UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

DIPARTIMENTO DI SCIENZE CHIMICHE

CORSO DI DOTTORATO DI RICERCA IN SCIENZE MOLECOLARI
CURRICOLO: SCIENZE FARMACAUTICHE
CICLO XXX

TWO COMPLEMENTARY STRATEGIES TO INHIBIT THE HIV-1 NUCLEOCAPSID PROTEIN

Coordinatore: Ch.mo Prof. Leonard Jan Prins

Supervisore: Ch.ma Prof.ssa Barbara Gatto

Dottorando: Elia Gamba

31 OTTOBRE 2017

ABSTRACT

This thesis work is devoted to the study of potential antiviral drugs to fight the human immunodeficiency virus type 1 (HIV-1). HIV-1 is responsible for the acquired immunodeficiency syndrome (AIDS), which represents a global concern affecting more than thirty-five million people worldwide. Since the discovery of the aetiological agent of AIDS in the mid-eighties, the pharmacological therapy had to deal with drug-resistant virus strains that are naturally selected by the available drugs as consequence of the inherent genetic variability of HIV-1. The current therapeutic approach, which focuses on the inhibition of key viral enzymes, must therefore employ concomitant administration of multiple drugs in order to suppress efficiently the infection. However, the high costs and the low compliance of these pharmacologic regimes represent major hurdles to the effectiveness of the therapy. In the past years, novel strategies able to interfere with non-canonical viral targets or host proteins have been proposed as possible venues toward the development of valuable drug candidates.

In this context, we focused on the HIV-1 nucleocapsid (NC), a small protein characterized by a cationic character and zinc fingers (ZFs) that confer nucleic acid chaperone properties. This protein is involved in many steps during the virus replication cycle, from the early reverse transcription to the late particle formation processes. In addition, NC functions rely on the strikingly conserved ZFs structure that has been observed to hardly mutate, differently from the validated drug targets exploited thus far. Therefore, NC represents a promising and novel pharmacological target to fight HIV-1, avoiding the problem of drug resistance.

Pursuing the discovery of inhibitors of the NC nucleic acid chaperone properties, in this research work we followed two complementary strategies. A direct approach aimed to the development of small molecules interacting with the ZFs domain of NC whereas an indirect approach employed binders of the nucleic acid substrates of NC. In the former approach, we studied the activity and mechanism of action of novel compounds able to bind NC, highlighting interesting hits that set the bases for the further development of this class of inhibitors. Series of 2,6-dipeptidyl-anthraquinone conjugates (AQs) were instead studied as intercalating agents able to freeze the dynamics of important viral nucleic acid structures and, in the process, inhibit indirectly the biological functions of NC. AQs have been extensively studied in our laboratory; my research activity contributed to explore the effect of the side chain orientation on the NC inhibitory activity observed for these compounds. In addition, we investigated the interaction of

AQs with novel viral RNA substrates, in order to prove the pleiotropic inhibitory effect exerted by these compounds on multiple steps of the virus replication cycle.

Altogether, the results achieved led us to propose a novel approach for the development of inhibitors of NC, which we think might represent a leap forward in the development of this potential therapeutic class to actual drugs.

RIASSUNTO

Questo lavoro di tesi è dedicato allo studio di potenziali farmaci antivirali per combattere il virus dell'immunodeficienza umana di tipo 1 (HIV-1). HIV-1 è responsabile dell'omonima sindrome da immunodeficienza acquisita (AIDS), un problema di ordine globale che interessa più di trentacinque milioni di persone in tutto il mondo. A partire dalla scoperta dell'agente eziologico dell'AIDS avvenuta alla metà degli anni ottanta, la terapia contro HIV-1 ha dovuto confrontarsi con virus farmaco-resistenti selezionate dai farmaci disponbili come conseguenza della variabilità genetica intrinseca di HIV-1. L'approccio farmacologico, che si concentra sull'inibizione di specifici enzimi virali, impiega la concomitante somministrazione di molteplici farmaci al fine di controllare efficacemente l'infezione. Tuttavia, i costi elevati e la scarsa compliance di questi regimi farmacologici rappresentano i maggiori ostacoli all'efficacia della terapia. Negli ultimi anni sono state proposte nuove possibili strade verso lo sviluppo di candidati farmaci promettenti, in grado di interferire con bersagli virali non-canonici o proteine della cellula ospite.

In questo contesto, ci siamo concentrati su nucleocapside (NC) di HIV-1, una piccola proteina cationica caratterizzata da due zinc fingers (ZFs) che conferiscono proprietà chaperoniche su acidi nucleici. Questa proteina è coinvolta in diversi punti del ciclo replicatvo del virus, a partire dalla retrotrascrizione fino al processo di formazione delle particelle virali. Inoltre, le funzioni di NC si basano sulle strutture ZFs che risultano altamente conservate e non tollerano mutazioni, diversamente dai bersagli canonici dei farmaci finora utilizzati. Pertanto, NC rappresenta un promettente e nuovo bersaglio farmacologico per combattere HIV-1, evitsndo così il problema della farmacoresistenza.

Con l'obbiettivo di scoprire nuovi inibitori delle proprietà chaperoniche di NC, in questo lavoro di ricerca abbiamo seguito due strategie complementari. Un approccio diretto è stato finalizzato allo sviluppo di piccole molecole interagenti con il dominio ZFs di NC, mentre un approccio indiretto impiegava leganti di acidi nucleici che sono substrati di questa proteina. Considerando l'appproccio diretto, abbiamo studiato l'attività e il meccanismo d'azione di nuovi composti in grado di legare NC, evidenziando interessanti molecole che possono rappresentare una nuova classe di inhibitori di NC con ulteriori potenzialità di sviluppo. Serie di coniugati 2,6-dipetidil-antraquinonici (AQs) sono stati invece studiati come agenti intercalanti in grado di bloccare la strutture di sequenze acidi nucleici virali e, in questo modo, inibire indirettamente le funzioni biologiche di NC. Gli AQs sono stati ampiamente studiati nel nostro laboratorio; la mia attività di ricerca ha contribuito ad esplorare le relazioni struttura-attività per l'inibizione di NC. Inoltre, l'interazione digli AQs con nuovi RNA virali è stata studiata al fine di dimostrare

l'effetto inibitorio pleiotropico esercitato da questi composti su diverse fasi del ciclo di replicazione del virus.

Complessivamente, i risultati raggiunti ci hanno portato a proporre un nuovo approccio per lo sviluppo di inibitori di NC, che pensiamo possa rappresentare un contributo allo sviluppo di questa potenziale classe terapeutica.

CONTENTS

1. INTRODUCTION

1.1 The human immunodeficiency virus type 1 and the acquired immunodeficiency syndrome

The human immunodeficiency virus type 1 (HIV-1) is a retrovirus of the Lentivirus genus, known to be the cause of the acquired immunodeficiency syndrome (AIDS) in humans. AIDS has been reported in the early 80s to describe the increased number of opportunistic infections and other malignity cases among the homosexual population; however, only in 1984 the syndrome was linked to the HIV-1 infection.[1]

HIV-1 infects immune cells exposing the CD4 receptor such as T-cells and macrophages (CD4+ cells) leading to depletion of cells and CD4-mediated functions and thus strong immunosuppression. AIDS emerge at advanced infection stages characterized by low CD4+ cells counts and concomitant opportunistic infections as well as others AIDS-related malignity. No cure completely eradicates the infection, which leads to death in a period ranging from 6 months to more than 15 years if not treated.[2, 3]

As the name *Lentivirus* (from the Latin: *Lente, "slow"*), AIDS is characterized by a long incubation time after the early virus infection burst and three different phases can be distinguished (**Figure 1.1**). During the acute primary infection, viral spreading occurs quickly and high viral loads can be observed in plasma weeks after contact with the virus. However, little depletion of CD4+ cells is observed and the virus is cleared by a strong immune response to levels where the virus is no longer detected in blood. In this asymptomatic phase, the virus remains latent mainly in resting T-cells that acts as a reservoir unaffected by the immune system and the antiretroviral therapy (ART).[4] Over the years after infection, chronic viral replication slowly depletes immune cells leading AIDS-related compliancy due to immunosuppression.[5]

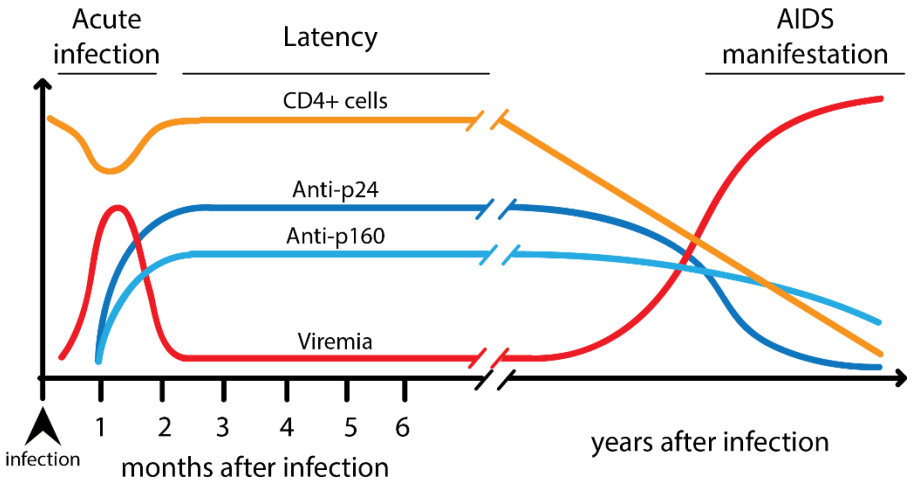


Figure 1.1. Correlation between virus production, immune response and number of CD4+ cells in a typical HIV-1 infected patient. During acute infection, HIV-1 blood titer increase rapidly and first symptoms appear such as fever, enlarged lymph nodes, weight loss, malaise, diarrhea, and oral candidiasis. Immune response suppresses viral spreading and neutralizing antibody against *gag* (p24) and *env* (p160) are detected. However, the virus is not completely eradicated and remains latent in resting T-cells during the latent phase. Chronic virus production and immunosuppression lead to strong depletion of CD4+ cells and AIDS-defining diseases such as opportunistic infections, lymphomas and central nervous system complication occurs. Adapted from Clark, S. J. et al., *N. Engl. J. Med.*, 324:14 (**1991**), 954-60.[5]

During late HIV-1 infection stages, strong depletion of CD4+ cells leads to characteristic complications collectively referred as AIDS-defining illness. Opportunistic infections (OIs) are the most frequent complications since many microbes can easily overgrow exploiting the lower immune system surveillance that usually controls their spreading. *Pneumocystis jiroveci* pneumonia, a fungal parasite of lungs that is usually symptomatic only during childhood, is frequently observed in patients with AIDS. Besides fungal infection, *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Mycoplasma pneumonia* and *Chlamydia pneumoniae*, are bacteria that also colonize the respiratory tract of immunosuppressed patients. The most common causes of diarrheal episodes in HIV-1 patients are non-typhi *Salmonella*, *Shigella* and *Campylobacter*, and *Clostridium difficile*, which are usually well controlled in healthy patients. Cryptosporidia and microsporidia are protozoan parasites that also commonly cause chronic diarrhea during advanced HIV-1 infection. Reactivation of latent cerebral toxoplasmosis and Cryptococcosis are dangerous parasitic and fungal infection occurring in the central nervous system of AIDS patients.[6]

Lymphomas represent other important AIDS-defining illnesses. The incidence of non-Hodgkin types such as diffuse large B-cell lymphoma, Burkitt's lymphoma and primary central nervous system lymphoma is dramatically 60 to 200-fold higher in patients with AIDS. Also the incidence Hodgkin lymphoma is substantially increased to 8 to 10-fold compared to healthy population, even if it is not considered AIDS-defining. AIDS-related lymphomas has been

linked to the concomitant immunosuppression and infection with oncogenic viruses such as Epstein-Barr and Kaposi's sarcoma related Herpes.[7, 8]

After infection, HIV-1 readily reaches the central nervous system; however, it is usually not replicative as long as no systemic infection occurs. Passing the blood-brain barrier within infected macrophages, HIV-1 colonize the brain that became a viral sanctuary protected from the pharmacological treatment. HIV-dementia that induces cognitive, behavioural and motor dysfunction, usually occurs in advanced AIDS. Chronic activation of infected macrophages into the central nervous system and dysregulated release of pro-inflammatory cytokines has been linked to the development of this type of dementia.[9]

Despite the numerous complications occurring in late HIV-1 infection stages, antiretroviral therapy has been shown to reduce risks associated with AIDS-related illness. Adopting correct timing in ART and antimicrobial therapy has greatly reduced the deaths related to opportunistic infections and by half the incidence of HIV-dementia;[9] however, an only little decrease in cancer morbidity is observed in ART-treated patients. Hence, the recent advances in ART permitted to increase life expectancy of patients with AIDS and shift the number of deaths to causes not directly related to HIV-1 infection.[10]

In the fact sheet 2016, the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimated 36.7 million people living with AIDS worldwide and around 2.1 million new HIV-1 infections in 2015. Considering the 1.1 million people died of AIDS-related illnesses in the same year, AIDS remains a worldwide global concern. Major incidence is observed in the sub-Saharan Africa, where access to ART is more difficult and HIV-1 pandemic strikes the most. About 75% of infected people are concentrated in the sub-Sahar Africa, South America and Caribbean areas whereas North America and Eastern Europe account for about 6.5%, highlighting a social concern regarding the disparity between developing and developed countries. Good results have been achieved passing from about 2.6% in the 2000 to 46% in 2015 of people accessing to ART worldwide; covering a similar percentage of patients in the different world areas permitted to drastically reduce the number of AIDS-related deaths in the post-ART era (**Table 1.1**).

2015 worldwide UNAIDS data on HIV	people living with HIV	people living with HIV (% relative to total 2015)	people accessing to ART	people accessing to ART (% related to geographical area)
Eastern and Southern Africa	19.0 million	51.8	10.0 million	52.6
Latin America and the Caribbean	2.0 million	5.4	1.1 million	55.0
Western and Central Africa	6.5 million	17.7	1.8 million	27.7
Western and Central Europe and North America	2.4 million	6.5	1.4 million	58.3
Total 2015	36.7 million	-	17.0 million	46.3

Table 1.1. Worldwide UNAIDS data on HIV infection for 2015. Data were taken from UNAIDS Fact sheet 2016. Source: http://www.unaids.org/

1.1.1 The HIV-1 genome

Discovery of the HIV-1 virus occurred thanks to the human immunodeficiency virus type 2 (HIV-2), a distantly related virus that is more genetically related to the simian immunodeficiency virus affecting Sooty mangabey monkey. The latter was observed to induce a syndrome similar to AIDS in Western Africa monkey hunters. A later genetic analysis set the origin of HIV-1 to isolated cases of blood-borne cross-species transmission of the chimpanzee immunodeficiency virus to humans. Hence, HIV originated from the simian immunodeficiency virus that occasionally infected humans and adapted to the new host specie creating a distinct phylogenetic tree.[1]

Due to the nature of isolated cross-species transmissions of HIV-1 origin, different viral genotypes are known today. Currently circulating viruses originated from three independent transmission events in the early 20th century that formed three main groups: *major* (M), *outlier* (O), and *nonmajor-nonoutlier* (N). Group M is the predominant circulating HIV-1 group and it has been divided into the subtypes or clades (letters) and sub-subtypes (numbers) A1, A2, A3, A4, B, C, D, F1, F2, G, H, J, and K, basing on genetic variation. Origin of such genetic variability in HIV-1 has been related to the low fidelity of viral polymerases that can lead to the production of million virus variants per day in each patient (see **Chapter 1.4**). Co-infection with different viral clades may also account for further genetic variability thanks to recombination events, leading to the formation of hybrid variants that are called recombinant forms.[11]

HIV-2 is less virulent compared to type 1 virus and usually does not leads to AIDS manifestation. HIV-2 is restricted to small Western African areas where infection incidence is even decreasing due to the spreading of HIV-1. For this reason, HIV-2 will not be considered in this work focusing on the virus type-1 that represents the cause of current AIDS pandemic worldwide.

HIV-1 genome is encoded by an RNA strand with positive polarity about 9.7kb in length (gRNA) and, as common in retroviruses, two strands are packed into each viral particle. HIV-1 gRNA could be directly translated into viral protein; however, first events after viral entry in the host cell is the reverse transcription of the gRNA into a complementary DNA genome (cDNA) by the viral enzyme reverse transcriptase (see **Chapter 1.4**). cDNA is later transported into the nucleus and integrated into host's genome where it is transcribed and translated by the cell's machinery. Translation initially produces spliced mRNA that encodes for different regulatory and accessory proteins whereas later, longer unspliced or partially spliced mRNAs are exported from the nucleus to the cytosol for translation of the large polyproteins characteristic of retroviruses and necessary for viral particle assembly.[12]

Even if linear, the structure of retrovirus genome is highly complex adopting high order structures with short- and long-distance base pair interactions that lead to the formation of hairpins and large loops (**Figure 1.2**). Several secondary structures have been recognized as direct or indirect regulatory elements that participate in gRNA genome dimerization and selection during viral particle formation, control of reverse transcription, transcription and translation processes. The two untranslated terminal regions (UTRs) appeared to be the most structured and conserved; however, secondary structured elements are also found in coding regions. Within the latter, hairpins are usually located at the level of linking sequences between the different domains of large polyproteins genes to introduce translational pausing sites helping the folding of the individual domains during protein synthesis. Other secondary structures such as RNA pseudoknots have been shown to guide transcription modulating the open reading frames or to isolate specific genes that thus remains unstructured and more prone to deletions and insertions events, responsible for surface antigens variability.[13]

Important known regulatory elements of the HIV-1 genome are:

- **Primer binding site** (PBS): located in the 5'-UTR, PBS RNA hairpin is involved in the reverse transcription initiation serving as annealing site for the $tRNA^{Lys3}$ primer.[14] The complementary (-)PBS and (+)PBS DNA sequences found in the HIV-1 DNA are involved in the plus strand transfer event during reverse transcription of the gRNA.[15] These sequences will be further discussed in the **Chapters 1.5.4** and in the **Appendix Chapter 7.1**;

- **Trans-activation response element** (TAR): found in the R region of UTR, TAR forms a hairpin structure that takes part, together with its complementary DNA (cTAR) sequence, to the minus strand transfer event during reverse transcription.[15] TAR RNA hairpin also enhance transcription of viral genes serving as the binding site for the trans-activator of transcription.[16] These sequences will be further discussed in the **Chapters 1.5.4** and **Appendix Chapter 7.2**;
- **Dimerization initiation site** (DIS): located downstream the PBS and upstream the major splicing donor, DIS adopts a stem-bulge-loop with a palindromic apical loop that mediates the dimerization of gRNA pairing with the DIS loop from another gRNA molecule.[17] For this reason, it is also referred as SL1 in the context of the packaging signal RNA. Thanks to the chaperone activity of NC, the formation of a DIS/DIS extended duplex participate in the maturation process of the viral particle stabilizing the gRNA dimeric form.[18] This sequence will be further discussed in **Appendix Chapter 7.3**;
- **Major splice donor site** (SD1): it is a stem-loop structure also called SL2 located downstream the DIS and upstream the SL3. SD1 represent the slicing site for the production of all spliced viral mRNA using different 3' acceptor sites (yielding more than 40 different coding transcripts) and it is important in maintaining the right balance in the expression of different HIV-1 proteins;[19]
- **Packaging signal sequence** (Ψ-RNA): collectively it encompasses four stem-loop structures (SL1-4) upstream the *gag* start codon in HIV-1. SL3 is selectively recruited by the NC domain of *gag* among the multitude of cellular mRNA during genome packaging into viral particle;[20]
- ***gag-pol* frame shift sequence**: it is a epta-nucleotide (UUUUUUA) slippery sequence located in the p6 coding region of the *gag* gene that promotes the shift of the open reading frame of -1 nucleotide. The new reading frame leads to the *gag-pol* transcript with a frequency of one *gag-pol* every 10-20 *gag* transcripts;[13]
- **Rev responsive element** (RRE): found in the *env* gene, RRE is a highly ordered stem-loop region that acts as a binding site for the Rev protein. RRE interaction with multiple Rev monomers promotes the export of larger unspliced and partially spliced viral mRNA from the nucleus to the cytosol where they serve for the synthesis or large polyprotein used during assembly of the viral particle;[21]
- **Polypurine tract** (PPT): in HIV-1 genome two purine-rich RNA sequences are resistant to the RNase H activity of reverse transcriptase. The central and 3'-polypurine tract are

used as primer for the synthesis of the plus DNA strand since they remain associated to the newly synthetized minus DNA strand after the degradation of other gRNA regions;[15]

- **Poly-adenylation signal** (PAS): for the correct translation, viral mRNA needs to undergo post-transcriptional modification similar to those of host's cell mRNA. These encompass the 5'-capping and 3'-poly-adenylation. HIV-1 possess poly-adenylation signals at both termini but the one located at the 3'-end is selectively used to produce 3'-polyadenylated transcripts.[22]

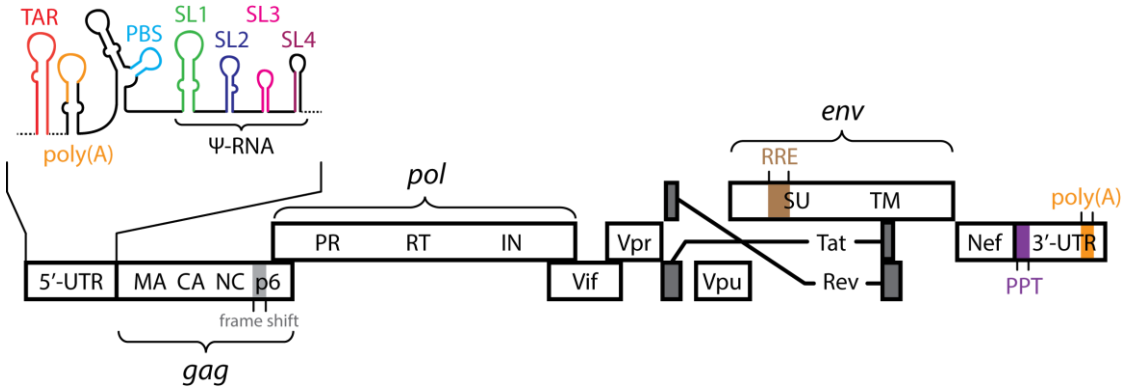


Figure 1.2. Schematic representation of gene organization into the HIV-1 genome and secondary structures regulatory sequences. 5'/3'-UTR are the terminal untranslated regions; MA, CA, NC and p6 are the matrix, capsid, nucleocapsid and p6 domains of the *gag* polyprotein, respectively; PR, RT and IN are the protease, reverse transcriptase and integrase domains of the *pol* polyprotein, respectively; SU and TM are the surface (Gp120) and transmembrane (Gp41) domains of the *env* (Gp160) polyprotein, respectively. Tat (trans-activator of transcription) and Rev (Regulator of the expression of the virus) are HIV-1 regulatory proteins; Vpu (Viral protein unique), Nef (Negative regulatory factor), Vif (Viral infectivity factor), Vpr (Viral protein R) are HIV-1 accessory proteins. A schematic representation of 5'-UTR structure and regulatory sequences is shown on top left. TAR, transactivation response element (red); poly(A), poly-adenylation signals (yellow); PBS, primer binding site (cyan), SL1: dimerization initiation site (also known as DIS) (green), SL2: major splice donor (also known as SD1) (blue); SL3: major packaging signal (pink); SL4: AUG start codon for *gag* open reading frame (purple), frame shift: gag-pol frame shift sequence (light grey), RRE: Rev response element (brown); 3'-PPT, polypurine tract (purple). See the "Genome Browser" tool at www.hiv.lanl.gov for an interactive map of the HIV-1 genome.

The HIV-1 genome encodes nine open reading frames and three of them are used for the production of the large *gag*, *pol* and *env* polyproteins. After catalytic activity of the viral protease, HIV-1 produces fifteen distinct proteins that can be divided into different groups basing on their function:

Envelope proteins that originate from *env* polyprotein (Gp160):[23]

- **Surface** (SU): SU is a 120 kDa glycoprotein (Gp120) exposed on the exterior of the viral membrane associated with TM and it represents the major surface antigen in the viral particle interacting with CD4 receptor;

- **Transmembrane** (TM): TM is a 41 kDa glycoprotein (Gp41) present in the viral membrane as a transmembrane protein. TM mediates interaction between the *env* and *gag* polyproteins during formation of the viral particle and also recruits SU protein at the budding site;

Structural proteins that originate from *gag* polyproteins:

- **Matrix** (MA): MA represents the N-terminal domain of *gag* and is responsible for its recruitment to the plasmatic membrane thanks to its basic N-terminal and three myristoylated residues. Onto the lipid bilayer, MA assembles as trimer where holes between adjacent trimers can locate the cytoplasmic C-terminal tail of TM protein to promote colocalization of *env* proteins to the budding site; [24]
- **Capsid** (CA): CA is the central domain of *gag* and form the characteristic cone-shaped capsid of mature HIV-1 particles after release by the viral protease;[24]
- **Nucleocapsid** (NC): NC is responsible for the interaction of *gag* with gRNA and is thus necessary during the genome packaging and viral particle formation. After release by viral protease, NC takes part in many important processes for viral replication such as particle maturation, reverse transcription and genome integration. NC will be further discussed in details being the focus of this thesis work;
- **p6**: p6 represents the C-terminal domain of *gag* and it is the binding site for the HIV-1 accessory protein Vpr, ensuring its incorporation into newly formed viral particles.

Enzymes that originates from *pol* polyprotein:

- **Protease** (PR): PR is a viral encoded dimeric aspartic-protease having a conserved Asp-Thr-Gly catalytic tirade on each subunit. PR is responsible for the cleavage of large precursor polyproteins at specific sites releasing the free independent domains after viral budding. For this reason, PR is highly important for the maturation of newly formed viral particles;
- **Reverse transcriptase** (RT): RT is a key enzyme for every retrovirus and gives the name to the *retroviridae* family. RT is responsible for the reverse transcription of the single-stranded gRNA to a cDNA genome having both DNA- and RNA-dependent 5' → 3' polymerase activity. RT is a heterodimer formed by the p51 and a p66 subunit that originate from the same viral transcript. Characterized by typical finger-palm-thumb structure with an additional connecting region, the p51 subunit is inactive whereas the p66 retains the polymerase activity and it also possesses RNase H activity able to degrade RNA/DNA hybrids;[15]

- **Integrase** (IN): IN is the enzyme responsible for the integration of the HIV-1 DNA into the host's cell genome. The 32 kDa HIV-1 IN has a zinc-finger domain involved in protein multimerization and a C-terminal DNA binding domain. The central domain is conserved among different retroviruses and possesses three characteristic residues (D64, D116, E152; DDE domain) that catalyses the transesterification reaction needed for integration of the cDNA.[25]

Regulatory proteins:

- **Trans-activator of transcription** (Tat): the promoter region located in the 5'-UTR has wake efficacy in recruiting the host's cell RNA polymerase II for transcription of the long viral mRNA. In absence of Tat, RNA polymerase terminate transcription at random levels downstream the TAR RNA hairpin. Tat serves as scaffold for transcriptional co-factors recruitment by binding to the TAR hairpin in the nascent mRNA. The TAR-Tat assembly increases the processivity of RT that thus can afford full genome transcription.[16] In addition, Tat has been related to several other functions useful for virus replication such as nucleic acid chaperone, mRNA capping and splicing and suppression of anti HIV-1 cellular RNA interference;[26-29]
- **Regulator of the expression of the virus** (Rev): Rev is the putative ligand of the RRE sequence and regulates the export of different viral mRNA from the nucleus to the cytosol for translation. During early infection stages, fully sliced mRNA lack the RRE and are exported to the cytosol in absence of Rev. After formation of the Rev protein, longer unspliced mRNA coding for large polyproteins are exported to the cytosol thanks to the interaction of RRE with multiple Rev monomers.[21]

Accessory proteins:[12]

- **Viral protein unique** (Vpu): the *env* polyprotein is produced in the endoplasmic reticulum and its migration to the plasmatic membrane may be prevented by premature interaction with a newly synthetized CD4 receptor in the host T cell. Vpr is an integral membrane protein that promotes degradation of CD4 thus aiding the export of envelope proteins to the membrane. Vpu also decreases the exposition of MHC-I receptors helping to escape the surveillance by the immune system;
- **Negative regulatory factor** (Nef): Nef is a protein associated with the membrane of endoplasmic reticulum having myristoylated N-terminal and, similarly to Vpu, it promotes the degradation of CD4 receptors. Interacting with Src-family kinases, which

regulates intracellular vesicular trafficking, Nef promotes the export of CD4 to lysosomes.

- **Viral Infectivity Factor** (Vif): Vif acts as down-regulator of the cellular APOBEC3G protein expression, a cytidine deaminase that introduces lethal dC → dU mutations into HIV-1 gRNA leading to the production of inefficient and unstable replicative intermediates if incorporated into viral particles;[30]
- **Viral protein R** (Vpr): introduced into viral particles thanks to the interaction with the p6 domain of *gag*, Vpr possesses nuclear localization sequences that promote nucleus import of the cDNA for integration. Vpr also induces cell cycle arrest at the G2 phase promoting resting of T-cells.

1.1.2 The HIV-1 virion structure

There are two main morphologically distinct forms of HIV-1, namely immature and mature particles (**Figure 1.3**). The latter is formed after virus budding through a process called "maturation". Maturation occurs thanks to the stepwise cleavage of polyproteins by the activity of PR. Maturation is necessary for viral infectivity as immature particles are not infective.[31] During such a process, a transition from the immature donut-shaped particle to a mature cone-like capsid is observed.[32]

Immature HIV-1 virion

In the immature virion, *gag* and *gag-pol* polyprotein form a protein lattice anchored to the viral envelope through their MA domain. gRNA forms a loose dimer through the loop-loop interaction involving DIS and remains bound to the protein lattice through the NC domain of *gag*. This disposition yields a characteristic donut shape with an electron dense area around the viral envelope and an electron poor core by electron microscopy observation. On the viral envelope, surface proteins protrude forming the viral spikes whereas in the particles interior are found only the accessory protein Nef, Vpr and Vif and other cellular components such as the $tRNA^{Lys3}$ primer.[23]

Mature HIV-1 virion

In the mature viral particle, only MA remains bound to the viral envelope together with the surface proteins whereas about 1500 CA monomers reassemble forming a fullerene-like capsid with characteristic hexameric architecture with pentameric declinations. The capsid assumes a cone shape, which forms an electron dense core by electron microscopy observation.[24] The dimeric gRNA form a stable nucleoprotein core coated by 1500-2000

molecules of NC, after the activity of PR.[33] During this process, the HIV-1 nucleocapsid plays an important role in the hybridization of the two gRNA strands by catalysing the structural conversion from the kissing-loop to the extended-duplex form of the DIS dimer (see **Appendix Chapter 7.3**). Together with other components, in the mature viral particles also the free essential enzymes are found released by the *gag-pol* polyprotein.[12]

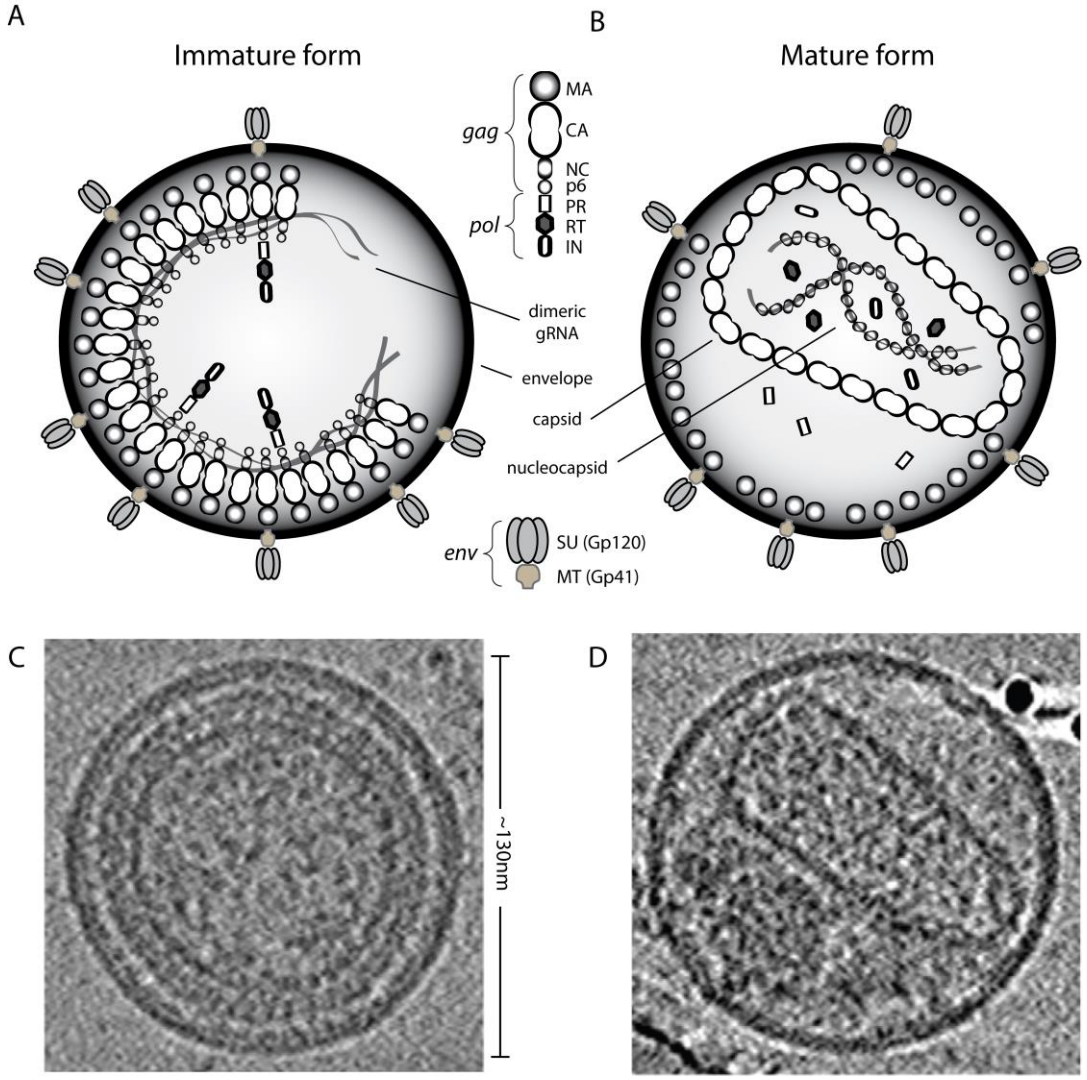


Figure 1.3. Schematic representation of HIV-1 virion. Schematic representation and Cryo-electron tomography images of the immature (A and C) and mature (B and D) virus morphology. Ganser-Pornillos, B. K. et al., *Curr. Opin. Struct. Biol*, 18:2 (**2008**) 203-217.[24]

1.2 The HIV-1 infection cycle

HIV-1 infection cycle starts with recognition and internalization of the viral particles by CD4+ cells and ends with budding and maturation of the novel viral particles. The processes occurring throughout the whole cycle can be distinct in early and late events. Starting from the virus entry until integration of the viral genome are considered early infection events whereas from transcription of viral mRNA to the viral particle formation characterize late stages of the infection cycle.

The entire HIV-1 infection cycle can be summarized by the following key steps (**Figure 1.4**):

1. Virus entry: Virus entry into cells is mediated by the viral envelope proteins and the cellular CD4 receptors and co-receptors. Therefore, HIV-1 can only infect CD4+ cells represented by T-cells and primary macrophages immune cells. Onto the viral envelope, Gp120 and Gp41 forms trimers of Gp120/Gp41 heterodimers where Gp120 mediates the interaction with the CD4 receptor. The latter event is thought to promote exposition of the binding site for the cellular co-receptor, usually a chemokine receptor.[34] Different HIV-1 lines use different co-receptor to penetrate the cell membrane and the CXCR4 and CCR5 represent the two major receptors, characterizing the lymphocyte-trophic (L-trophic) and monocyte-macrophage-trophic (M-trophic) viruses, respectively.[35, 36] Proper interaction between Gp120/CD4 and co-receptor leads to the exposition of Gp41 and its insertion into cellular membrane, promoting fusion of the viral and the cellular membranes.[37] The viral capsid is thus released into the cellular cytosol, where it disassembles in a process called uncoating;[38]

2. Reverse transcription: By capsid disassembly, the gRNA finally enter the cell cytosol and forms a reverse transcription complex (RTC) with different viral components, such as MA, CA, IN, RT and Vpr, that initiate reverse transcription. Once the retrotranscription of gRNA to cDNA is completed, the RTC becomes preintegration complex (PIC) characterized by directly repeated ends competent for integration called long terminal repeats (LTR). The PIC is imported into host cell nucleus thank to the interaction with microtubules and several nuclear localization signals, which confer to HIV-1 the capability to infect also nondividing cells;[39]

3. Integration: After the entrance of PIC into the cell nucleus, the viral DNA integrates into the host cells genome by the activity of IN. The integration process starts with the 3'-processing of viral DNA where IN recognizes specific sequences located into each LTR and cuts two nucleotides at the 3'-end leaving free oxydril groups. The second step of the integration process is the strand transfer. Into the PIC complex, IN catalyses attach of the two free 3'-OH simultaneously into the host cell genome, at a distance that is specific for different retroviruses. Such distance is very short and is 5-bp for HIV-1. Such mechanism leads to the

formation of 5'-nick into the host's genome that must be repaired by the cellular repair machinery, which copies the 5-nt sequence located between the two integration sites leading to the formation of two inverted repeats flanking the proviral DNA. The replication competent integrated HIV-1 DNA is referred as "*provirus*". How HIV-1 selects the integration site into host cell genome is not fully understood; however, highly transcribed regions (i.e., housekeeping genes) have been highlighted as preferred sites, probably to enhance efficiency of the viral replication;[25, 40]

4. Transcription of viral genome: The provirus is transcribed by the host cell RNA polymerase II forming a great variety of mRNA with different length. Early transcripts are spliced mRNA coding for several HIV-1 accessory proteins such as Tat and Rev. Differently, longer unspliced mRNA code for larger polyproteins and viral genome, necessary for viral particle formation are synthetized at late stages thanks to the trans-activation properties of Tat. The latter protein binds the TAR hairpin formed in nascent mRNA and recruits the positive-transcription elongation factor to phosphorylate the RNA polymerase II C-terminal domain increasing the transcription efficiency.[16, 41] The control of mRNA splicing thus represents a finely regulated process that controls the viral replication cycle;[19]

5. Translation of viral mRNA: After synthesis, the viral mRNA coding for proteins are usually processed as cellular transcripts introducing the 5'-cap and the 3'-poly(A) tail. These post-transcriptional modifications are necessary since the virus exploits host cell ribosomes for protein synthesis. Usually, free ribosomes translate *gag* and *gag-pol* mRNAs whereas *env* is processed into the endoplasmic reticulum to insert the envelope proteins SU and TM into the plasmatic membrane. Although the post-transcriptional modifications characteristic of human transcripts ensure translation of viral mRNA, HIV-1 RNAs also possesses several internal ribosomes entry sites, which guarantee translation in absence of the 5'-cap and poly-A;[42]

6. Virus assembly: After production, structural polyproteins such as *gag* and *gag-pol* migrate at the host cell membrane and colocalized into regions rich in phosphatidylinositol-(4-5)-diphosphate (which engages specific interaction with MA), cholesterol and sphingolipids (i.e., lipid rafts). The inherent affinity of *env* proteins for the lipid composition of these membrane regions, as well as the interaction with the MA domain of *gag* polyproteins, are the two major factors promoting colocalization of *env* and *gag* polyproteins. *Gag* selectively recruits two gRNA copies by the interaction of the NC domain with the packaging signals into the 5'-UTR.[20, 43] *Gag-gag* interaction through the CA and SP1 domains as well as coating of dimeric gRNA promotes the formation of the *gag* lattice onto the inner leaflet of the plasmatic membrane. This nucleoprotein assembly represents the site of viral budding and release;[23, 24]

7. Virus release and maturation: Interaction of the p6 domain of *gag* with proteins of the endosomal sorting complex is involved in the scission of membrane and release of the immature HIV-1 particle.[24] After release, the immature donut-shaped particle undergoes a complex mechanism of particle maturation to became fully infective, which has been already discussed in the **Chapter 1.1.2**.

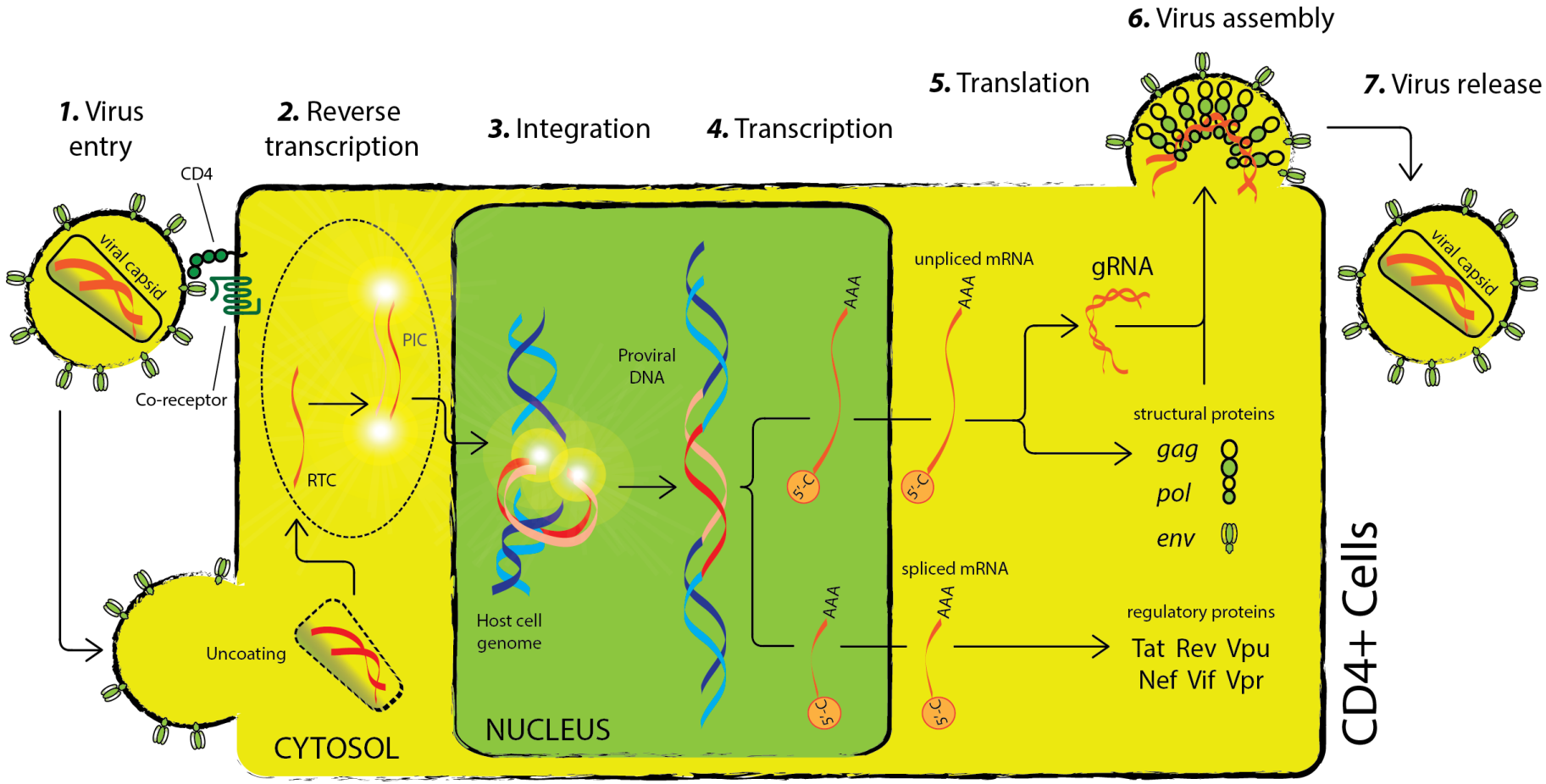


Figure 1.4. Schematic representation of HIV-1 replication cycle.

1.3 Antiretroviral therapy

There are 43 drugs used to treat infection by HIV-1 approved by the European Medicines Agency (source: http://www.ema.europa.eu/ema/, September 2017). Main classes are listed in **Table 1.2** and encompass nucleoside and non-nucleoside reverse transcriptase inhibitors (NRTIs and NNRTIs, respectively) and protease inhibitors (PIs). IN is another well-established pharmacological target and early antiretroviral drugs focused on the inhibition of these key viral enzymes. More recently, drugs targeting viral envelope proteins blocking either the virion entry or the fusion processes have been developed. Combinations of NRTIs, NNRTIs and PIs are the first antiretroviral drugs developed and remain the first line therapy against HIV-1.

Single-drug regimens usually fail to control HIV-1 infection and co-administration of drugs is more effective. As early as 1996, first highly active antiretroviral therapy (HAART) regimes were introduced combining two NRTIs plus one PI. NRTIs and PIs block early and late stages of the viral replication cycle respectively and this probably accounts for the lower chances to select drug-resistant strains.

Despite the plethora of available drugs against HIV-1 and the development of HAART, drug resistance remain a major concern in antiretroviral therapy. Development of drug resistance has been observed with virtually all drug classes. This phenomenon relies on the low fidelity of viral RT enzyme and the frequent recombination events occurring during reverse transcription (see **Chapter 1.4**). Single point mutations in the catalytic site or neighbouring residues in key viral enzymes are responsible for resistance of HIV-1 to NRTIs and NNRTIs, PIs and integrase inhibitors. Mutations in both Gp120 and Gp41 are responsible for resistance to fusion inhibitors such as enfuvirtide and entry inhibitors such as maravirock, even if the latter compound target CCR5 co-receptor which is a non-viral protein.[44]

Hence, antiretroviral therapy is currently focused on combination of multiple drugs to avoid selection of resistant viral strains. Although ART greatly improved the life expectancy of AIDS patients as discussed in **Chapter 1.1**, HAART is highly expensive, patient's compliance is low and adherence to the therapy remains one of the major cause of treatment failures.[45]

To overcome the hurdle of drug resistance, recent efforts focused on the validation of novel pharmacological targets to fight HIV-1. Several potential targets have been reported so far encompassing viral regulatory and accessory proteins and cellular co-factor components.[46] In this context, HIV-1 nucleocapsid is getting more and more interesting due to its striking conservation among different viral strains and the multiple roles played throughout all the viral replication cycle. Being the object of this thesis work, biological activity and potential novel drugs targeting NC will be discussed in details later on. (see **Chapter 1.5.5**)

An evolving field in the treatment and prevention of HIV-1 infection is the development of vaccines. If traditional methods delivering inactivated particles or subunits well worked for other viruses, this approach failed in the case of HIV-1. Neutralizing antibodies targeting envelope proteins are not sufficient as the virus escape immune response by mutation of surface antigens. Stimulation of cell-mediated immune response may represent a more valuable approach since has been shown that the initial viral burst decay is due to T-cell mediated immunity.[47, 48] Delivery of DNA vector to express viral antigens and *ex vivo* stimulation or reprogramming of T-cells against HIV-1 are all potential approaches to explore.[49] In the field of HIV-1 vaccines, it worth mention the good results recently achieved by Italian researchers using a Tat-based vaccine.[50]

Despite the multiple pharmacological approaches exploited against HIV-1 infection, yet no cure exists since none of them completely eradicates the infection. Antiretroviral therapy efficiently reduce viral loading in blood under level of detection (about <50 RNA copy/ml) greatly reducing the risk of transmission, but HIV-1 replicates continuously in long-living reservoir of memory T-cells at low rate.[4] For this reason, novel pharmacological targets and development of new antiretroviral drug is of primary interest, also to avoid selection of resistant strains by the chronic pharmacological regimes.

Medicine Name	Active Substance	Authorisation date
Multi-class combination products		
Atripla	efavirenz / emtricitabine / tenofovir disoproxil fumarate	13/12/2007
Descovy	emtricitabine / tenofovir alafenamide	21/04/2016
Eviplera	emtricitabine / rilpivirine hydrochloride /tenofovir disoproxil fumarate	28/11/2011
Genvoya	elvitegravir / cobicistat / emtricitabine / tenofovir alafenamide	19/11/2015
Odefsey	emtricitabine / rilpivirine hydrochloride / tenofovir alafenamide	21/06/2016
Stribild	elvitegravir / cobicistat / emtricitabine / tenofovir disoproxil fumarate	24/05/2013
Triumeq	abacavir sulfate / dolutegravir sodium / lamivudine	01/09/2014
Truvada	emtricitabine / tenofovir disoproxil fumarate	21/02/2005
Trizivir	abacavir (as sulfate) / lamivudine / zidovudine	28/12/2000
Nucleoside Reverse Transcriptase Inhibitors (NRTIs)		
Combivir	lamivudine / zidovudine	18/03/1998
Emtriva	emtricitabine	24/10/2003
Epivir	lamivudine	08/08/1996
Kivexa	abacavir / lamivudine	17/12/2004
Trizivir	abacavir (as sulfate) / lamivudine / zidovudine	28/12/2000
Ziagen	abacavir	08/07/1999
Nonnucleoside Reverse Transcriptase Inhibitors (NNRTIs)		
Edurant	rilpivirine hydrochloride	28/11/2011
Intelence	etravirine	28/08/2008
Stocrin/Sustiva	efavirenz	28/05/1999
Viramune	nevirapine	05/02/1998
Zerit	stavudine	08/05/1996
Viramune	nevirapine	05/02/1998
Zerit	stavudine	08/05/1996
Protease inhibitors (PIs)		
Aptivus	tripranavir	25/10/2005
Crixivan	indinavir sulphate	04/10/1996
Evotaz	atazanavir sulfate / cobicistat	13/07/2015
Invirase	saquinavir	04/10/1996
Kaletra	lopinavir / ritonavir	20/03/2001
Norvir	ritonavir	26/08/1996
Prezista	darunavir	12/02/2007
Reyataz	atazanavir sulphate	02/03/2004
Rezolsta	darunavir / cobicistat	19/11/2014
Telzir	fosamprenavir calcium	12/07/2004
HIV integrase strand transfer inhibitors		
Isentress	raltegravir	20/12/2007
Tivicay	dolutegravir	16/01/2014
Fusion inhibitors		
Fuzeon	enfuvirtide	27/05/2003
Entry Inhibitors - CCR5 co-receptor antagonist		
Celsentri	maraviroc	18/09/2007

Table 1.2. Antiretroviral drugs used in the treatment of HIV infection approved by European Medicines Agency. Equivalent and withdrawn drugs are not included. Source: http://www.ema.europa.eu/ema/; September 2017

1.4 The HIV-1 Reverse Transcription

As described in **Chapter 1.2**, reverse transcription is a crucial event of the infection cycle. Reverse transcription of viral genomic RNA starts after viral entry in the cell cytosol and may start before complete uncoating of the virus capsid (**Figure 1.4**). The product of the reverse transcription process is a double-stranded cDNA that integrates into the genome of the infected cell. The cDNA synthesis requires the RT as a key viral enzyme, which has both RNA-dependent DNA-polymerizing and RNase H (degradation of RNA in RNA/DNA duplexes) activities. The reverse transcription is of primary interest for the development of inhibitors of NC as most of its nucleic acid chaperone properties are required along this process. The entire process can be summarized in the following key steps (**Figure 1.5**):

1. The human $tRNA^{Lys,3}$, which is incorporated into the viral particle, acts as primer and binds the PBS at the 5'-UTR. After primer-annealing, RT polymerize a short DNA strand corresponding to the 180nt upstream the PBS reaching the 5'-end. This short DNA is called minus strand strong stop DNA ((-)ssDNA). As the (-)ssDNA synthesis proceeds, RT degrades the gRNA by its RNase H activity;
2. gRNA ends are direct repeats and for this reason the (-)ssDNA can transfer and anneal to the 3'-end. This event is known as minus strand transfer and involves either the same gRNA molecule or a different gRNA molecule, from the same virus or from a HIV-1 with different genotype, contributing to the genetic variability of the virus.[15] Minus strand transfer is necessary to complete polymerization of the (-)DNA strand and the annealing of (-)ssDNA to the 3'-UTR is regulated by the TAR hairpin structure in a process that require the activity of NC (see **Chapter 1.5.4** and **Appendix Chapter 7.2**);
3. After minus strand transfer, RT completes synthesis of the (-)DNA strand while the RNase H activity degrades the copied gRNA. During the latter process, the PPTs resist to the RNase H activity of RT and serves as primer for the synthesis of the plus strand strong stop DNA ((+)ssDNA). Between the two HIV-1 PPTs, the one located near the 3'-end is selectively used during the reverse transcription process;
4. Synthesis of (+)DNA starting from 3'-PPT extends to the tRNA primer coping the 18-nt sequence complementary to PBS. After this, the tRNA is cleaved and released leaving a single-stranded DNA having the same sequence of PBS ((+)PBS);
5. In a step called plus strand transfer, (+)PBS anneals with the complementary (-)PBS in the minus DNA strand. Starting from double-stranded (-)/(+)PBS, synthesis of the (+)DNA strand can be completed;

After reverse transcription, HIV-1 DNA genome results longer compared to the starting gRNA due to the LTR sequences and for a pre-integration complex competent for integration into host cell genome.

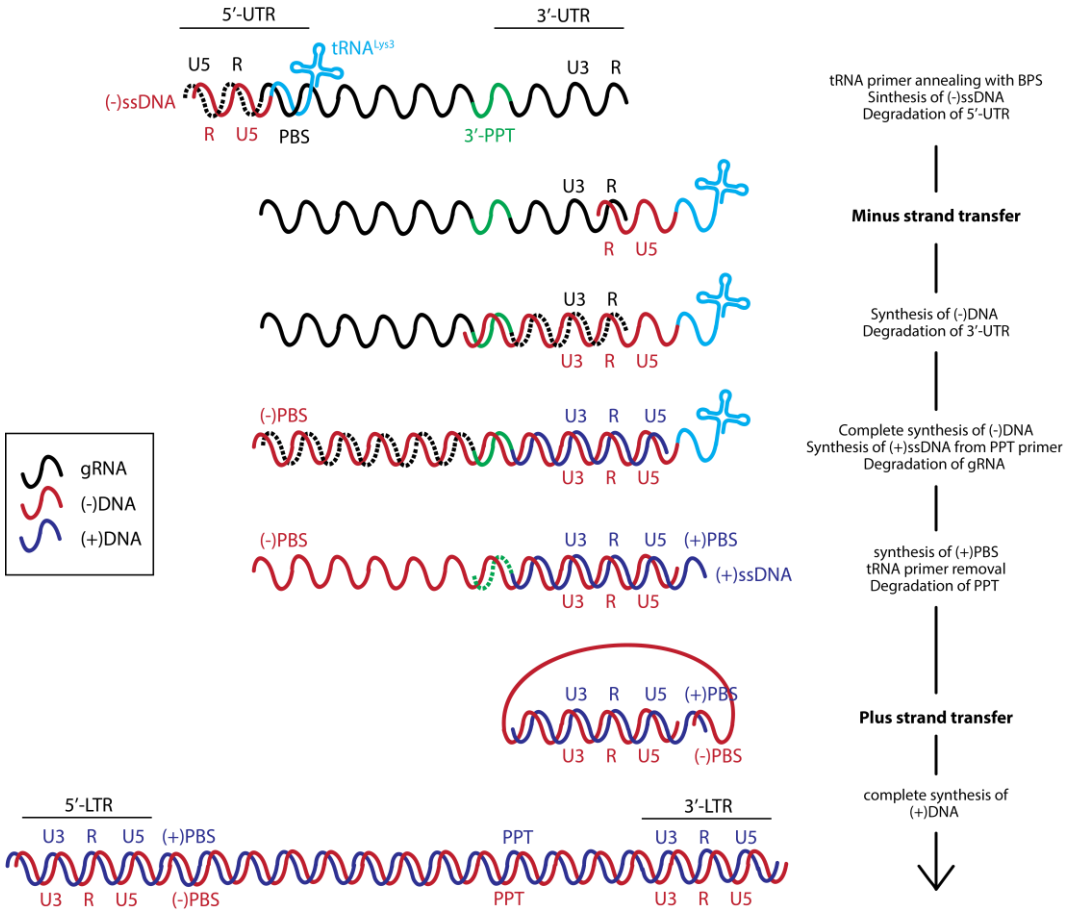


Figure 1.5. Scheme of the HIV-1 reverse transcription process.

Due to the reverse transcription mechanism, HIV-1 is inherently prone to mutate. Fidelity of RT during polymerization is low and spontaneously introduce mutations into gRNA due to the lack of proofreading activity. Expression of host cell factors such as APOBEC proteins also contributes introducing single point mutation into the provirus. In addition to these modifications at the level of the cDNA, errors may occur also during synthesis of the gRNA transcription as RNAP II also has no proofreading activity. Cumulative mutation rate has been estimated to be of one nucleotide every 50˙000 in each replication cycle.[39] In addition to single nucleotide mutations, the strand transfers contribute to recombination events where the (-)ssDNA anneals either to the same gRNA molecule or to a different one. Altogether, these

processes contribute to the genetic variability of the HIV-1 virus, and its ability to efficiently escape the pharmacological therapy.

