Characterization and targeting the dynamic recognition of Trp37-G during the interaction of NCp7 protein of HIV-1 with nucleic acids

Rajhans Sharma

THÈSE

présentée par :

Rajhans SHARMA

soutenue le : **10th Avril 2018**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Sciences du vivant/Biophysique

Caractérisation et ciblage de la reconnaissance dynamique de Trp37-G lors de l'interaction de la protéine NCp7 de HIV-1 avec des acides nucléiques

THÈSE dirigée par :

  M. MELY Yves   Professeur, Université de Strasbourg

RAPPORTEURS :

  Mme. MAHUTEAU-BETZER Florence   Chargée de recherche, Institut Curie,Paris

  Mme. GATTO Barbara   Associate Professor, University of Padova

AUTRES MEMBRES DU JURY :

  M. LEONARD Jérémie   Chargé de recherche, Universitié de Strasbourg

UNIVERSITY OF STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

UMR CNRS 7021 Laboratory of Bioimaging and Pathologies

THESIS

presented by :

Rajhans SHARMA

Defense on : **10th April, 2018**

For obtaining the degree of : **Doctorate of University of Strasbourg**

Discipline/ Specialization : Life sciences/Biophysics

Characterization and targeting the dynamic recognition of Trp37-G during the interaction of NCp7 protein of HIV-1 with nucleic acids

THESIS directed by :

   Mr. MELY Yves   Professor, University of Strasbourg

REPORTERS :

   Mrs. MAHUTEAU-BETZER Florence   Chargée de recherche, Institut Curie, Paris

   Mrs. GATTO Barbara   Associate Professor, University of Padova

OTHER MEMBERS OF JURY :

   Mr. LEONARD Jérémie   Chargé de recherche, University of Strasbourg

Table of Contents

Acknowledgement

I am immensely grateful to Prof. Yves Mély for giving me this opportunity to pursue the PhD research work under his supervision. Thank you for your guidance, discussions, sharing knowledge, pushing me to higher limits, improving my skills, encouragements and help me grow as a researcher. I admire your positivity, dynamism, sense of humor and winning attitude. Your 'perfectionism' has always inspired me to excel my inputs.

I am grateful to all the members of jury, Dr. Florence Mahuteau-Betzer, Dr. Barbara Gatto and Dr. Jerémié Leonard for accepting the proposal to evaluate this thesis work. I am grateful for your comments and suggestions.

I would like to thank our collaborators Prof. Yitzhak Tor, Dr. Mattia Mori and Dr. Roberto Importa for providing their scientific inputs. I would like to thank Prof. Alain Burger for careful reading of the thesis and providing important suggestions.

I am thankful to the permanent members of our lab for their suggestions during this work: Dr. Julien Godet for scientific discussions and designing the project; Dr. Pascal Didier for helping me solve complex mathematical calculations and playing football with us; Dr. Nicolas Humbert for teaching peptide purification (*on papurus*) and rewarding with *choco pain*; Dr. Ludovic Richert for giving me experimental guidance that made the long nights working with laser and *Brochon* more bearable; Dr. Christian Boudier for his valuable insight on ITC and Stopped flow results and additional thoughts on Game of thrones and Himalayas; Dr. Frédéric Przybilla for his suggestions and motivation, Dr. Real Elenore for teaching biological experiments; and Dr. Andrey Klymchenko and Dr. Youri Arntz for their excellent company and sharing never ending hilarious stories. Special thanks to the grand-father figure of our lab, Dr. Guy Duportail, who took care of the French administrative paper work and conundrums, and his huge support during the time when I underwent a surgery. I am thankful to Marlyse Wernert, Ingrid Barthel, Ludovic Four and Michel Morciano for their administrative and IT assistance.

I appreciate contribution from my PhD and Post-doc colleagues. I am hugely thankful to Dr. Marianna Sholokh, the first person who taught me experimental techniques, methods and critical analysis of results. I am thankful to the queen of Thinpad project and member of our 'Thinpad patient team', Dr. Lesia Kovalenko, for the advices, discussions and definitely the *cakes*. I am also grateful to Dr. Kamal kant Sharma, Dr. Avisek Ghose and Dr. Krishna Gavvala for the scientific discussions and excursions.

I appreciate the time spent with the members of our group making memories: Redouane (many thanks for the help), Liliyana (the best neighbor), Bogdan, Yosuke, Faisal, Katya, Alex, Manu, Jagganath, Caterina, Dima and others. As the leader of ''4 o'clock tea time'' group, I acknowledge contribution from its members: Sasha, for a wonderful friendship that certainly made life lot easier during these days and Taras for enriching our discussions with explicit contents. I would like to thank Nina for all the laughs, loads of cookies *(free food)* and particularly helping me in recover my health. Specially, I would like to thank Waseem and Tanveer for their true friendship. Back home in India, I am thankful to my friends Nachiketa, Himanshu and Aakriti for their kind support.

I will always be grateful to my parents. They have been a constant source of motivation and happiness in my life.

Thank you

Rajhans Sharma.

List of Abbreviations

A

AIDS: Acquired immunodeficiency syndrome
ALIX: ALG2-interacting protein X
ART: Antiretroviral therapy
d2Ap7(–)/T$_{12}$(+)PBS: duplex with 2Ap-T match

C

CA or p24: Capsid protein
CTD: C'-terminal domain

E

Env: Envelope proteins
ESCRT: Endosomal sorting complex required for transport
ESIPT: Excited state intramolecular proton transfer
λ_{ex}: Excitation wavelengths

G

G: Guanine
gRNA: genomic RNA

H

HAART: Highly Active Anti-Retroviral Therapy
3HC: 3-hydroxychromones
3HCnt: 2-thienyl-3-hydroxychromone

I

IN: Integrase
ITC: Isothermal titration calorimetry
IC: Internal conversion
ICS: Inter system crossing
IRES: Internal ribosome entry segment

K

K_d: Binding constant

M

MA or p17: Matrix protein
MD: Molecular dynamics
MM-PBSA: Molecular Mechanics Poisson-Boltzmann Surface Area

N

NA: Nucleic acids
NC: Nucleocapsid protein
nt: nucleotide
NC-ODN: NC-Oligonucleotide
NTD: N'-terminal domain
Nef: Negative regulatory factor

NRTIs/NtRTIs: Nucleoside/nucleotide reverse transcriptase inhibitors
NNRTIs: Non–nucleoside reverse transcriptase inhibitors

P

PBS: Primer binding site
(–)PBS: minus strand Primer binding site
Δ(–)PBS: truncated (–)PBS without the overhangs
(–)/(+)PBS or (+)/(–)PBS: Duplex PBS
PPT: Poly purine Tract
PPTC: Poly purine Tract central
PR: Protease
PIC: pre-integration complex
Phe or F: Phenylamine

Q

QY or Φ: Quantum yield

R

RT: Reverse transcriptase
RRE: Rev response element
Rev: Regulator of virion Expression
RTC: Reverse transcription complex

S

SU or gp120: Surface protein
SL: Stem loop
SNP: Single nucleotide polymorphism
s: 2-amino-6-(2-thienyl)purine

T

Trp or W: Tryptophan
TM or gp41: Transmembrane protein
TAR: Transactivator region
TSG101: Tumor susceptibility gene 101
thG: thienoguanosine
dthG: deoxy-thG as substituted in DNA (or mentioned with "underscore"DNA, like thG7(–)PBS_DNA)
tRNA: Transfer RNA

U

UTR: Untranslated region

V

Vpr: Viral protein R
Vif: Viral infectivity factor
Vpu: Viral protein U
VPS4: Vacuolar protein sorting 4

Z

ZF: Zinc finger

1. Bibliographic Review

1.1. Human Immunodeficiency Virus Type 1 (HIV-1)

1.1.1. Overview

Human immunodeficiency virus (HIV) causes Acquired Immunodeficiency Syndrome (AIDS) that impairs the individual's immune system during the lifetime. Since the beginning of the epidemic in early 1980's, more than 70 million people have been infected, and a toll of 35 million people have died with HIV, that currently represents the fourth leading cause of death worldwide and has been projected to become the third leading cause by 2030 (UNAIDS and WHO). The rising epidemics led to a large influx of attention and funding among the scientific and medical community. For example, the US federal agency has spent an all-time maximum of 25 billion dollars only in 2015, 19 billion dollars in 2016 and an estimated 26 billion will be required in response to HIV in 2020.

1.1.2. Origin Timeline

The causative effect of HIV in humans was first detected in United States of America in early 1980's. The patients diagnosed showed infections and rare malignancies. In 1983, Institut Pasteur, Paris became the first laboratory to isolate the lymphoadenopathy associated virus (LAV) (then termed name for HIV) from a culture derived from a lymph node biopsy sample of a patient with generalized lymphadenopathy (Barre-Sinoussi et al., 1983). Later in 2008, the Nobel Prize in Physiology and Medicine was awarded to Françoise Barré-Sinoussi and Luc Montagnier for this discovery. Their research led to the first diagnostic ELISA test for HIV infections that FDA allowed to sell commercially in 1985. Later, CD4 cell surface molecule was identified as the main receptors, and CXC- chemokine receptor 4 (CXCR4) and CC-chemokine receptor 5 (CCR5) as co-receptors for HIV. By the end of twentieth century, researchers gained insight into the life cycle of HIV, identifying targets of antiretroviral drugs. In 1989, AZT became the first commercially available drug that prevents HIV transfer from infected mother to child. Within next decade, combination antiretroviral therapy appeared and in 2009, doctors cured a HIV-infected patient from Berlin. However, the situation worsened with the discovery of the drug-resistant property of HIV virus that notably arises because of the error prone nature of its reverse transcription and the recombination occurring between the viruses.

1.1.3. Classification and Transmission

HIV is classified as a lentivirus that belongs to the family of the retroviruses. In general, retroviruses are spherical particles, enveloping a RNA genome, that infect the vertebrates. HIV

structure is similar to that of the Simian Immunodeficiency Virus (SIV), another lentivirus. Lentiviruses have the ability to infect dividing as well as non-dividing cells, ultimately causing neurological and immunological disorders.

Two types of HIV viruses, HIV-1 and HIV-2 were identified (Clavel et al., 1986; Guyader et al., 1987). They have similar gene organization, but their genome differs by about 55%. HIV-1 is more diverse, infectious and transmissible than HIV-2. They differ also geographically as HIV-1, which is widely spread, is mostly reported in North America, central Africa, Europe and Asia. HIV-2 is mainly localized in west Africa. HIV-1 and HIV-2 resemble to the strains of SIV in chimpanzees and sooty mangabeys, respectively. Hence, the virus might have cross transitioned during some point from primates to humans. HIV-2 infected areas are mostly being replaced by HIV-1 because the viral loads of HIV-2 is lesser than HIV-1, with a lower transmission rate and a nearly complete absence of mother to child transmission.

HIV-1 stains are further divided into four lineages, M (major), N (non M or non O), O (outliner) and P (new type) (Tebit and Arts, 2011). The group M is widely spread across the world and thus is largely studied. The other three groups represent less than 1 percent. Group M is further branched into nine subtypes (A-D, F-H, J, K) and more than 40 Circulating Recombinant Forms (CRFs) that were generated when multiple subtypes infected the same populations. HIV-2 exists in only eight distinct lineages, (A-H) in which only A and B have been identified in humans (Sharp and Hahn, 2011).

1.1.4. Structural Organization

HIV-1 forms a spherical, membrane enveloped virion, 100–150 nm in diameter, which contains two copies of single-stranded positive sense RNA genome (Ganser-Pornillos et al., 2012). The mature HIV-1 particle has several spikes on its lipid bilayer membrane that are composed of Envelope proteins (Env) (Checkley et al., 2011), which are trimers of Surface protein (SU, gp120) and Transmembrane protein (TM, gp41) . A canonically structured capsid membrane is formed of approximately 1500–2000 copies of Capsid protein (CA, p24), with an outer coating of 2000 copies of Matrix protein (MA, p17) separating it from the envelope membrane (Niedrig et al., 1994). The capsid membrane serves as a core that encloses two copies of unspliced viral RNA genome, coated with 2000 copies of Nucleocapsid protein (NCp7), 10–12 copies of $tRNA_3^{Lys}$ (Isel et al., 2010), as well as with Protease (PR), Reverse transcriptase (RT) and Integrase (IN). The immature virus is characterized by a thick layer of uncleaved Gag polyprotein that lies within the virion membrane. After proteolytic processing

of the Gag polyproteins, the specific structures of the mature viral particles are generated (Ganser-Pornillos et al., 2008).
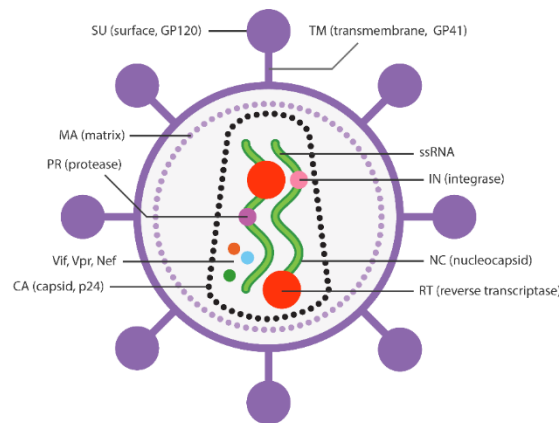
Figure 1. Schematic representation of HIV-1 virion.

1.1.5. Genomic Organization

The *viral genome* of HIV-1 is single stranded positive sense RNA of 9,200 nucleotides long that exists in a dimer form within mature viral particles. The genomic RNA (gRNA) has 100–200 adenine residues at its 3' end while at the 5' end it has capping of trimethylguanosine (TMG) (Yedavalli and Jeang, 2010). The gRNA is poorly structured with several folding domains or motifs. These motifs form serval mismatches, bulges and hairpins. The gRNA has nine Open Reading Frames (ORFs) (Figure 2): *gag, pol, env, tat, rev, vpr*, *vif*, *vpu* and *nef*. The *gag, pol, env* encodes for three polyproteins that are Gag, Pol and Env; *tat* and *rev* encode Tat and Rev regulatory proteins; and *vpr*, *vif*, *vpu*, *nef* encode Vpr, Vif, Vpu and Nef accessory proteins, respectively. The proteolytic cleavage of Gag polyprotein further give rise to MA (Matrix protein), CA (Capsid protein), NC (Nucleocapsid protein), and p6; while Env proteins gives SU (Surface protein, gp120) and TM (Transmembrane protein, gp41), all these proteins are considered as structural components of HIV-1 particle. The three Pol proteins: PR (Protease), RT (Reverse Transcriptase), and IN (Integrase), provide essential enzymatic functions and encapsulated within the particle. HIV-1 encodes six additional proteins, often called accessory proteins, three of which (Vif, Vpr, and Nef) found in the viral particle. Two other accessory proteins, Tat and Rev, provide essential gene regulatory functions, and the last protein, Vpu, indirectly assists in assembly of the virion. The gRNA is divided into two regions, the coding region that codes for proteins and the non-coding region that plays crucial role in structural organization.
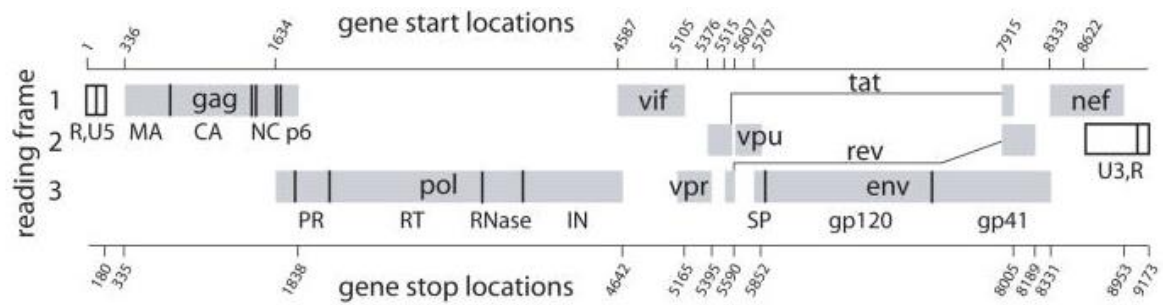
Figure 2. Genomic organization of HIV-1. The open reading frames are shown in shaded rectangles and the black lines corresponds to the connections between domain in polyprotein. Adapted from (Watts et al., 2009).

The **Untranslated Region (UTR)** of retrovirus gRNA is present on both 5' and 3' end. It comprises a number of folded sequences that play an important role in dimerization, genome packaging, and translation (Abbink et al., 2005a; Boeras et al., 2017; Wilkinson et al., 2008). The 5'UTR exists in two conformations that are in equilibrium: one that promotes the translation of Gag gene by Internal Ribosome Entry Segment (IRES) mechanism (Brasey et al., 2003); and second is dimmer promoting conformation (Abbink et al., 2005b; Keane and Summers, 2016).
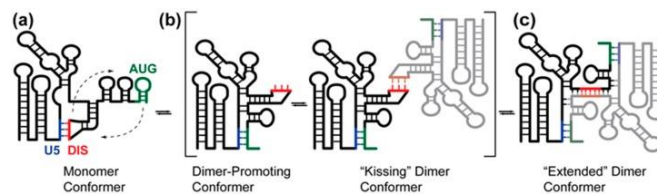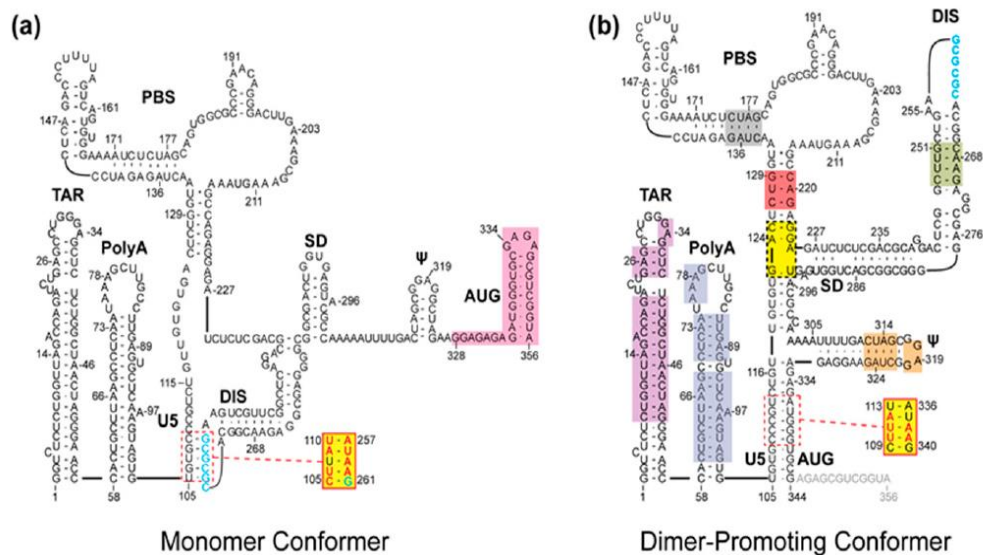
Figure 3. Secondary structure of HIV-1 5′-UTR in monomer (a) and dimer-promoting (b) conformations (Top section). The schematic diagram showing the formation of dimer conformer from monomer conformer (Below section). Adapted from (Keane et al., 2015).

The UTR is composed of:

A. The R region

The R region (repeat) is a 98 nucleotides region present at both ends of the HIV-1 gRNA. It is further sub-divided into two regions.

- Transactivator Region (**TAR**): This region consists of nucleotides from +1 to +59 in the HIV-1 UTR. Having a highly structured hairpin form, TAR is required for viral expression. The structural integrity of this region is critical for TAT activation (Heinicke et al., 2009; Wang et al., 1993). Moreover, TAR serves as a binding site for several cellular binding proteins, that regulate the basal expression of HIV-1 UTR (Henriet et al., 2005; Wang et al., 1993).

- **poly(A)**: The poly(A) signal contains a polyadenylation signal, the cleavage site and part of the GU-rich downstream element (Edmonds, 2002; Klasens et al., 1999; Sükösd et al., 2015)

B. U5 Region

The U5 region has 83 nucleotides and represents the first part of the gRNA to be retro-transcribed. This region also contains the 18-nucleotide primer binding site (**PBS**) region which is present at the 3' end of the U5 region. The role of PBS is crucial as it anneals to tRNA^{Lys3}, initiating reverse transcription (Sleiman et al., 2012).

C. ψ-site

The psi-site (ψ-site) is a 120 nt region, present between the PBS and the start codon. It has four stem loop motifs (SL1–SL4) that plays a crucial role in genome recognition and gRNA dimerization. Selective packaging of HIV-1 gRNA requires all four cis-elements (Lever et al., 1989; Lever, 2007; Pak et al., 2017).

- **SL1**: The structure of SL1 consists of an upper stem that has GC-rich loop and a lower stem having an intervening bulge which is highly conserved and required for packaging (Chung et al., 2010; Laughrea and Jetté, 1994) (Figure 4a). The eleven-nucleotide upper stem has a six-nucleotide palindromic sequence, usually GCGCGC (*e.g.*, in the Lai

isolate) and GUGCAC (*e.g.*, in the Mal isolate) that facilitates dimerization among two gRNAs (Lodmell et al., 2000). In addition, the portion having a GAAG bulge and the four nucleotide stem are mostly conserved among different HIV isolates (Mujeeb et al., 2007a). The function of SL1 is to ensure effective packaging by promoting dimer formation, through "kissing complex" interaction (Kieken et al., 2006), that facilitates the packaging of two copies of genome during virus assembly (Berkowitz et al., 1996; Lawrence et al., 2003; Lu et al., 2011) (Figure 4b,c). The importance of SL1 motif towards assuring effective packaging is assessed by introduction of mutations that resulted in a large decrease in packaging efficiency (Amarasinghe et al., 2000; McBride and Panganiban, 1997).
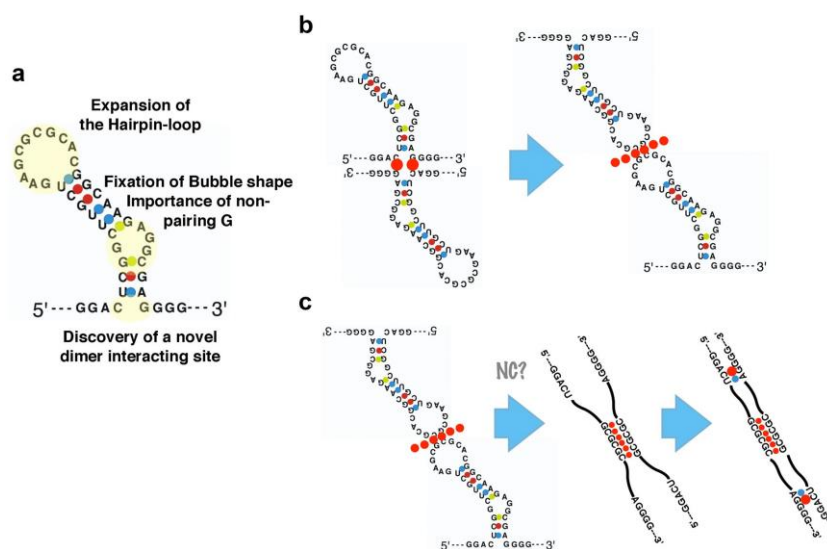
Figure 4. Stem loop structure of SL1(a). Proposed model for dimerization b) "pre-initiation model", where the bases at the end of the stem mediates the primary interaction of DLS dimer formation and after the two molecules are sufficiently proximal, DIS interaction occurs. c) "Finalization model", after interaction with DIS, two DLSs start to "melt" together by the viral nucleocapsid protein to form an extended duplex (Sakuragi et al., 2016)

- **SL2:** The SL2 region is a 19-nucleotide sequence that exists in the form of a hairpin (Amarasinghe et al., 2000). It has a tetraloop and a long stem with a single nucleotide bulge. SL2 loop contains the Splice Donor (SD) sequence that is required for the spliced mRNAs production. (Bazzi et al., 2012; Cruceanu et al., 2006b)**.** SL2 plays an important role in genome packaging (McBride and Panganiban, 1997), by interacting with the NC domain of Gag (Amarasinghe et al., 2000; Belfetmi et al., 2016).

- **SL3** is a 20-nucleotide hairpin having a tetraloop and a perfect double helix stem (Pappalardo et al., 1998). The loop of SL3 belongs to the A-family and is quite flexible. The structure of SL3 is highly conserved among the different strains of HIV-1 and its interaction with the NC domain of Gag during virus assembly is crucial for gRNA selection (Abd El-Wahab et al., 2014; Lever, 2007; Tanwar et al., 2017). Moreover, SL3 by itself is sufficient to drive the gRNA selection (Abd El-Wahab et al., 2014; Athavale et al., 2010; Belfetmi et al., 2016; Paoletti et al., 2002). Detailed discussion on SL3 is presented in section 4.3 of this thesis.

- **SL4 or AUG:** SL4 hairpin contains a stem with G-U pairs and a GAGA tetraloop (Kerwood et al., 2001). The SL4 tetraloop adopts a conformation similar to a classical GNRA form (Jucker et al., 1996) and the stem is quite unstable due to G-U pairs. The SL4 tetraloop participates in long range RNA-RNA interaction with the U5 element that connects polyA and PBS stem-loop (Abbink and Berkhout, 2003; Damgaard et al., 2004; Jucker et al., 1996; Lu et al., 2011; Watts et al., 2009).

D. U3 region

The U3 region at the 3' end of gRNA is required for the regulation of transcription of the proviral DNA into viral RNA by the transcription machinery of the host cell (Calzado et al., 2004; Gaynor, 1992; Sgarbanti et al., 2002). The U3 region includes:

- The **PPT** (Poly Purine Tract) and **PPTC** (Poly Purine Tract Central): They are the purine rich domains of the gRNA that are resistant to RNase H activity. The PPT is present upstream of U3 sequence of 3'UTR of proviral genome, whereas the PPTC sequence is present in the ORF of the Pol gene (Charneau et al., 1992; Pollom et al., 2013; Rausch and Le Grice, 2004). Their resistance towards RNase H activity of reverse transcription is instrumental because these sequences serve as primers for the synthesis of the plus-strand DNA.

- Rev Response Element (**RRE**): it is a ~ 350 nucleotide-long sequence that contains several stem-loops and bulges (Watts et al., 2009). It binds with Rev protein to overcome the intron surveillance mechanism and facilitate the nucleo-cytoplasmic export of the mRNAs corresponding to the viral genome (Bartel et al., 1991; Rausch and le Grice, 2015; Sherpa et al., 2015).

1.1.6. Viral Proteins

The viral proteins are encoded from nine ORF that are in the coding sequence of gRNA.

Gag gene encodes for 55 kDa *Gag* polyprotein precursor (Pr55^{Gag}) that upon cleavage by protease (PR, p12) generates the MA, CA, NC and p6 protein.

Pol gene together with *Gag* gene code for a 160 kDa GagPol polyprotein precursor (Pr160^{GagPol}) that upon cleavage by PR results in all the three HIV-1 enzymes, namely PR (cleaved by auto-processing), RT and IN. The production of GagPol arises due to a -1 frame shift in the translation frame which occurs approximately 5% of the time. (Karn and Stoltzfus, 2012)

Env gene encodes for a 160 kDa Env polyprotein precursor (gp160) that upon cleavage by a cellular furin-type protease in the Golgi, results in the synthesis of gp120 and gp41 glycoproteins.

In addition, the HIV-1 genome also encodes for regulatory proteins (Tat, Rev) and accessory (Vpr, Vif, Vpu, Nef) proteins. Figure 5 represents the 3D structures of functional domain of all the HIV-1 proteins.
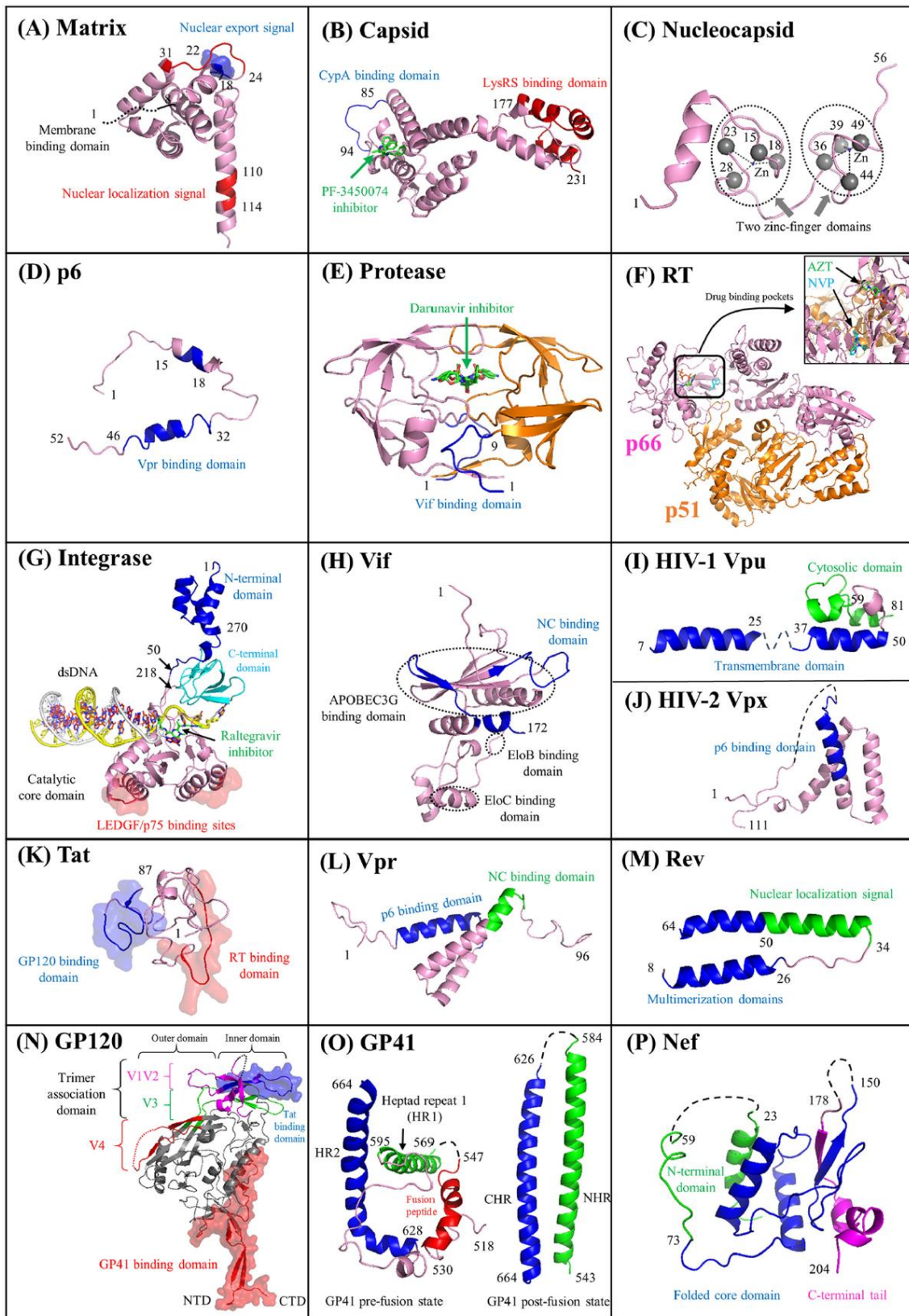
Figure 5. Functional domains of HIV proteins. Cartoon representations of HIV proteins (matrix, capsid, nucleocapsid, p6, protease, RT, integrase, Vif, Vpu, Vpx, Tat, Vpr, Rev, GP120, GP41, and Nef). In each panel, the protein domains involved with HIV

pairwise protein interactions are marked accordingly. Surface representations indicate protein interaction interfaces. The five small molecules shown in green correspond to protein inhibitors such as the capsid inhibitor PF-3450074 (503) (B), the protease inhibitor darunavir (E), the nucleoside analogue reverse transcriptase inhibitor zidovudine (AZT) (F), the nonnucleoside analogue reverse transcriptase inhibitor nevirapine (NVP) (F), and the integrase inhibitor raltegravir (G). Adapted from (Li and De Clercq, 2016)

Based on the structural and functional characteristics, the HIV-1 viral protein are classified in following sections:

A. Structural Proteins

- **Matrix Protein (MA)**: The 17 kDa and 132 amino acid MA protein consists of 5 α-helices; two short α-helical stretches; and a three-stranded mixed β-sheet (Verli et al., 2007; Zeinolabediny et al., 2017). Its N-terminal myristate group and basic residues located within the first 50 amino acids are involved in membrane targeting residues (Alfadhli and Barklis, 2014; Tedbury and Freed, 2015). The MA proteins assemble into trimers that are required for viral assembly. Other than targeting the Gag and GagPol to the plasma membrane, MA also accommodates the cytoplasmic tail of the full length Env glycoprotein and incorporate it into the viral particles (Frankel and Young, 1998; Ganser-Pornillos et al., 2012)

- **Capsid Protein (CA)**: It consists of two domains: N'-terminal domain NTD and C'-terminal domain CTD, that are linked together via a flexible linker. The 3D structure of NTD shows seven α-helices while the CTD has four α-helices (Gamble et al., 1997; Gitti et al., 1996; Pornillos et al., 2011). The CTD domain promote efficient viral assembly through CA dimerization and Gag oligomerization (Gamble et al., 1997), while the NTD domain participate in viral uncoating, through association with cyclophilin A (Luban, 2012). Moreover, CA protects the gRNA from cellular immune factors and ensures that RT is close enough to gRNA, in order to perform the reverse transcription (Campbell and Hope, 2015; Freed, 1998).

- **Nucleocapsid Protein (NC)**: NC is a 7 kDa, 55 amino acid protein that has two highly conserved CCHC zinc finger motifs which are connected by a basic linker (RAPRKKG). The NC protein exerts many functions throughout the virus life cycle, mainly through its

nucleic acid chaperone activity (Darlix et al., 2007, 2014, 2011). The properties and functions of NC are discussed in detail in section 1.2.

- **p6:** It is a 6 kDa, 52 amino acid protein that has two helices connected by a flexible linker. One of the role of p6 is to bind and recruit Vpr (Votteler and Sundquist, 2013). Moreover the two late domains of p6 are responsible for recruitment of the cellular ESCRT (endosomal sorting complex required for transport) machinery that enables viral budding (Freed, 1998; Votteler and Sundquist, 2013).

B. Enzymatic Proteins

- **Protease (PR):** PR is a 6 kDa, 99 amino acid protein that exists in a homodimer form. The N- and C- terminal domains are arranged in a four-stranded antiparallel β-sheet, one loop and one helix (Navia et al., 1989). The catalysis site is highly conserved and comprises of amino acids Asp25-Thr26-Gly27 (Wlodawer et al., 1989). Protease is generated due to self-catalysis from $Pr160^{Gag-pol}$ and is responsible for the cleavage of Gag and GagPol polyprotein. This cleavage of Gag and GagPol by PR is necessary to form mature viral particle (Sundquist and Kräusslich, 2012).

- **Reverse Transcriptase (RT):** RT is a heterodimer of two subunits p66 and p51. The p66 subunit is a 66 kDa, 560 amino acid that has a catalytic function while the p51 domain is a 51 kDa, 440 amino acid domain that has structural function (Esnouf et al., 1995; Hsiou et al., 1996; Rodgers et al., 1995). The RNase H domain of RT cleaves the RNA portion of the RNA-DNA hybrids generated during reverse transcription (Das and Arnold, 2013; Hu and Hughes, 2012). RT also possess RNA-dependent and DNA-dependent polymerase activities and actively catalyze the reverse transcription process by synthesizing double strand DNA starting from a single stranded gRNA.

- **Integrase (IN):** It is a 32 kDa, 288 amino acid protein that has three domains: the N-terminal domain (NTD), the catalytic core domain and the C-terminal domain (CTD) (Cai et al., 1997; Dyda et al., 1994; Hare et al., 2010; Lodi et al., 1995). The role of IN is to insert the proviral DNA into the host cell DNA, by performing 3' end processing activity (cleavage of two nucleotides from the two 3' ends of the viral DNA) and strand transfer activity (Lesbats et al., 2016; Lusic and Siliciano, 2017).

C. Envelope Proteins

- **Surface Glycoproteins (SU or gp120):** It is a 515 amino acid highly glycosylated protein that is located at the surface of the viral particle. Structurally it has ten segments, five of them are conserved domains (C1–C5) and the rest are variable loops (V1–V5). It has three structural units, the inner unit that has two α-helices, five-stranded β-sandwichs and loops; the outer domain that has stacked barrel; and the bridging domain that has four-stranded β-sheets (Kwong et al., 1998; Pancera et al., 2010). The role of gp120 during the viral infection is to interact with the CD4 receptors of the target cell (Checkley et al., 2011; Haqqani and Tilton, 2013).

- **Transmembrane Protein (TM or gp41):** It is a 345-amino acid glycosylated protein that exists in a trimer form. It has a N-terminal ectodomain, a transmembrane domain, and a C-terminal cytoplasmic tail that orient themselves in antiparallel helices to form a trimer (Chan et al., 1997; Weissenhorn et al., 1997). The gp41 is involved in the fusion of virion membrane and plasma membrane during viral entry (Haqqani and Tilton, 2013).

D. Regulatory Proteins

- **Trans activator of Transcription (Tat):** It is a 9-11 kDa protein that exist either as an 86 amino acid or a 101 amino acid form. It has six regions: region I between 1–21 amino acid, region II between 22–37 which is Cys rich: region III between 38–48 amino acid which forms the core of the protein, region IV between 49–59 amino acids that is mostly occupied by basic amino acids, region V between 60–72 amino acids which is Gln rich region and region VI is between 73–101 amino acid which forms the C'-terminal of the Tat (Bayer et al., 1995; Peloponese Jr. et al., 2000). The main role of Tat is to promote efficient transcription of the viral genome (Johri et al., 2011).

- **Regulator of Virion Expression (Rev):** Rev is a 13 kDa, 116 amino acid protein with two domains, the N'-terminal domain having two α-helices and a loop, and the disordered C'-terminal domain. The NTD has nuclear localization signal, RNA binding domain and Rev multimerization domain. On the other hand the CTD domain has a Leu-rich motif needed for nuclear export (Daugherty et al., 2010; DiMattia et al., 2010). Rev facilitates the nuclear export of single-spliced and unspliced RNAs by binding to their RRE element (Pollard and Malim, 1998; Sherpa et al., 2015).

E. Accessory Proteins

- **Viral Protein R (Vpr):** It is a 14 kDa, 96 amino acid protein having three α-helices and flexible N'- and C'- terminal domains which tend to multimerize (Morellet et al., 2003). The N'- terminal domain of Vpr take part in G2 phase cell cycle arrest and apoptosis (Guenzel et al., 2014). While the C'-terminal domain functions in virion incorporation, nuclear localization and Vpr oligomerization (Freed, 1998; Ganser-Pornillos et al., 2012).

- **Viral Infectivity Factor (Vif):** It is a 23 kDa, 96 amino acid protein that has three domains: a highly ordered N'-terminal domain, the central domain having Zn fingers motifs, and the disordered C'-terminal domain (Barraud et al., 2008). Vif play protective role against a cellular deaminase APOBEC3 (A3G and A3F). A3G halts viral progression as it mutates C to U residues during reverse transcription. Moreover, Vif also show nucleic acid chaperoning activity during virion assembly (Ennifar et al., 2007; J G Levin et al., 2010; Sleiman et al., 2014).

- **Viral Protein U (Vpu):** It is a 16 kDa, 81 amino acid protein that bind to internal membrane of the cell. It has two domains, N'-terminal domain which span through the membrane and C'-terminal domain which is present inside the cytoplasm (Zhang et al., 2015). The role of Vpu is to degrade CD4 receptors complexed with Env protein. Vpu also promote budding of the virus particle from the infected cell (Gonzalez, 2015).

- **Negative Regulatory Factor (Nef):** It is a 27–35 kDa, 206 amino acid protein with three domains: the N'-terminal domain, the core and the C'-terminal domain (Arold et al., 1997; Grzesiek et al., 1997). The role of Nef is to degrade the cell receptors, like CD4, CD28 and CxCR4 receptors. This in turn promote incorporation of Env protein into virion. It is also responsible for degradation of histocompatibility complex I and II on antigen presenting cells, and in recognition of infected cells by the immune system (Pereira and daSilva, 2016).

1.1.7. Replication Cycle of HIV-1

The life cycle of the HIV virus can be divided in two major phases, the early phase and the late phase.

1.1.7.1. Early Phase

Infection of the cell by HIV-1 particle start with protein-protein interaction on their surface (Figure 6). The CD4 receptors of the cell **binds** to HIV-1 gp120 surface glycoproteins. This interaction induce conformational change in CD4 receptors, which in turn promote the

interaction with the co-receptors i.e. chemokine receptors 5 (CCR5) chemokine receptor 4 (CXCR4) present on the surface of the target cell (Dean et al., 1996; Feng et al., 1999). Then, the conformational change in gp41 further promote its hydrophobic N'-terminal domain to cross into the plasma membrane and reach the cytoplasm of the cell target (Archin et al., 2014; Dr. Rainer Seitz, 2016). This step marks the completion of **fusion** between the cell membrane and viral envelope.

Figure 6. Early phase of the HIV-1 viral life cycle. Env, envelope glycoprotein; MA, matrix protein; CA, capsid protein; NC, nucleocapsid protein; RTC, reverse transcription complex; NPC nuclear pore complex; PIC, pre-integration complex. Adapted from (Campbell and Hope, 2015).

Fusion of the cell and virion membranes results in the translocation of the viral capsid inside the cytoplasm. ***Reverse transcription*** begins in cytoplasm and convert single strand gRNA to double strand DNA. The steps of reverse transcription are discussed in section 1.2.3. During reverse transcription, the Reverse Transcription Complex RTC transform into pre-integration complex (PIC) for its migration to nucleus through nucleoporin embedded in the nuclear membrane. The disassembly of capsid as well as migration of RTC/PIC into nucleus are still debated (Campbell and Hope, 2015).
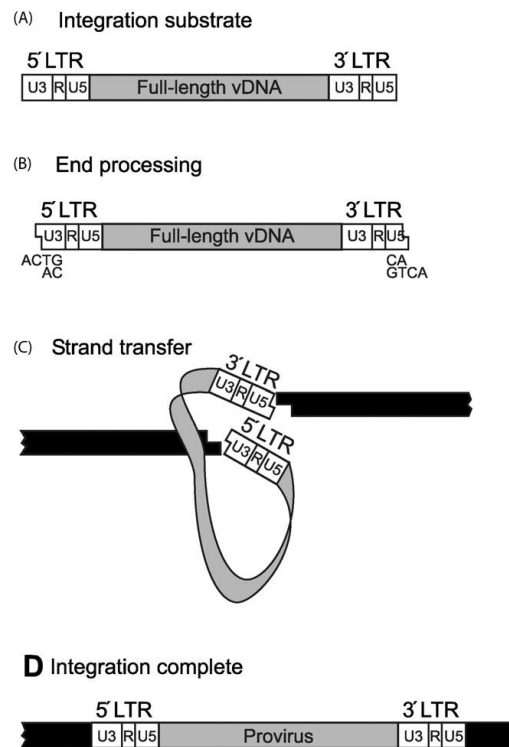
Figure 7. Integration of vDNA into the host chromosome is the final step of early infection. Adapted from (Thomas and Gorelick, 2008)

Integration of viral DNA into host genome performed by IN enzyme include two main steps (Figure 7). First step occurs in the cytoplasm, where IN binds to two specifically recognized nucleotides at 3' end of the UTR. Due to its endonuclease activity, these two nucleotides were removed creating a sticky end. This step is called as 3' processing. The second step occurs in nucleus where IN catalyze the insertion of 3' end processed viral DNA in to the host DNA (Campbell and Hope, 2015).

1.1.7.2. Late Phase

The late phase begins with transcription of viral DNA by the cellular *transcription* enzyme RNA polymerase II (Figure 8). Transcription starts from the R region at the 5' end of UTR and Tat protein promotes the elongation. The transcription gives mRNAs with multiple splicing degrees that includes: unspliced, incompletely spliced and spliced mRNAs. Unspliced mRNA is used as gRNA (9 kb) which encodes both Gag and GagPol proteins. Incompletely spliced mRNAs (4 kb) encodes Env, Vpu, Vif, and Vpr proteins. Interaction between RRE and Rev protein facilitate transportation of unspliced and incompletely spliced mRNAs to the cytoplasm. The completely spliced mRNA has the simplest transfer mechanism as it is directly transferred to the cytoplasm (Karn and Stoltzfus, 2012; Pollard and Malim, 1998).
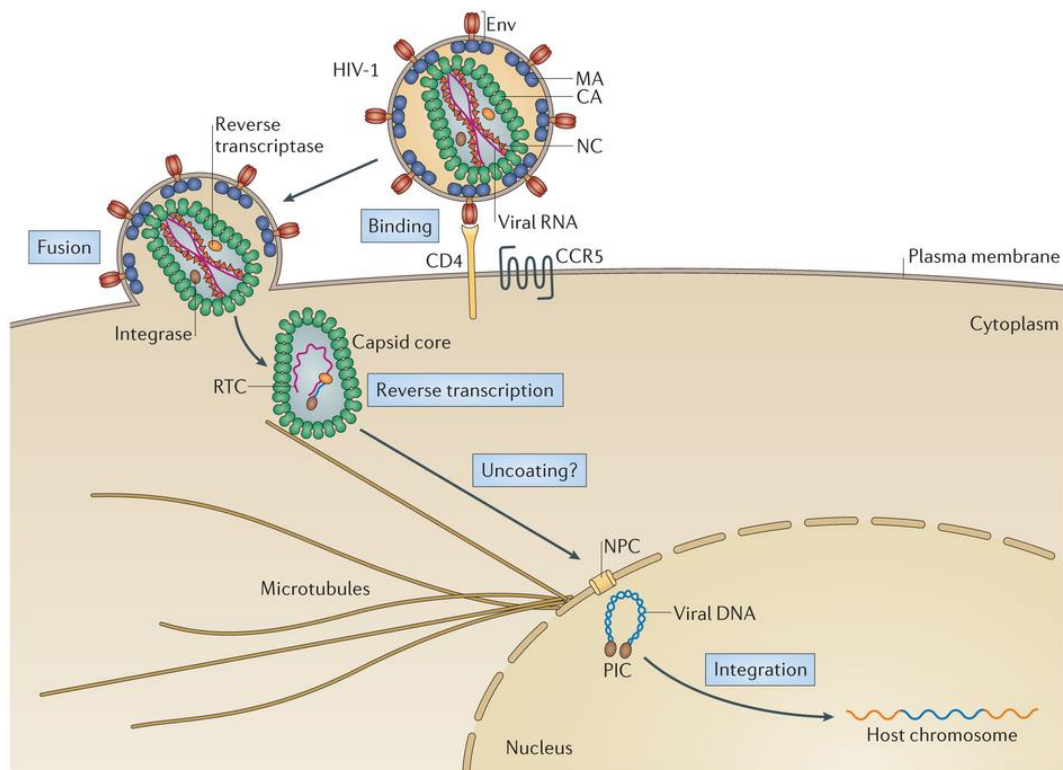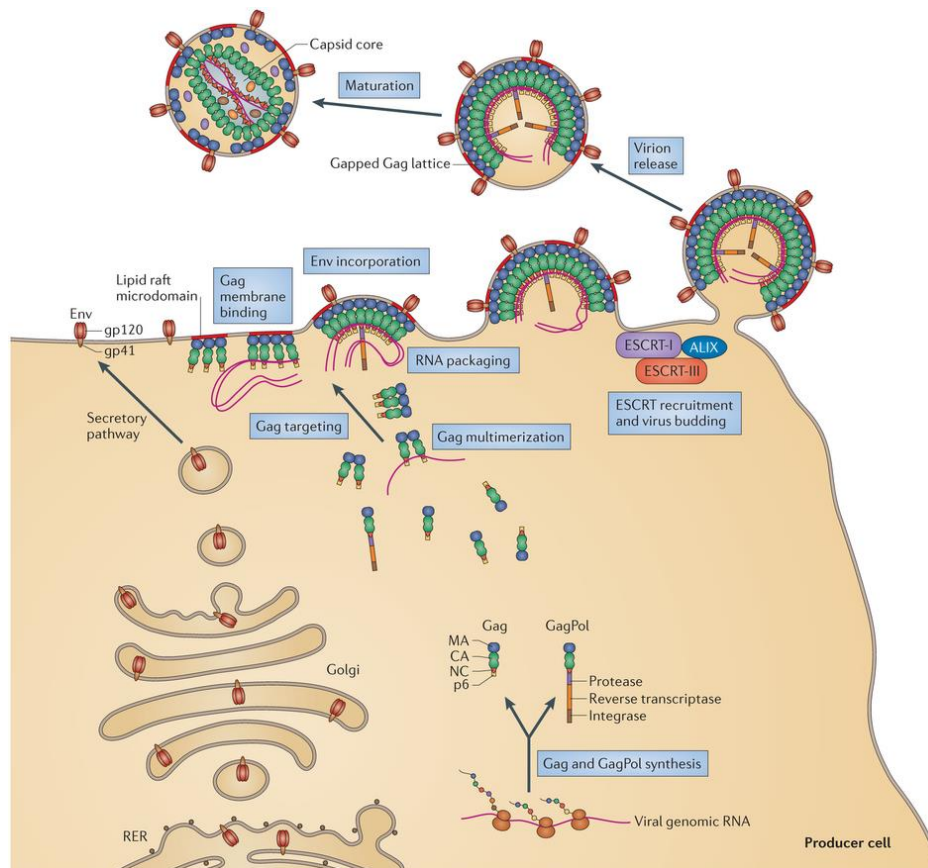
Figure 8. Late phase of the HIV-1 viral life cycle. Env, envelope glycoprotein; MA, matrix protein; CA, capsid protein; NC, nucleocapsid protein; ESCRT, endosomal sorting complex required for transport; ALIX, ALG2-interacting protein X; RER, rough endoplasmic reticulum. Adapted from (Freed, 2015).

Either free or membrane bound ribosomes execute the ***translation*** of mRNA in the cell. Translation of Env gene occur at the rough endoplasmic reticulum (Checkley et al., 2011). The unspliced mRNA encodes for Gag and GagPol polyproteins which further give rise to structural, functional and enzymatic proteins of the virus.

The ***assembly*** of viral particles starts with two important events that occurs in close time lapse of each other. First is the Gag oligomerization on gRNA and the second is the dimerization of the gRNA that is crucial for selection and packaging (Muriaux and Darlix, 2010; Paillart et al., 2004). Out of the pool of mRNAs in the cytoplasm, the Gag proteins select two unspliced gRNA. The dimerization of gRNA occurs through a series of correlated events happening at the 5' UTR (D'Souza and Summers, 2005). Dimerization of two gRNA strands begins with NC-promoted annealing of SL1 sequences (Bernacchi et al., 2017; Grigorov et al., 2007; Lu et al., 2011). Due to this specific recognition between SL sequences in gRNA and NC as a domain of Gag, two competent gRNAs are selected for packaging. On the other hand,

Gag oligomers are responsible for trafficking the gRNA-Gag complex to plasma membrane (El Meshri et al., 2015). The number of Gag oligomers on gRNA increases as the Gag-gRNA complex moves closer to plasma membrane (Kutluay and Bieniasz, 2010; Mailler et al., 2016). At the plasma membrane, MA protein of the Gag domain interacts with negatively charged lipids, and notably with phosphatidylinositol-(4,5)-biphosphate (PIP2) phospholipids (Ono et al., 2004; Saad et al., 2006). This starts the formation of the immature virion (Briggs et al., 2009).

Budding of viral particle require p6 and NC domain of the Gag protein, which interact with the cellular endosomal sorting complex required for transport (ESCRT) machinery. This recruitment process requires two-specific domains of p6, PTAP and YPXL domains, that binds to TSG101 (a tumor susceptibility gene 101) and ALIX (a ALG2-interacting protein X), respectively. The interaction of NC with Bro domain of ALIX facilitates the binding of Gag to ALIX. Later, ESCRT-I and ALIX interacts with ESCRT-III and AAA ATPase vacuolar protein sorting 4 (VPS4), respectively, which are required for membrane fission and virion release (Freed, 2015; Sundquist and Kräusslich, 2012).
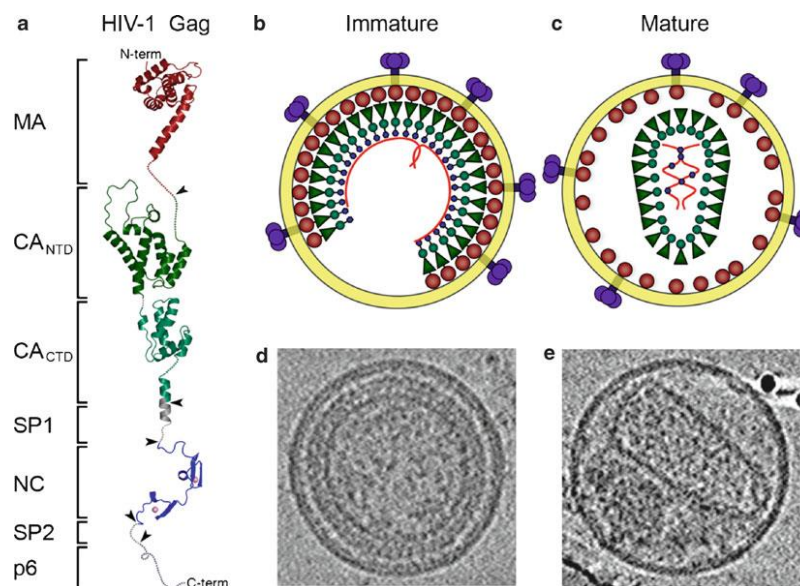
Figure 9 Organization of the immature and mature HIV-1 virions. Schematic structural model of full length HIV-1 Gag (a). Individual domains are in different colors and are labeled on the *left*. Schematic model and images of the immature virion (b, d) and mature virion (c, e). Adapted from (Ganser-Pornillos et al., 2008)

Maturation of the virion particle begins after it is released from the infected cell. An immature virion particle has radially aligned Gag molecules with the N'-terminal MA domain

being bound to the inner viral membrane and the C'-terminal of NC domain stretched towards the core of the virion. Transformation into the mature particle requires catalysis from the PR enzymes which are released from auto-processing of GagPol. PR cleaves the Gag and GagPol polyproteins to release MA, CA, NC, p6, PR, IN, and RT. As a result, the morphology inside the virion particle changes, giving rise to a conical shaped capsid (Freed, 2015; Ganser-Pornillos et al., 2012; Sundquist and Kräusslich, 2012).

1.1.8. Antiretroviral Therapy (ART)

The ever-increasing understanding on HIV-1 life-cycle is beneficial for antiviral drug development. Till date, FDA approved 28 drugs are in market that belongs to distinct classes. These drugs are classified based on their molecular mechanisms and resistance profiles: (1) nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs/NtRTIs), (2) non–nucleoside reverse transcriptase inhibitors (NNRTIs), (3) integrase inhibitors, (4) protease inhibitors (PIs), (5) fusion inhibitors, and (6) coreceptor antagonists.

a) Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs/NtRTIs)

These were the first classes of drugs approved by FDA. Before enacting their antiviral effect, NRTIs require host cell entry and phosphorylation by cellular kinases (Mitsuya et al., 1985). NRTIs are 2'-deoxynucleosides analogues that lack 3'-hydroxyl at the sugar moiety. Reverse transcriptase is unable to differentiate between NRTIs and dNTPs, and thus they are taken up and incorporated in the nucleic acid chain. Due to lack of 3'-hydroxyl at the sugar moiety, it is unable to build the 3'-5'-phosphodiester bond that occurs between two natural dNTPs in the classical DNA synthesis. This results in termination of the growing viral DNA chain (Arts and Hazuda, 2012). Currently, there are eight FDA-approved NRTIs: abacavir (ABC, Ziagen), didanosine (ddI, Videx), emtricitabine (FTC, Emtriva), lamivudine (3TC, Epivir), stavudine (d4T, Zerit), zalcitabine (ddC, Hivid), zidovudine (AZT, Retrovir) and Tenofovir disoprovil fumarate (TDF, Viread). Two mechanisms mediate resistance to NRTIs: removal of NRTIs from the 3′-end of the nascent chain through an ATP-dependent pyrophosphorolysis and reversal of chain termination (Boyer et al., 2001).

b) Non–Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

NNRTIs are the backbone of ART because they show unique antiviral activity, high specificity and low toxicity that makes them a frequent choice for first-line therapy (Li et al., 2014). NNRTIs inhibit HIV-1 RT binding as it binds at the allosteric sites and changes the

spatial conformations for the substrate binding site. This eventually reduces the polymerase activity. The well-known structure of the hydrophobic pocket of RT allowed to tailor the design for substituting NNRTIs. The so-called NNRTI-binding pocket consists of hydrophobic residues (Y181, Y188, F227, W229, and Y232), and hydrophilic residues such as (K101, K103, S105, D192, and E224) of p66 and p51 subunit (Arts and Hazuda, 2012). Currently, there are five approved NNRTIs: etravirine, delavirdine, efavirenz, nevirapine and rilpivirine. NNRTI resistance generally results from the substitutions of L100, K101, K103, E138, V179, Y181, and Y188 amino acids in the NNRTI-binding pocket of RT (Rodgers et al., 1995).

c) Integrase Inhibitors

They are recently developed interfacial inhibitor drugs targeting intasome which consist of catalytic site for one IN polypeptide, the vDNA, and the two catalytic magnesium cations (Arts and Hazuda, 2012; Métifiot et al., 2013). Since 2007, FDA has approved three integrase inhibitors: Raltegravir, Elvitegravir, Dolutegravir. Resistance to Raltegravir is associated with different pathways or sets of mutations in the integrase at Y143, N155, or Q148 residues that reduce the susceptibility to Raltegravir (Arts and Hazuda, 2012; Fransen et al., 2009)

d) Protease Inhibitors

The role of HIV-1 protease is to cleave the Gag and GagPol polyproteins precursors during virion maturation. Thus, to prevent the transformation of immature virion into mature virion, the activity of protease is targeted by protease inhibitors. Ten protease inhibitors are currently approved: amprenavir (APV, Agenerase), atazanavir (ATZ, Reyataz), darunavir (TMC114, Prezista), fosamprenavir (Lexiva), indinavir (IDV, Crixivan), lopinavir (LPV), nelfinavir (NFV, Viracept), ritonavir (RTV, Norvir), saquinavir (SQV, Fortovase/Invirase), and tipranavir (TPV, Aptivus). More than 20 substitutions are associated with resistance of protease inhibitors (Arts and Hazuda, 2012; Rhee et al., 2010). Other pathways for resistance are based on mutations at the protease cleavage site in Gag and GagPol proteins. (Anderson et al., 2009; Fun et al., 2012)

e) Entry Inhibitors

Entry inhibitors target the interactions of gp120 with CD4 receptors and gp120 with coreceptors (CCR5 and CXCR4). Entry inhibitors are divided into two categories: Fusion inhibitors and small CCR5 antagonists. *Fusion inhibitors*: The two domains of gp41 interact each other to promote the fusion process. Thus, fusion inhibitors are designed, as synthetic

peptides that mimic the C'-helices of gp41 and disrupts the intermolecular interactions of the virus protein. Enfuvirtide, a 36 amino acid drug is the only FDA approved drug available as a fusion inhibitor. Resistance to fusion inhibitors appeared due to mutations in the amino-terminal heptad repeat region of gp41 (Rimsky et al., 1998). *Small CCR5 antagonists* are allosteric inhibitors that binds to the hydrophobic pockets within the transmembrane helices of CCR5 (Dhami et al., 2009). They stabilize the conformation of CCR5, preventing its recognition. Three antagonists (VCV, MVC, and Aplaviroc) have been shown to inhibit virus replication (Dorr et al., 2005). Binding of CCR5 coreceptors to gp41 can also be inhibited with natural chemokines that compete for binding. Natural ligands of CCR5, like MIP-1α, MIP-1β, RANTES (and its derivatives: Met-RANTES, AOP-RANTES, PSC-RANTES and 5P12-RANTES), have been tested for their antiviral activity.

f) Combinational Therapy

The ability of HIV-1 subtype to overcome drug resistance through mutation allowed clinicians to use instead of one drug, a combination of several drugs. These drugs may belong to same class or different classes of anti-retrovirals. A Highly Active Anti-Retroviral Therapy (HAART) is administered with combination of three drugs in which atleast two drugs should have two different targets. Although HAART is not a cure, but it has helped in prolongation of patient's life expectancy.

1.2. Structure-Function Relationship of Nucleocapsid Protein

1.2.1. Introduction

Generated by proteolytic cleavage of the 55 kDa Gag precursor, there exists 1500–2000 copies of mature HIV-1 nucleocapsid protein NCp7 (NC) surrounding the viral genome. Multiple steps of proteolytic cleavage of the Gag, as shown in Figure 10, generate first a 71 amino acid intermediate (p9) containing NC (p7) and spacer peptide 2, and further a 55 amino acid mature form of NC (p7) (Thomas and Gorelick, 2008).
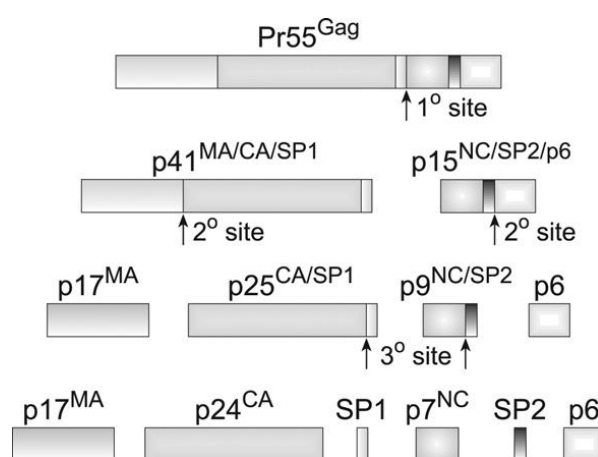
Figure 10 Proteolytic processing of HIV-1 Gag by PR (Thomas and Gorelick, 2008).

NC is a 55 amino acid, small (5–7 kD) retroviral protein consisting of a highly basic N-terminal region and two strictly conserved $CX_2CX_4 HX_4 C$ (CCHC) Zinc finger (ZF) motifs connected by a basic linker (Berg, 1986; Mély et al., 1996; Rice et al., 1995; Summers et al., 1990) (Figure 11). The ZF motifs contain conservatively substituted hydrophobic and aromatic residues that forms a hydrophobic cleft on the surfaces of both the knuckles which plays crucial roles in binding to nucleic acids (NA) (specifically and non-specifically). Consequently, NC displays multifunctional roles during several steps of HIV-1 virus life cycle that are basically driven by its nucleic acids chaperoning properties (Darlix et al., 2011). It means that NC effectively remodels the NA so that the most thermodynamically stable conformations are gained. Such remodeling are achieved through certain properties of NC namely, (i) nucleic acid aggregation; (ii) duplex destabilization; (iii) rapid on-off kinetics. Nucleic acid aggregation is mainly driven by the basic residues while the duplex destabilization is largely administered through ZFs.
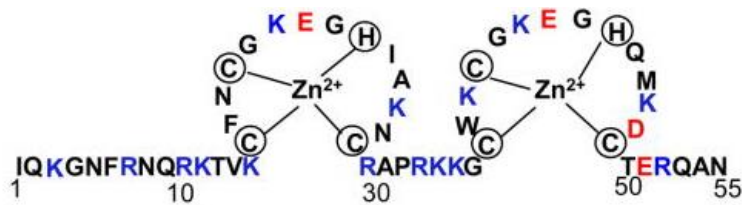
Figure 11 Amino acid sequence of HIV-1 NC(1–55). The zinc coordinating residues shown in red and the basic residues highlighted in blue. Adapted from (Levin et al., 2005).

1.2.2. Nucleic Acid Chaperoning of NC

NC has the ability to bind specifically and nonspecifically to nearly any NA sequence. However, the binding affinies varies with the nature, sequence and folding of the interacting sequences. Interestingly, through nonspecific low affinity interaction, NC is able to bind nearly any 5–8 nucleotide sequence due to its basic nature (Vuilleumier et al., 1999). Consequently, this binding mode is majorly defined in virus, where due to the small core, the concentration of NC and the relevant NAs are in millimolar range. Therefore, an extensive coating of 1500–2000 copies of NC can be achieved ensuring its protection against cellular nucleases (Darlix et al., 2011).
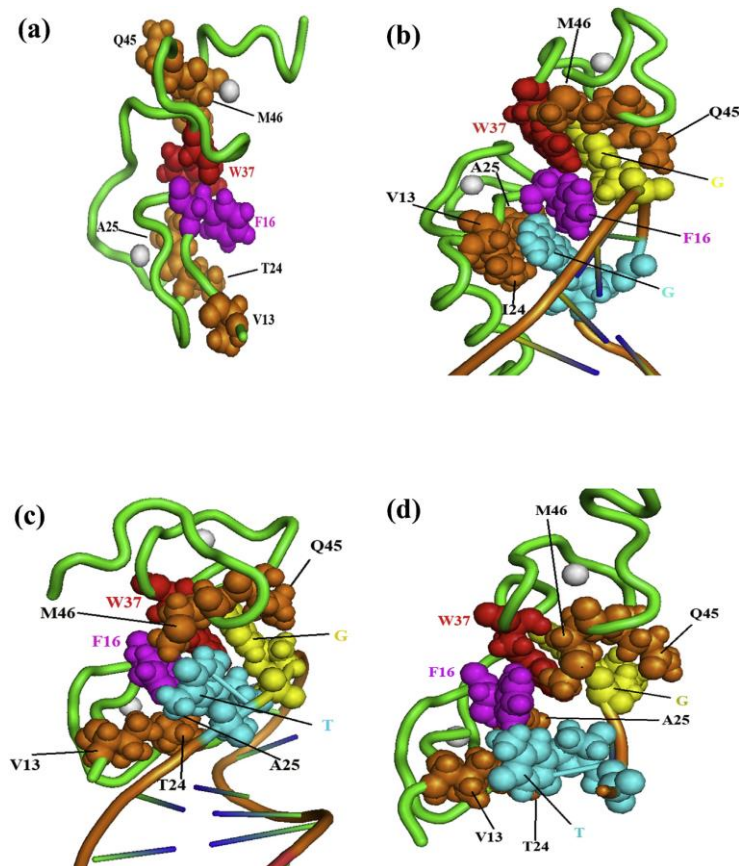
Figure 12 Three-dimensional structure representations of free HIV-1 NC and bound to various RNA and DNA molecules. For the sake of clarity, only the major structural determinants of NC are shown here. The NC protein backbone is shown as a green tube and the two zinc atoms coordinated by the CCHC tetrad in each zinc finger are indicated as white spheres. Key hydrophobic interactions taking place between NC and the NA are highlighted here. In addition, the hydrophobic residues contacting the nucleic acid are illustrated as spheres. The F16 and W37 aromatic residues are in magenta and red, respectively, while other residues in ZF1 (V13, T24, A25) and ZF2 (Q45, M46) are in orange. The two nucleotides with the tightest hydrophobic interaction with the protein side chains are 2G residues in blue and yellow in (b) NC-SL3 RNA complex (De Guzman, 1998), T and G in (c) NC-(–)PBS_DNA complex (Sarah Bourbigot et al., 2008), and (d) NC-mini cTAR_DNA complex (Bazzi et al., 2011). Relative to free NC (a) (Morellet et al., 1998)), note that the hydrophobic interface is reorganized in the NC–NA complexes (b–d). In (b) the two guanines in blue and yellow are deeply inserted within the hydrophobic pocket formed by the zinc fingers. In (c) and (d), one guanine (G in yellow) is inserted in the ZF2 hydrophobic pocket while the thymine (T in blue) is on the exterior of the hydrophobic pocket. Also note that the T24, V13 andA25 residues are in close contact with the double stranded part of (–)PBS_DNA in (c), while there is no such contact in the NC-SL3 RNA complex in (b). Adopted from (Darlix et al., 2014)

NC exhibits also sequence specific binding to numerous defined single strand NA containing unpaired guanine residues (Levin et al., 2005). RNA apical loops with GNG sequences (De Guzman, 1998) and DNA apical loops with TG, GNG and TNG sequences (Avilov et al., 2012; Sarah Bourbigot et al., 2008; Vuilleumier et al., 1999) display high binding affinity for NC. The reason behind this strong binding was inferred from the resolved 3D structure of NC-NA complexes that demonstrated an active involvement of both the zinc fingers (Figure 12). The two ZFs are not functionally equivalent (Gorelick et al., 1993), ZF1 being more critical than ZF2 for the nucleic acid chaperone activity (Zargarian et al., 2014). But the specific binding of NC to these sequences are largely mediated through the hydrophobic platform formed at the top of two ZFs by the V13, F16, T24, A25, W37, Q45 and M46 residues. Importantly, all solved structures demonstrate that guanine binds deeply into the ZF1 and ZF2 hydrophobic pockets and stacks extensively with W37. While F16 in ZF1 stacks with guanines (in case where two guanines are available in NA sequences) or stacks partly with thymine or cytosine residues that remain outside of the hydrophobic pocket of ZF. Thus, such specific and

strong binding properties of NC result in recognition of target Stem loop (SL) ψ-encapsidation signal of genomic RNA among a large excess of cellular RNAs during virus assembly.

These interesting internal dynamics of NC protein that is essential for NA recognition through W37-G interaction has been studied in section 4.2 and 4.3 of results and discussion. Section 4.2 focuses on comprehending the kinetics of NC-promoted annealing of (–)PBS with its complementary (+)PBS during the reverse transcription step of HIV-1 life cycle. Whereas, the section 4.3 investigate the orientation of ZF1 and ZF2 along the NA chain in NC-RNA and NC-DNA complexes of SL3 and (–)PBS model.

1.2.3. Role of NC During HIV-1 Reverse Transcription

When a mature HIV-1 virion infects a susceptible target cell, interactions of the envelope glycoproteins with the coreceptors on the surface of the cell bring about a fusion of the two membranes of the host cell and the virion. This fusion introduces the contents of the virion into the cytoplasm of the cell, setting the stage for reverse transcription. Reverse transcription is a multistep process that leads to the synthesis of linear double stranded DNA copy from the single stranded viral RNA genome that is finally integrated to the host chromosomes (Herschhorn and Hizi, 2010). This process is an early phase event occurring in the cytoplasm of the cell and catalyzed by the virus encoded transcriptase (RT) enzyme. The sequential steps of reverse transcription are shown in Figure 13.

The multistep process begins with the annealing of a specific cellular tRNA to the primer binding site (PBS) and the role of NC at every step is discussed.

A. Initiation

Step 1: "tRNA primer placement" this step includes the hybridization of the 3' end of 18 nt cellular tRNA^{Lys} primer to the complementary primer binding site present at the 5' end of the viral genome. This reaction is catalyzed by the NC domain in Gag (Cen et al., 1999; Feng et al., 1999) or NCp7 in its mature form (De Rocquigny et al., 1992; Saadatmand and Kleiman, 2012; Seif et al., 2015; Sleiman et al., 2012). The intrinsic dynamics of NC plays a dual role as the basic amino acids bring the complementary tRNA and PBS sequences together, whereas the two zinc fingers are responsible for destabilizing the organized viral RNA genome (Levin et al., 2005). There has been several studies suggesting the role of Gag and NC during the annealing of tRNA (Campbell and Rein, 1999; Cen et al., 1999). A two step annealing process was suggested in which the initial annealing is promoted by Gag while the final

primer/template remodeling is promoted by mature NC (Feng et al., 1999; Seif et al., 2015; Sleiman et al., 2013).

Step 2: RT catalyzes the synthesis of primer and extends it to 5' end of the vRNA to form minus-strand strong-stop DNA (–)ssDNA. This (–)ssDNA contained the unique 5' genomic sequence (U5) and R regions. The RNase activity of RT degrades the vRNA sequence that has already been reverse transcribed.
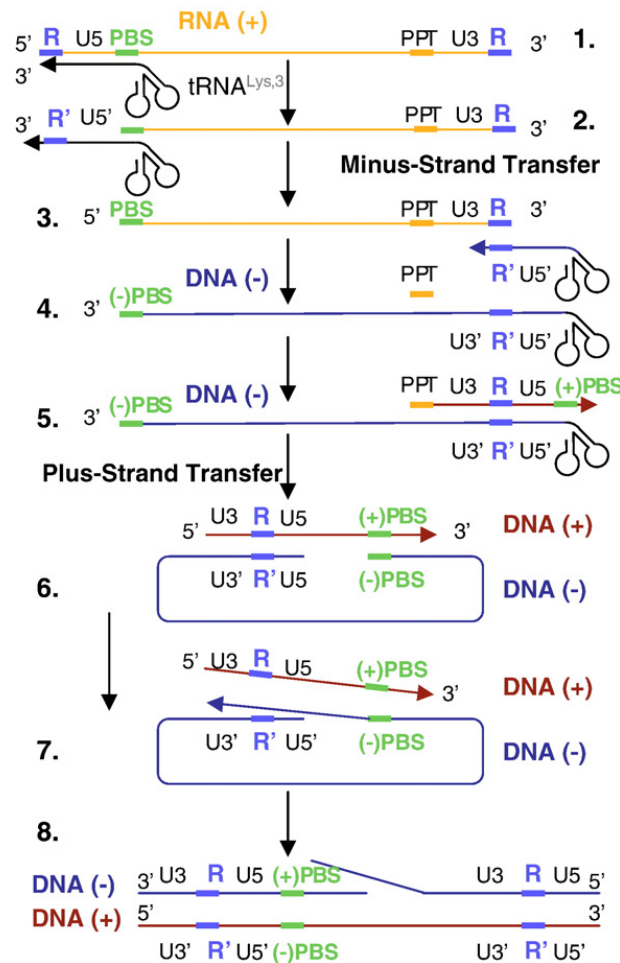
Figure 13 Schematic representation of reverse transcription (Sarah Bourbigot et al., 2008).

B. Minus strand transfer

Step 3: In order to extend this small (–)ssDNA in a full length (–)ssDNA copy of the viral RNA genome the "minus strand transfer" step is essential. This transfer is facilitated by the annealing of R region of short (–)ssDNA to the complementary R region at the 3' end of the genome. Both TAR and polyA sequences are imperfect SLs having several conserved bulges, mismatches and internal double-stranded segments (Bazzi et al., 2011; Godet et al., 2006; Mougel et al., 2009). The role of NC is significant during annealing of cTAR TAR

sequences as it chaperones the NA and dramatically accelerates the annealing by 3,000-fold (Guo et al., 2000, 1998; Lapadat-Tapolsky et al., 1995; Vo et al., 2009, 2006; You and McHenry, 1994). The annealing of cTAR-TAR in the presence or absence of NC proceeds through a two-step reaction mechanism (via intermediate complex). In the absence of NC, the annealing intermediate of the kissing pathway is a loop–loop kissing complex involving six base-pairs (Kanevsky et al., 2011). Interestingly, the NC-assisted pathway can proceed through multiple pathways depending on the solution concentration and substrate used. At high concentrations of NC, the dominant zippering pathway involves annealing through the 3' and 5' ends of the stem-loop (Zeng et al., 2007). The reaction pathway is likely selected based on the available reactive intermediates which require least base pair melting before annealing (Kanevsky et al., 2005; Vo et al., 2009, 2006).

Step 4. In continuation of the minus strand transfer, elongation of (–)ssDNA and RNase degradation continues except for the PPT sequences that are RNase resistant.

Step 5. RT synthesizes the (+)ssDNA that is initiated from the PPT sequence. Thereafter, the RNase activity of RT degrades the tRNA and PPT primer regions from (–)ssDNA and (+)ssDNA, respectively.

C. Plus strand transfer

Step 6. Plus strand transfer is supported by annealing of the 18 nt PBS region of (–)DNA (or also called as (–)PBS) and the 3' end of PBS region of (+)DNA.

Step 7, 8. Thereafter, RT resumes the formation and circularization of the DNA, resulting in linear double strand (ds) DNA strands. Annealing of short and stable (–)PBS and (+)PBS SLs is facilitated by NC. NMR resolved structure of Δ(–)PBS (truncated (–)PBS without the overhangs) and NC(12–55) proves crucial in comprehending the pivotal role of NC in successfully destabilizing and stretching the 5'-CTG-3' NA in the loops thus making them available as an intermediate for annealing reaction (Sarah Bourbigot et al., 2008) (Figure 14). The 3D structure of NC-(–)PBS complex demonstrated the interaction of hydrophobic residues of ZF1, namely Val13, Phe16, Thr24, and Ala25, with the C5 and T6 bases, while the Trp37, Gln45, and Met46 residues of the second zinc finger interact with the G7 base. These interactions drive the T6 and G7 bases to the exterior of the loop, making them competent for annealing with (+)PBS loop.
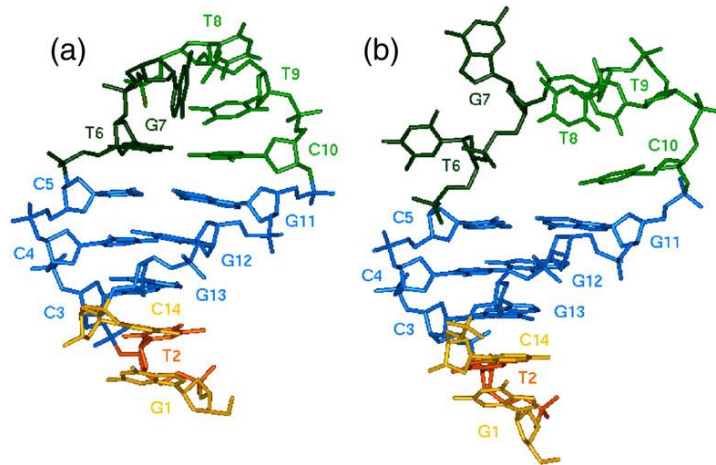
Figure 14 NMR resolved structure of Δ(–)PBS (a), and complex with NC(12–55) (b).

Taking an insight into the annealing of (–)/(+)PBS reveals that the reaction proceeds through two different pathways (Figure 15). In the absence of NC, the nucleation occurs through the single strand overhangs of (+)PBS and spontaneously results in duplex formation. However, in the presence of NC the nucleation occurs via loop-loop interaction and the reaction kinetics is increased by 60-fold (Ramalanjaona et al., 2007). Moreover, NC mutants (SSHS) with unfolded ZFs were able to increase the annealing kinetics but were unable to modify the annealing pathway, clearly highlighting the need of the hydrophobic plateau at the top of ZF for specific loop-loop interaction.
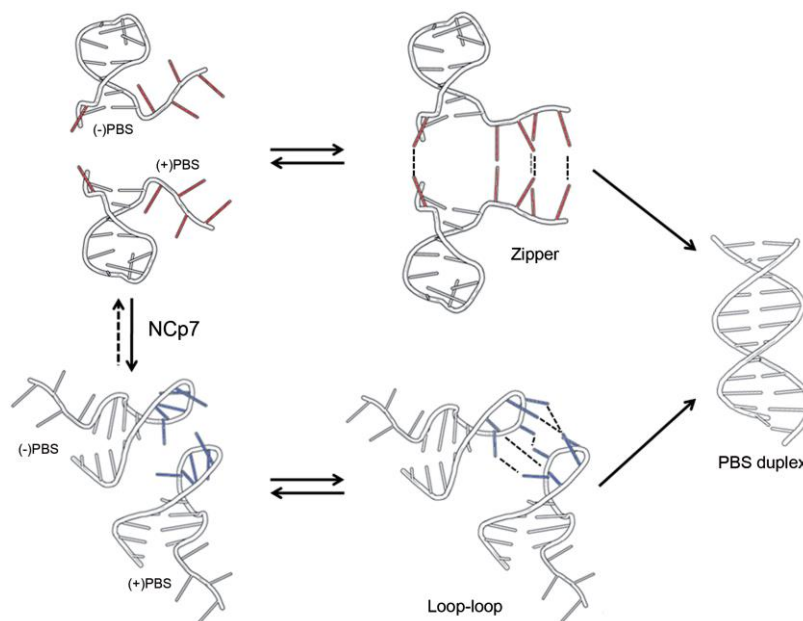
Figure 15 Proposed (–)/(+)PBS annealing pathways. Adapted from (Julien Godet et al., 2011)

1.2.4. Role of NC Domain of Gag in Viral Assembly

Assembly of HIV-1 particles requires selection of two copies of genomic RNA (gRNA) (Nikolaitchik et al., 2013) by the Gag polyprotein that contains all the signals and domains which orchestrate this process from the nucleation step to virus release by budding (Kutluay and Bieniasz, 2010; Muriaux and Darlix, 2010; Sundquist and Kräusslich, 2012). However, the mechanism implemented by Gag for selection of gRNA amidst large excess of cellular and spliced viral RNA is still unknown. Gag likely uses its NC domain to bind in nanomolar affinity to the psi specific region of gRNA (Cimarelli and Darlix, 2002; Lever, 2007; Muriaux and Darlix, 2010). The psi region corresponds to the SL1, SL2, SL3 and SL4 sequences present at the 5' end of gRNA, and each SL sequence plays a crucial role in assembly. This nucleation event can certainly favor Gag-Gag interaction on dimeric RNA (Bernacchi et al., 2017; Cimarelli and Darlix, 2002). This is supported by the recent series of works presented by the Group of J.C. Paillart and R. Marquet in which the binding affinity of Gag to gRNA has been shown to be 10-fold stronger than to spliced RNA sequences (Abd El-Wahab et al., 2014; Houzet et al., 2007) (Figure 16). Due to the long-range interaction of NC and MA domains of Gag to the upstream and downstream of the psi region, SL1 is optimally exposed to Gag recognition. Moreover, SL1 was shown to be the only SL that binds with high affinity, that was surprising outcome since prior to this study SL3 was considered as the main packaging signal (Aldovini and Young, 1990; Lever, 2007).
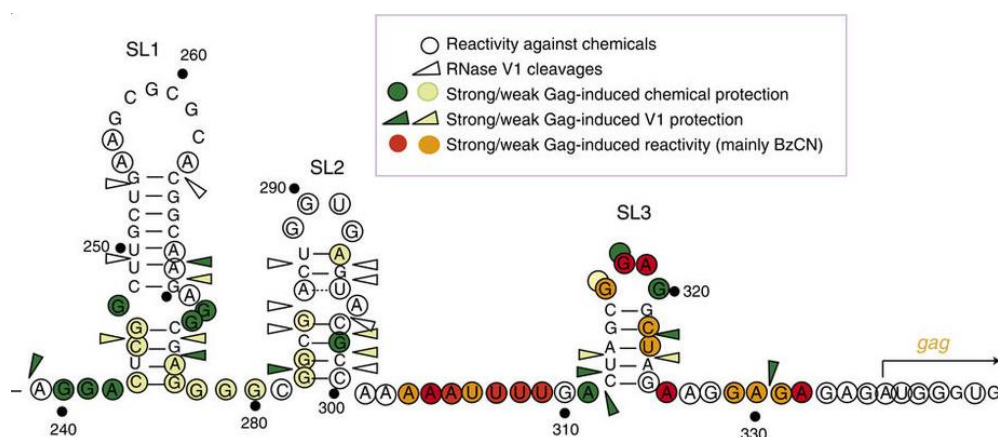
Figure 16 Encapsidation signal (SL1-3) and Gag induced activity changes on them. Adapted (Abd El-Wahab et al., 2014)

1.2.4.1. NC Promoted RNA Dimerization

The Dimer initiation site in gRNA is a highly conserved and stable sequence of SL1. The process of dimer formation is characterized as a two step mechanism. The dimerization

begins via annealing of the auto complementary apical loop to form a metastable complex via loop-loop kissing mechanism. NC protein facilitates by refolding the NA in the stem thus resulting in its melting (Figure 17) (Muriaux et al., 1996, 1995; Shubsda et al., 2002; Yuan et al., 2003). By destabilizing the intramolecular base pairing, NC shifts the thermodynamics of the metastable kissing complex toward the stable dimer (Paillart et al., 2004; Rist and Marino, 2002; Weixlbaumer et al., 2004). Both RNA-RNA interaction and NC-promoted annealing of SL1 dimers are crucial steps required for dimerization, efficient packaging, viral maturation and, as a consequence, optimal viral infectivity (Mujeeb et al., 2007b).
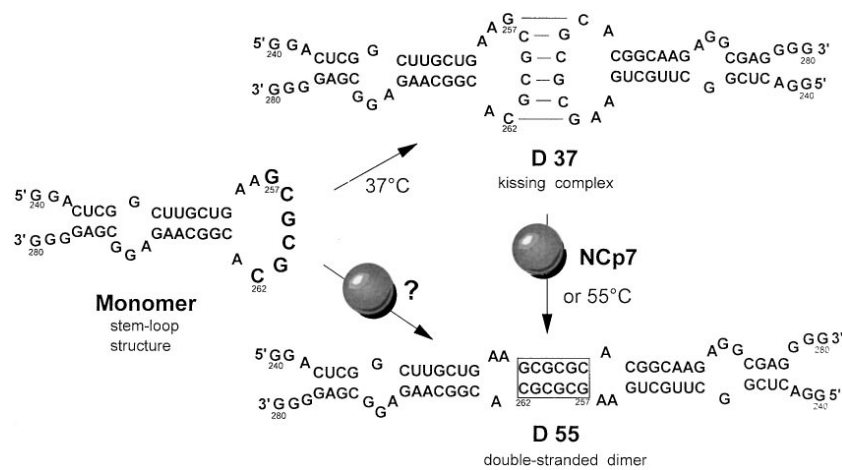
Figure 17 NC mediated Dimerization. (Muriaux et al., 1996)

1.2.4.2. NC - SL3 Recognition

The 3D structure of NC bound to SL3 recognition element was solved by group of M.F. Summers in 1998 (Figure 12b) (De Guzman, 1998). Later, the group of P. Borer investigated the role of NC in specific recognition of SL3 motif of gRNA for encapsidation. The folding of SL3 GGAG tetraloop is quite flexible. As a result, the G10 and G12 residues are in unstacked orientations while the G9 and A11 residues are tightly constrained in the minor grove. In NC-SL3 structure, the N1H and C6O of unstacked 5' end G10 and 3' end G12 make hydrogen bonds with the N terminal W37 of ZF2 and the C-terminal F16 of ZF1, respectively. Meanwhile, A11 residue is in the exterior of the loop establishing hydrophobic contacts with the protein. Sequence specific binding of NC to SL3 tetra-loop was shown to lose its affinity by a factor 15–120 if A11 residue was replaced by any of the U, G or C residues (Paoletti et al., 2002). NC binds with SL3 in 1:1 binding stoichiometry. Their binding affinities were determined at several salt concentrations, but 100–200 mM of salt concentrations were preferred as the binding stoichiometry remained conserved (Athavale et al., 2010; Warui and Baranger, 2012). In recent work from the Group of J. Mak, first time a thermodynamic

investigation (binding affinity) of different domains of Gag protein, namely Gag, GagΔP6, GagΔCA, p15 and NCp7, were established with their interaction with SL3 and mutant sequences (Tanwar et al., 2017). They demonstrated specific interaction between Gag and Adenosine containing RNA motifs; and oligomerization of Gag; altogether they were found to be energetically favorable reactions that facilitate virion particle formation.

1.2.4.3. Gag-Gag oligomerization and trafficking to plasma membrane

Gag-Gag oligomerization and its interaction with the plasma membrane initiates step towards formation of virus particle. Biophysical studies have reported that oligomerization of Gag is facilitated by its NC domain (El Meshri et al., 2015; Ott et al., 2009). The NC-RNA scaffold formation is likely a support for assembling of Gag molecules. Moreover, in recent work it has also been identified that the ribosomal protein RPL7 contribute in Gag oligomerization process by interaction with NC domain of Gag (Mekdad et al., 2016).

Binding of Gag to the plasma membrane is also thought to be driven by the interaction of NC with the lipids of plasma membrane (Kempf et al., 2015). Several works have highlighted the fact that by mutations in zinc finger of NC result in unsuccessful oligomerization and its trafficking to the plasma membrane (Cimarelli et al., 2000; Dawson and Yu, 1998; Robert J. Gorelick et al., 1999; Kafaie et al., 2008).

