed that the tested c-exNDI potently stabilizes G4s found in the HSV-1 genome by increasing their melting temperatures. Moreover we demonstrated that c-exNDI is able to induce a conformational change in the *gp054a* sequence from its hybrid-type structure to a parallel one, therefore confirming the preference of c-exNDI for parallel sequences [118]. Additionally we confirmed the c-exNDI binding to HSV-1 G4s by *Taq* polymerase stop assay: a significant concentration-dependent polymerase stalling on DNA, counterbalanced by a concentration-dependent reduction in the full-length synthesis, was measured upon binding to all sequences. These results confirmed that the presence of a G4 structure within a sequence directly inhibits polymerase activity and this *in vitro* inhibition has the potential to be maintained also *in vivo*. Interestingly, by Mass Spectrometry competition assay (MS), we confirmed the c-exNDI selectivity to the HIV-1 LTR III G4; even if a net preference for HSV-1 G4s over some cellular G4s was not observed, nevertheless c-exNDI displayed selectivity for HSV-1 G4s over the telomeric G4 (the

most represented cellular G4 in a cellular environment) therefore suggesting a preferential binding of viral G4s in infected cells.

We showed that treatment of HSV-1 infected cells with c-exNDI in a low nanomolar range induced a complete inhibition of virus production with a minimal/low effect on cell viability.

We were able to ascribe this behavior to two main reasons: first the massive presence of G4s within the viral genome and during HSV-1 cycle already demonstrated by our group. By an antibody- based approach in cells, we were able to demonstrate that viral G4s forms massively in the infected cell nucleus during viral replication [197].

During viral replication G4s exert a double role by (i) controlling viral replication itself and by (ii) controlling other cellular mechanisms of infection, with dramatic consequences once stabilized by c-exNDI. In fact, we demonstrated that viral replication is the exact step targeted by c-exNDI therefore confirming its G4-mediated mechanism of action on HSV-1.

Secondly, the higher affinity for c-exNDI for HSV-1 G4s over telomeric G4s, which are the most abundant G4s present within infected cells. Even if they are highly represented in the nuclei, c-exNDI does not optimally recognize these structures in favor of viral G4s: this is likely the main reason for the observed mild/low cellular cytotoxicity at the antiviral effective dose used.

Impairment of viral replication led to the decrease of viral genes in all phases of the viral life cycle (IE, E and L genes). This effect was shared between c-exNDI and ACV [249] therefore confirming our model of c-exNDI G4s-mediated activity toward viral DNA.

We cannot exclude the fascinating possibility that c-exNDI reacts with additional G4s, such as those in key regulatory IE and E genes distributed, beside the extended G4s repeats in the terminal and inverted repeats, in some promoter and coding regions of some fundamental genes [87]. Even if c-exNDI is characterized by a less wide therapeutic window than ACV, it is active in the low nanomolar range which can be considered a promising starting point for its development as anti-herpetic drug.

It has to be noted that all G4-ligands are mainly characterized by a large aromatic surface which makes them not easily druggable and sometimes not easily uptaken by cells. This consideration prompted us to investigate the activity of an in-house library of compounds with drug-like characteristics such as a limited surface size.

All compounds tested in this second part of the thesis were able to significantly stabilize HSV-1 G4s by inducing conspicuous increase in their melting temperatures and, again, the conformational change in the *gp054a* already observed with c-exNDI. All compounds induced polymerase stalling easily visible in the strong reduction of the full-length amplicon, therefore confirming a direct inhibition of polymerase activity due to a steric hindrance.

All G4-ligands from the in-house library displayed a promising antiviral activity in HSV-1 infected cells, being active at low nanomolar concentrations and not cytotoxic. For all compounds promising selectivity indexes were obtained ranging from 15.5 to 101.5. The most potent compound, GSA-0932, with an IC_{50} about 214.85 ± 21nM and a CC_{50} about 21.8 ± 3.9 μM was used for the investigation the mechanism of action assessed by this family of compounds. A perfect superimposable trend of GSA-0932 and ACV confirmed that GSA-0932 could act on HSV-1 by impairing its viral DNA replication with a consequent reduction in viral particles production. Preliminary evidences suggested that since compounds of the in-house library can stack on G4-tetrads, as previously demonstrated [102], they can interfere with DNA replication by a G4-mediated mechanism.

All results presented in this part of the thesis strongly suggested that the general antiviral activity of G4-ligands against HSV-1 can be ascribed to (i) the extensive presence of G4s-folding sequences within HSV-1 genome, (ii) the massive presence of G4s within infected cells at the time of replication and (iii) their optimal recognition over cellular G4s, therefore opening to the theoretical possibility that all G4-ligands can induce a decrease in the viral production (**Figure 69**).

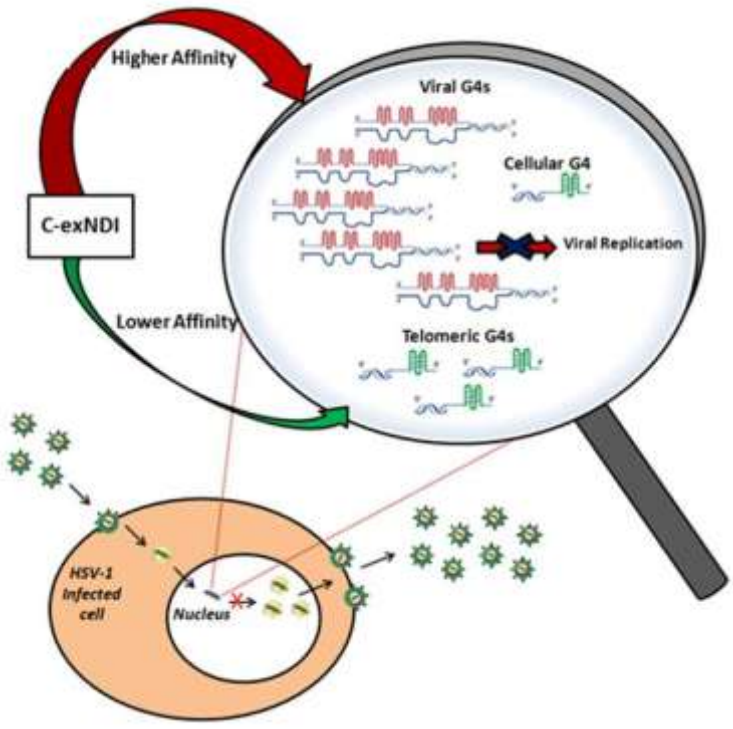


Figure 69: Scheme of the proposed c-exNDI mechanism of action of anti-HSV-1 activity. This mechanism can be enlarged to other G4-ligands useful for the treatment of HSV-1 related infection (i.e. G4-ligands of the in-house library of compounds tested in this thesis). Adapted from [243].