1.5 The HIV-1 nucleocapsid protein

The HIV-1 nucleocapsid is a nucleic binding protein located into viral capsid where 1500-2000 copies coat and protect gRNA (**Figure 1.4**).[51] After release of the virus capsid into the newly infected cell, NC takes part in several key steps of the viral replication cycle, both in the early (from entry to integration) and late phases (from integration to genome packaging). For these reasons, NC can be considered both a structural element of viral capsid and an HIV-1 regulatory protein.

1.5.1 Structure of NC

NC originates from the *gag* polyprotein by the proteolytic activity of PR. In its mature form, NC is a 7 kDa 55-mer highly basic protein bearing ten lysine and five arginine residues.[52] The core domain of the protein is characterized by two highly conserved zinc finger (ZFs) motifs, each binding a Zn(II) ion with high affinity.[53, 54] Three cysteine and a histidine residues form the chelating motif (CCHC motif) of each zinc finger. Together with the aromatic residues in position 2 of each zinc finger, these positions are virtually invariant and no mutation are tolerated to retain the biological function of NC. The N-terminal and C-terminal zinc fingers (ZF1 and ZF2, respectively) have similar folding and are flanked by unstructured regions (**Figure 1.6A**).[55]

As shown by NMR, NC preferentially adopts a conformation in which the two ZFs are in close proximity due to a kink induced by the proline residue present into the linker region. This conformation allows stacking interaction between the two conserved residues Phe16 e and Trp37, located in position two of each zinc finger (**Figure 1.6B**).[55] However, recent studies also showed that NC has a highly flexible nature and ZFs experience a wide conformational variability being free to move almost independently.[56] When the two zinc finger are close to each other, a small hydrophobic pocket is formed by residues Val13, Thr24, Ala25, Trp37, Gln45 and Met46. This hydrophobic pocket plays a crucial role for nucleic acid chaperone properties of NC as it drives the specific interaction with unpaired guanine bases (**Figure 1.6C**).[20, 43, 57, 58]

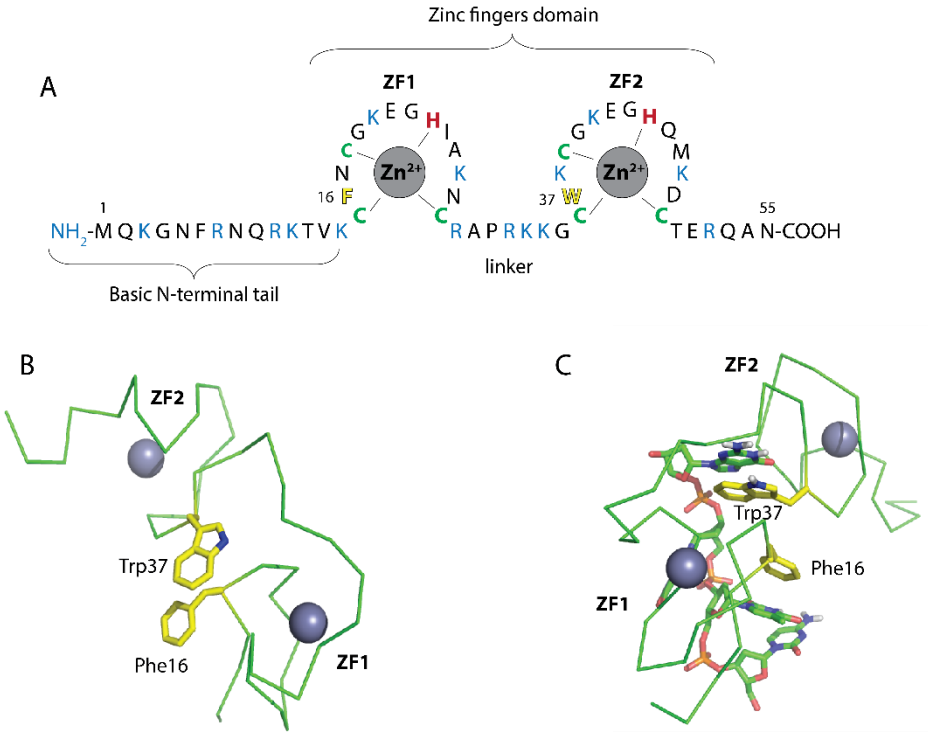


Figure 1.6. Structure of the HIV-1 nucleocapsid. NC primary sequence highlighting basic amino terminal, zinc fingers domain and linker sequence. Basic residues (blue), conserved aromatic residues (yellow) and zinc chelating cysteine (green) and histidine (red) (A). NMR structure of NC zinc finger domain free in solution showing the spatial proximity of the two zinc fingers and interaction between Trp37 and Phe16 (PBD entry: 1ESK) (B). Structure of NC in complex with the DNA oligonucleotide CTGG showing the staking interaction between aromatic residues and an unpaired nucleobases (PDB entry: 2L4L) (C).

1.5.2 NC as nucleic acid binding protein

NC binds a great variety of nucleic acids being able to interact with both RNA and DNA, both single and double-stranded sequences. Although this property reflects the fact that NC plays multiple roles throughout all the viral replication cycle, nevertheless how a single protein perform such different functions is still not fully understood.[59]

Binding of NC to nucleic acid is afforded by two main properties. Firstly, NC engages electrostatic interaction with the phosphate backbone by means of the several basic residues present in the amino terminal and linker regions. These interactions results quite unspecific and are responsible for the aggregating and charge screening properties of NC, enabling interaction with virtually any oligonucleotide sequence with low affinity.[60, 61] These properties makes NC an efficient RNA annealer[62, 63] and are useful in the context of viral capsid where this protein coats and give protection from physical and chemical injuries to the gRNA.[64] Secondly, the zinc fingers domain afford specific interaction with oligonucleotides by recognition of purine nucleobases.

Early studies with isolated ZFs showed that both electrostatic and hydrophobic interaction are involved in the interaction with guanine nucleobases into short oligonucleotides.[65] NC was shown to preferentially bind unpaired guanine bases placed into a sequence length of at least 5-nt.[66, 67] This binding specificity was shown to rely on the presence of the coordinated Zn(II) ions and the correct folding of each zinc finger. Among the zinc fingers residues, Trp37 have been shown to provide the highest contribute to the stabilization of the complex engaging direct stacking interaction with the base.[68] The dTdG DNA sequence was shown to represent a preferred motif for NC binding, irrespectively of its position in the context of a longer sequence.[69]

Correct folding of ZF1 have a high impact on overall structure and function of NC. Little mutation such as H23C, which retains the zinc binding ability of ZF1 but alter conformation of the peptidic backbone, leads to loss of NC biological activity.[70] Moreover, ZF1 appeared to affect NC binding stronger compared to ZF2 since little mutation in N-terminal ZF and NC mutants bearing two C-terminal zinc finger have lower affinity for nucleic acids.[61, 70] Order of the zinc fingers does not affect strongly NC binding to oligonucleotides and little affects biological functions of NC as observed by studies swapping the entire ZFs or the conserved aromatic residues. These results suggests that the two NC zinc finger acts as independent nucleic acids binding modules and that important residues are located in conserved positions in each ZFs.[61, 69, 71]

The first high-resolution NMR structure of NC in complex with a viral RNA substrate is the NC-SL3 assembly.[20] In line with previous studies, NC was shown to interact with nucleotides into the SL3 apical loop and aromatic residues were involved in staking interaction with guanine residues. Additional hydrophobic interaction was provided by zinc finger aliphatic residues and sugar the backbone of RNA (**Figure 1.7**).

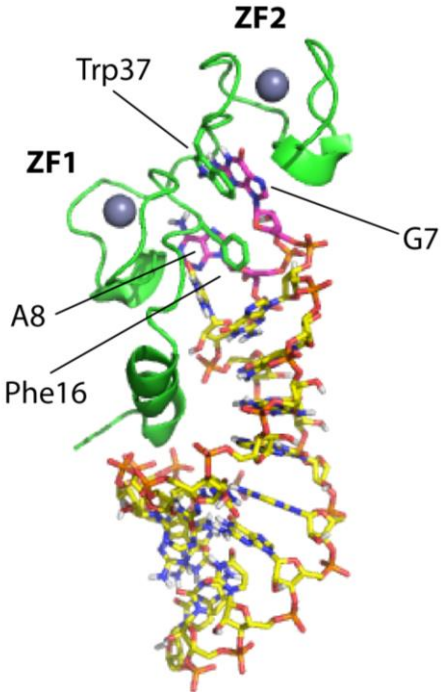


Figure 1.7. Structure of NC complexed to the SL3 from HIV-1. The picture shows the specific interaction engaged by the aromatic residues placed at position two of each zinc finger and purine nucleobases. NC is represented as green cartoon whereas SL3 is shown as stick with standard colours. Specifically, Phe16 from ZF1 and Trp37 from ZF2 of NC engages stacking interaction with A8 and G7 (pink coloured) of SL3, respectively. It worth noting that the purine bases in the SL3 apical loop interacting with NC are in a flipped out conformation. The basic amino terminal of NC fold in a 3_{10} helix structure interacting with the RNA major groove in the stem region of SL3 (PDB entry: 1A1T).

1.5.3 NC as nucleic acids annealer and chaperone

Early reports highlighted the ability of NC to catalyse refolding of metastable nucleic acid structures to more stable ones. NC was shown to promote strand transfer of complementary sequences favouring the formation of the duplex characterized by the higher number of base pairs and stability.[72] These properties are at the base of the many biological functions played by NC during important processes of the virus replication cycle.

Reflecting the specific and unspecific binding properties of NC, there are two main mechanisms by which this protein exerts its functions on nucleic acids. NC can be considered both a nucleic acid annealer and chaperone.[26] The high flexibility and basic character of NC are minimal requirements responsible for its nucleic acid annealing activity. A nucleic acid annealer catalyse annealing reaction between complementary sequences that usually occurs at slow kinetic rate due to electrostatic repulsion between the two complementary strands or the presence of metastable secondary structures. To overcome these barriers, nucleic acid annealers act through different mechanisms such as i) molecular crowding, increasing the apparent concentration of the two complementary strands; ii) charge screening, which shield the

electrostatic repulsion between the phosphate backbones of complementary sequences and iii) conformational selection of annealing-competent RNA forms among the structures already explored by the normal thermal conformational equilibria.[62, 63] The latter mechanism resemble the "entropy exchange" model proposed for other basic nucleic acid binding proteins.[73]

In addition to the nucleic acid annealer properties, NC also possesses duplex destabilization properties promoting melting of base pairs. NC does not promote complete melting of the nucleic acid secondary structure but acts lowering the unfolding energy barrier and base pairs cooperativity that hold the duplex structure together. This destabilization ability strictly relies on the interaction of specific purine nucleobases with the hydrophobic pocket located at the top of zinc fingers of NC. The contribution given by NC to the activation energy of unfolding has been calculated to be around -0.5 kcal/mol per base pair. Specifically, this value originates from two opposite contributes. The higher affinity of NC for single-stranded versus duplex oligonucleotides has been estimated to destabilize the duplex structure for -1 kcal/mol per base pair. On the opposite, the charge screening effect of NC basic residues stabilize the duplex structure for +0.5 kcal/mol per base of duplex stabilization due to reduction of phosphates repulsion.[74] The net effect is that NC interaction with a structured oligonucleotide do not promote direct base pairs melting but rather favour scrambling of base pairs and conformational sampling for the most stable structure. This specific nucleic acid destabilization property of NC is characteristic of nucleic acid chaperones. For this reason, chaperones can initiate conformational rearrangements that would not occur in absence of external stimuli, differently from protein having only nucleic acid annealer properties.[62, 63]

Although nucleic acid annealers and chaperones identify specific properties and activities, nevertheless they coexist within the same protein in the case of NC. These properties cannot be completely distinguished but rather together participate to the different biological functions of NC. However, the different properties may prevail in different contexts based on the different nucleic acid substrate as will be discussed in the **Appendix** discussing of different key viral nucleic acid sequences that regulate important steps of the virus replication cycle.

1.5.4 Roles of NC protein during the HIV-1 reverse transcription

Along with the structural functions of NC within the HIV-1 capsid, NC is necessary during several steps of the reverse transcription process occurring within the infected cell. Several steps involving NC have been reproduced *in vitro* by using model oligonucleotide constructs in order to study the activity of NC in different biologically relevant contexts. What emerged from these studies is that the mechanism of action of NC changes adapting to the different oligonucleotide substrate and function during different step of the virus replication

cycle. In this paragraph, we present an overview of the activity of NC in three different steps of the reverse transcription process that have been extensively studied, showing how the same protein behave differently in different contexts. This studies have a direct impact on the search of potential inhibitors of NC as the same model systems used to study the activity of NC can be used to probe the inhibitory activity of compounds.

The initial step of the reverse transcription process is the annealing of the $tRNA^{Lys,3}$ to the PBS RNA. During this step, RT polymerize the (-)ssDNA starting from the duplex formed by the PBS and the 3'-acceptor stem of the tRNA. PBS fold in a stable hairpin structure with four base pair stem and a penta-loop and for this reason annealing to the tRNA primer occurs at slow rate. Presence of NC strongly accelerate the process enabling efficient initiation of reverse transcription. In this context, NC has been suggested to act as an RNA annealer by means of it molecular crowding and charge screening properties. The zinc finger and the nucleic acid destabilization ability of NC appeared to play a minor role,[74, 75] differently to what reported for other steps of the reverse transcription process. This information was used in this work to design a novel assay to assess the inhibitory activity of compounds on the HIV-1 primer-annealing step and on the RNA annealing activity of NC. The reader is referred to the **Chapter 7.1.1** in the **Appendix** and reference therein for further details on the NC-mediated primer-annealing step.

The two obligatory strand transfer are other events of the reverse transcription process that have been intensively studied. During the minus strand transfer, the cTAR DNA sequences in the newly synthetized (-)ssDNA must transfer to 3'-UTR and anneal with the complementary TAR RNA in order to complete the synthesis of the (-)DNA. Without external stimuli, the process is highly inefficient since both TAR and cTAR form stable hairpin structures that impair annealing. The chaperone activity have been shown to strongly catalyse the minus strand transfer by promoting fraying of the stem region of TAR and cTAR structures.[15] Several biophysical studied involving isolated TAR and cTAR hairpin structures showed indeed that NC strongly activate the annealing of these two complementary sequences through a zippering mechanism involving the hairpins stem. The NC-mediated annealing is in contrast with the slower mechanism observed in absence of NC protein, which involves a loop-loop interaction. Such process is strongly dependent on the nucleic acid destabilization ability mediated by the ZFs of NC, Different to what described earlier for the NC-mediated primer-annealing. By using these pieces of information, the NC-mediated TAR/cTAR annealing reaction have been used in our laboratory to develop an electrophoretic assay that enabled the identification of several compounds with inhibitory activity on the chaperone properties of NC.[76-80] The reader is referred to the **Chapter 7.2** in the **Appendix** and reference therein for further details on the NC-mediated TAR/cTAR annealing.

The plus strand transfer involves PBS(+) and PBS(-) complementary sequences located in the plus and minus DNA strand, respectively. Similarly to the minus strand transfer, annealing of the two complementary sequences is hampered by the secondary structure of the PBS(-). Plus strand transfer thus requires the chaperone activity of NC. Similarly to what observed for the TAR/cTAR annealing, NC catalyses annealing of the PBS(-) to the complementary PBS(+) by activating alternative pathways not explored in absence of the protein. However, in this case NC has been shown to switch the pathway from a zippering mechanism involving the terminal overhangs to a kissing loops recognition. The reader is referred to the **Chapter 7.1.2** in the **Appendix** and reference therein for further details on the NC-mediated PBS(+)/(-) annealing.

Despite the apparent similarity of different viral nucleic acid substrates of NC, these examples demonstrate that different properties of NC are involved during different contexts and that its activity changes adapting to the specific substrate. The molecular crowding and charge screening properties of NC are the major requirement for the primer-annealing reaction whereas the nucleic acid destabilization property result necessary in both the strand events. Among the latter processes, NC showed to catalyse the annealing through opposite annealing pathways in different contexts (i.e., zippering versus kissing-loops recognition), strongly supporting the flexible nature of NC.

1.5.5 The HIV-1 nucleocapsid protein: a novel pharmacological target

In early 90', several reports emphasized the important role of nucleocapsid protein for the HIV-1 replication cycle. Mutants impairing NC zinc finger structure, as well as basic residues, resulted in the formation of non-infectious viral particles due to gRNA packaging deficiency despite having "mature" morphology (i.e., they retain protease activity).[70, 81-83 84] Supported by these findings, impairment of NC activity was proposed as potential novel pharmacological strategy to fight HIV-1 in the mid-90'.[85]

There are two main reasons that make NC an appealing novel pharmacological target. The first is the strict conservation of key residues within the NC sequence, which is thought to prevent the development of drug resistance while representing a major concern for current validated antiretroviral pharmacological targets (**Figure 1.8**). Secondly, the multiple roles played by NC can potentially block the viral replication cycle at different stages with fewer chances for the virus to escape the therapy.[86] However, many efforts have been done to target HIV-1 nucleocapsid but no drug reached the market yet. The lack of a valid drug candidate suggests that there is still need to explore the chemical space of NC-nucleic acids interaction,

which represents the aim of the this thesis work focused on the identification and study of novel chemotypes targeting NC.

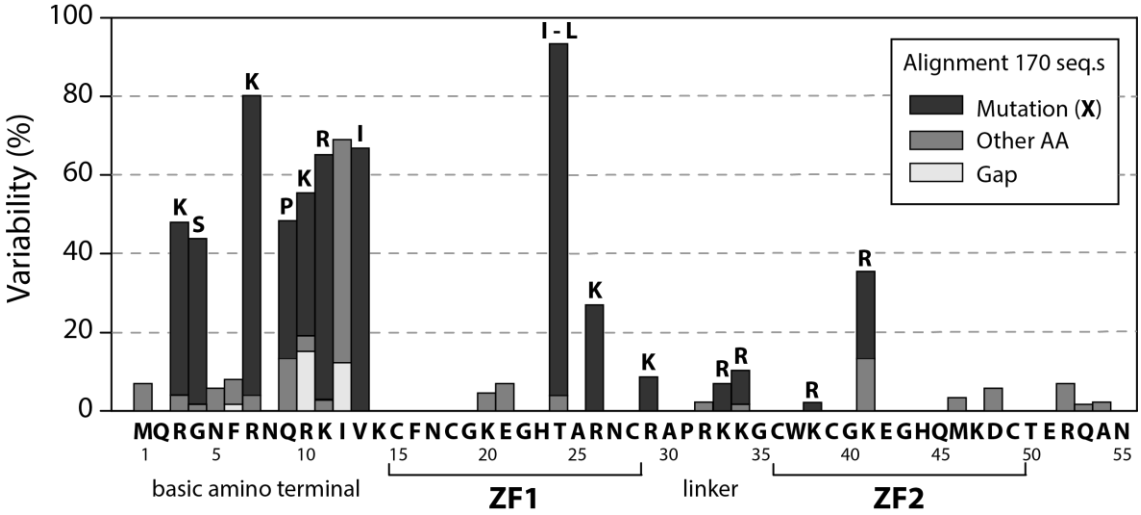


Figure 1.8. Alignment of NC sequence from 170 HIV-1 strains. Plot showing the variability of each amino acid residue highlighting gaps in the sequence (Gap), mutation to a generic (Other AA) or a specific (Mutation (**X**), as indicated on top of each column) residue; in the latter case the residue is indicated on top of the relative bar. The zinc fingers domain is highly conserved whereas mutations cluster into the basic amino terminal. Note how mutations appear to conserve the chemical features in each position. Source: http://hivmut.org/index.php

The first strategy followed to inhibit the biological functions of NC is the ejection of the chelated zinc. However, the specificity for the viral zinc fingers represented a major hurdle to develop further this class of compounds. Other strategies explored the feasibility of competitive inhibitors by creating oligonucleotide decoys or peptides mimicking NC. More recently, the development of small molecule that directly binds NC within the NC zinc finger domain to compete for nucleic acid binding represent the most exploited approach for the development of a potential drug candidate. To inhibit NC indirectly, nucleic acid binders of the substrates of NC represent another valuable approach to inhibit the biological functions of this viral protein.

Covalent zinc ejecting agents

First compounds reported as inhibitors of NC were molecules that covalently bind the zinc finger domain of NC. Alkylating agents are generally electrophile molecules that, reacting with the chelating thiolate groups of cysteines promoting the ejection of the divalent ion from the zinc fingers of NC. Loss of Zn(II) lead to unfolding of the zinc fingers and impairment of most biological functions of NC. These class of compounds showed to interfere with genome packaging and to promote accumulation of unprocessed cross-linked *gag* polyproteins, due to exposition of cysteine residues. These compounds appeared optimal virucidal agents as the

production of non-infectious virus particles have been observed in cell-based assay as well as loss of infectivity in newly synthetized virions.

Early molecules developed were the nitroso compounds NOBA and NOBP, firstly identified as inhibitors of the poly(ADP-ribose) polymerase, a cellular enzyme that also present a CCHC zinc finger motif. These compounds showed to effectively reduce infectivity of the viral particles but suffered of low specificity.[87] Similarly, disulfide substituted benzamindes (DIBAs, **Figure 1.9A**) also showed to covalently attack the zinc finger of NC promoting formation of intra and inter strand crosslinking through disulfides. A promising aspect of these early compounds was that they showed activity against a wide panel of HIV-1 strains and that the virus failed to develop resistance against these compounds even after several infection rounds.[88] Azodicarbonamide (ADA, **Figure 1.9A**) was later shown to be more selective for the viral zinc fingers of NC having virucidal activity on isolated particles similarly to DIBAs, even if direct alkylation of NC by ADA was not demonstrated. ADA has been the first molecule reaching phase II clinical trial in Europe.[89] Starting from DIBAs, structure-activity relationships optimization led to the identification of three pyridinioalkanoyl thiolesters (PATEs, **Figure 1.9A**) as promising molecules acting as covalent zinc ejecting agents able to impair the activity of the *gag* precursor and NC into newly formed viral particles.[90] Rational drug design recently lead to the bio-isoteric substitution of the disulfide functionality in DIBAs to obtain selenium-containing compounds (DISeBAs, **Figure 1.9A**), which mimic the mechanism of action of the precursor compounds with improved potency.[91]

Among the zinc ejectors group, the S-acyl-2-mercaptobenzamide thioesters (SAMTs, **Figure 1.9A**) represent nowadays the most promising class of compound.[92-97] SAMTs alkylate cysteine and lysine residues in the NC C-terminal zinc finger by trans-acetylation reaction.[98] In addition, these compounds where shown to act as prodrug when tested in living cells since re-acetylation of the drug occurred by cellular enzymes.[99] Among SAMTs, the molecule MDH-1-38 and its prodrug NS1040 yielded promising results in a preclinical study showing activity against different drug-resistant HIV-1 strains with low toxic effect in rats and to possess synergic effect together with other approved antiretroviral drugs in both *in vitro*[100] and *ex vivo*[101] models. Encouraged by these promising results, novel SAMTs compounds have been screened in order to afford structure-activity relationships on the recognition of the zinc finger of NC and to offer a greater variety of molecules for future clinical studies.[102]

Non-covalent zinc ejecting agents

Besides molecules able to eject the zinc ions by covalently attack chelating residues, compounds able to extract zinc without alkylation of the protein have been also reported. N,N-

bis(4-ethoxycarbonyl-1,2,3-thiadiazol-5-yl)benzene-1,2-diamine (NV038, **Figure 1.9B**) and 2-methyl-3-phenyl-2H-(1,2,4)thiadiazol-5-ylideneamine (WDO-217, **Figure 1.9B**) are two examples of zinc chelating agents having antiviral activity similarly to the covalent zinc ejecting agents but that do not engage covalent bond with NC.[103, 104]

Nucleomimetics and aptamers

Another strategy explored to inhibit NC function has been the use of nucleomimetics and aptamer decoys. Peptidic dinucleotides showed to compete with other nucleic acid for binding to NC. Not surprisingly, the TG motif appeared the most active compound in *in vitro* binding assays (T*G, **Figure 18C**).[105] RNA aptamers with nanomolar affinity for NC similar to that of SL3 sequence have been selected by SELEX and they showed to impair genome packaging when constitutively expressed in infected cells.[106, 107] Although these reports showed the effect of nucleic acid-based molecules against NC, no study in cells have been reported and the use of these compounds *in vivo* remains challenging due to problems related to the stability of these compounds in living organisms.

Peptides

Different to nucleic acid mimetics, peptides that mimic NC binding to its nucleic acid substrates have been proposed. NC-analogue peptides has been selected by phage display technique to bind the Ψ-site and antagonize selection of the gRNA by NC.[108] These selected peptides were short sequences rich in tryptophan residues and later showed to binds different stem-loop structures during reverse transcription such as TAR and PBS, inhibiting the NC nucleic acids destabilization activity.[109] The hexapeptide HKWPWW showed to reduce the production of virus particles when constitutively expressed in HIV-1 infected cells.[110]

Small molecules targeting the NC zinc finger domain

The most recent and explored approach to inhibit NC functions employs small molecules modulators of the NC-nucleic acid interaction, targeting the hydrophobic pocket located at the top of the zinc fingers domain, without ejection of zinc to avoid unwanted toxicity. Large chemical libraries have been extensively screened searching for compounds fitting the small and flexible hydrophobic pocket of NC. A first study identified a tetrachlorogallein compound able to inhibit NC-nucleic acid interaction and also displayed antiviral activity (tetrachlorogallein, **Figure 18D**).[111] Using a FRET based assay to probe the ability of NC to destabilize the structure of cTAR, five hit compounds all showing similar catechol chemical substructure has been selected among 4800 molecules (A10, **Figure 18D**).[112]

Trying to avoid the potentially toxic catechol moiety, compound A10 has been later optimized achieving a 2-amino-4-phenylthiazole valuable hit compound (AN3, **Figure 18D**).[113] The preselection of compounds by virtual screening thus turned out to be a valuable approach for the identification of novel inhibitors. By using this approach, two compounds with antiviral activity in the micromolar region and able to impair NC chaperone activity were selected among large chemical libraries encompassing more than three-hundred-thousand compounds (8, **Figure 18D**).[114] Other hit compounds were identified by screening of a 14400 compounds library by means of a fluorescence polarization assay involving a GST-p2-NC construct and a fluorescein labelled SL2 oligonucleotide. In this study the compounds showed to reduce level of viral particle formation from HIV-1 transfected cells with EC_{50} in the low micromolar region (CMPD1, **Figure 18D**).[115] In addition, natural products with NC-inhibitory activity have been discovered in our laboratory. We demonstrated that the polyphenolic compound acutissimin, found in oaked-barrel aged wine, inhibits the NC chaperone activity *in vitro* by direct binding to the NC protein.[77]

The molecule developed by Boehringer Ingelheim laboratories (Canada), who solved the NMR structure of a 2:1 inhibitor-NC complex, represents a breakthrough in the field. Starting from a capsid assembly assay, imino-oxindole derivatives showed to impair the formation of capsid lattice onto TG repeats DNA by interaction with the NC domain of *gag*. The two molecules showed to bind NC in a cooperative manner where one molecule serves as scaffold for binding of the second one. This peculiar mechanism of action confer high potency and specificity to these compounds (1, Gourdeau *et al., 2013*, **Figure 18D**).[116]

Taking together, these results shows that molecules characterized by aromatic rings to engage staking with Phe16 and Trp37 residues of NC, as well as hydroxyl groups able to mimic the hydrogen bonds network engaged by purine nucleobases, represent basic structural requirements to target the hydrophobic pocket formed by the zinc fingers domain of NC.[117]

Nucleic acid binders

The search for small molecules inhibitors of NC led also to the identification of binders of its nucleic acid substrates. Several molecules interacting with specific RNA and DNA viral sequences has been studied and the Ψ-site has been one of the most exploited substrate to this purpose. An early screening of more than 200˙000 molecules yielded no inhibitors of the NC-induced maturation of gRNA dimer (i.e., kissing-loop to extended duplex transition of the DIS), but rather lead to the identification of an activator molecule able to chaperone the maturation *in vitro*, mimicking the activity of NC.[118] The molecule identified was characterized by a cationic dipeptide bound to a methylcoumarin moiety and was shown to bind the SL1 stem-loop

engaging both electrostatic and staking interaction, especially in the region of the internal bulge (RR-AMC, **Figure 1.9E**). Such molecules was proposed to induce premature gRNA maturation into the cell cytosol instead of the budding particle.[119] On the opposite, aminoglycoside antibiotics were found to bind SL1 with sub-micromolar affinity preventing conversion of the SL1 dimer from the kissing-loop to extended duplex conformation. Although no NC-inhibitory activity was assessed and no antiviral activity was observed, aminoglycoside binding to the SL1 dimer was proposed as model system for the development of nucleic acid binders specific for this viral RNA.[120-124]

Within the Ψ-site, SL3 is a binding site with high affinity for NC and has been used as model substrate for the development of nucleic acid binders interfering with the genome packaging process. Compounds able to bind SL3 *in vitro* and to compete for NC binding have been identified, despite being not tested for antiviral activity (3, **Figure 18E**).[125] The quinoline derivate NSC260594 was shown to compete for NC binding to the apical loop of SL3 and also showed to reduce infectivity and viral particle release from HIV-1 infected cells (**Figure 18E**).[126] Another class of compounds reported to inhibit the NC chaperone activity *in vitro* are dipeptidyl-anthraquinone compounds able to intercalate dynamic stem-bulge-loop structures of TAR and cTAR, which will be further discussed in this thesis work.[76, 78, 79]

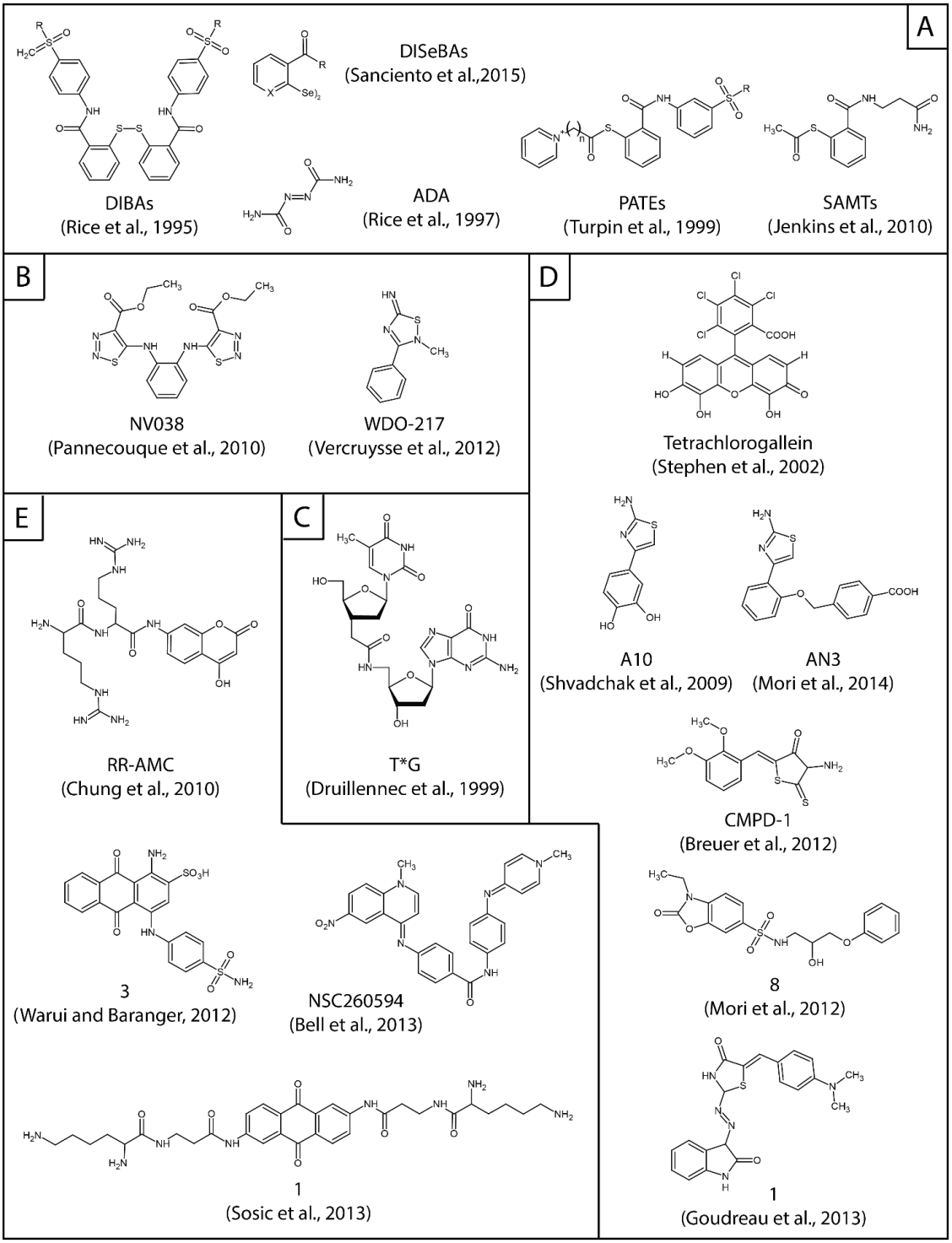


Figure 1.9. Classes and chemical structure of inhibitors of NC reported in the literature. Names are kept as the same as those reported in the original paper. Covalent zinc ejecting agents (A); non-covalent zinc ejecting agents (B); nucleomimetics (C); small molecules targeting the zinc finger domain of NC (D); nucleic acid binder inhibitors of NC (E).

2. AIM OF THE WORK

TWO COMPLEMENTARY STRATEGIES TO INHIBIT THE HIV-1 NUCLEOCAPSID PROTEIN

This research work is devoted to the discovery and study of small molecules inhibitors of the HIV-1 nucleocapsid protein as novel antiretroviral agents. We investigated two different approaches to inhibit NC by studying molecules which are either able to directly bind the protein or behave as intercalating agents targeting the nucleic acid substrates of NC, thus interfering indirectly with its activity.

To target NC protein directly, we studied molecules directed to the ZFs domain, which represents the most conserved and structured region of the protein. The residues Val13, Thr24, Ala25, Trp37, Gln45 and Met46 from both zinc fingers form a hydrophobic pocket at the top of the two zinc fingers that accommodate guanine nucleobases and this interaction is responsible for the nucleic acid destabilization properties of NC (see **Chapters 1.5.1-3**). Potential inhibitors were tested for their ability to compete with the binding of purine nucleobases to this site and to inhibit the chaperone properties of NC.

Identification of a promising lead NC inhibitor represents a challenging task due to the high flexibility of the protein structure, the high solvent exposition of the binding pocket and the lack of extended structural information on known protein-nucleic acid complexes. No defined pharmacophore model has been proposed yet, thus lacking a structural consensus that defines this potential class of antivirals.[127] Therefore, there is an important need to explore the chemical space and identify those molecular determinants necessary to maximize the compound-NC binding affinity and potent inhibitors. Part of this thesis work is devoted to addressing these specific needs, by investigating the NC inhibitory activity and the mechanism of action of two different libraries of small molecules (BN and BM series), originating from a virtual screening of commercially available compounds predicted to bind the hydrophobic pocket of NC (**Figure 2.1** and **2.2**).

To inhibit NC indirectly, we employed molecules targeting its nucleic acid substrates. Acting as nucleic acid chaperone, NC catalyses refolding of metastable secondary structures such as hairpin to duplex conversion and other annealing reactions that involve different viral nucleic acid sequences (see **Chapters 1.5.4** and **Appendix-Chapter 7**). By over-stabilization of nucleic acid secondary structure, nucleic acid binders may prevent the NC-mediated refolding necessary to important viral processes such as reverse transcription and genome dimerization and packaging.

In our laboratory, series of intercalating agents having an anthraquinone core have been studied extensively as inhibitors of NC acting by freezing the secondary structure of nucleic acids (**Figure 2.3**). We designed and tested different libraries focused on the anthraquinone scaffold exploring the structure-activity relationships (SARs) for the optimal interaction with viral oligonucleotide substrates such as the TAR RNA and its complementary DNA sequence (cTAR).[76, 79]

Besides further pursuing the investigation of anthraquinone SARs, during the last period of my doctoral research project we extended the study of the NC-inhibitory activity on novel RNA substrates involved in early events of the viral life cycle. Specifically, we investigated the ability of these molecules to inhibit the primer-annealing step of the HIV-1 reverse transcription process by interaction with the HIV-1 primer binding site (PBS) sequence.

2.1 Small molecules targeting NC: BN and BM compounds series as direct NC inhibitors

Selection of compounds from the two different unrelated series, respectively the so-called "BN" and the so-called "BM" compounds, was conducted starting from a virtual screening analysis performed by the group of Professor Maurizio Botta (University of Siena, Department of Biotechnology, Chemistry and Pharmacy). Representative structures of NC complexed to viral RNA[128] and DNA[20] structures refined by prior computational studies,[129] as well as the NMR structure of NC in complex with a small molecule inhibitor,[116] were used as rigid receptors to screen molecules with the desired features to mimic interaction of nucleobases within the hydrophobic pocket of NC. After visual inspection of the predicted binding poses, theoretical affinity ranking and filtering by structure diversity to maximize the chemical space explored, the most interesting theoretical binders were subject to biological evaluation of their NC inhibitory activty.[114]

- The BN series encompassed thirty-five small molecules (**BN1**−**35**) characterized by a great chemical diversity. After a selection of a promising hit compound from the BN series,[130] my research activity focused on the evaluation of the biological activity of a series of analogue molecules (**BN5-01**−**BN5-19**, **Figure 2.1**) and on the clarification of their mechanism of action (MOA).
- The BM series represents an explorative small library encompassing ten small molecules (**BM1**−**BM10**, **Figure 2.2**). Biological activity evaluation and study of the MOA for this small series of compounds led to the identification of one chemotype displaying an unusual reactivity toward the ZFs domain of NC.

The NC chaperone activity inhibition on both series was investigated experimentally employing different biological and biophysical assays developed in our lab. Most promising hit compounds emerged from each series were chosen for further studies aimed to depict the mechanism of the NC inhibition. Electrophoretic mobility shift assay, surface plasmon resonance and mass spectrometry techniques were used to assess the effect of the compounds on protein-nucleic acid complexation and their interaction with the NC putative target. Results and discussion are reported in **Chapter 4**.

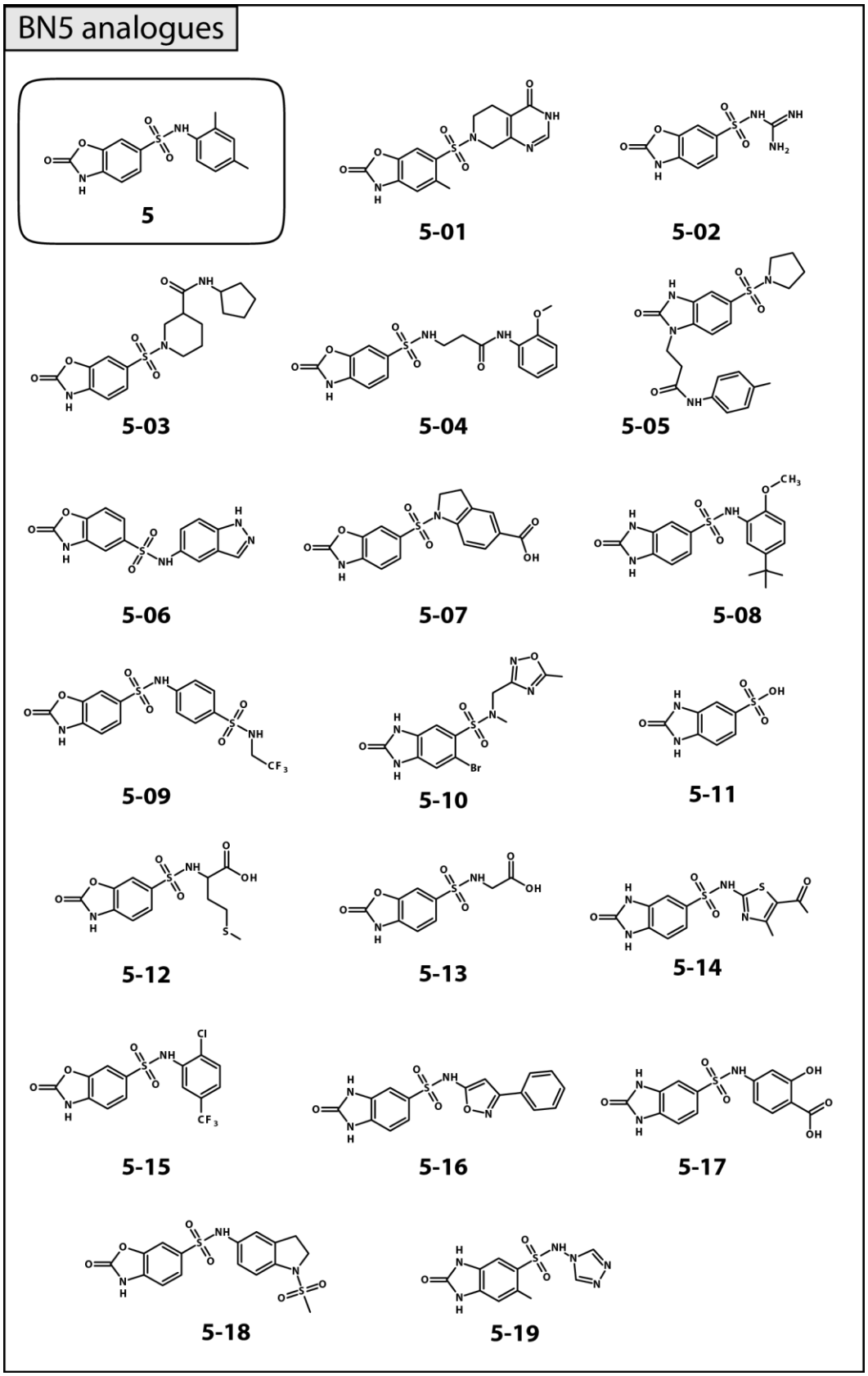


Figure 2.1. Chemical structures of the BN5 analogues series

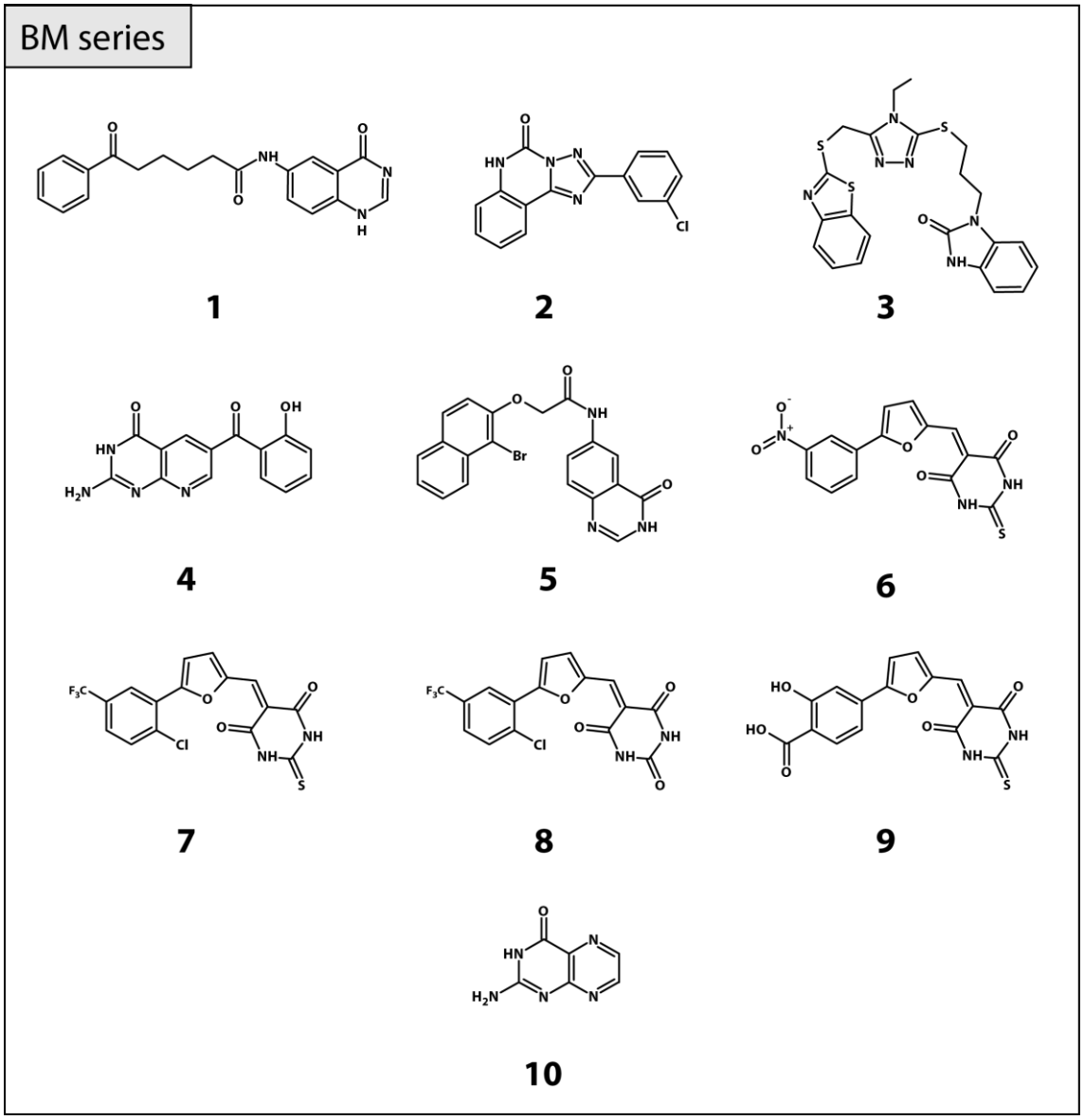


Figure 2.2. Chemical structures of the BM compounds series

2.2 2,6-dipeptidyl-anthraquinone compounds as indirect NC inhibitors

2,6-dipeptidyl-anthraquinones are nucleic binders: their anthraquinone scaffold intercalates nucleic acids by stacking with the paired bases. Binding is enhanced by appropriately positioned cationic side chains that interact with the phosphate backbone of the oligonucleotide substrate. Therefore, these anthraquinone conjugates (AQs) are characterized by two structural elements, namely anthraquinone intercalating scaffold and symmetric cationic side chains in positions 2 and 6. The side chains in the series analysed in this thesis are formed by two amino acids referred respectively as linker and terminal residue. Different AQs bear different combinations of the latter residues in each side chains. Compounds names are composite and characterized by a letter code based on the linker residues (GSF, G, LA and DA) and a number specific for the terminal residue, as exemplified in **Figure 2.3**.

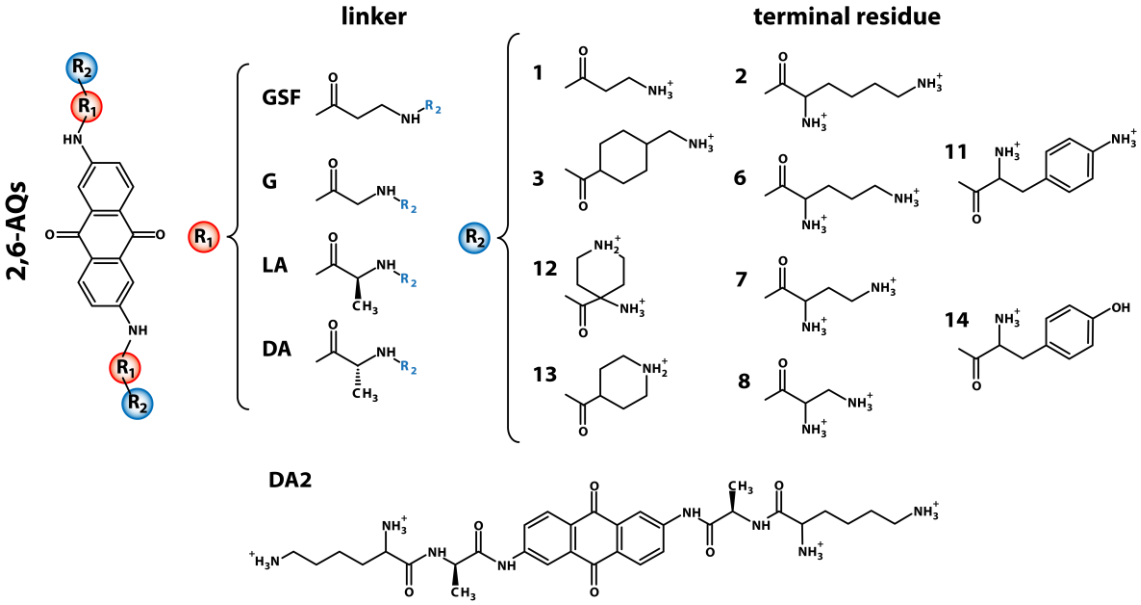


Figure 2.3. Chemical structures of the 2,6-dipeptidyl-anthraquinone compounds. The four compound series that were object of this study were differentiated by the linker residue, namely **GSF** (β-alanine), **G** (glycine), **LA** (L-alanine) and **DA** (D-alanine). Within each series, different compounds were identified by numbers corresponding to the specific terminal residue as follows: **1**, β-alanine; **2**, lysine; **3**, 4-methylene-amino-cyclohexyl-carboxylic acid; **6**, ornithine; **7**, 2,4-diamino-butanoic acid; **8**, 2,3-diamino-propionic acid; **11**, p-amino-phenylalanine; **12**, 4-amino-carboxyl-piperidine; **13**, 4-carboxyl-piperidine.

The GSF, G and LA series have been reported in previous research works and shed light on the molecular determinants and mechanism of action of AQs.[76, 79] Important SARs were found regarding optimal side chain hindrance and length. These results were obtained

investigating the interaction of the anthraquinone conjugates with the viral TAR RNA and its complementary DNA sequence cTAR substrates. During this research work, I contributed to the study of the new DA compound series, exploring the effect of the side chain orientation on the activity of anthraquinone conjugates as inhibitors of NC.

Starting from the data obtained with the TAR/cTAR model system, we then extended our study by employing other viral oligonucleotide substrates. Prompted by the multifaceted functions of NC, we aimed to assess the potential interaction of AQs with novel nucleic acid sequences that are substrates of NC, to probe the effect of these compounds in different biologically relevant contexts. Focusing on the HIV-1 PBS, we investigated the ability of AQs to inhibit the first step of the reverse transcription process, which requires the annealing activity of NC. To this purpose, we developed a novel biological assay to investigate inhibition of the NC annealing activity *in vitro* whereas mass spectrometry was employed to explore their MOA. Results and Discussion are reported in **Chapter 5**.

3. MATERIALS AND METHODS

3.1 Compounds, oligonucleotides, and NC protein preparation.

Compounds identified by the virtual screening process from the BN and BM were provided by Prof. M. Botta (Università degli Studi di Siena, Dipartimento di Biotecnologie, Chimica e Farmacia, Siena, Italy) and purchased from MolPort (Riga, LV). Series of 2,6-dipeptidyl-anthraquinone conjugates were synthetized by Proff. V. Santagada and G. Caliendo (Università di Napoli "Federico II", Dipartimento di Farmacia, Napoli, Italy). All the compounds were dissolved in pure DMSO and stored in a freezer at -20°C. All solutions were freshly prepared by diluting the initial DMSO stock into water to the desired concentration.

The following oligonucleotide constructs were employed in the study (**Figure 3.1**): TAR, 5'-GGC AGA UCU GAG CCU GGG AGC UCU CUG CC-3' (RNA); cTAR, 5'-GGC AGA GAG CTC CCA GGC TCA GAT CTG CC-3'(DNA); PBS, 5'-UGG CGC CCG AAC AGG GAC-3' (RNA); Lys3, 5'-GUC CCU GUU CGG GCG CCA-3' (RNA). TAR, PBS and Lys3 were purchased from Metabion (Steinkirchen, D), cTAR from Eurogenetec (Seraing, B) and used without further purifications. When needed, TAR and cTAR sequences were purchased labeled with 6-carboxyfluorescein (FAM) as fluorophore and 4-(4'-dimethylaminophenylazo) benzoic acid (DABCYL) as quencher at the 5' and 3'-end, respectively. TAR and cTAR correspond to the top half hairpin structure of the trans-activator response element from HIV-1 LAI strain. PBS correspond to the 18-nt HIV-1 primer binding site whereas Lys3 is the 18-nt complementary sequences found in the human $tRNA^{Lys,3}$. An extinction coefficient of 268900 $M^{-1}cm^{-1}$, 271800 $M^{-1}cm^{-1}$, 171300 $M^{-1}cm^{-1}$ and 162300 $M^{-1}cm^{-1}$ at 260nm were used to determine the concentration of TAR, cTAR, PBS and Lys3 sequences, respectively.

The 55-residue, full-length recombinant NC from the LAI strain (IQKGN FRNQR KTVKC FNCGK EGHIA KNCRA PRKKG CWKCG KEGHQ MKDCT ERQAN) was expressed in E. coli, purified by ion-exchange and size exclusion chromatography, as reported elsewhere,[18, 131, 132] and stored in ammonium acetate 100mM in the freezer at -20°C (**Figure 3.1**). The NC(12-55) (TVKCF NCGKE GHIAK NCRAP RKKGC WKCGK EGHQM KDCTE RQAN) corresponding to the NC sequence lacking eleven residues from the NC amino terminal was synthetized by EspIkem Peptides (Polo scientifico e tecnologico di Sesto Fiorentino, Firenze, IT) and was stored in the freezer at -20°C in Tris buffer (Tris 10mM, pH 7.5) (**Figure 3.1**). An extinction coefficient of 6410 $M^{-1}cm^{-1}$ and 5700 $M^{-1}cm^{-1}$ at 280 nm were used to determine the concentration of the full-length NC and NC peptides, respectively. Tat48-57

(GRKKR RQRRR) and Tat37-72 (CFTTK ALGIS YGRKK RRQRR RPPQG SQTHQ VSLSK Q) peptides were stored in Tris buffer (Tris 10 mM, pH 7.5) in a freezer at -20°C (**Figure 3.1**).

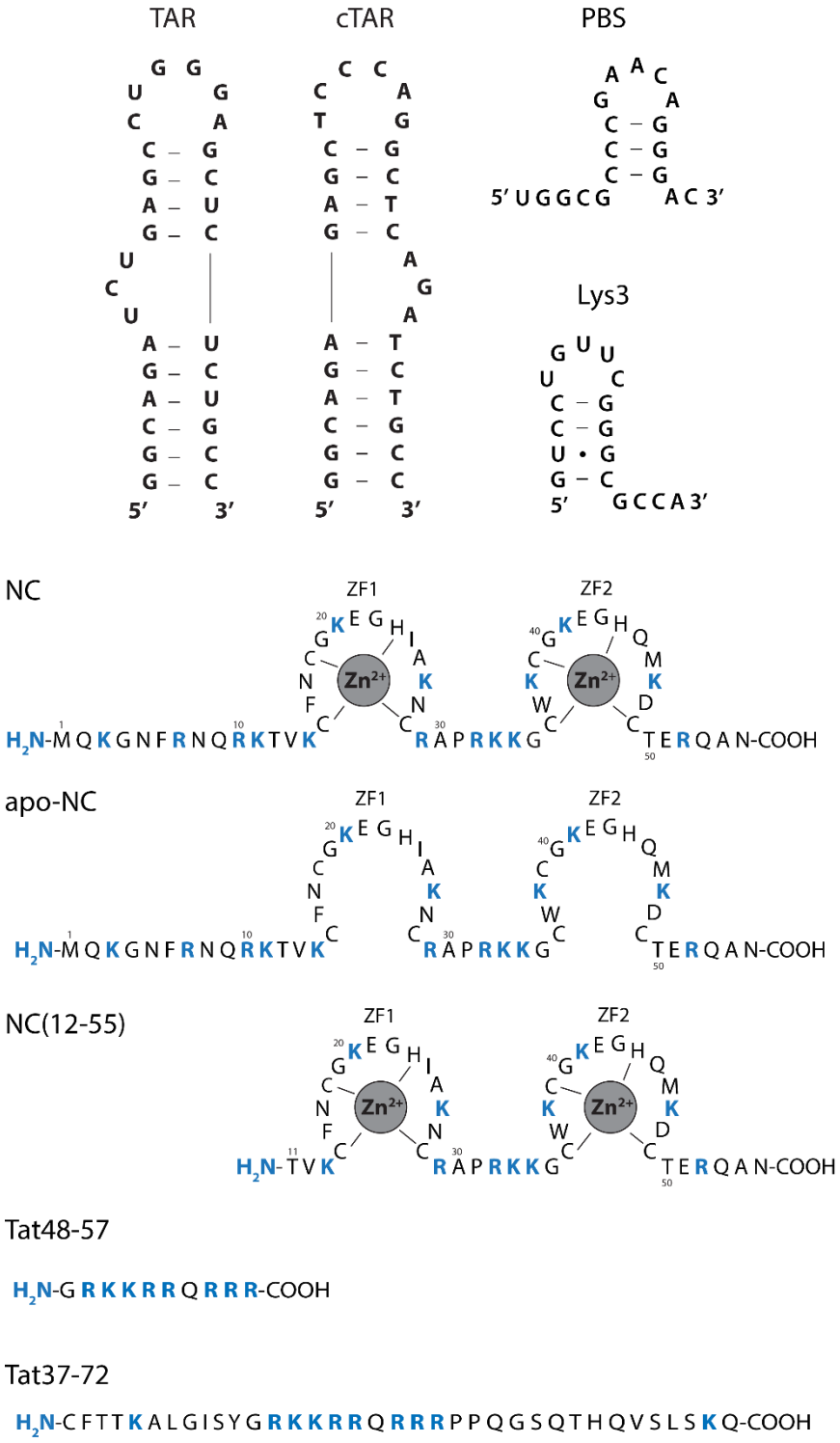


Figure 3.1. Oligonucleotides and proteins used in the study. Oligonucleotides secondary structures were calculated by mFold software.[133-135] Cationic residues into the NC and Tat sequences are blue colored.