1.3. Fluorescence and Nucleic Acids

1.3.1. Introduction to Fluorescence

Fluorescence is a form of luminescence with an emission of a photon from an excited state (S_1) to the ground state (S_o). The process of fluorescence can be illustrated using the Jablonski diagram (Figure 18).
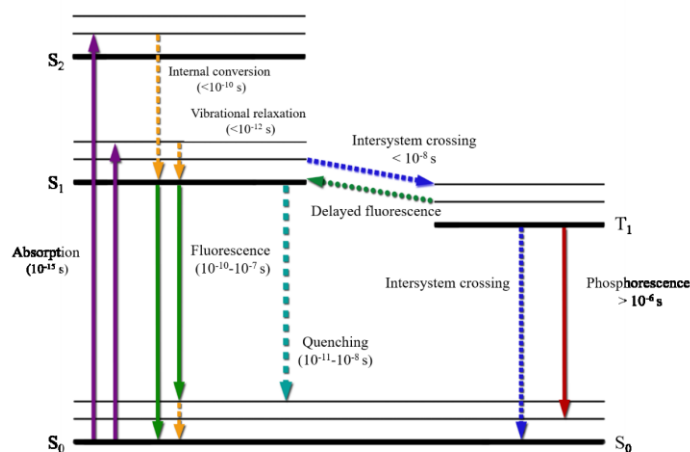
Figure 18. Jablonski diagram.

An organic molecule is generally found in its lowest electronic singlet state or ground state (S_o). It stays in its ground state until it is excited with an energy higher than the energy barrier between the ground and excited state. With an energy higher than the threshold energy, it absorbs the photons (absorption) to reach the excited state (S_n with n $\geq$ 1) within 10^{-15} s, generating *Franck Condon* states. Subsequently, through vibrational relaxation the vibrationally excited molecule is quickly reduced (10^{-12}–10^{-10} s) to the lowest vibrational state through collision with the solvent molecules, generating a stokes shift. The molecules in the S_1 state decay to the ground state through two processes, either *fluorescence* or *phosphorescence*. When the molecule from the S_1 state decays to the S_0 state, by emitting a photon (of lower energy than absorbed), this radiative process *(k_r)* is termed as fluorescence. Along with it, some part of energy is lost in form of heat that accounts for non-radiative processes. When the molecule is in S_1 state, there is another probability that results in *Intersystem crossing (k_{ISC})* of the molecule from S_1 to T_1 state, by changing the spin of an electron. Again, the molecule undergoes vibrational relaxation to the lower vibrational state in T_1 and thereafter, it reaches the ground state by emitting a photon and this process is called as phosphorescence. However, phosphorescence does almost never happen at room temperature, due to the much higher

probability of the molecule to go back to the ground state though non-radiative processes, that include molecular collisions, photoinduced electron transfer, bond rotation or vibration.

The characteristic fluorescence of a fluorophore is measured in terms of quantum yield (Φ), which is the ratio of the number of photons emitted to the number of photons absorbed. The maximum value of quantum yield is close to unity, but generally, its value is lower due to the non-radiative losses. Thus, the quantum yield takes in to account both radiative and non-radiative decay constant.

$$\phi_f = \frac{k_r}{k_r + k_{nr}} \quad (1)$$

k_r is the rate constant for fluorescence emission and k_{nr} is the sum of non-radiative rate constants. The relative quantum yield is determined by using the following equation.

$$\phi_f = \phi_{ref} \left(\frac{Abs^{\lambda}_{Std} \times Intg.Intensity_x}{Abs^{\lambda}_x \times Intg.Intensity_{Std}} \right) \left(\frac{\eta_x}{\eta_{Std}} \right) \quad (2)$$

ϕ_{ref} is the fluorescence quantum yield of the reference, Abs^{λ}_{Std} and Abs^{λ}_x are the absorbance of the standard and the sample at the characteristic wavelength, *Intg Intensity$_x$* and *Intg Intensity$_{std}$* are the integral intensities of the fluorescence spectra, while η_x and η_{Std} are the refractive indexes.

Fluorophores are also compared in terms of their brightness, which is the product of their fluorescence quantum yield and absorption coefficient (usually at the absorption maximum)

$$Brightness = \phi_f \times \varepsilon(\lambda) \quad (3)$$

Fluorescence lifetime is another characteristic parameter that is used to describe a fluorophore. Fluorescence lifetime is the average time spent by the molecule in the excited state before coming to back to the ground state. Fluorescence lifetime is a parameter that is mostly unaffected by inner filter effects, static quenching and variations in the fluorophore concentration.

$$\tau_f = \frac{1}{k_r + k_{nr}} \quad (4)$$

The fluorescence properties of the molecule can also be related to its absorbance by means of the radiative rate constant, k_r, and the absorption coefficient, and can be expressed with the Strickler-Berg equation (Strickler and Berg, 1962)

$$k_r = 2.880 \times 10^{-9} \eta^2 \frac{g_m}{g_n} < \vartheta_f^{-3} >^{-1} \int \varepsilon(\vartheta) d \ln\vartheta \quad (5)$$

where g_m and g_n are the degeneracies of the ground (m) and upper (n) electronic state, η is the refractive index of the medium and ϑ_f is the wavenumber of emission. This equation is applicable for strong, broad-banded transitions in molecules.

The fluorescence intensity of a fluorophore can decrease as the excited molecules depopulate via radiation less processes, resulting from either static or collisional quenching. Collisional quenching occurs when an excited fluorophore interacts with quenching molecules or groups that deactivate the fluorophore. The intensity changes due to dynamic quenching can be fitted to the Stern-Volmer equation.

$$\frac{F_0}{F} = 1 + K[Q] = 1 + k_q \tau_0 [Q] \quad (6)$$

where K is the Stern-Volmer quenching constant, k_q is the bimolecular quenching constant, τ_0 is the unquenched lifetime, and $[Q]$ is the quencher concentration. Typically, dynamic quenching results in a linear dependency on the quencher concentration. The bimolecular quenching constant represents the extent by which the fluorophore is affected by the quencher. If the fluorophore is well accessible to the quencher, a large value of k_q is expected. Various molecules are utilized as quenchers such as oxygen, halogens, amines, and electron-deficient molecules like acrylamide. However, if the Stern-Volmer plot is not linear, this suggests the presence of static quenching, due to formation of non-fluorescent ground state complex or a quencher being in close vicinity to the fluorophore. Quenching experiments have been used to study accessibility of Trp residues in proteins, conformational and dynamic properties of proteins (Alemán et al., 2014), microdomains in membranes (Silvius and Nabi, 2006), and RNA folding dynamics.

Resonance energy transfer (RET) is another process that occurs in the excited state. This process involves non-radiative transfer of the excited state energy of the donor chromophore to an acceptor chromophore. Various mechanisms facilitate this energy transfer such as Dexter mechanism (electron exchange) and Förster mechanism. Dexter mechanism operates when the donor and acceptor are in short distance and requires an intermolecular

orbital overlap. On the other hand, the Förster mechanism operates at larger distances and requires an overlap of the emission spectrum of the donor with the absorption spectra of the acceptor, this is known as FRET. The extent of energy transfer is calculated based on the distance between the donor and acceptor, and also on the extent of their spectral overlap. The rate of energy transfer $k_T(r)$ is given by

$$k_T(r) = \frac{1}{\tau_D} \left(\frac{R_0}{r}\right)^6 \quad (7)$$

where the Förster distance is (R_0), r is the distance between the donor (D) and acceptor (A) and τ_D is the lifetime of the donor in the absence of energy transfer. The efficiency of energy transfer for a single donor–acceptor pair at a fixed distance is

$$E = \frac{R_0^6}{R_0^6 + r^6} \quad (8)$$

The exquisite capability of FRET has been widely exploited in studying protein folding, protein-protein interaction and cellular signaling events in living cells (Ma et al., 2017; Michalet et al., 2006; Prinz et al., 2008; Schmid and Birbach, 2007; Schuler and Eaton, 2008). FRET has also been used to illustrate folding and dynamics of RNA (Bernat and Disney, 2015; Chung et al., 2017; Li et al., 2008; Xu et al., 2016).

ESIPT (excited state intramolecular proton transfer) is a process of proton transfer between a proton acceptor and a proton donor group in the same photoexcited molecule (Chevalier et al., 2013; Demchenko et al., 2013; Sengupta and Kasha, 1979). This is a commonly found process that is associated with ratiometric probes such as 3-hydroxychromones (3HC). It is an excited state process that changes the structure of the excited-state fluorophore, typically an enol form that converts into a keto form. Specific to 3HC, the absorption of light changes its electron distribution in a 3-OH group and carbonyl group (Figure 19). A tautomer excited form T* has a lower energy than a normal excited state N* form. This favors the proton transfer from N* to T*, in a picosecond timescale. After generation of the N* and T* forms, both the excited state tautomers undergo relaxation to ground state, and back proton transfer (BPT) leads to the reconversion to the normal ground state (T → N).

The rate and equilibrium of ESIPT is dependent on the proton energy transfer barrier and the energy difference between the normal (N*) and tautomer (T*) forms of the excited

molecule. This equilibrium between the N* and T* form can be influenced by the solvent molecules that may favor one of the tautomers. ESIPT chromophores exhibit a large stokes shift. The two tautomer forms are thus well separated and provide a large spectral window convenient for spectroscopic measurements. However, 3HC may exhibit low fluorescence emission, depending on the charge and microenvironment of the fluorophore as well as the exposure to water. ESIPT has been used for a large range of applications, including the fabrication of optoelectronics solid state emitters (Padalkar and Seki, 2016), investigation of protein interactions and lipid membrane dynamics among many others (Klymchenko, 2017; Matos et al., 2010; Shvadchak et al., 2017; Zamotaiev et al., 2011).
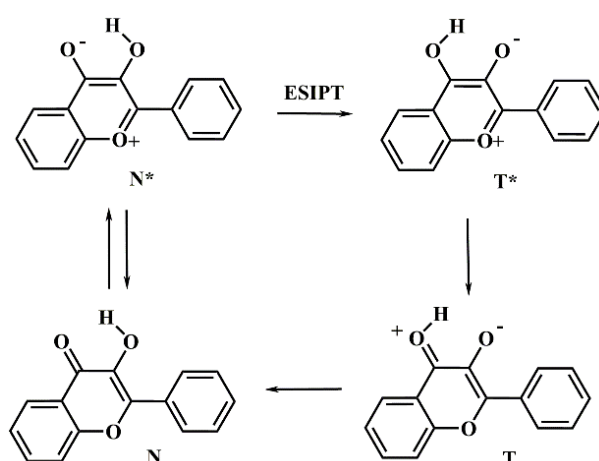
Figure 19. Representation of ESIPT reaction in 3HC.

1.3.2. Introduction to Nucleic Acids

1.3.2.1. Nucleic Acid Structures

DeoxyriboNucleic Acid (DNA) and and RiboNucleic Acid (RNA) are the two major types of nucleic acids. They are natural macromolecular polymers composed of deoxynucleotides or nucleotides as the building monomers. A (deoxy)nucleotide is a (deoxy)nucleoside attached to one or more phosphates. The phosphate is the connecting unit of DNA and RNA polymers. A nucleoside itself is composed of a sugar and a nucleobase. Numbering of the base and (deoxy)ribose is recalled in Figure 20. The carbon and nitrogen of the base ring are numbered normally (1, 2, 3 ...) while those of the (deoxy)ribose, are identified by the prime exponent (1', 2', 3 '...). For DNA, the sugar is the 2'-deoxyribose connected with one of the four canonical nucleobases (Figure 20b). Nucleobases are planar aromatic heterocycles divided into two groups: purines for adenine (A) and guanine (G); and pyrimidines for thymine (T) and cytosine (C) (Figure 20a). Epigenetic marks such as 5-methylcytosine and

5-hydroxymethylcytosine are also found into DNA. In RNA the sugar keeps the hydroxyl group at the 2' position of the ribose. RNA is mainly composed of A, G, C and uracil (U) instead of T but can contain also some minor nucleosides as for instance cytosine, pseudo-uridine and modified nucleosides (e.g. N6-methyladenine). The N9 of purines and N1 of pyrimidines link the nucleobase to the pentose at position C1'. This position is commonly called the anomeric position. Its configuration is beta, meaning that both the nucleobase and the CH$_2$OH at C-4' are on the same side and above the sugar plane.
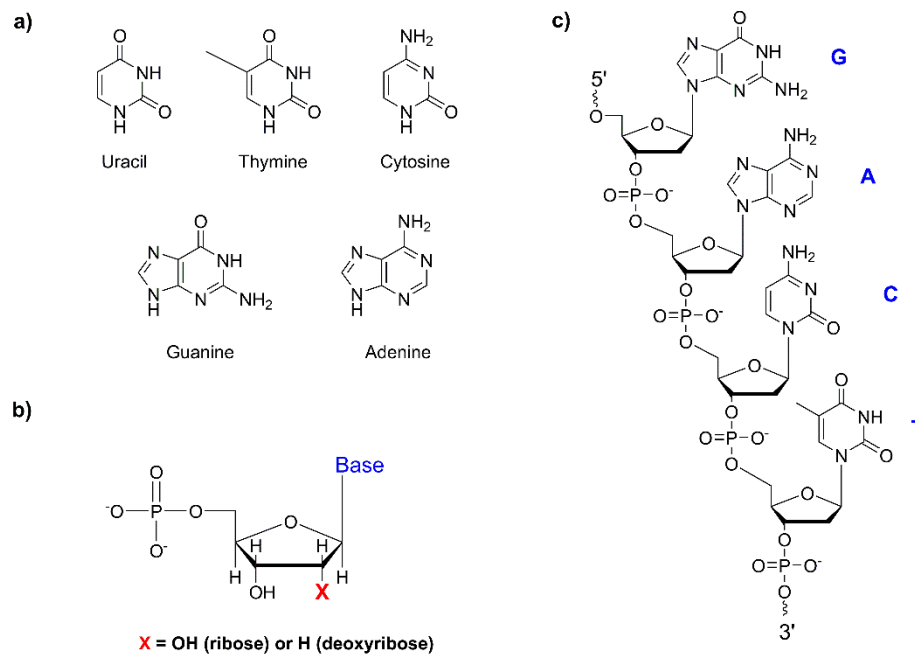
Figure 20. a) Structures of purine and pyrimidine nucleobases. b) Structure of ribose and deoxyribose sugar. c) Linear structure of the DNA fragment 5'-GACT

Phosphate diester linkages at the 3' and 5' hydroxyls of the (deoxy)nucleoside sugars give rise to the DNA/RNA linear chain. The primary structure of DNA/RNA consists of a linear sequence of (deoxy)nucleotides. Sequences are presented from the 5' to 3' and are determined by a series of letters designing the nucleobase composition (e.g. GACT and GACU for DNA and RNA, respectively) (Figure 20c).

The secondary structure of DNA/RNA is the result of non-covalent interactions between bases. In DNA double helix, the two strands of DNA are held together by specific hydrogen bonds as defined by Watson-Crick (Watson, J. D.; Crick, 1953): A base pairs with T (A•T), and G with C (G•C). The two base pairs are almost identical in dimensions. A•T has two hydrogen bonds while G•C has three hydrogen bonds (Figure 21a). The additional hydrogen bond imparts extra stability to the G•C base pair. The free energy of both base pairs

in aqueous solution is −4.3 and −5.8 kcal/mol for A•T and G•C base pairs, respectively (Stofer et al., 1999). Base stacking are additional non-covalent interactions contributing to DNA/RNA folding. In water, these hydrophobic and electrostatic interactions between adjacent bases are major contributors to the double strand stability.

a)

Adenine (anti)   Thymine (anti)

Guanine (anti)   Cytosine (anti)

b)

Adenine (syn)   Thymine (anti)

Guanine (anti)   Cytosine (anti)
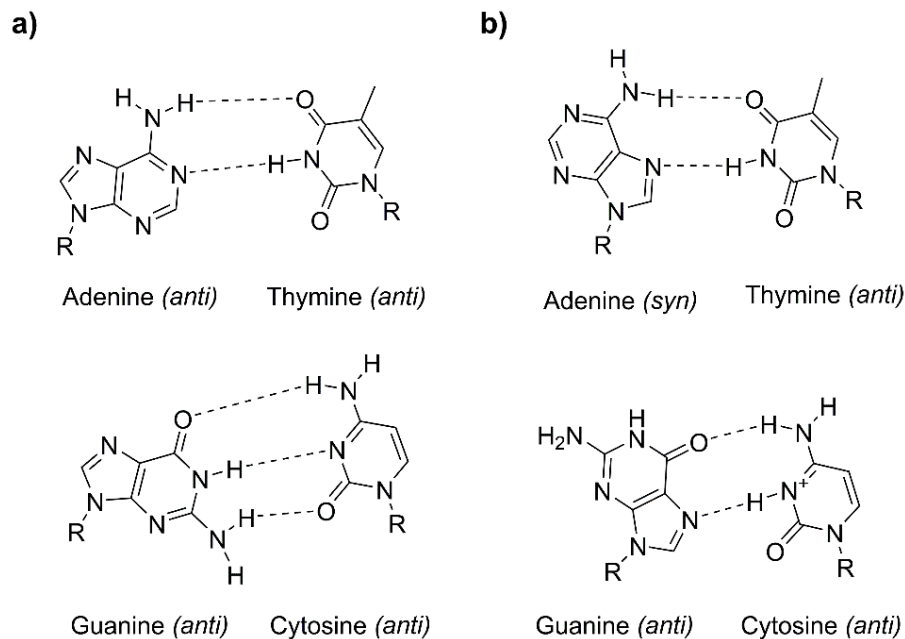
Figure 21. Watson–Crick hydrogen bonding: (a) A•T and G•C; Hoogsteen hydrogen bonding: A•T and G•C base pairs (b).

RNA can fold between complementarity sequences as for DNA. DNA/RNA can fold into various non-canonical secondary structures including triplex, G-quadruplex, i-motif, cruciform, Holliday junction, hairpin, and other non-canonical DNA/RNA forms. These structures involve one to four strands of nucleic acids stepping away from the traditional Watson-Crick base pairing, non-Watson-Crick base pairs can also be observed. Karl Hoogsteen described in 1963 the first example of a base pair involving a different edge of the adenine in interactions with thymine: the N6 and N7 atoms instead of the N1 and N6 atoms of the Watson-Crick base pair (Figure 21b). Later G was found to similarly base pair with a protonated C. These two base pairs were defined as the Hoogsteen base pairs. Since this seminal work, various other modes of pairing were discovered and listed. Non Watson-Crick base pairing are commonly involved in RNA folding and account for about 40% of structured RNA (Leontis and Westhof, 2001)

Folding of the linear chain(s) into a 3-D shape defines the tertiary structure of the nucleic acids and brings a supplementary degree of complexity. The different forms adopted

by DNA originate mainly from the orientation of the base with respect to the sugar (*anti* or *syn*) and the conformation of the sugar (C2'-*endo* or C3'-*endo*) (Figure 22a) (Saenger, 1984). The *anti* conformation is the most stable conformation and allows the base to establish Watson-Crick hydrogen bonds with the complementary base. On the other hand, the *syn* orientation of the base is required to establish Hoogsteen interactions and other non-canonical interactions. For purine nucleoside, the *syn* conformation is energetically close to the *anti* conformation and equilibrates quickly with the *anti* conformation at room temperature in solution. By contrast, the *syn* conformation is unfavorable for pyrimidine bases because of steric contacts between the O2 atom of the base and the CH_2-OH group of the sugar.
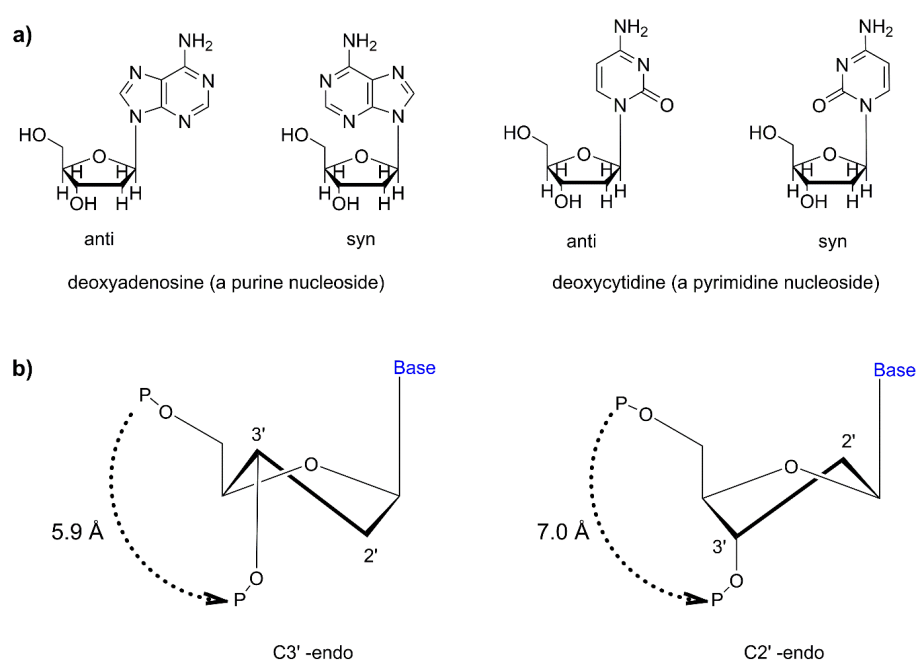
Figure 22. Anti- and syn- conformation of the nucleoside. B) C3'-endo and C2'-endo conformations of the pentose.

The two types of sugar pucker most commonly found in nucleic acids are the C3'-*endo* and C2'-*endo* conformations (Figure 22b). They are defined when the C3' and C2' are placed upwards of the mean plane of the sugar ring. C3'-*endo* is the sugar conformation prevalent in RNA double helix while C2'-*endo* is the one in canonical B double stranded DNA. The shorter phosphate-phosphate distance of C3'-*endo* compared to C3'-*exo* results in a more compact double helix (Figure 22).

The tertiary arrangement of a DNA helix is of threefold and defines the A, B and Z DNA forms (Table 1 and Figure 23). They are classical examples to illustrate structural polymorphism of DNA. They differ in all the structural parameters:

a. Handedness - right or left

b. Length of the helix turn

c. Axis and diameter of the helix

d. Number of base pairs per turn

e. Depth of the grooves

f. Intrastrand phosphate-phosphate distance
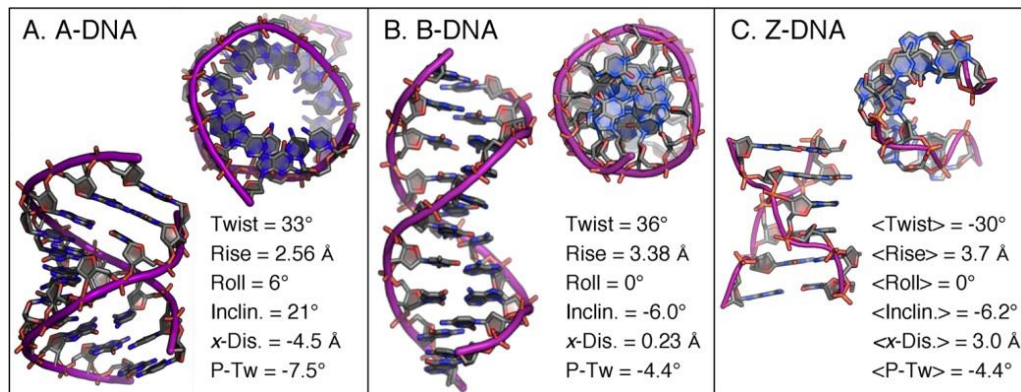
Figure 23. Representative double-helical structures of A, B and Z DNAs.

The most common form under physiological conditions is form B. The two chains run in antiparallel direction around a common axis to form a right-handed helix. The purine and pyrimidine bases of both polynucleotide chains face towards the inner core of the helix, while the sugar and phosphate groups are on the outside of the coil. The propeller has 10 nucleotides per turn, all of C2'-*endo* / *Anti* conformations. The bases are located in a plane perpendicular to the axis of the DNA helix. Two deep grooves are distinguished as major and minor.

Form A is sometimes observed in some regions of DNA especially in presence of high concentration of Mg^{2+} or under dehydrating conditions. It has physiological functions and is observed in several DNA / protein complexes. It has 11 nucleotides per turn of the helix, all conformations are C3'-*endo* / *Anti*. Compared to B DNA, it is shorter and wider than B DNA, the major groove being narrower and deeper while the minor one is broader and shallow.

As for the Z-shape, for zigzag, Z DNA is observed in DNA sequences rich in G•C base pairs. Its biological function is unclear. The unstable Z-DNA form prevails at high salt concentrations and in the presence of polyvalent metal ions. Z-DNA has a left handed helix and a single deep groove resulting from the alternation of purine (C3'-*endo* / *syn*) and pyrimidine (C2'-*endo* / *anti*) nucleotides (Shing and Carter, 2011). The Z helix is more elongated and narrower than A and B helixes.

Table 1. Average structural parameter for the 3 helical forms of DNA.

Parameter		A-DNA	B-DNA	Z-DNA
Helix handedness		right	right	Left
Residues per turn		11	10.5	12
Rotation per residue (°)		32.7	36	-10, -50
Axial rise [Å]		2.9	3.4	-3.9, -3.5
Helix pitch (°)		32	34	45
Base pair inclination (°)		12	2.4	-6.2
Diameter of helix [Å]		23	20	18
Base orientation	dA,dT,dC	anti	anti	anti
	dG	anti	anti	syn
Sugar pucker	dA,dT,dC	C3'-endo	C2'-endo	C2'-endo
	dG	C3'-endo	C2'-endo	C3'-endo
Intrastrand phosphate-	dA,dT,dC	5.9	7.0	7.0
phosphate distance [Å]	dG	5.9	7.0	5.9

The quaternary structure is basically the high-level organization of DNA into chromosomes.

1.3.2.2.   Non-canonical structure of nucleic acids

A number of non-canonical structures of DNA/RNA were reported. Polymorphism of DNA (Figure 24) extend from a simple mismatch to more complicated structures like quadraplexes and iMotifs, that take part in cellular processes like replication, transcription, recombination and repair. Some of these structures are highlighted below.

Cruciform structure formation requires regular inverted repeats of 6 or more nucleotides in the DNA sequence. Formation of cruciform is dependent on various factors like activation energy, salt concentration and temperature. They are involved in DNA supercoiling, positioning of the nucleosomes, replication and regulation of gene expression (Brázda et al., 2011).

Hairpin arise from single-strand DNA or RNA sequences having complementary sequences at the two ends of the sequence, which hybridize to form the stem, while the non-hybridizing residues in the middle of the sequences form the loop. Depending on the salt concentration, sugar moiety and oligomer concentration, they can exist in equilibrium with a duplex structure. Hence, they are molded either like a cruciform structure is formed or from a single strand DNA (ssDNA) which is produced during various cellular process like replication (Bikard et al., 2010). Thermodynamically, hairpins with long stems are genetically unstable, and play crucial roles in gene expression, recombination and transcription.
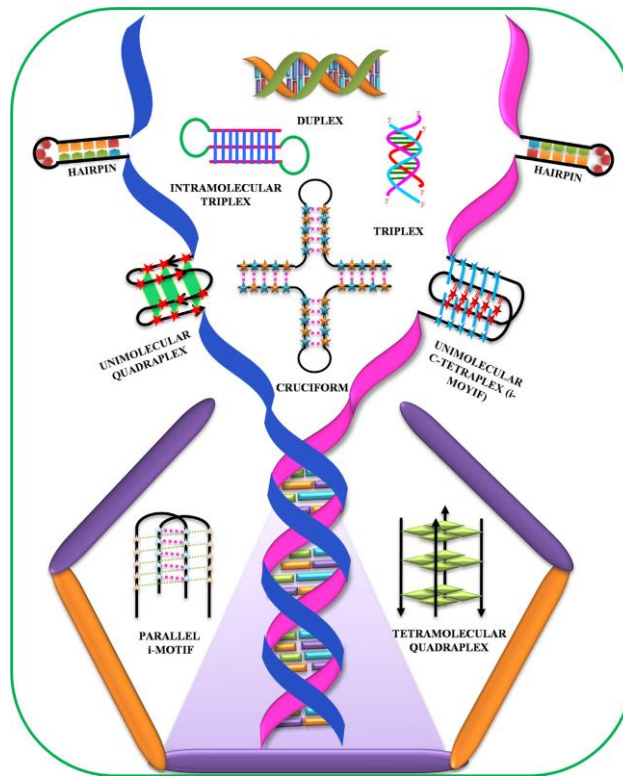
Figure 24. Bouquet of non-canonical structures. Adapted from (Kaushik et al., 2016)

Bubble and Bulges appear in DNA duplexes in the form of unpaired nucleotides on one strand. They are located within one strand or on both strands (internal loop) or at a junction. These structures are prone to protein interactions. Moreover, they appear in bent DNA to release the strain and thereby enhance the process of wrapping DNA around histones (Grinevich et al., 2015).

Bending is required to achieve the structural compactness of DNA/RNA. The formation of bends is driven through protein binding or specific sequences. DNA packaging into the nucleosomes require 147 base pairs of DNA wrapped around a histone octamer and it is achieved through DNA bending (Crick and Klug, 1975).

Triplex is formed when a single strand triplex forming oligonucleotide (TFO) recognizes an oligopurine•oligopyrimidine duplex and binds itself along the axis in a specific manner through Hoogsteen or reverse Hoogsteen hydrogen bonds (Felsenfeld et al., 1957). However, the formation of triplexes is dependent on factors such as the length of the oligonucleotide, base composition, divalent cation, pH and temperature.

iMotifs are four stranded, cytosine rich, DNA secondary structures that are comprised of two parallel-stranded DNA duplexes held together in an antiparallel orientation by

intercalated, cytosine–cytosine$^+$ base pairs (Day et al., 2014). These structures are stabilized by acidic pH and are used in the field of nanotechnology as pH switches.

Guanine quadruplexes, discovered in 1962 by Gellert et al, are formed based on the cyclic arrangement of G-tetrads through Hoogsteen hydrogen bonding (Gellert et al., 1962; Kaushik et al., 2016). These are highly stable structures due to the stacking of the hydrophobic quartets upon one another, hydrogen bonding within each quartet and coordination of monovalent counterions (Na^+, K^+) (Kaushik et al., 2016). The classification of H-quadruplexes can be based on its trans stoichiometry (uni-, bi- and tetramolecular), orientation (parallel, anti-parallel and mixed), shapes (chair or basket) and loops (lateral, propeller, diagonal, V-shaped) (Karsisiotis et al., 2013).

1.3.3. Fluorescence Nucleobase Analogues

The naturally occurring nucleobases are practically non-emissive in neutral aqueous conditions, having a fluorescence quantum yield less than 3×10^{-4} along with a sub-picosecond excited state lifetime (Leonard and Tolman, 1975; Sinkeldam et al., 2010). It is not surprising, because Nature preserves the genetic material by making them non-reactive, thus even if the bases are photoexcited they rapidly decay to ground state (Serrano-Andrés et al., 2006).

Thus, in order to fluorescently monitor biomolecular interactions involving nucleic acids, a reporter fluorophore has to be introduced in nucleic acids either by covalently attaching a fluorophore to nucleobases or designing a fluorescent nucleobase. A timeline displaying the developments made in fluorescent nucleobase analogues is presented in Figure 25. Fluorescent labeling of nucleic bases can be achieved through external or internal modifications. Modifying nucleic acids by covalently attaching fluorophores to the backbone at the end of or within an oligonucleotide sequence, but outside the actual base stack, is referred to *external modification* (Wilhelmsson, 2010). The attached external fluorophores such as fluorescein and rhodamine derivatives, Alexa dyes and Cy dyes have very high molar absorptivities and high fluorescence quantum yield. Consequently, the obtained photostable and highly bright fluorescent nucleic acids are useful in gel electrophoresis experiments and different microscopy techniques. *Internal modification* relates to the covalent attachment of a fluorophore replacing a specific nucleobase. These planar aromatic fluorophores are called fluorescent nucleobase analogues (Wilhelmsson, 2010; Wilson and Kool, 2006). These fluorophores are meant to preserve the DNA/RNA structure whereas, they may or not have the possibility to establish a Watson-crick base pairing. Fluorescent nucleobase analogues are used at specific positions in order to track

preferential binding sites of biomolecules. In comparison to external modification, these fluorophores are less bright, have the possibility of being environment sensitive, and are prone to change the folding of DNA/RNA structure. Different classes of fluorescent nucleoside analogues are described below.
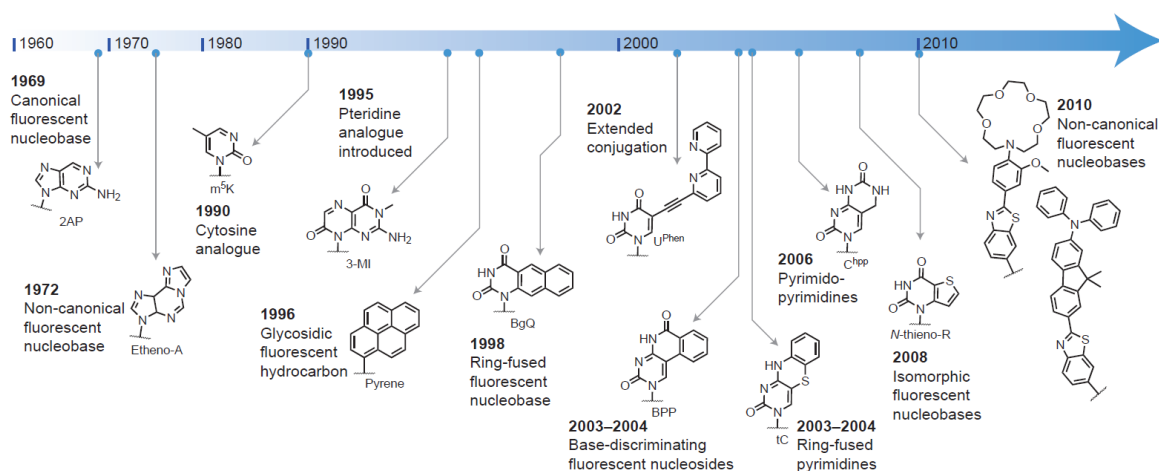
Figure 25. Timeline of fluorescent nucleobase development. In 1969, Stryer reported the first canonical fluorescent nucleobases, featuring 2Ap and formycin; in 1972, Leonard synthesized the first non-canonical fluorescent nucleobase, etheno-dA. The fluorescent nucleobase family gradually expanded during the following two decades, including m5K reported in 1990 by McLaughlin and pteridines by Hawkins in 1995. C-glycosidic nucleobases with hydrocarbon fluorophores directly attached to the sugar were first reported by Kool in 1996. Since 2000, the number of fluorescent nucleobases has increased markedly, featuring notable examples such as BPP by Saito and Okamoto, tC by Wilhelmsson, dUPhen (Hocek) and thieno-appended analogues of Tor. Dozens of non-canonical nucleobases have been introduced since 2010. Adapted from (Xu et al., 2017).

Chromophoric base analogues are fluorescent analogues that replace natural nucleobases by polycyclic aromatic hydrocarbons (PAH) which do not form Watson-Crick base pairs (Figure 26). On the bright side, they have close to unity fluorescence quantum yield and an isolated absorption band (> 345 nm) facilitating selective excitation in the presence of the natural nucleobases (Wilson and Kool, 2006). The applications of such fluorophores are well defined, such as PAH that has been used to investigate enzyme-substrate recognition (Wilson and Kool, 2006). The photophysics of the probe is sensitive to the neighboring bases and pH, hence depending on the environment, different quantum yield is obtained. Another example is the coumarin 120-containing nucleoside analogues that are designed to pair with an

abasic site in DNA (Sakata et al., 2011; Sen et al., 2005). They have been used to explore the DNA dynamics (Andreatta et al., 2005; Berg et al., 2008) and electron transfer in DNA (Grigorenko and Leumann, 2008; Properties, 2009).
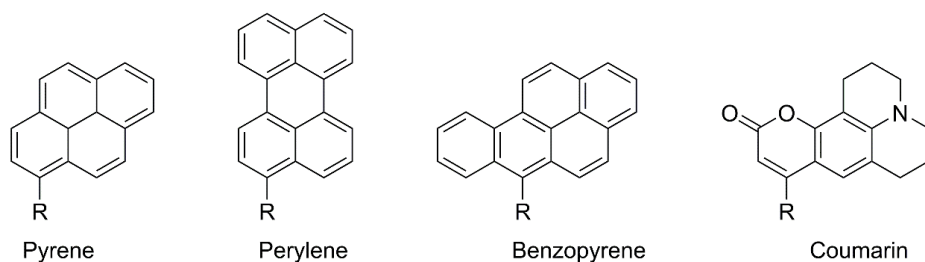
Figure 26. Selected examples of chromophoric base analogues, where R is 2′-deoxyribose.

Pteridines are the naturally occurring and widely used fluorophore that are mostly developed by the Hawkins and co-workers (Hawkins, 2008a, 2001) (Figure 27). They absorb in the blue region (300 nm) and emit in the visible region (~430 nm), also have an intense fluorescence quantum yield (0.39–0.88). The most prominent G (3-MI and 6-MI) and A (6-MAP and DMAP) analogues are shown in Figure 27. These are environment sensitive probe, hence when incorporated in oligonucleotide they show varying quantum yield which is dependent on the nature of the neighboring residues. The drawback of these residues is that they slightly hampers the stability of the oligonucleotide duplex structure which was observed due to the deviation in melting temperature of the duplex structures (Hawkins, 2001). Pteridines are successfully utilized for several applications such as, single molecule detection purpose (3-MI) (Hawkins, 2008b); to study the mechanism of the RecA-mediated DNA strand exchange (6-MI) (Roca and Singleton, 2003); premelting transition studies of DNA A-tracts (Augustyn et al., 2006) and as a probe for base flipping by DNA photolyase (Wilhelmsson, 2010; Yang and Stanley, 2006).
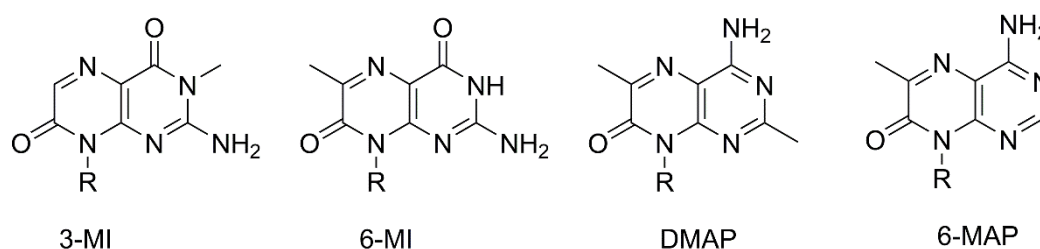
Figure 27. Selected examples of pteridines (R) 2′-deoxyribose or ribose). (MI = methyl isoxanthopterin; DMAP = dimethylaminopteridine)

Nucleosides containing expanded nucleobases are obtained by adding aromatic rings onto the purines or pyrimidines, which results in an extended conjugation with natural nucleobases (Figure 28). Having an expanded structure basically does not impair the hydrogen bonding face of the nucleobase, but the increased size of the nucleobase might obstruct the resulting oligonucleotide. The most beneficial aspect is the improvement of the photophysical properties. With the expanded base, the absorption becomes red-shifted compared to their natural counterpart and so does the emission band. This also results in favorable increment in quantum yield and brightness. One of the crucial fluorophore that proved useful for single nucleotide polymorphism (SNP) analysis was developed by Saito and co-workers (Okamoto et al., 2003a), and was called base discriminating fluorescent nucleoside. These probes were either having an extended base or an expanded base. Benzopyridopyrimidine (BPP) and naphthopyridopyrimidine (NPP) are examples of cytosine analogues that forms wobble base pairs with A, and Watson-Crick base pair with G, and can report SNP with a respectable quantum yield of ~ 0.1–0.3 (Okamoto et al., 2003a). Later Saito and co-worker synthesized analogues of pyrimidine-discriminating fluorescent nucleotides (Okamoto et al., 2003a; Sinkeldam et al., 2010).
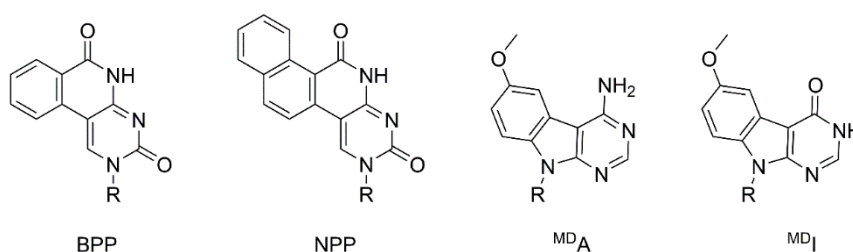
Figure 28. Expanded nucleobase analogues (R = 2′-deoxyribose).

Nucleosides containing extended nucleobases are obtained by extending purines and pyrimidines by conjugating them with known fluorescent moieties (Figure 29). The resulting chromophores possess enhanced photophysical properties resembling to the parent fluorophore. Basically, the two partners, purine/pyrimidyne and fluorophore, are joined together via non-conjugating linkers. As an example of such system, a xanthene-type fluorophore was attached to dideoxynucleotides using an ethynyl linker, and was used for sequencing application (Thoresen et al., 2003; Whitcombe et al., 1999). In addition to the fluorophore-nucleobase linkage, they have also a quencher-nucleobase linkage that has been developed for molecular beacon and sequencing application (McKeen et al., 2003; Seela et al., 2001). Another application to monitor SNP with extended nucleosides has been shown by Hocek and co-workers (Vrábel et al., 2007). They used a system in which the pyrimidine and

purine analogues were conjugated with bipyridine, terpyridine, and phenanthroline moieties using phenyl and ethynyl linkages (Vrábel et al., 2008).

Recently, A. Burger and coworkers developed 2-thienyl-3-hydroxychromone (3HCnt), an extended fluorescent nucleobase analogue that comes from the family of 3-hydroxychromones (Figure 29). 3HCnt acts as a universal nucleoside analogue able to substitute any of the natural nucleobases (Dziuba et al., 2012). 3HCnt is up to 50-fold brighter than 2Ap and exhibits a dual emission as a result of an excited-state intramolecular proton transfer (ESIPT) reaction (Demchenko et al., 2013; Sengupta and Kasha, 1979) between the initially excited normal form (N*) and its tautomer (T*). The ratio between the two bands is sensitive to the polarity and H-bond donor ability of the probe microenvironment. The photophysical properties of this probe as substituted in oligonucleotides are discussed in section 4.2 of results and discussion.
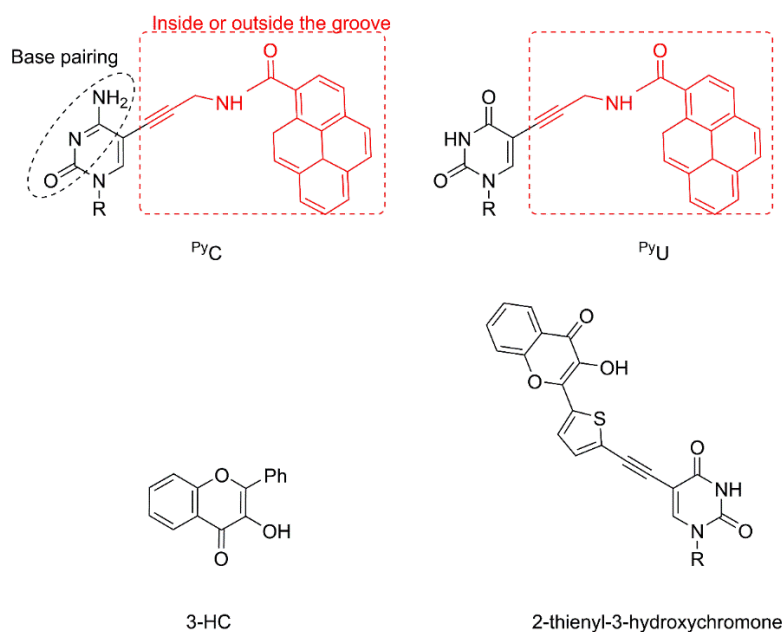
Figure 29. Extended nucleobase analogues (R = 2′-deoxyribose).

Isomorphic Nucleobases are the most advantageous fluorescent nucleobases, but they are difficult to design. They are heterocyclic compounds resembling to their native nucleobases; structurally in shape and size, and functionally in Watson-Crick base pairing. The clear advantage of these analogues is their structural similarity to natural nucleosides that portray minimal perturbation upon substitution.

Since last five decades, 2-Aminopurine (2Ap) has become the most widely used fluorescent nucleoside, and is thus regarded as the 'Gold standard' (Ward et al., 1969). It is an analogue of adenine, with enhanced photophysical properties. As a free probe it has isolated

absorption (303 nm), minimal sensitivity to pH change and a high quantum yield of 0.68 in water (Ward et al., 1969). Upon incorporation in oligonucleotides, it is quenched through neighboring residues, resulting in a strong drop in quantum yield (0.02–0.1). 2Ap establishes stable Watson-crick base pairs with T/U and in some cases with C (Nordlund et al., 1989; Sowers et al., 2000, 1987). The quenching of 2Ap's fluorescence upon incorporation in oligonucleotides has been exploited to develop a number of fluorescence based assays. Theoretical and experimental approaches have investigated the origin of its unique photophysical properties (Jean and Hall, 2002; Lobsiger et al., 2014; Nir et al., 2001; Serrano-Andrés et al., 2006)
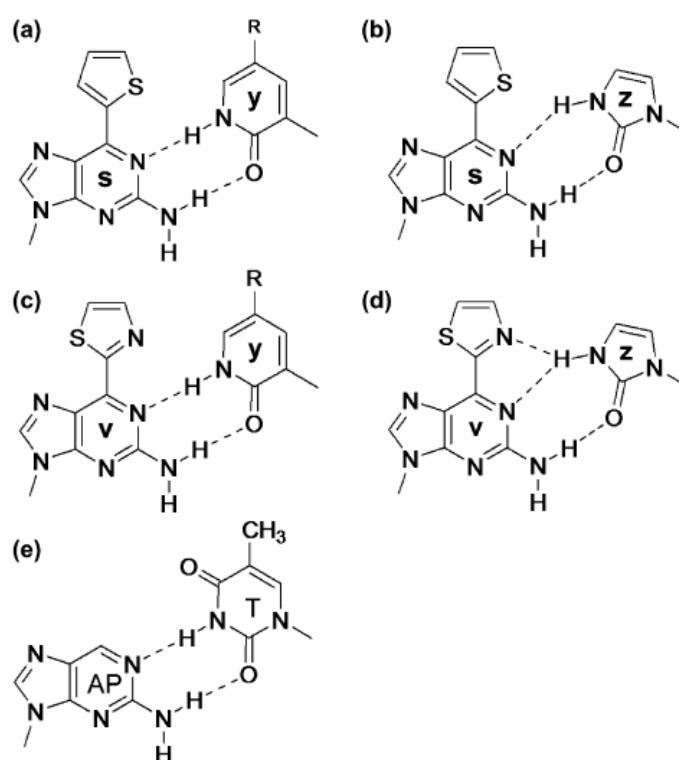
Figure 30. Unnatural base pairs: (a) **s–y**, (b) **s–z**, (c) **v–y**, (d) **v–z**, and (e) the **AP–T** base pair. 2-amino-6-(2-thienyl)purine (**s**) and 2-oxopyridine (**y**), 2-amino-6-(2-thiazolyl)purine (**v**) and **y**, and **s** and imidazolin-2-one (**z**) (Mitsui et al., 2007).

A unique application of 2Ap derivatives was demonstrated by Hirao and co-worker who designed (Figure 30) (Hirao et al., 2002; Kimoto et al., 2007; Morohashi et al., 2012) purine analogues, like 2-amino-6-(2-thienyl)purine (**s**) and 2-amino-6-(2-thiazolyl)purine (**y**), that form an unnatural base pair. The photophysical characterization of 2-amino-6-(2-thienyl)purine as substituted in oligonucleotide have been discussed in section 8 of Appendix. These derivatives of 2Ap have red-shifted absorption band in comparison to 2Ap (~355 nm),

and a strong emission ($\Phi \approx 0.4$) in the visible range ($\sim$450 nm) (Mitsui et al., 2007). These unnatural base pairs are an expansion to the genetic alphabets which provide platform for the site-specific, enzymatic incorporation of extra, functional components into nucleic acids (Hirao et al., 2012). Functionalized unnatural base pairs having fluorescent dyes are successfully incorporated into the nucleic acids by polymerase. Furthermore, a modified unnatural base pair with a combination of fluorophore-quencher have been developed for their function in PCR amplification and their applications as sensing and diagnostic tools (Kimoto et al., 2011).
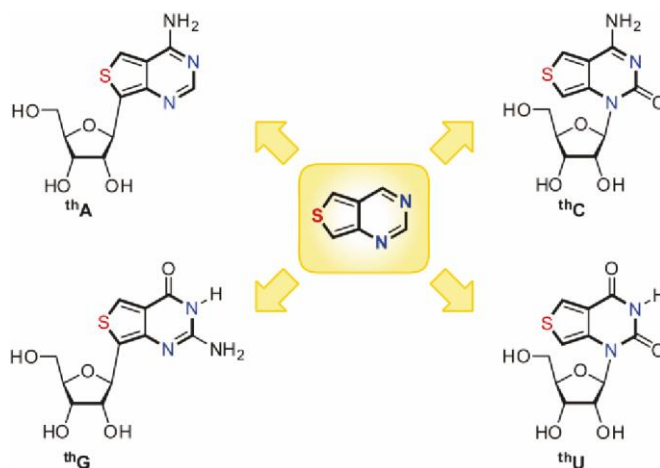
Figure 31. Emissive RNA Alphabet. Adapted from (Shin et al., 2011).

The main limitation with 2Ap and their derivatives is their low quantum yield when incorporated in oligonucleotides. Tor and coworkers succeeded to overcome this limitation by designing a series of thieno-appended analogues (Shin et al., 2011) (Figure 31). Our interest was focused on the guanine analogue, called as Thieno[3,4-d]-guanosine (thG). As a free probe, in comparison to 2Ap, it exhibits a slightly lower quantum yield of 0.48, but an improved red-shifted absorption maximum and an emission centered around 460 nm. When incorporated in oligonucleotides, thG was found to perfectly replace G residues and form stable Watson-Crick base pairs. In addition, thG shows high quantum yields and rather simple fluorescence decays (Sholokh et al., 2015). thG can be incorporated in oligonucleotides either by solid phase synthesis or enzymatic syntheses. The T7 RNA polymerase can initiate and maintain the transcription mechanism with thG nucleobases, which results in the formation of fully modified and highly emissive RNA transcripts in which all the natural guanosines are replaced by the thG analogue (McCoy et al., 2014). Tor and co-workers also designed further improved isomorphic nucleoside analogues derived from the isothiazolo[4,3- d]pyrimidine (^{tz}X, X= nucleobase) core that has a basic nitrogen corresponding to N7 of the purine core (Rovira et al., 2015). They are even closer to the native nucleobases and appear useful for monitoring

biomolecular interactions that depend upon the basicity and coordinating ability of this basic nitrogen. However, in comparison to the [th]G-analogues, the [tz]G-analogues have somewhat weaker photophysical properties, having more blue-shifted absorption and emission profile, and somewhat lower quantum yields. The [tz]G-analogues were notably used to monitor the deamination activity of adenosine deaminase.

1.3.4. Emerging Applications of Fluorescent Nucleobase Analogues

Fluorescent nucleobase analogues have been employed in a large number of chemical, structural, biophysical and biochemical implementations, including single nucleotide polymorphism detection (Okamoto et al., 2003b, 2003c), microenvironment monitoring (Brauns et al., 1999; Jeong et al., 2009), and monitoring structural and morphological changes (Börjesson et al., 2009; Godde et al., 2000; Hwang et al., 2005), as well as enzyme activity (Kirk et al., 2001; Raney et al., 1994; Sandin et al., 2009; Xu et al., 2017). A recent example illustrating the potency of the fluorescent nucleobases is the application of [th]A to monitor the activity of adenosine-to-inosine RNA editing adenosine deaminase acting on RNA (ADAR) enzymes (McCoy et al., 2014; Rovira et al., 2015; Xu et al., 2017). As this fluorescent nucleobase fits into the enzyme pocket, it is not only able to monitor the deamination process but also to probe the enzyme editing site (Rovira et al., 2015; Xu et al., 2017).
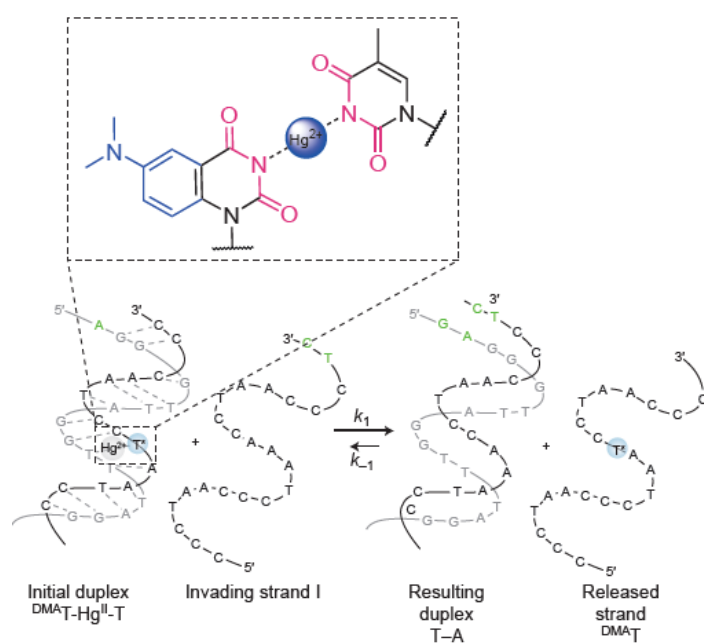
Figure 32. Kinetic and thermodynamic investigation of the effects of mercury on DNA metabolism. The fluorescent thymine can chelate mercury with another thymine ring and link DNA strands. This can be used to probe mercury metabolism in vivo and to

study the effects of mercury on DNA status. k_1, strand displacement rate constant;k_{-1}, reverse reaction rate constant. Adapted from (Xu et al., 2017)

Application of fluorescent nucleobase analogues are now observing a drift from *in vitro* towards *in vivo* applications. As negatively charged nucleobases are not taken up by the cells, they have to be tailored into neutral nucleosides. For example, the recently developed fluorescent thymidine analogue [DMA]T has been used to monitor in real-time the mercury-induced DNA metabolism *in vivo* (Schmidt et al., 2016) (Figure 32). The kinetic and thermodynamic parameters of Hg binding to T-T mismatches were monitored through fluorescent sensitive [DMA]T-A base pairs positioned few base pairs away. In another example, two fluorescence nucleoside analogue (uridine-furan and etheno-cytidine) were used as tools to study *in vivo* nucleoside transporter-related activity (Claudio-Montero et al., 2015). Human concentrative nucleoside transporters (hCNTs) transport most of the nucleoside derived drugs across the plasma membrane thus they are of important interest for predicting drug sensitivity. Both the fluorescent nucleoside analogues showed high binding affinity to the hCNTs and thus allowed visualization of transporters, through confocal microscopy and flow cytometry, in living cells and prediction of nucleoside-derived drug sensitivity.

2. Research Objectives

The objective of this research work is to characterize and target the interaction of HIV-1 nucleocapsid protein (NC) with its target nucleic acids (NA) using fluorescence nucleobase analogues as a tool to monitor these biomolecular interactions. Among the HIV proteins, the NC is projected a potential target for drug development because of its several key functions in the whole viral life cycle and its mutation intolerant nature. Due to the nucleic acid chaperoning property of NC, it effectively remodels the NA so that the most thermodynamically stable conformations are gained. NC exhibits also sequence specific binding to numerous defined single strand NA containing unpaired guanine residues that involve a dynamic interaction between the F16 residue of ZF1 and the W37 residue of ZF2 with unpaired guanines. Although, the structural and mechanistic aspects of these interactions have gained advances, its dynamics aspects are difficult to understand due to limitation of suitable tools.

In this context, the primary objective of the research work was to characterize the NC promoted annealing of (–)PBS with (+)PBS sequences as it plays critical role in the second strand transfer during reverse transcription. Till now, using fluorescence-based approach, vital insights have been gained into the annealing pathways, but they were not successful in providing the complete information regarding the rate constants associated with the reaction pathways. This is because the probes were either insensitive to the annealing of the loops or the poor brightness. Herein, we use new fluorescent nucleobase analogues in order to determine the full set of kinetic rate constants.

Second objective of the work is to target a unique interaction where NC binds in opposite direction to the nucleic acid chain in NC-RNA complexes as compared to NC-DNA complexes. Critical analysis of the 3D resolved structure made across several complexes of NC with nucleic acids sequences highlighted the role of sugar moiety to pivot the NC's binding direction. Herein, we use fluorescence based approach in order to observe whether the NC's opposite binding direction also follows within the RNA and DNA variant of the same nucleic acid model sequence (PBS and SL3). thG was site specifically incorporated by replacing the key residue in (–)PBS (G7) and SL3 (G10 and G12) sequences. Binding affinity of NC to each labeled and non-labeled ODNs were determined by using ITC and fluorescence titration. NC induced conformational changes in the probe were monitored through fluorescence lifetime changes. MD simulations provided structural insight in the binding mode adopted by NC for RNA and DNA.

Third objective is to participate in the high throughput screening of NC inhibitors and determine their IC_{50} through NCinh assay and to develop a site-specific assay for screening inhibitors able to compete with NC for its specific binding sites. This work is performed within the THINPAD (**T**argeting the **HI**V-1 **N**ucleocapsid **P**rotein to fight **A**ntiretroviral **D**rug Resistance) consortium that is organized from leading scientists in the field of NC. This project is supported by The European Union's Seventh Framework Programme (FP7) for Research and Innovation.

3. Materials and Methods

The reagents used in this work are of spectroscopic grade and purchased from Sigma-Aaldrich, Fluka, Carl Roth and Fischer scientific unless mentioned otherwise. Mass spectra were recorded on Bruker HCT Ultra and Agilent Technologies Accurate-Mass Q-TOF LC/MS 6520 mass spectrometers.

3.1.   Fluorescent Nucleobase Analogues

2-Aminopurine (2Ap): Free 2Ap and 2Ap-substituted in ODN were synthesized and purified by IBA GmbH Nucleic Acids Product Supply (Germany).

2-Amino-6-(2-thienyl)purine (s): The deoxyribonuceleoside **s** was synthesized by the research group of Hirao, RIKEN, Japan (Fujiwara et al., 2001) (Figure 33)
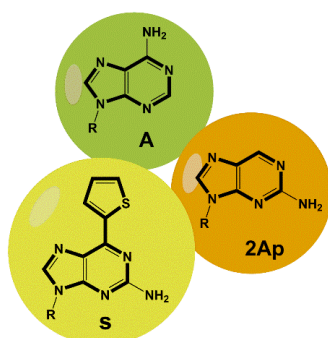
Figure 33 Chemical structures of A: Adenine, 2Ap: 2-Aminopurine and s: 2-Amino-6-(2-thienyl)purine.

Thienoguanosine (^{th}G): Thienoguanosine was synthesized by the group of Prof. Yitzhak Tor (UCSD) as previously described (Shin et al., 2011; Sholokh et al., 2015) (Figure 34).
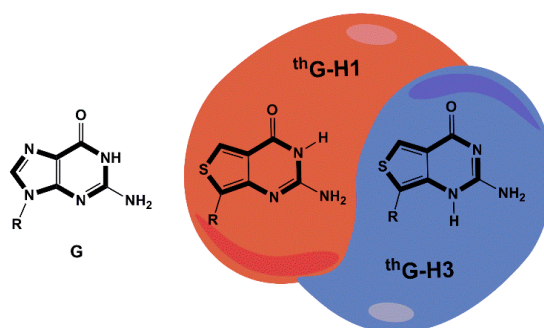
Figure 34 Chemical structures of G: Guanine; and the tautomers of Thienoguanine thG-H1 and thG-H3.

3.2.   Peptides

A. General Synthesis Method

The fundamental principle of this technique involves the incorporation of N-a-amino acids into a peptide of any desired sequence with one end of the sequence remaining attached to a solid support matrix. While the peptide is being synthesized usually by stepwise methods, all soluble reagents can be removed from the peptide-solid support matrix by filtration and washed away at the end of each coupling step. After the desired sequence of amino acids has been obtained, the peptide can be removed from the polymeric support.

- Synthesis

Solid state peptide synthesis was performed on a 433a peptide synthesizer (ABI, Foster city, CA). The synthesis was performed at a 0.1 mmole scale using the standard side-chain protected 9-fluorenylmethoxycarbonyl (Fmoc) amino acids and HBTU/HOBt coupling protocol. Fmoc-Asn(trt)-Wang resin LL (100–200 mesh) was used as a solid support. Deprotection steps were performed by piperidine and automatically controlled by UV absorbance at 301 nm. At the end of the synthesis, the peptidylresin was isolated, and washed with NMP, methanol and dichloromethane.

- Cleavage and Deprotection

Cleavage of the peptidylresin and deprotection were performed for 2 hours using 20 mL trifluoroacetic acid (TFA) solution containing water (5%, v/v), phenol (1%, v/v), thioanisole (5%, v/v), triisopropylsilane (2.5%, v/v) and ethanedithiol (2.5%, v/v). The peptide was precipitated using ice-cold diethyl ether and pelleted by centrifugation. The pellet was air dried for about 15 min, solubilized with aqueous TFA (0.05%, v/v) and lyophilized.

- Purification

Purification by HPLC was carried out on a C18 column (NucleosilVarioPrep 100A, 5μm; 250X10) in water/acetonitrile mixture containing 0.05% TFA. The labeled peptide was monitored at 230 nm and 370 nm (M3HFaa dye absorption). After purification, the fractions containing the pure peptide were combined and lyophilized.

B. Preparation of Zn-bound NC Peptides

The lyophilized peptides were dissolved in MilliQ water, at a concentration of approximatively 1 mg/mL. The concentration of stock solution was calculated by using the

extinction coefficient of $\varepsilon_{280} = 5700\ M^{-1}cm^{-1}$. As cysteines are rapidly oxidized, it is suggested to quickly estimate the concentration of peptide and then add 2.5 molar equivalents of $ZnSO_4$ to the stock solution. Later pH was raised to its final value by adding buffer. The stock solution was aliquoted (15 µL) in a low binding eppendorfs and stored at -20 °C.

C. Protein Activity Test

To check the activity of prepared peptide, we used cTAR sequence, a 55 nucleotide DNA hairpin, whose 5' and 3' ends are fluorescently labeled with Rhodamine6G and Dabcyl respectively. NC has the ability to destabilize the cTAR structure (Beltz et al., 2003; Bernacchi et al., 2002). In absence of NC, the cTAR hairpin induces a strong fluorescence quenching of the Rh6G emission by the Dabcyl group. When NC is added in 10-fold excess, it melts the bottom of the cTAR stem, thereby increasing the distance between the two fluorophores and increasing the fluorescence emission of Rh6G. Based on this destabilization mechanism, it is observed that after interaction with NC, fluorescence of Rh6G increases by 6-7 times.

3.3.   Oligonucleotides

Oligonucleotides used: (–)PBS_DNA, (–)PBS_RNA, (+)PBS_DNA, (+)PBS_RNA, 2Ap6(–)PBS DNA, 2Ap7(–)PBS DNA, SL3_RNA, SL3_DNA were synthesized and purified by IBA GmbH Nucleic Acids Product Supply (Germany) (Figure 35). Their stock solutions were prepared in MQ water and concentrations were calculated from their absorbance using the molar extinction coefficients at 260 nm provided by the supplier.
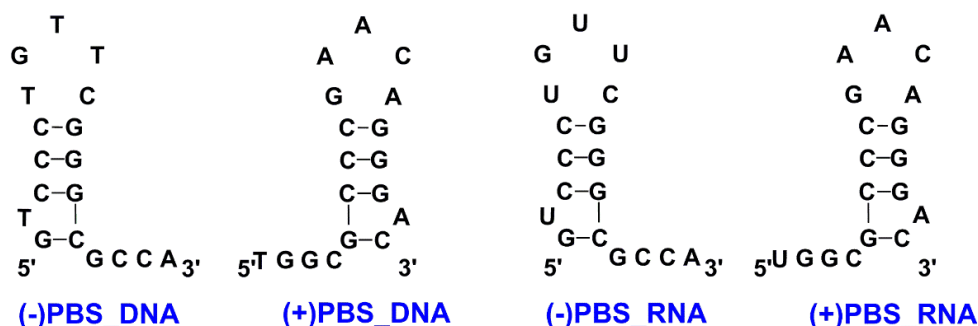
Figure 35 Oligonucleotides used in the thesis.

The (–)PBS oligonucleotides site specifically substituted by **s** at position 6 and 7 were synthesized with an Applied Biosystems 392 DNA synthesizer. The DNA fragments were purified with a Gilson HPLC system, using an analytical column (Synchropac RPP, 250 x 4.6 mm, Eichrom Technologies). The detailed protocol is described in (Hikida et al., 2010; Mitsui et al., 2007) (Figure 36).
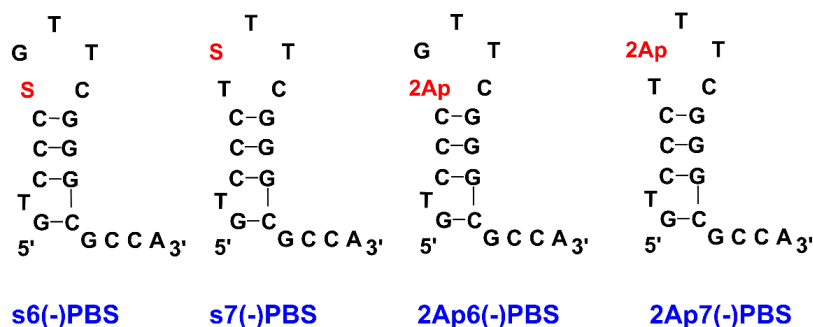
Figure 36 (–)PBS with **s** and 2Ap substituted at positions 6 and 7.

thG- substituted RNA and DNA copies of (–)PBS and SL3 sequences were synthesized by group of Prof. Tor (UCSD) as described in (Sholokh et al., 2015)(Figure 37).
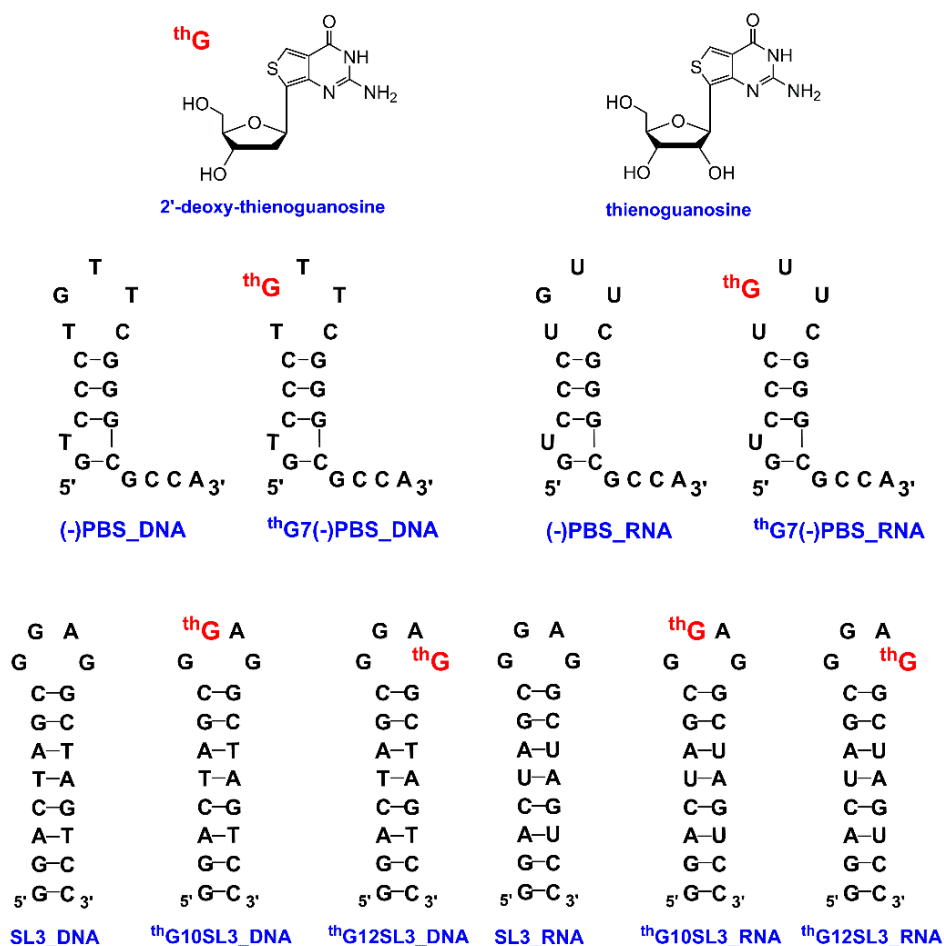
Figure 37 Oligonucleotides used in the thesis.

Double stranded oligonucleotides were prepared by the heat-cooling method, with water bath heating at 90 °C for 3 mins for concentrated solution of oligonucleotide and its complementary sequence in the ratio of 1:3 and then, cooling down to room temperature for 45–60 mins.

3.4.   UV/Vis Absorption and Fluorescence Spectroscopy.

UV spectrophotometry uses the absorbance of the light by an analyte at a certain wavelength to determine the analyte concentration. The component diagram of a typical spectrophotometer is shown below Figure 38.
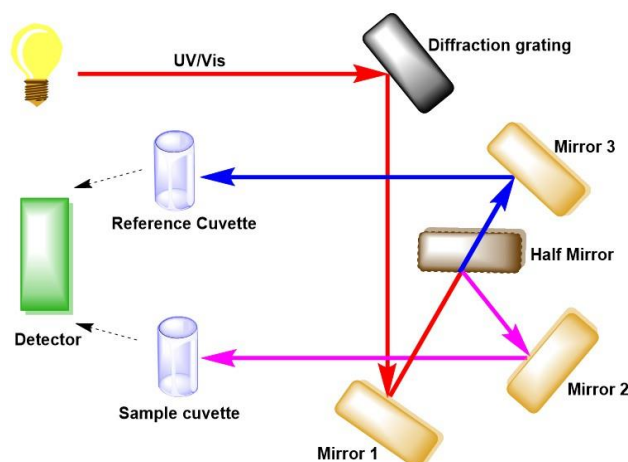
Figure 38 UV/Vis diagram.

Light from source (visible region: Tungsten lamp; UV region: Hydrogen/Deuterium lamp) of UV and/or visible gets separated into the component wavelengths by a prism or a diffraction grating. Each monochromatic beam in turn splits into two equal intensity beams by the half mirror. One beam, passes through the sample solution (*I*) whereas the other passes through the reference solvent cuvette (*I$_o$*). The intensities of both beams are measured by the detector. Absorbance was determined using the beer lambert equation:

$$A = -log_{10}\left(\frac{Io}{I}\right) = \varepsilon cl \qquad (9)$$

where *A* is the absorbance of light through a material of path length *l* and concentration *c*; *ε* is the extinction coefficient of the absorbing material, *I$_0$* and *I* correspond to the intensities of the incident and transmitted light. Absorption spectra were recorded on a Cary 400 or 4000 UV-visible spectrophotometer (Varian).

Thermal Denaturation Experiment

Melting temperatures were determined by measuring absorbance changes at 260 nm as a function of temperature using a Varian Cary 400 spectrophotometer equipped with a Peltier temperature controller. Absorbance was recorded in the forward and backward directions from 20 to 80 or 90 °C at a rate of 0.5 °C/min. Prior to the melting experiment, the double stranded

ODNs were prepared. Thermal denaturation experiments were performed either in 25 mM TRIS-HCl, 30 mM NaCl, 0.2 mM MgCl$_2$ pH 7.5, or in 10 mM cacodylate buffer, 150 mM NaCl pH 7.5. To avoid the evaporation of the samples they were overlaid with mineral oil (Sigma-Aldrich). Melting temperatures were determined from the first derivative of thermal denaturation curves

The excitation spectrum is defined as the fluorescent intensity measured as a function of excitation wavelength at a constant emission wavelength; the emission spectrum is the fluorescent intensity measured as a function of emission wavelength at a constant excitation wavelength (So and Dong, 2002). A typical fluorometer includes a light source (Xenon lamp), a specimen chamber with integrated optical components, and high sensitivity detectors. The optical paths of the excitation and the detection light paths are along the orthogonal axis. The orthogonal arrangement ensures minimal leakage of excitation light into the detection side. The component diagram shown below illustrates the principle of a fluorimeter Figure 39.
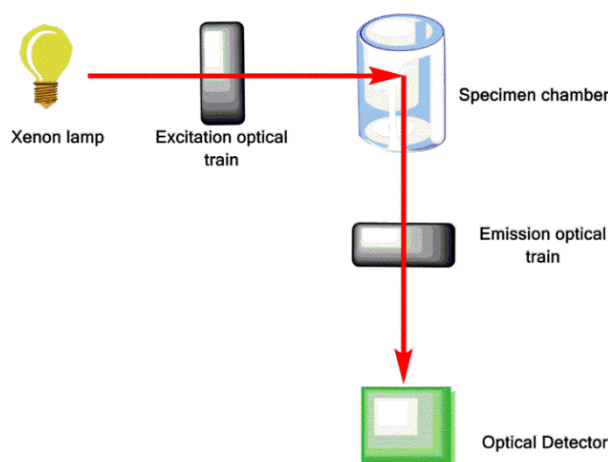
Figure 39 Fluorescence spectrofluorometer block diagram.

Fluorescence spectra were recorded on a FluoroMax 3 or 4 spectrofluorimeter (Horiba) equipped with a thermostated cell compartment at 20 ± 0.5 °C. Fluorescence spectra were corrected for Raman scattering, lamp fluctuations and instrumental wavelength dependent bias. Quantum yield of **s-** and [th]G-labeled sequences were calculated using quinine sulphate in 0.5 M sulphuric acid (Φ = 0.546) (Melhuish, 1961) or free 2-Aminopurine nucleoside (0.68)(Ward et al., 1969). Equation used for calculation fluorescence quantum yield

$$Fluorescence\ Quantum\ Yield = F.Q.Y._{Std}\ \left(\frac{Abs_{Std}^{\lambda} \times Intg.Intensity_x}{Abs_x^{\lambda} \times Intg.Intensity_{Std}} \right)\left(\frac{\eta_x}{\eta_{Std}} \right)$$

Where, $F.Q.Y._{Std}$ is the fluorescence quantum yield of the reference, Abs^{λ}_{Std} and Abs^{λ}_{x} is the absorbance of the standard and the sample at the characteristic wavelength, $Intg\ Intensity_{x}$ and $Intg\ Intensity_{std}$ is the integral intensity of the fluorescence spectra curves, while η_{x} and η_{Std} are the refractive index of the solution and the standard. All the fittings were carried using origin 8.6.

Photodegradation experiments

Photodegradation experiments of thG nucleoside and thG substituted in ODN were carried out in dark quartz micro-cuvettes with 50 μL. To check the photostability, the solutions of thG in 25 mM TRIS-HCl buffer (pH=7.5), 30 mM NaCl and 0.2 mM MgCl$_2$ were continuously illuminated at 325, 350, 360, 370 or 380 nm during 2000 s. The samples were illuminated using the xenon lamp of a FluoroMax 4 spectrofluorimeter. Excitation and emission slits were of 8 nm (or mentioned specifically Ex 1 nm and Em 29 nm). The emission signal was detected as a function of time.

Binding affinities

To determine the affinity of the labeled ODNs for the peptide, fixed amounts of labeled ODN (1 μM) were titrated with increasing concentrations of NC(1–55) or NC(11–55). Excitation wavelength for thG10SL_RNA, thG12SL_RNA and thG10SL_DNA was at 345 nm and thG12SL_DNA was excited at 360 nm. The binding data were fitted with the help of the Scatchard equation:

$$I = I_o - \frac{(I_o - I_t)}{N} \times \left\{ \frac{[1 + (P_t + nN_t)K_a] - \sqrt{[1 + (P_t + nN_t)K_a]^2 - 4P_t nN_t K_a^2}}{2K_a} \right\} \quad (10)$$

where I and I_t are the signal at a given and a saturating peptide concentration, respectively, I_0 is the signal in the absence of peptide, N_t is the total ODN concentration, P_t is the total concentration of peptide, K_a is the apparent affinity constant, and n is the number of binding sites. Noticeably, the apparent dissociation constant $K_d = K_a^{-1}$.

Anisotropy titrations were performed on the same instrument. Excitation wavelength for thG10SL_RNA, thG12SL_RNA and thG10SL_DNA was at 345 nm and emission was collected at 475 nm. For thG10SL_DNA, excitation was at 360 nm and emission was collected at 475 nm. Anisotropy values were obtained by averaging 10 measurements. The affinity of the labeled ODNs for the peptide was calculated by titrating fixed amounts of labeled ODN (1

µM) with increasing concentrations of NC. The affinity constants were determined by fitting the fluorescence anisotropy changes to the following equation:

$$r = \frac{vRr_t - r_d(v-1)}{1 + Rv - v} \qquad (11)$$

where r and r_t are the anisotropy values at a given and a saturating NC concentration respectively, and r_d is the anisotropy in the absence of protein. R is the ratio of the QYs of the bound to free forms, K_a is the apparent affinity constant, v is the fraction of bound NC calculated as follows:

$$v = \left\{ \frac{(K_a^{-1} + nL_t + P_t) - \sqrt{(K_a^{-1} + nL_t + P_t)^2 - 4nP_tL_t}}{2L_t} \right\} \qquad (12)$$

where P_t and L_t are the concentrations of NC and ODN, respectively, and n is the number of NC proteins bound per ODN

Deconvolution of Spectrum

As the emission spectrum of the red-shifted tautomer can be obtained at excitation wavelengths > 350 nm, the emission spectrum of the blue-shifted isomer in protic solvents was extracted from the emission spectrum recorded at λ_{ex} = 283 nm, by subtracting the emission spectrum of the red-shifted tautomer normalized at wavelengths > 525 nm. The individual absorption spectra of the two tautomers were deduced from the absorption spectrum of [th]G, by normalizing the excitation spectra at wavelengths (> 350 nm) where only the red-shifted form absorbs.

3.5.   Time Correlated Single Photon Counting

Time-resolved fluorescence measurements were performed with the time-correlated single-photon counting technique. Excitation pulses at 315 nm with a repetition rate of 4 MHz were generated by a pulse-picked frequency-tripled Ti-sapphire laser (Tsunami, Spectra Physics) pumped by a Millenia X laser (Spectra Physics) (Bernacchi et al., 2002; Godet et al., 2013). The fluorescence emission was collected through a polarizer set at magic angle and a 16 nm band-pass monochromator (Jobin Yvon) at 370 nm for 2Ap and **s** fluorophores. The fluorescence emission was collected at 500 or 550 nm in order to collect emission only from [th]G-H1 tautomer. Single-photon events were detected with a micro-channel plate photomultiplier (Hamamatsu) coupled to a pulse pre-amplifier HFAC (Becker-Hickl GmbH) and recorded on a time-correlated single photon counting board SPC-130 (Becker-Hickl

GmbH). The instrumental response function (IRF) recorded with a polished aluminum reflector was characterized by a ~ 50 ps full-width at half-maximum. The mean lifetime $<\tau>$ was calculated from the individual fluorescence lifetimes (τ_i) and their relative amplitudes (α_i) according to:

$$<\tau> = \sum \alpha_i \, \tau_i \qquad (13)$$

The population of dark species (α_0) were calculated by:

$$\alpha_0 = 1 - \tau_{free}/(\tau_{ODN} \times R_m) \qquad (14)$$

where, τ_{free} is the lifetime of the free nucleoside, τ_{ODN} is the measured mean lifetime of the probe within the ODN and R_m is the ratio of their corresponding quantum yields. The amplitudes of the fluorescent populations $\alpha_{ic}\alpha_{ic}$ were recalculated according to

$$\alpha_{ic} = \alpha_i \times (1 - \alpha_0) \qquad (15)$$

Time-resolved intensity data were fitted using the maximum entropy method (Pulse 5 software) (Brochon, 1994; Livesey and Brochon, 1987). In all cases, the χ^2 values were close to 1, indicating an optimal fit. The component diagram of TCSPC is shown below Figure 40:
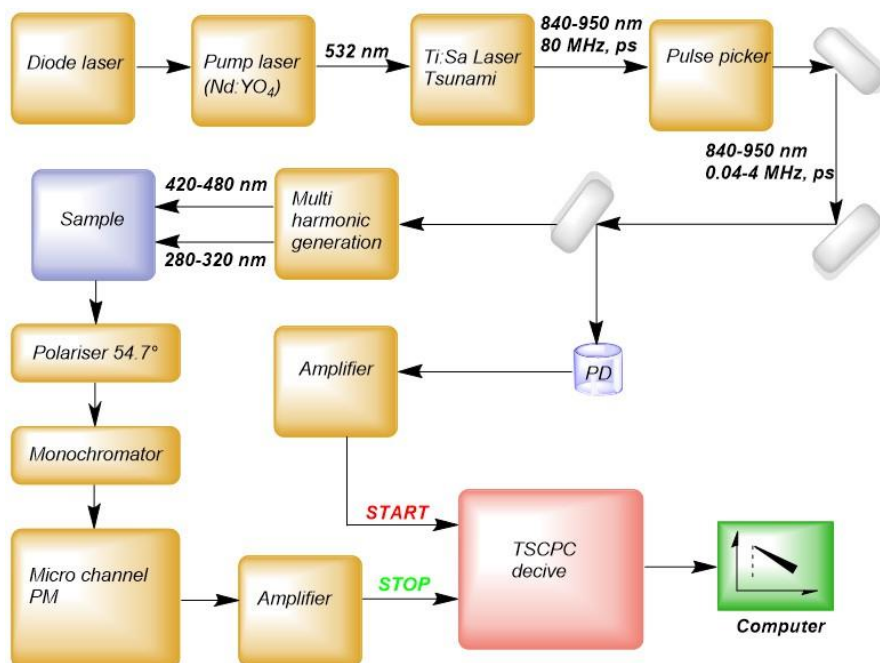
Figure 40 Schematic diagram of TCSPC.

3.6.   Kinetic Measurements

The stopped-flow technique is a kinetic method of analysis, designed to rapidly mix samples and reagents when using reactions with fast kinetics. Sample and reagents are loaded into two syringes (Figure 41), and precisely measured volumes are dispensed by the action of a syringe drive. The two solutions are rapidly mixed in a mixing chamber before flowing through an observation cell. The flow of sample and reagent is stopped by applying back pressure with the stopping syringe. The back pressure completes the mixing, and after that, the reaction is monitored spectrophotometrically. The dead time for the mixing of the sample and the reagent and initiation of the kinetics typically takes around 0.5 - 3 ms.
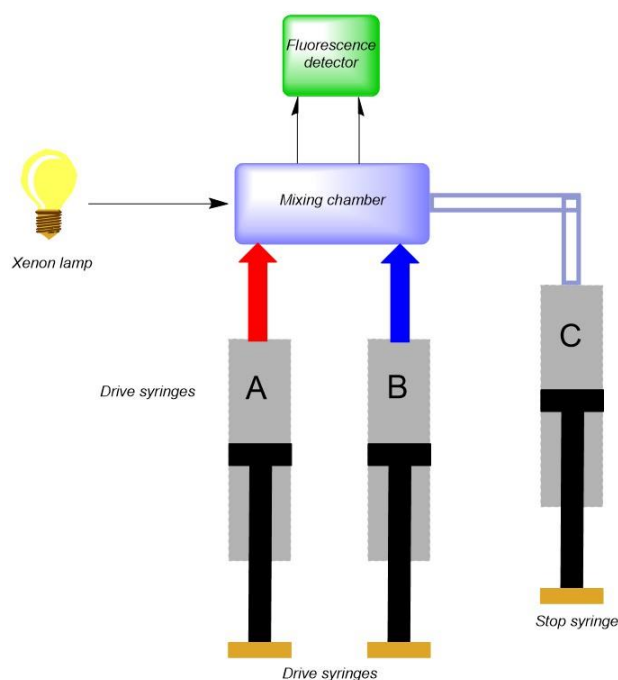
Figure 41 Schematic diagram of Stopped flow.

The kinetic curves were fitted using the following bi-exponential function:

$$I(t) = I_f - \left(I_f - I_0\right)(a \exp\left(-k_{obs1}(t - t_0)\right) + (1 - a) \exp\left(-k_{obs2}(t - t_0)\right)) \qquad (16)$$

where I_0 and I_f are the initial and final intensities of thG7($-$)PBS, respectively; a is the relative amplitude of the fast component; t_0 is the dead time; and k_{obs1} and $k_{o.bs2}$ are the apparent pseudo-first order rate constants.

Assuming a two-step model, the kinetic rate constants were then determined from the dependence of $k_{obs1,2}$ on the (+)PBS concentration according to:

$$k_{obs1} = k_1[(+)\text{PBS}] + k_{-1} \qquad (17)$$

$$k_{obs2} = (k_f K_m[(+)\text{PBS}])/(1 + K_m[(+)\text{PBS}]) + k_{-f} \qquad (18)$$

where k_1 and k_{-1} are the kinetic rate constants for the formation and dissociation of the intermediate complex (IC); K_m is the equilibrium constant (k_1/k_{-1}) for the formation of the IC ; while k_f and k_{-f} are the forward and backward rate constants for the conversion of the IC into the final duplex.

A numerical approach using the Dynafit software (Kuzmič, 2009, 1996) was used to confirm the kinetic rate constants obtained using the analytical approach and to determine the intensities of the intermediate products formed in the (−)/(+)PBS annealing reaction. A typical script for the thG7(−)/(+)PBS system in the absence of NC is presented below:

```
[task]
  task = fit  ;  simulate | design
  data = progress  ;  progress | rates | equilibria | generic
[mechanism]
  A + B <===> AB  :  k1  k-1
  AB ===> C  :  kf
[constants]
  k1  = 8500  ?
  k-1  = 0.003 ?
  kf = 3.5e-4  ?
[concentrations]
[responses]
[progress]
  directory C:\Users\Rajhans\Desktop\DynaFit4\PBS6
  extension txt
  file    1000nM
  concentration   A = 100e-9 , B = 1000e-9
  responses  A = 0.18e+6 , AB = 0.4e+6 ? , C = 0.4e+6
  file    2000nM
  concentration   A = 100e-9 , B = 2000e-9
  responses  A = 0.18e+6 , AB = 0.4e+6 ? , C = 0.4e+6
  file    3000nM
  concentration   A = 100e-9 , B = 3000e-9
  responses  A = 0.18e+6  , AB = 0.4e+6 ? , C = 0.4e+6
  file    4000nM
  concentration   A = 100e-9 , B = 4000e-9
  responses  A = 0.18e+6 , AB = 0.4e+6 ? , C = 0.4e+6
  file    5000nM
  concentration   A = 100e-9 , B = 5000e-9
  responses  A = 0.18e+6  , AB = 0.4e+6 ? , C = 0.4e+6
  file    6000nM
  concentration   A = 100e-9 , B = 6000e-9
  responses  A = 0.18e+6  , AB = 0.4e+6 ?  , C = 0.4e+6
[output]
  directory ./FP/out_rtpt
[end]
```

3.7. Isothermal Titration Calorimetry (ITC)

ITC is a label free technique that determines the binding affinity and the thermodynamic parameters in a single experiment. It has two chambers: a reference chamber and a reaction chamber that are connected to the heat sensing device. It directly measures the heat released or absorbed during a gradual titration of the ligand into the sample cell containing the biomolecule of interest. The component set up of an ITC instrument is shown below.
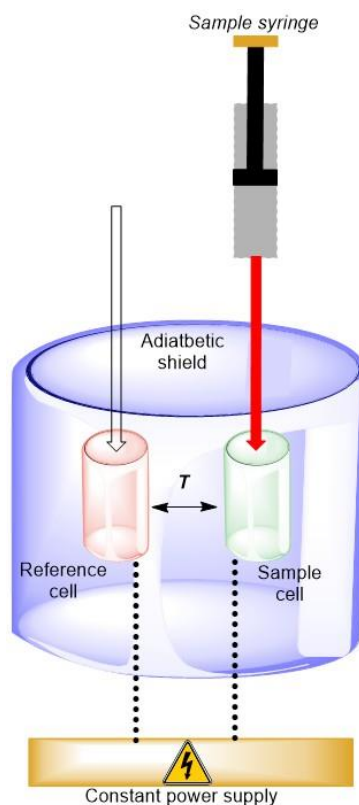
Figure 42 Isothermal calorimetry.