The effectiveness against HSV-1 makes c-exNDI and other G4-ligands suitable for the treatment of HSV-1 related infections. Since its first introduction in the 1980s, ACV has been the antiviral drug of choice for the treatment of HSV-1 infections [247], but the emergence of strains resistant to ACV

has created an important obstacle for the treatment of HSV-1-related infections, common illness associated to immunosuppression [260]. Since the circulation of resistant strains and their related infections is dangerous for AIDS, cancer and transplanted patients, there is an urgent need for novel anti HSV-1 therapies able to overcome viral resistance and effective in eradicating HSV-1.

In the third part of this thesis we focused our attention in the elucidation of the potential role of G4s within HSV-1 life cycle. To this purpose we adopted a pull-down approach for the identification of G4 DNA binding proteins able to interact with *un3long3* G4-folding sequence (extensive form of the *un3* G4 folding sequence clustered in the "*a*" terminal region of the viral genome). In preliminary results we identified both cellular and viral proteins, some of them being already identified as G4-binding proteins [19]. Concerning cellular proteins, we identified hnRNPU, nucleolin and nucleophosmin, strongly confirming that G4s are involved in a replication process and that their presence can require specific proteins (hnRNP) for their unfolding or their stabilization for transcription switch on or off, respectively. Identification of histone proteins, already previously reported for their ability to bind G4s [250], cannot exclude a potential role of G4s in viral DNA maintenance within latency mechanisms but further investigations must be done. Concerning viral proteins, we identified ICP4, ICP8, VP5 and PAP1, all proteins being involved in the viral replication. ICP4, ICP8 and PAP1 are all localized at the replication forks during a replication process [238,251,252], and this confirms a possible implication of G4s folding in the regulation of this mechanism.

Additionally, we identified the major capsid protein VP5, therefore opening the possibility of an implication of G4s as a driving force in the genome condensation during the virion assembly.

Finally in the last part of this thesis, since many reports described the presence of G4s not only in DNA genomes but also in RNA genomes [132], we decided to investigate the theoretical possibility of the presence of G4-folding sequences within the RNA genome of Measles Virus (MV).

Six sequences were identified within the single-stranded negative-sensed RNA while one sequence was found in the positive strand; all sequences were located in the coding region of viral genes therefore suggesting a potential role in the regulation of transcription mechanisms. Additionally all sequences (i) were conserved among 55 MV strains aligned, (ii) were located in the most conserved regions of the viral genome, and (iii) displayed a "GGG" pattern conservation of more than 80%, strongly confirming that G4s play essential roles within the viral life cycle and therefore should be highly conserved across viral evolution. We demonstrated that all sequences, except one in the negative-stranded RNA with a hairpin-like conformation, displayed a G4 signature at physiological concentrations of K^+ and a pronounced stability (with T_m in some cases above >90°C). Moreover, the addition of the G4-ligand B-19 induced an increase in the thermal stability of all sequences and the conversion to a G4 structure of the hairpin-like sequence. Preliminarily,

we demonstrated that treatment of MV infected cells with B-19 induced a decrease in the viral genome copies accumulation up to 60%.

The decline in the confidence for vaccination programs has caused the re-emergence of infections, such as the one caused by MV, that were previously strongly eradicated. Additionally, the absence of a specific therapy against MV prompted us to investigate the possible employment of G4-ligands against this virus.

Taken together all the results obtained across this thesis project, we can conclude that since G4s are present in many different viruses (i.e. HIV-1, HSV-1 and MV studied by our group), and are strongly conserved among different viruses and different strains, they possess important roles in the regulation of key processes during host cells infection. Additionally we importantly confirmed that targeting G4s with G4-ligands can represent an innovative therapeutic approach applicable for a broad range of microorganisms. For this reason we can hypothesize an application against strains resistant to current therapy (HSV-1-ACV resistant strains) and against viruses that still do not have a reference drug for their treatment (MV).

References

1. Gellert, M., Lipsett, M. & Davies, D. Helix Formation By Guanylic Acid. *Proc. Natl. Acad. Sci.* **48,** 2013–2018 (1962).

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

3. Zhang, S., Wu, Y. & Zhang, W. G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem* **9,** 899–911 (2014).

4. Bang, I. Untersuchungen uber die Guanylsaure. *Biochem. Z.* **26,** 293–231 (1910).

5. Bryan, T. & Baumann, P. G-quadruplexes: From guanine gels to chemotherapeutics. *Mol. Biotechnol.* **49,** 198–208 (2011).

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

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

8. Parrotta, L. *et al.* Targeting unimolecular G-quadruplex nucleic acids: a new paradigm for the drug discovery? *Expert Opin. Drug Discov.* **9,** 1167–1187 (2014).

9. Yuwei, C. & Danzhou, Y. Sequence, Stability, and Structure of G-Quadruplexes and Their Interactions with Drugs. *Curr. Protoc. Nucleic Acid Chem.* 1–26 (2012).

10. Balasubramanian, S. & Neidle, S. G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **13,** 345–353 (2009).

11. Qin, Y. & Hurley, L. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90,** 1149–1171 (2008).

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

13. Frees, S., Menendez, C., Crum, M. & Bagga, P. QGRS-Conserve: a computational method for discovering evolutionarily conserved G-quadruplex motifs. *Hum. Genomics* **8,** 1–13 (2014).

14. Di Antonio, M., Rodriguez, R. & Balasubramanian, S. Experimental approaches to identify cellular G-quadruplex structures and functions. *Methods* **57,** 84–92 (2012).

15. Burge, S., Parkinson, G., Hazel, P., Todd, A. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34,** 5402–5415 (2006).

16. Tóthová, P., Krafčíková, P. & Víglaský, V. Formation of Highly Ordered Multimers in G-Quadruplexes. *Biochemistry* **53,** 7013–7027 (2014).

17. Huppert, J. Hunting G-quadruplexes. *Biochimie* **90,** 1140–1148 (2008).

18. Harris, L. & Merrick, C. G-Quadruplexes in Pathogens: A Common Route to Virulence

Control? *PLOS Pathog.* **11,** 1–15 (2015).