3.2 High throughput screening

High throughput screening assay (HTS) assesses the inhibitory activity of compounds on the nucleic acid destabilization ability of NC employing the 29-nt TAR and cTAR sequences, [76, 79] whose annealing take place during the minus strand transfer, an obligated step of the HIV-1 reverse transcription process that requires the chaperone activity of NC (see **Chapter 1.4**). Both TAR and cTAR folds into a hairpin structure characterized by a 3-nt bulge and a hexa-loop (**Figure 3.1**) and the first requirement of the TAR/cTAR annealing reaction is destabilization of the TAR and cTAR stem region by the activity of NC. Binding of NC leads to the opening of the terminal base pairs in the stem region, shifting the conformational equilibria between the open and the closed form that already occur by thermal fluctuations.[57, 58, 64, 136-141] Nucleic acid destabilization relies on the interaction of specific guanine residues in the terminal stem region with the hydrophobic pocket located at the top of the zinc fingers of NC.[61] TAR and cTAR stem destabilization also occurs for the 29-nt sequences used here, which represent the top-half region of the full-length TAR hairpin structure (see **Chapter 7.2**). By using the doubly labelled TAR and cTAR sequences with FAM and DABCYL, the nucleic acid destabilizing ability of NC can be measured by means of the fluorescence-resonance energy transfer (FRET) effect (**Figure 3.2**).[142]

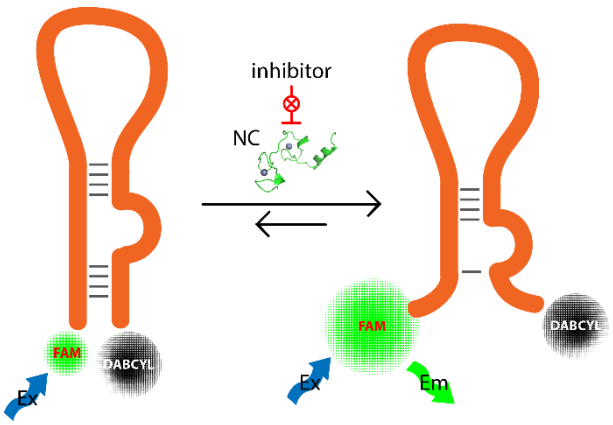


Figure 3.2. Schematic representation of the HTS assay. By using doubly labelled TAR or cTAR sequences with a fluorophore and quencher at the 5' and 3'-ends, fraying of the stem region can be probed by FRET effect. The proximity of the FAM fluorophore to the DABCYL quencher allows the formation of an exciton interaction with efficient quenching of the fluorescence signal in the hairpin closed conformation. Addition of NC favour the hairpin open conformation and restoration of the fluorescence according to the higher distance between the two labels. Inhibitory activity of compounds is characterized by a decrease in the fluorescence signal of the oligonucleotide-NC mixture.

The experiments were conducted at 25°C in a 96-well plate format and fluorescence signal was measured using a Virctor3 plate reader (Perkin Elmer, Waltham, MA, USA) using the 485 nm and 535 nm as excitation and emission filters, respectively. Firstly, labelled TAR

and cTAR sequences were folded separately in folding buffer (Tris HCl 10 mM, NaCl 20 mM, $Mg(ClO_4)_2$ 1 mM, pH 7.5) for 5 minutes at 95°C followed by slow chilling to room temperature in order to obtain the desired hairpin structure. In the following order, each well was added with 80 µL Tris buffer (Tris HCl 10 mM, NaCl 20 mM, pH 7.5), 5 µL compound dilutions (from 0 to 100 µM final), 5 µL of NC (0.8 µM final) and 10 µL of oligo folding solution (0.1 µM final). Control experiments were conducted in absence of both NC and compounds to obtain the fluorescence of the oligo in the hairpin closed state. The DMSO concentration was kept the same in each well. The percentage of inhibition of NC destabilization ($Inh_{\%}$) was obtained by using the relative fluorescence intensity of the oligo in presence of NC and the inhibitor. Intensity obtained in either presence or absence of NC and in absence of the inhibitor were set as references for zero and maximal inhibition, respectively. Half-inhibitory concentration (IC_{50}) values were calculated according to the dose-response equation (1):

$$Inh_{\%} = A_1 + \frac{(A_2 - A_1)}{1+10^{(\log(IC_{50}) - \log(C)) \times p)}} \quad (1)$$

in which A_1 and A_2 represent the inhibition observed in the absence and presence of saturating concentrations of inhibitor, respectively. C corresponds to the concentration of inhibitor, and p denotes the Hill coefficient. The half-maximal inhibitory concentration (IC_{50}) was calculated by using Prism 5.0 (GraphPad, La Jolla, CA, USA) to obtain the best possible fit to triplicate experimental data.

3.3 Nucleocapsid annealing mediated electrophoresis

By nucleocapsid annealing mediated electrophoresis (NAME) assay,[76-80] we assess the inhibitory activity of the compounds on the annealing properties of NC on the TAR/cTAR annealing reaction, without the need of sequence labelling (**Figure 3.3**). Through the NC-mediated annealing, NC initially destabilizes the structure of TAR and cTAR promoting exposition of the complementary single-stranded regions of the stems that can thus anneal forming the nucleation step of the process. NC then destabilizes additional base pairs in the stems to promote the formation of the extended TAR/cTAR hybrid duplex catalyzing the complete annealing though a zippering mechanism (see **Chapter 6.2**). In addition to the HTS, the NAME assay thus mimic the entire minus strand transfer event occurring during the reverse transcription process enabling a more comprehensive observation of the NC chaperone activity inhibition.

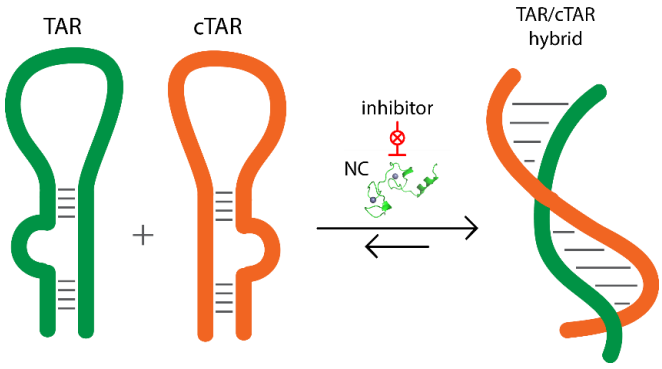


Figure 3.3. Schematic representation of the NAME assay. By employing TAR and its complementary sequence cTAR, the annealing reaction occurs slowly due to the formation of stable secondary structures that freeze the monomeric species. Addition of NC strongly activates annealing of the two sequence by destabilization of the stem regions of TAR and cTAR. Electrophoretic separation of the reaction mixture reveals the annealing activity of NC by observation of a band having lower electrophoretic mobility compared to those of the monomeric TAR and cTAR, which correspond to the hybrid duplex specie. The activity of compounds on the annealing activity of NC is assessed by observing the inhibition of formation of the duplex specie by incubation of oligonucleotide with NC in presence of compounds.

For experiments conducted with BN and BM compounds, TAR and cTAR constructs were folded separately in folding buffer (Tris 25 mM, NaCl 30 mM, $Mg(ClO_4)_2$ 0.2 mM, pH 7.5) by first heating for 5 minutes at 95°C, then slowly chilled at room temperature to produce the desired monomeric hairpins. Annealing reactions were carried by mixing equimolar amounts of TAR and cTAR (final 1 µM each) with an 8-fold amount (final 8 µM) of NC protein and then incubating at room temperature for 15 minutes. Experiments were conducted using the full-length NC from the HIV-1 LAI strain. For compounds from the BN series, all the experiments were also conducted using the shorter NC(12-55) peptide, which affords less aggregation of oligonucleotides under the assay conditions while retaining all the chaperone functions mediated by the zinc fingers domain.[64, 143] Control experiments were carried out by heat-refolding an equimolar mixture of TAR and cTAR to obtain the TAR/cTAR hybrid in the absence of chaperone. Inhibitory activity was determined by pre-incubating NC with increasing compound final concentrations (as indicated in the figure) at room temperature for 15 min, before adding the solution to the above mixture of TAR and cTAR (NC pre-incubation protocol). For experiments with AQs compounds, the procedure was similar but TAR and cTAR were folded in a different folding buffer (Tris 10 mM, NaCl 20 mM, $Mg(ClO_4)_2$ 1 mM, pH 7.5). Moreover, the order of addition was reverted by pre-incubating compounds with TAR and cTAR separately, before mixing the two solutions and adding NC protein to start the annealing reaction (oligo pre-incubation protocol). We made sure that the concentration of DMSO coming from the compound stocks was the same in all reaction wells. The annealing reactions were stopped by adding SDS gel-loading buffer (100 mM Tris, 4 mM EDTA, 2% SDS, 50% glycerol, 0.05% bromophenol blue, pH 7.5), which induced denaturation of the zinc finger structures

without causing dissociation of the nucleic acid duplex. The reaction mixtures were loaded on a 12% polyacrylamide/bis-acrylamide (19:1) gel in TBE buffer (Tris-Borate 89 mM, EDTA 2 mM) and analyzed at 200 V for 3h. The gels were stained with SYBR Green II and visualized on a Geliance 600 Imaging System (Perkin Elmer, Waltham, MA, USA). Band quantification was performed by using the GeneSnap software (SynGene, Cambridge, UK). The percentage of inhibition of annealing activity ($Inh_{\%}$) was obtained by using relative band intensity corresponding to the TAR/cTAR hybrid in the presence of inhibitor setting as a baseline the intensity that was observed in the absence of inhibitor. Half-inhibitory concentration (IC_{50}) values were calculated according to the dose-response equation (1) by using Prism 5.0 (GraphPad, La Jolla, CA, USA), from experimental data conducted in triplicate.

3.4 Fluorescence quenching assay

By fluorescence quenching assay (TQA) we measure the stabilization of the nucleic acid secondary structure induced by ligands binging. TQA permits to calculate the melting temperature (T_m) of oligonucleotides by monitoring the fluorescence signal of labelled sequences while rising the temperature to promote heat-denaturation of the structure. Secondary structure stabilization is then evaluated based on the different T_m (ΔT_m) observed for the oligonucleotide in either presence or absence of ligands (**Figure 3.4**). The assay was used to assess the stabilizing effect of AQs on the secondary structure of TAR and cTAR as well as the TAR/cTAR hybrid. Information on the stabilization effect on these viral nucleic acid sequences induced by AQs compounds provides important hints on the mechanism of NC inhibition since the first step of the NC-mediated TAR/cTAR annealing reaction require the destabilization of the stem region in both oligonucleotide substrates.

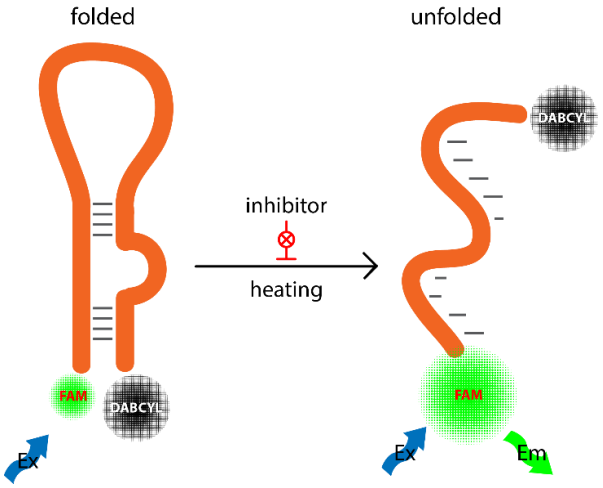


Figure 3.4. Schematic representation of the FQA assay. By using doubly labelled TAR or cTAR sequences with a fluorophore and quencher at the 5' and 3'-end, melting of the secondary structure can be probed by FRET effect. The proximity of the FAM fluorophore to the DABCYL quencher allows the formation of an exciton interaction with efficient quenching of the fluorescence signal in the hairpin closed conformation. Heating favour melting of the secondary structure and restoration of the fluorescence according to the higher distance between the two labels. Melting temperature is calculated from the midpoint of the fluorescence signal curve and the nucleic acid stabilization effect of compounds is characterized by an increase in the melting temperature observed for the oligonucleotide-compound complex.

All the experiments were conducted on a Roche LC480 II instrument in a 96-well plate format. FAM-DABCYL doubly labelled TAR and cTAR constructs were folded separately in folding buffer (Tris 10 mM, NaCl 20 mM, $Mg(ClO_4)_2$, 1 mM, pH 7.5) by first heating for 5 minutes at 95°C, and then slowly cooled down to room temperature to produce the desired monomeric hairpins. The TAR/cTAR hybrid duplex was obtained similarly by folding an equimolar mixture of doubly labelled TAR and cTAR. After chilling, the oligonucleotide folding solutions were diluted 10-fold with Tris-EDTA buffer (Tris 10 mM, NaCl 20 mM, EDTA 1 mM, pH 7.5). Each microplate well was filled with 18 µL of diluted oligonucleotide folding solution (1 µM final) and 2 µL of compounds dilution (1, 10 or 100 µM final). Control experiments were conducted in absence of compounds to obtain the melting temperature of the oligonucleotide in absence of compound. The samples were heated from 25°C to 99°C at a 0.02°C/s heating rate and the fluorescence signal from each well was measured using the 465 nm and 510 nm as excitation and emission filters, respectively. The T_m values were taken at the midpoint of the flex observed in the plots showing the fluorescence intensity over time using the LC 480 software. The stabilizing effect of compounds was expressed as melting temperature shift (ΔT_m), calculated as the difference between the melting temperatures observed in either presence or absence of the compound, respectively.

3.5 Surface plasmon resonance-competition assay

By using surface plasmon resonance (SPR) technique, we evaluate the inhibitory activity of compounds on the binding of NC to the TAR and cTAR. SPR is widely applied to the investigation of bio-molecular recognition finding wide application in the field of protein-protein and protein-nucleic acid interactions. The technique permits to monitor in real time the association and dissociation phases of a binding interaction with no need of labelling of the two interacting counterparts. Usually, a receptor molecule is immobilized onto a surface of a noble metal (e.g., gold) and the analyte flows free in solution above the receptor. The analyte-receptor interaction is monitored by measuring the change in the refractive index of the metal surface. Being sensitive to the mass of the functionalized metal layer, SPR is widely used to study the interaction of macromolecules such as proteins and oligonucleotides with a molecular mass in the range of kilodaltons. The analysis of small molecules binding to immobilized macromolecular receptors requires a higher sensitivity and correction factors due to the co-solvent commonly used in the analysis.[144] For this reason, SPR was used here to develop a competition assay rather than used to directly assess the binding of compounds to NC protein. In the SPR-competition assay format, folded TAR or cTAR biotinylated at the 3'-end was immobilized onto a gold chip through a biotin-streptavidin coupling and served as receptors for the NC analyte in solution. The compounds were incubated with NC protein and the mixture used as analyte flowing onto the oligonucleotide-functionalized surface. Under these conditions, only the free protein that does not bind the inhibitor drug is expected to interact with the immobilized receptors (**Figure 3.5**).

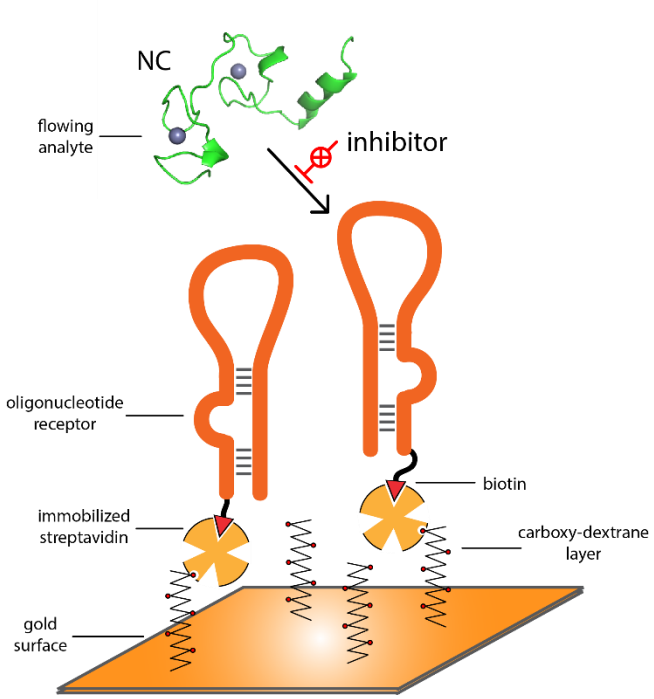


Figure 3.5. Schematic representation of the SPR-competition assay. By immobilization of TAR or cTAR sequences onto a sensor chip, binding of NC to the oligonucleotide can be assessed by SPR. When flowing onto the functionalized surface, the NC analyte is free to interact with the immobilized receptors and the binding can be monitored by SPR signal due to a change in the refractive index of the surface. The inhibitory activity of compounds can be assessed by a decrease in the SPR signal registered for a compound-NC mixture used as the analyte.

The experiments were conducted on a Biacore X100 (GE Healthcare), using SA sensor chips immobilized with either 400 RU of 3'-biotinylated TAR or cTAR hairpins. All the analysis were conducted at 25°C using Hepes buffer as the running solution (Hepes 10 mM, NaCl 150 mM, Tween 20 0.005%), adding DMSO to a variable concentration to match the composition of samples containing compounds. Analyte samples were prepared mixing NC (200 nM final) with compound dilutions (from 0 to 25 µM final) by mixing equal volumes of concentrated solution. DMSO concentration was kept the same in each sample. The analysis was performed using 10 µl/min flowrate and 60 s for both the association and dissociation phases. Between each injection, the surface was regenerated by flushing 0.05% SDS + 3 mM EDTA and 2 M NaCl solutions in running buffer for 30 s each. Raw sensorgrams were double referenced subtracting the signal from the reference cell and blank injections consisting of running buffer alone. The SPR signal obtained at the plateau of the curve 5 s before the dissociation phase was used for calculation of binding. The percentage of inhibition of NC binding ($Inh_{\%}$) was obtained from the SPR signal corresponding to the NC binding in presence of inhibitor setting as a baseline the signal observed in the absence of inhibitor. Half-inhibitory

concentration (IC_{50}) values were calculated according to the dose-response equation (1) by using Prism 5.0 (GraphPad, La Jolla, CA, USA), from experimental data conducted in triplicate.

3.6 Electrophoresis mobility shift assay

By electrophoresis mobility shift assay (EMSA), we evaluate the inhibitory activity of compounds on the binding of NC to the TAR and cTAR. TAR and cTAR constructs were folded separately in folding buffer (Tris 25 mM, NaCl 30 mM, $Mg(ClO_4)_2$ 0.2 mM, pH 7.5) by first heating for 5 minutes at 95°C, followed by slow chilling at room temperature to produce the desired monomeric hairpins. Binding of NC to TAR and cTAR was assessed by incubating an aliquot of the protein (8 µM final) with TAR or cTAR folding solution (1 µM final) for 1 hour at 20°C. Inhibition of NC binding was evaluated by incubation of the protein-nucleic acid complexes with compounds (200 µM final) for an additional hour at 20°C. Gel loading buffer (Tris 100 mM, Glycerol 40%, Bromophenol blue 0.05%, pH 7.5) was added and samples loaded on a 14% native polyacrylamide/bis-acrylamide (29:1) gel in 0,5X TBE buffer (Tris-Borate 45 mM, EDTA 1 mM) and analysed at 100 V for 5.5 hours at 6°C. The gels were stained with SYBR Green II and visualized on a Geliance 600 Imaging System (Perkin Elmer, Waltham, MA, USA).

3.7 UV-spectroscopy

UV-absorption and circular dichroism (CD) spectroscopy were used to assess the secondary structure of the oligonucleotide sequences. The oligonucleotide was diluted in a 1 cm path quartz cuvette with water to a concentration of about 1.5 µM. Spectra was acquired in the 210-350nm spectra region at 20°C on a J-810 spectrophotometer (Jasco, Oklahoma City, OK, USA) equipped with a thermostated cell compartment. The concentrated folding buffer was added to the cuvette to achieve final concentrations (Tris-HCl 10 mM, NaCl 20 mM, $Mg(ClO_4)_2$ 1mM, pH 7.5) and a spectrum was acquired at 20°C. The temperature was gradually increased at a rate of 50°C/h and a spectrum at 95°C was acquired after 5 minutes wait. Spectra were corrected for dilution factor. The dichroic signal was expressed in molar ellipticity using the oligonucleotide strand concentration whereas the absorption signal was reconstructed from the CD spectra using the instrument software.

3.8 Kinetic of PBS/Lys3 annealing

The experiments aimed to investigate the kinetic of the annealing between two complementary folded RNAs in absence of proteins by native polyacrylamide gel electrophoresis. PBS and Lys3 were folded independently in folding buffer (Tris 10 mM, NaCl

20 mM, $Mg(ClO_4)_2$ 1 mM, pH 7.5) by first heating at 95°C followed by a chilling step to produce the desired monomeric hairpins through different procedures. The different conditions tested were either 5 or 10 minutes heating at 95°C, followed by either snap cooling on ice or slow chilling at room temperature. After folding, the oligonucleotide solutions were mixed together and incubated at room temperature. The amount of RNA duplex formed was analysed over time by gel electrophoresis. At the indicated time-point, SDS gel-loading buffer (100 mM Tris, 4 mM EDTA, 2% SDS, 50% glycerol, 0.05% bromophenol blue, pH 7.5) was added to samples, loaded on a 12% polyacrylamide/bis-acrylamide (19:1) gel in TBE buffer (Tris-Borate 89 mM, EDTA 2 mM) and analyzed at 250 V for 2h. Control experiments were carried out by heat-refolding an equimolar mixture of PBS and Lys3 to obtain the PBS/Lys3 duplex in the absence of protein. The gels were stained with SYBR Green II and visualized on a Geliance 600 Imaging System (Perkin Elmer, Waltham, MA, USA). Band quantification was performed by using the GeneSnap software (SynGene, Cambridge, UK).

3.9 Assessment of the annealing activity of NC and Tat48-57

The experiments were conducted to assess the catalyzing ability of different viral proteins on the primer-annealing reaction within a fixed reaction time. PBS and Lys3 were folded independently in folding buffer (Tris 10 mM, NaCl 20 mM, $Mg(ClO_4)_2$ 1 mM, pH 7.5) by first heating for 5 minutes at 95°C, followed by snap cooling on ice for at least 5 minutes. Annealing reactions were carried by mixing equimolar amounts of PBS and Lys3 (final 1 μM each) with increasing concentration of either NC or Tat48-57 (as indicated in the Figure), followed by incubation at room temperature for 10 minutes. Control experiments were carried out by heat-refolding an equimolar mixture of PBS and Lys3 to obtain the PBS/Lys3 duplex in the absence of protein. The annealing reactions were stopped by adding SDS gel-loading buffer (100 mM Tris, 4 mM EDTA, 2% SDS, 50% glycerol, 0.05% bromophenol blue, pH 7.5) to promote detachment of the protein from the oligonucleotides while preserving the PBS/Lys3 duplex. The reaction mixtures were loaded on a 12% polyacrylamide/bis-acrylamide (19:1) gel in TBE buffer (Tris-Borate 89 mM, EDTA 2 mM) and analyzed at 250 V for 2h. The gels were stained with SYBR Green II and visualized on a Geliance 600 Imaging System (Perkin Elmer, Waltham, MA, USA). Band quantification was performed by using the GeneSnap software (SynGene, Cambridge, UK). The percentage of annealing activity was obtained by using the relative intensity of the band observed in presence of the protein. The band corresponding to the PBS/Lys3 duplex obtained in absence of the proteins was set as the baseline while the band corresponding to the control heat-refolded duplex as maximal annealing activity.

To assess the importance of the zinc fingers of NC in the PBS/Lys3 annealing reaction, control experiments were conducted using a zinc-less version NC, consisting of the 55-mer NC sequence without the two Zn(II) (apo-NC). apo-NC was obtained by overnight incubation at 25°C of an aliquot of NC (77 μM final) with 350 mM EDTA in 17% methanol and 1% acetic acid at 900rpm shaking. The day after, the precipitated EDTA was discarded and the supernatant was extensively desalted against 10mM ammonium acetate using Pierce concentrators with 3 kDa molecular weight cut-off (Thermo Scientific). Ejection of Zn(II) was assessed by mass spectrometry on a Mariner mass spectrometer (PerSeptive Biosystems). The activity of apo-NC on the PBS/Lys3 annealing reaction was carried under the same conditions to those used for the native NC protein. To assess the importance of the residues flanking the basic core of Tat protein, control experiments were conducted using Tat37-72 testing its annealing activity under the same experimental conditions to those used for NC and Tat48-57.

3.10 RNA-annealer mediated primer-annealing

By RNA-annealer mediated primer-annealing (RAMPA) assay we assess the inhibitory activity of compounds on the annealing of PBS to the complementary sequence in the 3'-acceptor arm of the $tRNA^{Lys,3}$ primer mediated by NC or Tat. Annealing of these two complementary sequences take place during the primer-annealing, which is an obligate step of the HIV-1 reverse transcription process, and require the RNA annealing activity of NC. Differently, from the TAR/cTAR annealing reaction, NC catalyses the primer-annealing by charge screening and molecular crowding effects and thus the specific nucleic acid destabilization ability of the zinc finger domains are not mandatory to achieve complete annealing. Differently, the annealing of TAR to the complementary cTAR sequence used by NAME assay is strictly dependent on the nucleic acid destabilization ability mediated by the zinc fingers of NC. For this reason, RAMPA and NAME assays provide complementary information permitting to assess inhibition of different properties of NC occurring in different biologically relevant contexts of the HIV-1 reverse transcription process. In addition, RAMPA assay permits also to evaluate the inhibitory activity of compounds on the same process while using other viral RNA annealers such as Tat.

Experimentally, PBS and Lys3 were folded independently in folding buffer (Tris 10 mM, NaCl 20 mM, $Mg(ClO_4)_2$ 1 mM, pH 7.5) by first heating for 5 minutes at 95°C, followed by snap cooling on ice for at least 5 minutes to produce the desired monomeric hairpins. Annealing reactions were carried by mixing equimolar amounts of PBS and Lys3 (final 1 μM each) with either NC (final 3 μM) or Tat48-57 (final 6 μM), and then incubating at room temperature for 10 minutes. Control experiments were carried out by heat-refolding an

equimolar mixture of PBS and Lys3 to obtain the PBS/Lys3 duplex in the absence of protein. Inhibitory activity was determined by pre-incubating PBS and Lys3 with increasing compound final concentrations (as indicated in the Figure) at room temperature for 15 min, before mixing and the addition of the NC solution. We made sure that the concentration of DMSO was the same in all the reaction wells. The annealing reactions were stopped by adding SDS gel-loading buffer (100 mM Tris, 4 mM EDTA, 2% SDS, 50% glycerol, 0.05% bromophenol blue, pH 7.5) to promote detachment of the protein from the oligonucleotide while preserving the PBS/Lys3 duplex. The reaction mixtures were loaded on a 12% polyacrylamide/bis-acrylamide (19:1) gel in TBE buffer (Tris-Borate 89 mM, EDTA 2 mM) and analyzed at 250 V for 2h. The gels were stained with SYBR Green II and visualized on a Geliance 600 Imaging System (Perkin Elmer, Waltham, MA, USA). Band quantification was performed by using the GeneSnap software (SynGene, Cambridge, UK). The percentage of inhibition of annealing activity ($Inh_{\%}$) was obtained by using the relative intensity of the band corresponding to the PBS/Lys3 duplex observed in the presence of protein and compound. The band corresponding to the duplex obtained in absence of compound and in presence of the protein was set as zero inhibition while that obtained in absence of both protein and compound as maximal inhibition. Half-inhibitory concentration (IC_{50}) values were calculated according to the dose-response equation (1) fitting data from triplicate experiments.

3.11 Mass spectrometry

The stoichiometry and strength of the binding interactions between NC and the various compounds were evaluated by using electrospray-ionization mass spectrometry (ESI-MS) under non-denaturing conditions. The detection of intact non-covalent complexes can be achieved by appropriate tuning of the ion source conditions to minimize the desolvation energy.[77, 79, 131, 145-148] This type of analysis can reveal the accurate partitioning between free and bound species in solution, which in turn provide the binding constant of a given complex.[131, 145] Moreover, performing experiments in which a mixture of ligand must compete simultaneously for the same substrate can effectively provide a relative scale of binding affinities by comparing the signal intensities of the complexes detected in the same spectrum.[146-148]

In a typical experiment aimed to assess the interaction of a compound with NC, a final 5 μM sample of NC was mixed with 50 μM concentration of the compound in 150 mM ammonium acetate at pH 7.5. Binding competition experiments instead, included 5 μM concentration of each test compound in the same solvent. Full-scan ESI-MS spectra were acquired after 30 minutes incubation at room temperature to allow for the establishment of the binding equilibrium. Tandem MS analysis was performed to verify the reversible nature of the

protein-ligand interactions. Protein fractional occupancy was calculated according to equation (2):

$$f_b = \frac{\Sigma\left(\frac{I_{NC+L}}{z}\right)}{\Sigma\left(\frac{I_{NC+L}+I_{NC}}{z}\right)} \qquad (2)$$

in which f_b represents the fraction of NC bound by the ligand (L), z is the charge state of the specie, I_{NC+L} and I_{NC} are the areas under the curve of the signal corresponding to protein-ligand complexes and free protein, respectively. All the detected charged states were included in the calculation to minimize any quantification bias.[131] All determinations were conducted in positive ion mode on a Thermo Fisher Scientific LTQ-Orbitrap Velos mass spectrometer. Typically, 6 μL samples were sprayed in nanoflow ESI mode by using quartz emitters produced in-house. Source conditions were set to maximize the detection of intact holo-NC containing full complement of coordinated Zn(II) ions[149, 150] which were typically achieved at an ionizing voltage of 0.8 kV, a desolvation voltage between 0 and 50 V, and a source temperature of 200°C. Data were analyzed by using Xcalibur 2.1 software (Thermo Fisher Scientific).

In a typical experiment aimed to assess the interaction of a compound with oligonucleotides, a final 2 μM sample of oligonucleotide was mixed with 10 μM concentration of the compound in 150 mM ammonium acetate at pH 7.5. Binding competition experiments instead, included 2 μM concentration of each test compound in the same solvent. Full-scan ESI-MS spectra were acquired after 30 minutes incubation at room temperature to allow for the establishment of the binding equilibrium. Oligonucleotide fractional occupancy was calculated according to equation (2). All determinations were conducted in negative ion mode on a Waters Synapt G2 HRMS mass spectrometer equipped with a heated capillary source built in-house.[151] Typically, 6 μL samples were sprayed in nanoflow ESI mode by using quartz emitters produced in house. Source conditions were set to maximize the detection of non-covalent nucleic acid complexes[18, 131, 132, 145] which were typically achieved at an ionizing voltage of 0.8 kV, a desolvation voltage of 50 V, and a source temperature of 30°C. Data were analyzed by using MassLynx 4.1 software (Waters). Tandem mass spectrometry aimed to the investigation of the compound binding site on the oligonucleotide structure.[152] The experiments were conducted by using nitrogen as collision gas and activation of the precursor ion was achieved by increasing the voltage in the transfer ion trap of the Synapt G2 instrument.

4. RESULTS AND DISCUSSION

SMALL MOLECULES TARGETING NC

In this part of the thesis work, we investigated two small libraries of compounds identified by means of virtual screening techniques to interact with the zinc fingers domain of NC (see **Chapter 2.1**). For each series, we firstly identified the potential hit molecules assessing inhibition of the NC chaperone properties *in vitro*. Assessment of the NC inhibitory activity was conducted by means of two biological assays. A FRET-based HTS assay was used firstly to evaluate the inhibitory activity of the compounds on the ability of NC to destabilize the hairpin structures of TAR and cTAR. As described in details in **Chapter 6.2**, the annealing of TAR to the complementary cTAR sequence represents an obligated step of the reverse transcription process and destabilization of the structure of TAR and cTAR by NC is a necessary requirement during the process. Using doubly labelled oligonucleotides (see **Materials and methods**), by addition of NC we observed about 4-fold and 8-fold fluorescence increase for TAR and cTAR, respectively. In presence of a compound inhibitor, a decrease in the fluorescence signal from the oligo/NC complex as the compound concentration increases indicated inhibition of the nucleic acids destabilization ability of NC.

To better characterize our *in vitro* biological evaluation, compounds from the BM and BN series were then tested by means of NAME assay. NAME assay employs the same TAR/cTAR model system used in the HTS assay and enables evaluation of the inhibitory activity of compounds on the NC-mediated annealing reaction. The assay well mimics the process occurring *in vivo* and does not require labelling of the oligonucleotide substrates (see **Materials and methods**). Electrophoresed on a native polyacrylamide gel, a mixture of TAR and cTAR yields two well-resolved bands corresponding to the monomeric oligonucleotides, indicating that the stable secondary structure of the two sequences prevents the annealing. Addition of NC to the oligonucleotides mixture leads to the formation of a band corresponding to the TAR/cTAR hybrid duplex having lower electrophoretic mobility compared to the monomeric species. Conducting the NC-mediated annealing in presence of increasing compound concentration, the NC-inhibitory activity was observed by a decrease in the formation of the band corresponding to the TAR/cTAR duplex and a concomitant increment of the bands with higher electrophoretic mobility corresponding to the monomeric TAR and cTAR.

In a first step, HTS and NAME assays thus served to identify the hit compounds displaying interesting NC inhibitory activity. The following step has been the clarification of the

mechanism by which these molecules exerts their inhibitory activity. To achieve a comprehensive study of the MOA, we designed different experiments aimed to assess the interaction of the compounds with the hydrophobic pocket of NC, which represented the rationale of the initial virtual screening process. To evaluate the ability of compounds to inhibit the specific NC-oligonucleotide interaction, we develop a SPR-competition assay meant to measure the inhibitory activity of the hit compounds on the binding of NC to the TAR and cTAR oligonucleotide receptors. EMSA was run in parallel using the same substrates to qualitatively corroborate the results observed by the SPR-competition assay. Finally, we employed ESI-MS techniques to assess the interaction of the compounds with NC, in order to confirm their putative target. Careful optimization of the experimental conditions, mass spectrometry provided a direct observation of the different species present at the equilibrium, thus offering the opportunity to unambiguously evaluate binding of ligands to the protein. The latter part of the work was conducted in collaboration with Professor Dan Fabris, head of the mass spectrometry facility at the RNA Institute (SUNY at Albany, NY, US).

4.1 BN compounds series: Identification and evaluation of analogues of BN5 as NC inhibitors

4.1.1 Inhibition of NC nucleic acid chaperone activity by BN5 analogues

Preliminary biological evaluation of compounds from the BN series led to identification of **BN5** as a promising hit compound optimal for further development. By HTS, **BN5** showed a good dose-response curve although no complete inhibition was observed in the concentration range tested.[130] NAME assay conducted with the NC(12-55) peptide (see **Materials and methods**), also supported the positive results obtained for **BN5** by HTS. By NAME assay, a clear inhibition of the NC chaperone activity was observed by the decrease in the band corresponding to the TAR/cTAR hybrid duplex and a concomitant increase in the bands corresponding to the monomeric TAR and cTAR (**Figure 4.1**).

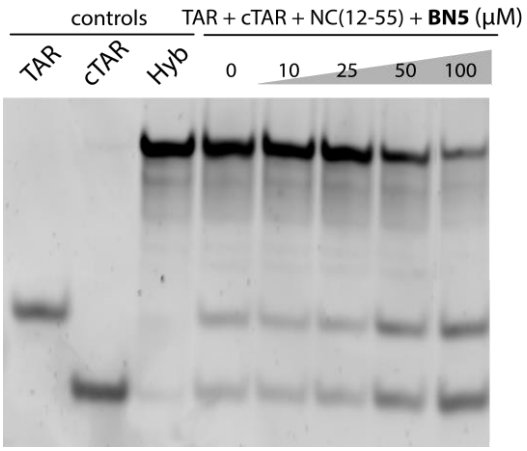


Figure 4.1. NAME assay conducted with BN5. Monomeric TAR and cTAR, as well as the heat-refolded hybrid duplex (Hyb), were included as controls. A clear dose-dependent inhibition of the NC nucleic acid chaperone properties was observed by the decrease in the formation of the band corresponding to the hybrid duplex specie with lower electrophoretic mobility compared to the TAR and cTAR monomeric species (see **Materials and methods**).

The results obtained by both HTS and NAME assays agreed in highlighting the ability of **BN5** molecule to affect the NC chaperone functions, although lacking the desired potency to afford any pharmacological application. Interestingly, we observed striking structural similarities between the positive hit found here and other reported NC inhibitor compounds. Specifically, the 2-benzoxazolinone scaffold of **BN5** well resembled the 2-oxindole sub-structure present in the compound inhibitor published by the Boehringer Ingelheim laboratories (Canada)[116] and the compound **8** identified in a previous virtual screening campaign[114] (**Figure 1.9**). The fact that independent studies converged to the same chemotype prompted us to hypothesize that this specific molecular structure might have caught the optimal structural

requirements for inhibition of NC. These structural similarities convinced us that optimization of the 2-oxindole molecular scaffold by proper chemical decoration might improve the potency observed for the **BN5** compound. We thus underwent a second round of virtual screening searching for potential binders of NC shearing the 2-oxindole molecular scaffold. The results of the second screening yielded nineteen analogues of **BN5** (**BN5-1**–**BN5-19**), which displayed both 2-benxozaxolinone and 2-benzimidazolinone substructures (**Figure 2.1**). In order to compare the activity of the analogues and to afford a rough SARs study, the analogue series was subjected to biological activity evaluation by means of NAME assay.

Testing the analogue series by NAME assay, four molecules (**BN5-01**, **BN5-02**, **BN5-07** and **BN5-15**) displayed inhibition of the NC chaperone activity in the high concentration range tested, whereas **BN5-06** appeared the most potent inhibitor of the NC(12-55) chaperone properties among the test set (supplementary information, **Figure S1**). Inhibition of the NC-mediated TAR/cTAR annealing in presence of **BN5-06** (**Figure 4.2**) was evident at concentration that were remarkably lower compared to **BN5** (**Figure 4.1**), indicating a clear improvement in the NC inhibitory activity for the analogue compound. **BN5-03** was not analysed due to its poor water solubility.

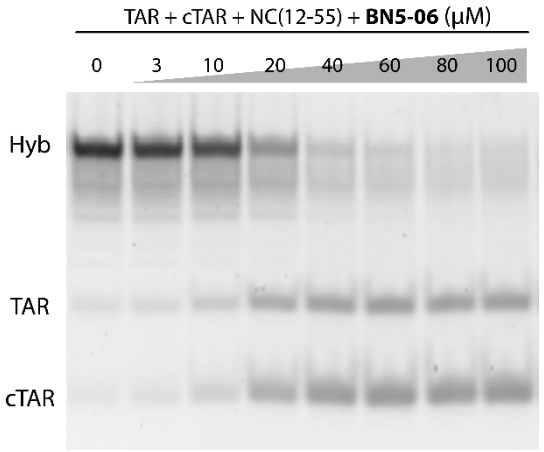


Figure 4.2. NAME assay conducted with BN5-06. A clear dose-dependent inhibition of the NC nucleic acid chaperone properties was observed by a decrease in the formation of the band corresponding to the hybrid duplex specie (Hyb) with lower electrophoretic mobility compared to the TAR and cTAR monomeric species (see **Materials and methods**).

By NAME assay, we calculated an IC_{50} value of (20 ± 2) μM (**Table 4.1**) for the compound **BN5-06**. As observed in **Figure 4.2**, the band at lower electrophoretic mobility nearly disappeared at high concentrations of **BN5-06**, indicating almost full inhibition of the NC-mediated TAR/cTAR annealing reaction. Similarly, **BN5-01**, **BN5-02**, **BN5-07** and **BN5-15** were ranked slightly below having calculated IC_{50} values in the ~200 μM range. Other

compounds showing no activity and for which, the IC_{50} value was clearly above 200 μM, were considered not active (**Table 4.1**).

ID	NAME IC_{50} (μM)	ID	NAME IC_{50} (μM)
5-01	170 ± 5	**5-11**	N.A.
5-02	203 ± 8	**5-12**	N.A.
5-03	N.D.	**5-13**	N.A.
5-04	N.A.	**5-14**	N.A.
5-05	N.A.	**5-15**	179 ± 7
5-06	20 ± 2	**5-16**	N.A.
5-07	163 ± 4	**5-17**	N.A.
5-08	N.A.	**5-18**	N.A.
5-09	N.A.	**5-19**	N.A.
5-10	N.A.		

Table 4.1. Activity ranking (IC_{50}) of analogues of BN5 assessed by NAME assay. N.A. = not active; N.D. = not determined

The results afforded rough SARs for the NC inhibitory activity by this analogue compound series. 2-benzimidazolinones appeared not suitable to inhibit NC as highlighted by the inactive compounds **BN5-08**, **BN5-10** and **BN5-11**, all sharing the same molecular scaffold. On the opposite, the 2-benzoxazolinone scaffold found in compound **BN5-01**, **BN5-06**, **BN5-07**, and **BN5-15**, all showing a different degree of activity by NAME assay, emerged as optimal structure endowed with NC-inhibitory activity. Among these active molecules, the 5-sulfonamide substitution proper of the most potent inhibitor **BN5-06** was optimal for interaction with NC. Differently, all other active compounds bear the sulfonamide substitution in position 6 of the benzoxazolinone ring indeed, similar to the precursor **BN5** compound. Introduction of bulkier sulphonamides (e.g. compounds **BN5-04**, **BN5-05**, and **BN5-18**) was detrimental for NC inhibition, as well as the negatively charged aliphatic carboxylic group of amino acids (**BN5-12** and **BN5-13**). In contrast, the aromatic carboxylic function (see **BN5-07**) seemed to be allowed in this series of NC inhibitors, at least to a limited extent. These rough SAR could facilitate the design of improved NC inhibitors.

Although these results are highly promising and validated the goodness of the workflow adopted for the identification of novel NC inhibitor hit compounds, to study the biological activity of the BN series by NAME we employed the NC(12-55) peptide (see **Materials and methods**). When the same assay was conducted using the full-length NC protein, we observed that the potency of **BN5-06** was rather low (supplementary information, **Figure S2**). This observation was later supported also by SPR-competition and EMSA experiments involving

BN5-06 and NC. These results suggested that the unspecific charge screening and crowding effect mediated by the NC positive residues contribute to defining the protein functions and they should be considered as well as those involved in the specific nucleic acid destabilization activity given by the ZFs domain. Our result thus demonstrated that **BN5-06** efficiently inhibit the chaperone functions afforded by the hydrophobic pocket located at the top of the ZFs of NC. However, under the experimental conditions used, which cannot closely reproduce those occurring in a real *in vivo* context, aggregation of oligonucleotide might prevent us from highlighting the inhibitory activity of the compound. Different properties of NC thus appear to play together in a concerted nucleic acid chaperoning mechanism, which should be considered for the development of more potent inhibitors, further challenging the identification of a valid inhibitor of NC.

4.1.2 Inhibition of the NC-oligonucleotide interaction by BN5 analogues

To study compounds inhibitory activity on the NC interaction with oligonucleotides we employed the SPR technique. By SPR, we assess the ability of compounds to inhibit the binding of NC on the TAR and cTAR sequences. In the competition assay format, an aliquot of NC was let free to interact with the oligonucleotide immobilized onto the sensor chip, monitoring the steady state SPR signal (i.e., the signal obtained at the top of the sensorgram trace during the association phase). By conducting experiments in the presence of a compound, the inhibitory activity on the binding interaction was assessed by a decrease in the SPR signal as an effect of the lower binding of NC to the immobilized receptor (see **Materials and methods**).

Control experiments were performed before the analysis to ensure that the compounds tested produced no SPR signal when flowed onto the oligonucleotide-functionalized sensor chip. No detectable signal was detected when compounds were let free to interact with the TAR and cTAR receptors, indicating no interference with the SPR signal and no detectable interaction of compounds with the immobilized oligonucleotides (data not shown). In a second step, binding studies were conducted to determine the optimal concentration of NC for the solution-competition assay. NC binding to TAR and cTAR hairpins led to the distortion of the sensorgrams at high protein concentration, probably due to aggregation and conformational changes of the immobilized oligonucleotides occurring at the surface. Although this prevented us from an accurate quantitative estimation of the NC-oligonucleotide affinity, a dissociation constant in the range of hundreds nanomolar was calculated with both substrates (supplementary information, **Figure S3**). Based on these results, the NC concentration adopted in the solution-competition assay was set to 200 nM, which resulted in a stable and reproducible SPR trace originating from the protein-oligonucleotide interaction.

When **BN5-06** was tested by SPR-competition assay, the BN hit molecule turned out to be barely active against the NC binding to the TAR and cTAR structures (supplementary information, **Figure S4**), in line with the weak inhibitory activity observed by the previous experiments conducted using the full-length NC protein. Control EMSA experiments also showed similar results (supplementary information, **Figure S4**). The SPR-competition assay thus showed that **BN5-06** weakly compete for binding of the full-length NC to TAR and cTAR and that this likely accounts for the weak inhibition of the TAR/cTAR annealing observed by NAME assay (supplementary information, **Figure S2**). As mentioned above, the discrepancy observed when comparing experiments conducted with the full-length NC or the NC(12-55) peptide, might be related to the higher tendency of the former protein to aggregate oligonucleotides, preventing us from highlight the inhibitory activity of the **BN5-06** compound on the specific function afforded by the zinc fingers domain of NC.

4.1.3 Binding of BN5 analogues to NC

In order to strengthen our working hypothesis and to investigate further the MOA of the active benzoxazolinone derivatives, we studied the interaction of compounds with the NC protein target by means of mass spectrometry. Accurate optimization of the source conditions enabled detection of intact protein-compound complexes, providing a representative observation of the binding equilibria occurring in solution.[77, 79, 131, 145-148] Identity of the protein was firstly confirmed by the experimental mass value of 6488.91 u, in agreement with the monoisotopic mass of 6488.91 u calculated by the elemental composition of the full-length NC protein incorporating two Zn(II) ions (**Table S1**, supplementary information).[149, 150] A typical result of ESI-MS analysis of NC is shown in **Figure 4.3**. In positive ion mode, multiply charged species were observed under the experimental conditions used. The most abundant specie was the quintuple charged ions whereas species bearing four, six and seven charges were still observable with lower abundance.

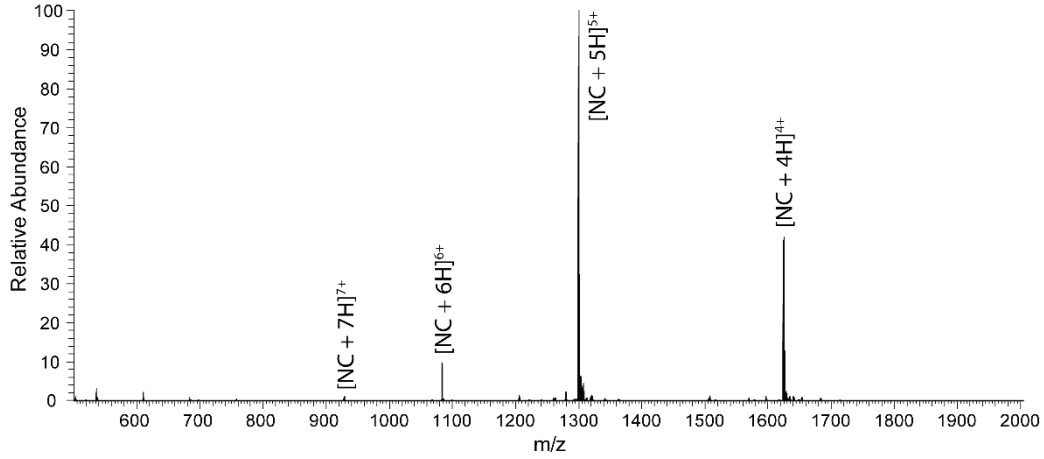


Figure 4.3. ESI-MS spectrum of NC obtained under typical experimental conditions (see **Materials and Methods**). Charge states ranging from 4+ to 7+ were readily detected in positive ion mode. These signals provided an experimental mass of 6488.91 u, which matched very closely a mono-isotopic mass of 6488.91 u calculated from the protein sequence with addition of two Zn(II) ions.[149, 150] Low-intensity signals observed at slightly higher m/z than the corresponding NC signal consist of typical sodium, potassium and ammonium adducts.

To study the MOA of compounds belonging to the BN series by mass spectrometry, we focused on the **BN5-06** hit and the **BN5** precursor molecules, as well as on the other BN analogues that displayed a different degree of inhibitory activity by NAME assay (**Table 4.1**). Binding to NC of **BN5-06** was compared directly to the parent compound **BN5** by incubating ten-fold compound concentration with the protein and acquiring a full scan spectrum after 30 minutes equilibration at room temperature. For both benzoxazolinone molecules, peaks corresponding to the NC-compound complex were readily observed indicating interaction of the selected hits with their putative target. A comparison of NC protein in complex with either **BN5** or **BN5-06** is shown in **Figure 4.4**, focusing on the spectra region containing the quintuply charged ions that represent the most abundant signals in the spectra.

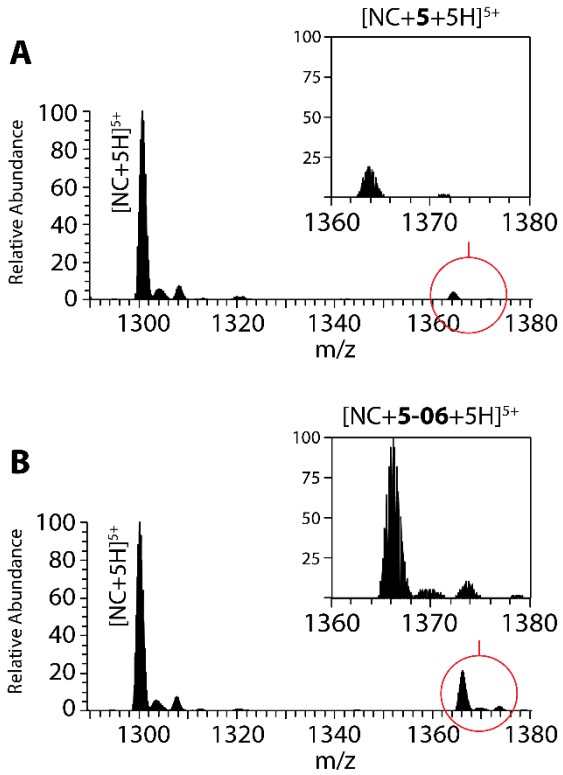


Figure 4.4. ESI-MS direct binding assay with BN5 and BN5-06. Spectra were acquired after incubation of NC with 10-fold molar concentration of either **5** or **5-06** (see **Materials and methods**). All spectra show the region containing signals from the quintuply charged ions that represented the most abundant species. Magnified spectra region containing the peak corresponding to NC in complex with **5** (A) and **5-06** (B); relative intensity scale was fixed to that of the full scan spectra enabling direct comparison of the relative intensity between the two complex species. Lower intensity signals at higher m/z compared to the main peak consist in typical sodium, potassium and ammonium adducts.

No peak corresponding to the mass of NC without zinc was detected after incubation with each compound, thus ruling out that the inhibitory activity observed may derive from ejection of Zn(II) of NC. By tandem MS analysis, the compound-protein peaks dissociated yielding the mass corresponding to the native NC form, suggesting the non-covalent character of the interaction (data not shown).

For both **BN5** and **BN5-06**, a peak corresponding to the sum of the masses of NC protein and compound species was detected, indicating the formation of a 1:1 complex (**Figure 4.4**). This result confirmed the interaction of selected inhibitors with NC and corroborated the molecular modelling results. Hence, these molecules effectively inhibited NC chaperone activity by direct interaction with the protein. As observed by comparing the relative intensity of peak corresponding to the NC in complex with **BN5** and **BN5-06** (**Figure 4.4A** and **4.4B**, insets), the latter compound appeared to bind NC with greater affinity compared to the precursor molecule. By quantification of the partitioning between the free and bound fraction of NC (see **Material and methods**), we found a fractional occupancy of (0.06 ± 0.02) for the precursor compound

BN5 and (0.16 ± 0.04) for the analogue **BN5-06** when a ten-fold compound concentration was incubated with NC. These results indicated that **BN5-06** binds NC with two to three-fold higher affinity the target protein compared to **BN5**, in line with the greater inhibitory activity observed by NAME assay.

To compare the affinity of different analogues of **BN5**, the interaction of selected molecules with NC was assessed in a binding competition assay format. NC and compounds **BN5**, **BN5-01**, **BN5-06**, **BN5-07**, **BN5-15**, all showing varying degrees of inhibition by NAME assay, were incubated together at the same concentration and analysed by ESI-MS. These experimental conditions promote competition among ligands that bind to the same site so that the relative intensities of the peaks are directly related to the relative affinity of the different ligands for the protein.

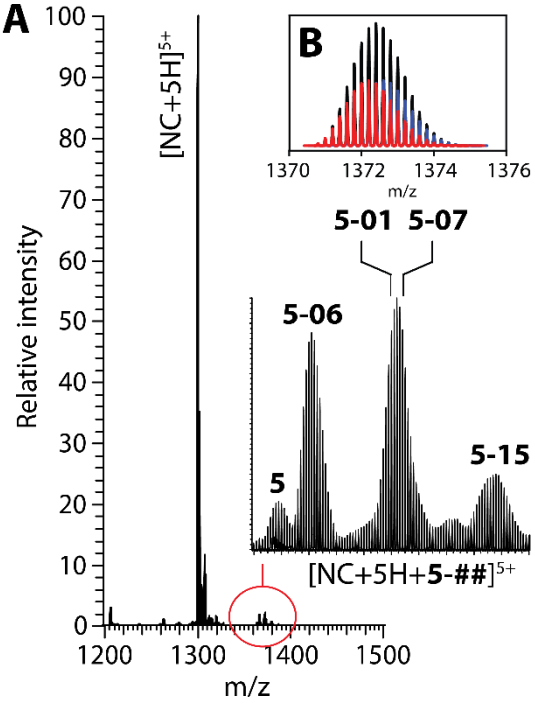


Figure 4.5. ESI-MS binding competition assay with BN5, BN5-01, BN5-06, BN5-07 and **BN5-15.** Spectra were acquired after incubation of NC with a 1-fold concentration of each compound (see **Materials and methods**) (A). Enlarged spectra focusing on the protein-compound adducts region. Lower intensity signals at lower m/z compared to main peaks consists in typical sodium, potassium and ammonium adducts (B). Calculated isotopic distribution of NC in complex with **BN5-01** (red) and **BN5-07** (blue); the combination of the two signals was calculated by the sum of the intensity of the two signals (black) assuming a similar affinity for the two compounds (C).

By ESI-MS competition assay, a signal from each protein-compound complex was readily observable within the same spectra, with the only exception of those containing the nearly isobaric species **BN5-01** and **BN5-07** (**Figure 4.5A**). For the latter compounds, two partially overlapping signals distributions were observed. In order to discriminate contributes

from each compound-protein complexes to the overall signal, we calculated the theoretical distribution of the two distinct species (blue and red traces in **Figure 4.5B**) and merged setting the affinity of the two compounds arbitrary equal (black trace in **Figure 4.5B**). After this partitioning operation, we were able to estimate the actual scale of NC affinity for the tested compounds. The results showed that **BN5-06**, **BN5-07** and **BN5-01** have a high affinity for NC while **BN5** and **BN5-15** bind the protein to a much lower extent. Based on these results, the affinity ranking of the tested compounds was **BN5-06** > **BN5-01** ≈ **BN5-07** > **BN5-15** > **BN5**.

Overall, the affinity of the compounds for NC protein is well consistent with their NC inhibitory activity. A major improvement in the binding affinity for NC protein was observed comparing the precursor compound **BN5** to the analogue **BN5-06**, which also appeared one of the best binders among other analogues series. Considering the ability of such molecules to inhibit the activity of NC(12-55) in the NAME assay,[61] binding of the benzoxazolinone compounds to the full-length NC observed by ESI-MS supports the expected MOA consisting in the competition with guanine nucleobases for binding to the ZFs domain of NC.

4.2 Evaluation of BM compound series

4.2.1 Inhibition of NC nucleic acid chaperone activity by BM compounds

Results from the HTS assay of the BM series highlighted four inactive compounds (**BM1**, **BM2**, **BM3** and **BM10**) for which no fluorescence decrease was observed. On the opposite, an upward trend was observed especially for the **BM3** molecule when tested using cTAR as substrate (supplementary information, **Figure S5**), which was likely related to the precipitation of the compound into the reaction well. **BM5** was not analysed since it appeared poorly soluble in the assay conditions prior to the analysis. Three compounds (**BM4**, **BM7** and **BM8**) showed to barely affect the NC function as a little downward trend in the fluorescence signal was observed at higher compound concentration (supplementary information, **Figure S5**). Clearly, two compounds (**BM6** and **BM9**) appeared the most potent inhibitors of the series displaying a dose-dependent decrease in the fluorescence signal observed for both TAR and cTAR substrates (supplementary information, **Figure S5**). The two positive hits were found having similar potency with IC_{50} value of (15.03 ± 4.19) µM and (12.76 ± 1.61) µM using TAR as substrate and (13.96 ± 0.64) µM and (8.21 ± 0.42) µM using cTAR as substrate for **BM6** and **BM9**, respectively (**Figure 4.6** and **Table 4.2**).