Measuring heat transfer during binding enables determination of binding constants (K_d), reaction stoichiometry (n), enthalpy (ΔH) and entropy (ΔS). Using the relation:

$$\Delta G = \Delta H - T\Delta S \qquad (19)$$

$$\Delta G = RT \ln K_d \qquad (20)$$

R is the universal gas constant 8.31447 $J \cdot K^{-1} \cdot mol^{-1}$ and T is the temperature in Kelvin.

The binding parameters of NC for native and labeled (–)PBS sequences in 50 mM HEPES, pH 7.5, 30 mM NaCl and 0.2 $MgCl_2$ were determined using NanoITC microcalorimeter. The heat of reaction was measured at 20 °C for 25 injections of 2.5 µL NC into 300 µL of oligonucleotide sequences. The syringe mixing speed was 310 rpm. The heat of

dilution was obtained by titrating the NC into a cell of buffer and was subtracted from the raw data. Binding curves were analyzed by non-linear least-squares fitting of the baseline-corrected ITC data to appropriate binding models, using origin 8.6 program.

3.8. Molecular Dynamics (MD)

The NMR structure of SL3 was retrieved from the protein data bank under the PDB ID 1BN0, while the structure of its complex with the NC was coded 1A1T (De Guzman, 1998; Pappalardo et al., 1998). In both cases, the first NMR model was used as starting structure in MD simulations. According to prior studies (Sholokh et al., 2016), MD simulations were performed with Amber14 program (ff14SB force field) (D.A. Case, 2015), on molecular species included in a rectilinear box of TIP3P-type water molecules, and for each system up to 500 ns of unrestrained MD trajectories were produced in the NPT ensemble to monitor structural and energetic changes due G replacement with [th]G in position 10 or 12 of SL3. The [th]G-H1 isomer was considered in this analysis.

For each system, the whole simulation consisted of the following consecutive steps: i) the solvent including water and counterions was energy minimized for 500 steps by the steepest descent algorithm (SD), followed by 1500 steps by the conjugate gradient algorithm (CG) while keeping the solute as frozen (*ibelly* flag); ii) the whole system was then energy minimized for 1000 steps by the SD, and 4000 steps by the CG; iii) the temperature was increased from 0 to 300 K in 100 ps using the Langevin thermostat in the NVT ensemble; iv) density was equilibrated in the NPT ensemble in 100 ps using the Berendsen barostat; iv) unrestrained MD trajectories were finally produced for 500 ns in the NPT ensemble. MD analysis has been performed by the *cpptraj* script and the *MMPBSA.py* python program of Amber14.

4. Results and Discussions

4.1. CHAPTER 1: Photophysical Investigation of Fluorescent Nucleobase Analogues.

4.1.1. Publication 1: Tautomers Of A Fluorescent G Surrogate and Their Distinct Photophysics Provide Additional Information Channels.

In search of a fluorescent nucleoside analogue that provides minimal structural and functional perturbation, we came across a set of newly synthesized purine (thA, thG) and pyrimidines (thU, thC) fluorescent analogues that were derived from a single heterocyclic core of thieno[3,4-d]-pyrimidine by the partner group of Yitzhak Tor (Shin et al., 2011). They demonstrated that these analogues, as a free probe and substituted in short oligonucleotide sequences, showed structural isomorphicity to G and high fluorescence quantum yield. This urged us to utilize thG as a tool for investigating the mechanisms of (–)PBS and (+)PBS annealing and protein-nucleic acids interactions. Prior to this, we characterized the photophysical properties of thG as a free probe and substituted at G7 residue of (–)PBS.

In-depth photophysical characterization of thG nucleoside (Publication 1), showed that in buffer solution thG exists in two ground state tautomers with significantly shifted absorption and emission wavelength maxima but with similar quantum yields (0.49 ± 0.03) (Publication 1, Figure 1). Through TD-DFT calculations, the two tautomers were identified as thG-H1 (red

shifted) and thG-H3 (blue shifted). Absorption spectra of thG-H1 resembled with absorption spectra of thG nucleoside in 1,4-dioxane. Using the extinction coefficient of thG in 1,4-dioxane, we deduced the molar extinction coefficient and proportion of thG-H3 tautomer. This helped us to deduce that the proportions and molar extinction coefficients of both tautomers in buffer were similar. By examining the photostability of thG tautomers, the thG-H1 tautomer was found more photostable than the thG-H3 tautomer (Publication 1, Figure S2). Moreover, the equilibrium between the two thG tautomers was found to be dependent on the hydrogen-bond donor ability of the solvent.

Using quantum mechanical calculation (in collaboration with Roberto Improta), the tautomers were identified as thG-H1 and thG -H3 (keto-amino) tautomers (Publication 1, Figure 3). thG-H1 was found to be dominant in gas phase and 1,4-dioxane. In water, its energy was found to be only 0.11 eV higher than thG-H3, which was assigned to the blue shifted tautomer. Using TD-DFT calculation the computed absorption and emission spectra for both tautomers confirmed that thG-H1 and thG-H3 were responsible of the photophysical features of thG.

We then examined the ground state equilibrium between the tautomers of dthG (deoxy-thG) as substituted at position 7 (G7 residue) of (–)PBS. In single stranded (–)PBS, the photophysical characterization of dthG appeared similar to thG nucleoside in buffer revealing the existence of both tautomers. However, when the dthG7(–)PBS was annealed with its complementary (+)PBS, the equilibrium shifted completely towards dthG-H1 tautomer. Comparing the emission spectra of matched and mismatched (–)/(+)PBS duplex revealed that thG-H3 tautomer is absent in matched duplex while its distribution increased depending on opposite base (T<G<A). Thus, the relative emission of the two dthG tautomers was found highly sensitive to nature of opposite base and can be used to detect single nucleotide polymorphisms. In addition, investigating the response of both tautomers in dthG7(–)PBS upon interaction with NC(11–55) protein, revealed that dthG-H1 is more sensitive towards protein binding.

Further, molecular dynamic (MD) simulations were performed on NMR structure of Δ(–)PBS to compare the unbiased MD trajectories and thermodynamic parameters for dG and dthG. They showed no difference in the behavior of dG and dthG in Δ(–)PBS. In the duplex structure, as substituted in matched (–)/(+)PBS, thG-H1 (three hydrogen bonds) behaves similar to dG and is comparatively more stable than thG-H3 (two hydrogen bonds). In addition, MD simulations showed that the two tautomers were mainly in anti-conformation in both stem-loop and duplex structures.

Tautomers of a Fluorescent G Surrogate and Their Distinct Photophysics Provide Additional Information Channels

Marianna Sholokh[+], Roberto Improta[+], Mattia Mori[+], Rajhans Sharma, Cyril Kenfack, Dongwon Shin, Karine Voltz, Roland H. Stote, Olga A. Zaporozhets, Maurizio Botta, Yitzhak Tor,* and Yves Mély*

In memory of David Shugar

Abstract: *Thienoguanosine (thG) is an isomorphic nucleoside analogue acting as a faithful fluorescent substitute of G, with respectable quantum yield in oligonucleotides. Photophysical analysis of thG reveals the existence of two ground-state tautomers with significantly shifted absorption and emission wavelengths, and high quantum yield in buffer. Using (TD)-DFT calculations, the tautomers were identified as the H1 and H3 keto-amino tautomers. When incorporated into the loop of (−)PBS, the (−)DNA copy of the HIV-1 primer binding site, both tautomers are observed and show differential sensitivity to protein binding. The red-shifted H1 tautomer is strongly favored in matched (−)/(+)PBS duplexes, while the relative emission of the H3 tautomer can be used to detect single nucleotide polymorphisms. These tautomers and their distinct environmental sensitivity provide unprecedented information channels for analyzing G residues in oligonucleotides and their complexes.*

$\boldsymbol{T}$he structure, acid-base features, and tautomeric equilibria of the canonical and non-canonical nucleobases found in nucleic acids have been the subject of intense investigation for decades.[1] While the role of minor tautomers in mutagenesis has been one of the primary foci,[2] recent observations suggest that such isomeric nucleobases also play key roles in regular nucleic acid structure and function.[3] As the population of distinct tautomeric forms is impacted by their micro-

environment, this added level of complexity also provides opportunities to further advance our understanding of nucleic acid structure and dynamics.

In this context, emissive nucleoside analogues, which have become powerful biophysical tools,[4] provide unique prospects. A tautomerizable nucleoside analogue, where the tautomers would have distinct absorption and emission spectra, could be instrumental for investigating the microenvironment of a nucleobase with greater insight compared to tautomerically stable probes. Herein we analyze the photophysics of thienoguanosine, thG, a highly useful G surrogate,[5] and identify two environmentally sensitive ground-state tautomeric forms (Figure 1) which display distinct absorption and emission spectra. The equilibrium between the two tautomers is mainly governed by the hydrogen-bond donor properties of the solvent. Their observed sensitivity to the microenvironment was rationalized by ab initio calculations.
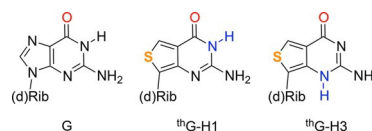
Figure 1. Structures of guanosine (G) and the two emissive tautomers of thienoguanosine (thG–H1 and thG–H3). (d)Rib = D-ribose or 2′-deoxy-D-ribose.

[*] M. Sholokh,[+] R. Sharma, Dr. C. Kenfack, Prof. Dr. Y. Mély
Laboratoire de Biophotonique et Pharmacologie
Faculté de Pharmacie, UMR 7213 CNRS, Université de Strasbourg
74 route du Rhin, 67401 Illkirch (France)
E-mail: yves.mely@unistra.fr

Dr. D. Shin, Prof. Dr. Y. Tor
Department of Chemistry and Biochemistry
University of California, San Diego
9500 Gilman Dr, La Jolla, CA 92093-0358 (USA)
E-mail: ytor@ucsd.edu

M. Sholokh,[+] Prof. Dr. O. A. Zaporozhets
Department of Chemistry
Kyiv National Taras Shevchenko University
60 Volodymyrska street, 01033 Kyiv (Ukraine)

Dr. R. Improta[+]
Consiglio Nazionale delle Ricerche
Istituto di Biostrutture e Bioimmagini
Via Mezzocannone 16, 80134 Napoli (Italy)

Dr. M. Mori,[+] Prof. Dr. M. Botta
Dipartimento di Biotecnologie, Chimica e Farmacia
Università degli Studi di Siena
Via Aldo Moro 2, 53100 Siena (Italy)

Dr. M. Mori[+]
Center for Life Nano Science@Sapienza
Istituto Italiano di Tecnologia
viale Regina Elena 291, 00161 Roma (Italy)

Dr. C. Kenfack
Laboratoire O'Optique et Applications, Centre de Physique Atomique
Moléculaire et Optique Quantique, Université de Douala
BP 8580 Douala (Cameroon)

Dr. K. Voltz, Dr. R. H. Stote
Department of Integrative Structural Biology, Institut de Génétique et
de Biologie Moléculaire et Cellulaire, INSERM U964 UMR 7104
CNRS, Université de Strasbourg, 67400 Illkirch (France)

[+] These authors contributed equally to this work.

Supporting information for this article can be found under:
http://dx.doi.org/10.1002/anie.201601688.

By exploring single- and double-stranded [th]G-containing oligonucleotides, as well as DNA–protein complexes, we illustrate that this probe provides compelling biophysical information and greater insight compared to monochromatic or ratiometric fluorescent nucleosides.

The emission spectra of either [th]G or d[th]G in water and methanol are surprisingly complex.[5a,b] Excitation at both $\lambda = 360$ and 380 nm gives a similar emission spectrum (Figure 2a, orange) centered at $\lambda = 468$ nm. When the excitation energy is progressively increased, a blue-shifted emission with a maximum at $\lambda = 400$ nm appears and becomes dominant for excitation below $\lambda = 300$ nm (Figure 2a, magenta and blue). The simplest interpretation is that [th]G exhibits two ground-state species with shifted emission spectra. This hypothesis is supported by recording excitation spectra for various emission wavelengths (see Figure S1 in the Supporting Information). Since sample purity was rigorously maintained, the two ground-state species likely correspond to two tautomers, differing by their excitation and emission spectra. This conclusion is highly likely, since tautomers have also been observed for guanosine itself.[6]

Spectral deconvolution yields well separated emission and absorption spectra of the individual tautomers in buffer (Figure 2b,c). Thus, by judiciously selecting the excitation and emission wavelengths, each tautomer can be individually excited and observed. The fluorescence quantum yield (QY) is found to be constant (0.49 ± 0.03) over a large range of excitation wavelengths ($\lambda = 290–375$ nm) and close to earlier reported values,[5a,c] thus indicating that the two forms possess similar QY values. The individual absorption spectrum of the red-shifted form (Figure 2c) is very similar to the spectra of [th]G and d[th]G in 1,4-dioxane.[5a,b] By using the molar extinction coefficient of [th]G in 1,4-dioxane ($\varepsilon^{333} = 4530\,\text{M}^{-1}\,\text{cm}^{-1}$),[5a] it is possible to calculate the concentration of the red-shifted tautomer in buffer and deduce the proportion (44%) and the molar extinction coefficient ($\varepsilon^{313} \approx 4600\,\text{M}^{-1}\,\text{cm}^{-1}$) of the blue-shifted tautomer. Importantly, the red-shifted tautomer, excited at $\lambda = 350–380$ nm, is highly photostable (see Figure S2). In contrast, extended illumination at higher energies (e.g., $\lambda = 325$ nm), where both tautomers absorb, show continuously diminished emission, thus suggesting that the blue-

shifted tautomer is less photostable and that the two tautomers are equilibrating.

The spectroscopic properties of [th]G were comparatively characterized in various solvents (see Table S1). In methanol, ethanol, and *n*-butanol (see Figures S3 and S4) spectra comparable to those in buffer are obtained, thus indicating that [th]G also exists in the two tautomeric forms in these solvents. In all other tested solvents, the emission spectra are independent on the excitation wavelength and indicate that only the red-shifted tautomer is present (see Figure S3c). Its emission maximum correlates well with the empirical polarity index $E_T(30)$ of the solvent (see Figure S5). Interestingly, although *N,N*-dimethylformamide, acetonitrile, and methanol are all rather polar [$E_T(30) > 43\,\text{kcal mol}^{-1}$, $\varepsilon > 32$], the blue-shifted tautomer is seen only in methanol, thus strongly suggesting hydrogen-bonding stabilization. This proposal is further substantiated by the deconvoluted absorption spectra in polar protic solvents (see Figure S6), as they show that the relative concentration of the blue-shifted isomer linearly increases with solvent proticity (see Figure S6d and Table S1). Thus, the equilibrium between the two [th]G tautomers is dependent on the hydrogen-bond donor ability of the solvent.

To assist in identifying the two emissive isomers, the ground-state energy minima of five hypothetical [th]G tautomers were optimized (Figure 3) in the gas phase, 1,4-dioxane, and water at the DFT level, by using PBE0 and M052X functionals and including solvent effects with the Polarizable Continuum Model (PCM; see Tables S2 and S3). The keto-amino [th]G–H1 tautomer (Figure 3a) appears largely dominant over the other tautomers (Figures 3b–e), with the exception of water where it is only 0.11 eV more stable than

Figure 3. Schematic drawing of the [th]G tautomers which have been calculated: a) keto-amino [th]G–H1, b) keto-amino [th]G–H3, c) enol-amino, d) keto-imino, and e) enol-imino.
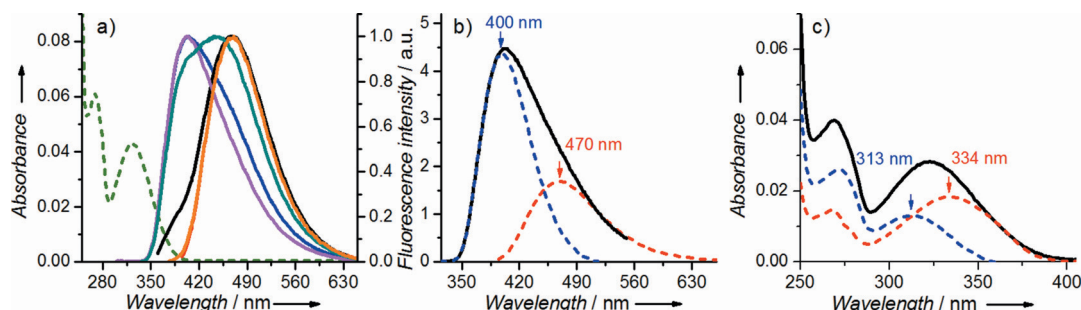
Figure 2. Absorption and emission spectra of [th]G in TRIS-HCl buffer 25 mM, pH 7.5, 30 mM NaCl, 0.2 mM MgCl₂. a) Absorption (green dashed line) and emission spectra of [th]G at different excitation wavelengths: $\lambda = 283$ nm (magenta line); $\lambda = 298$ nm (blue); $\lambda = 320$ nm (green); $\lambda = 345$ nm (black); and $\lambda = 360$ nm (orange). The emission spectra were normalized at their maxima. The normalized emission spectrum at $\lambda_{exc} = 380$ nm fully overlaps that at $\lambda_{exc} = 360$ nm and is not shown. b) Deconvoluted emission spectrum of [th]G, obtained at $\lambda_{ex} = 283$ nm. c) Deconvolution of the absorption spectrum of [th]G (black line) in its two ground-state forms (colors as in b).

thG–H3, when including only bulk solvent effects. Therefore, the two thG keto-amino tautomers, analogous to the most stable tautomer of guanine in solution,[7] are likely populated in water.

Independent of the inclusion of one solvent molecule in the computational model (see Figures S7 and S8, and Table S2), the thG–H1 tautomer appears as the main contributor to the observed spectroscopic properties of thG in 1,4-dioxane and is therefore assigned to the red-shifted isomer.[8] The thG–H1 and thG–H3 tautomers are found to be almost isoenergetic in water when solute–solvent hydrogen bonds are considered (see Figure S7 and Table S2). These data suggest that both tautomers likely contribute to the spectroscopic properties of thG in water and that the blue-shifted isomer corresponds to the thG–H3 tautomer. The computed energy difference between the two tautomers (< 0.05 eV, that is, < 400 cm^{-1}) is beyond the expected accuracy of our method, thus explaining why the molar fraction of the thG–H3 tautomer (see Table S2) does not perfectly match with the experimental value (0.44).[9] The computed vertical absorption and emission energies (see Table S4)[10] indicate that, independent of the functional, the lowest-energy transition in water for both tautomers corresponds to a bright $\pi\pi^*$ S$_0\rightarrow$S$_1$ transition attributed to a HOMO→LUMO excitation.[11]

Interestingly, small differences in the shape of the frontier orbitals involved in the electronic transition (see Figure S9) result in fairly large differences in the computed vertical excitation energy, so that the absorption maximum of the thG–H1 tautomer in water is red-shifted by 30–40 nm, with respect to that of thG–H3. The absorption maxima predicted for the two forms, namely $\lambda = 330\sim350$ nm (depending on the solvation model) for thG–H1 and $\lambda = 300\sim310$ nm for thG–H3, are very close to the experimental ones (see Figures S10a, 2c and Table S1).[12] TD-DFT excited-state geometry optimizations (employing either PBE0 or M052X) predict a stable S$_1$ minimum for both tautomers in all examined solvents.[13] This S$_1$ minimum is characterized by a fairly large oscillator strength, of about 80% of the value computed for absorption. This minimum contrasts guanosine, for which the same functionals predict a barrierless decay to S$_0$, through an effective conical intersection.[14] Both thG tautomers therefore show promising electronic features with potentially robust emissive states. The computed emission wavelengths

of $\lambda = 448$ and 383 nm for the thG–H1 and thG–H3 tautomers, respectively,[15] (see Figure S10b), are in good agreement with the spectroscopic data in buffer (Figure 2b). Moreover, by weighting the contribution of the different tautomers with a simple Boltzmann equation, the computed fluorescence spectra (see Figure S10b) are consistent with the experimental ones. Taken together, PCM/(TD)DFT calculations indicate that the thG–H1 and thG–H3 tautomers are responsible of the observed photophysical features of thG.

To examine the ground-state equilibrium between the tautomers of thG in oligonucleotides, the DNA equivalent of the 18-mer primer binding site of HIV-1 was selected as a biologically relevant model (see Figure S11). It forms a stem-loop of known three-dimensional structure [16] and is involved in the second strand transfer of HIV-1 reverse transcription.[17] Deoxy-thG (dthG), which exhibits spectroscopic properties very similar to thG,[5b,c] substitutes the dG7 residue in the loop [labeled dthG7(−)PBS; Figure 4a inset]. Comparing the emission spectra at various excitation wavelengths clearly shows that both dthG tautomers are present in the (−)PBS loop (Figure 4a). In contrast, when dthG7(−)PBS is annealed to its complementary (+)PBS strand (see Figure S11), forming the dthG7(−)/(+)PBS duplex (Figure 4c, inset), the normalized emission spectra obtained at different excitation wavelengths all overlap, thus indicating that the dthG–H1 tautomer is predominant in the double-stranded
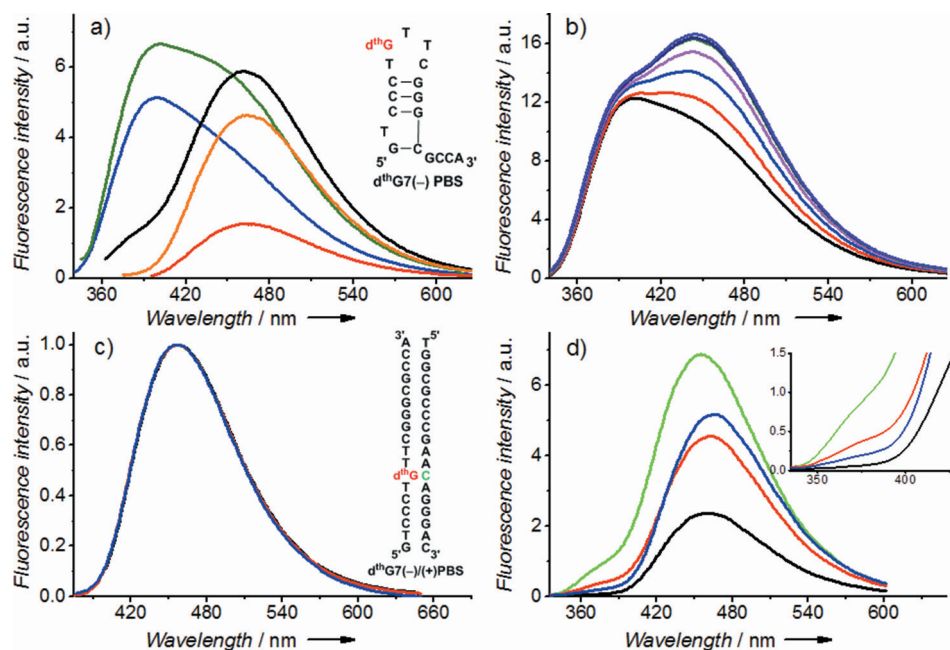
Figure 4. Emission spectra of dthG7(−)PBS (a, b) and dthG7(−)/(+)PBS duplexes (c, d). a) Emission spectra of dthG7(−)PBS recorded at different excitation wavelengths: $\lambda = 298$ nm (blue), $\lambda = 320$ nm (green), $\lambda = 345$ nm (black), $\lambda = 360$ nm (orange), and 380 nm (red). Inset: structure of dthG7(−)PBS, the G7 residue (red) is replaced by dthG. b) Emission spectra of dthG7(−)PBS in the absence (black) and in the presence of 1 to 6 equivalents of NC(11-55) protein (red to violet) at $\lambda = 320$ nm excitation wavelength. c) Emission spectra of the matched dthG7(−)/(+)PBS duplex at the same excitation wavelengths as in (a). Inset: structure of dthG7(−)/(+)PBS duplex. In mismatched duplexes, the C residue in green is replaced by a A, G or T. (d) Emission of the matched and mismatched dthG7-labeled (−)/(+)PBS duplexes at $\lambda = 310$ nm excitation wavelength. The base opposite to dthG is C (black, native duplex), T (blue), G (red), or A (green). Inset: zoom of the blue part of the spectra. The buffer was as in Figure 2.

form (Figure 4c). Although not attributed to the two tautomers disclosed here, a similar switch from a two-band to a single-band emission was previously observed upon transition from single- to double-stranded structures in model thG- and dthG-labeled sequences,[5a,b] thus indicating that the tautomeric shift reported here is not unique for (−)PBS.

Distinct behavior was seen for mismatched duplexes between dthG7(−)PBS and complementary (+)PBS oligonucleotides, where dthG was placed opposite A, T, or G (Figure 4d). In contrast to the fully complementary duplex, where the dthG–H3 tautomer is nearly absent, its relative contribution as estimated by the ratio of the fluorescence intensities at $\lambda = 375$ and 550 nm, I_{375}/I_{550}, increases by a factor of three and five in the mismatched duplexes with opposite dG and dA, respectively (see Table S5). For the mismatched duplex with opposite dT, the difference with the matched duplex is marginal, but the two duplexes can be easily discriminated by the twofold difference in their quantum yield (see Table S5). This difference likely results both from a change in polarity (as suggested by the changes in the positions of the dthG–H1 emission maximum; see Table S5) and in the quenching by the flanking nucleobases, as a result of the different geometries adopted by dthG and its neighbors in the two duplexes. The relative emission of the two dthG tautomers and the dthG quantum yield are therefore highly sensitive to the nature of the opposite base and can thus be used in combination to detect single nucleotide polymorphisms.

To further illustrate the potential applications of the two spectrally distinct dthG tautomers when in oligonucleotides, we investigated their response to binding of the HIV-1 nucleocapsid NC(11-55) peptide to the (−)PBS loop.[16b] Titration with NC(11-55) protein resulted in a strong increase of the dthG–H1 peak of dthG7(−)PBS, but only a marginal increase in the dthG–H3 peak (shift in the H3/H1 emission ratio from 1.1 to 0.8), thus indicating that the relative emission of the two tautomers is sensitive to protein binding (Figure 4b). As NC(11-55) was reported to direct the G7 base toward the exterior of the loop[16b] and restrict its collisions with the neighboring bases,[18] it appears that the dthG–H1 tautomer is more sensitive than dthG–H3 to these changes.

To shed light on the biophysical observations, molecular dynamics (MD) simulations using the ff12SB AMBER force field were performed on the NMR structure of ΔP(−)PBS DNA,[16b] a truncated form of (−)PBS lacking the 3′ overhang (see Figure S11). Analysis of MD trajectories (0.2 μs of unbiased MD trajectory production) and thermodynamic parameters unequivocally shows that there are no differences in the behavior of either dG or the two dthG tautomers in two representative ΔP(−)PBS structures (see Figures S12, S13a,b, and S14). In contrast, analysis of local motion within the Watson–Crick base pair established by either dG or dthG at position 7 in the (−)/(+)PBS DNA duplex clearly shows that dthG–H1 has the same behavior as dG, whereas dthG–H3 pairs with the counterpart dC with lower stability (see Figure S13d). A local structural analysis of MD trajectories further confirms that dthG–H1 forms the three canonical hydrogen-bonds with dC as observed for guanine (Figure S15a,b), while dthG–H3 contacts dC in multiple non-

canonical complexes (see Figures S15c and S16). Overall, and consistently with experimental observations, the replacement of dG with dthG–H3 in the (−)/(+)PBS DNA duplex is noticeably less favorable than the replacement with dthG–H1, from a thermodynamic and conformational viewpoint.[19] Finally, MD simulations reveal that the two tautomers are mainly in the *anti*-conformation in both the stem-loop and the duplex (see Table S6).

In summary, through a careful analysis of its spectroscopic properties as a free nucleoside and when incorporated into oligonucleotides, thienoguanosine thG was observed to exhibit two ground-state tautomers with significantly shifted absorption and emission spectra. Quantum mechanical calculations unambiguously identified the two tautomers as being the keto-amino tautomers, thG–H1 and thG–H3. MD studies further suggested thG–H1 behaves similarly to its native counterpart in both the single- and double-stranded structures studied here, whereas the thG–H3 tautomer behaves comparably to G only in the loop of a stem-loop DNA. When incorporated into double-stranded sequences, thG–H3 tautomerizes to the favorable and benign thG–H1 tautomer, which forms a stable Watson–Crick base pair. The ratio of the two tautomers and their relative emission were found to be highly sensitive to the nucleic acid strandedness, to the nature of the opposite base in DNA duplexes, as well as to protein binding. The tautomerism of the isomorphic thG, which is associated with distinct and highly emissive states, thus constitutes a highly useful additional channel of information that provides an unprecedented window into features of substituted G residues in oligonucleotides.

[1] a) H. T. Miles, *Proc. Natl. Acad. Sci. USA* **1961**, *58*, 791; b) Y. P. Wong, K. L. Wong, D. R. Kearns, *Biochem. Biophys. Res.*

Commun. **1972**, *49*, 1580; c) G. C. Lee, S. I. Chan, *J. Am. Chem. Soc.* **1972**, *94*, 3218; d) H. Robinson, Y. G. Gao, C. Bauer, C. Roberts, C. Switzer, A. H. J. Wang, *Biochemistry* **1998**, *37*, 10897; e) J. R. Blas, F. J. Luque, M. Orozco, *J. Am. Chem. Soc.* **2004**, *126*, 154.

[2] a) J. W. Drake, R. H. Baltz, *Annu. Rev. Biochem.* **1976**, *45*, 11; b) M. D. Topal, J. R. Fresco, *Nature* **1976**, *263*, 285; c) D. Shugar, B. Kierdaszuk, *Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci.* **1985**, *8*, 657; d) W. N. Wang, H. W. Hellinga, L. S. Beese, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17644.

[3] a) E. Westhof, *FEBS Lett.* **2014**, *588*, 2464; b) V. Singh, B. I. Fedeles, J. M. Essigmann, *RNA* **2015**, *21*, 1; c) I. Kimsey, H. M. Al-Hashimi, *Curr. Opin. Struct. Biol.* **2014**, *24*, 72.

[4] a) R. W. Sinkeldam, N. J. Greco, Y. Tor, *Chem. Rev.* **2010**, *110*, 2579; b) A. A. Tanpure, M. G. Pawar, S. G. Srivatsan, *Isr. J. Chem.* **2013**, *53*, 366; c) K. Phelps, A. Morris, P. A. Beal, *ACS Chem. Biol.* **2012**, *7*, 100; d) L. M. Wilhelmsson, *Q. Rev. Biophys.* **2010**, *43*, 159; e) J. N. Wilson, E. T. Kool, *Org. Biomol. Chem.* **2006**, *4*, 4265; f) M. E. Hawkins, *Methods Enzymol.* **2008**, *450*, 201.

[5] a) D. Shin, R. W. Sinkeldam, Y. Tor, *J. Am. Chem. Soc.* **2011**, *133*, 14912; b) S. Park, H. Otomo, L. Zheng, H. Sugiyama, *Chem. Commun.* **2014**, *50*, 1573; c) M. Sholokh, R. Sharma, D. Shin, R. Das, O. A. Zaporozhets, Y. Tor, Y. Mely, *J. Am. Chem. Soc.* **2015**, *137*, 3185.

[6] a) C. Colominas, F. J. Luque, M. Orozco, *J. Am. Chem. Soc.* **1996**, *118*, 6811; b) E. Nir, C. Janzen, P. Imhof, K. Kleinermanns, M. S. de Vries, *J. Chem. Phys.* **2001**, *115*, 4604; c) C. M. Marian, *J. Phys. Chem. A* **2007**, *111*, 1545.

[7] Y. J. Lee, Y. H. Jang, Y. Kim, S. Hwang, *Bull. Korean Chem. Soc.* **2012**, *33*, 4255.

[8] Interestingly, the stability of the thG–H3 tautomer increases with polarity and particularly with the hydrogen-bonding ability of the solvent.

[9] For the examined solvents, PBE0 and M052X provide similar indications, thus suggesting that our conclusions are robust with respect to the choice of the functional. Inclusion of vibrational effects does not substantially affect the conformational equilibrium between thG–H1 and thG–H3 tautomers.

[10] Computed at the PCM/TD-PBE0 and PCM/TD-M052X level.

[11] a) P. K. Samanta, A. K. Manna, S. K. Pati, *J. Phys. Chem. B* **2012**, *116*, 7618; b) P. K. Samanta, S. K. Pati, *New J. Chem.* **2013**, *37*, 3640.

[12] In addition to the relative position of the lowest energy peak, the general shape of the absorption spectra of the thG-H1 and thG–H3 tautomers was found to be very close to experimental ones, thus supporting their assignment.

[13] a) M. Gedik, A. Brown, *J. Photochem. Photobiol. A* **2013**, *259*, 25; b) P. K. Samanta, S. K. Pati, *Phys. Chem. Chem. Phys.* **2015**, *17*, 10053.

[14] a) V. Karunakaran, K. Kleinermanns, R. Improta, S. A. Kovalenko, *J. Am. Chem. Soc.* **2009**, *131*, 5839; b) R. Improta, *Chem. Eur. J.* **2014**, *20*, 8106.

[15] According to PCM/6-311 + G(2d,2p) calculations on a thG·6H$_2$O model.

[16] a) P. E. Johnson, R. B. Turner, Z. R. Wu, L. Hairston, J. Guo, J. G. Levin, M. F. Summers, *Biochemistry* **2000**, *39*, 9084; b) S. Bourbigot, N. Ramalanjaona, C. Boudier, G. F. J. Salgado, B. P. Roques, Y. Mély, S. Bouaziz, N. Morellet, *J. Mol. Biol.* **2008**, *383*, 1112.

[17] J. L. Darlix, J. Godet, R. Ivanyi-Nagy, P. Fossé, O. Mauffret, Y. Mély, *J. Mol. Biol.* **2011**, *410*, 565.

[18] J. Godet, N. Ramalanjaona, K. K. Sharma, L. Richert, H. De Rocquigny, J. L. Darlix, G. Duportail, Y. Mély, *Nucleic Acids Res.* **2011**, *39*, 6633.

[19] To rule out any possible bias by the selected force field used for the MD simulations (AMBER ff12SB force field) similar computations were performed using the CHARMM all-atom force field for nucleic acids. The results of these calculations are highly comparable and are detailed in the Supporting Information.

Angewandte Chemie

Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Tautomers of a Fluorescent G Surrogate and Their Distinct Photophysics Provide Additional Information Channels

Marianna Sholokh[+], Roberto Improta[+], Mattia Mori[+], Rajhans Sharma, Cyril Kenfack, Dongwon Shin, Karine Voltz, Roland H. Stote, Olga A. Zaporozhets, Maurizio Botta, Yitzhak Tor,* and Yves Mély*

anie_201601688_sm_miscellaneous_information.pdf

Table of contents

1. Materials

Thienoguanosine (thG) and 2′-deoxy-thienoguanosine (dthG)-labeled (−)PBS DNA were synthesized as previously described by Shin et al.[1] and Sholokh et al.,[2] respectively. Stock solutions of thG were prepared in spectroscopic grade DMSO. The complementary native and mismatched (+)PBS DNA were purchased from IBA Nucleic Acids Product Supply (Germany). dthG-labeled duplexes (−)/(+)PBS were prepared by hybridization of dthG7(−)PBS and (+)PBS samples (concentration ratio 1:3), denatured at 85 °C for 3 min and then slowly cooled down to the room temperature. NC(11-55) peptide was synthesized on a Applied Biosystems A433 peptide synthesizer, as described by de Rocquigny[3] and prepared using 2.2 equivalents of Zn(II). All experiments with aqueous solutions were done in 25 mM TRIS-HCl buffer (pH=7.5), 30 mM NaCl and 0.2 mM MgCl$_2$.

2. Spectroscopic measurements

Spectroscopic grade solvents were used for absorption and fluorescence spectroscopy measurements. Absorption spectra were recorded on a Cary 4000 UV-visible spectrophotometer (Varian). Fluorescence excitation and emission spectra were recorded on a FluoroMax 4 spectrofluorimeter (Jobin Yvon) equipped with a thermostated cell compartment at 20 ± 0.5 °C. Fluorescence spectra were corrected for Raman scattering, lamp fluctuations and instrumental wavelength-dependent bias. [th]G concentration in the various solvents was 6 μM, with a final DMSO concentration of 0.1 v/v %. Photostability measurements were performed in cuvettes with 50 μL total volume under continuous illumination at 325, 350, 360, 370 or 380 nm during 2000 s.

Deconvolution procedure

As the emission spectrum of the red-shifted tautomer can be obtained at excitation wavelengths > 350 nm, the emission spectrum of the blue-shifted isomer in protic solvents was extracted from the emission spectrum recorded at λ_{ex} = 283 nm, by subtracting the emission spectrum of the red-shifted tautomer normalized at wavelengths > 525 nm. The individual absorption spectra of the two tautomers were deduced from the absorption spectrum of [th]G, by normalizing the excitation spectra at wavelengths (> 350 nm) where only the red-shifted form absorbs.
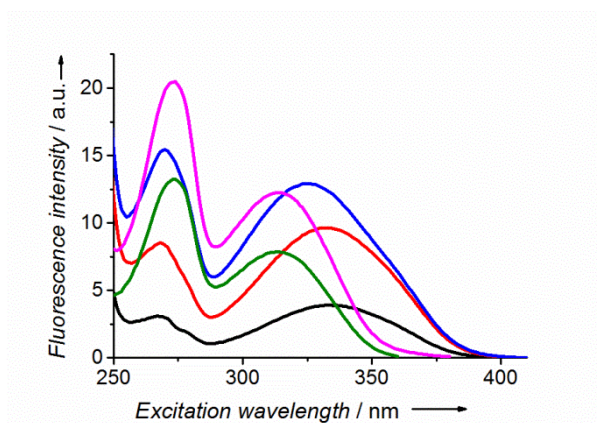
Figure S1. Excitation spectra of thG in TRIS-HCl buffer (25 mM, pH=7.5, 30 mM NaCl, 0.2 mM MgCl2) at different emission wavelengths: 550 nm (black line), 500 nm (red), 450 nm (blue), 400 nm (magenta), 375 nm (green). A maximum at 334 nm is observed when emission is recorded in the 500–550 nm range, and the excitation maximum is progressively blue-shifted down to 313 nm, for shorter wavelength emission.
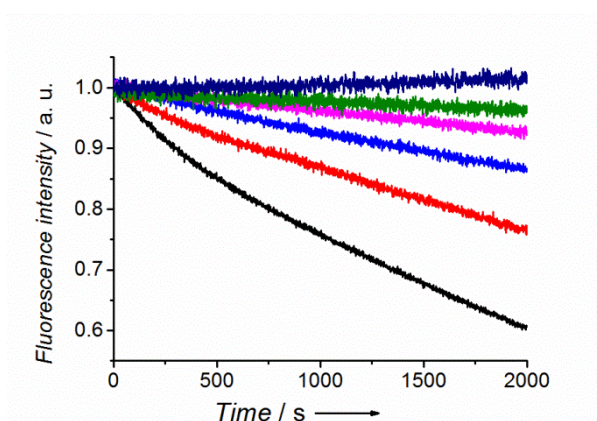
Figure S2. Photostability of the thG ground-state tautomers in buffer. Kinetics at different excitation and emission wavelengths, respectively: 325 and 400 nm (black); 325 and 525 nm (red); 350 and 525 nm (blue); 360 and 525 nm (magenta); 370 and 525 nm (green); 380 and 525 nm (dark blue). Concentration of thG was 1 µM.

Table S1. Photophysical Properties of thG in Buffer and Organic Solvents [a]

Solvent	$E_T(30)$	ε	α	β	λ_{abs}	λ_{em}^{325}	λ_{em}^{380}
Buffer	63.1	78.35	1.17	0.18	322	454	468
Methanol	55.4	32.61	0.93	0.62	326	458	459
Ethanol	51.9	24.85	0.83	0.77	329	458	458
n-Butanol	49.7	17.33	0.79	0.88	330	457	457
2-Methyl-2-butanol	41.0	5.82	0.28	0.93	335	446	445
N,N-dimethylformamide	43.2	37.21	0.00	0.69	338	440	440
Acetonitrile	45.6	35.68	0.19	0.31	330	432	432
Ethyl acetate	38.1	5.98	0.00	0.45	331	426	427
1,4-Dioxane	36.0	2.20	0.00	0.37	330	424	426

[a] $E_T(30)$ is the empiric polarity index[4] reported in kcal mol^{-1}; ε is the dielectric constant at 298 K; α is the Kamlet-Taft solvent hydrogen bond acidity; β is the Kamlet-Taft solvent hydrogen bond basicity;[5] λ_{abs} is the absorption maxima in nm; λ_{em}^{325} and λ_{em}^{380} are the fluorescence emission maxima at 325 nm and 380 nm excitation wavelengths, respectively, reported in nm. TRIS-HCl buffer 25 mM, pH=7.5, 30 mM NaCl, 0.2 mM MgCl$_2$ was used.
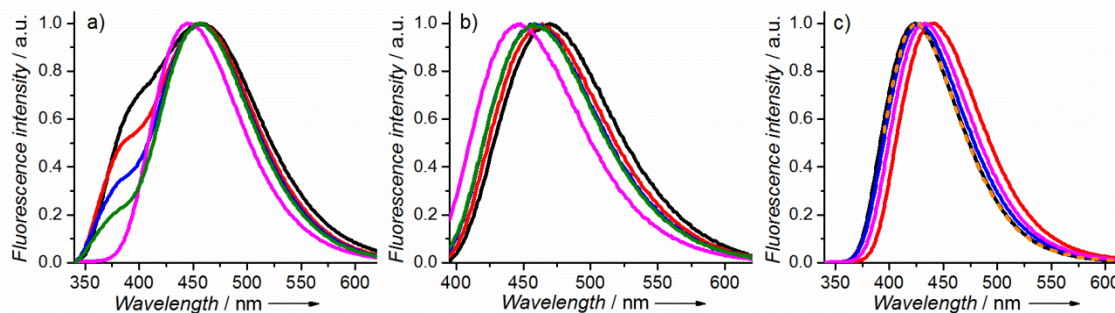
Figure S3. Normalized emission spectra of thG in various solvents. (a) Buffer (black), methanol (red), ethanol (blue), *n*-butanol (green), 2-methyl-2-butanol (magenta), λ_{ex} = 325 nm; (b) Emission spectra in the same solvents as in (a), but with λ_{ex} = 380 nm; (c) Emission spectra in 1,4-dioxane (black), *N,N*-dimethylformamide (red), ethyl acetate (blue), acetonitrile

(magenta), λ_{ex} = 325 nm; superimposed normalized thG emission spectra in 1,4-dioxane obtained at λ_{ex} = 380 nm (orange dash).
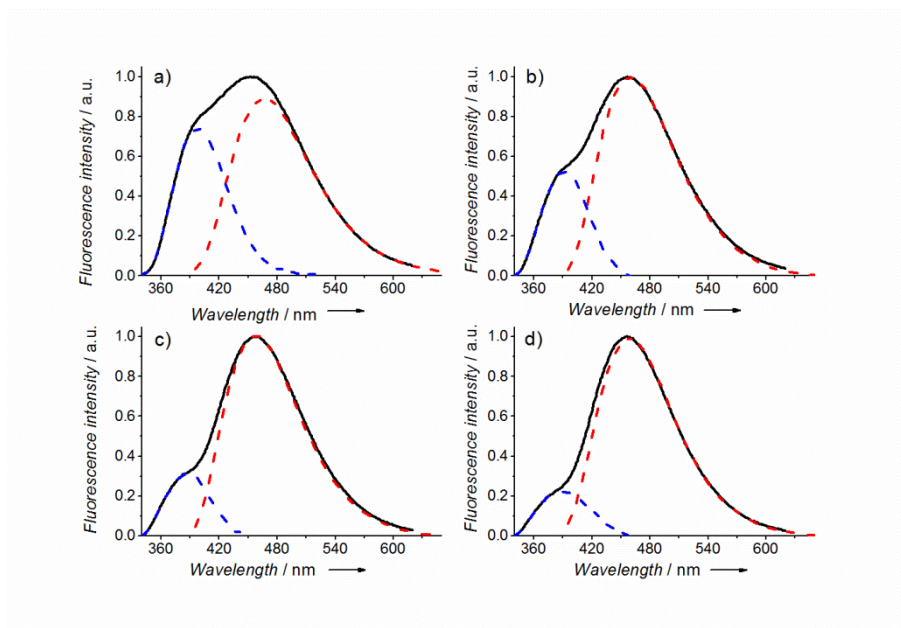
Figure S4. Deconvolution of the emission spectra of thG nucleoside in buffer (a), methanol (b), ethanol (c), and *n*-butanol (d). Excitation wavelength was 325 nm.
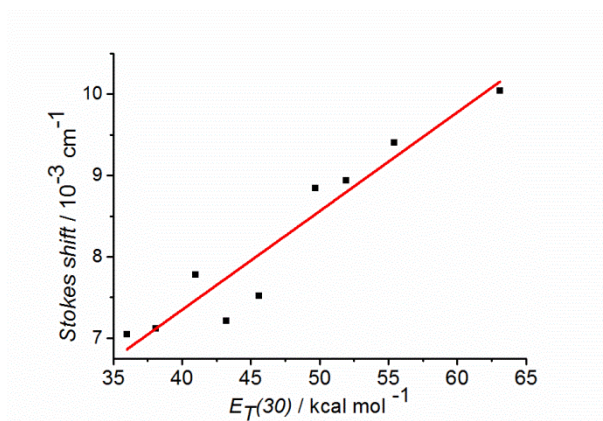
Figure S5. Dependence of the Stokes shift of the red-shifted tautomer on the empiric polarity index $E_T(30)$ (symbols). The red line represents the linear fit to the data with a slope of 0.12 $\pm$ 0.01 and an intercept of 2.4 $\pm$ 0.6, R^2 = 0.90. The $E_T(30)$ values of the solvents are given in

Table S1. The Stokes shift was calculated from the absorption and emission maxima, after correction of the emission spectra according to: Intensity $(\bar{v})$ = Intensity $(\lambda) \times \lambda^2$.[6]
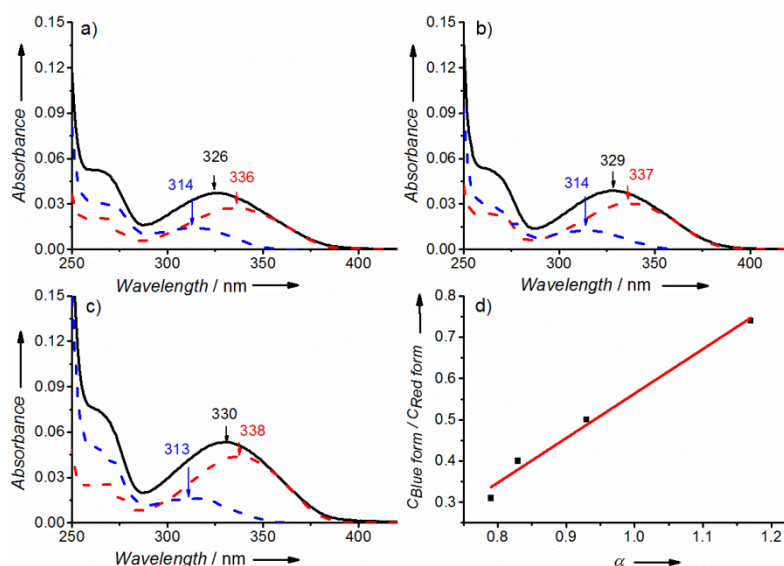
Figure S6. Deconvolution of the absorption spectra of thG in methanol (a), ethanol (b) and *n*-butanol (c) (black line) in its two ground state forms. Absorption spectra of the blue-shifted (blue dashed line) and red-shifted (red dashed line) forms. (d) Dependence of the concentration ratio of the two forms on the Kamlet-Taft's hydrogen bond acidity α. The straight line which was fitted to the experimental points shows a slope of 1.08 ± 0.08 and an intercept of -0.51 ± 0.08, $R^2 = 0.98$.

3. Quantum chemical calculations

The analysis is based on Density Functional Theory (DFT) and on its time dependent extension (TD-DFT), by using two different functionals, PBE0[7] and M052X.[8] PBE0 is a parameter-free functional, which accurately describes the bright states of guanine[9] and of other nucleobases,[10] providing vertical excitation and emission energies within ~0.15 eV of the corresponding experimental absorption maxima. Since PBE0 might overestimate the stability of charge transfer (CT) transitions[11] and to avoid any possible artifacts due to small

contribution of solute–solvent CT excitations, we have verified the results by using M052X functional, which is particularly effective for the treatment of non-bonding interactions and is not biased by the traditional failure of TD-DFT in describing CT states.[12]

Geometry optimizations have been performed at the less computationally demanding 6-31G(d) level, refining the vertical absorption vA and emission energies vE by single-point calculations employing more extended 6-31+G(d,p) and 6-311+G(2d,2p) basis sets.

Table S2. Relative Energy (in eV) of the thG-H3 Tautomer with respect to the thG-H1 Tautomer (taken as 0) Computed at the PCM/DFT/6-311+G(2d,2p) Level by Using Different Basis Sets and Geometries Optimized at the PCM/DFT/6-31G(d) Level [a]

	PBE0	M052X
	Gas phase	
6-31G(d)	0.441 (0.411)	0.439 (0.400)
6-311+G(2d,2p)	0.399	0.396
	Dioxane (PCM)	
6-31G(d)	0.349 (0.318)	0.345 (0.331)
6-311+G(2d,2p)	0.297	0.292
	Dioxane (PCM + 1 Dioxane)	
6-31G(d)	0.321 (0.299)	0.291 (0.226)
6-311+G(2d,2p)	0.277	0.250
	Water (PCM)	
6-31G(d)	0.199 (0.160)	0.181 (0.186)
6-311+G(2d,2p)	0.114	0.103
	Water (PCM+ 2H_2O)	
6-31G(d)	0.214 (0.124)	0.179 (0.088)
6-311+G(2d,2p)	0.067	0.054
	Water (PCM + 6H_2O)	
6-31G(d)	0.138 (0.141)	0.105 (0.141)

	6-311+G(2d,2p)	0.049 [b]	0.023 [c]

[a] The values obtained by including vibrational corrections are in parentheses. At 300 K, the calculated molar fraction of the thG-H3 tautomer in water is 0.13 [b] and 0.29 [c].

Table S3. Relative Energy (in eV) of the Different thG Tautomers Computed at the PCM/PBE0/level by Using Different Basis Sets and Geometries Optimized at the PCM/PBE0/6-31G(d) Level [a]

	thG-H3	Enol-amino	Keto-imino	Enol-imino
		Gas Phase		
6-31G(d)	0.44 (0.41)	0.33 (0.33)	0.31 (0.31)	0.83 (0.83)
6-31+G(d,p)	0.43	0.25	0.31	0.75
6-311+G(2d,2p)	0.40	0.24	0.29	0.72
		Dioxane (only PCM)		
6-31G(d)	0.35 (0.32)	0.39 (0.38)	0.31 (0.29)	0.85 (0.86)
6-31+G(d,p)	0.32	0.32	0.29	0.78
6-311+G(2d,2p)	0.30	0.31	0.27	0.75
		Water (only PCM)		
6-31G(d)	0.20 (0.16)	0.46 (0.45)	0.32 (0.35)	0.89 (0.91)
6-31+G(d,p)	0.13	0.41	0.30	0.82
6-311+G(2d,2p)	0.11	0.39	0.29	0.79

[a] The values obtained by including vibrational corrections are in parentheses. The energy of the thG-H1 tautomer is taken as 0.

Bulk solvent effects on the electronic states are accounted for with the polarizable continuum model (PCM).[13] The excitation and emission energies, νA and νE, are computed with the "standard" linear response (LR) implementation of PCM/TD-DFT, which has been also used in the excited-state geometry optimizations.[14] To verify the effect of the explicit inclusion of solute–solvent interaction on the optical properties of thG, we have included six

H$_2$O (Figure S7a, b) or one dioxane (Figure S7c) molecule. In the excited-state geometry, the first solvation shell was fully optimized, i.e., was treated like solute degrees of freedom.
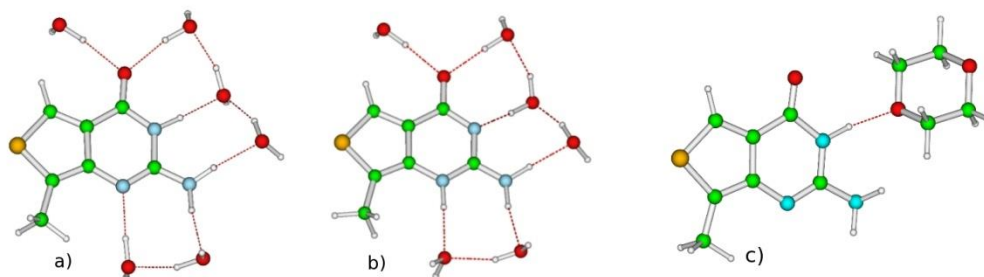
Figure S7. Computational models used to study the absorption and emission spectrum of the thG-H1 and thG-H3 tautomers in water (with six water molecules of the first solvation shell) (a and b, respectively) and thG-H1 in dioxane, with one explicit dioxane molecule (c).

In order to support the solidity of our computational analysis in water, we checked that our prediction do not qualitatively depend on the number of water molecules explicitly included in the computational model. In Table S2, we report the data obtained when only two water molecules, namely those H-bonded to the carbonyl group, are included in our model. In comparison with the calculations using six water molecules, the minimal model with two molecules of water already accounts for ~70% of the difference with respect to pure PCM calculations. Thus, the inclusion of only two water molecules, whose presence in the first solvation shell is extremely likely, is sufficient to provide similar stabilities for thG-H3 and thG-H1 tautomers.

Furthermore, analysis of solute-solvent H-bond interactions performed by G, dthG-H1 or dthG-H3 within the context of the (−)PBS molecule (Figure S12 structure #10, solvent exposed base) clearly confirmed that the solvation shell used in QM calculations may reflect the solvation shell observed in the DNA environment. Indeed, analysis of the solvation shell issuing from the MD simulations in the (−)PBS molecule shows the existence of an average of

~3.5 and ~2.8 H-bonds between the solvent molecules and the d^{th}G-H1 and d^{th}G-H3 tautomers, respectively (Figure S8). Considering that the thresholds for the definition of the solute-solvent H-bonds are rather strict (distance between the heavy atoms < 3 Å and angle > 135° (CPPTRAJ default values, see Roe et al. [15]) and that our QM calculation shows that 2 solute-solvent hydrogen bonds are sufficient to get the stability of the two tautomers very close, we can infer that the conclusions provided by QM calculations on the isolated bases apply also to the study of d^{th}G in a DNA environment, such as in (−)PBS.
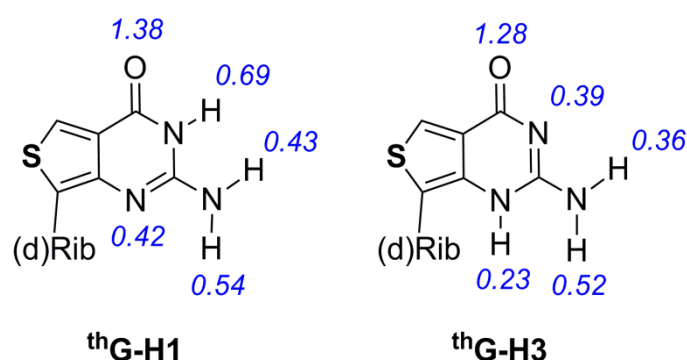
Figure S8. Average number of H-bonds calculated along MD trajectories for d^{th}G-H1 and d^{th}G-H3 in (−)PBS.

Finally, to further confirm the relevance of the computational approaches used for thG, we applied them to 9-methyl-guanine for which estimates based on alternative solvation models (MD/FEP) are available. For 9-methyl-guanine•$5H_2O$, including bulk solvent effect at the PCM level, the keto-amino (G-H1) was found to be more stable in aqueous solution than the enol (G-OH) tautomer by 7.1 kcal/mol according to PCM/PBE0/6-311+G(2d,2p)//PCM/PBE0/6-31G(d) calculations and by 6.2 kcal/mol according to PCM/M052X/6-311+G(2d,2p)//PCM/M052X/6-31G(d) calculations. Both values are very close to the estimates (5.7 - 7.5 kcal/mol) obtained by computational methods adopting MD/TI approaches to compute hydration energy. [16]

Table S4. Vertical Absorption and Emission Wavelengths Computed for the [th]G-H1 and [th]G-H3 Tautomers at the PCM/TD-DFT/6-311+G(2d,2p) Level in Different Solvents and Solvation Models by Using M052X and PBE0 Functionals [a]

| [th]G-H1 tautomer | | | | [th]G-H3 tautomer | | | |
| PBE0 | | M052X | | PBE0 | | M052X | |
Abs	Em	Abs	Em	Abs	Em	Abs	Em
Gas phase							
317	406	288	368	286	340	261	330
(0.08)	(0.06)	(0.11)	(0.08)	(0.11)	(0.08) [b]	(0.15)	(0.09) [b]
Dioxane (PCM)							
325	416	292	376	292	352	267	323
(0.10)	(0.08)	(0.14)	(0.11)	(0.14)	(0.11)	(0.19)	(0.16)
Dioxane (PCM + 1 Dioxane)							
330	418	295	378	293	353	269	332
(0.11)	(0.08)	(0.15)	(0.11)	(0.13)	(0.12)	(0.20)	(0.14)
Water (PCM)							
331	422	296	381	299	362	273	333
(0.09)	(0.08)	(0.16)	(0.11)	(0.12)	(0.10)	(0.16)	(0.13)
Water (PCM + 6H₂O)							
350	448	308	402	309	383	279	349
(0.08)	(0.07)	(0.12)	(0.10)	(0.10)	(0.08)	(0.14)	(0.11)

[a] Oscillator strengths are given in parentheses. Geometry optimizations are at the PCM/TD-DFT/6-31G(d) level. [b] 6-31+G(d,p) optimized geometries were used.
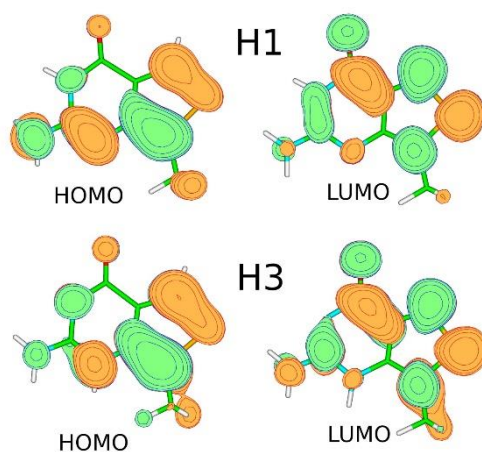
Figure S9. Frontier orbitals involved in the $S_0 \rightarrow S_1$ transition of the thG-H1 and thG-H3 tautomers.
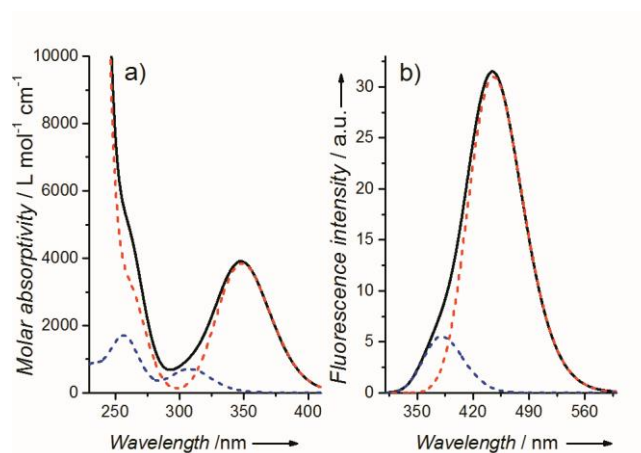
Figure S10. Absorption (a) and emission (b) spectra of the thG-H1 (red) and thG-H3 (blue) tautomers. The individual spectra and their sum weighted for their molar fraction were computed at PCM/TD-PBE0/6-311+G(2d,2p)/PCM/ PBE0/6-31G(d) level. The transitions have been convoluted by a phenomenological Gaussian with a HWHM of 0.25 eV.
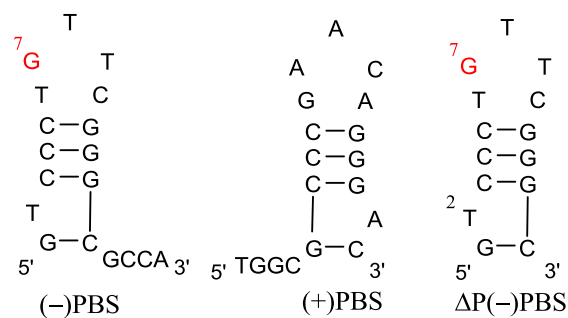
Figure S11. Structures of the primer binding site (PBS) oligonucleotides used in this study.

Table S5. Spectroscopic Parameters and Free Energy of Binding of dthG-containing Matched and Mismatched Duplexes [a]

Opposite base to dthG7	I_{375}/I_{550}	QY^{H1}	$\lambda_{em}^{max\,(H1)}$	ΔG^{H1}	ΔG^{H3}
dC	0.13 ± 0.02	0.20 [b]	461 ± 1	-71 ± 1	-58 ± 1
dT	0.14 ± 0.01	0.38 [b]	467 ± 1	-59 ± 1	-48 ± 1
dG	0.35 ± 0.01	-	462 ± 1	-54 ± 1	-53 ± 1
dA	0.65 ± 0.03	-	454 ± 1	-52.5 ± 0.8	-50 ± 1

[a]Fluorescence emission spectra of dthG-containing matched and mismatched duplexes were recorded at 310 nm excitation wavelength; I_{375}/I_{550} is the ratio of emission intensities at 375 and 550 nm; QY^{H1} is the fluorescence quantum yield of the dthG-H1 tautomer; $\lambda_{em}^{max(H1)}$ (in nm) is the fluorescence emission maxima of the dthG-H1 tautomer; $\Delta G^{H1/H3}$ is the free energy of binding (kcal/mol) of (+)PBS to (−)PBS where dG7 was substituted by dthG-H1 or dthG-H3 tautomer estimated using Molecular Mechanics Poisson-Boltzmann Surface Area approach along MD trajectories. The I_{375}/I_{550} and $\lambda_{em}^{max(H1)}$ values are the means from three experiments. [b] Values reported in [2].

4. Molecular Dynamics simulations

4.1 Molecular dynamics simulations using the AMBER force field

The family of ΔP(−)PBS DNA NMR structures[17] was provided by Nelly Morellet (personal communication). Two of these structures, namely structures #1 and #10, were selected as representative examples where the nucleobase at position 7 in the loop was in a π-π stacked or solvent-exposed conformation, respectively (Figure S12). The (−)/(+)PBS duplex structure was built with the Nucleic Acid Builder (NAB) molecular manipulation language, and the dG nucleotide at position 7 was manually replaced with dthG-H1 or dthG-H3.[18] For both dthG tautomers, bond lengths and partial charges were obtained by full geometry optimization at the DFT level, using the B3LYP functional in conjunction with the 6-311++G(d,p) basis set. The remaining parameters were taken from the Amber ff12SB force field[19] which adds novel torsional parameters for backbone and side chain torsions to the ff99SB force field for proteins, and includes the torsional modifications already validated in the ff99bsc0 force field for nucleic acids. Each macromolecular system investigated by MD was solvated by a cubic box of TIP3P-typed water[20] buffering 8 Å from the molecular surface and the total charge was neutralized by the addition of Na$^+$ counter ions. Water molecules were energy minimized for 500 steps using the Steepest Descent algorithm (SD) and a further 1500 steps using the Conjugate Gradient algorithm (CG), while keeping the solute as fixed. Removing the constraints, the solvated solute was energy minimized for 1000 steps using the SD and 4000 steps using the CG before being heated at constant volume from 0 to 300 K over 100 ps and using the Langevin thermostat. A density equilibration was carried out at constant pressure (NPT ensemble) for 100 ps, before running the production of unbiased MD trajectories for 200 ns. The frame with the lowest Root Mean Square Deviation (RMSD) with respect to the average structure of each system was extracted and used for

graphical representations of (−)/(+)PBS DNA, while the representative structure of the most populated cluster of ΔP(−)PBS DNA trajectories was selected for graphical representation. The binding free energy between (−)PBS and (+)PBS DNA was estimated by means of the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) approach implemented in Amber12.[21] All MD simulations were performed with Amber12 on two Nvidia K20 GPUs.[19b]
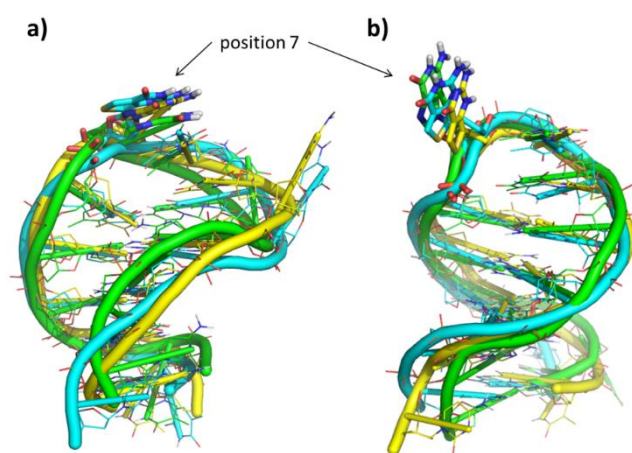
Figure S12. Superimposition of representative structures of the most populated clusters extrapolated from MD trajectories of ΔP(−)PBS single-stranded DNA. Representative structures from MD simulations of structure #1 (a) and structure #10 (b). The nucleotide at position 7 highlighted by a black arrow is shown as sticks. ΔP(−)PBS DNA bearing dG at position 7 is in green, that bearing d^{th}G-H1 is in cyan and that bearing d^{th}G-H3 is in yellow.

MD simulations of single-stranded ΔP(−)PBS DNA. The introduction of d^{th}G-H1 or -H3 at position 7 in the single-stranded loop of the ΔP(−)PBS does not affect the overall conformation of the oligomer (Figure S13a, b). The total energy of each ΔP(−)PBS system was nearly constant in both structures #1 and #10 during MD simulations, whereas structure #1 proved to be thermodynamically more stable than #10 (compare Figure S14a and b). Moreover, we checked whether the d^{th}G tautomers could influence the conformational

population of ΔP(−)PBS DNA. MD trajectories of the single-stranded loop of wild-type ΔP(−)PBS and ΔP(−)PBS bearing dthG-H1 or dthG-H3 at position 7 were clustered. Results clearly show that in all MD simulations one cluster of loop conformations is predominant, occurring between 45 and 73% of MD frames (Figure S12). Notably, the conformation of the nucleoside at position 7 is highly comparable in the representative structures of the predominant cluster of wild-type ΔP(−)PBS and the two dthG-bearing systems, thus suggesting that dG replacement with dthG-H1 or dthG-H3 in single stranded ΔP(−)PBS does not alter the conformational preferences of ΔP(−)PBS.
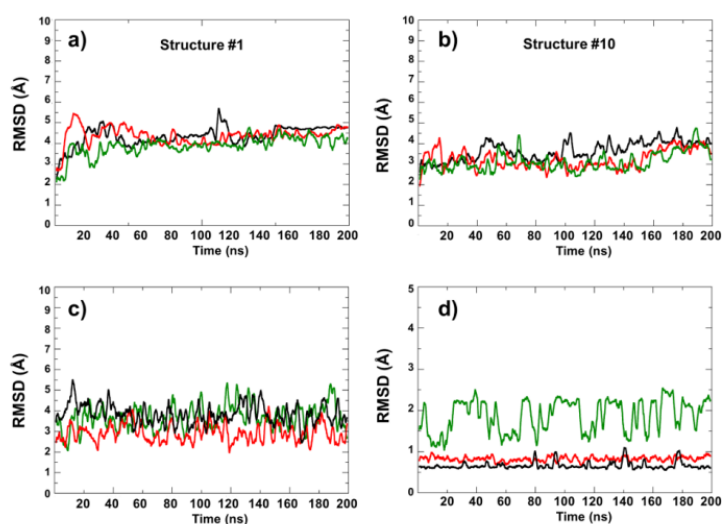
Figure S13. RMSD variation of (−)PBS and (−)/(+)PBS duplex. (a-b) RMSD variation of the single-stranded (−)PBS DNA structure #1 (a) and structure #10 (b) along 200 ns of unbiased MD trajectories. In black: unmodified (−)PBS having dG at position 7; in red: dthG7(−)PBS bearing dthG-H1 tautomer at position 7; in green: dthG7(−)PBS bearing dthG-H3 tautomer at position 7. (c-d) RMSD variation of the whole (−)/(+)PBS DNA duplex (c) and the dG (or dthG)-C Watson-Crick base pair at position 7 (d) along 200 ns of unbiased MD trajectories. Color codes for c-d are the same as for a-b.
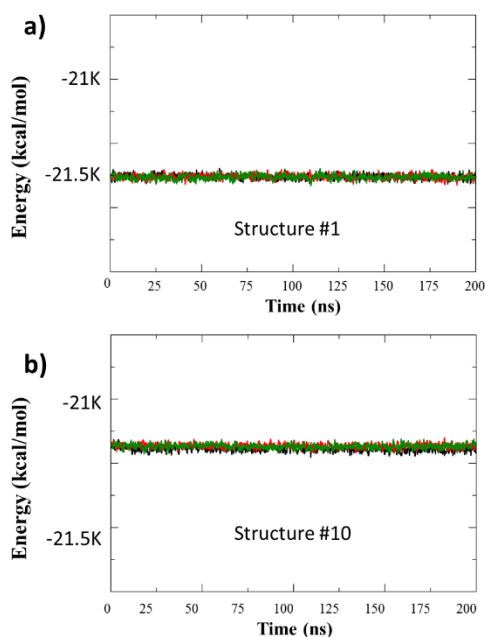
Figure S14. Total energy of the NMR-based structure #1 (a) and structure #10 (b) of ΔP(−)PBS DNA along 200 ns of unbiased MD simulation, bearing dG (black line), dthG-H1 tautomer (red line) or dthG-H3 tautomer (green line) at position 7.

MD simulations of (−)/(+)PBS DNA duplex. In the (−)/(+)PBS DNA duplex, the dG base at position 7 was kept unchanged or was replaced with dthG-H1 or dthG-H3, and each duplex was investigated by means of 200 ns of unbiased MD simulations. Analysis of RMSD along each MD trajectory unequivocally shows that replacing dG with dthG-H1 or dthG-H3 does not impact the overall geometry of the DNA duplex (Figure S13c). To further investigate the effects of replacing dG with dthG on the stability of the duplex, the free energy of binding of (−)PBS to (+)PBS in the duplex was estimated using MM-PBSA approach along MD trajectories. In line with the conformational findings described above and in comparison with the wild type (−)/(+)PBS (ΔG = −71 ± 1 kcal mol^{-1}), the introduction of dthG-H1 does not affect the stability of the DNA duplex (ΔG = −71.3 ± 0.8 kcal mol^{-1}), whereas the replacement of dG at position 7 with dthG-H3 induces a significant change (ΔG = −57.9 ± 0.8 kcal mol^{-1}).

This thermodynamic effect is probably due to the inability of dthG-H3 to establish canonical H-bond interactions with the opposite dC nucleobase in the duplex.
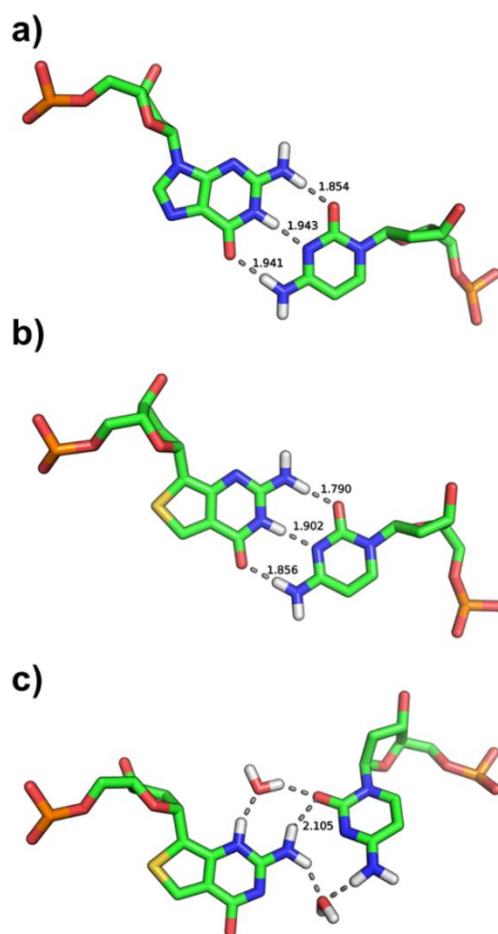
Figure S15. Structure of the base pair involving dG (a), dthG-H1 tautomer (b), and dthG-H3 tautomer (c) in the (−)/(+)PBS DNA duplex. Frames corresponding to the most representative structures of unbiased MD trajectories are shown. H-bond interactions are shown as grey dashed lines. Distances for direct and water-bridged H-bonds between bases are reported (in Å).
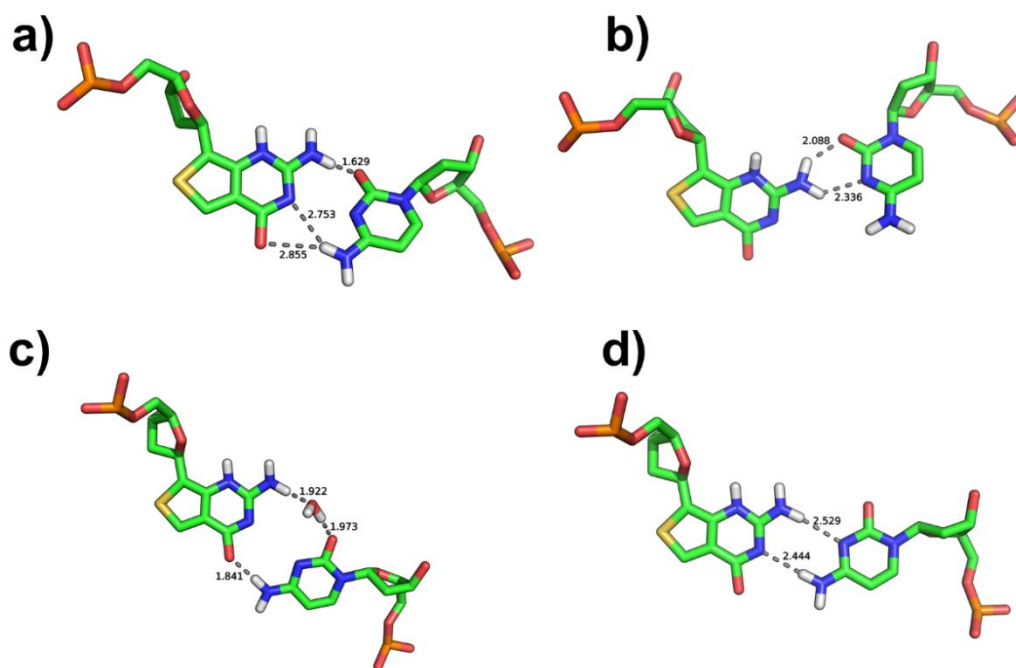
Figure S16. Four clusters of the dthG-H3 tautomer in various base-pair conformations with the opposite dC nucleobase. The most populated and stable cluster is shown in Figure S15c. H-bond interactions are showed as grey dashed lines. The distances of direct and water-bridged H-bonds between dthG and dC are shown.

Syn-anti conformation of the two dthG tautomers

To investigate the conformation adopted by the two dthG tautomers, the dihedral angle at the glycosidic bond was measured in MD simulations for the nucleotide at position 7 (dG, dthG-H1, or dthG-H3) of the (–)/(+)PBS duplex and the single stranded (–)PBS. Dihedral angle values comprised between -90° and +90° were assigned to the *anti* conformation of the nucleotide, whereas values comprised between -180° and -90° or +90° and +180° were assigned to the *syn* conformation. Results of this analysis (Table S6) unequivocally show that the preferential conformation of these nucleotides is the *anti*, particularly for dG, and dthG-H1 in both (–)/(+)PBS and (–)PBS systems. For dthG-H3, a slightly lower abundance of the *anti* nucleotide was observed in the (–)/(+)PBS duplex, whereas a more important variation was

observed in single stranded (–)PBS. In (–)PBS, the NH at position 3 of dthG-H3 can establish H-bond interactions with the phosphate group, which likely enhances the abundance of the *syn* conformation with respect to dG, and dthG-H1.

Table S6. Percentage of the syn or anti conformation of dG, dthG-H1, or dthG-H3 in MD simulations of (–)/(+)PBS and (–)PBS

	(–)/(+)PBS		(–)PBS	
	anti (%)	*syn* (%)	*anti* (%)	*syn* (%)
dG	92.5	7.5	89.4	10.6
dthG-H1	94.3	5.7	91.3	8.7
dthG-H3	87.5	12.5	63.2	36.8

Thus, our data strongly suggest that the *anti* conformation is preferred for dthG-H1 and dthG-H3 in the systems investigated in this work.

4.2 Molecular dynamics simulations using the CHARMM force field

Structure Preparation: To build the initial structures of the wild-type ΔP(–)PBS and ΔP(–)PBS bearing dthG-H1 or dthG-H3 at position 7, common heavy atom positions were retained from the experimental structures and the remaining heavy atoms were placed using the tools present in the CHARMM program.[22] The topology and parameters for the dthG tautomers were constructed based on similarity to groups and parameters existing in the CHARMM27 all-atom nucleic acid force field.[23] An initial energy minimization consisting of 1000 steps using the Steepest Descent method followed by 1000 steps of the Adapted Basis Newton-Raphson minimization method was realized in order to eliminate strong steric contacts prior to system solvation.

All-atom MD simulations set-up: Molecular dynamics simulations of the nucleic acids were done using the all-atom force field CHARMM27.[23] The system preparation and the analysis were done using the CHARMM program, while the simulations themselves were done using the NAMD program.[22] Periodic boundary conditions were used and the long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) algorithm.[24] All hydrogen-covalent bonds were constrained using the SHAKE algorithm[25] and an atom-based switching function with a cutoff of 12 Å was applied to the van der Waals non-bonded interactions. A 2 fs integration time step was used for all simulations. The water molecules were initially relaxed around a harmonically-constrained DNA by 5000 steps of Conjugate Gradient (CG) energy minimization, followed by 10000 steps without constraints. Subsequently, a molecular dynamics-based heating to 300 K over the course of 600 ps with the DNA harmonically constrained was done, followed by an equilibration phase, where the harmonic constraints were gradually removed over the course of 6 ns (the constraints were reduced every 500 ps). Pressure control was introduced during equilibration using a Berendsen piston[26] with a relaxation time of 400 fs and a rescaling of the atomic positions every 4 fs. The temperature was maintained using Langevin dynamics with a damping coefficient of 1 ps^{-1} applied to each atom. Finally, a 50-ns production simulation was performed in the isothermal–isobaric (NPT) ensemble at 1 bar and 300 K.

Simulations of the wild-type ΔP(−)PBS, as well as the dthG-H1- and dthG-H3-labelled ΔP(−)PBS, were run for 50 ns. The analysis included the calculation of the RMSD in order to quantify the structural behavior of the different tautomers. The same protocol was applied to the (−)/(+)PBS DNA duplex constructed with the same tautomers.

ΔP(−)PBS Simulation Results: The overall structure of the ΔP(−)PBS was reoriented over the backbone of bases 5-9 and the RMSD calculations of the base 7, which was dG, dthG-H1 or dthG-H3 tautomer (excluding the main chain), were performed. Average RMSDs at

the end of the simulations were: wild-type = 0.9 Å, d^{th}G-H1=0.9 Å and d^{th}G-H3=1.0 Å. The RMSD values show for the three constructs that there is no significant preference for the conformation of the base and that both tautomers can be present in the single-stranded ΔP(−)PBS.

(−)/(+)PBS *Simulation Results:* RMSD calculations were done in order to assess the structural deformation introduced by the substitution of dG7 by either d^{th}G-H1 or d^{th}G-H3 tautomers. From the simulations, the RMSDs were calculated in the same manner that is superposition over bases 5–9 and the calculation of the RMSD of base 7. An average RMSD was calculated over the final segment of the simulation yielding the following values for the three constructs: 0.5 Å for the wild-type, 0.6 Å for the d^{th}G-H1 and 0.7 Å for the d^{th}G-H3 tautomer in the final part of the trajectory.

While all the RMSDs are relatively small, the d^{th}G-H3 tautomer shows the largest deviation from the wild-type, which is qualitatively consistent with what was observed in the simulations done using the Amber force field. While the results do not contradict each other, the structural perturbations observed in the Amber force field simulations are somewhat larger, likely due to differences in the force fields.

When examining average energy minimized structures obtained from the end of the simulations, it was interesting to note that the d^{th}G-H3 took on a non-canonical conformation in its base pairing with dC similar to that of Figure S16d, in contrast to dG and d^{th}G-H1, which displayed a canonical H-bond network.

Further analysis was carried out using an MM-PBSA approach for free energy decomposition. We employed a protocol based on the MM-PBSA method described by Lafont et al,[27] which has proven to yield an accurate description of binding energetics. We looked at the by-base contribution to the binding energy. Individual contributions of each base to the complex formation were estimated. From the analysis, the wild-type dG7 and d^{th}G-H1

tautomer make larger contributions to binding, -3.8 ± 1.5 and -3.5 ± 0.7 kcal mol^{-1}, respectively, than the dthG-H3 tautomer which contributed -1.1 ± 0.6 kcal mol^{-1}. This suggests that the alternative dthG-H3 tautomer contributes less to the total binding free energy, in full line with the results presented in the main text, where the AMBER force field was used.

The results presented here confirm that the dthG-H1 tautomer mimics very well the dG base in both single-stranded and double-stranded DNA. The dthG-H3 tautomer does not show any distinct structural dynamics characteristic in the single-stranded DNA, but exhibits a non-canonical hydrogen bonding pattern in its base pairing with dC, so that it contributes less to the overall free energy of duplex formation than either the wild-type dG or the dthG-H1 tautomer.

References:

[1]   D. Shin, R. W. Sinkeldam, Y. Tor, *J. Am. Chem. Soc.* **2011**, *133*, 14912.

[2]   M. Sholokh, R. Sharma, D. Shin, R. Das, O. A. Zaporozhets, Y. Tor, Y. Mely, *J. Am. Chem. Soc.* **2015**, *137*, 3185.

[3]   H. de Rocquigny, D. Ficheux, C. Gabus, M. C. Fournie-Zaluski, J. L. Darlix, B. P. Roques, *Biochem. Biophys. Res. Commun.* **1991**, *180*, 1010.

[4]   C. Reichardt, *Chem. Rev.* **1994**, *94*, 2319.

[5]   M. J. Kamlet, J. L. M. Abboud, M. H. Abraham, R. W. Taft, *J. Org. Chem.* **1983**, *48*, 2877.

[6]   J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd ed., Springer, New York, **2006**.

[7]   a) C. Adamo, G. E. Scuseria, V. Barone, *J. Chem. Phys.* **1999**, *111*, 2889; b) M. Ernzerhof, G. E. Scuseria, *J. Chem. Phys.* **1999**, *110*, 5029; c) C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158.

[8]   a) Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Theory Comput.* **2006**, *2*, 364; b) Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157.

[9]   a) V. Karunakaran, K. Kleinermanns, R. Improta, S. A. Kovalenko, *J. Am. Chem. Soc.* **2009**, *131*, 5839; b) Y. Y. Zhang, R. Improta, B. Kohler, *Phys. Chem. Chem. Phys.* **2014**, *16*, 1487.

[10]   R. Improta, V. Barone, *Top. Curr. Chem.* **2015**, *355*, 329.

[11]   a) A. Dreuw, M. Head-Gordon, *Chem. Rev.* **2005**, *105*, 4009; b) R. Improta, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2656.

[12]   a) F. Santoro, V. Barone, R. Improta, *J. Am. Chem. Soc.* **2009**, *131*, 15232; b) R. Improta, V. Barone, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 12016; *Angew. Chem.* **2011**, *123*, 12222.

[13]   a) S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117; b) J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.

[14]   G. Scalmani, M. J. Frisch, B. Mennucci, J. Tomasi, R. Cammi, V. Barone, *J. Chem. Phys.* **2006**, *124*.

[15]   D. R. Roe, T. E. Cheatham, 3rd, *J. Chem. Theory Comput.* **2013**, *9*, 3084.

[16]   a) M. Kabelac, P. Hobza, *Phys. Chem. Chem. Phys.* **2007**, *9*, 903; b) M. Hanus, F. Ryjacek, M. Kabelac, T. Kubar, T. V. Bogdan, S. A. Trygubenko, P. Hobza, *J. Am. Chem. Soc.* **2003**, *125*, 7678; c) C. Colominas, F. J. Luque, M. Orozco, *J. Am. Chem. Soc.* **1996**, *118*, 6811; d) J. R. Blas, F. J. Luque, M. Orozco, *J. Am. Chem. Soc.* **2004**, *126*, 154.

[17] S. Bourbigot, N. Ramalanjaona, C. Boudier, G. F. J. Salgado, B. P. Roques, Y. Mély, S. Bouaziz, N. Morellet, *J. Mol. Biol.* **2008**, *383*, 1112.

[18] T. J. Macke, D. A. Case in *Molecular Modeling of Nucleic Acids*, (Eds.: N. B. Leontes, J. Santa Lucia Jr.), American Chemical Society, Washington DC, **1998**, pp. 379-393.

[19] a) V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712; b) D.A. Case, T. A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER12, University of California, San Francisco, **2012**.

[20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.

[21] a) M. Mori, U. Dietrich, F. Manetti, M. Botta, *J. Chem. Inf. Model.* **2010**, *50*, 638; b) B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, A. E. Roitberg, *J. Chem. Theory Comput.* **2012**, *8*, 3314.

[22] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545.

[23] N. Foloppe, A. D. MacKerell, *J. Comput. Chem.* **2000**, *21*, 86.

[24] H. G. Petersen, *J. Chem. Phys.* **1995**, *103*, 3668.

[25]   J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.

[26]   H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.

[27]   V. Lafont, M. Schaefer, R. H. Stote, D. Altschuh, A. Dejaegere, *Proteins* **2007**, *67*, 418.