19. Brázda, V., Hároníková, L., Liao, J. & Fojta, M. DNA and RNA Quadruplex-Binding Proteins. *Int. J. Mol. Sci.* **15,** 17493–17517 (2014).

20. Lopes, J. *et al.* G-quadruplex-induced instability during leading-strand replication. *EMBO J.* **30,** 4033–4046 (2011).

21. Maizels, N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.* **13,** 1055–1059 (2006).

22. Lam, E., Beraldi, D., Tannahill, D. & Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* **4,** 1796–1812 (2013).

23. Yang, D. & Okamoto, K. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.* **2,** 619–646 (2010).

24. Valton, A. & Prioleau, M. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet.* **32,** 697–706 (2016).

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

26. Mohaghegh, P., Karow, J., Brosh, R., Bohr, V. & Hickson, I. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.* **29,** 2843–2849 (2001).

27. Sanders, C. Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem. J.* **430,** 119–128 (2010).

28. London, T. *et al.* FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.* **283,** 36132–36139 (2008).

29. Castillo Bosch, P. *et al.* FANCJ promotes DNA synthesis through G-quadruplex structures. *EMBO J.* **33,** 2521–2533 (2014).

30. Mirkin, S. DNA replication: Driving past four-stranded snags. *Nature* **497,** 449–450 (2013).

31. Bochman, M., Paeschke, K. & Zakian, V. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13,** 770–780 (2012).

32. Bailey, S. & Murnane, J. Telomeres, chromosome instability and cancer. *Nucleic Acids Res.* **34,** 2408–2417 (2006).

33. Roderick, J. & Karlseder, J. Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **11,** 171–181 (2010).

34. Onel, B., Lin, C. & Yang, D. DNA G-quadruplex and its potential as anticancer drug target. *Sci. China Chem.* **57,** 1605–1614 (2014).

35. Li, G., Eller, M., Firoozabadi, R. & Gilchrest, B. Evidence that exposure of the telomere 3′ overhang sequence induces senescence. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 527–531 (2003).

36. O'Connor, B. Telomeres of Human Chromosomes. *Nat. Educ.* **1,** 166 (2008).

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

38. Phatak, P. & Burger, A. Telomerase and its potential for therapeutic intervention. *Br. J. Pharmacol.* **152,** 1003–1011 (2009).

39. McLuckie, K. *et al.* G-Quadruplex DNA as a Molecular Target for Induced Synthetic Lethality in Cancer Cells. *J. Am. Chem. Soc.* **135,** 9640–9643 (2013).

40. Ying, L., Green, J., Li, H., Klenerman, D. & Balasubramanian, S. Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 14629–14634 (2003).

41. Phan, A., Kuryavyi, V., Luu, K. & Patel, D. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K + solution †. *Nucleic Acids Res.* **35,** 6517–6525 (2007).

42. Parkinson, G., Lee, M. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417,** 876–880 (2002).

43. Phan, A. & Patel, D. Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics. *J. Am. Chem. Soc.* 15021–15027 (2003). doi:10.1021/JA037616J

44. Alcaro, S. *et al.* The polymorphisms of DNA G-quadruplex investigated by docking experiments with telomestatin enantiomers. *Curr. Pharm. Des.* **18,** 1873–1879 (2012).

45. Sattin, G. *et al.* Conformation and stability of intramolecular telomeric G-quadruplexes: sequence effects in the loops. *PLoS One* **8,** e84113 (2013).

46. Artese, A. *et al.* Toward the design of new DNA G-quadruplex ligands through rational analysis of polymorphism and binding data. *Eur. J. Med. Chem.* **68,** 139–149 (2013).

47. Zahler, A., Williamson, J., Cech, T. & Prescott, D. Inhibition of telomerase by G-quartet DMA structures. *Nature* **350,** 718–720 (1991).

48. Neidle, S. Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **277,** 1118–1125 (2010).

49. Balasubramanian, S., Hurley, L. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **10,** 261–275 (2011).

50. Neidle, S. The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* **19,** 239–250 (2009).

51. Capra, J., Paeschke, K., Singh, M. & Zakian, V. G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces cerevisiae. *PLoS Comput. Biol.* **6,** e1000861 (2010).

52. Perrone, R. *et al.* Mapping and characterization of G-quadruplexes in Mycobacterium tuberculosis gene promoter regions. *Sci. Rep.* **7,** 5743–5754 (2017).

53. Todd, A. & Neidle, S. The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Res.* **36,** 2700–2704

(2008).

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

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

56. Siddiqui-Jain, A., Grand, C., Bearss, D. & Hurley, L. 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.* **99,** 11593–11598 (2002).

57. Dexheimer, T., Sun, D. & Hurley, L. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the bcl-2 P1 Promoter. *J. Am. Chem. Soc.* 5404–5415 (2006). doi:10.1021/JA0563861

58. Sun, D., Guo, K., Rusche, J. & Hurley, L. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **33,** 6070–6080 (2005).

59. Rankin, S. *et al.* Putative DNA Quadruplex Formation within the Human c-kit Oncogene. *J. Am. Chem. Soc.* 10584–10589 (2005). doi:10.1021/JA050823U

60. Palumbo, M., SunMi, L., Ebbinghaus, S. & Hurley, L. Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the hTERT Core Promoter with Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands. *J. Am. Chem. Soc.* **131,** 10878–10891 (2009).

61. González, V. & Hurley, L. The c- *MYC* NHE III $_1$: Function and Regulation. *Annu. Rev. Pharmacol. Toxicol.* **50,** 111–129 (2010).

62. Brown, R., Danford, F., Gokhale, V., Hurley, L. & Brooks, T. Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *J. Biol. Chem.* **286,** 41018–41027 (2011).

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

64. Adrian, M. *et al.* Structure and Conformational Dynamics of a Stacked Dimeric G-Quadruplex Formed by the Human CEB1 Minisatellite. *J. Am. Chem. Soc.* **136,** 6297–6305 (2014).

65. Dunnick, W., Hertz, G., Scappino, L. & Gritzmacher, C. DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res.* **21,** 365–372 (1993).

66. Duquette, M., Handa, P., Vincent, J., Taylor, A. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* **18,** 1618–1629 (2004).

67. Wallgren, M. *et al.* G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures *in vitro* and are unwound by the Pfh1 DNA helicase. *Nucleic Acids Res.* **44,** 6213–6231 (2016).

68. Huppert, J., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* **36,** 6260–6268 (2008).

69. Broxson, C., Beckett, J. & Tornaletti, S. Transcription Arrest by a G Quadruplex Forming-Trinucleotide Repeat Sequence from the Human c-myb Gene. *Biochemistry* **50,** 4162–4172 (2011).