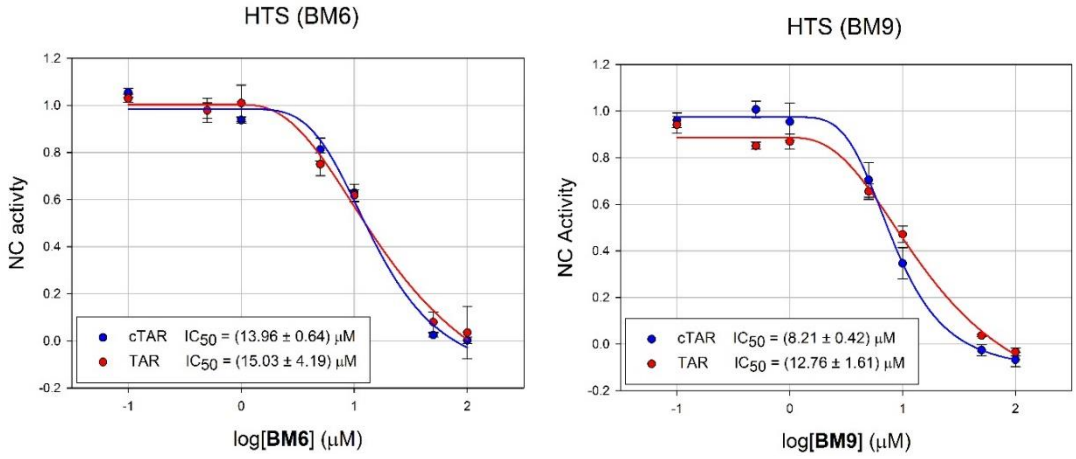


Figure 4.6. HTS assay employing TAR and cTAR substrates conducted with BM6 and BM9. Inhibition of the NC-induced TAR (red) and cTAR (blue) destabilization by **BM6** (left panel) and **BM9** (right panel). Lines correspond to the best fit to the experimental data using Equation 1 (see **Materials and methods**). A clear dose-dependent inhibition of the NC nucleic acid destabilization properties was observed by a decrease in the fluorescence signal from the NC-oligonucleotide mixture.

By NAME assay, compounds **BM1** and **BM2** showed no activity against the NC-mediated TAR/cTAR hybrid formation as no change in the band intensity with the lowest electrophoretic mobility was observed (**Figure S6**). Although the little activity observed by HTS assay, NAME revealed no activity for compounds **BM4**, **BM7** and **BM8** in the

concentration range used, indicating that these compounds weakly inhibits TAR and cTAR destabilization and do not impair the NC-mediated annealing. **BM5** and **BM10** were not analysed due to their low water solubility. **BM6** showed to interfere with the annealing process as several additional bands were observed having intermediate electrophoretic mobility between those corresponding to the monomeric and duplex species, which can be considered as partially annealed species. Although no complete inhibition of the hybrid duplex formation was observed under our experimental conditions, **BM6** likely reduced the rate of TAR/cTAR annealing reaction by interfering with the chaperone properties of NC (supplementary information, **Figure S6**). Differently, **BM9** was highlighted as the best inhibitor of the series by NAME assay, displaying complete inhibition of the NC chaperone activity. Bands corresponding to the monomeric TAR and cTAR species clearly appeared at higher compound concentration while the band corresponding to the hybrid duplex was strongly impaired with an IC_{50} value of (95.4 ± 5.5) μM (**Figure 4.7** and **Table 4.2**).

Based on the results obtained by HTS and NAME assays, **BM6** and **BM9** were highlighted as hit compounds among the BM series. Their activity was investigated further in order to elucidate their MOA.

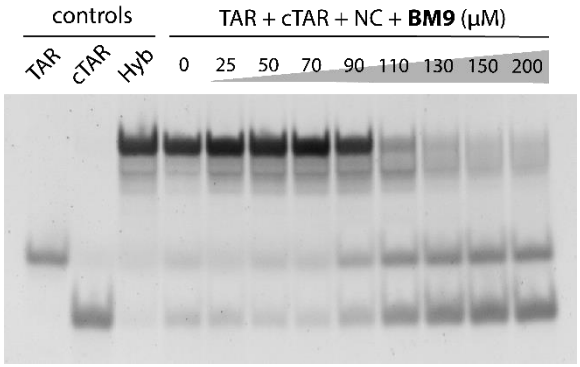


Figure 4.7. NAME assay conducted with BM9. Monomeric TAR and cTAR, as well as the heat-refolded hybrid duplex (Hyb), were included as controls. A clear dose-dependent inhibition of the NC nucleic acid chaperone properties is observed by a decrease in the formation of the band corresponding to the hybrid duplex specie with lower electrophoretic mobility compared to the TAR and cTAR monomeric species (see **Materials and methods**).

ID	HTS IC_{50} (μM)		NAME IC_{50} (μM)
	TAR	cTAR	
BM1	N.A.	N.A.	N.A.
BM2	N.A.	N.A.	N.A.
BM3	N.A.	N.A.	N.A.
BM4	Barely active	Barely active	N.A.
BM5	N.D.	N.D.	N.D.
BM6	15.03 ± 4.19	13.96 ± 0.64	Barely active
BM7	Barely active	Barely active	N.A.
BM8	Barely active	Barely active	N.A.
BM9	12.76 ± 1.61	8.21 ± 0.42	95.4 ± 5.5
BM10	N.A.	N.A.	N.D.

Table 4.2. Activity ranking (IC_{50}) of compounds from the BM series assessed by HTS and NAME assays. N.A. = not active; N.D. = not determined

4.2.2 Inhibition of the NC-oligonucleotide interaction by BM compounds

When NC was let free to interact with the receptors in presence of **BM6** and **BM9** by SPR-competition assay, a clear dose-dependent decrease of the SPR signal was observed, indicating lower binding of the protein to the immobilized oligonucleotides (**Figure 4.8**). These results highlighted the ability of the two BM hits to inhibit binding of NC to the TAR and cTAR sequences. By SPR-competition assay, **BM6** displayed a similar potency when tested against the two oligonucleotide receptors, with IC_{50} values of (1.20 ± 0.03) μM and (1.50 ± 0.02) μM using the two substrates, respectively. Differently, **BM9** showed slightly different potencies yielding IC_{50} values of (1.00 ± 0.02) μM and (1.93 ± 0.06) μM using TAR and cTAR, respectively.

These results were in line with the inhibitory activity observed for **BM6** and **BM9** by HTS and NAME assays. Preventing NC interaction with the TAR and cTAR structure, the two hit molecules prevented the functions of NC thus explaining the inhibition of the stem destabilization and the annealing reaction observed by HTS and NAME assays, respectively. By analysis of the quantitative activity data obtained for **BM9** by the different assays, we observed that the IC_{50} values obtained by both HTS and NAME assays were both close to the value of 10-fold compound ratio based on the protein concentration used in each specific assay. A similar result was also observed by the SPR-competition assay, suggesting that the results yielded by the different assays were consistent with each other. Therefore, the inhibition of NC chaperone

activity by **BM9** likely resided on the inhibition of protein binding to the TAR and cTAR oligonucleotides.

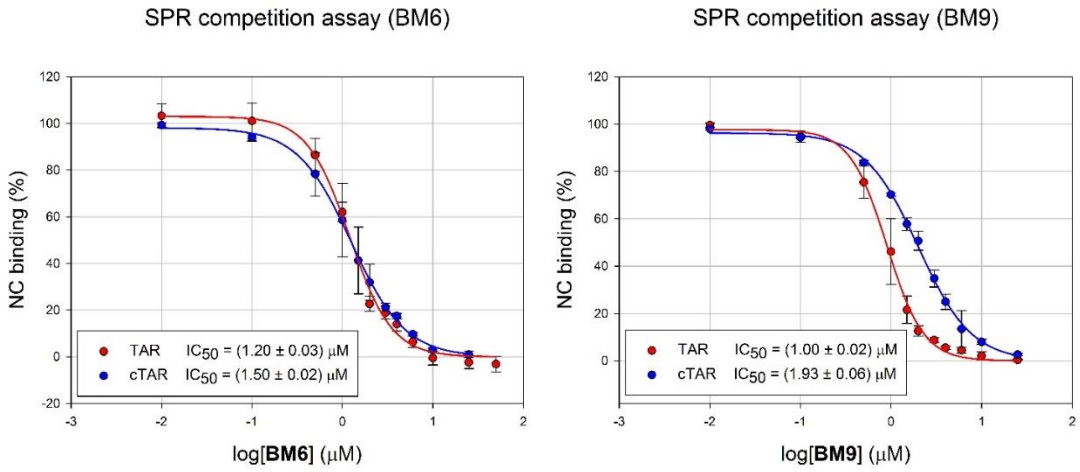


Figure 4.8. SPR-competition assay conducted with BM6 and BM9. Inhibition of the NC binding to TAR (red) and cTAR (blue) by **BM6** (left panel) and **BM9** (right panel). Lines correspond to the best fit to the experimental data using Equation 1 (see **Materials and methods**). A clear dose-dependent inhibition of the NC binding to both TAR and cTAR was observed by a decrease in the SPR signal obtained by flowing a mixture of NC in presence of increasing compounds concentration onto the oligonucleotide-functionalised sensor chip.

SPR technique has the great advantage of high sensitivity enabling analysis of binding interactions in the nanomolar concentration range and permits to efficiently quantify the inhibitory activity of tested molecules on the binding interaction by competition assays. However, SPR requires the immobilization of the receptor molecule that may potentially constrain important molecular motions that might be relevant to the binding process. In order to confirm qualitatively the data observed by the SPR-competition assay, we set EMSA experiments employing the same protein and oligonucleotide substrates. Moreover, the experimental condition adopted for EMSA closely matched those used for the NAME assay (e.g., concentration of oligonucleotide substrates and NC to nucleotides molar ratio) permitting a direct comparison of the results obtained.

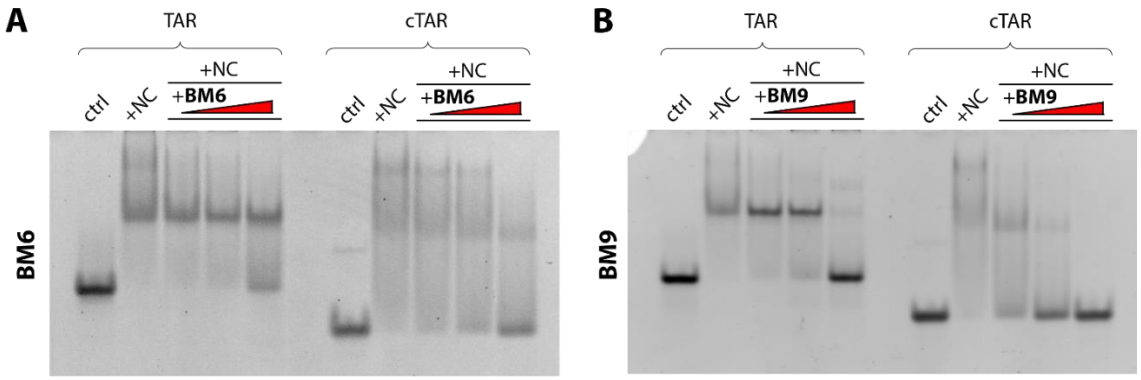


Figure 4.6. EMSA conducted with BM6 and BM9. TAR and cTAR controls (ctrl); TAR/NC and cTAR/NC complexes controls (+NC). Addition of **BM6** (A) and **BM9** (B) (50, 100 and 200 μM) to the protein-oligonucleotide complexes (see **Materials and methods**). Incubation of TAR and cTAR with NC led to a major band shift to lower electrophoretic mobility compared to free oligonucleotides, likely corresponding to the formation of a 1:1 complex. Higher order assemblies were also observable at lower electrophoretic mobility although they appeared less resolved. Addition of **BM9** led to complete disruption of NC-oligonucleotides interaction at high concentration for both TAR and cTAR substrates. **BM6** appeared instead less potent; nevertheless, a dose-dependent inhibition was still observable.

EMSA enabled detection of NC-oligonucleotides interaction (**Figure 4.6**). In the case of the TAR sequence, a major band shift, likely corresponding to the formation of a 1:1 complex by binding of NC to a high-affinity site, was observed. Higher order protein-nucleic acid complexes were also observed by the additional shift of the band to lower electrophoretic mobility suggesting that TAR substrate afforded additional binding sites with lower affinity (compare ctrl and +NC lanes in **Figure 4.6A** and **4.6B**). Similar results were obtained in the case of cTAR. However, little difference was observed between the intensity of the bands corresponding to the 1:1 and higher stoichiometry assemblies, suggesting the presence of multiple binding sites with similar affinities.

The results obtained for **BM6** and **BM9** by EMSA, clearly indicated that both compounds inhibited the NC-oligonucleotide interaction involving TAR and cTAR substrates. Complete inhibition of NC-oligo interaction was observed at high **BM9** concentration in the case of cTAR whereas little presence of protein-oligo complex was still observed in the case of TAR (**Figure 4.6B**). Differently, **BM6** showed to inhibit NC-oligo interaction with lower potency as protein-complexed oligonucleotide was still present at high compound concentration (**Figure 4.6A**). **BM9** displayed greater potency than the **BM6** in preventing NC-oligonucleotide interaction, in line the results obtained by NAME assay. Altogether, SPR-competition assay and EMSA indicated that **BM6** and **BM9** inhibited NC binding to oligonucleotides, a mode of action compatible with that predicted *in silico*.

4.2.3 Binding properties of BM compounds to NC

By ESI-MS analysis, no peak was detected at the calculated mass corresponding to the protein-compound complex for inactive compounds from the BM series (see **Table 4.2**). These results indicated that, in line with NAME and HTS assays, **BM1**, **BM2**, **BM3** and **BM10** did not interact with the protein and therefore had no NC inhibitory activity. For **BM4**, **BM7** and **BM8**, which showed to be barely active by HTS assay, no binding to the protein was observed. On the opposite, different signals corresponding to the NC-compound complexes were detected incubating NC with **BM6** and **BM9** compounds (**Figure 4.7**). For the latter compounds, we observed a peak corresponding to the sum of the masses of the intact NC protein and one compound molecule, indicating that the two BM hit molecules form 1:1 complexes with NC. Additional peaks were also detected corresponding to complexes that involved NC lacking one Zn(II) ion (zinc-less NC). In the latter cases, higher stoichiometry complexes were observed displaying up to three compound molecules bound to the zinc-less NC. Based on the shift from the predicted mass values, oxidized forms of the free cysteine residues of NC were observed such as the formation of disulfide bonds (-2 Da) and cysteic acid residues (+48 Da).

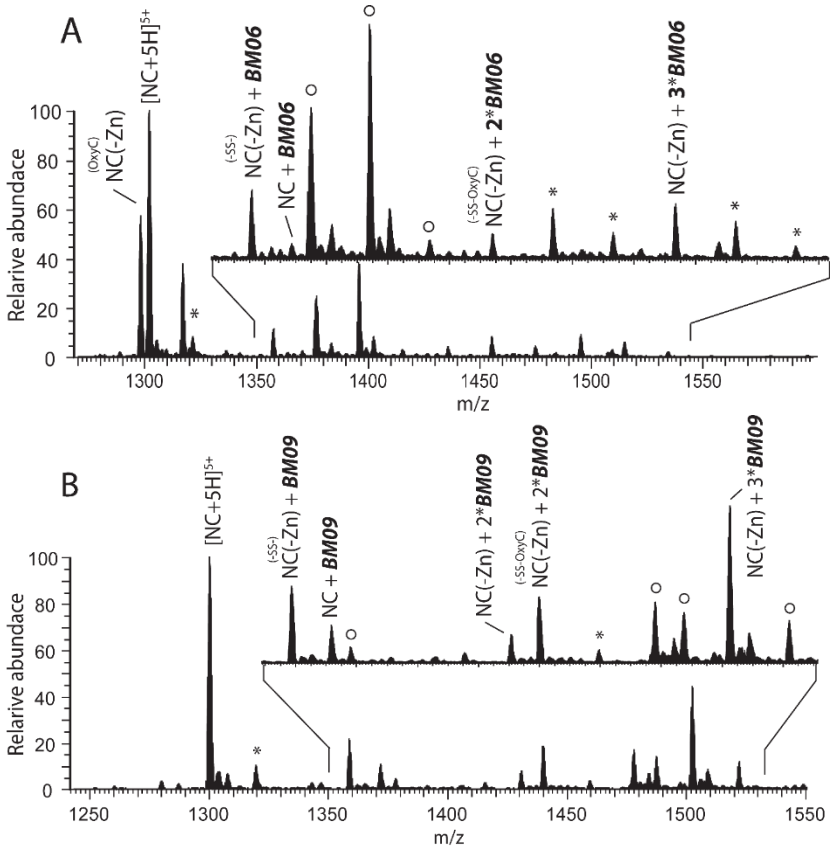


Figure 4.7. ESI-MS spectra of NC in presence of compound BM6 and BM9. Spectra obtained after incubation of NC with a 10-fold concentration of **BM6** (A) and **BM9** (B) (see **Materials and Methods**). For the sake of clarity, only the region containing the quintuply charged ions that represented the most

abundant species of the spectra is shown. NC, native NC protein; NC(-Zn), zinc-less NC; –SS–, formation of a disulfide bond; OxyC, oxidation to sulfonic acid; ○, unidentified peaks related to an impurity found in the compound tested; *, sulfuric acid adduct. Both **BM6** and **BM9** yielded the formation of a 1:1 complex with NC. However, complexes of compounds and zinc-less NC appeared to be predominant, indicating that the compound promoted ejection of one Zn(II) ion. After zinc ejection, oxidation of the free cysteine to disulfides and cysteic acid occurred, especially in the case of **BM6**. The zinc less-NC afforded multiple binding sites and complexes up to 3:1 compound-NC stoichiometry were observed. Low-intensity signals observed at slightly higher m/z than the corresponding NC signal consist of typical sodium, potassium and ammonium adducts.

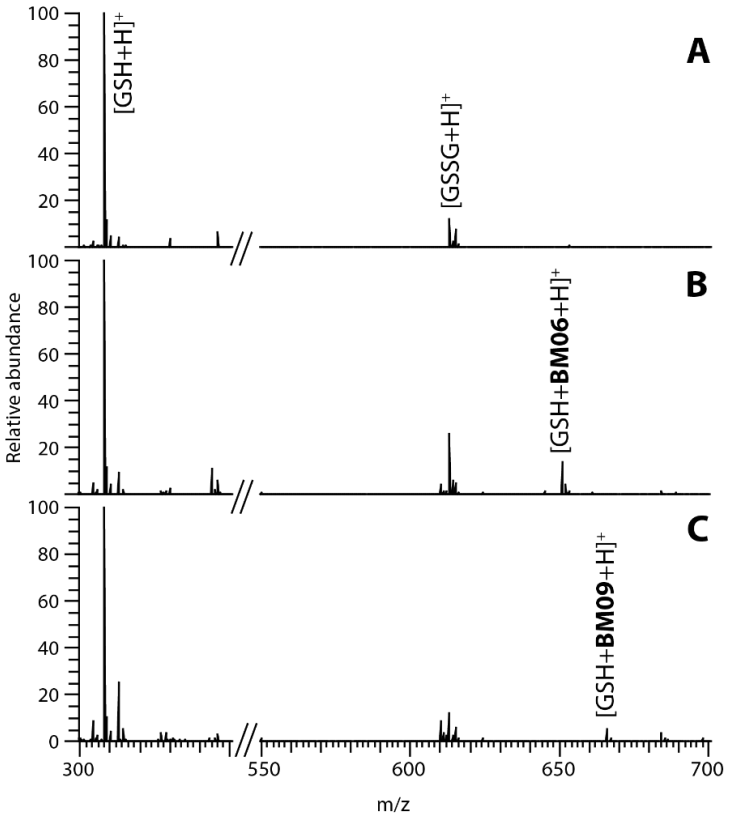
The MS results showed in **Figure 4.7** indicated that **BM6** and **BM9** interacted with NC and subsequently promoted the ejection of the divalent ion from one of the two finger motifs. Detection of compounds bound to the zinc-less protein suggested that such interaction is not mediated by the predicted binding mode involving the hydrophobic pocket located at the top of the zinc fingers of NC. The results resembled the mechanism of action of the reported DIBAs compounds (see **Chapter 1.5.5**).[93] Similarly, we hypothesized that **BM6** and **BM9** might react covalently with the zinc chelating cysteine residues. Interestingly, the 3:1 stoichiometry observed for the compound-protein complexes matched the number of cysteine residues in each zinc finger motif indeed. Moreover, both **BM6** and **BM9** molecules were characterized by a sulfur atom that could potentially react with the thiol group of cysteine residues (**Figure 2.2**).

Ejection of Zn(II) from NC can be easily detected by the characteristic mass shift (-64 Da), therefore mass spectrometry was a suitable technique to probe the mechanism of action of zinc-ejecting agents.[93] Two experiments were designed to elucidate further the interaction of **BM6** and **BM9** with NC. Firstly, to fully prevent any compound-NC interaction mediated by the zinc finger hydrophobic pocked, we evaluated the binding of **BM6** and **BM9** to the fully denatured NC (apo-NC), where any secondary structure was eliminated by ejection of both Zn(II) ions from the ZFs. Secondly, we assessed potential covalent reaction between the compounds and a model thiol group using reduced glutathione, which bears one free thiol group mimicking the free cysteine residues of NC after ejection of zinc.

To assess whether the interaction of **BM6** and **BM9** with NC required the correct folding of the zinc fingers domain, ESI-MS spectra of NC in presence of the compounds were acquired after denaturation of NC structure with a mixture water/methanol/acetic acid 49:49:2. The organic solvent together with the acidic pH promoted ejection of both Zn(II) ions and complete unfolding of protein structure. Therefore, any non-covalent interaction mediated by the hydrophobic pocket of NC was lost under these denaturing conditions. In the latter experiment, protein-compound complexes were still observed confirming that the binding did not involve the hydrophobic pocket of NC (supplementary information, **Figure S7**). Interestingly, higher compound-protein stoichiometry up to four **BM9** molecules bound to apo-

NC was found, indicating that additional binding sites were available after the loss of both Zn(II) ions.

ESI-MS spectra of an equimolar mixture of reduced glutathione (GSH) with either **BM6** or **BM9** showed a new peak corresponding to the sum of the masses of GSH and the compound (**Figure 4.8**). As observed for the NC-compound assemblies, we did not observe the characteristic -2 Da mass shift typical of disulfide formation; therefore, we were unable to assess the covalent or non-covalent nature of such interaction based solely on the measured mass values. We thus used tandem MS to investigate the nature of the compound-glutathione interaction, which was previously not feasible for the larger NC-compound complexes. Isolation and activation of ions corresponding to the GST-BM assemblies (see **Materials and methods**) yielded a series of fragments that unambiguously indicated the formation of a disulfide bond between the thiol group of GST and the sulfur atom in the thiobarbituric ring of **BM6** and **BM9** (**Figure 4.9**).



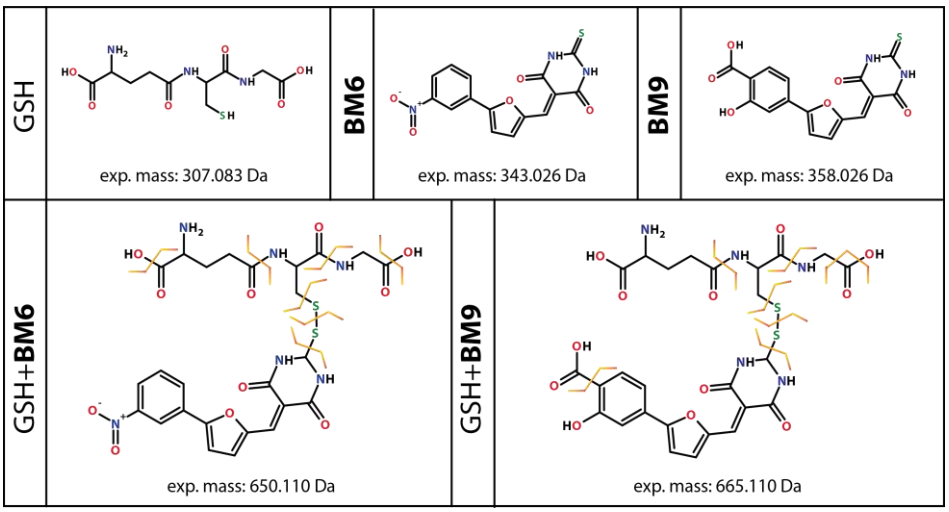


Figure 4.8. ESI-MS spectra of GSH in presence of BM6 and BM9 and structures of the detected complexes. ESI-MS analysis of 10 µM reduced glutathione alone (A) and after incubation with a 1-fold concentration of **BM6** (B) or **BM9** (C) compounds for two hours at room temperature in ammonium acetate 150 mM, pH 7.5. Presence of **BM6** and **BM9** led to the formation of a new peak corresponding to the mass of the GSH-compound complex. Bottom panel: structure and experimental masses of GSH, **BM6**, **BM9** and the detected GSH-BM complexes. Similarly to what observed for the NC-compound complexes, the assemblies formed by **BM6** and **BM9** with glutathione were characterized by masses that matched the exact sum of the two components. Isolation and activation of the latter species yielded series of fragment ions that unambiguously identified the formation of a disulfide bond. Sites of fragmentation are indicated by yellow lines.

Altogether, ESI-MS results indicated that **BM6** and **BM9** interacted with both NC and apo-NC proving that such interaction is not dependent on the hydrophobic pocket formed at the top of the ZFs domain. The stoichiometry observed suggested that the binding interaction was related to the number of free cysteine residues instead. Only one ligand molecule was observed bound to NC protein whereas up to three ligands were accommodated by the zinc-less protein which exposes three cysteine residues (**Figure 4.7**); higher stoichiometry was then observed when the second Zn(II) ion was ejected, at least in the case of **BM9** (supplementary information, **Figure S7**). Binding of the two ligands to the thiol group of glutathione indicated that both **BM6** and **BM9** engage disulfide bonds with cysteine residues without the characteristic -2 Da mass shift (**Figure 4.8**).

Combining all these information, we depicted the MOA of the two BM hit compounds (**Figure 4.9**). One ligand molecule first binds NC protein, probably interacting with the zinc fingers domain (**Figure 4.9A-B**), placing the thiobarbituric ring in close proximity to the zinc coordination centre. Subsequently, the thiourea sulfur atom in the thiobarbituric moiety reacts with one of the chelating thiol groups from the ZFs motifs forming a covalent disulfide bond (**Figure 4.9C**). As a consequence, the Zn(II) ion is released leading to the unfolding of the ZF. The other chelating cysteine residues thus became available for reaction with additional ligand molecules. If available, the cysteine residues from the second ZF may react similarly, as observed in experiments conducted under denaturing conditions to intentionally unfold both ZFs of NC (**Figure 4.9D**).

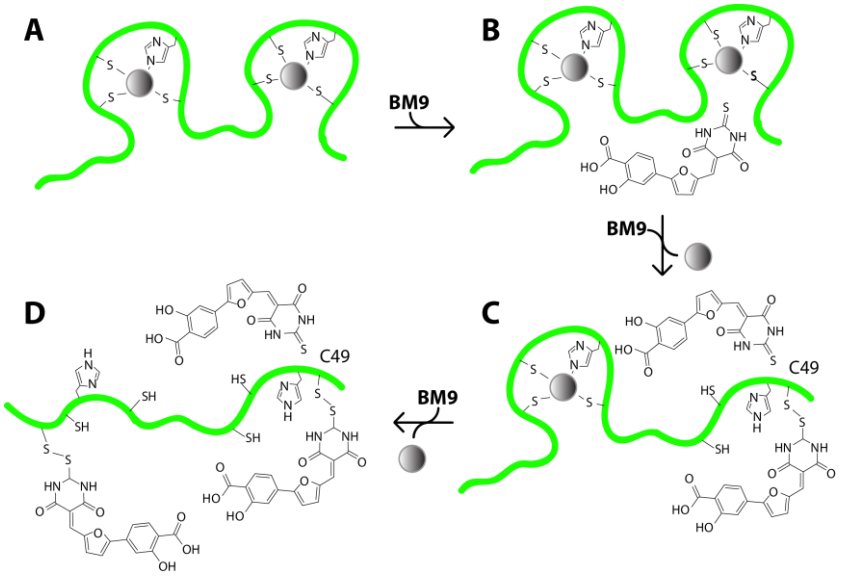


Figure 4.9. Proposed mechanism of action of the BM9 compound. NC is represented as a green cartoon and zinc ions as grey spheres; C49, cysteine residue 49. Binding of the compound to the protein (A-B); formation of a covalent disulfide bond and ejection of Zn(II) by the compound (B-C); unfolding of the second ZF and alkylation of NC by additional compound molecules (C-D).

A potential reaction mechanism between the thiobarbituric ring and thiol groups may resemble the mechanism of disufides formation from thiols and thiourea, by donation of the lone pair electrons from the neighbouring thiourea nitrogen atoms.[153] The C-terminal zinc finger motif (ZF2) is a good candidate to be the primary target of **BM6** and **BM9** compounds as it is more susceptible to the attach of zinc ejecting agents,[154] although the two zinc fingers share a similar structure.[155] Specifically, the cysteine at position 49 (C49, **Figure 4.9C**) appeared the less protected residue and has been shown to be the most reactive among the chelating cysteines.[93, 150, 156]

The complex between the zinc-bound NC protein and the BM ligand can be assumed to be an intermediate specie before cysteine reaction and zinc ejection as the peak intensity observed decreased by prolonging the incubation time before ESI-MS analysis (data not shown). In the latter case, the detected complex would correspond to the formation of a pre-reaction complex, where the ligand first poses onto the ZF2 with a specific orientation of the reactive moiety toward the zinc-chelating centre. This scenario has been proposed as mechanism of action for other NC alkylating agents such as PATEs, DIBAs and transition metal complexes (**Figure 1.9**).[90,91, 157, 158] Nevertheless, the peak corresponding to the sum of NC and one ligand molecule detected by ESI-MS could also correspond to the alkylated protein that still retains the zinc ion since we showed that alkylation of NC by **BM6** and **BM9** does not change the mass of the complex.

The results obtained by mass spectrometry analysis of compound-protein interaction supported the results of the biological evaluation of BM compounds. However, ESI-MS experiments depicted a different scenario from the predicted mechanism of action based on binding competition with the nucleic acid to the hydrophobic pocket of NC. Unfolding of ZFs by **BM6** and **BM9** disrupts the nucleic acid binding and chaperone activity of NC,[61] as observed experimentally by SPR-competition and NAME assays, respectively (**Figures 4.8** and **4.7**). According to this MOA, the IC_{50} values obtained by HTS and SPR-competition assays using the two different nucleic acid substrates, namely TAR (RNA) and cTAR (DNA), were found reasonably similar. Under binding competition conditions for binding to NC instead, different IC_{50} values would be expected[76, 79] due to the different affinity and activity of NC for the two oligonucleotide substrates.[159] ESI-MS experiments thus afforded a comprehensive biophysical investigation of the mechanism of action of the BM compounds, permitting also a more accurate interpretation of our biological data.

5. RESULT AND DISCUSSION

2,6-DIPEPTIDYL-ANTHRAQUINONE CONJUGATES AS INDIRECT NC INHIBITORS

As already discussed in the **Aim of the work**, another valuable approach to inhibit the NC chaperone functions is to target its nucleic acid substrates. Freezing the dynamics of nucleic acids secondary structure may indeed prevent the important conformational changes induced by NC that are necessary during the HIV-1 replication cycle. Viral RNA sequences that have been proposed as a possible target for nucleic acid binder inhibitors of NC are TAR, PBS, DIS and Ψ-RNA.[76-79, 125-128] The mechanism and function played by NC on these viral nucleic acid substrates have been discussed in details in **Chapter 7**. This section focuses on 2,6-dipeptidyl-anthraquinone compounds developed in our lab as binders of nucleic acids substrates of NC, which showed to inhibit its chaperone activity *in vitro*.

Series of 2,6-dipeptidyl anthraquinones were discovered by screening of an in-house chemical library encompassing different groups of compounds, which led to the identification of two interesting chemotypes. Along with the identification of natural polyphenolic C-glucoside ellagitannins,[77] the dipeptidyl-anthraquinone conjugates were highlighted as efficient inhibitors of the NC chaperone activity by blocking the ability of the protein to destabilize the secondary structures of TAR and cTAR. Structures of anthraquinone conjugates have been discussed in **Chapter 2.2** and are shown in **Figure 2.3**.

We expanded the study on these class of molecules creating different compounds series (GSF and G series) and investigated their MOA. AQs were shown to inhibit NC chaperone activity by over-stabilization of the TAR and cTAR stem-loop structures. Acting as threading intercalating agents, these compounds preferentially targeted dynamic nucleic acid secondary structures such as loops and bulges.[160, 161] For this reason, structured regions in the viral genomic RNA leader sequence (see **Chapter 1.1.1**) represent interesting targets for this class of compounds as potential antiviral agents. By testing different side chains, we identified preliminary structural requirements affording efficient inhibition of NC. We found that linear aliphatic side chains, characterized by 11-atom length and two positive charges provided good inhibitory activity. Shorter chains, as well as bulkier terminal residues, reduced the potency of AQs. Interestingly, aromatic residues appeared tolerated as side chain terminal residues, even if to a limited extent.[76]

Later, we observed that introduction of a chiral linker such as L-alanine (LA series) had a remarkable effect on the activity of these compounds, indicating that side chain orientation

also represented an important structural determinant for the NC inhibitory properties of AQs.[79] It appeared clear that compounds from the LA serie afforded lower activity, when compared to a similar compound from the G series (i.e., having a chiral and a non-chiral linker, respectively). These results pushed us to further investigate this chemical feature by creating a new series of conjugates.

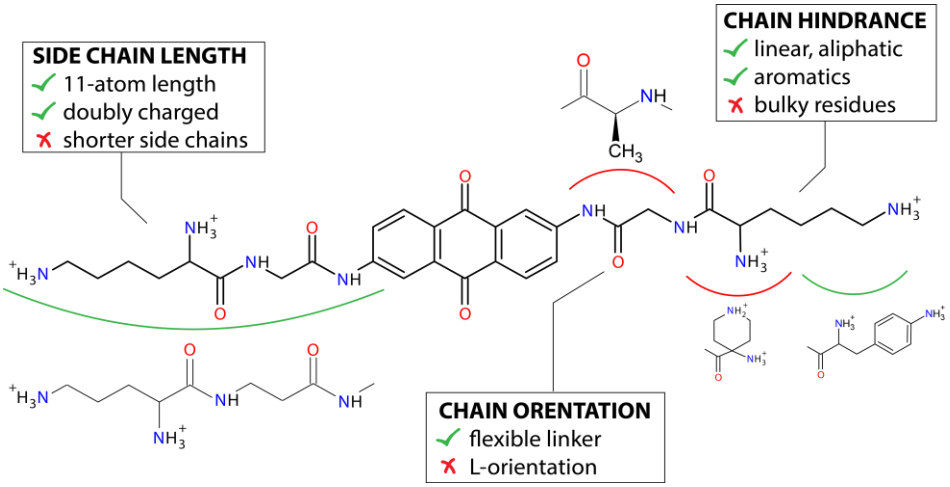


Figure 5.1. SARs for the inhibition of NC chaperone properties by AQs compounds known at the time of this study.[76,79] The symmetric side chains attached to the anthraquinone scaffold had an optimal length of eleven atoms with cationic terminal residues characterized by two positive charges separated by three to four carbon atoms. Linear aliphatic substituents provided the proper steric hindrance. Bulkier substituents such as cyclic aliphatic amines were detrimental for the activity although aromatic residues appeared tolerated in these positions. Introduction of a chiral linker residue such as L-alanine resulted detrimental for the activity. Although the new linker introduced an additional methyl group that might impair the activity due to its steric hindrance compared to the glycine linker, the L-alanine linker nevertheless provided a specific orientation to the side chains. The effect of this chemical feature was investigated further in this thesis work by studying the DA compound series bearing the novel D-alanine linker.

Novel AQs were designed having a D-alanine peptide linker (DA series), which provided an opposite configuration to each side chain compared to the compounds from the LA series. The results presented here aim to a direct comparison of the DA compounds series with the G and LA series, in order to evaluate the effect of side chain orientation on the NC inhibitory activity of these compounds. Moreover, special attention was also pointed to compounds bearing the tyrosine terminal residue. This new terminal moiety, not previously studied, was specifically introduced to better investigate the effect of aromatics into the side chains of AQs by studying compounds **G14**, **LA14** and **DA14** (**Figure 2.3**).

Prompted by the flexible nature of NC,[159] which confers to this protein the ability to interact with several nucleic acid substrates that take part to important steps of the virus replication cycle, we started here to investigate the effect of AQs on novel viral RNA substrates.

Starting from our knowledge on the chemical features found for optimal inhibition of the NC activity on the TAR/cTAR model system, we report here the result obtained from our study on the ability of AQs to inhibit the HIV-1 primer-annealing step, which involves the HIV-1 PBS and the RNA annealing activity of NC.

5.1 Effect of side chain orientation in AQs binders of the HIV-1 TAR and cTAR sequence

We compared directly the result obtained from the DA compounds to the previously reported LA series by conducting the same *in vitro* biological assays.[76, 79] The HTS assay was used to assess the inhibitory activity of compounds on the NC-mediated nucleic acid destabilization properties whereas the inhibitory activity of NC chaperone activity was analysed more in detail by means of NAME assay. The stabilizing ability of AQs on the TAR and cTAR stem-loop structure was then investigated by means FQA to determine the melting temperature of the nucleic acid secondary structure.

5.1.1 Inhibition of NC nucleic acid destabilization properties by DA compound series

The ability of compounds to inhibit the NC-mediated destabilization of TAR and cTAR stem-loop structures was assessed by HTS assay (see **Materials and methods**). Within the DA series, linear aliphatic side chains with doubly charged terminal residues (compounds numbered 2, 6, 7 and 8) supported different inhibitory activity according to the length of the side chain. The **DA2** compound, characterized by an 11-atom side chain in length, turned out as the most potent inhibitor of the series by HTS assay. Shorter side chains inhibited NC activity to a lower extent. Accordingly, compounds with linear aliphatic side chain were ranked as **DA2** > **DA6** > **DA7** >> **DA8**, strongly indicating a negative trend as the side chain length decreases. In line with the results from the previous G and LA series,[79] proper side chain length thus appeared as a key determinant for the NC inhibitory activity of these compounds. Compound with bulkier substituents (**DA3**, **DA12** and **DA13**), as well as singly charged and short side chains (**DA1**), showed lower potency. These data further supported the limited steric hindrance allowed in this series of inhibitors and the importance of the electrostatic interaction with the phosphate backbone of the oligonucleotide substrates (**Table 5.1** and **Figure 5.2**).

DA-series	TAR IC_{50} (µM)	cTAR IC_{50} (µM)	DA-series	TAR IC_{50} (µM)	cTAR IC_{50} (µM)
DA1	33.74 ± 2.05	19.98 ± 1.58	**DA8**	49.94 ± 3.11	46.25 ± 1.32
DA2	1.59 ± 0.39	1.41 ± 0,14	**DA11**	47.36 ± 1.34	20.17 ± 0.78
DA3	10.28 ± 1.00	9.12 ± 0.56	**DA12**	51.33 ± 5.01	32.08 ± 2.76
DA6	5.21 ± 0.66	3.99 ± 0.22	**DA13**	39.70 ± 3.99	24.86 ± 2.20
DA7	9.24 ± 0.49	6.52 ± 0.47	**DA14**	40.33 ± 8.33	21.75 ± 3.42
G14	11.80 ± 0.29	15.17 ± 1.49	**LA14**	12.68 ± 2.40	13.03 ± 0.77

Table 5.1. Activity ranking (IC_{50}) of DA anthraquinone conjugates assessed by HTS assay. (see **Materials and methods**)

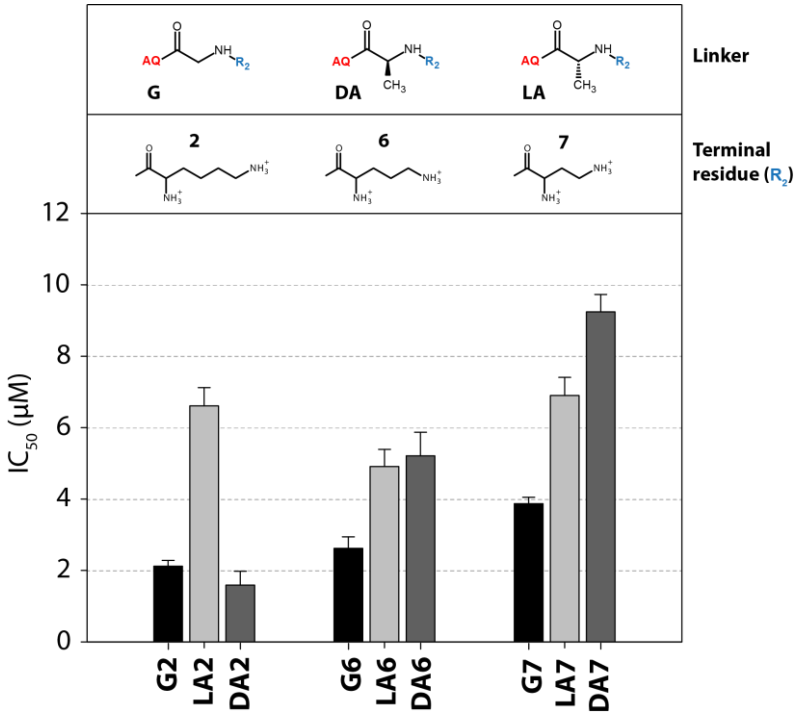


Figure 5.2. Effect of side chain orientation on the activity of linear aliphatic AQs conjugates. Plot showing the activity (IC_{50}) of linear aliphatic anthraquinone conjugates from different compound series obtained by HTS assay. Focusing on compounds with the lysine terminal residue (2), the side chain orientation showed to strongly affect the activity observed by HTS assay. In our previous works, the compound **LA2** appeared less active compared to the related compound **G2** indicating that the chiral linker provides a confrontation not favourable for optimal inhibition of NC activity. The compound **DA2** analysed here, which bear the same side chains but with opposite conformation compared to **LA2**, yielded a similar result to **G2**, indicating fully restoration of the activity loss due to the improper side chain orientation. These results highlighted a preferential side chain conformation for optimal interaction with the nucleic acid substrate to achieve the best inhibitory activity. Interestingly, the same steric effect was not observed for shorter and less potent compounds. The ornithine residue in **LA6** and **DA6** compounds afforded similar activity whereas the trend appeared reverted for compounds **LA7** and **DA7** characterized by a further shorter terminal residue. These results suggested that the activity of these compounds is finely tuned by a synergism between side chain length and orientation. Data for compounds from the G and LA series are taken from Sosic, A. et al, *J. Med. Chem.*, 59 (**2016**) 1914−1924.[79]

When comparing the results obtained for the DA series with the previously reported G and LA series, appeared clear that different orientations of the side chains have a strong impact on the biological activity of these compounds as inhibitors of NC (**Figure 5.2**). Considering compound with 11-atom linear aliphatic side chains (**G2**, **LA2** and **DA2**), which were among the most active compounds found in each series, the previous result showed that **LA2** is far less active compared to **G2**. The L-alanine linker was thus shown to impair the NC inhibitory activity of compounds.[80] By HTS assay, the compound **DA2** showed comparable activity to **G2**, strongly indicating that the D-alanine linker provides the desired side chain orientation for inhibition of NC activity. However, the fact that **DA2** did not appear more potent than the **G2**

molecule, which bears the flexible glycine linker, suggested that also the side chain flexibility might be important for the activity of these compounds. Differently, compounds with shorter side chains did not show the same steric features. Blocking the side chain orientation with a chiral linker in compounds with 10- and 9-atom chain length (**LA6**, **LA7**, **DA6** and **DA7**) led to lower activity, as observed comparing LA and DA compounds with the relative compound bearing the Gly linker (**G6** and **G7**) (**Figure 5.2**). The fact that the side chain orientation did not have the same effect on compounds with different side chain lengths suggested that there is an interplay between these two chemical features affecting the overall activity of compounds. The 11-atom side chain and D-orientation found in the compound **DA2** thus appeared as the most favourable molecular features in this class of inhibitors.

In our previous studies,[76, 79] for compounds bearing an aromatic terminal residue such as the p-amino phenylalanine of compound **G11** and **LA11** we expected a lower activity considering the steric hindrance of the aromatic ring. However, we observed that these compounds had a similar inhibitory activity to compounds bearing a linear aliphatic moiety. Based on this observation, we decided to better investigate the effect of aromatic substituents in these positions by expanding the series with compounds bearing the new tyrosine moiety (compound **G14**, **LA14** and **DA14**, **Figure 2.3**).

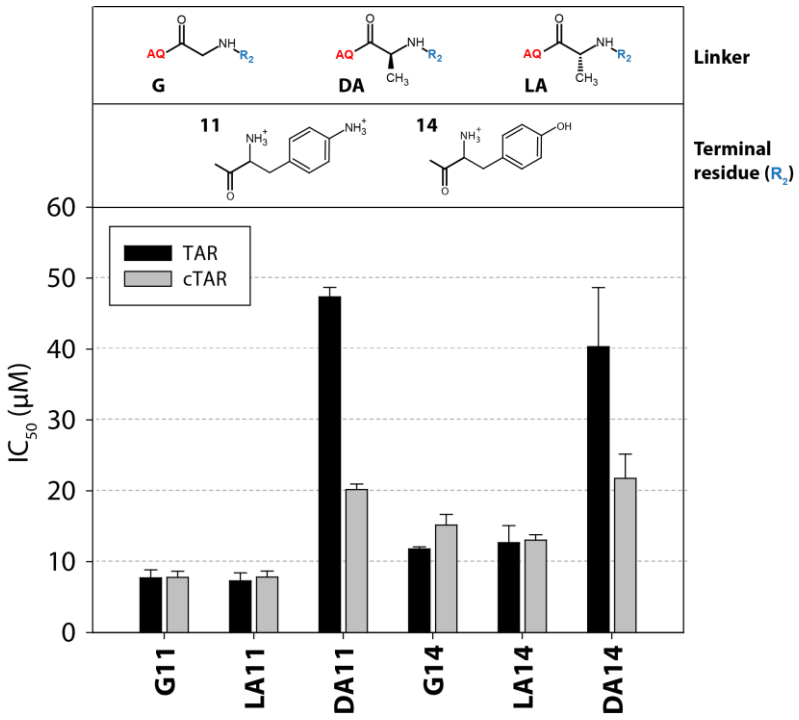


Figure 5.3. Effect of side chain orientation on the activity of aromatic AQs conjugates. Plot showing the activity (IC_{50}) of aromatic anthraquinone conjugates from different compound series obtained by HTS assay. Previous results highlighted that these compounds can efficiently inhibit NC despite the inherent steric hindrance afforded by the aromatic terminal residues compared to the linear aliphatic compounds

discussed above. Compounds **G11** and **LA11**, characterized by the p-amino-phenylalanine terminal residue, showed good inhibitory activity by HTS assay. The **DA11** compound tested here appeared far less active compared to the related compounds from the previous series. Similar results were obtained here analysing all the new conjugates bearing the tyrosine terminal residue (compounds **G14**, **LA14** and **DA14**). Similarly, the latter compounds from the G and LA series displayed similar activity whereas the DA conjugate was characterized by lower potency. Interestingly, the trend observed analysing the activity afforded by different linkers appeared reverted compared to that observed considering the linear aliphatic compounds described above (**Figure 5.2**). This observation suggested that the influence the side chain orientation is also modulated by the nature of the terminal residue, likely affording a different binding mode for aromatic conjugates compared to the other AQs studied here. Data for compounds **G11** and **LA11** are taken from Sosic, A. et al, *J. Med. Chem.*, 59 (**2016**) 1914−1924.[79]

Interestingly, compound **LA14** displayed similar activity to the **G14** whereas the compound **DA14** showed weaker potency by HTS assay (**Figure 5.3**). This behaviour was rather unexpected, as the activity trend appeared opposite to that observed for compounds bearing linear aliphatic terminal residues, as discussed earlier. In combination with an aromatic terminal residue such as tyrosine, the L-alanine linker provided better NC-inhibitory activity compared to the related compound from the DA series. The same result was also observed for compounds **G11**, **LA11** and **DA11**, all sharing the p-amino phenylalanine terminal residue but having different linker (**Figure 5.3**).

Altogether, the results from the HTS assay strongly indicated that when a chiral linker amino acid is introduced into the AQs chemical scaffold, the orientation of the side chain length strongly affects the activity of these conjugates as inhibitors of NC. In the case of the potent inhibitors having linear aliphatic side chains (i.e., terminal residues 2), the D-alanine linker tested here showed to restore the activity loss observed in compounds bearing the L-alanine linker. Differently, in comparison with the D-alanine linker, the L-alanine stereoisomer appeared a better option in the case of compounds with an aromatic terminal residue (i.e., terminal residues 11 and 14), suggesting that other modes of action might be involved for the latter molecules.

5.1.2 Inhibition of NC nucleic acid chaperone activity by DA compound series

In order to characterize further the inhibition of NC chaperone activity by AQs, we tested compounds from the DA series as well as the new tyrosine conjugates by NAME assay, and compared the results with the previously reported series of AQs. By NAME assay, we assessed the inhibitory activity of AQs against the annealing properties of NC on the TAR and cTAR complementary sequences (see **Materials and methods**), in addition to the data obtained by HTS assay.

DA-serie	IC_{50} (µM)	DA-serie	IC_{50} (µM)
DA1	>100	**DA11**	N.D.
DA2	22.49 ± 0.99	**DA12**	N.D.
DA3	87.07 ± 14.10	**DA13**	N.D.
DA6	40.16 ± 5.95	**G14**	78.40 ± 8.20
DA7	78.77 ± 10.82	**LA14**	89.31 ± 3.45
DA8	>100	**DA14**	>100

Table 5.2. Activity ranking (IC_{50}) of DA anthraquinone conjugates assessed by NAME assay. IC_{50} were calculated from experiments showed in supplementary information, **Figure S8**. (see **Materials and methods**)

The compound **DA2** was highlighted as the best inhibitor among the DA series (**Figure 5.4**), followed by other compounds having linear aliphatic side chain (**Table 5.2** and supplementary information, **Figure S8**). Similar to what found by HTS assay, these compound were ranked as **DA2** > **DA6** > **DA7**, whereas **DA8** showed no detectable activity in the concentration range used. The compound **DA3** showed to barely inhibit NC chaperone activity by NAME (**Table 5.2**). Differently, no activity was observed under the same conditions for the related compound **LA3**.[80] In line with the results obtained by HTS assay, NAME assay further indicated that the D-alanine linker affords strong inhibitory activity in AQs with aliphatic side chains, and the best activity when placed on the correct side chain length.

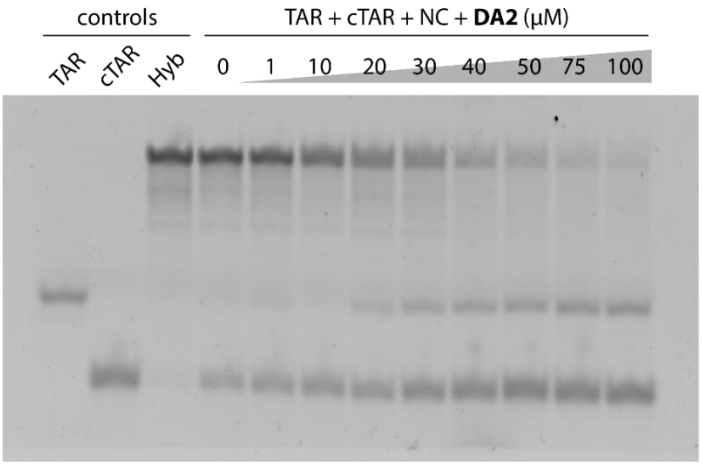


Figure 5.4. NAME assay conducted with DA2. Monomeric TAR and cTAR, as well as the heat-refolded hybrid duplex (Hyb), were included as controls. A clear dose-dependent inhibition of the NC nucleic acid chaperone properties was observed by a decrease in the formation of the band corresponding to the hybrid duplex specie with lower electrophoretic mobility compared to the TAR and cTAR monomeric species (see **Materials and methods**).

NAME assay strikingly matched also the activity trend observed for aromatic AQs by HTS assay. We were unable to observe any inhibitory activity testing **DA14** by NAME assay whereas **G14** and **LA14** yielded reproducible inhibition of NC chaperone activity under the

same experimental conditions (**Table 5.2**). Therefore, the results from both assays supported the fact that the L-alanine configuration favours the NC inhibitory activity of aromatic AQs.

5.1.3 Stabilization of nucleic acid secondary structure by DA compound series

The MOA of AQs from the DA series together with the new tyrosine derivatives was investigated by means of FQA. This assay permits to evaluate the stability of the TAR and cTAR stem-loop structures by measuring their melting temperature (T_m, see **Materials and methods**). Acting as threading intercalators, AQs interact with nucleic acid and stabilize their secondary structure, increasing the melting temperature observed. In order to estimate the stabilizing ability of AQs on the TAR and cTAR stem-loop structures, the T_m of the oligonucleotides in presence of a fixed compound concentration was compared to the control measured in absence of compound (**Table 4.5** and **Figure 5.5**).

DA-series	ΔT_m - TAR (°C)	ΔT_m - cTAR (°C)	ΔT_m - DUP (°C)
DA1	1.54 ± 0.04	4.16 ± 0.21	0.93 ± 0.37
DA2	16.36 ± 0.39	25.78 ± 0.28	6.98 ± 0.31
DA3	3.68 ± 0.21	9.89 ± 0.75	2.3 ± 0.04
DA6	12.73 ± 0.22	18.77 ± 0.75	4.81 ± 0.06
DA7	7.00 ± 0.13	16.35 ± 0.46	3.17 ± 0.07
DA8	1.12 ± 0.11	2.05 ± 0.23	0.53 ± 0.04
DA11	0.27 ± 0.06	0.53 ± 0.03	0.01 ± 0.04
DA12	0.24 ± 0.08	0.77 ± 0.15	0.35 ± 0.04
DA13	0.24 ± 0.03	1.21 ± 0.17	0.13 ± 0.03
G14	1.29 ± 0.04	2.38 ± 0.24	1.19 ± 0.28
DA14	0.21 ± 0.06	0.62 ± 0.17	0.01 ± 0.04
LA14	1.21 ± 0.04	2.26 ± 0.33	0.31 ± 0.04

Table 5.3. Activity ranking (T_m shift) of DA anthraquinone conjugates assessed by FQA. ΔT_m was calculated by the difference in the T_m observed for the oligonucleotide in either presence or absence of compound, respectively. (see **Materials and methods**)

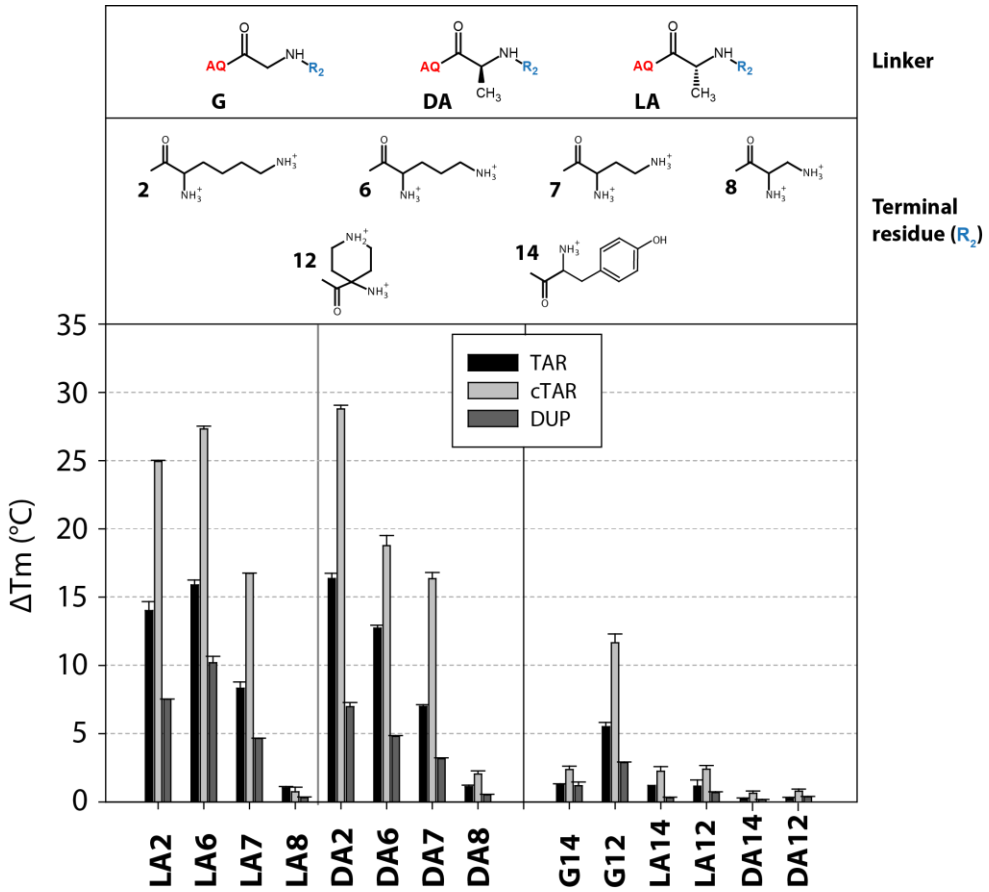


Figure 5.5. Effect of side chain orientation on the nucleic acid stabilization properties of AQs conjugates. Plot showing the stabilization effect (ΔT_m) of anthraquinone conjugates from different compound series obtained by FQA. Considering compounds with linear aliphatic side chains (compounds 2, 6, 7 and 8), also for the new DA series appeared clear that the side chain length strongly influences the nucleic acid stabilization properties of AQs. Longer compounds (compounds **DA2**, **DA6** and **DA7**) side chain length provided good stabilization whereas the shorter compound (**DA8**) appeared far less active. Differently, the side chain orientation seemed not to affect the properties of these compounds by FQA. Considering related compounds having the same side chain length, different linkers provided similar nucleic acid stabilization in all oligonucleotide substrates. These results indicated that proper side chain orientation is not strictly related to the stabilization properties of these compounds with the oligonucleotide substrate, but rather represent a feature that specifically modulates the NC inhibitory activity. Compounds with aromatic side chains (compounds 14), were highlighted as poor stabilizers of nucleic acid structure. Although compounds **G14** and **LA14** appeared to induce sensibly higher ΔT_m compared to the **DA14** compound, the little difference observed hardly explain the great difference in the NC inhibitory activity observed by NAME among these compounds (**Table 5.2**). Moreover, the stabilizing effect of **G14** and **LA14** appeared similar or even lower compared to inactive compound (e.g., **G12** and **LA12**), although the former compounds displayed much higher NC inhibitory activity (**Table 5.2**). These results suggested that aromatic AQs do not inhibit NC relying exclusively on the stabilization of nucleic acid structures. Data for compounds form the G and LA series are taken from Sosic, A. et al, *J. Med. Chem.*, 59 (**2016**) 1914−1924.[79]

For the new DA series, a clear trend was observed as lower secondary structure stabilization correlated with shorter side chain length. For all substrates, **DA2** showed the highest stabilization degree and compounds with aliphatic side chains were ranked as **DA2** > **DA6** > **DA7** > **DA8** (**Table 5.3** and **Figure 5.5**). Other AQs conjugates showed to poorly

stabilize the structure of the substrate, in line with the biological activity observed by HTS and NAME assays (**Table 5.1** and **5.2**). For each compound, the melting temperature shift observed with the TAR/cTAR duplex substrate is much lower compared to the monomeric TAR and cTAR species. This behaviour likely resides in the preferential intercalation of AQs into the dynamic hairpin structure of the two monomeric species compared to the rigid structure of the hybrid duplex, as suggested for other threading intercalating agents.[160, 161] These properties might provide specificity to AQs for viral nucleic acid substrates compared to the multitude of duplex DNA found in the cell nucleus, which might be source of unwanted effects.