4.1.2. Identifying the Fluorescence Lifetimes of [th]G Tautomers

We further determined the fluorescence lifetimes of [th]G tautomers in polar protic solvents (Water, methanol, ethanol) and apolar aprotic solvents (1,4 Dioxane and DMSO). All measurements were performed by exciting at 315 nm and emission was collected at emission wavelengths between 350 nm - 550 nm.
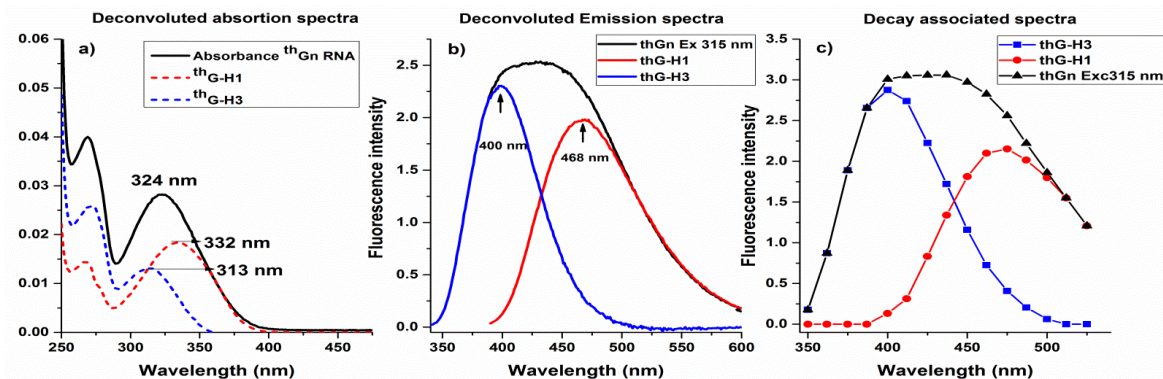
Figure 43. Deconvoluted a) absorption spectra, and b) emission spectra (excited at 315 nm), representing thG-H1 and thG-H3 tautomers for [th]G nucleoside in buffer. c) Decay associated spectra obtained from time resolved fluorescence spectroscopy

Analysis of the lifetime decay at each collected emission wavelength of [th]G in buffer, showed that the average lifetimes varies from 13.3 ns (350 nm) to 19.9 ns (550 nm) (Figure 43). When decay associated spectra (DAS) were compared with deconvoluted emission spectra, the 13.3 ns and 19.9 ns lifetimes were attributed to [th]G-H3 and [th]G-H1 tautomers, respectively. In the overlap region from 400 to 500 nm, we observed lifetime contribution from both the tautomers. Notably, QY of both the tautomers at different excitation wavelengths appears similar while the average lifetime changes from 13.3 ns to 19.9 ns with increasing emission wavelength. This suggests that both the radiative and non-radiative constants associated with them are different. Similarly, we performed the lifetime analysis of [th]G in other solvents of different polarity and proticity, namely methanol, ethanol, DMSO, and 1,4-Dioxane. Lifetimes and quantum yield of [th]G in these solvents were summarized in Table 2. In particular, the lifetime distribution of [th]G in methanol and ethanol also showed presence of a unstable <100 ps lifetime component that presently could not be assigned to any of the tautomers. Confirmation on existence of this short lifetime component is under investigation and it may likely represents to a fast-excited state process.

Table 2. Fluorescence quantum yield and lifetimes thG in various solvents.

	Solvent	QY	thG-H1, τ (ns)	thG-H3, τ (ns)
	Water	0.5	19.9	13.3
Polar protic	Methanol	0.5	14.7	8.2
	Ethanol	0.4	14.7	6.2
Polar Aprotic	DMSO	0.5	16.2	-
Apolar Aprotic	1,4-Dioxane	0.5	14.8	-

4.1.3. Photostability of thG as substituted in (–)PBS

Prior to study of the annealing kinetics of thG as substituted in (–)PBS, we investigated its photostability in free and substituted form. Initial investigation of the free probe revealed that the thG-H1 tautomer is comparatively more photostable than thG-H3 (Figure 44). Further, we examined the photobleaching of dthG as substituted in single stranded (–)PBS and (–)/(+)PBS duplex (Figure 45).
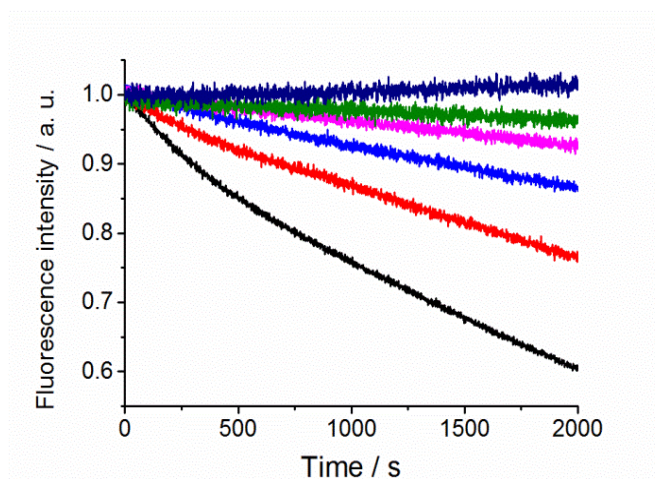
Figure 44. Photostability of the thG ground-state tautomers in buffer. Kinetics at different excitation and emission wavelengths, respectively: 325 and 400 nm (black); 325 and 525 nm (red); 350 and 525 nm (blue); 360 and 525 nm (magenta); 370 and 525 nm (green); 380 and 525 nm (dark blue). Concentration of thG was 1 µM.

The emission of thG-H1 can selectively be followed by exciting it from 350 nm to 400 nm. Meanwhile, at lower excitation wavelength < 350 nm both tautomers are observed. So, we decided to follow the photobleaching of thG-H1 by exciting at 325 & 360 nm whereas, collecting the emission at 525 nm.

As substituted in (–)PBS, a very sharp rate of photobleaching was observed at excitation of 325 nm as compared with 360 nm. In contrast, in duplex strands we observed comparable photobleaching at both excitation wavelengths leading to ~ 20% fluorescence loss in one hour This difference can be explained by the fact that in duplex (–)/(+)PBS only the thG-H1 tautomer is present, so that at both excitation wavelengths, we observe similar decays. In contrast, either for free thG or thG-labeled in (–)PBS, both tautomers are present. As the thG-H3 tautomer is much more sensitive than the thG-H1, we see a much faster decrease at 325 nm excitation wavelengths where both tautomers are excited than at 360 nm, where only the H1 tautomer absorbs light. Therefore, for the annealing kinetic reactions, the ~ 20% loss of fluorescence within one hour at 360 nm excitation wavelength appears as a limiting factor.
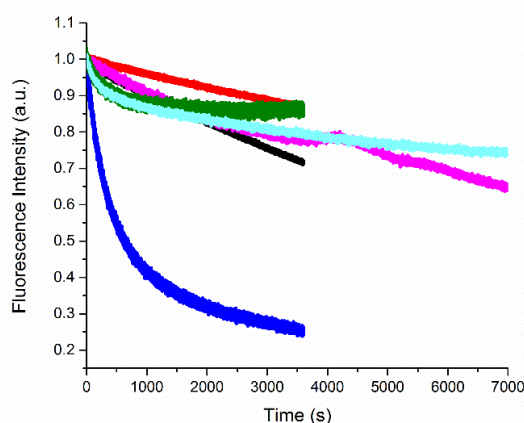
Figure 45. Photostability of thG nucleoside, dthG(–)PBS and dthG(–)/(+)PBS excited at 325 nm & 360 nm and the emission collected at 525 nm. thG nucleoside: 325 nm (Black), 360 nm (Red), dthG7(–)PBS: 325 nm (Blue), 360 nm (Magenta), dthG7(–)/(+)PBS: 325 nm (Green), 360 nm (Cyan). Concentration of thG in all samples 1μM.

4.1.3.1. Photobleaching During dthG7(–)(+)PBS Annealing Reaction.

We recorded the annealing kinetics of dthG7(–)/(+)PBS for three hours and compared it to the photobleaching kinetics of pre-prepared duplex as shown in Figure 46. The pre-prepared duplex was formed by heat annealing at 90°C for 3 mins and then cooling for 45 mins. For the annealing kinetic trace of dthG7 (–)/(+)PBS (instant mix), we observed an initial increase in fluorescence followed by a continuous decrease. In contrast, as the pre-prepared duplex is already formed, we only observe the decrease in fluorescence representing the photobleaching of the duplex. The temporal shift in the two kinetic traces is a consequence of the time needed to form the duplex (instant mix). Moreover, the final saturating fluorescence levels obtained from both samples ends up at same level at the end of long acquisition,

confirming that the second phase of the annealing kinetics trace is due to photobleaching of $d^{th}G7(-)/(+)PBS$.
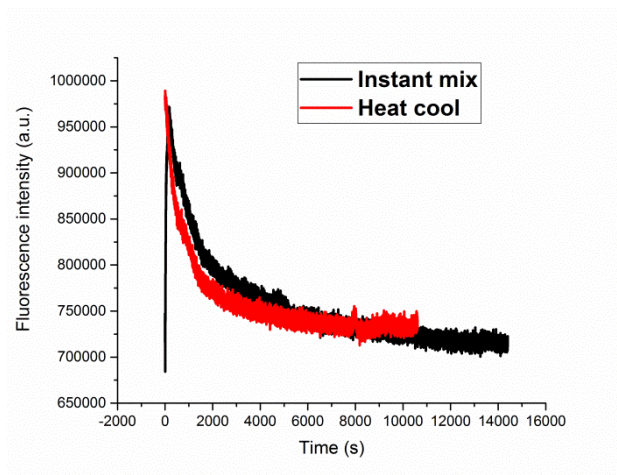
Figure 46. Fluorescence traces of $d^{th}G7(-)/(+)PBS$ obtained through annealing kinetics (Instant mix) and photobleaching of pre-formed duplex (Heat cool). Concentration of $d^{th}G(-)PBS$ 0.3µM and (+)PBS 3µM.

4.1.3.2.   Approaches to Minimize Photobleaching:

1. **Interval Scan**: In order to reduce the exposure of $d^{th}G7(-)/(+)PBS$, we recorded the annealing kinetics in a different way. First, we used a continuous kinetic scan till 200 seconds to catch the fast initial phase of annealing and then recorded the emission spectra with a time difference of 250 seconds. This kinetic trace was further compared to the continuous kinetic scan (Figure 47). We observed that by decreasing the exposure time, we reduced the bleaching effect confirming that the probe is sensitive to illumination in duplex form. Nevertheless, the fluorescence decrease is still significant, so that this approach cannot be used to extract the kinetic constants.
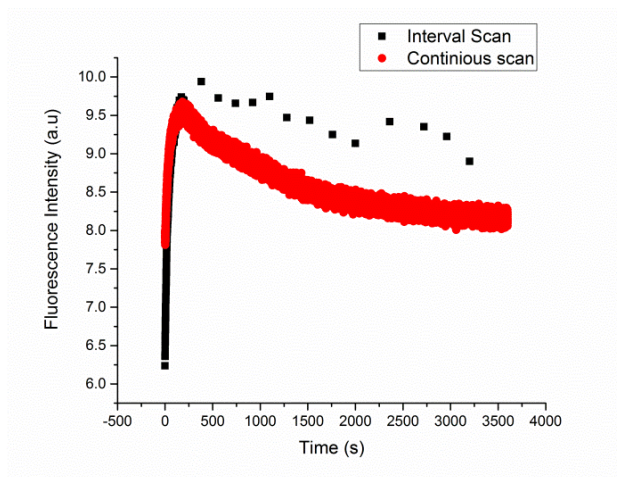
Figure 47. Interval scan and continuous can of dthG7(–)/(+)PBS. The annealing kinetics were performed using 0.3 µM thG7(–)PBS with 3µM (+)PBS for 3600 seconds.

2. **Chemical Approach:** To prolong the fluorescence in microscopy experiments, we use oxygen scavenging reagents that prevent oxidation of chemical structures. We used GLOX (glucose oxidase) reagents to investigate the effect of oxidation on free thG probe. We used 0.5 mg/ml gluocose oxidase, 40 µg/ml catalase and 10% glucose in TRIS buffer. Figure 48 shows the photobleaching of thG probe in TRIS buffer (black) and in presence of GLOX (red) reagents, excited at 360 nm. It appears that in presence of GLOX reagents the stability of thG is further decreased, indicating that this approach cannot be used to extract the kinetic constants.
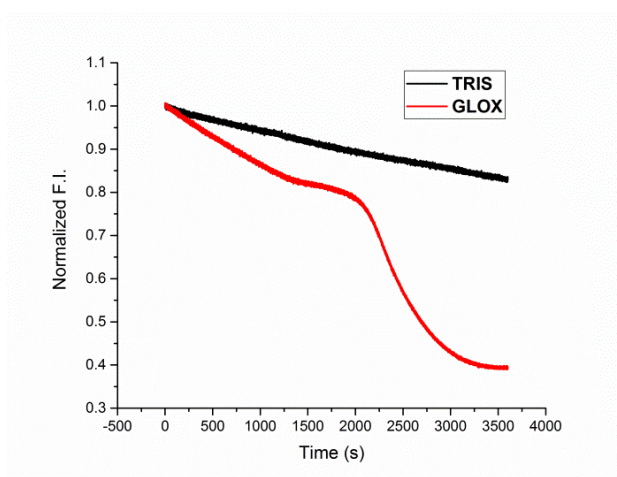
Figure 48. Photobleaching of 1 µM thG nucleoside in TRIS buffer(black) and in presence of GLOX reagents (red). Excitation 360 nm and emission 525 nm.

3. **Reducing Excitation Intensity:** Photobleaching can also be reduced by decreasing the excitation intensity. Lowering the excitation slits of the spectrophotometer reduces the

area of the excitation beam, so that lesser molecules within cuvette are excited and quickly replaced due to diffusion. To compensate for the reduced signal strength of sample from low excitation intensity, increasing the emission slits allows to detect more emitted photons. We monitored the annealing of thG7(–)/(+)PBS under these parameters, 1 nm (excitation) and 29 nm (emission) (Figure 49), and observed a significant decrease of photobleaching, that allows to monitor the annealing kinetics.
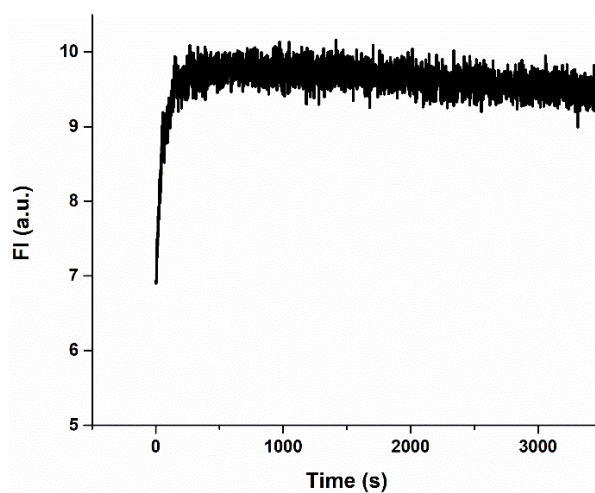
Figure 49. Annealing kinetics performed using 0.3 µM thG7(–)PBS with 3µM (+)PBS with reduced excitation slits (1 nm).

This photostability investigation shows that the dthG probe as substituted in duplex (–)/(+)PBS undergoes photochemical reaction leading to its fluorescence quenching. In order to successfully monitor the annealing reaction, the excitation intensity of the spectrophotometer was reduced to minimum that significantly lowered the rate of photobleaching.

4.1.4. Publication 2: Conquering 2-Aminopurine's Deficiencies: Highly Emissive Isomorphic Guanosine Surrogate Faithfully Monitors Guanosine Conformation and Dynamics in DNA

After establishing the ground rules to selectively maneuver the emission of both the tautomers as substituted at position 7 of (–)PBS stem-loop and its duplex (–)/(+)PBS, we compared its photophysics with 2-Aminopurine (2Ap), the "Gold standard" in the fluorescent nucleoside analogues. Publication 2 highlights this comparison by differentiating the ability of both probes in providing relevant structural and dynamic information on G7 residue of (–)PBS, together in its single and double-stranded structure.

The effect on these probes on the thermodynamic stability of (–)PBS and (–)/(+)PBS structures, was monitored through the melting temperatures (Publication 2, Table1). It revealed that replacement of G7 by $d^{th}G$ has minimal impact on the stability of stem-loop Similarly, identical melting temperatures of native (−)/(+)PBS and $d^{th}G7(−)/(+)$PBS indicate that $d^{th}G$ also perfectly substitutes for dG in the duplex by establishing a stable W-C base pair. In 2Ap labeled stem-loop sequences the melting temperature of d2Ap7(–)PBS was relatively similar to its native counterpart. But in the duplex structure, there was a deviation of 7 °C for d2Ap-C mismatch and 5 °C for the d2Ap-dT match.

Comparing the emission profiles prove interesting as the QY of $d^{th}G7(–)/(+)$PBS duplex was 2-fold higher than that of $d^{th}G7(–)$PBS and 10-fold higher than d2Ap7(–)/T_{12}(+)PBS (Publication 2, Table1). These differences show that the strong fluorescence quenching of dAp which occurs due to stacking with adjacent neighboring residues, was minimal for $d^{th}G$.

The effect of static and dynamic quenching could also be seen on the lifetime parameters of d2Ap- and $d^{th}G7$- labeled (–)PBS. In comparison to the four lifetime decay components of d2Ap7(–)PBS, $d^{th}G7(–)$PBS was only fitted with three components (Publication 2, Table1). In contrast to d2Ap7(–)PBS having 48 % of dark species, $d^{th}G7(–)$PBS shows less than 10 % of these species, indicating that all $d^{th}G$ conformations can be monitored. These differences are even more striking in (–)/(+)PBS duplex, where d2Ap7(–)/T_{12}(+)PBS showed four discrete lifetime components and a total of 98 % species having lifetime less than 0.5 ns, while $d^{th}G7(–)/(+)$PBS has only two lifetime components with marginal amount of dark species.

Lastly, by time-resolved anisotropy measurements, we observed two rotational correlation times for d2Ap7(–)PBS. The short 290 ps component likely represents the local rotation of the solvent-exposed extrahelical conformation (Publication 2, Table2, Figure S7) while, the longer correlation time of 1.9 ns represents a combination of segmental and tumbling motion of the loop. The anisotropy decay of dthG7(–)PBS exhibits only one component of 2.4 ns that matches with the theoretical correlation time of a folded loop. This indicates that the conformation of dthG, associated with 12.3 ns lifetime, is rigidly held in (−)PBS loop and only the tumbling of the entire (–)PBS is perceived. Consequently, the biomolecular quenching constant, k_q, observed for dthG7(–)PBS in iodine quenching experiment was found to be more than one order of magnitude lower than for free dthG nucleoside. In contrast, the k_q value for d2Ap7(–)PBS was observed to be only two-fold lower than for the free probe confirming its extrahelical positioning. In duplex structures, both dthG7(–)/(+)PBS and d2Ap7(–)/T$_{12}$(+)PBS show a single correlation time of 8.1 and 2.4 ns which represents the tumbling motion and segmental motion of the duplex, respectively.

Altogether these results showed that compared to d2Ap, conformations of dthG represent closer resemblance to native dG residue. In addition, dthG7(–)PBS shows higher fluorescence quantum yield and simpler lifetime decay with lesser dark species in both single-stranded and double-stranded forms. Importantly, thG7(–)PBS showed stable Watson-Crick base pairing and two-fold higher fluorescence emission upon duplex formation, this serves as a perfect tool for monitoring annealing kinetics. Conclusively, this publication clearly illustrates that dthG can faithfully substitute a key G7 residue in this HIV-1 construct, providing reliable information on its conformations and dynamics in both the (−)PBS stem loop and (−)/(+)PBS duplex.

Conquering 2-Aminopurine's Deficiencies: Highly Emissive Isomorphic Guanosine Surrogate Faithfully Monitors Guanosine Conformation and Dynamics in DNA

Marianna Sholokh,[†,§] Rajhans Sharma,[†] Dongwon Shin,[‡] Ranjan Das,[‖] Olga A. Zaporozhets,[§] Yitzhak Tor,*[‡] and Yves Mély*[†]

[†]Laboratoire de Biophotonique et Pharmacologie, Faculté de Pharmacie, UMR 7213 CNRS, Université de Strasbourg, 74 route du Rhin, 67401 Illkirch, France

[§]Department of Chemistry, Kyiv National Taras Shevchenko University, 01033 Kyiv, Ukraine

[‡]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093-0358, United States

[‖]Department of Chemistry, West Bengal State University, Barasat, Kolkata 700126, West Bengal, India

Ⓢ *Supporting Information*

ABSTRACT: The archetypical fluorescent nucleoside analog, 2-aminopurine (2Ap), has been used in countless assays, though it suffers from very low quantum yield, especially when included in double strands, and from the fact that its residual emission frequently does not represent biologically relevant conformations. To conquer 2Ap's deficiencies, deoxythienoguanosine (d[th]G) was recently developed. Here, steady-state and time-resolved fluorescence spectroscopy was used to compare the ability of 2Ap and d[th]G, to substitute and provide relevant structural and dynamical information on a key G residue in the (−) DNA copy of the HIV-1 primer binding site, (−)PBS, both in its stem loop conformation and in the corresponding (−)/(+)PBS duplex. In contrast to 2Ap, this fluorescent nucleoside when included in (−)PBS or (−)/(+)PBS duplex fully preserves their stability and exhibits a respectable quantum yield and a simple fluorescence decay, with marginal amounts of dark species. In further contrast to 2Ap, the fluorescently detected d[th]G species reflect the predominantly populated G conformers, which allows exploring their relevant dynamics. Being able to perfectly substitute G residues, d[th]G will transform nucleic acid biophysics by allowing, for the first time, to selectively and faithfully monitor the conformations and dynamics of a given G residue in a DNA sequence.

Figure 1. Structure of d2Ap (**1**), d[th]G (**2**), and the naturally occurring purines (R = 2′-deoxy-D-ribose).

For almost five decades, 2-aminopurine (2Ap, **1**) has been the fluorescent nucleoside of choice for the community interested in nucleic acid structure, dynamics and recognition.[1] Despite its isomerized base-pairing face, numerous fluorescence-based assays have used this isomorphic nucleoside analog as an emissive replacement for adenosine and guanosine (Figure 1), due to its small footprint, high emission quantum yield (QY = 0.68), and availability.[1a,2] However, challenges have been recognized, including 2-Ap's propensity to mispair with C and its tendency to perturb the dynamics and structure of DNA.[3] Additionally, 2-Ap's strong emission quenching upon incorporation into single-stranded and particularly double-

stranded oligonucleotides (ODNs) has been commonly observed.[2b,4] What appears to have been largely neglected is that the residual emission observed for such DNA and RNA constructs, although sufficient for numerous biophysical applications, frequently does not represent biologically relevant conformations of the native nucleoside replaced. The structural and dynamics information thus gathered might not actually reflect the behavior of the native system of interest. Here we demonstrate that this is indeed the case for the primer binding site (PBS) of the human immunodeficiency virus type 1 (HIV-1), and present 2-aminothieno[3,4-*d*]pyrimidin-4(3*H*)-one-7-*β*-D-2′-deoxyribofuranoside[5] (deoxythienoguanosine, d[th]G) (**2**) as a truly faithful emissive and responsive surrogate for G in single- and double-stranded ODNs, which actually reproduces the structural context and dynamics of the parent native nucleoside.

The PBS DNA sequence is an 18-mer stem-loop ODN of known 3D structure,[6] which is involved in the second strand transfer of HIV-1 reverse transcription (Figure 2).[7] This strand transfer, relying on the annealing of (−)PBS with its complementary (+)PBS sequence,[8] is required for completing the viral DNA synthesis. To compare the ability of d2Ap and d[th]G to provide structural and dynamic information on the stem-loop and the corresponding perfect and mismatched
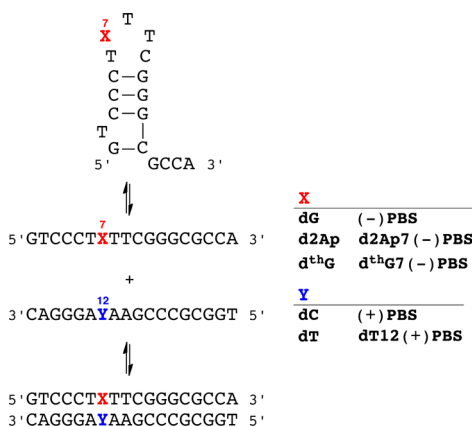
Figure 2. DNA sequence of the HIV-1 primer binding site (−)PBS shown as a single strand (middle), stem loop (top), and duplex with (+)PBS (bottom). Also shown are the site-specifically modified sequences containing d2Ap and dthG.

Figure 3. Emission spectra of (a) dthG- and (b) d2Ap-labeled (−)PBS sequences. Emission spectra of (a) dthG free nucleoside (black), dthG7(−)PBS (red), dthG7(−)/(+)PBS (blue), and dthG7(−)/T12-(+)PBS (green); (b) d2Ap free nucleoside (black), d2Ap7(−)PBS (red), and d2Ap7(−)/T12(+)PBS (blue). Excitation was at 380 nm for dthG and 315 nm for d2Ap. Nucleoside and ODN concentration was 6 μM for dthG and 4 μM for d2Ap in 25 mM TRIS-HCl buffer (pH 7.5), 30 mM NaCl, and 0.2 mM MgCl$_2$.

duplexes, we substitute the critical G7 loop residue with the two emissive deoxynucleosides and thoroughly analyze the biophysical and photophysical features of all constructs (see Supporting Information for synthetic and additional experimental details).

Thermal denaturation experiments reveal that replacement of G7 by dthG has a minimal impact on the stability of the (−)PBS stem-loop (50 $\pm$ 1 and 51 $\pm$ 1 °C, respectively) (Table 1). Similarly, the identical melting temperature of the native and the dthG7(−)/(+)PBS duplexes (67 $\pm$ 1 and 67 $\pm$ 2 °C, respectively) indicate that dthG also perfectly substitutes for dG in the duplex. Additionally, replacement of the pairing C12 by T in (+)PBS, forming a dthG-dT mismatch, results in a 6 °C decrease in the T_m, in excellent agreement with the $\Delta T_m = -7$ °C observed for the corresponding dG-dT mismatch.[5b] While substitution by d2Ap only slightly affects the stability of the (−)PBS stem-loop, it decreases the stability of the (−)/(+)PBS duplex by 7 °C, likely due to the formation of an unstable d2Ap-dC mismatch.[9] Notably, the "perfect duplex" d2Ap7(−)/T12(+)PBS is still 5 °C less stable than the native or the dthG7(−)/(+)PBS duplex, indicating that in contrast to dthG, d2Ap does not faultlessly substitute for dG in this context.

The free dthG nucleoside (2) emits in the blue with a QY of 0.46 $\pm$ 0.02 in buffer (Figure 3 and Table 1). When incorporated into position 7 in the (−)PBS loop, the QY drops to 0.10 $\pm$ 0.01, but increases 2-fold upon hybridization to its perfect complement to form dthG7(−)/(+)PBS (Table 1). In sharp contrast to dthG, the near UV emission of d2Ap (1) is severely quenched upon incorporation into ODNs, with 8-fold decrease for the stem-loop, and above 50-fold decrease upon forming the d2Ap7(−)/T12(+)PBS duplex (Table 1). Although displaying a significantly higher QY as a free nucleoside, its short emission wavelength and dramatic quenching in ODNs makes 2Ap a rather inferior emissive surrogate for G. The high QY of dthG in duplexes constitutes, therefore, an obvious asset over d2Ap for monitoring the single to double strand transition and for characterizing the dynamic properties of the substituted base, as discussed below. Moreover, while nearly no wavelength shift is observed for d2Ap in its distinct states, shifts of 5 and 12 nm were observed in the emission maxima of dthG7(−)PBS and dthG7(−)/(+)PBS, respectively (Figure 3), as compared to the free nucleoside 2. This dthG's responsiveness provides an additional spectroscopic handle for monitoring the biomolecular environment of this surrogate nucleoside.

While the two emissive nucleosides exhibit a single exponential decay, the corresponding modified ODNs display a more complex behavior (Table 1). Four decay components are observed for d2Ap7(−)PBS, indicating a large conformational heterogeneity of d2Ap in this loop position, as already described for other positions in the loop.[10] The three short lifetimes ($\tau_1 - \tau_3$) likely correspond to conformations where dynamic fluctuations of the loop facilitate dynamic quenching of d2Ap by its neighbors, through a charge transfer mechanism[11] or relaxation into a low-lying nonemissive electronic state.[12] The long-lived lifetime ($\tau_4 = 7.4$ ns), being close to that of the free nucleoside, likely corresponds to a

Table 1. Time-Resolved Fluorescence Parameters of d2Ap- and dthG-Labeled ODNs[a]

	T_m	QY	τ_1	α_1	τ_2	α_2	τ_3	α_3	τ_4	α_4	$\langle\tau\rangle$	α_0
d2Ap		0.68[b]							10.2	1	10.2	
d2Ap7(−)PBS	48 $\pm$ 1	0.08	0.15	0.16	0.66	0.10	2.6	0.15	7.4	0.11	2.4	0.48
d2Ap7(−)/T12(+)PBS	62 $\pm$ 1	0.013	0.18	0.20	0.44	0.27	1.4	0.01	5.2	0.01	0.4	0.51
dthG		0.46							19.6	1	19.6	
dthG7(−)PBS	51 $\pm$ 1	0.10			0.5	0.32	2.8	0.40	12.3	0.28	4.7	<0.1
dthG7(−)/(+)PBS	67 $\pm$ 2	0.20			1.1	0.17			11.3	0.83	9.6	<0.1
dthG7(−)/T12(+)PBS	61 $\pm$ 1	0.38			0.8	0.07	3.9	0.09	28.2	0.57	22.3	0.27

[a]T_m is the melting temperature (°C), QY is the fluorescence quantum yield, τ_i are the fluorescence lifetimes (ns), and α_i are their amplitudes. The amplitude α_0 of the dark species as well as the amplitudes α_i were calculated as described in the Supporting Information. $\langle\tau\rangle$ is the mean fluorescence lifetime (ns). Excitation and emission wavelengths were 315 and 370 nm for d2Ap and 315 and 500 nm for dthG. SDs for the lifetimes and amplitudes are <20%. SDs for QY are <10%. [b]Data from ref 1a.

conformation where d2Ap is extrahelical and distant from potential quenchers.[2b,13] Since the difference in the mean lifetime of d2Ap7(−)PBS as compared to the free d2Ap is markedly smaller than the difference seen for the QY (4.3- vs 8.5-fold), nonemissive "dark species", with lifetimes shorter than the detection limit of our setup (∼30 ps), are present.[10,11,14] This population, resulting from either static quenching or very fast dynamic quenching, represents a total of 48% (calculated from eq (1) in the Supporting Information).

Only three components are needed to fit the intensity decay of dthG7(−)PBS (Table 1). The long lifetime is close to the component measured for the free nucleoside in methanol (12.3 vs 13.7 ns),[5b] reflecting a minimally quenched dthG in the less polar environment of the (−)PBS loop.[15] The two other components are markedly shorter (0.5 and 2.8 ns), suggesting that they correspond to conformations where dthG is dynamically quenched by its neighboring nucleobases, likely through mechanisms comparable to those of d2Ap.

In contrast to d2Ap, however, comparison of the QY and mean lifetimes of dthG7(−)PBS with those of the free nucleoside reveals that the two evolve in parallel. Dark species are therefore negligible (<10%), which is a distinctive advantage over d2Ap, since all conformations of dthG in (−)PBS can therefore be monitored by the time-resolved measurements.

Differences between d2Ap and dthG become more pronounced in the (−)/(+)PBS duplex. The decay of d2Ap in d2Ap7(−)/T12(+)PBS is best fitted with four discrete lifetime components, ranging from 0.18 to 5.2 ns (Table 1). When comparing the duplex to the stem loop, a dramatic decrease in the amplitudes associated with the two long-lived lifetimes τ_3 and τ_4 is seen. A total of 98% of the species and thus of the d2Ap conformations in the d2Ap-labeled duplex exhibit lifetimes shorter than 0.5 ns, explaining its extremely low QY. These commonly observed features,[2b,13a,16] which severely limit the use of 2Ap in duplexes, likely originate from the destabilization induced by 2Ap in its own base pair and its immediate adjacent base pairs.[3] In line with the key role of conformational motions of DNA bases in charge transfer based quenching mechanisms,[17] the resulting increased dynamics likely favor efficient 2Ap quenching by its neighbors, explaining the multiple and mainly short-lived fluorescence lifetimes observed for 2Ap in double-stranded DNA.

In sharp contrast to the complex decay of d2Ap7(−)/T12(+)PBS, the decay of the corresponding dthG7(−)/(+)PBS duplex appears very simple, being characterized by only two lifetimes (1.1 and 11.3 ns) and a marginal fraction of dark species. This indicates that in contrast to d2Ap, dthG adopts better defined conformations, due to its ability to form a stable Watson−Crick base pair with C.[18] Thus, we attribute the major conformation (>80%) associated with the 11.3 ns component to the paired dthG in the rather apolar environment created by the stacked base pairs within the duplex.[15] This interpretation is further supported by the mismatched duplex dthG7(−)/T12(+)PBS, where the three lifetimes (0.8, 3.9, and 28.2 ns) and the significant amount of dark species (27%) reflect a greater conformational heterogeneity of dthG, as expected from the reduced constraints imposed by the dthG-dT mismatch compared to the Watson−Crick dthG-dC base pair. Similarly, the dramatic increase in the long-lived lifetime value (28.2 vs 11.3 ns, respectively), which is comparable to the lifetime value of dthG in water, suggests higher accessibility to water, as a result of the lesser constraints imposed by the dthG-dT mismatch in the duplex.

Table 2. Fluorescence Anisotropy Decay Parameters and Quenching Constants[a]

	θ_1	β_1	θ_2	β_2	k_q
d2Ap	0.08[b]	1.00			6.7
d2Ap7(−)PBS	0.29	0.52	1.9	0.48	3.5
d2Ap7(−)/T12(+)PBS			2.7	1.00	<10^{-3}
dthG	0.12	1.00			1.3
dthG7(−)PBS			2.4	1.00	0.09
dthG7(−)/(+)PBS			8.1	1.00	<10^{-3}
dthG7(−)/T12(+)PBS			8.4	1.00	<10^{-3}

[a]θ_i are the rotational correlation times (in ns) and β_i their amplitudes. The reported values are the means from three experiments. SDs for θ_i and β_i are <20%. k_q is the bimolecular quenching rate constant for the quenching by iodide (in 10^9 M^{-1} s^{-1}). The k_q values are the means from two experiments. SDs are <10% for this parameter. [b]Data from ref 10.

To further cement the picture painted above, we performed time-resolved anisotropy to provide information about the local, segmental and global motions of the labeled ODNs, as well as KI quenching experiments to quantitatively assess the solvent exposure of the emissive nucleosides within the ODNs (Table 2 and Figure S7). The free nucleosides d2Ap and dthG exhibit single rotational correlation times of 80 and 120 ps, respectively. Two correlation times were observed for d2Ap7(−)PBS. The short one (θ_1 = 290 ps) likely describes the local rotation of the solvent-exposed extrahelical d2Ap conformation, associated with the long-lived lifetime τ_4 = 7.4 ns, which contributes to more than 60% of the labeled ODN emission (as calculated by $\alpha_4\tau_4/\langle\tau\rangle$). This conclusion is further substantiated by the very high bimolecular quenching constant, k_q, observed for d2Ap7(−)PBS in iodide quenching experiments (Table 2). Indeed, this k_q value being only 2-fold lower than that of the free d2Ap nucleotide, unambiguously confirms that this extrahelical conformation is highly accessible to the solvent. The long correlation time (θ_2 = 1.9 ns) observed for d2Ap7(−)PBS was significantly shorter than the theoretical correlation time (2.5 ns) calculated for the tumbling of a sphere representing the stem-loop structure. Therefore, this θ_2 = 1.9 ns component may correspond to a combination of the (−)PBS tumbling motion and a segmental motion, likely associated with the loop.[10] In contrast, the anisotropy decay of dthG7(−)PBS is adequately fitted to only one component (2.4 ns) that matches with the theoretical correlation time of the folded ODN. This indicates that the conformations of dthG, associated with the 12.3 ns lifetime, are rigidly held in the (−)PBS loop and only the tumbling of the entire ODN is perceived. This behavior is fully consistent with the NMR structure of (−)PBS, showing that the G7 residue is directed toward the loop interior and well constrained by its neighbors.[6b] The internal orientation of dthG with poor solvent accessibility is further supported by the low k_q value observed with dthG7(−)PBS, that was more than 1 order of magnitude lower than that of the free nucleoside. Thus, time-resolved anisotropy and iodide quenching data confirm that dthG mimics the native G residue much more closely than d2Ap in the stem loop.

The anisotropy decay of d2Ap7(−)/T12(+)PBS could be fitted with a single component (2.7 ns) that is much shorter than the theoretical correlation time (9.6 ns) calculated for the tumbling motion of this duplex.[19] This likely reflects the segmental motions associated with the partially stacked d2Ap conformations that dominate the emission of d2Ap7(−)/

T12(+)PBS. In contrast, the anisotropy decay of the d^{th}G7(−)/(+)PBS, while also displaying a single correlation time, matches well with the theoretical correlation time of the tumbling duplex. This absence of segmental motion is fully consistent with the attribution of the dominant 11.3 ns lifetime component to the d^{th}G-dC base pair in its optimally stacked configuration. In this highly stable configuration, only the tumbling motion could be detected. Interestingly, a single correlation time (8.4 ns) describing the overall tumbling of the duplex was also observed for d^{th}G7(−)/T12(+)PBS, indicating that the major d^{th}G conformation associated with the 28.2 ns lifetime component is probably not extrahelical. Thus, in line with the high stability (>100 ms) of internal G-C base pairs and the absence of intrahelical dynamics (in the μs−ms range) in the central part of duplexes,[3,20] our data indicate that only d^{th}G but not 2Ap can be used to obtain relevant information on the oligonucleotide dynamics and size. Noticeably, for both d2Ap7- and d^{th}G7-labeled duplexes, the k_q values are at least 3 orders of magnitude below those of the free nucleosides, suggesting that the emissive nucleosides predominantly adopt an intrahelical conformation.

Taken together, our data clearly illustrate that d^{th}G can faithfully substitute a key G residue in this HIV-1 construct, providing reliable information on its conformations and dynamics in both the (−)PBS stem loop and (−)/(+)PBS duplex. Particularly beneficial are d^{th}G's reliable base pairing and its high emission QY, which is maintained in single- and double-stranded ODNs. As a result, and in sharp contrast to the corresponding d2Ap labeled ODNs, the species detected by d^{th}G fluorescence techniques, actually reflect the predominantly populated conformers as determined by other means, such as NMR. These features make this new emissive analog a perfect tool to faithfully monitor the conformations and dynamics of G residues in oligonucleotides. This will undoubtedly open a new era with the promise of properly addressing unsolved problems in nucleic acid biophysics.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Synthesis and fluorescence spectroscopy details. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors
*ytor@ucsd.edu
*yves.mely@unistra.fr

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) (a) Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem.* **1969**, *244*, 1228. (b) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579. (c) Hall, K. B. *Methods Enzymol.* **2009**, *469*, 269.

(2) (a) Rachofsky, E. L.; Osman, R.; Ross, J. B. A. *Biochemistry* **2001**, *40*, 946. (b) Guest, C. R. *Biochemistry* **1991**, *30*, 3271. (c) Jean, J. M.; Hall, K. B. *Biochemistry* **2002**, *41*, 13152.

(3) Dallmann, A.; Dehmel, L.; Peters, T.; Mügge, C.; Griesinger, C.; Tuma, J.; Ernsting, N. P. *Angew. Chem., Int. Ed.* **2010**, *49*, 5989.

(4) (a) Nag, N.; Ramreddy, T.; Kombrabail, M.; Krishna Mohan, P. M.; D'souza, J.; Rao, B. J.; Duportail, G.; Mely, Y.; Krishnamoorty, G. In *Reviews in Fluorescence*; Geddes, C., Lakowicz, J. R., Eds.; Springer Science: New York, 2006; pp 311−340. (b) Stivers, J. T. *Nucleic Acids Res.* **1998**, *26*, 3837. (c) Law, S. M.; Eritja, R.; Goodman, M. F.; Breslauer, K. J. *Biochemistry* **1996**, *35*, 12329. (d) Avilov, S. V.; Godet, J.; Piémont, E.; Mély, Y. *Biochemistry* **2009**, *48*, 2422. (e) Avilov, S. V.; Piemont, E.; Shvadchak, V.; de Rocquigny, H.; Mély, Y. *Nucleic Acids Res.* **2008**, *36*, 885.

(5) (a) Shin, D.; Sinkeldam, R. W.; Tor, Y. *J. Am. Chem. Soc.* **2011**, *133*, 14912. (b) Park, S.; Otomo, H.; Zheng, L.; Sugiyama, H. *Chem. Commun.* **2014**, *50*, 1573.

(6) (a) Johnson, P. E.; Turner, R. B.; Wu, Z. R.; Hairston, L.; Guo, J.; Levin, J. G.; Summers, M. F. *Biochemistry* **2000**, *39*, 9084. (b) Bourbigot, S.; Ramalanjaona, N.; Boudier, C.; Salgado, G. F. J.; Roques, B. P.; Mély, Y.; Bouaziz, S.; Morellet, N. *J. Mol. Biol.* **2008**, *383*, 1112.

(7) (a) Thomas, J. A.; Gorelick, R. J. *Virus Res.* **2008**, *134*, 39. (b) Poeschla, E. *Virology* **2013**, *441*, 1.

(8) Basu, V. P.; Song, M.; Gao, L.; Rigby, S. T.; Hanson, M. N.; Bambara, R. A. *Virus Res.* **2008**, *134*, 19.

(9) Sowers, L. C.; Fazakerley, G. V.; Eritja, R.; Kaplan, B. E.; Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 5434.

(10) Godet, J.; Ramalanjaona, N.; Sharma, K. K.; Richert, L.; De Rocquigny, H.; Darlix, J. L.; Duportail, G.; Mély, Y. *Nucleic Acids Res.* **2011**, *39*, 6633.

(11) (a) Larsen, O. F. A.; Van Stokkum, I. H. M.; Gobets, B.; Van Grondelle, R.; Van Amerongen, H. *Biophys. J.* **2001**, *81*, 1115. (b) Wan, C.; Fiebig, T.; Schiemann, O.; Barton, J. K.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14052. (c) Fiebig, T.; Wan, C.; Zewail, A. H. *ChemPhysChem* **2002**, *3*, 781. (d) O'Neill, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2002**, *124*, 13053. (e) Kelley, S. O.; Barton, J. K. *Science* **1999**, *283*, 375.

(12) Jean, J. M.; Hall, K. B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 37.

(13) (a) Neely, R. K.; Daujotyte, D.; Grazulis, S.; Magennis, S. W.; Dryden, D. T. F.; Klimašauskas, S.; Jones, A. C. *Nucleic Acids Res.* **2005**, *33*, 6953. (b) Neely, R. K.; Magennis, S. W.; Parsons, S.; Jones, A. C. *ChemPhysChem* **2007**, *8*, 1095.

(14) Godet, J.; Kenfack, C.; Przybilla, F.; Richert, L.; Duportail, G.; Mély, Y. *Nucleic Acids Res.* **2013**, *41*, 5036.

(15) (a) Klymchenko, A. S.; Shvadchak, V. V.; Yushchenko, D. A.; Jain, N.; Mély, Y. *J. Phys. Chem. B* **2008**, *112*, 12050. (b) Ogawa, A. K.; Abou-Zied, O. K.; Tsui, V.; Jimenez, R.; Case, D. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 9917.

(16) (a) Gaied, N. B.; Glasser, N.; Ramalanjaona, N.; Beltz, H.; Wolff, P.; Marquet, R.; Burger, A.; Mély, Y. *Nucleic Acids Res.* **2005**, *33*, 1031. (b) Kenfack, C. A.; Piémont, E.; Ben Gaied, N.; Burger, A.; Mély, Y. *J. Phys. Chem. B* **2008**, *112*, 9736. (c) Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G. *J. Phys. Chem. B* **2007**, *111*, 5757.

(17) O'Neill, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2004**, *126*, 13234.

(18) Samanta, P. K.; Pati, S. K. *New J. Chem.* **2013**, *37*, 3640.

(19) Ortega, A.; Garcia De La Torre, J. *J. Chem. Phys.* **2003**, *119*, 9914.

(20) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., III *Nat. Commun.* **2014**, *5*, No. 5152.

Supporting Information

Conquering 2-Aminopurine's Deficiencies: Highly Emissive Isomorphic GuanosineSurrogate Faithfully Monitors Guanosine Conformation and Dynamics in DNA

Marianna Sholokh[†,§], Rajhans Sharma[†], Dongwon Shin[¶], Ranjan Das[‖], Olga A. Zaporozhets[§], Yitzhak Tor[¶*], and Yves Mély[†*]

[†]Laboratoire de Biophotonique et Pharmacologie, UMR 7213 CNRS, Université de Strasbourg, Faculté de Pharmacie, 74 route du Rhin, 67401 Illkirch, France

[§]Department of Chemistry, Kyiv National Taras Shevchenko University, 01033 Kyiv, Ukraine

[¶]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, 92093-0358 California, United States

[‖]Department of Chemistry, West Bengal State University, Barasat, Kolkata 700126, West Bengal, India

Table of contents

1. Synthesis

Reagents were purchased from Sigma-Aldrich, Fluka, TCI, Acros and Synchem, Inc. (Elk Grove, IL), and were used without further purification unless otherwise specified. Solvents were purchased from Sigma-Aldrich and Fisher Scientific, and dried by standard techniques. NMR solvents were purchased from Cambridge Isotope Laboratories (Andover, MA). All reactions were monitored with analytical TLC (Merck Kieselgel 60 F254). All experiments involving air and/or moisture sensitive compounds were carried out under an argon atmosphere. Column chromatography was carried out with silica gel particle size 40-63 μm by CombiFlash® Rf 200 (Teledyne Isco). NMR spectra were obtained on Varian Mercury 400 MHz, Varian VX 500 MHz and Jeol ECA 500 spectrometers. Mass spectra were obtained on an

Agilent 6230 HR-ESI-TOF MS at the Molecular Mass Spectrometry Facility at the UCSD Chemistry and Biochemistry Department. Modified oligonucleotide (ODN) was quantified by Shimadzu UV 2450 at 70 °C. MALDI-TOF spectra were recorded on a PE Biosystems Voyager-DE STR MALDI-TOF spectrometer in negative-ion, delayed-extraction mode.

1.1. dthGphosphoramidite synthesis
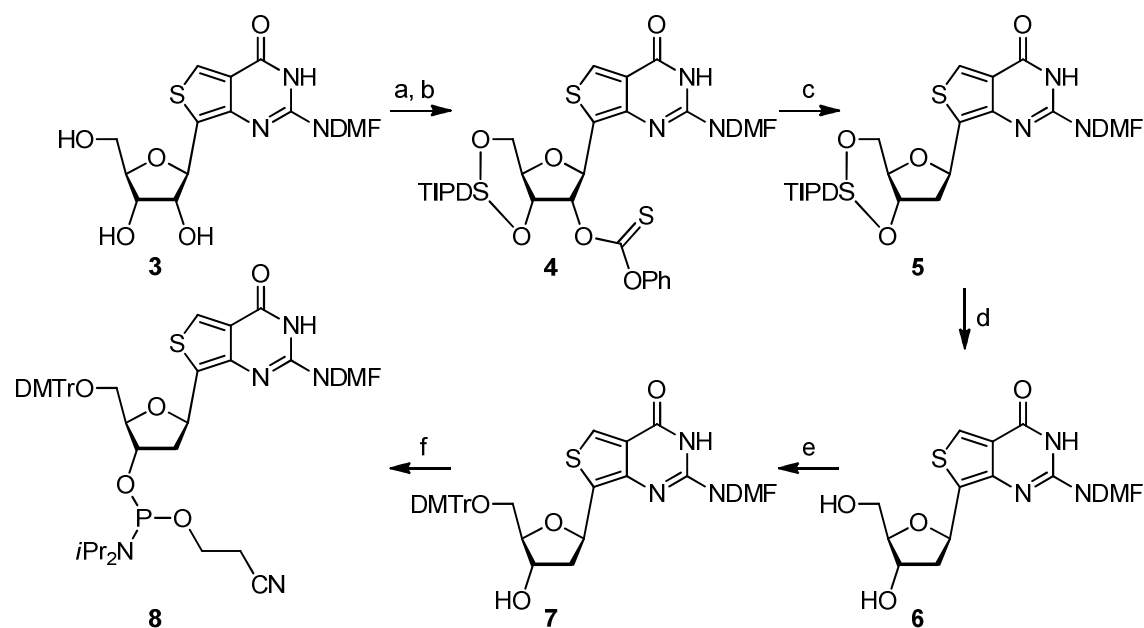
Figure S1. dthG-phosphoramidite synthesis[a]

[a] Reagents and conditions: (a) TIPDSiCl$_2$, Py, 89%; (b) PhOC(S)Cl, DMAP, Py, 78%; (c) Bu$_3$SnH, AIBN, toluene, 110 °C, 75%; (d) TEA•3HF, THF, 0 °C~RT, 94%; (e) DMTrCl, Py, 50 %; (f) 2-cyanoethyl N,N-diisopropylchlorophophoramidite, iPr$_2$NEt, DCM, 0 °C~RT, 56 %.

Synthesis of $O^{3',5'}$-TIPDS-$O^{2'}$-PTC-N^2-DMF-thG (4)

3 (0.22, 0.62 mmol), as prepared previously reported,[1] was co-evaporated with dry pyridine (2 × 3 mL) and dissolved in dry Py (4 mL). TIPDSCl$_2$ (0.20 mL, 0.62 mL) was added to the solution dropwise at 0 °C and stirred at RT for 16 hours. All volatiles

were evaporated and the residue was partitioned between DCM (50 mL) and saturated aq. NaHCO$_3$ (50 mL). The aq. layer was extracted with DCM (2 × 10 mL) and combined organic layer was dried over Na$_2$SO$_4$ then evaporated. The residue was purified by column chromatography with 0~10 % gradient MeOH in DCM to afford an off-white solid. Yield 0.32 g, 89 %. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.70 (s, 1H), 8.15 (s, 1H), 8.12 (s, 1H), 5.60 (d, J = 3.3 Hz, 1H), 4.48 (t, J = 6.2 Hz, 1H), 4.27 – 4.20 (m, 1H), 4.12 – 3.96 (m, 3H), 3.17 (s, 3H), 3.07 (s, 3H), 1.14 – 0.98 (m, 28H); ESI-MS calculated for C$_{26}$H$_{45}$N$_4$O$_6$SSi$_2$ [M+H]$^+$597.26, found 597.19; ESI-HRMS calculated for C$_{26}$H$_{45}$N$_4$O$_6$SSi$_2$ [M+H]$^+$597.2593, found 597.2591.

The intermediate (0.22 g, 0.37 mmol) was co-evaporated with dry Py (2 × 4 mL) and dissolved in Py/ACN (2 mL/6 mL). DMAP (93 mg, 0.76 mmol) and O-phenyl chlorothionoformate (PTCCl, 76 µL, 0.56 mmol) were successively added to the solution and stirred for 16 hour at RT. All volatiles were evaporated, the residue was dissolved in DCM (20 mL), washed with saturated aq. NaHCO$_3$ (20 mL), dried over Na$_2$SO$_4$ then evaporated. The residue was purified by column chromatography with 0~1.5 % gradient MeOH in DCM to afford a yellow solid. Yield 0.21 g, 78 %. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.78 (s, 1H), 8.60 (s, 1H), 8.11 (s, 1H), 7.46 – 7.37 (m, 2H), 7.33 – 7.27 (m, 1H), 7.15 – 7.08 (m, 2H), 6.48 (d, J = 4.4 Hz, 1H), 4.58 (dd, J = 8.8, 4.7 Hz, 1H), 4.20 – 3.94 (m, 4H), 3.08 (s, 3H), 3.05 (s, 3H), 1.10 – 1.05 (m, 14H), 1.05 – 0.99 (m, 14H); ^{13}C NMR (126 MHz, CDCl$_3$) δ 194.29, 158.45, 153.46, 129.60, 129.49, 127.84, 126.72, 125.67, 125.52, 121.89, 121.22, 86.95, 81.44, 78.10, 70.90, 61.00, 41.38, 35.19, 29.84, 17.61, 17.49, 17.45, 17.36, 17.34, 17.28, 17.25, 17.17, 13.53, 13.19, 12.99, 12.91; ESI-MS calculated for C$_{33}$H$_{49}$N$_4$O$_7$S$_2$Si$_2$ [M+H]$^+$733.26, found 733.22; ESI-HRMS calculated for C$_{33}$H$_{49}$N$_4$O$_7$S$_2$Si$_2$ [M+H]$^+$733.2576, found 733.2573

Synthesis of $O^{3',5'}$-TIPDS-2'-deoxy-N^2-DMF-thG (5)

A solution of **4** (0.20 g, 0.27 mmol), Bu$_3$SnH (0.22 mL, 0.81 mmol) and AIBN (22 mg, 0.14 mmol) in freshly distilled toluene over sodium (5 mL) was degased with argon bubbling for 20 min at RT, and then the mixture was heated at 110 °C for 2 hours. Solvent was evaporated and residue was dissolved in DCM (20 mL), washed with saturated aq. NaHCO$_3$ (20 mL), dried over Na$_2$SO$_4$ then evaporated. The residue was purified by column chromatography with 0~2 % gradient MeOH in DCM to afford a yellow foam. Yield 0.12 g, 75 %. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.70 (s, 1H), 8.50 (s, 1H), 8.10 (s, 1H), 5.87 (s, 1H), 4.64 – 4.53 (m, 1H), 4.13 – 4.05 (m, 1H), 3.92 – 3.82 (m, 3H), 3.19 (s, 3H), 3.08 (s, 3H), 2.49 (s, 1H), 2.43 – 2.23 (m, 1H), 1.69 – 1.59 (m, 1H), 1.18 – 0.98 (m, 28H); ^{13}C NMR (126 MHz, CDCl$_3$) δ 159.41, 157.98, 153.97, 131.34, 129.50, 125.45, 121.25, 86.09, 76.59, 73.61, 72.11, 63.66, 43.00, 41.55, 17.71, 17.67, 17.61, 17.57, 17.44, 17.28, 17.21, 17.13, 13.59, 13.48, 13.09, 12.69; ESI-HRMS calculated for C$_{26}$H$_{45}$N$_4$O$_5$SSi$_2$ [M+H]$^+$581.2644, found 581.2645.

Synthesis of 2'-deoxy-N^2-DMF-thG (6)

To a solution of **5** (0.11 g, 0.19 mmol) in dry THF (2 mL) was added TEA•3HF (0.14 mL, 0.86 mmol) dropwise at 0 °C then was stirred for 7 hours at RT. All volatiles were evaporated and the residue was purified by column chromatography with 0~10 % gradient MeOH in DCM to afford an off-white foam. Yield 60 mg, 94 %. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.82 (s, 1H), 8.48 (s, 1H), 8.05 (s, 1H), 5.56 (dd, J = 11.1, 5.4 Hz, 1H), 4.85 (s, 1H), 4.59 (d, J = 5.1 Hz, 1H), 4.13 (d, J = 1.0 Hz, 1H), 3.86 (dd, J = 12.0, 2.7 Hz, 1H), 3.71 (d, J = 11.7 Hz, 1H), 3.14 (s, 3H), 3.03 (s, 3H), 2.70 (td, J = 13.1, 5.2 Hz, 1H), 2.57 (s, 1H), 2.21 (dd, J = 13.2, 5.5 Hz, 1H); ^{13}C NMR (126 MHz, CDCl$_3$) δ 159.52, 158.16, 154.63, 146.20, 127.78, 126.30, 124.50, 88.31, 75.57, 75.27, 63.80, 43.52, 41.30, 35.04; ESI-MS calculated for C$_{14}$H$_{19}$N$_4$O$_4$S [M+H]$^+$339.11,

found 399.19; ESI-HRMS calculated for $C_{14}H_{19}N_4O_4S$ [M+H]$^+$339.1122, found 339.1125.

Synthesis of $O^{5'}$-DMT-2'-deoxy-N^2-DMF-thG (7)

6 (59 mg, 0.17 mmol) was coevaporated with dry Py (2 × 2 mL) and was dissolved in dry Py (1 mL). DMTrCl (69 mg, 0.20 mmol) was added to the solution at RT, and was stirred for 16 hours. The reaction was quenched by addition of MeOH (1 mL) and evaporated. The residue was purified by column chromatography with 0~2 % gradient MeOH in DCM with 1 % Py to afford an off-white foam. Yield 54 mg, 50 %. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.84 (s, 1H), 8.57 (s, 1H), 8.03 (s, 1H), 7.45 – 7.37 (m, 2H), 7.37 – 7.24 (m, 4H), 7.24 – 7.04 (m, 3H), 6.73 (d, J = 8.4 Hz, 4H), 5.91 (dd, J = 9.9, 5.2 Hz, 1H), 4.42 (d, J = 5.4 Hz, 1H), 4.01 (dd, J = 6.6, 4.5 Hz, 1H), 3.69 (s, 3H), 3.69 (s, 3H), 3.20 (qd, J = 9.9, 4.7 Hz, 2H), 3.02 (s, 3H), 2.93 (s, 3H), 2.52 (s, 1H), 2.33 (dd, J = 13.1, 5.2 Hz, 1H), 2.18 (ddd, J = 13.1, 10.5, 5.8 Hz, 1H); ^{13}C NMR (126 MHz, CDCl$_3$) δ 159.82, 158.49, 158.48, 157.92, 154.00, 146.49, 144.93, 136.14, 136.06, 130.85, 130.20, 130.18, 129.10, 128.55, 128.29, 127.92, 126.85, 125.33, 125.23, 113.19, 86.35, 86.24, 74.81, 73.14, 64.48, 55.32, 43.97, 41.35, 35.12; ESI-MS calculated for $C_{35}H_{37}N_4O_6S$ [M+H]$^+$641.24, found 640.86; ESI-HRMS calculated for $C_{35}H_{37}N_4O_6S$ [M+H]$^+$641.2428, found 641.2426

Synthesis of $O^{3'}$-(2-Cyanoethyldiisopropylphosphoramidite)-$O^{5'}$-DMT-2'-deoxy-N^2-DMF-thG (8)

7 (49 mg, 76 μmol) was coevaporated with dry Py (2 × 1 mL) and dried under high vacuum overnight then dissolved in dry DCM (1 mL). DIPEA (53 μL, 0.30 mmol) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (25 μL, 0.11 mmol) were

successively added to the solution at 0 °C and the mixture was stirred at RT for 2 hours. All volatiles were evaporated and the residue was purified by column chromatography with 0~2 % gradient MeOH in DCM with 1 % Py to afford a white foam. Yield 36 mg, 56 %. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.70 (s, 1H), 8.69 (s, 2H), 8.17 – 8.15 (m, 2H), 8.11 (s, 1H), 8.11 (s, 1H), 7.52 – 7.45 (m, 4H), 7.41 – 7.33 (m, 8H), 7.32 – 7.23 (m, 4H), 7.23 – 7.16 (m, 2H), 6.85 – 6.75 (m, 8H), 5.99 – 5.91 (m, 2H), 4.64 – 4.53 (m, 2H), 4.24 – 4.17 (m, 2H), 4.17 – 4.09 (m, 1H), 3.91 – 3.82 (m, 1H), 3.82 – 3.76 (m, 18H), 3.76 – 3.67 (m, 2H), 3.67 – 3.43 (m, 5H), 3.40 – 3.30 (m, 2H), 3.27 – 3.12 (m, 7H), 3.08 (s, 6H), 2.81 – 2.72 (m, 2H), 2.62 (t, J = 6.5 Hz, 2H), 2.56 (dd, J = 13.3, 5.1 Hz, 1H), 2.46 (dd, J = 12.1, 5.7 Hz, 3H), 2.37 – 2.20 (m, 2H), 1.30 – 1.23 (m, 9H), 1.19 – 1.14 (m, 9H), 1.05 (d, J = 6.8 Hz, 6H); ^{31}P NMR (202 MHz, CDCl$_3$) δ 148.58, 148.29; ESI-MS calculated for C$_{44}$H$_{54}$N$_6$O$_7$PS [M+H]$^+$841.35, found 840.86; ESI-HRMS calculated for C$_{44}$H$_{54}$N$_6$O$_7$PS [M+H]$^+$841.3507, found 841.3503.

1.2. Synthesis of non-labeled and dthG/d2Ap-labeled ODNs

Non-modified (-)PBS, (+)PBS, T12(+)PBS and d2Ap7(-)PBS were synthesized and purified by IBA GmbH Nucleic Acids Product Supply (Germany). For dthG7(-)PBS, solid-phase ODN synthesis was performed on an Expedite 8909 synthesizer using commercially available reagents and phosphoramidites (Glen Research). The modified phosphoramidite was chemically synthesized as described above and incorporated into ODN with coupling efficiency comparable to the commercially available phosphoramidites. The solution of the modified phosphoramidite was dried for 16 hours over molecular sieve 3A (dried for 2 days at 300 °C under high vacuum) and was filtered using syringe filterright before use. ODNs were synthesized (with trityl-off) on a 500 Å CPG solid support column (1 µmol scale). Cleavage from the

solid support and deprotection were accomplished with AMA (ammonium hydroxide/methylamine in water = 1/1) at 65 °C for 30 min. The oligonucleotides were purified by 20 % preparative polyacrylamide gel electrophoresis (PAGE) using the crush and soak method; the desired band was cut out, pulverized, extracted with 50 mM TEAA (pH 7.0) for 12 hours (while shaking) and decanted. The buffer containing the purified ODN was lyophilized and the residue was taken up in 0.2 M TEAB (pH 7.0) buffer and desalted on a Sep-pak C-18 (Waters). The ODNs were eluted with 40 % acetonitrile in water. The $d^{th}G$ containing (-)PBS DNA was>98 % pure as determined by analytical high resolution PAGE. The purified ODN was quantified by UV absorbance at 260 nm at 70 °C with the following extinction coefficients ($M^{-1}cm^{-1}$): dCMP, 7050; dTMP, 8840; dGMP, 12010; dAMP, 15200; and $d^{th}G$, 5500, and confirmed by MALDI-TOF mass spectrometry: calculated M 5483.52, found 5482.70 $[M-H]^{+}$.

2. UV/visible absorption and steady-state fluorescence measurements. Spectroscopic grade solvents were used for absorption and fluorescence spectroscopy. To determine the concentrations, extinction coefficients of 4150 $M^{-1}cm^{-1}$ at 321 nm and 157280 $M^{-1}cm^{-1}$ at 260 nm for $d^{th}G$ and $d^{th}G7(–)PBS$, respectively and 6800 $M^{-1}cm^{-1}$ at 303 nm and 143550 $M^{-1}cm^{-1}$ at 260 nm for d2Ap and d2Ap7(–)PBS, respectively, were used. All experiments were performed in 25 mM TRIS-HCl (pH = 7.5), 30 mMNaCl, 0.2 mM $MgCl_2$ at 20 °C.

Absorption spectra were recorded on a Cary 4000 UV-visible spectrophotometer (Varian). Fluorescence spectra were recorded on a FluoroMax 4 spectrofluorimeter (JobinYvon) equipped with a thermostated cell compartment at 20±0.5 °C. Fluorescence spectra were corrected for Raman scattering, lamp fluctuations and instrumental wavelength-dependent bias. QY of $d^{th}G$- and d2Ap-labeled ODNs were

calculated using quinine sulfate in 0.5 M sulfuric acid (QY=0.546),[2] and free 2Ap deoxyriboside (QY=0.68)[3] as references, respectively. Excitation wavelength was 380 nm for $d^{th}G$ and 315 nm for d2Ap. Melting temperatures were determined by measuring absorbance changes at 260 nm as a function of the temperature using a Varian Cary 400 spectrophotometer equipped with a Peltier temperature controller. Absorbance was recorded in the forward and backward directions from 20 to 80 °C at a rate of 0.5 °C/min. Prior to the melting experiment, (+)PBS and (–)PBS samples were denatured at 90 °C for 3 min and then slowly cooled down to allow their annealing. For melting experiments, the complementary ODNs were at 1 µM in 25 mM TRIS (pH = 7.5), 30 mM NaCl and 0.2 mM $MgCl_2$. Melting temperatures were determined from the first derivative of thermal denaturation curves.

3. Time-resolved fluorescence intensity decays

Time-resolved fluorescence measurements were performed with the time-correlated single-photon counting technique. Excitation pulses at 315 nm with a repetition rate of 4 MHz were generated by a pulse-picked frequency-tripled Ti-sapphire laser (Tsunami, Spectra Physics) pumped by a Millenia X laser (Spectra Physics).[4] The fluorescence emission was collected at 500 nm through a polarizer set at magic angle and a 16 mm band-pass monochromator (Jobin Yvon). The single-photon events were detected with a micro-channel plate photomultiplier (Hamamatsu) coupled to a pulse pre-amplifier HFAC (Becker-Hickl GmbH) and recorded on a time correlated single photon counting board SPC-130 (Becker-Hickl GmbH). The instrumental response function (IRF) recorded with a polished aluminum reflector was characterized by a ≈ 50 ps full-width at half-maximum. The mean lifetime <τ> was calculated from the individual fluorescence lifetimes ($τ_i$) and their relative

amplitudes (α_i) according to $<\tau>= \sum \alpha_i \times \tau_i$. The population of dark species (α_0) was calculated by:

$$\alpha_0 = 1 - \tau_{free}/(\tau_{ODN} \times R_m) \ (1),$$

where τ_{free} is the lifetime of the free nucleoside, τ_{ODN} is the measured mean lifetime of the probe within the ODN and R_m is the ratio of their corresponding QYs. The amplitudes of the fluorescent populations α_{ic} were recalculated according to $\alpha_{ic} = \alpha_i \times (1 - \alpha_0)$. Time-resolved intensity data were fitted using the maximum entropy method (Pulse 5 software).[5] In all cases, the χ^2 values were close to 1, indicating an optimal fit.

4. Time-resolved fluorescence anisotropy decays

Time-resolved fluorescence anisotropy was obtained from the fluorescence decay curves recorded in directions parallel $I_\parallel$ and perpendicular $I_\perp$ alternatively, to the excitation beam polarization and was analyzed by:

$$r(\tau) = \frac{I_\parallel(\tau) - G \times I_\perp(\tau)}{I_\parallel(\tau) + 2G \times I_\perp(\tau)} = r_0 \sum_i \beta_i \times \exp\left(-\frac{\tau}{\theta_i}\right) \ (2),$$

where β_i are the amplitudes of the rotational correlation times θ_i, r_0 is the initial anisotropy, and G is the geometry factor at the emission wavelength, determined in an independent experiment. Time-resolved anisotropy data were fitted using the maximum entropy method (Pulse 5 software) or according to a non-linear least-square analysis using an iterative reconvolution method (software provided by G. Krishnamoorthy). The r_0 values were found to be 0.32-0.33 for the 2Ap-containing sequences, while those (r_0) for the $d^{th}G$-containing sequences were 0.23-0.25. In all cases, the χ^2 values were close to 1, indicating an optimal fit.
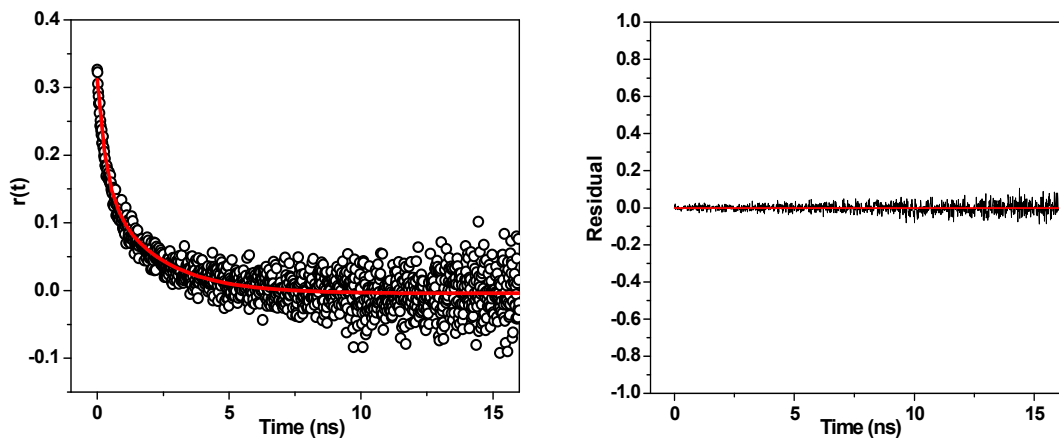
Figure S2. Anisotropy decay and corresponding residual plot for d2Ap7(-)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.
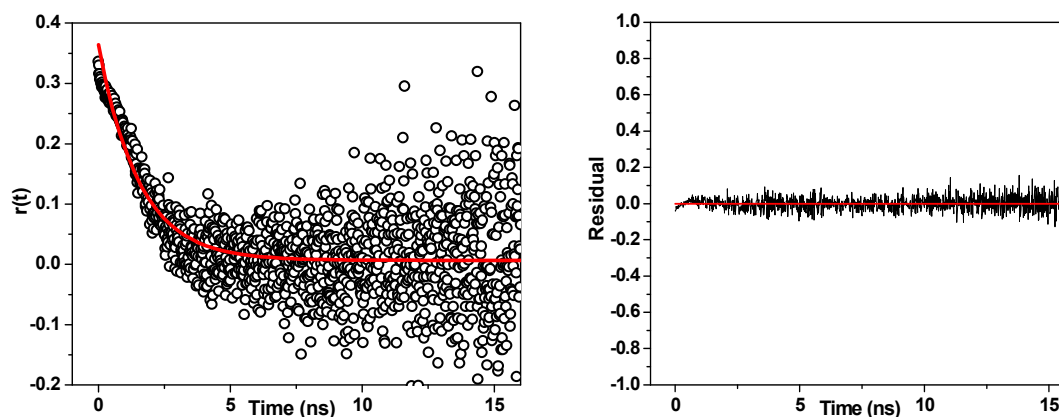
Figure S3. Anisotropy decay and corresponding residual plot for d2Ap7(-)/T12(+)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.
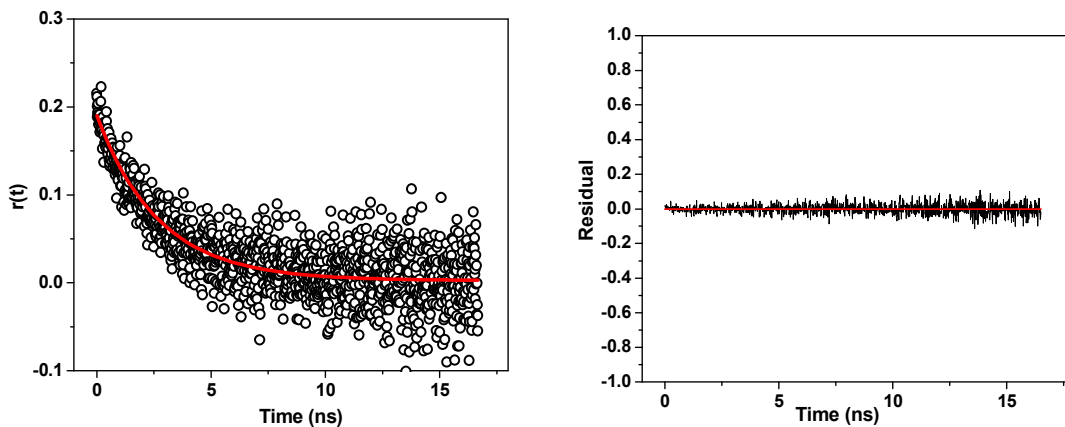
Figure S4. Anisotropy decay and corresponding residual plot for dthG7(-)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.
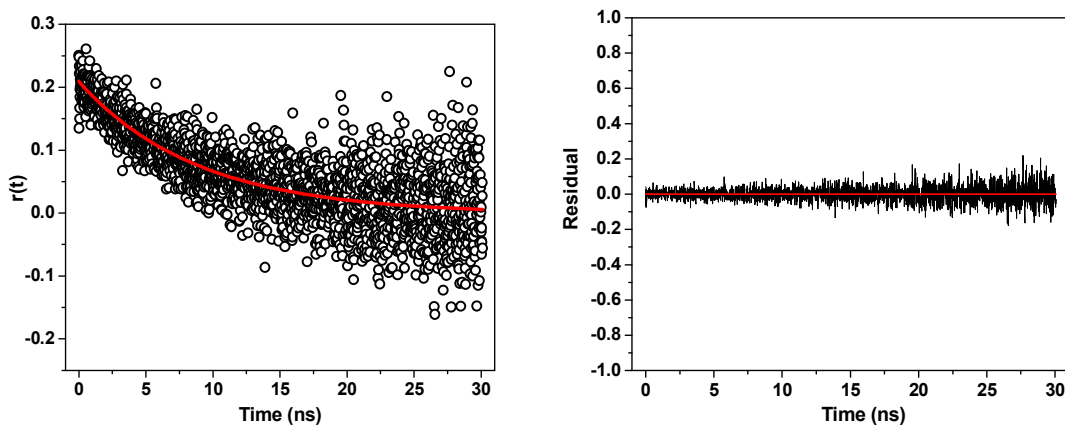
Figure S5. Anisotropy decay and corresponding residual plot for dthG7(-)/(+)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.
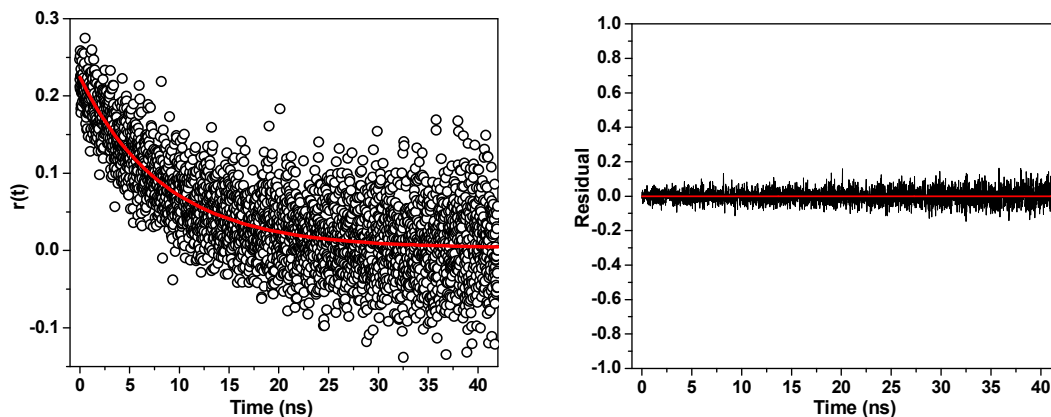
Figure S6. Anisotropy decay and corresponding residual plot for dthG7(-)/T12(+)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.

5. Quenching measurements

Fluorescence quenching by potassium iodide (KI) was carried out by adding aliquots of a concentrated aqueous stock of KI to the labeled ODNs. $Na_2S_2O_3$ was added to the KI stock solution to prevent its oxidation. Fluorescence intensity was corrected for dilution. The change in fluorescence intensity as a function of quencher concentration was fitted by the Stern-Volmer equation:

$$\frac{F_0}{F} = 1 + k_q \tau_0 [Q] \ (3),$$

where F and F_0 are the intensities in the presence and absence of quencher, respectively, k_q is the diffusion-controlled quenching rate, and τ_0 is the lifetime in the absence of the quencher.
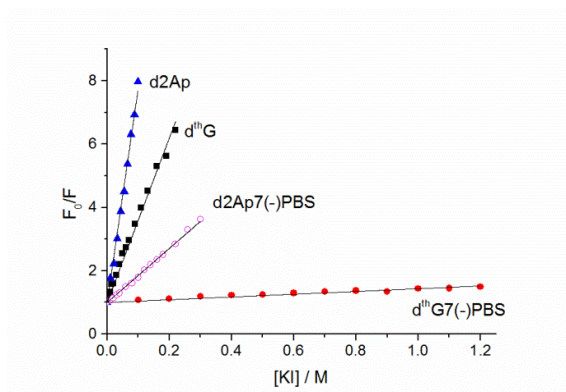
Figure S7. Stern-Volmer plots for KI quenching of $d^{th}G$ (black squares), $d^{th}G7(-)PBS$ (red disks), d2Ap (blue triangles), and d2Ap7(-)PBS (magenta circles). Black lines represent their linear fits.

References:

(1) Shin, D.; Sinkeldam, R. W.; Tor, Y. *J. Am. Chem. Soc.* **2011**, *133*, 14912.

(2) Melhuish, W. H. *J. Phys. Chem.* **1961**, *65*, 229.

(3) Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem.* **1969**, *244*, 1228.

(4) (a) Godet, J.; Kenfack, C.; Przybilla, F.; Richert, L.; Duportail, G.; Mély, Y. *Nucleic Acids Res.* **2013**, *41*, 5036.(b) Bernacchi, S.; Stoylov, S.; Piémont, E.; Ficheux, D.; Roques, B. P.; Darlix, J. L.; Mély, Y. *J. Mol. Biol.* **2002**, *317*, 385.

(5) (a) Brochon, J. C. *Methods Enzymol.* **1994**, *240*, 262. (b) Livesey, A. K.; Brochon, J. C. *Biophys. J.* **1987**, *52*, 693.

4.2. CHAPTER 2: Fluorescent Nucleobase Analogues as a Tool for Characterizing Dynamics of Nucleocapsid Protein Promoted Annealing of (–)/(+)PBS.

4.2.1. Publication 3: Environment-Sensitive Fluorescent Nucleoside Analogues for Surveying Dynamic Interconversion of Nucleic Acid Structures.

In this work, we used two versatile environment-sensitive fluorescent nucleobases 3HCnt at position 9 and thG at position 7 into the (–)PBS loop to directly monitor the role of the loops and to recover for the first time the whole set of kinetic rate constants governing (+)/(–)PBS annealing both in the absence and the presence of NC(11–55), NC(1–55) or (SSHS)$_2$NC(11–55) (Publication 3, Figure 9). 3HCnt acts as a universal nucleoside analogue able to substitute any of the natural nucleobases (Dziuba et al., 2012) and importantly, it has recently been shown to preserve the intercalation and stacking with the surrounding bases in a DNA helix (Zargarian et al., 2017). thG has already been shown to perfectly substitute the single G residue at position 7 in (–)PBS loop and exhibit a two-fold increase in quantum yield when the (–)PBS stem-loop in converted into the (+)/(–)PBS duplex. Its competence has surpassed over 2-Aminopurine's restraints by faithfully substituting a key G7 residue in (–)PBS, providing reliable information on its conformations and dynamics in both the (−)PBS stem loop

and (−)/(+)PBS duplex (Sholokh et al., 2015).Our data show that the 3HCnt nucleobase can sensitively monitor at nearly all positions, the binding of NC(11–55) and the stem-loop to duplex conversion. We evaluated the photophysical protease of 3HCnt by substituting it at different positions within (–)PBS. As 3HCnt at position 9 marginally affects the affinity for NC(11–55) and shows a high sensitivity to the binding of NC(11–55) and formation of the duplex, this position was further selected for monitoring the (+)/(–)PBS annealing reaction.

In absence of NC, (+)/(–)PBS annealing is governed by a two-step pathway nucleated via the overhangs (Figure 9a, upper pathway). We evidenced also for both probes a large fluorescence increase associated with the formation of the IC, that clearly indicated the existence of loop-loop contacts in the IC that were never seen before. These contacts are probably the driving force for the slow annealing reaction observed when $\Delta P()PBS$ was substituted for PBS. In comparison to the reaction in the absence of peptide, $(SSHS)_2NC(11–55)$ induces a two orders of magnitude increase of all three kinetic parameters (k_1, k_{-1} and k_f) suggesting the efficient electrostatic attraction between the complementary strands permitted by the high charge density and flexibility of this mutant, as well as by its ability to rapidly bind and dissociate from the ODNs. Our data showed that 3HCnt but not ^{th}G could be used to monitor the NC-promoted annealing reaction. As in the absence of NC, the full set of kinetic parameters could be retrieved, giving a clearer picture on the mechanism by which NC(11–55) promotes the annealing reaction. The formation of a loop/loop IC is thought to result from the specific interaction of the NC(1–55) or NC(11–55) hydrophobic platform, which solvent-exposes the nucleobases of both (–)PBS and (+)PBS loops, making them competent for annealing. This solvent exposure of the loop nucleobases induced is fully supported by the large increase in the 3HCnt and ^{th}G fluorescence on binding of both proteins.

FULL PAPER

Environmentally-Sensitive Fluorescent Nucleoside Analogues for Surveying Dynamic Interconversions of Nucleic Acid Structures

Marianna Sholokh,[a,c] Rajhans Sharma,[a] Natalia Grytsyk,[a] Lyes Zaghzi,[a] Viktoriia Y. Postupalenko,[a] Dmytro Dziuba,[b] Nicolas P.F. Barthes,[b] Benoît Y. Michel,[b] Christian Boudier,[a] Olga A. Zaporozhets,[c] Yitzhak Tor,[d] Alain Burger[b] and Yves Mély*[a]

Abstract: Nucleic acids are characterized by a variety of dynamically interconverting structures that play a major role in transcriptional and translational regulation as well as recombination and repair. To monitor these interconversions, FRET-based techniques can be used, but require two fluorophores that are typically large and can alter the DNA/RNA structure and protein binding. Additionally, events that do not alter the donor/acceptor distance and/or angular relationship are frequently left undetected. A more benign approach relies on fluorescent nucleobases that can substitute their native counterparts with minimal perturbation, such as the recently developed 2-thienyl-3-hydroxychromone (3HCnt) and thienoguanosine (thG). To demonstrate the potency of 3HCnt and thG in deciphering interconversion mechanisms, we used the conversion of the (–)DNA copy of the HIV-1 primer binding site (–)PBS stem-loop into (+)/(–)PBS duplex, as a model system. When incorporated into the (–)PBS loop, the two probes were found to be highly sensitive to the individual steps both in the absence and the presence of a nucleic acid chaperone, providing the first complete mechanistic description of this critical process in HIV-1 replication. The combination of the two distinct probes appears to be instrumental for characterizing structural transitions of nucleic acids under various stimuli.

Introduction

Structural polymorphism is a key characteristic of nucleic acids. Beyond the canonical B-DNA and A-RNA, nucleic acids can fold into diverse and more complex structures such as triplexes, G-quadruplexes, i-motifs, cruciforms, Holliday junctions, hairpins, and other alternative forms.[1] These non-canonical structures can be formed from single and/or double stranded nucleic acid sequences and can frequently dynamically interconvert. For instance, a Holliday junction is a branched DNA or RNA structure composed of four double-stranded arms joined together that may exist in a variety of forms and play a central role in genetic recombination.[2] Another illustrative example is found in G/C rich promoters and telomeric regions in DNAs. Dissociation of a G/C-rich double stranded DNA produces two single G- and C-rich strands, which can fold into a G-quadruplex and an i-motif, respectively.[3] Alternatively, a single stranded DNA/RNA sequence that contains complementary sequences separated by three or more nucleotides can fold into a stem-loop (hairpin). Hairpins may be produced during DNA repair, lagging-strand synthesis, transcription, rolling-circle replication and reverse transcription.[1]

Non canonical structures are thought to take part in major cellular processes such as transcriptional and translational regulation, recombination, repair or viral infections.[1] Though substantial progress has been made in our understanding of their physiological significance and dysfunctions in a variety of human diseases, major mechanistic questions remain unanswered. Elucidating the dynamics, conformational changes and interconversion of nucleic acids is needed not only for unraveling their molecular mechanisms and cellular functions, but also for proposing new strategic therapeutic approaches and structure-specific targets. Elucidating these structural transitions is challenging, however, because many of these non-canonical structures are transient and unstable, and their interconversions depend on a multitude of factors, including base composition, salt concentration and solvation, as well as small and large ligands and superhelical stress.

X-ray crystallography and cryo-electron microscopy are invaluable tools to access 3D structures but are inherently limited to the solid state. Spectroscopic techniques, as NMR and fluorescence spectroscopy, are complementary methods able to shed light on the dynamics and structures of nucleic acids in solution. Fluorescence spectroscopy is appealing due to its high sensitivity and exquisite spatio-temporal resolution. In nucleic acids, the interconversions are mainly monitored using Förster Resonance Energy Transfer (FRET) experiments in which the oligonucleotides (ODN) are labeled with a donor and acceptor pair. Because FRET efficiency depends on the distance and angular relationship between the donor and acceptor, FRET techniques are well suited to analyze global conformational transitions in nucleic acids. For instance, they were employed to characterize the dynamics and interconversion between different conformations of Holliday junctions and G-quadruplexes.[4,5] However, they failed to detect DNA/RNA changes that do not significantly impact the distance between the donor and acceptor. Moreover, FRET-based experiments require two chromophores

[a]   Dr. M. Sholokh, Dr. R. Sharma, Dr. N. Grytsyk, L. Zaghzi, Dr. V.Y. Postupalenko, Dr. C. Boudier, Prof. Dr. Y. Mély
      Laboratoire de Bioimagerie et Pathologies, UMR 7021 CNRS, Université de Strasbourg,
      Faculté de Pharmacie, 74 route du Rhin, 67401 Illkirch, France
      E-mail: yves.mely@unistra.fr
[b]   Dr. D. Dziuba, Dr. N.P.F. Barthes, Dr. B.Y. Michel, Prof. Dr. A. Burger
      Institut de Chimie de Nice, UMR 7272 CNRS, Université Côte d'Azur,
      Parc Valrose, 06108 Nice, France
[c]   Prof. Dr. O.A. Zaporozhets
      Department of Chemistry, Kyiv National Taras Shevchenko University,
      60 Volodymyrska street, 01033 Kyiv, Ukraine
[d]   Prof. Dr. Y. Tor
      Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0358, USA

      Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.2018xxxxx..

and the commonly used ones tend to be bulky (fluorescein, cyanines, rhodamines, Alexas, etc.), and can alter the DNA/RNA structure and the binding of proteins and ligands. The use of DNA/RNA site-specifically labeled with fluorescent nucleobase analogues that are environmentally sensitive[6,7] is an attractive alternative that can give additional insights compared to FRET experiments. These fluorescent probes should minimally disturb the DNA/RNA structure and function, and should respond sensitively and site selectively to different molecular events occurring during the interconversion process.

2-aminopurine (2AP), an isomer of adenine, was for a long time the gold standard of fluorescent nucleoside analogues.[6–8] 2AP absorbs in the red edge of the nucleic acid/protein absorption domain and thus can be selectively excited. It exhibits high quantum yield in water and fluoresces with relatively large Stokes shift.[9] However, it is strongly quenched in DNA and RNA oligonucleotides[10] and also locally perturbs the structure and dynamics of the modified nucleic acid.[11] As a consequence, its residual fluorescence in ODNs may reflect the structural and dynamic information of minor or non-biologically relevant conformers.[12] Quenching, structural disturbance or poor environment sensitivity are common drawbacks of fluorescent nucleobase mimics. These severe limitations have stimulated over the two past decades the design of new fluorescent nucleoside surrogates that minimally perturb the labeled nucleic acids and show improved brightness and site-specific responses to intermolecular interactions and conformational changes.[13,14] According to their chemical structure, these fluorescent nucleoside analogues can be classified in two major families: nucleobase mimics possessing base pairing properties and non-canonical nucleobases lacking the H-bonding face. Among them, we have selected a recently developed representative from each category with improved properties for surveying dynamic interconversions of nucleic acid structures.