70. Belotserkovskii, B. *et al.* Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12816–12821 (2010).

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

72. Tornaletti, S. Transcriptional processing of G4 DNA. *Mol. Carcinog.* **48,** 326–335 (2009).

73. Read, M. *et al.* Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4844–4849 (2001).

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

75. Fedoroff, O. *et al.* NMR-Based Model of a Telomerase-Inhibiting Compound Bound to G-Quadruplex DNA†. *Biochemistry* 12367–12374 (1998). doi:10.1021/BI981330N

76. Fiel, R., Howard, J., Mark, E. & Datta Gupta, N. Interaction of DNA with a porphyrin ligand: evidence for intercalation. *Nucleic Acids Res.* **6,** 3093–118 (1979).

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

78. Kim, M., Vankayalapati, H., Shin-Ya, K., Wierzba, K. & Hurley, L. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *J. Am. Chem. Soc.* **124,** 2098–2109 (2002).

79. Han, H., Hurley, L. & Salazar, M. A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.* **27,** 537–542 (1999).

80. Ungvarsky, J. *et al.* Novel trisubstituted acridines as human telomeric quadruplex binding ligands. *Bioorg. Chem.* **57,** 13–29 (2014).

81. Gowan, S. *et al.* Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with G-quadruplexes. *Mol. Pharmacol.* **60,** 981–988 (2001).

82. White, E. *et al.* Structure-specific recognition of quadruplex DNA by organic cations: influence of shape, substituents and charge. *Biophys. Chem.* **126,** 140–153 (2007).

83. Burger, A. *et al.* The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.* **65,** 1489–1496 (2005).

84. Huppert, J. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* **37,** 1375–1384 (2008).

85. Norseen, J., Johnson, F. & Lieberman, P. Role for G-Quadruplex RNA Binding by Epstein-Barr Virus Nuclear Antigen 1 in DNA Replication and Metaphase Chromosome Attachment. *J. Virol.* **83,** 10336–10346 (2009).

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

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

88. Wei, C., Han, G., Jia, G., Zhou, J. & Li, C. Study on the interaction of porphyrin with G-quadruplex DNAs. *Biophys. Chem.* **137,** 19–23 (2008).

89. Wheelhouse, RT. Sun, D. Han, H. Han, FX. Hurley, L. Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(N-methyl-4-pyridyl)porphine with Quadruplex DNA. *J. Am. Chem. Soc.* 3261–3262 (1998). doi:10.1021/JA973792E

90. Grand, C. *et al.* The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.* **1,** 565–573 (2002).

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

92. Izbicka, E. *et al.* Effects of Cationic Porphyrins as G-Quadruplex Interactive Agents in Human Tumor Cells. *Cancer Res.* **59,** (1999).

93. Han, F., Wheelhouse, R. & Hurley, L. Interactions of TMPyP4 and TMPyP2 with Quadruplex DNA. Structural Basis for the Differential Effects on Telomerase Inhibition. *J. Am. Chem. Soc.* 3561–3570 (1999). doi:10.1021/JA984153M

94. Dixon, I. *et al.* A G-Quadruplex Ligand with 10000-Fold Selectivity over Duplex DNA. *J. Am. Chem. Soc.* **129,** 1502–1503 (2007).

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

96. Berangere, G., Schultes, C., Hazel, P., Mann, J. & Neidle, S. Synthesis and evaluation of analogues of 10H-indolo[3,2-b]quinoline as G-quadruplex stabilising ligands and potential inhibitors of the enzyme telomerase. *Org. Biomol. Chem.* **2,** 981–988 (2004).

97. Ou, T. *et al.* Stabilization of G-Quadruplex DNA and Down-Regulation of Oncogene c-myc by Quindoline Derivatives. *J. Med. Chem.* **50,** 1465–1474 (2007).

98. Dai, J., Carver, M., Hurley, L. & 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.* **133,** 17673–17680 (2011).

99. Kaiser, C., Gokhale, V., Yang, D. & Hurley, L. Gaining Insights into the Small Molecule Targeting of the G-Quadruplex in the c-MYC Promoter Using NMR and an Allele-Specific Transcriptional Assay. in *Topics in current chemistry* **330,** 1–21 (2012).

100. Boddupally, P. *et al.* Anticancer Activity and Cellular Repression of c-MYC by the G-

Quadruplex-Stabilizing 11-Piperazinylquindoline Is Not Dependent on Direct Targeting of the G-Quadruplex in the c-MYC Promoter. *J. Med. Chem.* **55,** 6076–6086 (2012).

101. Smith, L. *et al.* A targeted oligonucleotide enhancer of SMN2 exon 7 splicing forms competing quadruplex and protein complexes in functional conditions. *Cell Rep.* **9,** 193–205 (2014).

102. Kendrick, S., Muranyi, A., Gokhale, V., Hurley, L. & Rimsza, L. Simultaneous Drug Targeting of the Promoter *MYC* G-Quadruplex and *BCL2* i-Motif in Diffuse Large B-Cell Lymphoma Delays Tumor Growth. *J. Med. Chem.* **60,** 6587–6597 (2017).

103. Tian, E. *et al.* Ellipticine derivative NSC 338258 represents a potential new antineoplastic agent for the treatment of multiple myeloma. *Mol. Cancer Ther.* **7,** 500–509 (2008).

104. Shin-ya, K. *et al.* Telomestatin, a Novel Telomerase Inhibitor from Streptomyces anulatus. *J. Am. Chem. Soc.* **123,** 1262–1263 (2001).

105. Doi, T., Yoshida, M., Shin-ya, K. & Takahashi, T. Total Synthesis of (R)-Telomestatin. *Org. Lett.* **8,** 4165–4167 (2006).

106. Han, H., Cliff, C. & Hurley, L. Accelerated Assembly of G-Quadruplex Structures by a Small Molecule†. *Biochemistry* **38,** 6981–6986 (1999).

107. Taka, T. *et al.* Telomere shortening and cell senescence induced by perylene derivatives in A549 human lung cancer cells. *Bioorg. Med. Chem.* **21,** 883–890 (2013).

108. Sissi, C. *et al.* Tri-, tetra- and heptacyclic perylene analogues as new potential antineoplastic agents based on DNA telomerase inhibition. *Bioorg. Med. Chem.* **15,** 555–62 (2007).

109. Di Antonio, M. *et al.* Novel Naphthalene Diimides as Activatable Precursors of Bisalkylating Agents, by Reduction and Base Catalysis. *J. Org. Chem.* **72,** 8354–8360 (2007).