Altogether, the results from the DA series obtained by FQA resemble those obtained with the previous reported LA series. We highlighted indeed the same structural requirements for optimal stabilization of nucleic acid structure (**Table 5.3**).[80] Within the same compound series, the nucleic acid stabilization effect induced by AQs well correlated with the biological activity observed against NC, which likely result from over stabilization of the nucleic acid secondary structure. However, when comparing the LA and DA series, we observed similar nucleic acid stabilizing effect for compounds characterized by remarkably different inhibitory activities on NC. Considering the best inhibitor of each series, **LA2** resulted poorly active by HTS and NAME assay compared to **DA2**, but the two compounds displayed similar results by FQA (**Table 5.3** and **Figure 5.5**).

These observations suggested that the different side chain orientation of LA and DA compounds is not critically determinant for the interaction with nucleic acids, but rather for the inhibition of NC chaperone activity. Proper linker potentially allows the side chains to interact with the phosphate backbone in a region that better compete for NC binding. Alternatively, different side chains orientation may afford different binding modes on the TAR and cTAR hairpin structures or exert a differently localized stabilizing effect on the regions of the stem that is crucial for the activity of NC (see **Chapter 6.2**). Therefore, FQA provided important information on the MOA of compounds showing that the nucleic acid stabilization properties of AQs partially correlate with the biological activity observed by HTS and NAME assays. However, from the comparison between stereoisomerically related compounds from the LA and DA series emerged that chemical features such as side chain orientation specifically affect the NC inhibitory activity of these compounds. The relevance of these structural findings will be the object of future studies in order to address such differences from a molecular mechanistic point of view.

The absence of clear correlation between the NC inhibitory activity and TAR and cTAR stabilization was even more apparent in the case of aromatic conjugates. The **G14** and **LA14** compounds showed to stabilize the TAR and cTAR secondary structure very weakly in spite of

an efficient inhibition of NC. As observed comparing the tyrosine conjugates with a compound from the same series having similar TAR and cTAR stabilizing ability by FQA (e.g. compare compounds 14 and 12, **Figure 5.5**), the aromatic conjugates showed higher NC-inhibitory activity by both HTS (**Table 5.1**) and NAME (**Table 5.2**) assays. These results suggested that interaction with nucleic acid may not be the sole mechanism of NC inhibition exert by the aromatic compounds. We hypothesized that, thanks to the tyrosine terminal residues, **G14** and **LA14** compounds may potentially interact with the hydrophobic pocket of NC as hydroxylated aromatic rings appear to be a common feature found in several hit compounds reported as NC inhibitors.[112-114] Investigation of potential NC-AQ conjugates interaction and evaluation of this proposed alternative mechanism of action will be the object of future studies.

Finally, the study done on the DA compound series add an important information to the SARs of this class of compounds collected in our laboratory. Together with the information on the side chain length required for optimal inhibition of NC chaperone activity, the work done here investigated the conformational space of the cationic side chains. In the case of linear aliphatic AQs, the DA linker showed unambiguously to provided stronger NC-inhibitory activity compared to the LA stereoisomers previously investigated. Interestingly, this trend was reverted considering compounds characterized by aromatic terminal residues. Therefore, we demonstrated here that the steric features of these compounds play a specific role in the inhibition of NC chaperone properties and these results will contribute to the design of more potent inhibitors. These SARs information set the bases for the investigation of AQs activity on novel viral RNA sequences described in **Chapter 5.2**.

5.2 Investigation of the NC-mediated primer-annealing inhibition by AQs

By testing the activity of AQs as binders of the TAR and cTAR sequences, we proved the ability of these compounds to inhibit the NC-mediated minus strand transfer event of the reverse transcription process *in vitro*. As discussed earlier (see **Chapter 1.5.4**), the HIV-1 nucleocapsid protein plays a key role during several other steps throughout the entire virus replication cycle while being able to interact with many different viral nucleic acid substrates. Similar to the TAR hairpin, other viral regulatory sequences are found in the genomic RNA leader region (**Figure 1.2**), where annealing reactions and conformational switching events regulate important processes such as reverse transcription and genome dimerization (see **Chapter 7**). We hypothesized that these viral sequences may represent novel substrates for the binding of the 2,6-dipeptidyl-anthraquinone conjugates, to potentially impair important steps of the viral replication cycle in addition to the previously studied minus strand transfer. We thus pursuit the investigation of anthraquinone activity on novel HIV-1 nucleic acid substrates.

We focused on the PBS sequence that is a viral RNA spanning nucleotide 636 to 653 within the 5'-UTR genomic RNA (**Figure 1.2**) and is involved during the reverse transcription process. PBS is fully complementary and anneals with the 3'-end of the human $tRNA^{Lys,3}$ acceptor stem. Primer-annealing is the first step of the reverse transcription process where the tRNA, acting as a primer, form an RNA duplex with PBS that is a substrate of the reverse transcriptase. Since it is necessary for the initiation of the complementary DNA synthesis, primer-annealing represents a crucial step to block the entire reverse transcription process at early stages.

The NC annealing properties catalyses the primer-annealing step because the PBS sequence folds into a stem-loop structure[162-164] that prevents complete annealing to the complementary region in the tRNA. In this context, NC has been shown to act as a nucleic acid annealer[74, 75] promoting the reaction by unspecific charge screening and molecular crowding effects.[62, 63] Differently from the case of TAR/cTAR annealing, the NC nucleic acid destabilization ability mediated by the ZFs plays a marginal role during primer-annealing. For a detailed overview of the PBS annealing to the tRNA primer, the reader is referred to the **Chapter 7.1** in the Appendix.

Two important reasons make the primer-annealing an interesting *in vitro* biologically relevant model system to test the NC inhibitory activity of anthraquinone conjugates. Firstly, the small 18-nt PBS substrate characterized by a single stem-loop structure in addition to a single-stranded region permits to evaluate binding of the anthraquinone conjugates to a smaller and more dynamic RNA secondary structure compared to the previously used TAR. Moreover, PBS

must anneal to a complementary RNA sequence, different from the RNA-DNA annealing system represented by TAR and cTAR. Secondly, the mechanism of the NC-mediated PBS annealing to the tRNA appeared remarkably different to that required for the TAR/cTAR annealing. By studying the inhibitory activity of AQs on the primer-annealing step we focus on the charge screening and aggregating ability of NC as RNA annealer, in contrast to the nucleic acid destabilization ability critical for the annealing of TAR to cTAR. Therefore, the PBS model system provides complementary information to those obtained by studying TAR and cTAR, giving a broad overview of the inhibitory activity of compounds against the same viral protein acting in different biologically relevant contexts.

To model the annealing reaction, we employed sequences spanning the complementary region of both PBS and tRNA species. As reported in **Figure 5.6**, we will refer as PBS to the 18-nt RNA sequence corresponding to the HIV-1 PBS and as Lys3 to its 18-nt complementary sequence found at the 3'-end of the human $tRNA^{Lys,3}$.

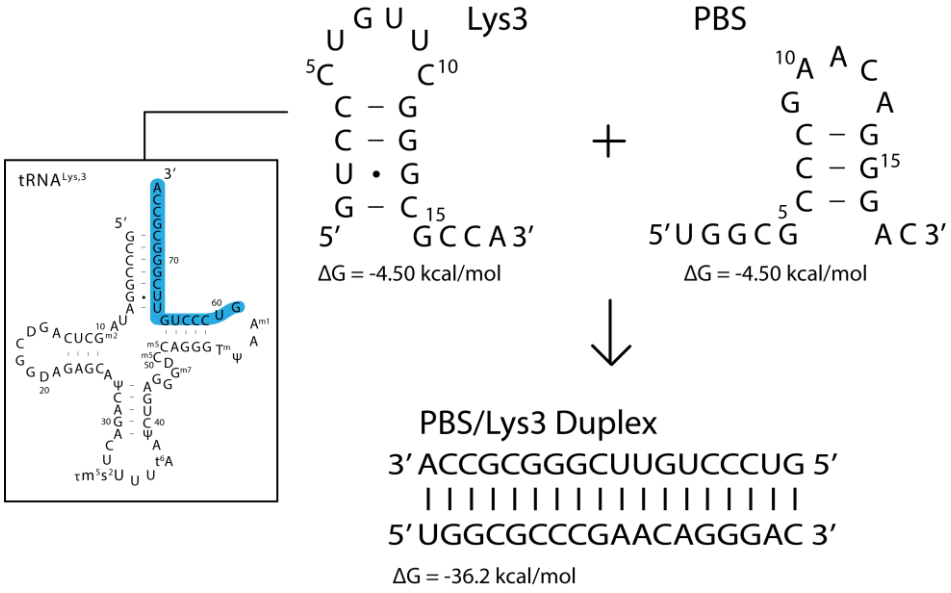


Figure 5.6. RNA oligonucleotides used to model the HIV-1 primer-annealing. PBS, primer binding site of HIV-1 (nt 636-653 in the genomic RNA); Lys3, 18-nt RNA sequence complementary to PBS located at the 3'-end of the acceptor arm in the human $tRNA^{Lys,3}$. Secondary structure and thermodynamic stability of the PBS and Lys3 hairpins were calculated by mFold,[133-135] stability of the PBS/Lys3 duplex was calculated using DINAMelt.[165, 166] Inset: $tRNA^{Lys,3}$ secondary structure highlighting the location of the region complementary to PBS (cyan).

The inhibitory activity of AQs on the NC-mediated primer-annealing reaction was investigated by means of the RAMPA assay to separate the PBS and Lys3 monomeric species from the PBS/Lys3 duplex RNA formed by the annealing activity of NC. Along with the NC protein, the assay was also optimized to assess the annealing mediated by the HIV-1 Tat

peptide, which has been recently reported to be involved in catalysing the annealing of different viral RNA strands.[167-170] The MOA of anthraquinone conjugates was then investigated by means of ESI-MS, in order to study the interaction of selected compounds with the putative RNA substrates.

5.2.1 Development of the RNA-annealer mediated primer-annealing assay

The *RNA-annealer mediated primer-annealing* (RAMPA) assay aims to separate the fraction of duplex RNA formed by the activity of viral RNA-annealers from the precursor monomeric species, in our case the PBS and Lys3 oligonucleotides, in order to assess the inhibitory activity of compounds. To better mimic the process occurring at physiological conditions, the experimental protocol was optimized to fulfil three important prerequisites: i) PBS should adopt a stable secondary structure similar to that reported in the context of the full genomic viral RNA; ii) annealing between PBS and its complementary sequence should occur at slow rate whereas iii) presence of NC or Tat proteins should strongly accelerate the reaction. We employed UV-melting spectroscopy to probe the secondary structure of PBS RNA whereas the annealing to the complementary Lys3 sequence was assessed by native gel electrophoresis.

i) PBS RNA folds into a hairpin structure.

UV-absorption spectroscopy coupled with melting experiments were conducted in order to evaluate the presence of base pairs and estimate their number, thus revealing structured regions in the PBS model sequences used here (see **Materials and methods**). By UV-VIS spectroscopy, base pairing leads to a lower absorptivity compared to that of the sum of the free bases, thus the number of base pairs melted can be calculated from the difference in the absorption spectra corresponding to the folded and unfolded states. Spectra of PBS were firstly acquired at 20°C as a control for the folded state. In buffer conditions, a sensible decrease in the absorbance was observed compared to the oligo in water solution and absence of salts, suggesting that the ionic strength and the divalent ions present in solution stabilize the PBS secondary structure (compare solid and dotted lines, **Figure 5.7A**). Rising the temperature to 95°C to promoted unfolding, the absorbance increased by 15% indicating that PBS is structured under the conditions used. The hyperchromic effect observed accounts for a 13.5% reduction upon folding of the PBS sequence (compare dotted and dashed lines, **Figure 5.7A**), which is consistent with the formation of 3- to 4-base pairs considering a 30% loss of absorbance for each pairing.[136] Reproducible spectra for the folded and unfolded states were obtained by several heating-cooling cycles, indicating the reversibility of the folding process.

Circular dichroism (CD) spectra of the same PBS sample was acquired in parallel to the absorption spectra to obtain information on the secondary structure adopted by PBS. Far-UV circular dichroism provides indeed a characteristic positive signal around 260 nm and a negative signal at 210 nm originating from the stacked base pairs in RNA helixes.[171] These characteristic features were used here to probe the stem region of the hairpin structure predicted for the PBS sequence. The two characteristic signals were readily observable in the CD spectra of PBS in aqueous solution at 20°C (solid line, **Figure 5.7B**). The intensity of the signals increased in buffer solution similarly to what observed by absorption spectroscopy, further supporting that PBS folds into a hairpin structure under the experimental conditions used (compare solid and dotted lines, **Figure 5.7B**). Both characteristic CD signals disappeared in the spectra acquired at 95°C indicating the absence of helix structures, in agreement with the melting of the PBS hairpin (dashed line, **Figure 5.7B**).

Altogether, the results obtained by UV spectroscopy indicated that PBS folds into to the desired hairpin structure predicted by mFold software[133-135] showed in **Figure 5.6**. The 18-nt PBS model used here thus folds similarly to the sequence in the context of longer RNA strands, mimicking the conditions found in the context of the full viral genome

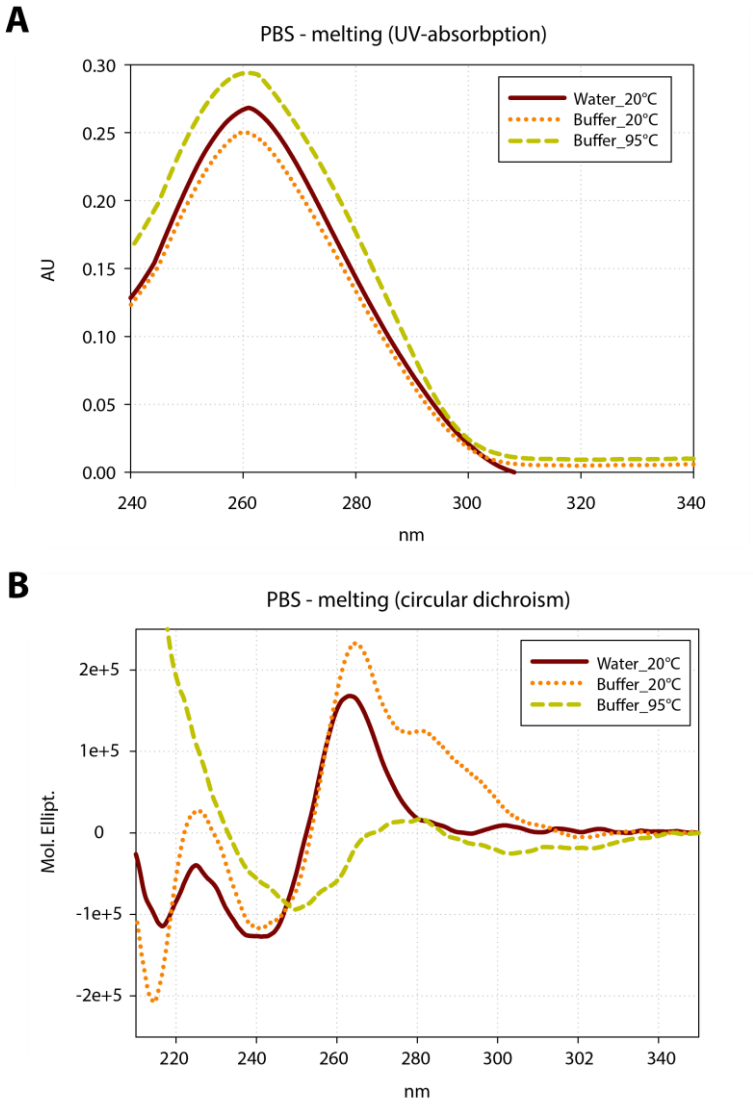


Figure 5.7. UV-VIS spectroscopy of PBS-RNA sequence. Absorption (A) and CD (B) spectra of a PBS sample diluted in water at 20°C (solid line) and buffer at 20°C (dotted line) or 95°C (dashed line) (see **Materials and methods**). PBS appeared more structured in the buffer as observed by the hypochromic effect and increase of the CD signal at 260 nm, compared to the spectra obtained in water at 20°C. Raising the temperature to 95°C led to the full denaturation of PBS structure as observed by the hyperchromic effect and almost loss of CD signal at 260 nm.

ii) PBS secondary structure prevents annealing to the complementary Lys3 RNA.

In order to assess the rate of annealing in absence of annealer proteins, we monitored the spontaneous formation of the PBS/Lys3 duplex RNA by native gel electrophoresis (see **Materials and methods**). The monomeric PBS and Lys3 sequences can thus be resolved from the PBS/Lys3 duplex RNA, which runs with a much lower electrophoretic mobility. The two complementary sequences were incubated together and the RNA mixture was loaded on the gel at different time points. Prior the incubation, PBS and Lys3 were heat-denatured to ensure the

formation of the desired hairpin structures. Different folding procedures were tested employing either 5 or 10 minutes heating at 95°C of the PBS and Lys3 solutions followed by either snap cooling on ice or slow chilling at room temperature, to select the best procedure that minimizes formation of the duplex specie. As observed in **Figure 5.8**, the bands corresponding to the monomeric PBS and Lys3 were not visible since they appeared poorly stained by the dye used to visualize RNA in the gels. On the opposite, the duplex RNA specie appeared very well stained (control lanes). Based on the amount of the RNA duplex formed over time, the folding procedure employing 5 minutes heating followed by snap cooling on ice afforded the best results showing the lowest amount of RNA duplex formed, indicating stability of the monomeric PBS and Lys3 secondary structures. The other folding protocols showed similar results or promoted the formation of slightly more duplex form. Under the selected conditions, incubation of folded PBS and Lys3 sequences yielded less than 40% of the annealing product after incubation at room temperature up to one-hour. After 10 minutes, the amount of duplex formed did not exceed 10% and only the presence of a faint band was observable on the gel (upper-left panel, **Figure 5.8**). Based on this result, the two complementary RNA appeared stable as monomeric species in the timeframe of the assay and 10 minutes was chosen as annealing reaction time to test the catalysing ability of NC on the PBS/Lys3 annealing.

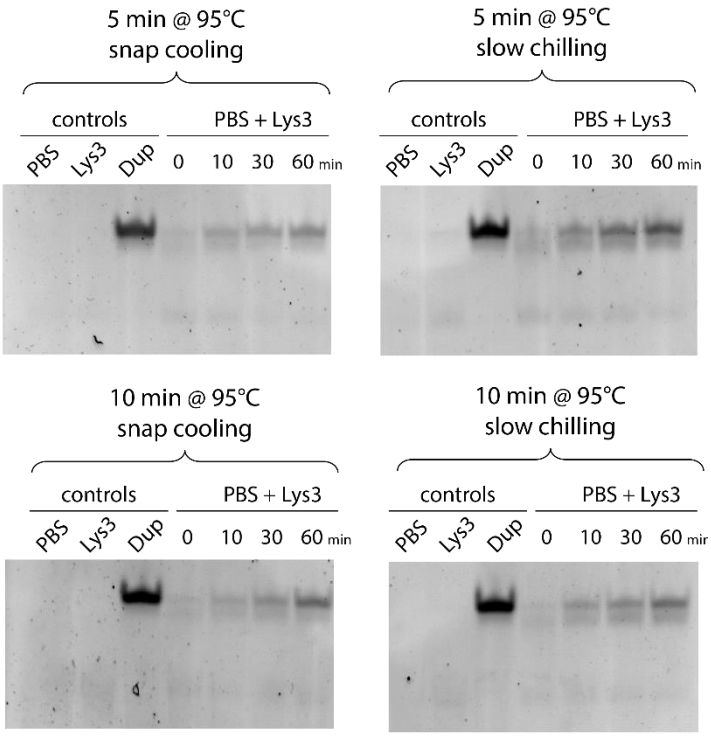


Figure 5.8. Kinetic of PBS RNA annealing to the complementary Lys3 sequence. Pictures showing gels obtained from electrophoresis separation of a mixture of folded PBS and Lys3 after the incubation time indicated on top of each lane (see **Materials and methods**). Before mixing the two oligonucleotides, PBS and Lys3 were folded employing different heat-denaturation protocols performing either 5 or 10

minutes followed by either snap cooling on ice or slow chilling at room temperature, as indicated on top of each panel. Monomeric PBS and Lys3, as well as the heat-refolded duplex (Dup), were included as controls. The experiment aimed to the quantification of the PBS/Lys3 duplex formed in absence of protein to identify the best folding procedure that stabilizes the monomeric species. Bands corresponding to PBS and Lys3 (see control lanes) are barely visible as the sequences were poorly stained by the dye used. Based on these results, the protocol employing heating for 5 minutes at 95°C followed by snap cooling yielded the best result and was employed in the following experiments aimed to test the catalytic activity of viral proteins.

iii) NC and Tat possess strong RNA-annealing activity.

Using the optimized experimental conditions that stabilize the RNA secondary structure, we finally tested the annealing properties of NC by conducting the PBS/Lys3 annealing reaction in presence of increasing protein concentration (see **Materials and methods**).

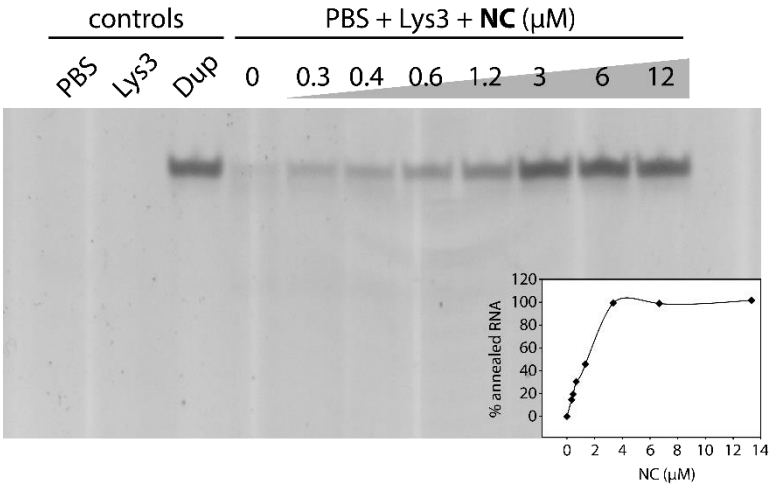


Figure 5.9. Effect of NC protein on the annealing of PBS RNA to the complementary Lys3 sequence. Pictures showing a gel obtained from electrophoretic separation of a mixture of folded PBS and Lys3 after the addition of increasing concentrations of NC and incubation for 10 minutes (see **Materials and methods**). Monomeric PBS and Lys3, as well as the heat-refolded duplex (Dup), were included as controls. Bands corresponding to the PBS and Lys3 (see control lanes) were barely visible as the sequences were poorly stained by the dye used. The annealing of PBS to Lys3 appeared scarce in the absence of protein and almost no duplex formation was observed. Differently, the addition of NC led to a dose-dependent formation of an intense band corresponding to the PBS/Lys3 duplex. Insets: plot showing the percentage of duplex product formed at different protein concentrations. Within the annealing reaction time used, the complete annealing was observed at 3 µM protein concentration.

As observed in **Figure 5.9**, the bands corresponding to the monomeric PBS and Lys3 were not visible since they appeared poorly stained by the dye used to visualize the RNA in the gels whereas the duplex RNA appeared very well stained (control lanes). In absence of NC, annealing was negligible. On the opposite, a strong band having the same mobility of the control duplex appeared when NC was added to the oligonucleotides mixture indicating that the

protein strongly catalyses the annealing of the two complementary RNAs. Within the reaction time used, complete annealing was achieved by the addition of protein at 3 μM concentration. Considering about three maximum NC binding sites on both the 18-nt single-stranded RNA such as the PBS and Lys3 substrates used here (see **Chapter 1.5.2**), the results indicated that sub-saturating concentration of NC was necessary to achieve complete annealing of the two RNAs. This observation is in line with previous studies where the same amount of NC was found to be the optimal concentration to catalyse complete annealing of longer viral RNA templates and the intact $tRNA^{Lys,3}$.[74, 75]

Primer-annealing has been proposed to proceed by nucleation of the single-stranded overhangs in PBS and the tRNA structures. The catalytic effect of NC mainly relies on the screening of the electrostatic repulsion between the two strands increasing the productive collisions that initiate formation of the duplex specie.[74, 75] Therefore, in the context of primer-annealing, NC works as an RNA annealer rather than a true chaperone protein since the nucleic acid destabilization ability mediated by the ZFs domain plays only a marginal role (see **Chapter 7.1**).[62, 63] To test the importance of the ZFs and thus compare the mechanism of the NC-mediated PBS/Lys3 annealing reaction used here to that reported for longer RNA constructs, we assessed the RNA-annealing capability of a zinc-less version of NC (apo-NC). By depleting the Zn(II) ions, zinc fingers structure is lost as well as the nucleic acid destabilization ability characteristic of RNA chaperones (see **Materials and methods**).[61] We found no difference comparing the annealing activity of apo-NC to that of the native protein under our experimental conditions (**Figure 5.10**). These results are in good agreement with studies showing that NC mutants unable to coordinate zinc catalyse the reaction even better compared to the zinc-bound form, indicating that the model system used here resemble the primer-annealing mechanism occurring *in vivo*.[74, 75]

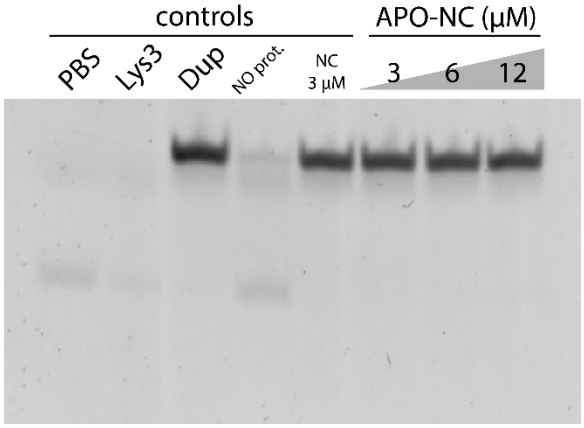


Figure 5.10. Effect of apo-NC protein on the annealing of PBS RNA to the complementary Lys3 sequence. Pictures showing a gel obtained from electrophoretic separation of a mixture of folded PBS and Lys3 after the addition of increasing concentrations of apo-NC and incubation for 10 minutes (see **Materials and methods**). Monomeric PBS and Lys3, as well as the heat-refolded duplex (Dup), were included as controls. Bands corresponding to the PBS and Lys3 (see control lanes) were barely visible as the sequences were poorly stained by the dye used. The annealing of PBS to Lys3 appeared scarce in the absence of protein and almost no duplex formation was observed (no prot.). Differently, the addition of apo-NC led to the formation of an intense band corresponding to the PBS/Lys3 duplex. A control lane employing the native protein at 3 μM concentration was introduced as a reference (NC 3 μM). Within the annealing reaction time used, the complete annealing was observed at 3 μM concentration for apo-NC indicating that the latter catalyses the annealing as efficiently as the native protein under the experimental conditions used.

The HIV-1 Tat protein has been recently shown to possess strong RNA annealing properties and that this activity mainly relays on the stretch of eight positive charges in the basic domain located at residues 47-58. Although lacking the nucleic acid destabilization ability typical of the ZFs of NC, Tat can be similarly considered an efficient RNA annealer. Tat was indeed shown to accelerate the annealing of different RNA strands and to specifically catalyse the primer-annealing step *in vitro*.[167, 169, 170, 172] Prompted by the structural and functional similarities between these two viral proteins in the latter context, we decided to include the Tat basic core (Tat48-57) in this study. However, the relevance of the RNA annealing properties of Tat to the biology of HIV-1 is still a matter of debate because transcripts coding of this protein are not present during early steps of the virus replication cycle, where most of the nucleic acid remodelling processes take place. The peculiar ability of Tat to permeate cell membranes[173] and the experimental evidence revealing its presence into the virus particle[174] nevertheless could support a potential role of this protein during early phases of viral infection.

Preliminary experiments aimed to determine the concentration of Tat48-57 necessary to achieve complete annealing to optimize conditions of the RAMPA assay for the Tat-mediated annealing reaction (see **Materials and methods**). Similarly, the electrophoretic analysis of a mixture of PBS and Lys3 sequences incubated with Tat peptide led to the appearance of a band

corresponding to the PBS/Lys3 duplex indicating that Tat48-57 strongly accelerates annealing of the two complementary sequences (**Figure 5.11A**). By performing the reaction with increasing concentration of peptide, we found that the addition of Tat48-57 at 6 μM concentration was necessary to achieve complete annealing within the same conditions and reaction time used for NC. The same experiment was also conducted employing a longer Tat peptide (Tat37-72) to assess the influence of the sequences flanking the basic stretch of lysine and arginine residues. We observed a slightly higher annealing efficiency for Tat37-72 compared to the shorter peptide, comparable to that of NC, indicating that the flanking residues play only a marginal role to the overall annealing activity (**Figure 5.11B**). Therefore, our results showed that both proteins efficiently promote the annealing reaction mimicking the primer-annealing *in vitro*. The optimized conditions found by these preliminary experiments were used to develop the RAMPA assay, which permitted to evaluate the ability of compounds to inhibit the RNA-annealing properties of NC and Tat.

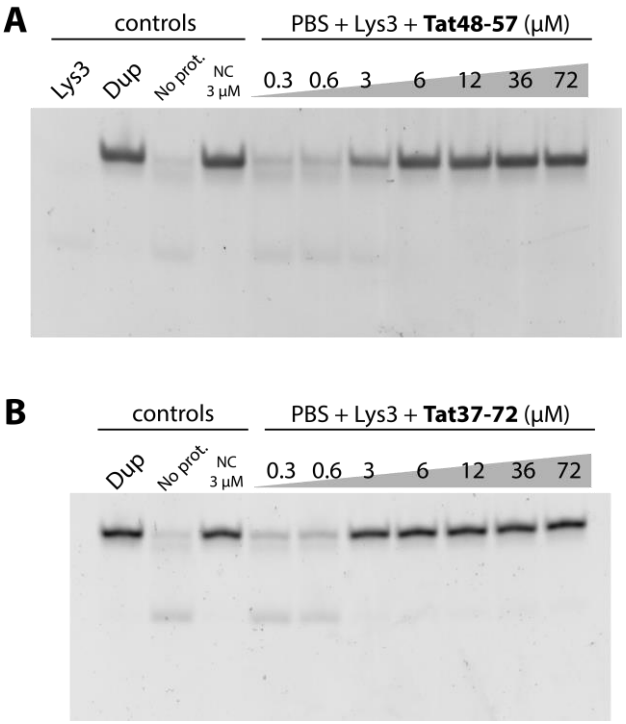


Figure 5.11. Effect of Tat48-57 and Tat37-72 on the annealing of PBS RNA to the complementary Lys3 sequence. Pictures showing a gel obtained from electrophoretic separation of a mixture of folded PBS and Lys3 after the addition of increasing concentrations of Tat48-57 (A) or Tat37-72 (B) and incubation for 10 minutes (see **Materials and methods**). Monomeric Lys3, as well as the heat-refolded duplex (Dup), were included as controls. Bands corresponding to Lys3 (see control lanes) were barely visible as the sequence was poorly stained by the dye used. The annealing of PBS to Lys3 appeared scarce in the absence of protein and almost no duplex formation was observed (no prot.). A control lane employing the native protein at 3 μM concentration was introduced as a reference (NC 3 μM). Addition of Tat peptides led to a dose-dependent formation of an intense band corresponding to the PBS/Lys3 duplex. Within the annealing reaction time used, the complete annealing was observed at 6 μM concentration of Tat48-57 and 3 μM concentration of Tat37-72.

5.2.2 Effect of 2,6-dipeptidyl-anthraquinones on the primer-annealing step

We distilled our previous works on AQs into SARs for optimal inhibition of the chaperone properties of NC.[76, 79] These pieces of information were used here as criteria for the selection of interesting molecules to be tested against the HIV-1 primer-annealing from the available series of compounds (**Figure 2.3**). We aimed to assess important chemical features such as side chain length, orientation and steric hindrance. The bunch of tested compounds was thus divided into sub-groups according to the nature of the side chains. Two groups encompassed compounds having 11- or 10-atom side chain length, including different peptidic linkers in order to assess the effect of the side chain orientation and the flexibility. Other representative compounds with shorter side chains were chosen to better investigate the effect of the side chain length. Finally, compounds with bulkier terminal residues were included to assess the effect of the steric hindrance. The compounds selected for this study are reported in **Table 5.4** along with their chemical features. The potential ability of AQs to inhibit the HIV-1 primer-annealing *in vitro* was tested by RAMPA assay.

ID	side chain	ID	side chain	ID	side chain
GSF6	11-atom side chain	**GSF7**	10-atom side chain	**DA7**	9-atom side chain
G2		**G6**		**DA8**	8-atom side chain
LA2		**LA6**		**LA13**	bulky side chain
DA2		**DA6**		**DA12**	

Table 5.4. Anthraquinone conjugates that were the object of the study on the HIV-1 primer-annealing step. Structure of the compounds are reported in **Figure 2.3**

Testing different compounds by RAMPA assay conducted with NC showed that AQs can inhibit the NC-mediated primer-annealing reaction. Different inhibitory activities were observed based on the features of the specific side chains of the compounds. **Figure 5.12A** depicts typical RAMPA assay experiment employing NC for the RNAs annealing reaction and conducted in presence of increasing concentration of the compound **DA2**. Although the band corresponding to the monomeric species PBS and Lys3 are poorly stained, the band corresponding to the control PBS/Lys3 duplex appeared clearly visible. As before, the formation of the duplex specie was remarkably slow when mixing equimolar amounts of PBS and Lys3 in absence of NC (control lanes, **Figure 5.12A**). On the opposite, the addition of the protein led to the complete duplex formation. By pre-incubating the oligonucleotides with increasing concentration of the compound prior to the NC-mediated annealing, the band corresponding to the duplex gradually disappeared revealing a clear-dose depended inhibition of the RNA-annealing properties of NC. By quantification of the RNA duplex bands (see **Materials and methods**), the activity of the different anthraquinone conjugates was ranked

according to the calculated IC_{50} observed by RAMPA assay. The compound **DA2** was highlighted as the best inhibitor among the test set having a calculated IC_{50} value of (12.5 ± 1.5) μM (**Table 5.5**). Compounds with similar but shorter side chains (compounds **DA6** and **DA7**) showed that shortening the side chains impair the inhibitory activity. Despite a proper side chain length, compound **LA2** appeared poorly active compared to the **DA2** isomer, strongly suggesting that the side chain orientation also affect the activity of AQs. Compounds characterized by bulker terminal residues (compounds **LA13** and **DA12**) displayed no detectable activity, indicating that limited steric hindrance is allowed.

cmpd	IC_{50} (μM) - NC[a]	IC_{50} (μM) – Tat48-57[b]
GSF6	20.2 ± 1.6	20.4 ± 0.4
G2	22.6 ± 1.4	18.0 ± 0.2
LA2	54.0 ± 8.4	48.5 ± 9.4
DA2	12.5 ± 1.5	15.1 ± 0.1
GFS7	30.2 ± 3.4	24.4 ± 3.1
G6	29.6 ± 2.3	34.3 ± 1.4
LA6	54.8 ± 4.8	54.3 ± 6.3
DA6	20.9 ± 0.9	16.1 ± 0.9
DA7	87.8 ± 1.5	100.0 ± N.D.
DA8	>200	>200
DA12	>200	>200
LA13	>200	>200

Table 5.5. Inhibition of the NC and Tat48-57 primer-annealing activity (IC_{50}) by AQs compounds assessed by RAMPA assay. [a]: IC_{50} values were calculated from the representative experiments showed in Supplementary information, **Figure S9**. [b]: IC_{50} values were calculated from the representative experiments showed in Supplementary information, **Figure S10**. (see **Materials and Methods**)

When AQs were tested against the Tat-mediated PBS/Lys3 annealing by RAMPA assay, the results closely matched those obtained for the inhibition of NC (**Table 5.5**). These results indicated that AQs not only inhibited the NC annealing activity but also displayed similar activity against Tat peptide. Similarly to what found by RAMPA assay conducted with NC, **DA2** displayed the best inhibitory activity against the Tat-mediated PBS/Lys3 annealing with an IC_{50} of (15.1 ± 1.0) μM (**Figure 5.12B**). In order to ensure that the region flanking the basic core of Tat do not affect the inhibitory activity observed for AQs, we run RAMPA assay control experiments employing the Tat37-72 peptide. **DA2** compound displayed the same inhibitory activity against the longer Tat peptide compared to that found by using Tat48-57, indicating that the additional residues do not participate in the annealing process and do not affect compounds inhibitory activity (supplementary information, **Figure S11**).

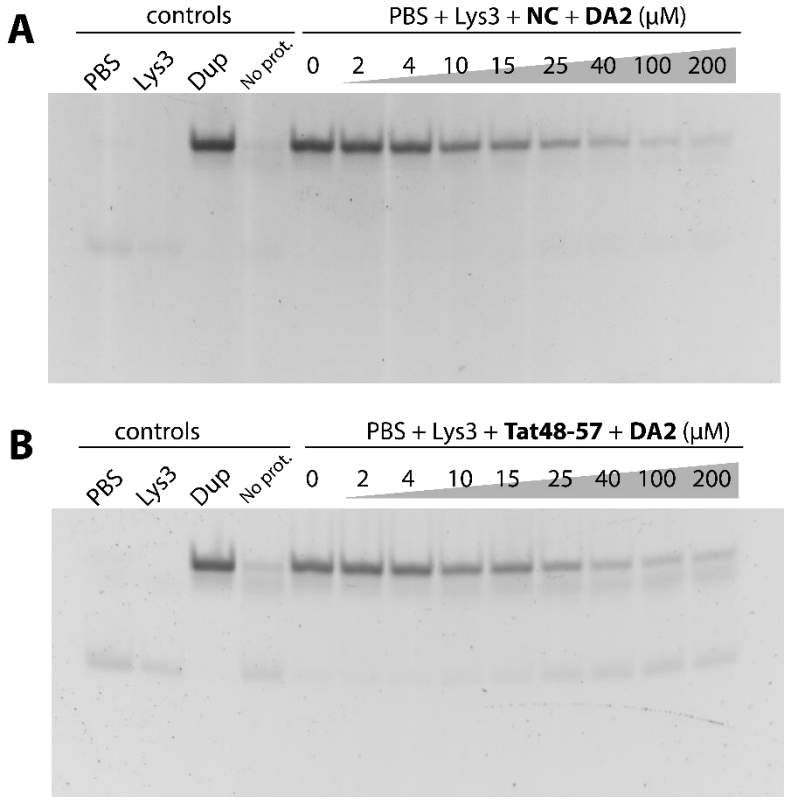


Figure 5.12. RAMPA assay conducted with DA2. RAMPA assay employing the annealing activity of NC protein (A) or Tat48-57 peptide (B). Monomeric PBS and Lys3, as well as the heat-refolded duplex (Dup), were included as controls. Bands corresponding to the PBS and Lys3 (see control lanes) are barely visible as the sequences were poorly stained by the dye used. The annealing of PBS to Lys3 appeared scarce in the absence of protein and almost no duplex formation was observed (no prot.) whereas the addition of proteins led to complete annealing. A clear dose-dependent inhibition of the RNA annealing properties of NC and Tat48-57 was observed by a decrease in the formation of the band corresponding to the duplex specie with lower electrophoretic mobility compared to the PBS and Lys3 monomeric species (see **Materials and methods**). The results indicated that the compound inhibited the annealing properties of both proteins with similar efficiency likely as result of interaction with the RNA substrates.

Inhibition of the primer-annealing by AQs appeared independent of the viral protein used. Comparable results were obtained for each compound by RAMPA assay comparing inhibition of the annealing activity of NC and Tat. The inhibition thus likely resides on favourable recognition of the RNA substrates by AQs. The results obtained from the analysis of all the compounds tested are summarized in **Figure 5.13**. Careful examination of the data obtained depicted the following SARs:

- 11-atom side chain length (compounds **GSF6**, **G2** and **DA2**) affords the best inhibitory activity on both NC- and Tat-mediated PBS/Lys3 annealing;
- The inhibitory activity decreases for compounds having 10-atom side chains length (compounds **GSF7**, **G6** and **DA6**);
- Activity of compounds results strongly impaired further shortening the cationic side chains to 9 or 8 atoms (compounds **DA7** and **DA8**);

- The D-alanine confers more activity compared to the L-alanine linker in compounds bearing the same terminal residue, and strong activity when placed on the correct side chain length;
- Bulkier terminal residues (compounds **LA13** and **DA12**) afford poor inhibitory activity compared to compounds having cationic linear aliphatic side chains (terminal residues 2, 6 and 7).

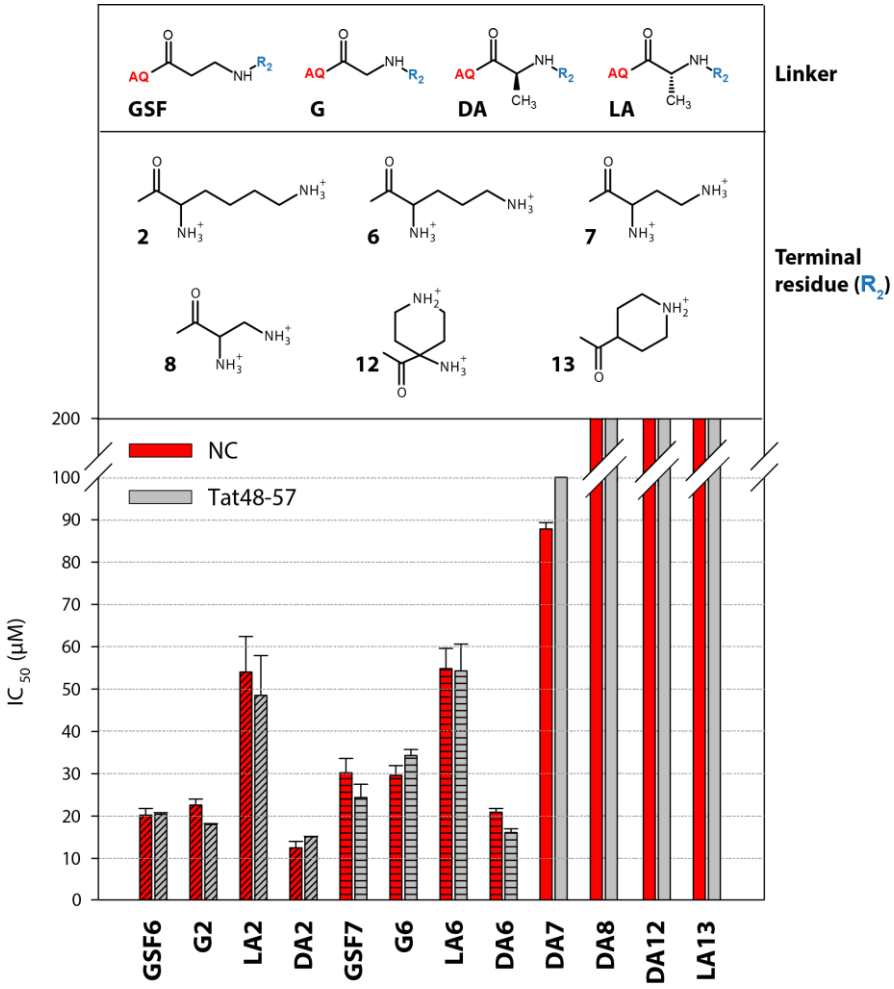


Figure 5.13. Inhibition of NC and Tat48-57 activity on the PBS RNA annealing to the complementary Lys3 sequence by AQs assessed by RAMPA assay. Plot showing the IC_{50} values obtained by RAMPA assay employing NC and Tat48-57 conducted with different compounds as reported in **Table 5.5**. Inhibitory activity against NC (red) and Tat48-57 (grey). Oblique lines, compounds with 11-atom side chain length; horizontal lines, compounds with 10-atom side chain length. Structures of the linker and terminal residues are shown on the top panels.

Together with our previous works showing inhibition of the NC-mediated TAR annealing to cTAR,[76, 79] our results showed that appropriately substituted anthraquinones are capable of interaction with multiple viral RNA substrates, with the potential ability to inhibit important viral replication processes at different stages. The compounds did not only inhibited

the activity of NC protein but also affected other viral proteins that mimic the RNA annealing properties of NC, further extending the activity spectrum of these compounds. Therefore we showed that 2,6-dipeptidyl-anthraquinones possesses pleiotropic activities on the viral replication cycle increasing the potentiality of this class of compounds as novel antiviral agents.

We observed that the SARs for optimal inhibition of the primer-annealing by AQs resemble those found in the context of the NC-mediated TAR/cTAR annealing model system, which was previously employed to characterize these compounds. The side chains length appeared again to strongly affect the activity of compounds as well as their orientation.[79] Observing the same structural requirements for optimal inhibition of PBS and TAR annealing to their complementary sequences, our biological data altogether suggest that AQs might share similar binding modes on different viral RNA substrates and that the secondary structure of the oligonucleotide substrates drives the affinity of such compounds, rather than a sequence specificity. According to the threading intercalation mechanism, stacking of anthraquinones in the PBS stem as well as bases in the penta-loop of the hairpin structures might represent preferred binding sites for this class of compounds. In order to prove our hypothesis, the binding mode of these compounds on the putative RNA substrates was investigated by mass spectrometry.

5.2.3 RNA binding mode of 2,6-dipeptidyl-anthraquinone conjugates

ESI-MS was used to investigate the interaction of AQs with the PBS and Lys3 substrates employed in the RAMPA assay. The experimental conditions were optimized for the detection of non-covalent oligonucleotide complexes (see **Materials and methods**). Control experiments conducted in absence of compounds, provided an experimental mass of 5805.86 u and 5696.82 u for PBS and Lys3 sequences, respectively, confirming the identity of the two RNAs having a calculated mass of 5805.85 u and 5696.76 u, based on the elemental composition of the two species (**Figure 5.14** and **Table S2** in supplementary information).

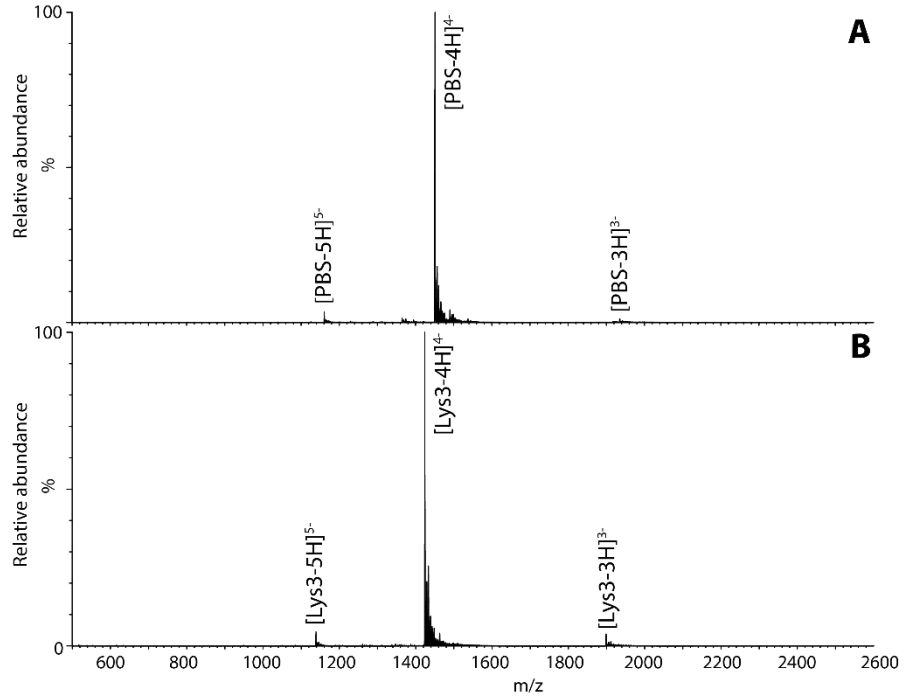


Figure 5.14. ESI-MS spectrum of PBS and Lys3 obtained under typical experimental conditions (see **Materials and Methods).** Charge states ranging from 3- to 5- where readily detected in negative ion mode for both PBS (A) and Lys3 (B). Low-intensity signals observed at slightly higher m/z than the corresponding oligonucleotide signal consist of typical sodium, potassium and ammonium adducts.

To assess the compound-nucleic acid interactions, ESI-MS direct binding experiments were conducted after pre-incubation of the RNAs with a 5-fold concentration of the compound (see **Materials and methods**). The representative spectra of **DA2** interaction with PBS and Lys3 shown in **Figure 5.15** revealed that, although with different affinity, AQs interact with both RNA substrates. The formation of a 1:1 RNA-compound complex was readily observable whereas also a second binding site was present for stronger binders (**Table S2** in supplementary information). These results indicated that the 18-nt PBS and Lys3 sequences easily accommodate the binding of two AQs molecules. The observation of a 2:1 complex in presence of the free substrate suggested that the two binding sites share similar affinities with no binding cooperativity.[79] In the latter case, saturation of the first binding site, and the disappearance of the unbound receptor specie should occur before the accumulation of a higher stoichiometry complex, differently to what observed in **Figure 5.15**, where all these species were observed at the same time.

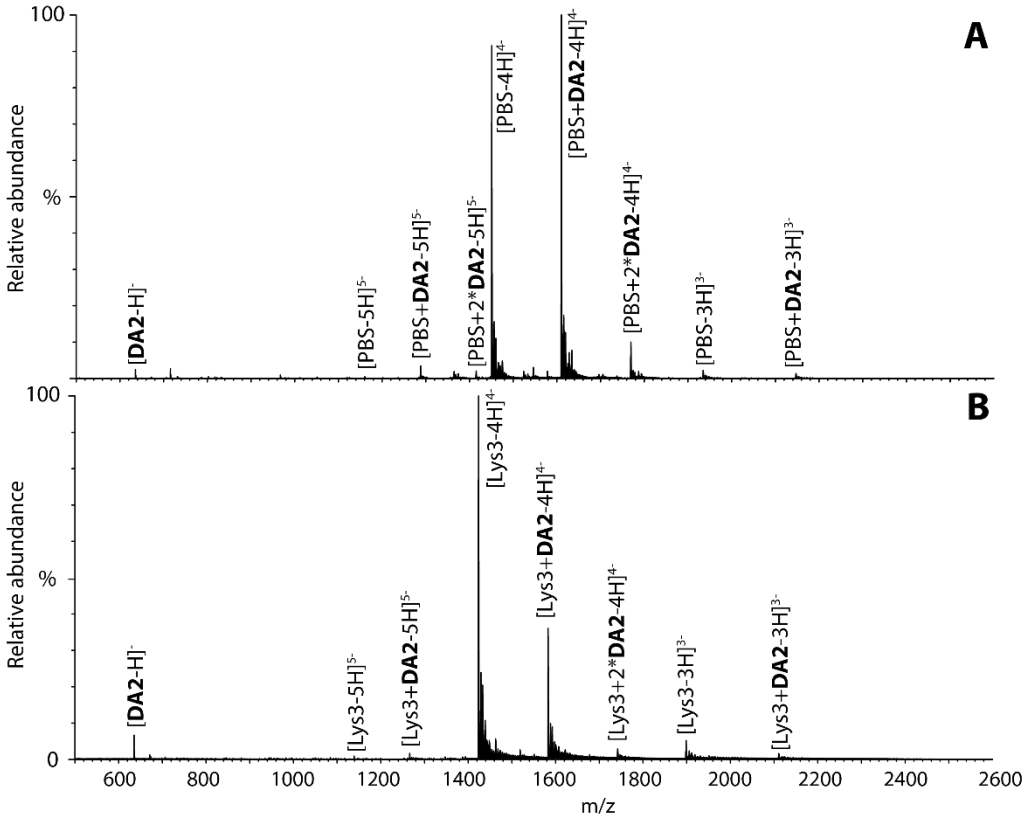


Figure 5.15. ESI-MS spectra of PBS and Lys3 in complex with DA2. ESI-MS spectra were acquired after 30 minutes incubation of compound with PBS (A) and Lys3 (B) (see **Materials and Methods).** For both RNA substrates, the formation of both 1:1 and 2:1 compound-RNA complexes was observed, confirming the RNA putative target of AQs compounds. **DA2** bound stronger the PBS viral sequence compared to the human Lys3 sequence. Low-intensity signals observed at slightly higher m/z than the corresponding oligonucleotide signal consist of typical sodium, potassium and ammonium adducts.