2-thienyl-3-hydroxychromone (3HCnt, Figure1) is a recently developed heterocycle that acts as a universal nucleobase.[15,16] 3HCnt exhibits a dual emission due to an excited-state intramolecular proton transfer (ESIPT) reaction between the initially excited normal (N*) and its tautomeric state (T*). The ratio between the two bands is sensitive to the polarity and the H-bond donor ability of the probe's microenvironment. This chromophore is up to 50-fold brighter than 2AP with minimal perturbation of the duplex structure.[17] 3HCnt has been used to monitor the local conformational changes of a single-stranded ODN upon binding of an HIV-1 protein[15] and the conformational changes in DNA associated to Endonuclease VIII processing.[18]

Another recently introduced nucleobase is thienoguanine (thG, Figure1), an isomorphic guanine surrogate,[19] which remains highly emissive when incorporated into single- and double-stranded ODNs.[20] thG has been shown to faithfully substitute a key G residue in the (–)DNA copy of the HIV-1 primer binding site (−)PBS. Finally, both 3HCnt and thG were introduced into hemimethylated DNAs and successfully used to monitor the flipping of a neighbor methylcytosine as the result of the binding of ubiquitin-like containing PHD and RING finger domains 1 (UHRF1) that plays a key role in the replication of the DNA methylation pattern.[21]

To demonstrate the utility of both 3HCnt and thG, we used the annealing process of the (–)PBS stem-loop into the (+)/(–)PBS duplex as a model system (Figure 1a). This pathway was previously examined using a PBS sequence labeled at its 5' and 3' ends with rhodamine 6G (Rh6G) and Dabcyl, respectively.[22] The (+)/(–)PBS hybridization was found to proceed spontaneously through the flexible single-stranded overhangs, while it was promoted through a loop-loop kissing mechanism by the HIV-1 nucleocapsid protein (NC), acting as a nucleic acid chaperone.[22–24] However, due to the insensitivity of the FRET pair to the initial steps of the annealing reaction, only one rate constant could be extracted from the kinetic measurements[22,23] and a complete kinetic picture of the two pathways remains missing. When either 3HCnt or thG were inserted into the (–)PBS loop, the two probes were found to be highly responsive, providing the first complete mechanistic description of the (+)/(–)PBS annealing reaction, which is a key step occurring during the early stage of HIV infection.[25]
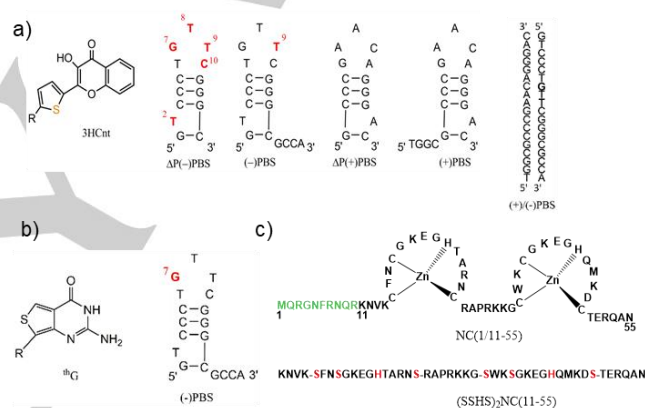
Figure 1. Structure of the 3HCnt and thG nucleobases (R = 2'-deoxyribose) and sequences of the PBS oligonucleotides (a and b), and NC peptides (c) used in this study. The residues replaced by 3HCnt or thG in the oligonucleotides are shown in red. The N-terminal domain (green) is missing in the truncated NC(11–55) mutant. In the (SSHS)$_2$NC(11–55) mutant, the SSHS residues (red) that replace the naturally occurring CCHC residues are unable to bind zinc.

Results and Discussion

Characterization of the thermodynamic, binding and spectroscopic properties of the thG-and 3HCnt-labeled oligonucleotides

To demonstrate the utility of thG and 3HCnt for deciphering the mechanism of the spontaneous and NC-promoted (+)/(–)PBS annealing processes, our first objective was to identify the best labeling position within the (–)PBS loop. Ideally, the emissive nucleoside should exhibit a respectable quantum yield within the oligonucleotide and be responsive to NC binding and duplex formation, but should not alter the binding to NC or the structure of the duplex with (+)PBS. thG has previously been shown to perfectly substitute the G residue at position 7 in (–)PBS loop and exhibit a two-fold increase in quantum yield when the (–)PBS stem-loop is converted into the (+)/(–)PBS duplex.[20,26] Thus, this

position appears as a good candidate for shedding light on the mechanisms of the (–)PBS stem-loop conversion into the (+)/(–)PBS duplex.

To identify the best labeling position for the 3HCnt nucleobase, we used the ΔP(–)PBS stem-loop sequence (Figure 1), whose 3D structure in both free and NC-bound forms has been solved by NMR and whose binding properties have been previoulsy characterized.[27] The ΔP(–)PBS sequence was labeled by 3HCnt at positions 7, 8, 9 or 10 in the loop (Figure 1), with labeling at position 2 in the stem serving as a control. To assess the effect of 3HCnt on the stability of the duplexes with the complementary ΔP(+)PBS sequence, thermal denaturation experiments were performed (Figure S1, Table S1 in Supplementary Information). With the exception of position 10, the melting temperatures differed only by ± 2°C from that of the non-labeled ΔP(+)/(–)PBS duplex, suggesting that the 3HCnt nucleobase only moderately perturbs the duplexes, in line with the behavior of 3HCnt as a universal nucleobase and its limited effect on the structure of duplexes.[15,17]

Next, we characterized the absorption and fluorescence properties of the 3HCnt-labeled ΔP(–)PBS stem-loops. All 3HCnt-labeled ODNs showed absorption maxima at 372–375 nm (Table S2), being 5–8 nm red-shifted as compared with the free chromophore in buffer,[16] in line with its stacking interactions with the neighboring bases.[17,28,29] Each labeled 3HCnt-ΔP(–)PBS sequence displayed a two-band emission spectrum characteristic for 3-hydroxychromone (3HC) derivatives (Figure 2a). While the emission maxima of the N* band (434–437 nm) were similar to that of the free probe (440 nm), a 30–37 nm red-shift was observed for the T* band (Table S2), indicating that the chromophore is partially desolvated through stacking interactions.
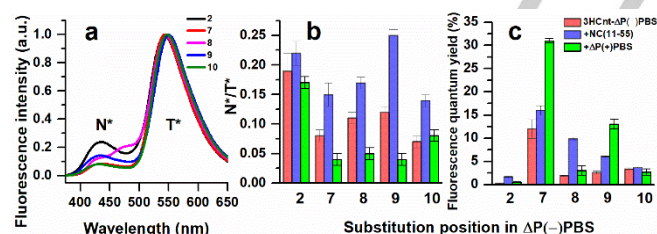

Figure 2. Spectroscopic properties of the ΔP(–)PBS sequences labeled by 3HCnt at different positions. (a) Fluorescence spectra were recorded for 2 μM ΔP(–)PBS labeled by 3HCnt at position 2 (black line), 7 (red), 8 (magenta), 9 (blue) and 10 (green). All spectra were normalized at the T* band. (b) N*/T* ratio and (c) fluorescence quantum yield of the 3HCnt-labeled ΔP(–)PBS sequences. The 3HCnt-labeled sequences at 2 μM were either in the form of stem-loops in the absence (red bars) or the presence of 8 μM NC(11–55) (blue bars) or in the form of duplexes with the complementary ΔP(+)PBS sequence (green bars). Buffer: 10 mM phosphate buffer, 30 mM NaCl, pH = 6.5. Excitation was at 360 nm.

This conclusion was further substantiated by the low values of the N*/T* ratios (0.07–0.19, Figure 2b), indicating a low polarity environment of the chromophore in all 3HCnt-labeled sequences.[15] Noticeably, the emission spectrum of 3HCnt-8ΔP(–)PBS showed an additional band between the N* and T* bands, with an emission maximum close to 475 nm. Such an

additional band has already been observed with other 3HC derivatives and is attributed to the anionic form of the probe, where the 3-OH group is deprotonated.[30–32] The values of the fluorescence quantum yield (QYs) at position 2, 8, 9 and 10 were rather low (0.4–3.4%) (Figure 2c and Table S2), due to the neighboring G and C residues, which are efficient quenchers of 3HCnt fluorescence.[15] A much higher QY value (12%) was observed for 3HCnt at position 7, because the two flanking T residues do not quench efficiently the 3HCnt fluorescence.

To investigate the fluorescence changes induced upon NC binding, we used the truncated NC(11–55) peptide, whose structure with ΔP(–)PBS has been solved.[27] This peptide retains the zinc finger domain responsible of the specific nucleic acid binding and destabilizing properties of NC.[25,33–38] Upon addition of NC(11–55) at saturating concentration (Figure 2 and Figure S2 and Table S2), the values of the N*/T* ratios and QYs of 3HCnt were found to increase at all positions, confirming that NC(11–55) binds all along the (–)PBS loop as well as at the 5' end of the stem.[27] At positions 7 and 8, the increase in the N*/T* ratio is consistent with NMR data[27] showing that NC(11–55) bound to its preferential binding site at the 5' end of the loop stretches the loop and exposes the bases to the solvent. Similarly, the increased solvent exposure of bases 9 and 10 is probably caused by NC(11–55) binding at its second site at the 3'-end of the loop.[27] As reported for other 3HC-based fluorophores,[39,40] the NC(11–55)-induced increase in the QY of 3HCnt is likely due to restricted rotation and co-planarization of the chromophore. This restricted motion is fully in line with the NC(11–55)-induced restriction in the local motion of nucleobases, previously evidenced with 2AP-labeled (–)PBS sequences.[23]

To determine whether 3HCnt alters the binding of NC(11–55) to ΔP(–)PBS, competition experiments measuring the relative affinity of NC(11–55) for labeled ΔP(–)PBS versus non-labeled ΔP(–)PBS, were performed.[41] To this end, NC(11–55) was mixed with equimolar concentrations of labeled and non-labeled ΔP(–)PBS. If 3HCnt does not modify the binding of NC(11–55) to ΔP(–)PBS, the peptide should be equally distributed between non-labeled and labeled ΔP(–)PBS. This was found to be the case for 3HCnt-8ΔP(–)PBS and 3HCnt-9ΔP(–)PBS (Figure 3 and Figure S3), indicating that substitutions at positions 8 and 9 marginally alter the affinity of the ODN for the peptide. In contrast, a substantial preference of NC(11–55) for the non-labeled sequence was observed when 3HCnt was replacing G7, indicating that 3HCnt decreases by at least 3-fold the affinity for the protein, likely due to the key role of G7 in the binding process.[27,42] An opposite effect was observed when C10 was replaced by 3HCnt, suggesting an about two-fold increase in ΔP(–)PBS affinity for the peptide. Notably, the rather low binding observed with 3HCnt-2ΔP(–)PBS is likely a consequence of the low affinity of the NC(11–55) peptide for the stem, so that this binding site is easily depopulated in the presence of the competing non-labeled sequence.

Finally, we investigated the changes in the fluorescence properties of the labeled ΔP(–)PBS sequences on conversion from the stem-loop to the extended duplex (Figure S4). Duplex formation was accompanied by a decrease in the N*/T* ratio (Figure 2b) and a substantial increase in the QY (Figure 2c) at

nearly all positions. The observed drop of the N*/T* ratio is in line with the expected environmental polarity decrease upon 3HCnt. Similarly, the observed QY increase is likely a consequence of the diminished rotation between the thienyl and 3HC moieties upon intercalation of 3HCnt. The minimal changes in the N*/T* ratio and QY of 3HCnt at position 10 are probably due to its tight stacking with its neighboring G11 in the ΔP(−)PBS stem-loop,[27] so duplex formation only marginally changes its exposure to solvent and its planarity.
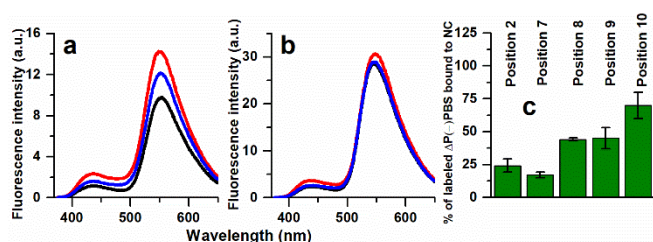

Figure 3. Binding competition experiments. Fluorescence spectra of 2 µM 3HCnt-9ΔP(−)PBS (a) and 3HCnt-7ΔP(−)PBS (b) were recorded in the absence (black line) or in the presence of NC (11−55) added at a ratio 1:1 (red). To this mixture, one equivalent of non-labeled ΔP(−)PBS (blue) was added. (c) Fraction of NC(11−55) protein bound to labeled ΔP(−)PBS after addition of non-labeled ΔP(−)PBS. A fraction of 50% corresponds to an equal distribution of NC(11−55) between labeled and non-labeled ΔP(−)PBS. Buffer was as in Figure 2. Excitation wavelength was 360 nm.

Taken together, our data show that the 3HCnt nucleobase can sensitively monitor the binding of NC(11−55) and the stem-loop to duplex conversion at most positions. As 3HCnt at position 9 marginally affects the affinity for NC(11−55) and shows a high sensitivity to the binding of NC(11−55) and formation of the duplex, this position was further selected for monitoring the (+)/(−)PBS annealing reaction. In comparison to 2AP that is almost non-emissive in double-stranded DNA, the QY of thG or 3HCnt increases upon duplex formation, reaching values above 10%. This key advantage over 2AP is beneficial as it facilitates the real time monitoring of the (+)/(−)PBS annealing reaction at much lower concentrations.

Real-time monitoring of (−)PBS stem-loop conversion into (+)/(−)PBS duplex

Annealing of (−)PBS with (+)PBS in the absence of NC: To monitor the annealing kinetics of (−)PBS with (+)PBS, we switched from the truncated ΔP(−)PBS to the full-length (−)PBS to take into account the contributions of the overhangs. As the stem-loop to duplex conversion was found to provide a larger QY increase with 3HCnt at position 9 (6-fold) than for thG at position 7 (2-fold), we first used the labeled 3HCnt-9(−)PBS sequence to monitor the annealing kinetics. By mixing this sequence with the complementary non-labeled (+)PBS in pseudo first-order conditions, we could follow the formation of the extended (+)/(−)PBS duplex in real-time through the fluorescence increase displayed by 3HCnt (Figure 4a).
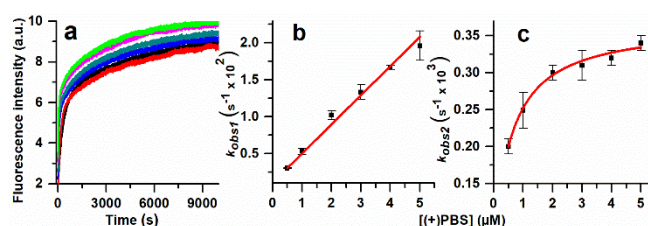
Figure 4. Kinetics of 3HCnt-9(−)PBS annealing with (+)PBS in the absence of NC. (a) Kinetic traces of 0.05 µM 3HCnt-9(−)PBS with (+)PBS concentrations of 0.5 µM (black), 1 µM (red), 2 µM (blue), 3 µM (green), 4 µM (magenta) and 5 µM (light green), respectively. The kinetic curves were fitted to equation 1 to recover the $k_{obs1,2}$ values. Dependence of the observed kinetic rate constants k_{obs1} (b) and k_{obs2} (c) as a function of (+)PBS concentration. The $k_{obs1,2}$ values are the means for two experiments. The red lines in (b) and (c) correspond to the fit of the data points (black squares) to equations 2 and 3, respectively. Excitation and emission wavelengths were 360 and 540 nm. Buffer was as in Figure 2.

The kinetic traces of (+)/(−)PBS annealing (Figure 4a) could be appropriately fitted with equation (1). The observed rate constants k_{obs1} and k_{obs2} obtained from the fits were found to show respectively, a linear (Figure 4b) and hyperbolic dependence (Figure 4c) on the (+)PBS concentration, in consistency with a two-step model where an intermediate complex (IC) is further converted through a rate-limiting step into an extended duplex (ED)[22,23]:

$$(-)PBS + (+)PBS \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} [IC] \underset{k_{-f}}{\overset{k_f}{\rightleftharpoons}} ED$$
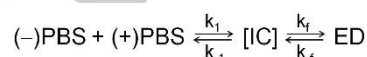
This model is characterized by k_1 and k_{-1}, the kinetic rate constants for the formation and dissociation of the IC, and by k_f and k_{-f}, the forward and backward rate constants for the conversion of the IC into the final duplex. The fit of the k_{obs1} (Figure 4b) and k_{obs2} (Figure 4c) dependence on the (+)PBS concentration to equation 2 and 3, respectively gave the k_1, k_{-1}, K_m, k_f and k_{-f} values reported in Table 1. Notably, the K_m value was consistent with the value obtained from the k_1/k_{-1} ratio, highlighting the internal consistency of our data and the appropriate selection of the kinetic model. Additionally, a negligibly small value was obtained for k_{-f}, indicating that the conversion of the IC into the ED is nearly irreversible, as expected from the high stability of the ED. To support the results of the analytical calculations, we used the Dynafit numerical solving software,[43,44] which makes no assumption on the relative values of the rate constants in the selected kinetic model. The kinetic rate constants obtained by Dynafit and the analytical approach corresponded perfectly (Table1), improving the confidence in the values obtained. In addition, the numerical approach revealed that the fluorescence intensity of the IC was about 70% of the ED, so that it corresponds to a QY of 9.1%. This clearly suggests that the loops also interact with one another in the IC, in contrast to the previously proposed model where the interaction was limited to the overhangs.[23] To further evaluate the contribution of the overhangs to the (+)/(−)PBS annealing, we monitored the annealing kinetics using 3HCnt-9ΔP(−)PBS and ΔP(+)PBS, where the overhangs were missing (Figure 1).

Table 1. Kinetic parameters of the annealing of 3HCnt-9(–)PBS with (+)PBS in the absence and presence of NC derivatives.

Protein	Probe	Approach	k_1 ($M^{-1}s^{-1} \times 10^{-3}$) [a]	k_{-1} ($s^{-1} \times 10^3$) [a]	k_f ($s^{-1} \times 10^4$) [a]	k_1/k_{-1} (K_m) ($M^{-1} \times 10^{-6}$)
-	3HCnt	Analytical	3.5 ± 0.2	1.4 ± 0.2	3.5 ± 0.1	2.5 (2.5)
		Numerical	4.5 ± 0.4	1.1 ± 0.1	3.0 ± 0.1	4.1
	thG	Numerical	11 ± 1	2.3 ± 0.4	2.4 ± 0.1	4.8
NC(1–55)	3HCnt	Analytical	230 ± 4	20 ± 1	930 ± 200	11.5
		Numerical	220 ± 10	41 ± 10	940 ± 30	5.4
NC(11–55)	3HCnt	Analytical	75 ± 5	20 ± 9	90 ± 3	3.7 (0.9)
		Numerical	75 ± 5	54 ± 1	70 ± 1	1.4
(SSHS)₂NC(11–55)	3HCnt	Numerical	470 ± 80	210 ± 20	1000 ± 400	2.2

[a] k_1 and k_{-1} are the kinetic rate constants for the formation and dissociation of the IC; K_m is the equilibrium constant for IC formation calculated from the k_1/k_{-1} ratio or equation 3 (in brackets); k_f is the forward rate constant for the conversion of the IC into the final duplex. For the analytical approach, the reported rate constant values are given as mean ± standard error of the mean for three experiments. For the Dynafit approach, the reported values and their errors were calculated from the least-squares fit of experimental data using the Levenberg-Marquardt algorithm as implemented by Reich.[45]

The k_{obs1} and k_{obs2} values of the ΔP(–)/ΔP(+)PBS kinetic trace (Figure S5) were one order of magnitude smaller than for (+)/(–)PBS, confirming that the overhangs play a pivotal role in the annealing reaction.

Our approach, using a fluorescence nucleobase surrogate, appears far superior to the previous one based on the double labeling of (–)PBS at its 5' and 3' ends by the rhodamine6G/Dabcyl FRET pair,[22,23] where only a single kinetic parameter ($k = 3.8 \times 10^3$ M^{-1} s^{-1}) could be recovered. Moreover, as this k value is in reasonable agreement with the product (k_1/k_{-1})×k_f = 1.2 × 10³ M^{-1} s^{-1}, this further indicates that the two approaches consistently describe the annealing reaction.

To substantiate the results obtained with 3HCnt-9(–)PBS and further validate the effectiveness of the tested nucleoside surrogates in deciphering the mechanisms of nucleic acid transitions, we monitored the same annealing kinetics by using thG7(–)PBS (Figure 5). Fit of the kinetic traces by DynaFIT provided k_1, k_{-1} and k_f values in good agreement with the values obtained using 3HCnt (Table 1). In addition, the value of the product (k_1/k_{-1})×k_f = 1.1 × 10³ $M^{-1}s^{-1}$ corresponds to that obtained with the 3HCnt-labeled (–)PBS and the k value of the doubly labeled (–)PBS,[22] indicating that all three approaches are consistent. Noticeably, the fluorescence intensity of the IC was almost identical to the ED one, indicating that thG's microenvironment was similar in the IC and ED context, confirming an interaction between the loops in the IC. This close similarity of IC and ED fluorescence intensities also readily explains the observed kinetic curves. At high concentrations of (+)PBS, well above the K_d value ($K_d = k_{-1}/k_1 = 0.2$ µM) of the IC, the rapid reach of the fluorescence plateau is due to the fast and

nearly total conversion of the free (–)PBS into the IC. The subsequent conversion of the IC into ED being spectroscopically silent, no further increase in fluorescence can be observed. At lower (+)PBS concentration, only a fraction of the free (–)PBS is converted to IC according to the mass action law, so that a smaller initial increase in fluorescence is observed. Then, the subsequent slower fluorescence increase is the result of the slow conversion of the IC to ED, that progressively shifts the thG7(–)PBS from its free form toward the ED.
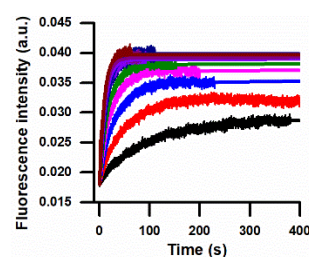
Figure 5. Annealing kinetic traces of thG7(–)PBS with (+)PBS and their fit using Dynafit. The kinetic traces (thick lines) were recorded by stopped-flow technique with 0.1 µM thG7(–)PBS and either 1 µM (black), 2 µM (red), 3 µM (blue), 4 µM (magenta), 5 µM (green), 6 µM (navy), 7 µM (violet), 8 µM (purple) or 9 µM (brown) (+)PBS, respectively. The kinetic traces were fitted by Dynafit (thin lines) with the parameters given in Table 1. Excitation wavelength was 360 nm. Buffer 25 mM TRIS, 30 mM NaCl and 0.2 mM MgCl₂.

Taken together, both 3HCnt and thG were highly capable of monitoring nucleic acid conversions, allowing us for the first time

to determine the full set of kinetic rate constants of the (+)/(–)PBS annealing reaction and to evidence that the IC is nucleated both through the overhangs and the loops.

Annealing of (–)PBS with (+)PBS promoted by NC(11–55): We next investigated the (+)/(–)PBS annealing kinetics in the presence of NC(11–55). Using the labeled 3HCnt-9(–)PBS sequence, a nearly two-fold increase in fluorescence was found to accompany the NC(11–55)-promoted annealing reaction. The fluorescence plateau was reached much faster than in the absence of NC (compare Figure 6a to Figure 4a), confirming the ability of NC(11–55) to promote annealing of complementary strands.[22,23,46–49] The k_{obs1} and k_{obs2} values were obtained by stopped-flow (Figure 6b) and steady-state (Figure 6a) monitoring of the annealing reactions.
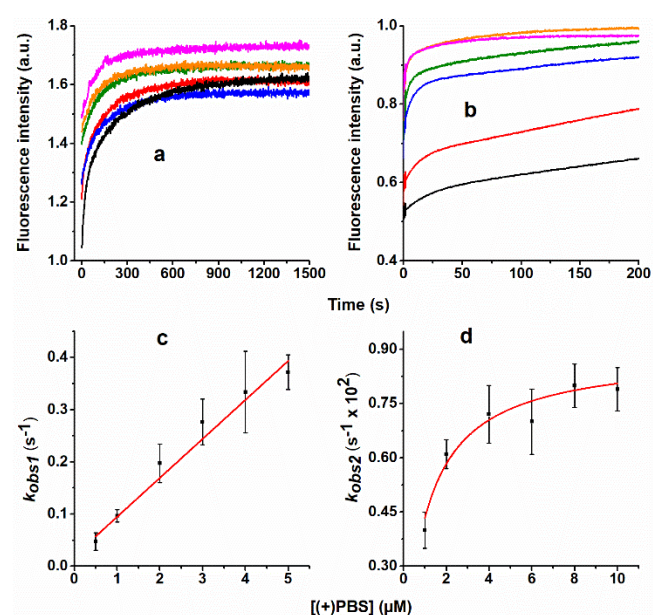
Figure 6. Kinetics of (+)/(–)PBS annealing in the presence of NC(11–55). (a) Kinetic traces monitored by steady-state spectroscopy of 0.05 µM 3HCnt-9(–)PBS with 0.5 µM (black), 1 µM (red), 2 µM (blue), 3 µM (green), 4 µM (orange) and 5 µM (magenta) (+)PBS, respectively. (b) Stopped-flow kinetic traces of 0.1 µM 3HCnt-9(–)PBS with 1 µM (black), 2 µM (red), 4 µM (blue), 6 µM (green), 8 µM (orange) and 10 µM (magenta) (+)PBS, respectively. NC(11–55) was added at a 3:1 peptide:ODN ratio. Buffer was as in Figure 2. Excitation wavelength was 360 nm. Dependence of the observed kinetic rate constants k_{obs1} (c) and k_{obs2} (d) as a function of (+)PBS concentration. The k_{obs1} and k_{obs2} values were obtained respectively, from the fits of the curves in (a) and (b) to equation 1. The $k_{obs1,2}$ values are the means +/- standard deviation for two experiments. Red lines in c) and d) correspond to the fit of the data points (black squares) to equations 2 and 3, respectively.

As in the absence of NC(11–55), the linear (Figure 6c) and hyperbolic (Figure 6d) dependence of the k_{obs1} and k_{obs2} values, respectively, on the (+)PBS concentration indicated a two-step annealing reaction. The k_1 and k_{-1} values were found to be 20- and 14-fold faster, respectively, than in the absence of the protein (Table 1). Moreover, the k_f value was 26-fold larger than in the absence of peptide. Noticeably, the consistency of the K_m values calculated from the k_1/k_{-1} ratio or from equation 3, as well as the

very similar values for all parameters obtained by using Dynafit (Table 1), further enhanced the confidence in the model and the data. The fluorescence intensity of the IC was found to be 87% of the value observed for the ED, confirming the involvement of the loops in the NC-promoted (+)/(–)PBS annealing reaction.[22,23] Finally, the product $(k_1/k_{-1}) \times k_f = 1.0 \times 10^4$ M^{-1}s^{-1} is in good agreement with the single kinetic parameter $k = 1.8 \times 10^4$ M^{-1}s^{-1} obtained by using (–)PBS species doubly labeled by Rh6G and Dabcyl,[22] confirming the suitability of 3HCnt for monitoring the annealing reaction.

The annealing kinetics with the truncated 3HCnt-9ΔP(–)PBS and ΔP(+)PBS ODNs in the presence of NC(11–55) added at a 3:1 ratio (Figure 7a, red curve) was almost superimposable with that obtained for the full-length (+)/(–)PBS (Figure 7a, black curve). This suggests that in contrast to the peptide-free conditions, the NC(11–55)-promoted annealing reaction is mediated through a loop-loop mechanism.[22,23]
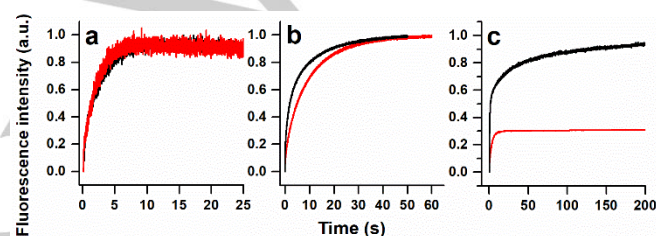
Figure 7. Role of the PBS overhangs in the NC(11–55)- (a), NC(1–55)- (b) and (SSHS)$_2$NC(11–55)- (c) promoted (+)/(–)PBS annealing. In (a) and (c), 0.1 µM 3HCnt-9(–)PBS (black) or 3HCnt-9ΔP(–)PBS (red) were reacted with 6 µM (+)PBS or ΔP(+)PBS in the presence of respectively, NC(11–55) and (SSHS)$_2$NC(11–55) added at a peptide:ODN ratio 3:1. In (b), 0.1 µM 3HCnt-9(–)PBS (black) or 3HCnt-9ΔP(–)PBS (red) were reacted with 1 µM (+)PBS or ΔP(+)PBS in the presence of NC(1–55) added at a peptide:ODN ratio 1:1. The kinetic parameters for the various progress curves are given in Table S3. Excitation wavelength was 360 nm. Buffer was as in Figure 2.

Next, we used thG7(–)PBS to characterize the NC(11–55)-promoted (+)/(–)PBS annealing mechanism. As NC(11–55) binding to thG7(–)PBS and conversion of thG7(–)PBS into the duplex provided nearly the same fluorescence increase, no significant change in fluorescence was observed when NC(11–55)-coated thG7(–)PBS was mixed with (+)PBS (Figure S6). This prevented us from monitoring the annealing reaction and confirming the values of the different parameters obtained by using the 3HCnt-9(–)PBS sequence.

Taken together, our data showed that 3HCnt but not thG could be used to monitor the NC-promoted annealing reaction using fluorescence intensity measurements. As in the absence of NC, the full set of kinetic parameters could be retrieved, giving a clearer picture on the mechanism by which NC(11–55) promotes the annealing reaction.

Annealing of 3HCnt-9(–)PBS with (+)PBS promoted by NC(1–55) and (SSHS)$_2$NC(11–55): To further assess the ability of 3HCnt to decipher the kinetic mechanisms of the conversion of the (–)PBS stem-loop into the (+)/(–)PBS duplex, we analyzed the annealing kinetics of 3HCnt-9(–)PBS with (+)PBS promoted by the full-length NC(1–55) and the (SSHS)$_2$NC(11–55) mutant.

FULL PAPER

The full-length NC(1–55) was added at a 1:1 ratio to limit the NC(1–55)-induced aggregation of ODNs.[23] The stopped-flow kinetic traces (Figure 8a) were much faster than with NC(11–55). Interestingly, the k_{obs1} values showed a linear dependence on the (+)PBS concentrations (Figure 8b), while the k_{obs2} values were nearly constant, suggesting that a plateau has been reached in the expected hyperbolic dependence on the (+)PBS concentration (Figure 8c). From equation 3 and the (+)PBS concentration range, this suggests that $K_m >> 10^6$ M^{-1} and thus, that the plateau value (0.093) corresponds to the k_f value. Using Dynafit and assuming a two-step reaction scheme, values of k_1, k_{-1} and k_f were obtained (Table 1). Both the agreement of k_1 and k_{-1} values with the fitted values of the dependence of k_{obs1} on (+)PBS concentration (Figure 8b) and the match of the k_f value with the plateau value of Figure 8c enhanced the confidence in the obtained values and the selected binding scheme. Moreover, the high K_m value (5.4×10^6 M^{-1}) calculated from the k_1/k_{-1} ratio is consistent with the rapid reach of the plateau observed in Figure 8c. The product $(k_1/k_{-1}) \times k_f = 5.0 \times 10^5$ M^{-1}s^{-1} is in line with the k value of 1.8×10^5 M^{-1}s^{-1} obtained using doubly labeled (–)PBS species,[22] confirming the consistency of the obtained rate constants. Interestingly, the kinetics of the NC(1–55)-promoted annealing of 3HCnt-9ΔP(–)PBS/ΔP(+)PBS and 3HCnt-9(+)/(–)PBS are very similar (Figure 7b and Table S3), indicating that the annealing reaction is mainly promoted through the PBS loops and that the overhangs play only a minor role in this reaction. Lastly, we investigated the promotion of 3HCnt-9(+)/(–)PBS annealing kinetics by the (SSHS)$_2$NC(11–55) mutant (Figure 1). This mutant peptide, which does not bind zinc and remains unfolded, is thought to promote nucleic acid annealing through a different pathway compared to NC(1–55) and NC(11–55).[47,50,51] The stopped-flow kinetic traces (Figure S7) were adequately fitted with Dynafit, assuming a two-step reaction. The (SSHS)$_2$NC(11–55) peptide was found to efficiently promote the association of the IC and its conversion into the ED, as can be seen from the k_1 and k_f values in Table 1. Moreover, the fluorescence intensity of the IC was found to be 95% of the value observed for the ED, indicating that the loops are involved in the IC formation. Furthermore, the product $(k_1/k_{-1}) \times k_f = 2.2 \times 10^5$ M^{-1}s^{-1} is consistent with the k value (1×10^5 M^{-1}s^{-1}) reported for the doubly labeled (–)PBS species,[23] giving credit to the selected two-step model and the values of the rate constants. Finally, a large decrease in both k_{obs1} and k_{obs2} values was observed when ΔP(–)PBS was substituted for PBS (Figure 7c and Table S3), indicating that the PBS overhangs play a key role in the (SSHS)$_2$NC(11–55)-promoted annealing process and that this peptide uses the same pathway as in the absence of NC.[23]
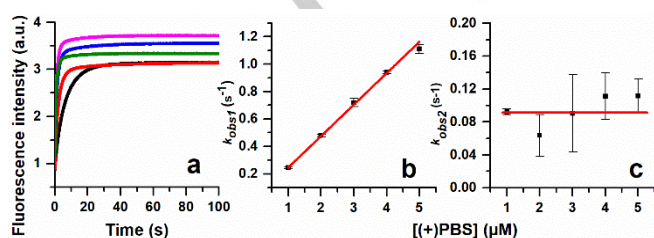
Figure 8. Kinetics of (+)/(–)PBS annealing in the presence of NC(1–55). (a) Stopped-flow kinetic traces of the NC(1–55)-promoted annealing of 0.1 μM 3HCnt-9(–)PBS with different concentrations of (+)PBS: 1 μM (black), 2 μM (red), 3 μM (green), 4 μM (blue), and 5 μM (magenta), respectively. NC(1–55) was added at a 1:1 peptide:ODN ratio. The progress curves were fitted with equation 1 and the $k_{obs1,2}$ values given in panels (b) and (c). Dependence of the observed kinetic rate constants k_{obs1} (b) and k_{obs2} (c) as a function of (+)PBS concentration in the presence of NC(1–55). The $k_{obs1,2}$ were obtained from the fits of the progress curves in panel (a) to equation 1. The red line in (b) corresponds to the fit of the data points (black squares) to equation 2 with $k_1 = 2.3 \times 10^5$ M^{-1}s^{-1} and $k_{-1} = 0.02$ s^{-1}. The red line in (c) indicates the mean value (0.093 s^{-1}) of k_{obs2}. Excitation wavelength was 360 nm. Buffer was as in Figure 2.

Taken together, our data with the different NCp7 variants show that 3HCnt at position 9 can confidently disclose the mechanism of (–)PBS conversion into (+)/(–)PBS duplex and recover the associated rate constant values.

Discussion

Here we used two versatile environmentally sensitive fluorescent nucleobase analogs, 3HCnt at position 9 and thG at position 7 of the (–)PBS loop, to directly monitor the role of the loops in (+)/(–)PBS annealing both in the absence and the presence of NC(11–55), NC(1–55) or (SSHS)$_2$C(11–55) (Figure 9). This allowed us to recover for the first time the entire set of kinetic rate constants governing these processes and to quantify the effects of all NC variants on each step.
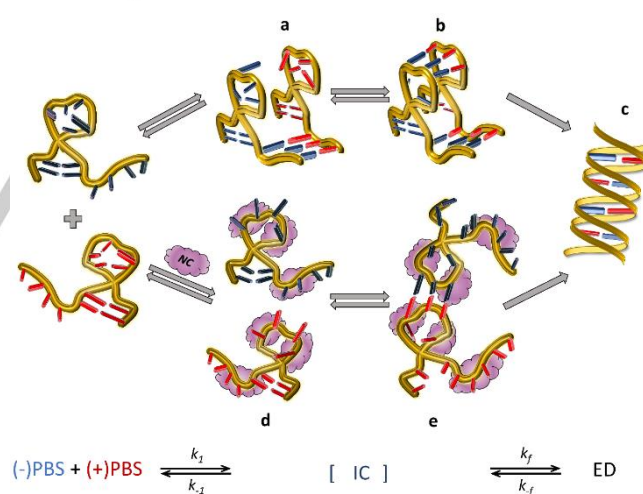
Figure 9. Proposed mechanism for (+)/(–)PBS annealing in the absence and the presence of NC derivatives, as revealed by the use of 3HCnt- and thG-labeled oligonucleotides. The nucleobases of the (+) and (–)PBS strands are shown in blue and red, respectively. In the absence of peptide or in the presence of 3 equivalents of (SSHS)$_2$NC(11–55) (upper pathway), an intermediate complex is initiated through annealing of the (–)PBS and (+)PBS overhangs (a). Due to the high flexibility of the loops and their close proximity, the loops can base pair rapidly, which further stabilizes the intermediate complex (b) and facilitates the conversion into the extended duplex (c). In contrast, through a specific binding to the PBS loops, both NC(1–55) and NC(11–55) (magenta clouds) expose the nucleotides of PBS loops to the solvent (d), which allows the formation of a loop-loop kissing complex (e). Moreover, both NC(1–55) and NC(11–55) destabilize the stem (e), which facilitates the conversion of the loop-

loop intermediate into the final duplex (c). The kinetic rate constants associated to these pathways are given in Table 1.

We found that without NC, (+)/(−)PBS annealing was governed by a two-step pathway, nucleating via the overhangs (Figure 9, upper pathway). The large fluorescence increase associated with the formation of the IC for both probes evidenced the existence of loop-loop contacts in the IC that have never been reported before. These contacts were likely the cause for the slow annealing reaction observed when ΔP(−)PBS was substituted for PBS (Figure S5). They likely consisted of Watson-Crick base pairs formed between the complementary loops as a result of their high flexibility.[27,42] In support of this hypothesis, molecular dynamics calculations have recently shown that ΔP(−)PBS can adopt conformations where the loop nucleobases are projected into the solvent[12] (Figure 9a). As these minor conformations are transient, the probability of productive collisions is low, which explains the slow ΔP(+)/ΔP(−)PBS annealing reaction. In contrast, the single-stranded PBS overhangs being unfolded expose their nucleobases and are thus prone to productive collisions (Figure 9a), explaining the faster IC formation with PBS as compared to ΔP(−)PBS. In the (+)/(−)PBS IC, the annealing of the overhangs likely holds the complementary loops in close proximity (Fig. 9a). As a result of this proximity and the fast interconversion between loop conformations,[12] the complementary loops can base pair rapidly (Figure 9b), which stabilizes the IC and thus favors the conversion into the ED (Figure 9c) by preventing the IC dissociation.

The (SSHS)$_2$NC(11−55) mutant was found to strongly promote the conversion of (−)PBS into the (+)/(−)PBS duplex, through the same pathway as in the absence of peptide (Figure 9, upper pathway). In comparison to the reaction in the absence of peptide, (SSHS)$_2$NC(11−55) induces a two orders of magnitude increase of all three kinetic parameters (k_1, k_{-1} and k_f). The increase in k_1 and k_f values can be rationalized by the efficient electrostatic attraction between the complementary strands permitted by the high charge density and flexibility of this mutant, as well as by its ability to rapidly bind and dissociate from the ODNs.[37,51] The large increase in the IC dissociation rate k_{-1} may be a collateral effect of the strong attraction between the strands induced by the peptide. Indeed, this nucleic acid aggregation effect may not only favor productive IC but also nonspecific and unstable ICs that cannot convert into ED and thus, readily dissociate. In contrast, in the absence of peptide an IC can likely form only if all four nucleobases of the overhangs anneal together, as at least three to four subsequent base pairs are required for stable duplex nucleation.[47,52,53] This is a low probability event, but it generates a specific and more stable IC, prone to convert into the ED.

Both NC(1−55) and NC(11−55) were found to direct (+)/(−)PBS annealing through a loop/loop pathway (Figure 9, lower pathway), that is nearly independent of the PBS overhangs. The formation of a loop/loop IC is thought to result from the specific interaction with the NC(1−55) or NC(11−55) hydrophobic platform, which solvent-exposes the nucleobases of both (−)PBS and (+)PBS loops (Figure 9d), making them competent for annealing.[23,27] This solvent exposure of the loop nucleobases is fully supported by the large increase in the 3HCnt and thG fluorescence on

binding of both proteins and further substantiated by the increase of the N*/T* intensity ratio of 3HCnt. Together with the electrostatic attraction between the strands permitted by the high charge density of NC(11−55) and NC(1−55), this exposure of the loop nucleobases readily explains the efficient annealing of the loops (Figure 9e), evidenced by the increase by one and two orders of magnitude, respectively, in the k_1 value in respect to the same reaction in the absence of a protein. The lower activity of NC(11−55) as compared to NC(1−55) can be attributed to its lower electrostatic strand attraction capability due to the lack of the highly positively charged N-terminal domain.[37] It can also be explained by its approximately 10-fold lower affinity for nucleic acids as compared to NC(1−55),[25,37,50,54,55] so that NC(11−55) is more shortly bound to the PBS loop than NC(1−55), decreasing the probability of finding the proper base pairing required for IC nucleation. We also observed a 26- and 170-fold increase in the k_f value for NC(11−55) and NC(1−55), respectively, as compared to the reaction in the absence of protein. This can be rationalized by their ability to destabilize the upper[27] and lower base pair[24,56] of the (−)PBS stem (Figure 9e), and thus favor the invasion by the complementary (+)PBS species. Finally, the one to two orders of magnitude increase in the IC dissociation rate constant k_{-1} promoted by NC(11−55) and NC(1−55) as compared with the reaction in the absence of any protein, can be explained by the nucleic acid destabilization associated with the folded zinc fingers.[51,54,57–59] Though this effect is too small to dissociate long duplexes such as the ED, it can efficiently dissociate the five base pairs stretch between the loops in the IC. The increased destabilizing activity of NC(1−55) as compared to NC(11−55) was already described[37,59] and attributed to the destabilizing property of the N-terminal domain of NC[23,37,54] and/or to its higher affinity as compared to NC(11−55), which increases the probability for the folded zinc fingers to dissociate the IC through a prolonged contact with the ODNs.

Interestingly, the k_1 and k_f values (Table 1) were about 7-fold higher with (SSHS)$_2$NC(11−55) as compared to NC(11−55) (Figure 9a and b). This is likely related to the high flexibility of the former, which ensures an efficient neutralization of the ODN negative charges by its positively charged basic amino acids. Noticeably, the nonspecific electrostatic-driven effect of the (SSHS)$_2$NC(11−55) mutant has a stronger effect on the annealing rate constants than the specific destabilizing activity of NC, which confirms that the notions of efficiency and specificity in the promotion of (+)/(−)PBS annealing are disconnected.[23]

Both 3HCnt-9(−)PBS and thG7(−)PBS were found to be superior to Rh6G-5'-(−)PBS-3'-Dabcyl for deciphering the mechanism of (−)PBS stem-loop conversion into the (+)/(−)PBS duplex. The main reason is that IC formation both in the pathway through the loops and the pathway through the overhangs does not induce a significant change in the distance between the 5' and 3' ends of the doubly labeled species. As a result, IC formation does not yield any change in FRET, and only its conversion into the ED can be monitored, explaining that only a single overall kinetic constant can be obtained. In contrast, as loop-loop interactions are observed in both pathways, a substantial fluorescence increase of 3HCnt and thG inserted in the loop accompanies the IC formation, so that the kinetic rate constants of both the IC

formation and conversion into ED steps can be extracted. The two probes were found also superior to 2Ap, which exhibits very low quantum yields in both (–)PBS stem-loop and (+)/(–)PBS duplex.[23] Therefore, higher concentrations of at least an order of magnitude would be required with 2Ap-labeled (–)PBS, which would lead to kinetics too fast to resolve.

Since we obtained the same kinetic rate constants using two different probes with different structural and photophysical features, located at two different positions of (–)PBS, confidence in the data obtained is elevated. As 3HCnt is an universal nucleobase and as [th]G can replace any G residue, it is in principle possible to site-specifically monitor the annealing process at nearly any position of (–)PBS. However, as illustrated in Figure 2 and Figure 3 for 3HCnt, it is necessary to carefully select the positions in the oligonucleotide where the fluorescent nucleobase minimally disturbs the system, displays a significant quantum yield and provides a large amplitude response to the process under study. 3HCnt and [th]G appear complementary with respect to these criteria, at least within the system investigated here. While [th]G was shown to exhibit a good quantum yield and perfectly substitute G at position 7, it gave similar fluorescence changes on NC binding and duplex formation, so that only the annealing in the absence of NC could be investigated. In contrast, 3HCnt fulfilled well all criteria at position 9, but exhibited low QY and/or perturbed NC binding at several other positions. Thus, having both fluorescent nucleobases increases the likelihood of finding appropriate positions that permit faithful extraction of the kinetic rate constants of a given structural transition in a nucleic acid sequence.

Conclusions

The combination of 3HCnt and [th]G appears well complementary to FRET approaches for monitoring and characterizing a large number of structural transitions under various environmental stimuli. FRET approaches are invaluable for monitoring large structural transitions, such as for instance between different conformations of Holliday junctions and G-quadruplexes.[4,5] Due to the dependence of FRET on the distance and/or orientation of the FRET donor-acceptor pair, the FRET approach can give unique information on the geometry of the initial and final species. On such systems, the approach with the environmentally-sensitive fluorescent nucleobases described here could be helpful for confirming and completing the description of the mechanism of these transitions. By selectively monitoring structural changes at defined locations, these nucleobases can provide key information on intermediate species and the involvement of their close neighbors in the structural transitions. Our approach will even be more useful in systems where transitions induce only local structural changes, as for instance the single step branch migration of a DNA Holliday junction, which implies the disruption of two base pairs on opposite arms, rotation around, and pairing to form two new terminal base pairs.[2] Finally, it should be mentioned that our approach is not limited to thG and 3HCnt. A number of other fluorescent nucleobases[13,14] that minimally perturb the labeled nucleic acids and exhibit improved brightness and site-specific responses to structural changes can likely be used for the same purpose. The choice of the most appropriate fluorescent nucleobase will probably depend on the chosen system, but because of the wide variety of chemical classes of fluorescent nucleobases and labeling methods, it is expected that the dynamics of almost all types of structural transitions in nucleic acids could be solved.

Experimental Section

All solvents and chemicals were purchased from Sigma-Aldrich Chemical Company. Non-labeled ΔP(–)PBS, ΔP(+)PBS, (–)PBS DNA, and (+)PBS DNA were purchased from IBA GmbH (Germany). Concentrations of ODNs were calculated from their absorbance using the molar extinction coefficients at 260 nm provided by the supplier. NC(1–55), NC(11–55) and (SSHS)$_2$NC(11–55) were synthesized by solid-phase peptide synthesis, as described in.[60] An extinction coefficient ε_{280} = 5700 M^{-1}cm^{-1} was used to determine their concentrations by absorption. NC stock solutions were prepared with 2.0 eq Zn(II) in: a) 10 mM phosphate buffer, 30 mM NaCl (pH = 6.5) for 3HCnt-labeled sequences and, b) 25 mM TRIS, 30 mM NaCl and 0.2 mM MgCl$_2$ for [th]G-labeled sequences (pH 7.5)

Synthesis and preparation of the labeled ODNs

The 3HCnt nucleobase was synthesized as described in.[15] Then, using solid-support ODN synthesis, 3HCnt was selectively introduced at position 2, 7, 8, 9 or 10 in the ΔP(–)PBS DNA sequences. For the kinetic measurements, 3HCnt was introduced at position 9 of (–)PBS DNA. [th]G nucleobase and [th]G-substituted at position 7 in (–)PBS were synthesized as described previously.[20] Lyophilized labeled ODNs were dissolved in deionized water and their concentrations were determined by absorption spectroscopy using the molar extinction coefficients at 260 nm calculated by using the ODN calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html) and taking into account the molar absorption coefficient of 3HCnt (ε_{260} = 12000 M^{-1}cm^{-1}) and [th]G (ε_{260} = 4150 M^{-1}cm^{-1}). The double stranded ODNs were prepared by mixing the labeled sequences with their non-labeled complementary sequences at a 1:3 molar ratio in PCR tubes (200 µL size), annealing them for 3 min at 85°C in a water bath and then, slowly cooling them down to room temperature.

Spectroscopic measurements

Absorption spectra were recorded on a Cary 4000 UV-visible spectrophotometer (Varian). Fluorescence spectra were recorded on a FluoroMax4 spectrofluorimeter (Horiba) equipped with a thermostated cell compartment at 20 ± 0.5 °C. Fluorescence spectra were corrected for Raman scattering, lamp fluctuations and instrumental wavelength-dependent bias. Excitation wavelength was 360 nm for the 3HCnt fluorophore. For quantum yield (QY) measurements, free 2-thienyl-3-hydroxychromone fluorophore in phosphate buffer (QY = 0.046)[16] or quinine sulfate in 0.5 M H$_2$SO$_4$ (QY = 0.546)[61] was used as a reference. All spectroscopic measurements for 3HCnt-labeled sequences were performed in 10 mM phosphate buffer (pH 6.5), 30 mM NaCl, at 20 °C. A pH value of 6.5 was chosen to limit the deprotonation of the 3-OH group of the 3HCnt fluorophore. Spectroscopic measurements for [th]G-labeled sequences were performed in 25 mM TRIS, 30 mM NaCl and 0.2 mM MgCl$_2$ (pH 7.5) at 20°C.

Competition experiments

To determine the relative affinity of the labeled ODNs comparatively to the non-labeled ODNs for NC(11–55), competition experiments were performed. To this end, 2 µM of 3HCnt-labeled ΔP(–)PBS was mixed with 2 µM of NC(11–55), and then with 2 µM of the non-labeled ΔP(–)PBS. The distribution expressed in % of NC(11–55) between the two sequences was calculated by: $(I_2-I_0)/(I_1-I_0)\times100$, where I_0 is the initial fluorescence intensity of 3HCnt-ΔP(–)PBS, while I_1 and I_2 are the fluorescence intensities after addition of NC(11–55) and the non-labeled sequence, respectively. If the peptide has an equal affinity for the labeled and non-labeled ΔP(–)PBS sequences, the calculated parameter should be close to 50%. Preferential affinity to one of the sequences will result in deviations from this value. Excitation wavelength was 360 nm.

Thermal denaturation experiments

Melting temperatures were determined by measuring the absorbance changes at 260 nm as a function of the temperature using a Varian Cary 400 spectrophotometer equipped with a Peltier temperature controller. Absorbance was recorded in the forward and backward directions from 20 to 90°C at a rate of 0.5°C/min. Prior to the melting experiment, the double stranded ODNs were prepared by mixing two complementary sequences (labeled and non-labeled) at equimolar concentrations (1 µM) in a PCR tube (200 µL size), heated for 3 min at 85°C in the water bath and slowly cooled down to room temperature. Thermal denaturation experiments were performed in 10 mM cacodylate buffer, 150 mM NaCl, pH 7.0. To avoid evaporation, the samples were overlaid with mineral oil (Sigma-Aldrich). Melting temperatures were determined from the first derivative of the thermal denaturation curves.

Kinetic measurements

Kinetic measurements were performed under pseudo-first-order conditions by mixing 50 or 100 nM of 3HCnt-9(–)PBS DNA or 3HCnt-9ΔP(–)PBS DNA with a 10- to 100-fold molar excess of non-labeled (+)PBS DNA or ΔP(+)PBS, respectively. Kinetic measurements with [th]G-labeled sequences were performed by mixing 100 or 300 nM of [th]G7(–)PBS with (+)PBS in pseudo-first order conditions. The experiments were performed either in the absence or in the presence of NC(11–55) or (SSHS)₂NC(11–55) added at a 3:1 ratio to each complementary ODN or NC(1–55) added at a 1:1 ratio prior to the start of the annealing reaction. Measurements were performed using a Fluoromax 4 spectrofluorometer or a stopped-flow instrument (SFM-3, Bio-Logic) with 2.8 ms dead time.[62] For the steady-state fluorescence experiments, excitation and emission wavelengths for 3HCnt-labeled sequences were 360 and 540 nm, respectively. For [th]G-labeled sequences, the excitation was at 360 nm and emission was collected at 500 nm. For the stopped-flow experiments with 3HCnt-labeled sequences, excitation wavelength was 365 nm and fluorescence emission was detected above 500 nm through a cutoff filter (Wratten N°12, Kodak). For [th]G-labeled sequences, excitation was at 355 nm and fluorescence emission was detected above 420 nm through a Kodak cutoff filter.

The kinetic curves were fitted using the following bi-exponential function:

$$I(t) = I_f - (I_f - I_0)[a\exp(-k_{obs1}(t - t_0)) + (1 - a)\exp(-k_{obs2}(t - t_0))] \quad (1)$$

where I_0 and I_f are the initial and final intensities of 3HCnt-9(–)PBS or 3HCnt-9ΔP(–)PBS, respectively; a is the relative amplitude of the fast component; t_0 is the dead time; and k_{obs1} and k_{obs2} are the apparent pseudo-first order rate constants.

Assuming a two-step model, the kinetic rate constants were then determined from the dependence of $k_{obs1,2}$ on the (+)PBS concentration according to:

$$k_{obs1} = k_1[(+)PBS] + k_{-1} \quad (2)$$

$$k_{obs2} = (k_f K_m[(+)PBS])/(1 + K_m[(+)PBS]) + k_{-f} \quad (3)$$

where k_1 and k_{-1} are the kinetic rate constants for the formation and dissociation of the intermediate complex (IC); K_m is the equilibrium constant (k_1/k_{-1}) for the formation of the IC ; while k_f and k_{-f} are the forward and backward rate constants for the conversion of the IC into the final duplex.

A numerical approach using the Dynafit software[43,44] was used to confirm the kinetic rate constants obtained through the analytical approach and to determine the intensities of the intermediate products formed in the (+)/(–)PBS annealing reaction. DynaFit performs a least-squares fit of experimental data using the classic Levenberg-Marquardt algorithm as implemented by Reich.[45]

[1] M. Kaushik, S. Kaushik, K. Roy, A. Singh, S. Mahendru, M. Kumar, S. Chaudhary, S. Ahmed, S. Kukreti, *Biochem. Biophys. Reports* **2016**, *5*, 388–395.

[2] D. M. Lilley, *Q. Rev. Biophys.* **2000**, *33*, 109–59.

[3] S. Dzatko, M. Krafcikova, R. Hänsel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J.-L. Mergny, S. Foldynova-Trantirkova, L. Trantirek, *Angew. Chemie Int. Ed.* **2018**, *57*, 2165–2169.

[4] M. Aznauryan, S. Søndergaard, S. L. Noer, B. Schiøtt, V. Birkedal, *Nucleic Acids Res.* **2016**, *44*, 11024–11032.

[5] S. A. McKinney, A.-C. Déclais, D. M. J. Lilley, T. Ha, *Nat. Struct. Biol.* **2003**, *10*, 93–97.

[6] R. W. Sinkeldam, N. J. Greco, Y. Tor, *Chem. Rev.* **2010**, *110*, 2579–2619.

[7] L. M. Wilhelmsson, *Q. Rev. Biophys.* **2010**, *43*, 159–183.

[8] A. C. Jones, R. K. Neely, *Q. Rev. Biophys.* **2015**, *48*, 244–279.

[9] D. C. Ward, E. Reich, L. Stryer, *J. Biol. Chem.* **1969**, *244*, 1228–1237.

[10] E. L. Rachofsky, E. Seibert, J. T. Stivers, R. Osman, J. B. A. Ross,

Biochemistry **2001**, *40*, 957–967.

[11]   A. Dallmann, L. Dehmel, T. Peters, C. Mügge, C. Griesinger, J. Tuma, N. P. Ernsting, *Angew. Chemie Int. Ed.* **2010**, *49*, 5989–5992.

[12]   K. Voltz, J. Léonard, P. T. Touceda, J. Conyard, Z. Chaker, A. Dejaegere, J. Godet, Y. Mély, S. Haacke, R. H. Stote, *Nucleic Acids Res.* **2016**, *44*, 3408–3419.

[13]   W. Xu, K. M. Chan, E. T. Kool, *Nat. Chem.* **2017**, *9*, 1043–1055.

[14]   M. Wilhelmsson, Y. Tor, Eds. , *Fluorescent Analogs of Biomolecular Building Blocks*, John Wiley & Sons, Inc, Hoboken, NJ, USA, **2016**.

[15]   D. Dziuba, V. Y. Postupalenko, M. Spadafora, A. S. Klymchenko, V. Guérineau, Y. Mély, R. Benhida, A. Burger, *J. Am. Chem. Soc.* **2012**, *134*, 10209–10213.

[16]   M. Spadafora, V. Y. Postupalenko, V. V. Shvadchak, A. S. Klymchenko, Y. Mély, A. Burger, R. Benhida, *Tetrahedron* **2009**, *65*, 7809–7816.

[17]   L. Zargarian, A. Ben Imeddourene, K. Gavvala, N. P. F. Barthes, B. Y. Michel, C. A. Kenfack, N. Morellet, B. René, P. Fossé, A. Burger, et al., *J. Phys. Chem. B* **2017**, *121*, 11249–11261.

[18]   A. A. Kuznetsova, N. A. Kuznetsov, Y. N. Vorobjev, N. P. F. Barthes, B. Y. Michel, A. Burger, O. S. Fedorova, *PLoS One* **2014**, *9*, e100007.

[19]   D. Shin, R. W. Sinkeldam, Y. Tor, *J. Am. Chem. Soc.* **2011**, *133*, 14912–14915.

[20]   M. Sholokh, R. Sharma, D. Shin, R. Das, O. A. Zaporozhets, Y. Tor, Y. Mély, *J. Am. Chem. Soc.* **2015**, *137*, 3185–3188.

[21]   V. Kilin, K. Gavvala, N. P. F. Barthes, B. Y. Michel, D. Shin, C. Boudier, O. Mauffret, V. Yashchuk, M. Mousli, M. Ruff, et al., *J. Am. Chem. Soc.* **2017**, *139*, 2520–2528.

[22]   N. Ramalanjaona, H. de Rocquigny, A. Millet, D. Ficheux, J.-L. Darlix, Y. Mély, *J. Mol. Biol.* **2007**, *374*, 1041–1053.

[23]   J. Godet, N. Ramalanjaona, K. K. Sharma, L. Richert, H. de Rocquigny, J.-L. Darlix, G. Duportail, Y. Mély, *Nucleic Acids Res.* **2011**, *39*, 6633–6645.

[24]   C. Egélé, E. Schaub, N. Ramalanjaona, E. Piémont, D. Ficheux, B. Roques, J.-L. Darlix, Y. Mély, *J. Mol. Biol.* **2004**, *342*, 453–466.

[25]   J.-L. Darlix, J. Godet, R. Ivanyi-Nagy, P. Fossé, O. Mauffret, Y. Mély, *J. Mol. Biol.* **2011**, *410*, 565–581.

[26]   M. Sholokh, R. Improta, M. Mori, R. Sharma, C. Kenfack, D. Shin, K. Voltz, R. H. Stote, O. A. Zaporozhets, M. Botta, et al., *Angew. Chemie Int. Ed.* **2016**, *55*, 7974–7978.

[27]   S. Bourbigot, N. Ramalanjaona, C. Boudier, G. F. J. Salgado, B. P. Roques, Y. Mély, S. Bouaziz, N. Morellet, *J. Mol. Biol.* **2008**, *383*, 1112–1128.

[28]   U. Asseline, M. Delarue, G. Lancelot, F. Toulme, N. T. Thuong, T. Montenay-Garestier, C. Helene, *Proc. Natl. Acad. Sci.* **1984**, *81*, 3297–3301.

[29]   N. Amann, R. Huber, H.-A. Wagenknecht, *Angew. Chemie Int. Ed.* **2004**, *43*, 1845–1847.

[30]   P. K. Mandal, A. Samanta, *J. Phys. Chem. A* **2003**, *107*, 6334–6339.

[31]   V. V. Shynkar, A. S. Klymchenko, Y. Mély, G. Duportail, V. G. Pivovarenko, *J. Phys. Chem. B* **2004**, *108*, 18750–18755.

[32]   V. I. Tomin, *Opt. Spectrosc.* **2011**, *110*, 550–556.

[33]   J. A. Thomas, R. J. Gorelick, *Virus Res.* **2008**, *134*, 39–63.

[34]   J. G. Levin, J. Guo, I. Rouzina,, K. Musier-Forsyth, in *Prog. Nucleic Acid Res. Mol. Biol.*, **2005**, pp. 217–286.

[35]   J. G. Levin, M. Mitra, A. Mascarenhas, K. Musier-Forsyth, *RNA Biol.* **2010**, *7*, 754–774.

[36]   C. Tisné, B. P. Roques, F. Dardel, *J. Biol. Chem.* **2004**, *279*, 3588–3595.

[37]   M. Cruceanu, *Nucleic Acids Res.* **2006**, *34*, 593–605.

[38]   J. Darlix, J. L. Garrido, N. Morellet, Y. Mély, H. de Rocquigny, in *Adv. Pharmacol.*, **2007**, pp. 299–346.

[39]   M. Sholokh, O. M. Zamotaiev, R. Das, V. Y. Postupalenko, L. Richert, D. Dujardin, O. A. Zaporozhets, V. G. Pivovarenko, A. S. Klymchenko, Y. Mély, *J. Phys. Chem. B* **2015**, *119*, 2585–2595.

[40]   V. Kilin, O. Glushonkov, L. Herdly, A. Klymchenko, L. Richert, Y. Mely, *Biophys. J.* **2015**, *108*, 2521–2531.

[41]   V. V. Shvadchak, A. S. Klymchenko, H. de Rocquigny, Y. Mély, *Nucleic Acids Res.* **2009**, *37*, e25–e25.

[42]   P. E. Johnson, R. B. Turner, Z. R. Wu, L. Hairston, J. Guo, J. G. Levin, M. F. Summers, *Biochemistry* **2000**, *39*, 9084–9091.

[43]   P. Kuzmič, *Anal. Biochem.* **1996**, *237*, 260–273.

[44]   P. Kuzmič, in *Methods Enzymol.*, **2009**, pp. 247–280.

[45]   J. G. Reich, *C Curve Fitting and Modeling for Scientists and Engineers*, McGraw-Hill, New York, **1992**.

[46]   J. C. You, C. S. McHenry, *J. Biol. Chem.* **1994**, *269*, 31491–5.

[47]   M. R. S. Hargittai, R. J. Gorelick, I. Rouzina, K. Musier-Forsyth, *J. Mol. Biol.* **2004**, *337*, 951–968.

[48]   M.-N. Vo, G. Barany, I. Rouzina, K. Musier-Forsyth, *J. Mol. Biol.* **2006**, *363*, 244–261.

[49]   M.-N. Vo, G. Barany, I. Rouzina, K. Musier-Forsyth, *J. Mol. Biol.* **2009**, *386*, 789–801.

[50]   S. V. Avilov, E. Piemont, V. Shvadchak, H. de Rocquigny, Y. Mely, *Nucleic Acids Res.* **2007**, *36*, 885–896.

[51]   M. C. Williams, I. Rouzina, J. R. Wenner, R. J. Gorelick, K. Musier-Forsyth, V. A. Bloomfield, *Proc. Natl. Acad. Sci.* **2001**, *98*, 6121–6126.

[52]   M. E. Craig, D. M. Crothers, P. Doty, *J. Mol. Biol.* **1971**, *62*, 383–401.

[53]   D. Pörschke, M. Eigen, *J. Mol. Biol.* **1971**, *62*, 361–381.

[54]   H. Beltz, C. Clauss, E. Piémont, D. Ficheux, R. J. Gorelick, B. Roques, C. Gabus, J.-L. Darlix, H. de Rocquigny, Y. Mély, *J. Mol. Biol.* **2005**, *348*, 1113–1126.

[55]   C. Vuilleumier, E. Bombarda, N. Morellet, D. Gérard, B. P. Roques, Y. Mély, *Biochemistry* **1999**, *38*, 16816–16825.

[56]   C. Égelé, E. Schaub, É. Piémont, H. de Rocquigny, Y. Mély, *C. R. Biol.* **2005**, *328*, 1041–1051.

[57]   J. Azoulay, J.-P. Clamme, J.-L. Darlix, B. P. Roques, Y. Mély, *J. Mol. Biol.* **2003**, *326*, 691–700.

[58]   H. Beltz, J. Azoulay, S. Bernacchi, J.-P. Clamme, D. Ficheux, B. Roques, J.-L. Darlix, Y. Mély, *J. Mol. Biol.* **2003**, *328*, 95–108.

[59]   M. C. Williams, I. Rouzina, V. A. Bloomfield, *Acc. Chem. Res.* **2002**, *35*, 159–166.

[60]   H. de Rocquigny, D. Ficheux, C. Gabus, M.-C. Fournié-Zaluski, J.-L. Darlix, B. P. Roques, *Biochem. Biophys. Res. Commun.* **1991**, *180*, 1010–1018.

[61]   W. H. Melhuish, *J. Phys. Chem.* **1961**, *65*, 229–235.

[62]   S. Maillot, A. Carvalho, J.-P. Vola, C. Boudier, Y. Mély, S. Haacke,
         J. Léonard, *Lab Chip* **2014**, *14*, 1767.

Entry for the Table of Contents

FULL PAPER

Newly synthesized fluorescent nucleobase analogues: 2-thienyl-3-hydroxychromone and thienoguanosine have been used to monitor the interconversion dynamics of nucleic acids structures. Both the probes are highly sensitive to report each step of HIV-1 (+)/(-)PBS interconversions both in absence and presence of protein.

Increase in Fluorescence
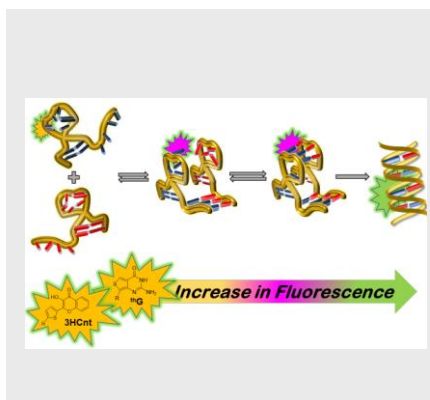
Marianna Sholokh, Rajhans Sharma, Natalia Grytsyk, Lyes Zaghzi, Viktoriia Y. Postupalenko, Dmytro Dziuba, Nicolas P.F. Barthes, Benoît Y. Michel, Christian Boudier, Olga A. Zaporozhets, Yitzhak Tor, Alain Burger and Yves Mély*

Environmentally-Sensitive Fluorescent Nucleoside Analogues for Surveying Dynamic Interconversions of Nucleic Acid Structures

Environmentally-Sensitive Fluorescent Nucleoside Analogues for Surveying Dynamic Interconversions of Nucleic Acid Structures

Marianna Sholokh,[a,c] Rajhans Sharma,[a] Natalia Grytsyk,[a] Lyes Zaghzi,[a] Viktoriia Y. Postupalenko,[a] Dmytro Dziuba,[b] Nicolas P.F. Barthes,[b] Benoît Y. Michel,[b] Christian Boudier,[a] Olga A. Zaporozhets,[c] Yitzhak Tor,[d] Alain Burger,[b] and Yves Mély*[a]

[a]   Dr. M. Sholokh, Dr. R. Sharma, Dr. N. Grytsyk, L. Zaghzi, Dr. V.Y. Postupalenko, Dr. C. Boudier, Prof. Dr. Y. Mély
        Laboratoire de Bioimagerie et Pathologies, UMR 7021 CNRS, Université de Strasbourg,
        Faculté de Pharmacie, 74 route du Rhin, 67401 Illkirch, France
        E-mail: yves.mely@unistra.fr
[b]   Dr. D. Dziuba, Dr. N.P.F. Barthes, Dr. B.Y. Michel, Prof. Dr. A. Burger
        Institut de Chimie de Nice, UMR 7272 CNRS, Université Côte d'Azur,
        Parc Valrose, 06108 Nice, France
[c]   Prof. Dr. O.A. Zaporozhets
        Department of Chemistry, Kyiv National Taras Shevchenko University,
        60 Volodymyrska street, 01033 Kyiv, Ukraine
[d]   Prof. Dr. Y. Tor
        Department of Chemistry and Biochemistry, University of California,
        San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0358, USA

        Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.2018xxxxx.

Melting temperatures of the native and 3HCnt-labeled ΔP(–)/ΔP(+)PBS duplexes
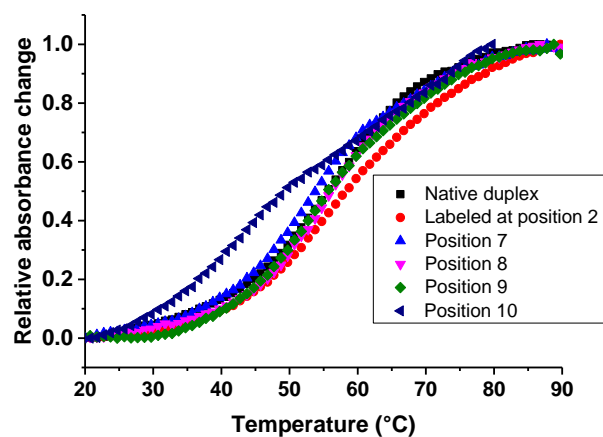
Figure S1 Normalized thermal denaturation curves of the non-labeled and 3HCnt-labeled ΔP(–)/ΔP(+)PBS duplexes. The absorbance was recorded in the forward and backward directions from 20 to 90 °C at a rate of 0.5 °C/min. The concentration of ODNs was 1 µM in 10 mM cacodylate buffer, 150 mM NaCl (pH 7.0).

Table S1. Melting temperatures of the PBS duplexes [a]

Sample	T_m (°C)
ΔP(–)PBS/ΔP(+)PBS	56 ± 1
3HCnt-2ΔP(–)PBS/ΔP(+)PBS	58 ± 2
3HCnt-7ΔP(–)PBS/ΔP(+)PBS	54 ± 1
3HCnt-8ΔP(–)PBS/ΔP(+)PBS	56 ± 1
3HCnt-9ΔP(–)PBS/ΔP(+)PBS	55 ± 2
3HCnt-10ΔP(–)PBS/ΔP(+)PBS	44 ± 2

[a] The melting temperatures, T_m, are expressed as mean ± standard error of the mean for at least two experiments.

Spectroscopic properties of the 3HCnt-labeled ΔP(–)PBS derivatives

Table S2. Spectroscopic properties of the 3HCnt-labeled ΔP(–)PBS derivatives in their free form, and in complex with NC(11–55) peptide and in duplex with ΔP(+)PBS [a]

	λ_{ABS} (nm)	λ_{N^*} (nm)	λ_{T^*} (nm)	N*/T*	QY (%)
3HCnt-2ΔP(–)PBS	373	436	545	0.19	0.4
+ NC(11–55)	373	432	547	0.19	1.7
+ ΔP(+)PBS	375	436	544	0.17	0.5
3HCnt-7ΔP(–)PBS	372	437	545	0.08	12.0
+ NC(11–55)	372	435	550	0.15	16.0
+ ΔP(+)PBS	373	431	544	0.04	31.0
3HCnt-8ΔP(–)PBS	375	434	548	0.11	1.9
+ NC(11–55)	373	436	547	0.17	9.9
+ ΔP(+)PBS	375	432	546	0.05	3.0
3HCnt-9ΔP(–)PBS	373	436	552	0.12	2.6
+ NC(11–55)	373	436	546	0.25	6.1
+ ΔP(+)PBS	374	435	545	0.04	13.0
3HCnt-10ΔP(–)PBS	373	434	549	0.07	3.4
+ NC(11–55)	373	435	549	0.14	3.7
+ ΔP(+)PBS	372	432	546	0.08	2.7

[a]λ_{ABS}, λ_{N^*}, λ_{T^*} are the maxima of the absorption band, and the N* and T* emission bands, respectively; N*/T* is the intensity ratio of the two emission bands measured at the peak maxima; QY is the fluorescence quantum yield. The reported data are the mean of n ≥ 2 experiments. Standard deviations are < 15% for N*/T* and QY values. Excitation wavelength was 360 nm. Experiments were performed in 10 mM phosphate buffer (pH 6.5), 30 mM NaCl at 20°C.

Fluorescence spectra of 3HCnt-labeled ΔP(–)PBS derivatives in the presence of increasing NC(11–55) concentrations
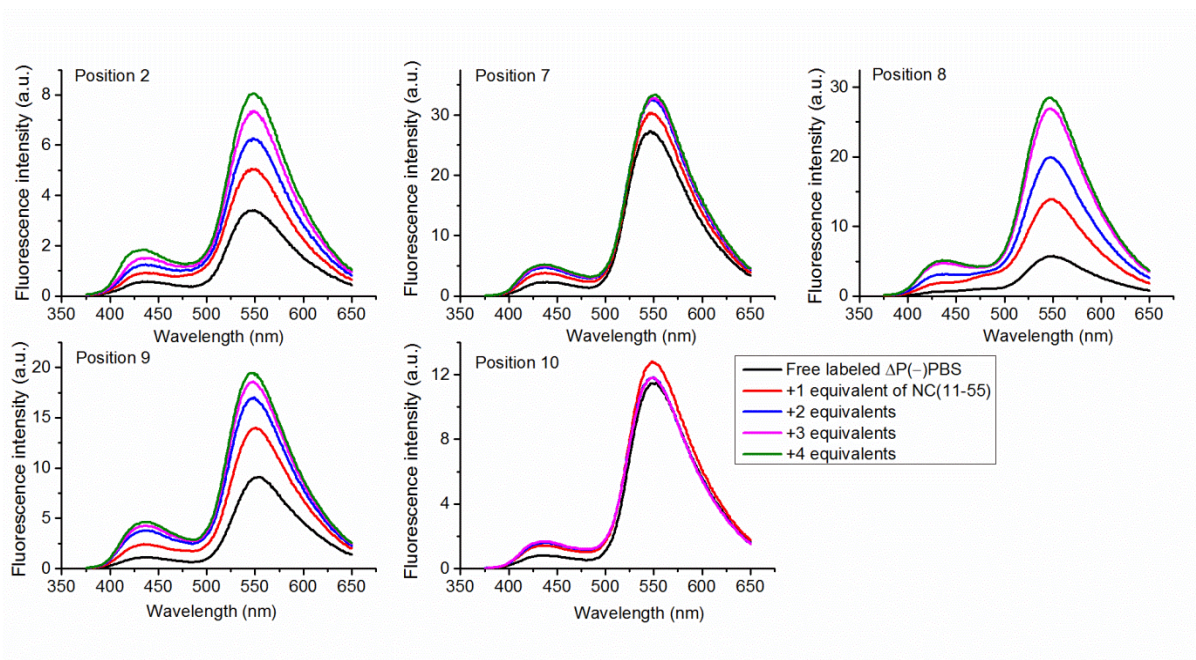
Figure S2 Fluorescence emission spectra of labeled 3HCnt-ΔP(–)PBS in the absence and in the presence of NC(11–55) added at 1:1; 1:2; 1:3 and 1:4 ODN : peptide ratios. Concentrations of labeled ODNs were 2 µM. Excitation wavelength was 360 nm. Experiments were performed in 10 mM phosphate buffer (pH 6.5), 30 mM NaCl at 20°C.

Binding competition experiments
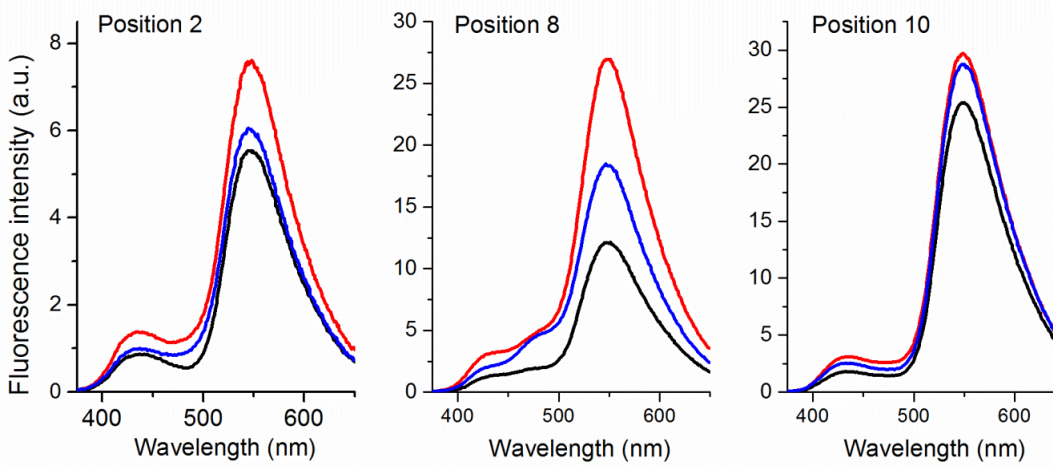
Figure S3 Binding competition experiments. Fluorescence spectra of the labeled 3HCnt-ΔP(−)PBS ODNs (2 μM) were recorded in the absence (black line) or in the presence of NC (11−55) added at a 1:1 ratio (red). To this mixture, one equivalent of non-labeled ΔP(−)PBS (blue) was added. Excitation wavelength was 360 nm. Experiments were performed in 10 mM phosphate buffer (pH 6.5), 30 mM NaCl at 20°C.

Fluorescence spectra of 3HCnt-labeled ΔP(+)/(−)PBS duplexes
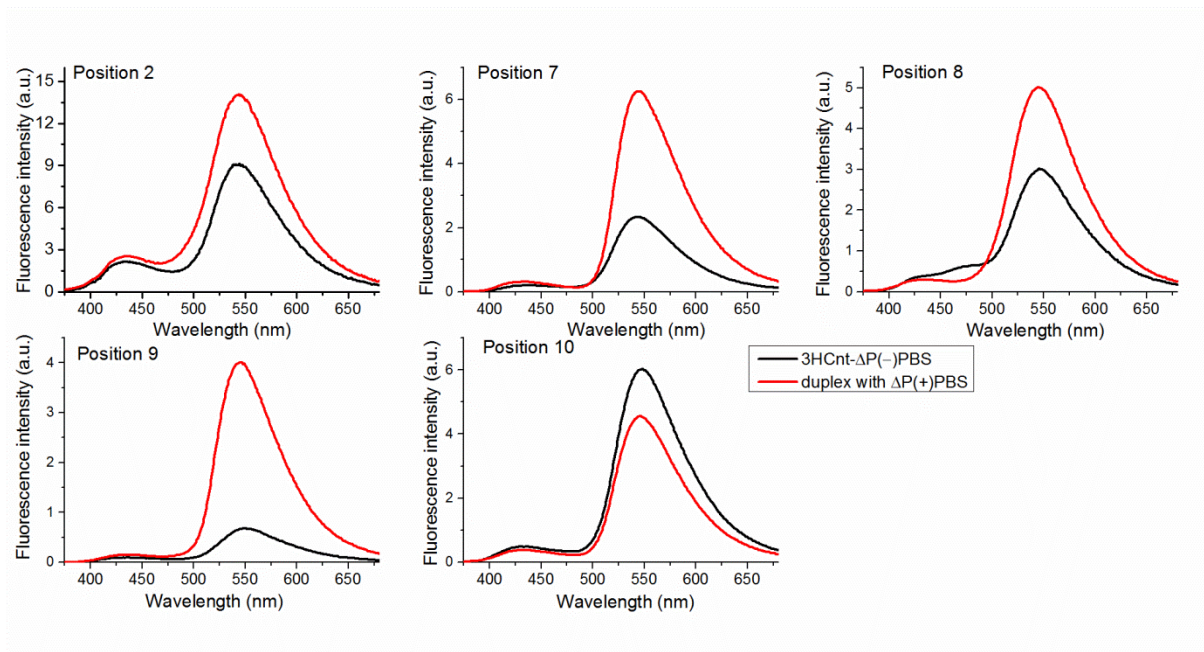
Figure S4. Fluorescence spectra of labeled 3HCnt-ΔP(−)PBS derivatives in their stem-loop and extended duplex forms. The duplexes were obtained by heating at 85°C a mixture of 2 µM 3HCnt-ΔP(−)PBS with ΔP(+)PBS added at 1:3 molar ratio, and then cooling down to room temperature. Excitation wavelength was 360 nm. Experiments were performed in 10 mM phosphate buffer (pH 6.5), 30 mM NaCl at 20°C.

(+)/(−)PBS annealing kinetics
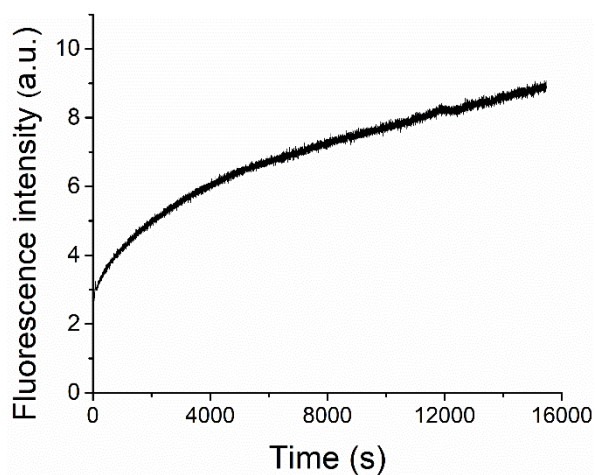
Figure S5. Annealing kinetics of 0.05 µM 3HCnt-9ΔP(−)PBS with 5 µM ΔP(+)PBS. The kinetic trace was fitted using equation (1). The fluorescence intensity at the end of the reaction was estimated

from the corresponding heat-annealed ΔP(−)/ΔP(+)PBS duplex. Calculated k_{obs1} and k_{obs2} values are 8×10^{-4} s^{-1} and 6×10^{-5} s^{-1}, respectively. The reported values are the means +/- standard deviations for two experiments. Excitation and emission wavelengths were 360 and 540 nm, respectively. Buffer was 10 mM phosphate (pH 6.5), 30 mM NaCl at 20°C.

Kinetics of (+)/(−)PBS annealing in the presence of NC derivatives
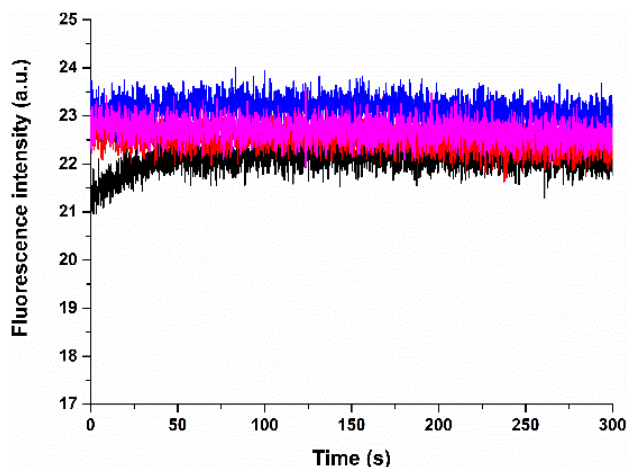
Figure S6. Annealing kinetic traces of thG7(−)PBS with (+)PBS in the presence of NC(11−55). Kinetic traces monitored by steady-state spectroscopy of 0.3 μM thG7(−)PBS with 1 μM (black), 3 μM (red), 5 μM (blue), and 7 μM (magenta) (+)PBS, respectively. NC(11−55) was added at a 3-fold excess over the PBS concentrations.

Table S3. Observed rate constants for the 3HCnt-9(+)/(−)PBS and 3HCnt-9ΔP(+)/(−)PBS annealing promoted by NC(11−55), NC(1−55) and (SSHS)$_2$NC(11−55).[a]

ODN	rate cst (s^{-1})	NC(11−55)		NC(1−55)		(SSHS)$_2$NC(11−55)	
			a		a		a
PBS full length	k_{obs1}	0.39	1	0.63	0.45	1.1	0.64
	k_{obs2}	-		0.082		0.021	
ΔP(−)PBS	k_{obs1}	0.53	1	0.41	0.24	0.31	0.29
	k_{obs2}	-		0.083		9×10^{-5}	

[a] The progress curves in Figure 6 were fitted with a mono exponential function for NC(11−55) as explained in the main text, or with equation (1) for NC(1−55) and (SSHS)$_2$NC(11−55).
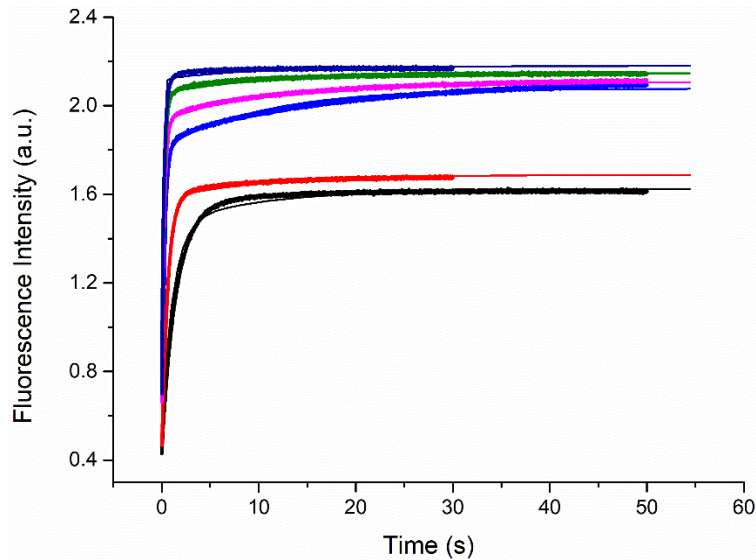
Figure S7. Stopped-flow kinetic traces of the (SSHS)$_2$NC(11–55)-promoted annealing of 3HCnt-9(–)PBS with (+)PBS and their fit using Dynafit. The kinetic traces (thick lines) were recorded with 0.1 μM 3HCnt-9(–)PBS and either 1 μM (black), 2 μM (red), 4 μM (blue), 6 μM (magenta), 8 μM (green) or 10 μM (navy) (+)PBS, respectively. (SSHS)$_2$NC(11–55) was added at a 3:1 peptide:ODN ratio. The fitted curves (thin lines) to the kinetic traces provided the k_1, k_{-1} and k_f values given in Table 1. Buffer was 10 mM phosphate buffer (pH 6.5), 30 mM NaCl at 20°C.

4.3. CHAPTER 3: Investigating the Role of Sugar Deoxyribose in Binding Polarity of HIV-1 Nucleocapsid Protein.

4.3.1. Introduction

The human immunodeficiency virus type 1 nucleocapsid protein (NC) is a small, basic protein with highly conserved zinc binding CX2CX4HX4C (CCHC) motifs called zinc fingers (ZFs) that are flanked by basic residues (Campbell and Rein, 1999; Gorelick et al., 1993; Lapadat-Tapolsky et al., 1995; Levin et al., 2005). Alongside it has a poorly folded highly basic N-terminal region and a basic flexible linker region separating the two ZFs. The ZF motifs contain conservatively substituted hydrophobic and aromatic residues that form a hydrophobic platform on the surface of both the knuckles. Interaction of NC with nucleic acids forms the basis of its crucial role played during the replication cycle, which notably includes selective packaging of unspliced viral genomic RNA and chaperoning of nucleic acid strands during reverse transcription (Amarasinghe et al., 2000; Darlix et al., 2007; Godet and Mély, 2010; Judith G Levin et al., 2010; Levin et al., 2005; Muriaux and Darlix, 2010; Thomas and Gorelick, 2008). The nucleic acid chaperone activity of NC allows it to remodel and rearrange the nucleic acids into their most stable conformations (Beltz et al., 2003). This activity results from the rapid on/off kinetics of NC binding to nucleic acids and its ability to favor nucleic acid aggregation/annealing and duplex destabilization (Cruceanu et al., 2006a, 2006b; Godet and Mély, 2010; Judith G Levin et al., 2010). NC can bind specifically or non-specifically to

any RNA or DNA (Darlix et al., 2011). At low degree of occupancy (1:1000 nt), NC mainly binds with high affinity to specific sequences. At higher occupancy degrees (1:15 nt to 1:7 nt), NC rearranges their conformation to promote annealing or duplex formation (Sarah Bourbigot et al., 2008; Lapadat-Tapolsky et al., 1995). At very high occupancy (1:5 nt to 1:1 nt) NC condenses and aggregates the nucleic acids. Non-specific binding results mainly from the interaction with the basic N-terminal residues, whereas, specific interactions are driven through the hydrophobic ZFs residues (Beltz et al., 2003; Julien Godet et al., 2011; Kanevsky et al., 2011; Lapadat-Tapolsky et al., 1995).