110. Cuenca, F. *et al.* Tri- and tetra-substituted naphthalene diimides as potent G-quadruplex ligands. *Bioorg. Med. Chem. Lett.* **18,** 1668–1673 (2008).

111. Nadai, M. *et al.* Naphthalene diimide scaffolds with dual reversible and covalent interaction properties towards G-quadruplex. *Biochimie* **93,** 1328–1340 (2011).

112. Gunaratnam, M. *et al.* Targeting human gastrointestinal stromal tumor cells with a quadruplex-binding small molecule. *J. Med. Chem.* **52,** 3774–3783 (2009).

113. Nadai, M. *et al.* Assessment of gene promoter G-quadruplex binding and modulation by a naphthalene diimide derivative in tumor cells. *Int. J. Oncol.* **46,** 369–380 (2015).

114. Collie, G. *et al.* Structural Basis for Telomeric G-Quadruplex Targeting by Naphthalene Diimide Ligands. *J. Am. Chem. Soc.* **134,** 2723–2731 (2012).

115. Micco, M. *et al.* Structure-Based Design and Evaluation of Naphthalene Diimide G-Quadruplex Ligands As Telomere Targeting Agents in Pancreatic Cancer Cells. *J. Med. Chem.* **56,** 2959–2974 (2013).

116. Di Antonio, M. *et al.* Quinone Methides Tethered to Naphthalene Diimides as Selective G-Quadruplex Alkylating Agents. *J. Am. Chem. Soc.* **131,** 13132–13141 (2009).

117. Doria, F. *et al.* Targeting loop adenines in G-quadruplex by a selective oxirane. *Chemistry* **19,** 78–81 (2013).

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

119. Doria, F. *et al.* Water soluble extended naphthalene diimides as pH fluorescent sensors and G-quadruplex ligands. *Org. Biomol. Chem.* **10,** 3830–3840 (2012).

120. Hwang, H., Buncher, N., Opresko, P. & Myong, S. POT1-TPP1 Regulates Telomeric Overhang Structural Dynamics. *Structure* **20,** 1872–1880 (2012).

121. Soldatenkov, V., Vetcher, A., Duka, T. & Ladame, S. First Evidence of a Functional Interaction between DNA Quadruplexes and Poly(ADP-ribose) Polymerase-1. *ACS Chem. Biol.* **3,** 214–219 (2008).

122. Hanakahi, L., Sun, H. & Maizels, N. High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.* **274,** 15908–15912 (1999).

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

124. Chiarella, S. *et al.* Nucleophosmin mutations alter its nucleolar localization by impairing G-quadruplex binding at ribosomal DNA. *Nucleic Acids Res.* **41,** 3228–3239 (2013).

125. Gallo, A. *et al.* Structure of nucleophosmin DNA-binding domain and analysis of its complex with a G-quadruplex sequence from the c-MYC promoter. *J. Biol. Chem.* **287,** 26539–26548 (2012).

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

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

128. Teng, F. *et al. E. coli* DNA polymerase I can disrupt G-quadruplex structures during DNA replication. *FEBS J.* (2017). doi:10.1111/febs.14290

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

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

131. Sundquist, W. & Heaphy, S. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 3393–3397 (1993).

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

133. Patel, P., Bhavesh, N. & Hosur, R. NMR observation of a novel C-tetrad in the structure of

the SV40 repeat sequence GGGCGG. *Biochem. Biophys. Res. Commun.* **270,** 967–971 (2000).

134. Tlučková, K. *et al.* Human Papillomavirus G-Quadruplexes. *Biochemistry* **52,** 7207–7216 (2013).

135. Kutok, J. & Wang, F. SPECTRUM OF EPSTEIN-BARR VIRUS–ASSOCIATED DISEASES. *Annu. Rev. Pathol. Mech. Dis.* **1,** 375–404 (2006).

136. Murat, P. *et al.* G-quadruplexes regulate Epstein-Barr virus–encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* **10,** 358–364 (2014).

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

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

139. Göertz, G., Abbo, S., Fros, J. & Pijlman, G. Functional RNA during Zika virus infection. *Virus Res.* (2017). doi:10.1016/j.virusres.2017.08.015

140. Tan, J. *et al.* The SARS-Unique Domain (SUD) of SARS Coronavirus Contains Two Macrodomains That Bind G-Quadruplexes. *PLoS Pathog.* **5,** e1000428 (2009).

141. Woo, H. *et al.* Single-stranded DNA aptamer that specifically binds to the influenza virus NS1 protein suppresses interferon antagonism. *Antiviral Res.* **100,** 337–345 (2013).

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

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

144. Perrone, R., Lavezzo, E., Palù, G. & Richter, S. Conserved presence of G-quadruplex forming sequences in the Long Terminal Repeat Promoter of Lentiviruses. *Sci. Rep.* **7,** 2018 (2017).

145. Wertheim, J., Smith, M., Smith, D., Scheffler, K. & Kosakovsky Pond, S. Evolutionary Origins of Human Herpes Simplex Viruses 1 and 2. *Mol. Biol. Evol.* **31,** 2356–2364 (2014).

146. Zaichick, S., Bohannon, K. & Smith, G. Alphaherpesviruses and the Cytoskeleton in Neuronal Infections. *Viruses* **3,** 941–981 (2011).

147. Kieff, E., Hoyer, B., Bachenheimer, S. & Roizman, B. Genetic Relatedness of Type 1 and Type 2 Herpes Simplex Viruses. *J. Virol.* **9,** 738–745 (1972).

148. Higgins, C., Schofield, J., Tatnall, F. & Leigh, I. Natural history, management and complications of herpes labialis. *J. Med. Virol.* **Suppl 1,** 22–26 (1993).

149. Whitley, R. & Roizman, B. Herpes simplex virus infections. *Lancet (London, England)* **357,** 1513–1518 (2001).

150. Held, K. & Derfuss, T. Control of HSV-1 latency in human trigeminal ganglia—current overview. *J. Neurovirol.* **17,** 518–527 (2011).

151. Lieberman, P. Epigenetics and Genetics of Viral Latency. *Cell Host Microbe* **19,** 619–628 (2016).

152. Beswick, T. The origin and the use of the word herpes. *Med. Hist.* **6,** 214–232 (1962).

153. Strang, B. & Stow, N. Circularization of the herpes simplex virus type 1 genome upon lytic infection. *J. Virol.* **79,** 12487–12494 (2005).