The ESI-MS binding analysis can reveal the accurate partitioning between any free and bound species in solution, providing a good estimate of the affinity of each ligand for the substrate (see **Materials and methods**). By quantification of the partition between free and bound RNA in presence of different AQs, we ranked compounds based on the relative affinity scale observed for the PBS and Lys3 RNAs (**Table 5.6** and **Figure 5.16**). The results revealed the following structural requirement for the optimal interaction of AQs with the RNA substrates:

- 11-atom side chains length (compounds **GSF6**, **G2** and **DA2**) provides the best affinity interaction on both PBS and Lys3 substrates;
- The binding affinity decreases for compounds having 10-atom side chains length (compounds **GSF7**, **G6** and **DA6**);
- Affinity results strongly impaired further shortening the cationic side chains to 9 and 8-atom (compounds **DA7** and **DA8**);

- The D-alanine provides better interaction with both RNA substrates compared the L-alanine linker in compounds bearing the same terminal residue, and strong interaction when placed on the correct side chain length;
- Bulkier terminal residues (compounds **LA13** and **DA12**) afford poor binding affinity compared to compounds having cationic linear aliphatic side chains (terminal residues 2, 6, 7 and 8).

cmpd	f_b - PBS	f_b - Lys3
GSF6	54.2 ± 4.6	42.2 ± 1.9
G2	53.5 ± 2.6	38.9 ± 3.2
LA2	30.8 ± 3.9	17.9 ± 1.5
DA2	60.2 ± 1.4	40.4 ± 2.6
GSF7	43.8 ± 0.8	29.0 ± 6.1
G6	44.2 ± 7.0	32.6 ± 0.6
LA6	17.6 ± 4.9	21.4 ± 3.5
DA6	52.0 ± 2.7	31.9 ± 3.1
DA7	10.5 ± 3.0	7.5 ± 0.3
DA8	7.8 ± 0.3	12.3 ± 0.3
DA12	3.2 ± 1.3	2.2 ± 0.3
LA13	5.3 ± 0.1	3.5 ± 0.6

Table 5.6. Affinity of compounds for PBS and Lys3 substrates calculated from ESI-MS binding experiments. f_b: fractional occupancy, calculated according to equation 2 (see **Materials and methods**)

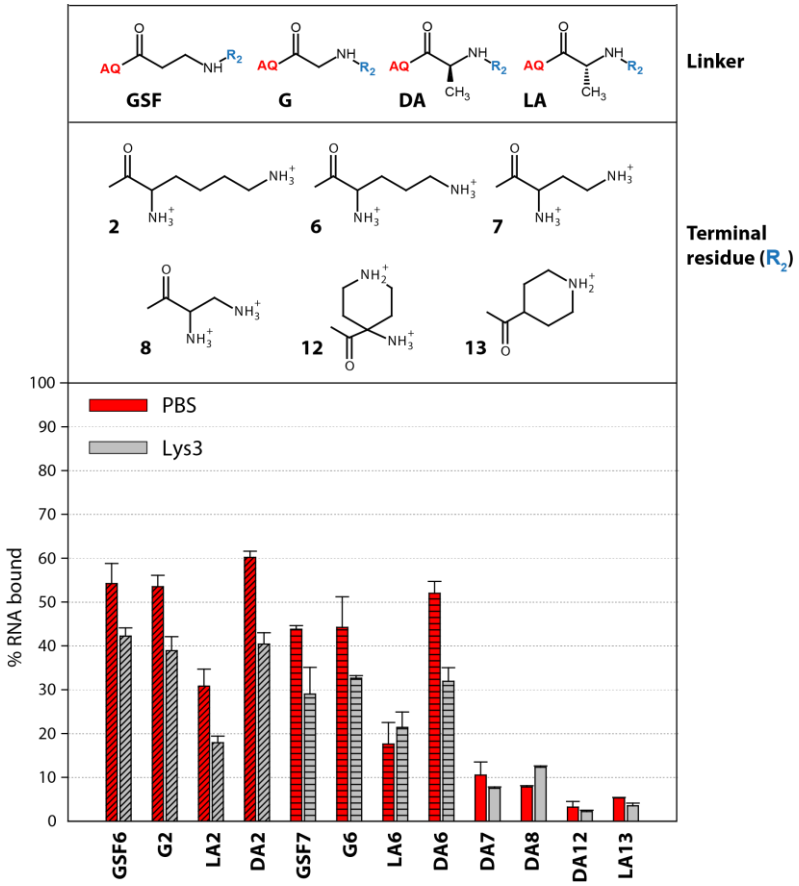


Figure 5.16. Relative affinity of anthraquinone conjugates for PBS and Lys3 RNA substrates. Plot showing the fractional occupancy values obtained by ESI-MS experiments employing the PBS and Lys3 RNAs conducted with different compounds as reported in **Table 5.6**. Affinity for PBS RNA (red) and Lys3 RNA (grey). Oblique lines, compounds with 11-atom side chain length; horizontal lines, compounds with 10-atom side chain length. Structures of the linker and terminal residues are shown on the top panels.

ESI-MS binding experiments revealed that the biological activity observed by RAMPA assay strikingly matches the binding affinity of compounds for the two RNA substrates (**Figure 5.17**). These results proved that the inhibitory activity of AQs on the annealing activity of NC and Tat proteins resides on the optimal recognition of the RNA substrates. Stronger binders likely provided higher RNA secondary structure stabilization preventing complete annealing of the two complementary strands. Binding interaction appeared finely tuned by the chemical features of the side chains, which in turn inhibited the annealing properties of NC and Tat, as observed with the TAR model sequence. The side chain length, orientation and hindrance thus have a synergic effect to maximize contacts between the anthraquinone conjugate and the RNA substrates. The optimal combination of these molecular features was found in the compound **DA2**, which was highlighted as the best binder of the series as well as the best inhibitor. For most of the compounds tested, ESI-MS experiments also showed that the relative affinity for the

viral PBS substrate is sensibly higher compared to the Lys3 sequence from the human $tRNA^{Lys,3}$, suggesting little specificity for the viral sequence by AQs.

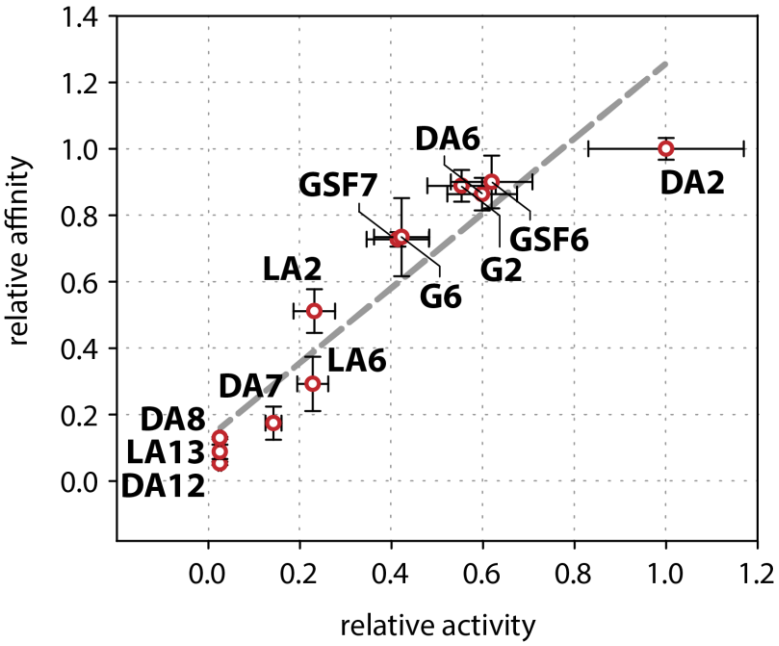


Figure 5.17. Correlation between the NC-annealing inhibitory activity and PBS binding affinity for tested AQs. The relative activity in the x-axis was calculated from the data reported in **Table 5.5** by normalizing the IC_{50} values of compounds to the IC_{50} of the best inhibitor **DA2**, obtained by the RAMPA assay conducted with NC protein. The relative affinity in the y-axis was calculated from the data reported in **Table 5.6** by normalizing the binding values of compounds to the value of the best binder **DA2**, obtained by ESI-MS binding experiments conducted with PBS-RNA. Appeared clear the correlation between the inhibitory activity and the affinity of the compounds for the RNA substrate.

To support further the influence of the side chain length on the nucleic recognition by AQs, a binding competition assay was conducted to compare directly compounds **DA2**, **DA6**, **DA7** and **DA8**. The latter compounds all share the same D-alanine linker, which provided the best inhibitory and binding activity, but bear liner gradually shorter and shorter terminal residues (**Figure 2.3**). Under this assay format, the PBS substrate was incubated with a 1-fold concentration of each compound and the different molecules thus competed for binding to the same substrate (**Figure 5.18**). All the RNA-compound complexes species can be detected within the same spectra providing an actual reading of the relative binding affinity scale by a simple comparison of the peak intensities. According to their ability to bind PBS, the signal corresponding to each RNA-compound specie was observed. Within the test set, we observed a clear trend showing lower affinity as the side chain length decreases. In line with previous results, the compounds ranking **DA2** > **DA6** >> **DA7** > **DA8** clearly indicated that 11- and 10-

atom length support strong interaction with the RNA substrate whereas shorter side chains afford much less RNA-compound affinity (inset, **Figure 5.18**).

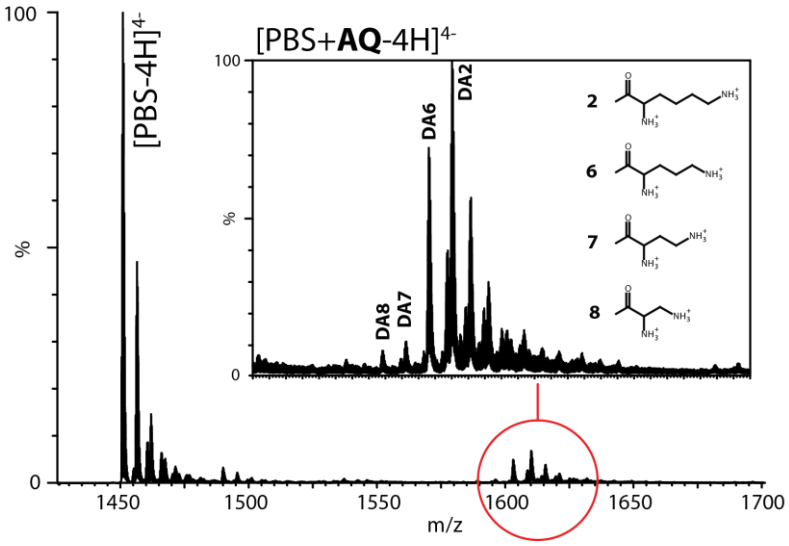


Figure 5.18. ESI-MS binding competition assay with DA2, DA6, DA7 and DA8. ESI-MS spectra were acquired after 30 minutes incubation of compounds mixture with PBS (see **Materials and Methods**). For the sake of clarity, only the region of the spectrum corresponding to the quintuply charged ions is shown. Inset: enlarged spectra focusing on the compound-oligonucleotide assemblies region. Lower intensity signals at lower m/z compared to main peaks consists in typical sodium, potassium and ammonium adducts. **DA2** and **DA6** compounds showed strong binding to the PBS substrate whereas **DA7** and **DA8** appeared to be the worst binders. These results indicated that presence of 11 to 10 atoms in the side chain represented the optimal length for interaction with the RNA substrate.

Our biological and biophysical data indicated **DA2** as compound endowed with proper side chain length, orientation and steric hindrance that afford optimal activity against the NC RNA-annealing properties. To get more insight into the binding mode of this compound, we performed tandem MS experiments aimed to assess the binding site of **DA2** on the PBS stem-loop structure. By tandem MS, isolation and activation of the precursor ions induced cleavage n the RNA phosphate backbone yielding a characteristic fragmentation footprint (**Figure 5.19A**). Typical ions from the c (also referred as d-H_2O) and y series were readily observed and analysis of the mass of the fragments enabled reconstruction of the cleavage sites.[175-178] Preliminary experiments were conducted to determine the minimal instrumental settings to achieve full coverage in the PBS sequence fragmentation (see **Materials and methods**). Under the same conditions, activation of the PBS-**DA2** complex induced no direct ligand dissociation whereas fragmentation of the oligonucleotide was still observable. However, a broad gap in the fragmentation pattern was observed for the sequence C7:C12, indicating a protective effect induced by the ligand in the 5'-side of the PBS stem and loop (**Figure 5.19B**). Interestingly, the same experiment conducted with **LA2**, which differ from **DA2** only for the orientation of the

side chains, revealed similar results. For PBS in complex with the latter compound, a protective effect was observed in the C8:A11 sequence, indicating that **LA2** bound to the same region in the PBS although afforded a slightly narrower protected site (**Figure 5.19C**).

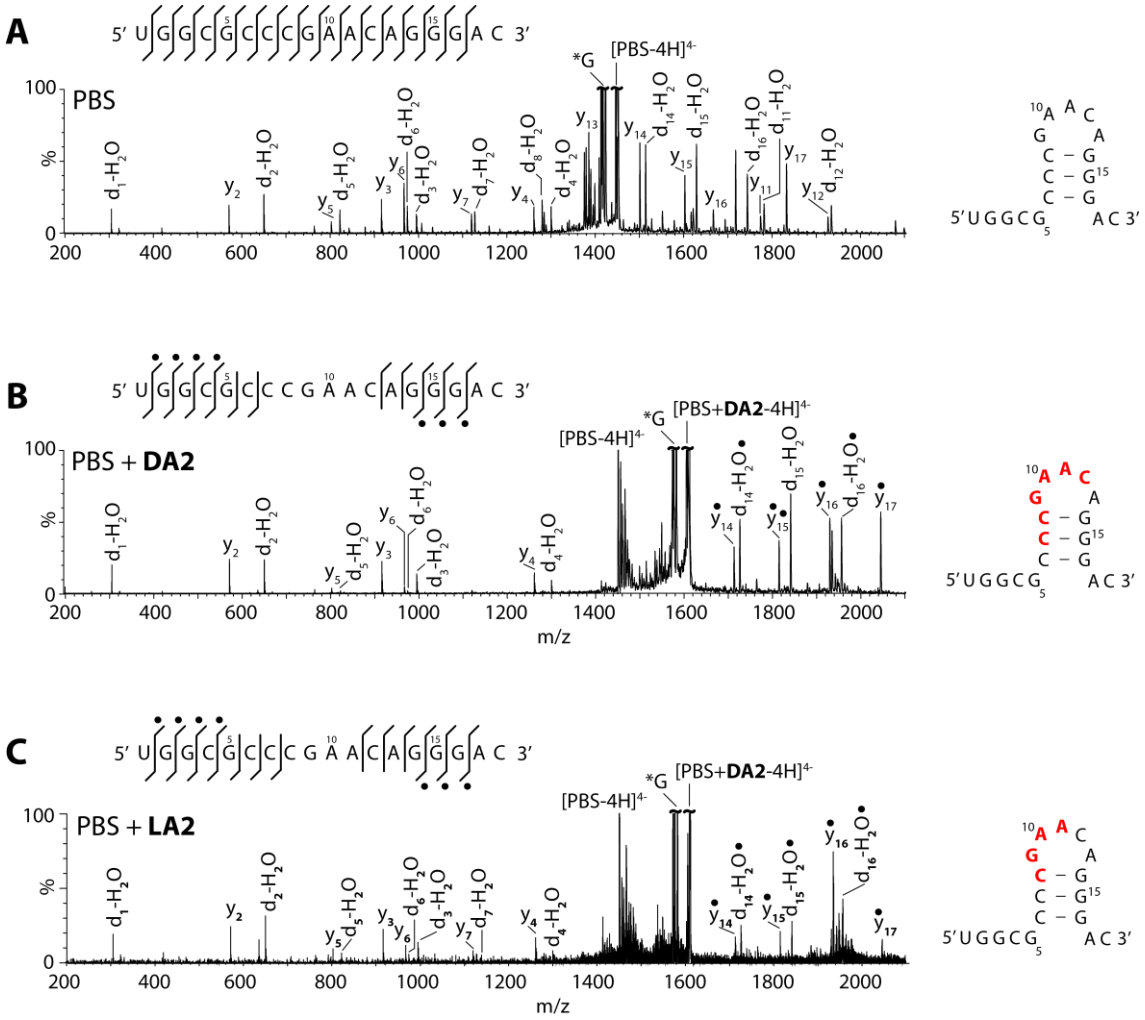


Figure 5.19. Tandem MS experiments with PBS and its complexes with DA2 and LA2. Collisional induced fragmentation of the isolated PBS (A), PBS-**DA2** (B) and PBS-**LA2** (C) complexes (see **Materials and methods**). The PBS sequence and the indicated sites of cleavage are reported on top of each spectrum, oblique rods indicate the formation of d-H_2O (bottom) and y (upper) fragments. Only the most abundant ions are labelled in the spectra. G*, loos of guanine by the precursor ions. Fragment ions containing bound ligand are marked by a dot. **DA2** binding afforded protection from fragmentation in the sequence C7:C12, which corresponds to the loop and the 5' side of the stem of PBS. **LA2** afforded similar results but the protected sequence was narrower compared to that observed in the PBS-**DA2** complex. The sequences interested by the binding of the ligand are highlighted in the red in the cartoons adjacent to each spectrum.

The results obtained by tandem MS suggested that the binding of a first anthraquinone molecule on the PBS sequence involves the penta-loop and part of the stem region in the hairpin structure of PBS. On the opposite, the single-stranded overhang of PBS appeared unaffected by

the binding of the ligand. A possible binding mode thus may occur by stacking of the anthraquinone scaffold with the apical G-C pair in the stem and the nucleobases of the loop, which can fold back to engage stacking interactions.[179] Bulge and loops have been already reported as preferential binding sites for threading intercalating agents like disubstituted AQs;[160] similarly, we recently reported that the apical loop in the TAR hairpin structure is a high-affinity binding site for the closely related 2,6-tripeptidyl-anthraquinone compound series.[78] This binding mode differs from that observed for classic DNA intercalators, which prefer well-stacked base-paired regions rich in G-C pairs. In the latter case, we would expect a clearer protection of both sides of the PBS stem and little effect on the bases into the loop. Differently, the binding mode observed here likely accounts for a non-canonical intercalation of AQs, which exploits dynamic structural motifs such as bulges and loops for efficient binding to different RNA structures.[160, 161] These RNA binding properties may contribute to the preferential recognition of viral RNA sequences, in which have been shown that loops and bulges play an important role in regulating different key processes of the virus.[18, 57, 58, 64, 112, 128, 132, 136-141, 180-182] This binding mode supports a mechanism by which AQs binding over-stabilizes the PBS secondary structure and, in the process, inhibits the NC-mediated annealing. No great difference was observed in the protective effect provided by **LA2** and **DA2** on PBS structure, suggesting that the side chains orientation, while affecting binding affinity and NC-inhibitory activity, does afford similar footprinting on the RNA substrates. These results further corroborate the hypothesis that orientation of the side chains specifically affects the NC inhibitory activity of these compounds while having little effect on their binding mode to the RNA.

The results obtained here from ESI-MS experiments permitted to highlight a strong correlation between the inhibition of the annealing properties of NC and Tat by the AQs and their affinity for the RNAs substrates. These results support the activity ranking observed by RAMPA assay and provide important information on the possible effect of these compounds on the HIV-1 primer-annealing step. In addition, ESI-MS enabled analysis of the binding mode of AQs on the PBS structure, suggesting that these compounds also interact with small structured sequences engaging non-canonical stacking interactions.

Therefore, we proved here that AQs are efficient inhibitors of the primer-annealing step in the HIV-1 reverse transcription process. Considering these data together with our previous investigation of AQs inhibition of the minus strand transfer, we highlighted the pleiotropic effect of these compounds on multiple key processes of the virus involving structured nucleic acid substrates and NC protein. In the future, AQs might find application as potential antiviral agents capable of blocking the HIV-1 reverse transcription at multiple stages, having a synergic effect on different steps within the same viral process.

6. CONCLUSIONS AND PERSPECTIVES

In the search for novel antiretroviral strategies, the HIV-1 nucleocapsid protein represents an interesting and so far challenging pharmacological target. More than two decades elapsed since inhibition of NC chaperone properties has been proposed as a potential strategy for the development of novel antiretroviral drugs[85] but no lead compound has reached the market yet. In this context, the research efforts led to the identification two main approaches to develop potential antiretroviral agents targeting this small virus protein, to which we refer here as "direct" and "indirect" strategies as illustrated in **Figure 7.1**. Direct NC inhibitors are molecule that target NC, usually by direct binding to the ZFs domain (see **Chapter 1.5.5**). Within this group fall both zinc ejecting agents and small molecule modulators of the NC-oligonucleotide interactions. On the opposite, different substrates of NC have been proposed as suitable targets for nucleic acid binders in order to inhibit indirectly the biological functions of NC. Among the indirect NC inhibitors, different nucleic acid binders have been identified to simply compete for NC binding to crucial substrates or to modulate its nucleic acid destabilization properties (see **Chapter 1.5.5**). This thesis work is devoted to a broad overview of potential NC-inhibitors antiretrovirals by exploring potentials and pitfalls of both approaches.

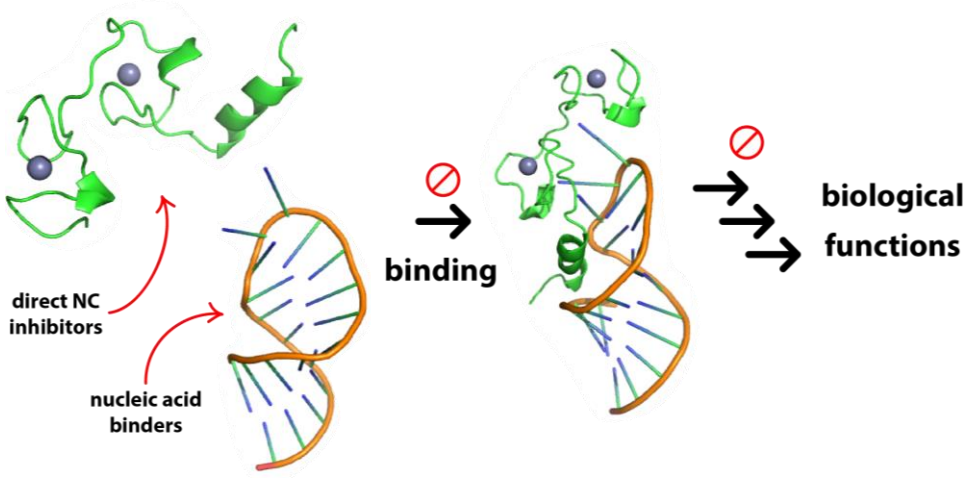


Figure 6.1. The two strategies to inhibit the HIV-1 nucleocapsid explored in this thesis work. Small molecules targeting the NC zinc fingers domain and nucleic acid binders of the substrates of NC.

We employed a panel of biological and biophysical assays to screen and rank the activity of different compounds, as well as aimed to assess their MOA. The NAME and HTS assays, already used for the identification of inhibitors of NC,[76-79] were confirmed a valuable screening platform to identify active compounds, enabling a fast and reliable evaluation of compound libraries. In addition, we developed a SPR-competition assay to assess the effect of small molecule compounds of the NC-nucleic acid interaction, which was corroborated by

EMSA results. Overall, ESI-MS technique gave a major contribution to the determination of compounds MOA providing an accurate direct observation of the species present in solution equilibria. The latter work was possible thanks to the University institutions of SUNY at Albany (NY, US) and Padova (IT), as well as Prof. Dan Fabris and Prof. Barbara Gatto, whose collaboration support international exchange of people, ideas and research possibilities.

In this thesis work, we investigated two new compound libraries designed to target the ZFs of NC (BN and BM series) as well as completed the study of novel chemical decorations in previously reported series of 2,6-dipeptidyl-anthraquinone conjugates as nucleic acid binder inhibitors of NC.

From the BN series, the compound **BN5** previously identified prompted the in-depth investigation of the 2-benzoxazole molecular scaffold in a series of analogue compounds as direct inhibitors of NC. Optimization of the chemical structure led to the identification of the **BN5-06** analogue (**Figure 6.2**), characterized by a strong improvement in the inhibitory activity observed against NC compared to the precursor molecule. Biological evaluation of BM compounds yielded the interesting compound **BM9** (**Figure 6.2**), characterized by a thiobarbituric moiety that showed to inhibit the nucleic acid destabilization and annealing properties of NC by HTS and NAME assays. These data were supported by the positive results from the SPR-competition assay and EMSA, showing that **BM9** disrupts the interaction of NC with the oligonucleotide substrates.

ESI-MS analysis of the BN hits confirmed the MOA predicted *in silico* showing their ability to bind NC protein, highlighting the non-covalent nature of the interaction. Mass spectrometry experiments also supported the compounds activity ranking obtained by NAME assay and elected **BN5-06** as the best binder among all the analogues. In the case of active BM compounds, ESI-MS analysis revealed instead that the mechanism of NC inhibition by **BM9** was strongly different to what predicted *in silico*. Despite targeting the zinc finger domain of NC as observed by molecular modeling, the compound did not only interact with NC but also promoted the ejection of Zn(II). Although the NAME, HTS, SPR-competition and EMSA assays have been used to rank the activity of compounds from the BM series, these experiments did not highlight such mechanism of action. Hence, ESI-MS experiments also enabled a more accurate interpretation of our previous results. The combination of biological assays with mass spectrometry adopted here thus appeared as a powerful approach to achieve a comprehensive evaluation of small molecule compounds inhibitors of NC.

The 2-benzoxazolinone substructure characterized in the first part of this research project focused on NC direct inhibitors represent an important finding because it unifies the structural requirements characteristic of small molecule binders of NC that have been already

reported by recent independent works.[114, 116] This allowed us to propose a structural consensus giving a pharmacophore model that contributes at establishing a definition of NC inhibitors. We strongly believe that benzoxazolinone derivatives provide a useful platform for the development of more potent inhibitors that might lead to future drugs. The second finding achieved in the first part the thesis work has been the identification of the zinc ejecting properties of thiobarbituric compounds. These compounds might have application as virucidal agents for topical application, which limit the toxicity that might occur by systemic administration.[104] Future works focused on the substituents of the thiobarbituric scaffold will aim to assess the specificity of such compounds for the viral ZFs of NC. Attention will also point to the effect of different chemical substitutions to the reactivity of the thiobarbituric moiety with cysteine thiol groups and the correlation with their NC inhibitory activity observed.

In the second part of this research project focused on indirect inhibitors of NC, we continued the investigation of the structural requirements of AQs as nucleic acid binders. We focused on the effect of the side chains orientation given by different peptidic linkers, making a direct comparison between the LA and DA series. By combining the results obtained by NAME and HTS assay, we demonstrated that the stereochemistry of the linker strongly affects the ability of these compounds to block the chaperone properties of NC, in strict correlation with the side chains length. Interestingly, this effect was also modulated by the nature of the terminal linker (i.e., aliphatic or aromatic), indicating a complex interplay between the different structural features that determine the activity of AQs as inhibitors of NC. The molecular bases of this behaviour will be the object of further investigation in future works. Based on the detailed SARs information available from our study of AQs, we then moved to the investigation of the effect of these compounds on novel viral RNA substrates of NC. Focusing on the PBS sequence, whose annealing to the $tRNA^{Lys,3}$ primer represents the first obligated step of the reverse transcription process, we demonstrated that AQs efficiently inhibit the NC-mediated annealing reaction *in vitro*. Together with our previous data, we highlighted here the pleiotropic effect of these compounds on multiple NC substrates able to potentially block the HIV-1 reverse transcription process at multiple stages. In this context, we also demonstrated that these compounds not only inhibited the NC-mediated primer-annealing, but also the activity of other viral RNA annealers acting on the same process such as the Tat protein, extending further the activity spectrum of these compounds. The ESI-MS technique played also here an important role in supporting the biological data showing the strict correlation between the affinity of the RNA-AQ complex and the NC-inhibitory activity observed. The results obtained validated the PBS/Lys3 electrophoretic annealing assay as a novel and valuable tool to investigate the activity of potential NC inhibitors. According to the different mechanism of the NC-mediated PBS/Lys3

annealing compared to that of TAR to the complementary cTAR sequence, this novel assay provides complementary results to that afforded by the NAME assay, giving a more comprehensive evaluation of the activity of potential inhibitors on the different properties of NC.

The results from our investigation of 2,6-dipeptidyl-anthraquinone conjugates converge on the compound **DA2** (**Figure 6.2**), which was highlighted here as the best inhibitor of the NC biological functions in both the TAR and PBS model systems. This compound mixed a proper side chain length together with a correct orientation to maximize the NC inhibitory activity. The linear aliphatic side chains afforded the minimal steric hindrance while providing optimal interaction with the phosphate backbone placing two positive charges in the right location. Interacting with the substrates of NC, this compound over-stabilizes the oligonucleotide secondary structure, which in turn inhibits the chaperone activity of the protein.

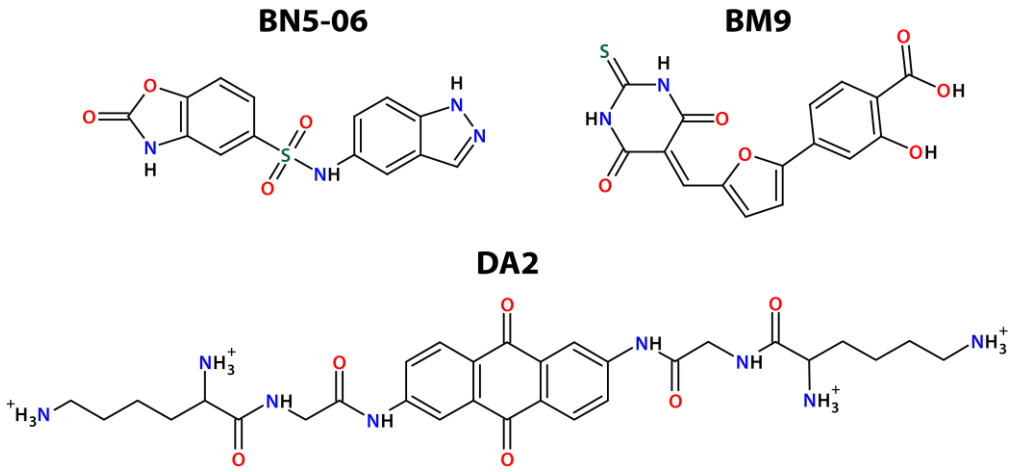


Figure 6.2. Structure of the three hit compounds identified in this thesis work. BN5-06, small molecule binder of NC; **BM9**, NC-zinc ejecting agent; **DA2**, dipeptidyl-anthraquinone conjugate binder of the substrates of NC.

Despite these encouraging results, there is still an urgent need to improve the current strategies against the HIV-1 nucleocapsid protein, in order to achieve a promising lead inhibitor compound. The direct NC binders identified here, as well as other molecules reported in the literature, still suffer from poor binding affinity affording insufficient antiviral activity to be tested in an *in vivo* context.[91, 112-116, 183] Although the development of this class of compounds should focus on novel strategies to increase compound-NC affinity, nevertheless, this venue is hurdled by the inherent flexibility of the NC structure, which hampers the design of stronger binders.[159] On the other side, nucleic acid binders of the substrates of NC may suffer from poor specificity, as the binding motifs recognized by these molecules hardly represent viral specific targets.[76, 108-110, 118, 119, 125, 126] The number of structural RNA motifs known is increasing

exponentially but the chemical space explored appears highly redundant indeed.[179] Structured RNAs important for cellular function thus might be targeted by these compounds as well, leading to unwanted off-target effects. Aminoglycosides antibiotics represent a clear example of the RNA structural redundancy. These drugs showed to bind the HIV-1 dimerization initiation site and other viral stem-loops impairing NC activity[145-148, 152] while being well known as antibacterial agents thanks to their strong binding to the bacterial ribosomal RNA.

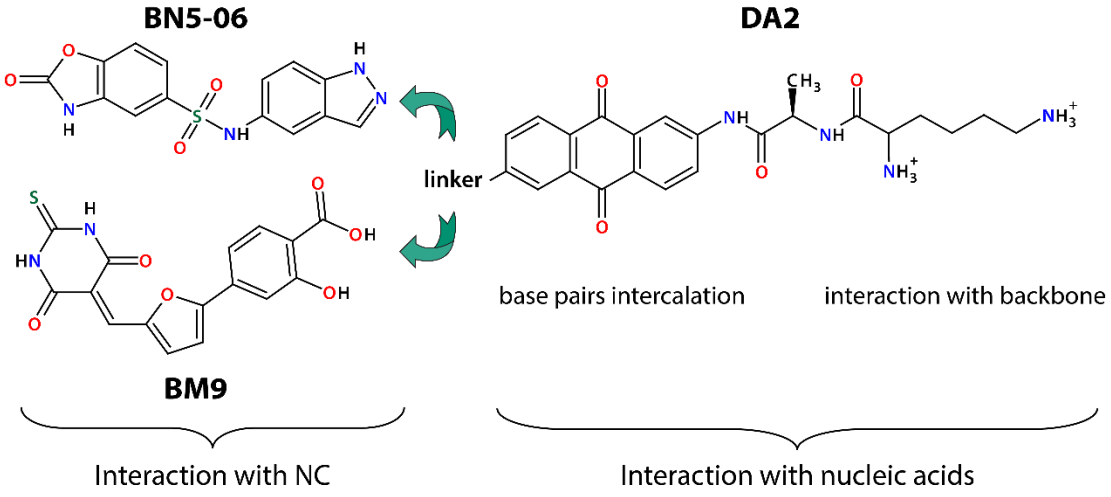


Figure 6.3. Strategy for the development of novel inhibitors of NC. Development of bifunctional inhibitors by conjugation of small molecules able to bind NC with nucleic acid binders of the substrates of NC.

We hypothesize that a possible leap forward in the field of NC inhibitors might be the development of bifunctional compounds, able to both directly interact with NC and with its nucleic acids substrates (**Figure 6.3**). By combining these two strategies within a single molecule, we try to catch the positive features from both NC and nucleic acid binders. A small molecule able to interact with the ZFs domain of NC may provide specificity toward the viral protein, as well as increase the selectivity of the nucleic acid binder moiety for the viral RNA. The latter may then increase the potency of the conjugate compound, overcoming the problem of the sub-optimal affinity of the NC binding moiety. The development of multiple ligands, able to affect different targets at the same time have been proposed as possible a venue for the development of more potent drugs indeed.[184] By taking advantage of the two strategies to inhibit NC chaperone function presented here, the anthraquinone molecular scaffold could be decorated asymmetrically, with one side chain optimized for nucleic acid backbone interaction, and the other bearing specific ligands able to engage contacts with the zinc fingers domain of NC. Different linking strategies and linker moieties, as well as different NC-binding terminal

residues, could be investigated. Special attention should be given to those chemical features such as the side chain length, orientation and hindrance, which we stressed here as important structural determinants. By developing conjugate compounds, we seek a synergism between the NC- and the nucleic acid-binding domains within the compound. We believe that this strategy may represent a valuable approach for the development of stronger NC inhibitors, to proceed toward the identification of a promising lead compound for the development of novel drugs.

7. APPENDIX

NUCLEIC ACIDS ANNEALING AND CHAPERONE PROPERTIES OF NC ADAPT TO SPECIFIC FUNCTIONS ALONG THE HIV-1 REPLICATION CYCLE

In this section, we provide an overview of specific function played by NC during crucial steps of the virus replication cycle to understand the mechanisms by which NC exerts its function on important nucleic acid substrates. In-depth investigation of these processes provided those model systems that later have been used to discover molecules able to inhibit NC biological functions. In this Appendix, we review the most relevant findings on NC catalysing the primer-annealing (**Chapter 7.1.1**) as well as the two obligatory strand transfers (**Chapters 7.1.2 and 7.2**) occurring during the reverse transcription, and the gRNA dimerization process (**Chapter 7.3**).

7.1 NC-mediated primer-annealing and the plus strand transfer

Commonly to all polymerases, RT uses a primer to start reverse transcription. As a primer, the HIV-1 uses the cellular $tRNA^{Lys,3}$ that incorporates into the viral particle before budding from the infected cell. The 3'-end of the tRNA primer anneals to a specific 18-nt sequence named primer binding site (PBS) located in the 5'-UTR region in the viral gRNA (**Figure 1.2**). The RNA-RNA duplex between the PBS and the tRNA primer represents the starting site for synthesis of the (-)ssDNA.[14] Although represents a necessary requirement for reverse transcription ignition, the spontaneous annealing of PBS to the tRNA primer is hampered by the secondary structure of PBS sequence and the process require the annealing activity of NC.[185]

Besides the primer-annealing, PBS is later involved also during following steps along the reverse transcription process. After synthesis of the (-)ssDNA and its transfer the 3'-end of the gRNA, reverse transcription must overcome another transfer event to synthetize the plus strand DNA. Using the 3'-PPT as a primer, synthesis of the (+)ssDNA proceed in the 3' direction using the minus strand DNA as template (**Figure 1.5**) and the process ends transcribing the 18-nt sequences in the tRNA primer complementary to the PBS. The (+)PBS thus forms during this step, which consists of the DNA sequence complementary to the PBS(-) located in the minus strand. At this point, the tRNA is still bound to the DNA genome forming an 18-bp DNA-RNA duplex with PBS(+). The reverse transcriptase cleaves the first ribonucleotide of the tRNA and NC promotes its release in a process known as primer removal.

Although NC assists the primer removal, has been shown that in its absence the reverse transcriptase can remove the tRNA primer by introducing additional cleavage sites in the tRNA/PBS(+) hybrid duplex.[186] After primer removal, the PBS(-) and PBS(+) are located at opposite side of the nascent genomic DNA and must anneal to each other in order to complete synthesis of the plus strand DNA. This step is known as plus strand transfer and leads to the formation of the 18-nt PBS(-)/(+) DNA duplex. Plus strand transfer requires the activity of NC because, similarly to the precursor PBS-RNA, both PBS DNA sequences form stable stem-loop structures that do not anneal spontaneously.

Primer-annealing to the PBS and the plus strand transfer are important steps along the reverse transcription process that requires the activity of NC protein to chaperone remodelling of nucleic acid secondary structures. They represent interesting and potential pharmacological targets for anti-retroviral drugs. Study of this viral sequence appears interesting and challenging since PBS is involved at different levels of the reverse transcription process both in the form of RNA and DNA, rising the complexity of a potential pharmacological approach targeting this viral sequence.

7.1.1 PBS-RNA and the primer-annealing

First studies conducted on the primer-annealing to the PBS focused on the $tRNA^{Lys,3}$. Being the most structured partner in the annealing reaction, unfolding of the tRNA was thought to be the rate-limiting step of the process while NC to catalyse the annealing reaction by promoting melting of its secondary structure (**Figure 7.1A**). However, measuring the 3'/5'inter-distance in a doubly fluorescently labelled tRNA molecule, NC showed little effect on the FRET efficiency indicating that the tRNA acceptor stem is not destabilized by NC.[187]

Later studies focused on the annealing reaction involving the two complementary RNAs. The PBS/$tRNA^{Lys,3}$ annealing reaction was described by a bimolecular process, consistent with a rate-limiting step represented by a nucleation complex involving few base pairs, which quickly proceeds to the more stable duplex form. Assuming a zippering mechanism, the most probable nucleation site was proposed to be the pairing of the 3'-end in the $tRNA^{Lys,3}$ and the 5'-end in the PBS. Such mechanism would require melting of the 5'-end of PBS, which form four base pairs with a complementary sequence located upstream in the Mal isolate genome. Mutation of the 3'-tRNA acceptor stem supported this annealing mechanism showing that length of the 3'-single-stranded region finely modulates the annealing rate. Moreover, the nucleation intermediate has been recently observed in virus, which appeared to form spontaneously in absence of NC or by the presence of *gag* before release of the nucleocapsid domain by the viral protease.[188]

Following the annealing reaction at different temperatures permitted to calculate a transition enthalpy of about 20 kcal/mol. According to the structure of PBS that encompass 3-4 base pairs (**Figure 7.1A**) and the destabilization ability of NC accounting for -0.5 kcal/mol per base pair, the latter property of NC alone was insufficient to explain the strong catalysing effect on the annealing reaction observed. The destabilizing ability of NC, that contributes to decrease the activation energy of only 1.5-2 kcal/mol, was proposed to accounts for no more than two out of the total of five orders of magnitude increment in the annealing rate observed in presence of NC compared to that observed in its absence. The data thus suggested that the PBS annealing to the tRNA primer strongly relies on other properties of NC, namely crowding and charge screening effect. The fact that NC mutants without active zinc fingers showed even better catalytic activity on the primer-annealing compared to the native protein further supported these findings. The results thus strongly indicated that the destabilizing ability of NC plays little role in the context of the primer-annealing step, differently to what observed for the TAR/cTAR annealing.[74]

Therefore, according to the available data, PBS/$tRNA^{Lys,3}$ annealing reaction appeared governed by the formation of a duplex involving few base pairs between the single-stranded 3'-acceptor arm in the tRNA primer and the 5' region of PBS. Formation of the nucleation complex thus results the rate-limiting step of the process. NC greatly catalyse the process by promoting the formation of the nucleation intermediate by means of its aggregating ability and reducing repulsions between the two oligonucleotide sequences.[74, 75] NC thus works as an RNA annealing protein without the need of the destabilizing ability characteristic of chaperones (**Figure 7.1B**). In line with this, has been shown that also viral proteins other than NC have the ability to catalyse the primer-annealing reaction, namely the trans-activator of transcription (Tat). Despite possessing zinc finger motifs, this protein has no chaperone activity but possess a basic core domain endowed with nucleic acid annealing properties being rich in basic residues. Tat basic core showed to catalyse annealing of several model RNA sequences[169] and specifically to anneal PBS to the tRNA primer.[167] However, the biological relevance of the annealing activity of Tat is still a matter of debate, as Tat is expressed from the proviral DNA only at later stages of the viral replication cycle compared to that at which the reverse transcription takes place.

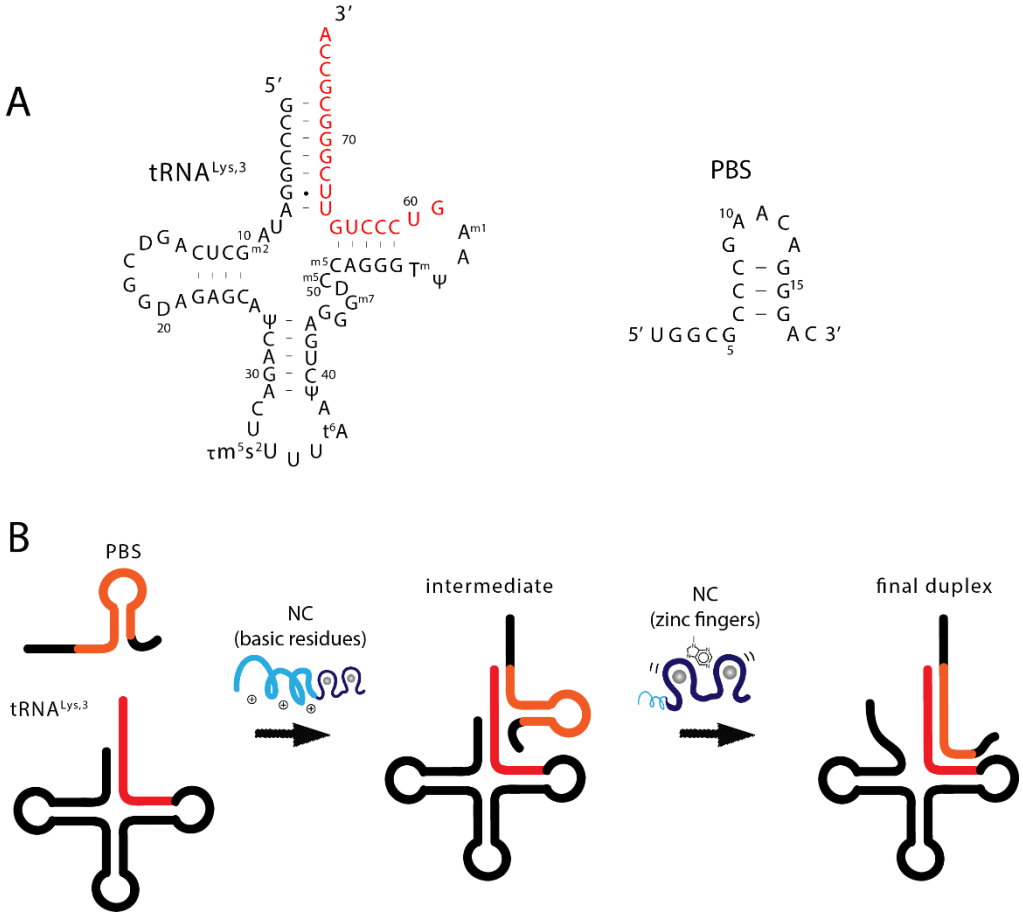


Figure 7.1. The HIV-1 primer binding site and the human tRNA^{Lys,3} primer and their annealing mechanism. The region in the tRNA complementary to the PBS in (A) is red coloured. In (B), a nucleation step involving the single-stranded 5'-end of PBS and 3'-end of tRNA likely represent the rate-limiting step of the annealing and NC catalyse the process through a crowding and charge screening effect mediated by the basic residues. Secondly, complete hybridization occurs and the nucleic acid destabilization of NC might help the umpiring of base pairs in the duplex regions.

Role of the ZFs domain during the primer-annealing has been more recently revealed by the identification of secondary interactions engaged between the tRNA^{Lys,3} and the HIV-1 gRNA. Recent studies using the entire tRNA primer and longer viral RNA acceptors showed that also sequences other than the PBS engages interaction with the tRNA primer to achieve a stable initiation complex. Specifically, the sequences upstream the PBS called A-rich loop, C-rich region and primer activation sequence, showed to make contact with complementary regions with the tRNA primer. These secondary interactions have little effect on the PBS/tRNA^{Lys,3} annealing, but modulate the activity of RT and are necessary to the formation of the initiation complex. According to earlier studies showing that NC has the ability to destabilize tRNA core structure,[75] the formation of these secondary interactions has been shown to require the specific nucleic acid destabilizing ability of NC zinc finger domain.[185]

Hence, NC promotes primer-annealing to the PBS mainly by its nucleic acid aggregating ability screening the electrostatic repulsion between phosphate groups. In this context, NC thus works as an RNA annealer protein. The specific nucleic acid destabilization ability of NC mediated by the ZFs domain only later occurs to destabilize the tRNA core structure and permits the secondary interactions needed to form a productive reverse transcription initiation complex.

7.1.2 PBS-DNA and the plus strand transfer

Investigating the plus strand transfer, model sequences mimicking the extremity of minus and plus strand DNA showed that annealing between the (-)PBS and the (+)PBS is highly inefficient. Addition of NC protein was able to catalyse the reaction *in vitro*, indicating that NC is likely involved in chaperoning the plus strand transfer *in vivo* similarly to what observed in the case of the minus strand transfer.[186]

Early NMR studies showed that the slow kinetics of plus strand transfer is related to the stable structure of the PBS DNA. The 18-nt (-)PBS was shown to fold into a stem-loop structure characterized by four base pairs, a penta-loop and a five-nucleotide overhang. A thymine nucleobase is flipped out of the stem whereas the nucleotide in the loop where found to be partially ordered, probably staking with the apical G-C base pair in the stem (**Figure 7.2**). Following the NMR signal of base pairs in the (-)PBS stem upon addition of NC, the results indicated that NC was able to destabilize the stem-loop structure of the (-)PBS by means of the ZFs domain. Even if the structure of the complex with NC was not solved yet, authors proposed that the protein was able to chaperone the plus strand transfer by lowering the energy required to melt the stem region of the two PBS DNA hairpin structures, similarly to what observed for TAR/cTAR annealing.[180]

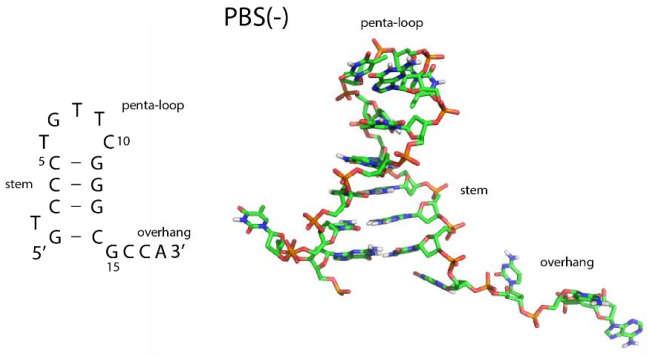


Figure 7.2. Structure of PBS(-) DNA. Similarly to the RNA counterpart, PBS(-) forms a hairpin structure characterized by a single-stranded overhang, stem and loop regions (A). NMR structure of PBS(-) (PDB entry: 1EN1) (B). Note how nucleobases in the loop are folded back to engage stacking interactions.

More recent NMR data revealed the structure of the complex between the ZFs of NC and the (-)PBS showing two major binding sites (**Figure 7.3**). Nucleotides ^{5}CTG^{7} where identified as a major binding site for NC and the T6 and G7 nucleobases appeared flipped out of the loop engaging staking interaction with the aromatic residues Phe16 and Trp37 of the ZFs. This interaction was shown to destabilize the C5-G11 base pair adjacent to the loop and to expose the ^{8}TTC^{10} bases by stretching the loop. A second minor binding site was identified in upper part of the region corresponding to the nucleotide ^{10}CGG^{12}.[128]

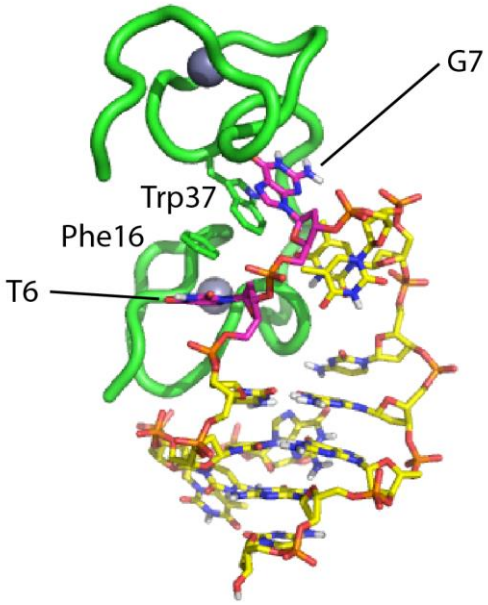


Figure 7.3. Mechanism of PBS(-) DNA destabilization by NC. The ZFs of NC binds strongly to the ^{6}TG^{7} site in the PBS(-) loop inducing a flipped-out conformation that destabilizes the hairpin structure (PDB entry: 2EXF).

Later, by mutational analysis was proved that the annealing reaction between (+)PBS and (-)PBS proceed through a mechanism involving the overhangs in absence of NC whereas a mechanism involving loops interaction prevails in absence of the protein.[189] Combining this results with the available structural data emerged that NC chaperone the plus strand transfer by activating the loop-loop interaction between (-)PBS and (+)PBS. In absence of NC, the complementary loops in the PBS DNA hairpins are not available to form a productive nucleation site because the nucleobases are in a folded conformation stacking to each other (**Figure 7.2**). Although at a slow rate, annealing reaction thus takes place starting from the PBS overhangs. In presence of NC, the stretching of the loop induced by protein binding indices a flipped out conformation in the ^{8}TTC^{10} residues of the loop, which represents good nucleation site for the annealing reaction. The kissing-loop intermediate then quickly proceeds to the DNA-DNA extended duplex conformation thanks to the destabilization of the hairpins stem.

Although NC activates the loop-loop interaction with the zinc finger domain, the mechanism involving the overhangs resulted also catalysed by the aggregating ability of NC, resembling the mechanism observed for the PBS-RNA during the primer-annealing. The removal of the NC cationic amino terminal was shown to decrease the chaperone ability by ten-fold whereas the NC SSHS mutant lacking the zinc finger domain resulted in only two-fold less active *in vitro* indeed (**Figure 7.4**).[190] As discussed in the **Chapters 1.5.2** and **1.5.3**, these results indicate that there is always an interplay between the unspecific nucleic acid aggregation properties and the specific ZFs-mediated nucleic acid destabilization ability of NC and that both contribute to the overall protein functions.

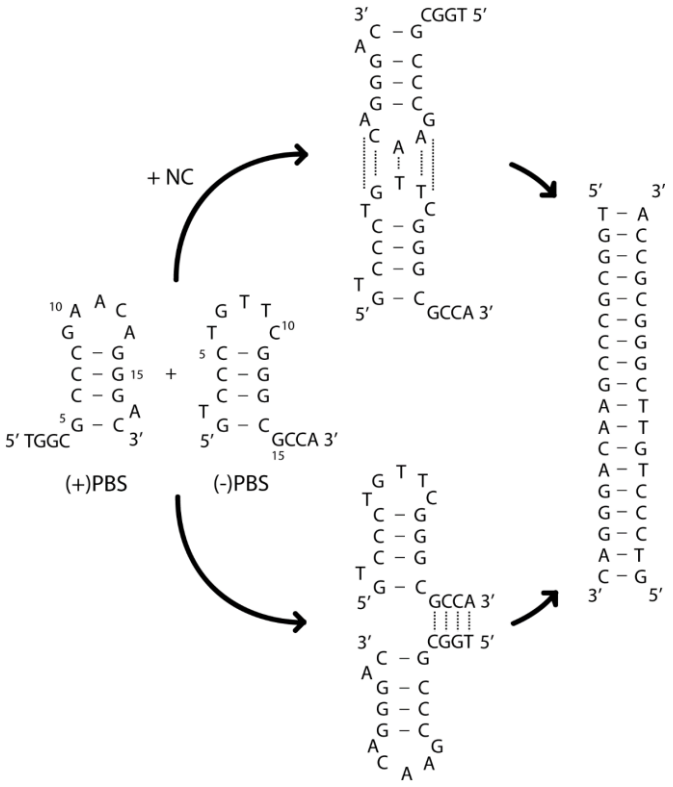


Figure 7.4. Possible mechanistic pathways proposed for the plus strand transfer. In absence of NC, annealing reaction mainly follows a zippering mechanism starting from the single-stranded region that slowly proceeds to the extended duplex form (bottom pathway). The NC-induced destabilization of PBS(-) structure leads to the exposition of the ^{8}TTC^{10} nucleobases enabling interactions with the complementary bases in the loop of PBS(+). The kissing-loop intermediate represents the nucleation step of the annealing process that quickly extends the duplex (upper pathway). Adapted from Godet, G. *et al.*, NAR, 39:15 (2011) 6633-45.[190]

7.2 NC-mediated minus strand transfer

The minus strand transfer consists in the migration of the newly synthetized (-)ssDNA from the 5' to the 3'-UTR of the gRNA (see **Chapter 1.4**). The (-)ssDNA contains the cTAR sequence, which represents the DNA counterpart of the TAR sequence located in the R region of the 5'-UTR (**Figure 1.2**). During the minus strand transfer, cTAR anneals with the 3'-TAR but the process is highly inefficient due to the stable secondary structure of both sequences.[15] In this context, NC promotes melting of cTAR secondary structure and strand exchange, enabling the minus strand transfer and complete synthesis of the (-)ssDNA by RT.[85, 191]

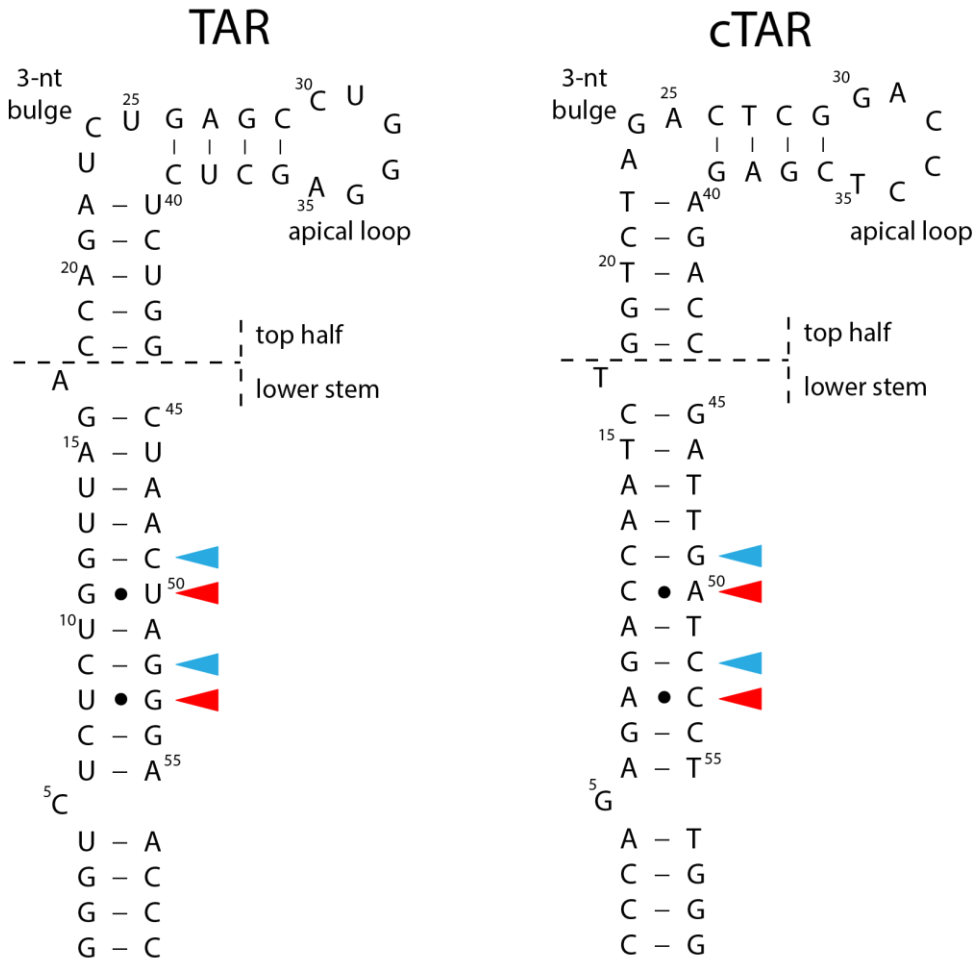


Figure 7.5. Secondary structure of TAR (RNA) and cTAR (DNA) sequences from the HIV-1 MAL strain. TAR and cTAR hairpins structures were calculated by mFold.[133-135] G-C pairs (blue) located next to base mismatching (red) appeared conserved structural requirements the enable the nucleic acid destabilization ability of NC. NC binding destabilizes the hairpin structure promoting fraying of the lower stem region and annealing through hybridization of the single-stranded termini. Propagation of the hybridization quickly proceeds to the extended hybrid TAR/cTAR duplex specie. In the case of the isolated cTAR top half, NC showed to act similarly by destabilizing the ^{18}GGT^{20} nucleobases.