3D structure analysis of NC-Oligonucleotide (NC-ODN) complexes has shown that NC binds preferentially to unpaired guanine residues in loop or bulge regions (Amarasinghe et al., 2000; Darlix et al., 2011; Pappalardo et al., 1998). More specifically, NC was shown to bind with high affinity to GNG (N = A, U, T, or C) sequences of RNA (Amarasinghe et al., 2000; De Guzman, 1998), and TG, GNG and TNG sequences of DNA (Avilov et al., 2012; Sarah Bourbigot et al., 2008; Vuilleumier et al., 1999). In all NC-ODNs structures, guanine binds deeply into the ZF1 and ZF2 hydrophobic plateau. The hydrophobic platform of NC formed of the residues V13, F16, T24, A25, W37, Q45 and M46, undergoes important structural changes upon binding with ODNs sequences. Specific binding is thought to involve a dynamic interaction between the F16 residue of ZF1 and the W37 residue of ZF2 with unpaired guanines in the internal or apical loops or in single-stranded domains distributed among the RNA genome and its DNA copies. As observed for NC complexes with RNA (SL1, SL3) or DNA (PBS, mini-cTAR) sequences, ZF2 (having W37) recognizes an unpaired guanine. In DNA sequences, ZF1 interacts with the stem and destabilize it (as seen with (–)PBS, mini-cTAR) (Beltz et al., 2003; Sarah Bourbigot et al., 2008; Ramalanjaona et al., 2007). In RNAs, ZF1 contributes to specific binding by interacting with the second guanine of the GNG sequence (as seen with SL1, SL3)(De Guzman, 1998; Pappalardo et al., 1998).

An additional important difference that has been emphasized from the high resolution structures of NC-DNA complexes (Bazzi et al., 2011; Sarah Bourbigot et al., 2008; Vuilleumier et al., 1999) and NC-RNA complexes (Amarasinghe et al., 2000; De Guzman, 1998; Spriggs et al., 2008) is the different binding polarity of NC on the NA sequence. In NC-DNA complexes, NC binds parallel to the 5'-3' direction of the NA chain, while in NC-RNA complexes, NC binds in the opposite direction (3'-5'). The sugar moiety (ribose or deoxy ribose) is responsible for the orientation of NC. Comparing the contacts established between NC-DNA and NC-RNA complexes, there exist differences at the level of hydrophobic contacts

made through the sugar bases. Hydrophobic contacts are seen for NC-DNA complexes in which the hydrophobic residues (F16, T24, W37, Q45 and M46) of NC interact with the C1' and C2' carbon atoms. However, the hydrophobic contacts are missing in NC-RNA complexes. Instead, O4', C4' and C5' of G10 and A11 sugars are in close contact with F16. Noteworthy, more residues of NC are involved in DNA than in RNA binding. These characteristic contacts are the basis for having a different binding polarity of NC to RNA or DNA sequences.

As the binding polarity rule of NC to NAs was deduced from a limited number of different sequences, our work was to further investigate this polarity by comparing the binding modes adopted by NC while interacting with the RNA and DNA versions of the minus strand primer binding site (–)PBS and the ψ-Stem loop SL3 sequences. Firstly, we review here the NMR determined structures of (–)PBS and SL3 alone; complexed with NC; and their respective binding affinities.

4.3.1.1.   (–)PBS Binding and Destabilization by Nucleocapsid Protein

The structure of the 18-nucleotide (–)PBS has been determined both free in solution (Johnson et al., 2000) and bound specifically to NC protein (Sarah Bourbigot et al., 2008). The free (–)PBS oligonucleotide in solution exists in a monomeric form, which has also been observed for several other stem-loop structure like SL1, SL2 and SL3 sequences (Amarasinghe et al., 2000; De Guzman, 1998; Pappalardo et al., 1998). The stem residues G1, C3, C4, and C5 form Watson-Crick base pairs with residues C14, G13, G12, and G11, respectively, and exhibit an overall B-helical structure (Johnson et al., 2000). However, the position of T2 residue relative to the rest of the stem is conformationally mobile as it is looped out of the stem and forms a bulge. The nucleotides T6, G7, T8, T9, and C10 in the loop forms a partially ordered penta-loop. The first nucleotide of the penta-loop, T6, stacks on C5 but this is not the case for the last nucleotide as there is no evidence of C10 stacking with T11. The nucleobase G7 along with T8 point towards the stems, whereas T9 points across the loop towards base T6 (Johnson et al., 2000). Interestingly, in spite of their limited number of base-pairs (four), the thermodynamic stabilities of (–)PBS and (+)PBS are comparable to that of cTAR DNA, which contains about 20 bp (Beltz et al., 2003; Egelé et al., 2004; Sholokh et al., 2015). This structural stability of (–)PBS internal hairpin likely accounts for its inefficient annealing in the absence of NCs (Guo et al., 2000; Wu et al., 1999).

The structures of free Δ(–)PBS and complexed with NC(12–55) was solved by *Bourbigot et al.* (Sarah Bourbigot et al., 2008). The structure of complex was solved by

replacing NCp7 with the truncated NC(12–55) that has been shown to have similar chaperone activity (Beltz et al., 2003; Godet et al., 2006). Instead of full of length (–)PBS, a truncated Δ(–)PBS sequence (deleted of the single strand overhang) was used (Sarah Bourbigot et al., 2008) to reduce the number of bound NC(12–55) molecules (Egelé et al., 2004). The NMR data showed that the V13, F16, T24, A25, W37, Q45, and M46 residues of NC(12–55) are involved in the binding to ΔP(−)PBS. The V13, F16, T24, and A25 residues of ZF1 were found to interact with the C5 and T6 bases, and the W37, Q45, and M46 residues of ZF2 interact with the G7 base (Sarah Bourbigot et al., 2008). These interactions elongate the phosphodiester backbone and stretch the entire loop, thus pulling T6 and G7 out of the loop (Sarah Bourbigot et al., 2008).

The nucleobases 5'-$C_5T_6G_7$-3' of (–)PBS act as a preferential binding site for the NC protein. ZF2 interacts with the T6 and G7 residues and forces the N-terminal ZF1 to interact with the C5 residue that is paired with G11 on the stem. This interaction alters the C5-G11 base pair but does not melt it. NC initiates the destabilization of the stem-loop structure but the melting of (–)PBS probably needs the presence of the complementary sequence (+)PBS (Sarah Bourbigot et al., 2008). The second binding site for NC on (–)PBS was identified at the 5'-$C_{10}G_{11}G_{12}$-3' segment, but its structure was not solved. In our buffer conditions, 25 mM Tris–HCl (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$, NC(12–55) binds the high affinity 5'-$C_5T_6G_7$-3' and low affinity 5'-$C_{10}G_{11}G_{12}$-3' binding sites with binding constants of 0.083 μM and 0.90 μM, respectively (Sarah Bourbigot et al., 2008; Egelé et al., 2004).

The role of the deoxyribose sugar of (–)PBS is to establish hydrophobic contacts from its carbon C1' and C2' atoms of sugar rings with the hydrophobic residues (F16, T24, W37, Q45 and M46) of NC. Interestingly, in comparison to the carbon C1' and C2' atoms, the C3' and C4' atoms of the deoxyribose sugar rings have lesser hydrophobics contacts. Contacts are observed between (i) F16 and C2' of T6, (ii) M46 and C2' of G7, and (iii) W37 and O4' and C1' of G7 (Bazzi et al., 2011; Sarah Bourbigot et al., 2008; Morellet et al., 1998). This network of interactions indicate that the 'side' of the deoxyribose sugar bearing C1' and C2' atoms is positioned in close proximity with the protein residues while the other 'side' with C4' and C5' atoms is directed towards the outside of the complex (Bazzi et al., 2011). A similar pattern of hydrophobic contacts has also been identified with the C1' and C2' atoms of the deoxyribose sugar of mini-cTAR and NC hydrophobic residues (Bazzi et al., 2011; Tolstorukov et al., 2004; Zargarian et al., 2009).

4.3.1.2.　SL3 Recognition by Nucleocapsid Protein

Structural studies have reported the folding of SL3 hairpin alone (Pappalardo et al., 1998) and complexed with NC protein (De Guzman, 1998). The 3D structure of SL3 shows that the stem has a A-type structure and the tetra-loop 5'-$G_9G_{10}A_{11}G_{12}$-3' is rather flexible (Pappalardo et al., 1998). On superimposition of all NMR structures, the base-pairs of the stem stack nicely, but there are variations occurring at the ends. The flexible nature of the tetra-loop allows the nucleobase G10 and G12 to be in unstacked orientation. Importantly, the G12 residue is exposed out of the loop and is entirely solvent exposed. The nucleobase G9 and G11 are tightly constrained even without having strong stacking with their neighboring residues and extend into the minor groove (Pappalardo et al., 1998). The first base-pair of the loop C8-G13 is also tightly constrained (Pappalardo et al., 1998). Due to its flexible nature, the tetraloop can bind NC with minimal energy requirements.

The 3D structure of NC(1–55) bound to SL3 RNA shows an interaction of NC hydrophobic residues to the G10 and G12 residues of the flexible SL3 tetraloop. The G12 residue interacts specifically with the V13, F16, I24, and A25 residues of ZF1, whereas the G10 residue interacts with the hydrophobic cleft formed by the W37, Q45, and M46 side chains of the ZF2 (De Guzman, 1998). The A11 residue is exposed to the solvent establishing contacts with the protein surface. In the stem part, A8 residue makes hydrophobic contacts with A25, F16, and N17 groups of the ZF1 knuckle and forms a hydrogen bond with the side-chain of R32 (De Guzman, 1998). Thus, in contrast to NC-(–)PBS, where ZF1 interacts with the stem to destabilize it, both ZFs in the NC-SL3 complex bind to the exposed guanines of the loop and thus do not interact strongly with the stem. In addition, the NC-SL3 complex is stabilized by electrostatic interactions of the highly conserved K3, R7, R10 and K11 residues with the phosphodiester backbone of the RNA stem. Another conserved N5 residue of the NC protein is the only residue that forms a specific hydrogen bonding contact with the 2'Hydroxy of the G10 residue of RNA.

NC binds to SL3 RNA in a 1:1 complex. Its binding affinity has been determined using ITC (Amarasinghe et al., 2000; Tanwar et al., 2017), gel electrophoresis (Shubsda et al., 2000), surface plasmon resonance (Druillennec et al., 1999; Fisher et al., 1998) and fluorescence (Athavale et al., 2010; Paoletti et al., 2002; Shubsda et al., 2002). The binding affinity is highly dependent on the nature and concentration of the salt (Athavale et al., 2010). Most studies were performed at 150 mM to 200 mM, where problems of aggregation and increase in binding stoichiometry are limited. Similarly, regarding the nature of salt, using Mg^{2+} ion has been

shown to stabilizes the backbone of the helix thus disfavoring the protein binding site. Considering these parameters, NCp7 binds to SL3 with a binding affinity of 28–48 nM (Athavale et al., 2010; Tanwar et al., 2017).

The binding parameters were also determined for the DNA variant of SL3 (Vuilleumier et al., 1999) through a fluorescence titration monitoring the quenching of the intrinsic fluorescence of W37 residue upon its stacking with a guanine residue of SL3_DNA (at 100 mM NaCl concentration). Binding constants of 1000 nM and 70 nM were observed for NC(12–55) and NC(1-72), respectively, evidencing the role of the basic terminal side chain of NC in binding with SL3 DNA. The strong W37 fluorescence quenching is in line with an interaction with a guanine residue, but it needs to be confirmed that the quenching results from a specific interaction with the G10 and/or G12 residues. Binding affinities were also measured with mutants of SL3 DNA in which G10 and G12 residues were omitted. Significant decrease in affinity was observed for mutants with A_9AAG_{12}, A_9GAA_{12} and A_9AAA_{12} sequences, confirming an interaction with both the G10 and G12 residues of the SL3 DNA similar to SL3 RNA.
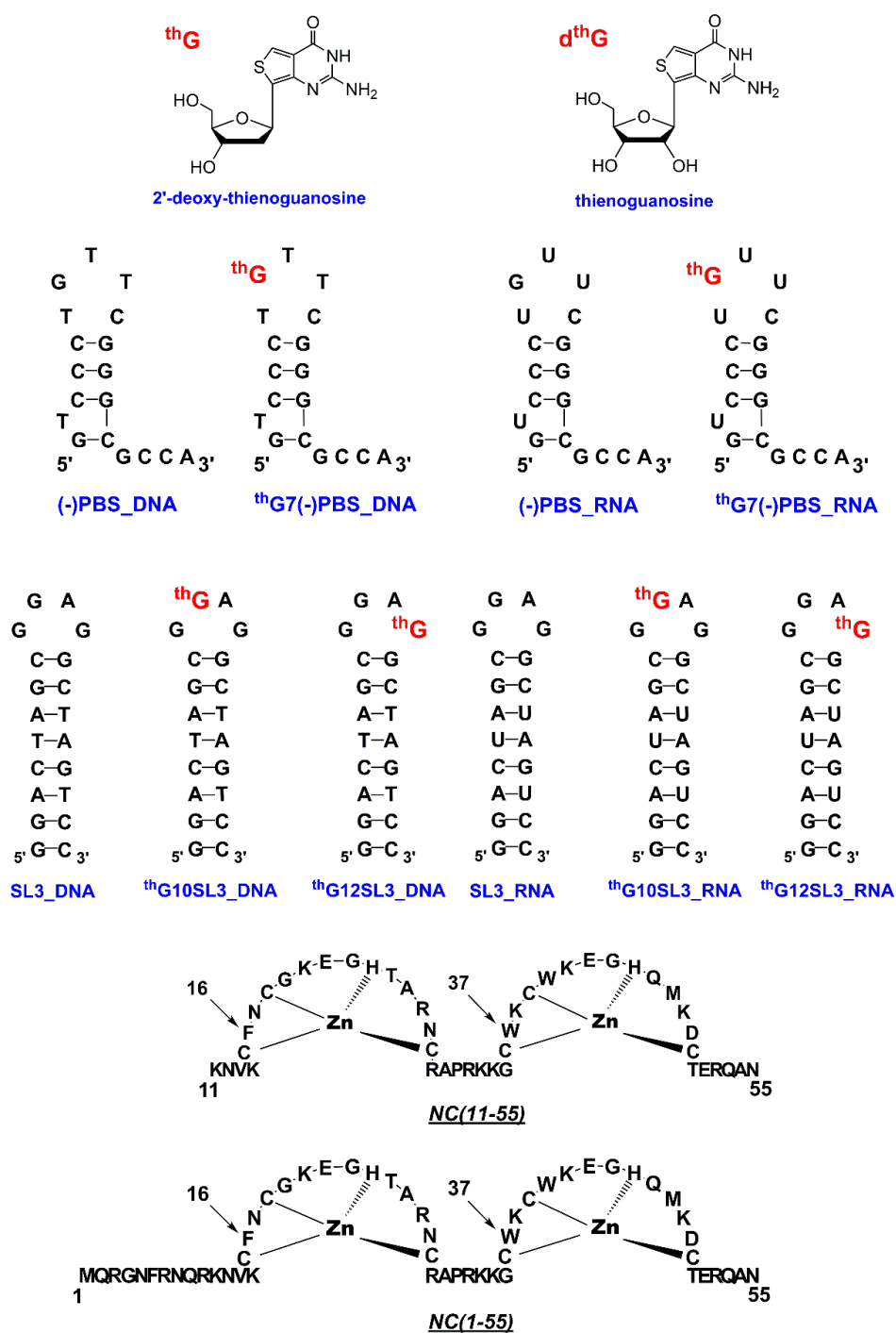
Figure 50. Structures of the oligonucleotides and peptides used in this study.

In this work, we aimed to provide further insight into the role of the sugar on the W37-G recognition in both NC-DNA and NC-RNA complexes. To achieve this aim, we site-specifically incorporated fluorescent nucleobase analogue thienoguanosine (thG) (Deoxyribose or ribose) (Shin et al., 2011) within the DNA and RNA copies of (–)PBS and SL3. Previous work (Sholokh et al., 2015) has shown that thG can faithfully substitute the G7 residue in (–)PBS, providing reliable information on the conformations and dynamics of the G residue in

both the (–)PBS stem loop and (–)/(+)PBS duplex. Consequently, G7 residue is an ideal position in (–)PBS to have thG substituted in both of its DNA [thG7(–)PBS_DNA] and RNA [thG7(–)PBS_RNA] copies for monitoring the interaction of this residue with NC. Similarly, we selected two positions in SL3, position 10 and 12 and substitute them with thG in their respective DNA [thG10SL3_DNA & thG12SL3_DNA] and RNA [thG10SL3_RNA & thG12SL3_RNA] copies. Using ITC, we found that for the non-labeled sequences, NC binds with similar affinity to both the DNA and RNA copies of (–)PBS. A similar trend was also observed for thG-labeled DNA and RNA copies of (–)PBS, suggesting that substitution by thG does not affect the binding process and NC possesses similar binding affinity for DNA and RNA copies of (–)PBS. Interestingly, the photophysical investigation of both the thG-labeled sequences showed more than a two-fold increase in quantum yield associated with a parallel change in lifetimes distribution upon NC interaction showing that the thG7 conformational changes due to its binding with NC appear similar in both DNA and RNA copies of (–)PBS. Investigation of NC binding to non-labeled SL3 sequences again showed similar binding constants to both of its DNA and RNA copies. However, NC-induced fluorescence changes are observed only for thG10SL3_RNA and not for thG12SL3_RNA, suggesting that W37-thG10 interaction led to larger conformational changes than the F16-thG12 interaction. The absence of thG fluorescence change for both positions 10 and 12 in the NC-SL3_DNA complex suggesting that the conformations of thG are already solvent exposed and not sensitive to NC binding. Interestingly, thG substitution in SL3_DNA induces a decrease in the binding affinity at both positions 10 and 12, indicating that both positions are effectively involved in NC binding. Molecular dynamic calculations evidence that substitution of thG in (–)PBS and SL3 is energetically favorable and mimics the conformation of native guanine residues.

4.3.2. Results and Discussions

4.3.2.1. NC-PBS Complex

4.3.2.1.1. Photophysical Characterization of thG-labeled (–)PBS.

Photophysical characterization of free thG nucleoside and as substituted at G7 residue of (–)PBS DNA (thG7(–)PBS_DNA) (Sholokh et al., 2016, 2015) has showed the presence of two ground state keto-amino tautomer identified as, thG-H1 and thG-H3. In free thG nucleoside, both tautomers are in equilibrium and have significantly shifted absorption and emission spectrum. When incorporated in (–)PBS, both tautomers are observed, but in duplex (–)/(+)PBS, the thG-H1 tautomer is favored.

Foremost task is to map the distribution of both thG tautomers in site-specifically labeled DNA and RNA copy of (–)PBS sequences. Both tautomers in thG7(–)PBS_DNA can be evidenced by progressively exciting from 380 nm to 283 nm (Figure 51a). Excitation at higher wavelengths (λ_{ex} = 360 to 380 nm) shows a single band red shifted emission centered at 468 nm. While gradually exciting at wavelengths below 360 nm, a blue shifted emission shoulder centered at 400 nm appears. This shoulder increases progressively and becomes dominant at excitations below 320 nm. On deconvolution (Figure 51c) of thG(–)PBS_DNA emission spectrum (λ_{ex} = 298 nm), the emission maxima (468 and 400 nm) matched with those obtained previously from the deconvoluted emission spectra of thG nucleoside in buffer (Sholokh et al., 2016), which corresponds to the emission from thG-H1 and thG-H3 tautomers, respectively. This is also supported by the excitation spectra (Figure 51e) of thG7(–)PBS_DNA at different emission wavelengths. From the normalized emission and excitation spectra, it becomes clear that both tautomers are excited at wavelengths below 360 nm. Whereas, for excitations above 360 nm only thG-H1 tautomer is selectively excited.
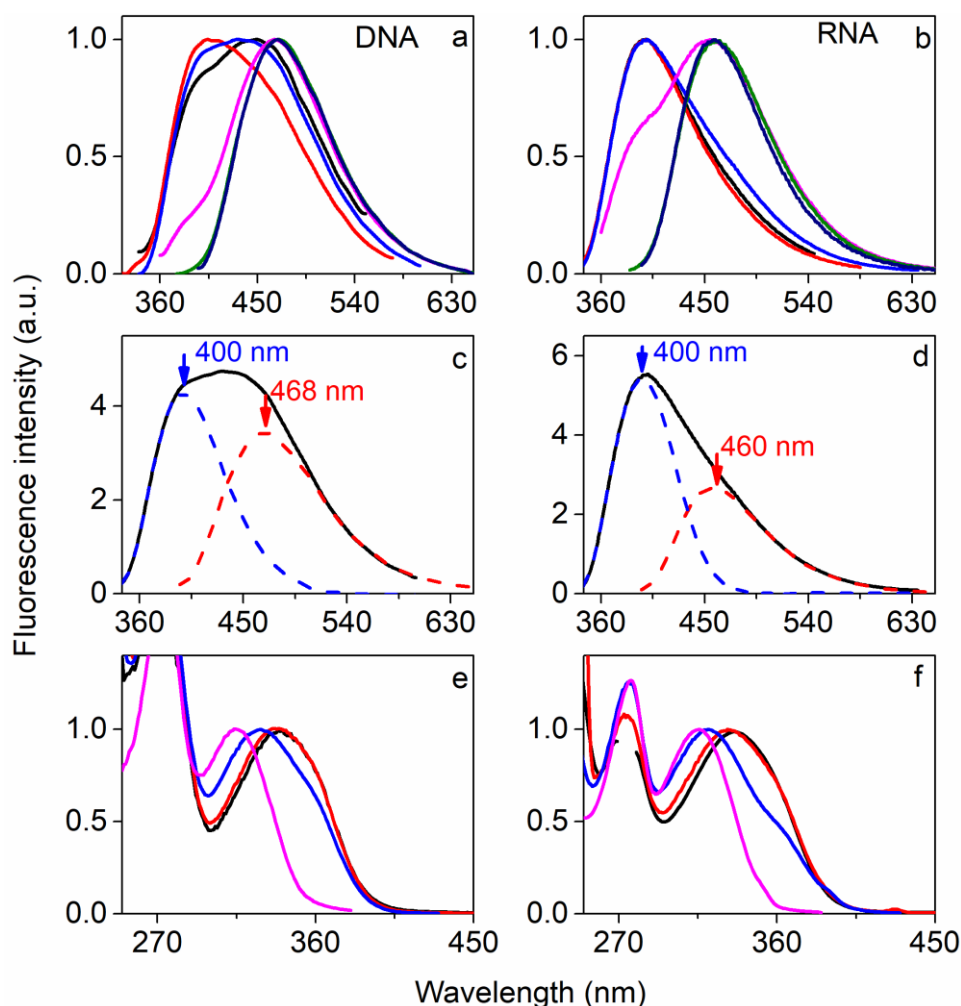
Figure 51. Spectral profile of thG7(–)PBS_DNA (a,c & e) and thG7(–)PBS_RNA (b,d & f). Emission spectra (a, b) at different excitation wavelengths: λ_{ex} = 283 nm (black line); λ_{ex} = 298 nm (red); λ_{ex} = 325 nm (blue); λ_{ex} = 345 nm (magenta); λ_{ex} = 360 nm (green) and λ_{ex} = 380 nm (navy). Emission spectra were normalized at their maxima. Deconvoluted emission spectrum (c, d) obtained at λ_{ex} = 325 nm. Normalized excitation spectra (e, f) at different emission wavelengths: 550 nm (black line), 500 nm (red), 450 nm (blue), 400 nm (magenta).

Next, we investigated the photophysics of thG substituted at G7 residue of (–)PBS_RNA. Similar to its DNA counterpart, thG7(–)PBS_RNA absorbs in the blue region having a similar absorption maxima (324 nm). Exciting at lower energy wavelength (λ_{ex} = 360 to 380 nm) showed a single band emission centered at 466 nm (Figure 51b), likely corresponding to thG-H1 tautomer. And at lower wavelengths (below 360 nm), we observed also contribution from thG-H3 tautomer. Spectral deconvolution (Figure 51d) shows the emission of both tautomers with maxima at 400 and 460 nm, respectively. This conclusion is also supported from the normalized excitation spectra (Figure 51f) obtained at different

emission wavelengths. Therefore, similar to [th]G7(–)PBS_DNA, the [th]G-H1 tautomer of [th]G7(–)PBS_RNA can selectively be excited at wavelengths above 360 nm. For excitation below 360 nm, both tautomers are excited.

Table 3. Time-Resolved Fluorescence Parameters of [th]G labeled ODNs[a].

	QY	τ_1(ns)	α_1	τ_2(ns)	α_2	τ_3(ns)	α_3	$<\tau>$(ns)
[th]G7(–)PBS_DNA	0.10	0.39	0.18	3.0	0.61	13.3	0.21	4.7
+NC	0.18	0.57	0.22	4.2	0.34	17.0	0.44	9.1
[th]G7(–)PBS_RNA	0.06	0.47	0.49	3.0	0.32	12.7	0.19	3.6
+NC	0.15	0.56	0.36	3.9	0.30	17.3	0.34	7.2

[a]QY is the fluorescence quantum yield, τ_i are the fluorescence lifetimes (ns), and α_i are their amplitudes. $<\tau>$ is the mean fluorescence lifetime (ns). Excitation and emission wavelengths were 315 and 500 nm. SDs for the lifetimes and amplitudes are <20%. SDs for QY are <10%.

We then investigated the heterogeneity of [th]G as substituted in DNA and RNA variants of (–)PBS using time-resolved fluorescence spectroscopy (Table 3). The lifetime decays of [th]G7(–)PBS_DNA were fitted with three exponentials and the lifetime components are in good agreement with the earlier reported values (Sholokh et al., 2015). The long-lived lifetime (13 ns) is close to that of the free probe in methanol and is thus, thought to correspond to a [th]G conformation that is minimally quenched in the less polar environment of (–)PBS loop (Sholokh et al., 2015). The two short lifetime components 0.39 ns and 3.0 ns likely represent conformations of [th]G that are dynamically quenched by the neighboring residues. The fluorescence quantum yield of [th]G7(–)PBS is largely reduced as compared to the free probe (Shin et al., 2011), confirming the strong fluorescence quenching from the neighboring residues.

The lifetime decays of [th]G7(–)PBS_RNA are found to be triexponential. Interestingly, the lifetime components are similar to the DNA copy of [th]G7(–)PBS with a long-lived lifetime (12.7 ns) indicating a minimally quenched conformation of [th]G7(–)PBS_RNA in a less polar environment and two short lifetimes of 0.47 ns and 3.0 ns, possibly representing [th]G conformations stacked with neighboring residues. The two times lower fluorescence quantum yield of [th]G7(–)PBS_RNA as compared to [th]G7(–)PBS_DNA can be explained by the larger contribution from the weakly emitting species (τ_1). Altogether, the lifetime distribution of [th]G7

in RNA and DNA copies of (–)PBS strongly suggests comparable conformations of thG7 in respect with its neighboring residues (T6-T8), but with a different distribution.

4.3.2.1.2. Binding of NC(11–55) with (–)PBS

Isothermal titration calorimetry

We use Isothermal titration calorimetry (ITC) to characterize NC(11–55) binding with labeled and non-labeled (–)PBS ODNs. Application of ITC is extensively used (Ball and Maechling, 2009; Crane-Robinson et al., 2009; Freisz et al., 2012) to determine the complete set of binding thermodynamics that includes, dissociation constant (K_d), enthalpy change (ΔH), entropy change (ΔS) and the stoichiometry (n), between two molecules in a single solution based experiment. Typical ITC experiment is based on the sequential addition of a ligand from the syringe to a thermostated cell (containing ODN) that is capable of sensing the heat release or absorbed during the mixing. Hence, it directly measures millimolar to nanomolar affinities between the molecules.

Table 4. ITC analysis of (–)PBS and NC(11–55) binding.

	K_d (μM)	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)	ΔS (cal/mol* K)	n
(–)PBS_DNA	0.38	-3.08	-5.51	-8.60	18.80	1.0
thG7(–)PBS_DNA	0.44	-0.59	-7.92	-8.52	27.05	1.8
(–)PBS_RNA	0.92	-3.68	-4.40	-8.09	15.02	0.9
thG7(–)PBS_RNA	1.12	-1.52	-6.45	-7.98	22.02	1.0

Calculated (K_d) dissociation constant, (ΔH) binding enthalpy, (-TΔS) binding entropy (ΔG), Gibb's Free Energy and (n) stoichiometry from ITC binding curves fitted to an independent model.

We first characterized the binding parameters of NC(11–55) with non-labeled (–)PBS_DNA and (–)PBS_RNA (Figure 52, Table 4). Small aliquots (2.5 µL of 150 µM) of NC(11–55) are injected to the cell containing ODN (7 µM). This order of titration is used since the reverse titration (ODN into protein) led to aggregation of protein-ODN complex in the cell as judged by abnormal signal and baseline drift. Mixing NC(11–55) to (–)PBS DNA and RNA is energetically favorable (ΔG < 0). Both interactions show a favorable entropy variation, suggesting that the interaction is driven by the displacement or release of counter ions and water molecules (Velkov, 2009). We obtained 1:1 binding stoichiometry for both (–)PBS DNA and RNA, in variance with the 1:3 (ODN:peptide) stoichiometry reported previously for (–

)PBS DNA (Sarah Bourbigot et al., 2008; Égelé et al., 2005). This discrepancy is probably due to the limited heat exchange that accompany the formation of the NC-(–)PBS ODNs complexes and that prevent the observation of all three binding sites.

(–)PBS	DNA	RNA
Wild type		
thG7		
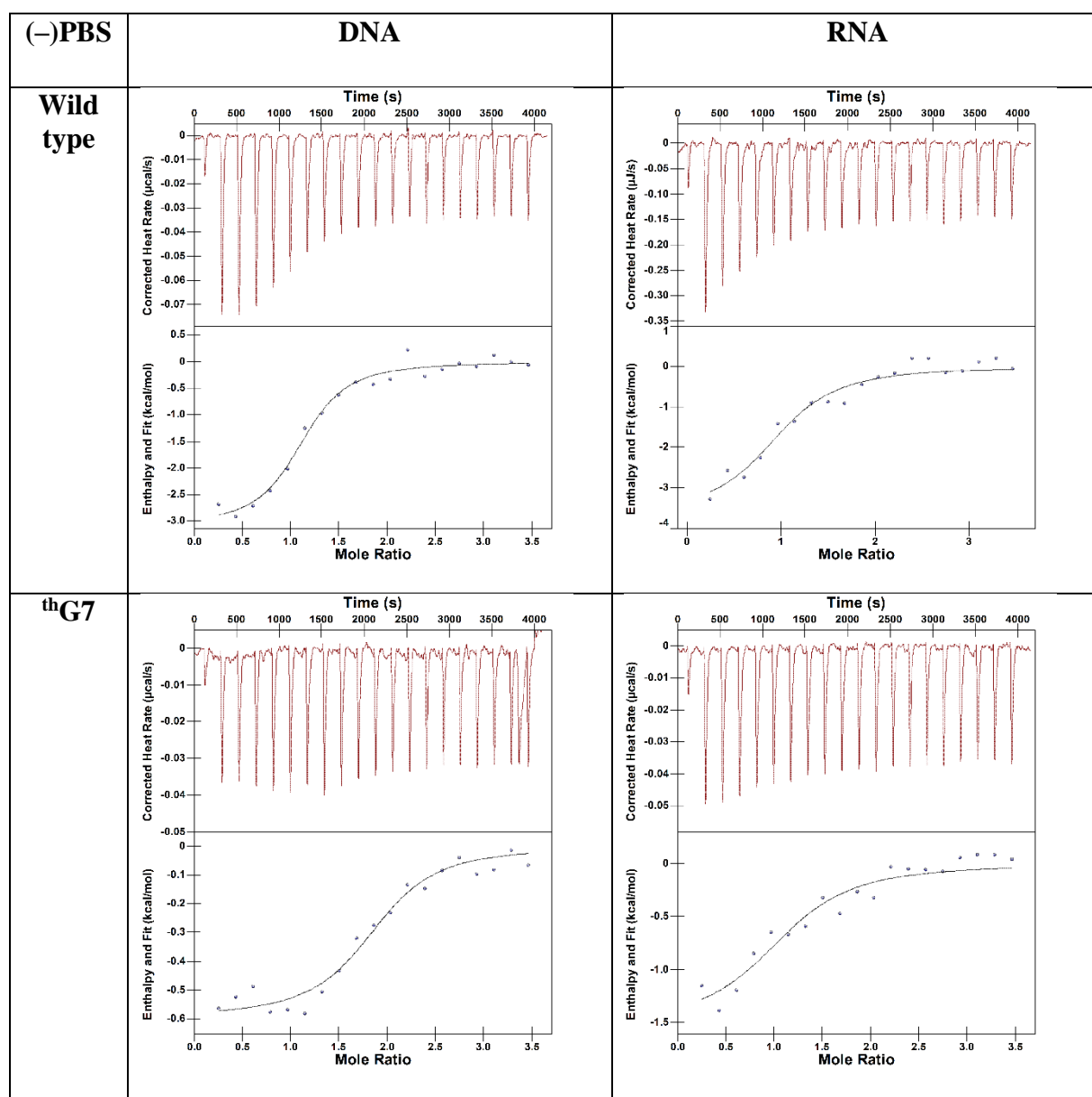
Figure 52. ITC titration of non-labeled and labeled (–)PBS DNA and RNA by NC(11–55). 150 µM of NC(11–55) in the syringe was added to 7 µM ODNs in the cell. The continuous line represents the best fit of the binding curves. Buffer: 50 mM Hepes, 30 mM NaCl, 0.2 mM $MgCl_2$ (pH 7.5).

A model of independent binding sites was used to fit the heat release signal. Interestingly, the binding affinities calculated for (–)PBS_DNA (K_d = 0.38 µM) and (–)PBS_RNA (K_d = 0.98 µM) upon interaction with NC(11–55) did not differ significantly. The

binding affinity between (–)PBS_DNA and NC(11–55) is in good agreement with the earlier reported values for NC(12–55) binding with Δ(–)PBS_DNA (K_d = 0.4 µM) (Sarah Bourbigot et al., 2008) and (–)PBS_DNA (K_d = 2 µM) (Egelé et al., 2004), thus giving confidence in our approach. Altogether, ITC data suggest that binding affinity of NC(11–55) to (–)PBS is poorly dependent on the sugar nature.

After benchmarking the binding affinities with wild type (–)PBS DNA and RNA, we performed ITC measurements with thG-labeled ODN to evaluate if substituting the G7 residue of (–)PBS by thG affects the overall binding affinity to NC(11–55). The interaction of thG7(–)PBS_DNA and (–)PBS_RNA with NC(11–55) showed a significantly lower heat release in comparison to their non-labeled counterparts, but compensated by an increased entropy contribution. The decrease in heat release suggests that the key contact established by G7 is not reproduced by thG. From the close similarity of the structures of G and thG, it may have speculated that the key contact (likely a H bond) could be established by the N7 atom that is missing in thG. Due to poor heat release, the binding affinity with thG7(–)PBS_DNA was difficult to determine. The obtained binding constant for thG7(–)PBS_RNA (K_d = 1.1 µM) is comparable to non-labeled NC-(–)PBS_RNA complex. In all, through ITC measurements with labeled (–)PBS, we evidenced that substitution of G7 residue with thG does only moderately perturb the NC-(–)PBS complex formation.

Fluorescence titration

Now though we know that NC binds to both DNA and RNA copy of (–)PBS, it would be interesting to further investigate and compare the extent with which NC induces conformational and dynamic changes in the loop of both the (–)PBS sequences. These conformational changes in (–)PBS loop upon NC interaction were followed through the fluorescence of site-specifically incorporated thG at position 7 of (–)PBS. We titrated thG7(–)PBS_DNA and thG7(–)PBS_RNA with increasing concentrations of NC(11–55) (Figure 53) and observed a slight blue shift (3–4 nm) in the emission maxima. At saturating NC(11–55) concentration, we observed a 1.8 and 3.2 fold increase in fluorescence of thG7(–)PBS_DNA and thG7(–)PBS_RNA, respectively. The data was fitted with Scatchard equation to determine an apparent binding constant K_d that was found to be similar for both ODNs, 4.6 ± 0.7 µM (n=1), (DNA) and 5.6 ± 0.6 µM (n=1) (RNA). These binding constants differ by a factor 3 to 10 from the affinities calculated through ITC experiments, suggesting that the binding site associated with heat release differs from the binding site(s) giving a change in thG fluorescence.
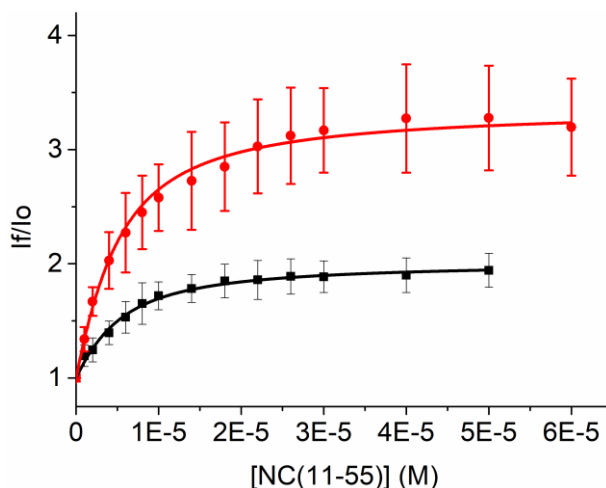
Figure 53. Binding curves of NC(11–55) with thG7(–)PBS_DNA (black) and thG7(–)PBS_RNA (red). Oligonucleotide concentration was 1 µM in 50 mM Hepes, NaCl 30 mM, MgCl$_2$ 0.2 mM, pH 7.5. Solid lines correspond to the fit of the experimental points with equation (10).

For both thG7(–)PBS_DNA and thG7(–)PBS_RNA, the time-resolved intensity decays revealed that the increase in quantum yield of both thG labeled ODNs is mainly due to the significant increase of the τ_3 lifetime (from 13 ns to 17 ns) and its associated population (Table 3). The increase in lifetime from 13 ns (a value close to that in methanol) to 17 ns (a value close to that in water) suggests that NC exposes thG to the solvent. For thG7(–)PBS_DNA, the available NMR structure of the complex with NC (Sarah Bourbigot et al., 2008) evidences the ability of NC to stretch the (–)PBS loop and expose the T6 and G7 residues towards the exterior of the loop. As for thG7(–)PBS_RNA, we observed similar changes in the fluorescence lifetimes, this indicates that NC also exposes the thG7 to the solvent. Interestingly, the lifetimes and the associated amplitudes are very similar for both ODNs in complex with NC, indicating a similar mechanism of NC on both ODNs, with an exposure of the G7 residue to the solvent.

Molecular modelling

Next, we use the molecular dynamics (MD) approach to investigate the possible structural consequence of the introduction of thG in position 7 of (–)PBS_DNA. Molecular modelling is an added tool to provide greater insight into the structural determinants involved in the interactions between NC and nucleic acids. In earlier studies (Mori et al., 2010a), unrestrained molecular dynamics on the NMR structure of (–)PBS DNA and NC complex has been useful in determining the network of hydrophobic and hydrogen bond interactions that

stabilize the complex. As the NMR structure of NC complex with (–)PBS_RNA is not yet resolved, our molecular modeling results will be confined to (–)PBS_DNA.

Both tautomers of the thG probe, namely thG-H1 and thG-H3, were inserted in position 7 of the (–)PBS_DNA to replace the G nucleotide that is stacked over the key W37 in multiple experimental structures. Starting from the MD-refined NC-(–)PBS_DNA structure generated in a prior work (Mori et al., 2010a), 500 ns of unrestrained MD simulations were performed using the protocols described in materials and methods, upon replacing G7 with thG isomers.
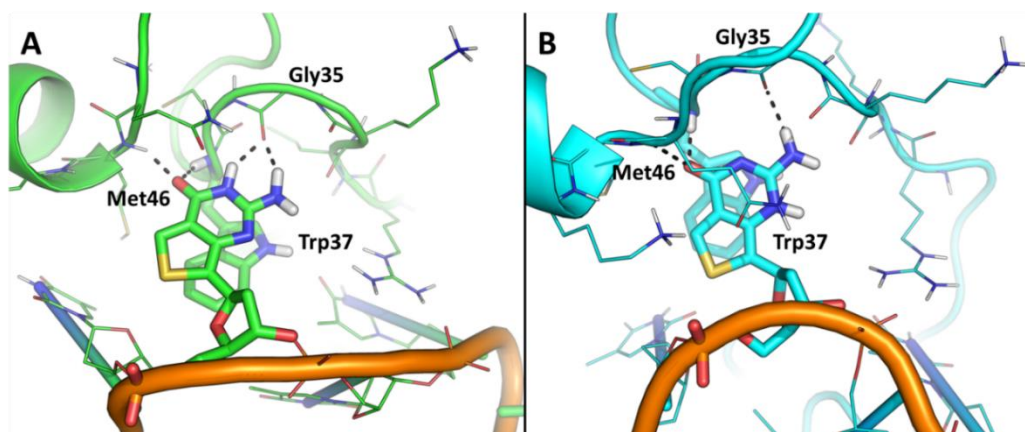
Figure 54. thG does not impair the binding of PBS to the hydrophobic platform of NC. Representative snapshots from MD trajectories showing A) the binding of thG-H1 within the hydrophobic platform of NC, stacked over the W37 residue and H-bonded to M46, W37, and G35 (two H-bonds), and B) the binding of thG-H3 within the hydrophobic platform, stacked over the W37 residue and H-bonded to M46, W37, and G35 (a single H-bond). thG-H1 is colored green and thG-H3 is colored cyan. The probe and the W37 residues are shown as sticks. H-bonds are shown as black dashed lines.
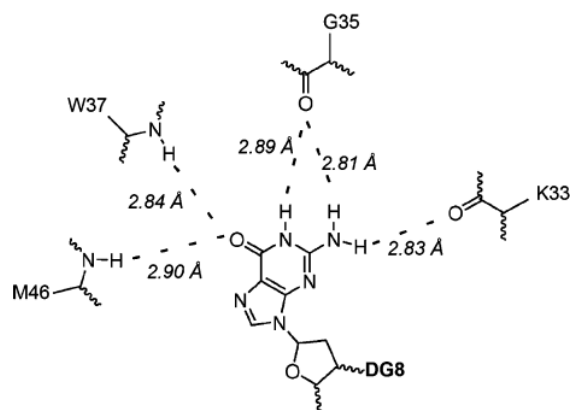
Figure 55. Schematic representation of the binding site for the central guanosine G7 of the (–)PBS_DNA. Hydrogen bonds are displayed as dashed blue lines (Mori et al., 2010b)

Results clearly show that, despite the expected global structural differences due to the large flexibility of the NC protein, [th]G-H1 and -H3 do not affect the overall binding of the (–)PBS_DNA to NC, at least from a conformational standpoint (Figure 54A and Figure 54B, respectively). While [th]G-H1 behaves as an efficient guanine-mimetic, establishing the same bonds with NC as the native G (Figure 55), the [th]G-H3 tautomer also fits into the hydrophobic platform of NC and establishes H-bond interactions with key residues (Figure 54B). However, [th]G-H3 is unable to establish the H-bond with the backbone of G35, which is persistently observed with G or [th]G-H1. From a conformational standpoint, the RMSD plots (Figure 56) show that both systems achieved geometric convergence during MD time, but this convergence was slower with [th]G-H3 and is accompanied by larger conformational fluctuations.
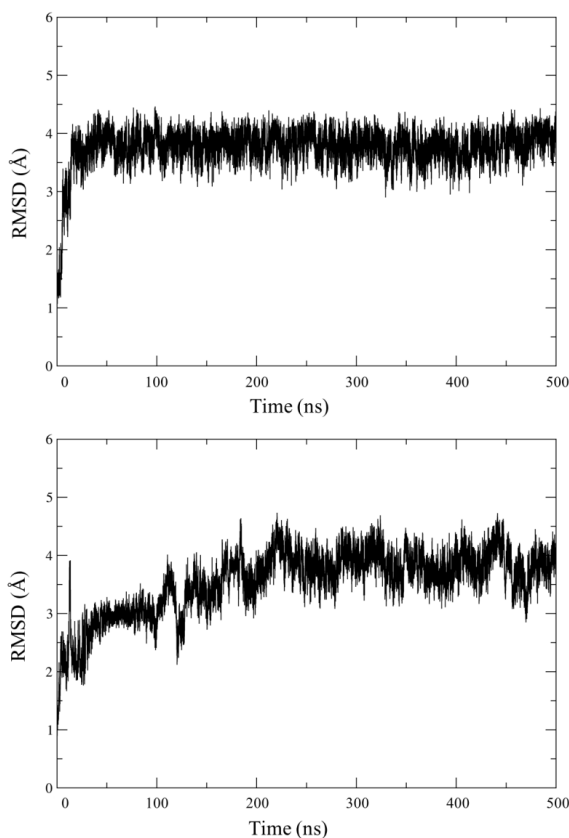
Figure 56. RMSD deviation of the NC-(–)PBS complex bearing thG-H1 (top panel) or thG-H3 (bottom panel) during MD simulation time.

Finally, the thermodynamic effect of thG insertion in position 7 of (–)PBS complexed to the NC was estimated along MD trajectories by means of the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) approach. Results of Table 5 suggest that (–)PBS bearing thG-H1 binds tighter to NC as compared to (–)PBS bearing thG-H3. In summary, thG-H1 is the preferred tautomer for binding to the hydrophobic pocket of the NC, when replaced to G7 in the (–)PBS structure as it is accompanied by a lower free energy of binding (Table 5).

Table 5. Theoretical affinity of (–)PBS for the NC estimated in studied systems

System	ΔG binding MMPBSA (kcal/mol) ± SEM	ΔG binding MMGBSA (kcal/mol) ± SEM
NC-thG7-H1(–)PBS	-45.67 ± 1.73	-87.29 ± 2.1
NC-thG7-H3(–)PBS	-30.43 ± 1.52	-74.68 ± 1.64

Large contribution of thG-H3 in thG7(–)PBS_DNA; difference in binding contacts; and, low theoretical binding affinity with NC compared to thG-H1 and native G, explains the deviation in experimental binding affinity of thG labeled PBS_DNA from native PBS_DNA.

4.3.2.2. NC-SL3 complex

4.3.2.2.1. Photophysical characterization of thG labeled SL3

Photophysical characterization was performed to reveal the distribution of thG tautomers as substituted at position 10 and 12 of SL3 DNA and RNA. All the four ODNs show an absorption maximum around 324 nm. The emission profile (Figure 57) of each thG labeled ODN was obtained by exciting at different wavelengths, from 380 nm to 298 nm. For all labeled SL3 ODNs, except for thG12SL3_DNA, a single emission peak centered (Figure 57a,c,d) at 464 nm is observed, indicating that the emission is dominated by the thG-H1 tautomer. When excited at lower wavelength (< 325 nm) a minor hump (<10 %) in the emission spectrum is observed, indicating the presence of a small amount of the thG-H3 tautomer. This is confirmed by recording the excitation spectra (Figure 58a,c,d), which showed a blue shift in excitation maximum, when recorded at 400 nm emission wavelength. In contrast, the emission (Figure 57b) and excitation (Figure 58b) spectra of thG12SL3_DNA clearly show that both thG-H1 and thG-H3 tautomers are significantly populated in this case.
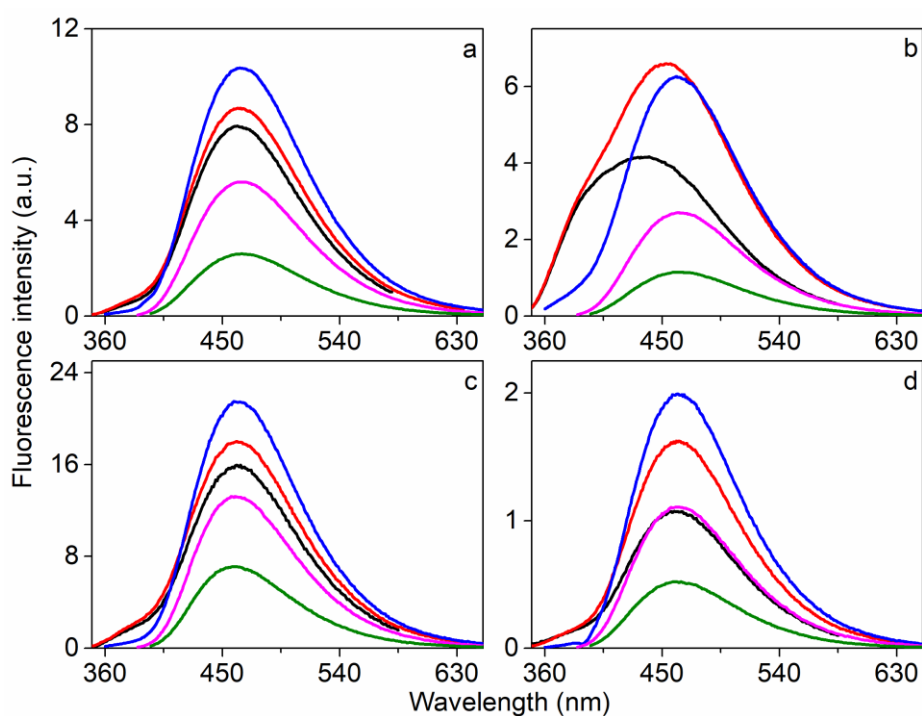
Figure 57. Emission spectra of thG10SL3_DNA & thG12SL3_DNA (a, b) and thG10SL3_RNA & thG12SL3_RNA (c, d). Emission spectra were recorded at different excitation

wavelengths: λ_{ex} = 298 nm (black line); λ_{ex} = 325 nm (red); λ_{ex} = 345 nm (blue); λ_{ex} = 370 nm (magenta); λ_{ex} = 380nm (green).
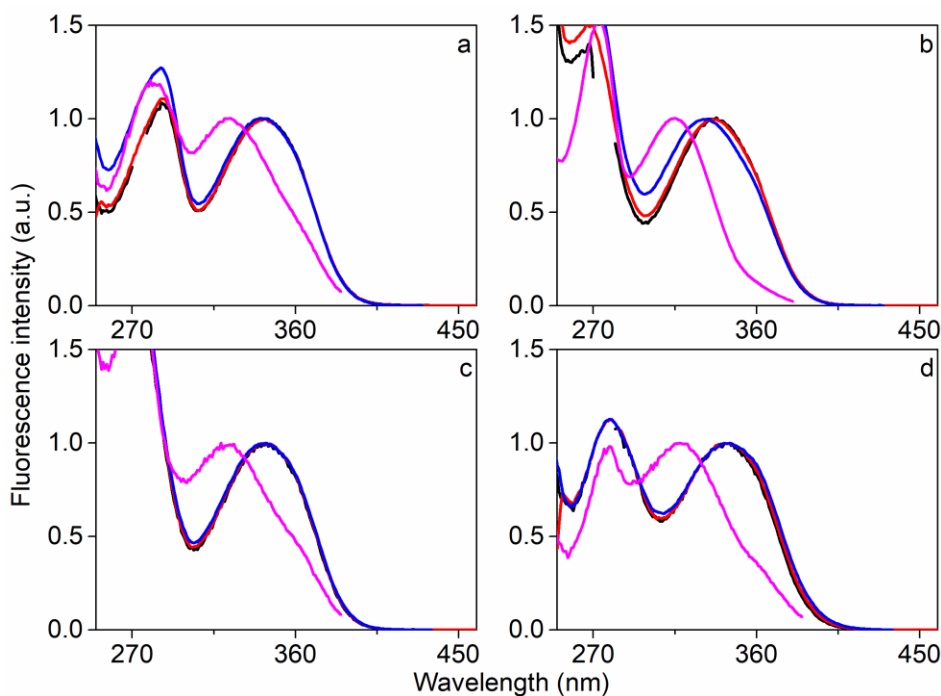
Figure 58. Excitation spectra of thG10SL3_DNA & thG12SL3_DNA (a, b) and thG10SL3_RNA & thG12SL3_RNA (c, d). Excitation spectra recorded at different emission wavelengths: λ_{em} = 550 nm (black line); λ_{em} = 500 nm (red); λ_{em} = 450 nm (blue); λ_{em} = 400 nm (magenta).

The thG-labeled SL3 RNAs show high fluorescence quantum yields (0.23 and 0.33, Table 6). The large quantum yields of thG at both positions 10 and 12 of SL3 RNA indicate that thG is only weakly quenched by its neighboring bases. The time-resolved intensity decays of thG-labeled SL3_RNAs show multiexponential decays that are best fitted with three discrete lifetime components (Table 6). In both labeled ODNs the lifetime components vary from 0.3 ns to 25 ns and the majority of the population (> 70 %) is distributed along the long-lived lifetime. The amplitude of the quenched species in thG10SL3_RNA (0.4 and 1.8 ns) is 22 % while for thG12SL3_RNA (0.33 ns) the percentage reduces to 8 %, evidencing that larger populations of thG experience collisional quenching at position 10. The long-lived lifetime component of thG10SL3_RNA (16.3 ns) is in between the lifetime of thG nucleoside in water (19.9 ns) and in methanol (13 ns) (Shin et al., 2011), while the 25 ns component seen for thG12SL3_RNA, is even higher than the lifetime of thG in water. Such a high lifetime value was already reported for a thG residue in a mismatched duplex (28 ns) and may be explained by an increase in the value of the radiative rate constant as compared to the free nucleoside in

water. Our lifetime data are consistent with the NMR structure (Pappalardo et al., 1998) of SL3_RNA, showing that [th]G at position 12 is largely solvent exposed while [th]G at position 10 may interact with the neighboring bases in a less polar environment.

Table 6. Photophysical parameters of [th]G labeled at position 10 of SL3 DNA and RNA.

	QY	τ_1(ns)	α_1	τ_2(ns)	α_2	τ_3(ns)	α_3	$<\tau>$(ns)
[th]G10SL3_DNA	0.27	-	-	0.8	0.31	22.6	0.69	15.8
[th]G12SL3_DNA	0.26	-	-	3.6	0.17	16.4	0.83	14.2
[th]G10SL3_RNA	0.23	0.3	0.04	1.8	0.18	16.3	0.78	13.0
+NC	0.31	0.7	0.19	5.3	0.18	22.6	0.63	15.2
[th]G12SL3_RNA	0.33	0.4	0.08	9.2	0.21	24.9	0.71	19.6

[a]QY is the fluorescence quantum yield, τ_i are the fluorescence lifetimes (ns), and α_i are their amplitudes. $<\tau>$ is the mean fluorescence lifetime (ns). Excitation and emission wavelengths were 315 and 500 nm. SDs for the lifetimes and amplitudes are <20%. SDs for QY are <10%.

The picture is somewhat different for the [th]G-labeled SL3_DNAs. At both positions, we obtained high and similar quantum yields (~ 0.26). Interestingly, only two components were needed to fit the intensity decay of the [th]G-labeled SL3_DNAs. The values of the long-lived lifetime were inverted as compared to SL3_RNA, because it was 16.4 ns at position 12 and 22.6 ns at position 10. This suggests that the folding of the SL3 loop may be different in the DNA as compared to the RNA, leading to different environments at position 10 and 12. We observed a larger contribution of fluorescently quenched species for [th]G10SL3_DNA than for [th]G12SL3_DNA, suggesting that [th]G experiences more collisional quenching with its neighboring residues at position 10 as compared to position 12.

4.3.2.2.1. Binding of NC(1–55) with SL3

Next, we investigated the interaction of NC and SL3 ODNs and for that we shifted from NC(11–55) to NC(1–55) and increased salt to 150 mM. We used wild-type NC, because the N-terminal chain of NC has shown direct effect on binding parameters (Didier et al., 2009; Vuilleumier et al., 1999). Moreover, the majority of previous studies (Amarasinghe et al., 2000; De Guzman, 1998; Paoletti et al., 2002; Pappalardo et al., 1998; Tanwar et al., 2017) have been carried out with full length NC at high salt concentration (> 100 mM) to avoid non-specific interaction, that artificially increases the binding stoichiometry (Athavale et al., 2010; Shubsda et al., 2000).

Isothermal calorimetry

We performed ITC measurements for the binding of native and labeled SL3 ODNs to NC(1–55) in order to determine the binding parameters and the effect of ^{th}G substitution on the NC-SL3 interaction (Figure 59). Binding of NC(1–55) to both the native DNA and RNA variants of SL3 is exothermic ($\Delta H < 0$) (Table 7). Comparative to the interaction of unmodified NC-PBS (DNA or RNA) ($\Delta H \sim 3.1$ kcal/mol), the interaction between NC-SL3 ($\Delta H \sim 1.7$ kcal/mol) is characterized with lower heat release. However, the enthalpic penalties (-1.75 & -1.00 kcal/mol) are subsided with higher entropic variations (-8.05 & -8.87 kcal/mol) suggesting that the interaction is driven by release of ordered counter ions and water molecules. A 1:1 binding stoichiometry was observed for both SL3 RNA and DNA, in agreement with the previously reported 1:1 stoichiometry of NC with SL3_RNA (Vuilleumier et al., 1999) and SL3_DNA (De Guzman, 1998). The calculated binding affinities of NC(1–55) for SL3_DNA (48 nM) and SL3_RNA (43 nM) did not differ significantly and are in good agreement with the previously reported values (Athavale et al., 2010; Tanwar et al., 2017).
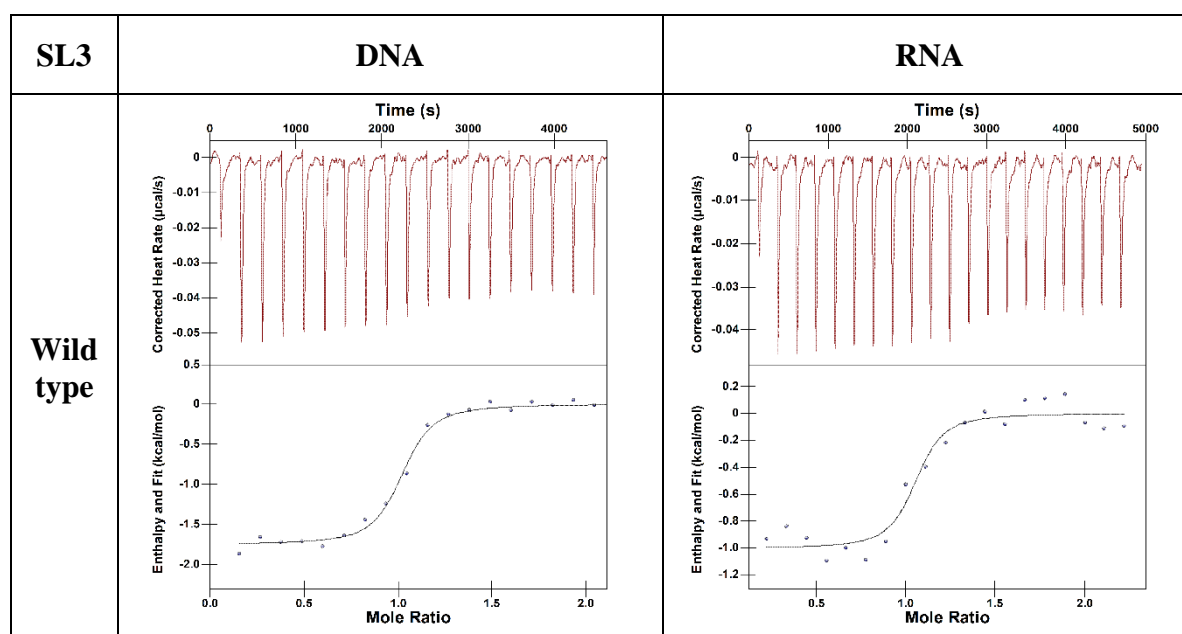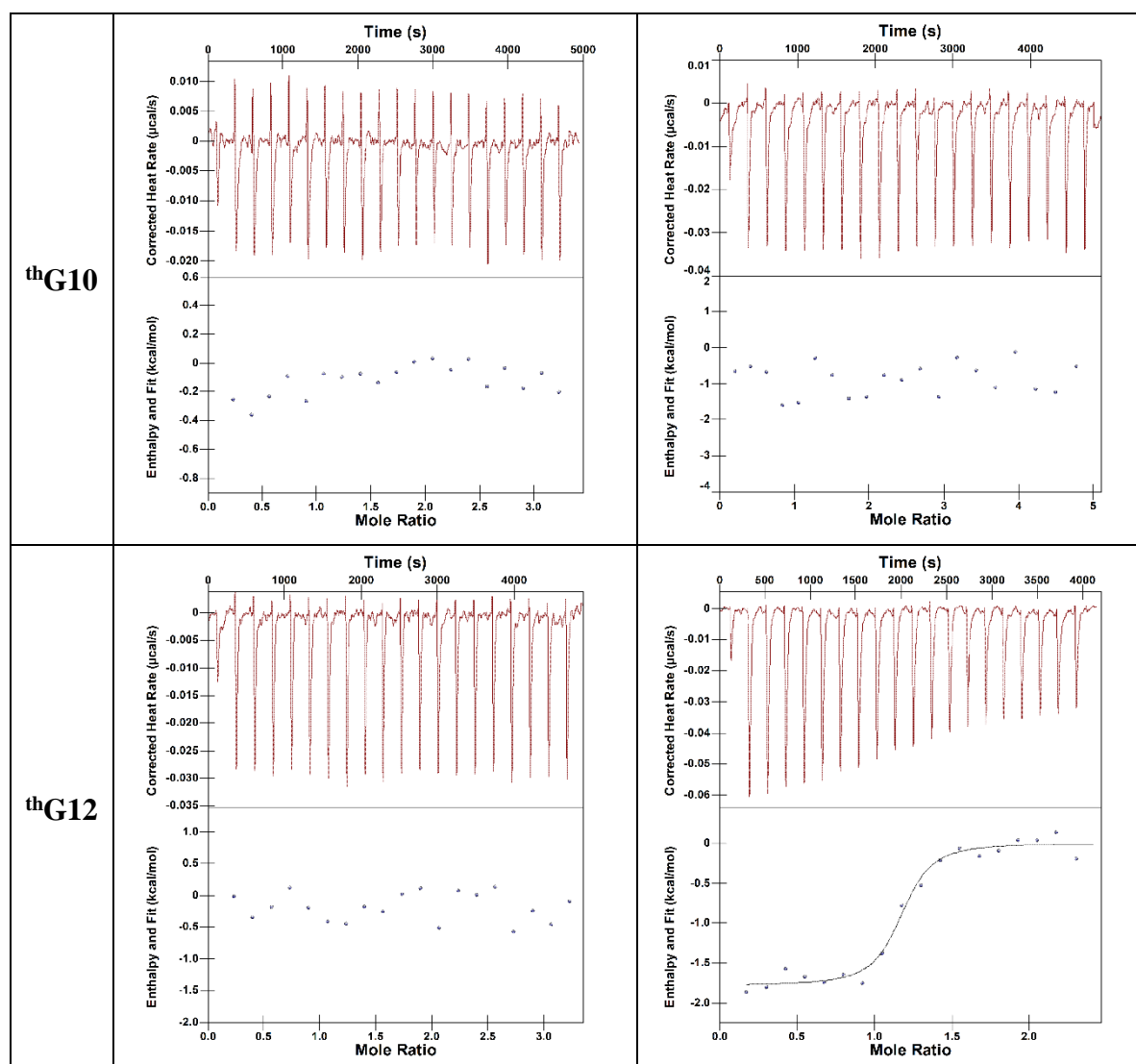
Figure 59. Binding parameters of NC(1–55) with non-labeled and labeled SL3 DNA and RNA. 100-150 µM of NC(11–55) in the syringe was added to 6-8 µM of ODNS in the cell. The continuous line represents the best fit to the binding curves. Buffer: 50 mM Hepes, 150 mM NaCl, 1 mM MgCl$_2$ (pH 7.5).

Next, we performed the ITC experiments for interaction of thG labeled SL3 ODNs and NC(1–55). Unfortunately, mixing of NC(1–55) into the thG labeled SL3 ODNs only led to a weak exothermic signal ($\Delta H < 1$ kcal/mol), that could not be fitted using a single binding model. However, we manage to extract some information with thG12SL3_RNA that shows binding parameters comparable to the native SL3_RNA. The loss of heat signal with thG substituted SL3 ODNs is similar to that seen for thG substituted (–)PBS ODNs. The calculated binding affinities for SL3 DNA (K_d = 48.4 nM), SL3_RNA (K_d = 42.7 nM) and thG12SL3_RNA (K_d = 66 nM) did not differ significantly. In addition, we observed a 1:1 binding stoichiometry between thG12SL3_RNA and NC(1–55), which is similar to the native

counterpart. Altogether, ITC measurements evidenced the binding of NC to SL3 is independent of the sugar moiety. Moreover, the native NC-SL3 interaction is rightly mimicked by the [th]G-labeled SL3 sequences. At this point, it cannot be confirmed if the binding polarity of NC is opposite for DNA and RNA. Further information on this point can be obtained by using the fluorescence of [th]G to follow the conformation and dynamic changes on NC interaction.

Table 7. ITC titration of SL3 to NC(11–55).

	K_d (nM)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	ΔG (kcal/mol)	ΔS (cal/mol*K)	n
SL3_DNA	48.4	-1.75	-8.05	-9.81	27.4	1.0
[th]G10SL3_DNA			N.D.			
[th]G12SL3_DNA			N.D.			
SL3_RNA	42.7	-1.00	-8.87	-9.88	30.29	1.0
[th]G10SL3_RNA			N.D.			
[th]G12SL3_RNA	66.1	-1.78	-7.84	-9.63	26.78	1.1

N.D. means not determined due to low signal.

Fluorescence Titration

Binding affinities and NC-induced conformational changes in the loop of [th]G-labeled SL3 ODNs were then investigated through fluorescence titration and fluorescence lifetime measurements. We titrated all the labeled SL3 RNA and DNA ODNs with NC(1–55) and monitored the fluorescence changes. Except for [th]G10SL3_RNA, we did not observe any change in fluorescence for [th]G-labeled SL3 ODNs upon NC(1–55) addition. At saturating concentrations of NC(1–55), we recorded a 1.3-fold increase in fluorescence of [th]G10SL3_RNA (Figure 60). The titration data was fitted with the help of Scatchard equation (Equation 10) that yielded a 1:1 binding stoichiometry. The calculated binding affinity (310 ± 20 nM) differs by a factor of six from the affinities obtained through ITC measurements, probably as the result of the replacement of G7 by [th]G.
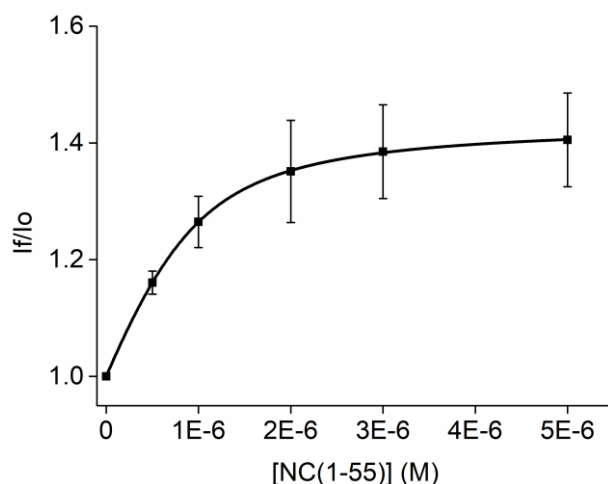
Figure 60. Binding curve of NC(1–55) with thG10SL3_RNA. Oligonucleotide concentration was 1 µM in 50 mM Hepes, NaCl 150 mM, MgCl$_2$ 1 mM in pH 7.5. Solid line corresponds to the fit of the experimental points with Equation 10.

Time resolved intensity decays showed an increase in the average fluorescence lifetimes of thG10SL3_RNA upon addition of NC(1–55). The increase in fluorescence quantum yield of thG10SL3_RNA upon addition of NC(1–55) was mainly due to the increase of the long-lived τ_3 lifetime component from 16 ns (a value close to that in methanol) to 22 ns (a value close to that in water), suggesting that NC exposes thG10 to the solvent. This solvent exposed (22 ns) lifetime component is comparable to the long lived (24.7 ns) lifetime component of thG12SL3_RNA in absence of NC. As the long-lived lifetime of thG12SL3_RNA is already high in the absence of NC, this may readily explain why NC cannot further increase its value and thus that NC has no effect on the QY of thG12 in this case.

To determine the binding constants for thG-labeled SL3 ODN that did not show fluorescence change upon NC(1–55) addition, we changed our approach and performed fluorescence anisotropy titration (Figure 61). Fluorescence anisotropy (Figure 61a) changes for thG-labeled SL3_DNA upon NC(1–55) addition suggest weak binding affinities ($K_d > 10\,\mu M$). As much better affinity constants were reported when the binding of NC to SL3 DNA was monitored through the W37 residue of NC (Vuilleumier et al., 1999), substitution of G10 and G12 residues by thG appears in this is case highly detrimental for binding. This could also explain why we do not observe heat release through ITC measurements.

On the other hand, using a 1:1 binding stoichiometry, the calculated binding affinity of NC to thG10SL3_RNA through anisotropy titration ($K_d = 340 \pm 70$ nM) was found to be comparable to the affinity obtained from fluorescence intensity titration ($K_d = 340$ nM)

experiments but six times less than from ITC measurements (Figure 61b black). The binding affinity of thG12SL3_RNA ($K_d = 2600 \pm 100$ nM) was found to be 7 times less favorable than the affinity of thG10SL3_RNA, suggesting that replacement of G by thG is more detrimental for binding at position 12 than 10. To further gain insight into the structure of NC-SL3 (only RNA) complex, we used molecular dynamics approach.
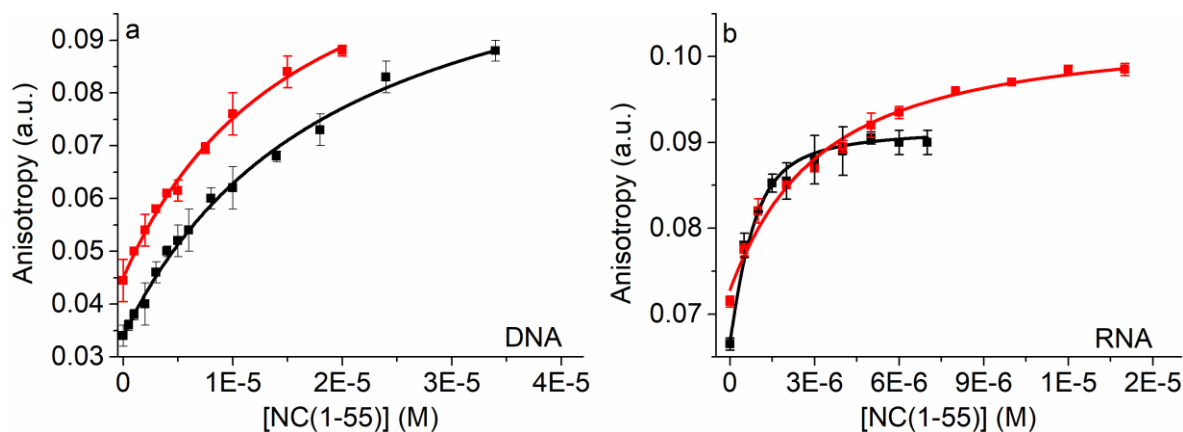
Figure 61. Binding curves of NC(1–55) with thG-labeled SL3 DNA (a) and RNA (b). Anisotropy titration curves for thG-labeled SL3 at position 10 (black) and position 12 (red). Oligonucleotide concentration was 1 µM in 50 mM Hepes, NaCl 150 mM, MgCl$_2$ 1 mM in pH 7.5. Solid lines correspond to the fit of the experimental points with Scatchard equation for anisotropy titration (Equation 12).

Molecular dynamics

Using molecular dynamics, we further investigated the possible structural consequences of the introduction of thG probe in position 10 or 12 of SL3_RNA. In particular, the effect of G replacement with the thG-H1 conformer was investigated both in SL3 alone in water and in its complex with NC. The most representative conformations of SL3 in the wild-type form, bearing thG-H1 in position 10 or 12 are superimposed in Figure 62A and, overall, are comparable to each other. Interestingly, the introduction of the thG probe in position 10 does not seem to affect the conformational landscape of SL3 (Figure 62B) and the nucleobase occupies the same position towards the solvent as in the wild type system (Figure 62B). Though, introduction of thG-H1 in position 12 does not impact the overall conformation of the phosphate backbone, it is projected towards the solvent in a conformation that is different from that of the nucleobase G (Figure 62C). This difference results from the fact that thG-H1 is mainly in an *anti* conformation, in contrast with the *syn* conformation of G12 in the wild type system and establishes H-bond interactions with the phosphate backbone near the 5' terminal

end. This difference in conformation may well explain the lower affinity of the [th]G12-labeled SL3 as compared to its native counterpart. The high solvent exposure of [th]G12 is also in line with the high value of the long-lived fluorescence lifetime (24 ns).
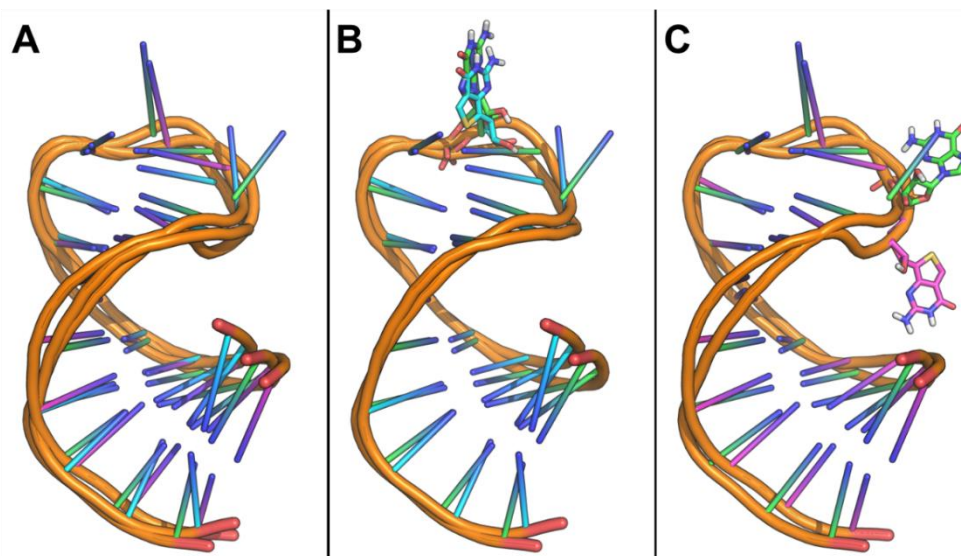
Figure 62. Representative conformation of SL3 wild-type, and incorporating [th]G-H1 in position 10 or in position 12. A) Superimposition of the most representative conformation of SL3 wt (green) SL3 with [th]G-H1 in 10 (cyan) and SL3 with [th]G-H1 in 12 (magenta). B) Superimposition of SL3 wt and SL3 with [th]G-H1 in 10: the nucleotides in position 10 are showed as sticks. C) Overimposition of SL3 wt and SL3 with [th]G-H1 in 12: the nucleotides in position 12 are showed as sticks.

From an energy standpoint, the introduction of [th]G-H1 in position 10 or in position 12 does not significantly modify the total energy of the SL3 system, as underlined in Figure 63.
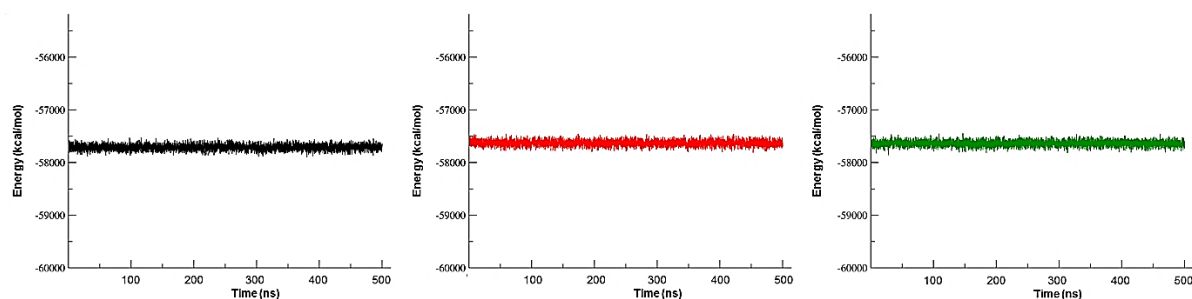
Figure 63. Plot of the total energy of the SL3 wild-type (A, black line), the SL3 bearing [th]G-H1 in position 10 (B, red line), and SL3 bearing [th]G-H1 in position 12 (C, green line) along MD simulation.

Comparison between the three SL3 systems in complex with NC reveals that introduction of [th]G-H1 in position 10 or 12 of SL3 does not impair significantly the structure

of the complexes. After 500 ns of MD simulations, all complexes are stable and at the geometric convergence (data not shown). Moreover, though the overall structures of the systems slightly differ from each other (particularly when comparing wildtype NC-SL3_RNA with NC-thG12SL3, a behavior that may be attributable to the intrinsic flexibility of the NC), the local effect due to the replacement of G with thG-H1 is negligible (Figure 64). This is further confirmed by the estimation of the binding affinity through the MM-PBSA methods, which highlights a highly comparable binding energy of SL3 to NC in all the three cases (Table 8), in variance with the affinity differences observed experimentally. Figure 64 shows the local environment in the positions where G was replaced by thG-H1, and offers the possibility to compare the behavior of the wild-type SL3 (top panels, Figure 64A, and Figure 64C) with that of the SL3 bearing thG-H1 in position 10 (Figure 64B) or in position 12 (Figure 64D).
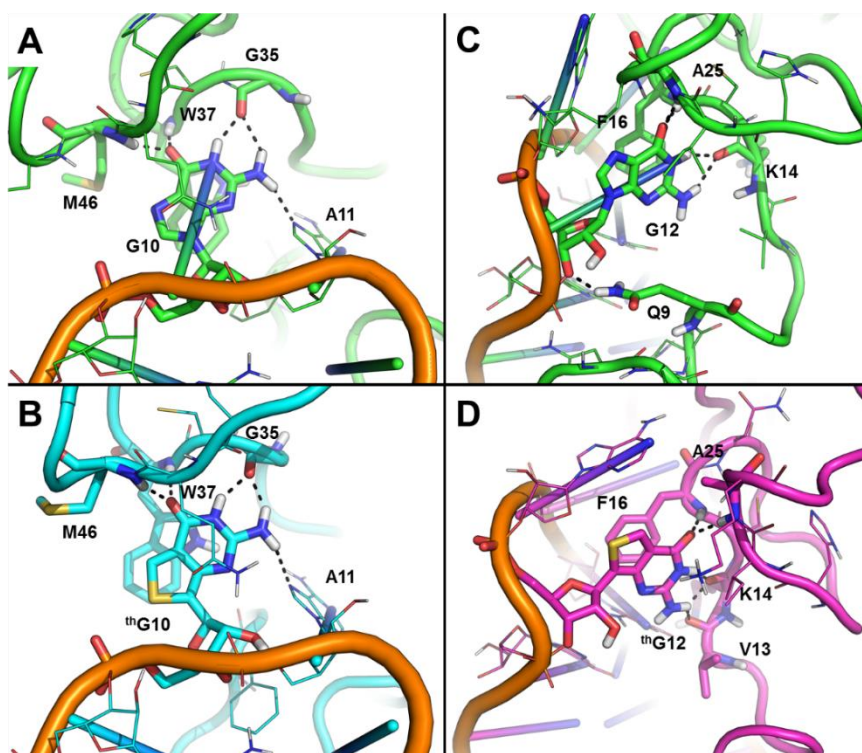
Figure 64. Analysis of the local environment surrounding positions 10 and 12 in the SL3/NC complexes studied herein. NC and SL3 are showed as cartoon, residues and nucleotides within 5 Å from the nucleotide in position 10 or 12 are showed as lines, while NC residues contacted by the nucleotide in position 10 or 12 are showed as sticks. Polar contacts are highlighted by black dashed lines. The wild-type NC-SL3 system (A and C) is colored green, the NC-thG10 SL3 (B) is colored cyan, while the NC-thG12 SL3 (D) is colored magenta.

121

In more details, G and [th]G-H1 in position 10 are stacked over the indole side chain of the key residue W37, and establish the same network of H-bond interactions that involves NC residues G35, W37 and M46, in agreement with literature data (De Guzman et al., 1998). An additional intramolecular H-bond is established with the flanking nucleobase A11 (Figure 64A-B). This analysis suggests that introduction of [th]G in position 10 should not significantly affect the binding properties of SL3 to the NC. Similar to the native G (Figure 65), [th]G12 is stacked to the phenyl ring of F16 and H-bonded to K14 and A25. The sugar moiety of G12 establishes an additional H-bond with the side chain of Q9, whereas [th]G-H1 at position 12 contacts the backbone of V13. Although there is small difference in the interaction pattern, the local affinity of G12 and [th]G in this position should be comparable. Contrary to the natural G12, combining the differences in orientation (*anti*) shown on substituting of [th]G12 in SL3_RNA and later in terms of binding contacts made with NC, it infers the poor binding affinity obtained through experimentally approach.
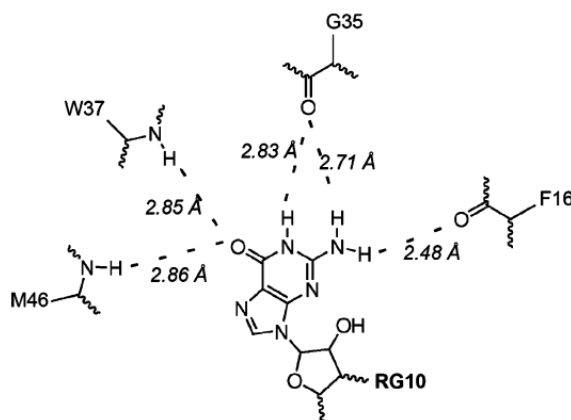
Figure 65. Schematic representation of the binding site for the central guanosine G10 of the SL3_RNA hydrogen bonds are displayed as dashed blue lines (Mori et al., 2010b)

This conclusion was substantiated by the theoretical affinity of SL3 to NC in these systems estimated by the MM-PBSA approach. Results of Table 8 suggest that the SL3 affinity for NC is preserved when [th]G is inserted in position 10 or in position 12, even though a higher affinity was experimentally measured for the wild-type SL3 system as compared to the [th]G-labeled SL3 derivatives.

Table 8. Theoretical affinity of SL3 for the NC estimated in studied systems

System	ΔG binding MMPBSA (kcal/mol) ± SEM	ΔG binding MMGBSA (kcal/mol) ± SEM
SL3/NC	-61 ± 1.2	123 ± 2
NC- thG10SL3	-59 ± 1.3	122 ± 2.5
NC-thG12SL3	-57.8 ± 1.3	119.5 ± 1.5

Next, the thG-H3 conformation was inserted in two different positions of free SL3, namely 10 and 12. At both positions, the introduction of thG-H3 has limited impact on the total energy of the system, that is comparable to the total energy calculated for wildtype SL3 and bearing thG-H1 (data not shown). thG-H3 in position 12 did not impact on the overall structure of SL3, as described in Figure 67. The main difference between thG-H1 and thG-H3 in position 12 is the conformation of the thG nucleotide that is in a *syn* and less solvent-exposed structure in thG-H3 compared to the *anti* form observed preferentially in thG-H1. Apparently, conformations of thG-H3 resembles closer to the natural G12 residue but the experimental approach (Figure 66c,d) shows that the equilibrium between the two tautomers majorly favors the thG-H1 along with a minor contribution from thG-H3. Further, this contribution pattern is also seen in the complex of thG12SL3_RNA with NC which confirms that thG-H3 participates in the binding and probably with a binding mode different than thG-H1 it would affect the overall binding affinity.
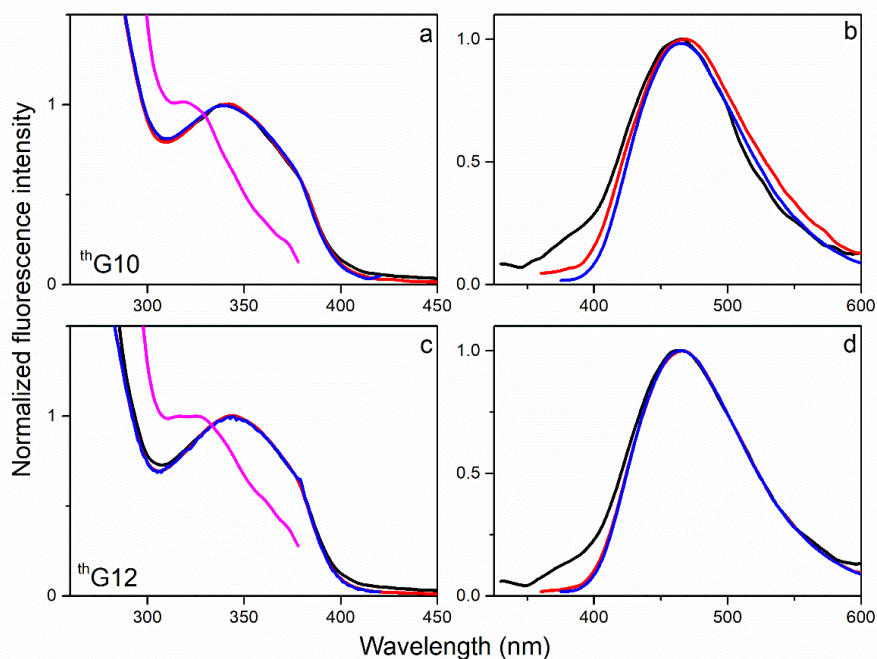
Figure 66. Normalized excitation (a,c) and emission (b,d) spectrum of thG10 (a,b) and thG12 (c,d) labeled SL3_RNA complexed with NC. Excitation spectra recorded at different emission wavelengths: λ_{em} = 550 nm (black line); λ_{em} = 500 nm (red); λ_{em} = 450 nm (blue); λ_{em} = 400 nm (magenta). Emission spectra were recorded at different excitation wavelengths: λ_{ex} = 315 nm (black line); λ_{em} = 345 nm (red); λ_{em} = 360 nm (blue).
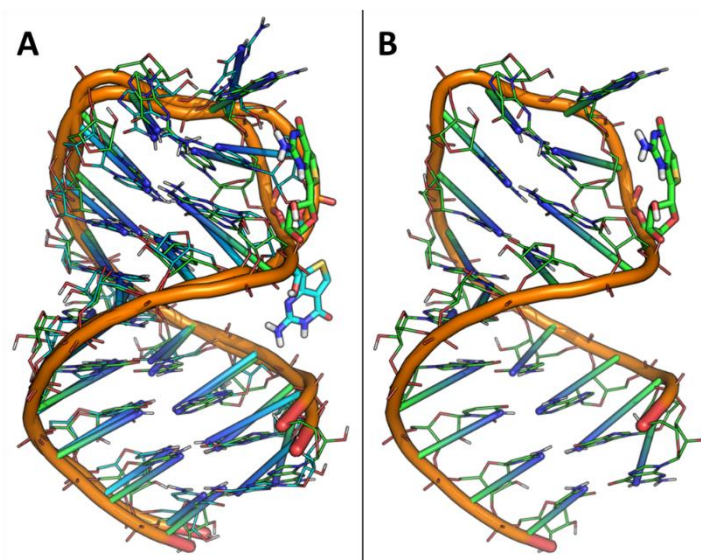
Figure 67. Introduction of of thG-H3 in position 12 of SL3. A) structural superimposition of SL3 bearing thG-H1 (cyan) and thG-H3 (green). B) representative structure of SL3 bearing thG-H3 in position 12.