154. De Silva, S. & Bowers, W. Herpes Virus Amplicon Vectors. *Viruses* **1,** 594–629 (2009).

155. Balliet, J. & Schaffer, P. Point mutations in herpes simplex virus type 1 oriL, but not in oriS, reduce pathogenesis during acute infection of mice and impair reactivation from latency. *J. Virol.* **80,** 440–450 (2006).

156. Weller, S. & Coen, D. Herpes simplex viruses: mechanisms of DNA replication. *Cold Spring Harb. Perspect. Biol.* **4,** a013011 (2012).

157. Ouyang, Q. *et al.* High GC content of simple sequence repeats in Herpes simplex virus type 1 genome. *Gene* **499,** 37–40 (2012).

158. Brown, J. High G+C Content of Herpes Simplex Virus DNA: Proposed Role in Protection Against Retrotransposon Insertion. *Open Biochem. J.* **1,** 33–42 (2007).

159. Zhao, X. *et al.* The genome of herpes simplex virus type 1 is prone to form short repeat sequences. *J. Biosci. Med.* **1,** 26–30 (2013).

160. Argnani, R., Lufino, M., Manservigi, M. & Manservigi, R. Replication-competent herpes simplex vectors: design and applications. *Gene Ther.* **12,** S170–S177 (2005).

161. Whitley, R., Kimberlin, D. & Roizman, B. Herpes Simplex Viruses. *CID* **26,** 861–873 (1998).

162. Fan, Q., Kopp, S., Connolly, S., Muller, W. & Longnecker, R. Mapping sites of herpes simplex virus type 1 glycoprotein D that permit insertions and impact gD and gB receptors usage. *Sci. Rep.* **7,** 43712 (2017).

163. Šedý, J., Spear, P. & Ware, C. Cross-regulation between herpesviruses and the TNF superfamily members. *Nat. Rev. Immunol.* **8,** 861–873 (2008).

164. Akhtar, J. & Shukla, D. Viral entry mechanisms: cellular and viral mediators of herpes simplex virus entry. *FEBS J.* **276,** 7228–7236 (2009).

165. Copeland, A., Newcomb, W. & Brown, J. Herpes simplex virus replication: roles of viral proteins and nucleoporins in capsid-nucleus attachment. *J. Virol.* **83,** 1660–1668 (2009).

166. Abrisch, R., Eidem, T., Yakovchuk, P., Kugel, J. & Goodrich, J. Infection by Herpes Simplex Virus 1 Causes Near-Complete Loss of RNA Polymerase II Occupancy on the Host Cell Genome. *J. Virol.* **90,** 2503–2513 (2015).

167. Samaniego, L., Neiderhiser, L. & DeLuca, N. Persistence and expression of the herpes simplex virus genome in the absence of immediate-early proteins. *J. Virol.* **72,** 3307–3320 (1998).

168. Sandri-Goldin, R. Replication of the herpes simplex virus genome: does it really go around in circles? *Proc. Natl. Acad. Sci. U. S. A.* **100,** 7428–7429 (2003).

169. Heming, J., Conway, J. & Homa, F. Herpesvirus Capsid Assembly and DNA Packaging. in *Advances in anatomy, embryology, and cell biology* **223,** 119–142 (2017).

170. Mettenleiter, T. Budding events in herpesvirus morphogenesis. *Virus Res.* **106,** 167–180 (2004).

171. Mettenleiter, T. Herpesvirus assembly and egress. *J. Virol.* **76,** 1537–1547 (2002).

172. Granzow, H. *et al.* Egress of alphaherpesviruses: comparative ultrastructural study. *J. Virol.* **75,** 3675–3684 (2001).

173. Beard, P., Faber, S., Wilcox, K. & Pizer, L. Herpes simplex virus immediate early infected-cell polypeptide 4 binds to DNA and promotes transcription. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 4016–4020 (1986).

174. Cai, W. & Schaffer, P. Herpes simplex virus type 1 ICP0 plays a critical role in the de novo synthesis of infectious virus following transfection of viral DNA. *J. Virol.* **63,** 4579–4589 (1989).

175. Galocha, B. *et al.* The active site of ICP47, a herpes simplex virus-encoded inhibitor of the major histocompatibility complex (MHC)-encoded peptide transporter associated with antigen processing (TAP), maps to the NH2-terminal 35 residues. *J. Exp. Med.* **185,** 1565–1572 (1997).

176. Lee, C. & Knipe, D. An immunoassay for the study of DNA-binding activities of herpes simplex virus protein ICP8. *J. Virol.* **54,** 731–738 (1985).

177. Gottlieb, J., Marcy, A., Coen, D. & Challberg, M. The herpes simplex virus type 1 UL42 gene product: a subunit of DNA polymerase that functions to increase processivity. *J. Virol.* **64,** 5976–5987 (1990).

178. Collot, M. *et al.* High conservation of herpes simplex virus UL5/UL52 helicase-primase complex in the era of new antiviral therapies. *Antiviral Res.* **128,** 1–6 (2016).

179. Trego, K., Zhu, Y. & Parris, D. The herpes simplex virus type 1 DNA polymerase processivity factor, UL42, does not alter the catalytic activity of the UL9 origin-binding protein but facilitates its loading onto DNA. *Nucleic Acids Res.* **33,** 536–545 (2005).

180. Gruffat, H., Marchione, R. & Manet, E. Herpesvirus Late Gene Expression: A Viral-Specific Pre-initiation Complex Is Key. *Front. Microbiol.* **7,** 869 (2016).

181. Nicoll, M., Proença, J. & Efstathiou, S. The molecular basis of herpes simplex virus latency. *FEMS Microbiol. Rev.* **36,** 684–705 (2012).

182. Wilson, A. & Mohr, I. A cultured affair: HSV latency and reactivation in neurons. *Trends Microbiol.* **20,** 604–611 (2012).

183. Lachmann, R. *The herpes simplex virus life cycle*. (Published by Cambridge University Press in association with the Clinical and Biomedical Computing Unit of the University of Cambridge School of Clinical Medicine, 1997).

184. Thellman, N. & Triezenberg, S. Herpes Simplex Virus Establishment, Maintenance, and Reactivation: In Vitro Modeling of Latency. *Pathogens* (2017). doi:10.3390/pathogens6030028

185. Vandevenne, P., Sadzot-Delvaux, C. & Piette, J. Innate immune response and viral interference strategies developed by Human Herpesviruses. *Biochem. Pharmacol.* **80,** 1955–1972 (2010).