First studies on the NC interaction with TAR and cTAR hairpins highlighted the ability of NC to destabilize the cTAR lower stem, shifting the conformational equilibria to an open form conformation. Most of these studies focused on the cTAR sequence while NC showed to destabilize the structure of the complementary RNA to a lower extent. Such difference was related to the greater stability of base pairing in RNA compared to DNA duplexes rather than a sequence discrimination by NC. NC was showed to destabilize a deoxyribose version of TAR to a similar extent indeed. This observation suggested that the nucleic acids destabilization ability of NC depends on the inherent stability of the substrate oligonucleotide structure.[136]

Fluorescence techniques and the FRET effect have been extensively used to probe the distance between the terminal ends in the cTAR structure studying the activity of NC. These studies revealed that NC at saturating conditions (i.e., protein to nucleotide ratios between 1:5 and 1:8)[64] does not melt the entire structure but only shift the equilibrium favouring the open conformation of the lower stem region. Appeared clear that the bulge and mismatches in the cTAR lower stem were determinant for NC functions and the lower stem region was thus proposed to be the primary target of the nucleic acid destabilization properties of NC (**Figure 7.5**).[136] [137] However, was not clear whether these sites represented preferential binding sites for NC.[138] Supporting these findings, the 1-nt bulge in the lower region of the cTAR stem was later shown to affect NC destabilizing ability the most, suggesting that fraying of the stem region starts from the bottom of the hairpin structure.[139]

Following these initial studies on the cTAR destabilization by NC, different studies investigated the role of the cTAR top half. Using shorter constructs consisting of the cTAR top half structure (**Figure 7.5**), the same stem destabilization ability was observed for NC suggesting that the protein always melts terminal base pairs regions. In this shorter version of cTAR, NC destabilization activity was strongly dependent on the 3-nt internal bulge, resembling the role of the smaller bulges in the lower stem of the full-length cTAR. Mutation or deletion of the apical loop was shown to have little effect on the NC-induced destabilization of cTAR instead. For these reasons, the 3-nt internal bulge and the apical loop were initially proposed as a strong and a weak NC binding site, respectively.[140] However, by NMR studies, the terminal nucleotides ^{18}GGT^{20} turned out to be the most affected site by binding of NC, suggesting that the protein preferentially binds to guanine bases in the lower stem.[57] Given the higher affinity of NC for single-stranded sequences, the rationale for such discrimination was unknown given the presence of several other potential binding sites in the 3-nt bulge and apical loop of cTAR. NMR showed that, although located in a double-stranded region, the terminal ^{18}GGT^{20} are the most dynamic nucleotides. On the opposite, guanine nucleobases in the bulge and loop appeared more constrained, probably due to a conformation that folds back on the

backbone structure. Therefore, the data indicated that the NC nucleic acid destabilization ability affects the least stable regions of the secondary structure, by interaction with specific guanine nucleobases,[58] similarly to what observed for the lower stem in the full-length cTAR (**Figure 7.5**).

Starting from the data available for the cTAR top-half, a model for NC destabilization of the full-length cTAR was later proposed. Sequence alignment of several cTAR sequences from different HIV-1 genotypes showed that specific guanine residues are highly conserved as well as mismatches in the cTAR lower stem (**Figure 7.5**, blue and red arrows). The model suggested that these mismatches, together with the bulges, selectively destabilize the adjacent base pair enabling interaction of NC with the conserved guanine residues. Hence, these positions served as base grips for binding of NC, driving the opening of the duplex structure. These results highlighted the fine modulation of NC activity by specific elements into the cTAR sequence, and the strict co-evolutionary relationship between NC protein and its nucleic acid substrates.[141]

Beside the nucleic acid destabilizing ability on TAR and cTAR sequences, the NC-mediated annealing reaction mechanism has been intensively studied. The NC-mediated annealing reaction was found to follow a bimolecular mechanism, indicating a two-step process. This observation was initially suggested to correspond to the formation of an unproductive TAR/cTAR complex with few inter-strand base, and a productive complex forming inter-strand base pairings beyond the second bulge of the lower stem.[192] A second model proposed instead fast formation of a nucleation intermediate followed by strand invasion and propagation of inter-strand base pairing promoted by the nucleic acid destabilization properties of NC. The nucleation intermediate was shown to involve almost exclusively the stem region of cTAR.[193] These observations were supported further by the fact that the NC-mediated TAR/cTAR annealing was strongly dependent on the bulge and mismatches located in the lower stem. This annealing mechanism thus appeared strongly correlated to the ability of NC to destabilize the cTAR structure. Promoting fraying of the stem region, NC was proposed to catalyse the formation of a productive intermediate annealed through the terminal regions. In contrast, the slow annealing observed in absence of NC was shown to involve also the interaction of the apical loops.[191] Despite the different interpretation of the kinetic data, all the studies indicated that NC strongly activates the annealing reaction by promoting TAR/cTAR hybridization starting from the terminal stem. This mechanism relies on the nucleic acid destabilization ability of NC mediated by the interaction of specific guanine with the hydrophobic pocket located at the top of the two zinc finger.[61] For this reason, in the context of the minus strand transfer, NC acts as an RNA chaperone protein.

Comparing the minus and plus strand transfer, both processes appear to rely on the interaction of specific guanine residues with the ZFs domain of NC, which induce local destabilization of the secondary structure. Based on this observation, NC can be considered to work as a real chaperone protein in both contexts. However, the high flexibility of NC permits adaptation of the NC activity to the different nucleic acid substrates. NC interacts with the guanine nucleobases located in the cTAR lower stem whereas it binds to the guanine of the loop of PBS(-) structure. Due to the interaction of the protein with guanine residues in different structural contexts, the NC-mediated minus and plus strand transfer have been shown to proceed through opposite mechanisms. Zippering from the terminal region is observed in the case of TAR/cTAR annealing whereas kissing-loop mechanism in the case of PBS(-)/(+) annealing. Altogether, these studies highlight the inherent flexibility of NC structure as an important feature that permits to this protein to play many different functions throughout the entire virus replication cycle by adaptation to specific viral nucleic acid substrates.

7.3 NC-mediated genome dimerization

Besides the reverse transcription process, other key steps of the viral replication cycle are regulated by structured oligonucleotide sequences. Examples are the viral genome dimerization and encapsidation processes, which rely on a highly structured region spanning about hundred nucleotides at the 5' terminal in the gRNA. The viral packaging sequence, also referred as Ψ-RNA encompasses four conserved stem-loops (SL1-4) (**Figure 1.2**), similarly to the TAR and PBS sequences. SL1, also referred as dimerization initiation site (DIS) represents the starting point for gRNA dimerization. Together with SL3, SL1 also represents a strong binding site for *gag* polyprotein that serves for genome recruitment during viral budding. SL2 contain the major splice donor (SD), which is the predominant 5' site for viral mRNA slicing. SL4 contains the starting codon for the *gag* gene and represents the 3' boundary of the 5'-UTR.

HIV-1 genome dimerization was initially shown to involve a discrete sequence at the 5'-end of the gRNA,[194] in a region located somewhere downstream the primer binding site and upstream the major splice donor. RNA probing and deletion analysis highlighted a six-nucleotide palindromic sequence within this region that served as a primary requirement for gRNA dimerization. The self-complementary enabled interaction of two gRNA molecules and this site was thus proposed to be the nucleation site of the dimerization process.[17] Later, the auto-complementary sequence was found to be part of a stem-loop structure, characterized by stem regions interrupted by a bulge and a 9-nt apical loop containing the hexa-palindromic sequence.[195] These initial studies highlighted two major DIS sequences, namely the GCGCGC and GUGCAC for the LAI and MAL isolates respectively (**Figure 7.6**).[195,17]

Basing on these early works, a kissing-loop model was proposed where the stable stem-loop structure adopted by the DIS exposes the auto-complementary sequence, enabling inter-strand loop-loop recognition promoting gRNA dimerization. In such model, the dimeric genome was thought to extend the inter-strand hybridization to the regions flanking the auto-complementary loops during the maturation process. The identification of gRNA dimers with different thermal stability later supported the kissing-loop model. Using *in vitro* dimerization assays with synthetic RNA mimicking the 5'-UTR, several studies showed that different dimeric species where obtained indeed upon incubating the RNA at 37°C or 55°C. The dimers formed at lower temperature was characterized by lower thermal stability and were able to convert into more stable dimer upon heating. Differently, the dimer formed at higher temperature was characterized by higher thermal stability and showed a single conformation. These two dimeric forms where suggested to correspond to a "loose dimer" connected by the loop-loop interaction (kissing loops, KL) whereas the "thigh dimer" was proposed to correspond

to the extended inter-strand duplex (extended duplex, ED) involving also the regions flanking the auto-complementary sequence (**Figure 7.6**).[196, 197]

In this context, NC was shown to promote stabilization of the gRNA dimer converting loose dimers into tight dimers.[197, 198] This activity was also demonstrated for the *gag* polyprotein, supporting the active role played by this polyprotein during the gRNA dimer maturation.[199] Therefore, NC was proposed to chaperone conversion from the kissing-loop to the extended duplex conformation, promoting stabilization of the gRNA dimer.[200]

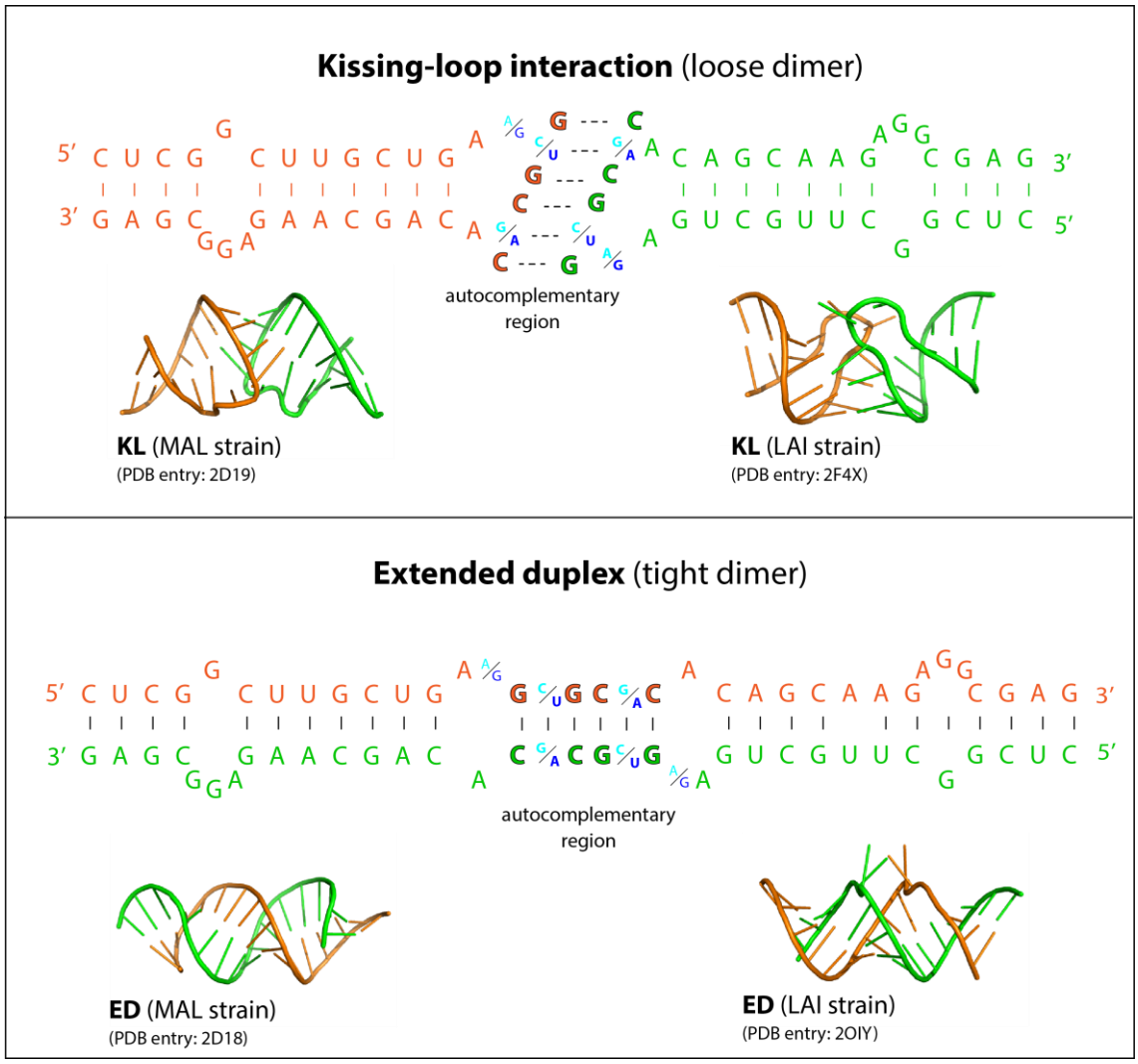


Figure 7.6. Structure of the HIV-1 dimerization initiation site dimer in the kissing-loop and extended duplex conformation. Auto-complementary sequence (bold letters). Base sequence from the LAI (cyan) and MAL (blue) HIV-1 strains, which represent the most prevalent variants. Kissing-loop interaction (KL, upper); extended duplex conformation (ED, bottom). Crystallographic structures of the auto-complementary and hinge regions for the KL and ED complexes from the two DIS variants.

Structural studies later confirmed the predicted model by solving the structure of the two distinct dimer conformations of DIS. Both crystallographic and NMR identified the KL

dimer characterized by six-base pairings between the two auto-complementary sequences, confirming the presence of the loop-loop interaction. The stem regions formed a conventional A-helix and stack together with the kissing loops in either a coaxial or tilted conformations.[201, 202] The adenine residues flanking the palindromic sequence at the 3' side showed to be highly dynamic experiencing both flipped-in and flipped-out conformations. The purines at the 5' side also were observed either stacked together outside the KL structure or independently adopting alternative conformations (**Figure 7.6**).[201-203] Differently, the residues within the 3-nt bulge were found staked into the stem region distorting the axes of the helixes.[204]

Structure of the ED dimer resulted similar to that of the KL complex. The palindromic sequences were shown to retain the similar conformations within the two different dimeric forms. In the ED observed for the LAI strain, the three purines flanking the auto-complementary sequence were found to form an unusual A-G base pair whereas a single adenine was observed protruding from the duplex structure.[205] The flipped out purines were proposed to represent "base grips" for NC binding and play a role in the genome dimerization and packaging processes (**Figure 7.6**).

The similarity between the heat treatment and the activity of NC in the maturation of DIS dimer prompted a comparison between the thermal- and NC-induced KL-to-ED conformational switching. Upon heating, the preformed KL complex was shown to undergo a concerted melting of the stem region in both hairpin structures and an inter-strand reannealing leading to the more stable ED dimer. Further temperature increment led to the dissociation of the ED dimer. The two transition temperatures presented enough difference to yield two well-resolved peaks by UV-melting analysis. Similar experiments conducted by NMR spectroscopy also showed that the six-nucleotide base pairing between the two auto-complementary sequences does not dissociate during the transition from the KL to the ED conformation.[182, 206] According to the dynamic conformation of the purine residues flanking the palindromic sequence, the first adenine at the 5'-side of the loop was shown to switch between a flipped in and flipped out conformation basing on its protonation state. This conformational switching was shown to destabilize the hinge region of the KL complex and thus suggested to participate in promoting conversion to the ED dimer during either thermal- or NC-induced conformational change.[207] Other pathways have been proposed arguing against the concerted melting mechanism, which evoked the formation of cruciform tetrameric intermediates or inter-strand trans-acylation for the conversion of the KL complex to the ED dimer conformation. [201, 208]

Following the NC-induced DIS conformational switching, NC protein at concentration higher than two equivalent per DIS dimer showed to efficiently chaperone the KL-to-ED transition. No effect of NC was observed on the loop-loop affinity suggesting that NC does not

promote the formation of the KL complex. Similarly to what observed in the thermal melting of the KL complex, the loop-loop interaction was not disrupted by the addition of NC, suggesting a that the NC- and thermal-mediated DIS conformational switching follow similar mechanistic pathways.[182, 209] Therefore, both in presence and absence of NC protein, remodelling of the KL complex was proposed to proceed through a strand exchange mechanism involving concerted melting of the stem region of both hairpin structures and an inter-strand reannealing guided by the complementarity of the two palindromic sequences.

The mechanism by which NC promotes the strand exchange in the KL complex and conversion to the ED conformation is still not clear. However, appear clear that the KL-to-ED transition requires the chaperone activity of NC mediated by both zinc fingers domain and basic residues.[210] NC was shown to selectively bind the 3-nt bulge in the stem of DIS and additional binding sites were observed in the apical loop of the monomeric DIS.[132] In the DIS from the LAI strain, the purine residue flanking the auto-complementary region showed to finely modulate the NC chaperone activity. Adenine to guanine mutation in this region was shown to afford binding of an additional NC molecule but this prevented transition to the ED conformation.[18] On the opposite, binding of NC to the hinge region in the KL complex from the MAL strain, which has a single guanine residue flanking the 5'-side of the loop (**Figure 7.6**), was proposed to be the driving force of the NC-mediated KL complex refolding to ED dimer.[210] Hence, although different mechanism has been proposed, these studies agree indicated that one equivalent of NC likely destabilizes the stem region of DIS by interaction with the 3-nt bulge whereas a second equivalent likely affects dynamic of the KL hinge region.

Although a single gRNA molecule contains all the genetic information necessary for the virus, all retroviruses pack two genomic strands. The DIS is highly conserved and mutations within this region shown to strongly interfere with the virus replication. Impairment of the dimerization process showed indeed to almost abolish infectivity due to the inability to pack the genetic material into the virus particles before budding.[211, 212] Hence, these results highlighted a link between genome dimerization and encapsidation processes, and virus infectivity. In line with these results, model HIV-1 viruses mutated within the genome 5'-leader region to over-stabilize the DIS hairpin structure of the DIS, were unable to convert the KL into ED genome dimers and unable replicate in cell cultures.[213] Interestingly, compensatory mutations that restored infectivity were observed after several infection cycles. These mutations were shown to restore the KL-to-ED transition, thus unambiguously linking the DIS conformational switching to vital functions for virus replication.[214] Based on the biological relevance and evolutionary conservation of the sequence, DIS has been proposed as a promising pharmacological target for potential antiretroviral nucleic acid binders.[215]

In a first attempt to target DIS, antisense strategy was employed trying to screen the auto-complementary sequence in the DIS apical loop thus preventing self-dimerization. This strategy showed to inhibit dimerization of gRNA both *in vitro*[216] and in a cellular system,[217] and highlighted DIS as a good target for antisense oligonucleotides within the HIV leader sequence.[218] Another approach targeting DIS used aminoglycosides antibiotics as specific binders of the DIS-DIS complex. The strict structural similarity between the bacterial 16S ribosomal RNA site A and the KL complex early suggested the retargeting of these molecules to the viral DIS sequence, which showed to effectively bind and stabilize the KL complex in infected cells and virions.[120] Further studies permitted to better characterize the interaction of aminoglycosides analogues with the DIS,[124] collecting structural-activity relationship important for further development of small molecules antiviral agents targeting the genome dimerization step.[121-123]

8. SUPPLEMENTARY MATERIAL

In this chapter, we report extensively those experimental results found in specific sections within "**Results and Discussion**" (**Chapters 4** and **5**) and named **Figure S#x** or **Table#y**, that were omitted to achieve a more fluid discussion and a clearer comprehension of our work.

Control experiments, as well as experiments conducted with the sole purpose to extend considerations originating from the main data, were also reported as supplementary material in order to keep the Results and Discussions sections as short as possible.

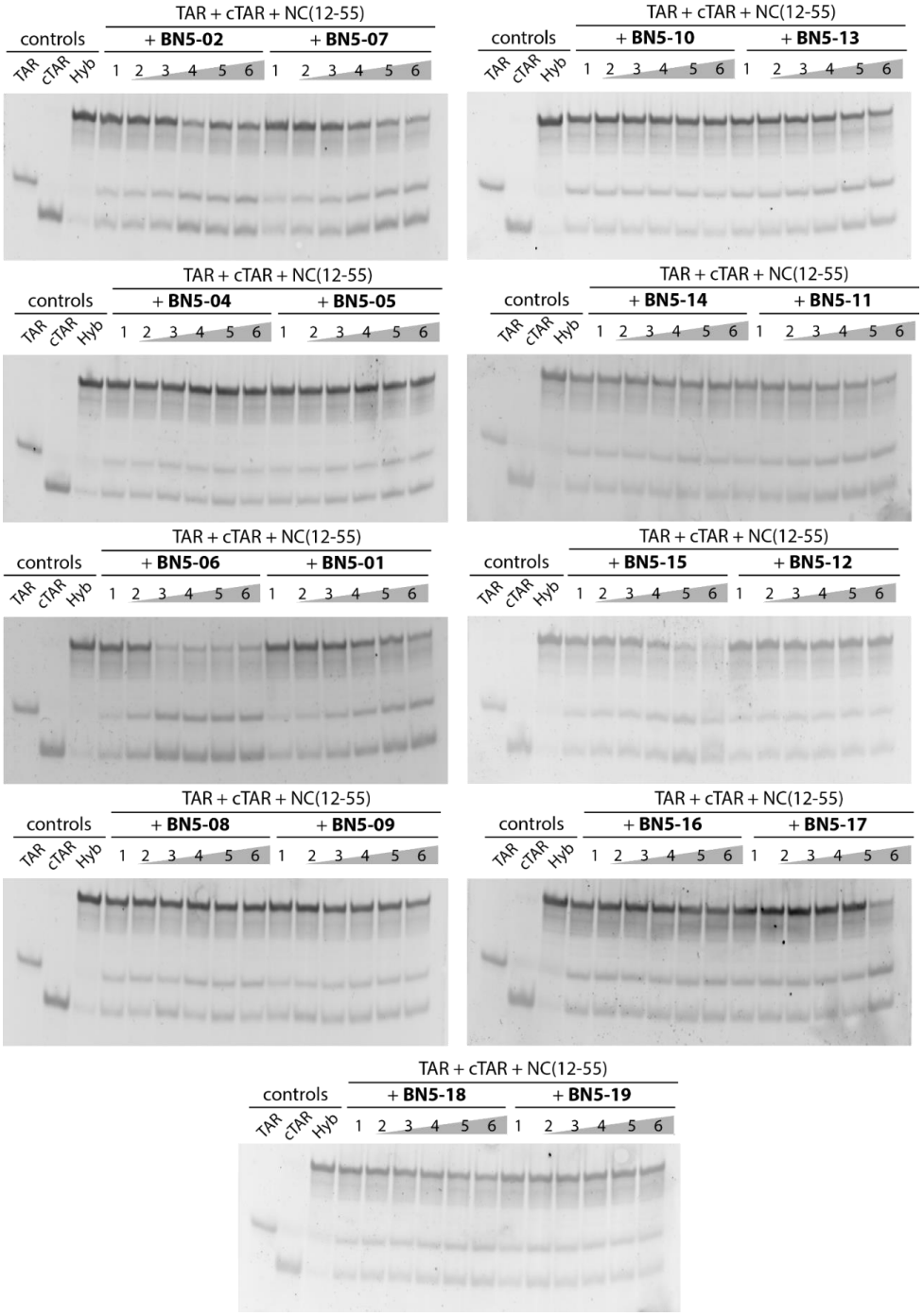


Figure S1. NAME assay conducted with BN compounds. Monomeric TAR and cTAR, as well as the heat-refolded hybrid duplex (Hyb), were included as controls. A clear inhibition of the NC nucleic acid chaperone properties was observed with **BN5-06**. The formation of the band corresponding to the hybrid duplex specie with lower electrophoretic mobility compared to the TAR and cTAR monomeric species appeared strongly impaired at higher compound concentrations indicating inhibition of NC(12-55) chaperone properties (see **Materials and methods**). **BN5-01**, **BN5-02**, **BN5-07** and **BN5-15** also showed to affect the formation of the TAR/cTAR hybrid duplex inhibiting the activity of NC(12-55). No detectable activity was observed for other compounds as the hybrid duplex was formed by NC despite the presence of compounds.

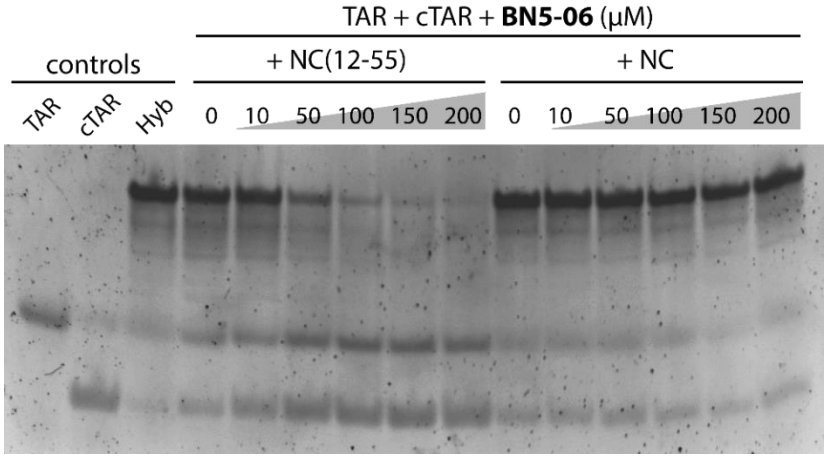


Figure S2. NAME assay employing NC and NC(12-55) conducted with BN5-06. The result showed that the **BN5-06** compound displayed lower potency by NAME assay when tested using the full-length NC compared to the truncated NC(12-55) peptide. Almost complete inhibition of NC(12-55) function was observed at higher compound concentration whereas no effect was observed against the full-length protein assessed under the same experimental conditions.

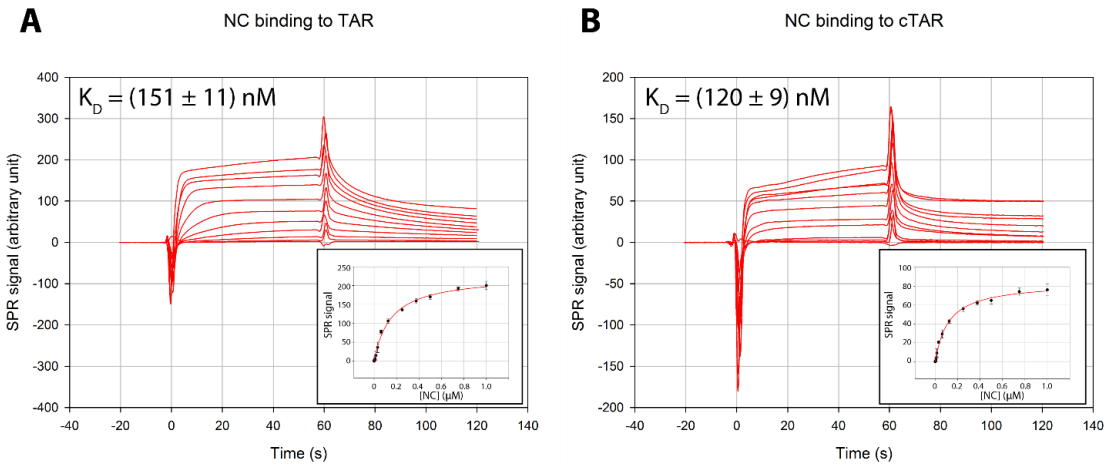


Figure S3. Binding affinity of NC for TAR and cTAR substrates assessed by SPR. Binding of NC was assessed by flowing increasing concentration of NC onto an SPR sensor chip. The latter was a commercially available streptavidin-coated chip functionalized with either TAR or cTAR that served as receptors (see **Materials and methods**). SPR curves showing association and dissociation phases of NC to TAR (A) or cTAR (B). Insets: binding isotherms plotted using SPR data.

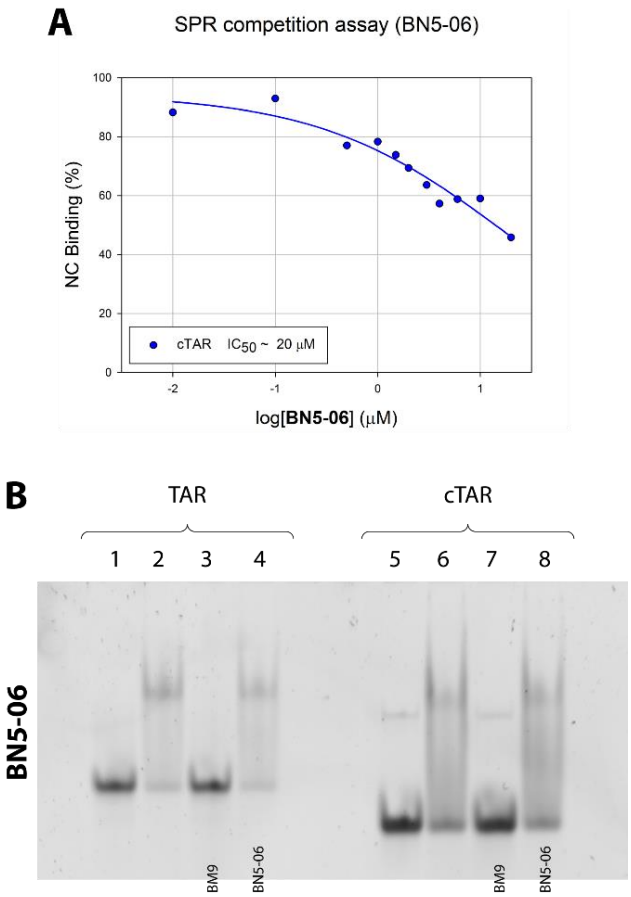


Figure S4. SPR-competition assay and EMSA conducted with BN5-06. BN5-06 was tested by SPR-competition assay (A) employing the cTAR functionalized sensor chip (see **Materials and methods**). Under these conditions, the compound showed little inhibition of binding of the full-length NC. **BN5-06** was tested by EMSA assay (B) that assesses the formation of NC-oligonucleotide complexes by gel electrophoresis analysis (see **Materials and methods**). NC binding to TAR and cTAR led to a band shift to lower electrophoretic mobility due to the formation of protein-oligonucleotide adducts (compare lanes 1 and 2, 5 and 6). Addition of **BN5-06** did not affect NC binding to TAR and cTAR (lanes 4 and 8) differently from the **BM9** compound, used here as a positive control (lanes 3 and 7).

ID	Experimental (u)	Calculated (u)	Accuracy (ppm)
NC	6488.91	6488.91	-1.02
BM6	343.03	343.03	-1.30
BM9	358.03	358.03	0.34
BN5	318.07	318.07	-0.91
BN5-01	362.07	362.07	5.62
BN5-06	330.04	330.04	8.76
BN5-07	360.04	360.04	-2.00
BN5-15	391.98	391.98	0.85
NC/**BM6**	6831.93	6831.95	-2.99
NC/**BM9**	6846.93	6846.97	-5.05
NC/**BN5**	6806.92	6806.97	7.35
NC/**BN5-01**	6850.99	6850.98	-1.46
NC/**BN5-06**	6818.96	6818.95	-1.47
NC/**BN5-07**	6848.96	6848.95	-1.46
NC/**BN5-15**	6880.91	6880.89	-2.91

Table S1. Experimental masses in mass units (u) observed for NC and NC in complex with small molecules considered in the study. Expected mono-isotopic masses were calculated by adding masses calculated from the sequence of NC(1-55), two coordinated Zn(II) ions, and the elemental composition of each ligand. Accuracy expressed in part per million (ppm) was calculated as ($mass_{calc}$ – $mass_{exp}$) / $mass_{calc}$.

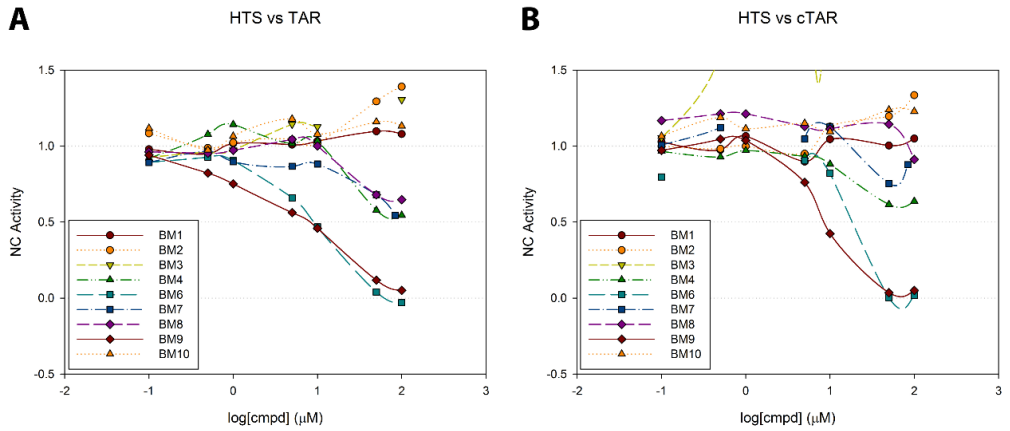


Figure S5. HTS assay employing TAR and cTAR substrates conducted with BM compounds. A clear dose-dependent inhibition of the NC nucleic acid destabilization properties was observed by a decrease in the fluorescence signal from the NC-oligonucleotide mixture for **BM6** and **BM9** (see **Materials and methods**). **BM7**, **BM8** and **BM4** displayed only weak activity at higher concentrations, other compounds appeared inactive. An upward trend in the curve was observed for **BM3**, especially with the cTAR substrate, indicating an increase in the fluorescence signal at higher compound concentrations that might be related to precipitation of compound in the assay conditions.

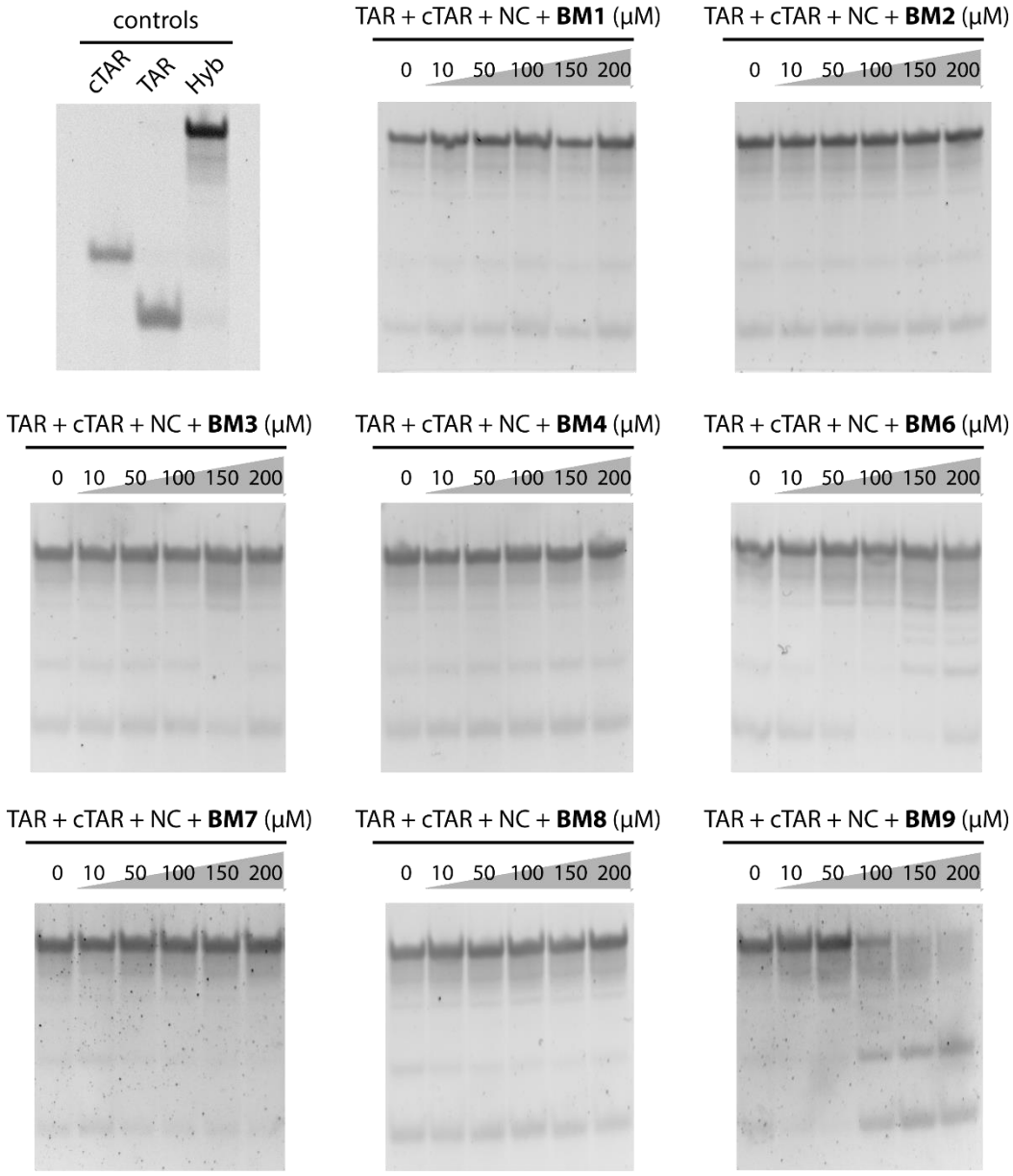


Figure S6. NAME assay conducted with BM compounds. Monomeric TAR and cTAR, as well as the heat-refolded hybrid duplex (Hyb), were included as controls (upper, left panel). A clear dose-dependent inhibition of the NC nucleic acid chaperone properties was observed with **BM9**. Formation of the band corresponding to the hybrid duplex specie with lower electrophoretic mobility compared to the TAR and cTAR monomeric species almost disappeared at higher compound concentrations indicating inhibition of NC chaperone properties (see **Materials and methods**). **BM6** showed little effect on the NC-mediated TAR/cTAR annealing as additional bands were observed with intermediate mobility between the monomeric and duplex species. These bands are compatible with the accumulation of partially annealed intermediates, indicating a decrease in the annealing rate. Other compounds displayed no activity as the hybrid duplex was formed by NC despite the presence of compounds.

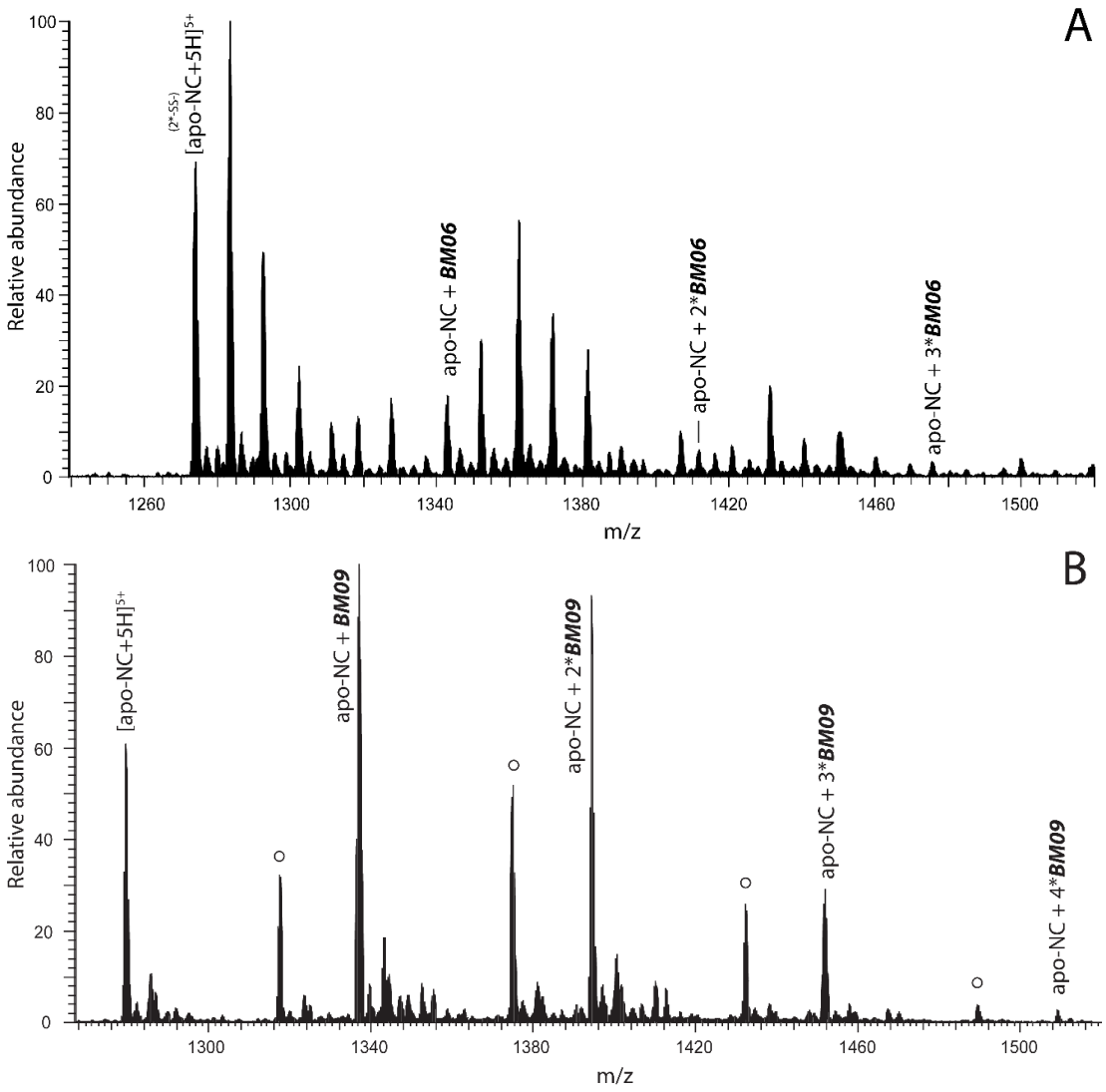


Figure S7. ESI-MS spectrum of NC in presence of BM6 and BM9 under denaturing conditions. NC was incubated with **BM6** and **BM9** for two hours and a mixture of water/methanol/acetic acid 49:49:2 was added 10 minutes before ESI-MS analysis in order to denature NC structure and eject the zinc ions (see **Materials and methods**). Under denaturing conditions, NC-compound complexes were still observable for both **BM6** (A) and **BM9** (B), indicating that binding is not afforded by the hydrophobic pocket of NC. In the case of **BM6**, additional peaks were observed corresponding to oxidation of the free cysteine to disulfides or cysteic acid. Peaks marked by cycles in (B) correspond to protein complexes involving an impurity present in the compound stock solution. Low-intensity signals observed at slightly higher m/z than the corresponding NC signal consist of typical sodium, potassium and ammonium adducts.

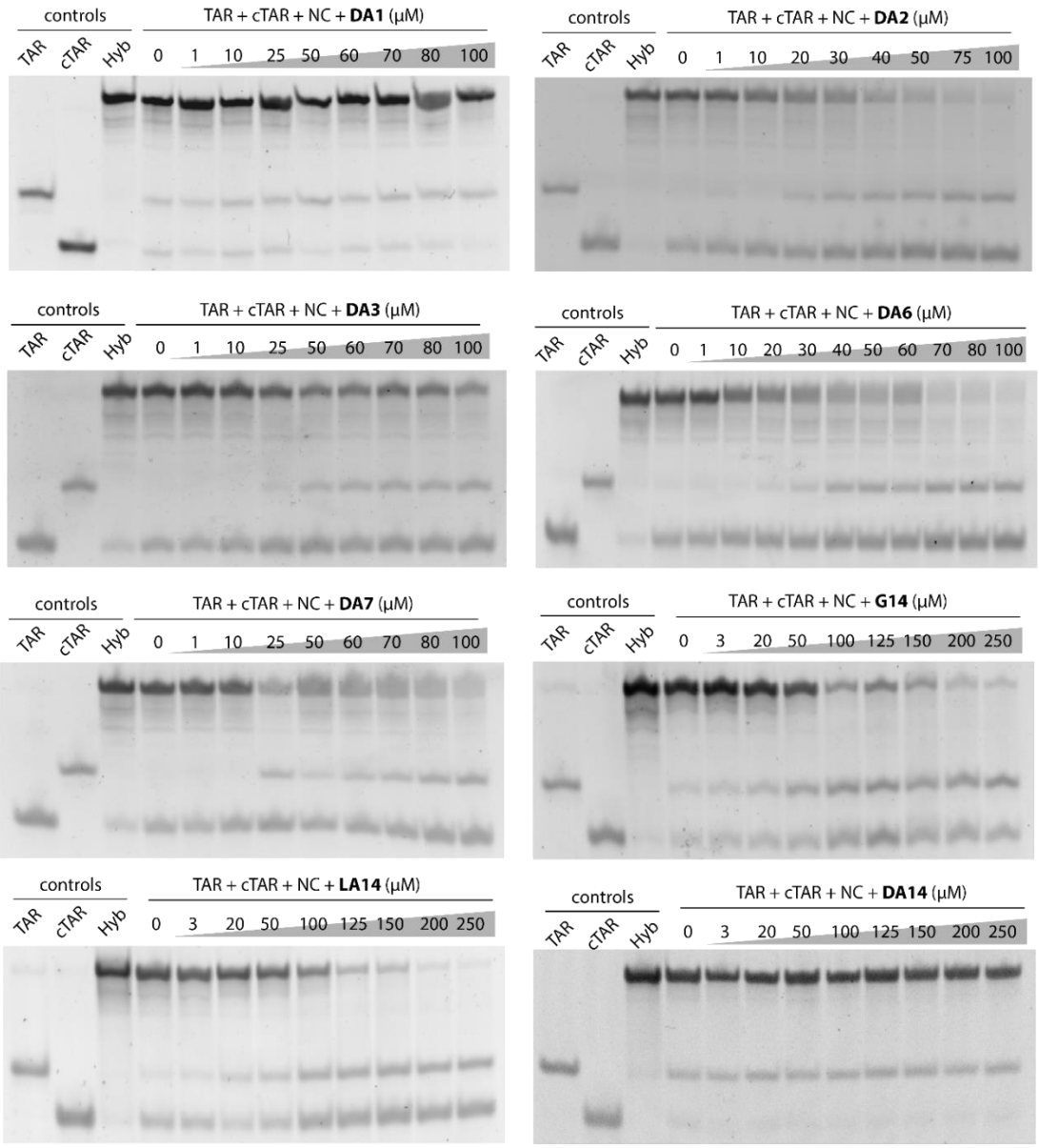


Figure S8. NAME assay conducted with AQs compounds from the DA series. Picture showing representative gels used to calculate IC_{50} values reported in **Table 4.4**. Monomeric TAR and cTAR, as well as the heat-refolded hybrid duplex (Hyb), were included as controls.

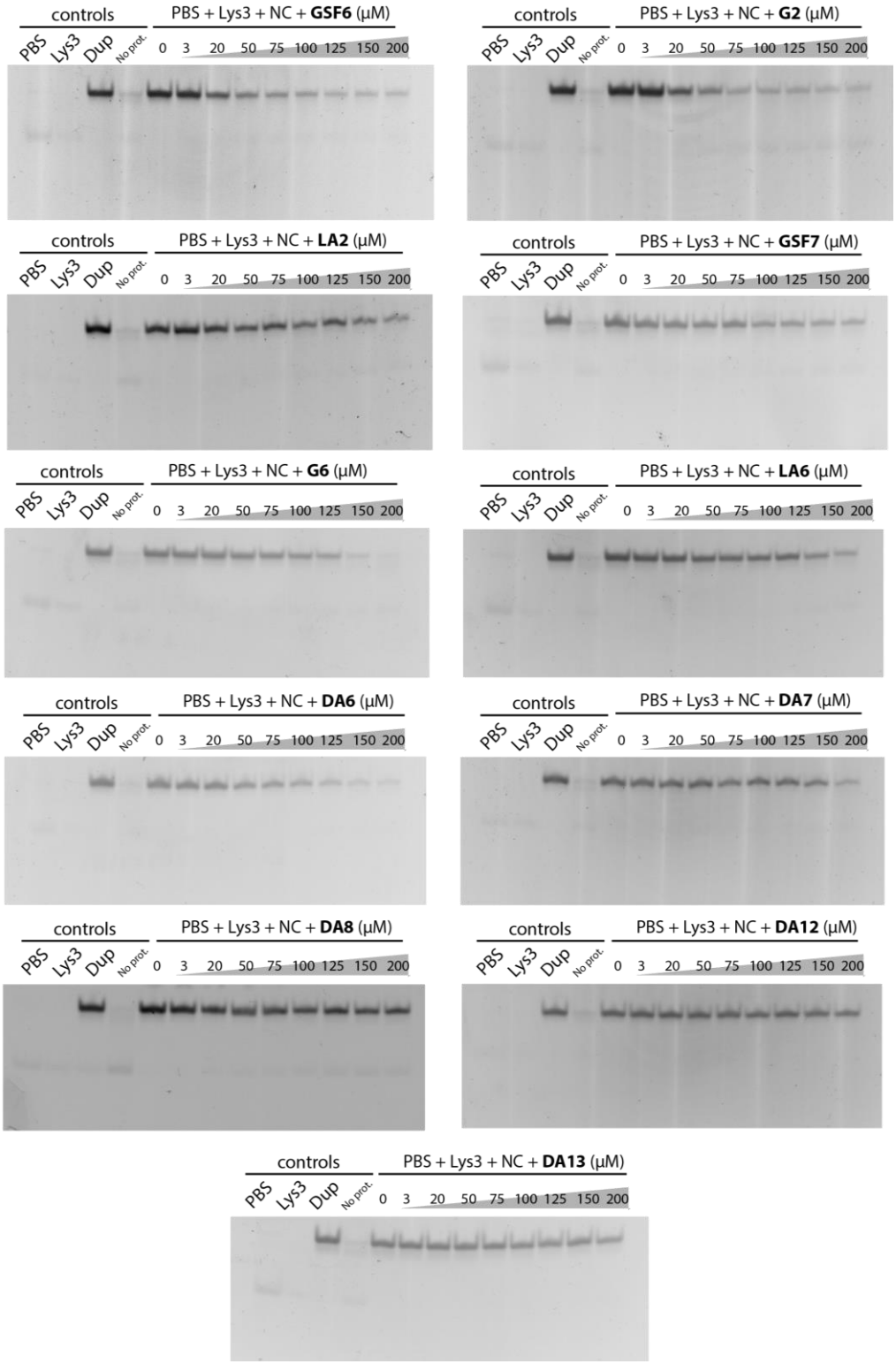


Figure S9. RAMPA assay employing NC conducted with AQs compounds. Monomeric PBS and Lys3, as well as the heat-refolded duplex (Dup), were included as controls. Bands corresponding to PBS and Lys3 (see control lanes) were barely visible as the sequences were poorly stained by the dye used. Annealing of PBS to Lys3 appeared scarce in the absence of protein and almost no duplex formation was observed (no prot.) whereas addition of proteins led to complete annealing. Picture showing representative gels used to calculate IC_{50} values reported in **Table 4.7** (see **Materials and Methods**).

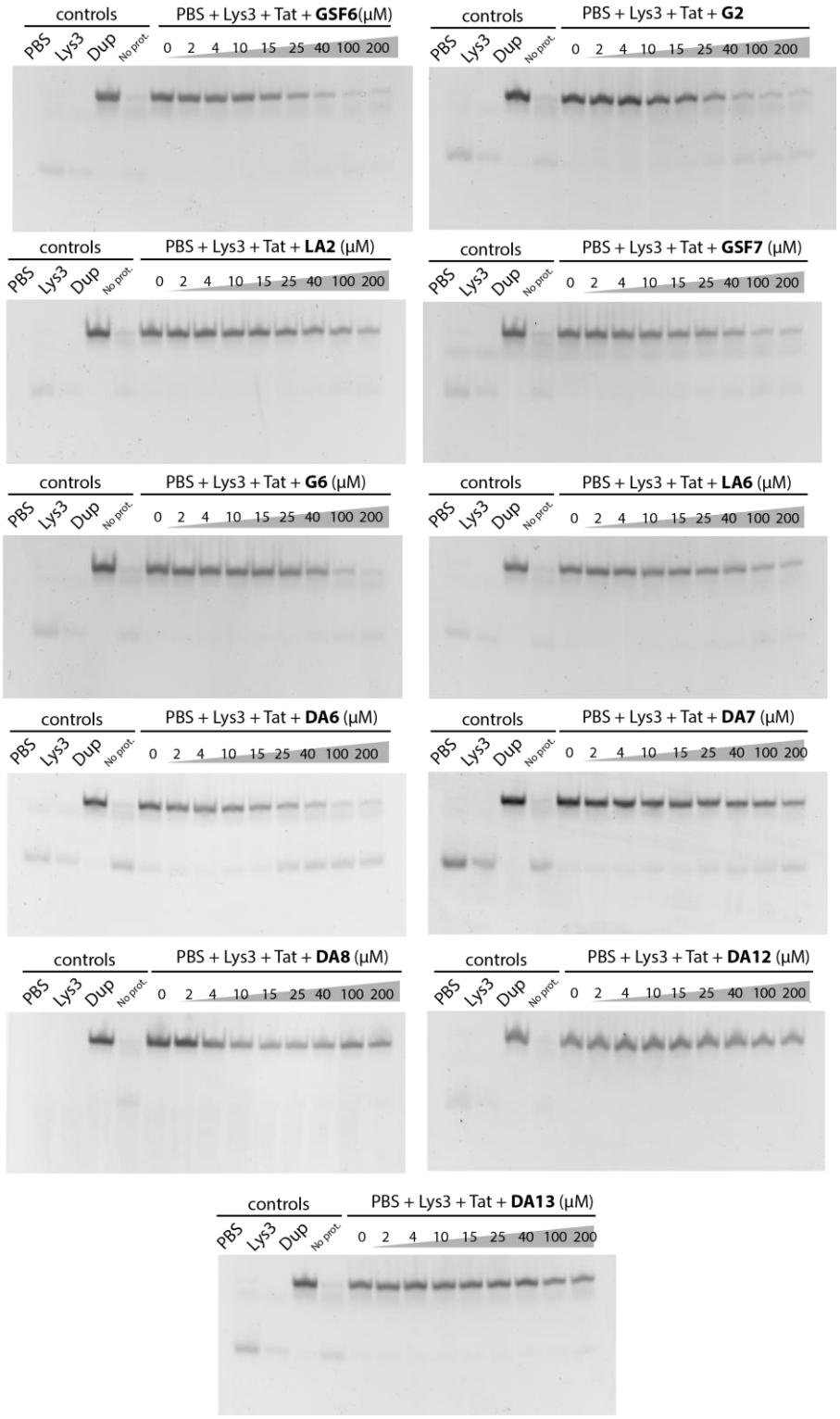


Figure S10. RAMPA assay employing Tat48-57 conducted with AQs compounds. Monomeric PBS and Lys3, as well as the heat-refolded duplex (Dup), were included as controls. Bands corresponding to PBS and Lys3 (see control lanes) were barely visible as the sequences were poorly stained by the dye used. Annealing of PBS to Lys3 appeared scarce in the absence of protein and almost no duplex formation was observed (no prot.) whereas addition of proteins led to complete annealing. Picture showing representative gels used to calculate IC_{50} values reported in **Table 4.7** (see **Materials and Methods**).

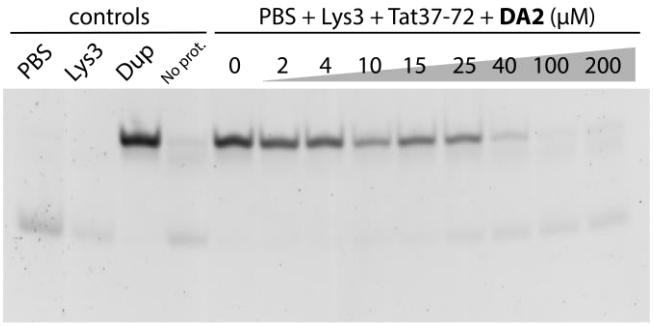


Figure S11. RAMPA assay employing Tat37-72 conducted with DA2. Monomeric PBS and Lys3, as well as the heat-refolded duplex (Dup), were included as controls. Bands corresponding to PBS and Lys3 (see control lanes) were barely visible as the sequences were poorly stained by the dye used. Annealing of PBS to Lys3 appeared scarce in the absence of protein and almost no duplex formation was observed (no prot., control lanes) whereas addition of proteins led to complete annealing. A clear dose-dependent inhibition of the RNA annealing properties of Tat37-72 was observed by a decrease in the formation of the band corresponding to the duplex specie with lower electrophoretic mobility compared to the PBS and Lys3 monomeric species (see **Materials and methods**). The results indicated that the compound inhibited the annealing properties of Tat37-72 peptide with similar efficiency to that observed with the shorter version Tat48-57.