124

When [th]G-H3 was inserted in position 10 of free SL3, the system evolved in two slightly different conformational states, namely conformations A and B, which are however not too different from each other as well as from the wild type SL3 and [th]G-H1-labeled SL3. As shown in Figure 68, and compared to the [th]G-H1 tautomer, [th]G-H3 can exist in the *syn* and *anti* forms, thus giving rise to the abovementioned conformational states. As expected, the *syn* form (conformation A, Figure 68A) is slightly different from the wild type SL3 or SL3 bearing [th]G-H1, compared to the *anti* form (conformation B, Figure 68B). Similar to observation made at position 12, experimental approach (Figure 66a,b) evidence that at position 10 a minor contribution from [th]G-H3 exists as substituted in SL3_RNA and its complex with NC.
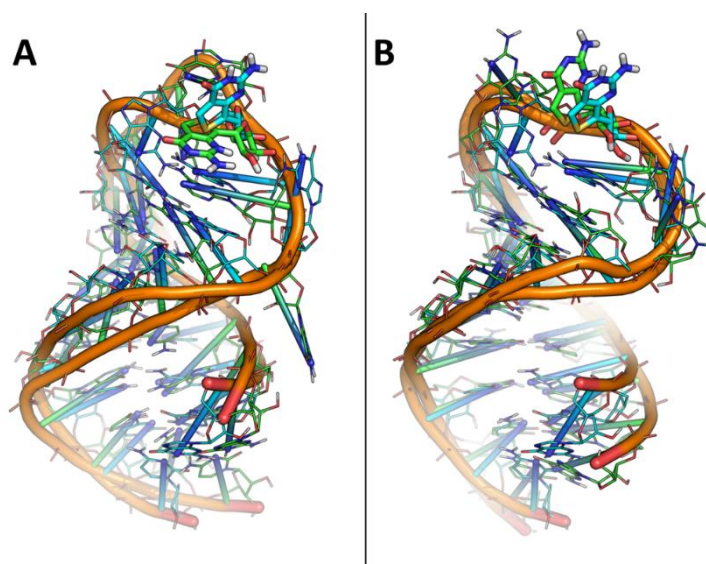
Figure 68. Introduction of [th]G-H3 in position 10 of SL3. A) structural superimposition of SL3 bearing [th]G-H1 (cyan) and [th]G-H3 conformation A (green). B) structural superimposition of SL3 bearing [th]G-H1 (cyan) and [th]G-H3 conformation B (green).

Looking at the two conformations in deeper details reveals that, in conformation A, [th]G-H3 is stacked with the two flanking bases, namely G9 and A11 (Figure 69A). In contrast, in conformation B it is stacked only to the G9 thus exposing the other edge of the probe to the solvent (Figure 69B). This solvent exposed conformation may represent the 16.3 ns long lived lifetime of [th]G10SL3_RNA. In comparison to the conformation diversity seen at position12 among both the [th]G tautomers, position 10 displays relatively lesser complications. Therefore, it could correlate with the fact that binding affinity of NC to thG10SL3_RNA is closer to the native SL3_RNA and poorly stated for [th]G12SL3_RNA.
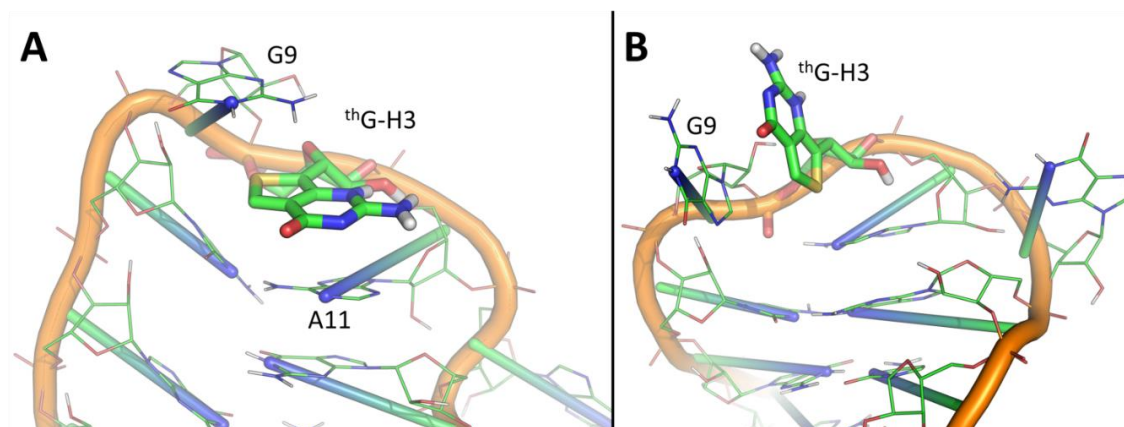
Figure 69. Different conformations of ^thG-H3 in position 10 of SL3. A) details of conformation A with ^thG-H3 in a *syn* form are showed. B) details of conformation B with ^thG-H3 in an *anti* form are showed. ^thG-H3 is shown as sticks; nucleotides stacked to ^thG-H3 base are labeled.

Docking of ^thG-H1 and ^thG-H3 to the binding site of NC

In order to provide further support to the binding conformations of ^thG-H1 and -H3 within the hydrophobic pocket observed via MD, molecular docking studies were conducted. To this end, the MD-refined structures of the NC in complex with SL3_RNA or PBS_DNA (Mori et al., 2010b) were used as rigid receptor in molecular docking, which was performed with GOLD and FRED programs (Jones et al., 1997; McGann, 2012, 2011; Verdonk et al., 2003). The ^thG base was sketched de novo, and as a preliminary approach docking is performed in absence of the sugar moiety that was replaced with a methyl group for the sake of simplicity. Docking results summarized in Figure 70 are rather consistent with MD observations (Figure 64) and previous computational studies (Mori et al., 2011).
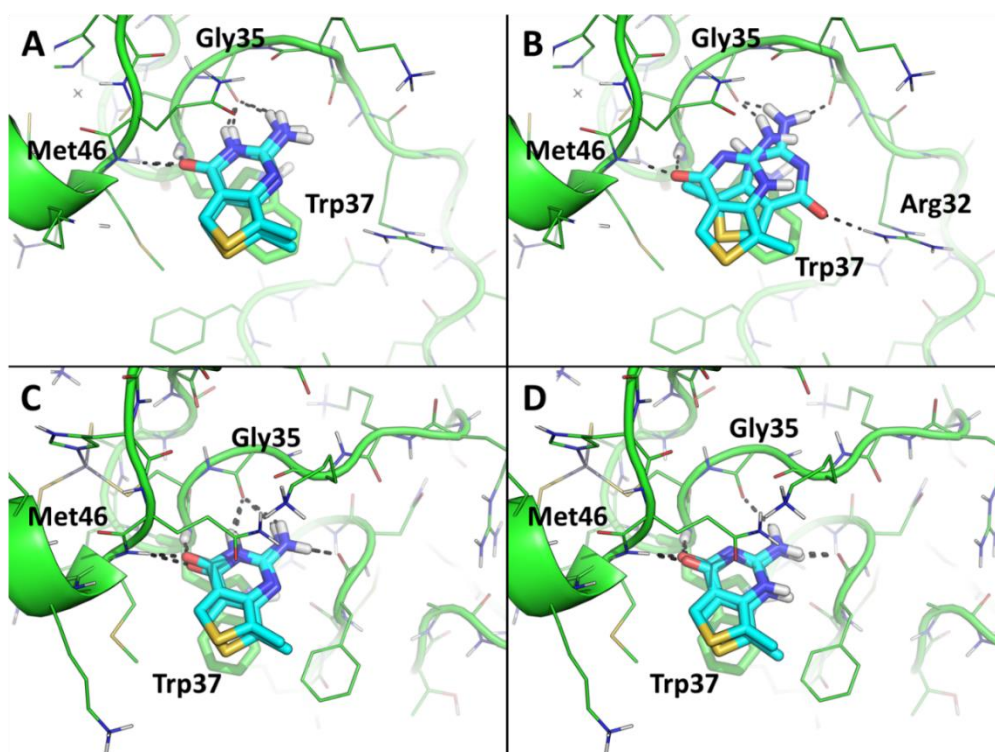
Figure 70. thG behaves as guanine-mimetic within the hydrophobic pocket of the NC. The predicted binding mode of thG-H1 and thG-H3 within the NC structure relaxed in complex with PBS_DNA (A,B) or SL3_RNA (C,D) is shown. The thG nucleosides (thG-H1 and thG-H3) and W37 are showed as sticks. The best docking pose predicted by GOLD and FRED are showed. In panel B), the pose not consistent with other predictions (that shifted right in the picture) was calculated by GOLD. In all panels, the protein is colored green and showed as lines and cartoon; the molecules are colored cyan. H-bond interactions are highlighted by black dashed lines, residues H-bonded by the ligands are labeled.

Docking simulations showed that both the thG tautomers forms binding contacts within the hydrophobic plateau of NC for PBS_DNA (Figure 70A,B) and SL3_RNA (Figure 70C,D) that are similar to contacts shown earlier using MD observation. However, an exception is observed for thG-H3 within NC-PBS_DNA complex (Figure 54B) in which thG-H3 makes one H-bond less than thG-H1 to the Gly35 backbone. This difference in H-bond networking is expected to provide a weaker affinity for thG-H3 within the NC plateau compared to thG-H1. As shown in Table 9, this hypothesis is consistent with, and supported by, score analysis of binding poses. For both programs, and towards both NC conformations used herein, thG-H1 shows a higher affinity for NC than thG-H3. Although theoretically, these data support the preference of thG-H1 tautomer with respect to the thG-H3 in binding to the NC.

Table 9. Docking scores of thG-H1 and -H3

	FRED (ChemGauss4)*		GOLD (ChemScore)**	
	NC-DNA	NC-RNA	NC-DNA	NC-RNA
thG-H1	-8.24	-7.87	26.78	30.33
thG-H3	-6.59	-5.97	23.42	25.05

* the lower the score, the higher the affinity

** the higher the score, the higher the affinity

4.3.3. Conclusions

In this work, we aimed to provide insight in an interesting mode of interaction adapted by HIV-1 NC protein to specifically bind with RNA and DNA sequences in opposite direction. Critical analysis of the 3D resolved structure made across several complexes of NC with nucleic acid sequences highlighted the role of sugar moiety to pivot the NC's binding direction. Herein, we used a fluorescence-based approach in order to observe whether the NC's opposite binding direction also follows within the RNA and DNA variant of the same nucleic acid sequence (namely (–)PBS and SL3). To monitor the interaction with NC, we incorporated thienoguanosine (thG) by replacing key residues of (–)PBS (G7) and SL3 (G10 and G12) in their DNA and RNA variants.

Firstly, we identified and mapped the distribution of thG tautomers as substituted in DNA and RNA copies of (–)PBS and SL3 ODNs. This proves instrumental in selectively exciting the thG-H1 tautomer and following its fluorescence change upon interaction with NC. Substitution of thG in (–)PBS and SL3 provides crucial information on folding of their loop as the NMR structures of (–)PBS_RNA and SL3_DNA are not resolved. In (–)PBS, it strongly suggests comparable conformational arrangement of thG7 with its neighboring residues (T6-T8) and it is highly probable that both the DNA and RNA variants of (–)PBS structures have similar penta-nucleotide loop folding. In SL3_RNA, thG10 residue is partially stacked with neighboring bases while thG12 residue is largely solvent exposed, in line with the NMR and MD calculation results. However, in SL3_DNA it appears that the tetra-loop may be more rigid, and both the residues thG10 and thG12 are mostly solvent exposed.

MD calculation showed that both the thG tautomers within PBS_DNA mimic the conformations of the natural G7 residue and form similar binding contacts with NC. On an energetic standpoint, the substitution preserves the (–)PBS_DNA-NC interaction as they do not modify the total energy of the system. We observed differences in orientation (*syn-anti form*)

of [th]GH1 and [th]GH3 tautomers at position 12 and 10, respectively. These changes in orientation of [th]G tautomers affect the binding interaction with NC, as differences in binding contacts were observed in comparison to native G residue. Hence, we observed a lower binding affinity than for the non-labeled SL3_RNA-NC complexes.

NC binds to labeled and non-labeled RNA and DNA copies of (–)PBS with similar binding affinities as shown by ITC titration. Fluorescence titrations also revealed similar binding affinities among RNA and DNA variants of [th]G labeled (–)PBS but they were found to be 3- to 10-fold weaker than the value through ITC titration. These differences in binding affinities showed that the binding site associated with heat release differs from the binding sites(s) giving a change in [th]G fluorescence. Interestingly, the photophysical investigation of [th]G-labeled sequences showed a two-fold increase in quantum yield associated with a parallel change in lifetimes distribution upon NC interaction. This suggests that the [th]G conformational changes due to its stacking with Trp37 appear similar in both DNA and RNA copies of (–)PBS. However, for [th]G-labeled SL3 ODNs, NC induced conformational changes in the loop could not be compared as we observed a fluorescence change only for [th]G10SL3_RNA.

The ITC titrations of NC to RNA and DNA variants of SL3 ODNs also display similar binding affinity. However, as all [th]G-labeled SL3 ODNs failed to produce sufficient heat exchange, it limited us to accurately determine their binding affinity. As NC has a single binding site for SL3_RNA, it becomes clear that modifying the preferential binding site significantly alters the thermodynamic parameters, which is contrary to NC-PBS interaction that has three binding sites. Although [th]G is an isomorph of guanosine, it is likely that the N7 atom which is missing in [th]G, may play a role in affecting the heat exchange. Although we used fluorescence titration to determine the binding affinity but other than [th]G10SL_RNA, none of the other labeled ODNs displayed fluorescence change upon NC addition. As for [th]G12SL3_RNA, MD simulation evidenced that [th]G12 residue is completely solvent exposed. For both the [th]G labeled SL3_DNA, the high fluorescence lifetime strongly indicates that in the SL3 penta-loop both the G10 and G12 residues are also solvent exposed.

Altogether, we confirmed that the interaction of the Trp37 residue with guanines was critical for the formation of complexes with both RNA and DNA variants of PBS and SL3. This work also accords with our earlier observations, which showed that NC shows similar binding affinity towards shortest pentanucleotide 5'-A_1CGCC$_5$-3' in their RNA and DNA variants (Morellet et al., 1998) At least with NC-PBS complex, we report that NC likely adapts

similar binding mode during the interaction. As shown earlier with (–)PBS model sequence (Sholokh et al., 2015), this work also presents a specimen in which thG as substituted at position 10 and 12 of SL3_RNA is able to faithfully mimic the conformations of native guanosine and preserves its interaction with NC.

4.4. CHAPTER 4: Screening the Inhibitors of Nucleocapsid Protein

4.4.1. Introduction

Highly active anti-retroviral therapy (HAART) is effectively employed to treat patients diagnosed with HIV/AIDS. The treatment of HIV was revolutionized in the beginning of 1990s by the development of inhibitors targeting reverse transcriptase and protease enzymes. Initially, HIV treatment was administered as a monotherapy. However, later on, HAART administered a combination or cocktail of drugs comprising three antiretroviral agents targeting two distinct enzymes. In comparison to monotherapy, combinational therapy proved beneficial as there have been observed reduction in morbidity and mortality of HIV-infected patient. Drug resistance resulted as HIV is prone to mutate under pharmacological pressure. Briefly, 1 to 10 mutations appear in the viral genome due to the highly error prone reverse transcription process. This generation of mutated species tend to develop low susceptibility towards antiretroviral agents. Nevertheless, several approaches were developed to tackle mutated drug resistant viral proteins (Ghosh et al., 2012; Kuritzkes, 2011) or to target viral structures that bind with host cellular proteins having a low tendency to mutate (Kellam, 2006) or to target host cell factors catalyzing the virus replication cycle (Garbelli et al., 2011; Maga et al., 2013).

Among all HIV proteins, the nucleocapsid protein is of particular interest because,

- It is highly conserved in diverse HIV subtypes and play an important role in the virus replication cycle (Darlix et al., 2016, 2011; Levin et al., 2005; Thomas and Gorelick, 2008)

- Point mutations in NC results into fully non-infectious viruses (Aldovini and Young, 1990; Demene et al., 1994; Dorfman et al., 1993; R J Gorelick et al., 1999)

- NC exhibits nucleic acid chaperoning activity that assist reverse transcriptase and integrase enzymes during reverse transcription and integration, respectively (Darlix et al., 2016, 2011)

Therefore, NC appears as a key target for developing inhibitors that may be able to alter several key processes during HIV lifecycle. Therefore, the THINPAD (**T**argeting the **HI**V-1 **N**ucleocapsid **P**rotein to fight **A**ntiretroviral **D**rug Resistance) consortium was organized from leading scientists in the field of NC and was supported by The European Union's Seventh Framework Programme (FP7) for Research and Innovation.

4.4.2. THINPAD

4.4.2.1. General Objectives of the THINPAD Project

- To develop potential NC inhibitors (NCIs) from discovery to preclinical trials.

- To tackle the appearance of possible drug resistance.

- To execute pharmacokinetics and ADMET investigations to optimize the NCIs till the evaluation of its preclinical efficacy.

4.4.2.2. The Partners of THINPAD

Partner 1. University of Siena, Siena, Italy (research groups of Prof. Maurizio Botta and Dr. Maurizio Zazzi)

Partner 2. IDIBAPS (Consorci Institut D'Investigacions Biomèdiques August Pi I Sunyer), Barcelona, Spain (research group of Prof. Jose M Gatell)

Partner 3. University of Strasbourg, Strasbourg, France (research group of Prof. Yves Mély)

Partner 4. IRBM Science Park, Rome, Italy

Partner 5. ViroStatics Srl, Alghero, Italy

4.4.2.3. Organisation of the Project

The organisational platform is briefly shown in Figure 71.
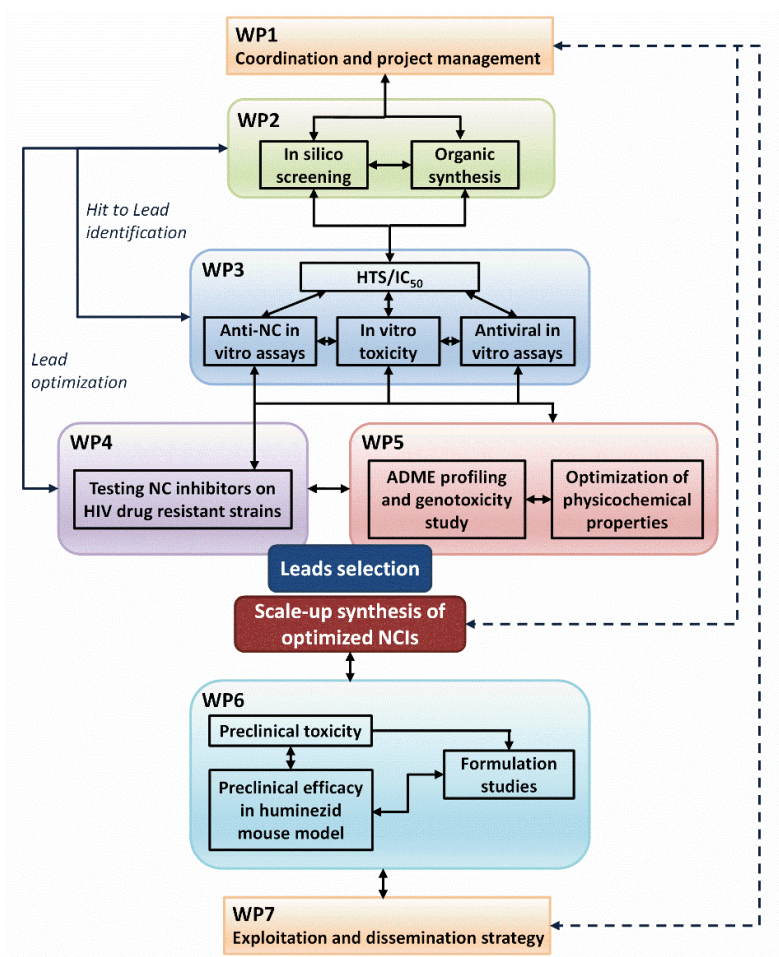
Figure 71. Schematic representation of THINPAD project organization.

4.4.2.4. Role of our Laboratory

Our laboratory has been assigned to test compounds via an *in vitro* assay and to determine the mechanism of action of the potential hits (WP3 in Figure 71). Our primary assay (NCinh) was based on the NC destabilization activity on the cTAR stem-loop sequence. cTAR is the complementary sequence of TAR, the trans activator response element.

My role:

- To participate in the high throughput screening of compounds and determine their IC_{50} through NC inh assay.

- To develop a site-specific assay for screening inhibitors that are able to compete with NCp7 for its specific binding sites.

4.4.3. NCinh assay
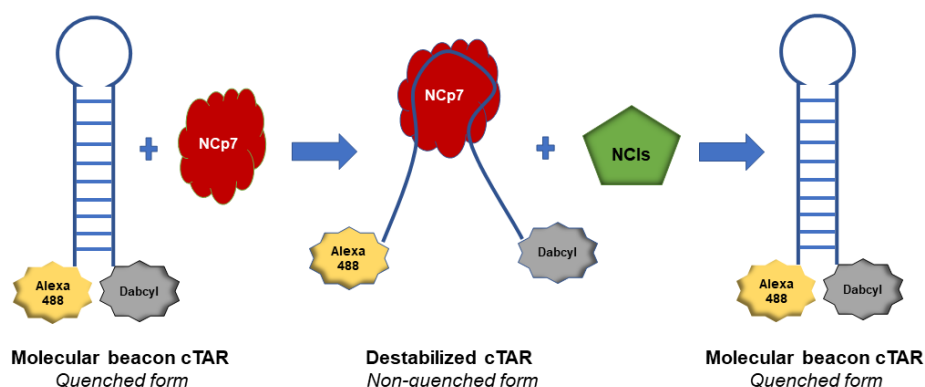
Figure 72. Scheme of NCinh assay.

Molecules selected by in silico screening were tested in the NCinh assay for their ability to inhibit the NC-nucleic acid chaperone activity through a highly specific high throughput screening assay. In this assay, we monitored the destabilizing activity of NC(11–55) on a cTAR oligonucleotide labeled at its 5' and 3' ends by Alexa488 and Dabcyl, respectively. In absence of NC, the cTAR hairpin induces a strong fluorescence quenching of the Alexa488 emission by the Dabcyl group. When NC is added in 10-fold excess, it melts the bottom of the cTAR stem, thereby increasing the distance between the two fluorophores and increasing the fluorescence emission of Alexa488 (Figure 72). Similar results have been reported earlier from our group by using Rhodamine instead of Alexa488 (Shvadchak et al., 2009). The destabilization activity of NC to cTAR is very sensitive to the zinc-bound finger motifs, as upon zinc ejection (through EDTA) the destabilizing activity of NC disappears. Therefore, NC inhibitors are expected to prevent the NC-induced destabilization of cTAR and thus, reverse the NC-induced increase in cTAR fluorescence. Compounds active in the destabilization test are further examined for their IC_{50} values, by adding increasing concentrations of the selected compounds to the pre-formed NC(11–55)-cTAR complex.

Using NCinh assay the Hit-to-lead optimization of inhibitor's IC_{50} are shown in Figure 73. Starting from the parent compound having an IC_{50} of 167 µM, we screened 208 analogues, 107 were positive and 101 negative. Among the positove compounds we succeed to achieve IC_{50} less than 10 µM and for the best analogues an IC_{50} of 1.5 µM.
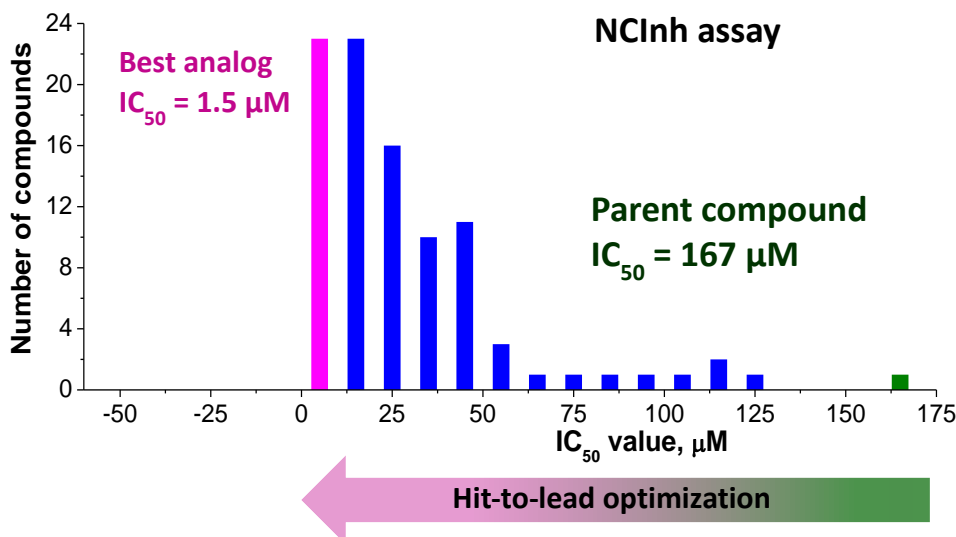
Figure 73. Optimization of the most active series of inhibitors discovered by THINPAD project.

4.4.4. Site Specific Binding Assay

Our group has also worked on developing an additional assay to monitor the competition of NCIs to the specific binding of NC(11–55) to the primer binding site DNA (–)PBS. NC(11–55) binds preferentially at the 5' end of the loop (5-CTG-7) (Sarah Bourbigot et al., 2008; J Godet et al., 2011)which induces the flipping of the G7 residue from the crowded inner environment of the loop toward the solvent. In order to observe the mechanism of interaction, 2-Aminopurine (2Ap), a fluorescent adenine mimic, was first used to substitute the G7 residue. However, 2Ap had major drawbacks namely a low quantum yields as well as a limited ability to conservatively substitute a G residue. To overcome these limits, a newly synthesized analogue of guanosine, thieno [3,4-d]–deoxyguanosine (thG), was used as an alternate for 2Ap. Unlike 2Ap, it was found to perfectly mimic the G7 residue in the loop. Moreover, the quantum yields of thG in (–)PBS [thG(–)PBS = 0.10] and in (–)/(+)PBS [thG(–)/(+)PBS = 0.20] were significantly higher as compared to 2Ap [2Ap7(–)PBS = 0.08, and 2Ap7(–)/(+)PBS = 0.01, respectively]. In addition, both the large number of fluorescence lifetimes and the strong contribution of the dark species in 2Ap was an obstacle for time-resolved fluorescence measurements. This was clearly avoided with thG, where lesser lifetime components and no dark species were observed. Interestingly, interaction of thG(–)PBS with NC(11–55) provided a 1.6 times increase in quantum yield that could be used to identify hits that are able to inhibit the selective binding of NCp7 to (–)PBS, and thus, the NCp7-promoted (–)/(+)PBS annealing in the second strand transfer (J Godet et al., 2011).

135

To validate our assay, we selected a few compounds: A05, D04, D05 and WDO that were identified through the NCinh assay. The validation of the assay required a few critical steps.

- To determine the emission and absorption profile of the inhibitor (autofluorescence).

- To test if the inhibitors interact with free thG7(–)PBS.

- To optimize the added concentration of the inhibitor in order to avoid aggregation.

We obtained autofluorescence spectrum of each A05, D04, D05 by exciting their 1 µM and 10 µM concentration solution at wavelength 370 nm. These emission spectrum were then compared to the emission spectrum obtained by exciting complex of NC-thG7(–)PBS (concentration ratio NC:PBS 1:4). Their respective comparison showed D04 (Figure 74c) and D05 (Figure 74e) to be highly emissive at these concentrations and thus cannot be used further for titration with NC-thG7(–)PBS. However, A05 (Figure 74a) at 1 µM was found to be least emissive and was further used for titration.
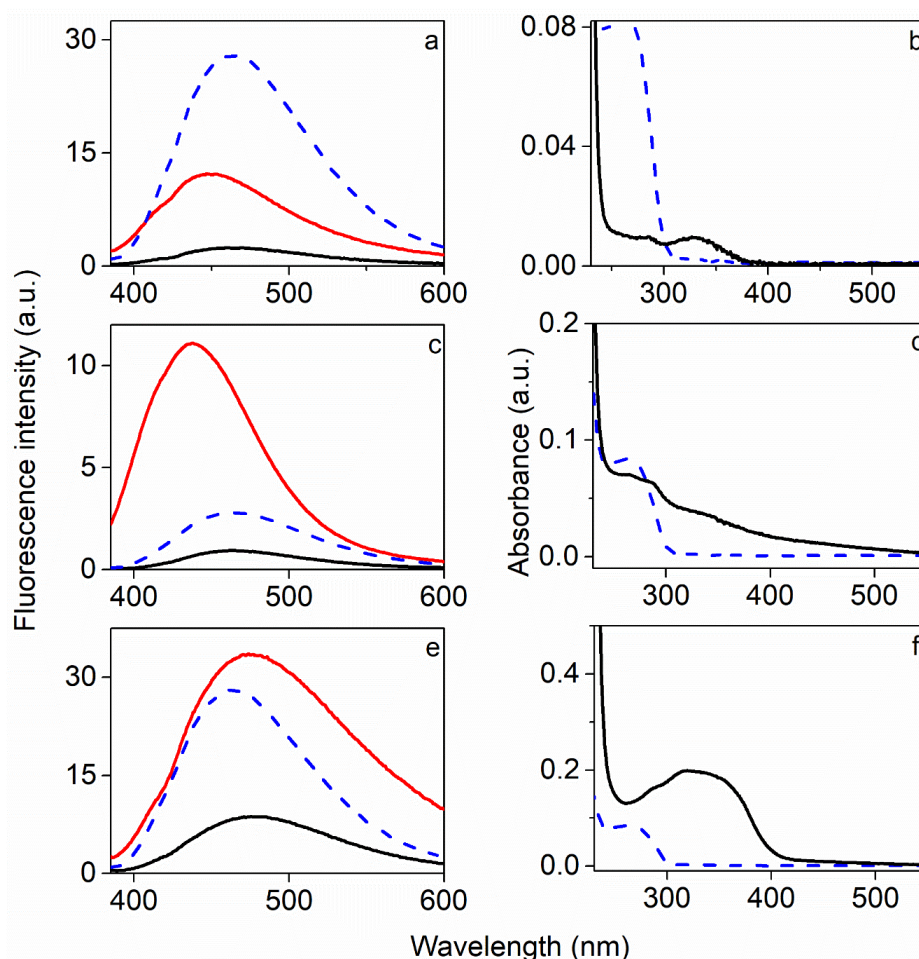
Figure 74. Comparison of autofluorescence spectra (a,c,e) of 1 µM (black) and 10 µM (blue) concentration of A05 (a); D04 (c); D05 (e) with emission spectra of NC(11–55)-thG7(–)PBS complex (blue dotted line). Comparison of absorption spectra of 1 µM (black) concentration of A05 (b); D04 (c); D05 (f); with absorption spectra of NC(11–55)-thG7(–)PBS complex (blue dotted line). Buffer 25 mM TRIS, 30 mM NaCl, 0.2 mM MgCl$_2$. Excitation wavelength 370 nm.

To observe activity of A05, we titrated it to NC-thG7(–)PBS complex (Figure 75). Firstly, we added 4 equivalents of NC(11–55) to thG7(–)PBS that resulted in 1.8 fold increase in its fluorescence. Further, upon addition of 1µM A05 to NC-thG7(–)PBS complex, the fluorescence intensity increased to more than two-fold (Figure 75, blue line).
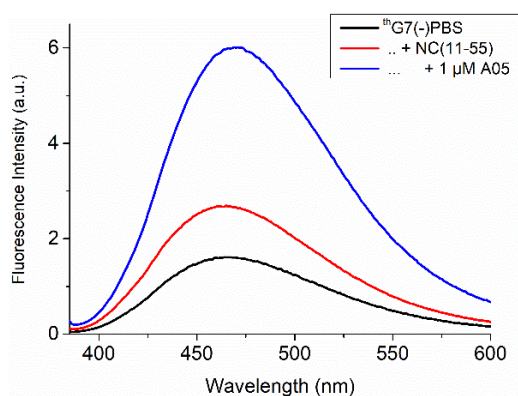
Figure 75. Fluorescent based activity test for 1 µM A05 on thG7(–)PBS/NC complex.

This behavior of A05 stand contrary to its activity observed in NCinh assay. To reason this behavior we followed the A05 interaction with thG7(–)PBS in absence of NC. Addition of A05 at 1 µM increased thG7(–)PBS emission by 2-fold, suggesting that A05 interacts with thG in the (–)PBS motif (Figure 76).
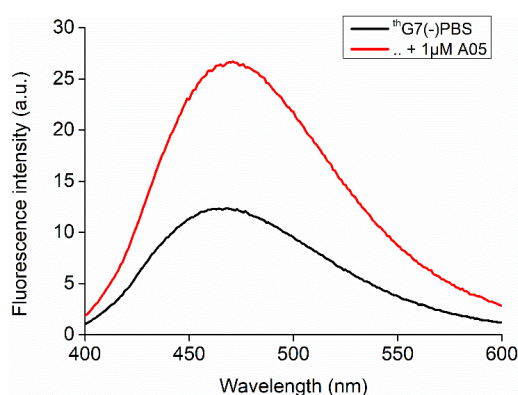
Figure 76. Spectrum of thG7(–)PBS in the absence (black) and the presence of 1 µM A05 (red). Excitation wavelength 370 nm.
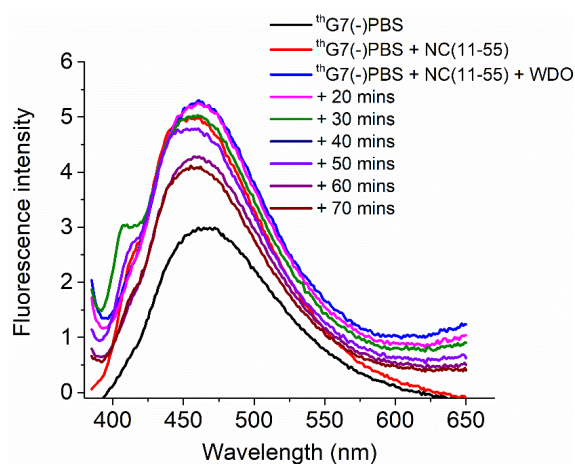
Figure 77. Activity of WDO on NC-thG7(–)PBS complex recorded after certain time interval.

138

Next, we tried WDO, a NCI acting as a zinc ejector. As usual, the NC-thG7(–)PBS complex was added with 20 µM of WDO and the fluorescence was recorded after every 10 mins. As seen from Figure 77, the fluorescence intensity decreases with time that confirms WDO activity. Additional experiments will be needed in order to optimize the concentration of WDO (and other compounds) in a way to avoid possibility of aggregation and enhanced sensitivity. The positive activity of WDO shows that this assay can be used for screening of inhibitors, provided inhibitors are non-emissive and non-interaction with thG. Nevertheless, this work is in perspectives for further development.

4.4.5. Conclusions

The primary task while participating within THINPAD consortium, was to screen inhibitors of HIV-1 nucleocapsid protein using NCinh assay.The inhibitors were examined by determining their IC$_{50}$ values. Hit-to-lead optimization resulted in discovery of positive compounds with IC$_{50}$ less than 10 µM. In addition to the NCinh assay, a new assay was developed in order to monitor the competition of NCIs to the specific binding of NC(11–55) to the (–)PBS. A site specifically incorporated fluorescent nucleoside analogues, thG, substituting G7 residue of (–)PBS was used as substrate for NCI and NC interaction. Validation of assay was performed by using inhibitors that were found active in NCinh assay. As a control, assay was found to be functional while examined with WDO, showing that this assay can be used for screening of compounds, provided inhibitors are non-emissive and non-interactive.

5. Conclusions

The aim of this thesis work was to characterize and target the dynamic tryptophan-guanine recognition during the interaction of HIV-1 nucleocapsid protein (NC) with target nucleic acid sequences. Among the numerous roles played by NC in HIV-1 life cycle, we focused on chaperoning of nucleic acid strands during reverse transcription and selective packaging of unspliced viral genomic RNA. We used fluorescence nucleobases base analogue as a tool for studying various aspects of RNA and DNA such as their structure, dynamics and interactions with NC protein. The primary task was to select a suitable fluorescent nucleobase analogue that can effectively report the biomolecular interaction. Photophysical characterization of different types of fluorescent base analogues as free probe and substituted in ODN and their sensitivity to peptide-ODN were examined. These fluorescent nucleobase analogues include isomorphic probes: 2-Aminopurine (2Ap), 2-Amino-6-(2-thienyl)purine (**s**) (analogue of Adenine) and thieno deoxyguanosine (thG7) (analogue of guanine), and extended probe: 2-thenyl,3-hydroxychromone derivative (3HCnt). The selected probe was used to i) monitor the annealing kinetics of (–)/(+)PBS in order to determine the full set of kinetic rate constants, and ii) to understand the differences in the binding modes adopted by NC towards RNA and DNA sequences.

First work included selection of a suitable probe for monitoring the annealing kinetics of (–)/(+)PBS. We investigated the photophysical properties of **s** labeled at positions 6 and 7 of (–)PBS and compared them to the respective 2Ap-labeled sequences. At both 6 and 7 positions, we observed notably higher fluorescent quantum yield of **s** labeled sequences for both the (–)PBS stem-loop and its duplex (–)/(+)PBS. Moreover, the time-resolved fluorescence properties of **s** are characterized by a lower amount of dark species along with a higher average lifetime as compared to 2Ap. Interaction of **s**6(–)PBS with NC provided a two-fold higher increase in fluorescence quantum yield as compared to 2Ap6(–)PBS, showing its better sensitivity towards interaction with protein. However, no change in fluorescence quantum yield was observed on NC binding when **s** was at position 7, so that this probe is inadequate for monitoring the specific interaction of this position with Trp37. Due to the limited improvement of **s** over 2Ap, its use was discontinued.

Next, we characterized the photophysics of thG both as a free nucleoside and included in (–)PBS. The photophysical analysis of free thG nucleoside in different solvent revealed the existence of two ground state tautomer and their structures were identified using TD-DFT calculation as keto-amino tautomer: thGH1 and thG-H3, differing by the position of the protonated nitrogen atom. The equilibrium between the two tautomers were observed

141

differently across various solvent, as both tautomers were observed in water and protic solvents, the thG-H1 tautomer was largely dominant in H-acceptor solvents. We also identified the lifetimes and quantum yield associated with each tautomer in different solvents, in order to better understand the microenvironment sensitivity of the label when substituted in different model sequences. When thG was substituted in (–)PBS, contribution from both the tautomers were observed. In contrast, Watson-Crick base-pairing in (–)/(+)PBS duplexes strongly favors the thG-H1 tautomer. Molecular dynamics further showed on the basis of conformational and energetic stability, that thG-H1 behaves similarly to its native counterpart in both the single- and double-stranded structures. The ratio of the two tautomers and their relative emission were found to be highly sensitive to the nucleic acids, to the nature of the opposite base in DNA duplexes as well as to protein binding. However, a long lived lifetimes component of thG7(–)/T$_{12}$(+)PBS duplex (24 ns) (thG-T mismatch) was found to be much higher than the longest lifetime of free thG nucleoside in any solvent (19.9 ns, buffer). Thus, further photophysical investigation will be required in the direction to examine the role of neighboring adjacent residue and opposite base on the fluorescence lifetimes of thG as substitute in (–)PBS duplex.

We performed a comparative photophysical investigation of thG and 2Ap substituted at position 7 in (–)PBS model. In contrast to 2Ap, substitution of thG was found to fully preserve the Watson-Crick base pairing in the (–)/(+)PBS duplex, thus serving as perfect surrogate of Guanosine residue. In addition, it show much higher fluorescent quantum yield and simpler fluorescence decay than 2Ap in both (–)PBS stem-loop and its duplex (–)/(+)PBS. Therefore, thG appears much superior than 2Ap in substituting G residues. Interaction with NC(11–55) resulted in a two-fold increase in emission of thG7(–)PBS, highlighting its sensitivity towards reporting the conformational changes of the NC-(–)PBS interaction.

Next, we proceed with thG7(–)PBS to monitor (–)/(+)PBS annealing in absence and presence of NC protein. In parallel, we also investigated the annealing kinetics of (–)PBS with (+)PBS by substituting the T9 residue of (–)PBS by 2-thienyl-3-hydroxychromone (3HCnt). In the absence of NC, the kinetic rate constants obtained separately from the two probes were found to be similar. It was confirmed that the reaction proceeds through a two-step mechanism with the formation of an intermediate complex (IC). The fluorescence intensity of the IC was found to be about 80 % of that of the extended duplex (ED) suggesting that the interaction proceeds not only from the overhang but also involves a loop-loop interaction. In presence of NC, the kinetic traces of thG-labeled (–)PBS reach the final plateau much more rapidly than in its absence, notably through a facilitated formation of the IC, illustrating the NC chaperone

activity. This limited us to obtain the kinetic rate constants. Nevertheless, the kinetics rate constants were determined by 3HCnt-labeled (–)PBS. Using fluorescent nucleobase analogues as site specifically incorporated at two different position in (–)PBS allows for the first time to get the whole set of kinetic rate constants of the annealing reaction with (+)PBS.

In the next work we investigated the role of sugar in NC-RNA and NC-DNA interactions. We examined the interaction of NC with DNA and RNA variants of non-labeled and thG-labeled (–)PBS and SL3 sequences. Firstly, through photophysical analysis we compared the folding of loop in RNA and DNA variants of (–)PBS and SL3 sequences. In (–)PBS, we observed that both the DNA and RNA variants have similar penta-nucleotide loop folding. In SL3_RNA, thG10 residue is partially stacked with neighboring bases while thG12 residue is largely solvent exposed, that is in line with the NMR and MD calculation. However, for SL3_DNA it appears that the tetra-loop is rigid, and both the thG10 and thG12 residues are mostly solvent exposed. Further, we determined the binding affinity using ITC and fluorescence titration. They revealed that NC binds to RNA and DNA copies of labeled and non-labeled (–)PBS with similar binding affinities. Interestingly, the photophysical investigation of thG-labeled (–)PBS sequences showed a two-fold increase in quantum yield associated with a parallel change in lifetimes distribution upon NC interaction. This suggests that the thG conformational changes due to its stacking with Trp37 appear similar in both DNA and RNA copies of (–)PBS. However, we only observed NC-induced fluorescence change for thG at position 10 of SL3_RNA, in line with its stacking with Trp37 of NC. In contrast, the absence of changes in the fluorescence properties of thG at position 12 in the NC-SL3_RNA complex, suggests that its interaction with Phe16 is too limited to perturb its photophysical properties. The absence of thG fluorescence changes in both positions 10 and 12 in the NC-SL3_DNA complex suggests that none of the positions may be involved in key interactions with NC. It may be speculated that in this case, NC binds to another part of the loop. We confirmed that the interaction of the Trp37 residue with guanines was critical for the formation of complexes with both RNA and DNA variants of PBS and SL3. This work is in progress and it would be needed to resolve the NMR structure of the PBS_RNA and SL3_DNA complexed with NC to confirm the change in orientation of the NC and characterize the binding contacts formed by sugar moiety.

Lastly, while participating within THINPAD project, the inhibitors of NC protein were screened using NCinh assay and their activity were examined by determining IC$_{50}$ values. Starting from parent inhibitor having a IC$_{50}$ of 170 µM, Hit-to-lead optimization resulted in

discovery of positive compounds with IC_{50} less than 10 µM and the best inhibitor has IC_{50} of 1.5 µM. In addition, taking into consideration the site specific binding of NC, we developed a NC inhibition assay using site-specifically labeled thG at position 7 of (–)PBS. The highly active compounds from NCinh assay were tested to validate the assay. The positive activity of WDO, as a control, showed that this assay can be used for screening of inhibitors, provided inhibitors are non-emissive and non-interacting with thG. Nevertheless, this work is in perspectives for further development.

6. Perspectives

We identified the lifetimes of both the thG tautomers through lifetime analysis of free probe in buffer. In less polar solvents (methanol, ethanol), we observed an additional short and unstable lifetime component (<100 ps) besides two long-lived lifetime components representing thG-H1 and thG-H3 tautomers. We found a growth component when lifetimes are collected at longer emission wavelengths (> 500 nm). This rise lifetime component was matching with the short lifetime component observed at the blue emission band. This indicated that some excited state reaction/conversion happens when one of the form is excited. It was difficult to assign this short lifetime component as no evidence of other tautomers was found in steady state spectra. We have not seen this kind of short lifetime in any other solvents except in the above mentioned polar protic solvents. However, we could not determine this third short lifetime component in more polar protic medium, aqueous solution. We speculated that the reason for the missing short lifetime component may be due to its ultrafast behavior below the time resolution of our instrumental setup (< 70 ps). Thus, femtosecond up-conversion should be used to identify short-lived lifetimes component of thG probe in buffer and other solvents. To identify the possibility of proton transfer reaction occurring in ground state or excited state, the photophysics of thG tautomer should be explored at different pH using steady state and time resolved spectroscopy. All these results together in exploring the microenvironment sensitivity of free thG nucleoside will help us better explain the behavior of thG as substituted in ODNs. Importantly, the unexplained huge variation in lifetimes (11 ns to 28 ns) of thG7(–)PBS depending on its opposite base will be benefitted. In addition, photophysical investigation will be required to examine the role of neighboring adjacent residues and opposite base on the fluorescence lifetimes of thG as substituted in single stranded (–)PBS, in its double stranded, and mismatched duplexes. To facilitate better understanding of the experimental results, quantum mechanical calculations will be performed to find the possible reasons for radiative and non-radiative decay channels of thG in different media and in ODNs. We investigated the interaction of NC to RNA and DNA variants of (–)PBS and SL3. We were able to support the experimental results through molecular dynamic simulations that used the available NMR resolved structures of NC complexed with (–)PBS_DNA and SL3_RNA. Molecular dynamic simulations helped in characterizing the binding contacts established by NC towards its labeled and non-labeled ODNs. However, in order to confirm the change in orientation of NC, NMR structures of (–)PBS_RNA and SL3_DNA complexed with NC should be resolved. In this way we will be able to compare the orientation of NC and its binding contacts with sugar moieties (ribose and deoxyribose) within both (–)PBS and SL3 complexes, and thus, better interpret our fluorescence and ITC data.

7. Résume en Français

Caractérisation et ciblage de la reconnaissance dynamique de Trp37-G lors de l'interaction de la protéine NCp7 de HIV-1 avec des acides nucléiques.

Introduction

Le contrôle de l'infection du VIH par la tri-thérapie a réussi à faire baisser le taux de la charge virale jusqu'à un niveau indétectable. Cependant, pour cause de mutations, des sous-types résistants du VIH-1 apparaissent au cours des essais cliniques. Des médicaments ciblant les protéines les plus conservées du VIH-1 sont donc requis pour surmonter cette résistance. Parmi les protéines du VIH, la protéine de la nucléocapside (NCp7) est considérée comme une cible potentielle pour le développement de médicaments en raison de ses fonctions pléiotropes tout au long du cycle viral et de son caractère réfractaire aux mutations. Exprimé en tant que domaine de la polyprotéine Gag et ensuite clivée par la protéase du VIH, la NCp7 est une protéine basique qui contient un ou deux motifs hautement conservés $CX_2CX_2HX_4C$ (CHCC) appelés doigts de zinc (ZFs) qui lient Zn avec une affinité de l'ordre du picomolaire. La structure 3D des protéines NC natives a été résolue par résonance magnétique nucléaire (RMN) (Figure 78a). Le domaine de la NC sous ses diverses formes (NC15, NCp9, Gag, NCp7) joue de nombreux rôles au cours du cycle de réplication du virus. Cela inclut notamment l'encapsidation sélective de l'ARN génomique viral non épissé et le chaperonnage des brins d'acide nucléique au cours de la transcriptase inverse. En utilisant la même technique de RMN, la structure de NC avec des séquences d'ARN et d'ADN a été résolue (Figure 78). En comparant les structures des complexes ainsi résolus, certaines caractéristiques communes ont été mises en évidence. La liaison de NC aux acides nucléiques (NA) est régie par des interactions électrostatiques, hydrophobes et hydrogène. Des changements structuraux importants surviennent au niveau du plateau hydrophobe de NC comprenant les résidus V13, F16, T24, A25, W37, Q45 et M46. Par ailleurs, dans les complexes tant avec l'ARN que l'ADN, l'activité chaperonne de l'acide nucléique implique une interaction dynamique entre le résidu W37 de ZF2 et des guanines non appariées dans les motifs TG, UG, TGG ou GXG présents dans les boucles internes ou apicales ou dans les domaines simples brins de l'ARN génomique et ses copies d'ADN.

Une telle interaction est critique lors du second saut de brin de la transcription inverse où NC déstabilise la boucle du brin négatif du site d'initiation de la transcription inverse (–)PBS et induit son hybridation avec le brin complémentaire (+)PBS. L'interaction avec le domaine en doigt de zinc de NCp7 mène à l'étirement et à la déstabilisation de la boucle (Figure

78b). Les nucléobases de la boucle sont exposées, ce qui permet d'initier leur hybridation avec les nucléobases complémentaires de la boucle de (+)PBS. Le mécanisme d'hybridation a été étudié par marquage fluorescent des extrémités du (+)PBS (fluorophore-inhibiteur) ou en substituant l'une des nucléobases naturelles de la boucle par un analogue fluorescent de nucléoside, la 2-aminopurine (2Ap). Il a été démontré qu'en absence de NC, l'hybridation (–)/(+)PBS procède via les extrémités protrudentes, alors que la réaction induite par NC s'effectue via une réaction boucle-boucle. Dans les deux cas, il a été proposé que la réaction d'hybridation implique deux voies distinctes, avec une étape de pré-équilibre rapide conduisant à un complexe intermédiaire (IC) partiellement hybridé, suivi d'une conversion cinétiquement limitante vers le duplex étendu (ED). Bien que ces études donnent un aperçu essentiel sur les voies d'hybridation, elles ne sont pas parvenues à procurer une information complète concernant les constantes de vitesse associées à ces réactions. Les résultats basés sur le (+)PBS doublement marqué ont été insensibles à l'hybridation des boucles, alors que la faible brillance de 2Ap a nécessité une forte concentration en acides nucléiques conduisant à des cinétiques trop rapides pour être résolues. Ainsi, l'un des objectifs de ma recherche fût de mesurer les cinétiques d'hybridation de (–)/(+)PBS à l'aide de nouveaux analogues fluorescents de nucléosides afin de déterminer l'ensemble des constantes cinétiques.
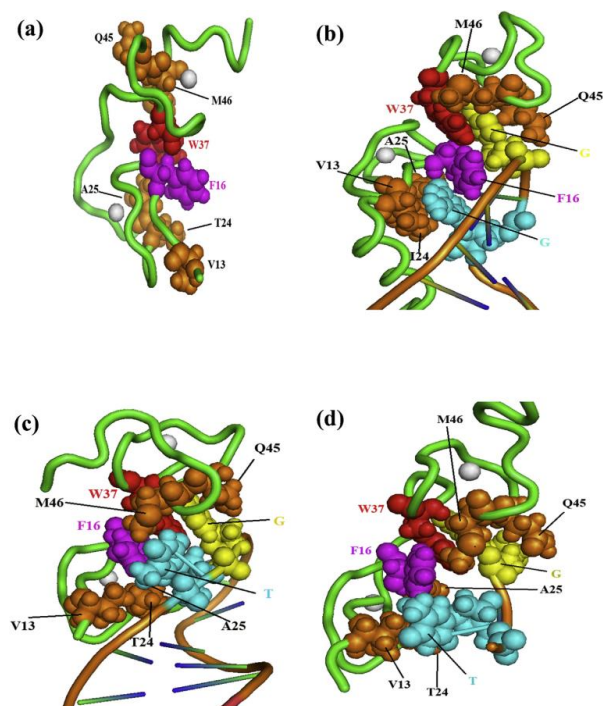
Figure 78 Structures tridimensionnelles de la NC de VIH-1 sous forme libre ou sous forme liée à différentes séquences d'ARN et d'ADN. Le squelette de la protéine NC est présenté en vert et les deux atomes de zinc liés par la tétrade CCHC dans chaque doigt

de zinc sont indiqués par des sphères blanches. Les résidus hydrophobes de la NC libre (a) et les résidus en contact avec les acides nucléiques sont également représentés par des sphères. Les résidus aromatiques F16 et W37 sont en magenta et rouge, respectivement, alors que les autres résidus hydrophobes de ZF1 (V13, T24, A25) et ZF2 (Q45, M46) sont en orange. Les deux nucléotides en interaction étroite avec les chaînes latérales de la protéine sont en bleu et jaune en (b) : (complexe NC-SL3 ARN, en (c) : (complexe NC-(–)PBS ADN) et en (d) : (complexe NC-mini TAR ADN).

L'interaction dynamique de NC avec les acides nucléiques joue également un rôle clé lors de l'encapsidation du génome viral dans la phase tardive du cycle viral du VIH. En tant que domaine de Gag, NC interagit avec les motifs SL1, SL2 et SL3 de la séquence d'encapsidation ψ de l'ARN génomique. Parmi ces motifs, SL3 présente un intérêt particulier en raison de sa structure hautement conservée dans les souches de VIH-1 et de sa capacité à gouverner la reconnaissance et l'encapsidation. Une distinction importante entre les complexes NC-ARN et NC-ADN réside dans l'orientation opposée de ZF1 et ZF2 le long de l'acide nucléique dans les complexes NC-ARN comparés aux complexes NC-ADN. Il a été proposé que le sucre joue un rôle clé dans la détermination de cette orientation. Dans les complexes ADN-NC, les atomes C1' et C2' du sucre forment des contacts hydrophobes avec les résidus hydrophobes (T24, F16, Q45, M46, W37) de NC. Au contraire, dans les complexes NC-ARN, les atomes O4', C4' et C5' des sucres sont en contact étroit avec Phe16, alors que les atomes C2' portant un groupement OH, sont loin des résidus protéiques. Ainsi, en raison de l'hydroxyle supplémentaire à la position C2' des riboses comparé aux déoxyriboses, les contacts hydrophobes des sucres avec la protéine diffèrent fortement dans les deux types de complexes. Le deuxième objectif de ma thèse fût de mieux comprendre les différences dans les modes de liaison adoptées par NC pour les séquences d'ARN et d'ADN. À cette fin, nous avons caractérisé comparativement l'interaction de NC avec les séquences d'ADN et d'ARN de PBS, ainsi qu'avec les versions ADN et ARN de SL3.

Résultats et discussion

Afin de suivre les processus dynamiques sur une échelle de temps s'étendant de la picoseconde à l'heure, les techniques basées sur la fluorescence sont fortement adaptées. Toutefois, ces techniques reposent sur la disponibilité de sondes fluorescentes à même de suivre ces interactions avec sensibilité et sélectivité sans les perturber. Etant fluorescent, le Trp37 était un candidat idéal à cette fin. Malheureusement, son empilement avec

151

les Gs inhibe presque totalement sa fluorescence, de sorte que peu d'informations peuvent être extraites concernant le processus de reconnaissance et sa dynamique. Une autre alternative aurait été de suivre les résidus G, mais ceux-ci ne sont pas fluorescents. En recherchant de meilleurs substituts que la 2Ap, nous avons testé deux nucléotides fluorescents isomorphes récemment développés, la 2-Amino-6-(thiényl-2) purine (**s**) et la thiéno-désoxyguanosine (dthG ou thG pour le sucre ribose) qui ont été spécifiquement incorporés dans la séquence (–)PBS de l'ADN. En premier, nous avons étudié les propriétés photophysiques de **s** marqué en positions 6 et 7 de (–)PBS et nous les avons comparées aux mêmes séquences marquées par la 2Ap. Aux deux positions 6 et 7, nous avons observé un rendement quantique plus élevé des séquences de la tige-boucle de (–)PBS et de son duplex (–)/(+)PBS marquées par **s**. En outre, les propriétés de fluorescence résolue en temps de **s**, en comparaison avec 2Ap, sont caractérisées par une quantité moindre d'espèces noires et une durée de vie moyenne plus élevée. L'interaction de **s**6(–)PBS avec NCp7 entraine une augmentation du rendement quantique de fluorescence d'un facteur 2 par rapport à 2Ap6(–)PBS, montrant une meilleure sensibilité envers l'interaction avec les protéines. Cependant, aucune modification du rendement quantique n'a été observée lors de la liaison de NCp7 pour **s** en position 7, de sorte que cette sonde se révèle insuffisante pour étudier l'interaction spécifique de cette position avec le Trp37. Etant donné le progrès limité de **s** par rapport à 2Ap, son utilisation a été abandonnée.

Dans l'étape suivante, nous avons étudié la photophysique de thG sous deux formes, comme nucléoside libre ou incorporé dans la séquence (–)PBS. L'analyse photophysique de thG libre dans différents solvants a révélé l'existence de deux tautomères à l'état fondamental présentant un déplacement significatif des spectres d'absorption et d'émission. Par des calculs DFT-TD, les structures de ces deux formes ont été identifiées comme tautomères céto-énoliques : thG-H1 et thG-H3. Alors que les deux tautomères ont été observés dans l'eau et les solvants protiques, le tautomère thG-H1 dominait largement dans des solvants apolaires. En outre, nous avons également déterminé les durées de vie et les rendements quantiques associés à chaque tautomère dans différents solvants, afin de mieux comprendre la sensibilité au microenvironnement du marqueur une fois substitué dans des acides nucléiques. Incorporé en position 7 dans (–)PBS, les deux tautomères de thG peuvent être observés. Par contre, l'appariement Watson-Crick dans les duplexes (–)/(+)PBS favorise fortement le tautomère thG-H1. L'existence de ces tautomères et leur sensibilité à leur microenvironnement procure ainsi un canal d'information très utile pour étudier de manière

152

site-sélective les propriétés d'un résidu G substitué dans un oligonucléotide. Ces données ont été publiées dans Angewandte Chemie.

Nous avons ensuite effectué une étude comparative de la photophysique de thG et de 2Ap substitués en position 7 de la séquence (–)PBS. Nous avons observé que thG n'affecte pas le repliement de la boucle et que, contrairement à 2Ap, il préserve l'appariement des bases Watson-Crick dans le duplex (–)/(+)PBS, servant ainsi de parfait substitut au résidu Guanosine. En outre, il présente un rendement quantique de fluorescence beaucoup plus élevé et un déclin de fluorescence plus simple que la 2Ap, dans la tige-boucle (–)PBS comme dans le duplex (–)/(+)PBS. Nous pouvons donc affirmer que thG s'avère bien meilleur que 2Ap en vue de substituer des résidus G. Ces résultats ont été publiés dans la revue J Am Chem Soc. Par ailleurs, l'interaction avec NC(11–55) entraîne une augmentation de l'émission de thG7(–)PBS d'un facteur 2, ce qui souligne sa sensibilité aux changements de conformation lors de l'interaction NC/(–)PBS. En raison de sa parfaite analogie avec G et son extrême sensibilité à l'environnement, thG apparaît comme un outil idéal pour suivre fidèlement l'interaction dynamique de la boucle (–)PBS avec NC et l'hybridation (–)/(+)PBS en absence et en présence de la protéine NC.

En l'absence de NC, les cinétiques d'hybridation (–)/(+)PBS ont été traitées à l'aide d'une approche numérique avec le logiciel DynaFit pour obtenir les constantes de vitesse. Il a été confirmé que la réaction procède par un mécanisme en deux étapes impliquant la formation d'un complexe intermédiaire (IC). L'intensité de fluorescence de l'IC s'est révélée être environ 80 % de celle du duplex étendu (ED). Ceci suggère que la formation de l'IC découle non seulement de l'interaction des extrémités protrudentes, mais également d'une interaction boucle-boucle. En présence de la NC, les cinétiques atteignent le plateau final beaucoup plus rapidement qu'en son absence, notamment par une formation accélérée de l'IC, illustrant l'activité chaperonne de NC. En parallèle, nous avons aussi étudié la cinétique d'hybridation de (–)PBS avec le (+)PBS en substituant le résidu T9 de (–)PBS par la base azotée 2-thiényl-3-hydroxychromone (3HCnt). Les mécanismes des réactions (+)/(–) PBS obtenus en absence et en présence de NC en suivant les sondes 3HCnt et thG se sont révélés très consistants.

Afin d'approfondir le rôle du sucre dans les interactions NC-ADN et NC-ARN, nous avons étudié l'interaction de NC avec les versions ADN et ARN de (–)PBS et SL3. Notre première tâche fût d'obtenir les constantes de liaison respectives des différents complexes par titration en calorimétrie isotherme (ITC) et titration en fluorescence avec les séquences

ADN_(–)PBS et ARN_(–)PBS marquées par thG en position 7 et SL3_ADN et SL3_ARN marquées en position 10 et 12. Nous avons observé que pour les séquences non marquées, NC se lie avec une affinité légèrement supérieure à la copie ADN de (–)PBS par rapport à son homologue ARN. Une tendance similaire a aussi été observée pour les copies ADN et ARN marquées par thG, ce qui suggère que la substitution par thG n'affecte pas le processus de liaison. Fait intéressant, l'étude photophysique des séquences marquées avec thG a montré une augmentation d'un facteur 2 du rendement quantique associée à une modification de la distribution des durées de vie lors de l'interaction avec la NC. Ceci suggère que les changements de conformation de thG dus à son empilement avec Trp37 paraissent identiques dans les copies ARN et ADN de (–)PBS. L'étude de l'interaction de NC avec les séquences SL3 non marquées montre une légère préférence pour la séquence d'ARN. Pour les séquences SL3 marquées avec thG, les mesures ITC suggèrent que le remplacement de G par thG affecte quelque peu le processus de liaison avec NC. Enfin, nous n'avons observé de changements de fluorescence induit par NC que pour thG en position 10 de l'ARN SL3, en accord avec son empilement avec le Trp37 de NC. Au contraire, la non-modification des propriétés de fluorescence de thG pour le complexe ARN/NC-SL3 marqué en position 12 suggère que son interaction avec Phe16 est trop limitée pour perturber ses propriétés photophysiques. De même, la non-modification de fluorescence de thG dans le complexe NC-SL3 ADN marqué en positions 10 et 12 suggère qu'aucune de ces deux positions ne semble être critique pour l'interaction avec NC. On peut spéculer que dans ce cas, NC se lie à une autre partie de la boucle. Afin de caractériser davantage les modes de liaison adoptés par NC sur les séquences ARN et ADN de (–)PBS et SL3, des études de dynamique moléculaire (MD) sont en cours. Cela devrait, combiné avec nos données de fluorescence, nous permettre de mieux comprendre le rôle de l'interaction Trp-G dans la formation des différents complexes.

8. Appendix

A Brighter 2-Aminopurine Derivative to Investigate Protein-Nucleic Acid Interaction

1. Abstract

2-Amonopurine (**2Ap**) has been widely used for almost five decades but its underlying drawback of poor emission fluorescence observed as substituted in oligonucleotide sequence compel us to quest for other isomorphic fluorophore that can serve in not only adopting the natural conformation of adenine but also report biomolecular interaction. Herein, we describe the photophysical characterization of a newly synthesized adenine analogue, 2-Amino-6-(2-thienyl)purine (**s**), and compare its properties with **2Ap**, when substituted in single and double strand of primer binding site (–)PBS of HIV-1. Both the isomorphs were compared at two different positions by site specifically incorporating at T6 and G7 residues of (–)PBS. We further compared the interaction of (–)PBS with HIV-1 nucleocapsid protein, which is shown to promote the annealing of (–)PBS with complementary (+)PBS sequence during reverse transcription of HIV-1 life cycle. At both the positions, we found that **s-** and **2Ap-** substitution in (–)PBS to equally preserves the stability and represent comparable photophysical properties both in single and duplex strand. At position 6, as compared to **2Ap**6(–)PBS, **s**6(–)PBS showed better sensitivity towards binding with NC but poor fluorescence emission when hybridized with complementary (+)PBS. At position 7, fluorescence from **s**7(–)PBS was found insensitive towards binding with NC, whereas in duplex form slighty better fluorescence was seen comparable to **2Ap**7(–)PBS,

2. Introduction

Synthetic fluorescent nucleobase analogues are predominantly used for *in vitro* characterization of biomolecular interactions, as the major components in living systems are inherently non-emissive. The need to develop fluorescent nucleoside analogues arise due to the extremely low fluorescent quantum yield of natural nucleobases ($\sim 10^{-4}$) along with picosecond order of lifetimes (Peon and Zewail, 2001). The key attribute of fluorescent isomorphs is the ability to mimic structurally and functionally their natural counterparts. Since last five decades, 2-Aminopurine (**2Ap**) has served as an ideal fluorescent nucleobase analogue finding its application in molecular becon based assays (Martí et al., 2006), helicase activity (Raney et al., 1994), base flipping (Nakano et al., 2005), abasic site structure/dynamics (Rachofsky et al., 2001; Stivers, 1998), interaction of DNA with RNA polymerase (Huang et al., 1997), G

quadruplex structure and ligand binding (Kimura et al., 2007). These application are brought forward because of 2Ap's ability to form Watson-Crickbase pair with T/U, high quantum yield (Φ = 0.68 in water), absorption band (303 nm), minimal sensitivity to pH change, and its noteworthy sensitivity to environment polarity (Ward et al., 1969).
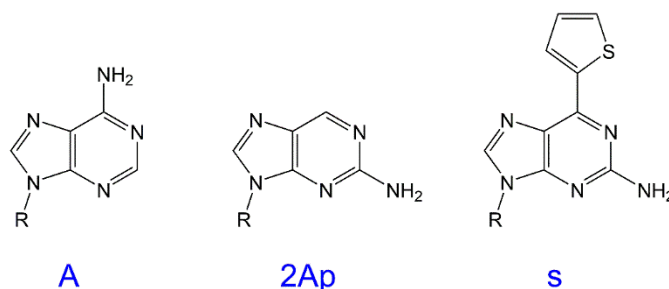
Figure 79. Structure of deoxyribo Adenine (**A**), 2-Amino purine (**2Ap**) and 2-Amino-6-(2-thienyl)purine (**s**).

The photophysical properties of any fluorescent nucleobase analogue incorporated in DNA or RNA are hampered by two key reasons, one arising due to the stacking of neighboring bases and second, by duplex formation with complimentary strands (Holz et al., 1998; Liu and Martin, 2001; Purohit et al., 2003). Likewise, photophysical properties of **2Ap** showed drawbacks where its quantum yield decreases to minimum (< 0.01) when incorporated in nucleic acid sequences. Although this makes it useful in developing assays, but importantly it displays poor sensitivity towards duplex formation or upon interaction with protein. Moreover, its fluorescence does not represent faithfully the actual conformation of the substituted residues. To overcome these limitation, Hirao and co-workers synthesized 2-Amino-6-(2-thienyl)purine (**s**) (Hikida et al., 2010; Hirao et al., 2006) (Figure 79), a derivative of **2Ap** having an additional thienyl ring, showed a red shifted absorption (355 nm) and strong emission (Φ = 0.41 in water) profile (Mitsui et al., 2007).
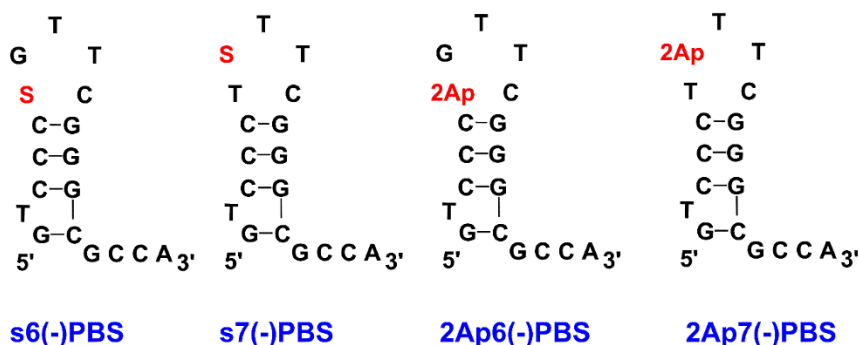
Figure 80. Substitution made on DNA sequence of HIV-1 primer binding site replacing T6 residue with s [**s**6(–)PBS] and **2Ap** [**2Ap**6(–)PBS]. Similarly, at position G7 by **s** [s7(–)PBS] and **2Ap** [**2Ap**7(–)PBS].

In HIV-1 virus life cycle, second strand transfer require annealing of minus stranded primer binding site (–)PBS, a 18-mer nucleic base Stem loop (SL) sequence, with its complimentary (+)PBS which is chaperoned by the nucleocapsid protein (NC) (S Bourbigot et al., 2008). NC is a 55 amino acid protein which destabilizes the loop of (–)PBS by flipping the thymine (T6) and guanine (G7) residues towards exterior environment and thus making them available for annealing with complementary (+)PBS sequence. This annealing mechanism was previously examined using the PBS sequences labeled at its 5' and 3' ends, by rhodamine 6G (Rh6G) and Dabcyl, a fluorescence quencher, respectively (Ramalanjaona et al., 2007) but it was limited as the PBS species doubly labeled by Rh6G and Dabcyl were insensitive to the annealing of the loops. Another approach was made using labeled PBS loop with the **2Ap** (Julien Godet et al., 2011), but this also provided limited information due to the poor brightness of **2Ap** (Guest et al., 1991; Jones and Neely, 2015; Nag et al., 2006; Stivers, 1998). These approaches revealed a single rate constant from kinetic measurement, thus a complete picture showing all the rate constants could be extracted due to limitation of probes.

In order to find a better replacement of **2Ap** to monitor annealing kinetics, herein, we compared the photophysical properties of **s** and **2Ap** as substituted at T6 and G7 residues of (–)PBS (Figure 80) and its duplex structure (–)/(+)PBS, and its sensitivity upon interaction with NC. As a free probe, **s** (Φ=0.41) has a lower quantum yield than **2Ap** (Φ=0.68), but as substituted in single strand (–)PBS and duplex (–)/(+)PBS, we observed improved quantum yields for **s** than **2Ap** labeled sequences. Moreover, it exhibits simpler lifetime decays with relatively lower amount of dark species than that of **2Ap**. However, as substituted at position 6, fluorescence of **s**-labeled (–)PBS was quenched due to stacking with neighboring residues,

which is comparable **2Ap**. In addition at position 7, **s**-labeled (–)PBS showed insensitivity towards binding with NC.

3. Results and Discussion

3.1. Thermodynamic Stability

We examined the thermodynamic stability of non-labeled and labeled (–)PBS in single stranded and duplex structure by determining their melting temperatures. This require monitoring the absorbance at 260 nm over varying temperature (20-90 °C) (Table 10). In stem loop sequences, at position 6, we observed similar melting temperatures for **s6-** and **2Ap**-labeled (–)PBS (49 °C) with their native non-labeled (50 °C) counterpart. At position 7, small deviation of 1-2 °C was seen for **s7**(–)PBS and **2Ap7**(–)PBS (50 °C and 48 °C respectively) as compared to the non-labeled (–)PBS. Thus at both the position we observe that substitution of **s** does not disturb the thermodynamics stability of the (–)PBS hairpin.

Table 10. Melting temperatures of single and double stranded labeled ODN measured at 260nm. The concentration of ODNs was 1µM in 20 mM Cacodylate buffer with 150mM NaCl.

ODN	T_m (°C)
(–)PBS	50±1
s6(–)PBS	49±1
s7(–)PBS	50±1
2Ap6(–)PBS	49±1
2Ap7(–)PBS	48±1
(–)/(+)PBS	68±1
s6(–)/(+)PBS	67±1
s7(–)/(+)PBS	60±1
2Ap6(–)/(+)PBS	66±1
2Ap7(–)/(+)PBS	60±1

In duplex form, both **s6**(–)/(+)PBS and **2Ap6**(–)/(+)PBS (67 °C and 66 °C respectively) showed a minor 1-2 °C deviation in melting temperature compared to non-labeled (–)/(–)PBS (68 °C) suggesting the formation of stable duplex. However, the effect of mismatch formation with complimentary C residue on the stability of duplex can be seen from melting temperatures

of both the **s7**(–)/(+)PBS and **2Ap7**(–)/(+)PBS, that deviated from non-labeled (–)/(+)PBS by a difference of 8 °C. The variability in thermal stability of **2Ap7**(–)/(+)PBS arise due to the 2Ap-C base pair which is stabilized by two hydrogen bonds instead of three (Goodman and Ratliff, 1983). And with likely the same reason, s7-C mismatch results in the deviation of melting temperatures compared (–)PBS.

3.2. Photophysical Characterization of s and 2Ap labeled (–)PBS Sequences

Photophysical characterization of **s** and **2Ap** as a free probe, as well as substituted at position 6 and 7 of single and double stranded (–)PBS are listed in Table S1 and Figure 81. As **s** being a **2Ap** derivative with an additional thienyl group, it results in slightly different photophysical properties. As a free nucleoside, **s** showed larger stokes shift with an absorption (348 nm) and emission maxima (434 nm) shifted to longer wavelength relative to **2Ap** (305 nm, 365 nm). This structural modification also affected the fluorescence quantum yield of free probe **s** (0.41) which is lower than that of **2Ap** (0.68).

As substituted in single stranded (–)PBS sequence, at position 6 they both possess a fairly comparable and low quantum yield (0.02). Poor quantum yields for both the probes results from effective quenching due to the stacking with neighboring G residue (S Bourbigot et al., 2008; Julien Godet et al., 2011). It was confirmed as at position 7, the quantum yield of **s**7(–)PBS and **2Ap7**(–)PBS gained 9-fold and 4-fold, respectively, as compared to their respective position 6 quantum yield. Moreover, **s**7(–)PBS (0.18) has more than two-fold higher quantum yield as compared to **2Ap7**(–)PBS (0.08).
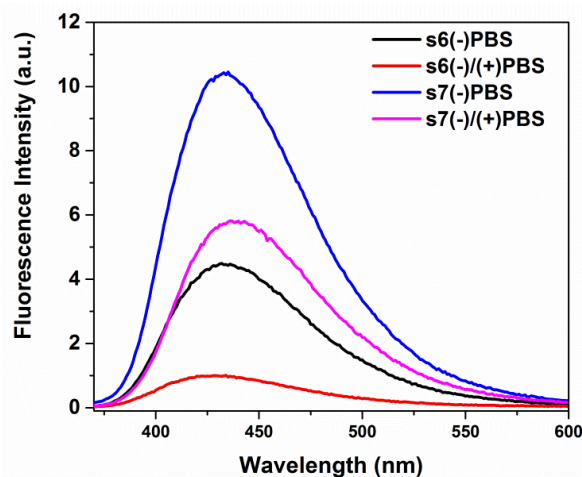
Figure 81. Steady state fluorescence emission of **s6**(–)PBS (3 μM, black) , **s6**(–)/(+)PBS (3 μM red), **s7**(–)PBS (1 μM, blue) and **s7**(–)/(+)PBS (1 μM, pink).

In (–)/(+)PBS duplex form, at position 6, we observed poor quantum yield for both the **s** and **2Ap** labeled (–)PBS (0.006) suggesting effective quenching due stacking with neighboring bases. While at position 7, fivefold higher quantum yield is obtained for **s7**(–)/(+)PBS (0.10) as compared to **2Ap7**(–)/(+)PBS (0.02). A relatively higher quantum yield of **s**-labeled than **2Ap**-labeled (–)PBS at position 7 for both single and duplex form suggest that the additional thienyl group is believed to protect **s** from quenching from neighboring nucleic acid residues. Especially, in duplex form it appears that due to s7-C mismatch, **s**7 residue skips the conventional stacking and likely to be solvent exposed. Based on these results as substituted at G7 residue, **s** can be considered as a slightly better probe for monitoring annealing reactions.

Table 11. Photophysical parameters of 2Ap and s labeled (–)PBS.

	QY	α_0	τ_1(ns)	α_1	τ_2(ns)	α_2	τ_3(ns)	α_3	τ_4(ns)	α_4	<τ>(ns)
s6(–)PBS	0.028	0.48	0.20	0.38	1.4	0.07	3.9	0.07	-	-	0.81
+ NC	0.094	0.23	0.36	0.32	1.9	0.28	4.5	0.17	-	-	1.84
s6(–)/(+)PBS	0.006	0.70	0.11	0.26	0.6	0.02	3.6	0.02	-	-	0.30
s7(–)PBS	0.18	0.14	0.41	0.15	-	-	3.7	0.71	-	-	3.14
+ NC	0.18	0.10	0.36	0.22	-	-	3.8	0.68	-	-	3.02
s7(–)/(+)PBS	0.10	0.36	0.56	0.08	-	-	2.6	0.56	-	-	2.33
2Ap6(–)PBS	0.02	0.85	0.08	0.08	0.62	0.02	3.5	0.02	8.0	0.03	2.04
+ NC	0.04	0.77	0.18	0.10	1.01	0.05	3.5	0.04	8.2	0.05	2.56
2Ap6(–)/(+)PBS	0.006	0.79	0.08	0.17	0.48	0.02	2.2	0.01	5.1	0.01	0.43
2Ap7(–)PBS	0.08	0.47	0.14	0.22	0.72	0.07	2.5	0.13	7.2	0.11	2.27
+ NC	0.14	0.38	0.29	0.25	-	-	2.2	0.15	7.5	0.22	3.36
2Ap7(–)/(+)PBS	0.02	0.84	0.25	0.07	0.80	0.03	2.3	0.03	6.3	0.03	1.83

τ_i are the fluorescence lifetimes (ns), α_i their amplitudes. The amplitude α_o of the dark species, as well as the amplitudes of the various lifetimes were calculated as described in the Material and Methods section. <τ> is the mean fluorescence lifetime (ns). Excitation and emission wavelengths were 315 and 370 nm for **2Ap** substituted sequences; 350 and 450 nm for **s** substituted sequences, respectively. SDs for the lifetimes and amplitudes are <20%. SDs for QY is < 10 %.

Next, we investigated the heterogeneity of the **s-** and **2Ap-**labeled (–)PBS sequences using time-resolved fluorescence spectroscopy (Table 11 and Figure 82). At position 6 for **s**-

labeled (–)PBS sequence, the lifetime decay was fitted with three exponential, suggesting that **s6-** experience three conformational states. Total contribution from the short-lived lifetime components (τ_1 and τ_2) and dark species (α_o) attribute more than 90 percent. Such effective dynamic quenching by neighboring residues has previously been reported for **2Ap**6(–)PBS, where the quenching arise through charge transfer mechanism (Fiebig et al., 2002; J Godet et al., 2011; Larsen et al., 2001; O'Neill and Barton, 2002; Wan et al., 2000). With likely the same mechanism, the dynamic quenching was observed for **s6-** by its neighbor residues and also results in poor fluorescence quantum yield. The lifetime decay of **2Ap**6(–)PBS was quite complex with presence of four discrete lifetime components. The lifetime components are in good agreement with the earlier reported values (J Godet et al., 2011). Similar to **s6**(–)PBS, majority of **2Ap**6(–)PBS population (~ 97%) comprising short-lived lifetime components (τ_1, τ_2 and τ_3) and dark species (α_o) represents the dynamically quenched conformations.
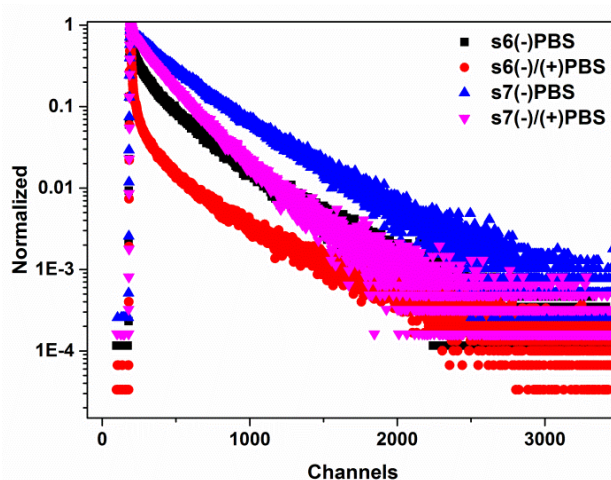
Figure 82. Normalized lifetime decay of **s6**(–)PBS (black), **s6**(–)/(+)PBS (red), **s7**(–)PBS (blue) and **s7**(–)/(+)PBS (magenta).

In duplex form, at position 6, the average lifetimes for both the **s-** and **2Ap-**labeled (–)PBS sequences reduces to minimum, 0.30 ns and 0.43 ns, respectively. The effect of quenching through neighboring residues become stronger as the total contribution from quenched species (dark species and short-lived lifetime species) increased to ~ 98 percent and significant decrease in the contribution from long-lived lifetimes. This further explained the poor quantum yield obtained for both the **s-** and **2Ap**-labeled (–)PBS in duplex form.

At position 7, within stem loop sequences an improvement in lifetime distribution was seen for both **s7**(–)PBS and **2Ap7**(–)PBS as compared to position 6. We observed only two lifetime components for **s7**(–)PBS compared to four components of **2Ap7**(–)PBS, suggesting

better conformational stability for **s-**labeled (–)PBS. The three short-lived lifetimes of **2Ap**7(–)PBS represents the dynamically quenched conformations while the long-lived lifetimes component (8 ns) represents conformation this extrahelical placed away from the quencher (Fiebig et al., 2002; Larsen et al., 2001; O'Neill and Barton, 2002; Wan et al., 2000). The lifetimes distribution of **2Ap**7(–)PBS appeared similar to the previously reported values showing the reproducibility of the work (Sholokh et al., 2015). It is very probable that the two short-lifetime components of **s**7(–)PBS represents the dynamically quenched conformation arising due to the flexibility of loop. Altogether, **s**7(–)/(+)PBS showed two conformations lesser than **2Ap**7(–)/(+)PBS, and hence slightly better average lifetime.

3.3. Interaction of NC on the Dynamics of s- and 2Ap- labeled (–)PBS

To investigate the dynamic changes occurring in the stem loop of (–)PBS upon interaction with NC, we examined the binding parameters of **s-** and **2Ap-** labeled (–)PBS sequences upon NC addition. Instead of full length NC, we used truncated NC(11–55) peptide as the missing basic N-terminal domain is responsible for the nucleic acid aggregation properties (Stoylov et al., 1997). Nevertheless, NC(11–55) still possess the zinc finger domain that is required for specific nucleic acid binding and destabilizing properties of NC (Cruceanu et al., 2006b; Darlix et al., 2007, 2011; Judith G Levin et al., 2010; Levin et al., 2005; Thomas and Gorelick, 2008; Tisné et al., 2004).

Conformational changes occurring in the loop of (–)PBS upon interaction with NC were followed through the fluorescence changes arising in **s-** and **2Ap-**labeled (–)PBS. At position 6, binding of NC to **s6**(–)PBS and **2Ap6**(–)PBS did not modify their wavelength of fluorescence emission maxima, but increased the fluorescence emission by three-fold and two-fold, respectively. These changes in fluorescence of **s6**(–)PBS and **2Ap6**(–)PBS are in line with the chaperoning activity of NC that exposes T6 residues to the exterior solvent thus reducing the level of quenching from neighboring residues without changing the environment (J Godet et al., 2011). Further, the lifetime analysis revealed that the increase in quantum yield of **s6**(–)PBS arise mainly from the decreased population of dark species which was benefitted largely by long-lived lifetime component (τ_3). Parallel changes in the fluorescence lifetime of **2Ap6**(–)PBS were observed as compared to **s6**(–)PBS. However, a higher average fluorescence lifetime is seen for **2Ap6**(–)PBS as its long-lived lifetime (τ_4) is significantly higher than that of **s6**(–)PBS (τ_3), thus benefitting more from the decrease in dark species. These changes in lifetime of **s6**(–)PBS and **2Ap6**(–)PBS suggests that NC disrupts the stacking of **2Ap** and **s**

with its neighboring residues, which is in line with the NMR resolved structure of NC with (–)PBS (S Bourbigot et al., 2008).
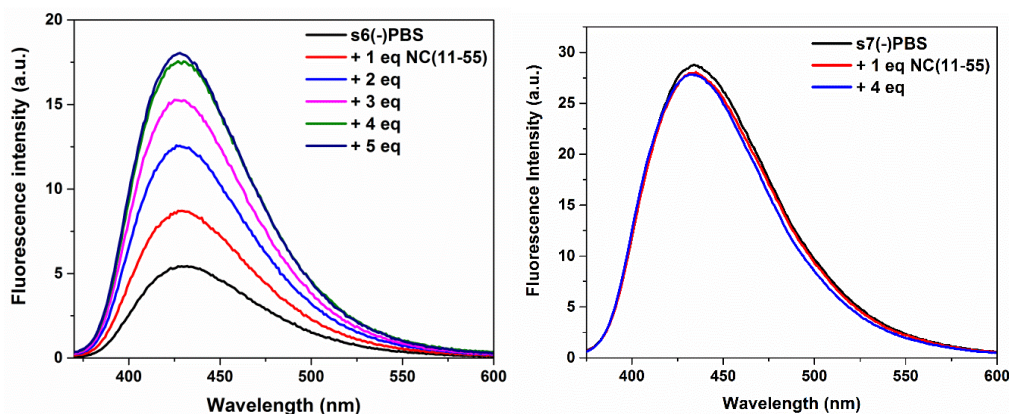
Figure 83. Fluorescence titration of s6(–)PBS (left) and s7(–)PBS (right) with NC(11–55)

At position 7, contrarily, **s7**(–)PBS does not show fluorescence change upon interaction with NC(11–55) (Figure 83). This could possibly arise either due to the loss of key Trp37-G recognition or due to absence of a strong quencher in the vicinity, thereby binding of NC(11–55) was unable to modify its fluorescence. However, we observed a two fold increase in its quantum yield of **2Ap**7(–)PBS upon interaction with NC which was also seen with the increase in mean lifetime arising mainly due to decrease in population of dark species.

4. Conclusions

In this study, we investigated the photophysical characteristics of a newly developed fluorescent nucleobase analogues **s** which was site specifically substituted at position T6 and G7 residue of (–)PBS in HIV-1 model, and further compared it to **2Ap** labeled sequences at similar position. The photophysical properties were compared in both stem loop (–)PBS and a duplex form of (–)PBS. We also examined the sensitivity conformational changes arising in **s** at both the positions of (–)PBS upon binding with NC(11–55).

The substitution of **s** at both the positions in (–)PBS stem loop conserves the stability of the loop as the melting temperature remains similar to the native sequences. Whereas, in duplex structures both the **s7**(–)/(+)PBS and **2Ap**7(–)/(+)PBS shows deviation in melting temperature as compared to native (–)/(+)PBS, due to unstable W-C base pairing arising for s-C mismatch. Twice higher fluorescence quantum yield is obtained for **s**7(–)PBS in comparison to **2Ap**7(–)PBS, even though the quantum yield of free probe **s** is lower than **2Ap.** Moreover in duplex structures, **s7**(–)/(+)PBS is a better substitute than **2Ap**7(–)/(+)PBS due to fivefold

higher quantum yield. Consequently, we have simple fluorescence lifetime decay and lesser amount of dark species for **s** labeled sequences with comparatively lesser lifetime component, suggesting the ability of **s** to adopt stable conformation in single and duplex structures. Especially, more than threefold increase in fluorescence of **s**6(–)PBS on NC(11–55) interaction makes it worthy to be used as a successful probe for establishment of fluorescence assay. However, at position 6, duplex of **s**6(–)/(+)PBS shows poor fluorescence emission comparable to **2Ap**. In addition at position, **s**7(–)PBS is insensitive towards binding with NC protein. Thus, **s** cannot be used a suitable probe to monitor annealing kinetics.

Supplementary Information

Table 12. Spectral parameters for **s** and **2Ap** probes in free from and when incorporated in single and double stranded PBS.