186. Marchi, S., Trombetta, C., Gasparini, R., Temperton, N. & Montomoli, E. Epidemiology of herpes simplex virus type 1 and 2 in Italy: a seroprevalence study from 2000 to 2014. *J. Prev. Med. Hyg.* **58,** E27–E33 (2017).

187. Shen, J. *et al.* Seroprevalence of Herpes Simplex Virus Type 1 and 2 in Taiwan and Risk Factor Analysis, 2007. *PLoS One* **10,** e0134178 (2015).

188. Arvin, A. *Human Herpesviruses. Biology, Therapy, and Immunoprophylaxis.* (2007).

189. Itzhaki, R. *et al.* Herpes simplex virus type 1 in brain and risk of Alzheimer's disease. *Lancet (London, England)* **349,** 241–244 (1997).

190. Pyles, R. The association of herpes simplex virus and Alzheimer's disease: a potential synthesis of genetic and environmental factors. *Herpes* **8,** 64–68 (2001).

191. Tan, W., Jaganath, I., Manikam, R. & Sekaran, S. Evaluation of Antiviral Activities of Four Local Malaysian Phyllanthus Species against Herpes Simplex Viruses and Possible Antiviral Target. *Int. J. Med. Sci.* **10,** 1817–1829 (2013).

192. Elion, G. Acyclovir: discovery, mechanism of action, and selectivity. *J. Med. Virol.* **Suppl 1,** 2–6 (1993).

193. Brady, R. & Bernestein, D. Treatment of herpes simplex virus infections. *Antiviral Res.* **61,** 73–81 (2004).

194. Darby, G. Acyclovir--and beyond. *J. Int. Med. Res.* **22 Suppl 1,** 33A–42A (1994).

195. De Clercq, E. Antivirals for the treatment of herpesvirus infections. *J. Antimicrob. Chemother.* **32 Suppl A,** 121–132 (1993).

196. Doria, F. *et al.* Hybrid ligand–alkylating agents targeting telomeric G-quadruplex structures. *Org. Biomol. Chem.* **10,** 2798–2806 (2012).

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

198. Holzmann, H., Hengel, H., Tenbusch, M. & Doerr, H. Eradication of measles: remaining challenges. *Med. Microbiol. Immunol.* **205,** 201–208 (2016).

199. Laksono, B., De Vries, R., McQuaid, S., Duprex, W. & de Swart, R. Measles Virus Host Invasion and Pathogenesis. *Viruses* **8,** 210 (2016).

200. Mühlebach, M. *et al.* Adherens junction protein nectin-4 is the epithelial receptor for measles virus. *Nature* **480,** 530–533 (2011).

201. Moss, W. Measles. *Lancet (London, England)* **0,** (2017).

202. Aref, S., Bailey, K. & Fielding, A. Measles to the Rescue: A Review of Oncolytic Measles Virus. *Viruses* **8,** 294 (2016).

203. Yu, X. & Ghildyal, R. Measles Virus Infection: Mechanisms of Immune Suppression. in *Immunosuppression - Role in Health and Diseases* (InTech, 2012). doi:10.5772/29662

204. Piccirilli, G. *et al.* Molecular detection and genetic characterization of circulating measles virus in northern Italy. *J. Clin. Virol.* **81,** 34–42 (2016).

205. Rota, P. *et al.* Improving Global Virologic Surveillance for Measles and Rubella. *J. Infect. Dis.* **204,** S506–S513 (2011).

206. Sowers, S. *et al.* High Concentrations of Measles Neutralizing Antibodies and High-Avidity Measles IgG Accurately Identify Measles Reinfection Cases. *Clin. vaccine Immunol.* **23,** 707–716 (2016).

207. Bankamp, B., Takeda, M., Zhang, Y., Xu, W. & Rota, P. Genetic Characterization of Measles Vaccine Strains. *J. Infect. Dis.* **204,** S533–S548 (2011).

208. Plumet, S., Duprex, W. & Gerlier, D. Dynamics of viral RNA synthesis during measles virus infection. *J. Virol.* **79,** 6900–6908 (2005).

209. Brandler, S. & Tangy, F. Recombinant vector derived from live attenuated measles virus: Potential for flavivirus vaccines. *Comp. Immunol. Microbiol. Infect. Dis.* **31,** 271–291 (2008).

210. Griffin, D., Lin, W. & Pan, C. Measles virus, immune control, and persistence. *FEMS Microbiol. Rev.* **36,** 649–662 (2012).

211. Mateo, M. & Lopez, M. Nectine-4, une protéine clé pour la transmission du virus de la rougeole. *médecine/sciences* **28,** 363–365 (2012).

212. Noyce, R. & Richardson, C. Nectin 4 is the epithelial cell receptor for measles virus. *Trends Microbiol.* **20,** 429–439 (2012).

213. Kutty, P. *et al.* VPD Surveillance Manual, 6th Edition, 2013 Measles: Chapter 7-1 Chapter 7: Measles. *VPD Surveill. Man.* (2013).

214. Takeda, M. Measles virus breaks through epithelial cell barriers to achieve transmission. *J. Clin. Invest.* **118,** 2386–2389 (2008).

215. Markel, H. Koplik's Spots: The Harbinger of a Measles Epidemic. *Milbank Q.* **93,** 223–229 (2015).

216. Abad, C. & Safdar, N. The Reemergence of Measles. *Curr. Infect. Dis. Rep.* **17,** 51 (2015).

217. Xavier, S. & Forgie, S. Koplik spots revisited. *Can. Med. Assoc. J.* **187,** 600–600 (2015).

218. Koenig, K., Alassaf, W. & Burns, M. Identify-Isolate-Inform: A Tool for Initial Detection and Management of Measles Patients in the Emergency Department. *West. J. Emerg. Med.* **16,** 212–219 (2015).

219. Orenstein, W., Perry, R. & Halsey, N. The Clinical Significance of Measles: A Review. *J. Infect. Dis.* **189,** S4–S16 (2004).

220. Cole, A. Nervous Complications of Measles. *Br. Med. J.* **1,** 1361–1362 (1938).

221. Ulivelli, A., Casini Raggi, G. & Lecchini, L. [The neurological complications of measles.

(Clinical case studies with special reference to the pathogenesis of post-exanthematous encephalitis)]. *G. Mal. Infett. Parassit.* **19,** 203–207 (1967).