ID	Experimental (u)	Calculated (u)	Accuracy (ppm)
PBS	5805.85	5805.86	-7.75E-07
Lys3	5696.76	5696.82	-1.05E-05
PBS/**GSF6** 1:1	6414.16	6414.21	-7.33E-06
PBS/**GSF6** 1:2	8778.08	8778.14	-6.89E-06
PBS/**G2** 1:1	6414.16	6414.17	-2.49E-06
PBS/**G2** 1:2	7022.47	7022.49	-3.99E-06
PBS/**LA2** 1:1	6442.19	6442.18	2.02E-06
PBS/**LA2** 1:2	7078.53	7078.51	2.33E-06
PBS/**DA2** 1:1	6442.19	6442.22	-4.24E-06
PBS/**DA2** 1:2	7078.53	7078.60	-9.03E-06
PBS/**GSF7** 1:1	6386.13	6386.12	1.32E-06
PBS/**GSF7** 1:2	6966.41	6966.39	2.80E-06
PBS/**G6** 1:1	6386.13	6386.12	1.08E-06
PBS/**G6** 1:2	6966.40	6966.34	8.50E-06
PBS/**LA6** 1:1	6414.16	6414.14	3.60E-06
PBS/**DA6** 1:1	6414.16	6414.19	-4.37E-06
PBS/**DA6** 1:2	7022.47	7022.50	-5.13E-06
PBS/**DA7** 1:1	6386.13	6386.13	-6.26E-07
PBS/**DA8** 1:1	6358.10	6358.15	-8.30E-06
PBS/**DA8** 1:2	6414.16	6414.21	-7.33E-06
Lys3/**GSF6** 1:1	6305.07	6305.17	-1.54E-05
Lys3/**GSF6** 1:2	6913.38	6913.48	-1.56E-05
Lys3/**G2** 1:1	6305.07	6305.09	-3.09E-06
Lys3/**G2** 1:2	6913.38	6913.41	-5.03E-06
Lys3/**LA2** 1:1	6333.10	6333.14	-5.37E-06
Lys3/**DA2** 1:1	6333.10	6333.13	-3.55E-06
Lys3/**DA2** 1:2	6969.44	6969.54	-1.36E-05
Lys3/**GSF7** 1:1	6277.04	6277.11	-1.08E-05
Lys3/**GSF7** 1:2	6857.32	6857.40	-1.30E-05
Lys3/**G6** 1:1	6277.04	6277.11	-1.08E-05
Lys3/**G6** 1:2	6857.32	6857.40	-1.30E-05
Lys3/**LA6** 1:1	6305.07	6305.06	1.52E-06
Lys3/**DA6** 1:1	6305.07	6305.06	1.52E-06
Lys3/**DA6** 1:2	6913.38	6913.41	-4.40E-06
Lys3/**DA7** 1:1	6277.04	6277.01	4.30E-06
Lys3/**DA8** 1:1	6249.01	6248.98	5.12E-06

Table S2. Experimental masses in mass units (u) observed for PBS and Lys3 RNAs and their complex with anthraquinone conjugates considered in the study. Expected mono-isotopic masses were calculated by adding masses calculated from the sequence of PBS or Lys3, and the elemental composition of each ligand. Accuracy expressed in part per million (ppm) was calculated as $(mass_{calc} - mass_{exp}) / mass_{calc}$.

ABBREVIATION LIST

(-)PBS	primer binding site (negative DNA strand)
(-)ssDNA	minus strand strong stop DNA
(+)PBS	primer binding site (positive DNA strand)
(+)ssDNA	plus strand strong stop DNA
AIDS	acquired immunodeficiency syndrome
apo-NC	NC lacking both Zn(II) ions
AQs	2,6-dipeptidyl-anthraquinones conjugates
ART	antiretroviral therapy
CA	capsid protein
CD	circular dichroism
cDNA	viral complementary DNA genome
CD4	cluster or diversity 4
DABCYL	4-(4'-dimethylaminophenylazo) benzoic acid
EMSA	electrophoresis mobility shift assay
ESI-MS	electrospray-ionization mass spectrometry
FAM	6-carboxyfluorescein
FQA	fluorescence quenching assay
gRNA	genomic RNA
GSH	reduced glutathione
HAART	highly active antiretroviral therapy
HIV-1	human immunodeficiency virus type 1
HTS	high throughput screening
IN	integrase
LTR	long terminal repeats
MA	matrix protein
MOA	mechanism of action
NAME	nucleocapsid annealing mediated electrophoresis
NC	nucleocapsid protein
NEF	negative regulatory factor
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
OI	opportunistic infection
PAS	poly-adenylation signal
PBS	primer binding site (RNA)
PI	protease inhibitors
PIC	preintegration complex
PPT	polypurine tract
PR	protease
RAMPA	RNA-annealer mediated primer-annealing
Rev	regulator of the expression of the virus
RRE	rev responsive element
RT	reverse transcriptase
RTC	reverse transcription complex
SAR	structure-activity relationship
SL1 or DIS	dimerization initiation site
SL2 or SD1	major splice donor site
SPR	surface plasmon resonance
SU	surface protein
TAR	trans-activation response element
Tat	trans-activator of transcription
TM	transmembrane protein
UNADS	Joint United Nations Programme on HIV/AIDS
UTRs	untranslated terminal regions
Vif	viral infectivity factor
Vpr	viral protein
Vpu	viral protein unique
ZF	zinc finger
zinc-less NC	NC lacking one Zn(II) ion
Ψ-RNA	packaging signal sequence

BIBLIOGRAPHY

1. Sharp, P. M.; Hahn, B. H. Origins of HIV and the AIDS pandemic. *Csh. Pespect. Med.* **2011,** 1, a006841.

2. Macher, A. M. The pathology of AIDS. *Public health Rep.* **1988,** 103, 246.

3. Lucas, S. The pathology of HIV infection. *Leprosy Rev.* **2002,** 73, 64-71.

4. Finzi, D.; Hermankova, M.; Pierson, T.; Carruth, L. M.; Buck, C.; Chaisson, R. E.; Quinn, T. C.; Chadwick, K.; Margolick, J.; Brookmeyer, R. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **1997,** 278, 1295-1300.

5. Clark, S. J.; Saag, M. S.; Decker, W. D.; Campbell-Hill, S.; Roberson, J. L.; Veldkamp, P. J.; Kappes, J. C.; Hahn, B. H.; Shaw, G. M. High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. *New Engl. J. Med.* **1991,** 324, 954-960.

6. Seddon, J.; Bhagani, S. Antimicrobial therapy for the treatment of opportunistic infections in HIV/AIDS patients: a critical appraisal. *HIV AIDS (Auckl)* **2011,** 3, 19-33.

7. Kaplan, L. D. HIV-associated lymphoma. *Best Pract. Res. Cl. Ha.* **2012,** 25, 101-117.

8. Gloghini, A.; Dolcetti, R.; Carbone, A. Lymphomas occurring specifically in HIV-infected patients: from pathogenesis to pathology. In *Seminars in Cancer Biology*, 6 ed.; Vincent, T., Ed. Elsevier: 2013; Vol. 23, pp 457-467.

9. McArthur, J. C.; Brew, B. J.; Nath, A. Neurological complications of HIV infection. *Lancet Neurol.* **2005,** 4, 543-555.

10. Hooshyar, D.; Hanson, D. L.; Wolfe, M.; Selik, R. M.; Buskin, S. E.; McNaghten, A. D. Trends in perimortal conditions and mortality rates among HIV-infected patients. *Aids* **2007,** 21, 2093-2100.

11. Taylor, B. S.; Sobieszczyk, M. E.; McCutchan, F. E.; Hammer, S. M. The challenge of HIV-1 subtype diversity. *New Engl. J. Med.* **2008,** 358, 1590-1602.

12. Frankel, A. D.; Young, J. A. T. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **1998,** 67, 1-25.

13. Watts, J. M.; Dang, K. K.; Gorelick, R. J.; Leonard, C. W.; Bess Jr, J. W.; Swanstrom, R.; Burch, C. L.; Weeks, K. M. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **2009,** 460, 711-716.

14. Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M. Nucleotide sequence of the AIDS virus, LAV. *Cell* **1985,** 40, 9-17.

15. Basu, V. P.; Song, M.; Gao, L.; Rigby, S. T.; Hanson, M. N.; Bambara, R. A. Strand transfer events during HIV-1 reverse transcription. *Virus Res.* **2008,** 134, 19-38.

16. Barboric, M.; Yik, J. H. N.; Czudnochowski, N.; Yang, Z.; Chen, R.; Contreras, X.; Geyer, M.; Matija Peterlin, B.; Zhou, Q. Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. *Nucleic Acids Res.* **2007,** 35, 2003-2012.

17. Skripkin, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *P. Natl. Acad. Sci. USA* **1994,** 91, 4945-4949.

18. Turner, K. B.; Hagan, N. A.; Fabris, D. Understanding the Isomerization of the HIV-1 Dimerization Initiation Domain by the Nucleocapsid Protein. *J. Mol. Biol.* **2007,** 369, 812-828.

19. Ocwieja, K. E.; Sherrill-Mix, S.; Mukherjee, R.; Custers-Allen, R.; David, P.; Brown, M.; Wang, S.; Link, D. R.; Olson, J.; Travers, K. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acid. Res.* **2012,** 40, 10345-10355.

20. De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F. Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element. *Science* **1998,** 279, 384-388.

21. Fernandes, J.; Jayaraman, B.; Frankel, A. The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. *RNA Biol.* **2012,** 9, 6-11.

22. Wilusz, J. Putting an 'End' to HIV mRNAs: capping and polyadenylation as potential therapeutic targets. *AIDS Res. Ther.* **2013,** 10, 1.

23. Freed, E. O. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* **2015,** 13, 484-496.

24. Ganser-Pornillos, B. K.; Yeager, M.; Sundquist, W. I. The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.* **2008,** 18, 203-217.

25. Craigie, R.; Bushman, F. D. HIV DNA integration. *Csh. Perspect. Med.* **2012,** 2, a006890.

26. Godet, J.; Boudier, C.; Humbert, N.; Ivanyi-Nagy, R.; Darlix, J.-L.; Mély, Y. Comparative nucleic acid chaperone properties of the nucleocapsid protein NCp7 and Tat protein of HIV-1. *Virus Res.* **2012,** 169, 349-360.

27. Chiu, Y.-L.; Ho, C. K.; Saha, N.; Schwer, B.; Shuman, S.; Rana, T. M. Tat stimulates cotranscriptional capping of HIV mRNA. *Mol. Cell* **2002,** 10, 585-597.

28. Jablonski, J. A.; Amelio, A. L.; Giacca, M.; Caputi, M. The transcriptional transactivator Tat selectively regulates viral splicing. *Nucleic Acid. Res.* **2009**, gkp1105.

29. Sanghvi, V. R.; Steel, L. F. A re-examination of global suppression of RNA interference by HIV-1. *PloS one* **2011,** 6, e17246.

30. Rose, K. M.; Marin, M.; Kozak, S. L.; Kabat, D. The viral infectivity factor (Vif) of HIV-1 unveiled. *Trends Mol. Med.* **2004,** 10, 291-297.

31. Peng, C.; Ho, B. K.; Chang, T. W.; Chang, N. T. Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. *J. Virol.* **1989,** 63, 2550-2556.

32. Ohishi, M.; Nakano, T.; Sakuragi, S.; Shioda, T.; Sano, K.; Sakuragi, J.-i. The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. *Nucleic Acid. Res.* **2011,** 39, 3404-3417.

33. Muriaux, D.; Darlix, J.-L. Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol.* **2010,** 7, 744-753.

34. Wilen, C. B.; Tilton, J. C.; Doms, R. W. HIV: cell binding and entry. *Csh. Perspect. Med.* **2012,** 2, a006866.

35. Oberlin, E.; Amara, A.; Bessia, C.; Virelizier, J.-L.; Arenzana-Seisdedos, F.; Schwartz, O.; Heard, J.-M.; Clark-Lewis, I.; Legler, D. F.; Loetscher, M. The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* **1996**.

36. Dragic, T.; Litwin, V.; Allaway, G. P.; Martin, S. R.; Huang, Y.; Nagashima, K. A.; Cayanan, C.; Maddon, P. J.; Koup, R. A.; Moore, J. P. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **1996**.

37. Haqqani, A. A.; Tilton, J. C. Entry inhibitors and their use in the treatment of HIV-1 infection. *Antiviral Res.* **2013,** 98, 158-170.

38. Lehmann-Che, J.; Saïb, A. Early stages of HIV replication: how to hijack cellular functions for a successful infection. *AIDS Rev.* **2004,** 6, 199-207.

39. Hu, W.-S.; Hughes, S. H. HIV-1 reverse transcription. *Csh. Perspect. Med.* **2012,** 2, a006882.

40. Suzuki, Y.; Chew, M. L.; Suzuki, Y. Role of host-encoded proteins in restriction of retroviral integration. *Front. Microbiol.* **2012,** 3, 227.

41. Wei, P.; Garber, M. E.; Fang, S.-M.; Fischer, W. H.; Jones, K. A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **1998,** 92, 451-462.

42. Balvay, L.; Lastra, M. L.; Sargueil, B.; Darlix, J.-L.; Ohlmann, T. Translational control of retroviruses. *Nat. Rev. Microbiol.* **2007,** 5, 128-140.

43. Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R.; Summers, M. F. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* **2000,** 301, 491-511.

44. Menéndez-Arias, L. Molecular basis of human immunodeficiency virus type 1 drug resistance: overview and recent developments. *Antiviral Res.* **2013,** 98, 93-120.

45. Dolgin, E. Long-acting HIV drugs advanced to overcome adherence challenge. *Nat. Med.* **2014,** 20, 323-324.

46. Adamson, C. S.; Freed, E. O. Novel approaches to inhibiting HIV-1 replication. *Antiviral Res.* **2010,** 85, 119-141.

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

48. Johnston, M. I.; Fauci, A. S. An HIV vaccine—evolving concepts. *New Engl. J. Med.* **2007,** 356, 2073-2081.

49. Archin, N. M.; Sung, J. M.; Garrido, C.; Soriano-Sarabia, N.; Margolis, D. M. Eradicating HIV-1 infection: seeking to clear a persistent pathogen. *Nat. Rev. Microbiol.* **2014,** 12, 750-764.

50. Ensoli, B.; Bellino, S.; Tripiciano, A.; Longo, O.; Francavilla, V.; Marcotullio, S.; Cafaro, A.; Picconi, O.; Paniccia, G.; Scoglio, A. Therapeutic immunization with HIV-1 Tat reduces immune activation and loss of regulatory T-cells and improves immune function in subjects on HAART. *PloS one* **2010,** 5, e13540.

51. Darlix, J.-L.; Lapadat-Tapolsky, M.; de Rocquigny, H.; Roques, B. P. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **1995,** 254, 523-537.

52. Henderson, L. E.; Bowers, M. A.; Sowder, R.; Serabyn, S. A.; Johnson, D. G.; Bess, J. W.; Arthur, L. O.; Bryant, D. K.; Fenselau, C. Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J. Virol.* **1992,** 66, 1856-1865.

53. South, T. L.; Blake, P. R.; Sowder Iii, R. C.; Arthur, L. O.; Henderson, L. E.; Summers, M. F. The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers. *Biochemistry* **1990,** 29, 7786-7789.

54. Green, L. M.; Berg, J. M. Retroviral nucleocapsid protein-metal ion interactions: folding and sequence variants. *P. Natl. Acad. Sci. USA* **1990,** 87, 6403-6407.

55. Morellet, N.; Jullian, N.; De Rocquigny, H.; Maigret, B.; Darlix, J. L.; Roques, B. P. Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR. *The EMBO journal* **1992,** 11, 3059.

56. Clore, M. G.; Deshmukh, L.; Schwieters, C.; Grishaev, A. Quantitative characterization of configurational space sampled by HIV-1 nucleocapsid using solution NMR, X-ray scattering and protein engineering. *ChemPhysChem* **2016,** 17, 1548-1552.

57. Bazzi, A.; Zargarian, L.; Chaminade, F.; Boudier, C.; De Rocquigny, H.; René, B.; Mély, Y.; Fossé, P.; Mauffret, O. Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: role of sugar deoxyriboses in the binding polarity of NC. *Nucleic Acids Res.* **2011,** 39, 3903-3916.

58. Bazzi, A.; Zargarian, L.; Chaminade, F.; De Rocquigny, H.; René, B.; Mély, Y.; Fossé, P.; Mauffret, O. Intrinsic Nucleic Acid Dynamics Modulates HIV-1 Nucleocapsid Protein Binding to Its Targets. *PLoS ONE* **2012,** 7, e38905.

59. Darlix, J.-L.; de Rocquigny, H.; Mauffret, O.; Mély, Y. Retrospective on the all-in-one retroviral nucleocapsid protein. *Virus Res.* **2014,** 193, 2-15.

60. Wu, H.; Mitra, M.; Naufer, M. N.; McCauley, M. J.; Gorelick, R. J.; Rouzina, I.; Musier-Forsyth, K.; Williams, M. C. Differential contribution of basic residues to HIV-1 nucleocapsid protein's nucleic acid chaperone function and retroviral replication. *Nucleic Acid. Res.* **2014,** 42, 2525-2537.

61. Beltz, H.; Clauss, C.; Piémont, E.; Ficheux, D.; Gorelick, R. J.; Roques, B.; Gabus, C.; Darlix, J.-L.; de Rocquigny, H.; Mély, Y. Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J. Mol. Biol.* **2005,** 348, 1113-1126.

62. Rajkowitsch, L.; Chen, D.; Stampfl, S.; Semrad, K.; Waldsich, C.; Mayer, O.; Jantsch, M. F.; Konrat, R.; Bläsi, U.; Schroeder, R. RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* **2007,** 4, 118-130.

63. Doetsch, M.; Schroeder, R.; Fürtig, B. Transient RNA–protein interactions in RNA folding. *The FEBS journal* **2011,** 278, 1634-1642.

64. Godet, J.; Mély, Y. Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol.* **2010,** 7, 687-699.

65. South, T. L.; Summers, M. F. Zinc-and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci.* **1993,** 2, 3-19.

66. Berglund, J. A.; Charpentier, B.; Rosbash, M. A high affinity binding site for the HIV-1 nucleocapsid protein. *Nucleic Acids Res.* **1997,** 25, 1042-1049.

67. Fisher, R. J.; Rein, A.; Fivash, M.; Urbaneja, M. A.; Casas-Finet, J. R.; Medaglia, M.; Henderson, L. E. Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. Virol.* **1998,** 72, 1902-1909.

68. Urbaneja, M. a. A.; Kane, B. P.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Casas-Finet, J. R. Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function1. *J. Mol. Biol.* **1999,** 287, 59-75.

69. Vuilleumier, C.; Bombarda, E.; Morellet, N.; Gerard, D.; Roques, B. P.; Mely, Y. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry* **1999,** 38, 16816-16825.

70. Demene, H.; Dong, C. Z.; Ottmann, M.; Rouyez, M. C.; Jullian, N.; Morellet, N.; Mely, Y.; Darlix, J. L.; Fournie-Zaluski, M. C. 1H NMR structure and biological studies of the His23. fwdarw. Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity. *Biochemistry* **1994,** 33, 11707-11716.

71. Wu, H.; Mitra, M.; McCauley, M. J.; Thomas, J. A.; Rouzina, I.; Musier-Forsyth, K.; Williams, M. C.; Gorelick, R. J. Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus Res.* **2013,** 171, 263-277.

72. Tsuchihashi, Z.; Brown, P. O. DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **1994,** 68, 5863-5870.

73. Tompa, P.; Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *The FASEB Journal* **2004,** 18, 1169-1175.

74. Hargittai, M. R. S.; Gorelick, R. J.; Rouzina, I.; Musier-Forsyth, K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.* **2004,** 337, 951-968.

75. Hargittai, M. R. S.; Mangla, A. T.; Gorelick, R. J.; Musier-Forsyth, K. HIV-1 nucleocapsid protein zinc finger structures induce tRNA Lys, 3 structural changes but are not critical for primer/template annealing. *J. Mol. Biol.* **2001,** 312, 985-997.

76. Sosic, A.; Frecentese, F.; Perissutti, E.; Sinigaglia, L.; Santagada, V.; Caliendo, G.; Magli, E.; Ciano, A.; Zagotto, G.; Parolin, C.; Gatto, B. Design, synthesis and biological evaluation of TAR and cTAR binders as HIV-1 nucleocapsid inhibitors. *MedChemComm* **2013,** 4, 1388-1393.

77. Sosic, A.; Cappellini, M.; Sinigaglia, L.; Jacquet, R.; Deffieux, D.; Fabris, D.; Quideau, S.; Gatto, B. Polyphenolic C-glucosidic ellagitannins present in oak-aged wine inhibit HIV-1 nucleocapsid protein. *Tetrahedron* **2015,** 71, 3020-3026.

78. Sosic, A.; Sinigaglia, L.; Cappellini, M.; Carli, I.; Parolin, C.; Zagotto, G.; Sabatino, G.; Rovero, P.; Fabris, D.; Gatto, B. Mechanisms of HIV-1 Nucleocapsid Protein Inhibition by Lysyl-Peptidyl-Anthraquinone Conjugates. *Bioconjugate Chem.* **2016,** 27, 247-256.

79. Frecentese, F.; Sosic, A.; Saccone, I.; Gamba, E.; Link, K.; Miola, A.; Cappellini, M.; Cattelan, M. G.; Severino, B.; Fiorino, F.; Magli, E.; Corvino, A.; Perissutti, E.; Fabris, D.; Gatto, B.; Caliendo, G.; Santagada, V. Synthesis and in Vitro Screening of New Series of 2,6-Dipeptidyl-anthraquinones: Influence of Side Chain Length on HIV-1 Nucleocapsid Inhibitors. *J. Med. Chem.* **2016,** 59, 1914-1924.

80. Sosic, A.; Cappellini, M.; Scalabrin, M.; Gatto, B. Nucleocapsid Annealing-Mediated Electrophoresis (NAME) Assay Allows the Rapid Identification of HIV-1 Nucleocapsid Inhibitors. *J. Vis. Exp.* **2015**, e52474.

81. Gorelick, R. J.; Nigida, S. M.; Bess, J. W.; Arthur, L. O.; Henderson, L. E.; Rein, A. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **1990,** 64, 3207-3211.

82. Aldovini, A.; Young, R. A. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* **1990,** 64, 1920-1926.

83. Dorfman, T.; Luban, J.; Goff, S. P.; Haseltine, W. A.; Göttlinger, H. G. Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **1993,** 67, 6159-6169.

84. Ottmann, M.; Gabus, C.; Darlix, J. L. The central globular domain of the nucleocapsid protein of human immunodeficiency virus type 1 is critical for virion structure and infectivity. *J. Virol.* **1995,** 69, 1778-1784.

85. You, J. C.; McHenry, C. S. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J. Biol. Chem.* **1994,** 269, 31491-31495.

86. Mori, M.; Kovalenko, L.; Lyonnais, S.; Antaki, D.; Torbett, B. E.; Botta, M.; Mirambeau, G.; Mély, Y. Nucleocapsid protein: a desirable target for future therapies against HIV-1. In *The Future of HIV-1 Therapeutics*, Torbett, B.; Goodsell, D., S.; Richman, D., Eds. Springer: 2015; pp 53-92.

87. Rice, W. G.; Schaeffer, C. A.; Harten, B.; Villinger, F.; South, T. L.; Summers, M. F.; Henderson, L. E.; Bess, J. W.; Arthur, L. O.; McDougal, J. S. Inhibition of HIV-1 infectivity by zinc-ejecting aromatic C-nitroso compounds. *Nature* **1993,** 361, 473-475.

88. Rice, W. G.; Supko, J. G.; Malspeis, L.; Buckheit, R. W.; Clanton, D.; Bu, M.; Graham, L.; Schaeffer, C. A.; Turpin, J. A.; Domagala, J. Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. *Science* **1995,** 270, 1194-1197.

89. Rice, W. G.; Turpin, J. A.; Huang, M.; Clanton, D.; Buckheit Jr, R. W.; Covell, D. G.; Wallqvist, A.; McDonnell, N. B.; DeGuzman, R. N.; Summers, M. F. Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. *Nat. Med.* **1997,** 3, 341-345.

90. Turpin, J. A.; Song, Y.; Inman, J. K.; Huang, M.; Wallqvist, A.; Maynard, A.; Covell, D. G.; Rice, W. G.; Appella, E. Synthesis and biological properties of novel pyridinioalkanoyl thiolesters

(PATE) as anti-HIV-1 agents that target the viral nucleocapsid protein zinc fingers. *J. Med. Chem.* **1999,** 42, 67-86.

91. Sancineto, L.; Mariotti, A.; Bagnoli, L.; Marini, F.; Desantis, J.; Iraci, N.; Santi, C.; Pannecouque, C.; Tabarrini, O. Design and synthesis of diselenobisbenzamides (DISeBAs) as nucleocapsid protein 7 (NCp7) inhibitors with anti-HIV activity. *J. Med. Chem.* **2015,** 58, 9601-9614.

92. Huang, M.; Maynard, A.; Turpin, J. A.; Graham, L.; Janini, G. M.; Covell, D. G.; Rice, W. G. Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J. Med. Chem.* **1998,** 41, 1371-1381.

93. Loo, J. A.; Holler, T. P.; Sanchez, J.; Gogliotti, R.; Maloney, L.; Reily, M. D. Biophysical characterization of zinc ejection from HIV nucleocapsid protein by anti-HIV 2, 2 '-dithiobis [benzamides] and benzisothiazolones. *J. Med. Chem.* **1996,** 39, 4313-4320.

94. Rice, W. G.; Turpin, J. A.; Schaeffer, C. A.; Graham, L.; Clanton, D.; Buckheit, R. W.; Zaharevitz, D.; Summers, M. F.; Wallqvist, A.; Covell, D. G. Evaluation of selected chemotypes in coupled cellular and molecular target-based screens identifies novel HIV-1 zinc finger inhibitors. *J. Med. Chem.* **1996,** 39, 3606-3616.

95. Schito, M. L.; Goel, A.; Song, Y.; Inman, J. K.; Fattah, R. J.; Rice, W. G.; Turpin, J. A.; Sher, A.; Appella, E. In vivo antiviral activity of novel human immunodeficiency virus type 1 nucleocapsid p7 zinc finger inhibitors in a transgenic murine model. *AIDS Res. Hum. Retrov.* **2003,** 19, 91-101.

96. Srivastava, P.; Schito, M.; Fattah, R. J.; Hara, T.; Hartman, T.; Buckheit, R. W.; Turpin, J. A.; Inman, J. K.; Appella, E. Optimization of unique, uncharged thioesters as inhibitors of HIV replication. *Bioorgan. Med. Chem.* **2004,** 12, 6437-6450.

97. Miller Jenkins, L. M.; Byrd, J. C.; Hara, T.; Srivastava, P.; Mazur, S. J.; Stahl, S. J.; Inman, J. K.; Appella, E.; Omichinski, J. G.; Legault, P. Studies on the mechanism of inactivation of the HIV-1 nucleocapsid protein NCp7 with 2-mercaptobenzamide thioesters. *J. Med. Chem.* **2005,** 48, 2847-2858.

98. Jenkins, L. M. M.; Hara, T.; Durell, S. R.; Hayashi, R.; Inman, J. K.; Piquemal, J.-P.; Gresh, N.; Appella, E. Specificity of acyl transfer from 2-mercaptobenzamide thioesters to the HIV-1 nucleocapsid protein. *J. Am. Chem. Soc.* **2007,** 129, 11067-11078.

99. Jenkins, L. M. M.; Ott, D. E.; Hayashi, R.; Coren, L. V.; Wang, D.; Xu, Q.; Schito, M. L.; Inman, J. K.; Appella, D. H.; Appella, E. Small-molecule inactivation of HIV-1 NCp7 by repetitive intracellular acyl transfer. *Nat. Chem. Biol.* **2010,** 6, 887-889.

100. Hartman, T. L.; Yang, L.; Helfrick, A. N.; Hassink, M.; Shank, N. I.; Rosenker, K. G.; Scerba, M. T.; Saha, M.; Hughes, E.; Wang, A. Q. Preclinical evaluation of a mercaptobenzamide and its prodrug for NCp7-targeted inhibition of human immunodeficiency virus. *Antiviral Res.* **2016,** 134, 216-225.

101. Yang, Y.; Zhu, J.; Hassink, M.; Jenkins, L. M. M.; Wan, Y.; Appella, D. H.; Xu, J.; Appella, E.; Zhang, X. A novel preventive strategy against HIV-1 infection: combinatorial use of inhibitors targeting the nucleocapsid and fusion proteins. *Emerg. Microbes Infec.* **2017,** 6, e40.

102. Saha, M.; Scerba, M. T.; Shank, N. I.; Hartman, T. L.; Buchholz, C. A.; Buckheit, R. W.; Durell, S. R.; Appella, D. H. Probing Mercaptobenzamides as HIV Inactivators via Nucleocapsid Protein 7. *ChemMedChem* **2017,** 12, 714-721.

103. Pannecouque, C.; Szafarowicz, B.; Volkova, N.; Bakulev, V.; Dehaen, W.; Mély, Y.; Daelemans, D. Inhibition of HIV-1 replication by a bis-thiadiazolbenzene-1, 2-diamine that chelates zinc ions from retroviral nucleocapsid zinc fingers. *Antimicrob. Agents Ch.* **2010,** 54, 1461-1468.

104. Vercruysse, T.; Basta, B.; Dehaen, W.; Humbert, N.; Balzarini, J.; Debaene, F.; Sanglier-Cianférani, S.; Pannecouque, C.; Mély, Y.; Daelemans, D. A phenyl-thiadiazolylidene-amine derivative ejects zinc from retroviral nucleocapsid zinc fingers and inactivates HIV virions. *Retrovirology* **2012,** 9, 95-95.

105. Druillennec, S.; Meudal, H.; Roques, B. P.; Fournié-Zaluski, M.-C. Nucleomimetic strategy for the inhibition of HIV-1 nucleocapsid protein NCp7 activities. *Bioorg. Med. Chem. Lett.* **1999,** 9, 627-632.

106. Kim, S. J.; Kim, M. Y.; Lee, J. H.; You, J. C.; Jeong, S. Selection and stabilization of the RNA aptamers against the human immunodeficiency virus type-1 nucleocapsid protein. *Biochem. Bioph. Res. Co.* **2002,** 291, 925-931.

107. Kim, M. Y.; Jeong, S. Inhibition of the functions of the nucleocapsid protein of human immunodeficiency virus-1 by an RNA aptamer. *Biochem. Bioph. Res. Co.* **2004,** 320, 1181-1186.

108. Pustowka, A.; Dietz, J.; Ferner, J.; Baumann, M.; Landersz, M.; Königs, C.; Schwalbe, H.; Dietrich, U. Identification of Peptide Ligands for Target RNA Structures Derived from the HIV-1 Packaging Signal ψ by Screening Phage-Displayed Peptide Libraries. *ChemBioChem* **2003,** 4, 1093-1097.

109. Raja, C.; Ferner, J.; Dietrich, U.; Avilov, S.; Ficheux, D.; Darlix, J.-L.; de Rocquigny, H.; Schwalbe, H.; Mély, Y. A tryptophan-rich hexapeptide inhibits nucleic acid destabilization chaperoned by the HIV-1 nucleocapsid protein. *Biochemistry* **2006,** 45, 9254-9265.

110. Dietz, J.; Koch, J.; Kaur, A.; Raja, C.; Stein, S.; Grez, M.; Pustowka, A.; Mensch, S.; Ferner, J.; Möller, L. Inhibition of HIV-1 by a Peptide Ligand of the Genomic RNA Packaging Signal Ψ. *ChemMedChem* **2008,** 3, 749-755.

111. Stephen, A. G.; Worthy, K. M.; Towler, E.; Mikovits, J. A.; Sei, S.; Roberts, P.; Yang, Q.-e.; Akee, R. K.; Klausmeyer, P.; McCloud, T. G. Identification of HIV-1 nucleocapsid protein: nucleic acid antagonists with cellular anti-HIV activity. *Biochem. Bioph. Res. Co.* **2002,** 296, 1228-1237.

112. Shvadchak, V.; Sanglier, S.; Rocle, S.; Villa, P.; Haiech, J.; Hibert, M.; Van Dorsselaer, A.; Mély, Y.; de Rocquigny, H. Identification by high throughput screening of small compounds inhibiting the nucleic acid destabilization activity of the HIV-1 nucleocapsid protein. *Biochimie* **2009,** 91, 916-923.

113. Mori, M.; Nucci, A.; Lang, M. C. D.; Humbert, N.; Boudier, C.; Debaene, F.; Sanglier-Cianferani, S.; Catala, M.; Schult-Dietrich, P.; Dietrich, U. Functional and structural characterization of 2-amino-4-phenylthiazole inhibitors of the HIV-1 nucleocapsid protein with antiviral activity. *ACS Chem. Biol.* **2014,** 9, 1950-1955.

114. Mori, M.; Schult-Dietrich, P.; Szafarowicz, B.; Humbert, N.; Debaene, F.; Sanglier-Cianferani, S.; Dietrich, U.; Mély, Y.; Botta, M. Use of virtual screening for discovering antiretroviral compounds interacting with the HIV-1 nucleocapsid protein. *Virus Res.* **2012,** 169, 377-387.

115. Breuer, S.; Chang, M. W.; Yuan, J.; Torbett, B. E. Identification of HIV-1 inhibitors targeting the nucleocapsid protein. *J. Med. Chem.* **2012,** 55, 4968-4977.

116. Goudreau, N.; Hucke, O.; Faucher, A.-M.; Grand-Maître, C.; Lepage, O.; Bonneau, P. R.; Mason, S. W.; Titolo, S. Discovery and structural characterization of a new inhibitor series of HIV-1 nucleocapsid function: NMR solution structure determination of a ternary complex involving a 2:1 inhibitor/NC stoichiometry. *J. Mol. Biol.* **2013,** 425, 1982-1998.

117. Mori, M.; Manetti, F.; Botta, M. Predicting the binding mode of known NCp7 inhibitors to facilitate the design of novel modulators. *J. Chem. Inf. Model.* **2010,** 51, 446-454.

118. Chung, J.; Mujeeb, A.; Jiang, Y.; Guilbert, C.; Pendke, M.; Wu, Y.; James, T. L. A Small Molecule, Lys-Ala-7-amido-4-methylcoumarin, Facilitates RNA Dimer Maturation of a Stem− Loop 1 Transcript in Vitro: Structure− Activity Relationship of the Activator†. *Biochemistry* **2008,** 47, 8148-8156.

119. Chung, J.; Ulyanov, N. B.; Guilbert, C.; Mujeeb, A.; James, T. L. Binding characteristics of small molecules that mimic nucleocapsid protein-induced maturation of stem-loop 1 of HIV-1 RNA. *Biochemistry* **2010,** 49, 6341-6351.

120. Ennifar, E.; Paillart, J.-C.; Bodlenner, A.; Walter, P.; Weibel, J.-M.; Aubertin, A.-M.; Pale, P.; Dumas, P.; Marquet, R. Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acid. Res.* **2006,** 34, 2328-2339.

121. Ennifar, E.; Paillart, J.-C.; Bernacchi, S.; Walter, P.; Pale, P.; Decout, J.-L.; Marquet, R.; Dumas, P. A structure-based approach for targeting the HIV-1 genomic RNA dimerization initiation site. *Biochimie* **2007,** 89, 1195-1203.

122. Bodlenner, A.; Alix, A.; Weibel, J.-M.; Pale, P.; Ennifar, E.; Paillart, J.-C.; Walter, P.; Marquet, R.; Dumas, P. Synthesis of a neamine dimer targeting the dimerization initiation site of HIV-1 RNA. *Org. Lett.* **2007,** 9, 4415-4418.

123. Bernacchi, S.; Freisz, S.; Maechling, C.; Spiess, B.; Marquet, R.; Dumas, P.; Ennifar, E. Aminoglycoside binding to the HIV-1 RNA dimerization initiation site: thermodynamics and effect on the kissing-loop to duplex conversion. *Nucleic Acid. Res.* **2007,** 35, 7128-7139.

124. Freisz, S.; Lang, K.; Micura, R.; Dumas, P.; Ennifar, E. Binding of Aminoglycoside Antibiotics to the Duplex Form of the HIV-1 Genomic RNA Dimerization Initiation Site. *Angew. Chem. Int. Edit.* **2008,** 47, 4110-4113.

125. Warui, D. M.; Baranger, A. M. Identification of small molecule inhibitors of the HIV-1 nucleocapsid–stem-loop 3 RNA complex. *J. Med. Chem.* **2012,** 55, 4132-4141.

126. Bell, N. M.; L'Hernault, A.; Murat, P.; Richards, J. E.; Lever, A. M. L.; Balasubramanian, S. Targeting RNA–protein interactions within the human immunodeficiency virus type 1 lifecycle. *Biochemistry* **2013,** 52, 9269-9274.

127. Mori, M.; Manetti, F.; Botta, M. Targeting Protein-Protein and Protein-Nucleic Acid Interactions for Anti-HIV Therapy. *Curr. Pharm. Design* **2011,** 17, 3713-3728.

128. Bourbigot, S.; Ramalanjaona, N.; Boudier, C.; Salgado, G. F. J.; Roques, B. P.; Mély, Y.; Bouaziz, S.; Morellet, N. How the HIV-1 nucleocapsid protein binds and destabilises the (−) primer binding site during reverse transcription. *J. Mol. Biol.* **2008,** 383, 1112-1128.

129. Mori, M.; Dietrich, U.; Manetti, F.; Botta, M. Molecular dynamics and DFT study on HIV-1 nucleocapsid protein-7 in complex with viral genome. *J. Chem. Inf. Model.* **2010,** 50, 638-650.

130. Cattelan, M. G. *Screening of small chemical entities as potential inhibitors of HIV-1 NCp7 chaperone*. Università di Padova: Padova, 2014.

131. Hagan, N.; Fabris, D. Direct Mass Spectrometric Determination of the Stoichiometry and Binding Affinity of the Complexes between Nucleocapsid Protein and RNA Stem−Loop Hairpins of the HIV-1 Ψ-Recognition Element. *Biochemistry* **2003,** 42, 10736-10745.

132. Hagan, N. A.; Fabris, D. Dissecting the Protein–RNA and RNA–RNA Interactions in the Nucleocapsid-mediated Dimerization and Isomerization of HIV-1 Stemloop 1. *J. Mol. Biol.* **2007,** 365, 396-410.

133. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acid. Res.* **2003,** 31, 3406-3415.

134. Waugh, A.; Gendron, P.; Altman, R.; Brown, J. W.; Case, D.; Gautheret, D.; Harvey, S. C.; Leontis, N.; Westbrook, J.; Westhof, E. RNAML: a standard syntax for exchanging RNA information. *RNA* **2002,** 8, 707-717.

135. Zuker, M.; Jacobson, A. B. Using reliability information to annotate RNA secondary structures. *RNA* **1998,** 4, 669-679.

136. Bernacchi, S.; Stoylov, S.; Piémont, E.; Ficheux, D.; Roques, B. P.; Darlix, J. L.; Mély, Y. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence1. *J. Mol. Biol.* **2002,** 317, 385-399.

137. Azoulay, J.; Clamme, J.-P.; Darlix, J.-L.; Roques, B. P.; Mély, Y. Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J. Mol. Biol.* **2003,** 326, 691-700.

138. Cosa, G.; Harbron, E. J.; Zeng, Y.; Liu, H.-W.; O'connor, D. B.; Eta-Hosokawa, C.; Musier-Forsyth, K.; Barbara, P. F. Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1 NC protein. *Biophys. Journal* **2004,** 87, 2759-2767.

139. Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J.-P.; Ficheux, D.; Roques, B.; Darlix, J.-L.; Mély, Y. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **2003,** 328, 95-108.

140. Beltz, H.; Piémont, E.; Schaub, E.; Ficheux, D.; Roques, B.; Darlix, J.-L.; Mély, Y. Role of the Structure of the Top Half of HIV-1 cTAR DNA on the Nucleic Acid Destabilizing Activity of the Nucleocapsid Protein NCp7. *J. Mol. Biol.* **2004,** 338, 711-723.

141. Godet, J.; Kenfack, C.; Przybilla, F.; Richert, L.; Duportail, G.; Mély, Y. Site-selective probing of cTAR destabilization highlights the necessary plasticity of the HIV-1 nucleocapsid protein to chaperone the first strand transfer. *Nucleic Acid. Res.* **2013,** 41, 5036-5048.

142. Bernacchi, S.; Mély, Y. Exciton interaction in molecular beacons: a sensitive sensor for short range modifications of the nucleic acid structure. *Nucleic Acid. Res.* **2001,** 29, e62.

143. Stoylov, S. P.; Vuilleumier, C.; Stoylova, E.; De Rocquigny, H.; Roques, B. P.; Gerard, D.; Mely, Y. Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers* **1997,** 41, 301-312.

144. Davis, T. M.; Wilson, W. D. Determination of the refractive index increments of small molecules for correction of surface plasmon resonance data. *Anal. Biochem.* **2000,** 284, 348-353.

145. Turner, K. B.; Hagan, N. A.; Fabris, D. Inhibitory effects of archetypical nucleic acid ligands on the interactions of HIV-1 nucleocapsid protein with elements of Ψ-RNA. *Nucleic Acid. Res.* **2006,** 34, 1305-1316.

146. Turner, K. B.; Kohlway, A. S.; Hagan, N. A.; Fabris, D. Non-covalent probes for the investigation of structure and dynamics of protein-nucleic acid assemblies: the case of NC-mediated dimerization of genomic RNA in HIV-1. *Biopolymers* **2009,** 91, 283-296.

147. Turner, K. B.; Brinson, R. G.; Yi-Brunozzi, H. Y.; Rausch, J. W.; Miller, J. T.; Le Grice, S. F. J.; Marino, J. P.; Fabris, D. Structural probing of the HIV-1 polypurine tract RNA: DNA hybrid using classic nucleic acid ligands. *Nucleic Acid. Res.* **2008,** 36, 2799-2810.

148. Brinson, R. G.; Turner, K. B.; Yi-Brunozzi, H. Y.; Le Grice, S. F. J.; Fabris, D.; Marino, J. P. Probing Anomalous Structural Features in Polypurine Tract-Containing RNA− DNA Hybrids with Neomycin B. *Biochemistry* **2009,** 48, 6988-6997.

149. Fabris, D.; Zaia, J.; Hathout, Y.; Fenselau, C. Retention of thiol protons in two classes of protein zinc ion coordination centers. *J. Am. Chem. Soc.* **1996,** 118, 12242-12243.

150. Hathout, Y.; Fabris, D.; Han, M. S.; Sowder, R. C. N.; Henderson, L. E.; Fenselau, C. Characterization of intermediates in the oxidation of zinc fingers in human immunodeficiency virus type 1 nucleocapsid protein P7. *Drug Metab. Dispos.* **1996,** 24, 1395-1400.

151. Lippens, J. L.; Mangrum, J. B.; McIntyre, W.; Redick, B.; Fabris, D. A simple heated-capillary modification improves the analysis of non-covalent complexes by Z-spray electrospray ionization. *Rapid Commun. Mass Sp.* **2016,** 30, 773-783.

152. Turner, K. B.; Hagan, N. A.; Kohlway, A. S.; Fabris, D. Mapping Noncovalent Ligand Binding to Stemloop Domains of the HIV-1 Packaging Signal by Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectr.* **2006,** 17, 1402-1412.

153. Sirakawa, K.; Aki, O.; Tsujikawa, T.; Tsuda, T. S-Alkylthioisothioureas. I. *Chem. Pharm. Bull.* **1970,** 18, 235-242.

154. Maynard, A. T.; Covell, D. G. Reactivity of zinc finger cores: analysis of protein packing and electrostatic screening. *J. Am. Chem. Soc.* **2001,** 123, 1047-1058.

155. South, T. L.; Blake, P. R.; Hare, D. R.; Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **1991,** 30, 6342-6349.

156. Maynard, A. T.; Huang, M.; Rice, W. G.; Covell, D. G. Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *P. Natl. Acad. Sci. USA* **1998,** 95, 11578-11583.

157. Anzellotti, A. I.; Sabat, M.; Farrell, N. Covalent and Noncovalent Interactions for [Metal (dien) nucleobase] 2+ Complexes with l-Tryptophan Derivatives: Formation of Palladium− Tryptophan Species by Nucleobase Substitution under Biologically Relevant Conditions. *Inorg. Chem.* **2006,** 45, 1638-1645.

158. Anzellotti, A. I.; Bayse, C. A.; Farrell, N. P. Effects of nucleobase metalation on frontier molecular orbitals: Potential implications for π-stacking interactions with tryptophan. *Inorg. Chem.* **2008,** 47, 10425-10431.

159. Darlix, J.-L.; Godet, J.; Ivanyi-Nagy, R.; Fossé, P.; Mauffret, O.; Mély, Y. Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **2011,** 410, 565-581.

160. Carlson, C. B.; Vuyisich, M.; Gooch, B. D.; Beal, P. A. Preferred RNA Binding Sites for a Threading Intercalator Revealed by In Vitro Evolution. *Chem. Biol.* **2003,** 10, 663-672.

161. Gooch, B. D.; Beal, P. A. Recognition of duplex RNA by helix-threading peptides. *Journal of the American Chemical Society* **2004,** 126, 10603-10610.

162. Baudin, F.; Marquet, R.; Isel, C.; Darlix, J.-L.; Ehresmann, B.; Ehresmann, C. Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.* **1993,** 229, 382-397.

163. Paillart, J.-C.; Dettenhofer, M.; Yu, X.-f.; Ehresmann, C.; Ehresmann, B.; Marquet, R. First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* **2004,** 279, 48397-48403.

164. Lu, K.; Heng, X.; Garyu, L.; Monti, S.; Garcia, E. L.; Kharytonchyk, S.; Dorjsuren, B.; Kulandaivel, G.; Jones, S.; Hiremath, A. NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. *Science* **2011,** 334, 242-245.

165. Markham, N. R.; Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acid. Res.* **2005,** 33, W577-W581.

166. Markham, N. R.; Zuker, M. UNAFold: software for nucleic acid folding and hybridization. In *Bioinformatics: structure, function and applications*, Keith, J., M. , Ed. Springer: 2008; pp 3-31.

167. Kameoka, M.; Morgan, M.; Binette, M.; Russell, R. S.; Rong, L.; Guo, X.; Mouland, A.; Kleiman, L.; Liang, C.; Wainberg, M. A. The Tat protein of human immunodeficiency virus type 1 (HIV-1) can promote placement of tRNA primer onto viral RNA and suppress later DNA polymerization in HIV-1 reverse transcription. *J. Virol.* **2002,** 76, 3637-3645.

168. Kuciak, M.; Gabus, C.; Ivanyi-Nagy, R.; Semrad, K.; Storchak, R.; Chaloin, O.; Muller, S.; Mely, Y.; Darlix, J.-L. The HIV-1 transcriptional activator Tat has potent nucleic acid chaperoning activities in vitro. *Nucleic Acid. Res.* **2008,** 36, 3389-3400.

169. Boudier, C.; Storchak, R.; Sharma, K. K.; Didier, P.; Follenius-Wund, A.; Muller, S.; Darlix, J. L.; Mely, Y. The mechanism of HIV-1 Tat-directed nucleic acid annealing supports its role in reverse transcription. *J. Mol. Biol.* **2010,** 400, 487-501.

170. Boudier, C.; Humbert, N.; Chaminade, F.; Chen, Y.; de Rocquigny, H.; Godet, J.; Mauffret, O.; Fossé, P.; Mély, Y. Dynamic interactions of the HIV-1 Tat with nucleic acids are critical for Tat activity in reverse transcription. *Nucleic Acid. Res.* **2013,** 42, 1065-1078.

171. Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acid. Res.* **2009,** 37, 1713-1725.

172. Doetsch, M.; Fürtig, B.; Gstrein, T.; Stampfl, S.; Schroeder, R. The RNA annealing mechanism of the HIV-1 Tat peptide: conversion of the RNA into an annealing-competent conformation. *Nucleic Acid. Res.* **2011,** 39, 4405-4418.

173. Frankel, A. D.; Pabo, C. O. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* **1988,** 55, 1189-93.

174. Chertova, E.; Chertov, O.; Coren, L. V.; Roser, J. D.; Trubey, C. M.; Bess, J. W.; Sowder, R. C.; Barsov, E.; Hood, B. L.; Fisher, R. J. Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J. Virol.* **2006,** 80, 9039-9052.

175. Kirpekar, F.; Krogh, T. N. RNA fragmentation studied in a matrix-assisted laser desorption/ionisation tandem quadrupole/orthogonal time-of-flight mass spectrometer. *Rapid Commun. Mass Sp.* **2001,** 15, 8-14.

176. Cerny, R. L.; Tomer, K. B.; Gross, M. L.; Grotjahn, L. Fast atom bombardment combined with tandem mass spectrometry for determining structures of small oligonucleotides. *Anal. Biochem.* **1987,** 165, 175-182.

177. Kellersberger, K. A.; Yu, E.; Kruppa, G. H.; Young, M. M.; Fabris, D. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Anal. Chem.* **2004,** 76, 2438-2445.

178. Tromp, J. M.; Schürch, S. Gas-phase dissociation of oligoribonucleotides and their analogs studied by electrospray ionization tandem mass spectrometry. *J. Am. Soc. Mass Spectr.* **2005,** 16, 1262-1268.

179. Schudoma, C.; May, P.; Nikiforova, V.; Walther, D. Sequence–structure relationships in RNA loops: establishing the basis for loop homology modeling. *Nucleic Acid. Res.* **2009,** 38, 970-980.

180. Johnson, P. E.; Turner, R. B.; Wu, Z. R.; Hairston, L.; Guo, J.; Levin, J. G.; Summers, M. F. A Mechanism for Plus-Strand Transfer Enhancement by the HIV-1 Nucleocapsid Protein during Reverse Transcription†. *Biochemistry* **2000,** 39, 9084-9091.

181. Takahashi, K. I.; Baba, S.; Chattopadhyay, P.; Koyanagi, Y.; Yamamoto, N.; Takaku, H.; Kawai, G. Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA* **2000,** 6, 96-102.

182. Mujeeb, A.; Ulyanov, N. B.; Georgantis, S.; Smirnov, I.; Chung, J.; Parslow, T. G.; James, T. L. Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: sequence effects and mechanism of RNA refolding. *Nucleic Acid. Res.* **2007,** 35, 2026-2034.

183. Kim, M.-J.; Kim, S. H.; Park, J. A.; Yu, K. L.; Jang, S. I.; Kim, B. S.; Lee, E. S.; You, J. C. Identification and characterization of a new type of inhibitor against the human immunodeficiency virus type-1 nucleocapsid protein. *Retrovirology* **2015,** 12, 90.

184. Morphy, R.; Kay, C.; Rankovic, Z. From magic bullets to designed multiple ligands. *Drug Discov. Today* **2004,** 9, 641-651.

185. Sleiman, D.; Goldschmidt, V.; Barraud, P.; Marquet, R.; Paillart, J.-C.; Tisné, C. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res.* **2012,** 169, 324-339.

186. Wu, T.; Guo, J.; Bess, J.; Henderson, L. E.; Levin, J. G. Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. *J. Virol.* **1999,** 73, 4794-4805.

187. Chan, B.; Weidemaier, K.; Yip, W.-T.; Barbara, P. F.; Musier-Forsyth, K. Intra-tRNA distance measurements for nucleocapsid proteindependent tRNA unwinding during priming of HIV reverse transcription. *P. Natl. Acad. Sci. USA* **1999,** 96, 459-464.

188. Seif, E.; Niu, M.; Kleiman, L. In virio SHAPE analysis of tRNA Lys3 annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. *Retrovirology* **2015,** 12, 40.

189. Ramalanjaona, N.; de Rocquigny, H.; Millet, A.; Ficheux, D.; Darlix, J.-L.; Mély, Y. Investigating the mechanism of the nucleocapsid protein chaperoning of the second strand transfer during HIV-1 DNA synthesis. *J. Mol. Biol.* **2007,** 374, 1041-1053.

190. Godet, J.; Ramalanjaona, N.; Sharma, K. K.; Richert, L.; de Rocquigny, H.; Darlix, J.-L.; Duportail, G.; Mély, Y. Specific implications of the HIV-1 nucleocapsid zinc fingers in the annealing of the primer binding site complementary sequences during the obligatory plus strand transfer. *Nucleic Acid. Res.* **2011,** 39, 6633-6645.

191. Vo, M.-N.; Barany, G.; Rouzina, I.; Musier-Forsyth, K. Mechanistic Studies of Mini-TAR RNA/DNA Annealing in the Absence and Presence of HIV-1 Nucleocapsid Protein. *J. Mol. Biol.* **2006,** 363, 244-261.

192. Godet, J.; de Rocquigny, H.; Raja, C.; Glasser, N.; Ficheux, D.; Darlix, J.-L.; Mély, Y. During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J. Mol. Biol.* **2006,** 356, 1180-1192.

193. Vo, M.-N.; Barany, G.; Rouzina, I.; Musier-Forsyth, K. HIV-1 Nucleocapsid Protein Switches the Pathway of Transactivation Response Element RNA/DNA Annealing from Loop–Loop "Kissing" to "Zipper". *J. Mol. Biol.* **2009,** 386, 789-801.

194. Lever, A.; Gottlinger, H.; Haseltine, W.; Sodroski, J. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* **1989,** 63, 4085-4087.

195. Laughrea, M.; Jette, L. A 19-Nucleotide Sequence Upstream of the 5' Major Splice Donor Is Part of the Dimerization Domain of Human Immunodeficiency Virus 1 Genomic RNA. *Biochemistry* **1994,** 33, 13464-13474.

196. Laughrea, M.; Jetté, L. Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248− 271 are dispensable for dimer formation. *Biochemistry* **1996,** 35, 1589-1598.

197. Muriaux, D.; Fossé, P.; Paoletti, J. A Kissing Complex Together with a Stable Dimer Is Involved in the HIV-1 Lai RNA Dimerization Process in Vitro. *Biochemistry* **1996,** 35, 5075-5082.

198. Feng, Y. X.; Copeland, T. D.; Henderson, L. E.; Gorelick, R. J.; Bosche, W. J.; Levin, J. G.; Rein, A. HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. *P. Natl. Acad. Sci. USA* **1996,** 93, 7577-7581.

199. Feng, Y.-X.; Campbell, S.; Harvin, D.; Ehresmann, B.; Ehresmann, C.; Rein, A. The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J. Virol.* **1999,** 73, 4251-4256.

200. Russell, R.; Liang, C.; Wainberg, M. Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* **2004,** 1, 23.

201. Ennifar, E.; Walter, P.; Ehresmann, B.; Ehresmann, C.; Dumas, P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Mol. Biol.* **2001,** 8, 1064.

202. Mujeeb, A.; Clever, J. L.; Billeci, T. M.; James, T. L.; Parslow, T. G. Structure of the dimer a initiation complex of HIV-1 genomic RNA. *Nat. Struct. Mol. Biol.* **1998,** 5, 432-436.

203. Dardel, F.; Marquet, R.; Ehresmann, C.; Ehresmann, B.; Blanquet, S. Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res.* **1998,** 26, 3567-3571.

204. Lawrence, D. C.; Stover, C. C.; Noznitsky, J.; Wu, Z.; Summers, M. F. Structure of the intact stem and bulge of HIV-1 Ψ-RNA stem-loop SL1. *J. Mol. Biol.* **2003,** 326, 529-542.

205. Ennifar, E.; Yusupov, M.; Walter, P.; Marquet, R.; Ehresmann, B.; Ehresmann, C.; Dumas, P. The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure* **1999,** 7, 1439-1449.

206. Theilleux-Delalande, V.; Girard, F.; Huynh-Dinh, T.; Lancelot, G.; Paoletti, J. The HIV-1Lai RNA dimerization. *Eur. J. Biochem.* **2000,** 267, 2711-2719.

207. Mihailescu, M.-R.; Marino, J. P. A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex. *P. Natl. Acad. Sci. USA* **2004,** 101, 1189-1194.

208. Bernacchi, S.; Ennifar, E.; Tóth, K.; Walter, P.; Langowski, J.; Dumas, P. Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site. *J. Mol. Biol.* **2005,** 280, 40112-40121.

209. Rist, M. J.; Marino, J. P. Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* **2002,** 41, 14762-14770.

210. Aduri, R.; Briggs, K. T.; Gorelick, R. J.; Marino, J. P. Molecular determinants of HIV-1 NCp7 chaperone activity in maturation of the HIV-1 dimerization initiation site. *Nucleic Acid. Res.* **2013,** 41, 2565-2580.

211. Clever, J. L.; Parslow, T. G. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **1997,** 71, 3407-3414.

212. Laughrea, M.; Jetté, L.; Mak, J.; Kleiman, L.; Liang, C.; Wainberg, M. A. Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J. Virol.* **1997,** 71, 3397-3406.

213. Abbink, T. E. M.; Ooms, M.; Haasnoot, P. C. J.; Berkhout, B. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry* **2005,** 44, 9058-9066.

214. van Bel, N.; Das, A. T.; Cornelissen, M.; Abbink, T. E. M.; Berkhout, B. A short sequence motif in the 5′ leader of the HIV-1 genome modulates extended RNA dimer formation and virus replication. *J. Biol. Chem.* **2014,** 289, 35061-35074.

215. Moore, M. D.; Hu, W.-S. HIV-1 RNA dimerization: It takes two to tango. *AIDS reviews* **2009,** 11, 91.

216. Schwergold, C.; Depecker, G.; Di Giorgio, C.; Patino, N.; Jossinet, F.; Ehresmann, B.; Terreux, R.; Cabrol-Bass, D.; Condom, R. Cyclic PNA hexamer-based compound: modelling, synthesis and inhibition of the HIV-1 RNA dimerization process. *Tetrahedron* **2002,** 58, 5675-5687.

217. Elmén, J.; Zhang, H.-Y.; Zuber, B.; Ljungberg, K.; Wahren, B.; Wahlestedt, C.; Liang, Z. Locked nucleic acid containing antisense oligonucleotides enhance inhibition of HIV-1 genome dimerization and inhibit virus replication. *FEBS letters* **2004,** 578, 285-290.

218. Jakobsen, M. R.; Haasnoot, J.; Wengel, J.; Berkhout, B.; Kjems, J. Efficient inhibition of HIV-1 expression by LNA modified antisense oligonucleotides and DNAzymes targeted to functionally selected binding sites. *Retrovirology* **2007,** 4, 29.