ODN	λ_{abs} (nm)	λ_{em} (nm)	Φ^1
s[a]	348	434	0.41
s6(–)PBS	353	434	0.028
s7(–)PBS	353	434	0.18
s6(–)/(+)PBS	357	428	0.006
s7(–)/(+)PBS	357	436	0.10
2Ap6	304	365	0.68
2Ap6(–)PBS	313	365	0.024
2Ap7(–)PBS	313	365	0.07
2Ap6(–)/(+)PBS	317	365	0.006
2Ap7(–)/(+)PBS	317	365	0.02

[1]standard deviation in quantum yield is less than 10%

9. References

Abbink, T.E.M., Berkhout, B., 2003. A Novel Long Distance Base-pairing Interaction in Human Immunodeficiency Virus Type 1 RNA Occludes the Gag Start Codon. J. Biol. Chem. 278, 11601–11611. doi:10.1074/jbc.M210291200

Abbink, T.E.M., Ooms, M., Haasnoot, P.C.J., Berkhout, B., 2005a. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. Biochemistry 44, 9058–9066. doi:10.1021/bi0502588

Abd El-Wahab, E.W., Smyth, R.P., Mailler, E., Bernacchi, S., Vivet-Boudou, V., Hijnen, M., Jossinet, F., Mak, J., Paillart, J.-C., Marquet, R., 2014. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. Nat. Commun. 5, 4304. doi:10.1038/ncomms5304

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

Alemán, E.A., De Silva, C., Patrick, E.M., Musier-Forsyth, K., Rueda, D., 2014. Single-molecule fluorescence using nucleotide analogs: A proof-of-principle. J. Phys. Chem. Lett. 5, 777–781. doi:10.1021/jz4025832

Alfadhli, A., Barklis, E., 2014. The roles of lipids and nucleic acids in HIV-1 assembly. Front Microbiol 5, 253. doi:10.3389/fmicb.2014.00253

Amarasinghe, G.K., De Guzman, R.N., Turner, R.B., Chancellor, K.J., Wu, Z.R., Summers, M.F., 2000. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. J. Mol. Biol. 301, 491–511. doi:10.1006/jmbi.2000.3979

Anderson, J., Schiffer, C., Lee, S.K., Swanstrom, R., 2009. Viral protease inhibitors. Handb Exp Pharmacol 85–110. doi:10.1007/978-3-540-79086-0_4

Andreatta, D., Pérez Lustres, J.L., Kovalenko, S.A., Ernsting, N.P., Murphy, C.J., Coleman, R.S., Berg, M.A., 2005. Power-law solvation dynamics in DNA over six decades in time. J. Am. Chem. Soc. 127, 7270–7271. doi:10.1021/ja044177v

Archin, N.M., Sung, J.M., Garrido, C., Soriano-Sarabia, N., Margolis, D.M., 2014. Eradicating HIV-1 infection: Seeking to clear a persistent pathogen. Nat. Rev. Microbiol. doi:10.1038/nrmicro3352

Arold, S., Franken, P., Strub, M.P., Hoh, F., Benichou, S., Benarous, R., Dumas, C., 1997. The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. Structure 5, 1361–1372.

Arts, E.J., Hazuda, D.J., 2012. HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med 2, a007161. doi:10.1101/cshperspect.a007161

Athavale, S.S., Ouyang, W., McPike, M.P., Hudson, B.S., Borer, P.N., 2010. Effects of the Nature and Concentration of Salt on the Interaction of the HIV-1 Nucleocapsid Protein with SL3 RNA. Biochemistry 49, 3525–3533. doi:10.1021/bi901279e

Augustyn, K.E., Wojtuszewski, K., Hawkins, M.E., Knutson, J.R., Mukerji, I., 2006. Examination of the premelting transition of DNA A-tracts using a fluorescent adenosine analogue. Biochemistry 45, 5039–5047. doi:10.1021/bi0518343

Avilov, S. V., Boudier, C., Gottikh, M., Darlix, J.L., Mély, Y., 2012. Characterization of the inhibition mechanism of HIV-1 nucleocapsid protein chaperone activities by methylated oligoribonucleotides. Antimicrob. Agents Chemother. 56, 1010–1018. doi:10.1128/AAC.05614-11

Ball, V., Maechling, C., 2009. Isothermal Microcalorimetry to Investigate Non Specific Interactions in Biophysical Chemistry. Int. J. Mol. Sci. 10, 3283–3315. doi:10.3390/ijms10083283

Barraud, P., Paillart, J.C., Marquet, R., Tisne, C., 2008. Advances in the structural understanding of Vif proteins. Curr HIV Res 6, 91–99.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (80-.). 220, 868–871.

Bartel, D.P., Zapp, M.L., Green, M.R., Szostak, J.W., 1991. HIV-1 rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell 67, 529–536. doi:10.1016/0092-8674(91)90527-6

Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R., Rosch, P., 1995. Structural studies of HIV-

1 Tat protein. J Mol Biol 247, 529–535. doi:10.1006/jmbi.1995.0158

Bazzi, A., Zargarian, L., Chaminade, F., Boudier, C., De Rocquigny, H., René, B., Mély, Y., Fossé, P., Mauffret, O., 2011. 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. 39, 3903–3916. doi:10.1093/nar/gkq1290

Bazzi, A., Zargarian, L., Chaminade, F., de Rocquigny, H., René, B., Mély, Y., Fossé, P., Mauffret, O., 2012. Intrinsic nucleic acid dynamics modulates HIV-1 nucleocapsid protein binding to its targets. PLoS One 7, 1–12. doi:10.1371/journal.pone.0038905

Belfetmi, A., Zargarian, L., Tisné, C., Sleiman, D., Morellet, N., Lescop, E., Maskri, O., René, B., Mély, Y., Fossé, P., Mauffret, O., 2016. Insights into the mechanisms of RNA secondary structure destabilization by the HIV-1 nucleocapsid protein. Rna 22, 506–517. doi:10.1261/rna.054445.115.506

Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J.P., Ficheux, D., Roques, B., Darlix, J.L., Mély, Y., 2003. 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. 328, 95–108. doi:10.1016/S0022-2836(03)00244-4

Berg, J.M., 1986. Potential metal-binding domains in nucleic acid binding proteins. Science (80-.). 232, 485–487. doi:10.1126/science.2421409

Berg, M.A., Coleman, R.S., Murphy, C.J., 2008. Nanoscale structure and dynamics of DNA. Phys. Chem. Chem. Phys. 10, 1229–1242. doi:10.1039/b715272h

Berkowitz, R., Fisher, J., Goff, S.P., 1996. RNA packaging. Curr. Top. Microbiol. Immunol. 214, 177–218. doi:10.1007/978-3-642-80145-7_6

Bernacchi, S., El-wahab, E.W.A., Dubois, N., Hijnen, M., Smyth, R.P., 2017. HIV-1 Pr55 Gag binds genomic and spliced RNAs with different affinity and stoichiometry Keywords. RNA Biol. 14, 1–37. doi:10.1080/15476286.2016.1256533

Bernacchi, S., Stoylov, S., Piémont, E., Ficheux, D., Roques, B.P., Darlix, J.L., Mély, Y., 2002. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. J. Mol. Biol. 317, 385–99. doi:10.1006/jmbi.2002.5429

Bernat, V., Disney, M.D., 2015. RNA Structures as Mediators of Neurological Diseases and as Drug Targets. Neuron. doi:10.1016/j.neuron.2015.06.012

Bikard, D., Loot, C., Baharoglu, Z., Mazel, D., 2010. Folded DNA in Action: Hairpin Formation and Biological Functions in Prokaryotes. Microbiol. Mol. Biol. Rev. 74, 570–588. doi:10.1128/MMBR.00026-10

Boeras, I., Seufzer, B., Brady, S., Rendahl, A., Heng, X., Boris-Lawrie, K., 2017. The basal translation rate of authentic HIV-1 RNA is regulated by 5'UTR nt-pairings at junction of R and U5. Sci. Rep. 7, 6902. doi:10.1038/s41598-017-06883-9

Börjesson, K., Preus, S., El-Sagheer, A.H., Brown, T., Albinsson, B., Wilhelmsson, L.M., 2009. Nucleic acid base analog FRET-pair facilitating detailed structural measurements in nucleic acid containing systems. J. Am. Chem. Soc. 131, 4288–4293. doi:10.1021/ja806944w

Bourbigot, S., Ramalanjaona, N., Boudier, C., Salgado, G.F., Roques, B.P., Mely, Y., Bouaziz, S., Morellet, N., 2008. How the HIV-1 nucleocapsid protein binds and destabilises the (–)primer binding site during reverse transcription. J Mol Biol 383, 1112–1128. doi:10.1016/j.jmb.2008.08.046

Boyer, P.L., Sarafianos, S.G., Arnold, E., Hughes, S.H., 2001. Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. J. Virol. 75, 4832–42. doi:10.1128/JVI.75.10.4832-4842.2001

Brasey, A., Lopez-Lastra, M., Ohlmann, T., Beerens, N., Berkhout, B., Darlix, J.L., Sonenberg, N., 2003. The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. J Virol 77, 3939–3949.

Brauns, E.B., Madaras, M.L., Coleman, R.S., Murphy, C.J., Berg, M.A., 1999. Measurement of local DNA reorganization on the picosecond and nanosecond time scales. J. Am. Chem. Soc. 121, 11644–11649. doi:10.1021/ja992456q

Brázda, V., Laister, R.C., Jagelská, E.B., Arrowsmith, C., 2011. Cruciform structures are a common DNA feature important for regulating biological processes. BMC Mol. Biol. doi:10.1186/1471-

2199-12-33

Briggs, J.A., Riches, J.D., Glass, B., Bartonova, V., Zanetti, G., Krausslich, H.G., 2009. Structure and assembly of immature HIV. Proc Natl Acad Sci U S A 106, 11090–11095. doi:10.1073/pnas.0903535106

Brochon, J.C., 1994. Maximum entropy method of data analysis in time-resolved spectroscopy. Methods Enzymol. 240, 262–311. doi:10.1016/S0076-6879(94)40052-0

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G.M., Gronenborn, A.M., 1997. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. Nat Struct Biol 4, 567–577.

Calzado, M.A., Sancho, R., Muñoz, E., 2004. Human immunodeficiency virus type 1 Tat increases the expression of cleavage and polyadenylation specificity factor 73-kilodalton subunit modulating cellular and viral expression. J. Virol. 78, 6846–54. doi:10.1128/JVI.78.13.6846-6854.2004

Campbell, E.M., Hope, T.J., 2015. HIV-1 capsid: the multifaceted key player in HIV-1 infection. Nat Rev Micro 13, 471–483. doi:10.1038/nrmicro3503

Campbell, S., Rein, A., 1999. In Vitro Assembly Properties of Human Immunodeficiency Virus Type 1 Gag Protein Lacking the p6 Domain. J. Virol. 73, 2270–2279.

Cen, S., Huang, Y., Khorchid, A., Darlix, J.L., Wainberg, M.A., Kleiman, L., 1999. The role of Pr55(gag) in the annealing of tRNA3Lys to human immunodeficiency virus type 1 genomic RNA. J. Virol. 73, 4485–8.

Chan, D.C., Fass, D., Berger, J.M., Kim, P.S., 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell 89, 263–273.

Charneau, P., Alizon, M., Clavel, F., 1992. A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. J. Virol. 66, 2814–20.

Checkley, M.A., Luttge, B.G., Freed, E.O., 2011. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. J Mol Biol 410, 582–608. doi:10.1016/j.jmb.2011.04.042

Chevalier, K., Grün, A., Stamm, A., Schmitt, Y., Gerhards, M., Diller, R., 2013. ESIPT and photodissociation of 3-hydroxychromone in solution: Photoinduced processes studied by static and time-resolved UV/Vis, fluorescence, and IR spectroscopy. J. Phys. Chem. A 117, 11233–11245. doi:10.1021/jp407252y

Chung, H.S., Meng, F., Kim, J.-Y., McHale, K., Gopich, I. V., Louis, J.M., 2017. Oligomerization of the tetramerization domain of p53 probed by two- and three-color single-molecule FRET. Proc. Natl. Acad. Sci. 114, E6812–E6821. doi:10.1073/pnas.1700357114

Chung, J., Ulyanov, N.B., Guilbert, C., Mujeeb, A., James, T.L., 2010. Binding characteristics of small molecules that mimic nucleocapsid protein-induced maturation of stem-loop 1 of HIV-1 RNA. Biochemistry 49, 6341–6351. doi:10.1021/bi100660r

Cimarelli, A., Darlix, J.L., 2002. Assembling the human immunodeficiency virus type 1. Cell Mol Life Sci 59, 1166–1184. doi:10.1007/s00018-002-8495-6

Cimarelli, A., Sandin, S., Höglund, S., Luban, J., 2000. Basic Residues in Human Immunodeficiency Virus Type 1 Nucleocapsid Promote Virion Assembly via Interaction with RNA. J. Virol. 74, 3046–3057. doi:10.1128/JVI.74.7.3046-3057.2000.

Claudio-Montero, A., Pinilla-Macua, I., Fernández-Calotti, P., Sancho-Mateo, C., Lostao, M.P., Colomer, D., Grandas, A., Pastor-Anglada, M., 2015. Fluorescent nucleoside derivatives as a tool for the detection of concentrative nucleoside transporter activity using confocal microscopy and flow cytometry. Mol. Pharm. 12, 2158–2166. doi:10.1021/acs.molpharmaceut.5b00142

Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.A., Santos-Ferreira, M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C., al., et, 1986. Isolation of a new human retrovirus from West African patients with AIDS. Science (80-.). 233, 343–346.

Crane-Robinson, C., Dragan, A.I., Read, C.M., 2009. Defining the thermodynamics of protein/DNA complexes and their components using micro-calorimetry. Methods Mol. Biol. 543, 625–651. doi:10.1007/978-1-60327-015-1_37

Crick, F.H.C., Klug, A., 1975. Kinky helix. Nature 255, 530–533. doi:10.1038/255530a0

Cruceanu, M., Gorelick, R.J., Musier-Forsyth, K., Rouzina, I., Williams, M.C., 2006a. Rapid Kinetics of Protein-Nucleic Acid Interaction is a Major Component of HIV-1 Nucleocapsid Protein's Nucleic Acid Chaperone Function. J. Mol. Biol. 363, 867–877. doi:10.1016/j.jmb.2006.08.070

Cruceanu, M., Urbaneja, M.A., Hixson, C. V., Johnson, D.G., Datta, S.A., Fivash, M.J., Stephen, A.G., Fisher, R.J., Gorelick, R.J., Casas-Finet, J.R., Rein, A., Rouzina, I., Williams, M.C., 2006b.

Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. Nucleic Acids Res. 34, 593–605. doi:10.1093/nar/gkj458

D.A. Case, 2015. Amber15.

D'Souza, V., Summers, M.F., 2005. How retroviruses select their genomes. Nat Rev Micro 3, 643–655.

Damgaard, C.K., Andersen, E.S., Knudsen, B., Gorodkin, J., Kjems, J., 2004. RNA Interactions in the 5′ Region of the HIV-1 Genome. J. Mol. Biol. 336, 369–379. doi:10.1016/j.jmb.2003.12.010

Darlix, J.-L., de Rocquigny, H., Mauffret, O., Mély, Y., 2014. Retrospective on the all-in-one retroviral nucleocapsid protein. Virus Res. 193, 2–15. doi:10.1016/j.virusres.2014.05.011

Darlix, J.-L., de Rocquigny, H., Mély, Y., 2016. The multiple roles of the nucleocapsid in retroviral RNA conversion into proviral DNA by reverse transcriptase. Biochem. Soc. Trans. 44, 1427–1440. doi:10.1042/BST20160101-T

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

Darlix, J.L., Godet, J., Ivanyi-Nagy, R., Fossé, P., Mauffret, O., Mély, Y., 2011. Flexible nature and specific functions of the HIV-1 nucleocapsid protein. J. Mol. Biol. 410, 565–581. doi:10.1016/j.jmb.2011.03.037

Das, K., Arnold, E., 2013. HIV-1 reverse transcriptase and antiviral drug resistance. Part 1. Curr Opin Virol 3, 111–118. doi:10.1016/j.coviro.2013.03.012

Daugherty, M.D., Liu, B., Frankel, A.D., 2010. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. Nat Struct Mol Biol 17, 1337–1342. doi:10.1038/nsmb.1902

Dawson, L., Yu, X.F., 1998. The role of nucleocapsid of HIV-1 in virus assembly. Virology 251, 141–157. doi:10.1006/viro.1998.9374

Day, H.A., Pavlou, P., Waller, Z.A.E., 2014. I-Motif DNA: Structure, stability and targeting with ligands. Bioorganic Med. Chem. 22, 4407–4418. doi:10.1016/j.bmc.2014.05.047

De Guzman, R.N., 1998. Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 -RNA Recognition Element. Science (80-.). 279, 384–388. doi:10.1126/science.279.5349.384

De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., Summers, M.F., 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. Science (80-.). 279, 384–388.

Dean, M., Carrington, M., Winkler, C., Huttley, G.A., Smith, M.W., Allikmets, R., Goedert, J.J., Buchbinder, S.P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R., O'Brien, S.J., 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Science (80-.). 273, 1856–1862. doi:10.1126/science.273.5283.1856

De Rocquigny, H., Gabus, C., Vincent, A., Fournié-Zaluski, M.C., Roques, B., Darlix, J.L., 1992. Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. Proc. Natl. Acad. Sci. U. S. A. 89, 6472–6476. doi:10.1073/pnas.89.14.6472

Demchenko, A.P., Tang, K.-C., Chou, P.-T., 2013. Excited-state proton coupled charge transfer modulated by molecular structure and media polarization. Chem. Soc. Rev. 42, 1379–1408. doi:10.1039/C2CS35195A

Demene, H., Dong, C.Z., Ottmann, M., Rouyez, M.C., Jullian, N., Morellet, N., Mely, Y., Darlix, J.L., Fournie-Zaluski, M.C., 1994. 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 33, 11707–11716. doi:10.1021/bi00205a006

Dhami, H., Fritz, C.E., Gankin, B., Pak, S.H., Yi, W., Seya, M.J., Raffa, R.B., Nagar, S., 2009. The chemokine system and CCR5 antagonists: potential in HIV treatment and other novel therapies. J Clin Pharm Ther 34, 147–160. doi:10.1111/j.1365-2710.2008.00978.x

Didier, P., Godet, J., Mély, Y., 2009. Two-Photon Two-Focus Fluorescence Correlation Spectroscopy with a Tunable Distance Between the Excitation Volumes. J. Fluoresc. 19, 561–565. doi:10.1007/s10895-008-0424-0

DiMattia, M.A., Watts, N.R., Stahl, S.J., Rader, C., Wingfield, P.T., Stuart, D.I., Steven, A.C., Grimes, J.M., 2010. Implications of the HIV-1 Rev dimer structure at 3.2 A resolution for multimeric binding to the Rev response element. Proc Natl Acad Sci U S A 107, 5810–5814.

doi:10.1073/pnas.0914946107

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

Dorr, P., Westby, M., Dobbs, S., Griffin, P., Irvine, B., Macartney, M., Mori, J., Rickett, G., Smith-Burchnell, C., Napier, C., Webster, R., Armour, D., Price, D., Stammen, B., Wood, A., Perros, M., 2005. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. Antimicrob. Agents Chemother. 49, 4721–4732. doi:10.1128/AAC.49.11.4721-4732.2005

Dr. Rainer Seitz, 2016. Human Immunodeficiency Virus (HIV). Transfus. Med. Hemotherapy 43, 203–222. doi:10.1159/000445852

Druillennec, S., Meudal, H., Roques, B.P., Fournié-Zaluski, M.C., 1999. Nucleomimetic strategy for the inhibition of HIV-1 nucleocapsid protein NCp7 activities. Bioorganic Med. Chem. Lett. 9, 627–632. doi:10.1016/S0960-894X(99)00052-9

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R., Davies, D.R., 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science (80-.). 266, 1981–1986.

Dziuba, D., Postupalenko, V.Y., Spadafora, M., Klymchenko, A.S., Guérineau, V., Mély, Y., Benhida, R., Burger, A., 2012. A universal nucleoside with strong two-band switchable fluorescence and sensitivity to the environment for investigating DNA interactions. J. Am. Chem. Soc. 134, 10209–10213. doi:10.1021/ja3030388

Edmonds, M., 2002. A history of poly A sequences: from formation to factors to function. Prog. Nucleic Acid Res. Mol. Biol. 71, 285–389. doi:10.1016/S0079-6603(02)71046-5

Égelé, C., Schaub, E., Piémont, É., de Rocquigny, H., Mély, Y., 2005. Investigation by fluorescence correlation spectroscopy of the chaperoning interactions of HIV-1 nucleocapsid protein with the viral DNA initiation sequences. C. R. Biol. 328, 1041–1051. doi:10.1016/j.crvi.2005.06.005

Egelé, C., Schaub, E., Ramalanjaona, N., Piémont, E., Ficheux, D., Roques, B., Darlix, J.L., Mély, Y., 2004. HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions. J. Mol. Biol. 342, 453–466. doi:10.1016/j.jmb.2004.07.059

El Meshri, S.E., Dujardin, D., Godet, J., Richert, L., Boudier, C., Darlix, J.L., Didier, P., Mély, Y., De Rocquigny, H., 2015. Role of the nucleocapsid domain in HIV-1 gag oligomerization and trafficking to the plasma membrane: A fluorescence lifetime imaging microscopy investigation. J. Mol. Biol. 427, 1480–1494. doi:10.1016/j.jmb.2015.01.015

Ennifar, E., Paillart, J.C., Bernacchi, S., Walter, P., Pale, P., Decout, J.L., Marquet, R., Dumas, P., 2007. A structure-based approach for targeting the HIV-1 genomic RNA dimerization initiation site. Biochimie 89, 1195–1203. doi:10.1016/j.biochi.2007.03.003

Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., Stuart, D., 1995. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. Nat Struct Biol 2, 303–308.

Felsenfeld, G., Davies, D.R., Rich, A., 1957. Formation of a Three-Stranded Polynucleotide Molecule. J. Am. Chem. Soc. doi:10.1021/ja01565a074

Feng, Y.X., Campbell, S., Harvin, D., Ehresmann, B., Ehresmann, C., Rein, A., 1999. 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. 73, 4251–6.

Fiebig, T., Wan, C., Zewail, A.H., 2002. Femtosecond charge transfer dynamics of a modified DNA base: 2-Aminopurine in complexes with nucleotides. ChemPhysChem 3, 781–788. doi:10.1002/1439-7641(20020916)3:9<781::AID-CPHC781>3.0.CO;2-U

Fisher, R.J., Rein, A., Fivash, M., Urbaneja, M.A., Casas-Finet, J.R., Medaglia, M., Henderson, L.E., 1998. Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. J. Virol. 72, 1902–9.

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

Fransen, S., Gupta, S., Danovich, R., Hazuda, D., Miller, M., Witmer, M., Petropoulos, C.J., Huang, W., 2009. Loss of raltegravir susceptibility by human immunodeficiency virus type 1 is conferred

via multiple nonoverlapping genetic pathways. J. Virol. 83, 11440–11446. doi:10.1128/JVI.01168-09

Freed, E.O., 2015. HIV-1 assembly, release and maturation. Nat Rev Microbiol 13, 484–496. doi:10.1038/nrmicro3490

Freed, E.O., 1998. HIV-1 Gag Proteins: Diverse Functions in the Virus Life Cycle. Virology 251, 1–15. doi:10.1006/viro.1998.9398

Freisz, S., Mezher, J., Hafirassou, L., Wolff, P., Nominé, Y., Romier, C., Dumas, P., Ennifar, E., 2012. Sequence and structure requirements for specific recognition of HIV-1 TAR and DIS RNA by the HIV-1 Vif protein. RNA Biol. 9, 37–41. doi:10.4161/rna.20483

Fujiwara, T., Kimoto, M., Sugiyama, H., Hirao, I., Yokoyama, S., 2001. Synthesis of 6-(2-thienyl)purine nucleoside derivatives that form unnatural base pairs with pyridin-2-one nucleosides. Bioorganic Med. Chem. Lett. 11, 2221–2223. doi:10.1016/S0960-894X(01)00415-2

Fun, A., Wensing, A.M., Verheyen, J., Nijhuis, M., 2012. Human Immunodeficiency Virus Gag and protease: partners in resistance. Retrovirology 9, 63. doi:10.1186/1742-4690-9-63

Gamble, T.R., Yoo, S., Vajdos, F.F., von Schwedler, U.K., Worthylake, D.K., Wang, H., McCutcheon, J.P., Sundquist, W.I., Hill, C.P., 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science (80-.). 278, 849–853.

Ganser-Pornillos, B.K., Yeager, M., Pornillos, O., 2012. Assembly and architecture of HIV. Adv. Exp. Med. Biol. 726, 441–465. doi:10.1007/978-1-4614-0980-9_20

Ganser-Pornillos, B.K., Yeager, M., Sundquist, W.I., 2008. The structural biology of HIV assembly. Curr Opin Struct Biol 18, 203–217. doi:10.1016/j.sbi.2008.02.001

Garbelli, A., Radi, M., Falchi, F., Beermann, S., Zanoli, S., Manetti, F., Dietrich, U., Botta, M., Maga, G., 2011. Targeting the human DEAD-box polypeptide 3 (DDX3) RNA helicase as a novel strategy to inhibit viral replication. Curr Med Chem 18, 3015–3027.

Gaynor, R., 1992. Cellular transcription factors involved in the regulation of HIV-1 gene expression. AIDS 6, 347–363.

Gellert, M., Lipsett, M.N., Davies, D.R., 1962. HELIX FORMATION BY GUANYLIC ACID. Proc. Natl. Acad. Sci. 48, 2013–2018. doi:10.1073/pnas.48.12.2013

Ghosh, A.K., Anderson, D.D., Weber, I.T., Mitsuya, H., 2012. Enhancing protein backbone binding--a fruitful concept for combating drug-resistant HIV. Angew Chem Int Ed Engl 51, 1778–1802. doi:10.1002/anie.201102762

Gitti, R.K., Lee, B.M., Walker, J., Summers, M.F., Yoo, S., Sundquist, W.I., 1996. Structure of the amino-terminal core domain of the HIV-1 capsid protein. Science (80-.). 273, 231–235.

Godde, F., Toulmé, J.J., Moreau, S., 2000. 4-amino-1H-benzo[g]quinazoline-2-one: a fluorescent analog of cytosine to probe protonation sites in triplex forming oligonucleotides. Nucleic Acids Res. 28, 2977–85. doi:10.1093/nar/28.15.2977

Godet, J., De Rocquigny, H., Raja, C., Glasser, N., Ficheux, D., Darlix, J.L., Mély, Y., 2006. 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. 356, 1180–1192. doi:10.1016/j.jmb.2005.12.038

Godet, J., Kenfack, C., Przybilla, F., Richert, L., Duportail, G., Mély, Y., 2013. Site-selective probing of cTAR destabilization highlights the necessary plasticity of the HIV-1 nucleocapsid protein to chaperone the first strand transfer. Nucleic Acids Res. 41, 5036–5048. doi:10.1093/nar/gkt164

Godet, J., Mély, Y., 2010. Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. RNA Biol. 7, 687–99. doi:10.4161/rna.7.6.13616

Godet, J., Ramalanjaona, N., Sharma, K.K., Richert, L., de Rocquigny, H., Darlix, J.-L., Duportail, G., Mély, Y., 2011. 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 Acids Res. 39, 6633–6645. doi:10.1093/nar/gkr274

Gonzalez, M.E., 2015. Vpu Protein: The Viroporin Encoded by HIV-1. Viruses 7, 4352–4368. doi:10.3390/v7082824

Goodman, M.F., Ratliff, R.L., 1983. Evidence of 2-aminopurine-cytosine base mispairs involving two hydrogen bonds. J. Biol. Chem. 258, 12842–12846.

Gorelick, R.J., Chabot, D.J., Rein, A., Henderson, L.E., Arthur, L.O., 1993. The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. J.

Virol. 67, 4027–36.

Gorelick, R.J., Fu, W., Gagliardi, T.D., Bosche, W.J., Rein, A., Henderson, L.E., Arthur, L.O., 1999. Characterization of the Block in Replication of Nucleocapsid Protein Zinc Finger Mutants from Moloney Murine Leukemia Virus. J Virol 73, 8185–8195.

Gorelick, R.J., Gagliardi, T.D., Bosche, W.J., Wiltrout, T.A., Coren, L. V., Chabot, D.J., Lifson, J.D., Henderson, L.E., Arthur, L.O., 1999. Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences. Virology 256, 92–104. doi:10.1006/viro.1999.9629

Grigorenko, N.A., Leumann, C.J., 2008. Electron transfer through a stable phenanthrenyl pair in DNA. Chem. Commun. 5417. doi:10.1039/b810751c

Grigorov, B., Décimo, D., Smagulova, F., Péchoux, C., Mougel, M., Muriaux, D., Darlix, J.-L., 2007. Intracellular HIV-1 Gag localization is impaired by mutations in the nucleocapsid zinc fingers. Retrovirology 4, 54. doi:10.1186/1742-4690-4-54

Grinevich, A.A., Ryasik, A.A., Yakushevich, L. V., 2015. 130 Modeling the DNA bubbles dynamics. J. Biomol. Struct. Dyn. 33, 84–84. doi:10.1080/07391102.2015.1032763

Grzesiek, S., Bax, A., Hu, J.S., Kaufman, J., Palmer, I., Stahl, S.J., Tjandra, N., Wingfield, P.T., 1997. Refined solution structure and backbone dynamics of HIV-1 Nef. Protein Sci 6, 1248–1263. doi:10.1002/pro.5560060613

Guenzel, C.A., Herate, C., Benichou, S., 2014. HIV-1 Vpr-a still "enigmatic multitasker." Front Microbiol 5, 127. doi:10.3389/fmicb.2014.00127

Guest, C.R., Hochstrasser, R. a., Sowers, L.C., Millar, D.P., 1991. Dynamics of mismatched base pairs in DNA. Biochemistry 30, 3271–3279. doi:10.1021/bi00227a015

Guo, J., Wu, T., Anderson, J., Kane, B.F., Johnson, D.G., Gorelick, R.J., Henderson, L.E., Levin, J.G., 2000. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. J. Virol. 74, 8980–8. doi:10.1128/JVI.74.19.8980-8988.2000

Guo, J., Wu, T., Bess, J., Henderson, L.E., Levin, J.G., 1998. Actinomycin D inhibits human immunodeficiency virus type 1 minus-strand transfer in in vitro and endogenous reverse transcriptase assays. J. Virol. 72, 6716–24.

Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., Alizon, M., 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature 326, 662–669. doi:10.1038/326662a0

Haqqani, A.A., Tilton, J.C., 2013. Entry inhibitors and their use in the treatment of HIV-1 infection. Antiviral Res. 98, 158–170. doi:http://dx.doi.org/10.1016/j.antiviral.2013.03.017

Hare, S., Gupta, S.S., Valkov, E., Engelman, A., Cherepanov, P., 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature 464, 232–236. doi:10.1038/nature08784

Hawkins, M.E., 2008a. Fluorescent Pteridine Probes for Nucleic Acid Analysis. Methods Enzymol. doi:10.1016/S0076-6879(08)03410-1

Hawkins, M.E., 2008b. Fluorescent Pteridine Probes for Nucleic Acid Analysis. Methods Enzymol. doi:10.1016/S0076-6879(08)03410-1

Hawkins, M.E., 2001. Fluorescent pteridine nucleoside analogs: a window on DNA interactions. Cell Biochem. Biophys. doi:10.1385/CBB:34:2:257

Heinicke, L.A., Wong, C.J., Lary, J., Nallagatla, S.R., Diegelman-Parente, A., Zheng, X., Cole, J.L., Bevilacqua, P.C., 2009. RNA Dimerization Promotes PKR Dimerization and Activation. J. Mol. Biol. 390, 319–338. doi:10.1016/j.jmb.2009.05.005

Henriet, S., Richer, D., Bernacchi, S., Decroly, E., Vigne, R., Ehresmann, B., Ehresmann, C., Paillart, J.C., Marquet, R., 2005. Cooperative and specific binding of Vif to the 5′ region of HIV-1 genomic RNA. J. Mol. Biol. 354, 55–72. doi:10.1016/j.jmb.2005.09.025

Herschhorn, A., Hizi, A., 2010. Retroviral reverse transcriptases. Cell. Mol. Life Sci. doi:10.1007/s00018-010-0346-2

Hikida, Y., Kimoto, M., Yokoyama, S., Hirao, I., 2010. Site-specific fluorescent probing of RNA molecules by unnatural base-pair transcription for local structural conformation analysis. Nat. Protoc. 5, 1312–23. doi:10.1038/nprot.2010.77

Hirao, I., Kimoto, M., Mitsui, T., Fujiwara, T., Kawai, R., Sato, A., Harada, Y., Yokoyama, S., 2006.

An unnatural hydrophobic base pair system: Site-specific incorporation of nucleotide analogs into DNA and RNA. Nat. Methods 3, 729–735. doi:10.1038/nmeth915

Hirao, I., Kimoto, M., Yamashige, R., 2012. Natural versus artificial creation of base pairs in DNA: Origin of nucleobases from the perspectives of unnatural base pair studies. Acc. Chem. Res. doi:10.1021/ar200257x

Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T., Kodama, K., Yokogawa, T., Nishikawa, K., Yokoyama, S., 2002. An unnatural base pair for incorporating amino acid analogs into proteins. Nat. Biotechnol. 20, 177–182. doi:10.1038/nbt0202-177

Holz, B., Klimasauskas, S., Serva, S., Weinhold, E., 1998. 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases. Nucleic Acids Res. 26, 1076–1083. doi:10.1093/nar/26.4.1076

Houzet, L., Paillart, J.C., Smagulova, F., Maurel, S., Morichaud, Z., Marquet, R., Mougel, M., 2007. HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. Nucleic Acids Res. 35, 2695–2704. doi:10.1093/nar/gkm153

Hsiou, Y., Ding, J., Das, K., Clark Jr., A.D., Hughes, S.H., Arnold, E., 1996. Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms. Structure 4, 853–860.

Hu, W.-S., Hughes, S.H., 2012. HIV-1 reverse transcription. Cold Spring Harb. Perspect. Med. 2, a006882-. doi:10.1101/cshperspect.a006882

Huang, Y., Eckstein, F., Padilla, R., Sousa, R., 1997. Mechanism of ribose 2'-group discrimination by an RNA polymerase. Biochemistry 36, 8231–8242. doi:10.1021/bi962674l

Hwang, G.T., Seo, Y.J., Kim, B.H., 2005. Pyrene-labeled deoxyuridine and deoxyadenosine: Fluorescent discriminating phenomena in their oligonucleotides. Tetrahedron Lett. 46, 1475–1477. doi:10.1016/j.tetlet.2005.01.015

Isel, C., Ehresmann, C., Marquet, R., 2010. Initiation of HIV reverse transcription. Viruses. doi:10.3390/v2010213

Jean, J.M., Hall, K.B., 2002. 2-Aminopurine electronic structure and fluorescence properties in DNA. Biochemistry 41, 13152–13161.

Jeong, H.S., Kang, S., Lee, J.Y., Kim, B.H., 2009. Probing specific RNA bulge conformations by modified fluorescent nucleosides. Org. Biomol. Chem. 7, 921. doi:10.1039/b816768k

Johnson, P.E., Turner, R.B., Wu, Z.R., Hairston, L., Guo, J., Levin, J.G., Summers, M.F., 2000. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. Biochemistry 39, 9084–9091. doi:10.1021/bi000841i

Johri, M.K., Mishra, R., Chhatbar, C., Unni, S.K., Singh, S.K., 2011. Tits and bits of HIV Tat protein. Expert Opin Biol Ther 11, 269–283. doi:10.1517/14712598.2011.546339

Jones, A.C., Neely, R.K., 2015. 2-aminopurine as a fluorescent probe of DNA conformation and the DNA–enzyme interface. Q. Rev. Biophys. 48, 244–279. doi:10.1017/S0033583514000158

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R., 1997. Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. J. Mol. Biol. 267, 727–748. doi:10.1006/jmbi.1996.0897

Jucker, F.M., Heus, H.A., Yip, P.F., Moors, E.H.M., Pardi, A., 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. J. Mol. Biol. 264, 968–980. doi:10.1006/jmbi.1996.0690

Kafaie, J., Song, R., Abrahamyan, L., Mouland, A.J., Laughrea, M., 2008. Mapping of nucleocapsid residues important for HIV-1 genomic RNA dimerization and packaging. Virology 375, 592–610. doi:10.1016/j.virol.2008.02.001

Kanevsky, I., Chaminade, F., Chen, Y., Godet, J., René, B., Darlix, J.L., Mély, Y., Mauffret, O., Fossé, P., 2011. Structural determinants of TAR RNA-DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. Nucleic Acids Res. 39, 8148–8162. doi:10.1093/nar/gkr526

Kanevsky, I., Chaminade, F., Ficheux, D., Moumen, A., Gorelick, R., Negroni, M., Darlix, J.L., Fossé, P., 2005. Specific interactions between HIV-1 nucleocapsid protein and the TAR element. J. Mol. Biol. 348, 1059–1077. doi:10.1016/j.jmb.2005.03.046

Karn, J., Stoltzfus, C.M., 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb. Perspect. Med. 2, a006916. doi:10.1101/cshperspect.a006916

Karsisiotis, A.I., O'Kane, C., Webba da Silva, M., 2013. DNA quadruplex folding formalism - A

tutorial on quadruplex topologies. Methods 64, 28–35. doi:10.1016/j.ymeth.2013.06.004

Kaushik, M., Kaushik, S., Roy, K., Singh, A., Mahendru, S., Kumar, M., Chaudhary, S., Ahmed, S., Kukreti, S., 2016. A bouquet of DNA structures: Emerging diversity. Biochem. Biophys. Reports 5, 388–395. doi:10.1016/j.bbrep.2016.01.013

Keane, S., Summers, M., 2016. NMR Studies of the Structure and Function of the HIV-1 5′-Leader. Viruses 8, 338. doi:10.3390/v8120338

Keane, S.C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J., Bolden, N.C., McCowin, S., Case, D.A., Johnson, B.A., Salemi, M., Telesnitsky, A., Summers, M.F., 2015. RNA structure. Structure of the HIV-1 RNA packaging signal. Science (80-.). 348, 917–921. doi:10.1126/science.aaa9266

Kellam, P., 2006. Attacking pathogens through their hosts. Genome Biol 7, 201. doi:10.1186/gb-2006-7-1-201

Kempf, N., Postupalenko, V., Bora, S., Didier, P., Arntz, Y., de Rocquigny, H., Mély, Y., 2015. The HIV-1 Nucleocapsid Protein Recruits Negatively Charged Lipids To Ensure Its Optimal Binding to Lipid Membranes. J. Virol. 89, 1756–1767. doi:10.1128/JVI.02931-14

Kerwood, D.J., Cavaluzzi, M.J., Borer, P.N., 2001. Structure of SL4 RNA from the HIV-1 packaging signal. Biochemistry 40, 14518–14529. doi:10.1021/bi0111909

Kieken, F., Paquet, F., Brulé, F., Paoletti, J., Lancelot, G., 2006. A new NMR solution structure of the SL1 HIV-1Lai loop-loop dimer. Nucleic Acids Res. 34, 343–352. doi:10.1093/nar/gkj427

Kimoto, M., Cox, R.S., Hirao, I., 2011. Unnatural base pair systems for sensing and diagnostic applications. Expert Rev. Mol. Diagn. 11, 321–31. doi:10.1586/erm.11.5

Kimoto, M., Mitsui, T., Harada, Y., Sato, A., Yokoyama, S., Hirao, I., 2007. Fluorescent probing for RNA molecules by an unnatural base-pair system. Nucleic Acids Res. 35, 5360–9. doi:10.1093/nar/gkm508

Kimura, T., Kawai, K., Fujitsuka, M., Majima, T., 2007. Monitoring G-quadruplex structures and G-quadruplex-ligand complex using 2-aminopurine modified oligonucleotides. Tetrahedron 63, 3585–3590. doi:10.1016/j.tet.2006.08.108

Kirk, S.R., Luedtke, N.W., Tor, Y., 2001. 2-Aminopurine as a real-time probe of enzymatic cleavage and inhibition of hammerhead ribozymes. Bioorganic Med. Chem. 9, 2295–2301. doi:10.1016/S0968-0896(01)00123-7

Klasens, B.I.F., Thiesen, M., Virtanen, A., Berkhout, B., 1999. The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure. Nucleic Acids Res. 27, 446–454. doi:10.1093/nar/27.2.446

Klymchenko, A.S., 2017. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. Acc. Chem. Res. 50, 366–375. doi:10.1021/acs.accounts.6b00517

Kuritzkes, D.R., 2011. Drug resistance in HIV-1. Curr Opin Virol 1, 582–589. doi:10.1016/j.coviro.2011.10.020

Kutluay, S.B., Bieniasz, P.D., 2010. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. PLoS Pathog. 6. doi:10.1371/journal.ppat.1001200

Kuzmič, P., 2009. DynaFit—A Software Package for Enzymology, in: Methods in Enzymology. pp. 247–280. doi:10.1016/S0076-6879(09)67010-5

Kuzmič, P., 1996. Program DYNAFIT for the Analysis of Enzyme Kinetic Data: Application to HIV Proteinase. Anal. Biochem. 237, 260–273. doi:10.1006/abio.1996.0238

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., Hendrickson, W.A., 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393, 648–659. doi:10.1038/31405

Lapadat-Tapolsky, M., Pernelle, C., Borie, C., Darlix, J.L., 1995. Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. Nucleic Acids Res. 23, 2434–41. doi:5c0062 [pii]

Larsen, O.F.A., Van Stokkum, I.H.M., Gobets, B., Van Grondelle, R., Van Amerongen, H., 2001. Probing the structure and dynamics of a DNA hairpin by ultrafast quenching and fluorescence depolarization. Biophys. J. 81, 1115–1126. doi:10.1016/S0006-3495(01)75768-2

Laughrea, M., Jetté, L., 1994. 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 33, 13464–13474. doi:10.1021/bi00249a035

Lawrence, D.C., Stover, C.C., Noznitsky, J., Wu, Z., Summers, M.F., 2003. Structure of the intact stem and bulge of HIV-1 Ψ-RNA stem-loop SL1. J. Mol. Biol. 326, 529–542. doi:10.1016/S0022-2836(02)01305-0

Leonard, N.J., Tolman, G.L., 1975. Fluorescent nucleosides and nucleotides. Ann. N. Y. Acad. Sci. 255, 43–58.

Leontis, N.B., Westhof, E., 2001. Geometric nomenclature and classification of RNA base pairs. RNA 7, 499–512. doi:10.1017/S1355838201002515

Lesbats, P., Engelman, A.N., Cherepanov, P., 2016. Retroviral DNA Integration. Chem Rev 116, 12730–12757. doi:10.1021/acs.chemrev.6b00125

Lever, A., Gottlinger, H., Haseltine, W., Sodroski, J., 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. J. Virol. 63, 4085–7.

Lever, A.M.L., 2007. HIV-1 RNA Packaging. Adv. Pharmacol. 55, 1–32. doi:10.1016/S1054-3589(07)55001-5

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

Levin, J.G., Mitra, M., Mascarenhas, A., Musier-Forsyth, K., 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. RNA Biol. 7, 754–774. doi:10.4161/rna.7.6.14115

Li, G., De Clercq, E., 2016. HIV Genome-Wide Protein Associations: a Review of 30 Years of Research. Microbiol. Mol. Biol. Rev. 80, 679–731. doi:10.1128/MMBR.00065-15

Li, P.T.X., Vieregg, J., Tinoco, I., 2008. How RNA Unfolds and Refolds. Annu. Rev. Biochem. 77, 77–100. doi:10.1146/annurev.biochem.77.061206.174353

Li, X., Zhang, L., Tian, Y., Song, Y., Zhan, P., Liu, X., 2014. Novel HIV-1 non-nucleoside reverse transcriptase inhibitors: a patent review (2011-2014). Expert Opin Ther Pat 24, 1199–1227. doi:10.1517/13543776.2014.964685

Liu, C., Martin, C.T., 2001. Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. J. Mol. Biol. 308, 465–475. doi:10.1006/jmbi.2001.4601

Livesey, A.K., Brochon, J.C., 1987. Analyzing the Distribution of Decay Constants in Pulse-Fluorimetry Using the Maximum Entropy Method. Biophys. J. 52, 693–706. doi:10.1016/S0006-3495(87)83264-2

Lobsiger, S., Blaser, S., Sinha, R.K., Frey, H.-M., Leutwyler, S., 2014. Switching on the fluorescence of 2-aminopurine by site-selective microhydration. Nat. Chem. 6, 989–993. doi:10.1038/nchem.2086

Lodi, P.J., Ernst, J.A., Kuszewski, J., Hickman, A.B., Engelman, A., Craigie, R., Clore, G.M., Gronenborn, A.M., 1995. Solution structure of the DNA binding domain of HIV-1 integrase. Biochemistry 34, 9826–9833.

Lodmell, J.S., Ehresmann, C., Ehresmann, B., Marquet, R., 2000. Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site. RNA 6, 1267–1276. doi:10.1017/S1355838200000698

Lu, K., Heng, X., Summers, M.F., 2011. Structural determinants and mechanism of HIV-1 genome packaging. J. Mol. Biol. doi:10.1016/j.jmb.2011.04.029

Luban, J., 2012. Innate immune sensing of HIV-1 by dendritic cells. Cell Host Microbe 12, 408–418. doi:10.1016/j.chom.2012.10.002

Lusic, M., Siliciano, R.F., 2017. Nuclear landscape of HIV-1 infection and integration. Nat Rev Micro 15, 69–82. doi:10.1038/nrmicro.2016.162

Ma, J., Yanez-Orozco, I.S., Rezaei Adariani, S., Dolino, D., Jayaraman, V., Sanabria, H., 2017. High Precision FRET at Single-molecule Level for Biomolecule Structure Determination. J. Vis. Exp. e55623–e55623. doi:10.3791/55623

Maga, G., Veljkovic, N., Crespan, E., Spadari, S., Prljic, J., Perovic, V., Glisic, S., Veljkovic, V., 2013. New in silico and conventional in vitro approaches to advance HIV drug discovery and design. Expert Opin Drug Discov 8, 83–92. doi:10.1517/17460441.2013.741118

Mailler, E., Bernacchi, S., Marquet, R., Paillart, J.C., Vivet-Boudou, V., Smyth, R.P., 2016. The life-cycle of the HIV-1 gag RNA complex. Viruses 8, 1–19. doi:10.3390/v8090248

Martí, A.A., Jockusch, S., Li, Z., Ju, J., Turro, N.J., 2006. Molecular beacons with intrinsically

fluorescent nucleotides. Nucleic Acids Res. 34, e50. doi:10.1093/nar/gkl134

Matos, P.M., Franquelim, H.G., Castanho, M.A.R.B., Santos, N.C., 2010. Quantitative assessment of peptide-lipid interactions. Ubiquitous fluorescence methodologies. Biochim. Biophys. Acta - Biomembr. doi:10.1016/j.bbamem.2010.07.012

McBride, M.S., Panganiban, A.T., 1997. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. J. Virol. 71, 2050–8.

McCoy, L.S., Shin, D., Tor, Y., 2014. Isomorphic emissive GTP surrogate facilitates initiation and elongation of in vitro transcription reactions. J. Am. Chem. Soc. 136, 15176–15184. doi:10.1021/ja5039227

McGann, M., 2012. FRED and HYBRID docking performance on standardized datasets. J. Comput. Aided. Mol. Des. 26, 897–906. doi:10.1007/s10822-012-9584-8

McGann, M., 2011. FRED pose prediction and virtual screening accuracy. J. Chem. Inf. Model. 51, 578–596. doi:10.1021/ci100436p

McKeen, C.M., Brown, L.J., Nicol, J.T.G., Mellor, J.M., Brown, T., Gut, I.G., Livak, K.J., Flood, S.A.J., Marmaro, J., Giusti, W., Deetz, K., Tyagi, S., Kramer, F.R., Whitcombe, D., Theaker, J., Guy, S.P., Brown, T., Little, S., Whitcombe, D., Kelly, S., Mann, J., Theaker, J., Jones, C., Little, S., Thelwell, N., Millington, T., Solinas, A., Booth, J., Brown, T., Solinas, A., Brown, L.J., McKeen, C., Mellor, J.M., Nicol, J.T.G., Thelwell, N., Brown, T., Roberts, L., Selvin, P.R., Mergny, J.-L., Lacroix, L., Teulade-Fichou, M.P., Houndsou, C., Guittat, L., Hoarau, M., Arimondo, P.B., Vigneron, J.-P., Lehn, J.-M., Riou, J.-F., Garestier, T., Helene, C., Mergny, J.-L., Garestier, T., Rougee, M., Lebedev, A. V., Chassignol, M., Thuong, N.T., Helene, C., Mergny, J.L., Boutorine, S.A., Garestier, T., Belloc, F., Rougee, M., Bulychev, N. V., Koshkin, A.A., Bourson, J., Lebedev, A. V., Valeur, B., Thuong, N.T., Helene, C., Ellouse, C., Piot, F., Takahashi, M., Yang, M., Ghosh, S.S., Millar, D.P., Yang, M., Millar, D.P., Fox, K.R., Sollogoub, M., Dominguez, B., Fox, K.R., Brown, T., Sojka, B., Piunno, P.A.E., Wust, C.C., Krull, U.J., Bartley, J.P., Brown, T., Lane, A.N., Haq, L., Chowdhry, B.Z., Brown, T., Lane, A.N., Rolland, V., Mitsuhara, K., Lhomme, J., Davoll, J., Kazimierczuk, Z., Cottam, H.B., Revanker, G.R., Robins, R.K., Buhr, C.A., Wagner, R.W., Grant, D., Froehler, B.C., Froehler, B.C., Matteucci, M.D., Davison, A., Duckworth, G., Vaman-Rao, M., McClean, J., Grzybowski, J., Brown, T., Dreschsler, G. V., Smagin, S., Rossi, F.M., Kao, P.Y., Darby, R.A.J., Sollogoub, M., McKeen, C., Brown, L.J., Risitano, A., Brown, N., Barton, C., Brown, T., Fox, K.R., Solinas, A., Thelwell, N., Brown, T., Brown, T., Brown, D.J.S., 2003. Synthesis of fluorophore and quencher monomers for use in Scorpion primers and nucleic acid structural probes. Org. Biomol. Chem. 1, 2267. doi:10.1039/b301859h

Mekdad, H. El, Boutant, E., Karnib, H., Biedma, M.E., Sharma, K.K., Malytska, I., Laumond, G., Roy, M., Réal, E., Paillart, J.-C., Moog, C., Darlix, J.L., Mély, Y., de Rocquigny, H., 2016. Characterization of the interaction between the HIV-1 Gag structural polyprotein and the cellular ribosomal protein L7 and its implication in viral nucleic acid remodeling. Retrovirology 13, 54. doi:10.1186/s12977-016-0287-4

Melhuish, W.H., 1961. QUANTUM EFFICIENCIES OF FLUORESCENCE OF ORGANIC SUBSTANCES: EFFECT OF SOLVENT AND CONCENTRATION OF THE FLUORESCENT SOLUTE 1. J. Phys. Chem. 65, 229–235. doi:10.1021/j100820a009

Mély, Y., De Rocquigny, H., Morellet, N., Roques, B.P., Gérard, D., 1996. Zinc binding to the HIV-1 nucleocapsid protein: A thermodynamic investigation by fluorescence spectroscopy. Biochemistry 35, 5175–5182. doi:10.1021/bi952587d

Métifiot, M., Marchand, C., Pommier, Y., 2013. HIV Integrase Inhibitors: 20-Year Landmark and Challenges. Adv. Pharmacol. 67, 75–105. doi:10.1016/B978-0-12-405880-4.00003-2

Michalet, X., Weiss, S., Jäger, M., 2006. Single-molecule fluorescence studies of protein folding and conformational dynamics. Chem. Rev. doi:10.1021/cr0404343

Mitsui, T., Kimoto, M., Kawai, R., Yokoyama, S., Hirao, I., 2007. Characterization of fluorescent, unnatural base pairs. Tetrahedron 63, 3528–3537. doi:10.1016/j.tet.2006.11.096

Mitsuya, H., Weinhold, K.J., Furman, P.A., St Clair, M.H., Lehrman, S.N., Gallo, R.C., Bolognesi, D., Barry, D.W., Broder, S., 1985. 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. Proc Natl Acad Sci U S A 82, 7096–7100.

Morellet, N., Bouaziz, S., Petitjean, P., Roques, B.P., 2003. NMR structure of the HIV-1 regulatory protein VPR. J Mol Biol 327, 215–227.

Morellet, N., Déméné, H., Teilleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournié-Zaluski, M.C., Roques, B.P., 1998. Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). J. Mol. Biol. 283, 419–434. doi:10.1006/jmbi.1998.2098

Mori, M., Dietrich, U., Manetti, F., Botta, M., 2010a. Molecular dynamics and DFT study on HIV-1 nucleocapsid protein-7 in complex with viral genome. J. Chem. Inf. Model. 50, 638–650. doi:10.1021/ci100070m

Mori, M., Dietrich, U., Manetti, F., Botta, M., 2010b. Molecular dynamics and DFT study on HIV-1 nucleocapsid protein-7 in complex with viral genome. J. Chem. Inf. Model. 50, 638–650. doi:10.1021/ci100070m

Mori, M., Manetti, F., Botta, M., 2011. Predicting the binding mode of known NcP7 inhibitors to facilitate the design of novel modulators. J. Chem. Inf. Model. 51, 446–454. doi:10.1021/ci100393m

Morohashi, N., Kimoto, M., Sato, A., Kawai, R., Hirao, I., 2012. Site-specific incorporation of functional components into RNA by an unnatural base pair transcription system. Molecules 17, 2855–76. doi:10.3390/molecules17032855

Mougel, M., Houzet, L., Darlix, J.-L., 2009. When is it time for reverse transcription to start and go? Retrovirology 6, 24. doi:10.1186/1742-4690-6-24

Mujeeb, A., Ulyanov, N.B., Georgantis, S., Smirnov, I., Chung, J., Parslow, T.G., James, T.L., 2007a. Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: Sequence effects and mechanism of RNA refolding. Nucleic Acids Res. 35, 2026–2034. doi:10.1093/nar/gkm097

Muriaux, D., Darlix, J.-L., 2010. Properties and functions of the nucleocapsid protein in virus assembly. RNA Biol. 7, 744–753. doi:10.4161/rna.7.6.14065

Muriaux, D., Girard, P.M., Bonnet-Mathonière, B., Paoletti, J., 1995. Dimerization of HIV-1 LAI RNA at low ionic strength. J. Biol. Chem. doi:10.1074/jbc.270.14.8209

Muriaux, D., Rocquigny, H. De, Roques, B., Paoletti, J., 1996. NCp7 Activates HIV-1 Lai RNA Dimerization by Converting a Transient Loop-Loop Complex into a Stable Dimer NCp7 Activates HIV-1 Lai RNA Dimerization by Converting a Transient Loop-Loop Complex into a Stable Dimer *. J Biol Chem. 271, 33686–33692. doi:10.1074/jbc.271.52.33686

Nag, N., Ramreddy, T., Kombrabail, M., Mohan, P.M.K., D'Souza, J., Rao, B.J., Duportail, G., Mely, Y., Krishnamoorthy, G., 2006. Dynamics of DNA and protein-DNA complexes viewed through time-domain fluorescence. Rev. Fluoresc. 3, 311–340. doi:10.1007/0-387-33016-X

Nakano, S.I., Uotani, Y., Uenishi, K., Fujii, M., Sugimoto, N., 2005. DNA base flipping by a base pair-mimic nucleoside. Nucleic Acids Res. 33, 7111–7119. doi:10.1093/nar/gki1018

Navia, M.A., Fitzgerald, P.M., McKeever, B.M., Leu, C.T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., Springer, J.P., 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 337, 615–620. doi:10.1038/337615a0

Niedrig, M., Gelderblom, H.R., Pauli, G., Marz, J., Bickhard, H., Wolf, H., Modrow, S., 1994. Inhibition of infectious human immunodeficiency virus type 1 particle formation by Gag protein-derived peptides. J. Gen. Virol. 75, 1469–1474. doi:10.1099/0022-1317-75-6-1469

Nikolaitchik, O.A., Dilley, K.A., Fu, W., Gorelick, R.J., Tai, S.H.S., Soheilian, F., Ptak, R.G., Nagashima, K., Pathak, V.K., Hu, W.S., 2013. Dimeric RNA Recognition Regulates HIV-1 Genome Packaging. PLoS Pathog. 9. doi:10.1371/journal.ppat.1003249

Nir, E., Kleinermanns, K., Grace, L., De Vries, M.S., 2001. On the photochemistry of purine nucleobases. J. Phys. Chem. A 105, 5106–5110. doi:10.1021/jp0030645

Nordlund, T.M., Andersson, S., Nilsson, L., Rigler, R., Gräslund, A., McLaughlin, L.W., 1989. Structure and Dynamics of a Fluorescent DNA Oligomer Containing the EcoRI Recognition Sequence: Fluorescence, Molecular Dynamics, and NMR Studies. Biochemistry 28, 9095–9103. doi:10.1021/bi00449a021

O'Neill, M.A., Barton, J.K., 2002. 2-Aminopurine: A probe of structural dynamics and charge transfer in DNA and DNA:RNA hybrids. J. Am. Chem. Soc. 124, 13053–13066. doi:10.1021/ja0208198

Okamoto, A., Tainaka, K., Saito, I., 2003a. Clear Distinction of Purine Bases on the Complementary

Strand by a Fluorescence Change of a Novel Fluorescent Nucleoside. J. Am. Chem. Soc. 125, 4972–4973. doi:10.1021/ja034090u

Okamoto, A., Tainaka, K., Saito, I., 2003b. Clear Distinction of Purine Bases on the Complementary Strand by a Fluorescence Change of a Novel Fluorescent Nucleoside. J. Am. Chem. Soc. 125, 4972–4973. doi:10.1021/ja034090u

Okamoto, A., Tanaka, K., Fukuta, T., Saito, I., 2003c. Design of base-discriminating fluorescent nucleoside and its application to T/C SNP typing. J. Am. Chem. Soc. 125, 9296–9297. doi:10.1021/ja0354081

Ono, A., Ablan, S.D., Lockett, S.J., Nagashima, K., Freed, E.O., 2004. Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. Proc Natl Acad Sci U S A 101, 14889–14894. doi:10.1073/pnas.0405596101

Ott, D.E., Coren, L. V., Shatzer, T., 2009. The nucleocapsid region of human immunodeficiency virus type 1 Gag assists in the coordination of assembly and Gag processing: role for RNA-Gag binding in the early stages of assembly. J. Virol. 83, 7718–27. doi:10.1128/JVI.00099-09

Padalkar, V.S., Seki, S., 2016. Excited-state intramolecular proton-transfer (ESIPT)-inspired solid state emitters. Chem. Soc. Rev. 45, 169–202. doi:10.1039/C5CS00543D

Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., Mak, J., 2004. Dimerization of retroviral RNA genomes: an inseparable pair. Nat. Rev. Microbiol. 2, 461–472. doi:10.1038/nrmicro903

Pak, A.J., Grime, J.M.A., Sengupta, P., Chen, A.K., Durumeric, A.E.P., Srivastava, A., Yeager, M., Briggs, J.A.G., Lippincott-Schwartz, J., Voth, G.A., 2017. Immature HIV-1 lattice assembly dynamics are regulated by scaffolding from nucleic acid and the plasma membrane. Proc. Natl. Acad. Sci. 114, 201706600. doi:10.1073/pnas.1706600114

Pancera, M., Majeed, S., Ban, Y.E., Chen, L., Huang, C.C., Kong, L., Kwon, Y.D., Stuckey, J., Zhou, T., Robinson, J.E., Schief, W.R., Sodroski, J., Wyatt, R., Kwong, P.D., 2010. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. Proc Natl Acad Sci U S A 107, 1166–1171. doi:10.1073/pnas.0911004107

Paoletti, A.C., Shubsda, M.F., Hudson, B.S., Borer, P.N., 2002. Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1. Biochemistry 41, 15423–15428. doi:10.1021/bi026307n

Pappalardo, L., Kerwood, D.J., Pelczer, I., Borer, P.N., 1998. Three-dimensional folding of an RNA hairpin required for packaging HIV-1. J. Mol. Biol. 282, 801–18. doi:10.1006/jmbi.1998.2046

Peloponese Jr.,J.M., Gregoire, C., Opi, S., Esquieu, D., Sturgis, J., Lebrun, E., Meurs, E., Collette, Y., Olive, D., Aubertin, A.M., Witvrow, M., Pannecouque, C., De Clercq, E., Bailly, C., Lebreton, J., Loret, E.P., 2000. 1H-13C nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. C R Acad Sci III 323, 883–894.

Peon, J., Zewail, A.H., 2001. DNA/RNA nucleotides and nucleosides: Direct measurement of excited-state lifetimes by femtosecond fluorescence up-conversion. Chem. Phys. Lett. 348, 255–262. doi:10.1016/S0009-2614(01)01128-9

Pereira, E.A., daSilva, L.L., 2016. HIV-1 Nef: Taking Control of Protein Trafficking. Traffic 17, 976–996. doi:10.1111/tra.12412

Pollard, V.W., Malim, M.H., 1998. The HIV-1 Rev protein. Annu Rev Microbiol 52, 491–532. doi:10.1146/annurev.micro.52.1.491

Pollom, E., Dang, K.K., Potter, E.L., Gorelick, R.J., Burch, C.L., Weeks, K.M., Swanstrom, R., 2013. Comparison of SIV and HIV-1 Genomic RNA Structures Reveals Impact of Sequence Evolution on Conserved and Non-Conserved Structural Motifs. PLoS Pathog. 9, e1003294. doi:10.1371/journal.ppat.1003294

Pornillos, O., Ganser-Pornillos, B.K., Yeager, M., 2011. Atomic-level modelling of the HIV capsid. Nature 469, 424–427. doi:10.1038/nature09640

Prinz, A., Reither, G., Diskar, M., Schultz, C., 2008. Fluorescence and bioluminescence procedures for functional proteomics. Proteomics. doi:10.1002/pmic.200700802

Properties, F., 2009. 2-Phenanthrenyl – DNA : Synthesis , Pairing , and Fluorescence Properties. Chem. Eur. J. 15, 639–645. doi:10.1002/chem.200801135

Purohit, V., Grindley, N.D.F., Joyce, C.M., 2003. Use of 2-aminopurine fluorescence to examine conformational changes during nucleotide incorporation by DNA polymerase I (Klenow fragment). Biochemistry 42, 10200–10211. doi:10.1021/bi0341206

Rachofsky, E.L., Seibert, E., Stivers, J.T., Osman, R., Ross, J.B.A., 2001. Conformation and dynamics of abasic sites in DNA investigated by time-resolved fluorescence of 2-aminopurine. Biochemistry 40, 957–967. doi:10.1021/bi001665g

Ramalanjaona, N., Rocquigny, H. de, Millet, A., Ficheux, D., Darlix, J.L., Mély, Y., 2007. Investigating the Mechanism of the Nucleocapsid Protein Chaperoning of the Second Strand Transfer during HIV-1 DNA Synthesis. J. Mol. Biol. 374, 1041–1053. doi:10.1016/j.jmb.2007.10.001

Raney, K.D., SOWERS, L.C., MILLAR, D.P., BENKOVIC, S.J., 1994. A Fluorescence-Based Assay for Monitoring Helicase Activity. Proc. Natl. Acad. Sci. U. S. A. 91, 6644–6648. doi:10.1073/pnas.91.14.6644

Rausch, J.W., le Grice, S.F.J., 2015. HIV Rev assembly on the Rev response element (RRE): A structural perspective. Viruses. doi:10.3390/v7062760

Rausch, J.W., Le Grice, S.F.J., 2004. "Binding, bending and bonding": Polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. Int. J. Biochem. Cell Biol. doi:10.1016/j.biocel.2004.02.016

Rhee, S.Y., Taylor, J., Fessel, W.J., Kaufman, D., Towner, W., Troia, P., Ruane, P., Hellinger, J., Shirvani, V., Zolopa, A., Shafer, R.W., 2010. HIV-1 protease mutations and protease inhibitor cross-resistance. Antimicrob Agents Chemother 54, 4253–4261. doi:10.1128/aac.00574-10

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., Gogliotti, R., Bader, J.P., Halliday, S.M., Coren, L., Sowder, R.C., Arthur, L.O., Henderson, L.E., 1995. Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. Science (80-.). 270, 1194–7.

Rimsky, L.T., Shugars, D.C., Matthews, T.J., 1998. Determinants of human immunodeficiency virus type 1 resistance to gp41-derived inhibitory peptides. J. Virol. 72, 986–993.

Rist, M.J., Marino, J.P., 2002. Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. Biochemistry 41, 14762–14770. doi:10.1021/bi0267240

Roca, A.I., Singleton, S.F., 2003. Direct Evaluation of a Mechanism for Activation of the RecA Nucleoprotein Filament. J. Am. Chem. Soc. 125, 15366–15375. doi:10.1021/ja0270165

Rodgers, D.W., Gamblin, S.J., Harris, B.A., Ray, S., Culp, J.S., Hellmig, B., Woolf, D.J., Debouck, C., Harrison, S.C., 1995. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. Proc Natl Acad Sci U S A 92, 1222–1226.

Rovira, A.R., Fin, A., Tor, Y., 2015. Chemical Mutagenesis of an Emissive RNA Alphabet. J. Am. Chem. Soc. 137, 14602–5. doi:10.1021/jacs.5b10420

Saad, J.S., Miller, J., Tai, J., Kim, A., Ghanam, R.H., Summers, M.F., 2006. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proc Natl Acad Sci U S A 103, 11364–11369. doi:10.1073/pnas.0602818103

Saadatmand, J., Kleiman, L., 2012. Aspects of HIV-1 assembly that promote primer tRNALys3 annealing to viral RNA. Virus Res. doi:10.1016/j.virusres.2012.06.001

Saenger, W., 1984. Principles of Nucleic Acid Structure, Springer Advanced Texts in Chemistry. Springer New York, New York, NY. doi:10.1007/978-1-4612-5190-3

Sakata, T., Kawashima, Y., Nakano, H., 2011. Solvent effect on the absorption spectra of coumarin 120 in water: A combined quantum mechanical and molecular mechanical study. J. Chem. Phys. 134, 14501. doi:10.1063/1.3506616

Sakuragi, S., Yokoyama, M., Shioda, T., Sato, H., Sakuragi, J., 2016. SL1 revisited: functional analysis of the structure and conformation of HIV-1 genome RNA. Retrovirology 13, 79. doi:10.1186/s12977-016-0310-9

Sandin, P., Stengel, G., Ljungdahl, T., Börjesson, K., Macao, B., Wilhelmsson, L.M., 2009. Highly efficient incorporation of the fluorescent nucleotide analogs tC and tCO by Klenow fragment. Nucleic Acids Res. 37, 3924–3933. doi:10.1093/nar/gkp266

Schmid, J.A., Birbach, A., 2007. Fluorescent proteins and fluorescence resonance energy transfer (FRET) as tools in signaling research. Thromb. Haemost. doi:10.1160/TH06-08-0472

Schmidt, O.P., Mata, G., Luedtke, N.W., 2016. Fluorescent Base Analogue Reveals T-HgII-T Base Pairs Have High Kinetic Stabilities That Perturb DNA Metabolism. J. Am. Chem. Soc. 138, 14733–14739. doi:10.1021/jacs.6b09044

Schuler, B., Eaton, W.A., 2008. Protein folding studied by single-molecule FRET. Curr. Opin. Struct. Biol. doi:10.1016/j.sbi.2007.12.003

Seela, F., Feiling, E., Gross, J., Hillenkamp, F., Ramzaeva, N., Rosemeyer, H., Zulauf, M., 2001. Fluorescent DNA: The development of 7-deazapurine nucleoside triphosphates applicable for sequencing at the single molecule level. J. Biotechnol. 86, 269–279. doi:10.1016/S0168-1656(00)00418-1

Seif, E., Niu, M., Kleiman, L., 2015. In virio SHAPE analysis of tRNALys3 annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. Retrovirology 12, 40. doi:10.1186/s12977-015-0171-7

Sen, S., Paraggio, N.A., Gearheart, L.A., Connor, E.E., Issa, A., Coleman, R.S., Wilson, D.M., Wyatt, M.D., Berg, M.A., 2005. Effect of protein binding on ultrafast DNA dynamics: Characterization of a DNA:APE1 complex. Biophys. J. 89, 4129–4138. doi:10.1529/biophysj.105.062695

Sengupta, P.K., Kasha, M., 1979. EXCITED STATE PROTON-TRANSFER SPECTROSCOPY OF 3-HYDROXYFLAVONE AND QUERCETIN. Chem. Phys. Lett. 68, 382–385. doi:10.1016/0009-2614(79)87221-8

Serrano-Andrés, L., Merchán, M., Borin, A.C., 2006. Adenine and 2-aminopurine: paradigms of modern theoretical photochemistry. Proc. Natl. Acad. Sci. U. S. A. 103, 8691–8696. doi:10.1073/pnas.0602991103

Sgarbanti, M., Borsetti, A., Moscufo, N., Bellocchi, M.C., Ridolfi, B., Nappi, F., Marsili, G., Marziali, G., Coccia, E.M., Ensoli, B., Battistini, A., 2002. Modulation of Human Immunodeficiency Virus 1 Replication by Interferon Regulatory Factors. J. Exp. Med. 195, 1359–1370. doi:10.1084/jem.20010753

Sharp, P.M., Hahn, B.H., 2011. Origins of HIV and the AIDS Pandemic. Cold Spring Harb. Perspect. Med. 1, a006841. doi:10.1101/cshperspect.a006841

Sherpa, C., Rausch, J.W., Le Grice, S.F.J., Hammarskjold, M.L., Rekosh, D., 2015. The HIV-1 Rev response element (RRE) adopts alternative conformations that promote different rates of virus replication. Nucleic Acids Res. 43, 4676–4686. doi:10.1093/nar/gkv313

Shin, D., Sinkeldam, R.W., Tor, Y., 2011. Emissive RNA alphabet. J. Am. Chem. Soc. 133, 14912–14915. doi:10.1021/ja206095a

Shing, P., Carter, M., 2011. DNA Structure: Alphabet Soup for the Cellular Soul, in: Seligmann, H.B.T.-D.N.A.R.-C.A. (Ed.), DNA Replication-Current Advances. InTech, Rijeka, p. Ch. 01. doi:10.5772/18536

Sholokh, M., Improta, R., Mori, M., Sharma, R., Kenfack, C., Shin, D., Voltz, K., Stote, R.H., Zaporozhets, O.A., Botta, M., Tor, Y., Mély, Y., 2016. Tautomers of a Fluorescent G Surrogate and Their Distinct Photophysics Provide Additional Information Channels. Angew. Chemie - Int. Ed. 55, 7974–7978. doi:10.1002/anie.201601688

Sholokh, M., Sharma, R., Shin, D., Das, R., Zaporozhets, O.A., Tor, Y., Mély, Y., 2015. Conquering 2-aminopurines deficiencies: Highly emissive isomorphic guanosine surrogate faithfully monitors guanosine conformation and dynamics in DNA. J. Am. Chem. Soc. 137, 3185–3188. doi:10.1021/ja513107r

Shubsda, M.F., Kirk, C.A., Goodisman, J., Dabrowiak, J.C., 2000. Binding of human immunodeficiency virus type 1 nucleocapsid protein to psi-RNA-SL3. Biophys. Chem. 87, 149–65.

Shubsda, M.F., Paoletti, A.C., Hudson, B.S., Borer, P.N., 2002. Affinities of Packaging Domain Loops in HIV-1 RNA for the Nucleocapsid Protein †. Biochemistry 41, 5276–5282. doi:10.1021/bi016045+

Shvadchak, V., Sanglier, S., Rocle, S., Villa, P., Haiech, J., Hibert, M., Van Dorsselaer, A., Mely, Y., de Rocquigny, H., 2009. Identification by high throughput screening of small compounds inhibiting the nucleic acid destabilization activity of the HIV-1 nucleocapsid protein. Biochimie 91, 916–923. doi:10.1016/j.biochi.2009.04.014

Shvadchak, V. V., Kucherak, O., Afitska, K., Dziuba, D., Yushchenko, D.A., 2017. Environmentally sensitive probes for monitoring protein-membrane interactions at nanomolar concentrations. Biochim. Biophys. Acta - Biomembr. 1859, 852–859. doi:10.1016/j.bbamem.2017.01.021

Silvius, J.R., Nabi, I.R., 2006. Fluorescence-quenching and resonance energy transfer studies of lipid microdomains in model and biological membranes (review). Mol. Membr. Biol. doi:10.1080/09687860500473002

Sinkeldam, R.W., Greco, N.J., Tor, Y., 2010. Fluorescent analogs of biomolecular building blocks:

design, properties, and applications. Chem. Rev. 110, 2579–619. doi:10.1021/cr900301e

Sleiman, D., Barraud, P., Brachet, F., Tisne, C., 2013. The Interaction between tRNA(Lys) 3 and the primer activation signal deciphered by NMR spectroscopy. PLoS One 8, e64700. doi:10.1371/journal.pone.0064700

Sleiman, D., Bernacchi, S., Xavier Guerrero, S., Brachet, F., Larue, V., Paillart, J.C., Tisne, C., 2014. Characterization of RNA binding and chaperoning activities of HIV-1 Vif protein. Importance of the C-terminal unstructured tail. RNA Biol 11, 906–920. doi:10.4161/rna.29546

Sleiman, D., Goldschmidt, V., Barraud, P., Marquet, R., Paillart, J.C., Tisné, C., 2012. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. Virus Res. 169, 324–339. doi:10.1016/j.virusres.2012.06.006

So, P.T., Dong, C.Y., 2002. Fluorescence Spectrophotometry. Encycl. Life Sci. 1–4. doi:10.1038/npg.els.0002978

Sowers, L.C., Boulard, Y., Fazakerley, G.V., 2000. Multiple structures for the 2-aminopurine-cytosine mispair. Biochemistry 39, 7613–7620. doi:10.1021/bi992388k

Sowers, L.C., Shaw, B.R., Veigl, M.L., David Sedwick, W., 1987. DNA base modification: Ionized base pairs and mutagenesis. Mutat. Res. - Fundam. Mol. Mech. Mutagen. 177, 201–218. doi:10.1016/0027-5107(87)90003-0

Spriggs, S., Garyu, L., Connor, R., Summers, M.F., 2008. Potential Intra- and Intermolecular Interactions involving the Unique-5'Region of the HIV-1 5'-UTR. Biochemistry 47, 13064–13073. doi:10.1021/bi8014373.Potential

Stivers, J.T., 1998. 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: Metal-ion and base sequence effects. Nucleic Acids Res. 26, 3837–3844. doi:10.1093/nar/26.16.3837

Stofer, E., Chipot, C., Lavery, R., 1999. Free energy calculations of watson-crick base pairing in aqueous solution. J. Am. Chem. Soc. 121, 9503–9508. doi:10.1021/ja991092z

Stoylov, S.P., Vuilleumier, C., Stoylova, E., De Rocquigny, H., Roques, B.P., Gérard, D., Mély, Y., 1997. Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. Biopolymers 41, 301–12. doi:10.1002/(SICI)1097-0282(199703)41:3<301::AID-BIP5>3.0.CO;2-W

Strickler, S.J., Berg, R. a., 1962. Relationship between Absorption Intensity and Fluorescence Lifetime of Molecules. J. Chem. Phys. 37, 814. doi:10.1063/1.1733166

Sükösd, Z., Andersen, E.S., Seemann, S.E., Jensen, M.K., Hansen, M., Gorodkin, J., Kjems, J., 2015. Full-length RNA structure prediction of the HIV-1 genome reveals a conserved core domain. Nucleic Acids Res. 43, 10168–10179. doi:10.1093/nar/gkv1039

Summers, M.F., South, T.L., Kim, B., Hare, D.R., Summers, M.F., South, T.L., Kim, B., Hare, D.R., 1990. High-Resolution Structure of an HIV Zinc Fingerlike Domain via a New NMR-Based Distance Geometry Approach. Biochemistry 29, 329–340. doi:10.1021/bi00454a005

Sundquist, W.I., Kräusslich, H.G., 2012. HIV-1 assembly, budding, and maturation. Cold Spring Harb. Perspect. Med. doi:10.1101/cshperspect.a006924

Tanwar, H.S., Khoo, K.K., Garvey, M., Waddington, L., Leis, A., Hijnen, M., Velkov, T., Dumsday, G.J., McKinstry, W.J., Mak, J., 2017. The thermodynamics of Pr55Gag-RNA interaction regulate the assembly of HIV. PLoS Pathog. 13, e1006221. doi:10.1371/journal.ppat.1006221

Tebit, D.M., Arts, E.J., 2011. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. Lancet Infect Dis 11, 45–56. doi:10.1016/s1473-3099(10)70186-9

Tedbury, P.R., Freed, E.O., 2015. HIV-1 gag: an emerging target for antiretroviral therapy. Curr Top Microbiol Immunol 389, 171–201. doi:10.1007/82_2015_436

Thomas, J.A., Gorelick, R.J., 2008. Nucleocapsid protein function in early infection processes. Virus Res. 134, 39–63. doi:10.1016/j.virusres.2007.12.006

Thoresen, L.H., Jiao, G.S., Haaland, W.C., Metzker, M.L., Burgess, K., 2003. Rigid, conjugated, fluoresceinated thymidine triphosphates: Syntheses and polymerase mediated incorporation into DNA analogues. Chem. - A Eur. J. 9, 4603–4610. doi:10.1002/chem.200304944

Tisné, C., Roques, B.P., Dardel, F., 2004. The Annealing Mechanism of HIV-1 Reverse Transcription Primer onto the Viral Genome. J. Biol. Chem. 279, 3588–3595. doi:10.1074/jbc.M310368200

Tolstorukov, M.Y., Jernigan, R.L., Zhurkin, V.B., 2004. Protein-DNA Hydrophobic Recognition in the

Minor Groove is Facilitated by Sugar Switching. J. Mol. Biol. 337, 65–76. doi:10.1016/j.jmb.2004.01.011

Velkov, T., 2009. Thermodynamics of lipophilic drug binding to intestinal fatty acid binding protein and permeation across membranes. Mol. Pharm. 6, 557–570. doi:10.1021/mp800227w

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., Taylor, R.D., 2003. Improved protein-ligand docking using GOLD. Proteins Struct. Funct. Genet. 52, 609–623. doi:10.1002/prot.10465

Verli, H., Calazans, A., Brindeiro, R., Tanuri, A., Guimarães, J.A., 2007. Molecular dynamics analysis of HIV-1 matrix protein: Clarifying differences between crystallographic and solution structures. J. Mol. Graph. Model. 26, 62–68. doi:10.1016/j.jmgm.2006.09.009

Vo, M.N., Barany, G., Rouzina, I., Musier-Forsyth, K., 2009. HIV-1 Nucleocapsid Protein Switches the Pathway of Transactivation Response Element RNA/DNA Annealing from Loop-Loop "Kissing" to "Zipper." J. Mol. Biol. 386, 789–801. doi:10.1016/j.jmb.2008.12.070

Vo, M.N., Barany, G., Rouzina, I., Musier-Forsyth, K., 2006. Mechanistic Studies of Mini-TAR RNA/DNA Annealing in the Absence and Presence of HIV-1 Nucleocapsid Protein. J. Mol. Biol. 363, 244–261. doi:10.1016/j.jmb.2006.08.039

Votteler, J., Sundquist, W.I., 2013. Virus budding and the ESCRT pathway. Cell Host Microbe 14, 232–241. doi:10.1016/j.chom.2013.08.012

Vrábel, M., Pohl, R., Klepetárová, B., Votruba, I., Hocek, M., 2007. Synthesis of 2'-deoxyadenosine nucleosides bearing bipyridine-type ligands and their Ru-complexes in position 8 through cross-coupling reactions. Org. Biomol. Chem. 5, 2849–57. doi:10.1039/b709245h

Vrábel, M., Pohl, R., Votruba, I., Sajadi, M., Kovalenko, S.A., Ernsting, N.P., Hocek, M., 2008. Synthesis and photophysical properties of 7-deaza-2'-deoxyadenosines bearing bipyridine ligands and their Ru(II)-complexes in position 7. Org. Biomol. Chem. 6, 2852–2860. doi:10.1039/b805632c

Vuilleumier, C., Bombarda, E., Morellet, N., Gérard, D., Roques, B.P., Mély, Y., 1999. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: A fluorescence study. Biochemistry 38, 16816–16825. doi:10.1021/bi991145p

Wan, C., Fiebig, T., Schiemann, O., Barton, J.K., Zewail, A.H., 2000. Femtosecond direct observation of charge transfer between bases in DNA. Proc. Natl. Acad. Sci. 97, 14052–14055. doi:10.1073/pnas.250483297

Wang, P., Rouyez, M.C., Ducamp, S., Saragosti, S., Ventura, M., 1993. Similarity of the 5' and 3'-TAR secondary structures in HIV-1. Biochem Biophys Res Commun 195, 565–573. doi:10.1006/bbrc.1993.2083

Ward, D.C., Reich, E., Stryer, L., 1969. Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. J. Biol. Chem. 244, 1228–1237.

Warui, D.M., Baranger, A.M., 2012. Identification of small molecule inhibitors of the HIV-1 Nucleocapsid-Stem-Loop 3 RNA Complex. J. Med. Chem. 55, 4132–4141. doi:10.1021/jm2007694

Watson, J. D.; Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids. Nature 171, 737–738. doi:10.1038/171737a0

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess Jr, J.W., Swanstrom, R., Burch, C.L., Weeks, K.M., 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460, 711–716. doi:10.1038/nature08237

Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., Wiley, D.C., 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature 387, 426–430. doi:10.1038/387426a0

Weixlbaumer, A., Werner, A., Flamm, C., Westhof, E., Schroeder, R., 2004. Determination of thermodynamic parameters for HIV DIS type loop-loop kissing complexes. Nucleic Acids Res. 32, 5126–5133. doi:10.1093/nar/gkh841

Whitcombe, D., Theaker, J., Guy, S.P., Brown, T., Little, S., 1999. Detection of PCR products using self-probing amplicons and fluorescence. Nat. Biotechnol. 17, 804–807. doi:10.1038/11751

Wilhelmsson, L.M., 2010. Fluorescent nucleic acid base analogues. Q. Rev. Biophys. 43, 159–183. doi:10.1017/S0033583510000090

Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., Weeks, K.M., 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA

strongly conserved across distinct biological states. PLoS Biol. 6, 883–899. doi:10.1371/journal.pbio.0060096

Wilson, J.N., Kool, E.T., 2006. Fluorescent DNA base replacements: reporters and sensors for biological systems. Org. Biomol. Chem. 4, 4265. doi:10.1039/b612284c

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., Kent, S.B., 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science (80-.). 245, 616–621.

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

Xu, W., Chan, K.M., Kool, E.T., 2017. Fluorescent nucleobases as tools for studying DNA and RNA. Nat. Chem. 9, 1043–1055. doi:10.1038/nchem.2859

Xu, X., Yu, T., Chen, S.-J., 2016. Understanding the kinetic mechanism of RNA single base pair formation. Proc. Natl. Acad. Sci. 113, 116–121. doi:10.1073/pnas.1517511113

Yang, K., Stanley, R.J., 2006. Differential distortion of substrate occurs when it binds to DNA photolyase: A 2-aminopurine study. Biochemistry 45, 11239–11245. doi:10.1021/bi060408u

Yedavalli, V.S., Jeang, K.T., 2010. Trimethylguanosine capping selectively promotes expression of Rev-dependent HIV-1 RNAs. Proc Natl Acad Sci U S A 107, 14787–14792. doi:10.1073/pnas.1009490107

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

Yuan, Y., Kerwood, D.J., Paoletti, A.C., Shubsda, M.F., Borer, P.N., 2003. Stem of SL1 RNA in HIV-1: Structure and nucleocapsid protein binding for a 1×3 internal loop. Biochemistry 42, 5259–5269. doi:10.1021/bi034084a

Zamotaiev, O.M., Postupalenko, V.Y., Shvadchak, V. V, Pivovarenko, V.G., Klymchenko, A.S., M??ly, Y., 2011. Improved hydration-sensitive dual-fluorescence labels for monitoring peptide-nucleic acid interactions. Bioconjug. Chem. 22, 101–107. doi:10.1021/bc100434d

Zargarian, L., Kanevsky, I., Bazzi, A., Boynard, J., Chaminade, F., Fossé, P., Mauffret, O., 2009. Structural and dynamic characterization of the upper part of the HIV-1 cTAR DNA hairpin. Nucleic Acids Res. 37, 4043–4054. doi:10.1093/nar/gkp297

Zargarian, L., Tisné, C., Barraud, P., Xu, X., Morellet, N., René, B., Meĺy, Y., Fossé, P., Mauffret, O., 2014. Dynamics of linker residues modulate the nucleic acid binding properties of the HIV-1 nucleocapsid protein zinc fingers. PLoS One 9, 27–30. doi:10.1371/journal.pone.0102150

Zeinolabediny, Y., Caccuri, F., Colombo, L., Morelli, F., Romeo, M., Rossi, A., Schiarea, S., Ciaramelli, C., Airoldi, C., Weston, R., Donghui, L., Krupinski, J., Corpas, R., García-Lara, E., Sarroca, S., Sanfeliu, C., Slevin, M., Caruso, A., Salmona, M., Diomede, L., 2017. HIV-1 matrix protein p17 misfolding forms toxic amyloidogenic assemblies that induce neurocognitive disorders. Sci. Rep. 7, 10313. doi:10.1038/s41598-017-10875-0

Zeng, Y., Liu, H.-W., Landes, C.F., Kim, Y.J., Ma, X., Zhu, Y., Musier-Forsyth, K., Barbara, P.F., 2007. Probing nucleation, reverse annealing, and chaperone function along the reaction path of HIV-1 single-strand transfer. Proc. Natl. Acad. Sci. U. S. A. 104, 12651–6. doi:10.1073/pnas.0700350104

Zhang, D., Li, W., Jiang, S., 2015. Peptide fusion inhibitors targeting the HIV-1 gp41: a patent review (2009 - 2014). Expert Opin Ther Pat 25, 159–173. doi:10.1517/13543776.2014.987752

Rajhans SHARMA

Caractérisation et ciblage de la reconnaissance dynamique de Trp37-G lors de l'interaction de la protéine NCp7 de HIV-1 avec des acides nucléiques

Résumé

La protéine de la nucléocapside (NC) possède un rôle important dans le cycle de viral du VIH-1 grâce à sa propriété chaperone des acides nucléiques (NA) qui implique la reconnaissance de son résidu Trp37 avec un résidu Guanine de l'acide nucléique cible. Nous avons caractérisé cette reconnaissance dynamique Trp37-G en utilisant des séquences impliquées dans la transcription inverse et l'assemblage de l'ARN génomique. En utilisant les analogues nucléosidiques fluorescents thienoguanosine et 2-thiényl-3-hydroxychromone, nous avons déterminé l'ensemble des constantes de vitesse cinétiques du mécanisme d'hybridation de la séquence (-)PBS avec (+)PBS en absence et en présence de NC. Nous avons également étudié le rôle du NA sucre dans les complexes NC-ARN et NC-ADN, puisque la protéine NC se lie avec la polarité opposée aux séquences d'ADN et d'ARN. Nous avons confirmé que l'interaction du résidu Trp37 avec les amino-acides de type guanines était critique lors de la formation des complexes avec les deux mutants d'ARN et d'ADN de PBS et de SL3. Enfin, nous avons réalisé un criblage de potentiels inhibiteurs de la protéine NC et examiné les touches identifiées à partir d'un test basé sur la fluorescence de la sonde thG.

Mots-clés : VIH-1, nucléocapside protéine, analogues fluorescents, thiéoguanosine, PBS, SL3, acides nucléiques dynamique, fluorescence, protéine - acides nucléiques interaction

Résumé en anglais

Nucleocapsid protein (NC) plays crucial roles in HIV-1 life cycle through its nucleic acid (NA) chaperoning property that involves recognition of its Trp37 residue with a Guanine residue of the target nucleic acid. Herein, we characterized this dynamic Trp-G recognition with sequences involved in reverse transcription and genomic RNA packaging. Using the fluorescent thienoguanosine (thG) and 2-thienyl-3-hydroxychromone (3HCnt) nucleoside analogues, we determined the whole set of kinetic rate constants for annealing of (-)PBS with (+)PBS in the absence and presence of NC. We also investigated the role of NA sugar in NC-RNA and NC-DNA complexes, as NC binds with opposite polarity to DNA and RNA sequences. We confirmed that the interaction of the Trp37 residue with guanines was critical for the formation of complexes with both RNA and DNA variants of PBS and SL3. Finally, we performed screening of NC inhibitors and tested the selected hits on a thG-based assay.

Key words: HIV-1, nucleocapsid protein, fluorescent nucleobase analogues, thieoguanosine, primer binding site, SL3, nucleic acid dynamics, fluorescence spectroscopy, protein-nucleic acids interaction.