222. Griffin, D. Measles virus and the nervous system. *Handb. Clin. Neurol.* **123,** 577–590 (2014).

223. Rota, P., Rota, J. & Goodson, J. Subacute Sclerosing Panencephalitis. *Clin. Infect. Dis.* **65,** 233–234 (2017).

224. Schneider-Schaulies, S. & ter Meulen, V. Pathogenic aspects of measles virus infections. *Arch. Virol.* **15,** 139–158 (1999).

225. Rima, B. & Paul, D. Molecular mechanisms of measles virus persistence. *Virus Res.* **111,** 132–147 (2005).

226. Morgan, E. & Rapp, F. Measles virus and its associated diseases. *Bacteriol. Rev.* **41,** 636–666 (1977).

227. Rota, P. *et al.* Global Distribution of Measles Genotypes and Measles Molecular Epidemiology. *J. Infect. Dis.* **204,** S514–S523 (2011).

228. Durrheim, D., Crowcroft, N. & Strebel, P. Measles – The epidemiology of elimination. *Vaccine* **32,** 6880–6883 (2014).

229. Ison, M. Antiviral Treatments. *Clin. Chest Med.* **38,** 139–153 (2017).

230. Ortac Ersoy, E. *et al.* Severe measles pneumonia in adults with respiratory failure: role of ribavirin and high-dose vitamin A. *Clin. Respir. J.* **10,** 673–675 (2016).

231. Miyazaki, M. *et al.* Apparent response of subacute sclerosing panencephalitis to intrathecal interferon alpha. *Ann. Neurol.* **29,** 97–99 (1991).

232. Rota, P. *et al.* Measles. *Nat. Rev. Dis. Prim.* **2,** 16049 (2016).

233. Kowalzik, F., Faber, J. & Knuf, M. MMR and MMRV vaccines. *Vaccine* (2017). doi:10.1016/j.vaccine.2017.07.051

234. Bester, J. Measles and Measles Vaccination. *JAMA Pediatr.* **170,** 1209 (2016).

235. Palù, G. & Pacenti, M. AGGIORNAMENTO IN TEMA VACCINALE. (2017).

236. Santibanez, S. *et al.* Molecular surveillance of measles and rubella in the WHO European Region: new challenges in the elimination phase. *Clin. Microbiol. Infect.* **23,** 516–523 (2017).

237. Lehman, I. & Boehmer, P. Replication of herpes simplex virus DNA. *J. Biol. Chem.* **274,** 28059–28062 (1999).

238. Lester, J. & DeLuca, N. Herpes simplex virus 1 ICP4 forms complexes with TFIID and mediator in virus-infected cells. *J. Virol.* **85,** 5733–5744 (2011).

239. Daelemans, D., Pauwels, R., De Clercq, E. & Pannecouque, C. A time-of–drug addition approach to target identification of antiviral compounds. *Nat. Protoc.* **6,** 925–933 (2011).

240. Waterhouse, A., Procter, J., Martin, D., Clamp, M. & Barton, G. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).

241. De Clercq, E. *et al.* Antiviral activities of 5-ethynyl-1-beta-D-ribofuranosylimidazole-4-carboxamide and related compounds. *Antimicrob. Agents Chemother.* **35,** 679–684 (1991).

242. Krajczyk, A. *et al.* Antivirally active ribavirin analogues – 4,5-disubstituted 1,2,3-triazole nucleosides: biological evaluation against certain respiratory viruses and computational modelling. *Antivir. Chem. Chemother.* **23,** 161–171 (2013).

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

244. Ambrus, A., Chen, D., Dai, J., Jones, R. & Yang, D. Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization†. *Biochemistry* **44,** 2048–2058 (2005).

245. Hsu, S. *et al.* A G-Rich Sequence within the c-kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics. *J. Am. Chem. Soc.* **131,** 13399–13409 (2009).

246. 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.* **78,** 8002–8014 (2004).

247. Vere Hodge, R. & Field, H. Antiviral Agents for Herpes Simplex Virus. in *Advances in pharmacology* **67,** 1–38 (2013).

248. Pavić, I. *et al.* Flow cytometric analysis of herpes simplex virus type 1 susceptibility to acyclovir, ganciclovir, and foscarnet. *Antimicrob. Agents Chemother.* **41,** 2686–2692 (1997).

249. Zheng, K. *et al.* Inhibition of herpes simplex virus type 1 entry by chloride channel inhibitors tamoxifen and NPPB. *Biochem. Biophys. Res. Commun.* **446,** 990–996 (2014).

250. Takahama, K. *et al.* G-Quadruplex DNA- and RNA-Specific-Binding Proteins Engineered from the RGG Domain of TLS/FUS. *ACS Chem. Biol.* **10,** 2564–2569 (2015).

251. Resnick, J., Boyd, B. & Haffey, M. DNA binding by the herpes simplex virus type 1 ICP4 protein is necessary for efficient down regulation of the ICP0 promoter. *J. Virol.* **63,** 2497–2503 (1989).

252. Gao, M. & Knipe, D. Genetic evidence for multiple nuclear functions of the herpes simplex virus ICP8 DNA-binding protein. *J. Virol.* **63,** 5258–5267 (1989).

253. Bera, A., Perkins, E., Zhu, J., Zhu, H. & Desai, P. DNA Binding and Condensation Properties of the Herpes Simplex Virus Type 1 Triplex Protein VP19C. *PLoS One* **9,** (2014).

254. Harris, L. & Merrick, C. G-Quadruplexes in Pathogens: A Common Route to Virulence Control? *PLOS Pathog.* **11,** 1–15 (2015).

255. Agarwala, P., Pandey, S. & Maiti, S. The tale of RNA G-quadruplex. *Org. Biomol. Chem.* **13,** 5570–5585 (2015).

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

257. Song, J., Perreault, J., Topisirovic, I. & Richard, S. RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* **4,** e1244031 (2016).

258. Bozza, M., Sheardy, R., Dilone, E., Scypinski, S. & Galazka, M. Characterization of the Secondary Structure and Stability of an RNA Aptamer That Binds Vascular Endothelial Growth Factor†. *Biochemistry* (2006). doi:10.1021/BI0521442

259. Perrone, R. *et al.* The G-quadruplex-forming aptamer AS1411 potently inhibits HIV-1 attachment to the host cell. *Int. J. Antimicrob. Agents* **47,** 311–316 (2016).

260. Bacon, T., Levin, M., Leary, J., Sarisky, R. & Sutton, D. Herpes simplex virus resistance to acyclovir and penciclovir after two decades of antiviral therapy. *Clin. Microbiol. Rev.* **16,** 114–128 (2003).